Molecular Recognition

Biotechnology, Chemical Engineering and Materials Applications

Jason A. McEvoy

Editor

NOVA

MOLECULAR RECOGNITION: BIOTECHNOLOGY, CHEMICAL ENGINEERING AND MATERIALS APPLICATIONS

CHEMICAL ENGINEERING METHODS AND TECHNOLOGY

Additional books in this series can be found on Nova's website under the Series tab.

Additional E-books in this series can be found on Nova's website under the E-books tab.

MOLECULAR RECOGNITION: BIOTECHNOLOGY, CHEMICAL ENGINEERING AND MATERIALS APPLICATIONS

JASON A. MCEVOY

EDITOR

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

CONTENTS

PREFACE

The term 'molecular recognition' refers to the specific interaction between two or more molecules through noncovalent bonding. This book presents research in the study of molecular recognition, including next generation molecular imprinted polymers; applications of molecular imprinting; recent advances in DNA-Ligand molecular recognition and allosteric interactions; the proteomic code and the molecular recognition of odorant-binding proteins in insect olfaction.

Chapter 1 - Synthetic molecules able for efficient and selective binding of carboxylic acids belong to very promising receptor structures. Special focus on hydroxy- and dicarboxylic acids is due to the central role of these molecules in metabolic paths of the living organisms and commercial importance in biotechnology. In addition, wide range of biological and organic molecules contain carboxylic group. For this reasons, great attention is paid to modeling of the synthetic receptors able to specifically bind carboxylic acids or their fragments. This chapter describes the current state-of-the-art in research and development of the methods for development of artificial receptors for carboxylic acids. The focus is on the structural and physical properties of synthetic receptors because the efficiency of interaction between the receptors and acids depends on various factors, i.e. nature of the substituents, their structural accepting characteristics, geometrical complementarity of the binding sites etc. In general, the development of artificial receptors for carboxylic acids has some difficulties and now full understanding of all the principles of molecular recognition of acids is still far from being completed. Due to their unique properties, the synthetic receptor molecules can contribute to solution of these problems. They offer new opportunities for modeling artificial living systems and physiological processes, designing therapeutic agents and sensitive elements of (bio)sensors and others diagnostic tools devoted to fast detection of pathogens and pathological stages in medicine. Active search in the area led to development of artificial receptors different in efficiency of recognition toward a number of hydroxy-, amino and dicarboxylic acids. In this chapter, general approaches to the design of synthetic receptor, their classification and performance in recognition of various substrates are considered.

Chapter 2 - Molecular imprinted polymers (MIPs) have been studied for a few decades. They enable the specific recognition of the molecule for which they have been prepared. They have been thoroughly studied in organic solvents and this showed that a good recognition was observed only when a crosslinking ratio above 70% was used. In this case, however, the capacity of the MIP was limited owing to a poor accessibility of the imprinted cavities. In the ongoing research on this subject, many teams assess the possibility of controlling the

recognition process and that of using the MIPs in aqueous systems. They present here their experience in liquid crystal MIPs as tunable systems and hydrogel-MIPs for the recognition of proteins.

In the first case, the MIP is composed of a liquid crystal elastomer which is built around the template. The difference from regular MIPs lies in the percentage of crosslinker agent, which is in the 5-10% range. This low range becomes possible owing to replacing a large part of the chemical crosslinking by a physical one, coming from the interactions between liquid crystal moieties attached to the material. This brings up a recognition ability similar to the regular MIPs, but with an increased accessibility to the cavities. Thus, the mass capacity of this LC-MIP is tremendously increased. Furthermore, since liquid crystal elastomers exhibit an organized/disorganized transition temperature and have a shape memory capacity, the LC-MIP can be controlled with external parameters, such as temperature or solvents. By this method, different types of materials have been examined and are presented here: MIPs able to specifically interact with an enantiomer, catalytic MIPs acting as artificial enzymes or MIPs able to interact with pesticides.

The second part of the manuscript describes the state of the art as well as the author's preliminary experiments aiming at developing MIPs made of hydrogels which will be able to selectively recognize a protein in solution. The goal is to fix the hydrogel MIP to a detection device for a future application in diagnostics. Due to this, a severe constraint exists for process: temperature and pH limits, process in less than 30minutes, porosity slightly lower than the size of the protein. Indeed, since proteins are large molecules, recognition at the surface of the MIP is sought. Several monomer formulations have been studied and the technological problems have been examined.

Chapter 3 - The concept of molecular recognition in supramolecules with different types of intermolecular interactions and in some biological processes is discussed. The features and manifestations of hydrogen bonding as one of the most important types of interactions participating in the molecular recognition are presented in geometric, spectroscopic and natural bond orbitals terms. Molecular recognition in co-crystals and polymorphs is object of discussion in order to give a general view of this matter in different technologies, particularly in the pharmaceutical one. The role of molecular conformation and association in solution in predicting crystalline structures is investigated. Attention is also given to the molecular recognition in solid/solution interfaces in some important processes of crystal growth.

Chapter 4 - Molecular recognition can be defined as recognition of molecules, i.e. polar molecules by other polar molecules, or recognition of chiral molecules by other chiral molecules. These molecules can interact together, for instance, by creating hydrogen bonding. This review shows using of microscopy, spectroscopy and other methods to investigate of chiral recognition (i.e. chiral surfaces).

The term of chiral surface can be defined as surface of metal (i.e Cu, Au) or non–metal (i.e graphite) covered by chiral molecules. The covering ratio of surfaces can be different, and it depends of method specification used in appropriate experiment. In this review the author want to show, so chiral molecule (i.e. α–amino acids) adsorbed on surface (i.e. metal) can recognize another chiral molecule also adsorbed on surface. The useful methods discussed in this paper are: STM (scanning tunneling microscopy), AFM (atomic force microscopy), electrochemistry and vibrational spectroscopy − infrared and Raman, including SERS effect (surface–enhanced Raman scattering) and others. Molecular recognition can be investigated, for instance, when molecules create i.e. the SAMs or Langmuir–Blodgett films.

The most sensitive methods are STM, AFM and Raman spectroscopy (SERS), because the most interesting results were obtained using these methods. The author would like to underline, so every method has its own limitation and specification, and results are strong depended from using method. Structural information obtained using various techniques for chiral surfaces can be useful in understanding of key of interaction of adsorbed chiral molecules.

Chapter 5 - Molecular imprinting has been widely studied and applied in the last two decades as an innovative tool for various technological and scientific fields. The first approach to molecular imprinting was not well known until 1949 when Dicky and his coworkers succeeded to create complementary molecular cavities for certain dye molecules. Dickey's silicates could be considered as the first molecularly imprinted material [1and2].

In general, molecular imprinting process involves the formation of molecular cavities inside the polymer matrix being imprinted which are of complementary structural, functional group orientation and geometrical features with respect to the molecule being imprinted. Specifically, the molecular cavities are created by the incorporation of the template molecule during the polymerization process in such a way that a three dimensional network is formed around the molecule being imprinted. Upon extraction of the molecular moiety (template) from the polymer matrix, molecular cavities with specific shape, size and electrostatic features, remain in the cross-linked host material [3-5].

Because of its unique properties, molecularly imprinted materials have been widely utilized for a lot of applications and in various fields. They were applied in high performance liquid chromatography [6], food analysis [7], capillary chromatography, solid phase extraction [8] and drug delivery techniques [9].

One of the most important applications of the imprinting technique is the molecular recognition. Sensors prepared by imprinting methods could introduce good solution for the recognition of a variety of biologically active molecules. Recognition mechanism of the molecules is largely similar to what happens in living organisms. In their bodies, there are a large number of different molecules, and cells, without which they cannot survive, which are able to work cooperatively in such a way that certain function are carried out very precisely and accurately. For example, the receptors on the surface of cell membranes bind hormones specifically and selectively. When the receptor binds a hormone, its conformation is changed and a message of the hormone is transferred in terms of a conformational change and as a result of that a specific function of that hormone is fulfilled. In molecular recognition, the molecules being imprinted can rebind to their molecular cavities with very high degree of selectivity and specificity, so that these materials may be named as artificial antibodies.

Chapter 6 - It is generally recognized that elucidating the molecular basis for recognition of specific sequences in target DNA by proteins and small synthetic molecules vitally underpins research on the modulation of gene expression. In this review they discuss the fundamental basis of DNA sequence recognition by small molecules at the atomic bonding level based on recent X-ray diffraction results together with circular dichroism spectra and footprinting experiments on DNA-small ligand binding. Monodentate (single) interactions, intrastrand bidentate interactions and interstrand bidentate interactions are considered not only central to the capabilities of small ligands and proteins to recognize DNA sequences, but also provide the means for allosteric communication between multiple DNA binding loci. Thermodynamic, kinetic and allosteric features of molecular recognition by drugs and small ligands within the minor groove of DNA are reviewed. Allosteric interactions between small

synthetic peptides and multiple DNA binding sites are discussed, and hypothetical models are proposed to interpret the complex allosteric communication process. In contrast to protein-protein interaction networks which have been extensively investigated, studies on small ligand-DNA interaction networks have only recently been commenced. In this review, three different types of novel allosteric interaction networks between peptides and DNA are considered, together with hypothetical models featuring monodentate interactions and interstrand bidentate interactions. The new concept of DNA-small ligand interaction networks illuminates some basic chemical rules of DNA-small ligand allostery and may find applications in future drug design as well as structural biology research.

Chapter 7 - Like other animals, insects sense chemicals evoke their behaviors, such as aggregation, mating, feeding and migration behaviors, through their olfactory and taste recognition systems. The olfactory molecules, in the air out of insect body, diffuse into sensillum lymph through the cuticle pores and at first step interact with the small soluble odorant-binding proteins (OBPs). The odor specificity of a given olfactory neuron is determined by expressed olfactory receptor (OR) genes along with other accessory proteins. The signals accepted by ORs then are sent to higher processing centers in the brain to elicit distinct behavioral outputs.

So far over 100 OBPs have been identified from more than 40 species of insects. However, only a few have been studied for their recognition with ligands during binding activity. Although OBPs certainly show binding ability to hydrophobic odorants, their physiological roles are still in controversy. There are three unclear points which people are very interested in. Firstly, whether or not insect OBPs have binding specificity and how do they perform such binding specificity? Secondly, how do the proteins recognize the odorants, and which binding sites are involved in recognition? Thirdly, how do OBPs interact with odorant receptors? To reveal mechanisms of OBPs molecular recognition is helpful in understanding their physiological functions and in designing interfering molecules to ruin normal recognition and control insect pests. Therefore this field is attracting more and more scientists' passions and interests.

Recently some interesting experiments on insect OBPs had been conducted to partly address these questions with the combination of three dimensional structure analysis, binding experiments, site-directed mutagenesis, and simulation and docking experiments. Here the authors discuss insect OBPs molecules recognition and review the progress in the field. It has been found that the conformational change of OBPs can be induced either by pH value alteration or by ligand binding. Some hydrophilic amino acids at the entrance of OBPs binding pocket, as well as the hydrophobic amino acids in the pocket, are involved in ligand binding and may contribute to the recognition specificity of OBPs.

Chapter 8 - The Proteomic Code is a set of rules by which information in genetic material is transferred into the physico-chemical properties of amino acids and determines how individual amino acids interact with each other during folding and in specific protein–protein interactions. The Proteomic Code is part of the redundant Genetic Code. The 25-year-old history of this concept is reviewed from the first independent suggestions by Biro and Mekler, through the works of Blalock, Root-Bernstein, Siemion, Miller and others, followed by the discovery of a Common Periodic Table of Codons and Nucleic Acids in 2003 and culminating in the recent conceptualization of partial complementary coding of interacting amino acids as well as the theory of the nucleic acid-assisted protein folding. A novel cloning method for the design and production of specific and with high affinity reacting proteins

(SHARP) is presented. This method is based on the concept of proteomic codes and is suitable for large-scale, industrial production of specifically interacting peptides.

Chapter 9 - The 'on/off'-switched molecular recognition by a smart imprinted polymer (MIP-T) was presented in this chapter. The smart MIP-T was prepared using PNIPA as the thermosensitive element and 2-aminopurine as the template. The thermal phase transition of PNIPA induced a self-switching ability in the prepared polymer. At a relatively low temperature (such as 20 $^{\circ}$C), the MIP-T was capable of highly specifically recognizing the imprint species, i.e., 2-aminopurine. However, above the transition temperature, the MIP-T did not demonstrate significant resolution for 2-aminopurine compared with its analogue 2-amino-6-methylpurine. Such temperature-responsive recognition, in nature, was comparable to an on/off-switched process, which allows an efficient self-regulation in the molecular recognition behavior.

Chapter 1

MOLECULAR RECOGNITION OF CARBOXYLIC ACIDS AND CARBOXYLATE ANIONS BY SYNTHETIC RECEPTOR

Ivan I. Stoikov[*], Maria N. Agafonova, Luidmila S. Yakimova, Igor S. Antipin and Alexander I. Konovalov

Organic Chemistry Department, A.M. Butlerov' Chemistry Institute,
Kazan (Volga Region) Federal University, 18 Kremlevskaya Street,
Kazan, 420008, Russian Federation

ABSTRACT

Synthetic molecules able for efficient and selective binding of carboxylic acids belong to very promising receptor structures. Special focus on hydroxy- and dicarboxylic acids is due to the central role of these molecules in metabolic paths of the living organisms and commercial importance in biotechnology. In addition, wide range of biological and organic molecules contain carboxylic group. For this reasons, great attention is paid to modeling of the synthetic receptors able to specifically bind carboxylic acids or their fragments. This chapter describes the current state-of-the-art in research and development of the methods for development of artificial receptors for carboxylic acids. The focus is on the structural and physical properties of synthetic receptors because the efficiency of interaction between the receptors and acids depends on various factors, i.e. nature of the substituents, their structural accepting characteristics, geometrical complementarity of the binding sites etc. In general, the development of artificial receptors for carboxylic acids has some difficulties and now full understanding of all the principles of molecular recognition of acids is still far from being completed. Due to their unique properties, the synthetic receptor molecules can contribute to solution of these problems. They offer new opportunities for modeling artificial living systems and physiological processes, designing therapeutic agents and sensitive elements of (bio)sensors and others diagnostic tools devoted to fast detection of pathogens and pathological stages in medicine. Active search in the area led to development of artificial

---

[*] Corresponding author: e-mail: Ivan.Stoikov@mail.ru, Tel: +7-8432-337462; Fax: +7-8432-752253;

receptors different in efficiency of recognition toward a number of hydroxy-, amino and dicarboxylic acids. In this chapter, general approaches to the design of synthetic receptor, their classification and performance in recognition of various substrates are considered.

Keywords: Calixarenes; Thiacalixarenes; Self-assembly, Molecular recognition.

1. INTRODUCTION

The development of synthetic receptors able to recognize specific substrates belongs to the trends of modern organic chemistry. The demands in design of high-selective synthetic complexing agents relates to various aspects of molecular recognition, including templated organic synthesis and separation (chromatography, membrane technologies), design of miniature receptor devices, information storage and transfer which are realized on molecular level [Mutihac et al. 2002; Sokoließ et al. 2003; Creaven et al. 2009]. Problems of modeling and synthesis of artificial receptor molecules are very actual. Thus, they refer also for biomimetic systems, e.g. those involved in biological systems functioning (coding and replicating genetic information, enzyme catalysis, immune response, active transmembrane transfer of specific ions and molecules, etc.) [Mutihac et al. 2002; Kubo et al. 2006; Consoli et al. 2007; Creaven et al. 2009]. Permeability and selectivity of artificial membranes are also achieved by introduction of lipophilic molecules-carriers which function as receptor forming complexes with the substrates [Oshima et al. 2008; Erdemir et al. 2009]. Membrane transport of biologically significant compounds, such as α-hydroxy and dicarboxylic acids, through lipophilic organic membranes can be used for separation and analysis of multi-component media, for modeling of the biological membranes operation, for development of drug delivery systems and purification of pharmaceuticals.

All natural and anthropogenic processes mentioned above assume molecular recognition as their indispensible step. Thus, the problem of directed synthesis of compounds able to molecular recognition is considered very actual not only for synthetic organic chemistry, but also for basic and applied research areas of science. The solution of the problem directly depends on the establishment of structure- property relationships both for the receptor (transporter, complexing agent, catalyst) and a substrate. The appropriate relationships can involve the parameters of transport, efficiency of detection, stereoselectivity of chemical reactions responsible for recognition and interface transfer of the initial components and supramolecular complexes formed [Li et al. 2008; Creaven et al. 2009].

The choice of a structure and synthesis paths for host molecules required for various substrates can differ. However, in last two decades, the most attention to such a choice is focused on macrocyclic compounds, among them, on metacyclophanes, named also calixarenes. They show many advantage over transitional ligands as molecular platform for synthetic receptors, i.e. variety of functionalization possibilities, the design of three-dimensional structures and high structural flexibility of host molecules. Calixarenes are the products of one-stage condensation of phenols and formaldehyde, available and inexpensive compounds. Followed by directed modification of available reaction sites, this makes it possible to obtain various modification of their structure, including geometry of a macrocycle cavity and its nearest environment. Functionalization of phenolic groups of a macrocycle

bridged fragments by appropriate organic and organoelemental reagents can multiply changes in efficiency and selectivity of the binding reaction with the ions and neutral organic molecules [Stibor 2005; Stoikov et al. 2010]. The application of such compounds as building blocks for the synthesis of highly organized supramolecular structures, like drug delivery systems, artificial molecular and ion channels, nanosized electronic and mechanochemical devices is of special interest [Stibor 2005; Kubo et al. 2006; Creaven et al. 2009].

Study of complexation of organic ligands with metal cations is a research base of molecular recognition by synthetic host receptors. The number of investigations in this area is extremely high. Meanwhile selective binding of organic anions and acids by molecular receptors has yet not been found adequate attention [Casnati et al. 2003; Späth and König 2010]. It should be noted that this aspect of molecular recognition seems very important because many of them are responsible for interactions with biologically active species, including metabolites and mediators of biochemical paths like α-amino and α-hydroxy acids, nucleotides, etc. To some extent, limited amount of publications in this area is related to some complications arising design of receptors toward anions and neutral organic acids. In particular, the achievement of geometric complementary of coordination sites required for molecular recognition of these substrates (especially, polyfunctional molecules, i.e. peptides, nucleosides, dicarboxylic, α-hydroxy and α-amino acids) calls for more complicated spatial organization of coordination site contrary to that for spherical inorganic ions of the metals [Blondeau et al. 2007]. Besides, interaction energy is much lower for neutral molecules than that of the ions [Stibor 2005].

In this connection, binding and recognition of uncharged molecules demands not only electrostatic, but also other types of interactions, i.e. hydrogen bonds and donor-acceptor interactions [Späth and König 2010]. These factors do the synthesis of receptors for carboxylic acids a complex problem, which is more difficult than that of recognition of cationic species [Stoikov et al. 2010].

Progress in understanding nature of chemical bonds, intra- and intermolecular interactions, interfacial processes, methods of directed organic synthesis which was achieved from late 1960-s, made it possible to design not only molecular systems but also supermolecules and organized poly-molecular (supramolecular) assembles [Lehn et al. 1988]. C. Pedersen, J.-M. Lehn and D. Cram, the pioneers in this direction, gave birth to a new area of the science which received the name of supramolecular chemistry [Lehn et al. 1988]. The terminology implemented first by D.Cram, described the host-guest or synthetic receptor-substrate components involved in the formation of supramolecular associates. Thus, host (a synthetic receptor) was denoted as "an organic molecule or an ion whose binding centers converge in a complex. Guest (substrate) is a molecule or an ion, whose binding centers disperse in a complex" [Dietrich 1981].

A wide number of biological and organic molecules contain carboxylic group in the molecular (neutral) or ionic form. It explains importance of the development of receptor structures for carboxylic and carboxylate groups to be recognized in such molecules or their fragments.

In our works, dicarboxylic, amino and hydroxy acids were specified as the substrates containing these groups. They are not only often presented in metabolic paths but also are used as important products in biotechnological manufacture.

2. SYNTHETIC RECEPTOR MOLECULES FOR THE SENSING OF CARBOXYLATE ANIONS AND CARBOXYLIC ACIDS

Despite of successes in the field of creation of receptors for carboxylic acids, the problem of effective recognition of these substrates has not yet been solved. Serious difficulties in the synthesis of the receptors for these substrates are caused by a number of the reasons, i.e. strong solvation of charged zwitterionic form of α-amino acids, self-association α-hydroxy acids in aqueous solutions, the necessity of the binding of the hydrophobic side chain often present in natural α-amino and α-hydroxy acids [Kimura 1981].

2.1. MACROCYCLIC POLYAMINES

Initially, synthetic receptors containing ammonium groups were used for carboxylate anions recognition. Kimura [Kumura et al. 1981] and Lehn [Dietrich et al. 1981; Hosseini et al. 1982; Breslow et al. 1981; Jazwinski et al. 1981; Dhaenens et al. 1993; Cudic et al. 1999; Miranda et al. 2004; Cruz et al. 2004] published their pioneer articles in 1981. Macrocyclic and polymacrocyclic polyamines 1-12 act as receptors for di- and tri-carboxylates (Figure 1). Reaction of polyamines 1-12 with guests is controlled by geometric complementary of molecules and charge value of a guest and host (see summary in the Table 1).
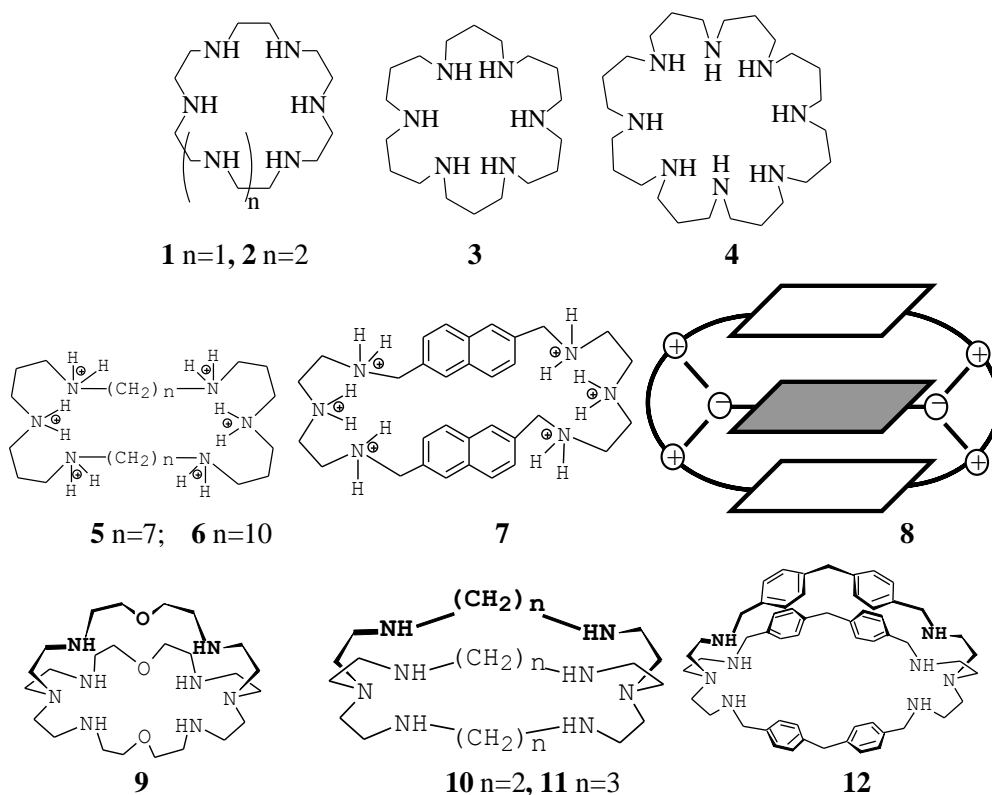

Figure 1. The structure of compounds 1-12.

Table 1. Association constants lgK_{as} for complexation of carboxylate anions with polyammonium macrocycles in aqueous solution (references are given below in the text)

anion	2 3H+	3 6H+	4 8H+
oxalate2-	-	3.8	3.7
malonate2-	1.5	3.3	3.9
succinate2-	1.3	2.4	3.6
maleate2-	1.5	3.7	4.1
fumarate2-		2.2	2.9
cutrate3-	2.4	4.7	7.6
1,3,5-benzenetricarboxylate3-	-	3.5	6.1

The affinity of carboxylate anions toward protonated macrocyclic oligoamines, e.g. compounds 1, 2 3H+, 3 6H+ and 4 8H+, is determined by Coulomb interactions because the association constant was found to be increased with the charge decrease of a receptor and substrate (Table 1). This was clearly observed in the range from double charged to triple charged anions and depended on the degree of charge localization. Thus stronger association was found for small anions with highly localized negative charge: oxalate > malonate > succinate > fumarate. Except electrostatic interaction, a macrocycle size (geometric complementary) plays significant role. Rather large dianions (succinate, fumarate anions) formed more stable complexes with the macrocycle 4 8H+, the biggest one in the range considered.

Macrocyclic hexa-amines 5 and 6 were synthesized for dicarboxylate anions binding [Hosseini et al. 1982; Breslow et al. 1981]. In neutral aqueous solutions, the compounds 5 and 6 exist as hexaprotonated cations. The analysis of the dependence of the association constant on the chain length (m=1-8) in dicarboxylates $^{-}O_2C-(CH_2)_m-CO_2^{-}$ showed definite selectivity of the receptors toward a substrate size. Thus, the macrocycle 5 formed more stable complexes with glutarate dianion (m=3, lgK_{as} = 4.4), and the macrocycle 6 with pimelate dianion (m=5, lgK_{as} = 4.4).

Macrocycle 7 containing 2,6-naphtyl bridge was synthesized for dianions binding. The formation of the complex is presented on the scheme 8 [Jazwinski et al. 1981; Dhaenens et al. 1993]. Two large aromatic cycles symmetrically located in receptor provide effective binding unsaturated substrates. The estimation of acid association constants 7 with maleate (lgK_{as} = 3.5), fumarate (4.4), phthalate (3.6), iso-phthalate (5.0) and terphthalate (5.2) showed that π-π interactions and complementary of carboxylate and ammonium groups in receptor plaid main role in effective binding of these species.

The development of pre-organized structures with rigid location of the binding sites is one of the approaches to increase of the complexation selectivity and stability. According to this strategy, series of bicyclic cryptands 9-12 which form hexaprotonated particles in acid solution were synthesized by Lehn's group [Cudic et al. 1999; Miranda et al. 2004].

The large value of association constants (lgK_{as} (9 6H+ / oxalate) = 4.95, (9 6H+/malonate) = 3.10; (10, 11 6H+/oxalate) = 4.50, (10, 11 6H+/malonate) = 2.85) [Cudic P. et al. 1999] and X-ray data [Miranda et al. 2004] made it possible to conclude that anion was fixed inside the molecular cavity by a number of hydrogen bonds. Conformation rigidity of bicycle affects complexation selectivity toward a guest size. Introduction of one methylene chain spacer in

the dianion caused a 100-fold decrease in the association constant. However, cryptand 12 6H[+] [Cudic et al. 1999] with diphenylmethane spacer interacted with aliphatic acid anions much worse than monocyclic polyammonium receptors and showed lower selectivity to a guest size (Table 2).

Table 2. Association constants lgK_{as} (M^{-1}) for complexation of acid dianions with receptors 19 and 20, 21 in aqueous solution (pH = 8.8, 27 ºC) (references are given below in the text)

Substrate	19	20, 21	K_{as}(20, 21)/ K_{as}(19)
22	208	714	3.4
26	62	322	5.1
23	556	2041	3.7
24	476	5265	11.1
25	1024	10000	9.6

According to X-ray data, each carboxylate group in crystal structure of cryptand binds with ammonium centers by three H-bonds. The decrease of selectivity could be due to large size and conformation lability of cryptand adapting to a guest size. No preferences in the binding of unsaturated and aromatic dianions by bicyclic receptor were found. Association constant of 9 6H[+] with maleate (lgK_{as} = 3.4), fumarate (3.6) and terephthalate (4.4) anions was significantly lesser than that obtained with monocyclic ligand 7. Binding selectivity also decreased. For example, it was equal to 8.3 for fumarate/maleate pair and monocycle 7 and 1.7 for cryptand 12 6H[+]. According to author's opinion, weak π-π and CH-π interactions are main factors affecting molecular recognition [Cudic et al. 1999]. Acridine derivative 13 selectively bounded *trans*-azobenzenedicarboxylates in D$_2$O [Cudic et al. 1999] (Figure 2).

Interaction of the pyrazole containing macrocyclic receptors with glutamate in aqueous solution was studied by potentiometric techniques [Miranda et al. 2004]. It was shown, that the compound 14 which contained benzyl groups at the central nitrogen atoms of the polyamine side chain exerted larger interaction with glutamate (K_{ass} = 2.04·10^4 M^{-1} at pH 7.4). Macrocycle 15 with pyridine fragments was used as receptor for the molecular recognition of aromatic and aliphatic carboxylate substrates (phthalate, isophthalate, terephthalate, 4,4'-dibenzoate, benzoate, 3- and 4-nitrobenzoate, and $^-$O$_2$C(CH$_2$)$_n$CO$_2$$^-$, with n = 1-4). No specific selectivity was found for the series of anions, association constants varied from lgK_{as} = 2.15 to 3.69. Single crystal X-ray showed alternating molecules of diprotonated receptor 15 with terephthalate and 4,4-dibenzoate anions. These anions were identified also as the building blocks of an extensive 3-D network of hydrogen bonds [Cruz et al. 2004].

Novel 18-mer cyclic oligopeptide 16 showed excellent receptor properties for carboxylates [Chakraborty et al. 2002]. The binding ability of 16 toward carboxylate anion was measured by the ^{1}H NMR titration with tetrabutylammonium acetate in CD$_3$CN. The stoichiometry of complex was 1:1 with association constant of 8.64·10^3 M^{-1}.

Application of protonated polyamines for anion binding is limited by acidic medium and base strength of the anions studied that determined their deprotonation. Quaternary ammonium bases [Schmidtchen 1981; Schmidtchen 1986; Lara et al. 2001; Sunamoto et al. 1982; Rebek et al. 1987; Rebek et al. 1990] do not show such a limitation. However, they do

not participate in specific interactions because they do not contain acidic N^+-H bonds. Indeed, this essentially reduces the energy of their association with anions, because hydrogen binding plays important role in interaction energy of protonated amines. Tetrahedral tetraazo derivatives 18 and 19 bind carboxylates, i.e. formiate, acetate, benzoate and some other anions with formation of inclusion complexes with 1:1 stoichiometry (K_{as}=3-70 M^{-1}) [Schmidtchen 1986]. Open-chained bifunctional receptors 20 and 21 involving two tetraaza groups connected by *p*-phenylene bridge were synthesized to increase the host-guest complex stability and selectivity.
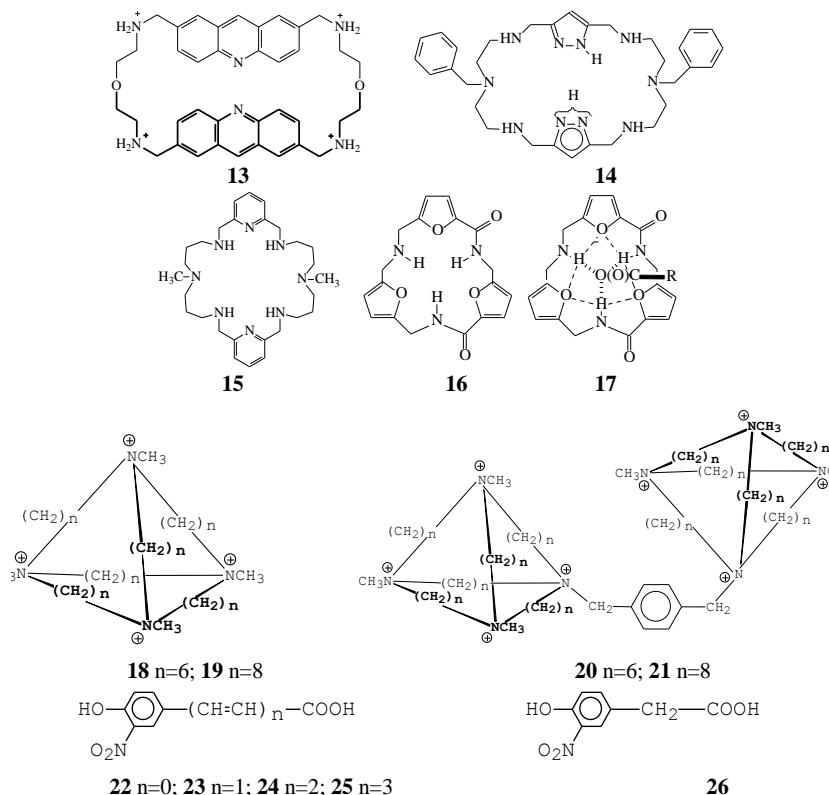
Figure 2. The structure of compounds 13-26.

Acid dianions 22-26 [Schmidtchen 1986] were used for investigation of the effect of the substrate side chain length effect on their complexation with the receptors studied [Schmidtchen F.P. 1986]. The *o*-nitrophenol substrates 22-26 exerted bathochromic shift of their long wave absorption band after the addition of the host compounds. The shifts were analyzed by a Benesi-Hildebrand treatment [Lara et al. 2001] (Table 2).

It is obvious from the Table 2, that the association constants of the monotopic receptor 19 are lower than those of ditopic hosts 20 and 21. This is expected from the comparison of electrostatic interactions which are much higher for anionic particles with 20, 21 than those of 18, 19, because of the doubled positive charge of the host molecule. Another trend yielding from Table 2 is that the association constant increases with elongating the side chain of the substrate with no respect of its nature even though electrostatic interactions are weakened, at

least in the case of 18, 19. Hydrophobic and dispersion interactions of host and guest species contribute to guest binding and compensate for the decay of electrostatic attractions.
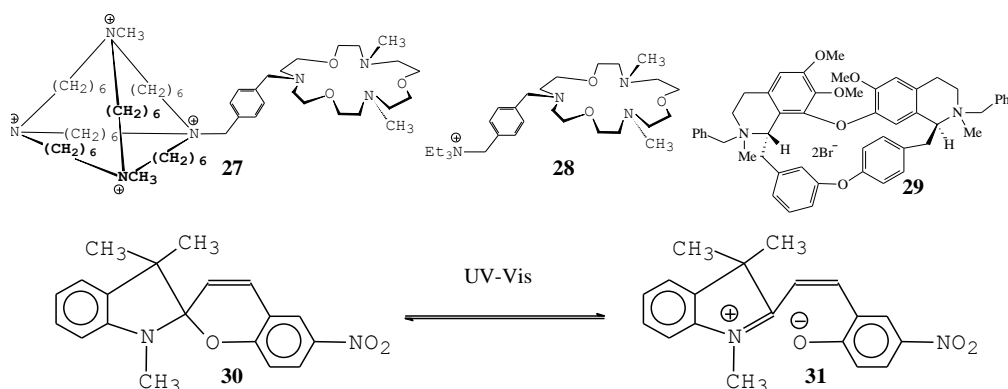

Figure 3. The structure of compounds 27-31.

Receptor 27 containing spacer linked binding centers for two functional groups of a guest was obtained for ω-amino acids [Schmidtchen et al. 1986] (Figure 3). The compound cannot self-associate. Probably this contributes to its complexation. However, its K_{as} values did not depend on the chain length of a substrate. An analog of 27, obtained by formal substitution of the tetrahedral tetraaza fragment by triethylammonium group (compound 28) effectively bounded ω-amino acids in aqueous methanol.

Binding bis-carboxylates by bis-quaternary salts 29 obtained from (*S,S*)-(+)-tetrandrine was described by Eliseev and Eatsimirsky. Receptor 29 bounded in aqueous solution phthalate ($K_{as} = 135$ M^{-1}) and terephralate ($K_{as} = 110$ M^{-1}) anions stronger than isophtalate ($K_{as} = 49$ M^{-1}) [Lara et al. 2001].

The first investigation on amino acids bonded by zwitterionic receptors was reported in [Sunamoto et al. 1982]. Upon UV irradiation in nonpolar organic solvents, the photospiran 30 was easily converted to the colored ring-opened zwitterionic form 31 which transferred phenylalanine across liposomal membrane by formation an ionic complex between the carrier and amino acid.

2.2. PROTONATED HETEROCYCLES

Conformation rigid chain-opened receptors based on Kemp's acid 32 and 33 were synthesized [Rebek et al. 1987; Rebek et al. 1990] for the binding of dicarboxylic acids. Oxalic acid ($pK_{a1} = 1.2$), insoluble in CDCl$_3$, was dissolves in the presence of compound 32 ($K_{as} = 103$ M^{-1}) (Figure 4). Complex with 1:1 stoichiometry is observed for malonic acid ($pK_{a1} = 2.9$) and its C-substituted derivatives, for maleic ($pK_{a1} = 1.8$) and phthalic acid ($pK_{a1} = 2.9$). The weaker acids such as fumaric (pK_{a1} 3.0), succinic ($pK_{a1} = 4.2$) and glutaric ($pK_{a1} = 4.3$) acids under these conditions did not interact with 32, thus the authors suggested the protonation was important for the binding.
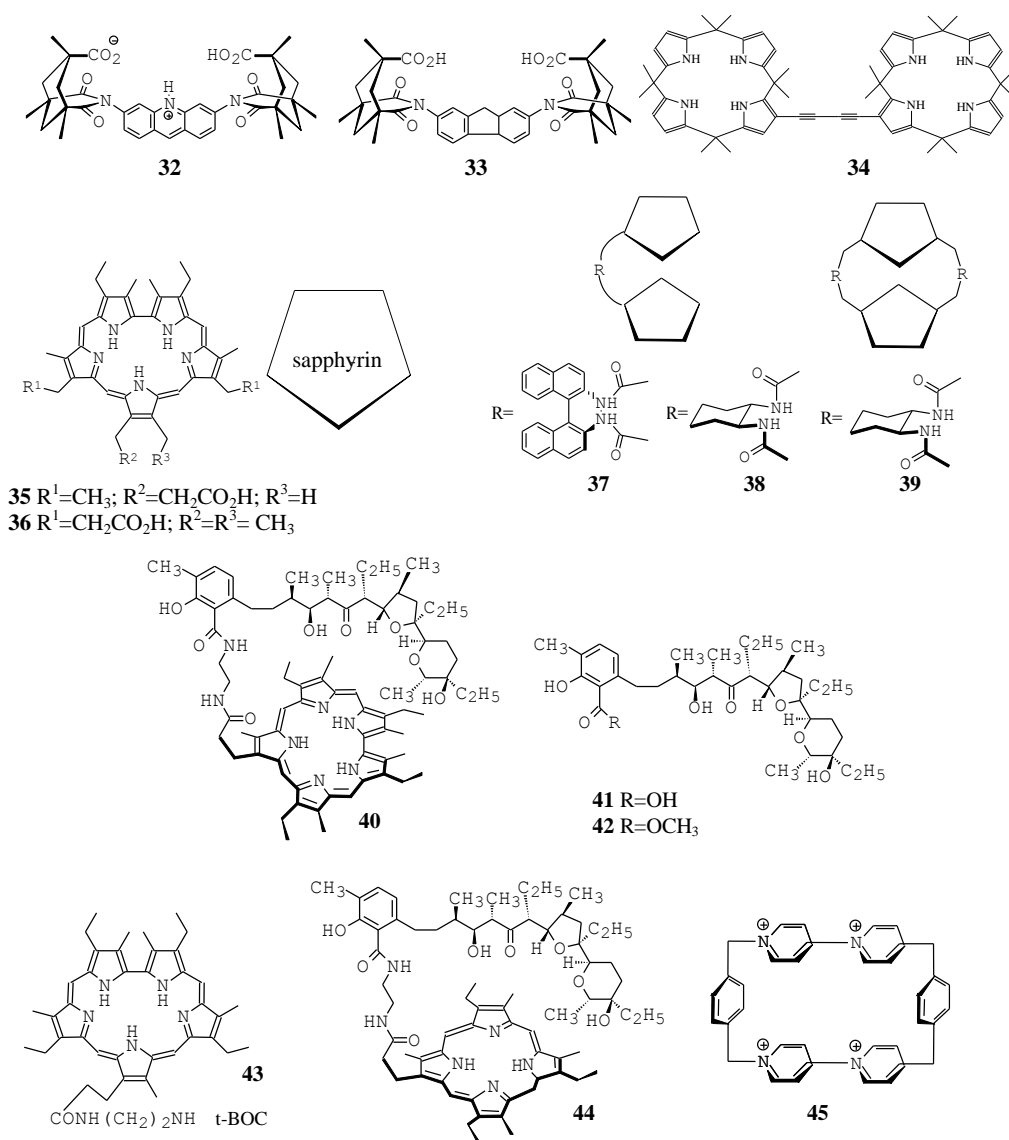
Figure 4. The structure of compounds 32-45.

A number of works is devoted to the development of pyrrole based receptors toward carboxylic acids [Sato et al. 2000; Sessler et al. 2003; Tomankova et al. 2003; Sessler et al. 1996]. Calix-pyrrole dimer with rigid butyliden-1,3-diine spacer 34 bounded isophtalate in CH_2Cl_2 with association constant $K_{as} = 4.2 \cdot 10^3$ M^{-1} [Tomankova et al. 2003]. Macrocycles, which demonstrated more efficient binding of carboxylate anions against halogenide anions were synthesized by condensation of bis-furan or pyrrole with thiophenes [Sessler et al. 2003]. In [Tomankova et al. 2003], the application of oligopyrrole macrocycles as sorbents for chromatography columns separation was investigated. It was proposed, association was controlled by a combination of electrostatic, π-π and hydrophobic interactions, with the latter two playing a predominant role.

Sapphyrin derivatives 35, 36 associated in acyclic (compounds 37, 38) and cyclic (39) dimers were used as building blocks in receptors for carboxylates [Tomankova et al. 2003]. Comparison of the association constants of the 1:1 complexes of 37 $2H^+$ in methanol showed that the aromatic substrates bounded more effectively than the aliphatic ones. The aliphatic monocarboxylates formed unstable complexes (*e.g.*, $K_{as} \leq 20$ M^{-1} for trifluoroacetate), meanwhile the stability of the complex with benzoate was higher by two orders of magnitude (1380 M^{-1}). Similar relation was observed for dicarboxylates. The only difference was that the association constant generally increased (phtalate – 310, isophthalate – 2400, terephtalate – 4600, oxalate – 260, malonate – 450 M^{-1}). The data of NMR spectroscopy and X-ray analysis revealed that this effect was not due to π-π interactions of aromatic rings in host and guest but because of the formation of hydrogen bond $C(sp^2)$-H$\cdots$N (distance H$\cdots$N: 2.38Å) and because of the coordination of the substrate aromatic ring perpendicular to the sapphyrin cycle in the complex formed. It was found that the cyclic dimer 39 bounded glutamate and aspartate anions with high affinity.

Aromatic amino acids, i.e. phenylalanine, tryptophan and tyrosine, were specified as the substrates for transport studies using receptors containing fragments of sapphyrin, porphyrin and lasalocid [Sessler et al. 1996; Tejeda et al. 2000]. At neutral pH, novel sapphyrin-lasalocid compound 40 was found to be very efficient carrier for phenylalanine and tryptophan through liquid membrane model system consisted of H_2O-CH_2Cl_2-H_2O layers placed in a U-tube. The organic layer contained the compounds 40-44 (Table 3). In direct competition experiments, L-phenylalanine was transported 1000 times faster than L-tyrosine. No one of the amino acids investigated was transported effectively by the compounds 41-43 representing structural "fragments" of the carrier 40. Substitution of sapphyrin (macrocycle 40) by a porphyrin fragment (carrier 44) led to suppression of the transport of aromatic amino acids. Authors reported the enantioselectivity of the membrane transport by the compounds studied was modest. Only receptor 40 showed selective carrier properties (k_l:k_d = 1.57) for D- and L-phenylalanine [Goodnow et al. 1991].

Table 3. Initial rates of amino acid transport (k_t)
[Goodnow T.T. et al. 1991; Oton et al. 2007]

	k_t ($\cdot 10^{-5}$, $mol \cdot cm^{-2} \cdot h^{-1}$)					
Carrier	L-Phe	D-Phe	L-Trp	D-Trp	L-Tyr	D-Tyr
40	20.0	12.7	5.0	4.2	0.02	0.02
44	0.7	0.9	0.3	0.2	<0.001	<0.001
43	6.9	-	1.4	-	<0.001	-
41	0.5	0.4	0.2	0.2	<0.001	<0.001
43+41	3.2	3.1	0.9	0.9	0.2	0.2
43+42	7.8	7.2	1.4	1.4	0.5	0.6
42	0.9	0.8	0.2	0.2	<0.001	<0.001
No carrier	0.05	-	0.01	-	<0.001	-

Water-soluble cyclophane 45 with lipophilic cavity bounded some aromatic α-amino acids, e.g., tryptophan and tyrosine, in acid medium [Goodnow et al. 1991].

2.3. RECEPTORS WITH GUANIDINE AND AMIDINE FUNCTIONS

The guanidine group of arginine participates in the binding of anionic substrates in enzymatic reactions as well as in stabilization of the tertiary structure of proteins due to formation of salt bridges with the carboxylate group. This offers opportunities to develop appropriate receptors based on guanidine fragments [Oton et al. 2007; Blondeau et al. 2007; Schmidtchen 1980; Echavarren et al. 1988; Muller et al. 1988; Echavarren et al 1989; Gleich et al. 1990; Schmidtchen 1990; Kurzmeier et al. 1990; De Mendoza et al. 1997; Metzger et al. 1995; Dietrich et al. 1978; Schmidtchen et al. 1997]. Strong interaction of guanidine cations with carboxylate anions is related to the formation of two strong hydrogen bonds accompanied with electrostatic interactions of charged particles (complex 46) [Blondeau et al. 2007] (Figure 5).
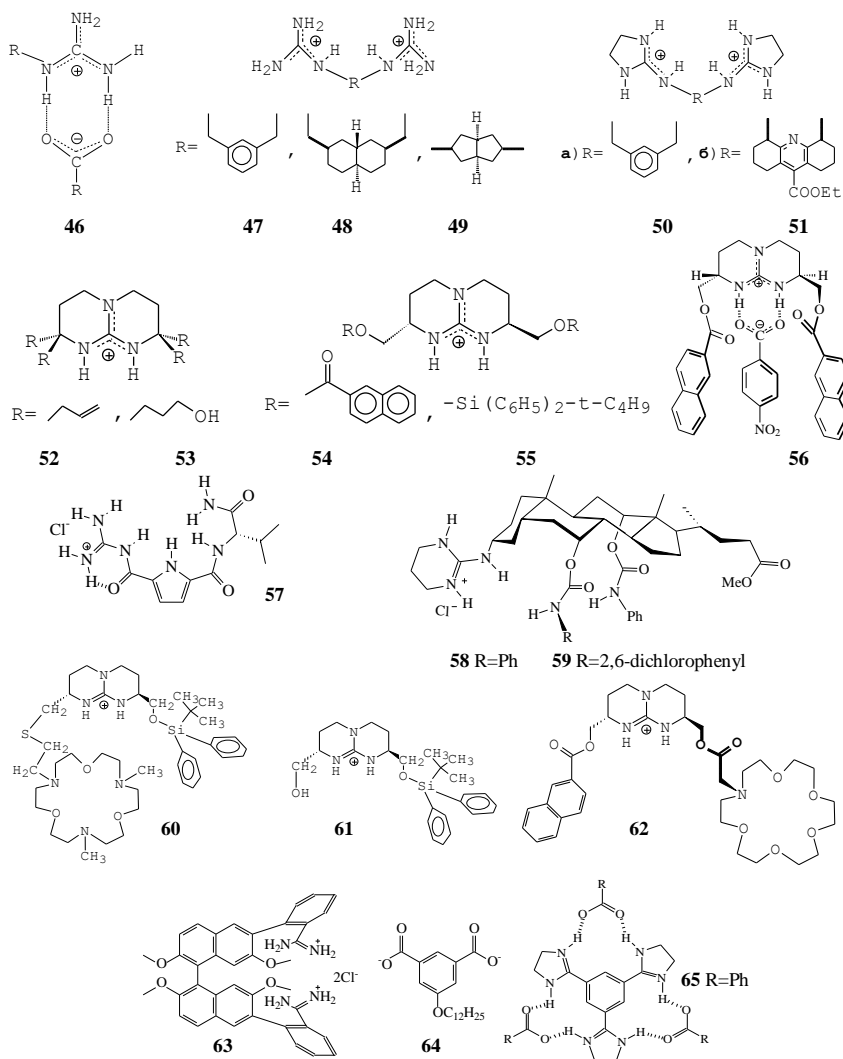

Figure 5. The structure of compounds 46-65.

High basicity of guanidine ($pK_a = 13.5$) makes it possible to keep it protonated in a wide pH range. Series of chain-ring bis-guanidine derivatives [Schmidtchen 1980; Echavarren et al. 1988] got together by bridging groups 47-51 [Muller et al. 1988; Echavarren et al 1989], and series of bicyclic compounds including guanidine function 52-55 [Kurzmeier et al. 1990; De Mendoza et al. 1997] were obtained. Bis-ligands [Metzger et al. 1995] showed high binding of carboxylates. Thus, the association constant 50 with fumarate was equal to $5 \cdot 10^4$ M^{-1} according to NMR-spectroscopy in DMSO-d_6.

Modification of bicyclic ligands allows enantioselective binding of aromatic α-amino acids [Dietrich et al. 1978]. Thus the naphtyl substituents of 54 promote effective binding of aromatic acids due to additional π-π interaction [Schmidtchen et al. 1997] (complex 56).

Effeciency of the binding of *N*-Ac-α-amino carboxylates by guanidine salts was enhanced by introduction of functional group able to form additional H-bonds, i.e. pyrrole [Katayev et al. 2007] and amide fragments in 57 [30], cholic fragment [Liu et al. 2007; Chahar et al. 2006] in 58 and 59 [Davis et al. 2001] into the receptor structure. The association of the receptor 57 with *N*-Ac-α-amino carboxylates in H_2O/DMSO mixture depended on lateral chain structure of the acid ($K_{as} = 3.5 \cdot 10^2$ $M^{-1} \div 5.3 \cdot 10^3$ M^{-1}) [Schmuck et al. 1999]. The efficiency of the extraction of *N*-Ac-α-amino acids from neutral and basic aqueous solutions by compounds 58 and 59 was quite high (52-93%) for the substrates containing nonpolar groups in side chains. The receptor 59 with the extraction ability (74-93%) higher than that of 58 due to increased acidity of its dichlorophenylcarbamoyl groups is of special interest [Metzger et al. 1996].

Molecular design of zwitterionic α-amino acids was performed in [Metzger et al. 1996]. In carrier 60, chiral bicyclic guanidinium salt functioned as an anchor of the carboxylate and triaza-crown ether as a binding moiety for ammonium cation. In combination with strong hydrophobic silyl ether, this contributed to complex formation and transfer of amino acids (phenylalanine, leucine, tryptophan, glycine) from water into dichloromethane with unprecedented efficiency at pH = 9. The quantification of the extraction by radiometry confirmed 1:1 stoichiometry of the complex. Te zwitterionic form of the species transferred was confirmed by decrease of the extraction coefficient of the substrate with increasing pH value. Modest enantioselectivity of the phenylalanine transfer was mentioned (40% of L-isomer). From the compounds 60 and 61, the latter one does not contain a crown ether fragment. The following range of extractability decrease was found in both cases: phenylalanine > leucine > tryptophan > glycine >> serine. The comparison of the receptors 60 and 62 [Galan et al. 1992] showed preferably extraction of most hydrophilic acids be 60. For serine and glycine, its efficiency was by 3000 times higher than that of 62 [Sell et al. 1995].

In amidinium salts suggested for carboxylate binding, high electrostatic and H-bonding interactions are mainly responsible for host-guest association [Sebo et al. 2000]. A series of receptors for dicarboxylate substrates was prepared by attachment of two phenylamidinium ions to 1,1'-binaphthalene scaffolds.

The 1,1'-binaphthalene derivative (±)-63 was found to be highly efficient receptor for glutarate ($K_{as} = 8.2 \cdot 10^3$ M^{-1}) and dianion 64 ($K_{as} = 1.0 \cdot 10^4$ M^{-1}) in CH_3OH with similar association constants [Lilienthal et al. 1996]. Low substrate selectivity was explained by the electrostatic interactions responsible for comlexation which did not have certain spatial orientation [Lilienthal et al. 1996].

The idea to use substituents structurally similar to guanidinium and amidinium groups was realized for nitrogen containing heterocycles [Gamez et al. 2007]. Charged or neutral compounds, such as *N,N'*-dimethyl-2,7-diazapyrenium [Lilienthal et al. 1996] and tris(imidazoline) [Kraft et al. 1998] were synthesized and examined as receptors for benzoate anions. Thus, the tris(imidazoline) formed complex with benzoate 65 (K_{as} = 990 M^{-1}) [Costero et al. 2007].

2.4. RECEPTORS CONTAINING UREA AND THIOUREA MOIETIES

Urea and thiourea derivatives are able to strong binding of carboxylates due to two H-bonds (complex 66) [Roussel et al. 2006; Costero et al. 2007] (Figure 6). Receptor 67 effectively interacts with tetrabutylammonium benzoate in CDCl$_3$ (K_{as} = 2.7·10^4 M^{-1}) [Webb et al. 1993].

Introduction of electron accepting substitutient in aromatic ring of 68 yielded in highly selective colorimetric sensor for acetate ion in aqueous solution that showed selectivity toward various other monovalent inorganic anions ($CH_3CO_2^- > H_2PO_4^- > Cl^-$, Br^-, I^-, SCN^-, NO_3^-, HSO_4^-, ClO_4^-) [Kato et al. 2001]. In the UV-vis. spectrum of 68 recorded in the absence of anions, a band with λ_{max} at 343 nm ($\varepsilon = 3.1 \times 10^4$ M^{-1} cm^{-1}) was observed which can be assigned as an intramolecular charge transfer. Addition of one equivalent of acetic acid resulted in remarkable changes of the band observed. Charge transfer absorption band shifted to 392 nm with shoulder absorption at around 450 nm. The color of solutions changed from colorless to yellow.

Compounds 69-71 containing urea, thiourea and guanidine groups bind carboxylates more effectively than carboxylic acids [Fan et al. 1993]. The energy ΔG_{295} strongly depended on the acidity of NH bonds in receptor structure. Increasing acidity in the range urea (pK_a = 26.9) < thiourea (pK_a = 21.0) < alkylguanidinium ($pK_a \approx 14$) led to increase of the association constant for the complexes of carboxylates and mono - and bidentate ligands 69-71 [Gunnlaugsson et al. 2005].

Substitution of butyl radical at nitrogen atoms by glucopyranose fragment in 72 led to dramatical decrease in the stability constant of 1:1 complexes with glutarate in CDCl$_3$. Probably this could be related to lower lipophilicity of 72 (391 M^{-1}) [Benito et al. 2001]. Introduction of two additional binding fragments of thiourea (compound 73) resulted in effective binding of glutarate (K_{as} > 10^6 M^{-1}) coordinated in pseudo cavity of a receptor. Replacement of *meta*-methylene bridge in 70 by anthracene fragment made it possible to use the receptor 74 as fluorescent molecular sensor for malonate [Gunnlaugsson et al. 2005].

Host molecules obtaining by modification of aminochromenone 75 were used as receptor of carboxylate group [Raposo et al. 1994]. Ligands containing symmetric fragments of urea 76 and sulfurylamides 77 can form four hydrogen bonds with carboxylate substrates.

The receptor with high effective molecular recognition of carboxylates of α-heterocyclic and α-keto acids was prepared by introduction of hydroxamic function in a neutral receptor with bis-chromenylurea fragments. This resulted in improved complexation of the host with the guests mentioned due to formation of the fifth hydrogen bond in the complexes between the OH donor of the host and the α-heteroatom acceptor of the guest molecules [De la Torre et al. 2001].
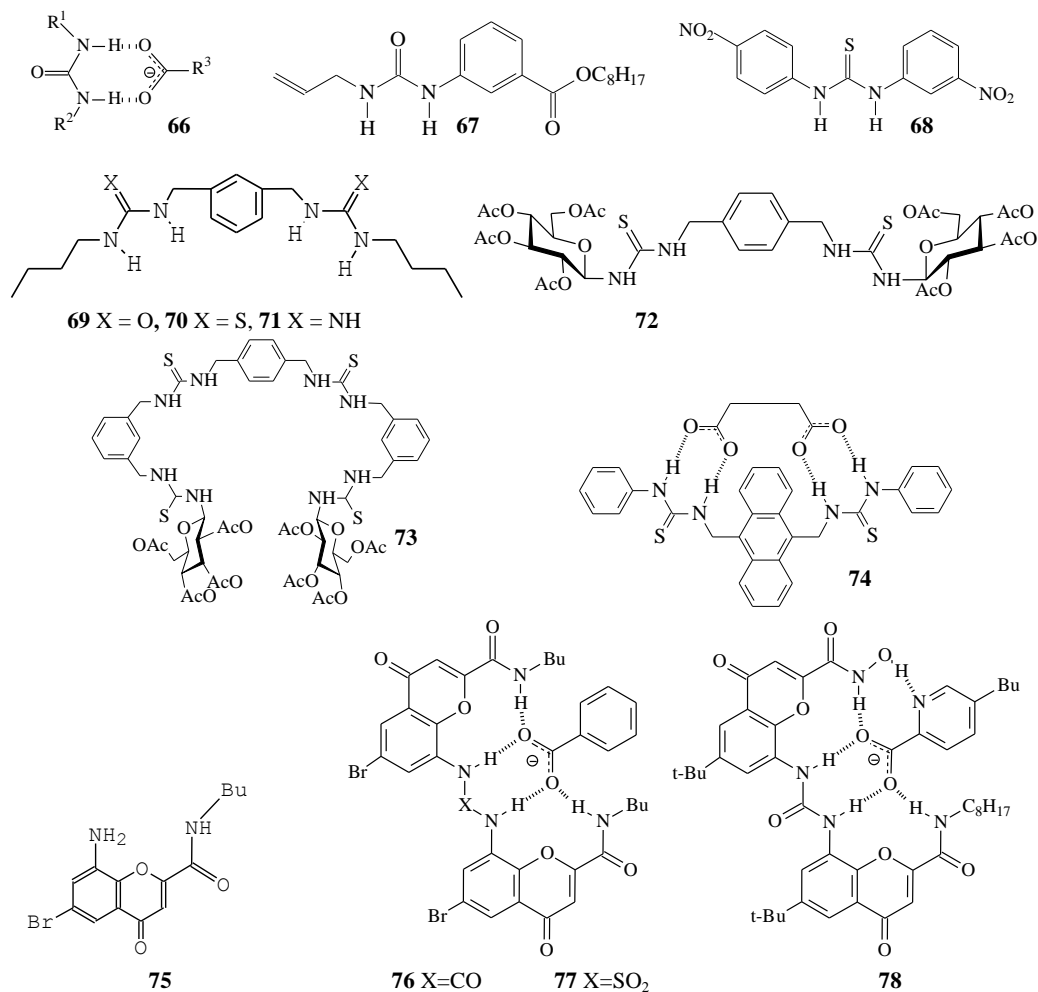
Figure 6. The structure of compounds 66-78.

Two binaphthyl crown receptors containing phenylboronic acid 79 and 2,4-dinitrophenylurea 80 as bridges were synthesized from the optically active binaphthyl crown alcohol [Tsubaki et al. 2003] (Figure 7). Receptor 79 selectively extracted γ-aminobutyric acid among other amino acids $H_2N-(CH_2)_n-COOH$ (n = 1-6). Contrary to the other hosts specific for γ-amino acids, the compound 80 is more effective for ε-aminohexanoic acid extraction.

For the complex of 80 and *tert*-butylammonium valerate it was established there are not only H-binding between urea fragment of the host and carboxylate group of the guest, but also interaction of ammonium groups with a crown ether ring in the complex 81.

Application of monosubstituted boronic acid for zwitterionic form for binding and recognition realized by the formation of three-component complex has been described [Reetz et al. 1994]. Arylboronic acid and proton-accepting host were used as receptors for ammonium and carboxylate group, respectively. The model proposed in [Jeonget al. 1996] was realized in 83-85 with crown ethers as donor groups. The use of crown ether modified

phenylboronic acid 83 decreased relative transport rates in comparison both with phenylboronic acid and 18-crown-6 probably due to steric limitations.
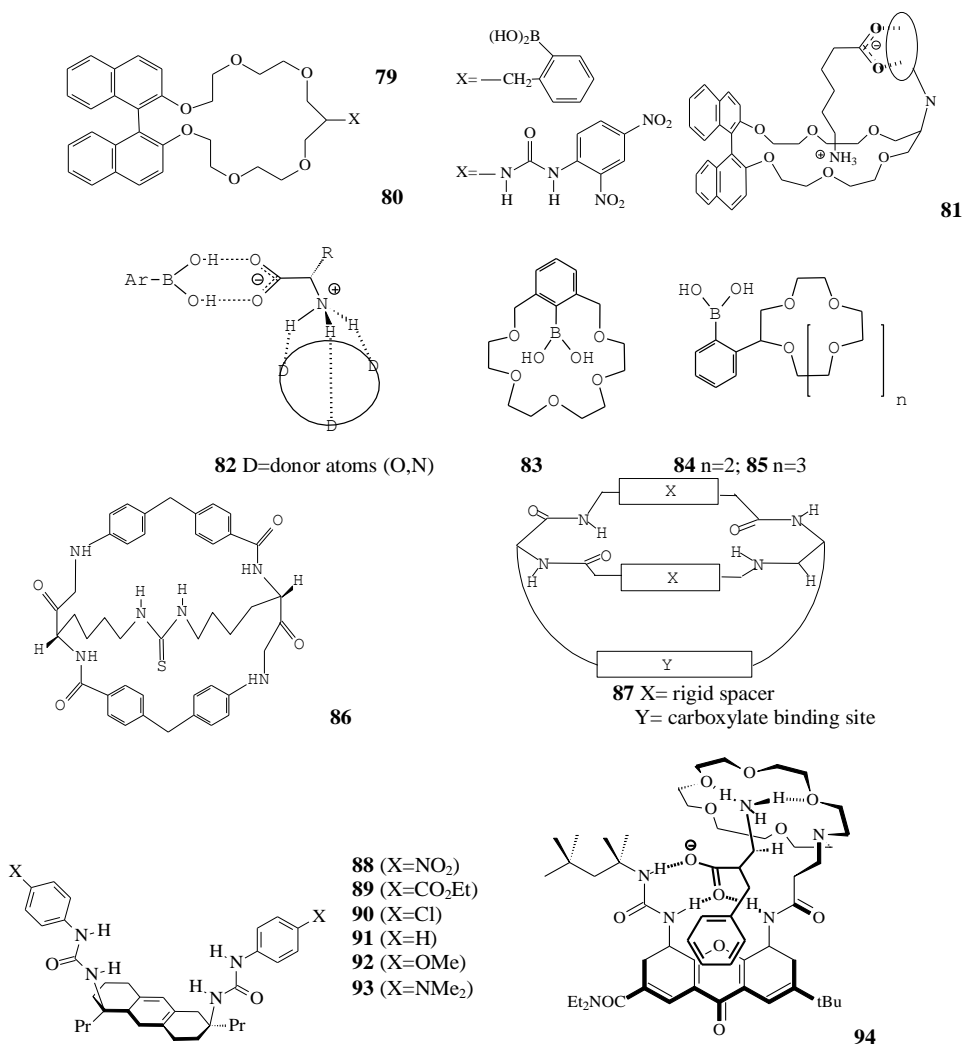
Figure 7. The structure of compounds 79-94.

Proton exchange in one of the macrocycle methylene groups on the arylboronic residue (receptors 84 and 85) resulted in increasing rate of phenylalanine transfer in comparison with initial crown-ethers. In case of the receptor 84 and the mixture of 15-crown-5 and phenylboronic acid, transport rate of substrate was found to be the same. Thus, the compound 85 appeared the most effective carrier for amino acids.

A novel macrobicycle 86 featuring a thiourea was prepared as a receptor for carboxylate according to the model presented on the scheme 87 [Pernia et al. 1995]. A novel macrobicycle 86 with a thiourea fragment was synthesized as receptor for carboxylate in accordance with the model 87 [Pernia et al. 1995]. Investigation of the binding with the receptor 86 was carried out using ^{1}H NMR with various N-acylated α-amino acids taken as

tetrabutylammonium salts. It was shown, that the compound 86 was an effective receptor and that binding via carboxylate-thiourea interaction dominated. The glycine derivative is bounded most strongly, but it was less effective for derivatives of alanine, phenylalanine, asparagine, histidine and glutamine. This assumes that the side chain of amino acids is less important, presumably due to steric reasons.

Table 4. Stability constants ($K_{as} \pm 10\%$, M^{-1}) of receptor complexes 88-93 with carboxylates (DMSO-d_6, 297 K) [Goodnow T.T. et al. 1991; Oton et al. 2007]

Receptor	Carboxylate	K_{as} (M^{-1})	Receptor	Carboxylate	K_{as} (M^{-1})
91 (X=H)	glutarate	1430	88 (X=NO$_2$)	adipinate	21800
	adamantane 1,3-dicaroxylate	2580	89 (X=CO$_2$Et)		6840
			90 (X=Cl)		2360
	isophtalate	920	91 (X=H)		1710
	adipinate	1710	92 (X=OMe)		1400
			93 (X=NMe$_2$)		510

New dicarboxylic acid was synthesized from 1,5-dichloroanthraquinone and modified to ditopic receptors 88-93 for binding of dicarboxylate salts by multiple hydrogen bonds [Jeong et al. 1996]. The stability constants of the complexes with various dicarboxylates have been quantified in DMSO-d_6 (Table 4). The derivative with -NHC(O)NH(4-C$_6$H$_4$-NO$_2$) fragment was found to be most effective receptor for adipinate. The efficiency of binding by host 91 increased in the following range of dicarboxylate range: isophtalate < glutarate <adamantane-1,3- dicarboxylate. This was explained [Poh et al. 1994] first by conformational rigidity of adamantane-1,3-dicarboxylate, and, second, by lower basicity of isophtalate in comparison with glutarate. On example of a series of *p*-substituted bis-phenylureas 88-93, the dependence of the stability constants on electron accepting effect of the substitutient in *para*-position of phenyl ring was found. The stability constant value for adipate was increased by 40 times from 88 with urea at nitrogroups to 93 with N, N-dimethylaminogroups. Both compounds mentioned form equal number of H-bonds with dicarboxylate.

Importance of additional π-π interactions in binding of zwitterionic amino acids by the receptor 94 was shown [Hernandez et al. 2003]. This receptor most effectively extracted the aromatic amino acids, e.g., phenylalanine and tryptophan, unlike others lipophilic amino acids (leucine or valine). Glycine with its small size well corresponded to a receptor cavity makes an exception. Others hydrophilic amino acids (serine, asparagine, tyrosine) do not extract by given receptor.

2.5. RECEPTORS CONTAINING AMIDE GROUP

Compounds 95 and 96 contain conformationally rigid (cyclic) sin-amide fragments, providing effective interaction with carboxylic group of a guest (Figure 8). Receptor 95 binds more effectively malonic acid and receptor 96 aromatic acids, e.g., *p*-ethoxybenzoic acid [Poh et al. 1994].

Introduction of additional amide group in receptor 98 with formation of 1-tetralone derivative made it possible to bind benzoic acid [Moore et al. 2001]. The appropriate

association constant was found to be 80 M^{-1} for 97 and 200 M^{-1} for 98 according to ^{1}H NMR titration.
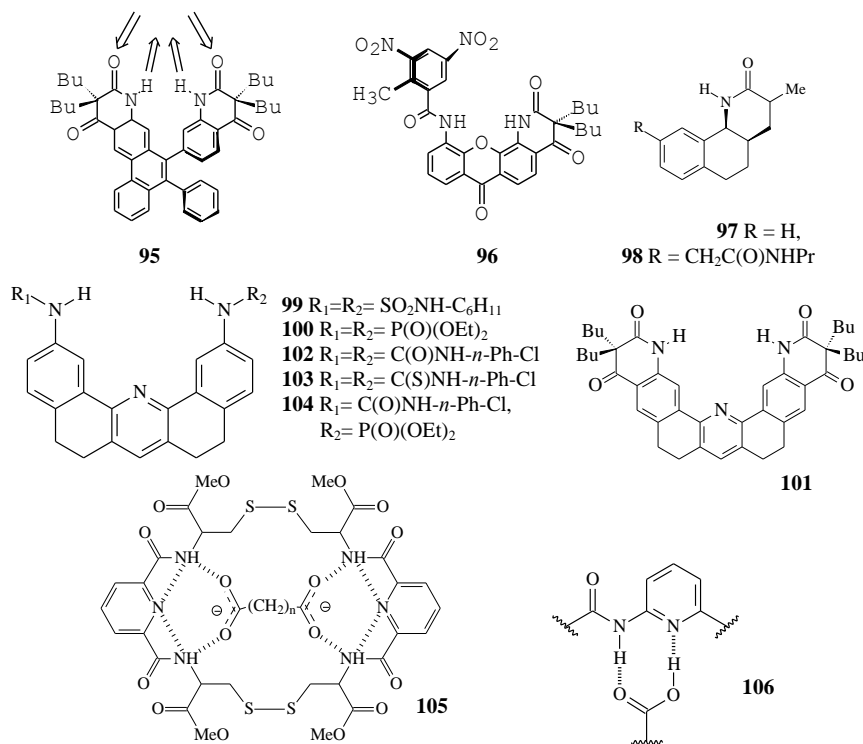

Figure 8. The structure of compounds 95-106.

Effective receptor based on secondary squaramide base (3,4-diaminocyclobutene-1,2-dion) for glutarate binding was described [Prohens et al. 1998]. In the receptor, the spatial arrangement of the groups participating in formation of H-bonds is similar to that in urea, thiourea and guanidine. As was shown, the presence of positively charged group near the center of H-bonds increased the stability constants of the complexes formed by 8-10 time.

New molecular receptors 99-104 with dibenz[c,h]acridine skeleton bearing functional groups complementary to malonic acids have been developed [Mussons et al. 1999]. Dibutylmalonic acid is more effectively bonded by receptor 102 containing cis-amide groups ($K_{as} = 2.8 \cdot 10^5$ M^{-1}).

Stability constant of malonic acids/dithiourea complex 103 was four times lower, than that with urea 102. According to [Gonzalez-Alvarez et al. 2008], it is caused by S atom in thiourea fragment which is weaker acceptor of H-bond than O atom. All the synthesized receptors catalyzed decarboxylation of dibutylmalonic acid [Gonzalez-Alvarez et al. 2008].

Simple one-step strategy for the synthesis of aromatic bridged cystine macrocycles (cystinophanes) for the binding of dicarboxylates [(CH$_2$)$_n$(COO$^-$)$_2$, n = 1 − 4] to form the complex 105 was described [Ranganathan et al. 1998]. The receptor ability of macrocycle was explained by preorganisation of its binding sites consisted of anion sensing amide protons required for complexation 105. Obviously, four intramolecular H-bonds between N-pyridine atoms and amide protons direct N-H groups into the macrocycle cavity. The

compound 105 showed maximum affinity (K_{as} = 3.69 × 10^2 M^{-1}) and selectivity for glutaric acid (*n* = 3) dianion.

2.6. RECEPTOR WITH AMIDOPYRIDINE FRAGMENT

Amidopyridines bind carboxylic group by two H-bonds, i.e. that of hydrogen atoms of carboxylic group with N-pyridine and carbonyl of carboxylic acid with amide proton (complex 106, Figure 8). It should be mentioned that N atom of a heterocyclic fragment can be protonated or form H-bonds with polar solvents. Thus amidopyridines are effective commonly in rather low-polar media [Gonzalez-Alvarez et al. 2008; Ghosh et al. 2007].

Implementation of amidopyridine fragments into the macrocycle structure 107 and acyclic compound 108 connected by conformationally rigid *para*-xylene spacer made it possible to realize binding of carboxylic acids [García-Tellado et al. 1990] (Figure 9). The macrocycle 107 interacts with ethylmalonic (K_{as} = 7.3·10^3 M^{-1}) and diethylmalonic (K_{as} = 1.1·10^3 M^{-1}) acids in CDCl$_3$ with rather low association constants [Geib et al. 1993]. For formation of inclusion complex 107, dicarboxylic acid should be in energetically unfavorable conformation. The acyclic receptor 108 forms stable complexes with dicarbonic acids in CDCl$_3$, the strongest complex was observed in case of adipinic acid (K_{as} > 10^5 M^{-1}).

It was shown, that bis-(2-amidopyridine) 109 with *meta*-xylene spacer stabilized S-cis-rotamer of dicarboxylic acids on a proline platform 110, whereas the derivative with naphtalene spacer 111 stabilized preferably S-trans-rotamer [Vicent et al. 1991]. Complementary of the amidopyridine fragments and dicarboxylic acids was also used to design of extended helical and planar ribbon supramolecular assembles [Geib et al. 1993].

2-Acrylamido pyridine 112 was shown to be a versatile receptor for a variety of carboxylic acid derivatives but it did not bind hydroxyl and amine groups [Steinke et al. 1999].

Monoamidopyridine derivatives with additional amide or urea groups bind both the carboxylic acid and amide groups. Thus they are considered as effective receptors for acylated amino acids, e.g. *N*-Ac-Pro-OH. Chiral receptor 113 bound ACE inhibitor captopril in CDCl$_3$ with a 2:1 enantioselectivity in favor of (*R*)-captopril (K_{as} = 500 M^{-1}) [Vicent et al. 1992]. The receptor 114 interacts with the maleimide acid 115 (K_{as} > 4.8 × 10^3 M^{-1}) in CDCl$_3$ and accelerates the 1,4-addition of a thiol to the maleimide [Fitzmaurice et al. 2002].

Ability of amidopyridines to bind carboxylic acids was enhanced by incorporation of additional hydrogen bond formed by *syn* lone pair of the carbonyl oxygen atom in molecular clefts, e.g., 116 [Bilz et al. 1997; Goswami et al. 1996]. For R = Ph or 1-napthyl, the complexation of aromatic carboxylic acids (*e.g.* naproxen, phenylacetic acid and hydratropic acid) led to upper field shift of the α -H signals in the ^{1}H NMR spectra in CDCl$_3$ (Δδ ≈ 0.28 ppm). This suggests π-π stacking interactions.

Receptor 117 [Hamilton et al. 1993] realizing three-point hydrogen bonding of the amide group in guest and nitrogen atom to the pyridine of 2-aminopyridine fragment in the host molecule was synthesized. According to IR, ^{1}H and ^{13}C NMR spectroscopy, the efficiency of binding of 117 with aliphatic acids was higher than that of 118, meanwhile aromatic acids cooperate with 118 more stronger. Such a behavior can be explained by steric hindrance of

the formation of third H-bond between N-H bond of host and one of *ortho*-H atoms at phenyl rings of a guest in case of the complex 117.
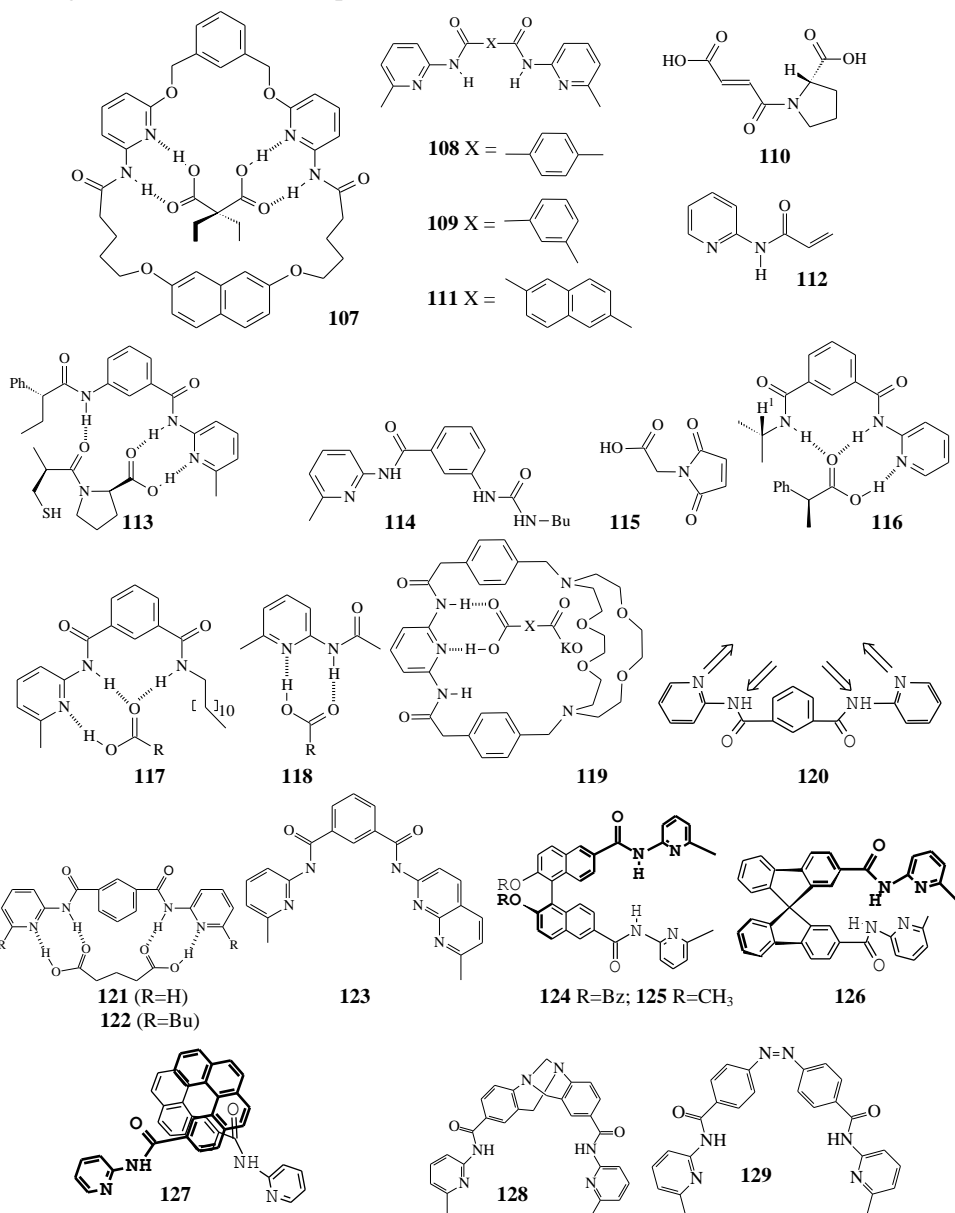
Figure 9. The structure of compounds 107-129.

Amidopyridines have been incorporated into a number of more complicated macrocyclic architectures for binding various carboxylic acid derivatives. For example, receptor 119 was synthesized to bind monoanions of various biscarboxylic acids [Flack et al. 1992] using a combination of H-bonding interactions and electrostatic association between the carboxylate anion and metal cation bonded to crown ether. The mode of binding proposed was supported by ROESY experiment and FAB mass spectrometry data. It is important that 119 bound the

monocarboxylate salt of maleic acid but not that of fumaric acid which was probably too large for the macrocyclic cavity of a host molecule.

A number of works is devoted to development of acyclic hosts with bisamidopyridine fragments [Geib et al. 1993; Steinke et al. 1999; Hamilton et al. 1993, Pant et al. 1988]. Thus, ligand 120 incorporates dicarboxylic acids, e.g., adipinic and glutaric acids, using H-bonds. The complex structure was confirmed by X-ray analysis [Geib et al. 1993; Steinke et al. 1999]. Substitution of the amidopyridine groups by urea, thiourea, and guanidine fragments led to more effective binding of carboxylates in comparison with carboxylic acids [Gunnlaugsson et al. 2005].

The compound 121 exerted good carrier properties for acetic, citric, succinic, glutaric, dimethylglutaric and adipinic acids. The experiments were performed in supported liquid membranes. 0.0022 M Receptor was dissolved in a kerosene/dodecanol mixture (10%) [Palet et al. 2000]. The use of butyl group instead of methyl one (compound 122) provides transport of glutaric, dimethylglutaric and adipinic acid. Other acids were not transferred through the liquid membrane. Receptor 123, which bound insoluble tartaric acid in $CHCl_3$, was produced by implementation of naphthylpyridine units in the structure 106 [Goswami et al. 1996].

For the synthesis of pre-organized receptors on the base of bispyrimidine 124-129, 9,9'-spirobifluorene, 1,1'-binaphtyl and hexahelicene [Owens et al. 1993], Troger's base receptor [Goswami et al. 1997, Goswami et al. 2000], azobenzene [Goswami et al. 1999] were used. The receptors form complexes with various dicarboxylic acids, from diethylmalonic to pimelic (heptanedioic) acid. Selective binding of various stereoisomers in some cases was reached with chiral molecular platforms of the hosts. As for all bis-ligands [Huggins et al. 2007], stability of the complexes formed depended on complementarity of cooperating groups. Conformationally labile compounds incorporate substrate with lower stability constants [Tabushi et al. 1977].

2.7. METAL-BASED RECEPTORS

In general, there are two possibilities of development of metal-based receptors for carboxylic acids and carboxylate anions sensing. The first one is direct coordination of a substrate with the metal center in the receptor molecule. This is typical for many natural biomacromolecules, e.g., proteins. For example, metalloenzymes can bind anionic substrates by three-component complexation: ligand (protein) - central metal cation -anionic guest. Second one is the use of metal cations as a template for assembling receptor from the fragments containing binding sites [Goel et al. 2007].

Metallo-cyclodextrins 130 and 131 illustrate first approach to carboxylate anion recognition (Figure 10). They bind cycloalkyl and aromatic carboxylates in buffered media (pH = 10) [Tabushi et al. 1977]. Interaction between carboxylate anions and macrocyclic receptors [Boiocchi et al. 2007] containing metal cations ($M^{2+} \equiv Cu^{2+}$, Co^{2+}, Ni^{2+}, Zn^{2+}) was shown to be responsible for inclusion of the oxalate dianion into the macrocycle cavity with formation of the μ-bridged structure (complex 132) [Schmidtchen et al. 1997; Martell et al. 1988]. Logarithm of the complexation constants for oxalate dianion (lgK_{as}) was found to be 5.59 with Cu^{2+} and 9.06 for Co^{2+} cations [Sell et al. 1995].
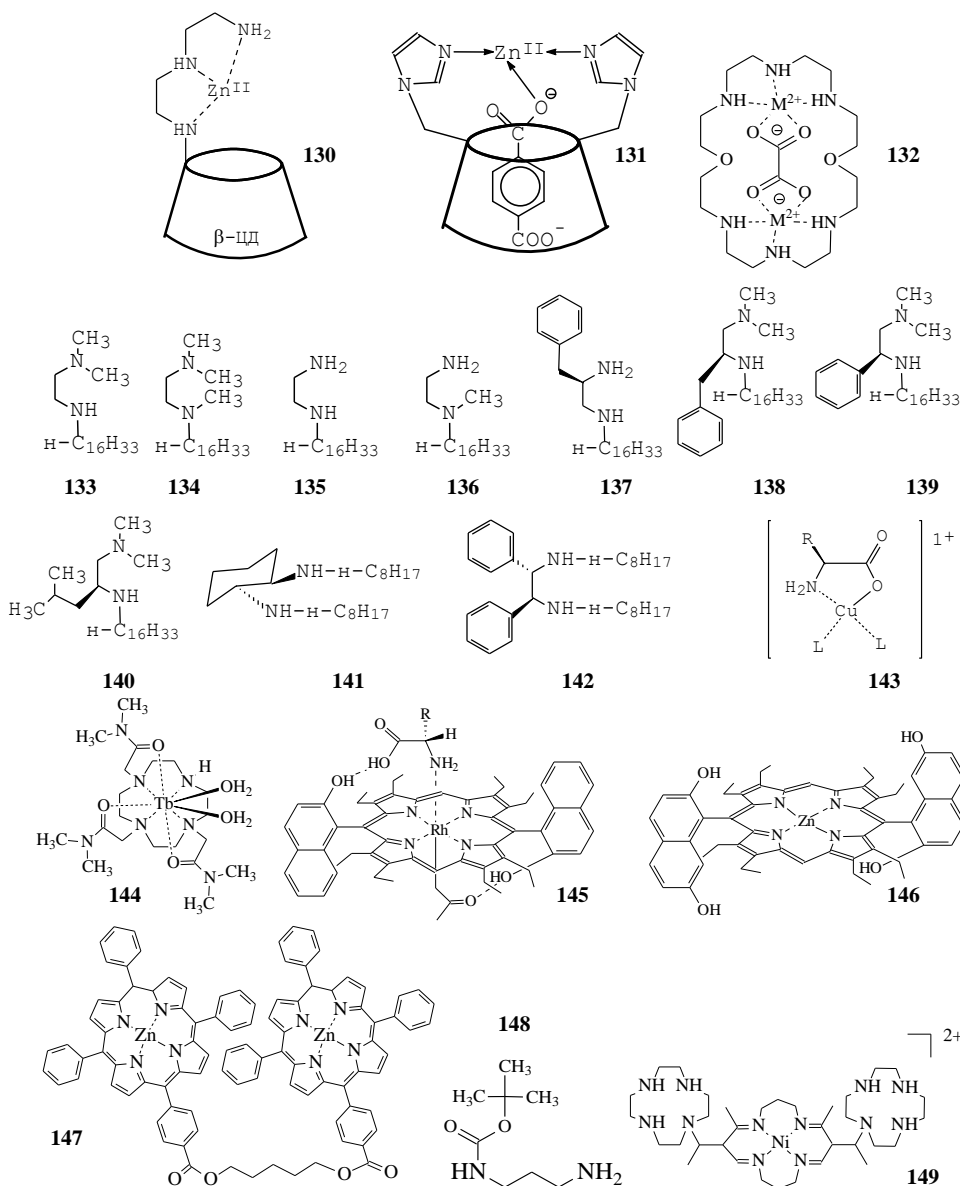
Figure 10. The structure of compounds 130-149.

Water-soluble cyclophanes with binding sites similar to that of vancomycin interact with carboxylate anions and can serve as receptors for the salts of aromatic, aliphatic acids and N-acylated α-amino acids [Hinzen et al. 1996].

Coordinating interaction with metal cations, similar to the function of metallo-enzymes, is used in molecular recognition of α-amino acids [Sell et al. 1995]. As shown [Potvin et al. 1987; Scrimin et al. 1988], various chelate complexes of Cu (II) with nitrogen-containing organic ligands, e.g. 133-142, can be applied for recognition of 2-aminoacetate fragments.

Achiral (133-136) and optically active (137-142) derivatives of 1,2-diaminoethane were synthesized. These receptors were successfully used for simultaneous transport of Cu (II)

cation and α-amino acids through the organic phase (CHCl$_3$) from buffer solution (pH=5.5). EDTA was added to the receiving phase [Scrimin at al. 1995]. In this case, transport is occurred by the formation of tertiary complex 143 consisting of metal cation, lipophilic diamine and amino acid.

It was shown, that the hydrophobic lanthanide complexes with tris-(β-diketonates) can transfer such α-amino acids as valine and leucine through lipophilic membranes from neutral aqueous solutions [Tsukube et al. 1996]. Contrary to the complexes of Eu (III), complexes of Tb (III), e.g. 144, could bind carboxylates [Gunnlaugsson et al.2002]. The substitution of two molecules of water by aromatic carboxylates (e.g., salicylate anion) in the complex of Tb (III) 144 led to remarkable enhancement of luminescence intensity against Tb (III) [Tsukube et al. 1996]. The phenomenon can be used for determination of the substrate concentration in different solutions.

Lipophilic amino acids are bonded by modified rhodium porphyrinate 145 in chloroform. The reaction involves coordinative interaction of the cation with the amino group and the hydrogen-bonding interaction between hydroxyl group of receptor and the carboxyl fragment of an α-amino acid [Redman et al. 1997]. Similar type of receptor based on modified zinc porphyrinate 146 with proton-donor groups in the side chains, was also used as a receptor for α-amino acid derivatives [Tsukube et al. 1996]. Most effective three-point interaction with dimethyl ester of aspartic acid is realized by two hydrogen bonds between the ester fragments of a guest and the hydroxyl groups of a host, as well as by coordination of the interaction between zinc cations and the amino group.

Chiral host-guest recognition was applied for the determination of absolute configuration of carboxylic acids with α-stereogenic center [Proni et al. 2002]. It was proved by IR spectroscopy, that complexation between receptor and carboxylic acid was guided by coordination of amide carbonyl and zinc ion C = O → Zn. The value of the stability constant of the complex of porphyrine "tweezers" 147, carrier 148 and L-tartaric acid with α-stereogenic center was determined ($4.65 \cdot 10^5$ M^{-1}). Although the O-nucleophiles are weaker axial ligands for Zn-porphyrins in comparison with N-nucleophiles, it was shown that the coordination between the neutral amide groups and metal cations take place on an oxygen atom [Potvin et al. 1987].

Metal-based molecular receptor 149 on 16-mer macrocyclic platform contained Ni(II) and two cyclic tetraaminoethylene fragments were synthesized. This compound was suggested for recognition of dicarboxylic acids [Kryatova et al. 2002]. The receptor in saddle conformation participates in two-point recognition of bifunctional substrates. It was demonstrated by NMR titration in DMSO-d_6 that this compound could interact with dicarboxylic acids with maximal selectivity for cis-1,2-dicarboxylates, e.g., malonate, maleate and orthophthalate (stability constants from 10 to 10^4 M^{-1})

The use of a metal cation as a matrix (template) for self-assembly of receptor fragments containing the binding sites is another interesting approach to design of metal-based receptors for carboxylic acids and carboxylate anions. It was applied for the synthesis of the receptors 150-152 [Potvinet al. 1987; Scrimin et al. 1988] (Figure 11). Assembling catechol molecules with thiourea and amidopyridine units around the central ion of the transition metal led to the formation of symmetric complexes able to efficient binding of dicarboxylates C$_4$-C$_7$ (lgK_{as} 5.7-6.9). Conformational mobility of the receptor leads probably to low selectivity of

interaction. Receptor 151 showed selectivity to C_4-C_5 acids, whereas more flexible receptor 152 preferably bonded C_7-C_8 acid.
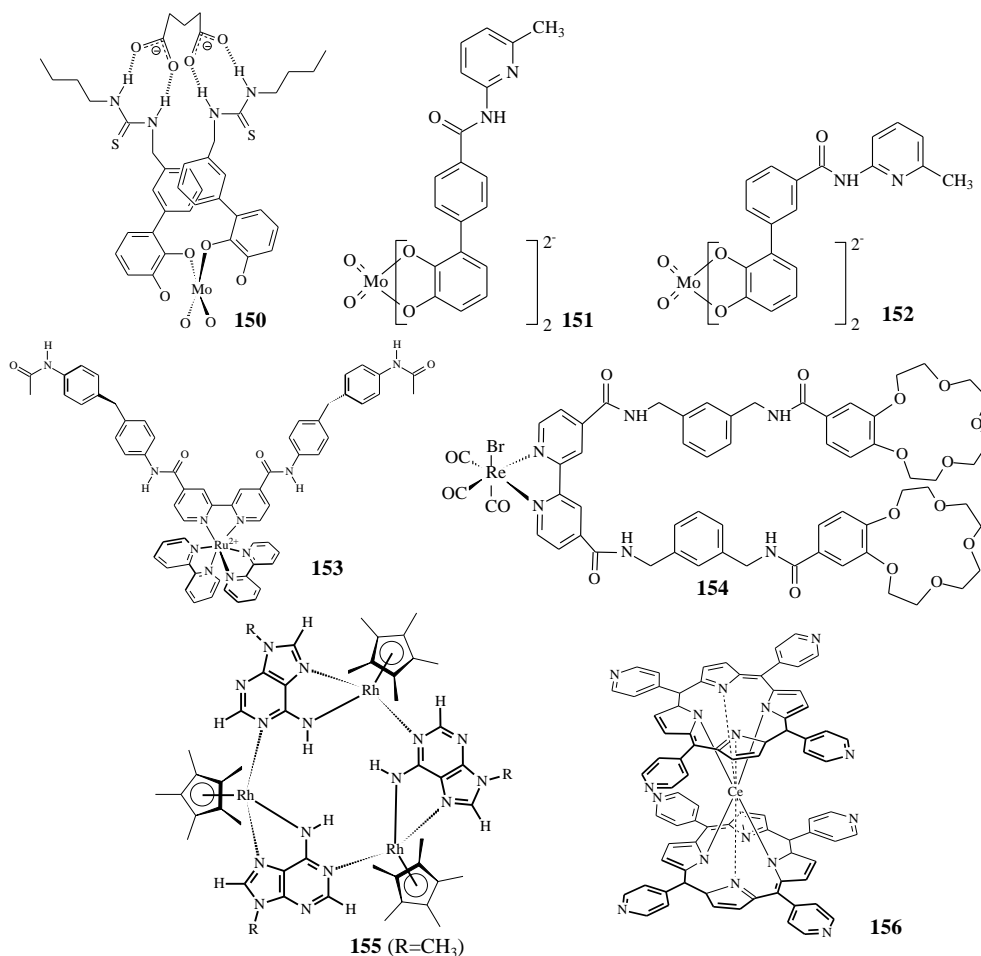
Figure 11. The structure of compounds 150-156.

Also, the bipyridine ruthenium complex 153 (Watanabe et al. 2000) exhibited pronounced efficiency but weak selectivity toward aromatic and aliphatic carboxylates and dicarboxylates ($K_{as} > 10^4$ M, UV, NMR, DMSO-d_6). Two-handed podand based on bipyridine rhenium (I) complex 154 with two amido and two crown ether moieties demonstrated allosterically controlled complexation of acetate anion. This molecule formed sandwich complex coordinated by potassium cation which could bind acetate [Uppadine et al. 2001].

As indicated by ^{1}H NMR spectroscopy, cyclic complex 155 can realize molecular recognition of aromatic amino acids in aqueous solutions (pH=7) [Chen et al. 1995; Chen et al. 1996]. Complex 155 can bind in neutral aqueous phase only lipophilic amino acids with aromatic fragments, e.g., L-tryptophan and L-phenylalanine [Takeuchi et al. 1998]. The study of organometallic complexes of receptors with aromatic and aliphatic acids in the gas phase by mass spectroscopy showed that non-covalent π-π interaction were the most important for complexation.

Strong positive cooperative effect (Hill coefficient 4) was detected for complexation of four molecules of dicarboxylic acids with receptor based on the bis[tetrakis(4-pyridyl)porphyrinate] Ce(IV) 156 [Takeuchi et al. 1998]. The cooperative effect is related to suppression of rotation of porphyrin rings of receptor against each other after the binding of the first guest molecule.

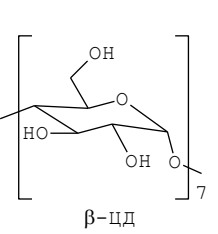## 2.8. Receptors Based on Cyclodextrins and Calixarenes

The compounds with large intramolecular cavity [Asfari et al. 2001] are very promising receptors interesting for potential use in living systems [Durmaz et al. 2007; Sun et al. 2007]. Meanwhile selective interaction of functionalized crown ethers and calixarenes with cations have been widely studied for almost five decades, the application of cyclophane based receptors for carboxylate anions and carboxylic acids recognition become of special interest rather recently. Such compounds can also be successfully used as building blocks and molecular scaffold in the construction of more sophisticated supramolecular systems. This review is focused on recent developments in the design and synthesis of calixarene-based carboxylate anions and carboxylic acids receptors.
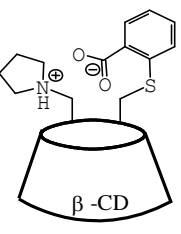
Highly specific binding of the substrates with hydrophobic centers of enzymes in water can be modeled by lipophilic cyclodextrin cavity 157 (cyclic α-glucose oligomers), framed by array of hydrophilic hydroxyl groups [Clark et al. 2006; Yuan et al. 2006] (Figure 12). The complexation between carboxylic acids, e.g., benzoic, 4-methylbenzoic, R-phenylpropionic, S-2-phenylpropionic acids and their conjugated bases with α-cyclodextrin and methylated α- and β-cyclodextrins was investigated by potentiometric titration (Hendrickson et al. 1995). The interaction of 4-methylbenzoic acid with hexakis(2,3,6-tri-O-methyl)-α-cyclodextrin is most successful example.
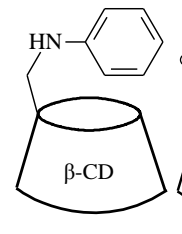
Receptor molecules based on crown ethers were proposed for recognition of the substrates with ammonium fragments (K_{as}=4.3-4.7). Similar receptors are used for interaction with lysine containing sites of small peptides [Späth and König 2010]. Modified β-cyclodextrin 158 was prepared and used for the binding of tryptophan (pH=9.8) [Tabushi et al. 1986].
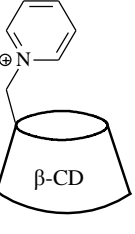
Thermodynamic characteristics (K_{as}, $\Delta H°$, $\Delta S°$) of the complexes of various amino acids and compounds 159 and 160 at 1:1 stoichiometry were quantified [Liu et al. 1997] by UV–vis titration in buffer (pH=7.2) at 25.0-40.0°C. The highest values of stability constants were obtained in the case of compound 160 with glutamic and aspartic acid (K_{as} 4.17 and 4.04, respectively). It is interesting that cyclodextrin 160 with positively charged groups showed rather weak complexation contrary to 159. Probably, this is due to electrostatic interaction. Meanwhile, the compound 160 is more enantioselective receptor (L/D=2.1 for valine and 1.8 for serine). Thus, it can be concluded that molecular flexibility and geometric complementarity play significant role in chiral recognition of amino acids by the receptors.

Calix[4]resorcinarenes are first examples of calixarenes applied for recognition of carboxylic acids. Choline, acetylcholine, carnithine and salts of the aromatic amines (dopamine, ephedrine, norephedrine, adrenaline and noradrenaline) were used as guests for resorcinarenes [Späth and König 2010]. Calix[4]resorcinarene 161 functionalized with hydrophobic hydrocarbon chains at the lower rim and exhibited selective binding of glutaric

acid in chloroform is another interesting example of receptor design. Macrocycle 161 is capable of two-point hydrogen bonding of guest molecules [Tanaka et al. 1990].
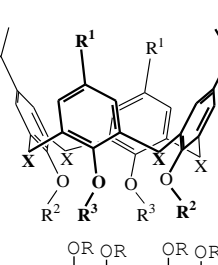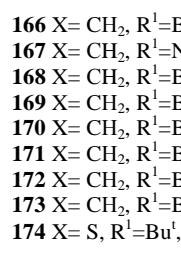
$X(CH_2)_nCOOH$
162 n = 3, X= COOH
163 n = 3, X= CH_3
164 n = 3, X= $COOCH_3$

161 R = $-(CH_2)_{10}CH_3$

165 R=$(CH_2)_3(OCH_2CH_2)_3OMe$

166 X= CH_2, R^1=Bu^t, R^2=$CH_2CH_2NHC(O)Ph$, R^3=H
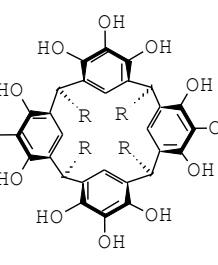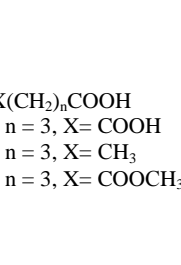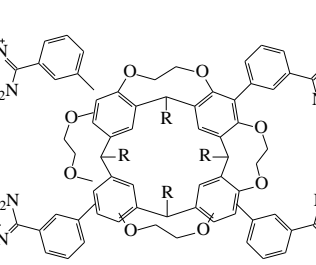**167** X= CH_2, R^1=NO_2, R^2=$CH_2CH_2NHC(O)Ph$, R^3=H
168 X= CH_2, R^1=Bu^t, R^2=$CH_2C(O)NH(CH_2)_7Me$, R^3=H
169 X= CH_2, R^1=Bu^t, R^2=4-CH_2Pyr, R^3=H
170 X= CH_2, R^1=Bu^t, R^2=$CH_2C_6F_5$, R^3=H
171 X= CH_2, R^1=Bu^t, R^2=$CH_2C(O)OEt$, R^3=H
172 X= CH_2, R^1=Bu^t, R^2=$CH_2C(O)OEt$, R^3= $CH_2C(O)OEt$
173 X= CH_2, R^1=Bu^t, R^2=$CH_2C(O)Me$, R^3= $CH_2C(O)Me$
174 X= S, R^1=Bu^t, R^2=$CH_2C(O)OEt$, R^3= $CH_2C(O)OEt$

175 R=H
176a R=$CH_2COOC_2H_5$
176b R=$CH_2CONHCH(CH_2C_6H_5)COOCH_3$
176c R=$CH_2CONHCH(C_6H_5)COOCH_3$
176d R=$CH_2COOCH(CH_3)C_6H_5$

Figure 12. The structure of compounds 157-176.

The complex structure has been confirmed by ^{13}C NMR, two-dimensional ^{1}H-^{13}C, ^{1}H-^{1}H COSY and [1]H-^{1}H NOESY spectra. In IR spectrum of the complex in CCl_4 the valence vibration stretch band ν_{CO} at 1725 cm[-1] was observed. It indicates carboxyl groups linked by hydrogen bonding. Complexation was not observed in polar solvents destroying hydrogen bonds, e.g., deutero acetone and the mixture $CDCl_3/CD_3OD$ (9:1). The stability constant for 161 with the glutaric acid was higher by two orders of magnitude that those of pimelic, malonic, valeric acids and monomethyl ester of glutaric acid. Based on the relationships of the energy of complexes $\Delta G°$ 161/162 < 2$\Delta G°$ 161/163 or 2$\Delta G°$ 161/164, the authors supposed that an ideal two-point interaction was more efficient than two independent single-point interactions.

Complexation of dicarboxylates with tetrakis(phenylimidine) salt 165 based on calix[4]resorcinarene was studied by Sebo research group [Sebo et al. 2000]. ^{1}H NMR

titration gave the stoichiometry 1:2 for the complex of ammonium 5-nitrosophthalate in methanol and water. The interaction of substrate with the upper rim of cavitand in methanol is realized. However, one of the two isophthalate molecules is coordinated in water in such a manner that hydrophobic aromatic ring is included into the receptor cavity [Casnati et al. 1996].

Further investigations demonstrated that design of the molecules able to bind carboxylic acids require two- or three-point interactions. In subsequent investigations, calix[4]arenes distributed at the lower rim were synthesized and examined. Membrane extraction and UV spectroscopy proved complexation of a wide range of amino and carboxylic acids by the receptors obtained [Antipin et al. 2001; Stoikov et al. 2004]. Their investigations made it possible to propose two models of molecular recognition of dicarboxylic, α-hydroxy, and α-amino acids by 1,3-disubstituted calix[4]arenes.

The schemes proposed mainly differ in orientation of a substrate against a host molecule, namely, i.e., "docking" and "tweezers" type orientation. In first model, an acid is recognized simultaneously by the phenolic hydroxyls of the macrocycle and functional groups of the lateral substituents (R). In the second scheme, the R substituents at the lower rim of calix[4]arene function as "tweezers" or a two-handed podand performing the ditopic interaction of the host binding site with both functions of a guest. Binding of the first or second type is mainly determined by the size and strength of acids.

As demonstrated by UV and ^{1}H NMR spectroscopy, chiral calix[4]arenes 176 can effectively bind hydrochlorides of α-amino acids esters and their Z-carboxylates (Z = PhCH$_2$OCO-) [Okada et al. 1995]. These compounds have similar efficiency and selectivity in transport of substrates through dichloromethane membrane containing receptors 176. The hydrochlorides of α-amino acid esters are transported more slowly than their Z-carboxylates, whereas the latter ones are extracted weaker. The authors suppose that transport is limited in this case by re-extraction. Enantioselectivity of membrane transport induced by macrocycles 176 increases in the range 176a<176d≤176b<176c . Macrocycles 176b-176d are more efficient in extraction of L-isomers of "guests" with the enantiomeric excess from 11.0 to 73.1%.

Upper rim substituted calix[4]arenes with one (177a, 177b) or two (178a, 178b) urea and thiourea units fixed by four propoxy-groups in *cone* conformation were synthesized (Figure 13). Receptors 177a and 177b formed stabile complexes with aromatic carboxylate anions, e.g. benzoate (K$_{as}$ = 118 and 170 M^{-1}, respectively) and propionate anions (K$_{as}$ = 128 and 339 M^{-1}). Complexation can be explained by hydrogen bonding and π/π or CH$_3$/π host-guest interactions. The compound 178a efficiently interacted with acetate anion with formation of four hydrogen bonds between the substrate and the two urea groups (K$_{as}$ = 2.2·10^3 M^{-1}) [Casnati et al. 1996].

The NMR and mass spectroscopy confirmed that *p-tert*-butylcalix[6]arenes 179a and 179b symmetrically functionalized in 1,3,5-positions at the lower rim by three butylurea and butylthiourea fragments, could successfully bind various anions, e.g., 1,3,5 -, 1,2,4- and 1,2,3-benzenetricarboxylates [Scheerder et al. 1995] (Figure 13). It was found that the receptor 179a formed more stable complexes with 1,2,4- and 1,2,3-benzeneltricarboxylate anions (K$_{as}$ = 2.3·10^4 and 4.7·10^4 M^{-1}, respectively) whereas the compound 179b bonded more effectively 1,3,5-three substituted guest with K$_{as}$ = 2.9·10^5 M^{-1}.
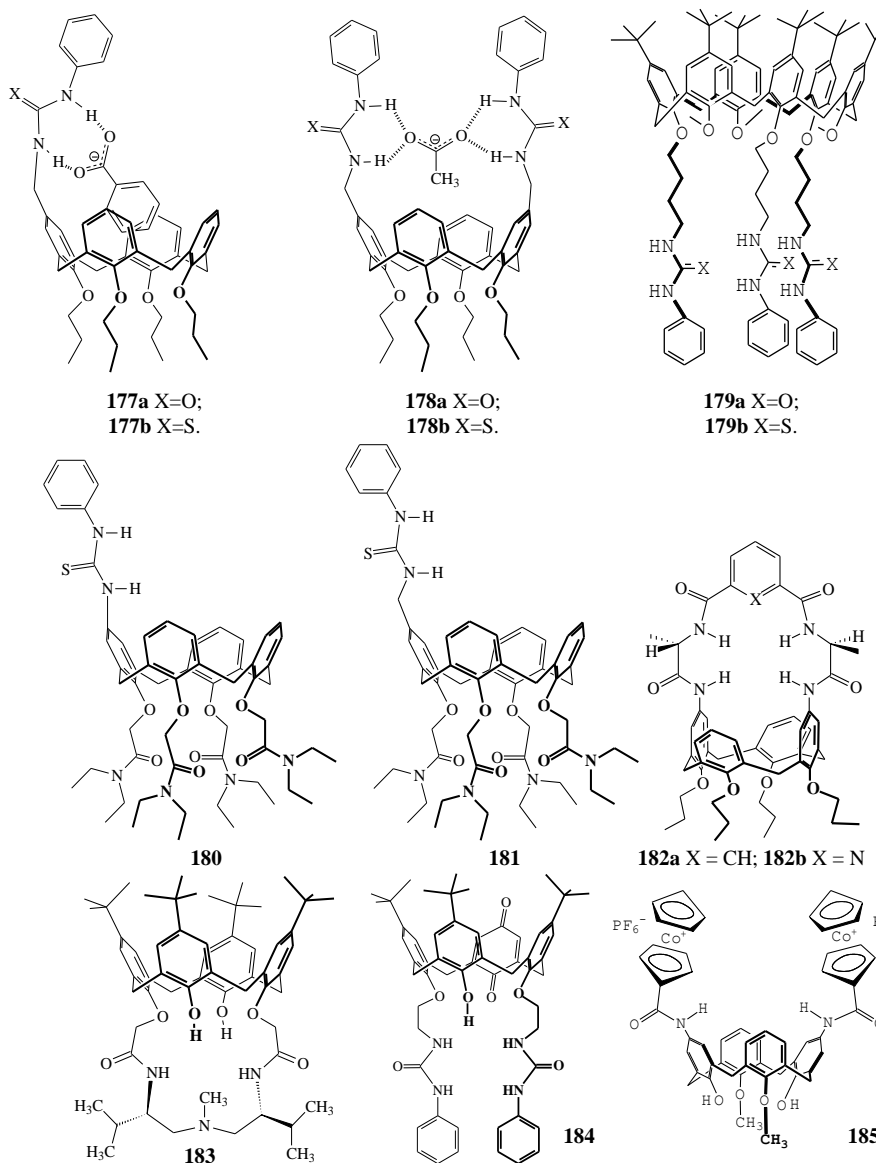
Figure 13. The structure of compounds 177-185.

Heterotopic receptors on calixarene platform 180 and 181 are able to dissolve sodium salts of monocarboxylic acids in chloroform [Pelizzi et al. 1998]. Four amide groups at the lower rim can bind the cation, while the interaction with the anion is realized by one urea or thiourea groups at the upper rim of the macrocycle. Contrary to 181, the affinity of 180 to anion depends from on the binding of the sodium cation. Probably this could be due to conformational changes of a host during the encapsulation of cation by substituents at the lower rim of calix[4]arene. In the case of thiourea group (compound 180) directly attached to the macrocycle, significant enhancement of anion binding was shown.

Interaction of receptors 182a and 182b containing two L-alanine and phtaloyl fragments or 2,6-diacyl pyridinic bridge at the upper rim of calix[4]arene in *cone* conformation with the

carboxylate anions is realized because of amide NH-groups serving as hydrogen bond donors [Sansone et al. 2002]. For all corresponding carboxylic acids, no evidences of such interactions were obtained from ^{1}H NMR spectra. The receptor 182b showed the ability to selectively and effectively bind carboxylate anions in such donor solvent as acetone (K_{as} 10500, 40100 and 33300 M^{-1} for acetate, benzoate and p- methoxybenzoate, respectively). According to the authors, this macrocycle could preferentially bind amino acids containing phenyl substituent because of cooperative stabilization of the complex by stacking interactions of the aromatic ring of guest and pyridinic or phenolic fragment of a host. Although both ligand (182a and 182b) are chiral, minor preference of d-stereoisomers of guests was observed only in some cases.

It was demonstrating by UV spectroscopy that the calix[4](aza)crown ethers with L-valine fragments 183 can interact with D- and L-tartaric acid. Complexation could be based on hydrogen bonds and OH-π interactions [He et al. 2002]. As defined by ^{1}H NMR titration, the receptor 183 formed complexes of 1:1 stoichiometry with D-and L-forms of guest (stability constants equal to $(9.83 \pm 0.43)\ 10^3\ M^{-1}$ and $(5.04 \pm 0.28)\ 10^3\ M^{-1}$, respectively).

In another case, complexation between calixarene 184 and acetate anion in acetonitrile was studied [Jeong et al.1999]. The addition of acetate or other inorganic anions, the intramolecular hydrogen bond between oxygen atom of quinone fragment and the proton of calixarene hydroxyl groups was broken. Also, blue shift in the absorption bands of UV spectrum was observed. The shift depended on the basicity of anions ($CH_3COO^- > H_2PO_4^- > Cl^- > Br^- > HSO_4^- > I^- > ClO_4^-$). Calix[4]arene 185 with two cobaltocenium fragments at the upper rim was proposed as receptor for various anions, e.g. adipinic acid dianion [Beer 1996].

The comparison of calix[4]diquinone containing receptors 186 and 187 with the calix[4]arene receptors 188 and 189 has shown that the first compounds bounded acetate anion more strongly (K_{as} 9990 and 1790 M^{-1} for 186 and 187; 4060 and 760 M^{-1} for 188 and 189, respectively) [Beer et al. 1999] (Figure 14). Moreover, electrostatic interaction played important role in this case. Uncharged complex Re (I) gave less stable complexes with anions than charged ruthenium-containing analogs.

In general, nature of macrocyclic platform substituents is one of the most important factors for complexation. The introduction of additional functional group or replacement of existed one can lead to dramatic changes in receptor properties of synthetic molecules. Thus, in this work, the molecular design of the receptor structures was performed for the recognition of biologically significant acids. For this purpose, *p-tert*-butyl calix[4]arenes 190–197 substituted at the lower rim were synthesized and their ability to efficient and selective transport of dicarboxylic, amino and α-hydroxy acids through a lipophilic membrane was investigated.

For the investigation of the transport characteristics, nature of substituents at the lower rim of calixarene macrocycle and the acidity of phenolic protons were varied in compounds 190–197. The relationships established in the experiments made it possible to directionally alter the receptor characteristics by variation of the substituents. Thus, introduction of nitro groups at the upper rim of calix[4]arene resulted in novel synthetic receptor for glutamic acid [Stoikov et al. 2009].

Charged sulfur-containing groups are also actively used in design of receptors for amino acids. The transport of tryptophan and phenylalanine from the acidic to the alkaline media against their concentration gradient was performed by negatively charged dinonyl sulfonate [Behr et al. 1973].

190 $R^1=Bu^t$, $R^2= CH_2C_6H_5$
191 $R^1=Bu^t$, $R^2=CH_2(CH_2)_6CH_3$
192 $R^1=Bu^t$, $R^2= CH_2C_6F_5$
193 $R^1=Bu^t$, $R^2= CH_2COOEt$
194 $R^1=NO_2$, $R^2= CH_2C_6H_5$
195 $R^1=NO_2$, $R^2= CH_2(CH_2)_6CH_3$
196 $R^1=NO_2$, $R^2= CH_2C_6F_5$
197 $R^1=NO_2$, $R^2= CH_2COOEt$

198a R = H, R_1 = CH_2COOH
198b R = SO_3H, R_1 = CH_2COOH

199 n = 4; R = -H
200 n = 4; R = -CH_2COOH
201 n = 4; R = -CH_2CONH_2
202 n = 4; R = -$CH_2CH_2NH_2$

203 n = 6; R = -H
204 n = 6; R = -CH_2COOH
205 n = 6; R = -CH_2CONH_2
206 n = 6; R = -$CH_2CH_2NH_2$

207 n = 8; R = -H
208 n = 8; R = -CH_2COOH
209 n = 8; R = -CH_2CN
210 n = 8; R = -$CH_2CH_2NH_2$

Figure 14. The structure of compounds 186-210.

The mass transfer of α-amino acids, e.g., arginine and histidine, from aqueous phase (pH=3) through chloroform membrane with di(2-ethylhexyl)sulfosuccinate in the receiving phase (pH=10) was described [Takeshima et al 1994]. The attempt to transport guests with nonionic units in the side chain, e.g. leucine, tryptophan and phenylalanine, in these conditions was unsuccessful. The acids are able to be transferred in cationic form at pH=1 in giving phase.

Also it was shown that calix[4]arene with sulfonate groups can bind basic amino acids, e.g., arginine and lysine (pH = 5) by strong electrostatic interactions (Späth, König 2010). Separation of the mixture of 15 amino acids was achieved on the column containing *p*-H-37-(2-carboxymethyloxy)calix[6]arene 198a and *p*-sulfonato-37-(2-carboxy-methyloxy)calix[6]arene 198b [Kalchenko et al.2002]. More polar sulfonatocalix[6]arene binds amino acids worse than calix[6]arene 198a. With the increase of the side-chain length of aliphatic amino acids, the decrease of stability constant was observed. The authors explained this fact by steric reasons. The highest stability constants (6585 M^{-1} and 4091 the M^{-1} for compounds 198a and 198b, correspondingly) for asparagine, the only amino acid with carboxyl group in the side chain, were found. The pK_a value of carboxyl group in the side

chain of this acid is equal 3.65. This additionally stabilizes the complex due to formation of carboxylic acid dimer linked by two hydrogen bonds.

A series of water-soluble *p*-sulfonato derivates of calix[n]arenes 199-210 was synthesized as receptors for α-amino acids [Silva et al. 2003]. Complexation ability for these compounds with the 11 α-amino acids at pH=8 was studied by ^{1}H NMR titration. The presence of substitutes at the lower rim of calixarenes has a major influence for values of binding constant (from K_{as}=17 M^{-1} for complex 203 with glycine to K_{as}=10083 M^{-1} for complex 207 with arginine). Cationic amino acids lysine and arginine strongly associated with all the derivatives, while the interaction between other acids and compounds 199-210 depended on the nature of substrate side chain.

Organophosphorous compounds provide other prospects in design of receptors and carriers for carboxylic acids or their derivatives. Thus, the compounds based on α-amino phosphonates 211-219 were successfully used as carriers of α-hydroxy-, carboxylic and dicarboxylic acids through liquid lipophilic membrane [Stoikov et al. 2004] (Figure 15). Such receptors realize molecular recognition of oxalic acid among other structurally similar substrates. The stoichiometry of 2:1 was determined by isomolar series method in methanol. The efficiency and selectivity of molecular recognition of the substrates depended on the lipophilicity of receptor molecules and nature of the substituents at the α-C atom.

Implementation of similar fragments at the upper or lower rim of the macrocyclic platform can provide effective binding of acids, peptides, biogenic amines and some other molecules [Späth, König 2010]. Membrane transport of a number of hydroxy-, and carboxylic acids was performed with tri-*n*-octylphosphine oxide as carrier [Syen et al. 1994; Juang et al. 1997]. The effective extraction of lipophilic α-amino acids, e.g., phenylalanine, leucine, isoleucine and valine, from acidic phase was realized by trialkyl phosphites [Thien et al. 1988].

Calix[4]arene 220 tetrasubstituted at the upper rim by phosphonate groups showed complexation ability toward some amino acids. The following values of the binding constants were obtained in methanol: 7.9 ×10^2 M^{-1} for Ac-Lys-OMe (Lys, K_{as} = 3 ×10^3 M^{-1}) and 1.9 ×10^4 M^{-1} for Ts-Arg-OMe (Arg, K_{as} = 7.9 × 10^2 M^{-1}) [Späth, König 2010].

Receptors for 2,4-dichlorophenoxyacetic acid based on calixarenes substituted at the upper rim by phosphonic acid fragments were synthesized [Kalchenko et al. 2003]. A key role of hydrophobic effects was proposed for complexation of 2,4-dichlorophenoxyacetic acid with calixarenes 221-223. However, these effects are suppressed due to electrostatic repulsion exerted in water by partially dissociated carboxyl group of a guest and dihydroxyphosphoryl group of a host. The authors explained more higher value of the stability constants obtained for 2,4- dichlorophenoxyacetic acid with bis- aminophosphonic acids 223 (5077 M^{-1}) against diacid 221 by larger volume of calixarene 223 molecular cavity which contained two phosphoryl-*N*-tolyl aminomethyl derivatives at the upper rim of the macrocycle.

As indicated by ^{1}H NMR spectroscopy (DMSO-*d$_6$*), *p-tert*-butylcalix[4]arenes 222 and 223 1,3-disubstituted at lower rim with trifluoroacetyl benzyl fragment are more selective in binding acetate anion than halogenide and HSO$_4^-$ anions. The values of stability constants were 1200 and 5800 M^{-1}, respectively [Whang et al. 2003]. The consideration of the magnitude of the shifts of receptor proton signals with acetate anion addition concluded that the acetate anion was bonded by trifluoroacetyl groups.

Replacement of substituents methylene groups by the carbonyl units (226 and 227) led to the disappearance of their ability to bind acetate anion because of the loss of complementarity

of the binding center. The absence of volume *tert*-butyl groups at the upper rim of the macrocycles (224 and 226) has negative effect on the binding ability of the receptor. Similar data were obtained for the binding of alkali cations by calix[4]arenes containing ester fragments at the lower rim [Arnaud-Neu et al. 1989].

	R^1	R^2	R^3	R^4
211	$CH_2CH(Et)Bu$	H	H	CH_2Ph
212	$CH_2CH(Et)Bu$	Me	Me	CH_2Ph
213	$CH_2CH(Et)Bu$	$(CH_2)_4$	$(CH_2)_4$	CH_2Ph
214	$C_{10}H_{21}$	H	C_7H_{15}	$CH(Et)CH_2OH$
215	$C_{10}H_{21}$	H	C_9H_{19}	$CH(Et)CH_2OH$
216	$C_{10}H_{21}$	H	Ph	$CH(Et)CH_2OH$
217	C_8H_{17}	H	Ph	$CH(Et)CH_2OH$
218	C_8H_{17}	H	$4\text{-}MeOC_6H_4$	$CH(Et)CH_2OH$
219	$C_{10}H_{21}$	H	$2\text{-}CH_2=CHCH_2OC_6H_4$	$CH(Et)CH_2OH$
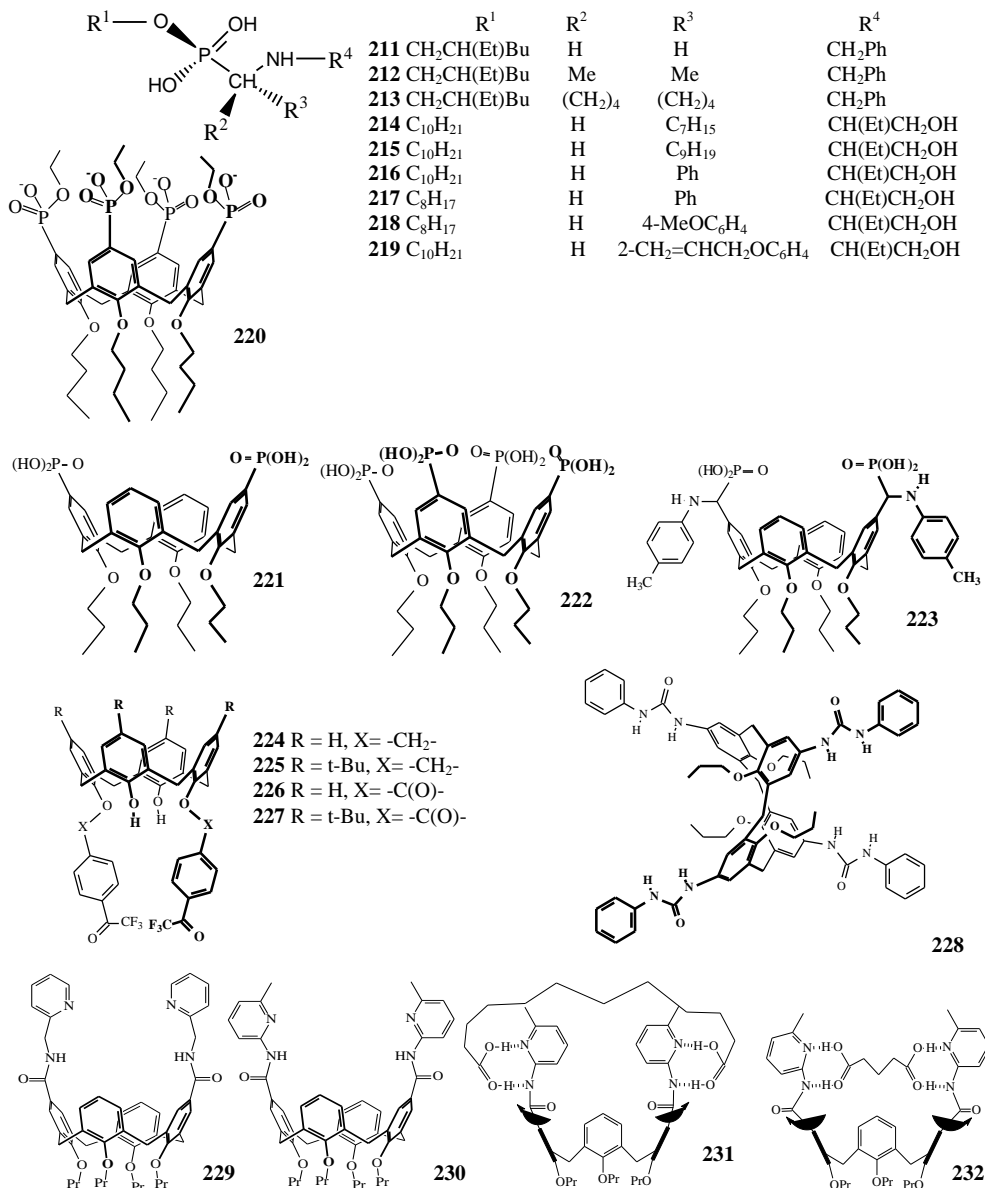
Figure 15. The structure of compounds 211-232.

Compound 228 based on calix [4] arene tetrasubstituted by urea fragments in *1,3-alternate* conformation was synthesized and showed strongly negative allosteric effect [Budka et al. 2001]. Conformational changes caused by the binding of the first guest molecule made impossible binding of the second molecule by substituents placed on the opposite side of the macrocyclic platform. As determined by ¹H NMR titration, the values of stability

constants of the complexes with acetate and benzoate anions were K_{as} = 2100 and 1800 M^{-1}, correspondingly. Most effective interaction was observed for benzoate anion and calix[4]arene 1,3-disubstituted at the upper rim. This could be due to possible inclusion of "guest" benzene ring in the cavity of the macrocycle. Substitution of urea units by amide groups in the compound 228 led to losses of complexation ability toward carboxylate anions.

As mentioned above, amidopyridine fragments can also serve as binding sites of the acids. For this reason, these functional groups are interesting for modeling of the receptor structures based on macrocyclic compounds. Positioned at the upper or lower rim of macrocycle, they determine receptor ability of calixarenes. Thus, the structures 229 and 230 functionalized with amidopyridine groups at the upper rim were synthesized as receptors for dicarboxylic acids [Miyaji et al. 2002]. Their complexation with a number of aliphatic and aromatic dicarboxylic acids was studied by ^{1}H NMR. The authors concluded that the binding ability of aliphatic diacid was mainly dependent on the length of hydrocarbon chain in the substrate (complexes 231, 232). For example, the highest value of binding constant (K_{as} 3000 M^{-1}, DMSO-d_6/CDCl$_3$) of the complex of 231 with dodecane dicarboxylic acid was observed. Meanwhile the acidity of substrate is mainly important for the stability of the complexes with aromatic acids. Thus, no π-π interactions between aromatic parts of acid and phenyl fragments of calixarene ring are possible for aromatic acids interacting with compound 230.

233 R = *o*-HNPy
234 R = *m*-HNPy
235 R = *p*-HNPy
235 R = *o*-HNCH$_2$Py
237 R = *m*-HNCH$_2$Py

238 R = *o*-HNPy
239 R = *m*-HNPy
240 R = *p*-HNPy
241 R = *o*-HNCH$_2$Py
242 R = *m*-HNCH$_2$Py
243 R = *p*-HNCH$_2$Py

244 R = *o*-HNPy
245 R = *m*-HNPy
246 R = *p*-HNPy
247 R = *o*-HNCH$_2$Py
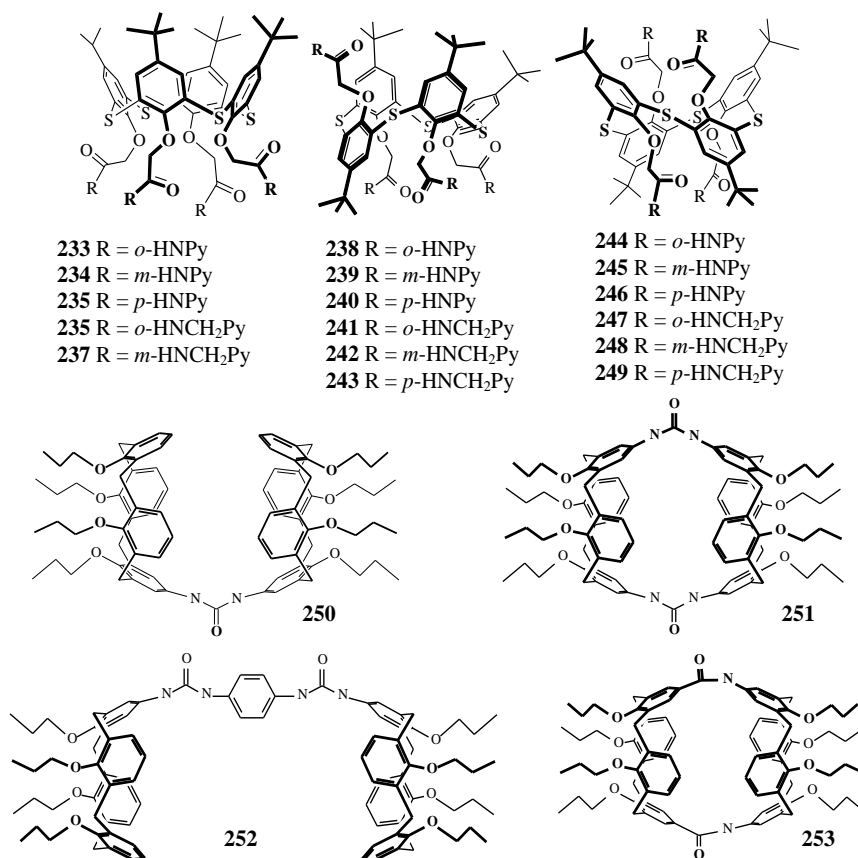**248** R = *m*-HNCH$_2$Py
249 R = *p*-HNCH$_2$Py

Figure 16. The structure of compounds 233-253.

Also, a series of new *p-tert*-butyl thiacalix[4]arenes with *o-*, *m-*, *p*-amido and *o-*, *m-*, *p*-(amidomethyl)pyridine substituents at the lower rim in *cone*, *partial cone*, and *1,3-alternate* conformations 233-249 were synthesized (Figure 16).

Their complexation with α-hydroxy (glycolic, tartaric) and dicarboxylic (oxalic, malonic, succinic, fumaric, and maleic) acids was investigated by UV–vis spectroscopy. The receptors demonstrated high ability to recognize the substrates studied. The efficiency and selectivity of interaction, the association constants $\lg K_{as}$ (10^2 to 10^7 M^{-1}) and the stoichiometry 1:1 were determined for the complexes of calix[4]arenes with the acids.

The receptors based on *p-tert*-butyl thiacalix[4]arenes with (amidomethyl)pyridine substitutes 235-237, 241-243, 247-249 are most efficient in complexation in many cases. It was shown that the complexation ability of the receptors crucially depends not only on the substituents and bridge fragments, but also on the size and configuration of calix[4]arene macrocycle [Stoikov et al. 2010].

The bismacrocyclic receptors 250-253 which selectively bind benzoate anion by one or two bridging urea fragments at the upper rim of calixarene were obtained [Stastny et al. 2002]. These compounds have pre-organized cavity able to bind anions. Stability constants of the complexes with benzoate anion (CDCl$_3$/DMSO-d_6, 4:1) varied in the range 252 > 251 > 250 (1280, 734 and 43 M^{-1}, correspondingly). Bis-calixarene 253 with amide units did not bind benzoate anion.

CONCLUSION

This review has highlighted the contribution of artificial receptor molecules in the chemistry of carboxylic acids and carboxylate anions sensing. Using numerous recent examples, we have demonstrated that synthetic receptors represent a very promising family of molecules with many potential applications in supramolecular chemistry, including molecular recognition. All the receptor molecules considered are very promising and can be combined with each other. This offers opportunities in development of new receptors with unusual properties. Because of their unique topology, properties, ability for pre-organization, conformational behavior, the employment of such molecules as building blocks/molecular scaffolds in design of novel selective receptros toward carboxylic acid and carboxylate anion seems to be very interesting. The topic reviewed here indicates some novel possibilities arising from the combination of the molecule skeleton with functional groups/moieties designed for complexation. It is evident that continuing progress in the chemistry of artificial receptors and related molecules could bring many fascinating discoveries in the near future, with many possible applications in the field of medicine, industry, for synthesis of nanoscale architectures and development of sensors.

ACKNOWLEDGMENTS

REFERENCES

Antipin, I. S., Stoikov, I. I., Khrustalev, A. A., Konovalov, A. I. (2001). Membrane extraction of organic compounds 3. A new receptor fragment for carboxylate group based of the calix[4]arene platform. *Russ.Chem.Bull., Int.Ed., 7,* 2134-2143.

Arnaud-Neu, F., Collins, E. M., Deasy, M., Ferguson, G., Harris, S. J., Kaitner, B., Lough, A. J., McKervey, M.A., Marques, E., Ruhl, B. L., Schwing-Weill, M. J. and Seward E. M. (1989). Synthesis, X-ray crystal structures, and cation-binding properties of alkyl calixaryl ester and ketones, a new family of macrocyclic molecular receptors. *J. Am. Chem. Soc.,111*, 8681-8691.

Asfari, Z., Bohmer, V., Harrowfield, J. and Vicens, J. (2001). Calixarenes 2001. *Netherlands: Kluwer Acad. Publishers*, 683.

Beer, P. D. (1996). Anion selective recognition and optical/electrochemical sensing by novel transition-metal reseptor systems. *Chem. Comm.*, 689-696.

Beer, P. D., Timoshenko, V., Maestri, M., Passaniti, P. and Balzani, V. (1999). Anion recognition and luminescent sensing by new ruthenium(II) and rhenium(I) bipyridyl calix[4]diquinone receptors. *Chem. Comm.*, 1755-1756.

Behr, J.-P. and Lehn, J.-M. (1973). Transport of amino acids through organic liquid membranes. *J. Am. Chem. Soc.*, *95*, 6108-6110.

Benito, J. M., Gómez-García, M., Blanco, J. L. J., Mellet, C. O. and Fernández, J. M. G. (2001). Carbohydrate-based receptors with multiple thiourea binding sites. Multipoint hydrogen bond recognition of dicarboxylates and monosaccharides. *J. Org. Chem.*, *66*, 1366-1372.

Bilz, A., Stork, T. and Helmchen, G. (1997). New chiral solvating agents for carboxylic acids: discrimination of enantiotopic nuclei and binding properties. *Terahedron Asymmetry*, *8*, 3999-4002.

Blondeau, P., Segura, M., Perez-Fernandez, R. and De Mendoza, J. (2007). Molecular recognition of oxoanions based on guanidinium receptors. *Chem. Soc. Rev.*, *36*, 198-210.

Boiocchi, M., Bonizzoni, M., Moletti, A., Pasini, D. and Taglietti, A. (2007). Linear recognition of dicarboxylates by ditopic macrocyclic complexes. *New J. Chem.*, *31*, 352-356.

Breslow, R., Rajagopalan, R. and Schwarz, J. (1981). Selective functionalization of doubly coordinated flexible chains. *J. Am. Chem. Soc.*, *103*, 2905-2907.

Budka, J., Lhoták, P., Michlová, V. and Stibor, I. (2001). Urea derivatives of calix[4]arene 1,3-alternate: an anion receptor with profound negative allosteric effect. *Tetrahedron Lett.*, 1583-1586.

Casnati, A., Fochi, M. and Minari, P. (1996). Upper-rim urea-derivatized calix[4]arenes as neutral receptors for monocarboxylate anions. *Gazz. Chim. Ital.*, *126*, 99-106.

Casnati, A., Sansone, F., Ungaro, R. (2003**).** Peptido- and glycocalixarenes: playing with hydrogen bonds around hydrophobic cavities. *Acc. Chem. Res*., *36*, 246-254.

Chahar, M., Upreti, Sh. and Pandey, P. S. (2006). Anion recognition by bisimidazolium and bisbenzimidazolium cholapods. *Tetrahedron*, *63*, 171-176.

Chakraborty, T. K., Tapadar, S. and Kumar, S.K. (2002). Cyclic trimer of 5-(aminomethyl).-2-furancarboxylic acid as a novel synthetic receptor for carboxylate recognition. *Tetrahedron Lett.*, *43*, 1317-1320.

Chen, H., Maestre, M. F. and Fish, R. H. (1995). Bioorganometallic chemistry. 5. Molecular recognition of aromatic amino acid guests by Cp*Rh-nucleobase/nucleoside/nucleotide cyclic trimer hosts in aqueous solution. *J. Am. Chem. Soc.*, *117*, 3631-3632.

Chen, H., Ogo, S. and Fish, R. H. (1996). Bioorganometallic chemistry. 8. The molecular recognition of aromatic and aliphatic amino acids and substituted aromatic and aliphatic carboxylic acid guests with supramolecular (eta(5)-pentamethylcyclopentadienyl) rhodium-nucleobase, nuceloside and nucleotide cyclic trimer hosts via non-covalent pi-pi and hydrophobic interactions in water: Steric, electronic, and conformational parameters. *J. Am. Chem. Soc.*, *118*, 4993-5001.

Clark, J. L., Peinado, J., Stezowski, J. J., Vold, R. L., Huang, Y. and Hoatson, G. L. (2006). Molecular recognition in cyclodextrin complexes of amino acid derivatives: the effects of kinetic energy on the molecular recognition of a pseudopeptide in a nonconstraining host environment as revealed by a temperature-dependent crystallographic study. *J. Phys. Chem. B.*, *110*, 26375-26387.

Consoli, G. M. L., Granata, G., Galante, E., Di Silvestro, I., Salafiab L., Geraci C. (2007). Synthesis of water-soluble nucleotide-calixarene conjugates and preliminary investigation of their in vitro DNA replication inhibitory activity. *Tetrahedron, 63,* 10758–10763.

Creaven, B. S., Donlon, D. F., McGinley, J. (2009). Coordination chemistry of calix[4]arene derivatives with lower rim functionalisation and their applications. *Coordination Chemistry Reviews, 253,* 893–962.

Costero, A. M., Gavina, P., Rodriguez-Muniz, G. M. and Gil, S. (2007). Relationship between ligand conformations and complexation properties in ditopic biphenyl thioureas. *Tetrahedron*, *63*, 7899-7905.

Cruz, C., Delgado, R., Drew, M.G.B. and Félix, V. (2004). Supramolecular aggregates between carboxylate anions and an octaaza macrocyclic receptor. *Org. Biomol. Chem.*, **2**, 2911-2918.

Cudic, P., Vigneron, J. P., Lehn, J.-M., Cesario, M. and Prange, T. (1999). Molecular recognition of azobenzene dicarboxylates by acridine-based receptor molecules; crystal structure of the supramolecular inclusion complex of trans-3,3-azobenzene dicarboxylate with a cyclo-bis-intercaland receptor. *Eur. J. Org. Chem.*, 2479-2487.

Davis, A. P., Lawless, L. J., Blackburn, A. G., Ayling, A. J., Perez-Payan, M.N. and Davis, A.P. (2001). Steroidal guanidines as enantioselective receptors for *N*-acyl-amino acids. Part 1. 3-Guanylated carbamates derived from cholic acid. *J. Chem. Soc., Perkin Trans. 1*, 1329-1335.

De la Torre, F., Campos, E. G., Gonzalez, S., Moran, J. R. and Caballero, C. (2001). Binding properties of an abiotic receptor for complexing carboxylates of α-heterocyclic and α-keto acids. *Tetrahedron*, *38*, 3945-3950.

De Mendoza, J., Alcazar, V. and Botana, E. (1997). Total synthesis of non-natural compounds for molecular recognition. The double challenge. *Pure Appl.Chem.*, *69*, 577-582.

Dhaenens, M., Lehn, J.-M. and Vigneron, J. P. (1993). Molecular recognition nucleosides, nucleotides and anionic planar substrates by a water-soluble bis-intercaland-type receptor molecule. *J. Chem. Soc., Perkin Trans.2*, 1379-1381.

Dietrich, B., Fyles, T. M. and Lehn, J.-M. (1978). Anion receptor molecules. Synthesis and some anion binding properties of macrocyclic guanidinium salts. *J. Chem. Soc., Chem. Commun.*, *21*, 934-936.

Dietrich, B., Hosseini, M. W. and Sessions, R. B. (1981). Anion receptor molecules. Synthesis and anion-binding properties of polyammonium macrocycles. *J. Am. Chem. Soc.*, *103*, 1282-1283.

Durmaz, M., Alpaydin, S., Sirit, A. and Yilmaz, M. (2007). Enantiomeric recognition of amino acid derivatives by chiral Schiff bases of calix[4]arene, *Tetrahedron: Asymmetry*, *18*, 900-905.

Echavarren, A., Galan, A. and de Mendoza, J. (1988). Anion-receptor molecules: synthesis of chiral and functionalized binding subunit, a bicyclic guanidinium group derived from l- or d-asparagine. *Helv. Chim. Acta*, *71*, 685-693.

Echavarren, A., Galan, A. and Lehn, J.-M. (1989). Chiral recognition of aromatic carboxylate anions by optically active receptor containing a rigid guanidinium binding subunit. *J. Am. Chem. Soc.*, *111*, 4994-4995.

Erdemir, S., Bahadir, M., Yilmaz M. (2009). Extraction of carcinogenic aromatic amines from aqueous solution using calix[*n*]arene derivatives as carrier. *Journal of Hazardous Materials, 168,* 1170–1176.

Fan, E., Arman, S.V. and Kincaid, S. (1993). Molecular recognition: hydrogen-bonding receptors that function in highly competitive solvents. *J. Am. Chem. Soc.*, *115*, 369-370.

Fitzmaurice, R.J., Kyne, G.M., Douheret, D. and Kilburn, J.D. (2002). Synthetic receptor for carboxylic acids and carboxylates. *J. Chem. Soc., Perkin Trans. 1*, 841-864.

Flack, S. S. and Kilburn, J. D. (1992). Synthesis and binding properties of a peptide receptor. *Tetrahedron Lett.*, *36*, 3409-3412.

Galan, A., Andreu, D., Echavarren, A. M., Prados, P. and de Mendoza, J. (1992). A receptor for enantioselective recognition of phenylalanine and tryptophan under neutral conditions. *J.Am.Chem.Soc.*, *114*, 1511-1512.

Gamez, P., Mooibroek, T. J., Teat, S. J. and Reedijk J. (2007). Anion Binding Involving π - Acidic Heteroaromatic Rings. *Accounts of Chemical Research*, *40*, 435-444.

García-Tellado, F., Goswami, S., Chang, S.K., Geíb, S.J. and Hamolton, A.D. (1990). Molecular recognition: a remarkably simple receptor for selective complexation of dicarboxylic acids. *J. Am. Chem. Soc.*, *112*, 7393-7394.

Geib, S.J., Vicent, C., Fan, E. and Hamilton, A. D. (1993). A self-assembling, hydrogen-bonded helix. *Angew. Chem. Int. Ed. Engl.*, *32*, 119-121.

Ghosh, K., Sen, T. and Froehlich, R. (2007). A naphthyridine-based receptor for sensing citric acid. *Tetrahedron Lett.*, *48*, 2935-2938.

Gleich, A. and Schmidtchen, F. P. (1990). Kunstliche molekulare anion-wirte. Die synthese eines chiralen bicyclischen guanidinium-salzes als funktionalisierte ankergruppe fur oxo-anionen. *Chem. Ber.*, *123*, 907-915.

Goel, A., Brennan, N., Brady, N. and Kenny, P.T.M. (2007). Electrochemical recognition of anions by 1,1'-N,N'-ferrocenoylbisamino acid esters. *Biosensors Bioelectronics*, *22*, 2047-2050.

Gonzalez-Alvarez, A., Alfonso, I., Diaz, P., Garcia-Espana, E., Gotor-Fernandez, V. and Gotor, V. (2008). A Simple Helical Macrocyclic Polyazapyridinophane as a Stereoselective Receptor of Biologically Important Dicarboxylates under Physiological Conditions. *J. Org. Chem.*, *73*, 374-382.

Goodnow, T. T. Reddington, M. V., Stoddart, J. F. and Kaifer, A. E. (1991). Cyclobis (paraquat-*p*-phenylene).: a novel synthetic receptor for amino acids with electron-rich aromatic moieties. *J. Am. Chem. Soc.*, *113*, 4335-4337.

Goswami, S., Ghosh, K. and Dasgupta, S. (1996). Molecular recognition: connection and disconnection of hydrogen bonds, a case study with dimeric and highly associated monocarboxylic acids with simple receptors *Tetrahedron*, *52*, 12223-12232.

Goswami, S., Ghosh, K. and Halder, M. (1999). Molecular recognition: hydrogen bonding induced configurational locking of a new photoresponsive receptor by dicarboxylic acids. *Tetrahedron Lett.*, *40*, 1735-1738.

Goswami, Sh. and Ghosh, K. (1997). Molecular recognition: chain length selectivity studies of dicarboxylic acids by the cavity of a new Troger's base receptor. *Tetrahedron Lett.*, *38*, 4503-4506.

Goswami, Sh., Ghosh, K., Dasgupta, S. (2000). Troger's base molecular scaffolds in dicarboxylic acid recognition. *J. Org. Chem.*, *65*, 1907-1914.

Gunnlaugsson, T., Davis, A. P., O'Brien, J. E. and Glynn, M. (2005). Synthesis and photophysical evaluation of charge neutral thiourea or urea based fluorescent PET sensors for bis-carboxylates and pyrophosphate. *Org. Biomol. Chem.*, *3*, 48-56.

Gunnlaugsson, Th., Harte, A. J., Leonard, J. P. and Nieuwenhuyzenb, M. (2002). Delayed lanthanide luminescence sensing of aromatic carboxylates using heptadentate triamide Tb(III). cyclen complexes: the recognition of salicylic acid in water. *Chem. Comm.*, 2134-2135.

Hamilton, A. D., Fan, E. and Arman, S. V. (1993). Molecular recognition. Design of new receptors for complexation and catalysis. *Supramol .Chem.*, *1*, 247-252.

He, Y., Xiao, Y., Meng, L., Zeng, Z., Wu, X. and Wu, C. T. (2002). New type chiral calix[4](aza)crowns: synthesis and chiral recognition. *Tetrahedron Lett.*, *43*, 6249-6253.

Hendrickson, K., Easton, C. J. and Lincoln, S. F. (1995). Cyclodextrin and termethylated cyclodextrin complexation of aromatic carboxylic acids and their conjugate bases in aqueous solution. The effect of size, hydrophobicity and charge. *Aust. J. Chem.*, *48*, 1125-1132.

Hernandez, J. V., Muniz, F. M., Oliva, A. I., Simon, L., Perez, E. and Moran, J. R. (2003). A xanthone-based neutral receptor for zwitterionic amino acids. *Tetrahedron Lett.*, *44*, 6983-6985.

Hinzen, B., Seiler, P. and Diederich, F. (1996). Mimicking the vancomycin carboxylate binding site: synthetic receptors for sulfonates, carboxylates, and N-protected alpha-amino acids in water. *Helv. Chim. Acta*, *79*, 942-960.

Hosseini, M. W. and Lehn, J.-M. (1982). Anion receptor molecules. Chain length dependent selective binding of organic and biological dicarboxylate anion by ditopic polyammonium macrocycles. *J. Am. Chem. Soc.*, *104*, 3525-3527.

Huggins, M. T., Musto, Ch., Munro, L. and Catalano, V. J. (2007). Molecular recognition studies with a simple dipyrrinone. *Tetrahedron*, *63*, 12994-12999.

Jazwinski, J., Blacker, J. and Lehn, J.-M. (1981). Cyclo-bisintercalands: synthesis and structure of an intercalative inclusion complex, and anion binding properties. *Tetrahedron Lett.*, *28*, 6057-6060.

Jeong, H., Choi, E. M., Kang, S. O., Nam, K.Ch. and Jeon, S. (1999). Electrochemistry and anion binding of urea functionalized calix[4]monoquinone. *Bull. Korean Chem. Soc.*, *20*, 1232-1234.

Jeong, K.-S., Park, J. W. and Cho, Y. L. (1996). Molecular recognition of dicarboxylate ions by bis-phenylureas derived from a new dicarboxylic acid. *Tetrahedron Lett.*, *37*, 2795-2798.

Juang, R. S. and Chen, L. J. (1997). Transport of citric acid across a supported liquid membrane containing various salts of a tertiary amine. *J. Memb. Sci.*, *123*, 81-87.

Kalchenko, O. I., da Silva, E. and Coleman A. W. (2002). Determination of the stability constants of inclusion complexes of p-H-37-(2-carboxymethyloxy)calix[6]arene and p-sulphonato-37-(2-carboxymethyloxy)-calix[6]arene with 15 amino acids by RP-HPLC. *J. Incl. Phenom. Macrocycl. Chem.*, *43*, 305-310.

Kalchenko. O. I., Solovyov, A. V., Cherenok, S.A., Starodub, N. F. and Kalchenko, V. I. (2003). Complexation of calix[4]arenephosphonous acids with 2,4-dichlorophenoxyacetic acid and atrazine in water. *J. Incl. Phenom. Macrocycl. Chem.*, *46*, 19-25.

Katayev, E. A., Boev, N. V., Khrustalev, V. N., Ustynyuk, Yu. A., Tananaev, I. G. and Sessler, J. L. (2007). Bipyrrole- and Dipyrromethane-Based Amido-imine Hybrid Macrocycles. New Receptors for Oxoanions. *J. Org. Chem.*, *72*, 2886-2896.

Kato, R., Nishizawa, S., Hayashita, T. and Teramae, N. (2001). A thiourea-based chromoionophore for selective binding and sensing of acetate. *Tetrahedron Lett.*, *42*, 5053-5056.

Kimura, E., Sakonaka, A. and Yatsunami, T. (1981). Macromonocyclic polyamines as specific receptors for tricarboxylate-cycle anions. *J. Am. Chem. Soc.*, *103*, 3041-3045.

Kraft, A. and Frohlich, R. (1998). Star-branched non-covalent complexes between carboxylic acids and a tris(imidazoline). base. *Chem. Commun.*, 1085-1092.

Kryatova, O. P., Kolchinski, A. G. and Rybak-Akimova, E. V. (2002). Molecular tweezers for dicarboxylic acids based on a saddle-shaped metallomacrocyclic platform. *J. Incl. Phen. Macroc. Chem.*, *42*, 251-260.

Kubo, M., Nashimoto, E., Tokiyo, T., Morisaki, Y., Kodama M., Hioki H. (2006). Development of calixarene-based host molecules for peptides in aqueous media. *Tetrahedron Letters, 47,* 1927–1931.

Kurzmeier, H. and Schmidtchen, F.P. (1990). Abiotic anion receptor functions. A facile and dependable access to chiral guanidinium anchor groups. *J. Org. Chem.*, *55*, 3749-3755.

Lara, K.O., Godoy-Alcantar, C., Rivera, I.L., Eliseev, A.V. and Yatsimirsky, A.K. (2001). Complexation of dicarboxylates and phosphates by a semisynthetic alkaloid-based cyclophane in water. *J. Phys. Org. Chem.*, *14*, 453-459.

Lehn, J.-M. (1988). Supramolecular chemistry − scope and perspectives molecules, supramolecules, and molecular devices (Nobel lecture). *Angew. Chem. Int. Ed. Eng.*, *27*, 89-112.

Li, W. Y., Li, H., Zhang G. M., Chao, J. B.,. Ling, L. X, Shuang, S. M., Donga C. (2008). Interaction of water-soluble calix[4]arene with l-tryptophan studied by fluorescence spectroscopy. *Journal of Photochemistry and Photobiology A: Chemistry, 197,* 389–393.

Lilienthal, N. D., Enlow, M. A. and Othman, L. (1996). N,N'-dimethyl-2,7-diazapyrenium: a redox-dependent receptor for aromatic carboxylates. *J. Electroanal. Chem.*, *414*, 107-114.

Liu, X.-L. and Zhao, Zh.-G. (2007). Design and synthesis of novel tweezer anion receptors based on deoxycholic acid and molecular recognition properties for anions. *Huaxue Yanjiu Yu Yingyong*, *19*, 670-674.

Liu, Y., Zhang, Y.-M., Sun, S.-X., Li, Y.-M. and Chen R.-T. (1997). Molecular recognition study of a supramolecular system. Part 4. Molecular recognition thermodynamics of amino acids by modified β-cyclodextrins. *J. Chem. Soc., Perkin Trans. 2*, 1609-1613.

Martell, A. E. and Motekaitis, R. J. (1988). Formation and degradation of an oxalato- and peroxo-bridged dicobalt BISDIEN dioxygen complex. Binuclear complexes as host for the activation of coordinated guests. *J. Am. Chem. Soc.*, *110*, 8059-8064.

Metzger, A. Gloe, K., Stephan, H. and Schmidchen, F. P. (1996). Molecular recognition and phase transfer of underivatized amino acids by a foldable artificial host. *J.Org.Chem.*, *61*, 2051-2055.

Metzger, A., Peschke, W. and Schmidtchen, F.P. (1995). A convenient access to chiral monofunctionalized bicyclic guanidinium receptor groups. *Synthesis-Stuttgart*, *5*, 566-570.

Miranda, C., Escarti, F., Lamarque, L., Yunta, J. R., Navarro, P., Garcia-Espana, E. and Jimeno, M.L. (2004). New 1*H*-pyrazole-containing polyamine receptors able to complex L-glutamate in water at physiological pH values. *J. Am. Chem. Soc.*, *126*, 823-833.

Miyaji, H., Dudic, M., Tucker, J., Prokes, I., Light, M., Hursthouse, M., Stibor, I. and Lhotak P. (2002). Bis(amidopyridine)-linked calix[4]arenes: a novel type of receptor for dicarboxylic acids. *Tetrahedron Lett.*, *43*, 873-878.

Moore, G., Levacher, V., Bourguignon, J. and Dupas, G. (2001). Synthesis of a heterocyclic receptor for carboxylic acids. *Tetrahedron Lett.*, *42*, 261-263.

Muller, G., Riede, J. and Schmidtchen, F.P. (1988). Host-guest binding of oxoanion to guanidinium anchor groups. *Angew. Chem. Int. Ed. Eng.*, *27*, 1516-1518.

Mussons, L., Raposo, C., de la Torre, F., Moran, J. R. and Caballero, C. (1999). Dibenz[c,h]acridine receptors for dibutylmalonic acid. Decarboxylative catalytic activity. *Tetrahedron*, *55*, 4077-4094.

Mutihac, L., Buschmannb, H. J., Diacuc, E. (2002). Calixarene derivatives as carriers in liquid membrane transport. *Desalination, 148,)* 253-256.

Okada, Y., Kasai, Y. and Nishimura, J. (1995). The selective extraction and transport of amino acids by calix[4]arene-derived esters. *Tetrahedron Lett.*, *36*, 555-558.

Oshima, T., Suetsugu A., Baba, Y., Shikaze Y., Ohto K., Inoue K. (2008). Liquid membrane transport of cytochrome *c* using a calix[6]arene carboxylic acid derivative as a carrier. *Journal of Membrane Science, 307,* 284–291.

Oton, F., Espinosa, A., Tarraga, A., Ramirez de Arellano, C. and Molina, P. (2007). [3.3]Ferrocenophanes with guanidine bridging units as multisignalling receptor molecules for selective recognition of anions, cations, and amino acids. *Chemistry--A European Journal*, *13*, 5742-5752.

Owens, L., Thilgen, C. and Deiderich, F. (1993). A new helicopodand: molecular recognition of dicarboxylic acids with diastereoselectivity. *Helv. Chim. Acta*, *76*, 2757-2774.

Palet, C., Munoz, M., Valiente, M., Cynkowski, T., Daunert, S. and Bachas, L. G. (2000). Selective membrane transport of dicarboxylic acids in their neutral form by a synthetic receptor containing amidopyridine groups. *Anal. Chim. Acta*, *41*, 4583-4586.

Pant, N. and Hamilton, A. D. (1988). Carboxylic acid complexation by a synthetic analogue of the "carboxylate-binding pocket" of vancomycin. *J. Am. Chem. Soc.*, *110*, 2002-2003.

Pelizzi, N., Casnati, A., Friggeri, A. and Ungaro, R. (1998). Synthesis and properties of new calixarene-based ditopic receptors for the simultaneous complexation of cations and carboxylate anions. *J. Chem. Soc., Perkin Trans. 2*, 1307-1311.

Pernia, G. J., Kilburn, J. D. and Rowley, M. (1995). A novel receptor for amino acid derivatives. *J. Chem. Soc., Chem. Commun.*, *3*, 305-306.

Poh, B. L. and Tan, C. M. (1994). Complexation of amino acids by cyclotetrachromotropylene in aqueous solution – importance of CH-pi and pi-pi interactions source. *Tetrahedron*, *50*, 3453-3462.

Potvin, P. G., Lehn, J.-M., Izzat, R. M. and Christensen, J.J. (1987). Design of cation and anion receptors, catalysts, carriers. Synthesis of macrocycles. *New York: Wiley-Interscience*, 167.

Prevot-Halter, I. and Weiss, J. (1998). Synthetic receptors for dicarboxylates and diammoniums: templated assembly of functionalized catechols around molybdenum and complexation studies. *New J. Chem.*, 869-874.

Prevot-Halter, I., Smith, T. J. and Weiss, J. (1997). Assembling organic receptors around transition metal templates: functionalized catechols and dioxomolybdenum(VI) for the recognition of dicarboxylic acids. *J. Org. Chem.*, 2186-2192.

Prohens, R., Tomas, S., Morey, J., Deya, P. M., Ballester, P. and Costa, A. (1998). Squaramido-based receptors: molecular recognition of carboxylate anions in highly competitive media. *Tetrahedron Lett.*, 39, 1063-1066.

Proni, G., Pescitelli, G., Huang, X., Quraishi, N. Q., Nakanishi, K. and Berova, N. (2002). Configurational assignment of a-chiral carboxylic acids by complexation to dimeric Zn-porphyrin: host-guest structure, chiral recognition and circular dichroism. *Chem. Comm.*, 1590-1591.

Ranganathan, D., Haridas, V. and Karle, I.L. (1998). Cystinophanes, a novel family of aromatic-bridged cystine cyclic peptides: synthesis, crystal structure, molecular recognition, and conformational studies. *J. Am. Chem. Soc.*, *120*, 2695-2702.

Raposo, C., Crego, M. and Mussons, M.L. (1994). Readily available chromenone receptors for carboxylates. *Tetrahedron Lett.*, *35*, 3409-3410.

Rebek, J. Jr. (1990). Molecular recognition with model systems. *Angew. Chem. Int. Ed. Eng.*,*29,* 245-255.

Rebek, J. Jr., Nemeth, D., Ballester, P. and Lin, F-T. (1987). Molecular recognition: size and shape specificity in the binding of dicarboxylic acids. *J. Am. Chem. Soc.*, *109*, 3474-3475.

Redman, P., Tecilla, and P. Tonellato, U. (1997). Chiral lipophilic ligands. 2. Rh-porphyrin complexation of amino acids in chloroform. *Tetrahedron*, *50*, 1220-1223.

Reetz, M. T., Huff, J., Rudolph, J., Tollner, K., Deege, A. and Goddard, R. (1994). Highly efficient transport of amino acids through liquid membranes via three-component supermolecules. *J. Am. Chem. Soc.*, *116*, 11588-11589.

Roussel, Ch., Roman, M., Andreoli, F., Del Rio, A., Faure, R. and Vanthuyne, N. (2006). Non-racemic atropisomeric (thio).ureas as neutral enantioselective anion receptors for

amino-acid derivatives: origin of smaller Kass with thiourea than urea derivatives. *Chirality*, *18*, 762-771.

Sansone, F., Baldini, L., Casnati, A., Lazzarotto, M., Ugozzoli, F. and Ungaro, R. (2002). Biomimetic macrocyclic receptors for carboxylate anion recognition based on C-linked peptidocalix[4]arenes. *Proceedings of the National Academy of Sciences*, *99*, 4842-4847.

Sato, W., Miyaji, H. and Sessler, J.L. (2000). Calix[4]pyrrole dimers bearing rigid spacers: towards the synthesis of cooperative anion binding agents. *Tetrahedron Lett.*, *41*, 6731-6735.

Scheerder, J., Engbersen, J. F. J. and Casnati, A. (1995). Complexation of halide anions and tricarboxylate anions by neutral urea-derivatized p-tert-butylcalix[6]arenes. *J. Org. Chem.*, *60*, 6448-6454.

Schmidtchen, F.P. and Berger, M. (1997). Artificial organic host molecules. *Chem. Rev.*, *97*, 1609-1646.

Schmidtchen, F. P. and Berger, M. (1997). Artificial organic host molecules. *Chem. Rev.*, *97*, 1609-1646.

Schmidtchen, F. P. (1980). Synthese symmetrisch substituierter bicyclischer guanidine. *Chem. Ber.*, *113*, 2175-2182.

Schmidtchen, F. P. (1981). Macrocyclishe quartare ammoniumsalze, II. Einschlubkom plexbildung mit anionen in losung. *Chem. Ber.*, *114*, 597-607.

Schmidtchen, F. P. (1986). Probing design a novel ditopic anion receptor. *J. Am. Chem. Soc.*, *108*, 8249-8255.

Schmidtchen, F. P. (1986). Tetrazac: a novel artifical receptor for binding ω-amino carboxylates. *Org. Chem.*, *51*, 5161-5168.

Schmidtchen, F. P. (1990). A novel synthesis of chiral guanidinium molecular hosts. *Tetrahedron Lett.*, *31*, 2269-2272.

Schmuck, C. (1999). Side chain selective binding of N-acetyl-amino acid carboxylates by a 2-(guanidiniocarbonyl).pyrrole receptor in aqueous solvents. *Chem. Commun.*, 843-846.

Scrimin, P., Tecilla, P. and Tonellato, U. (1995). Chiral lipophilic ligands. 2. Cu(II).-mediated transport of alpha-amino acids across a bulk chloroform membrane. *Tetrahedron*, *51*, 217-230.

Scrimin, P., Tonello, U. and Zanta, N. (1988). Cu(II).-mediated selective transport of α-amino acid across a bulk liquid membrane using a chiral lipophilic ligand as carrier. *Tetrahedron Lett.*, *29*, 4967-4970.

Sebo, L. and Diederich, F. (2000). Tetrakis(phenylamidinium)-substituted resorcin[4]arene receptors for the complexation of dicarboxylates and phosphates in protic solvents. *Helv. Chim. Acta*, *83*, 93-99.

Sebo, L., Schweizer, B. and Diederich, F. (2000). Cleft-type diamidinium receptors for dicarboxylate binding in protic solvents. *Helv.Chim.Acta*, *83*, 80-86.

Sell, C., Galan, A. and de Mendoza, J. (1995). Molecular recognition of organic acids and anions – receptor models for carboxylates, amino acids, and nucleotides. *Top. Curr. Chem.*, *175*, 101-132.

Sessler, J. L. and Andrievsky, A. (1996). Sapphirin-lasalocid conjugate: novel carrier for aromatic amino acid transport. *Chem. Commun.*, *10,* 1119-1120.

Sessler, J. L., D. An, Cho, W.-S. and Lynch, V. (2003). Calix[2]bipyrrole[2]furan and calix[2]bipyrrole[2]thiophene: new pyrrolic receptors exhibiting a preference for carboxylate anions. *J. Am. Chem. Soc*., *125*, 13646-13647.

Silva, E. D. and Coleman, A. W. (2003). Synthesis and complexation properties towards amino acids of mono-substituted p-sulphonato-calix[n]arenes. *Tetrahedron*, *59*, 7357-7364.

Sokoließ, T., Schönherr, J., Menyes, U., Roth, U., Jira, T. (2003). Characterization of calixarene- and resorcinarene-bonded stationary phases I. Hydrophobic interactions. *Journal of Chromatography A, 1021,* 71–82.

Späth, A., König, B. (2010). Molecular recognition of organic ammonium ions in solution using synthetic receptors. *Beilstein J. Org. Chem., 32,* 1-111.

Stastny, V., Lhotak, P., Michlova, V., Stibor, I. and Sykora, J. (2002). Novel biscalix[4]arene-based anion receptors. *Tetrahedron*, *58*, 7207-7211.

Steinke, J. H. G., Dunkin, I. R. and Sherrington, D. C. (1999). A simple polymerisable carboxylic acid receptor: 2-acrylamido pyridine. *Trends in analytical chemistry*, *18*, 159-163.

Stibor, I. (2005). Anion sensing. *Springer-Verlag Berlin Heidelberg,* 1-238.

Stoikov, I. I., Agafonova, M. N., Padnya, P. L., Zaikov, E. N., Antipin, I. S. (2009). New membrane carrier for glutamic acid based on *p-tert*-butylcalix[4]arene 1,3-disubstituted at the lower rim. *Mendeleev Commun.*, *19*, 163–164.

Stoikov, I. I., Zhukov, A. Yu., Agafonova, M. N., Sitdikov, R. R., Antipin, I. S., Konovalov, A. I. (2010). *p-tert*-Butyl thiacalix[4]arenes functionalized at the lower rim by *o*-, *m*-, *p*-amido and *o*-, *m*-, *p*-(amidomethyl)pyridine fragments as receptors for a-hydroxy- and dicarboxylic acids. *Tetrahedron*, *66,* 359–367.

Stoikov, I. I., Gafiullina, L. I., Ibragimova, D. Sh., Antipin, I. S., Konovalov A. I. (2004). Synthetic receptors based on calix[4]arene functionalized at the lower rim in molecular recognition of dicarboxylic, α-hydroxycarboxylic, and α-amino acids. *Russ.Chem.Bull., Int.Ed., 6*, 1172—1180.

Stoikov, I. I., Fitseva, N. A., Akhmetzyanova, L. R., Gafioullina, L. I., Antipin, I. S., Zheltukhin, V. F., Devyaterikova, A. I., Alrfonsov, V. A., Konovalov, A. I. (2004). Membrane transport of dicarboxylic and α_hydroxy carboxylic acids induced by α-amino phosphonates. *Russ.Chem.Bull., Int.Ed., 7,* 1517—1523.

Sun, X. H., Li, W., Xia, P. F., Luo, H.-B., Wei, Y., Wong, M. Sh., Cheng, Y.-K. and Shuang, Sh. (2007). Phenyl-calix[4]arene-based fluorescent sensors: Cooperative binding for carboxylates. *J. Org. Chem.*, *72*, 2419-2426.

Sunamoto, J., Iwamoto, K., Mohri, Y. and Kominato, T. (1982). Liposomal membranes. 13. Transport of amino acid across liposomal bilayers as mediated by a photoresponsive carrier. *J. Am. Chem. Soc.*, *104*, 5502-5504.

Syen, Y., Gronberg, L. and Jonsson, J. A. (1994). Experimental studies on the enrichment of carboxylic acids with tri-n-octylphosphine oxide as extractant in a supported liquid membrane. *Anal.Chim.Acta*, *293*, 31-39.

Tabushi, I., Kuroda, Y. and Mizutani, T. (1986). Artifical receptor for amino acids in water. Local environmental effect on polar recognition by 6-A-amino-6-B-carboxy- and 6-A-amino-6-B-carboxy-β-cyclodextrins. *J. Am. Chem. Soc.*, *108*, 4514-4518.

Tabushi, I., Shimizu, N. and Sugimoto, T. (1977). Cyclodextrin flexibly capped with metal ion. *J. Am. Chem. Soc.*, *99*, 7100-7102.

Takeshima, S. and Wada, S. (1994). Transport behavior of basic amino acids through an organic liquid membrane system. *Separ. Sci. Technol.*, *29*, 2117-2129.

Takeuchi, M., Imada, and T. Shinkai, S. (1998). A strong positive allosteric effect in the molecular recognition of dicarboxylic acids by a cerium(iv) bis[tetrakis(4-pyridyl)-porphyrinate] double decker. Angew. Chem. Int. Ed., 37, 2096-2099.

Tanaka, Y., Kato, Y. and Aoyama, Y. (1990). Two-point hydrogen-bonding interaction: a remarkable chain-length selectivity in the binding of dicarboxylic acids with resorcinol-aldehyde cyclotetramer as a multidentate host. *J. Am. Chem. Soc.*, *112*, 2807-2808.

Tejeda, A., Oliva, A. I., Simon, L., Grande, M., Caballero, C. and Moran, J. R. (2000). A macrocyclic receptor for the chiral recognition of hydroxycarboxylates. *Tetrahedron Lett.*, *41*, 4563-4570.

Thien, M. P., Hatton, T. A. and Wang D. J. C. (1988). Surfactant-mediated water transport in liquid emulsion membrane bio separation systems. *Biotechn. Bioeng.*, *32*, 604-615.

Tomankova, Z., Matejka, P., Sykora, D. and Kral, V. (2003). Interaction of oligopyrrole macrocycles with aromatic acids: spectroscopical, quantum chemical and chromatographic aspects. *Talanta*, *59*, 817-829.

Tsubaki, K., Tanaka, H., Morikawa, H. and Fuji, K. (2003). Synthesis and recognition of amino acids by binaphthyl-crown receptors. *Tetrahedron*, *59*, 3195-3199.

Tsukube, H., Shinoda, S. and J. Uenishi (1996). Efficient transport of aliphatic amino acids mediated by lanthanide complex carriers under neutral conditions. *Chem. Lett.*, *11*, 969-970.

Uppadine, L. H., Redman, J. E., Dent, S. W., Drew, M. G. B. and Beer, P. D. (2001). Ion pair cooperative binding of potassium salts by new rhenium bipyridine crown ether receptors. *Inorg. Chem.*, *40*, 2860-2869.

Vicent, C., Fan, E. and Hamilton, A. D. (1992). Molecular recognition: directed hydrogen bonding receptors for acylamino acid carboxylates. *Tetrahedron Lett.*, *33*, 4269-4272.

Vicent, C., Hirst, S. C. and Garcia-Tellado, F. (1991). Conformational selectivity in molecular recognition: the influence of artificial receptors on the cis-trans isomerization of acylprolines. *J. Am. Chem. Soc.*, *113*, 5466-5467.

Watanabe, S., Higashi, N., Kobayashi, M., Hamanaka, K., Takata, Y. and Yoshida, K. (2000). Stereoselective optical sensing of dicarboxylate anions by an induced-fit type Ru(II) receptor. *Tetrahedron Lett.*, *41*, 4583-4586.

Webb, T. H. and Wilcox, C. S. (1993). Enantioselective and diastereoselective molecular recognition of neutral molecules. *Chem. Soc. Rev.*, 383-395.

Whang, S. S., Ko, S. W., Oh, S. M., Cho, S. and Nam K. C. (2003). Highly carboxylate anion selective receptors containing trifluoroacetylbenzyl moieties at the lower rim of calix[4]arene. *Bull. Korean Chem. Soc.*, *24*, 165-166.

Yuan, Q., Xie, Zh., Fu, E., Wu, C. and Lei, J. (2006). Application of chiral selectors in molecular recognition. *Huaxue Yu Shengwu Gongcheng*, *23*, 1-3.

Chapter 2

NEXT GENERATION MOLECULAR IMPRINTED POLYMERS: EXAMPLES OF LIQUID CRYSTALLINE MATERIALS AND HYDROGELS FOR PROTEIN RECOGNITION

Mingotaud Anne-Françoise[1,2], Fitremann Juliette[1,2], Mauzac Monique[1,2], Rodriguez Vilches Seila[1,2], Doucet Jean-Baptiste[1,2,3], Séverac Childérick[3], Laurent Elisabeth[1,2], Binet Corinne[1,2], Marty Jean-Daniel[1,2], Palaprat Guillaume[1,2] and Weyland Marie[1,2]

[1]Université de Toulouse, Lab. des IMRCP, 118 Rte de Narbonne, 31062 Toulouse cedex 9, France
[2]CNRS, Lab. des IMRCP, UMR CNRS 5623, 118 Rte de Narbonne, 31062 Toulouse cedex 9, France
[3]LAAS-CNRS, Université de Toulouse, 7 Av. du Colonel Roche, F-31077 Toulouse, France
Université de Toulouse, UPS, INSA, INP, ISAE ; LAAS ; F-31077 Toulouse, France

ABSTRACT

Molecular imprinted polymers (MIPs) have been studied for a few decades. They enable the specific recognition of the molecule for which they have been prepared. They have been thoroughly studied in organic solvents and this showed that a good recognition was observed only when a crosslinking ratio above 70% was used. In this case, however, the capacity of the MIP was limited owing to a poor accessibility of the imprinted cavities. In the ongoing research on this subject, many teams assess the possibility of controlling the recognition process and that of using the MIPs in aqueous systems. We present here our experience in liquid crystal MIPs as tunable systems and hydrogel-MIPs for the recognition of proteins.

In the first case, the MIP is composed of a liquid crystal elastomer which is built around the template. The difference from regular MIPs lies in the percentage of crosslinker agent, which is in the 5-10% range. This low range becomes possible owing to replacing a large part of the chemical crosslinking by a physical one, coming from the interactions between liquid crystal moieties attached to the material. This brings up a recognition ability similar to the regular MIPs, but with an increased accessibility to the cavities. Thus, the mass capacity of this LC-MIP is tremendously increased. Furthermore, since liquid crystal elastomers exhibit an organized/disorganized transition temperature and have a shape memory capacity, the LC-MIP can be controlled with external parameters, such as temperature or solvents. By this method, different types of materials have been examined and are presented here: MIPs able to specifically interact with an enantiomer, catalytic MIPs acting as artificial enzymes or MIPs able to interact with pesticides.

The second part of the manuscript describes the state of the art as well as our preliminary experiments aiming at developing MIPs made of hydrogels which will be able to selectively recognize a protein in solution. The goal is to fix the hydrogel MIP to a detection device for a future application in diagnostics. Due to this, a severe constraint exists for process: temperature and pH limits, process in less than 30minutes, porosity slightly lower than the size of the protein. Indeed, since proteins are large molecules, recognition at the surface of the MIP is sought. Several monomer formulations have been studied and the technological problems have been examined.

INTRODUCTION

The molecular imprinting technique aims at mimicking the molecular recognition that exists in biological systems such as enzymes. This recognition is obtained in a synthetic material by creating cavities that are specific in shape, size and functionalities to the target molecule to be recognized. The creation of these specific cavities most often relies on three major steps (figure 1).
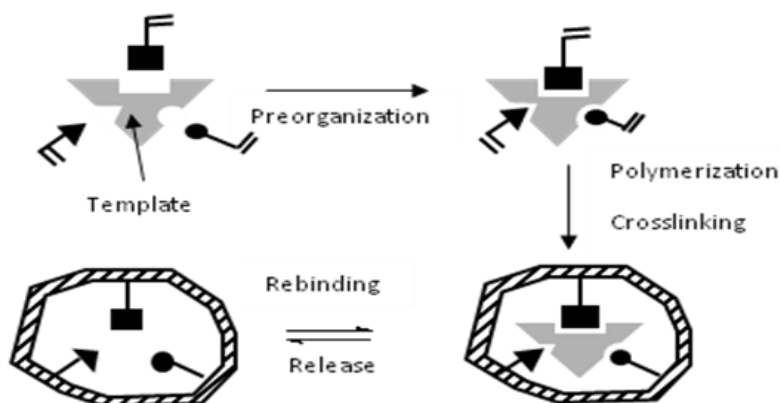

Figure 1. Molecular imprinting technique.

The first step consists in the pre-organisation phase, the template being complexed to an adequate molecule bearing polymerizable entities. Introduction of specific chemical groups in the molecular imprinted polymer (MIP) structure plays an important role in the recognition

process by functional affinities. Subsequently, polymerization and crosslinking are triggered to fix the material around the template. In a third step, the template is removed by washing steps. By this way, the material presents cavities that are complementary in shape and size to the template. Compared to enzymes, they are more robust and cheaper to produce. However, their main drawback remains lesser selectivities and specificities.

MIPs have been studied for a few decades and find their main applications in separation science, such as solid phase extraction (SPE) [1-3]. The stability and the possibility of using molecular imprinted materials in different environments have broadened their potential applications from separation to sensing [1,2,4] and catalysis. Also, a large variety of templating molecules has been studied such as pharmaceuticals [5-8] and pesticides [9-14]. If the chosen template is of biological importance, this class of molecular imprinted materials can be assessed as antibody analogues. Thus, drugs like theophylline (bronchodilator) or diazepam (tranquillizer) have been quantified in human serum using molecular imprinted polymers with results comparable to those obtained by a well established immunoassay method [15,16].

Compared to alternative techniques (involving biomolecules, abzymes, etc.), the molecular imprinting technique exhibits several advantages: low cost, effectiveness, mechanical, thermal and chemical stability of the support and long lifetime.

MIPs have been initially developed for recognition of small molecules, mainly in organic solvents. In such conditions, there is a strong need for a large amount of crosslinking agent, usually around 80 to 90mol%, to restrict distortion phenomena of the polymer backbones. This quantity is mandatory in order to obtain good recognition properties because the number of interacting sites between the template and the MIP is limited. Since the molecule is small, the interacting sites have to be precisely located in space. To achieve this, in regular MIPs, crosslinking is obtained through chemical bonds, which results in rigid networks. This leads to the desired recognition but limits template extraction and reinsertion, due to a poor accessibility.

In order to overcome this problem, several solutions have been proposed in the literature, including the synthesis of molecularly imprinted polymers as beads and films of fine-tuned porosity [17,18] or deposition of an MIP at the surface of already constituted membranes [19]. Based on our knowledge on liquid crystal elastomers, we decided to replace a part of this chemical crosslinking by a physical one owing to interactions between mesogens. The first part of this chapter describes our experience in this field for different problems: chiral recognition, catalytic MIPs acting as artificial enzymes or MIPs able to interact with pesticides.

Concerning the amount of crosslinker, the situation is quite different when MIPs are built around large molecules such as proteins. In such cases, the recognition sites are multiple and it is important that the MIP may adapt around the flexible molecule. In these cases, it has been found that good results can be obtained with lower crosslinking ratios. Furthermore, a crucial point when imprinting proteins is to avoid their premature denaturation. Thus, compatible methods of synthesis have to be employed, avoiding too high temperatures and most organic solvents.

The second part of this chapter deals with protein imprinted materials, presenting a critical overview of the present literature and, through our experience, enhancing the specific difficulties for developing such MIPs.

1. LIQUID CRYSTAL MIPS

The important parameter to obtain a good recognition process in MIPs is the maintaining of functional groups at essential locations. In order to do that, either physical or chemical crosslinking is needed. Physical crosslinking may be achieved by non covalent interactions, such as hydrogen bonds or Van der Waals interactions. Liquid crystals are typical examples of molecules which self organize by interacting with each other reversibly, depending either on the temperature for so called thermotropic liquid crystals or on the concentration for so called lyotropic liquid crystals. To obtain such properties, the molecules often exhibit two parts: one being very stiff such as aromatic rings and the other very soft such as aliphatic chains. The rigid core tends to organize to crystallize but a good crystallization is impeded by the presence of the soft aliphatic chains. This dual incompatible trend leads to the formation of mesophases.

The thermotropic liquid crystals may self-organize into nematic, smectic or cholesteric mesophases (figure 2). In nematic phases, the liquid crystal moieties are aligned following one direction only, whereas in smectic phases, a supplementary order exists and layers are formed. Cholesteric phases are formed when nematic mesogens exhibit a chiral center: this chirality may induce a torsion of the mesophase, leading to the formation of a helix.

To benefit from the possible self-assembly of the mesogens in MIPs, these have to be incorporated into the MIP matrix (figure 3) by synthesizing a liquid crystal elastomer around the template molecule leading to LC-MIPs. Liquid crystal elastomers present both chemical crosslinking together with physical crosslinking from the mesogens. The chemical crosslinking is needed to avoid sliding of the polymer chains but its ratio is very low (5-10%). An important point to benefit from the presence of mesogens is the necessity of crosslinking the material in the mesophase state, so that the self-organization gets fixed definitely. In such conditions, a shape memory effect is then obtained: the elastomer can be distorted either by addition of a solvent or temperature increase and it recovers the exact initial shape when the stimulus is stopped [20].
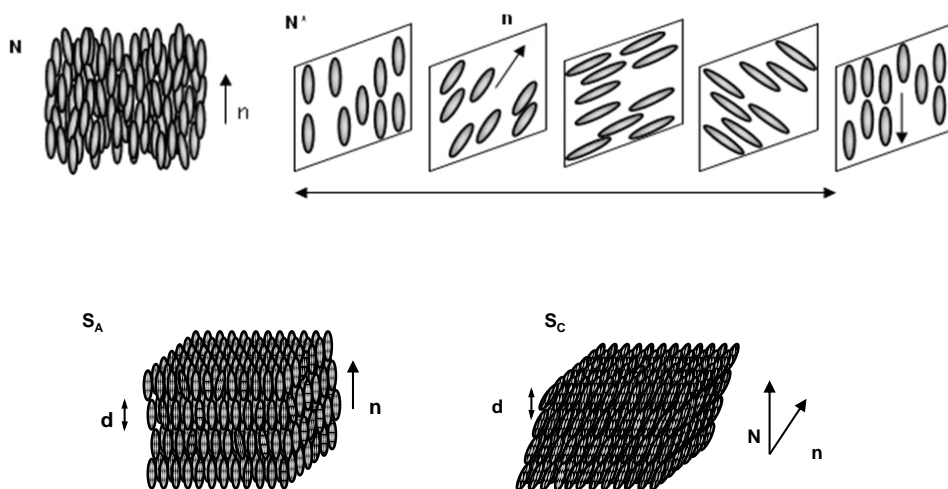

Figure 2. Examples of mesophases: nematic (N), cholesteric (N*), smectic A (S_A) and smectic C (S_C).
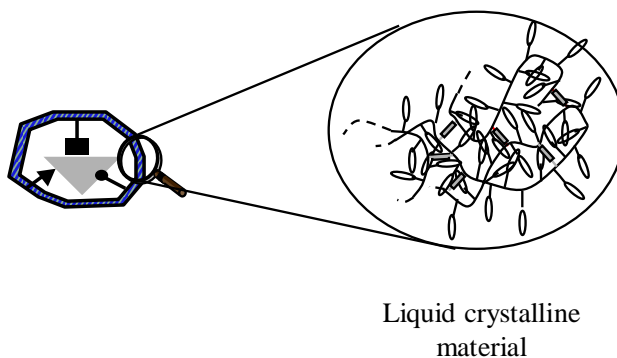
Liquid crystalline
material

Figure 3. Molecular imprinting technique using liquid crystalline materials.

In mesogenic polymers, it has been shown that the interactions between liquid crystalline side chain groups induced significant modifications of the mechanical behaviour when compared to 'classical' polymers. The side-chain mesogenic polymers appeared to behave like gels below their gelation point [21,22]. If crosslinks are introduced, the liquid crystalline networks display remarkable elastic properties [23-25] due to the mesomorphic order. For example, a transition from a poly-domain to a mono-domain can be induced by a strain field, which couples itself directly to the mesomorphic order. Thus, LC elastomers exhibit macroscopic morphology changes between the liquid crystalline and isotropic states. These changes include shape change and modification of the mechanical behaviour.

Different ways of processing LC-MIPs have been developed in our laboratory. The most common way for regular MIPs consists in a one step reaction polymerizing and crosslinking the system at the same time. In such a reaction, the interactions between the template and the functional monomers have to be already present at the beginning of the reaction and be stable during the process. In the case of LC-MIPs, beside this method, we have also developed a two-step technique leading to the synthesis of a liquid crystal polymer which is secondly crosslinked in the mesomorphous state. After having been washed and processed, the samples have been tested for molecular recognition or as specific catalysts.

In the following we will summarize the results obtained from three different studies in order to bring out some general behaviors of the LC-MIPs.

I.1. CHIRAL RECOGNITION IN LC-MIPS

If the template has a chiral structure, it can induce a helical structure inside a nematic liquid crystalline network, like a chiral dopant. Indeed it had been previously shown that a helical symmetry can be introduced by simply generating the liquid crystalline network in an oriented chiral solvent [26-29]. Y. Mao and M. Warner [30] discussed the 'robustness' of the imprinted phase chirality as a function of the helical twisting power of the dopant: the imprinting would be successful in the case of low helical twisting power or equivalently, a large pitch. Using this type of macroscopically oriented cholesteric elastomer, S. Courty and coll. [31] have presented a study of separation of chiral isomers. They have demonstrated the capacity of the material to preferentially absorb and retain the "correct" chirality molecule from a racemic solvent.
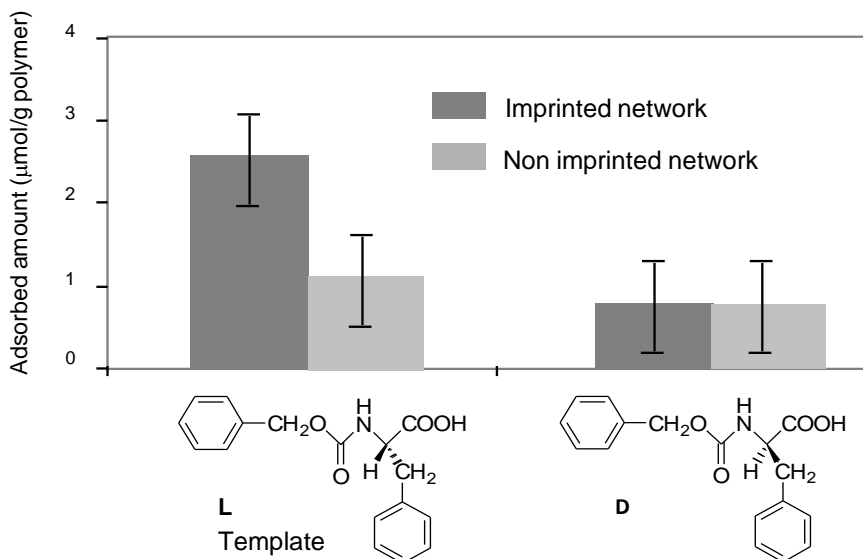
Figure 4. Absorbed amount of the L and D enantiomers of N-(carbobenzyloxy)-phenylalanine by the mesomorphous network imprinted around L-form compared to the non imprinted mesomorphous network. From Ref [34] with kind permission of the European Physical Journal.

Combining molecular imprinting and chiral induction, a cholesteric imprinted elastomer with a large helical pitch was obtained by crosslinking a nematic side chain polysiloxane around the chiral N-(carbobenzyloxy)-L-phenylalanine template [32]. The sample was macroscopically oriented during the synthesis. So, both a molecular chirality and a supramolecular phase chirality were topologically imprinted inside the network.

Batchwise adsorption tests, performed by putting the polymer sample in a solution containing the template or the other enantiomer, showed that the imprinted polymer had a pronounced stereo-selectivity towards the template enantiomer (Figure 4). The rebinding capacity (around 2.5 µmol/g of polymer) appeared to be greater than that of a non-imprinted mesogenic network (1 µmol/g of polymer) as well as that of an imprinted non mesogenic one (at best around 1 µmol/g of polymer [33]).

In such a system, the introduction of a chiral molecule during the synthesis of the MIP leads to two different levels of chirality. The first one is at the cavity level, exactly like in regular MIPs and the second one, at the mesomorphic structure level, is due to the presence of liquid crystal moieties. Therefore, in the system we were interested in, an enantiomer separation process could come either from the chiral cavity or from the chiral helix owing to the cholesteric phase. To analyze the contribution of each type of chirality, several MIPs were synthesized around (R) or (S)-methylbenzyl amine (figure 5).

The LC-MIP was synthesized in two steps. In a first one represented in figure 5, a linear polysiloxane was obtained by hydrosilylation. This polysiloxane was grafted with different entities: acidic aromatic groups which may interact with the template, the liquid crystal moiety and a benzophenone derivative. This last group was the initiator for the second step, which was a photocrosslinking under UV irradiation. This step was carried out at room temperature after having spread the polymer onto a solid support in order to obtain regular films with a thickness around 160µm.
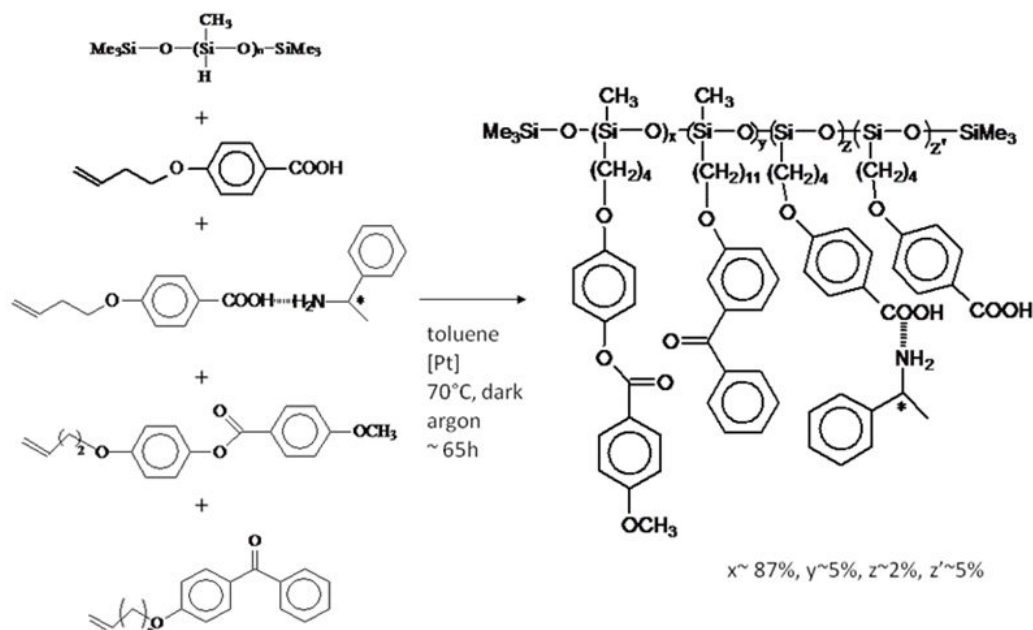
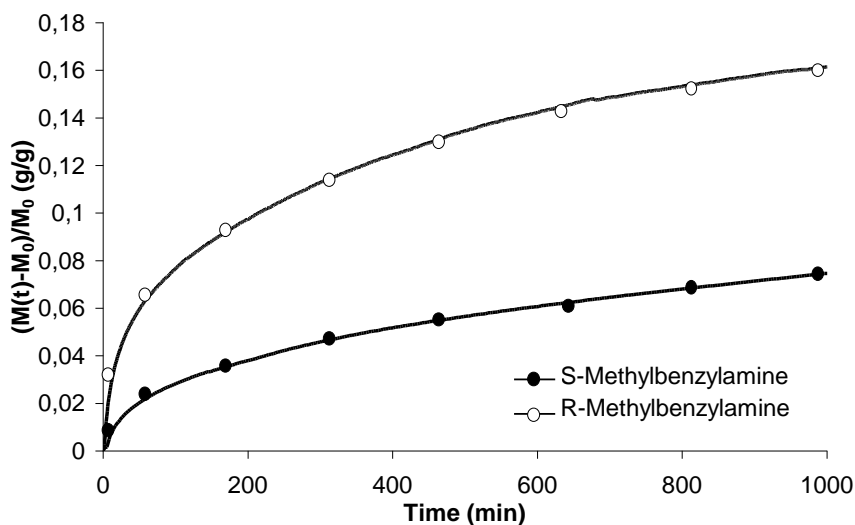Figure 5. Synthesis of liquid crystalline polysiloxane.

Figure 6. Absorption kinetics of (R)- and (S)-methylbenzylamine on a LC-MIP synthesized around (R)-methylbenzylamine.

Recognition properties for LC-MIPs for (R) or (S)-methylbenzylamine were furthermore characterized by using a microbalance technique [35,36]. Each sample was separately exposed to each enantiomer of the imprint molecule and the quantity of amine was recorded with time (figure 6). Non imprinted materials were also synthesized together with a

cholesteric material without any imprint (so called doped network). The results are summarized in figure 7. Figures 6 and 7 illustrate that a better recognition existed between the LC-MIP and its template compared to the opposite enantiomer.
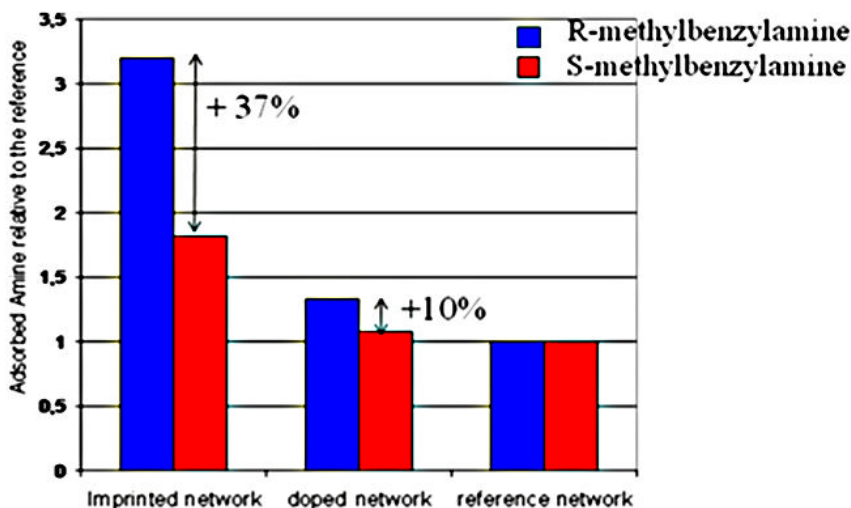

Figure 7. Summary of amine readsorption on LC-MIPs.

The adsorbed amine quantity was measured for each sample at the equilibrium and showed that if a non imprinted material is considered to adsorb equally both enantiomers, the so-called doped material, presenting only a chirality at the mesoscopic level coming from the cholesteric helix, showed only a limited effect in the recognition process, whereas the imprinted material presented a very high recognition. In the last case where the LC-MIP was synthesized, the comparison of the template and the other enantiomer led to a 37% recognition increase. This shows that the main element responsible for the recognition process is the chiral cavity.

The use of this system constituted a good opportunity to study the tunability of the LC-MIPs. By studying the complex formed between the template molecule and the interacting group, we were able to show that the nematic-isotropic transition induced modification of their interactions (figure 8) [35]. For that, the complex, formed by hydrogen bonds between the template and the functional groups, was introduced in a mesogenic solvent (4-(methoxy)-4'-(3-butenyloxy) phenylbenzoate). It was analyzed by infrared spectroscopy, either in the mesogenic phase (cholesteric phase) or in the isotropic state.

A small jump in the 1169 cm^{-1} and 1606 cm^{-1} intensities occurred at the transition between the cholesteric phase and the isotropic state (42 °C) proving that the sample was slightly oriented. At the same time, a clear decrease of the intensity of the associated amine functions was detected at 1378 cm^{-1} concomitant with a small increase of the free amine function intensity at 1368 cm^{-1}. This provided the direct evidence for the jump in concentration of the (R)-methylbenzylamine /4-(3-butenyloxy) benzoic acid complex at the transition temperature between the liquid crystalline phase and the isotropic state. It gave an unambiguous indication that the mesomorphous field has a significant effect on the bonding between the functional acid groups and the amine template; i.e. there is an effective "extra

bonding" energy due to the mesogenic interactions. At the isotropic transition, this "extra bonding" energy is broken. This fact might be used to facilitate the extraction of the template in the liquid crystalline imprinted materials simply by modifying the temperature of the material on both sides of the transition.
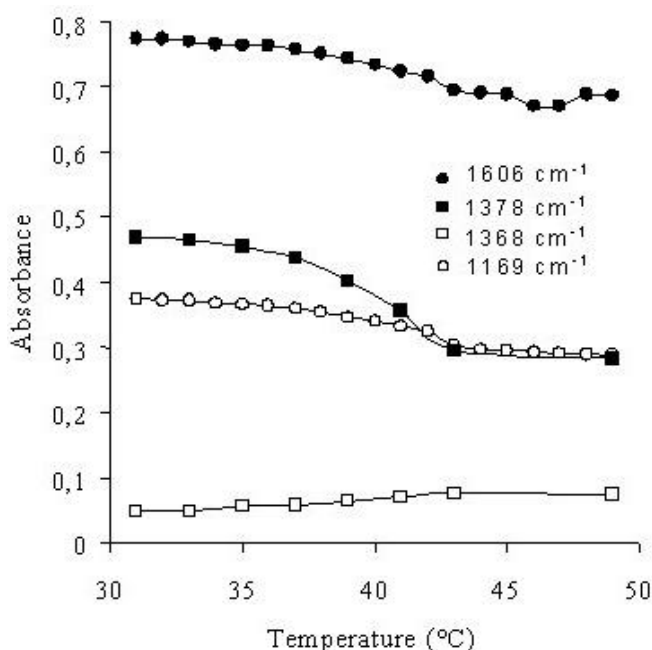

Figure 8. Evolution of IR bands intensity as a function of temperature. (●, o) bands of phenyl groups; (□) C-N stretching in free amine; (■) C-N stretching in the complex. From ref. [35], with permission from ACS.

I.2. CATALYTIC LC-MIPS

When the template molecule is a transition state analogue of a reaction, the final imprinted material can catalyze this reaction, exactly like enzymes do. This technique has been tested for several years now and several reviews can be found in the literature [1,2,4,37]. Various reactions have been catalyzed in this way: elimination [38-40], C-C bond formation [41-43], oxidation or reduction [44-46] and mainly hydrolysis of esters or carbonates [47-53]. The efficiency of the imprint on the catalysis can be described by comparing the kinetics of the MIP compared to a material synthesized with the same molecules (including the monomer that interacts with the template) but without the template. From this point of view, very efficient systems begin to appear, such as in the case of the hydrolysis of carbonate molecules where the reaction has been accelerated 50 to 80 times in the presence of the MIP [49]. These ratios become quite close to what can be obtained by bioimprinting, which uses denatured proteins as MIPs [54,55]. An even more outstanding system was described in 2008 by the combination of transition state stabilization and a defined orientation of the catalytic chemical groups in the MIP: a synthetic carboxypeptidase A was thus obtained leading to a 410 000 fold acceleration. These values are higher compared to artificial antibodies.
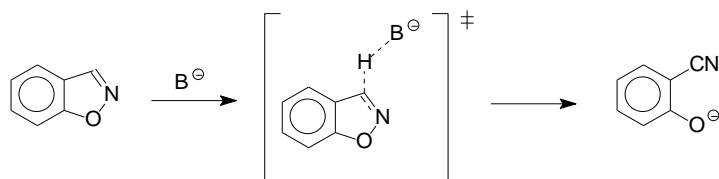
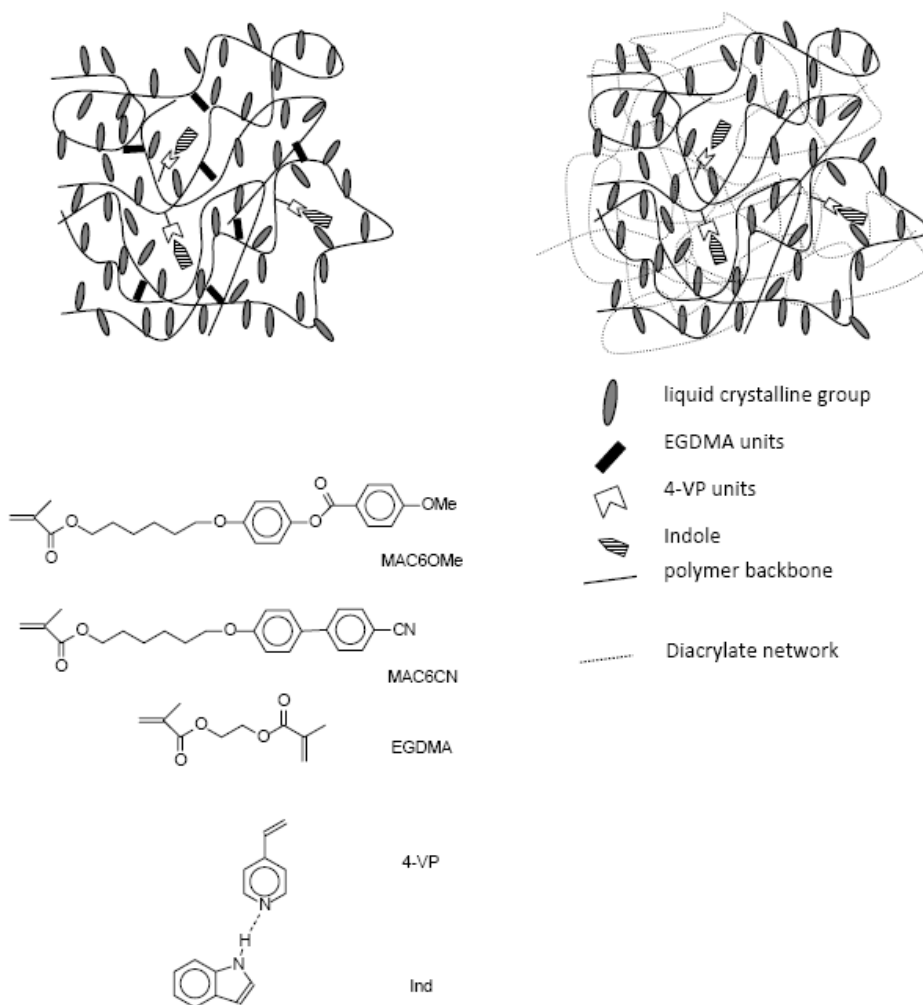Figure 9. Base-catalyzed isomerization of bensizoxazole.

Figure 10. Synthesis of the liquid crystalline MIPs.

In order to characterize the usefulness of liquid crystal MIPs in such systems, we chose a reaction for which a regular MIP had already been described by Mosbach's group. This consisted in benzisoxazole isomerization into 2-cyanophenol [40] (figure 9). Two types of imprinted materials have been synthesized: liquid crystalline elastomers by polymerization/crosslinking under UV irradiation at 4°C and semi-interpenetrated networks (figure 10) [56].
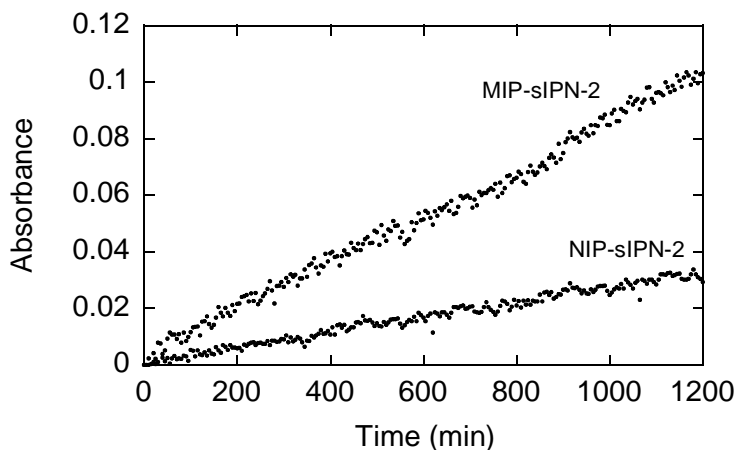
Figure 11. Example of UV signal evolution during benzisoxazole isomerization.

The imprinting was obtained through the interaction between 4-vinylpyridine acting as a functional monomer and indole which can be considered as a transition state analogue of the reaction. For each synthesized material, a similar non imprinted one was obtained. Also, a classical MIP without any liquid crystal was also synthesized in order to compare the catalytic activity of all MIPs exactly in the same conditions.

Since the appearance of cyanophenol can be followed by UV-spectroscopy, this technique was used for the kinetics study and this enabled a comparison between imprinted and non imprinted materials, as well as a comparison between classical MIPs, LC-MIP and the semi-interpenetrated (s-IPN) LC-MIP. In both liquid crystalline and non liquid crystalline materials, an imprinting effect was observed, as depicted in figure 11 as an example. For the classical MIPs, the increase factor was found close to 2.0, whereas for LC-MIPs, this factor was close to 2.8. For the s-IPNs, a higher imprinting effect of 3.5 was calculated. This means that the cavity shape was well kept in the semi-interpenetrating network, even if the functional monomers were not directly linked to the crosslinked system.
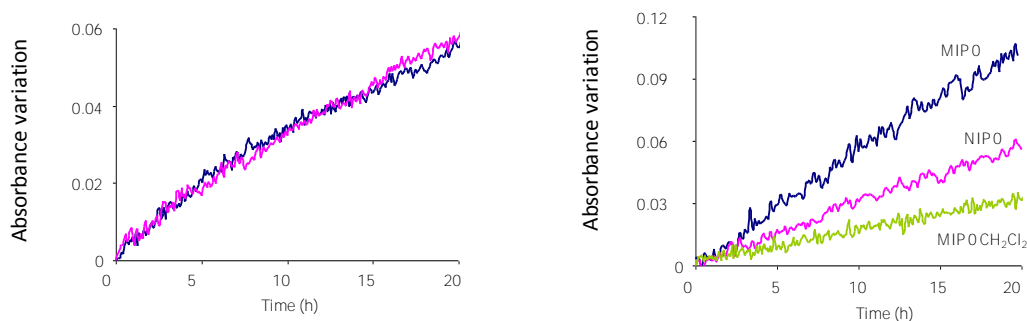

Figure 12. Kinetics curves for benzisoxazole isomerization with LC-MIPs before and after thermal treatment (left) and with classical MIPs before (MIP 0, NIP 0) and after swelling (MIP 0 CH_2Cl_2)(right).

All materials exhibited a limited reaction rate at high substrate concentration, which is typical of enzymes. This saturation occurs when all active sites are used for the catalysis. The

efficiency of catalysis was further characterized by evaluating the reactivity of each type of catalytic sites, namely the imprinted and the non-imprinted ones. Indeed, contrary to enzymes, MIPs exhibit two types of reactive sites, each having different activities. After calculations, these experiments show that the imprinted sites of the semi-interpenetrated network catalyze almost one hundred times more the reaction of isomerization compared to the non imprinted sites [56]. These results are undeniably better than those obtained for non liquid crystal MIP for which the imprinted sites catalysed 22 times more the reaction compared to the non imprinted sites. For LC-MIP, the result is in-between, since the imprinted sites are 37 more active than the non imprinted ones.

An advantage of liquid crystalline imprinted elastomers is also the possibility of using the memory effect of the mesomorphous material. This was predicted by De Gennes [20] and shown by Finkelmann [57-59] or Keller [60,61]. In such systems, a strong coupling between the liquid crystalline moieties and the polymer backbone exists, ensuring that if the elastomer is formed in the mesomorphic phase, the interactions between the liquid crystals impose a permanent arrangement of the backbone's conformation to which the material will come back whenever possible. This effect was assessed in our case using the catalytic LC-MIP developed for the isomerization of benzisoxazole. The kinetics was performed at room temperature on a material before and after thermal treatment above the nematic/isotropic transition. In the case of the LC-MIP, the evolutions were superimposed (figure 12, left). In the same manner, the kinetics was analyzed following strong swelling in a good solvent such as methylene chloride. In this case, the material was first swollen, then deswollen, dried and finally used in the same solvent than the first kinetics. Such a treatment led to an important loss of catalytic activity for classical non liquid crystalline MIPs (figure 12, right), whereas once again, the kinetics curves were superimposed in the case of LC-MIPs. This shape memory constitutes another strong advantage of liquid crystalline MIPs, added to the possibility of controlling the interactions depending on the temperature.

I.3. APPLICATION TO THE RECOGNITION OF PESTICIDES

The growing concerns about pesticides residues in water, air, soils and food have triggered much work on MIPs dedicated to them under different formats such as binding assays, sensors or stationary phases in solid phase extraction [62]. Among all pesticides, organophosphates are of peculiar interest because of their toxicity to human beings and their similarity with nerve agents. Several strategies have been presented in the literature to obtain MIPs for organophosphates. For example, complexes formed between phosphate groups and europium compounds that were covalently trapped inside the polymer were proposed by Jenkins [11,63,64]. Owing to the luminescence of the complexes, they were able to detect quantities of pesticides as low as 10ppt. Another team has synthesized a soluble and processable MIP able to detect dicrotophos at the ppb level [65]. Most often, either methylacrylic acid or 4-vinylpyridine is used as functional monomer for MIPs around phosphonates [10,12,66-69]. Recently, Palmas and coll. have described an elegant method to fully characterize by NMR the complex between pinacolyl methylphosphonate and methacrylic acid, leading to an optimization of MIPs [70,71]. Two original systems are first the use of a calixarene derivative associated to cyclic voltammetry as detection method [72-

74] and secondly the synthesis of MIP composite materials around nylon-6 preformed membranes [75,76]. Advincula's team has assessed the semi-covalent method where a MIP is built around covalent species whereas the recognition is performed through a non covalent one [77]. Several other studies have also appeared in 2009 and often use the interaction of phosphonates with acids [78-80].

All the studies published so far use MIPs which are highly crosslinked in order to have good recognition properties. This implies that the material is used under its glass transition temperature, in a zone where the mobility of the molecules is very low [81,82]. Since none of the above systems described is entirely satisfactory, we decided to examine whether liquid crystal MIPs could improve some of the characteristics, including the possibility to work above glass transition temperature.

The chosen template was a pesticide analogue, diethyl-4-nitrobenzylphosphonate (DE4NBP). The LC-MIPs were synthesized in two steps: first hydrosilylation between a polyhydrogenomethylsiloxane and all functional groups, followed by a photocrosslinking reaction (figure 13).

The interacting moiety was constituted by a fluorinated alcohol which was shown to form a stable complex with DE4NBP [83]. Here again, a benzophenone derivative was used as initiator of the photocrosslinking process. Therefore, the linear polymer was spread onto a Teflon support and the irradiation was performed in the liquid crystalline state.
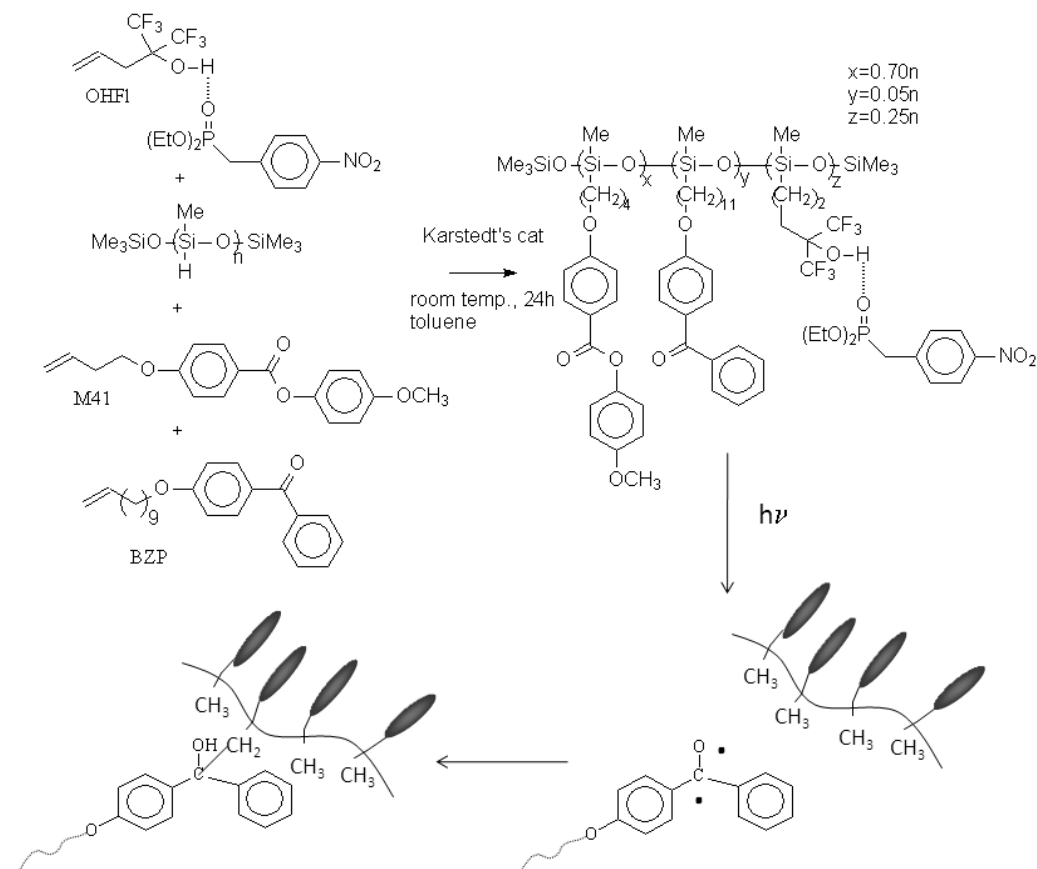

Figure 13. Synthesis of LC-MIP around DE4NBP.

After washing of the films, batch rebinding studies were performed in order to evaluate the recognition properties of the materials. The quantity of DE4NBP measured for each type of films was used for evaluating the general capacity, C_{ads} ($\mu mol.g^{-1}$).

$$C_{ads} = \frac{Q_{DE4NBP}(\mu mol)}{m_{washed\,film}(g)}$$

When comparing the imprinted material (MIP) with the non imprinted one (NIP), it appeared that the capacity of the MIP is higher than for the NIP (Figure 14). Since all films have the same weight and the same geometry (width and thickness), this difference in capacity can be attributed to an imprinted effect. The ratio between the MIP capacity and the NIP capacity gives an imprinting effect of 2.6. This value is comparable to other ones described in the literature for similar template molecules [9,10,77,84].

Although the specificity of the LC-MIP is comparable to the one of classical MIPs, an important point is their possible use above the glass transition of the network. For instance, in the system described in this part, the glass transition temperature was determined by DSC and was found around 21°C. The presence of DE4NBP did not influence this. This means that rebinding experiments can be performed at room temperature and above the glass transition temperature, therefore giving more mobility to the molecules and more permeability in the case of membrane applications. This optimal combination is possible owing to the concomitant use of very flexible polysiloxane chains and of physical crosslinking of the liquid crystals.
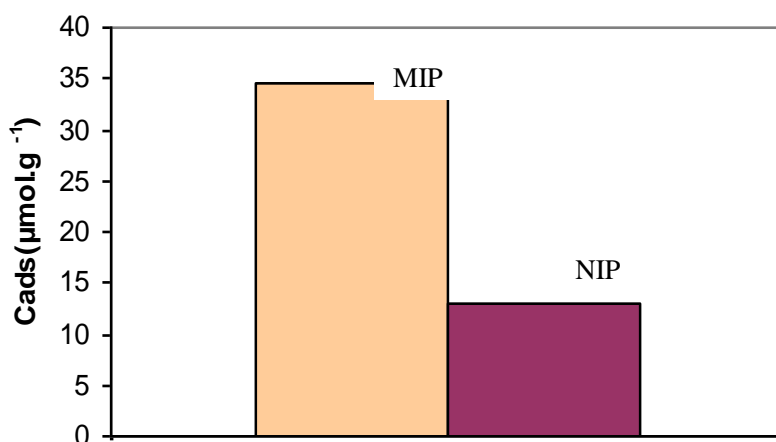

Figure 14. Imprinted (MIP) and non imprinted (NIP) films capacity.

To conclude this part, liquid crystalline MIPs can be synthesized either in a single or a two step process. They present the same type of specificity compared to classical MIPs. Although their synthesis is more demanding than for classical MIPs, they present significant advantages [36]:

– an increased mass capacity
– tunable interactions controlled by external stimuli (temperature, solvent…)

- shape memory
- use in the rubbery region.

II. HYDROGEL MIPS FOR PROTEIN RECOGNITION

Building MIPs in an organic solvent around a small template molecule requires tight interactions between the material and the template to obtain a good recognition process. For large molecules such as proteins, numerous interactions depending on the various amino acid units at their surface may be involved. In this case, in addition to strong and directional interactions, adaptive numerous ones should be present. Indeed, proteins are well known to change their shape and adapt to the interacting moiety. In the case of MIPs, the same phenomenon should happen in which both MIP and the protein should adjust to each other. Another challenge comes from the size of the template. In the case of regular MIPs, the template molecule is small compared to the pores of the MIP, enabling its diffusion inside the material. For molecules as large as proteins, this is not true anymore and the size of the pores has to be adapted to allow the diffusion of the protein or the process has to lead to a surface recognition. Thus, building MIPs for proteins brings new challenges because of the size of the template, its complexity, its conformational flexibility and its limited solubility. New specifications have to be met, because of their possible denaturation. Organic solvents should be generally avoided, as well as increased temperatures. Furthermore, whereas MIPs in organic solvents use hydrogen bonds or Van der Waals forces to get good interactions between the template and the MIP, using aqueous solutions strongly decreases their strength. Therefore, a combination of electrostatic, hydrogen bonds and hydrophobic interactions has to be used to ensure a good recognition process.

Because of the size and the complexity of the proteins, different approaches have been assessed, namely surface imprinting or the epitope approach [85,86]. The latter consists in building the MIP around a small peptide sequence which is as specific as possible of the aimed-at protein. This enables functional monomers to be best spatially organized in the MIP and the synthesis may be simpler. However, this can be very limiting, because other proteins may have the same sequence and also because the whole protein might not be able to reach the recognition site which is much smaller. In surface imprinting, the protein recognition sites are built on the surface of the MIP, therefore enabling a good accessibility. In the present work, we decided to study this last case. Therefore, this chapter will only deal with this type of imprinting. Persons interested in the epitope approach may refer to recent reviews [86-88]. Also, it is not our goal to present a complete overview of the literature on the subject of MIPs for proteins. Once again, the persons should refer to the aforementioned reviews to obtain a more complete outlook. We will mainly describe some of the most recent developments in this field before explaining our study. Another interesting approach is worthwhile citing, although not strictly speaking molecular imprinting. It is based on the synthesis of linear soluble polymers or their analogues grafted onto a solid surface, involving various methacrylamide monomers bearing different hydrophobic or hydrophilic chemical groups [89,90]. Contrary to molecular imprinting, the synthesis is performed without any template and the concept is that the polymer is flexible enough to adapt to the incoming protein and find by itself the most stable interaction.

II.1. RECENT ACCOMPLISHMENTS

Table 1 presents some of the recent studies published. They are arranged according to the crosslinker ratio and they are presented here in order to highlight some crucial parameters when dealing with protein MIPs.

Table 1. Examples of MIPs synthesized for protein recognition

Protein (MW, pI)	Monomers	Crosslinker (amount mol%)	Initiation	Use	Reference
Horseradish peroxidase Microperoxidase Hemoglobin A0 Lactoperoxidase	Aminophenylboronic acid	-	Redox	Modification of microtiter plates IF between 3 and 10	99
Papain	Aminophenylboronic acid	-	Redox	Grafting onto PS microbeads IF≈1.6	103
BSA IgG Fibrinogen	Hexafluoropropylene disaccharides	-	Radio-frequency glow discharge	Nanostructured surface	98
Bovine hemoglobin	Dopamine	-	Redox	Synthesis of superparamagnetic nanoMIPs	104
BSA	tBuAm, AAm, maleic acid	MBAAm (1.8)	Redox	Responsive MIP IF≈6	110
Melittin	NIPAM/AAm/AA/tBuAm	MBAAm (2.0)	Redox	Formation of MIP nanoparticles	111,112
α-fetoprotein	AAm/vinylConA Vinyl antiAFP/AAm	MBAAm (≈2.0)	Redox	Responsive biomolecular MIP	113
Lysozyme Cytochrome C	AAm , AA Me2N ethylmethacrylate	MBAAm (3.0)	Redox	Interaction studies IF~2	95
hemoglobin	tBuAm, AAm, itaconic acid	MBAAm (3.0)	Redox	Effect of pH IF ≈4	114
Bovine serum albumin	NIPAm AAm Me2Npropylmethacrylamide	MBAAm (3.3)	Redox	Thermo- and salt-sensitive MIP IF~2.5	115
Bovine hemoglobin	AAm Acrylamido methylpropanesulfonic acid Methacrylamidopropyl NMe3+	MBAAm (4.9)	Redox	Polyampholyte MIP IF max 13	94
Bovine hemoglobin	AAm Acrylamido methylpropanesulfonic acid or Methacrylamidopropyl NMe3+	MBAAm (4.9)	Redox	Effect of charge density washings IF max 15	93

Bovine hemoglobin FITC-albumin	AAm	MBAAm (5.0)	Redox	hydroMIPs	91,92
Human hemoglobin Bovine hemoglobin	AAm	MBAAm (6.5)	hν	Label free detection Use of photonic suspension array	107,108
Hemoglobin	Methacryloyl-L-histidine methyl ester HEMA	MBAAm (14.3)	Redox	Cryogel MIP for SPE IF max 38?	116
Carcinoembryonic antigen	Poly(allylamine HCl) M=15 000	Ethyleneglycol diglycidyl ether (15.0)		Cancer biomarker analysis IF≈5	117
BSA	AAm	MBAAm (16.0)	Redox	MIP on carbon nanotubes	105
Trypsin	MA, Styr	DVB (17.0)	Thermal followed by hν	Structuring of sensor materials	101
BSA Egg albumin Lysozyme	Saccharose MA	EGDMA (18.0)	hν	MIP photonic crystal from sacrificial silica particles IF ≈ 6	106
Myoglobin	MMA, 4-VPyr	TRIM (66.0)	hν	Surface bound nanofilaments IF≈9	109
Lysozyme cristallisé	AA 2-methacryloyl phosphorylcholine (MPC) PEG	MBAAm (71.0)	Redox	Recognition of protein cristals IF~5	102
Lysozyme	AA 2-methacryloyl phosphorylcholine (MPC)	MBAAm (71.0)	Redox	Surface plasmon resonance sensor " IF "~3	100
RNase A	MMA	EGDMA (73.0)	Redox	Miniemulsion polymerization Study of protein denaturation 3<IF<14	118
Lysozyme Ribonuclease A Myoglobin Ovalbumin	MA, HEMA Me$_2$Nethylmethacrylate, Styr MMA, 4-VP	EGDMA TEGDMA PEG400DMA PEG600DMA (>90)	hν	Selection of best monomers IF~3-16	96,97

IF imprinting factor, PS polystyrene, tBuAm tert-butyl acrylamide, AAm acrylamide, NIPAM N-isopropyl acrylamide, AA acrylic acid, HEMA hydroxyethyl methacrylate, MA methacrylic acid, Styr styrene, MMA methyl methacrylate, 4-VPyr 4-vinyl pyridine, PEG poly(ethylene glycol).

Process choice. The first element is the process used to increase the contact surface between the MIP and the template protein. The MIPs can be grounded and sieved, just as regular MIPs are [91-95] (figure 15, left). However, for protein MIPs, this often has to be a wet sieving process to avoid possible denaturation of the protein. Furthermore, sieving leads to very irregular particles, which is not always compatible with the intended application.
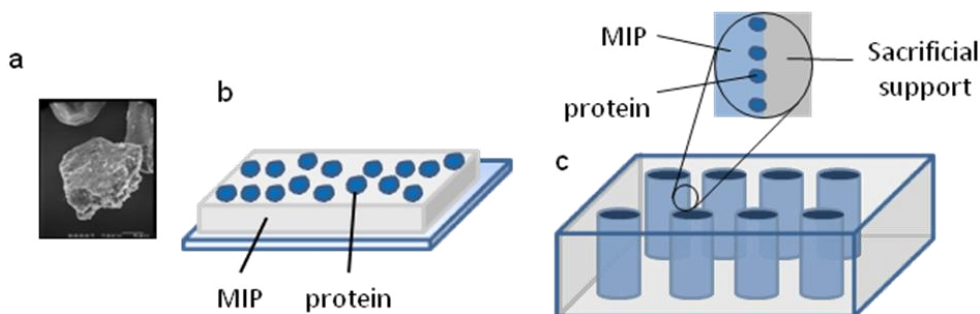
Figure 15. Schematic representation of processes used in protein MIP syntheses. a) 3D MIP, b) planar MIPs, c) surface recognition through sacrificial support.

The second process used consists in first adsorbing the protein onto a plane solid surface such as glass, synthesizing the MIP over it and peeling off the whole assembly [96-99] (figure 15, center). This is typically the case when sensors are looked for [100]. If this strategy is satisfactory in terms of the concept and the good accessibility of the imprinted sites, it presents two main shortcomings. First, the question has to be raised whether depositing the protein on a solid surface may lead to denaturation. This question is almost always overlooked in the literature and it is still not clear whether using a MIP with sites corresponding to a denatured protein might be detrimental to the recognition or not. Indeed, depending on each application and each protein studied, the protein in the rebinding experiments might be able to adapt to these sites and be recognized by the MIP. In a still unequaled study, Ratner proposed the use of mica as the solid surface first to minimize this effect of possible denaturation [98] and secondly to be sure that the observed surface topography reflected that of the template molecules. Dickert also adopted this strategy by first prepolymerizing the monomers, depositing the protein over them, and finishing the crosslinking [101]. The second shortcoming of this method is the quite low surface area. This strongly limits the number of possible recognition sites, thus the maximum sensitivity. To increase the number of sites, some authors have proposed the use of already existing solid templates, which also increases the mechanical properties of the MIP. Thus, supports such as cellulose membranes [102] or polystyrene beads [103] have been examined. Magnetic nanospheres [104] or carbon nanotubes [105] have also been assessed. They present the advantage of an easy recovery of the MIP associated to a good dispersion.

In a third process, a porous sacrificial solid such as silica or alumina is used. The protein is deposited into the pores before the monomer solution is poured (figure 15, right). After polymerization and destruction of the sacrificial solid, this leads to MIPs having a large external surface, with the protein recognition sites at their surface [106-109]. Following this elegant strategy, Haupt used porous alumina to obtain nanofilaments of MIPs grafted onto a glass plate. Two Chinese teams have presented the use of colloidal silica crystals to develop label-free chemical sensors [106-108]. After destroying the silica template, the MIP has the negative organization compared to silica, the consequence of which being that it exhibits different wavelengths absorption depending on the quantity of molecule readsorbed into the MIP. Therefore, readsorbing the protein template leads to a wavelength shift which is easily detected.

Finally, a special attention should be given to a new example developed by Shea and coll. They succeeded in synthesizing MIPs in the shape of 30-40 nm nanoparticles in the presence

of a peptide biotoxin, melittin, which is the main component of bee venom. By adopting this process, they were able to intravenously inject the MIPs which acted as an antidote in melittin-infected mice. This constitutes the first time where MIPs have been used in *in vivo* conditions. Biodistributions studies showed that following MIPs' injection, the nanoparticles were able to bind to melittin before being eliminated from the bloodstream by macrophages and led to the liver [111,112].

Synthesis method. To obtain MIPs, radical polymerization is the most frequent method, because very different monomers are compatible with this chemical process and can be reacted together. As already mentioned, in the case of a protein as template, very mild conditions should be sought. Thus, redox and photochemical initiations are mostly employed, since they enable polymerization at room temperature. Once again, a relevant question is whether a denaturation of the protein could occur during the process. Tong is the only author to have partially examined this point. He showed by circular dichroism that RNase A was denaturated after 24h of irradiation [118]. However, this time of irradiation is much longer than typical reaction times, which are closer to a few minutes [107-109] or an hour [106]. Unfortunately, Tong did not check the effect of redox conditions on denaturation.

This part on hydrogel MIPs began by explaining that contrary to classical MIPs, there is a strong need of mobility inside the MIP, and that the crosslinker ratio should be decreased consequently. The above table shows that the situation is far from simple, since crosslinker quantities as low as 1.8mol% [110] have been used, up to 90mol% [96,97], apart from the few examples without any crosslinker at all but which use non acrylate polymers [98,99,103,104]. A notable exception is the study from Schrader which examined the possibility of protein recognition even with soluble polymers or uncrosslinked soft polymer chains grafted onto a solid surface [89,90]. Contrary to the small molecules, there is no correlation between the crosslinker ratio and the imprinting factor IF. For all formulations used, this factor is typically between 2 and 20. This means that each case should be examined and optimized independently.

Another comment on the synthesis conditions is the quantity of solvent. Once again, this markedly varies between the different studies. Some do not use any solvent at all [96,97,109] whereas others prefer quite diluted systems [100]. In a general manner, concentrations of monomers in the 0.5-1M range are frequent. Most teams use either aqueous or buffer solutions for the synthesis, but a few have presented the use of alcohols such as methanol [110] or ethanol [106]. The choice of solvent quantity strongly depends on the application, since this can bring porosity into the MIP. An interesting case was described by Denizli who synthesized the MIP as a cryogel. In this system, the monomer solution was partially frozen, which enabled the polymerization to occur while the ice crystals acted as a porogen [116].

Most of the studies presented above used very well known and abundant proteins. This enables relatively high concentrations of the protein during the synthesis, typically between 10 and 50 mg/ml [100,115]. If this is mandatory for fundamental studies, this does not correspond to what will be needed to analyze specific proteins. Indeed, for unusual proteins, a concentration of 100µg/ml is very common. Only a few teams have published on such conditions [96,97,99,117,118]. Most of them correspond to the stiffest hydrogel MIPs with high crosslinker content or polymers other than acrylate types. Kofinas however described the use of a soft hydrogel for the quantification of a cancer biomarker [117]. One of the lowest concentration used (1.7µg/ml) was that in Shea's study on melittin [111,112].

Another important point is the nature of the monomers used. In most cases, hydrophilic ones are chosen for protein MIPs, the most frequent being (meth)acrylic acid and acrylamide [91,92,94,96,97,102,106-108,114]. Other hydrophilic monomers which have been used are hydroxyethylmethacrylate [96,97,116], acrylamido methylpropanesulfonic acid [94] or amine-bearing acrylates or acrylamides [95-97,115]. In his study on grafted polymers, Schrader used several types of either hydrophobic or hydrophilic monomers, such as methacryloyl glucosamide or cyclohexylmethyl methacrylamide [89,90]. In some cases, specific monomers have been developed, such as 2-methacryloylphosphoryl choline (MPC) [100,102] or methacryloyl-L-histidine methyl ester (MAH) [116]. MPC was designed to decrease non specific interactions and MAH to mimic possible interactions in proteins. Schrader described the use of a methacrylamide bearing two phosphonate groups that binds specifically to arginine residues, enabling selectivity in the final polymer [89,90]. Kofinas published an unusual MIP based on polyallylamine crosslinked in the presence of a diepoxide [117]. Only a few studies use hydrophobic monomers alone [118]. In these cases, the obtained MIP is stiff and the shape of the cavity plays an important role in the recognition process. In his study on mica, Ratner used two types of monomers, a very hydrophobic fluorinated one and a layer of disaccharides which become linked to the MIP during the polymerization process [98].

Table 2. Characteristics of proteins and peptides used in MIPs

Protein	Molecular weight	Isoelectric point
Bovine hemoglobin	67 000	6.8
Bovine Serum Albumin (BSA)	66 000	4.8
Carcinoembryonic antigen	180 000	4.7
Cytochrome C	12 300	10.2
Egg albumin	36 000	4.7
α-fetoprotein	70 000	4.7
Fibrinogen	340 000	5.5
Hemoglobin A0	66 000	7.2
Horseradish peroxidase	40 000	7.5
Human hemoglobin	65 000	7.2
Immunoglobulin G (IgG)	150 000	6.4-9.0
Lactoperoxidase	77 500	8.1
Lysozyme	14 400	10.5-11.0
Melittin	2 850	9.1
Microperoxidase	1 900	4.7
Myoglobin	16 900	7.0
Papain	23 000	9.5
RNase A	13 712	9.6
Trypsin	23 300	9.3

Since electrostatic interactions are strong in aqueous solutions, a special attention should be given to the isoelectric point of the template (table 2). So, if the protein bears a net positive charge, monomers leading to a net negative charge at pH 7 should be favored [100] and vice versa [115].

Protein detection. This is an essential part in developing MIPs, where a method should be found to detect and quantify the protein. Typically, three types of techniques are used: spectroscopic, physical and enzymatic ones. In most cases, simple UV-visible absorption is employed [91-95,99,103,110,114-116]. This naturally implies that the protein concentrations are high enough to be detected. Haupt has presented the quantification of rebinding by fluorescence using a modified protein [109]. He was able to quantify the fluorescence and compare different samples by using relative fluorescence intensities, after withdrawing of a baseline signal coming from the substrate itself. Other very elegant methods recently described used photonic nanocrystals [106-108]. These enable the recognition detection by simple UV-visible absorption, although the templates themselves do not necessarily absorb light. Other methods include radioactive labeling [98], Surface Plasmon Resonance [100,102] or Quartz Crystal Microbalance [101]. An interesting quantification was proposed by Bossi which used the enzymatic activity of the trapped protein. In the presence of an adequate substrate, this activity was detected by UV-visible absorption [99]. Chou is the only author who has adopted ELISA tests to the case of MIPs [96,97]. This has the tremendous advantage of detecting very small amounts of proteins, but the shortcoming is that it can be done only with very well known proteins for which antibodies have been developed.

Extraction methods. When the MIP is synthesized around the exact same template than the molecule to be recognized, it is essential to ensure that all templates have been removed before doing any batch rebinding tests. Else, two problems arise: first some sites are already occupied, which of course limits the loading of the MIP and secondly, it is not possible to discriminate whether new extracted molecules are from the beginning or from the rebinding studies. Thus, this is a crucial point. If extraction of small molecules is relatively simple, extraction of proteins on or inside a MIP is more complex. Some teams have described the extraction of the protein with simple washings in water or buffer solution [102,107,108,110,113,115]. Very often, in order to ensure a complete extraction of the protein, a combined washing solution of acetic acid and a surfactant such as SDS is used [91,92,96,97,99,107-109]. Besides the fact that SDS can lead to protein denaturation, it is very difficult to completely eliminate it from the MIP. This has led to a very interesting study from Kofinas which has shown that this remaining SDS may lead to an artificially increased imprinting factor, especially if the exact same washing procedure is not used for the non-imprinted analogue [93,94]. In order to avoid this, he has shown that, following SDS washing, the MIP should be rinsed thrice with NaCl solutions. This ensures that no SDS remains in the MIP.

Since the template molecule is a protein, some authors have assessed the use of proteases such as trypsin as a mild method to remove it from the MIP [86,91,99]. According to Bossi, there is no gain by using trypsin and, for Reddy, trypsin hydrolysis of the template leads to less efficient MIPs than SDS/AcOH washing. Although the template removal percent is higher for trypsin hydrolysis, the rebinding tests are better for SDS/AcOH washings [91]. This was attributed to the fact that trypsin may produce small peptide fragments that remain in the MIP. However, the washings with SDS were not performed with NaCl rinsing steps, so the problem mentioned by Kofinas might be a reason for this [93,94].

Visualization of the cavities. When synthesizing MIPs around proteins, the recognition sites should correspond to that of the template, i.e typically in the 5-100nm range. Therefore, it is not unrealistic to seek and observe these cavities by modern techniques such as electronic microscopy or AFM. Actually, only a few teams have been able to describe this. The first one

was Ratner in 1999 who has presented a tapping mode AFM image of the surface of fibrinogen imprint [98]. This was facilitated by the fact that the MIP polymer was a very hard one. Chou also depicted, in 2009, AFM images of MIP cavities before and after batch rebinding experiments with ovalbumin [97]. He showed that existing 34nm cavities were filled with the protein. Here again, the absence of solvent in the synthesis together with the high crosslinker ratio leading to a stiff MIP facilitated this observation. Dickert [101] also presented AFM images of both yeasts (several microns wide) and tobacco mosaic virus (300 nm x18 nm) in a rigid MIP. Reddy is the only author to have described a process to observe cavities in MIPs by transmission electron microscopy [119]. From a general point of view, observing the cavities remains a very difficult task, especially on softer materials.

Rebinding tests. This is naturally the central point for all MIPs, the reason why they are made. In most cases of protein MIPs, the rebinding tests are performed in the same solvent than the one used for the synthesis. This is to ensure that the shape of the cavity does not change before the recognition occurs [91,92,94,95,100,109,113]. A few authors have investigated the influence of washing and readsorbtion solutions on the rebinding results more thoroughly. Takeuchi [100] examined the effect of NaCl concentration in MIPs for lysozyme. He showed that the optimal NaCl concentration was 40mM during the polymerization, and 20mM for rebinding tests. In another study, Zhao also assessed the influence of NaCl concentration in thermosensitive MIPs around bovine serum albumin [115]. In this case, best results were obtained for a 1mM concentration. Another possibly important point is the pH of rebinding solution. However, most studies use either distilled water or PBS solutions to do the rebinding experiments. This latter ensures that the medium is as close as possible to the natural cellular medium. Exception are two studies of Caykara [110,114] where he characterized the influence of pH in the rebinding of hemoglobin and BSA. He found a different optimal pH for each protein. Although pH is a crucial parameter for the activity of proteins, the most frequent way of synthesizing MIPs for proteins is to use a neutral pH and to benefit from the residual charge of the protein to adjust the interactions with the MIP.

Characterizing rebinding capacities of MIPs implies performing several experiments. The first basic one consists in comparing rebinding of the template into a MIP to a non imprinted analogue. This leads to the imprinting factor (IF). Among all studies, common IFs are found between 1.5 and 20. A second test analyzes the selectivity of the MIP. To do this, rebinding of different proteins are measured and compared to that of the template protein. This is performed in most studies reported in table 1. The most realistic rebinding studies imply testing competitive rebinding, i.e. with several proteins present at the same time in the rebinding solution [95-99,104,106,115,116,118]. Analyzing the results obtained for the protein MIPs show that these materials do not have rebinding selectivities as high as natural proteins. This could be detrimental to applications in complex living systems. In order to overcome this, Takeuchi and coll. proposed a method of protein profiling by several MIP arrays [87,120]. By developing several MIPs against several different proteins and testing rebinding for all crossed systems, they were able to build 3D profiles with different areas. An answer obtained in a certain area points to a certain protein. There is no doubt that this is a very promising way to adapt shortcomings of MIPs.

II.2. PATTERNING OF MIPS FOR MICRO- AND NANOSYSTEMS APPLICATIONS

For applications in which the purpose is to make extractions, the protein MIPs can be used as bulk materials. However, in other occasions, such as microfluidics or diagnostic devices, a micro- or nanostructuration of the material becomes mandatory. This is the reason why we decided to assess the synthesis of protein MIPs in the shape of hydrogel films patterned onto a solid support (figure 16).
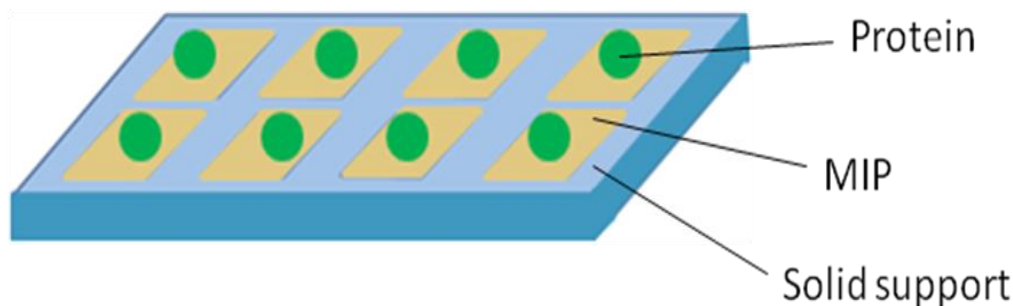

Figure 16. Patterning of protein MIP.

The following specifications were established:

- − Synthesis in protein-friendly conditions (buffer solution, room temperature if possible)
- − Transparent and homogenous material to enable applications in optics devices
- − Porosity compatible with the size of the protein
- − Process compatible with micro- and nanosystems established technologies
- − Size of the MIP patterns below 10 microns

This part of the chapter will mainly focus on the steps discussed earlier in a more general manner. This will give a more tangible idea about the difficulties to be solved when developing protein MIPs.

Process selection. The selected process has already been described above. It consists in first adsorbing the protein onto a plane solid surface such as glass, synthesizing the MIP over it and peeling off the whole assembly [96-99]. In our case, the process becomes more complicated due to the patterning required. Two types of strategies were considered (figure 17): in a first one, the pattern came from the micro-contact printing of the protein onto a solid surface and, in a second one, it was obtained from soft UV nanoimprint lithography (UV-NIL).

Micro-contact printing is a method using a siloxane stamp to transfer patterns onto a solid surface [96,121-126]. The molecules to be transferred are first wetted onto the stamp and deposited onto the support by simply putting the stamp in contact with the surface. In the areas where this contact occurred, the molecules transfer from the stamp to the surface. The

patterned protein can subsequently be covered with a monomer solution and, after crosslinking, the desired device can be obtained.

The second strategy in figure 17 uses a micro- or nanometric mould to form the patterns into the hydrogel. In this case, the protein can be dissolved in the monomer solution, deposited onto the surface and molded. In a similar strategy, micro-contact printing of a protein is used to obtain a PDMS mould having proteins at the bottom of the cavities. This mould can then be used in soft UV-NIL over the monomer solution to shape the MIP into the desired device. Soft UV-NIL is a lithography method in which a mould presses into a prepolymer before crosslinking is induced by UV irradiation [127-134]. Various patterns, including nanometric ones, can be obtained by this technique.
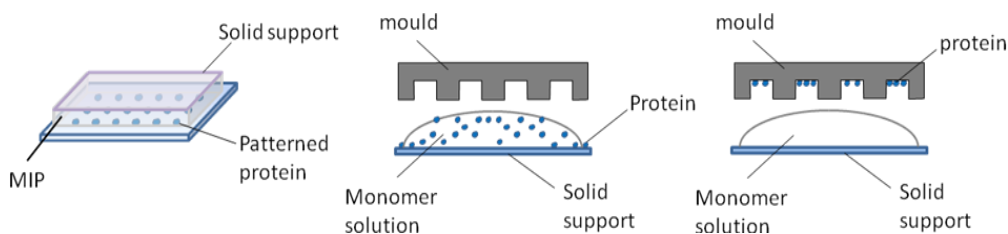
Figure 17. Different processes of patterning.

In such devices, the quality of the film as well as the interfaces play a major role. The one between the MIP and the solid support should be as strong as possible to avoid peeling of the MIP. This implies functionnalization of the surface of the solid support to create chemical bonds between the MIP and its support. In many cases, it consists in modifying the surface with an alkoxysilane bearing another reactive function that can react with the film constituents, such as acrylates. An example is methacryloxypropyltrimethoxysilane which can be grafted onto glass slides in toluene at 80°C.

The second interface is the one between the protein and the MIP, which is essential for the recognition process. This is controlled by the choice of the functional monomers involved in the formulation of the MIP. As already discussed earlier, several authors have proposed the use of saccharides as a protective layer between the protein and the MIP [90,98,106], based on the interactions known to occur in cells. It enables multiple interactions and is also known to lead to a minimum denaturation of the proteins. In our case, we decided to use a monomer having two characteristics: α, β-glucosyloxyethylmethacrylate (GEMA). First, its incorporation into a hydrogel is possible by polymerization of its methacrylate moiety. Secondly, it can interact with a protein by its saccharidic group (figure 18).
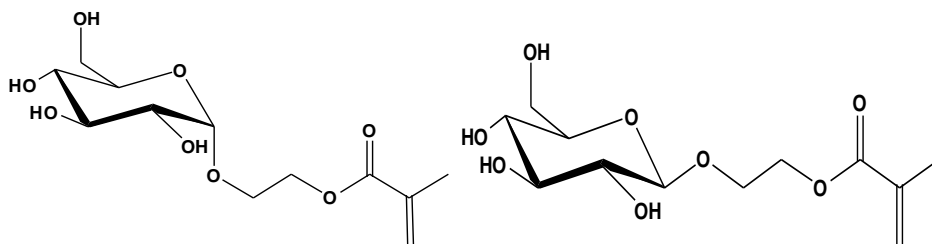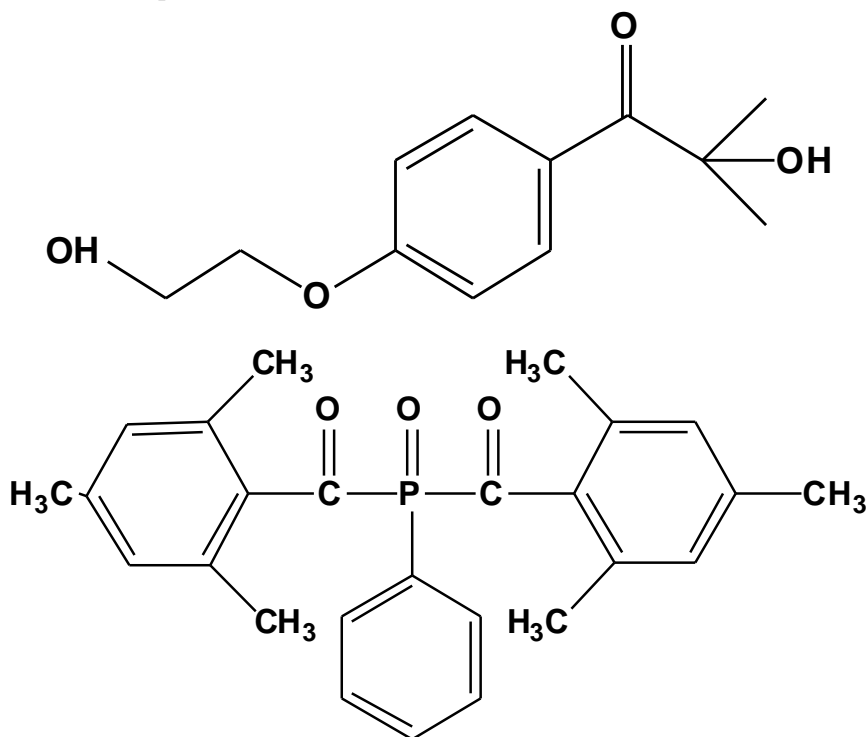

Figure 18. Structure of GEMA.

A third interface to be considered in such processes is the one between the mould and the MIP. Indeed, it is important that the monomer solution correctly wets the mould in order to reproduce the patterns. However, the interaction between the mould and the formed MIP should not be too strong to enable an easy peeling off the mould. This is a very common problem in microelectronics [126-128,130,134,135]. Frequently used chemical modifyers include alkylchlorosilanes or fluorinated molecules. They decrease the force of adhesion by hydrophobization of the surface. However, when the solution that must wet the mould is aqueous, they can lead to complete loss of the resolution of the patterns obtained owing to a poor wetting. In these cases, the modifyers have to be optimized to lead to a good wetting as well as a good peeling off. This can be controled by varying the length of the alkyl chain or that of the fluorinated part.

2-hydroxy 4'-(2-hydroxy-ethoxy)- Bis-acyl- phosphine oxide (BAPO) 2-methylpropiophenone (HHMP).

Figure 19. Structure of photoinitiators.

Synthesis method. We used radical polymerization to synthesize the MIPs, based on redox or photochemical initiation. The redox system we have chosen is a mixture of ammonium persulfate $(NH_4)_2S_2O_8$ and sodium thiosulfate $Na_2S_2O_3$ and two photochemical initiators were tested: a propiophenone derivative HHMP and a phosphine oxide BAPO (figure 19). BAPO forms a suspension in buffer whereas HHMP dissolves over 50°C with ultra-sonication. The photopolymerization was achieved by irradiating at a wavelength above 332nm. Both systems enable working at room temperature, but as far as patterning is concerned, photochemical initiation displays the advantage of a better control in time and space.

Beside GEMA, the formulation of the monomers included other hydrophilic molecules. Indeed, neutral, positively charged or negatively charged monomers were tested in the materials, such as hydroxyethylmethacrylate, 2-sulfoethyl methacrylate and aminoethyl methacrylate hydrochloride. Crosslinking agents were kept between 30 and 50 mol%. Two crosslinkers with very different lengths, namely N,N'- methylene bis acrylamide and poly (ethylene glycol) 1000 dimethacrylate, were used, in order to maintain some freedom in the gel while ensuring a certain mechanical resistance. As already mentioned, the choice of monomers is essential, since the interactions with the proteins are based on it. Therefore, for each protein, a new formulation has to be developed and optimized. This is why we decided to assess different types of monomers, which could interact with different proteins.

Our choice of the proteins was mainly determined by the possibility to observe the system by fluorescence. For this, several parameters have to be examined. The first one is the compatibility of the fluorophore used with the polymerization process. Indeed, many fluorescent probes that can be grafted onto proteins are not compatible with radical processes. By choosing naturally fluorescent proteins such as GFP or mCherry, this problem can be overcome, because the fluorescence comes from a spatial arrangement of several amino acids. The second parameter is the possible autofluorescence of the MIP itself. Even if no aromatic groups are present in the formulation, the material can exhibit autofluorescence which might completely mask that of the protein if happening at the same wavelength. This is particularly the case for fluorescence in the green wavelengths. Finally, another parameter is the choice of the equipment to analyze the fluorescence. If this might appear as an evidence, it is however very important to use an instrument which is compatible with the wavelengths used. This can become a problem for instance when using fluorescence scanners, because these instruments have been developed for the analysis of DNA chips working with only a few fluorophores such as Cy5. Since these cannot be used for the polymerization, it is then necessary to find a fluorescent protein which is compatible with the lasers and the filters of the scanner.

Based on this, three proteins were selected for this study: Green Fluorescent Protein (GFP), Glutathione S-transferase (GST) and mCherry.

Protein	Molecular weight	Isoelectric point
GFP	27-30 000	5.34
mCherry	28 800	6.2
GST	53 000	6.6

The first two are naturally fluorescent proteins which resist to radical processes and the third protein, GST, is not fluorescent by itself. However, anti-GST fluorescent antibodies are commercially available, which can lead to a visual detection.

In a first approach, millimetric patterns were formed on a glass plate and the polymerization was carried out by irradiation after deposition of the protein and the monomer solution. This enabled us to observe the fluorescent patterns by a fluorescence scanner and to check the difference in fluorescence intensity between patterns containing the fluorescent protein and the others. At this step, GFP proved to be a poor choice, because of an unexpected autofluorescence of the polymer, and had to be rejected. GST was then prefered for preliminary tests owing to its availability. This implied the use of the fluorescent antibody anti-GST and the passivation of the surface with albumin before GST-antiGST binding test.

Extraction method and rebinding tests. Based on the litterature, several washing methods were examined, either using only sodium chloride solutions or acetic acid or sodium dodecyl sulfate (SDS). Figure 20 presents an example of simple sodium chloride washing. In this case, the fluorescence intensity before and after washing did not change, showing that such a washing does not work in these conditions. Based on Valvanuz-L. A et al.'s work [109], a new protocol was applied to promote the protein-antibody removal. It consisted first in washing three times with deionised water, then four times with 20 ml of AcOH-SDS solution for 30 minutes, afterwards 3 washing series with NaCl 150mM and three final washings with deionised water. As a result, the red fluorescence significantly decreased. Thus, in this case, only addition of SDS in the presence of acetic acid led to a strong decrease of the fluorescence.
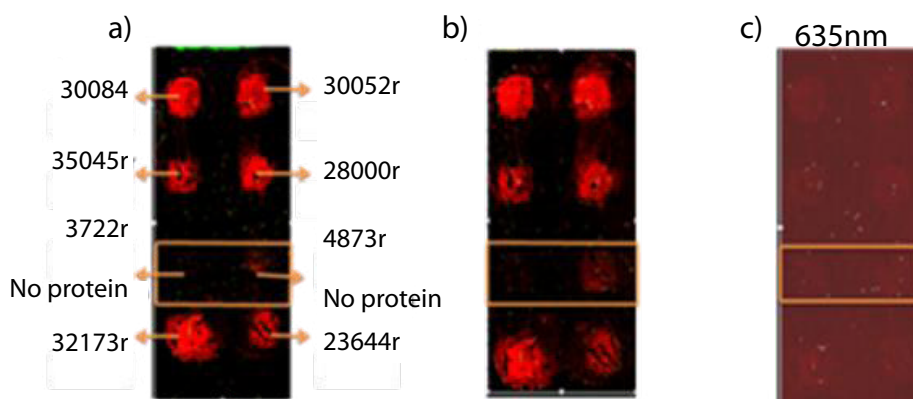
Figure 20. Fluorescence images of MIP after washing steps. a) washing with 3mM NaCl b) washing with 150mM NaCl, c) washing with SDS/AcOH. The rectangle represents the position of non imprinted material.

Following these washing steps, rebinding was assessed. For this, 4µl of the same GST solution used in the material synthesis were placed on each pattern for 1 hour. Afterwards, the material was rinsed with deionised water to remove protein excess which was weakly adsorbed. The film was then covered with 1ml of BSA solution ([BSA] = 1mg/ml). After 30 minutes, the film was washed three times with buffer for 10 minutes. Subsequently, 200µl of antiGST ([antiGST]=5.5µg/ml) were spread on the film for a period of 1 hour. This led to a slightly higher preference of GST towards MIP versus NIP of 1.4.

The use of fluorescence scanner is an efficient qualitative method. It enables to discriminate between the presence and the absence of the molecule. However, quantification is more difficult, because all images have to be performed in exactly the same conditions (spot size, scanning rate, gain and laser power) to compare the measured intensities. This is illustrated by comparing the images in figure 20 which were registered with the same parameters and the following ones (figure 21), which were registered in optimized conditions for each scan. In this latter, the gain was systematically changed in order to enhance the image result. Whereas figure 20 shows an efficient washing, figure 21 may lead to think that a lot of the protein remains in the polymer.
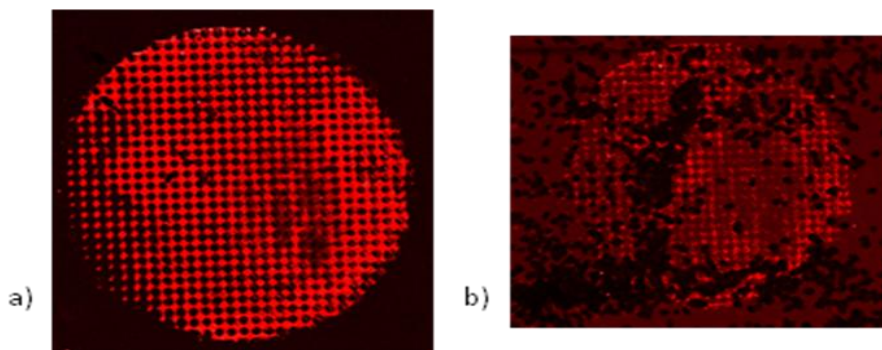
Figure 21. Fluorescence images of Streptavidin-Cy5 prints onto a polymer. a) initial print. b) after washing steps.

Furthermore, a slight difference in film thickness might lead to different fluorescence values. In our systems, the thickness is a parameter that cannot be neglected, since it is typically close to 100 μm. This is different from DNA chips where the molecules are directly deposited onto the solid surface. In our case, the focus position used by the scanner is essential and should be assessed, to determine whether it should be at a fixed value for all experiments or whether the instrument should be let optimizing it by itself.

Another problem in this system is also the quantity of antibodies necessary to quantify the GST protein. Indeed, in order to exploit the antibody quantitatively, a stoechiometric ratio between the antibody and the protein should be used, which would be far too expensive owing to the number of proteins on the surface.

Therefore, a lot of precautions should be taken using the method presented for quantitative analyses. Other complementary techniques such as QCM (quartz crystal micro balance) should be examined to guarantee a quantitative estimation of the imprinting effect.

CONCLUSION

Since its early development more than 20 years ago, the molecular imprinting technique has been the subject of tremendous evolutions, from the chemical but also the technical standpoint. The first very stiff and poorly accessible MIPS were only able to recognize small molecules in organic solvents. This was quite limiting and prevented many applications. Since then, many studies have been published to improve the accessibility of the MIPs either through the chemistry or through the development of surface MIPs. Another big challenge remains the use of MIPs in aqueous systems where interactions are more difficult to control. Finally, the use of MIPs as artificial antibodies to selectively interact with proteins is today a dream to be reached for chemists and this quest has completely altered some of the formerly established rules for a good design of MIPs. To achieve such a goal, it is important to bring together several science specialties. The development of nanoarrays of MIPs is a brilliant example of this.

ACKNOWLEDGMENTS

REFERENCES

[1]   Alexander, C.; Andersson, H. S.; Andersson, L. I.; Ansell, R. J.; Kirsch, N.; Nicholls, I. A.; O'Mahony, J.; Whitcombe, M. J. *Journal of Molecular Recognition* 2006, *19*, 106-180.

[2]   Marty, J.-D.; Mauzac, M. *Advances in Polymer Science* 2005, *172*, 1-35.

[3]   Tamayo, F. G.; Turiel, E.; Martin-Esteban, A. *Journal of Chromatography A* 2007, *1152*, 32-40.

[4]   Alexander, C.; Davidson, L.; Hayes, W. *Tetrahedron* 2003, *59*, 2025.

[5]   Andersson, L. I. *Analytical Chemistry* 1996, *68*, 111-117.

[6]   Sellergren, B.; Lepistö, M.; Mosbach, K. *Journal of the American Chemical Society* 1988, *110*, 5853-5860.

[7]   Svenson, J.; Karlsson, J.; Nicholls, I. A. *Journal of Chromatography A* 2004, *1024*, 39-44.

[8]   Idziak, I.; Benrebough, A.; Deschamps, F. *Analytica Chimica Acta* 2001, *435*, 137-140.

[9]   Hall, A. J.; Lanza-Sellergren, F.; Manesiotis, P.; Sellergren, B. *Analytica Chimica Acta* 2005, *538*, 9-14.

[10] Harvey, S. D. *Journal of Separation Science* 2005, *28*, 1221-1230.

[11] Jenkins, A. L.; Bae, S. Y. *Analytica Chimica Acta* 2005, *542*, 32-37.

[12] Lv, Y.; Lin, Z.; Feng, W.; Zhou, X.; Tan, T. *Biochemical Engineering Journal* 2007, *36*, 221-229.

[13] Takeuchi, T.; Ugata, S.; Masuda, S.; Matsui, J.; Takase, M. *Organic and Biomolecular Chemistry* 2004, *2*, 2563-2566.

[14] Yamazaki, T.; Yilmaz, E.; Mosbach, K.; Sode, K. *Analytica Chimica Acta* 2001, *435*, 209.

[15] Ramström, O.; Ye, L.; Mosbach, K. *Chemistry and Biology* 1996, *3*, 471.

[16] Vlatakis, G.; Andersson, L. I.; Müller, R.; Mosbach, K. *Nature (London)* 1993, *361*, 645.

[17] Schmidt, R. H.; Mosbach, K.; Haupt, K. *Advanced Materials (Weinheim, Federal Republic of Germany)* 2004, *16*, 719-722.

[18] Yilmaz, E.; Haupt, K.; Mosbach, K. *Angewandte Chemie, International Edition in English* 2000, *39*, 2115-2118.

[19] Ulbricht, M. *Polymer* 2006, *47*, 2217-2262.

[20] De Gennes, P. G. *Molecular Crystals and Liquid Crystals* 1971, *12*, 193.

[21] Mao, Y.; Warner, M.; Terentjev, E. M.; Ball, R. C. *Journal of Chemical Physics* 1998, *108*, 20.

[22] Martinoty, P.; Hilliou, L.; Mauzac, M.; Benguigui, L.; Collin, D. *Macromolecules* 1999, *32*, 1746.

[23] Degert, C.; Davidson, P.; Megtert, S.; Petermann, D.; Mauzac, M. *Liquid Crystals* 1992, *12*, 779.

[24] Pakula, T.; Zentel, R. *Makromolekulare Chemie* 1991, *192*, 2401.

[25] Sigel, R.; Stille, W.; Strobl, G.; Lehnert, R. *Macromolecules* 1993, *26*, 4226.

[26] Broer, D. J.; Hynderickx, I. *Macromolecules* 1990, *23*, 2474.

[27] Hasson, C. D.; Davis, F. J.; Mitchell, G. R. *Chemical Communications* 1998, *22*, 2515.

[28] Hasson, C. D.; Davis, F. J.; Mitchell, G. R. *Molecular Crystals and Liquid Crystals* 1999, *332*, 155.

[29] Kelly, S. M. *Journal of Materials Chemistry* 1995, *5*, 2047.

[30] Mao, Y.; Warner, M. *Physical Review Letters* 2000, *84*, 5335.

[31] Courty, S.; Tajbakhsh, A. R.; Terentjev, E. M. *Physical Review Letters* 2003, *91*, 085503-1.

[32] Marty, J.-D.; Labadie, L.; Mauzac, M.; Fournier, C.; Rico-Lattes, I.; Lattes, A. *Molecular Crystals and Liquid Crystals* 2004, *411*, 1603.

[33] Kempe, M. *Analytical Chemistry* 1996, *68*, 1948.

[34] Marty, J.-D.; Gornitzka, H.; Mauzac, M. *European Physics Journal E* 2005, *17*, 515.

[35] Palaprat, G.; Marty, J.-D.; Routaboul, C.; Lattes, A.; Mingotaud, A.-F.; Mauzac, M. *Journal of Physical Chemistry A* 2006, *110*, 12887.

[36] Palaprat, G.; Weyland, M.; T., P.; Binet, C.; Marty, J.-D.; Mingotaud, A.-F.; Mauzac, M. *Polymer International* 2006, *55*, 1191-1198.

[37] Wulff, G. *Chemical Reviews* 2002, *102*, 1.

[38] Brüggemann, O. *Analytica Chimica Acta* 2001, *435*, 197.

[39] Beach, J. V.; Shea, K. J. *Journal of the American Chemical Society* 1994, *116*, 379.

[40] Liu, J.; Mosbach, K. *Macromolecular Rapid Communications* 1998, *19*, 671.

[41] Visnjevski, A.; Yilmaz, E.; Brüggemann, O. *Applied Catalysis A: General* 2004, *260*, 169.

[42] Svenson, J.; Zheng, N.; Nicholls, I. A. *Journal of the American Chemical Society* 2004, *126*, 8554-8560.

[43] Liu, X.-C.; Mosbach, K. *Macromolecular Rapid Communications* 1997, *18*, 609.

[44] Biffis, A.; Wulff, G. *New Journal of Chemistry* 2001, *25*, 1537.

[45] Polborn, K.; Severin, K. *Chemistry. A European Journal* 2000, *6*, 4604.

[46] Locatelli, F.; Gamez, P.; Lemaire, M. *Journal of Molecular Catalysis A* 1998, *135*, 89.

[47] Lettau, K.; Warsinke, A.; Laschewsky, A.; Mosbach, K.; Yilmaz, E.; Scheller, F. W. *Chemistry of Materials* 2004, *16*, 2745-2749.

[48] Liu, J.; Wulff, G. *Angewandte Chemie, International Edition in English* 2004, *43*, 1287.

[49] Liu, J.; Wulff, G. *Journal of the American Chemical Society* 2004, *126*, 7452.

[50] Sagawa, T.; Togo, K.; Miyahara, C.; Ihara, H.; Ohkubo, K. *Analytica Chimica Acta* 2004, *504*, 37.

[51] Sellergren, B.; Karmalkar, R. N.; Shea, K. J. *Journal of Organic Chemistry* 2000, *65*, 4009.

[52] Kawanami, Y.; Yunoki, T.; Nakamura, A.; Fujii, K.; Umano, K.; Yamauchi, H.; Masuda, K. *Journal of Molecular Catalysis A* 1999, *145*, 107.

[53] Ohkubo, K.; Funakoshi, Y.; Urata, Y.; Hirota, S.; Usui, S.; Sagawa, T. *Journal of the Chemical Society, Chemical Communications* 1995, 2143.

[54] Rich, J. O.; Mozhaev, V. V.; Dordick, J. S.; Clark, D. S.; Khmelnitsky, Y. L. *Journal of the American Chemical Society* 2002, *124*, 5254.

[55] Liu, J.; Luo, G.; Gao, S. Z., K.; Chen, X.; Shen, J. *Journal of the Chemical Society, Chemical Communications* 1999, 199.

[56] Weyland, M.; Ferrère, S.; Lattes, A.; Mingotaud, A.-F.; Mauzac, M. *Liquid Crystals* 2008, *35*, 219-231.

[57] Finkelmann, H.; Wendorff, J. H. *Makromolekulare Chemie* 1978, *179*, 273.

[58] Küpfer, J.; Finkelmann, H. *Makromolekulare Chemie, Rapid Communications* 1991, *12*, 717.

[59] Kundler, I.; Finkelmann, H. *Macromolecular Chemistry and Physics* 1998, *199*, 677.

[60] Thomsen III, D. L.; Keller, P.; Naciri, J.; Pink, R.; Jeon, H.; Shenoy, D.; Ratna, B. *Macromolecules* 2001, *34*, 5868.

[61] Naciri, J.; Srinivasan, A.; Jeon, H.; Nikolov, N.; Keller, P.; Ratna, B. *Macromolecules* 2003, *36*, 8499-8505.

[62] Pichon, V.; Chapuis, F. *Analytica Chimica Acta* 2008, *622*, 48-61.

[63] Jenkins, A. L.; Uy, O. M.; Murray, G. M. *Analytical Chemistry* 1999, *71*, 373-378.

[64] Jenkins, A. L.; Yin, R.; Jensen, J. L. *Analyst* 2001, *126*, 798-802.

[65] Southard, G. E.; Van Houten, K. A.; Murray, G. M. *Macromolecules* 2007, *40*, 1395-1400.

[66] Le Moullec, S.; Bégos, A.; Pichon, V.; Bellier, B. *Journal of Chromatography A* 2006, *1108*, 7-13.

[67] Prasad, K.; Prathish, K. P.; Gladis, J. M.; Naidu, G. R.; Prasada Rao, T. *Electroanalysis (New York)* 2007, *19*, 1195-1200.

[68] Zhu, X.; Yang, J.; Su, Q.; Cai, J.; Gao, Y. *Annali di Chimica* 2005, *95*, 877-884.

[69] Pereira, L. A.; Rath, S. *Analytical Bioanalytical Chemistry* 2009, *393*, 1063-1072.

[70] Malosse, L.; Buvat, P.; Adès, D.; Siove, A. *Analyst* 2008, *133*, 588-595.

[71] Malosse, L.; Palmas, P.; Buvat, P.; Adès, D.; Siove, A. *Macromolecules* 2008, *41*, 7834-7842.

[72] Li, C.; Wang, C.; Guan, B.; Zhang, Y.; Hu, S. *Sensors and Actuators, B: Chemical Sensors and Materials* 2005, *107*, 411-417.

[73] Li, C.; Wang, C.; Wang, C.; Hu, S. *Analytica Chimica Acta* 2005, *545*, 122-128.

[74] Li, C.; Wang, C.; Wang, C.; Hu, S. *Sensors and Actuators, B: Chemical Sensors and Materials* 2006, *117*, 166-171.

[75] Zhu, X.; Yang, J.; Su, Q.; Cai, J.; Gao, Y. *Journal of Chromatography A* 2005, *1092*, 161-169.

[76] Zhu, X.; Cai, J.; Yang, J.; Su, Q.; Gao, Y. *Journal of Chromatography A* 2006, *1131*, 37-44.

[77] Taranekar, P.; Huang, C.; Advincula, R. C. *Polymer* 2006, *47*, 6485-6490.

[78] Alizadeh, T. *Electroanalysis (New York)* 2009, *21*, 1490-1498.

[79] Alves Pereira, L.; Rath, S. *Analytical Bioanalytical Chemistry* 2009, *393*, 1063-1072.

[80] Yang, Q.; Sun, Q.; Zhou, T.; Shi, G.; Jin, L. *Journal of Agricultural and Food chemistry* 2009, *57*, 6558-6563.

[81] Deppe, D. D.; Dhinojwala, A.; Torkelson, J. M. *Macromolecules* 1996, *29*, 3898-3908.

[82] Slark, A. T.; O'Kane, J. *European Polymer Journal* 1997, *33*, 1369-1376.

[83] Gué, A.-M.; Lattes, A.; Laurent, E.; Mauzac, M.; Mingotaud, A.-F. *Analytica Chimica Acta* 2008, *614*, 63.

[84]   Say, R. *Analytica Chimica Acta* 2006, *579*, 74.

[85]   Byrne, M. E.; Salian, V. *International Journal of Pharmaceutics* 2008, *364*, 188-212.

[86]   Janiak, D. S.; Kofinas, P. *Analytical Bioanalytical Chemistry* 2007, *389*, 399-404.

[87]   Takeuchi, T.; Hishiya, T. *Organic and Biomolecular Chemistry* 2008, *6*, 2459-2467.

[88]   Turner, N. W.; Jeans, C. W.; Brain, K. R.; Allender, C. J.; Hlady, V.; Britt, D. W. *Biotechnology Progress* 2006, *22*, 1474-1489.

[89]   He, D.; Sun, W.; Schrader, T.; Ulbricht, M. *Journal of Materials Chemistry* 2009, *19*, 253-260.

[90]   Koch, S. J.; Renner, C.; Xie, X.; Schrader, T. *Angewandte Chemie, International Edition in English* 2006, *45*, 6352-6355.

[91]   Hawkins, D. M.; Stevenson, D.; Reddy, S. M. *Analytica Chimica Acta* 2005, *542*, 61-65.

[92]   Hawkins, D. M.; Trache, A.; Ellis, E. A.; Stevenson, D.; Holzenburg, A.; Meininger, G. A.; Reddy, S. M. *Biomacromolecules* 2006, *7*, 2560-2564.

[93]   Janiak, D. S.; Ayyub, O. B.; Kofinas, P. *Macromolecules* 2009, *42*, 1703-1709.

[94]   Janiak, D. S.; Ayyub, O. B.; Kofinas, P. *Polymer* 2010, *51*, 665-670.

[95]   Kimhi, O.; Bianco-Peled, H. *Langmuir* 2007, *23*, 6329-6335.

[96]   Lin, H.-Y.; Hsu, C.-Y.; Thomas, J. L.; Wang, S.-E.; Chen, H.-C.; Chou, T.-C. *Biosensors and Bioelectronics* 2006, *22*, 534-543.

[97]   Su, W.-X.; Rick, J.; Chou, T.-C. *Microchemical Journal* 2009, *92*, 123-128.

[98]   Shi, H.; Tsai, W.-B.; Garrison, M. D.; Ferrari, S.; Ratner, B. D. *Nature (London)* 1999, *398*, 593-597.

[99]   Bossi, A.; Piletsky, S. A.; Piletska, E. V.; Righetti, P. G.; Turner, A. P. F. *Analytical Chemistry* 2001, *73*, 5281-5286.

[100]   Matsunaga, T.; Hishiya, T.; Takeuchi, T. *Analytica Chimica Acta* 2007, *591*, 63-67.

[101]   Dickert, F. L.; Hayden, O.; Lieberzeit, P.; Haderspoeck, C.; Bindeus, R.; Palfinger, C.; wirl, B. *Synthetic Metals* 2003, *138*, 65-69.

[102]   Matsunaga, T.; Takeuchi, T. *Chemistry Letters* 2006, *35*, 1030-1031.

[103]   Lu, Y.; Yan, C.-L.; Wang, X.-J.; Wang, G.-K. *Applied Surface Science* 2009, *256*, 1341-1346.

[104]   Ouyang, R.; Lei, J.; Ju, H. *Nanotechnology* 2010, *21*, 185502.

[105]   Zhang, M.; Huang, J.; Yu, P.; Chen, X. *Talanta* 2010, *81*, 162-166.

[106]   Hu, X.; Li, G.; Huang, J.; Zhang, D.; Qiu, Y. *Advanced Materials (Weinheim, Federal Republic of Germany)* 2007, *19*, 4327-4332.

[107]   Zhao, Y.; Zhao, X.; Pei, X.; Hu, J.; Zhao, W.; Chen, B.; Gu, Z. *Analytica Chimica Acta* 2009, *633*, 103-108.

[108]   Zhao, Y.-J.; Zhao, X.-W.; Hu, J.; Xu, W.-Y.; Gu, Z. *Angewandte Chemie, International Edition in English* 2009, *48*, 7350-7352.

[109]   Valvanuz Linares, A.; Vandevelde, F.; Pantigny, J.; Falcimaigne-Cordin, A.; Haupt, K. *Advanced Functional Materials* 2009, *19*, 1299-1303.

[110]   Demirel, G.; Ösçetin, G.; Turan, E.; Caykara, T. *Macromolecular Biosciences* 2005, *5*, 1032-1037.

[111]   Hoshino, Y.; Kodama, T.; Okahata, Y.; Shea, K. J. *Journal of the American Chemical Society* 2008, *130*, 15242-15243.

[112]   Hoshino, Y.; Koide, H.; Urakami, T.; Kanazawa, H.; Kodama, T.; Oku, N.; Shea, K. J. *Journal of the American Chemical Society* 2010, *132*, 6644-6645.

[113] Miyata, T.; Jige, M.; Nakaminami, T.; Uragami, T. *Proceedings of the National Academy of Sciences of the United States of America* 2006, *103*, 1190-1193.

[114] Uysal, A.; Demirel, G.; Turan, E.; Caykara, T. *Analytica Chimica Acta* 2008, *625*, 110-115.

[115] Hua, Z. D.; Chen, Z. Y.; Li, Y. Z.; Zhao, M. P. *Langmuir* 2008, *24*, 5773-5780.

[116] Derazshamshir, A.; Baydemir, G.; Andac, M.; Say, R.; Galaev, I. Y.; Denizli, A. *Macromolecular Chemistry and Physics* 2010, *211*, 657-668.

[117] Casey, B. J.; Kofinas, P. *Journal of Biomedical Materials Research Part A* 2008, *87*, 359-363.

[118] Tan, C. J.; Tong, Y. W. *Langmuir* 2007, *23*, 2722-2730.

[119] Hawkins, D. M.; Ellis, E. A.; Stevenson, D.; Holzenburg, A.; Reddy, S. M. *Journal of Materials Science* 2007, *42*, 9465-9468.

[120] Takeuchi, T.; Goto, D.; Shinmori, H. *Analyst* 2007, *132*, 101-103.

[121] Kumar, A.; Whitesides, G. M. *Applied Physics Letters* 1993, *63*, 2002-2004.

[122] Mrksich, M.; Whitesides, G. M. *Trends in Biotechnology* 1995, *13*, 228-235.

[123] Gates, B. D.; Whitesides, G. M. *Journal of the American Chemical Society* 2003, *125*, 14986-14987.

[124] Offenhausser, A.; Bocker-Meffert, S.; Decker, T.; Helpenstein, R.; Gasteier, P.; Groll, J.; Moller, M.; Reska, A.; Schafer, S.; Schulte, P.; Vogt-Eisele, A. *Soft Matter* 2007, *3*, 290-298.

[125] Perl, A.; Reinhoudt, D. N.; Huskens, J. *Adv. Mater.* 2009, *21*, 2257-2268.

[126] Qin, D.; Xia, Y. N.; Whitesides, G. M. *Nat. Protoc.* 2010, *5*, 491-502.

[127] Cattoni, A.; Cambril, E.; Decanini, D.; Faini, G.; Haghiri-Gosnet, A. M. *Microelectronic Engineering* 2010, *87*, 1015-1018.

[128] Hamouda, F.; Barbillon, G.; Gaucher, F.; Bartenlian, B. *Journal of Vacuum Science and Technology B* 2010, *28*, 82-85.

[129] Hamouda, F.; Barbillon, G.; Held, S.; Agnus, G.; Gogol, P.; Maroutian, T.; Scheuring, S.; Bartenlian, B. *Microelectronic Engineering* 2009, *86*, 583-585.

[130] Koo, N.; Plachetka, U.; Otto, M.; Bolten, J.; Jeong, J.-H.; Lee, E.-S.; Kurz, H. *Nanotechnology* 2008, *19*, 225304/1-4.

[131] Moran, I. W. M.; Cheng, D. F.; Jhaveri, S. B.; Carter, K. R. *Soft Matter* 2008, *168*, 168-176.

[132] Plachetka, U.; Bender, M.; Fuchs, A.; Vratzov, B.; Glinsner, T.; Lindner, F.; Kurz, H. *Microelectronic Engineering* 2004, *73-74*, 167-171.

[133] Viheriälä, J.; Tommila, J.; Leinonen, T.; Dumitrescu, M.; Toikkanen, L.; Niemi, T.; Pessa, M. *Microelectronic Engineering* 2009, *86*, 321-324.

[134] Ye, X.; Ding, Y.; duan, Y.; Liu, H.; Lu, B. *Journal of Vacuum Science and Technology B* 2009, *27*, 2091-2096.

[135] Choi, S.-J.; Yoo, P. J.; Baek, S. J.; Kim, T. W.; Lee, H. H. *Journal of the American Chemical Society* 2004, *126*, 7744-7745.

Chapter 3

MOLECULAR RECOGNITION AND CRYSTAL GROWTH

J. S. Redinha[1] and A. J. Lopes Jesus[1,2]

[1] Department of Chemistry, University of Coimbra, Coimbra, Portugal
[2] Faculty of Pharmacy, University of Coimbra, Coimbra, Portugal.

ABSTRACT

The concept of molecular recognition in supramolecules with different types of intermolecular interactions and in some biological processes is discussed. The features and manifestations of hydrogen bonding as one of the most important types of interactions participating in the molecular recognition are presented in geometric, spectroscopic and natural bond orbitals terms. Molecular recognition in co-crystals and polymorphs is object of discussion in order to give a general view of this matter in different technologies, particularly in the pharmaceutical one. The role of molecular conformation and association in solution in predicting crystalline structures is investigated. Attention is also given to the molecular recognition in solid/solution interfaces in some important processes of crystal growth.

1. INTRODUCTION: SUPRAMOLECULAR CHEMISTRY AND MOLECULAR RECOGNITION

In nature, a certain number of atoms join together by covalent bonds giving rise to molecules. Such bonds, consisting of electron sharing between the atoms involved, withdraw their discrete properties from those of the molecules. The quantum mechanical forces between the atoms, although varying in a wide range of strengths, are always a few hundreds of thermal energy, *kT*. Furthermore, these bonds are oriented at well defined angles and their number is determined by the electronic structure of the atoms.

Molecules in turn can be assembled into supramolecular structures by other type of forces without loosing their individuality. The intermolecular non-covalent bonds are fundamentally of electrostatic origin. The lack of a comprehensive theory for these interactions leads to their subdivision into different types: ionic, hydrogen bonding, dipole-dipole, dipole-induced

dipole, van der Waals interactions, π-π stacking, etc. As a first approximation their energies can be calculated by the classical electrostatic laws providing that the spatial distribution of the charges are known. All these interactions are much weaker than the covalent ones and they are also less specific and more flexible as far as orientation is concerned.

Supramolecular chemistry encompasses practically all chemical fields and plays a fundamental role in the interpretation of chemical and biochemical systems and processes [1,2]. Thus, the knowledge of non-covalent interactions between molecules or molecular groups has increased exponentially.

The replication of DNA is possible because the two helical strands are connected by hydrogen bonds. The leading motives for the separated strands to template a new supramolecule is the structural compatibility allowing the reconstruction of the primitive hydrogen bonding network. The protein folding that transforms an inert macromolecule into a living polypeptide is the result of internal non-covalent interactions between complementary sites. As scientists became familiar with a great number of such phenomena, they were encouraged to apply the observed motives to chemical synthesis. The synthesis of crown ethers by Pedersen in 1967 [3] initiated a new era in the supramolecular chemistry. The Nobel Prize of Chemistry awarded in 1987 to Donald J. Cram, Jean-Marie Lehn and Charles J. Pederson was the recognition of the perspectives opened by the research in that pioneering time [4].

A great number of molecular organic crystals are supramolecules. Let us take β-D-glucose as an example. A single crystal of this compound is an assembly of molecules ordered in such a way that four hydroxyl groups and the oxygen atom of the pyranose ring are at 1.69-1.78 Å of the polar groups of neighbour molecules, Figure 1(a). These distances are well within the criteria commonly accepted for hydrogen bonding [5]. Three hydroxyl groups act simultaneously as hydrogen donor and acceptor; the hydroxyl group connected the anomeric carbon as donor and the oxygen of the pyranose ring as acceptor, Figure 1(b). The position of the OH group connected to carbon four does not allow its interaction with any polar group of a neighbouring molecule. In a crystal, the molecules are self-assembled in such a way that the interactions are maximized, leading to a lower energy state of the supramolecule.

Carboxylic acids are strongly associated in solution by interaction of the –COOH groups. The hydroxyl of one molecule is the hydrogen donor to the carbonyl of the partner and *vice-versa*. The resulting structure is depicted in Figure 2(I). The structure originated by these acids in solution has been observed for the first time by Wolf [6] who used the term supramolecule to describe the type of structure formed.

The structure pattern I, also observed in crystals built by molecules containing carboxyl groups, is designated as a synthon. This term was introduced by Corey [7] in molecular chemistry and it was defined by the author as "structural units within molecules which can be formed and/or assembled by known or conceivable synthetic operations". This concept was later adapted to supramolecular chemistry by Desiraju [8] by just changing the word molecules by supramolecules and synthetic operations by intermolecular interactions.

Synthon I, the simplest supramolecule of a carboxylic acid, still persists in crystals, a rather complex supramolecular arrangement. The synthon is therefore a structural characteristic of a supramolecule. It is very useful in crystal engineering as far as it relates the molecule to the supramolecule and allows understanding of the structure of the latter.
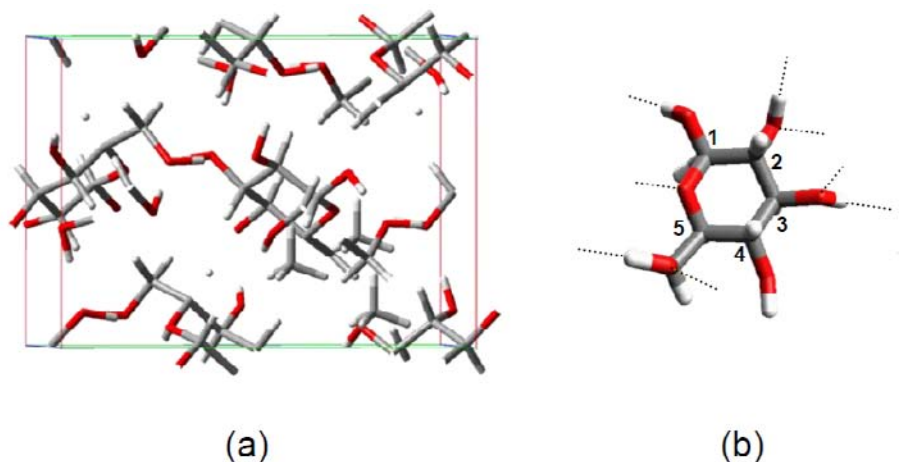
Figure 1. Unit cell of the crystalline structure of β-D-glucose (a) and detail of the intermolecular hydrogen bonds established by each molecule (b).
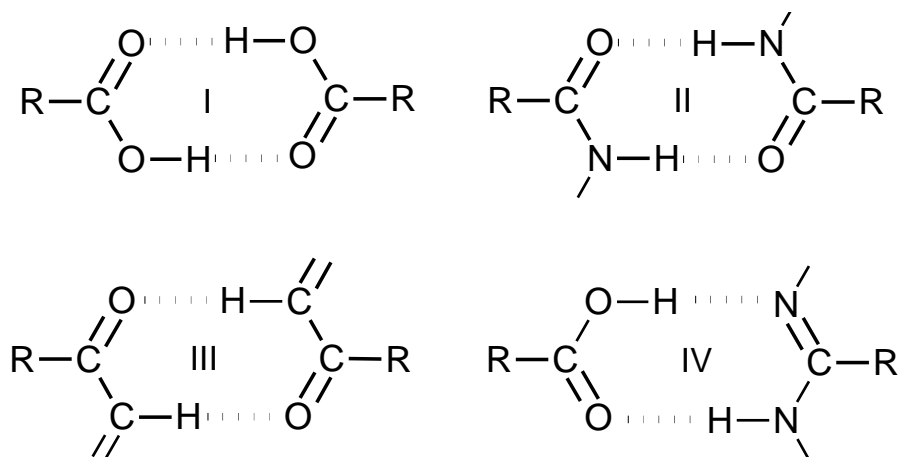

Figure 2. Examples of synthons in supramolecular chemistry.

In Figure 2 are illustrated some common supramolecular synthons. They are called homosynthons (I, II and III) or heterosynthons (IV), depending on whether the association involves identical or different functional groups, respectively. The NH and OH groups as hydrogen bond donors and oxygen and nitrogen as acceptors are important participants in many synthons. In some others, there is also the involvement of weak bonds such as CH···O (III).

One of the best known supramolecules in biological systems is deoxyribonucleic acid (DNA). It is formed by two macromolecular chains that are constituted by nucleotide monomers of deoxyribose joined by phosphodiester bridges. Bonded to each deoxyribose there is a purine or pyrimidine base. The link between the two DNA strands is established by hydrogen bonds between the bases, as illustrated in Figure 3. Cytosine is paired with guanine and thymine with guanine. The unaltered structure of the strands stores the genetic information that depends on the base composition and sequence. The genetic transmission is

due to the possibility that both strands can be separated and no other synthon can be formed by the bases besides those just referred.
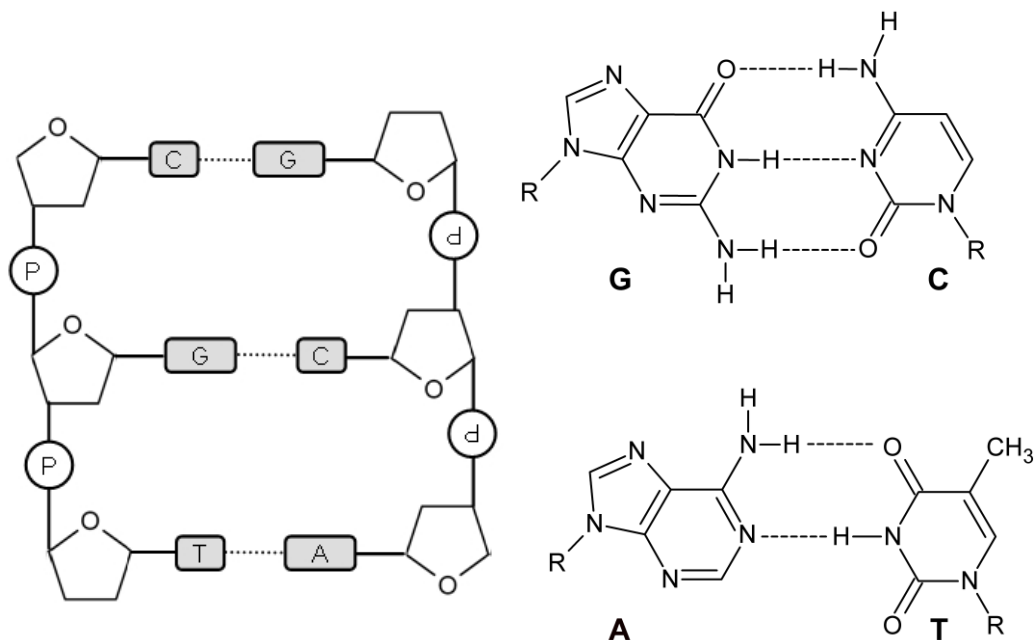
Figure 3. Primary structure of DNA (left) and hydrogen bonds between the guanine-cytosine and adenine-thymine base pairs (right).

The molecules of the amphiphilic compounds such as surfactants tend to be associated in aqueous solutions. Even at low concentrations they are self-assembled into micelles. These are supramolecules in which the molecules are held together by their non-polar chains, being the polar groups in the micelle/water interface. In Figure 4 is depicted a micelle of a negative surfactant in a dilute electrolyte aqueous solution. The negative charges at the surface are partially neutralized by counter-ions distributed across the particle-solution interface. Since the micelle has a colloidal size and the surface possesses an excess of negative charges, particle aggregation is avoided and a stable colloidal dispersion is obtained.

The explanation for this type of structure organization lies in the behaviour of non-polar and polar groups towards water. Thermodynamically, the solvation of both groups is accompanied by a decrease of enthalpy and entropy. However, while for the polar groups $|\Delta H^0| > |T\Delta S^0|$, hence $\Delta G^0 < 0$, for the non-polar ones $|\Delta H^0| < |T\Delta S^0|$ and $\Delta G^0 > 0$.

The unfavourable hydration entropy of the non-polar part tends to reduce as much as possible its contact with water. Molecular association accomplishes this target. In the micelle, the non-polar chains are associated originating a particle surrounded by solvent molecules. An explanation for the behaviour of the non-polar groups in aqueous solutions, either in thermodynamics or molecular terms, can be found elsewhere [9,10].

Multidentate ligands became the focus of interest after the discovery of macrocyclic polyethers, termed crown ethers, by Pedersen in 1967 [3,11]. They consist of a ring containing ether groups. The repeating unit in the most common crown ether is the

ethyleneoxy group. The crown ethers are ligands of cations forming complexes by interaction of the oxygen with the ion located in the interior of the ring (Figure 5).
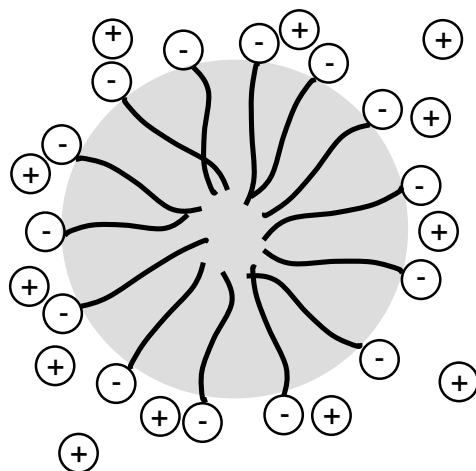
Figure 4. Schematic representation of a micelle formed by an anionic surfactant in a dilute electrolyte solution.

K^+/18-crown-6 (D_{3d}) Fe^{2+}/porphine (D_{4h})
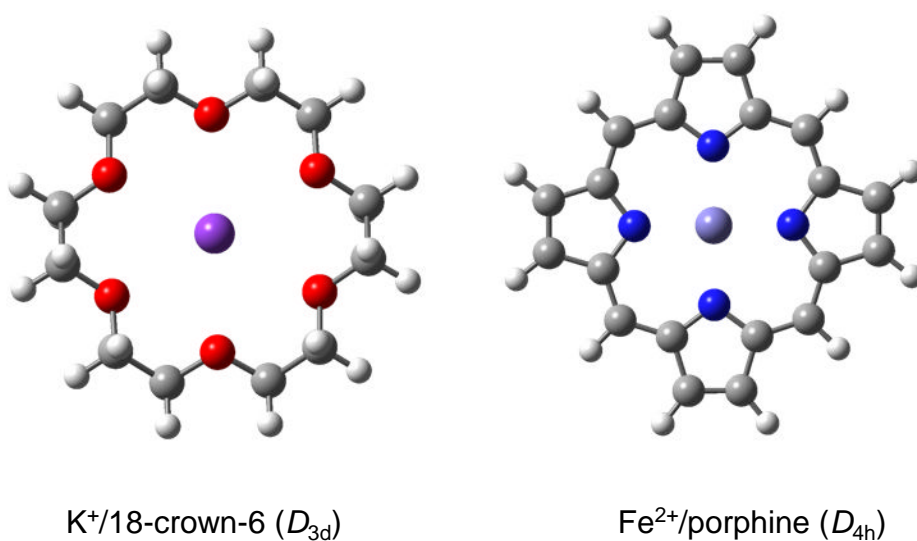
Figure 5. Structures of crown ether/potassium and porphine/iron complexes obtained by gas-phase optimization at the B3LYP/6-31G* level of theory.

Selectivity in complexation is based on the ratio of the radius of the ring and that of the ion. Since the ring can be tailored as desired, a diversity of crown ethers can be obtained according to the problem in hand. For example, a good complexing agent for the potassium ion is 18-crown-6 (18 stands for the number of the atoms of the ring and 6 for the number of oxygen atoms). The cavity size of this crown ether (radius 1.3-1.6 Å) is well adjusted to a

potassium ion (ionic radius 1.33 Å). This crown ether has a poor selectivity for sodium ion (ionic radius 0.95 Å). The adequate ligand for this ion is 15-crown-5 (ring size 0.85-1.1 Å).

An insight into the interaction between the oxygen and the metal ion can be obtained from Natural Bond Orbital Analysis [12,13]. An NBO calculation at the B3LYP/6-31G* level of theory reveals that in K^+/18-crown-6, electrons from one of the oxygen lone-pairs are transferred to a vacant s orbital of the metal ion. Oxygen has two lone-pairs, one that is practically pure p and the other that is $sp^{1.31}$. The latter is that involved in the interaction with potassium. The donor-acceptor energy per bond is 5.15 kJ mol^{-1}. The oxygen-metal bond is 43.27% of s-character and 56,68% of p-character; which means that the iconicity of the bond is low.

Crown ethers find applications in a wide range of chemical operations such as ion transport and separations [14,15], chemosensors for ion or neutral analytes detection [16] and in phase transfer catalysis [17].

Porphyrins are other representative structures of macrocyclic ligands containing four pyrrole groups linked by four methine bridges. They represent a remarkable group of ligands used as catalysts of oxidation and photooxydation organic reactions. Porphine, the simplest member of porphyrins (Figure 5), is able to form complexes with many metals. These complexes are particularly important as they constitute the heme group of hemoglobin and myoglobin (Fe-porphine) and the basic structure of the chlorophyll molecule (Mg-porphine).

From the examples of supramolecules just shown we can conclude that packing brings the molecular functional groups to positions that fulfil the conditions required for the establishment of non-covalent intermolecular bonds. The conditions deal with the complementary nature of the groups involved, distance between them and their relative orientations. Obviously, the interacting group should be complementary in nature: a positively charged group is attached to a negatively charged one; a hydrogen bond is only possible if one of the groups is a hydrogen donor and the other a hydrogen acceptor. It is also evident that the interaction is only possible if the groups are at a distance within the range of the attractive forces to be set up. The mutual orientation of the intervening groups is a very important factor and may play a decisive role. The maximum value for the interaction between two dipoles is reached when they are parallel and is minimum when they are anti-parallel. A X–H···Y hydrogen bond is only significant on the energetic point of view when the angle formed by the three atoms is higher than, say, 110º [5]. Although the dispersion forces between the molecules depend on the mutual orientation of the groups, this effect is in general insignificant [18] due to the anisotropy of the bond electronic polarizability.

The interaction between molecules through groups satisfying the conditions enunciated above is called molecular recognition. The concept is extended to the interactions between groups of the same molecule originating internal bonds.

Molecular recognition is a basic physico-chemical concept of utmost importance in all branches of chemistry. As will be presented later on in this chapter, the role played by molecular recognition in recent fields of chemistry with great technological impact reinforces the scientific interest for this matter. It is also the key to understanding a great number of relevant biochemical processes and provides useful data to design new products in chemical and biochemical technologies.

2. MOLECULAR RECOGNITION BY BIOLOGICAL RECEPTORS

2.1. DRUG ACTION MECHANISM

The cell walls of some biological tissues contain specific groups capable of linking molecules of toxins leading to undesirable effects in living organisms [19,20]. These groups are called receptors. Chemically, a receptor is a polypeptidic macromolecule containing side chains with groups able to bind messenger molecules. There are polar groups that have a complementary structure to that of the messenger.

As an example, the receptors in cardiac tissue interact with catecholamines giving a "complex" that increases the rate and strength of the heart contraction and consequently the blood pressure. The interaction between the *S*-(+)-epinephrine and a β receptor is schematized in Figure 6(a). The phenyl group and the cationic form of the amine are bonded to the aryl and carboxylate ion by aromatic stacking and ionic hydrogen bonds interactions, respectively. The molecule linked to a receptor giving rise to a pharmacological response is called agonist.
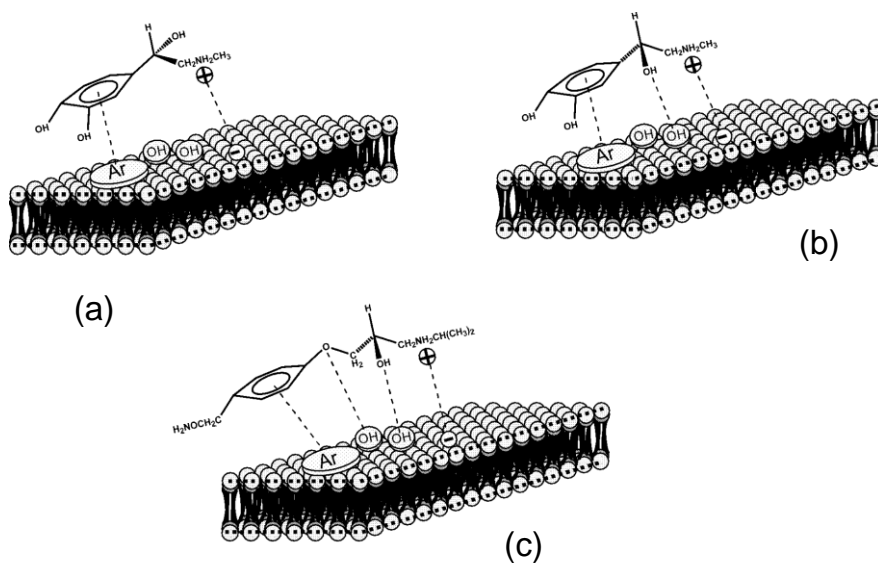
Figure 6. Interaction scheme of *S*-(+)-epinephrine (a) and of *R*-(-)-epinephrine (b) with a β receptor. (c) Interaction of the beta-adrenergic blocker atenolol with the receptor.

R-(-)-epinephrine can also interact with the β receptor but the OH group bonded to the asymmetric carbon atom is in such a position that allows its interaction with a hydroxyl group of the receptor through a hydrogen bond, Figure 6(b). That is, whilst a two-site agonist-receptor attachment is not able to discriminate enantiomers, a three-site attachment does.

The undesirable action of an agonist can be prevented by the presence of a molecule with a structure suitable to be linked to the active centres of the receptor available for the interaction with the agonist. Two requisites have to be fulfilled: the affinity of the receptor towards the molecule of the compound has to be higher than that corresponding to the agonist; no nocive effect can be initiated by the compound used. This compound is an

antagonist because it occupies the receptor sites, preventing its interaction with the agonist. For example, atenolol, 4-[2-hydroxy-3-[(1-methylethyl) amino] propoxy]- benzeneacetamide, because of the similitude of the structure of the hydroxylamine moiety with that of epinephrine, has an antagonist effect and is a drug used in the treatment of arterial tension. The attachment of atenolol to a β receptor is depicted in Figure 6(c).

Molecular chirality has a great importance in drug action [21]. For example, it is known that single-enantiomer formulations can provide greater selectivity for their biological targets. For this reason, the pharmaceutical industry has been replacing racemic mixtures for pure enantiomeric compounds, racemic switch [22]. In biological systems, the selectivity of a receptor for an agonist or antagonist is controlled by molecular recognition that plays a fundamental role in the activity of the drug and gives valuable contributions for the design of new ones.

2.2. SWEETNESS PERCEPTION

Numerous vital processes in living organisms become possible by molecular recognition of ligands by biological receptors. Great effort has been made to understand the complex mechanism of the ligand-receptor interaction. Just to have an idea about the research going on in this field, we leave here a short note about the relationship between the sweet taste and the structure of the sweetener. This matter has importance in the discovery of sweeteners by reasons of flavour and low calorie value.

Shallenberg and Acre [23] advanced a theory on sweetness perception according to which the molecular recognition is a consequence of two hydrogen bonds established between AH and B sites, localized either in the receptor or in the sweetener. The group AH of the sweetener is hydrogen donor to B of the receptor and *vice versa*. AH and B in any of the partners should be at 2.5–4.0 Å apart. The AH, B or two-point attachment theory gained acceptance and had the merit to arise interest for the sweet chemoreception.

It was noticed earlier on that this theory was only approximate. So far, it did not account for chirality. For example, in most of the aminoacids one enantiomer is sweet and the other tasteless or bitter. In some cases both enantiomers are sweet while in others both have a bitter flavour [24]. Furthermore, erythritol and threitol are 1,2,3,4-butanetetrol diatereomers in which the distance between the OH groups is identical. However, the former is sweet while the latter is bitter. The main difference between both structures lies in the configuration of the asymmetric carbon atoms: (*R*,*S*)- in erythritol and (*R*,*R*)- or (*S*,*S*)- in threitol. A third site was then introduced in the receptor accounting for chirality [25]. In this three-site model the third point is localized at 3.5 Å from A, 0.55 Å from B; A and B would be at 2.6 Å apart.

After intense research activity Nofre and Tinti concluded that the flavour receptor should be provided with a set of several interaction sites and proposed a multi-point-attachment (MPA) theory [26]. In humans, a set of eight sites, corresponding to the same number of aminoacids, and designated as B, AH, XH, G1, G2, G3 and G4 and D were admitted. All recognition sites but D can be unfolded into two interaction points or subsites, as the receptor is activated by a sweetener. The interaction sites, interaction points, recognition sites, as well as the type of interactions are indicated in Table 1.

Table1. Sweetener interaction sites, interaction points, interactions and receptor recognition sites [27].

Interaction Sites	Interaction points	Interactions	Recognition sites
B	B1	Ionic and/or	Lys
	B2	H-bonding	
AH	AH1	Ionic and/or	Asp/Glu
	AH2	H-bonding	
XH	XH1	Ionic and/or	Asp/Glu
	XH2	H-bonding	
G1	G1	Steric	Thr
	E1	H-bonding	
G2	G2	Steric	Thr
	E2	H-bonding	
G3	G3	Steric	Thr
	E3	H-bonding	
G4	G4	Steric	Thr
	E4	H-bonding	
D	D	H-bonding	Ser/Thr

B1 and B2 are assumed to be the hydrogen atoms of lysine ε-NH$_3^+$ group. AH1, AH2, XH1 and XH2 the oxygen atoms of the carboxylate group of aspartic or glutamic acids. E1, E2, E3 and E4 are OH groups of CH$_3$CHOH of threonine residue side chains. G1, G2, G3 and G4 are the methyl groups of the side chain of threonine residue. D can be the side chain of a serine (CH$_2$OH) or threonine (CH$_3$CHOH).
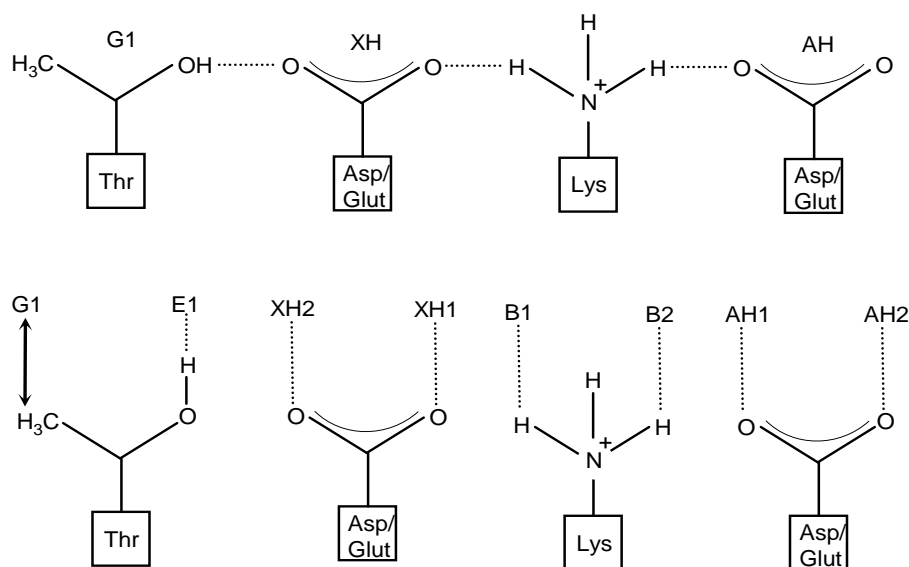
Figure 7. Sweetener-receptor molecular recognition in man. Top: Receptor fragment in rest; Bottom: After interaction with sweetener.

A fragment of the receptor at rest is shown in Figure 7. The sites are interconnected, particularly through hydrogen bonds. When the receptor is activated by the sweetener these bonds are broken and the groups are released and able to recognize the subsites of the sweet molecule.

According to Nofre and Tinti theory, the interaction between erythritol and the receptor is sketched in Figure 8. An identical picture is suggested for glucose, a B1, B2, AH1, AH2, XH1 and XH2 – type sweetener involving the OH groups 1 to 4. The sweetener molecule is recognized by three sites of the receptor and six hydrogen bonds are established between both entities. In sucrose, besides an identical recognition regarding the glucose unit there are four E, G pairs of sites accounting for the interaction with the pentose moiety [27].
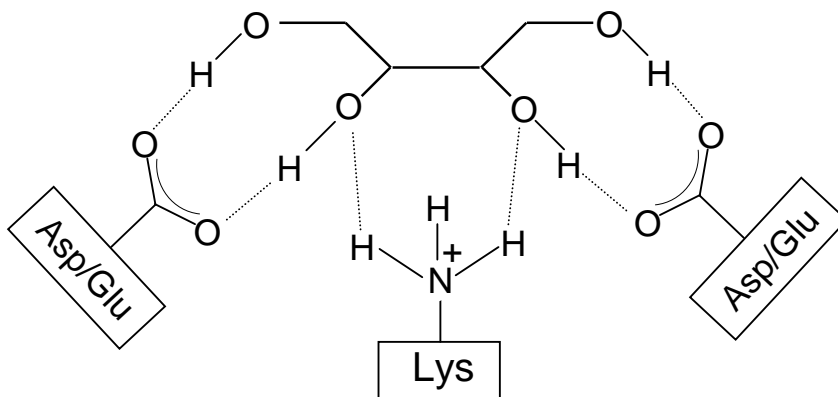
Figure 8. Molecular recognition of erythritol by the taste receptor.

3. HYDROGEN BONDING IN MOLECULAR RECOGNITION

As shown throughout this chapter, molecular recognition is accomplished by different types of forces according to the nature of the interacting sites. However, hydrogen bonding is by far the most frequent interaction. Indeed, the molecules of many organic compounds have polar groups capable of acting as a hydrogen donor or acceptor. Furthermore, the intermediate position of hydrogen bonding regarding the properties of the different types of interactions reinforces its importance. The knowledge of the main features of hydrogen bonding is therefore very useful in the study of molecular recognition. For this reason, this section is devoted to this type of interaction and to its detection by infrared spectroscopy.

3.1. GENERAL FEATURES OF HYDROGEN BONDING

A conventional hydrogen bond may be described as an attractive interaction between a hydrogen connected to an electronegative atom (X) and a lone-pair bearing atom (Y). This interaction can be represented as follows:

$$X–H + {:}Y \rightarrow X–H{\cdots}{:}Y$$
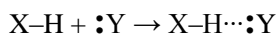
The energy of the cluster resulting from the hydrogen bond is stabilized by resonance effect:

$$X–H \cdots {:}Y \leftrightarrow {:}X^{-} \cdots H–Y^{+}$$
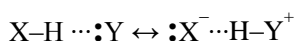
Using a Natural Bond Orbital (NBO) language [12,13], the complex formed by the establishment of a hydrogen bond results from an electron delocalization from a filled orbital (lone pairs of Y) to the antibonding X–H orbital (σ*XH). This effect can be considered as an electronic donor–acceptor process represented by $n(Y) \rightarrow \sigma*XH$. The energetic stabilization due to this donor–acceptor interaction can be estimated by second-order perturbation theory, providing an estimation for the hydrogen bond strength [12,13]. It is to be noted that the electron donor corresponds to the hydrogen acceptor group in the classical description and the electron acceptor to the hydrogen donor group. In Figure 9 is displayed the most stable conformer of 1,4-butanediol. This structure is stabilized by an O_A–H···O_D intramolecular hydrogen bond [28,29].

The lone pair, $n(O_D)$, of the oxygen atom of higher p-character interacts with the antiligant orbital of the O_AH bond, σ*O_AH. The overlap of both orbitals is quite perceptible in Figure 9. The stabilizing energy of the intramolecular hydrogen bond, calculated by second order perturbation, amounts to -42.8 kJ mol^{-1}, value high enough to bend the molecular backbone into a seven-member ring conformation.
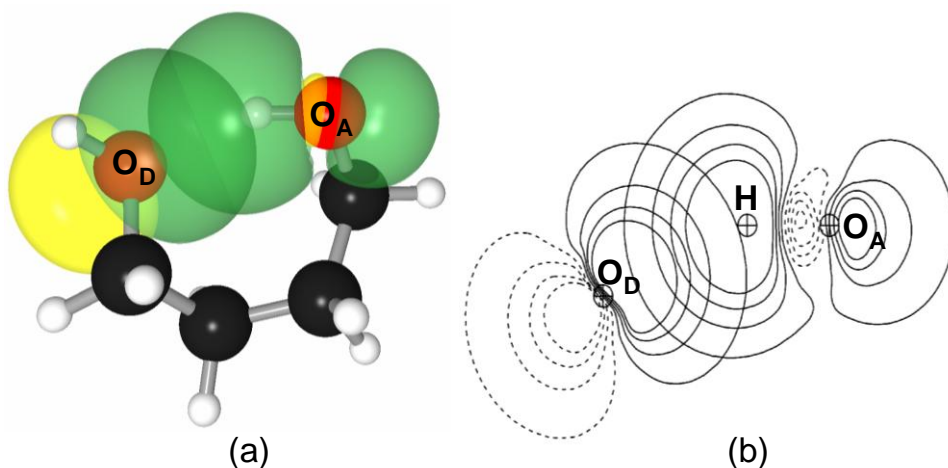
Figure 9. (a) 3D view of the of the preorthogonalized NBOs in the most stable conformer of 1,4-butanediol, showing the overlap of the stronger donor lone pair of the electron donor oxygen with the σ*O_AH orbital. (b) Contour plot of both orbitals in the O_DHO_A plane.

The $n(O_D) \rightarrow \sigma*O_A$H electron transference originates a rehybridization/repolarization of the σO_AH orbital, that in turn induces a variation of the hydrogen atom charge. To explain this variation in the case we are dealing with, let us consider the composition of σOH and σ*OH orbitals in terms of hybrid orbital components when the group is free and when it is bonded. The hybrid composition of the free σO_DH = 0.8581(sp$^{3.71}$)$_O$ + 0.5135(s)$_H$ and σ*OH = 0.5135(sp$^{3.71}$)$_O$ - 0.8581(s)$_H$. The establishment of the hydrogen bond leads to the following hybrids' composition: σO_AH = 0.8681(sp$^{3.20}$)$_O$ + 0.4963(s)$_H$ and σ*O_AH = 0.4963(sp$^{3.20}$)$_O$ - 0.8681(s)$_H$. The σO_AH orbital is repolarized towards the oxygen atom (the polarization coefficient increases from 0.8581 to 0.8681) and the p-character of the oxygen atom decreases from 78.65 to 76.05%. This repolarization reduces the electron population at the hydrogen atom by $(2e)[(0.5135)^2-(0.4963)^2)]=0.034e$. On the other hand, the charge transference to the σ*OH orbital increases the electron population at the hydrogen atom by

$0.025e$ $(0.8681)^2 = 0.019e$, where $0.025e$ is the occupancy of the σ^*O_AH orbital. Therefore, the net electron population reduction at the hydrogen atoms is $0.015e$. This is in agreement with the fact that the natural charge for the hydrogen atom involved in the hydrogen bond is $+0.491$, whereas that of the free OH group is $+0.469$. The charge increase of the hydrogen atom is, therefore, a manifestation of its participation in a hydrogen bond.

The NBO theory provides a suggestive and useful description of the hydrogen bond. Due to the s-character of the hydrogen atom and non-existence of an internal core in this atom, the H-terminus of the σ^*XH orbital is a nodeless spherical lobe of high amplitude. These features confer to this bond the unique character that we have been emphasizing. Indeed, the σ^*XH orbital is capable of overlapping with the lone pair orbitals of Y in a wide range of directions around the X–H axis. That is, the hydrogen bond has maximum overlap along the H–X direction but also exhibits significant overlaps in directions with pronounced deviations from this axis [12]. Moreover, the charge transference from the lone pair of Y to the antibonding σ^*XH orbital increases the occupancy of this orbital, which causes the weakening of the covalent X–H bond and consequently its elongation. This explains some geometrical and spectroscopic manifestations of the hydrogen bonding.

From the geometrical point of view, hydrogen bonding is characterized by an increase of the X–H length relatively to a free X–H bond. Also, the H···Y distance is shorter than the sum of the van der Waals radii of the H and Y atoms, and X–H···Y angle that allows the orbitals to overlap. The following geometrical criteria are commonly used in the hydrogen bonding diagnostic: H···Y distance smaller than 3.0 or even 3.2 Å and X–H···Y angle higher than 110º [5,30]. Certainly, these cut-off limits are arbitrary as the hydrogen bond energy varies from > 63 kJ mol^{-1} for strong bonds to very low values that are hardly distinguishable from a van der Waals interaction [31].

So far we have been considering a hydrogen bond established between uncharged donor–acceptor groups. However, hydrogen bond systems in which one of the pairs is charged are quite frequent. One example is betaxolol hydrochloride (Figure 10), a drug used as β-adrenergic antagonist in cardiovascular diseases and glaucoma treatment. To increase its solubility, it is available as hydrochloride.

The X-ray diffraction analysis shows that each chloride ion establishes three intermolecular hydrogen bonds: two of them with the NH_2^+ and OH groups of the same molecule and the other with the NH_2^+ group of a second molecule [32]. A detail of the hydrogen bond network in the crystalline structure of betaxolol hydrochloride is shown in Figure 10(b). They are hydrogen bonds as the hydrogen atoms of the donor groups participate in the bonds. The presence of the charges increases the bonds strength and for this reason they are called ionic hydrogen bonds or charge assisted hydrogen bonds [33,34]. This type of hydrogen bond is frequent in biological media since the polar groups of many biomolecules have acid-base behaviour and may be ionized at the physiological pH.

Weak interactions such as N–H···π, O–H···π, C–H··· O and C–H···π, that may play a role in protein folding, protein-ligand recognition and enzymatic activity, are considered as hydrogen bonds by some authors [31].
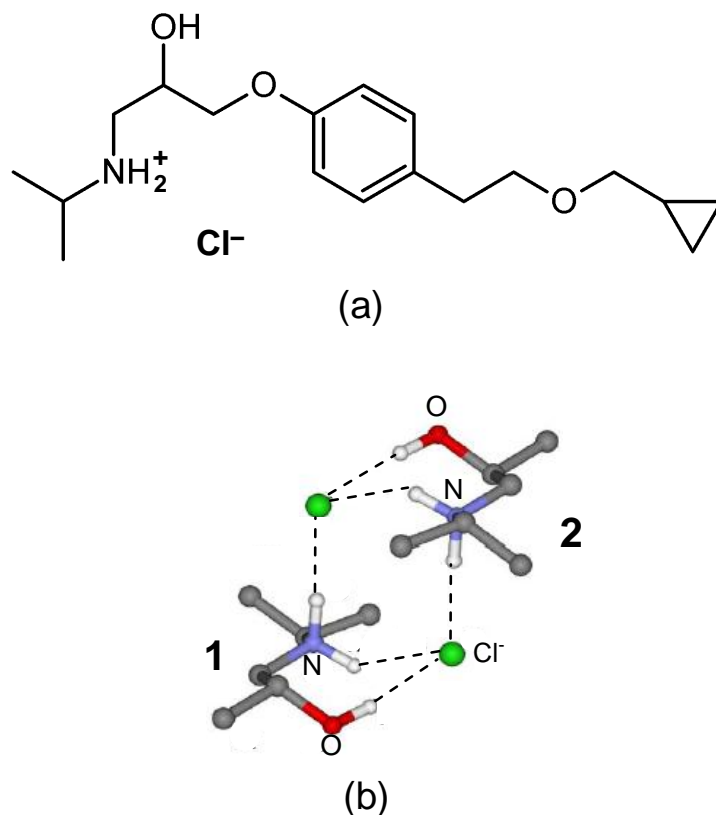
(a)

(b)

Figure 10. Structure of betaxolol hydrochloride (a) and detail of intra and intermolecular hydrogen bonds in the crystalline structure (b). For visualization simplicity only the molecular fragments of molecules 1 and 2 involved in hydrogen bonds are shown.

3.2. INFRARED SPECTROSCOPY AND HYDROGEN BONDING

Infrared spectroscopy is a very useful method for identification and characterization of hydrogen bonding. This interaction affects the bands corresponding to the participating groups, particularly those of the X–H donor group. The stretching vibration modes of this group are very sensitive to hydrogen bonding and therefore the corresponding absorption bands localized in the 4000-2500 cm^{-1} spectral region give valuable informations about this interaction. Useful data can also be obtained from the bands assigned to the bending vibration modes (<1600 cm^{-1}). However, in this spectral region, one can find bands of practically all molecular groups, making band assignment a very hard task. Our ongoing discussion about the infrared as a tool to study hydrogen bonding is confined to the analysis of the stretching region of the functional groups.

The interaction of the donor X–H and the acceptor Y is accompanied by the elongation and weakness of this bond. The corresponding stretching vibration band is then shifted to lower frequencies (red shift) relative to its position when X–H is free. The dipole moment of X–H is modified by hydrogen bonding owing to the elongation and repolarization of the bond as explained before. Since the intensity of an infrared absorption band is proportional to the

derivative of the dipole moment with respect to the normal coordinate that describes the vibration, the band intensity is drastically increased by this interaction [35]. Effort has been made to correlate shift and intensity with the enthalpy of the bond and various equations have been proposed [36-38].

In short, a hydrogen bond manifests itself in the stretching vibration spectra by a red shift of the X–H band and an increase of its width and intensity (area under the band profile).

In compounds with more than one functional group, the spectral interpretation is difficult due to the bands overlapping. Often, band assignment is made even more difficult by the Fermi resonance effect. Vibrational anharmonicity gives rise to bending overtones or combination bands whose frequency is close to the fundamental stretching vibration band. The coupling between the overtone or combination band and the fundamental one originates two bands at higher and lower frequency relatively to that expected for the fundamental band. This phenomenon is known as the Fermi effect and renders the spectra more complex.

Let us exemplify the methodology used in the characterization of overlapped bands by considering erythritol (Figure 11). At 298.15K, the infrared spectrum of the solid compound reveals the existence of three broad overlapped bands in the 3800-2500 cm^{-1} spectral region, Figure 12(a): one maximum at around 3250 cm^{-1} and two shoulders at higher and a lower frequencies [39].
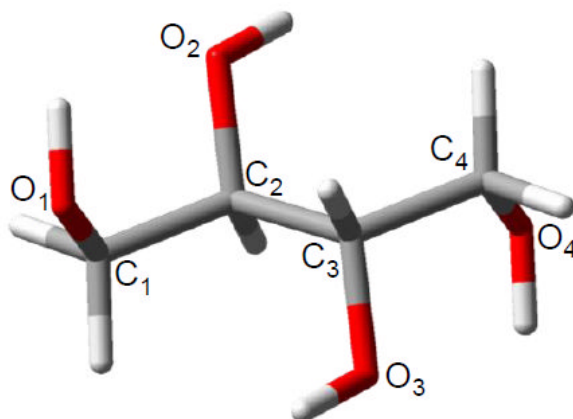
Figure 11. Structure of erythritol including atom numbering.

Since the bands become narrower as temperatures decrease, the first step to simplify the band assignment is to obtain a spectrum at a lower temperature, Figure 12(b). Four bands have been identified in this spectral region and by peak-fit analysis their maxima were located at 3373, 3234, 3190 and 3124 cm^{-1}, Figure 13(a).

The H/D isotopic substitution is a common procedure to help in band assignment. Since the mass of the deuterium is twice as much as that of hydrogen, the stretching frequency of the OD band is $1/\sqrt{2}$ times that of the corresponding OH band. Figure 13(b) displays the spectrum of deuterated erythritol obtained at 15 K in the OD stretching region. Comparing the spectra registered before and after deuteration one can conclude that three bands in the OH region were shifted to frequencies expected from H/D substitution, keeping their relative intensities. Therefore, these bands are unequivocally assigned to the OH stretching vibrations. The band located at 3373 cm^{-1} is not present in the OD region which means that it does not correspond to any deuterable group. Very likely, it is a result of a Fermi resonance effect.
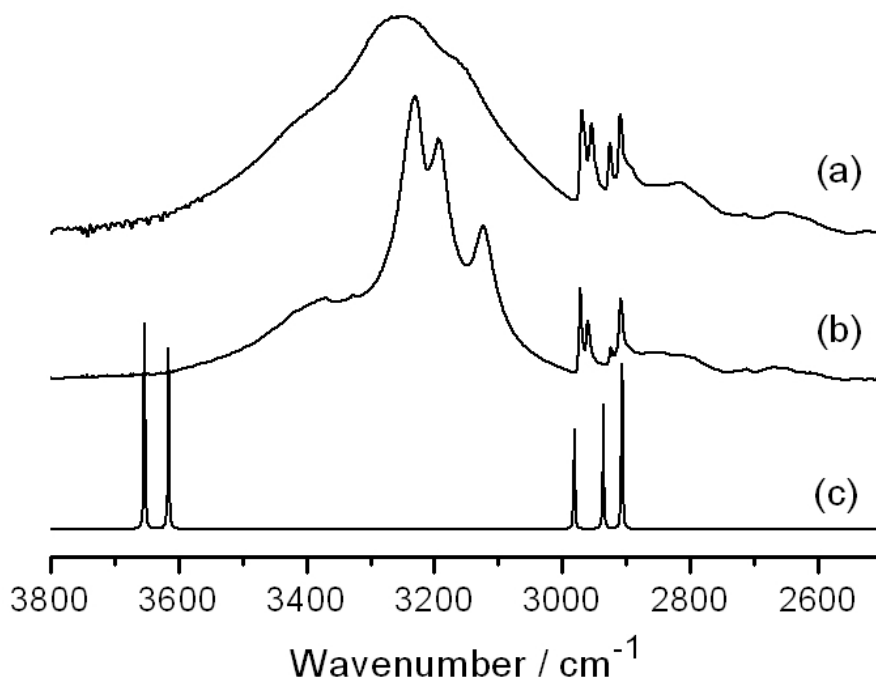
Figure 12. Experimental spectra of solid erythritol dispersed in a KBr pellet at 298.15 K (a) and 15 K (b) and calculated spectrum of the isolated erythritol molecule existent in the crystalline structure at the B3LYP/6-311++G** level of theory (c).

To estimate the strength of the intermolecular hydrogen bonds using the empirical correlation that relates the red shift with enthalpy [38], one has to know the stretching frequency of a free OH group. This value can be obtained from the experimental spectrum of the compound dissolved in an inert solvent or isolated in a noble gas matrix at very low temperatures. Alternatively, it can be obtained from the theoretical spectrum of the isolated molecule. In the present case, the calculated spectrum of the isolated erythritol molecule retaining the crystalline conformation was used. This spectrum, Figure 12(c), exhibits two bands: one centred at 3652 and the other at 3615 cm^{-1}, corresponding to the stretching vibration of the terminal and middle OH groups, respectively. Applying the empirical correlation to the data observed for the spectra at 15 K, values between 27 and 31 kJ mol^{-1} have been estimated for the enthalpy of hydrogen bond formation.

We can achieve deeper insight into the OH band assignment providing we have single crystal diffraction data as happens with erythritol. A neutron diffraction study shows that erythritol adopts two conformations in the crystalline structure, labelled as A and B, and three intermolecular hydrogen bonds can be identified [40]. Conformation A corresponds to that represented in Figure 11. The two conformations differ from one another just in the positions of the hydrogen atoms connected to O_2 and O_3. Assuming that a lower H···O distance and higher O–H···O angle corresponds to a larger red shift, which is quite a plausible assumption, a complete assignment of the OH stretching bands can be made (Table 2).
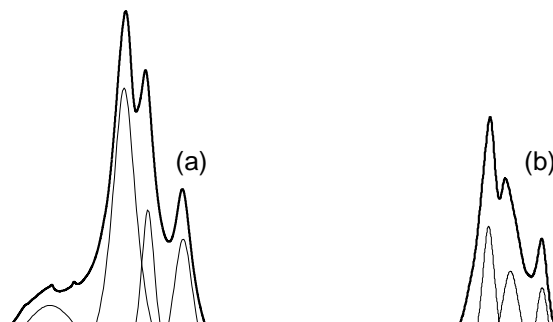
Figure 13. Result of the peak-fit analysis in the OH (a) and OD (b) stretching regions of the experimental spectrum of erythritol at 15 K.

Table 2. Correlation between the spectroscopic and structural data in erythritol.

Hydrogen bond	H···O distance /Å	O–H···O angle /°	ν (OH)/cm^{-1}
O_3–H···O_2 or O_2–H···O_3 (Conf. A)	1.85	163.9	3234
O_3–H···O_2 or O_2–H···O_3 (Conf. B)	1.79	164.5	3190
O_1–H···O_4 or O_4–H···O_1	1.70	173.2	3124

4. CO-CRYSTALS

A co-crystal is a supramolecule containing two or more components together. The designation is matter of some controversy regarding the term and the need to introduce it into science [41,42]. Almarsson and Zaworotko [43] proposed the following definition of co-crystal: "Multi-component crystal in which two or more molecules that are solid under ambient conditions coexist through a hydrogen bond".

One example of a co-crystal is quinhydrone, a supramolecule containing hydroquinone (benzene-1,4-diol) and quinone (p-benzoquinone) in 1:1 stoichiometry. It can exist in a monoclinic and triclinic modifications [44]. In both crystallographic forms the component molecules are linked by OH···O and CH···O hydrogen bonds.

In the monoclinic supramolecule the intermolecular OH···O bonds give linear chains interconnected by CH···O bonds (Figure 14). The triclinic structure differs from the monoclinic one in the relative orientation of the adjacent chains forming sheets parallel to (001) [45].

In aqueous solution quinhydrone gives equimolecular quantities of hydroquinone and quinone, an oxidation/reduction pair. It can be used as electrode to measure pH attending to the acid/base behaviour of hydroquinone and to well-defined quantities of the oxidant and reductor present in solution [46].
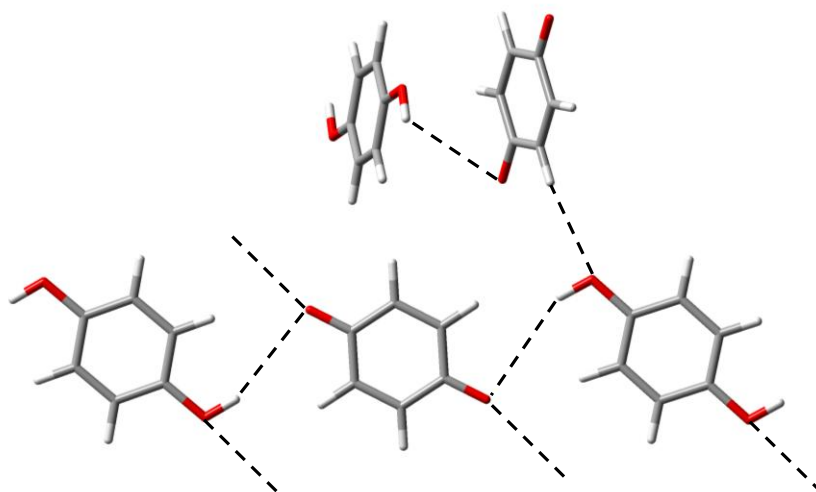
Figure 14. Structure pattern of monoclinic quinhydrone.

Potential applications of co-crystals in different technological fields to improve the properties of various substances have been described but it is in the pharmaceutical industry where they find a higher number of applications. A pharmaceutical co-crystal consists of an active pharmaceutical ingredient (API) and a co-crystal former. The last component may be an excipient, another drug or any other acceptable substance. Both components of the co-crystal are neutral molecules liked by non-covalent intermolecular bonds [47].

Pharmaceutical technology pays particular attention to drug formulation. It should have a higher bioavailability. For a drug substance administered as oral dosage form the bioavailability is the fraction of the amount administered that reaches the site of physiological activity. It depends on several properties such as solubility, the dissolution rate and its permeation through the biological membranes [48]. Solubility and permeation are the properties used by the U.S. Food and Drug Administration as basis of the Biopharmaceutical Classification System (BCS) [49].

To fulfil these demands the API is in some cases formulated as salt, solvate, clathrate, or inclusion crystal. Co-crystals are an extension of the list of pharmaceutical formulations that are being explored. The molecular structure of the API is not modified but the physical properties of the co-crystal may be improved relative to those of the single API. For example, an insoluble compound with pharmaceutical interest may be available as an adequate co-crystal. Also, an amorphous may be transformed into crystalline during the co-crystallization process.

In order to have an overview regarding pharmaceutical co-crystal assembly, we select carbamazepine [CBZ, *5H*-dibenzo (b,f) azepine-5-carboxamide, Figure 15(a)] owing to its potentiality to generate co-crystals [50]. CBZ is a drug used to treat epilepsy and trigeminal neuralgia. One of the multiples examples of co-crystals originated by CBZ is that formed with saccharin [SAC, Figure 15(b)], a synthetic sugar. Two structure modifications may exist: one is monoclinic and the other triclinic [51]. In the monoclinic co-crystal two inverted CBZ molecules are linked by a carboxamide supramolecular homosynthon leaving the peripheral N–H groups available as hydrogen bond donors to the S=O group of SAC [Figure 15(c)]. The

NH of SAC is hydrogen bonded to the anti carbonyl group of CBZ. Between two CBZ dimers there are two SAC molecules interconnected by CH···O bonds [Figure 15(e)].
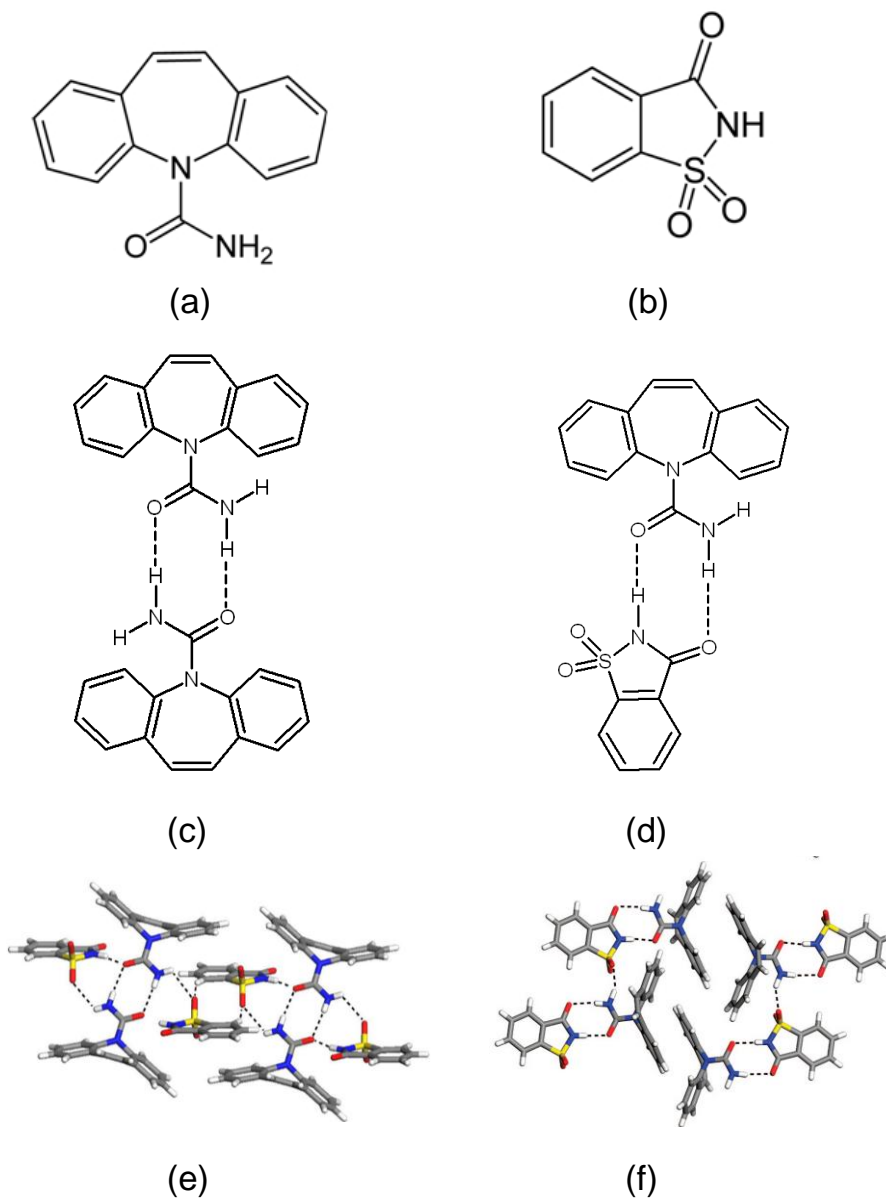
Figure 15. Carbamazepine-saccharin co-crystals. (a) and (b): molecular structures of API and co-crystal former; (c) and (d): supramolecular synthons; (e) and (f): detail of monoclinic and triclinic co-crystal structures.

The other type of structure consists of a heterosynthon established between CBZ and SAC as depicted in Figure 15(d). An additional hydrogen bond is set up between a NH of CBZ and the S=O of SAC giving chains growing along the crystallographic c-axis, Figure

15(f). Besides those just referred, other supramolecular synthons are generated by CBZ with other co-crystal formers [52,53].

5. POLYMORPHISM

Organic compounds can be present under more than one crystalline form. This phenomenon is called polymorphism and has been the object of great interest for the last fifty years in various technologies such as in pharmaceutical, food, dyes and pigments, explosives, etc [54-57]. Among all of these application fields, the practical interest of polymorphism is to obtain the most adequate crystalline form for the desired purposes. On the thermodynamic point of view, for a certain compound, only the lowest energy polymorph is stable. However, some of the thermodynamically unstable forms may have a high energy barrier into the most stable form, remaining as metastables for a long period of time.

One example of polymorphism is *L*-glutamic acid which can exhibit two polymorphs, commonly labelled as α and β (Figure 16). The latter is the stable form. The difference between both polymorphs lies in the conformation of their backbone defined by the C_1-C_2-C_3-C_4 and C_2-C_3-C_4-C_5 dihedrals. The values of these dihedrals are 59.2, 68.3° in form α [58] and -171.1°, -73.1° in form β [59]. As can be seen in Figure 16 the molecular backbone of the metastable form is bent whereas in the stable form it is more distended.
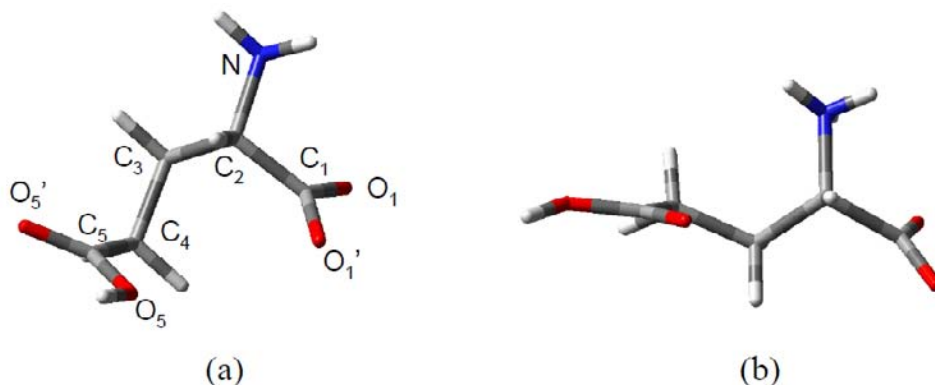
Figure 16. Molecular conformations of polymorphs α (a) and β (b) of *L*-glutamic acid in the zwitterionic form.

Both polymorphs crystallize in the orthorhombic space group with four molecules in the unit cell [58,59]. In both crystalline structures the molecules are in the zwitterionic form and all polar groups participate in intermolecular hydrogen bonds: the OH and NH_3^+ groups as donors and the carbonyl oxygen of the carboxylic and the carboxylate groups as acceptors. An OH···OCO$^-$ hydrogen bond originates chains which are interlinked by N–H···OCO$^-$ bonds yielding a three-dimensional structure. A detailed scheme of the hydrogen bond network found in both polymorphs is displayed in Figure 17 and the values of the hydrogen bond parameters are given in Table 3.

Table 3. Geometrical parameters of the intermolecular hydrogen bonds network existing in the crystalline structures of α and β polymorphs of *L*-glutamic acid.

H-bond	Form α			Form β		
	H···Y	X···Y	X–H···Y	H···Y	X···Y	X–H···Y
N–H···$O_{1'}$	1.774	2.785	111.7	–	–	–
N–H···O_1	1.794	2.823	168.8	1.845	2.868	167.8
				1.868	2.888	166.9
N–H···$O_{5'}$	1.926	2.895	154.9	1.895	2.919	174.1
O_5–H···O_1	–	–	–	1.475	2.519	171.3
O_5–H···$O_{1'}$	1.568	2.518	169.3	–	–	–
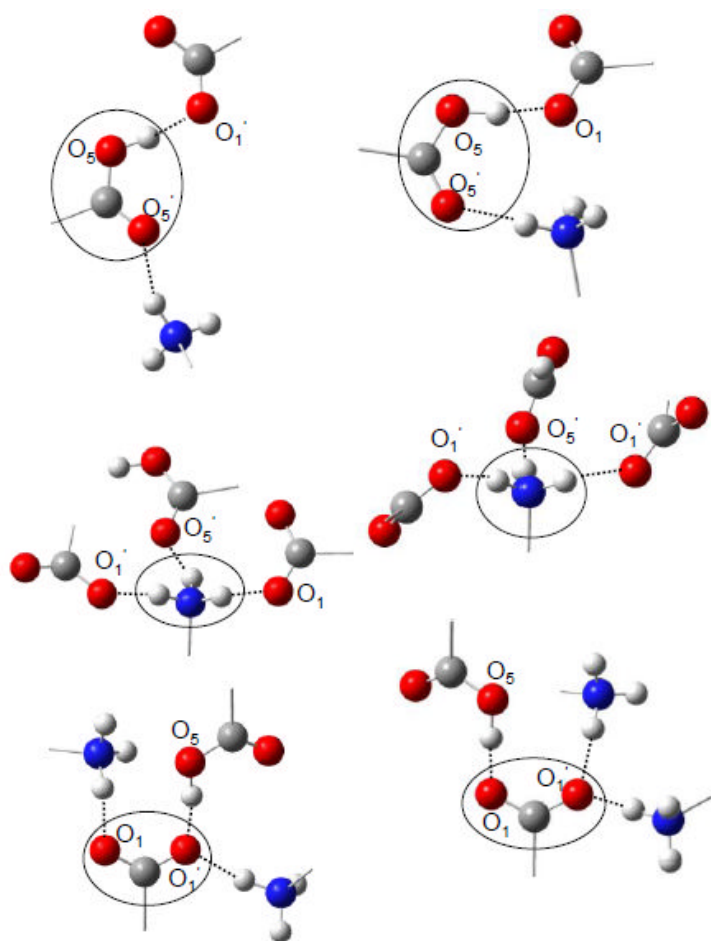
Figure 17. Molecular recognition pattern in the crystalline structures of α (left) and β (right) polymorphs of *L*-glutamic acid. The circled molecular groups belong to the same central molecule.

In general, the molecular recognition pattern in both polymorphs is of the same type. The most significant difference lies in the hydrogen bonds involving the carboxylate group. While in form α two NH_3^+ groups are bonded to different carboxylate oxygen atoms, in form β they are bonded to the same oxygen atom.

All hydrogen bonds have favorable directions and are reinforced by the positive or negative charges of the amine or carboxylate groups. Indeed, the short X···Y and H···Y distances indicate strong hydrogen bonds in the two polymorphs (Table 3).

One of the technological fields where polymorphism is particularly relevant is in the pharmaceutical industry. The search for drugs with high bioavailability, suitable for the desired formulation and stable for shelf life, is an important goal of this industry. These properties are emphasized through the examples that follow.

Chloramphenicol palmitate (Figure 18), a broad-spectrum antibiotic, may exhibit three polymorphs designated as A, B and C [60-62]. A, is the stable form, B, is metastable and C, is unstable and it is quickly transformed into B. The structure of the stable modification has been studied by X-ray diffraction [63] and vibrational spectroscopy [63,64]. Regarding form B, its X-ray powder diffraction pattern shows essentially a lack of long-range order relatively to A [65].
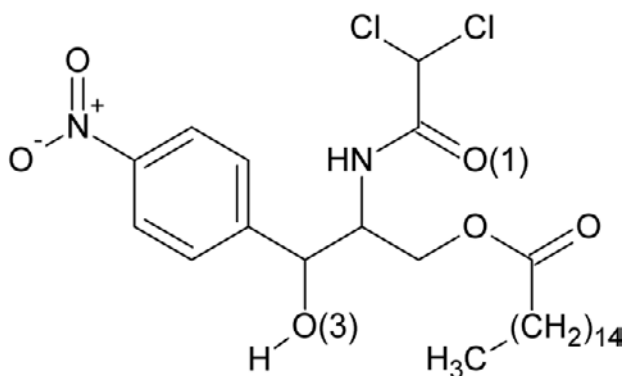
Figure 18. Chloramphenicol (2,2-dichloro-*N*-[(1*R*,2*R*)-2-hydroxy-1-(hydroxymethyl)-2-(4-nitrophenyl)ethyl]acetamide) palmitate.

The drug is formulated as palmitate because chloramphenicol has a pronounced bitter taste while the ester is tasteless. To be active, it needs to be liberated from the esther form. This happens with polymorph B that is hydrolyzed by enzymatic action in the small intestine. Conversely, polymorph A is not recognized by the enzyme and its solubility is too low, as well as its bioavailability. Because of the bioavailability difference between the two polymorphs, the drug formulation should not contain more than 10% of polymorph A.

From the available structural data one can conclude that the difference between both polymorphs lies in the strength of the main intermolecular interaction in the self-assembly of chloramphenicol palmitate, O(3)–H···O(1), stronger in A than in B. Thus, the reaction of the former polymorph with the enzyme is too slow to release a significant amount of chloramphenicol. The difference between the therapeutic efficiency of the two polymorphs, discovered by Aguiar *et al*. in 1967 [61], is an historical event in polymorphism and called the attention to the practical interest of this phenomenon.

Another reason to obtain a certain polymorph lies in its improved properties for formulation. Paracetamol (acetaminophen, Figure 19) is a widely used analgesic and antipyretic. Three polymorphic forms have been identified [66]: one stable (form I), another metastable (form II) and a third one that is unstable and only detected under certain

conditions (form III) [67]. X-ray diffraction studies show that the stable form is monoclinic while the metastable belongs to the orthorhombic system [68-70].
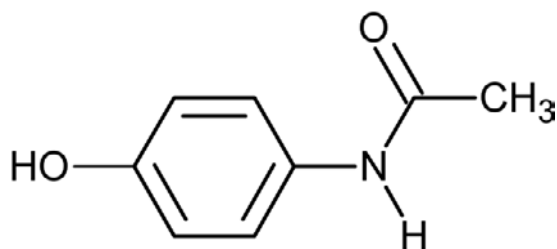

Figure 19. Paracetamol (acetaminophen).

The molecular recognition in the formation of the crystalline structures occurs through N–H⋯OH and O–H⋯O=C intermolecular hydrogen bonds. The spectroscopic characterization of these interactions, based on the stretching vibration of the donor groups, is given in Table 4 [71]. The higher stability of form I is due to the stronger O–H⋯O=C intermolecular hydrogen bond.

Table 4. Frequency of the NH and OH stretching vibrations in the paracetamol polymorphs.

Polymorph	v (OH)/cm^{-1}	v (NH)/cm^{-1}
I	3160	3324
II	3205	3325

Both types of intermolecular hydrogen bonds form chains, which in turn are structured in plated sheets, stacked together by van der Waals forces. The molecular sheets are flat and thin in the orthorhombic modification and rather rough and thick in the monoclinic form. This structural difference gives the polymorphs a different behavior as they are compressed. Form I breaks by brittle fractures whereas form II exhibits a certain plasticity [70]. In the latter, the adjacent sheets linked by weak interactions are slip planes when the crystal is compressed. This feature is technologically advantageous because this form can be tableted by direct compression, unlike form I that needs a binding agent. However, the product available in the market is prepared from I because it is easily obtained from crystallization in different solvents [68,72,73]. Conversely, no adequate method has yet been developed to prepare form II on an industrial scale [67]. Research is going on to reach this target, a step ahead to simplify the formulation of this drug.

Structure stability is surely an essential requirement for the practical interest of polymorphism. Any structural modification of a drug during its shelf-life is not acceptable. This is a matter deserving particular attention as many polymorphs used as drugs are metastable forms. An elucidative example of the importance of the structural stability of a polymorph is the ritonavir story. This drug is used in the treatment of patients infected with immunodeficiency virus type 1 (HIV-1). The product was launched in the market in 1996 in semi-solid capsules formulation. Less then two years later it was observed that some batches of capsules began failing the dissolution requirement. A large amount of the drug was

precipitating out of the semi-solid product. After investigation, it was found that a new polymorph (form II), with a solubility five times lower than that of form I has been formed [74]. Research carried out later showed that polymorph II was more stable than polymorph I but did not crystallize except in highly supersaturated solutions. After intense research a new formulation was developed and the drug was reintroduced to the market.

6. MOLECULAR RECOGNITION IN SOLUTION AND CRYSTAL STRUCTURE

Despite great advances in computational science, it is not yet possible to predict the structure of a crystal from that of a molecule. A great deal of research has been done from the early 1990's [75]. Most of the work has been performed on simple and rigid molecules but a complete structure prediction is not possible at the moment [76]. The computer is a valuable tool in structure prediction when it is used in combination with other techniques. Applying the Monte Carlo method to study powder diffraction data, Harris *et al*. [77] succeeded in determining the structure of some crystalline compounds. Blagden and co-workers [78] used the structure predictions for the 2-amino-4-nitrophenol to evaluate the packing modes and selected the solvents to promote the hydrogen bond network capable to stabilize a certain form.

Experimentally, a method commonly used to obtain polymorphs is crystallization from solvents or from melts. This fact requires knowledge of the molecular recognition in solution during the stages of crystal growth. The crystallization process can be understood as a balance between the molecular attractions that join the molecules together in a new phase and the creation of an interface between the two phases. The Gibbs energy variation per volume unit (Δg) is negative because the intermolecular forces in the second-phase are stronger than in solution. On the contrary, work has to be done to create an interface. The overall Gibbs energy variation corresponding to the molecular aggregation admitting a spherical geometry, can be written as:

$$\Delta G = \frac{4}{3}\pi\, r^3 \Delta g + 4\pi\, r^2 \gamma \tag{1}$$

where r is the particle radius of the aggregate and γ the interfacial tension. The first is called the volume term and the second the interfacial term [79,80].

The variation of ΔG with the particle radius, $\partial(\Delta G)/\partial r$, depends on the size of the particle. For small size particles $\partial(\Delta G)/\partial r$ is positive (interfacial term is dominant), while for larger sized particles it is negative (the volume term is dominant). The maximum value of ΔG as function of r, $\partial(\Delta G)/\partial r = 0$, gives the critical value for the radius (r^*), expressed by $r^* = -2\gamma/\Delta g$. This is a meaningful parameter in crystal growth from solution. In fact, aggregates with $r < r^*$ tend to redissolve while those with $r > r^*$ grow irreversibility towards the crystal. Hence, molecular aggregates with $r = r^*$ are the embryo or nucleus of the solid phase. ΔG for $r = r^*$ is given by $\Delta G^* = 16\pi\gamma^3/3(\Delta g)^2$. The variation of ΔG with r is shown in Figure 20.
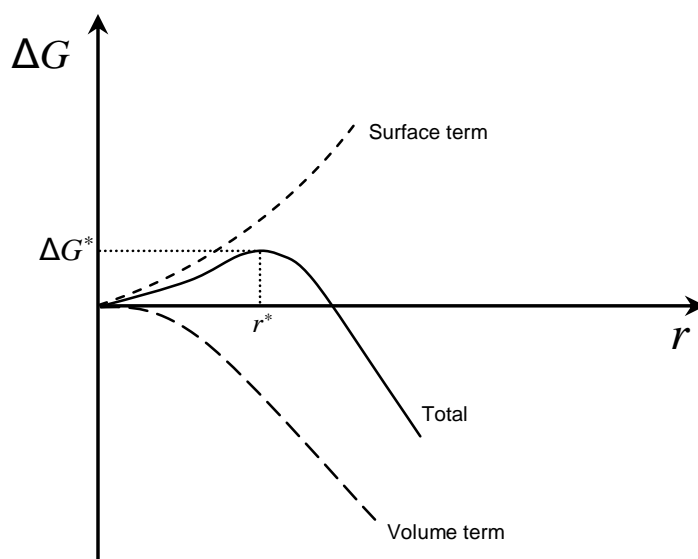
Figure 20. Variation of the total Gibbs energy, as well as the surface and volume terms as function of the particle size.

The screening of polymorphs by crystallization is a time-consuming method. Despite the knowledge of the molecular structure of the compound under investigation and the variety of solvents that can be used, and, moreover, the rules deduced from the accumulated data supplied by other systems, it is matter of experience. Crystallization is, therefore, a trial-and-error procedure. To save time, high-throughput methods have been developed allowing a large number of experiments to be processed simultaneously employing small amounts of the solid under investigation. The new phase is detected and characterized *in situ* by physical methods, as X-ray diffraction and Raman spectroscopy [81]. These automatic and monitored processes may be a useful guide but do not dispense the manual operations because the results they provide do not reach the quality of those obtained from individual experiments.

Crystallization can be considered as comprising the following stages:

Molecule → molecular aggregates → nucleus → crystal

Considering the importance of the nucleus as the birth of the generated solid forms, two main events are usually considered: from the molecule to the nucleus (nucleation) and from this to crystal (crystal growth).

A relevant question arises in respect with molecular recognition and that is whether or not the same pattern persists from the formation of the molecular aggregates to the final crystal form. In affirmative cases, the crystal structure can be inferred from the structure existing in aggregates of a few molecules, say a dimer. Since these structures can be studied by various methods including the computational ones, its study is on open door to investigate the crystalline forms. This is a matter that will be the object of discussion based on the results obtained for the crystallization of terfenadine from solvent carried out in the authors' research group.

Terfenadine, 1-(4-*tert*-butylphenyl)-4-[4´-(diphenyl-hydroxymethyl)-1´-piperidyl]butan-1-ol (Figure 21), was formerly used for allergy treatment as an antagonist of the H_1-receptor. Late in 1980s, because of the risk of cardiac arrhythmia, it was superseded by the metabolite fexofenadine.

Terfenadine prepared from solution results in various polymorphic forms being produced. A detailed study of crystallization in different media and using different techniques has been undertaken [82-85]. To have an idea of the type of molecular recognition that determines the self-assembly of terfenadine in solution, the study started by a molecular dynamics simulation of terfenadine monomers and dimers in gas phase, and in ethanol and methanol solutions [85], two commonly used solvents for terfenadine crystallization.
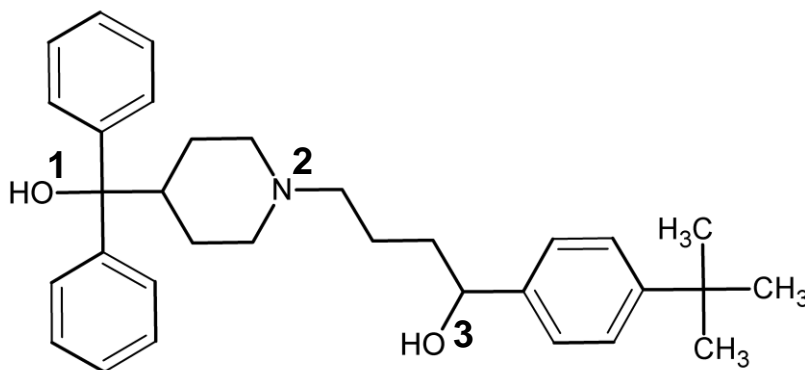
Figure 21. Molecular structure of terfenadine. Numbers represent the three polar sites of this molecule.

It is expected that terfenadine aggregation would be essentially determined by inter- and intramolecular hydrogen bonds, involving the three polar sites labelled in Figure 21. The bond criteria was based on the distance between the hydrogen atom and the acceptor group ($R_{i,j}$). $R_{i,j} \leq 2.5$ Å denotes a clearly formed hydrogen bond, while $R_{i,j} \geq 4.0$ Å indicates that no inter- or intramolecular hydrogen bond is found to exist. Hydrogen bonds were further characterized by their persistence time ($p_{i,j}$), being the ratio between the time during which the bond is established ($t_{i,j}$) and the total simulation time (t_{total}), $p_{i,j} = t_{i,j} / t_{total}$.

The results obtained for the persistence of the O–H···N intramolecular hydrogen bond (sites 2 and 3 of Figure 21) for the monomer and dimer of terfenadine are shown in Table 5.

Table 5. Persistence of the N–H···O intramolecular hydrogen bond for the monomeric and dimeric structures of terfenadine in different media [85].

Media	$p_{i,j}$ ($R_{i,j} \leq 2.5$ Å)	
	Monomer	Dimer
Gas phase	0.262	0.117
Methanol	0.018	0.005
Ethanol	0.002	0.000

This interaction is relatively frequent in the gas phase monomers. However, its persistence is diminished by the solute-solvent interactions. It is worth noting the difference between methanol and ethanol. The persistence time of the intramolecular hydrogen bond is

higher in methanol than in ethanol. The persistence of this interaction also decreases as a consequence of the intermolecular bonds between the monomers to originate dimers.

Regarding the intermolecular hydrogen bonds involved in the formation of dimers, their persistence times are summarized in Figure 22. Practically, all polar groups work as recognition active centres, with a preference for sites 3 and 1. In gas phase the persistence of the bond 1-3 (head-to-tail) is higher than that of bond 1-1 (head-to-head). The same sequence occurs in ethanol, whereas the reverse is found in methanol. These two main types of intermolecular hydrogen bonds are represented in Figure 23.
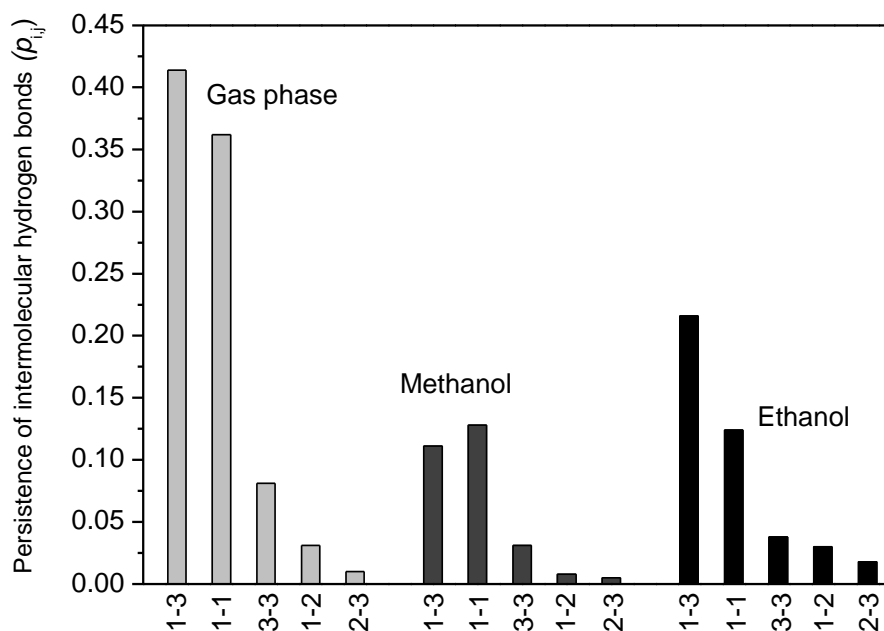
Figure 22. Persistence of the intermolecular hydrogen bonds between sites 1, 2 and 3 for the terfenadine dimer in gas phase, methanol and ethanol.
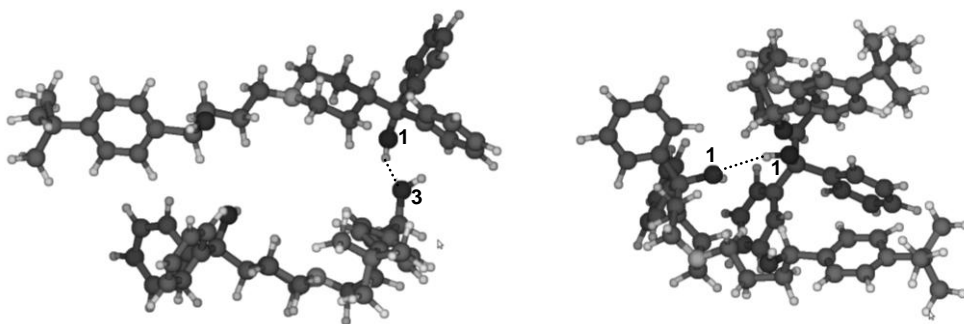
Figure 23. Snapshots of a molecular dynamics simulation of the gas phase terfenadine dimer showing the establishment of the two main intermolecular hydrogen: 1-3 (left) and 1-1 (right).

Considering that solvation competes with the solute self-association, the persistence of both interactions decreases considerably from the gas phase to ethanol or methanol solutions.

The two solvents show a different behaviour in respect to the most predominant interactions: bond between sites 1 of the two monomers is the most relevant in methanol while that involving sites 1 and 3 prevails in ethanol.

The results of the molecular dynamics simulations also provide important data on the type of molecular conformation exhibited by terfenadine in solution. In ethanol the molecules adopt preferentially a stretched backbone, whereas bent and stretched conformations co-exist in methanol solution. Therefore, methanol should favour the formation of more disordered solid state structures relatively to ethanol.

Temperature has a marked influence in the relative persistences. In methanol, at 25ºC, the persistences of the 1-1 and 1-3 bonds have close values while at 100ºC the persistence of the first decreases drastically relative to the second [82].

To compare the data obtained for the dimer formation with the polymorphism exhibited by terfenadine, a large number of specimens of this compound were prepared by crystallization from different solvents under different experimental conditions. The specimens were studied by differential scanning calorimetry. Three polymorphs have been identified and characterized by their fusion temperatures [83]: Form I (150.2 ± 0.1ºC), Form II (147.5 ± 0.3ºC) and Form III (146.2 ± 0.4ºC). By rule, the harvest of the crystallization is a mixture of two polymorphs in a wide range of compositions. Form I as a unique polymorph in the crystallization products, was obtained by the slow vaporisation of the ethanol-water co-solvent. This polymorph is likely a head-to-tail molecular association [82]. It is frequently the dominant form in specimens prepared from ethanol. Polymorphs II and III are occasionally near pure in specimens crystallized from methanol. The former by slow evaporation at 50ºC and the latter by slow evaporation at -3ºC. As predicted from molecular dynamics simulation, crystallization in ethanol yields a higher crystalline product than in methanol.

The mixtures of polymorphs that are observed are expected before the simultaneous existence of different forms in solution. They are the result of various combinations of forms during the crystal growth as predicted by molecular dynamics. However, the dominant molecular recognition in dimers apparently persists up to the crystal.

A good correlation between self-association of the glycine in a supersaturated solution and the polymorphic outcome was also observed [86]. We should be well-advised that sometimes the molecular recognition pattern in the crystal may be significantly different from the early stages of crystallization. As the aggregates growing the minimum energy conditions are reached at the expense of a conformational variation.

7. CONFORMATIONAL RECOGNITION IN SOLUTION AND CRYSTAL STRUCTURE

Molecules of organic compounds, even the simpler ones, adopt various conformations in solution. Thus, a solution may be considered as an assembly of conformers dispersed in the solvent with relative populations determined by the Boltzmann distribution of the Gibbs energy. Crystallization is therefore an aggregation of conformers determined by preferential recognition of the interacting groups in competition with the interaction of the conformers with the solvent. Like the interaction between molecules, we can call conformational recognition to the interaction between conformers.

The recent progress of the computational methods made the conformational study of medium sized molecules using electronic structure methods a matter of routine. Geometric and thermodynamic properties of the conformers, as well as their population can now be estimated with reasonable accuracy.

These calculations can be performed either in gas phase or in solution. Accurate solvation Gibbs energies may be estimated for a wide range of molecules at a relative low computational cost from continuum solvation models [87]. It is thus possible, for a given molecular system, to predict the relevant conformers in solution. To take advantage of the computational developments in this matter, it is important to compare the conformers prevailing in solution with those presented in the crystal. This question will be addressed by considering the crystallization of erythritol from water.

Despite the small number of atoms, erythritol can adopt a large number of conformations. Figure 24 shows the equilibrium populations of the most relevant conformers of erythritol in aqueous solution [10]. Just for comparison, their gas phase populations have also been included in the same Figure [88]. The calculations have been performed using the DFT(B3LYP) approximation method and the 6-311++G(d,p) basis set. Solvent effects were accounted for by the conductor-like polarizable continuum solvation model (CPCM) [10].
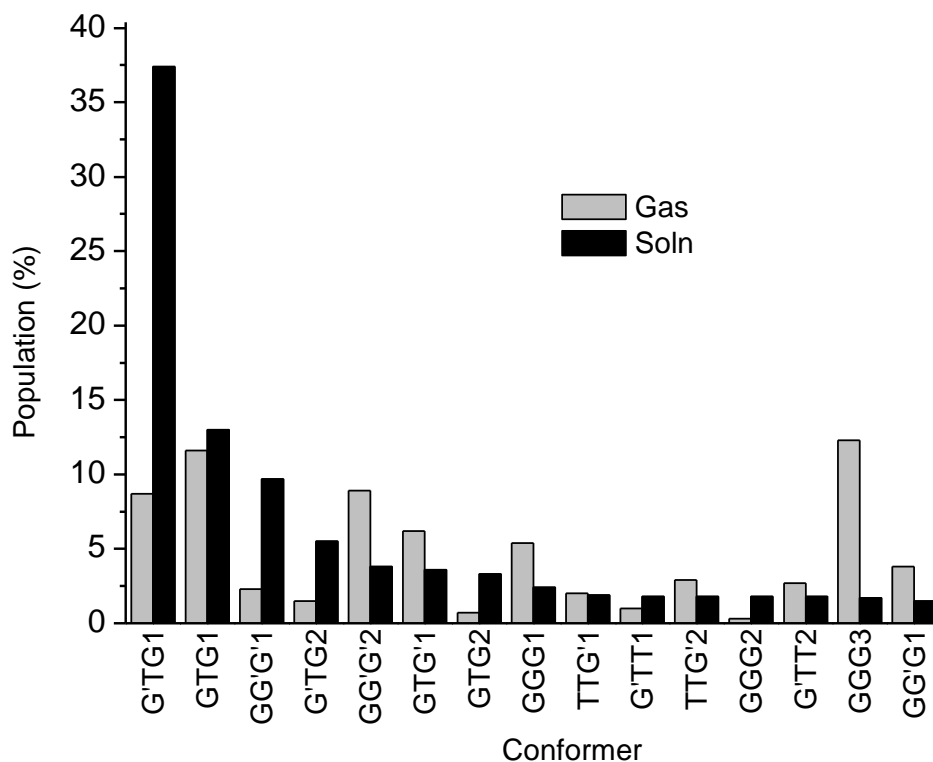
Figure 24. Conformational distribution of erythritol in aqueous solution and gas phase.

Erythritol conformations are identified by a set of three letters that characterize the orientation of the φ_1 ($O_1C_1C_2O_2$), φ_2 ($C_1C_2C_3C_4$) and φ_3 ($O_3C_3C_4O_4$) dihedral angles, respectively (see Figure 11 for the atom numbering scheme). The letters used are G, G' or T,

for dihedrals of around 60, -60 and 180º, respectively. Conformers with the same backbone conformation but different OH groups' orientation are discriminated by a number following the stability order.

The molecular transference from the gas to the solution changes drastically the conformation. In fact, no relationship between the populations in both media can be established as can be seen from Figure 24. The two most abundant conformers in solution have a distended hydrocarbon chain while the structures stabilized by intramolecular hydrogen bonds are those preferred in gas phase [88]. As referred before, crystalline erythritol presents two isomorphic structures. Both have a G'TG backbone and differ one from another just in the orientation of the middle OH groups. As the stable conformers in aqueous solution have the same backbone we conclude that the crystal is built up by the aggregation of these conformers. Hence, the knowledge of the conformational distribution for a certain compound in solution is undoubtedly a good starting point to interpret crystal growth. Conformational recognition is apparently an effective way to understand crystallization, particularly in the identification of the solvent effect. More data dealing with other solutes and solvents are required in order to improve our ability to evaluate the potentialities of the method.

8. MOLECULAR RECOGNITION AT INTERFACES

8.1. CRYSTAL GROWTH IN THE PRESENCE OF TAYLOR-MADE ADDITIVES

In a broad sense, the crystalline structure is not the only property that needs to be considered when we are looking for a desired crystal. Other features are technologically important such as morphology or crystalline habit. Filterability, flowability, compressibility and many other properties are relevant in industrial processing. They are dependent on the size and morphology of the crystal.

The morphology can be modified by adding to the mother liquid a small amount of a substance that can be selectively absorbed at one or a few crystal faces. The effective molecular additive should have molecular groups similar to those of the substract but different one another in the other moiety. As the additive is incorporated in one of the crystal surfaces the growth in a perpendicular direction is interrupted or decreases and the molecular geometry is modified. The additives with this behaviour are called tailor-made additives. The solvent has also a tailor-made effect.

The mechanism of the tailor-made additives can be exemplified with benzamide grown in the presence of benzoic acid, o-toluamide or p-toluamide [89]. Benzamide is a compound widely used in organic synthesis. It has also biological interest because it acts as a neuroprotectant inhibiting the polymerase action, an enzyme activated by nitric oxide. It can be crystallized from ethanol giving plate-like crystals belonging to the monoclinic system [90,91]. In the crystal the molecules are interlinked by hydrogen bonds along the b-axis as represented in Figure 25. The ribbons with b-direction are stacked along the a-axis in layers which, in turn, are juxtaposed along the c-axis through van der Waals interactions between the phenyl groups. The velocity of crystallization is higher along the b-axis than in the other directions.

Benzoic acid is a tailor-made impurity for benzamide. Both molecules have identical capacity for hydrogen bonding but while in benzamide the NH_2 group participates in two hydrogen bonds as donor, the OH group in the benzoic acid is involved just in one. Thus, when by chance, the benzoic acid molecule is incorporated into the growing benzamide structure, the crystallization front is interrupted. The link of a new benzamide molecule is rendered difficult by the absence of the NH_2 groups and also by the repulsion between the oxygen lone pairs, as illustrated in Figure 25. Since the crystallization is impeded along the direction of b-axis the crystal grows along the a-axis as palettes.
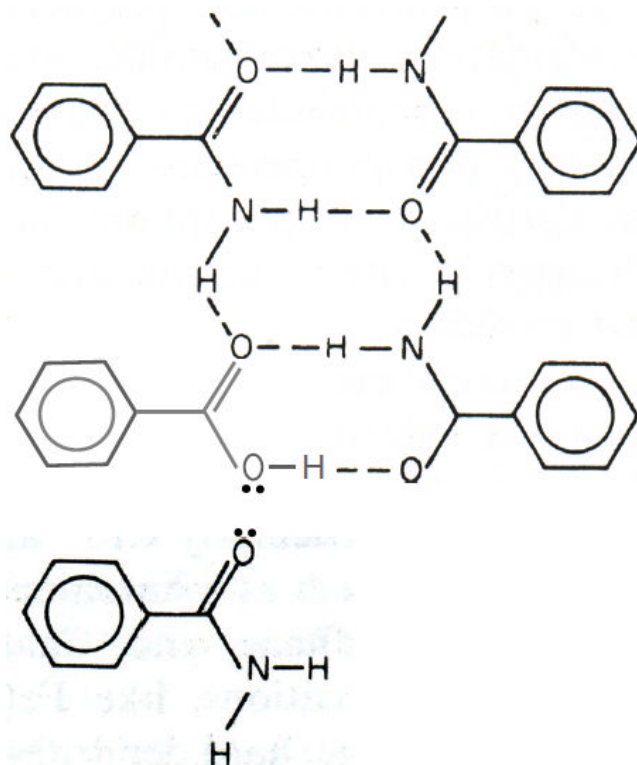
Figure 25. Crystal growth disruption in benzamide by the incorporation of benzoic acid.

O-toluamide is also used as tailor-made additive. Once the molecules are incorporated into the benzamide crystalline structure through a N–H···O bond, they disturb the crystal growth along the a-axis owing to the voluminous methyl group and the crystal adopts a form of bars along the b-axis. Moreover, a different habit modification is achieved with p-toluamide by retarding the growth along the c-axis.

In principle, the crystalline habit modification by tailor-made additives is dependent on the amount of the compound added to the mother liquor. In fact, the probability of its incorporation in the crystallization front depends on the ratio between the additive and the substract.

Tailor-made habit modifiers find wide applications in industry owing to their potentialities to modify size and morphology of crystalline solids [92,93]. It is possible to have an adequate molecule recognized by the crystallization front of the substract provided we know its structure.

8.2. EFFECT OF THE SOLVENT ON THE CRYSTALLINE HABIT

Besides the influence in the structure of the crystals, the solvent also affects their morphology. As was mentioned before, specimens of terfenadine precipitated from methanol, ethanol and ethanol-water, exhibit different polymorphic compositions. Furthermore, they present different morphologies too. The crystals grown from methanol by slow evaporation of solvent at 25ºC are platelets while those grown from the remaining solvents in the same experimental conditions are needle-like (see Figure 26).

(a)   (b)   (c)
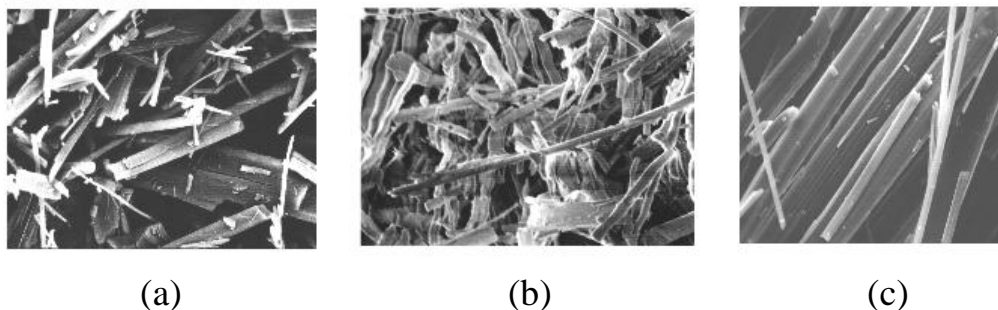
Figure 26. Electron microscopy micrographs (JEOL 5400, 2000x) of terfenadine specimens crystallized from methanol (a), ethanol (b) and ethanol-water (c) [94].

The solvent effect lies in the difference of interaction of the solvent molecules with the crystal surfaces. Polar crystals have faces with different groups exposed to the solvent. For example, the packing arrangement in crystalline alkylgluconamides, $C_nH_{2n+1}NHCO(CHOH)_4CHOH$, n=7-10, has a face where the groups exposed to the solvent are polar while those of the opposite face are non-polar [95]. Zwiterionic molecules of (R,S)-alanine give crystals in which the pheripherical group of one of the faces is the carboxylate whilst the amine group is in the opposite face [96].

The surface-solvent interaction depends on the solvent nature and the specific structure of each face. Various theories have been proposed to interpret the solvent effect on the crystal morphology. One of them is based on the argument that the interface growth implies the striping of the solute and of the interacting site of the substract from the solvating molecule. This event is as much unfavourable to the growth as stronger the solvent-surface interaction is. Therefore, a strong solvent binding decreases or inhibits the growth [97,98]. Another approach considers that a strong solvent-surface interaction provokes a roughened interface favouring faster growth [99]. A third interpretation considers the solvent a tailor-made additive and therefore its behaviour depends on the structure of the surface. Using as an example the (R,S)-alanine, in the face with the carboxylate exposed to the solvent one can distinguish neighbour pockets of one-molecule deep separated by one solute molecule [100]. Ethanol and methanol are solvents adsorbed at the surface. The former by CH···O interactions and the latter by OH···O and CH···O bonds with the substrate. The approaching alanine molecule is prevented from reaching the molecules within the pockets by these solvents. The growth through this face is rendered difficult or inhibited. When the solvent is water the repulsion between the oxygen lone-pairs of water and carboxylates at the pocket periphery

inhibits the presence of the solvent within the pocket. Thus, water favours a faster growth along the direction perpendicular to the face. It is to be noted that after a deposition of a new solute layer on a crystal interface the surface sites are reversed. What was previously a cavity is now a protrusion and *vice versa*. For this reason the authors call this type of growth a "relay" mechanism.

8.3. CONFORMATIONAL MIMICRY

Tailor-made inhibitors can also be used to prevent the crystallization of one polymorph, leaving the other or others free to be precipitated out of the solution. Davey *et al.* [101] succeeded in preventing the growth of the stable form of *L*-glutamic acid by adding trimesic acid (benzene-1,3,5-tricarboxylic acid) to the crystallizing solution.

As referred to before, *L*-glutamic acid is a dimorphic compound. By crystallization, the two pure forms (α and β) or mixtures of both polymorphs can be obtained, depending on the degree of supersaturation and temperature [102]. Both modifications give rise to orthorhombic crystals with different morphologies; the crystals of form α are rhombic while those of the latter are needle-like.

The data available on the structure and crystallization of *L*-glutamic acid [101] and the conformational analysis performed by the authors served as basis for selecting a tailor-made inhibitor for form β. As can be seen in Figure 27, the carboxylate groups of the trimesic acid are recognized by the groups of this polymorph as far as the $O_5 \cdots O_1$ and $O_5' \cdots O_1'$ distances in the acid are close to those on the fastest growth (101) face of the substrate. Hence, the hydrogen bonds involving the OH and C=O groups of the acid and the substrate are formed without great distortion. Two of the three carboxylic groups of the additive are recognized by that face of form β but the third one disrupts the addition of further substrate molecules.

Interatomic distances related to molecular recognition between the two forms of *L*-glutamic acid and the trimesic acid are given in Table 6 and illustrated in Figure 27. Due to the bent backbone of form α, the mimicry by the additive can not be possible. However, the interaction of the trimesic acid with this polymorph retards nucleation and modifies its crystalline habit.

Table 6. Distances between the carboxylic groups in the trimesic acid and the two polymorphs of *L*-glutamic acid.[a]

Molecule	$C_1 \cdots C_5$ / Å	longer O⋯O / Å	shortest O⋯O / Å
Trimesic acid	5.01	7.19	4.94
L-glutamic acid (form β)	4.55	6.88	4.61
L-glutamic acid (form α)	4.01	6.26	4.39

[a] These values were taken from fully optimized geometries at the B3LYP/6-31G* level. Input geometries for forms α and β of *L*-glutamic acid were obtained from X-ray and neutron diffraction data [58,59].
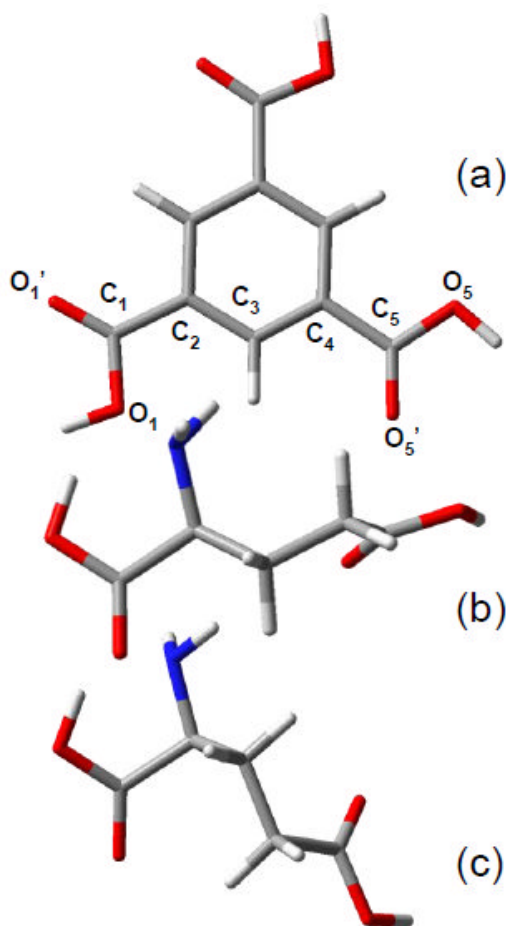
Figure 27. Comparison of the optimized conformations (B3LYP/6-31G*) of trimesic acid (a) and forms β (b) and α (c) of *L*-glutamic acid.

8.4. SEEDING CRYSTALLIZATION

Additives are also employed to control the structure of a crystal obtained from a solvent or from the melt. The inoculation into a supersaturated solution (or supercooled melt) of a small amount of the crystalline solid form that we want to obtain is a possible procedure to reach this target. The molecules at the surface of the crystals have part of their bonding sites free to interact with those in solution. As the surface preferentially recognizes molecules with the same conformation, the structure of the new phase corresponds to that of the primitive crystal. In this case the nucleation occurs at the surface of the crystal and is called heterogeneous nucleation to differentiate it from the nucleation taking place in the solution free from any solid contaminant, homogeneous nucleation.

Besides the structural control of the new solid phase the heterogeneous nucleation may result in forming an improved quality product since it requires lower supersaturation. Let us consider the rate of both nucleation types. The rate of critical nuclei formation (J), in random conditions as those existing in homogeneous nucleation, can be expressed by the following equation [103-105]:

$$J = A e^{\left(-\frac{\Delta G^*}{kT}\right)} \qquad\qquad (2)$$

ΔG^* is the Gibbs energy barrier to nucleation. The pre-exponential factor (A) can be given as follows:

$$A = N_V \left(\frac{kT}{h}\right) e^{(-\varepsilon/kT)} \qquad\qquad (3)$$

N_V is the number of molecules per volume unit and ε the activation energy for the jumps of the molecule across the boundary from solution to the new phase.

The difference between homogeneous and heterogeneous nucleation is that in the latter ΔG^* is lower because the interfacial energy is reduced; A is higher for the former because N_V is higher than the number of sites at the surface able to recognize the nucleus molecules. However, the differences between the ΔG^* values overcome that between the values of A. Hence, at a certain supersaturation, $J_{het} > J_{hom}$. Therefore, a noticeable value of J_{het} is reached at a lower supersaturation than that required for the homogeneous nucleation.

Single crystal growth is widely used in industry and several methods have been developed to obtain materials suited for different purposes [106]. The seed is not necessarily a solid with the composition corresponding to that we intend to obtain; isomorphous of the new phase can also induce crystallization. For example, phosphates can nucleate arsenates because they have identical crystalline lattices. Also, silver iodide is used in rain-making due to the similarity of its lattice of salt with that of ice. Thus, when particles of silver iodide are dispersed in a supercooled cloud, the molecular recognition between the (0001) substrate plane and the water molecules initiates the condensation process [107].

Entacapone (((2E)-2-cyano-3-(3,4-dihydroxy-5-nitrophenyl)-N,N-diethylprop-2-enamide) is a drug belonging to the nitrocatechols and is used in the treatment of Parkinson's disease. The Au (100) surface recognizes the stable polymorphic form of this compound in an acetone/water mixture. When the solution is in contact with that surface of the metal, a well ordered layer of a prismatic polymorph is formed. However, in the absence of seeding the crystallization of a metastable form, a needle-shaped crystal, is generally observed [108].

CONCLUDING REMARKS

One of the main objectives of the present paper was to answer the question of whether molecular recognition in the first stages of molecular aggregation and the conformational composition can be used as guidelines to infer the resulting crystal structure.

From a Molecular Dynamics study of terfenadine in various solvents, the sites involved in molecular recognition originating molecular backbone folding and aggregation were spotted. A polymorphic study of terfenadine in the solvents selected for the simulation was performed by thermal methods. The comparison between the results shows that the features of the dominant polymorphs could be expected from the molecular recognition of dimeric assembly in solution.

The conformational distribution of erythritol in solution was shown to be a successful approach to predict the structure of the crystalline form obtained from this solvent. The conformers existing in the crystal were found to be the most abundant in solution. Like molecular recognition, one can talk about conformational recognition when the starting point is the conformer.

The importance of molecular recognition in the first stages of crystallization should be pointed out as far as molecular self-assembly in solution can be studied computationally while the crystalline structure prediction from the molecular properties is rather difficult to undertake. An identical argument is valid for conformational recognition in so far available computational techniques permit the determination of the conformational Gibbs energy in solution at a reasonable level of theory.

Crystal growth is a matter of utmost importance in a wide range of industries. Its control making use of molecular recognition points out the importance of this concept in science and technology.

The molecular recognition at solid/solution interfaces has deserved particular attention as far as it is a means by which to prepare crystalline forms with morphology suited to specific applications, stabilize desired metastable polymorphs or promoting certain crystalline forms precipitating out from solution.

Besides the importance of molecular recognition in the crystal growth, this chapter calls the attention for the role it plays in the biological processes by considering the interaction of drugs and sweeteners with the receptors.

Among the non-covalent forces involved in molecular recognition, hydrogen bond occupies a special position. Firstly, it predominates in the interactions involving molecular recognition. Secondly, the directional character in a wide range of orientations confers it with unique properties that have been reviewed in the present chapter. Effects of the hydrogen bond such as the modification of the natural hybrid orbital composition of the interatomic bond in the donor group and the charge increase of the hydrogen atom bridging donor and acceptor groups were analysed by the Natural Bond Orbital Theory. This theory was also used to estimate the hydrogen bond strength.

Infrared spectroscopy was found to be a privileged method for hydrogen bonding identification. The interpretation of complex spectra in the stretching vibration region was presented as a valuable tool for molecular recognition diagnose.

REFERENCES

[1]   Lehn, J.-M. *J. Inclusion Phenom. Macrocyclic Chem.* 1988, *6*, 351-396.
[2]   Lehn, J.-M. *Supramolecular Chemistry*; VCH: Weinheim, 1995.
[3]   Pedersen, C. J. *J. Am. Chem. Soc.* 1967, *89*, 7017-7036.
[4]   Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. *Org. Process Res. Dev.* 2000, *4*, 413-417.
[5]   Steiner, T. *Angew. Chem., Int. Ed. Engl.* 2002, *41*, 48-76.
[6]   K. L. Wolf; H. Frahm; H. Harms *Z. Phys. Chem. Abt. B* 1937, *36*, 237-287.
[7]   Corey, E. J. *Pure Appl. Chem.* 1967, *14*, 19-38.

[8]   Desiraju, G. R. *Angew. Chem. Int. Ed.* 1995, *34*, 2311-2317.

[9]   Costa, F. S.; Eusébio, M. E.; Redinha, J. S.; Leitão, M. L. P. *J. Chem. Thermodyn.* 2000, *32*, 311-317.

[10]   Jesus, A. J.; Tomé, L. I. N.; Eusébio, M. E.; Redinha, J. S. *J. Phys. Chem. B* 2006, *110*, 9280-9285.

[11]   Pedersen, C. J. *J. Am. Chem. Soc.* 1967, *89*, 2495-2496.

[12]   Weinhold, F.; Landis, C. R. *Valency and bonding: A Natural Bond Orbital Donor–Acceptor Perspective*; Cambridge University Press: New York, 2005.

[13]   Weinhold, F. *NBO 5.0 Program Manual* Theoretical Chemistry Institute, Univ. Wisconsin: Madison, 2001.

[14]   Fabre, B.; Simonet, J. *Coord. Chem. Rev.* 1998, *178-180*, 1211-1250.

[15]   Demirel, N.; Bulut, Y.; Hosgören, H. *Tetrahedron: Asymmetry* 2004, *15*, 2045-2049.

[16]   Blakemore, J. D.; Chitta, R.; D'Souza, F. *Tetrahedron Lett.* 2007, *48*, 1977-1982.

[17]   Lee, S.-W.; Lee, H.-N.; Kim, H. S.; Beauchamp, J. L. *J. Am. Chem. Soc.* 1998, *120*, 5800-5805.

[18]   Israelachvili, J. N. *Intermolecular and surface forces*; Academic Press: London, 1992.

[19]   Silverman, R. B. *The Organic Chemistry of Drug Design and Drug Action*; Elsevier Academic Press: New York, 2004.

[20]   Coulson, C. J. *Molecular mechanisms of drug action*; 2 ed.; Taylor & Francis: London, 1994.

[21]   McConathy, J.; Owens, M. J. *Prim. Care Companion. J. Clin. Psychiatry* 2003, *52*, 62-67.

[22]   I. Szelenyi; G. Geisslinger; E. Polymeropoulos; W. Paul; M. Herbst; Brune, K. *Drug News Perspect.* 1998, *11*, 139-160.

[23]   Shallenberger, R. S.; Acree, T. E. *Nature* 1967, *216*, 480-482.

[24]   Lee, C.-K.; Horton, R. S. T. a. D. In *Adv. Carbohydr. Chem. Biochem.*; Academic Press: 1987; Vol. 45, p 199-351.

[25]   Kier, L. B. *J. Pharm. Sci.* 1972, *61*, 1394-1397.

[26]   Tinti, J. M.; Nofre, C. In *Sweeteners: Discovery, Molecular Design, and Chemoreception*; Walters, D. E., Orthoefer, F. T., Dubois, G. E., Eds.; American Chemical Society: Washington DC, 1991, p 209-213.

[27]   Nofre, C.; Tinti, J.-M. *Food Chem.* 1996, *56*, 263-274.

[28]   Jesus, A. J.; Rosado, M. T. S.; Leitão, M. L. P.; Redinha, J. S. *J. Phys. Chem. A* 2003, *107*, 3891-3897.

[29]   Jesus, A. J.; Rosado, M. T. S.; Reva, I.; Fausto, R.; Eusébio, M. E.; Redinha, J. S. *J. Phys. Chem. A* 2008, *112*, 4669-4678.

[30]   Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, 1997.

[31]   Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond In Structural Chemistry and Biology*; Monographs on crystallography; Oxford University Press: New York, 1999.

[32]   Pascard, C.; Tran Huu Dau, E.; Manoury, P.; Mompon, B. *Acta Crystallogr. C* 1984, *40*, 1430-1432.

[33]   Meot-Ner, M. *Chem. Rev.* 2005, *105*, 213-284.

[34]   Gilli, P.; Ferretti, V.; Bertolasi, V.; Gilli, G. In *Advances in Molecular Structure Research*; Hargittai, M., Hargittai, I., Eds.; JAI Press: Greenwich, 1996; Vol. 2, p 67.

[35] Maréchal, Y. *The hydrogen bond and the water molecule: the physics and chemistry of water molecule, aqueous and biomedia*; Elsevier B.V.: Amsterdan, 2007.

[36] Badger, R. M.; Bauer, S. H. *J. Chem. Phys.* 1937, *5*, 839-851.

[37] Andrey, A. S.; Mikhail, D. B.; Boris, N. S. *J. Phys. Org. Chem.* 1996, *9*, 241-251.

[38] Iogansen, A. V. *Spectrochim. Acta, Part A* 1999, *55*, 1585-1612.

[39] Jesus, A. J.; Redinha, J. S. *J. Mol. Struct.* 2009, *938*, 156-164.

[40] Ceccarelli, C.; Jeffrey, G. A.; McMullan, R. K. *Acta Crystallogr. B* 1980, *36*, 3079-3083.

[41] Desiraju, G. R. *Cryst. Eng. Commun.* 2003, *5*, 466-467

[42] Dunits, J. D. *Cryst. Eng. Commun.* 2003, *5*, 505-506.

[43] Almarsson, Ö.; Zaworotko, M. J. *Chem. Commun.* 2004, 1889-896.

[44] Sakurai, T. *Acta Crystallogr. C* 1968, *24*, 403-412.

[45] Sakurai, T. *Acta Crystallogr.* 1965, *19*, 320-330.

[46] Stock, J. T. *J. Chem. Educ.* 1989, *66*, 910-912.

[47] Vishweshwar, P.; McMahon, J. A.; Bis, J. A.; Zaworotko, M. J. *J. Pharm. Sci.* 2006, *95*, 499-516.

[48] Camenisch, G.; Folkers, G.; van de Waterbeemd, H. *Eur. J. Pharm. Sci.* 1998, *6*, 321-329.

[49] Amidon, G. L.; Lennernäs, H.; Shah, V. P.; Crison, J. R. *Pharm. Res.* 1995, *12*, 413-420.

[50] Vishweshwar, P.; McMahon, J. A.; Zaworotko, M. J. *Crystal Engineering of Pharmaceutical Co-crystals*; Frontiers in Crystal Engineering; John Wiley & Sons, Ltd, 2006.

[51] Porter Iii, W. W.; Elie, S. C.; Matzger, A. J. *Cryst. Growth Des.* 2008, *8*, 14-16.

[52] P. Vishweshwar; J. A. McMahon; Zaworotko, M. J. *Crystal Engineering of Pharmaceutical Co-crystals*; Frontiers in Crystal Engineering; John Wiley & Sons: Chichester, 2006.

[53] William Jones; W. D. Samuel Motherwell; Trask, A. V. *MRS bulletin* 2006, *31*, 875-879.

[54] Hilfiker, R. *Polymorphism in the Pharmaceutical Industry*; Wiley-VCH: Weihheim, 2006.

[55] McCrone, W. C. *Polymorphism in Physics and Chemistry of the Organic Solid State*; Wiley Interscience: New York, 1965; Vol. II.

[56] Bernstein, J. *Polymorphism in Molecular Crystals*; Clarendon Press: Oxford, 2002.

[57] Threlfall, T. L. *Analyst* 1995, *120*, 2435 - 2460.

[58] Lehmann, M. S.; Nunes, A. C. *Acta Crystallogr. B* 1980, *36*, 1621-1625.

[59] Lehmann, M.; Koetzle, T.; Hamilton, W. *J. Chem. Crystallogr.* 1972, *2*, 225-233.

[60] L. Borka; Backe-Hansen, K. *Acta Pharm. Suec.* 1968, *5*, 271-278

[61] Armando, J. A.; John, K., Jr.; Arlyn, W. K.; Joseph, C. S. *J. Pharm. Sci.* 1967, *56*, 847-853.

[62] Kanenewa, N.; Otsuka, M. *Chem. Pharm. Bull.* 1985, *33*, 1660-1668.

[63] Szulzewsky, K.; Kulpe, S.; Schulz, B.; Kunath, D. *Acta Crystallogr. B* 1981, *37*, 1673-1676.

[64] Gamberini, M. C.; Baraldi, C.; Tinti, A.; Rustichelli, C.; Ferioli, V.; Gamberini, G. *J. Mol. Struct.* 2006, *785*, 216-224.

[65] Szulzewsky, K.; Kulpe, S.; Schulz, B.; Fichtner-Scmitter, H. *Acta Pharm. Suec.* 1982, *19*, 457-470.

[66] Perlovich, G.; Volkova, T.; Bauer-Brandl, A. *J. Therm. Anal. Calorim.* 2007, *89*, 767-774.

[67] Di Martino, P.; Conflant, P.; Drache, M.; Huvenne, J. P.; Guyot-Hermann, A. M. *J. Therm. Anal. Calorim.* 1997, *48*, 447-458.

[68] Haisa, M.; Kashino, S.; Kawai, R.; Maeda, H. *Acta Crystallogr. B* 1976, *32*, 1283-1285.

[69] Haisa, M.; Kashino, S.; Maeda, H. *Acta Crystallogr. B* 1974, *30*, 2510-2512.

[70] Gary, N.; Christopher, S. F. *J. Pharm. Sci.* 1998, *87*, 684-693.

[71] Ivanova, B. B. *J. Mol. Struct.* 2005, *738*, 233-238.

[72] Sohn, Y.-T. *J. Korean Pharm.Sci.* 1990, *20*, 97-103.

[73] Nath, B. S.; Khalil, S. S. *Indian J. Pharm. Sci.* 1984, *46*, 106-110.

[74] Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharm. Res.* 2001, *18*, 859-866.

[75] Gavezzotti, A. *J. Am. Chem. Soc.* 1991, *113*, 4622-4629.

[76] Gavezzotti, A. *Acc. Chem. Res.* 1994, *27*, 309-314.

[77] Harris, K. D. M.; Tremayne, M.; Lightfoot, P.; Bruce, P. G. *J. Am. Chem. Soc.* 1994, *116*, 3543-3547.

[78] Blagden, N.; Cross, W. I.; Davey, R. J.; Broderick, M.; Pritchard, R. G.; Roberts, R. J.; Rowe, R. C. *Phys. Chem. Chem. Phys.* 2001, *3*, 3819-3825.

[79] Philips, R. *Crystals, Defects and Microstructures: Modeling Across Scales*; Cambridge University press: Cambridge, 2001.

[80] Abraham, F. F. *Homogeneous nucleation theory*; Academic Press: New York, 1974.

[81] Peterson, M. L.; Morissette, S. L.; McNulty, C.; Goldsweig, A.; Shaw, P.; LeQuesne, M.; Monagle, J.; Encina, N.; Marchionna, J.; Johnson, A.; Gonzalez-Zugasti, J.; Lemmo, A. V.; Ellis, S. J.; Cima, M. J.; Almarsson, A. *J. Am. Chem. Soc.* 2002, *124*, 10958-10959.

[82] Leitão, M.; Canotilho, J.; Sousa, A.; Pais, A.; Sousa, A.; Simões Redinha, J. *J. Therm. Anal. Calorim.* 2003, *73*, 763-774.

[83] Canotilho, J.; Costa, F.; Sousa, A.; Redinha, J.; Leitão, M. *J. Therm. Anal. Calorim.* 1998, *54*, 139-149.

[84] Leitão, M. L. P.; Canotilho, J.; Ferreira, S. C. R.; Sousa, A. T.; Simões Redinha, J. *Thermochim. Acta* 2004, *411*, 53-60.

[85] Sousa, A. F.; Canotilho, J.; Pais, A. A. C. C.; Leitão, M.; Redinha, J. S. *Mol. Phys.* 2003, *101*, 871-879.

[86] Erdemir, D.; Chattopadhyay, S.; Guo, L.; Ilavsky, J.; Amenitsch, H.; Segre, C. U.; Myerson, A. S. *Phys. Rev. Lett.* 2007, *99*, 115702.

[87] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* 2005, *105*, 2999-3094.

[88] Lopes Jesus, A. J.; Tomé, L. I. N.; Rosado, M. T. S.; Leitão, M. L. P.; Redinha, J. S. *Carbohydr. Res.* 2005, *340*, 283-291.

[89] Berkovitch-Yellin; Addadi, L.; Idelson, M.; Lahav, M.; Leiserowitz, L. *Angew. Chem. Suppl.* 1982, 1336-1345.

[90] Gao, Q.; Jeffrey, G. A.; Ruble, J. R.; McMullan, R. K. *Acta Crystallogr. B* 1991, *47*, 742-745.

[91] Blake, C. C. F.; Small, R. W. H. *Acta Crystallogr. B* 1972, *28*, 2201-2206.

[92] Garside, J.; Davey, R. J.; Jones, A. G. *Advances in industrial crystallization* Butterworth Heinemann: Oxford, 1991.

[93] Popovitz-Biro, R.; Weissbuch, I.; Jacquemain, D.; Leveiller, F.; Leizerowitz, L.; Lahav, M. In *Advances in industrial crystallization* Garside, J., Davey, R. J., Jones, A. G., Eds.; Butterworth Heinemann: Oxford, 1991, p 3-19.

[94] Canotilho, J. *Ph.D. thesis*; University of Coimbra: Coimbra, 1999.

[95] Zabel, V.; Müller-Fahrnow, A.; Hilgenfeld, R.; Saenger, W.; Pfannemüller, B.; Enkelmann, V.; Welte, W. *Chem. Phys. Lipids* 1986, *39*, 313-327.

[96] Shimon, L. J. W.; Wireko, F. C.; Wolf, J.; Weissbuch, I.; Addadi, L.; Berkovitch-yeelin, Z.; Lahav, M.; Leiserowitz, L. *Mol. Cryst. Liq. Cryst.* 1986, *137*, 67-86.

[97] Wireko, F. C.; Shimon, L. J. W.; Frolow, F.; Berkovitch-Yellin, Z.; Lahav, M.; Leiserowitz, L. *J. Phys. Chem.* 1987, *91*, 472-481.

[98] Davey, R. J. *J. Cryst. Growth* 1986, *76*, 637-644.

[99] Elwenspoek, M.; Bennema, P.; van der Eerden, J. P. *J. Cryst. Growth* 1987, *83*, 297-305.

[100] Shimon, L. J. W.; Vaida, M.; Addadi, L.; Lahav, M.; Leiserowitz, L. *J. Am. Chem. Soc.* 1990, *112*, 6215-6220.

[101] Davey, R. J.; Blagden, N.; Potts, G. D.; Docherty, R. *J. Am. Chem. Soc.* 1997, *119*, 1767-1772.

[102] Kitamura, M. *J. Cryst. Growth* 1989, *96*, 541-546.

[103] Turnbull, D.; Fisher, J. C. *J. Chem. Phys.* 1949, *17*, 71-73.

[104] Buckle, E. R. *Proc. R. Soc. London, A* 1961, *261*, 189-196.

[105] Bartell, L. S. In *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; Gavezzotti, A., Ed.; John Wiley & Sons: Chichester, 1997.

[106] Ropp, R. C. *Solid State Chemistry*; Elsevier Science & Technology Amsterdam, 2003.

[107] Boucher, E. A. In *nucleation*; Zettlemoyer, A. C., Ed.; Marcel Dekker, Inc: New York, 1969, p 527-572.

[108] Kwokal, A.; Nguyen, T. T. H.; Roberts, K. J. *Cryst. Growth Des.* 2009, *9*, 4324-4334.

Chapter 4

SPECTROSCOPIC AND MICROSCOPIC EXAMINATION OF CHIRAL RECOGNITION AT THE MOLECULAR LEVEL

Marek Graff

Department of Chemistry, Warsaw University, Poland

ABSTRACT

Molecular recognition can be defined as recognition of molecules, i.e. polar molecules by other polar molecules, or recognition of chiral molecules by other chiral molecules. These molecules can interact together, for instance, by creating hydrogen bonding. This review shows using of microscopy, spectroscopy and other methods to investigate of chiral recognition (i.e. chiral surfaces).

The term of chiral surface can be defined as surface of metal (i.e Cu, Au) or non–metal (i.e graphite) covered by chiral molecules. The covering ratio of surfaces can be different, and it depends of method specification used in appropriate experiment. In this review I want to show, so chiral molecule (i.e. α–amino acids) adsorbed on surface (i.e. metal) can recognize another chiral molecule also adsorbed on surface. The useful methods discussed in this paper are: STM (scanning tunneling microscopy), AFM (atomic force microscopy), electrochemistry and vibrational spectroscopy – infrared and Raman, including SERS effect (surface–enhanced Raman scattering) and others. Molecular recognition can be investigated, for instance, when molecules create i.e. the SAMs or Langmuir–Blodgett films.

The most sensitive methods are STM, AFM and Raman spectroscopy (SERS), because the most interesting results were obtained using these methods. I would like to underline, so every method has its own limitation and specification, and results are strong depended from using method. Structural information obtained using various techniques for chiral surfaces can be useful in understanding of key of interaction of adsorbed chiral molecules.

INTRODUCTION

The term 'chirality' comes from Greek word '*cheir*', what means '*hand*'. Louis Pasteur, French chemist and biologist, was a pioneer of chirality: he carried out first experiment with chiral species: manually separating of sodium–ammonium tartrate enantiomers [1-4]. The term 'chirality' was defined by Lord Kelvin [5]. According to his definition, chiral is 'any geometrical figure, or group of points', and chirality 'if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself' [5]. Chiral carbon – stereogenic centre of sp^3 hybridization – is that, which bears four different subsitituents. Chirality is important because most biomolecules, like amino acids, sugars, proteins and nucleic acids are chiral. However, every chiral molecule has two enantiomers, but only one can be biochemically active, and second – inactive or toxic. This is starting point for enantioseparation.

Important aspect is a chiral selector, used to separation of enantiomers, and creation of diastereoisomeric complex. Molecular recognition might be a result of free energy of Gibbs (ΔG) difference between two diasteroisomeric pairs (enantiomer–selector). Interaction between these pairs might be described by three–point model (Easson and Stedman[6], Ogston [7], Dalgliesh [8]), or selector's three point site [9] presented on Figure 1. In few words, interaction between enantiomer 1 and selector can be attractive in three points, but between enantiomer 2 and selector can be attractive only in two points [9]. In addition, interaction between enantiomer and selector can be attractive or repulsive, and *summa summarum* if attractive interaction are stronger than repulsive one, it might promote the creation diastereoisomeric pairs [9]. Some researches do not accept this model, however, this model in only geometrical model [9].
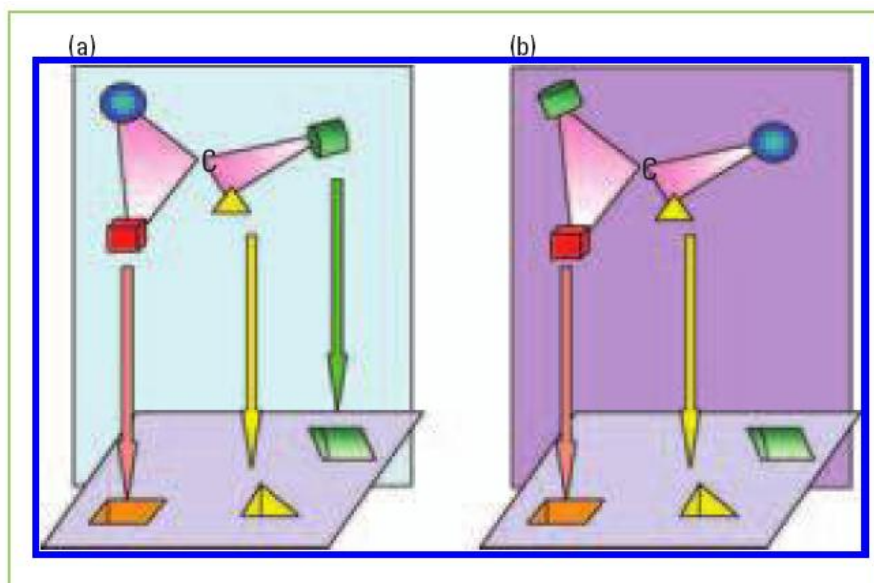
Figure 1. The three-point interaction model. (a) A chiral molecule with an asymmetric carbon atom can present three groups that can match three sites of the selector. (b) Mirror image of this molecule, after hypothetical rotations, could present a maximum of two groups, which can interact with only two sites of the selector [9].

The strongest interaction between molecules or part of them, are hydrogen bonding and coulomb/electric interaction. $\pi - \pi$ interaction are strong, and they are observed, if aromatic and $-NO_2$ groups are present (π–donating, π–basic, in general – where π–electron can delocalize) [9]. Characteristics of molecular interaction is presented in Table 1 [9].

Table 1. Characteristics of molecular interaction [9]

Type of interaction	Strength	Direction	Range (d)
Coulombic or electric	Very strong	Attractive or repulsive	Medium ($1/d^2$)
Hydrogen bond	Very strong	Attractive	Long
Steric hindrance	Very strong	Repulsive	Very short
$\pi-\pi$	Strong	Attractive or repulsive	Medium
van der Waals	Very weak	Attractive	Very short

In this paper some methods like STM, AFM, Raman (including SERS effect) and IR using for investigation of chiral surfaces will be presented. If it is possible (if current knowledge let do it), I try to show, why enantiomers and racemic mixture can interact with other enantiomers by different way.

INFRARED

The most known and widely used vibrational spectroscopy method is infrared absorption. This subset spectroscopy deals with infrared region of the electromagnetic spectrum. IR spectroscopy can be used to investigate sample composition and identify compounds. The infrared spectrum can be divided into 3 regions – for their relation to visible spectrum: near–IR, mid– and far– infrared. Near–IR, (14000–4000 cm^{-1}) can be used to study overtone and harmonic vibrations. Mid–IR region (4000–140 cm^{-1}), the most widely known and can be used to rotational–vibrational structure. The last region, far–infrared one (400–10 cm^{-1}), lying to microwave region, may be used in studying, for instance, of rotational structure.

Molecules can adsorb appropriate frequencies which are characteristic of their structure. The frequencies of absorbed radiation are related to vibrating bonds or groups. In general, molecules which absorb infrared radiation, change their own energy from lower vibrational state to higher vibrational state by $h\omega/2\pi$. In other words, adsorption is possible in that, when electric field of incident radiation resonates with electrical moment of vibration in the molecule. Molecular vibrations are active in IR spectra, if the dipole moment changes. For example, a symmetrical stretching of CO_2 molecule, or vibration of N_2 are inactive in IR spectra, but seen only in Raman spectra. Polar groups or bonds are corresponded to strong infrared absorption.

For molecule many vibrational ways are possible and these way are known as vibrational modes. For linear molecules, are possible $3N$–5 vibrational modes, for nonlinear molecules – $3N$–6 vibrational modes (or degree of vibrational freedom). 'N' means the number of atoms in molecule.

Sample preparation: a long path length is needed to acquire spectra of gas samples (diluteness compensation). For absorption IR spectra of liquid samples, they are sandwiched between two plates of salt (i.e. NaCl, KBr, CaF_2). These plates are transparent to IR radiation and do not interfere with lines onto the spectra. Solid samples are the most often prepared by grinding a quantity of the sample with a salt, i.e. KBr, specially purified. Using mechanical press a powder mixture is pressed to form a pellet, and beam in spectrometer can pass through it.

After acquiring IR spectra, Fourier transformation is used because it is easier and cheaper a building interferometer than the monochromator fabrication. Additionally, all frequencies are collected using FT.

RAMAN

If a sample is treated by monochromatic radiation frequency ω_0, it is occurred some scattering radiation. The scattered radiation can be divided on: radiation with the same frequency (ω_0), radiation with decreased frequency, known as Stokes bands ($\omega_0 - \omega$) and increased frequency, known as anti Stokes bands ($\omega_0 + \omega_0$). These frequencies are observed in Raman spectrum, with different intensity. In Raman experiments, usually Stokes bands are analyzed. It is necessary for energy conservation, so photon annihilation process with the photon energy of $h\omega_0/2\pi$ and photon creation process with the photon energy of $h(\omega_0 - \omega_0)/2\pi$ and in result, it is leaded a molecule transition and inelastic scattered is occurred. The molecule is transferred on higher energy state. It is possible observation of vibrational, rotational and electronic transitions but only vibrational transitions are most widely used [10].

The most difficulty in Raman spectroscopy is very small Raman cross section of molecules, for simply Raman scattering process. For that reason, Raman signals can be measured only after amplification by many orders of magnitude. One from way to increasing of Raman signal is using the laser excitation wavelength. More efficient inelastic scattering of radiation are known as resonance Raman (RR) scattering. Well–known effect of increasing of intensity of bands for molecules adsorbed on surfaces of noble metals (Cu, Ag, Au) is surface enhanced Raman scattering (SERS). The factor of surface enhancement can be as much as 10^{14}.

The SERS effect was discovered by M. Fleischmann and co. in 1974 [11]. They observed strong increasing of bands in Raman spectra of pyridine adsorbed on Ag surface. Anyway, they did not recognize that this was a major enhancement effect. They suggested so this effect is a matter of number of molecules adsorbed on metal surface. Later this effect was interpreted more correctly: concentration of scattering species did not account for enhanced signal. Jeanmaire and Van Dyune [12] and Albrecht and Creighton [13] plus co–workers proposed independently two mechanisms: chemical and electromagnetic, respectively [14–23]. It is possible, so SERS effect is related to existence of surface plasmon (Ritchie) [18]. Two theories are known of exact mechanism of the enhancement effect of SERS, substantially different from each other. Electromagnetic mechanism is connected with excitation of localized plasmon surface, as chemical effect is connected with formation of charge transfer complexes. Electromagnetic theory can be applied only to species physisorbed

to the metal surface, while chemical theory can be applied only to samples which formed a bond with the surface.

The electromagnetic enhancement comes from interaction electrons in the metal surface with incident electric field from incident radiation. The consequence this interaction is enhancement of the electric field at the surface of metal. For excitation of surface plasmons, the rough surface is needed, and metal surfaces should be specially prepared for SERS experiments.

Chemical mechanism is known as CT mechanism (charge transfer). The 'chemical' enhancement is similar to resonance Raman process, which occurs in complexes such as metal–ligand. It is offered new electronic transition for metal electrons, when molecules are adsorbed at metal surface. It can be considered a hypothetical situation, so electron at Fermi energy level could be excited into molecular orbitals of adsorbed molecule and back to metal state. It is possible also another situation: the electrons at highest occupied molecular orbital could be transferred into Fermi level of the metal and back to the molecule. It is possible to change of the Fermi level energy, i.e. in electrochemical systems it could be realized by change of applied potential.

It is important, so electric field enhancement strong decreases with the distance from metal surface. The enhancement of charge transfer is observed for molecules directly interacting or bands directly oriented to metal surface. However, enhancement factors for various bands have a different values. Additionally, if molecules are adsorbed to metal surface, the symmetry of system could be modified, and symmetry of molecule could be changed (i.e. lost of symmetry centre). In consequences, it leads to variation in modes selection. For example, loss of centre of symmetry reduce of necessity of the mutual exclusion rule, which says, that modes can be active only in R or IR.

In SERS spectra, the surface enhancement of Raman cross section of bands, which comes from adsorbed molecules, has been related to the morphology of metal surface. The SERS–active metals (i.e. Cu, Ag and Au) can be prepared in roughening procedures. It is accepted electrochemical roughening, what could be carried out in a three–electrode cell, with the silver (or gold/copper) electrode as working electrode, a platinum sheet as counter–electrode, and saturated calomel electrode (SCE) as the reference. All potentials presented below are quoted vs. SCE. The process of silver roughening is done by 3 cycles from –0.3 V to 0.3 V (sweep rate of 5 mV/s), with the end in –0.3 V, in 0.1 M KCl aqueous solution. At the end silver electrode is kept at –0.4 V for 5 min. The procedure of roughening of gold is different: it is done 20 cycles, from –0.6 V to 1,25 V (sweep rate of 200 V/s), to –0.6 V and at this potential gold electrode is kept for 5 min.

After that, the working electrode are removed from KCl solution, at open circuit potential, rinsed by distilled water and put into solution with SERS active species. Probably, the electrochemical roughening of silver in KCl, do enlarge in the fraction of facets similar to (100) and (110), and a diminish in those similar to (111) crystallites. Another ways of the obtain large SERS enhancement factor is preparing metal soles [24, 25]. The most disadvantage of this method is impossibility of carrying out the experiments with added potential.

STM

STM, scanning tunneling microscope, is a method for imaging of surfaces at atomic level [26–28]. With the STM resolution – 0.1 nm lateral and 0.01 nm depth, it is possible individual molecules imaging and manipulation. This method could be used not only in high vacuum in 10–20 K temperatures, but also in water, air, and in wide range of temperatures – from not far from 0 K to few hundred degrees of K.

The STM conception is based on quantum tunneling. If a metal tip is taken very close to surface for examination, a bias (voltage difference) applied between the two ones can let electrons tunnel through the vacuum between them. The consequential tunneling current is a function of: applied voltage, tip position, and local density of states (LDOS) of the sample. If the tip scans across the surface, information is acquired by supervising the current, and it is presented in image form. The sample of surfaces, which are examined using the STM, should be extremely clean and stable.

The procedure in STM experiment is as follow: a voltage is applied and the tip is moved close across the surface by some coarse surface–to–tip control. If tip and surface are adequately close, the voltage is switch off. Additionally, good control of the tip in all three dimensions, if typically piezoelectric not far from the surface, continuing tip–sample separation (distance in range 0,4–0,7 nm). We can say, so equilibrium position tip–surface is between attractive (0,3–1,0 nm) and repulsive (< 0,3 nm) position. If voltage bias cause electrons to tunneling between surface and tip, it will create a current, that could be measured. If electron tunneling is stable, the bias and position of tip to the surface, could be changed. The STM data is obtained as a result of current changing.

If the tip is brought to the surface (x–y plane), will cause a current varying the change in surface height and states density. As a result, the current changes are mapped in images. It is known as constant height method, CHM. Additionally, a constant current could be measured, if the tip is moved across z–plane. It is known as constant current method (CCM) or constant gap width mode (CGM). In the CCM mode, a height variation comes from the tip topography transversely the surface and constant charge density is obtained. In other words, variations in charge density due a contrast on the image. In constant height method (CHM), the height and voltage do not change, and current changes to keep voltage unchanged. An image obtained using CHM is made of current changes over the sample and it may be associated to charge density. The using CHM is more beneficial than CCM, i.e. it is faster. It is also possible with STM technique, to receive information on electronic structure, when the sample is scanned.

AFM

Atomic force microscopy (AFM) is very high resolution type of scanning probe microscopy, with resolution of 0,01–0,1 nm, what is 1000 times more than optical diffraction limit [29, 30]. The AFM consists of a beam (cantilever) with a sharp tip at this end for scanning the sample. If the tip is moved nearness of surface, forces between the tip and the surface direct to a deflection of the cantilever according to Hooke's law. Additionally, forces measured in AFM contain mechanical contact force, van der Waals force, capillary forces, chemical bonding, electrostatic forces, magnetic forces, etc. In AFM experiment, the

deflection is précised using laser spot reflected from the top surface of the cantilever into an array of photodiodes. Also others techniques are used − capacitive sensing or piezoresistive AFM cantilevers and optical interferometry.

If the surface was scanned by the tip at unchanged height, a risk of causing damage could happen when the tip collides the surface. For adjusting the tip–to–sample distance to keep on a constant force from the tip and sample mostly feedback mechanism is used. Typically, the sample is mounted on a piezoelectric tube moving the sample in the z–direction with unchanged force. The sample is scanned in x and y directions. It is also possible to employ a tripod configuration of three piezo crystals and are each responsible for examination in x, y and z direction. Number of AFM modes could be operated (it depends on application). First are static (known also contact) and dynamic or non–contact modes when beam is vibrated. If we operate a static mode of AFM, the not moving tip deflection is used as a response signal. Additionally, static mode of AFM is done in contact when overall force is foul. Summary, for this mode the force between the tip and sample is kept constant if the surface is scanned by continuing a constant deflection. Second mode are dynamic − the cantilever is oscillating on the surface and amplitude of vibration, phase and resonance frequency could be modified by tip–sample contact forces. Vibration frequency could be measured with extreme sensitivity and mode of frequency modulation permit for use very hard cantilevers, which offer stability very close to the surface. The AFM provide real atomic resolution in ultra–high vacuum conditions.

The AFM have advantages and disadvantages. First: it provides true three dimensional surface profile and samples do not need special treatment (i.e. metal or carbon coatings) what could destroy irreversibly the sample. The AFM modes can do perfectly in liquid environment or ambient air. Consequently, it is possible studying of biological macromolecules or living organisms. Second: it is scanning area only of about 150x150 μm and image area 10–20 μm.

CHIRAL SURFACES AND THEIR INVESTIGATION BY STM AND AFM

It is possible to examine enantioselectivity of molecules on surfaces, at molecular level. The determination of the enantiomers adsorbed on metal surface is possible using the methods like scanning tunneling microscopy or atomic force microscopy.

Chiral recognition and molecular interaction between cysteine molecules adsorbed on Au (110) was studied using STM by Kühnle et al. [31]. They reported so molecular pairs of cysteine were formed on metal surface. It is interesting, so it was observed protrusions of cysteine at rows created by gold atoms. No isolated molecules (unpaired protrusions) were observed at Au (110) (Figure 1 from [31][1]).

The L–cysteine pairs at Au(110) created double–lobe features, which is always rotated clockwise (20 degrees) due to 110 surface direction. The pairs of D–cysteine created similar to L–cysteine pairs, but they are rotated anticlockwise (the value of angle as the same, 20 degree). The described cysteine deposition on Au (110) should be a consequence of enantioselectivity of cysteine molecules themselves. If racemic cysteine (DL) was adsorbed on Au surface, it was identified pairs of cysteine, in LL form and DD form, and pairs of DL–

---

[1] available on http://www.nature.com/nature/journal/v415/n6874/fig_tab/415891a_F1.html#figure-title

cysteine were not observed. Dimerization of cysteine in strong enantioselective – and binding is possible only with molecule of the same chirality. The pairs of LL and DD cysteine adsorbed on Au (110) are related by mirror symmetry to each other. The pairs of enantiomeric cysteine (DD or LL) molecule are adsorbed at surface by sulphur and amino groups (binding with Au atoms), which stabilized both molecules additionally, and carboxyl groups of both molecules create double hydrogen bonding among these groups (Figure 3 from [31][2]). The hypothetical (but not observed) pairs of DL–cysteine would be less stable, because for DL pairs only interaction between sulphur and golds atoms, and the double hydrogen bonding between carboxylic groups would be preserve. Additionally, interaction between amino groups and gold atoms would be lost, and in consequence the stabilization of hypothetical LD–dimers is poorer in comparing to DD or LL dimers. Second way is that, so if interaction sulphur – gold and amino – gold is preserved, the double hydrogen bonding among carboxylic groups is lost.

The DFT ab initio calculation were performed for enantiomeric and raceminc pairs of cysteine. In results, heterodimer is less stable than homodimer by circa 0.2 eV (if amino group – gold interaction is lost) or 0.5 eV (when carboxylic groups do not interact).

The term 'chiral surfaces' should be defined – it is a surface of metal or no–metal (i.e. graphite), which has been covered by chiral species. Some experiments have been carried out and described in details and I would like to present results of them.

The electrodeposition films of monoclinic low–symmetry CuO on high–symmetry surfaces of Au(001) was carried out by Switzer and co. [32]. Both species were achiral, of course. In next step, epitaxial film of CuO on Au(001) was electrodeposited in chiral solution of Cu(II)(S,S)–tartrate. For determination of absolute configuration of both films the analysis of X–ray pole figures was needed. The sample was moved with sequence of tilt angles, χ, and additionally, sample was rotated (angles from 0° to 360°). The electrodeposition in solution of enantiomer of tartrate acid has changed orientation of film to [111]. Similar experiment was carried out with Cu(II)(R,R)tartrate soluition. The final film had a [111] orientation. Four peaks were observed for tilt angle of χ=30° (Figure 2 from [32][3]), in both experiments, independently.

R–film and *S*–film are enantiomorphs, and they are in relation to them as mirror image. Similar experiment was carried out, when CuO film on Au was electrodeposited in Cu(II)(R,S)–tartrate solution. For this chiral film also four peaks were observed and this film is mixture of equal amounts of *R*– and *S*– configurations.

Switzer et al. carried out electrochemical oxidation of CuO films incubated in tartrate solution [32]. The higher current density was registered, when CuO(II)(S,S)–tartrate was oxidized in (S,S)–tartrate solution comparing to CuO(II)(S,S)–tartrate film was oxidized in (R,R)–tartrate solution. The comparable voltamogramms were registered, if CuO(II)(R,R)–tartrate film was modified in (R,R) tartrate solution comparing to CuO(II)(R,R)–tartrate was oxidized in (S,S) tartrate solution. For CuO(II)(R,S)–tartrate film no enantioselectivity was observed during oxidation in (S,S) or (R,R) tartrate (Figure 4 from [32][4]).

---

[2] available on http://www.nature.com/nature/journal/v415/n6874/fig_tab/415891a_F3.html#figure-title
[3] available on http://www.nature.com/nature/journal/v425/n6957/fig_tab/nature01990_F2.html#figure-title
[4] available on http://www.nature.com/nature/journal/v425/n6957/fig_tab/nature01990_F4.html#figure-title
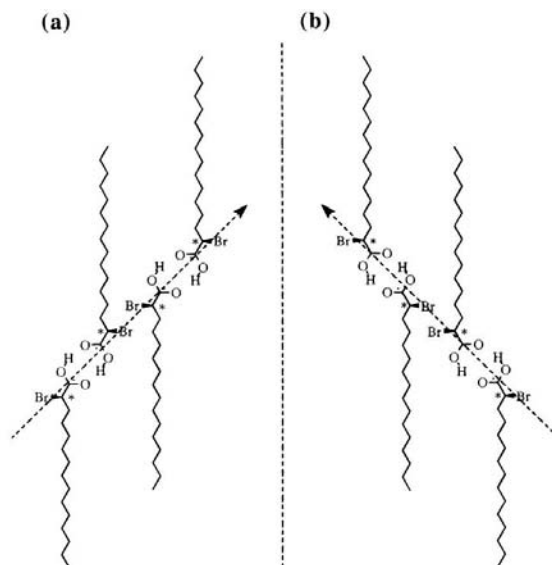
Figure 2. Schematic of the molecular structures of (*R*)-2-bromohexadecanoic acid (a) and (*S*)-2-bromohexadecanoic acid (b) physisorbed on a graphite surface with an all-trans conformation. The single chiral carbon atom on the molecules is indicated by an asterisk. Each molecule is hydrogen-bonded to the next one through the carboxyl groups [34].
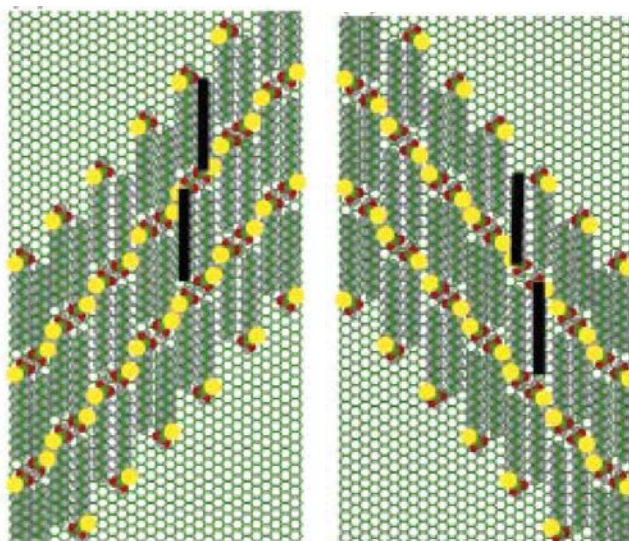
Figure 3. Top view of a model of (R)-2-bromohexadecanoic acid (a) and (S)-2-bromohexadecanoic acid (b) molecules on a graphite surface. Two black bars mark the chiral dimer formed by two molecules hydrogen-bonded through carboxyl groups. Yellow represents bromine atoms, green – carbon atoms, gray – hydrogen atoms, and red – oxygen atoms [34].

Lorenzo et al. adsorbed tartaric acid on Cu (110) surface and scanned this surface by the STM needle [33]. They observed, so tartaric acid (*S,S* or *R,R*) adsorbed on Cu surface created chains: the spatial alignment and hydrogen bonds between each neighboring molecules of enantiomeric tartaric acid on copper surface was present. The tartaric acid enantiomer chains

adsorbed on Cu surface are in relation as mirror image to each other. Similar effect was not observed for racemic tartaric acid (*R,S* or *S,R*) (Figure 2 and 3 from [33][5]).

Adsorption of enantiomeric 2–bromohexadecanoic (*S* and *R*) acid on graphite surface were carried out by Fang et al. [34]. The surface of graphite was investigated by the STM method. Similar to tartaric acid adsorbed on Cu(110), molecules of 2–bromohexadecanoic interact together: carboxylic groups of two neighboring molecules created double hydrogen bonding, and all structure is stabilized by this way. Molecular structures of *S* and *R* 2–bromohexadecanoic are mirror images of each other (Figure 2 and 3a and 3b).
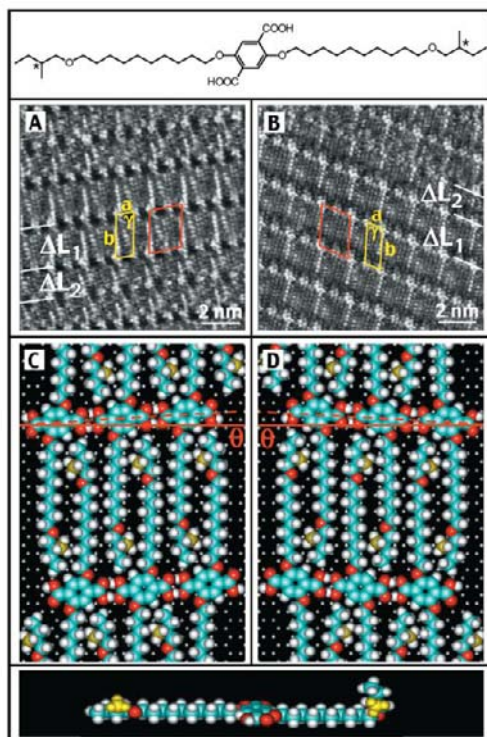
Figure 4. STM image and model of the (*S*)-enantiomer (left) and the (*R*)-enantiomer (right). ΔL_1 and ΔL_2 indicate large lamellae (2-methylbutoxy groups adsorbed) and small lamellae (2-methylbutoxy groups pointing away from the surface), respectively. Θ is the angle between the lamella axis and the graphite reference axis. (Bottom) The alkoxy chain at the left of the model represents the extended conformation, while the 2-methylbutoxy component of the alkoxy group at the right is rotated out of plane [36].

Orme et al. reported *in situ* observation using the AFM technique and molecular modelling studies of calcite ($CaCO_3$) covered by enantiomers of aspartic acid [35]. They observed enantiospecific binding of amino acid to surface what offers the optimal chemical and geometrical fit transforms the step–edge liberated energies. Summary, they go for results in macroscopic crystal shape conversion. These observation could better help to understand mechanism of stereochemical recognition and process of growing crystal connected with

---

effect of binding on the interfacial energies. They adsorbed of racemic aspartic acid on calcite surface and mirror symmetry and chirality of aspartic acid was observed. L–Asp binds to surface by negatively charged carboxylic groups (COO^-) by coordination of calcium ions, and positively charged ammonium ions (NH_3^+) can interact with $CaCO_3^{2-}$ ions. Neighbouring enantiomeric aspartic acid molecules are stabilised by interaction of polar groups – amino and carboxylic groups, which can interact together, for instance, amino group from molecule 1 and carboxylic group from molecule 2 (Figure 3 from [35][6]).

De Feyter et al. adsorbed chiral species on graphite surface and investigated them using the STM technique [36]. Their researches were joined with simultaneous analysis of chirality and conformation of enantiomer terephthalic acid derivative – 2,5–bis[10–(2–methylbutoxy)decyloxy]terephthalic acid. This species has two identical stereogenic centres, localised on both ends of carbon chains (Figure 4). For this terephthalic acid derivative, the width of the wide lamellae – $\Delta L_1 = 2.54 \pm 0.05$ – is the same like length extended alkoxy chains, which are localised (laid) parallel to graphite surface. The dimension of narrow lamellae – $\Delta L_1 = 1.9 \pm 0.1$ – shows, so decyloxy groups are lying horizontal on the surface, but end 2–methylbutoxy groups direction are away from the surface. The 2,5– bis[10–(2–methylbutoxy)decyloxy]terephthalic acid molecules are attached to the graphite surface by two carboxylic groups, which can interact with carboxylic groups of neighbouring molecules of this acid and double hydrogen bonding is created. Graphite surfaces covered by enantiomers 2,5– bis[10–(2–methylbutoxy)decyloxy]terephthalic acid are mirror images to each other.

Mahapatro et al. investigated chiral hydrophobic molecules adsorption by chemical force spectroscopy [37]. This is technique unites AFM and chemical recognition by chemical task of scanning sample. It is possible to investigate interaction (hydrophilic and hydrophobic) between adsorbed molecules, for instance, on surface and on tip. They observed so adhesion forces between 3,5–dinitrobenzoylphenylglycine adsorbed on tip and surface covered by phenyl–amine were different for pairs (4 permutations) of species. Adhesion force values were similar for *RS* pairs, and lower for *RR* or *SS* pairs, 0.5–0.7±0.1 nN, and 0.2±0.1 nN, respectively. According to Johnson, Kendal and Roberts (JKR) theory of contact mechanics, a small change in contact radius might in consequence increase the adhesive force. Unfortunately, Mahapatro et al. did not discuss in details the interaction between species adsorbed on surface and tip (hydrogen bonding, van der Waals, etc.).

The Atomic Force Microscopy (AFM) offers wide application in high resolution imaging of single biological molecules: proteins, nucleotides, membranes and cells (in near–physiological environment) and intermolecular forces present them. It is possible using AFM to resolve nanometer–dimension feature, and in consequence, recognition and topography of molecules are known. It is difficult to measure physical parameters of biomolecules using other method than the AFM. Kienberger et al showed, for instance, AFM images of antibodies used for recognition of lysozyme [38]. Molecular recognition is also possible in interfaces of water–air, and in these condition decisive role them play hydrogen bonding. This recognition might be useful, if host–guest combination is present: recognition is possible by nucleic acids, interfacial receptors, etc. [39].

Bombis et al. investigated adsorption of molecules, that become chiral upon adsorption because of symmetry reduction which go after from surface detention [40]. They used

---

[6] available on http://www.nature.com/nature/journal/v411/n6839/fig_tab/411775a0_F3.html#figure-title

STM as a technique and group of linear organic molecules, like oligophenyleneethynylenes. These species were adsorbed on Au(111) under ultrahigh vacuum. They used: 1,4–bis((5–*tert*–butyl–4–hydroxy–3–formylphenyl)ethynyl)benzene   (BHA),   1,4–bis((3– formyl–4–hydroxyphenyl–)ethynyl)benzene   (HA),   1,4–bis((5–tert–butyl–3–formyl–4– methoxyphenyl)ethynyl)benzene (BH) (Figure 5). BHA adsorbed on metal surface can exist in 3 conformers: LR/RL (*meso* form, achiral), RR and LL – they are mirror–image to other one (they are enantiomers) (Figure 6).
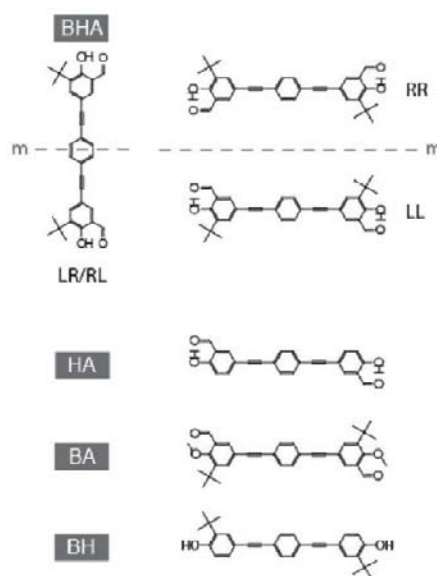
Figure 5. The formula of molecules: (BHA) 1,4-Bis((5-tert-butyl-4-hydroxy-3-formylphenyl)ethynyl)benzene; (HA) 1,4-bis((3-formyl-4-hydroxyphenyl-)ethynyl)benzene; (BA) 1,4-bis((5-tert-butyl-3-formyl-4-methoxyphenyl)ethynyl)benzene; (BH) 1,4-bis((2-*tert*-butyl-4-hydroxy)ethynyl)benzene [40].
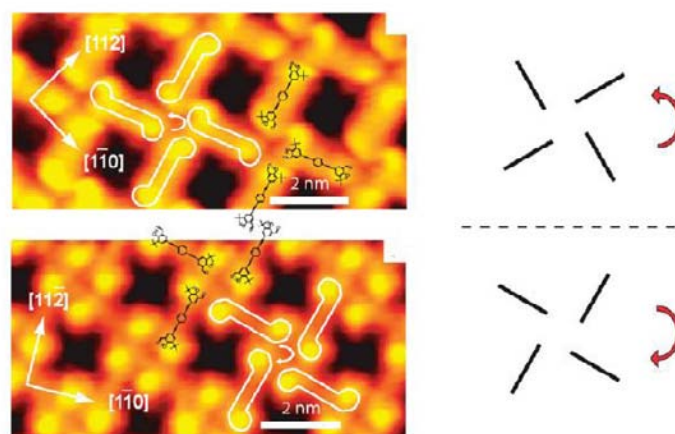
Figure 6. High-resolution STM images and structural models of grid structure formed by BHA [40].

Difference between RR and LL conformers is related to orientation of –CHO and *tert*–butyl groups on Au surface. Similar enantiomers (and *meso* form) are observed on metal surface after HA and BA species adsorption.

Both methods (STM and AFM) offer wide possibility in detecting chiral molecules adsorbed on surface. It is very difficult to detect intermolecular interaction and compare two surface samples covered by two opposite enantiomers, separately, using others methods than STM and AFM.

CHIRAL RECOGNITION AND IR AND SERS

Vibrational spectroscopy offers also wide possibility in detecting intermolecular interaction between chiral molecules adsorbed on surface as compared to STM and AFM. Additionally, the advantage of the vibrational spectroscopy to STM and AFM, is possibility of detecting, for instance, difference in molecular interaction between chiral species – adsorbed on surface – also in racemic mixture, not only between enantiomers of the same chirality, as ascribed above.

Baranska et al. analyzed vibrational specta (R and IR) of racemic and enantiomeric malic acid [41]. They observed, so malic acid spectra were different (racemic and enantiomeric), especially in range 1600–1700 cm^{-1}. For enantiomeric form was observed single band (from C=O vibration), and for racemic acid – doublet in this region. Authors argued, so in the case of racemic acid, two types of H–bonded carboxyl dimer rings were created, and in case of enantiomeric acid – only one type of H–bonded carboxyl dimer rings is present. Also interesting results were received by Ruperez and Laserna when β-cyclodextrines interacting with two enantiomers separately, but they did not discuss carefully their own results [42].
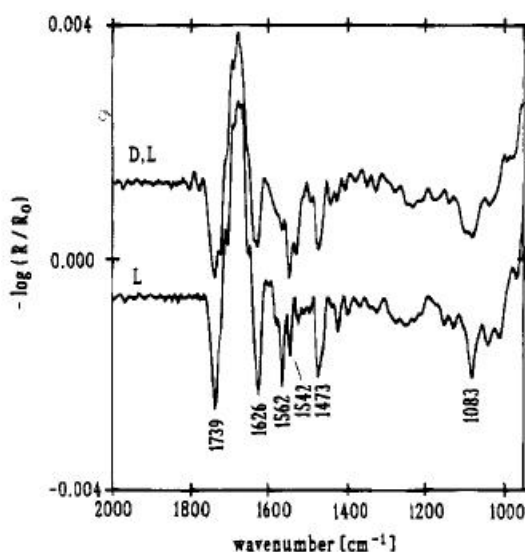
Figure 7. External infrared reflection-absorption spectra for a racemic and enantiomeric *N*-octadecanoylserine methyl ester monolayer in the range 2000-1000 cm^{-1} [43].

Gericke and Hühnerfuss investigated of chiral effects in SAM at air/water boundary. They observed chiral discrimination of *N*–octadecanoyserine methyl ester films at the air/water interface using external infrared reflection–adsorption spectroscopy (EIR–RAS): for enantiomeric form of adsorbed film, the alkyl chain was less ordered comparing to racemic one, in liquid phase [43]. The amide 1626 cm^{-1} band is broader in racemic film spectra than in enantiomeric one. Additionally, band from the C–O stretching vibration (circa 1080 cm^{-1}) is wider or split into two fragments (1095 cm^{-1} and 1079 cm^{-1}) for racemic monolayer. Additionally, the band representing stretching vibration of C–O group is narrower for enantiomeric film (Figure 7).
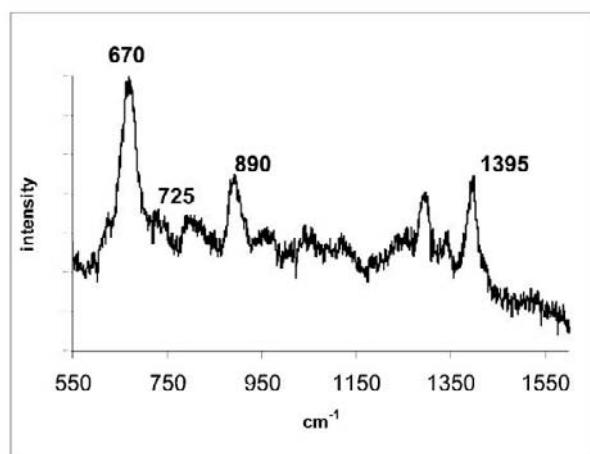
Figure 8. SERS spectrum of L-cysteine on Ag in pH 3 [44].
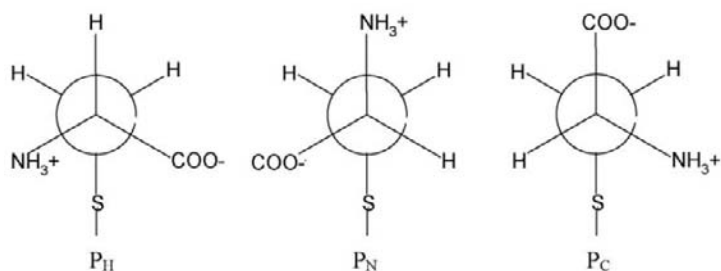
Figure 9. Cysteine conformers: P_H, P_N and P_C, according to Newman projections [44].

The idea of described below experiments [44] was verify of conclusion from paper [31].

It is widely known, so the amino acids exist in zwitterionic form in crystalline state, and in solution all polar groups are ionized (it is connected with pH). For the cysteine, amino acid containing sulphur atom in its the molecule, respective pK_a values are equal: 1.9 (for COOH), 8.4 (for SH) and 10.5 (for NH$_3^+$) [45].

Spectra of cysteine SAM on Ag surface were recordered at varying pH. Analyzing of SERS spectra can be initiated from spectra recordered for cysteine adlayer adsorbed from solution at pH 3 in 0.1 M KCl (Figure 8). Two bands (670 cm^{-1} and 725 cm^{-1}) are present in the frequency range of 650–750 cm^{-1} assigned as C–S stretching. Both bands could be

connected to C–S stretching vibration of cysteine rotamers (P_H and P_C or P_N) (Figure 9). Before cysteine molecule adsorption on Ag surface, the thiol group is ionized (–S⁻). The intensity of 725 cm⁻¹ band (related to 670 cm⁻¹) is decreased, if we compared the SERS spectra recordered after 10 min and 30 min of cysteine adsorption. Additionally, intensity of 1395 cm⁻¹ band is increased in the same interval of time (after 1 h of cysteine adsorption, the spectra is stable). Probably it is connected with % ratio of rotamers of cysteine adsorbed on Ag surface. For the rotamer P_H three polar groups of cysteine (thiol, carboxyl and amine) interact with Ag. For rotamers P_C / P_N, two groups interact with metal, thiol and amine / thiol and carboxyl, respectively. It is possible, so % ratio of P_C rotamer is decreased (for this rotamer carboxyl group does not interact with Ag), and % ratio of P_H rotamer increased (for this one carboxyl group does interact with metal). The intensity of bands in SERS spectra is connected with proximity to surface and enhancement factor exponentially decays, if the distance from the surface increases.
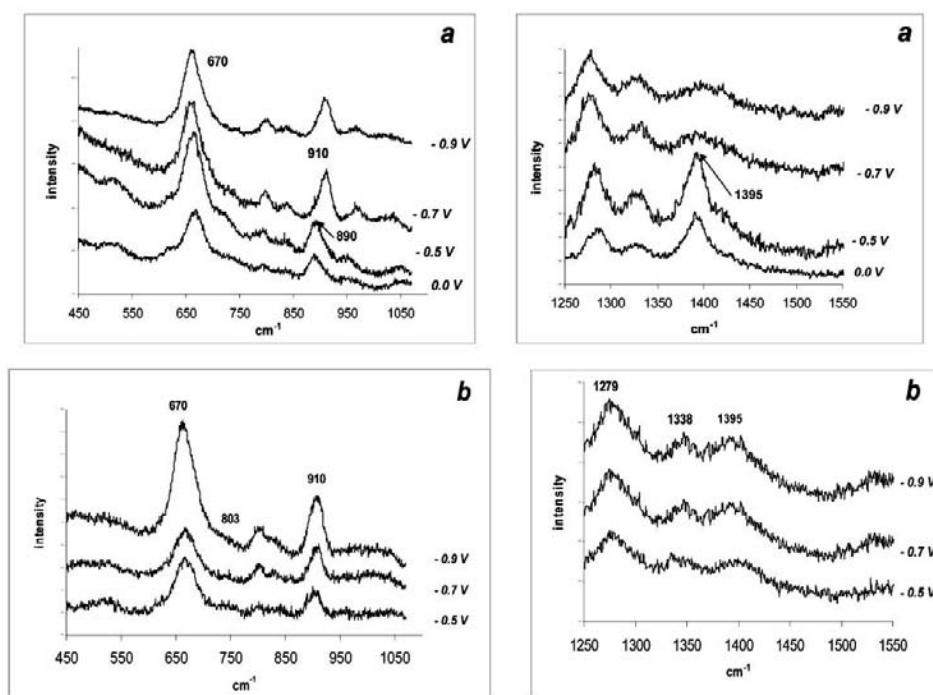
Figure 10. SERS spectra obtained for different electrode potentials from a 10^{-3} M solution of L-cysteine in 0.1 M KCl in H_2O at pH: (a) 3; (b) 13 [44].

The conformation of many molecules adsorbed on metal surface may be changed by adding potential. The potential–controlled conformation change was observed for L–cysteine using two methods: second–harmonic generation (SESHG) and SERS by Brolo [46]. To verify the conclusion presented by Brolo, the SERS spectra of cysteine were recordered at pH 3 and 13, in solution of 0.1 M KCl, from open circuit potential to − 0.9 V (representative spectra are presented on Figure 10). The most important bands of cysteine adsorbed on Ag surface are: 670 cm⁻¹ (stretching vibration of CS), 890–910 cm⁻¹ (mostly assigned to $\nu(CCOO^-)$) and ~1400 cm⁻¹ assigned to symmetrical band of COO^-, attached to Ag surface. If the added potential is lowered, the 890 cm⁻¹ and 1395 cm⁻¹ bands decrease and new 910

cm^{-1} band arises, at –0.7 V. The 890 cm^{-1} band is present only in acidic solution, at higher potential values. The 910 cm^{-1} band is observed in alkaline solution, in wide potentials range (Figure 10 b). If the experiment was carried out in heavy water solution (D$_2$O), it is observed the 975 cm^{-1} and 1395 cm^{-1} bands disappearing, and new 960 cm^{-1} band arising (Figure 11 a). Probably the 975 cm^{-1} band correspond to 890 cm^{-1} band in SERS spectra and it could be connected with labile hydrogen atoms from amino groups; thiol groups are dissociated, as mentioned above. The mode 890 / 910 cm^{-1} or 975 / 960 cm^{-1}, in H$_2$O or D$_2$O, respectively, could be ascribed as mixture (coupling) the C$_\alpha$–C stretching and N$^+$C$_\alpha$C$_\beta$ bending [47]. However, more probable is lowering the frequency, if the protons were exchanged to deuter atoms (mass effect). This effect is sometimes ascribed as 'inverse' isotope effect (increasing of frequency value of band on H/D exchange). Similar effect was observed for glicyne – the frequency was increased from 893 cm^{-1} to 1001 cm^{-1} (exchange NH$_3^+$ / ND$_3^+$) [48].
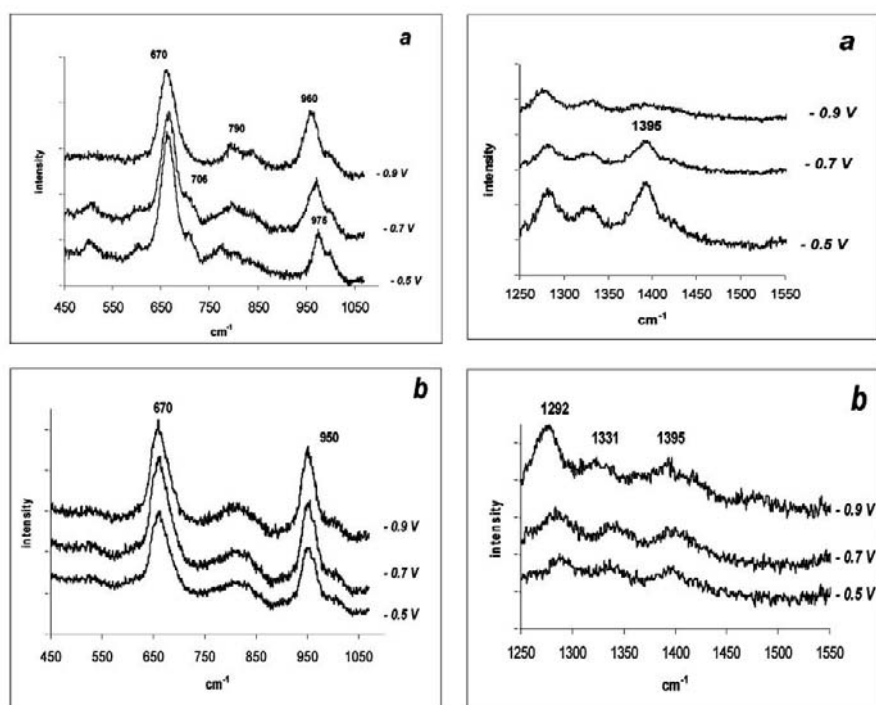
Figure 11. SERS spectra obtained for different electrode potentials from a 10^{-3} M solution of L-cysteine in 0.1 M KCl in D$_2$O at pH: (a) 3; (b)13 [44].

As it was mentioned above, if the added potential is lowered, the 1395 cm^{-1} band (carboxylic groups attached to Ag surface) is decreased. This band is strong, if potential values are higher, and disappears at lower potential values. This band is weak, if the spectra was recordered for cysteine SAM adsorbed from alkaline solution. The 1395 cm^{-1} band decreasing at applying potential if pH solution is 13 does not occur. Carboxyl groups of cysteine at pH 13 are dissociated (pK_a = 10.5) and cysteine molecule exist in this condition as

$^-$SCH$_2$CH(NH$_2$)COO$^-$. It could be possible, so similar form of cysteine is present at negative potentials at pH 3 (please compare the SERS spectra at pH 3 and 13). The observed transition in acidic solution could be linked not with protonating carboxylic groups, but it is

repulsion caused by negatively charged metal surface of them. If carboxylic groups were protonated in negative potentials at pH 3, it would be present the ~1700 cm^{-1} band in SERS spectra at these potentials. This band – 1700 cm^{-1} corresponding to C=O oscillation is usually very weak in SERS spectra. Extreme weak the C=O band was found in 1715 cm^{-1} in acidic solution in pH 2 and 1.6 (phosphate buffer), at several potential values (Figure 12). But the 1715 cm^{-1} and 890 cm^{-1} bands were present in the same condition, and the transition 890 cm^{-1} to 910 cm^{-1} may not be directly connected with protonation of carboxyl groups. Most logical assignment of 890 cm^{-1} band is $H_3N^+C_\alpha C_\beta$ and 910 cm^{-1} $H_2NC_\alpha C_\beta$ and this transition is connected with deprotonation of amino groups.
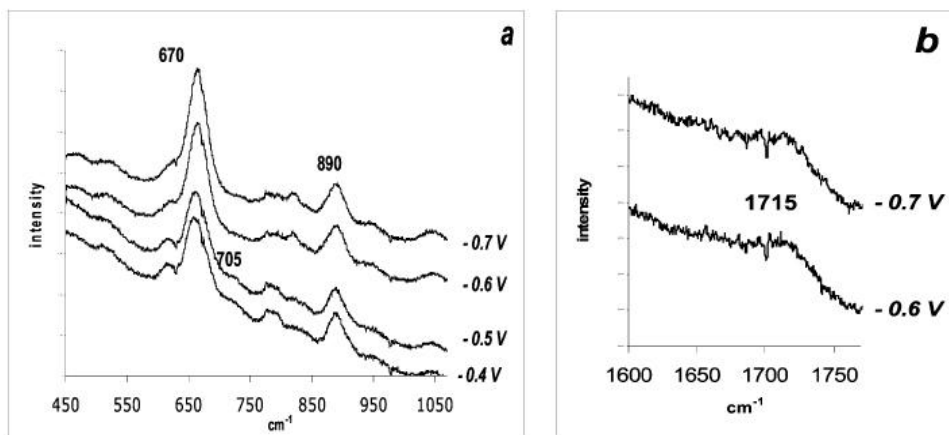
Figure 12. SERS spectra of L-cysteine adsorbed from 10^{-3} M solution in 0.1 M H_3PO_4 at pH 1.6 in the C-S and C-C stretching range (a) and in the C=O stretching range (b) [44].

The 910 cm^{-1} band was observed in spectra recordered for SAM cysteine adsorbed from neutral and alkaline solution. Probably the absence of 1715 cm^{-1} band in negative potentials at pH 3 and 0.1 M KCl solution is related to orientation of carboxylic groups – parallel to metal surface (it that case the band is not enhance by surface). However, the protonation of –COO$^-$ groups in 0.1 KCl solution at lower potentials does not allow to be definitely excluded. Similar results were also current for the D–cysteine.

Adsorption of racemic cysteine from acidic solution (pH 3) on Ag surface with added potential were carried out, but some variation were found in the spectra. The potential stimulated transition of cysteine adsorbed on metal surface at negative potential values in acidic solution, are also present in DL–cysteine spectra, similar to D– or L– cysteine spectra (Figure 13). However, the band shift value in racemic cysteine spectra is different: it is 10 cm^{-1} (the decreasing of 890 cm^{-1} band and arising 900 cm^{-1} band), and for enantiomeric cysteine it is 20 cm^{-1} (the decreasing of 890 cm^{-1} band and arising 910 cm^{-1} band). The 1395 cm^{-1} band, associated with symmetric oscillation of –COO$^-$, vanishes also in DL–cysteine SERS spectra. According to Kühnle et at. observation in STM experiment, LL or DD dimers at the metal surface are favored (they did not observe DL dimers on Au surface), because every cysteine molecule is adsorbed on the metal via suphur atom, amino group, and hydrogen bonds between both –COOH groups. As it is presented on Figure 3 from [31][7], this is most stable configuration. Additionally, similar configuration for racemic cysteine in less

---

[7] available on http://www.nature.com/nature/journal/v415/n6874/fig_tab/415891a_F3.html#figure-title

stable. If DL dimers of cysteine were created on Ag surface, it would be a mismatch in hydrogen bonds (between –COOH groups) two cysteine molecules, and probable these bonds would become weaker. But Kühnle et at. observed cysteine pairs after annealing, and agglomerates of cysteine molecules were not observed. We observe rather agglomerates of cysteine than cysteine pairs on metal surface in SERS experiment. If SAM from racemic cysteine solution is formed, the enantiomers of cysteine molecules will adsorb at Ag surface with identical probability, but interaction between molecules with opposite chirality will be weaker in comparing to molecules with the same chirality.
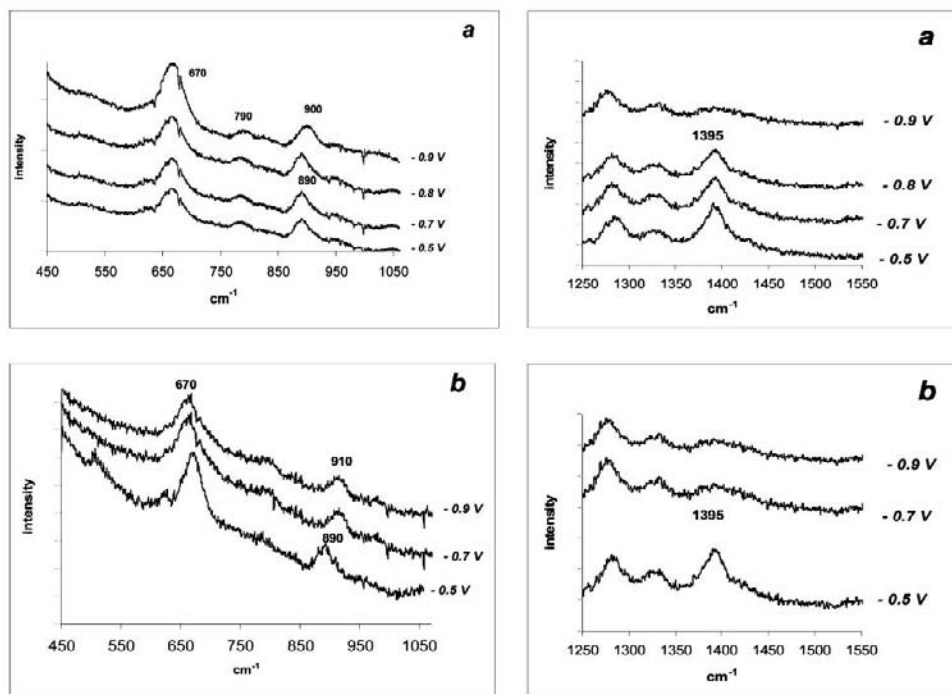
Figure 13. SERS spectra of a racemic mixture of D,L-cysteine (a) and of D-cysteine (b) adsorbed from a 10^{-3} M solution at pH 3 [44].

The 900 cm^{-1} band in cysteine SERS spectra is sensitive to interaction (inter–intramolecular): the mode form of this band is connected not only with state of carboxylic group (–COOH / –COO$^-$), but with hydrogen bond complex created by carboxylic groups. Then, shifting of frequency value for racemic cysteine (in comparing to racemic cysteine – 910 cm^{-1}) adsorbed on metal surface at low potentials, could be cleared as lowered of strength of molecular interaction between two neighboring cysteine molecules. It is assumed, so hydrogen bonding and carboxylic groups pay a decisive role in stabilizing of SAM structure of cysteine.

It is known, so cysteine molecules can interact with metal surface through sulphur atom (Ag–S or Au–S bond is created). Additionally, other polar groups – amine and carboxyl of cysteine can also interact with Ag / Au surface. Cysteine molecule can interact and can be stabilized by 3 polar groups (P_H rotamer) or two polar groups (P_N i P_C). It is possible, so metal surface covered by adsorbed enantiomer of cysteine could recognize an enantiomer other chiral species i.e. amino acid?

Silver electrode modified by L– or D– cysteine was removed to solution of phenylalanine (10^{-3} M) and SERS spectra were acquired at an interval of 30 min [49]. These series of spectra (four pairs of cysteine and phenylalanine) were presented on Figure 14. The upper spectrum in every pair cysteine and phenylalanine series is *ex–situ* spectrum. If cysteine SAM (built of L– or D– aminoacid) were maintained in contact with solution of phenylalanine enantiomer of opposite chirality, spectra was unchanged. If an electrode covered by SAM built from cysteine enantiomer and removed to solution of phenylalanine enantiomer of the same chirality, a new band appeared (730 cm^{-1}). The assignment of this band could be C–S stretching vibration of cysteine rotamers – P_N or P_C.
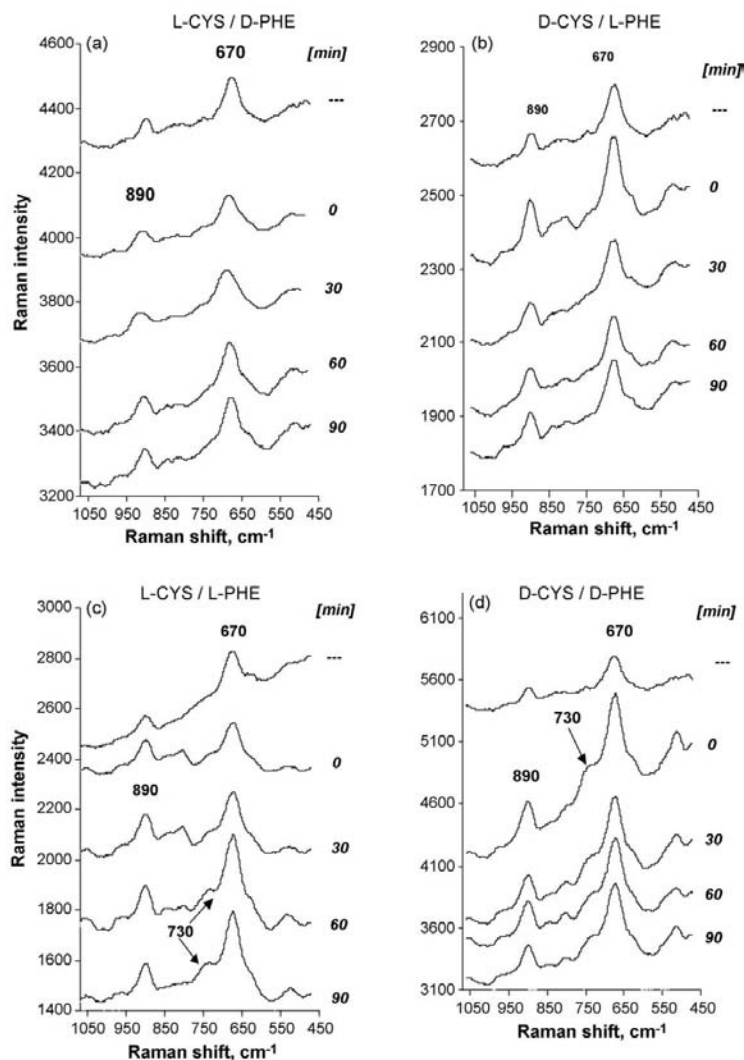
Figure 14. Series of the SERS spectra on Ag electrode covered with SAM of L- or D-cysteine in contact with a 0.001 M phenylalanine solution (L- or D-). Abbreviations: Cysteine—CYS, phenylalanine—PHE (the spectra were shifted along the vertical axis to enhance the clarity of presentation) [49].

It can be helpful analyzing higher–frequency part of spectra for deciding, which conformer is present beside the dominating P_H conformer on metal surface, if enantiomers of

cysteine and phenylalanine of the same chirality interact together. In the Figure 15 is presented the SERS spectra within 1250–1550 cm^{-1} of SAM of L–cysteine on Ag surface after immersing into L–phenylalanine solution. The band at about 1400 cm^{-1}, with an assignment to stretching vibration of the dissociated carboxyl groups, remains unchanged in time. In other experiments (corresponding to remaining of cysteine–phenylalanine pairs of enantiomers) were observed the same results.
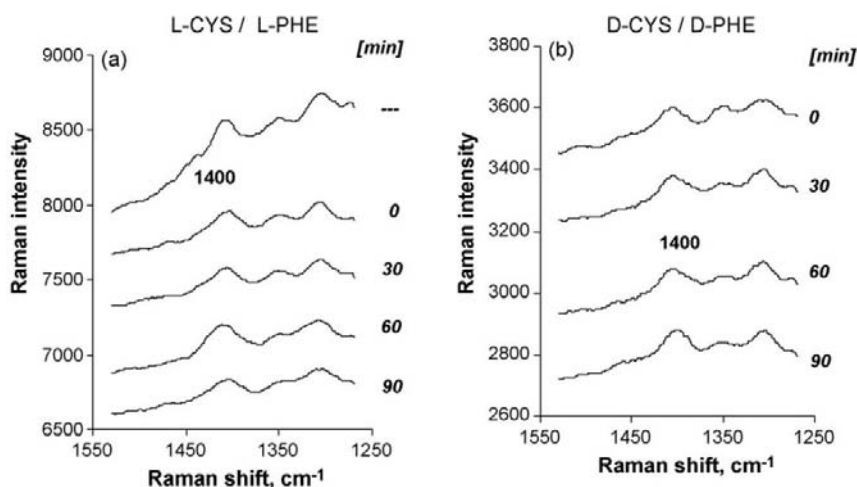
Figure 15. The SERS spectra of (L- or D-) cysteine SAM immersed into (L- or D-) phenylalanine solution within 1250–1550 cm^{-1} range. The upper spectra—cysteine SAM spectrum before transferring into phenylalanine solution. Abbreviations: Cysteine—CYS, phenylalanine—PHE (the spectra were shifted along the vertical axis to enhance the clarity of presentation) [49].
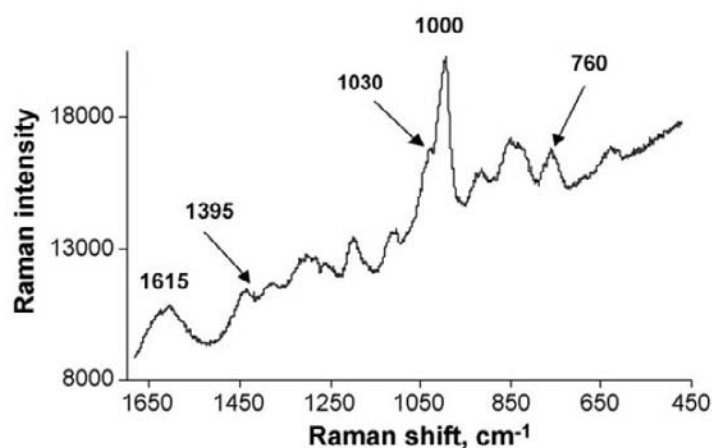
Figure 16. The SERS spectrum of L-phenylalanine adsorbed on the Ag surface at pH 3 [49].

If intensity of band at 1400 cm^{-1} is stable, although some part cysteine molecules have changed conformation. The 730 cm^{-1} band, observed in spectra if chirality of interacting of

cysteine and phenylalanine is the same, could be ascribed to conformer P_N. The conformer P_C may be excluded because the COO– groups are far away from the Ag surface. If the assignment of band at 730 cm^{-1} came from phenylalanine molecule adsorbed at Ag surface, the strongest band of phenylalanine (1000 cm^{-1} and 1615 cm^{-1}) would be also present in the spectra. The band at 760 cm^{-1} in spectra of phenylalanine adsorbed on Ag surface, has a lower intensity comparing to 1000 cm^{-1} and 1615 cm^{-1} (Figure 16). Both bands are not observed. Probably, metal surface is covered by cysteine monolayer and phenylalanine molecules do not directly interact with Ag surface. These enantioselective effects could be discussed as chiral discrimination in intermolecular interaction between cysteine and phenylalanine molecules on Ag surface. Probably, some cysteine molecules change their conformation from P_H do P_N and group $-NH_3^+$ of cysteine can interact with a COO$^-$ group of phenylalanine (Figure 17). The fact, so only DD or LL cysteine–phenylalanine pairs hydrogen bonding between COO$^-$ group of phenylalanine and NH_3^+ group of cysteine is possible, is connected with existence of bulky $C_6H_5CH_2-$ group bonded to chiral carbon atom of phenylalanine and hindrance effect.
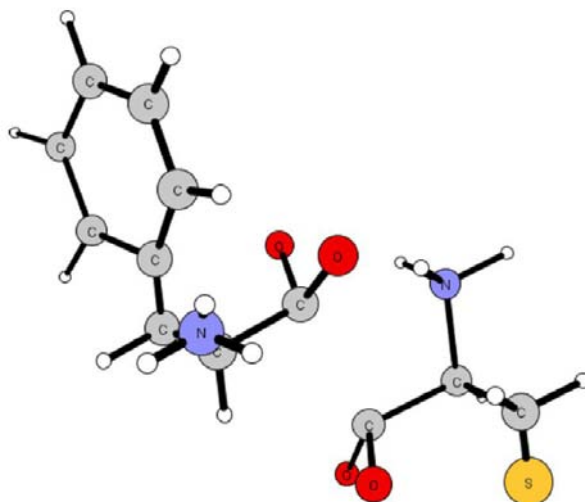
Figure 17. Model of interaction between two amino acids: L-cysteine and L-phenylalanine at the Ag surface [49].

Problem of chiral discrimination in interaction of phenylalanine and cysteine adsorbed on Ag surface could be discussed by considering the result of changing the chirality of one molecule. It is possible two ways of creation of phenylalanine–cysteine pairs, which contains molecule with different chirality (DL, LD). The first way is exchange the location of $C_6H_5CH_2-$ and NH_3^+ groups of phenylalanine. This arrangement may be less optimal as presented in Figure 17, because of stronger coulombic interaction between amino groups of phenylalanine with cysteine molecules and amino groups of neighboring cysteine molecule in SAM (it means decreasing the distance between both groups comparing to model in Figure 17). The second one consist in exchanging the location of hydrogen atom and benzyl group bonded to chiral carbon atom of phenylalanine. This could not be possible because of

hindrance cause without splitting of hydrogen bonds between amino groups of cysteine and carboxyl groups and phenylalanine.

Cysteine molecules adsorbed on Ag surface can recognize others chiral molecules – cysteine or phenylalanine, in experiments carried out with applied potential or not. Molecular interaction between two enantiomers (with the same or opposite chirality) can be detected by changes in vibrational spectra – presence or absence of the bands, bands shift or splitting and wider/narrower bands present in the spectrum.

CHIRAL SURFACES AND OTHERS METHODS

It is also possible to detect intermolecular interaction between chiral molecules using not only vibrational spectroscopy or STM/AFM techniques. Others methods are also used, but theirs versatility are not (in my opinion) so wide in comparison to STM/AFM and Raman.

Lu et co. observed, so borneol associated with Zn(II) and L–tryptophan in gas phase, do increase a nozzle potential (from 90 to 120 V) on lowered intensities of both forms of ions $[AB_3Zn(II)]^{2+}$ and $[ABZn(II)H]^+$ (included (+)–borneol and (–)–borneol, separately). Used symbols mean: A – L–tryptophan, B – borneol [50].

Sato et al. reported, so they carried out Monte Carlo simulations of studies on adsorption tris (1,10–phenantroline)metal(II) (racemic and enantiomeric form) by a clay [51]. They observed differences in free energy curves of chelate species along x, y, z and the axle Θ among the species for racemic and enantiomeric pairs, and intermolecular distance among these species. Similar calculation were carried out by Szabelski et al – they investigated chiral adsorption on nanostructured chiral surfaces and slit pores [52].

Theoretical calculation were done by Dodziuk et al. – they investigated interaction (molecular and chiral recognition) of decalin isomers by β–cyclodextins using force fields methods (AMBER, CVFF, CFF9I). They reported, so results were not so simple to interpretation and they suggest so molecular dynamics in solvent is optimal method of choice chiral and molecular recognition by cyclodextrins [53].

Sawada et al. reported so differences in height peaks were observed for complexes of host–guest determined by fast–atom bombardment (FAB) mass spectroscopy between crown ethers (*RRRR*–1 and *SSSS*–1–d$_6$) and amino acid ester ammonium ions (G) [54]. They defined the IRIS value as $I[(H_{RRRR} \cdot G)^+]/[H_{SSSS-1-d} \cdot G]^+ = I_R/I_{S-d6}$. The I_R/I_{S-d6} values were changed from, for instance, 2.0 through 1.0 to 0.5 in dependence from % ratio of enantiomeric form. They used two methods: FABMS/EL–Host method and FABMS/EL–Guest method.

Small differences in enantioselective recognition of alanine in solution on SAM adsorbed on gold, when chiral PAMAM dendrimers G4.0 were used [55]. They observed a lineary dependence the electro–oxidation peak current of Ala concentration over the range from 0 to 10 mM (for 143–187 µAcm^{-2}/µM). They suggested, so detection limit ($3\sigma[\mu M]$) was slight higher for PAMAM G4.0–D(+)Ala–L(–)Ala than for PAMAM G4.0– L(–)Ala– D(+)Ala.

Harvey and Arnett observed chiral recognition in monolayers created by methyl ester of stearoyserine (SSME) and palmitic acid in air/surface interface, if chiral species was in excess and if film arrangement was in a distorted and condensed form [56]. The Π (dyn/cm) for racemic film is higher in comparison for enantiomeric monolayer. The surface shear viscosity

film was also higher for racemic film, and equilibrium spreading pressure in prespread monolayers, too (Figure 18).

Small differences in peaks intensities of the ^{13}C NMR spectrum was observed by Belogi et al. They registered series of NMR spectra of pairs 5 chiral species (for instance, amine and carboxylic acid) in different concentration of them [57].

Sometimes in molecular recognition are used imprinted polymers: they are helpful in chromatographic separations, in pesticide and synthetic antibodies, in drug analysis, in discriminating absorption of molecular species and synthetic enzymes. The imprinted polymers is a template particle or ionic molecule joined together with the monomer to create a covalent or non–covalent–bonded complex.
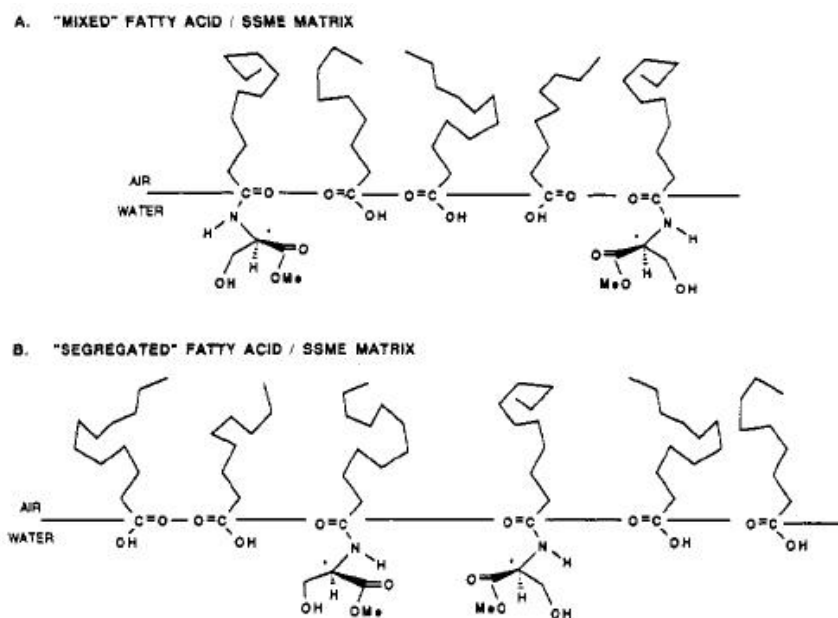
Figure 18. Model of molecular packing in a film containing (*R*)-(-) and (*S*)-(+)-SSME and an achiral fatty acid. Scheme A represents a separation of the chiral centres by an intervening fatty acid matrix. Scheme B represents segregation of SSME from the achiral matrix and direct contact between chiral headgroups [56].

Hollow spaces are formed in the polymer matrix (created from matrix–forming monomer) and these spaces can 'remember' the features. Additionally, these features have preferences in bonding and imprinted polymer might selectively join together with molecules from mixture of species: biological macromolecules, small organic molecules, inorganic cations and anions, etc. Generally, imprinted polymers are put as stationary phase for chemical discrimination of molecules. The idea of 'molecular imprinting' was born in 1940's as L. Pauling's conception of mechanism the antibodies formation [58]. At the present time, theory of molecular imprinting comes from Wulff and Sarhan idea [59]: they synthesized polymers with enantiomeric cavities and separated racemic mixture of chemical species. According to that, it is also possible employing non–covalent interaction among monomers and species, and using them, for instance, immunosay–type analyses as an alternative of

antibodies [60]. A glucose sensor could be also used as a ligand exchange process on an imprinted metal complex [60].

Neumann et al. investigated surface pressure/area (Π/A) isotherms created by Langmuir films of enantiomeric and racemic 2–hydroxyhexadecanoic acid (HHDA) in the presence of metal cations: Ca^{2+}, Pb^{2+}, Zn^{2+} [61]. They observed chiral discrimination in HHDA films from Pb^{2+} / Zn^{2+} to Ca^{2+}. Probably conformational status of the CH chain is autonomous of film condition (its compression) and minimal surface pressure (~1 mN/m) is needed for getting the finest order stimulated by presence Pb^{2+} and Zn^{2+}. The most probable explanation these effects is causing different complexes functional group of monolayer forming species and particular metal cation. Important role in these relations play electrostatic interaction, including hydrogen bonding and van der Waals interaction. Neumann et al. observed higher surface pressure/area (Π/A) at the comparable surface coverage ratio (nm^2/molecule) for racemic than enantiomeric HHDA. Similar effects was observed by Harvey et al. – equilibrium spreading pressure of stearoytyrosine for two kinds of films (film type I and type II) had higher values for racemic monolayer comparing to enantiomeric film [62]. Authors explained these differences as expression of repulsive electrostatic interaction between tyrosine polar groups, which is predisposed for fluidizing the structure of film.

Tran and Yu determined spectrophotometrically the mechanism of interaction between two chemical species: *tert*–butyl carbamoylated quinine (*t*–BuCQN) and *N*–derivative of leucyne (DNB–Leu) [63]. They strong suggest, so 3 probable interaction are possible: between ammonium group of of the *t*–BuCQN and carboxylate of the DNB–Leu (electrostatic interaction), between aromatic group of DNB–Leu and aromatic group of *t*–BuCQN (donor–acceptor CT transfer type of interaction), and amide group and carboxylic group (hydrogen bonding interaction). The dominant role play electrostatic interaction, and chiral recognition is possible, if electron is moved from donor quinoline group to acceptor dinitrophenyl group. It is important to use solvent with low polarity for eliminating of solvatation of donor and acceptor molecules by molecules of solvent.

According to Xu and McCarroll's results, between the average anisotropy and enantiomeric composition is linear relationship [64]. The average anisotropy is also connected with chiral selectivity and enantiomeric composition.

He et al. observed, so chiral sensors like binaphtol derivatives contain two chiral centres, can interact (and recognize them) with amino acids derivatives and obtained adducts, might be differentiated by the fluorescence intensity changing [65].

Similar experiments were carried out by Demirats et al. – they investigated interaction between chiral calyx[4]azacrown derivative (molecule with two chiral centres) and methyl ester of alanine and phenylalanine (D and L enantiomers) by UV spectroscopy [66]. They observed, so binding constant (*K*) and enthalpy (*ΔH*) values are higher, if L–enantiomer of amino acid is used. Authors suspect, so not only hydrogen bonding are responsible for intermolecular interaction and, in consequence, chiral discrimination, but also π–π stacking between aromatic groups.

Molecular recognition might be illustrated as simply model of Langmuir monolayer at the air–water interface: polar–moieties of Langmuir monolayer are 'host molecules' and hydrophilic species dissolved in aqueous subphase are 'guest molecules' [67]. The process of molecular recognition comes from the interaction between host–molecules (Langmuir monolayer) and dissolved guest–molecules. According to Leblanc conclusion, it is unknown a

nature host–guest interaction at molecular level. Additionally, these interaction in Langmuir monolayers were investigated using surface chemistry methods (isotherms of ΔV–A and π–A), microscopy (epifluorescence and BAM, Brewster angle microscopy), oscillating spectroscopy (infrared, UV–vis and fluorescence), imaging elipsometry and X and neutron–ray.

Chiral recognition is also important in understanding different smell of two enantiomers, like carvone. Nandi and Vollhardt argue, so difference in smell carvone enantiomers comes from different interaction with lipid molecules presented in nasal membrane – probably they interact with different part of chiral molecules [68]. (*S*)(+)–carvone has caraway–like smell, and (*R*)(–)–carvone smells as spearmint (Figure 19). They observed chiral discrimination of carvone in the surface potentials and π–A isotherms.
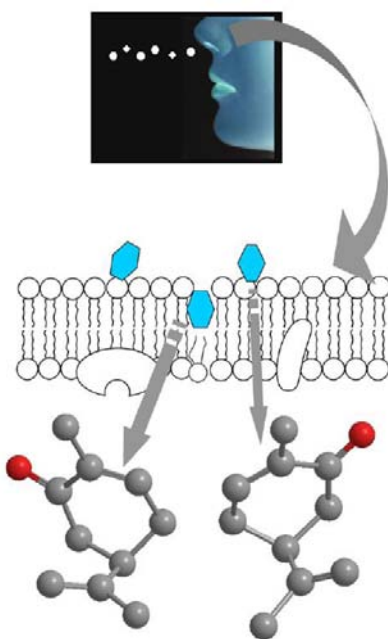
Figure 19. Presentation of the odor recognition process of chiral molecules at nasal membrane [68].

Ariga et al. divided chiral discrimination in air–water interface on two main categories: chiral discrimination among monolayer species and chiral discrimination between monolayers and aqueous molecules (guest) [69]. It is possible to use techniques in chiral discrimination like polarized IR reflection–absorption spectroscopy, BAM with GIXD, and also computational modeling and energy calculation, and nonliner optics and surface reactions.

Molecular imprinted polymers (MIP), similar to Langmuir monolayers at the air–water interface, might also be used to molecular recognition according to Rushton et al. [70]. MIP have poorer selectivity in comparison to natural recognition systems and their properties are strong connected from their concentration. If concentration of MIP is high, their selectivity is low and lead binding assets. In opposition, at low concentration of MIP, dominate selectivity. In conclusion, high selectivity of MIP for low concentration can be useful for detecting toxins in water contributor.

Qian, P. et al. investigated polymer Langmiur–Blodgett (LB) films built from chiral species (*N*–alkylacrylamides): (*S*)–2–metoxy–1,1'–binaphtalen–2'yl methacrylate (*S*–MeBN) or (*S*)–2–hydroxy–1,1'–binaphtalen–2'yl methacrylate (*S*–HOBN), with *N*–decyl– (DA), docecyl– (DDA), or tetradecylacrylamide (TDA) [71]. They measured surface pressure–area isotherms. The electrodes with adsorbed chiral species, induced very responsive potential change after addition to solution of enantiomer of 1–phenylethylamine (PEA): (*S*)–HOBN film can discriminate enantiomers of *R*– and *S*–PEA, but (*S*)–MeBN film can not discriminate these PEA enantiomers. Potential decrease measured after addition to solution of enantiomer of PEA is stronger for *R*– enantiomer than for *S*– one. Additionally, potential changes as a function of PEA concentration, is stronger for *R*–PEA. Authors argue, so *R*–enantiomer binds stronger in comparison to *S*–enantiomer. The lack of chiral discrimination for *S*–MeBN comparing to *S*–HOBN shows, so crucial role plays interaction between hydroxyl group (*S*–HOBN) and amino group (PEA). Probably steric hindrance is smaller for *R*– than for *S*–enantiomer of PEA and methyl group is distant from naphthyl group of (*S*)–HOBN (Figure 20). Interaction between π–π electrons of benzyl and naphthyl group can not be definitely excluded, and possibly might stabilize adduct PEA–(*S*)–HOBN.
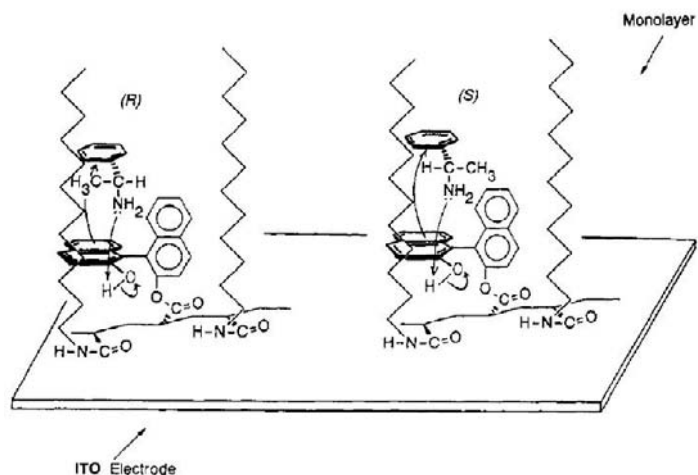
Figure 20. Schematic illustration for mechanism of chiral molecular recognition [71].

Gariani et al. tested a chiral tellurium ferrocene as a chiral species in NMR enantiomeric purity discrimination [72]. They synthesized a derivative of tellurium ferrocene (monosubstituted) in first step, and acquired ^{125}Te NMR spectrum of this species and (*S*)–(+)– and (*R*)–(–)– *N*–(3,5–dinitrobenzoyl)–a–methyl–benzamide, *S*–(+)–*N*–DNBMBA and *R*–(–)–*N*–DNBMBA. The ^{125}Te NMR chemical shift is related to enantiomeric excess of *N*–DNBMBA: the highest value (in ppm) is observed, if spectra was acquired at 100% ee of S–enantiomer, and the lowest – at 100% ee of *R*–enantiomer. If spectra was acquired for adduct of racemate of *N*–DNBMBA, the value of chemical shift is at half−way for *S* and *R* enantiomers (Figure 21). Gariani et al. also observed, so two peaks intensity – at 327–329 ppm region – is correlated with enantiomeric excess of *N*–DNBMBA – peak with lower ppm value has higher intensity than peak with higher intensity for *S* enantiomer, and for *R* enantiomer both intensities were inversed in comparing to *S* enantiomer. For racemic mixture

of *N*–DNBMBA, both peaks have comparable intensities. Unfortunatelly, Gariani et al. do not propose mechanism of creating adduct between both species (*N*–DNBMBA and derivative of tellurium ferrocene) and, do not explain why spectra for *S* and *R* enantiomers are different.

An example of chiral shapes of biological crystals are shapes of specimen of the foraminifer *Globigerina pachyderma*, a marine microorganism, offers a significant example of macroscopic symmetry. If a sea or oceanic water temperature is low (Antarctic, Arctic), the direction of calcium carbonate shells coil is counterclockwise (>98%) [73]. If water temperature is high (tropic regions), more than 98% coil in clockwise direction.

Attard observed electrochemical enantioselectivity at chiral metal surfaces [74]. He adsorbed enantiomer of 2–butanol (simplest chiral alcohol, $CH_3*CH(OH)CH_2CH_3$) on platinum surface (according to Attard nomenclature – Pt $\{643\}^S$ and Pt $\{643\}^R$). In next step, platinum electrode covered by 2–butanol, was immersed to glucose (D or L) solution in 0.1 M H_2SO_4 and electrochemically oxidized (CV was registered). In this condition were present:

– Pt $\{643\}^S$ and D–glucose = Pt $\{643\}^R$ and L–glucose (first pairing),

– Pt $\{643\}^S$ and L–glucose = Pt $\{643\}^R$ and D–glucose (second pairing).

For first pairing of species, the peak at 0.33V/Pd–H has higher intensity than peak registered for second pairing at 0.22V/Pd–H (Figure 22). If platinum surface was not cover by enantiomer of 2–butanol and removed to D– or L–glucose solution, and CV was registered, no electrochemical enantioselectivity was observed.
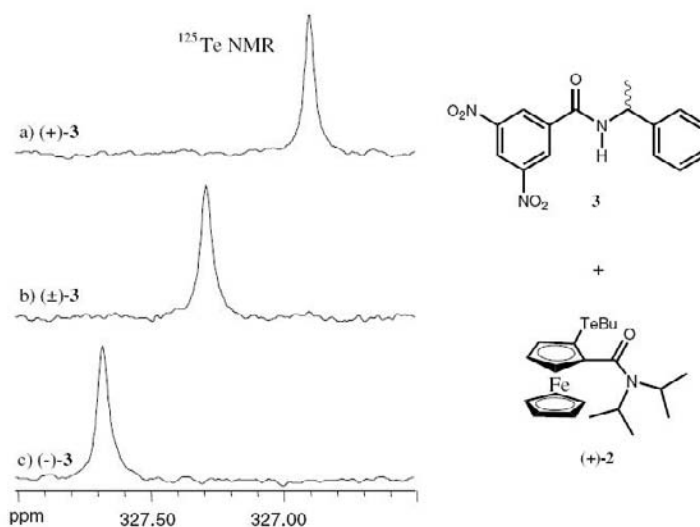
Figure 21. ^{125}Te NMR spectra of (+)-N,N-diisopropyl-(2-butyltellurium)ferrocenyl-carboxamide (+)-2 with: (a) (S)-(+)-N-DNBMBA (+)-3; (b) N-DNBMBA (±)-3; (c) (R)-(-)-N-DNBMBA (-)-3 [72].

Similar effects were observed for other carbohydrates – mannose, arabinose and xylose by Ahmadi and Attard [75]. Unfortunately, they does not explain, why these peaks occurred in different potential and reason of changed intensities. Probably, the platinum surface heterogeneity, covered by enantiomer of 2–butanol occurs, so more simple is oxidation of

second pairing [75]. Attard also argued, so important is kind of electrode – better enantioselectivity is using chiral electrode like Pt {643}[S] / Pt {643}[S] in comparing to Pt {211} and Pt {332}. According to his hypothesis, it is connected with surface density kink sites [75].

Similar experiments were carried out by McFadden et al. [76]. They adsorbed also 2–butanol on chiral Ag surfaces, like Ag {643}[S] and Ag {643}[R]. In next step, they desorbed of 2–butanol (by Temperature Programmed Desorption, TPD) and desorption rates were measured. Unfortunately, observed differences in TPD spectra for 4 pairing of species are rather small.
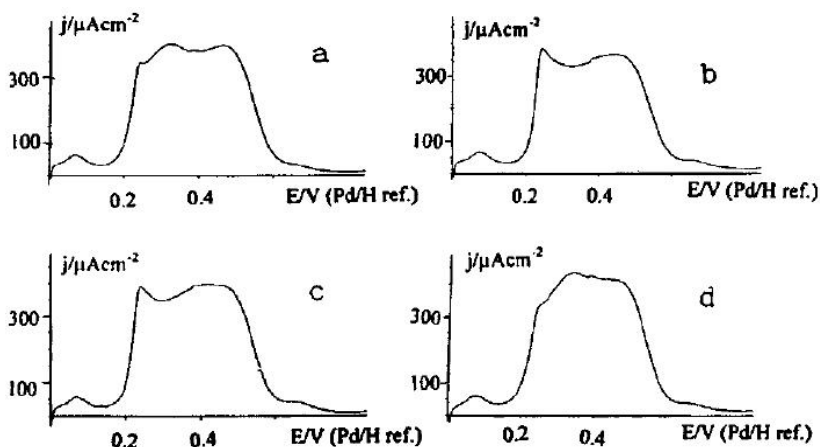
Figure 22. Pt{643}[S] voltammogram in (a) 0.05 M sulfuric acid + 5x10[-3] M D-glucose or (b) 0.05 M sulfuric acid + 5x10[-3] M L-glucose. Pt{643}[R] voltammogram in (c) 0.05 M sulfuric acid + 5x10[-3] M D-glucose or (d) 0.05 M sulfuric acid + 5x10[-3] M L-glucose [74].

ACKNOWLEDGMENT

CONCLUSION

The most important problem in investigation of chiral surfaces is the fact, so enantioselectivity has very subtle form and energy differences between two enantiomers are no more than few kJ (<10 kJ) [77]. Despite it, enantiospecific interaction are possible to detect using the technique like STM, AFM, SERS, electrochemical methods and others. If

STM or AFM are used, it is possible to observe the metal or non–metal surface covered by interacting chiral species, for instance, one enantiomer, then surface sample covered by second enantiomer is a mirror image to the previous one (this is logical – this is the definition of chiral species). Using others techniques, like Raman and SERS, IR, and electrochemical methods in addition, it is possible to characterize the interaction between enantiomer molecules, or molecules of racemic mixture. It is also feasible to determine interaction between diastereoisomers – i.e. on surface covered by one chiral species, are adsorbed other chiral species, for instance, of carbohydrate and chiral alcohol pairing enantiomers.

For practical use, significant quantities of chiral metal surfaces are needed, to enantioselective catalysis and preparation. In addition, for enantiospecific hydrogenation high thermal stability is needed [77]. In opposite to enantioselective catalysers, enantiospecific sensors require low surface area [77].

REFERENCES

[1]   Pasteur L. *C.R.Acad. Sci.* 1858, 46, 615–618.

[2]   Pasteur L., *Recherches sur les relations qui peuvent exister entre la forme crystalline et la composition chimique, et le sens de la polarisation rotatoire*, Ann. Chim. Phys., 24, 442-459 (1848).

[3]   Oevres de Pasteur. Réunies par Pasteur Vallery-Radot (Louis). Paris, Masson et CIE: part I - *Dyssymétrie moléculaire* (1922), part II - *Études sur le vinaigre et sur la vin* (1924).

[4]   Pasteur L. *Über die Asymetrie bei natürlichen vorkommenden organischen Verbindungen*. Von ... 2 Vorträge gehalten am 20. Januar und 3.Februar 1860 in der Société chimique zu Paris. Leipzig 1891.

[5]   Thomson W. T. (Lord Kelvin) *Baltimore Lectures on molecular Dynamics and the Wave Theory of Light*; Publication agency of the John Hopkins University: Baltimore, MD, 1904.

[6]   Easson E.H., Stedman E. *Biochem. J.* 1993, 27, 1257–1266.

[7]   Ogston A.G. *Nature* 1948, 162, 963–963.

[8]   Dalgliesh C.E. *J. Chem. Soc.* 1952, 137, 3940–3952.

[9]   Berthod A. *Analytical Chemistry*, 2093, April 2006.

[10]   Kudelski A. *Vibrational Spectroscopy* 39 (2005) 200–213.

[11]   Fleischman M., Hendra P.J., McQuillan A.J*., Chem. Phys. Lett.* 26 (1974) 163.

[12]   Jeanmaire D.L., Van Duyne R.P., *J. Electroanal. Chem.* 84 (1977) 1.

[13]   Albrecht M.G., Creighton J.A., *J. Am. Chem. Soc.* 99 (1977) 5215.

[14]   Kudelski A., Pettinger B., *Chem. Phys. Lett.* 321 (2000) 356.

[15]   Nie S., Emory S.R., *Science* 275 (1997) 1102.

[16]   Moskovits M., Such J.S., *J. Phys. Chem.* 88 (1984) 5526.

[17]   Allen C.S., Van Duyne R.P., *Chem. Phys. Lett.* 63 (1979) 455

[18]   Moskovits M., Such J.S., *J. Phys. Chem.* 92 (1988) 6327.

[19]   Pemberton J.E., in: Abruna H.D. (Ed.), *Electrochemical Interfaces: Modern Techniques for In–situ Interface Characterization*, VCH Publishers, New York, 1991, p. 193.

[20]   Tian Z.Q., Ren B., Wu D.Y., *J. Phys. Chem. B* 106 (2002) 9463.

[21] Garrell R.L., in: Hubbard T.A. (Ed.), *The Handbook of Surface Imaging and Visualization*, CRC Press, New York, 1995, p. 785.

[22] Pettinger B., in: Lipkowski J., Ross P.N. (Eds.), *Adsorption of Molecules at Metal Electrodes*, VCH Publishers, New York, 1992, p. 285.

[23] Kudelski A., *Vib. Spectrosc.* 33 (2003) 197.

[24] Podstawka E., Ozaki Y., and Proniewicz L. M. *Applied Spectroscopy*, Vol. 58, Issue 10, pp. 1147-1156 (2004).

[25] Podstawka E., Ozaki, Y., Proniewicz, L. M. *Applied Spectroscopy*, Vol. 58, Issue 10, Pages 294A-304A and 1141-1263 (October 2004), pp. 1147-1156(10).

[26] C. Julian Chen (1993). *Introduction to Scanning Tunneling Microscopy. Oxford University Press*. ISBN 0195071506.

[27] Oura K., Lifshits V. G., Saranin A. A., Zotov A. V., and Katayama M. (2003). *Surface science: an introduction*. Berlin: Springer–Verlag. ISBN 3540005455.

[28] Bonnell D. A. and Huey B. D. (2001). *Basic principles of scanning probe microscopy*.

[29] Bonnell D. A. *Scanning probe microscopy and spectroscopy: Theory, techniques, and applications* (2 ed.). New York: Wiley–VCH.

[30] Roiter, Y; Minko, S (Nov 2005). *AFM single molecule experiments at the solid–liquid interface: in situ conformation of adsorbed flexible polyelectrolyte chains*. *Journal of the American Chemical Society* 127 (45): 15688–9.

[31] Kühnle, A.; Linderoth, T. R.; Hammer, B.; Besenbacher, F. *Nature* 2002, 415, 891.

[32] Switzer J. A., Kothari H. M., Poizot, P. Nakanishi S. and Bohannan E. W. *Nature* 2003, 425, 490.

[33] Lorenzo, M. O.; Baddeley, C. J.; Muryn, C.; Raval, R. *Nature* 2000, 404, 376.

[34] Fang H., Giancarlo L. C., and Flynn G. W. *J. Phys. Chem. B* 1998, 102, 7311–7315.

[35] Orme C.A. Noy A., Wierzbicki A., McBride M.T., Grantham M., Teng H.H., Dove P.M., DeYoreo J.J. *Nature* Vol. 411, 775, 2001.

[36] Feyter S., Gesquière A., Mottaleb M. M. A. Grim P.C.M., de Schrynver F. C., Meiners C., Sieffert M., Valiyaveettil S., Mülen K. *Acc. Chem. Res.* 2000, 33, 520–531.

[37] Mahapatro M., Gibson C., Abell, C., Rayment T. *Ultramicroscopy* 97 (2003) 297–301.

[38] Kienberger F., Ebner A., Gruber H.J., Hinterdorfer P. *Acc. Chem. Res.* 2006, 39, 29–36.

[39] Ariga K., Kunitake T. *Acc. Chem. Res.* 1998, 31, 371–378.

[40] Bombis C., Weigelt S., Knudsen M.M., Nørrgard M., Busse C., Lægsgaard E., Besenbacher F., Gothelf K.V., Linderoth T.R. *ACS Nano* Vol. 4, No 1. 297–311, 2010.

[41] Barańska H., Kuduk–Jaworska J., Szostak R., Romaniewska A. *Journal of Raman Spectroscopy* 2003, 34, 68–76.

[42] Ruperez, A.; Laserna, J. J. Anal. Chim. Acta 1996, 335, 87.

[43] Gericke A., Hühnerfuss H., *Langmuir* 1994, 10, 3782–3786.

[44] Graff M., Bukowska J. *J. Phys. Chem. B* 2005, 109, 9567–9574.

[45] Fasman, G. D. *Practical Handbook of Biochemistry and MolecularBiology*; CRC Press: Brolo, A.; Germain, P.; Hager, G. *J. Phys. Chem. B* 2002, 106, 5982.

[46] Boca Raton, FL, 1990; p 3.

[47] Chakraborty, D.; Manogaran, S. *J. Mol. Struct.* 1998, 422, 13.

[48] Suzuki, S.; Shimanouchi, T.; Tsuboi, M. Spectrochim. Acta 1963,19, 1195.

[49] Graff M., Bukowska J. *Vibrational Spectroscopy* 52 (2010) 103–107.

[50] Lu. H.–J., Guo Y.–L. *Analytica Chimica Acta* 482 (2003) 1–7.

[51] Sato H., Yamagishi A. Kato S. H. *J. Phys. Chem.* 1992, 96, 9382–9387.

[52] Szabelski P., Panczyk T., Drach M. *Langmuir* 2008, 24, 12972–12980.

[53] Dodziuk H., Lukin O., Nowinski K.S., *Journal of Molecular Structure* 503 (2000) 221.

[54] Sawada M., Yamaoka H., Takai Y., Kawai Y., Yamada H., Azuma T., Fujioka T., Tanaka, T. *Int. J. Mass Spectrom.* 1999, 193, 123−130.

[55] Bustos E., Garcia J.E., Bandala Y., Godinez L.A., Juaristi E. *Talanta* 78 (2009) 1352–1358 (2009).

[56] Harvey N.G., Arnett E.M., *Langmuir* 1989, 5, 998–1005.

[57] Belogi G., Croce M., Mancini G., *Langmuir* 1997, 13, 2903–2904.

[58] Pauling L. *J. Am. Chem. Soc.* 1940, *60*, 2643.

[59] Wulff G., Sarhan A., *Angew. Chem. Int. Ed. Engl.* 1972, 11, 341.

[60] Maeda M., Bartsch R. *ACS Symposium Series*; American Chemical Society: Washington, DC, 1998.

[61] Neumann V., Gericke A., Hühnerfuss H., *Langmuir* 1995, 11, 2206–2212.

[62] Harvey N.G., Rose P.L., Mirajovsky D. and Arnett E.M., *J. Am. Chem. Soc.*, 1990, 112, 3547–3554.

[63] Tran C.D., Yu S., *J. Phys. Chem. B.* 2005, 109, 12627–12635.

[64] Xu Y., McCarroll M.E., *J. Phys. Chem. A.* 2004, 109, 12627–12635.

[65] He X., Cui. X., Li M., Lin L. Liu X., Feng X. *Tetrahedron Letters* 50 (2009) 5853–5856.

[66] Demirtas H.N., Bozkurt S., Durmaz M., Yilmaz M., Sirit A. *Tetrahedron* 65 (2009) 3014–3018.

[67] Leblanc R.M. *Science Direct, Current Opinion in Chemical Biology* 2006, 10, 529–536.

[68] Nandi N., Vollhardt D., *Current Opinion in Colloid and Interface Science* 13 (2008) 40–46.

[69] Ariga K., Michinobu T., Nakanishi T., Hill J. *Current Opinion in Colloid and Interface Science* 13 (2008) 23–30.

[70] Rushton G. T., Furmanski B., Shimizu K.D. *Journal of Chemical Education*, Vol. 82, No. 9. September 2005.

[71] Qian P., Matsuda M., Miyashita T., *J. Am. Chem. Soc.* 1993, 115, 5624–5628.

[72] Gariani R.A., Simonelli F., Oliveira A.R.M., Barison A., Comasetto J.V. *Assymetry* 17 (2006) 2930–2934.

[73] Addadi L., Weiner S. *Nature* Vol. 411, 753, 2001.

[74] Attard G. *J. Phys. Chem. B.* 2001, 105, 3158–3167.

[75] Ahmadi A., Attard G. *Langmuir* 1999, 15, 2420–2424.

[76] McFadden C., Cremer P.S., Gellman A.J. *Langmuir* 1996, 12, 2483–2487.

[77] Gellman A. *ACS Nano* Vol. 4, No 1. 5–10, 2010.

Chapter 5

MOLECULAR IMPRINTING: STATE OF THE ART AND APPLICATIONS

Nada F. Atta, Ahmed Galal and Ali M. Abdel-Mageed

Department of Chemistry, Faculty of Science, Cairo University
Giza 12613, Egypt

1.1. GENERAL INTRODUCTION

Molecular imprinting has been widely studied and applied in the last two decades as an innovative tool for various technological and scientific fields. The first approach to molecular imprinting was not well known until 1949 when Dicky and his coworkers succeeded to create complementary molecular cavities for certain dye molecules. Dickey's silicates could be considered as the first molecularly imprinted material [1and2].

In general, molecular imprinting process involves the formation of molecular cavities inside the polymer matrix being imprinted which are of complementary structural, functional group orientation and geometrical features with respect to the molecule being imprinted. Specifically, the molecular cavities are created by the incorporation of the template molecule during the polymerization process in such a way that a three dimensional network is formed around the molecule being imprinted. Upon extraction of the molecular moiety (template) from the polymer matrix, molecular cavities with specific shape, size and electrostatic features, remain in the cross-linked host material [3-5].

Because of its unique properties, molecularly imprinted materials have been widely utilized for a lot of applications and in various fields. They were applied in high performance liquid chromatography [6], food analysis [7], capillary chromatography, solid phase extraction [8] and drug delivery techniques [9].

One of the most important applications of the imprinting technique is the molecular recognition. Sensors prepared by imprinting methods could introduce good solution for the recognition of a variety of biologically active molecules. Recognition mechanism of the molecules is largely similar to what happens in living organisms. In our bodies, there are a large number of different molecules, and cells, without which we cannot survive, which are able to work cooperatively in such a way that certain function are carried out very precisely

and accurately. For example, the receptors on the surface of cell membranes bind hormones specifically and selectively. When the receptor binds a hormone, its conformation is changed and a message of the hormone is transferred in terms of a conformational change and as a result of that a specific function of that hormone is fulfilled. In molecular recognition, the molecules being imprinted can rebind to their molecular cavities with very high degree of selectivity and specificity, so that these materials may be named as artificial antibodies.

1.2. MOLECULAR IMPRINTING UNDERLYING PRINCIPLE

Now we will focus here on the basic idea of discussion, the imprinting process .we will talk in details about different types and mechanisms involved in molecular imprinting. Generally, there are two main strategies employed in the molecular imprinting processes. Both strategies are mainly based on the type of interaction with building units of the matrix being imprinted. The first strategy is covalent imprinting which involves covalent interaction between the functional monomer units and template molecules being imprinted. On the other hand the second strategy is non-covalent imprinting which involves non-covalent interaction between the functional monomers and template molecules. In the following paragraphs the two types are discussed in details.
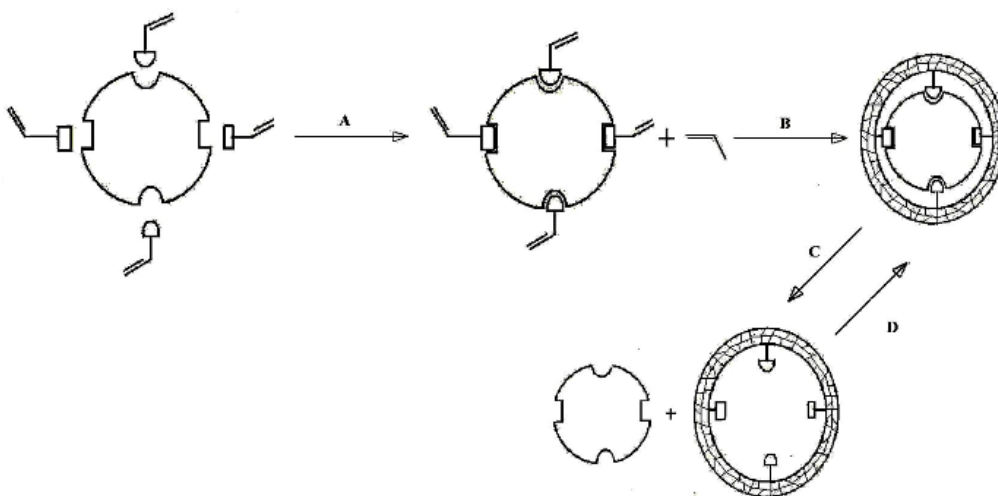
Figure 1. General idea underlying the molecular imprinting process. A) Molecular assembly between template molecule and functional monomers; B) polymerization of functional monomers around template molecule; C) template extraction from three-dimensional matrix; D) rebinding process to the molecular cavities.

It can be mentioned here that, there are four steps involved in tailoring and in using the molecularly imprinted polymers (MIP) whether the devoted mechanism is making use of covalent or non-covalent approach. The first step involves a contact between the template molecule and the pre-selected functional monomer so that a pre-polymerization complex (adduct) is formed between them. The stability of the complex structure will determine the behavior of the future molecularly imprinted polymer (MIP) with respect to the imprinted

molecules. The second step involves polymerization and cross-linking of the monomer building units around the complex formed in the first step. The last step involves leaching out (extraction) of the imprinted template molecules leaving well established molecular cavities in the polymeric material. The imprinted material now carries finger prints of the target molecule and is able to rebind it again with high selectivity even in the presence of molecular structures of comparable shape and size.

1.3. IMPRINTING MECHANISMS

The mechanism of molecular imprinting is almost determined by the type and strength of forces underlying the formation of a pre-polymerization complex at the preliminary step in the imprinting process. The importance of this step lies behind the fact that the recognition characteristics of the imprinted material rely on how efficient interaction between the template and the functional monomers would be achieved. Whenever the complex formed between the monomer and the template molecule is due to rearrangement of the bonds between the two species leading to a new adduct, then the mechanism would be covalent. On the other hand, non-covalent imprinting involves only arrangement of molecular species around each other in a three-dimensional network using the typical forces of attraction between the molecules like Van der Waals forces, electrostatic attraction or dipole-dipole interaction between the molecular species.

1.3.1. COVALENT IMPRINTING

This type of imprinting involves a preliminary step in which a template-monomer unit is covalently formed ,then these complex units copolymerize with additional polymer forming components, with a high proportion of cross linker, in a suitable solvent which is able to dissolve both the monomer and template molecules. A polymer network in which the template is covalently bound to the polymer matrix is finally obtained [9] with the template being encapsulated inside its matrix. Condensation reactions are the main processes involved in covalent imprinting approaches. It is found that moderate and reversible covalent bonds are formed under mild aqueous conditions while strong covalent bonds, notably ester bonds, are formed under fairly acidic or basic conditions.

The mild covalent imprinting process is found to be more advantageous when compared to strong covalent imprinting. On one hand the imprinting process itself is much easier and needs less processing steps. Most importantly, the extraction of template and its rebinding to the created molecular cavities would be easier in the first case. The basic routes for mild covalent imprinting process include the use of boronate esters, Schiff bases, ketals, and acetals. However, boronate ester imprinting turned out to be the most efficient and easily handled process among the above mentioned methods. Guntter Wulff et al had been the first group to use boronate esters for the purpose of molecular imprinting process [9].

A variety of template molecules which were imprinted using this approach included glyceric acid [10], derivatives of manose [11], galactose and fructose [12], sialic acid [13], castasterone [14] L-dopa [15], and some nucleosides [16]. Also, amino acid derivatives were

successfully imprinted by condensation with carbonyl compounds (mainly aldehydes) [15, 17] which is an example of Schiff bases. Moreover, ketals and/or acetals, formed between a diol and a carbonyl compound, have been employed in molecular imprinting protocol. Mono- and di-ketone template molecules have been extensively studied by Shea et al [18-20].

Damen et al [21-23] have contributed to improving the molecular imprinting methodologies by introducing for the first time the use of strong covalent bonds in the template-monomer assembly process. Strong covalent bonds, notably ester bonds, were used to assemble the polymerizable template monomer species, followed by incorporation into divinylbenzene-based matrices using suitable reagent or by hydrolysis. Rebinding to the polymers involved reaction of an acyl chloride with the alcohol component or displacement of bromide by a salt of the carboxylic acid.

Semi-covalent imprinting can be affected by making use of a condensation of amides and esters together. In the simplest form a (meth) acrylate ester of the template is copolymerized with the matrix-forming monomer mixture. The template is subsequently removed by hydrolysis, with rebinding of the unesterified template to the polymer. Rebinding process could be attributed to interaction of the template hydroxyl(s) with (meth)acrylic acid residues introduced into the imprinted site.[9]

1.3.2. NON-COVALENT IMPRINTING

Non-covalent imprinting approaches make use of the typical forces of attraction between molecules such as hydrogen bonds, dipole-dipole interactions, and Van der Waals weak forces to generate adducts between template molecules and functional monomers used in the polymerization process. In non-covalent imprinting mechanism, a pre-polymerization adduct (Complex) is formed between the molecular assembly being imprinted and the monomer molecule before polymerization process in the solution mixture. The functional monomers used in the imprinting process would be arranged around the template molecule according to the mutual functional groups on both of them. The polymeric material grows around the template assembly creating molecular cavity around it, so that after the molecule is extracted a complementary molecular cavity is created in the matrix. For example, the imprinting of a group of catecholamine compounds like dopamine, epinephrine and nor-epinephrine molecules has been done in a hybrid matrix formed of tetraethylorthosilicate and phenyltriethylorthosilicate precursor monomers. A computational study about the type of interactions occurring during the pre-polymerization step showed these molecules interact with the hydrolyzed precursors via the amino group which achieve the best interaction between the imprinted matrix and the rebinding template [24].

To the best of our knowledge it was referred for the first time to non-covalent imprinting in 1980 by Mosbach's group [25, 26]. Non covalent imprinting can be affected by a single monomer or a combination of monomers .The first type is the simplest and the most widespread approach. During the last two decades many functional monomers have been used in non-covalent imprinting for different applications. According to their nature these monomers were classified as acidic, basic or neutral. Carboxylic acid based monomers have been widely utilized. Their widespread comes from the fact that they have few bonds with few rotational degrees of freedom, so they have different ways to interact with the template as

H-bond donor, H-bond acceptor and through formal ion pair formation as well as weak dipole-dipole interactions.

For instance, the group of methacrylic acid derivatives of enhanced properties including, triflouromethyl acrylic acid, methacrylic acid, methyl methacrylate, etc are examples of monomers of acidic properties. Among the basic monomers, vinyl-pyridines, electron rich compounds, interact strongly with electron deficient molecules having aromatic rings [27]. A variety of neutral monomers include acrylamide which showed superior H-bonding ability compared to methacrylic acid under conditions of low polarity [28].

Imprinting with a combination of monomers have been very attractive to chemists as it introduced a probability of interaction with a wide varieties of molecular assemblies having wide range of chemical structures which in turn gives the chemist much more options in tailoring good recognition matrices able to rebind the template efficiently and selectively. In order that the imprinting process to be a successful one, the adducts formed between the template and the functional monomers need to be stronger than any interaction between the functional monomers themselves. Amino acid derivatives were imprinted in a mixture of methacrylic acid derivatives [29]. A variety of molecules were imprinted in a mixture of acrylamide and 2-vinyl pyridine [30].

1.4. FACTORS AFFECTING IMPRINTING PROCESS

Here in this section we shall present a concise discussion about the factors affecting the molecular imprinting process and the ability of the imprinted materials to respond to the template over repeated rebinding processes. On surveying the literature over the last two decades we could determine few parameters which govern the characteristics of the imprinted materials. The type of monomer used, molar ratio between template and monomer, degree of cross-linking, temperature under which the polymerization is conducted, and media in which rebinding process shall be carried out are all factors affecting the rebinding efficiency of the imprinted materials.

1.4.1. MONOMER SELECTION

The effect of the type of functional monomer used in the the process of molecular imprinting has been found to be of a very crucial role for the molecular recognition characteristics of the imprinted materials. The type and the strength of interaction between monomers and template molecules occurring primarily before polymerization affects the binding sites created inside the polymer matrix. The functional group orientation inside the imprinted sites controls the regioselectivity of the polymeric material for the template during the rebinding process. One example, quinine as a template was imprinted in a series of functional monomers. Methacrylic acid (MAA), acrylic acid (AA), and 2-vinylpyridine (2-Vpy) were synthesized in the presence of quinine. All the polymers, after leaching out of the template were analyzed for rebinding ability of the template molecule with respect to the molecular cavities created during the polymerization process[31]. It could be mentioned here that among the MIPs prepared using the three different monomers, only MAA-containing

polymer showed higher binding affinity for the template molecule. Contrarily, MIPs prepared with AA and 2-VP as functional monomers exhibited only very low binding ability for the template. Obviously, this is because of the absence of significant interaction between these functional monomers and the template molecules. In view of the structure of these functional monomers, it was found that MAA is not only a proton donor, but also it is a proton acceptor which augments its interaction ability.

Another example is a study carried out by Fiona Regan et al on the molecular imprinting of ibuprofen[32]. Ibuprofen was imprinted in a mixture of ethyleneglycoledimethacrylate (EGMDA) and one of three functional monomers including allyamine (MIP1), 2-vinlypyridine (MIP2), and methacrylic acid (MIP3). On assessing the imprinted polymers for the molecular recognition of Ibrupofen in the prescence of its structural similar compounds like naproxen and ketoprofen, the percentage rebinding was evaluated and selectivity was determined by solid phase extraction. The 2-vinylpyridine MIP as a monomer–showed high selectivity for Ibuprofen over structurally related analogue with reference to the molecularly non-imprinted polymer sameples(corresponding controls) as indicated in talble1 [32]. Moreover it was capable of quantitatively extracting Ibuprofen from a pharmaceutical preparation.

Table 1. Recoveries of Ibuprofen and analogues on MIPs with ibuprofen (MIP1 and MIP2)long with the corresponding controls (CON 2 and CON 3).[32] [cite MIP16]

Compound	MIP 2 % recovery	CON2% recovery	MIP 3 % recovery	CON3% recovery
Ibuprofen	83.34	2.81	21.20	2.51
Naproxen	30.96	3.25	14.32	2.57
Ketoprofen	17.94	1.44	4.52	1.08

1.4.2.TEMPLATE, FUNCTIONAL MONOMER AND CROSS-LINKER MOLAR RATIO

Template concentration is very critical in the creation of suitable number of the active sites (imprints). It can be mentioned here that, the functional monomer and the template molar ratio has been found of great importance with respect to the number and quality of MIP recognition sites [33]. Once the ratio of cross-linker to functional monomer has been determined, then the template (T) concentration can be optimized with respect to the functional monomer concentration (M) to the point producing the best recognition characteristics of the polymer [34]. For covalent imprinting this is not necessary because the template dictates the number of functional monomer that can be covalently attached in a stoichiometric manner forming a complex which then polymerizes producing the imprinted polymeric material. For non-covalent imprinting the optimized T/M ratio is achieved

empirically by evaluating several polymers made with different formulations with an increasing template concentration [35]. A very informative example about the role of the molar ratio between the template molecules is the imprinting of doxazosin mesylate into methacrylic acid as the functional monomer and triallylisocyanurate as the cross-linker in methanol solution using 2, 20-azobis-isobutyronitrile as the initiator for polymerization process. As indicated in table 2, the rebinding capacity Q for the imprinted polymer toward its template increases with the increase in the template concentration in the polymerization mixture reaching an optimum value at mixture P5 and levels off with higher concentrations. When the amount of template is low there is no chance for complexation with the functional monomer and as a result of that the binding sites formed in the polymer matrix would not be of appreciable importance. On the other hand using higher concentration of the template more than the optimum value has no benefit to the imprinting process, because any excessive amount would not have enough monomers to assemble with.

Table 2. Preparation of MIPs for doxazosin mesylate and binding capacity of MIPs [36]

No.	Doxazosin mesylate (mmol)	MAA (mmol)	TAIC (mmol)	Doxazosin mesylate:MAA:TAIC	Q (lmol/g)
1	0	50	200	0:1:4	9.8
2	20	0	200	1:0:10	6.4
3	20	10	200	1:0.5:10	25.7
4	20	40	200	1:2:10	48.3
5	20	120	200	1:6:10	76.2
6	20	140	200	1:7:10	77.4

Q is the rebinding capacity of the ratio of doxazosin mesylate (μmol/L) to the amount of imprinted polylmer(g) undergoing the rebinding process.

I.4.3. ROLE OF CROSS-LINKERS

During the synthesis of an imprinted polymer, cross-linkers generally act as a controller of the mechanical characteristics of the polymer matrix which would greatly affect the stability and three-dimensional shape of the imprinted binding sites after removal of the template molecules. Also, the type of cross-linker can greatly affect the selectivity of the imprinted polymer toward its counterpart molecules (template) which in turn affects the reversibility of the rebinding process. As an example, cholesterol was molecularly imprinted into methacrylic acid which is able to bind cholesterol through H-bonds and is able to repel bile acids in the intestinal mimicking solution in which cholesterol binding is tested. Various cross-linkers were tested for the synthesis process [37]. In view of large excess of hydrophobic cross-linker (ethylene glycol dimethacrylate), it is believed that a large portion of non-specific binding of cholesterol might be due to hydrophobic interaction between the template and cross-linker incorporated during the process. Substitution of ethyleneglycoldimethacrylate (EGDMA) by glyceroldimethacrylate (GDMA), bearing more

hydrophilic groups (OH) on the cross-linker backbone, decreases the non-specific binding of chlosterol, but at the same time more of the template molecules must be incorporated during the polymerization mixture to introduce larger degree of cross-linking. To overcome this problem new cross-linker, bearing cholesterol moiety was introduced into the process. This functional unit augmented both the selectivity of the imprinted polymer toward its template, and the degree of cross-linking. All in all, the percent and type of the cross-linker has to be determined experimentally for each template to reach the optimum values which guarantees the best recognition characteristics of the imprinted materials.

1.5.4. TEMPERATURE EFFECT

Always the effect of temperature is to accelerate the rate of the chemical reaction according to the famous Arrhenius formula. Regarding the role of temperature in the molecular imprinting process the situation is quite complicated. Low degrees of temperature would be an advantage to the stability of the template-functional monomer complex, however higher polymerization temperature is very favorable for the completeness of polymerization which in turn improves the quality of MIPs recognition sites created by the molecular assemblies during the polymerization process [38]. It was reported that the amount of quinine imprinted in thermally polymerized methacrylic acid, able to rebind to the imprinted film, is greatly affected by the temperature of polymerization being used. Polymerization at 15 $^{\circ}$C and 55 $^{\circ}$C respectively produced specific binding amounts of quinine equal to65.8 µmol/ g and 59.89 µmol/g. This observation might be explained by the fact that at low temperatures, the template-monomer complex is formed better than that at higher temperatures [31]. Moreover several studies have shown that polymerization at lower temperatures forms polymers with higher selectivity versus polymers made at elevated temperatures [39]. The reason for this has been postulated on the basis of Le Chatelier's principle, which predicts that lower temperatures will drive the pre-polymer complex toward complex formation, thus increasing the number and, possibly, the quality of the binding sites formed.

1.5.5. RECOGNITION MEDIA AND PROGEN

The solvent used in the process of molecular imprinting is always called as progen. The importance of progen in molecular imprinting mechanism comes from the fact that it plays the central role in formation of the pre-polymerization complexes between the template molecules and functional monomer units by dissolving them together. A good progen should i) dissolve both the template and the functional monomer to the same extent, ii) interaction between the progen and either the template and /or ought not to be stronger than the interaction between the template and the monomer; iii) does not undergo any chemical reactions with either the template and/or the monomer during the imprinting process; iv) be easily evaporated. One important observation to be mentioned in this section is that the molecularly imprinted polymers exhibited higher rebinding effect in the solvent from which it was polymerized (progen). As an example of that, Metsulfuron-methyl (MSM) imprinted in a copolymer of 2-(triflouromethyl) acrylic acid (TF-MAA) and divinylbenzene (DVB) showed

the highest binding capacity for MSM [40]. The rebinding experiments characterizing the rebinding capacity of MSM depicted that the MIP binds only MSM in dichloromethane, which was used as a progen during polymerization process.

Table 3. ΔE (a.u.) MAA in different solvents. [41][Cite paper MIP3]

Environment	Energy (a.u.)	ΔE (a.u.)	ΔE (kJ/mol)
Vacuum	−306.49774	-	-
Chloroform	−306.50841	0.01067	28.00253
THF	−306.51003	0.01228	32.25112
DMSO	−306.51305	0.01530	40.17724

One more example is the effect of solvent on the interaction between theophylline and methacrylic acid as a functional monomer studied by Zheng Liu and his coworkers [41]. The interaction energy between theophylline and chloroform, tetrahydrofuran (THF) and dimethyl sulfoxide (DMSO) were calculated using B3LYP/6-31+G** level of calculations as indicated in table 3. In parallel to these computations the imprinted polymer was synthesized from the three solvent. 1H NMR spectroscopic investigations of the polymerization mixture showed that theophylline interacts most strongly with methacrylic acid in chloroform which showed the lowest interaction energy with theophylline while DMSO which had the highest interaction energy, showed the minimum interaction in the NMR. This result could be justified by the fact that whenever the solvent interaction with the template is smaller, that would assist to increasing the interaction between the template and the functional monomer.

However, the role of the progen be visualized much more clearly as a controller of the morphological properties of the imprinted polymer like surface area and porosity of the imprinted material. Porosity controls the diffusion of the template molecules and rebinding efficiency of the imprinted materials. Almost in a lot of cases porosity arises from phase separation of the solvent (progen) and the growing polymer. Progens with low solubility separate early and tend to form larger pores with lower surface area. Conversely, progens with higher solubility phase were found to separate later in the polymerization providing materials with smaller pore radii distribution and greater surface area of the imprinted material [38].

1.4.6. PH OF THE IMPRINTING SOLUTION

The effect of the solution pH from which the polymerization process takes place is quite important when we consider its effect on the particles size and the porosity of the imprinted materials. In general, the effect of pH is mostly discussed for the inorganic imprinted materials prepared using sol-gel processes like silicates and titanates. The reason behind that the catalytic action of the acid or base during the hydrolysis process of the alkoxide precursors of the silicates and titanates. The pH value affects the physicochemical characteristics of these materials to a large extent by altering the porosity and particles size of the formed polymer.

The control of the porosity and particles size is very crucial because it affects the diffusion of template molecules into the imprinted sites during the rebinding process which would necessarily affect both sensitivity and selectivity of the recognition element going to be used later. More precisely, the nonspecific adsorption of the molecularly imprinted materials is affected by the porosity and particles size of the polymer. Whenever the polymer is very porous, then its permeability for the template molecules might increase regardless of the presence of imprinted sites. On the other hand, the very nonporous polymeric materials might also hinder the diffusion of the template molecules into the polymer and would prevent their approach to their imprinted cavities. Regarding the trend of the recognition ability of the imprinted materials, it is not a fixed trend but depends largely on the chemical structure of the template and its molecular size and how it would fit into the polymer matrix.

As an example, dopamine (DA) and ascorbic acid(AA) as templates were molecularly imprinted into a hybrid matrix of silica and alumina from solutions having different pH values. Figure 2(a) shows rebinding process of dopamine (DA) to dopamine molecularly imprinted polymer (DA-MIP) when the imprinting process was carried out under a controlled pH with different values: 0.7, 4.0, and 6.0. The DA-MIP was synthesized with the molar ratio of DA:Si:Al being 1:10:3. In recognition process of the DA-MIP polymer in phosphate buffer solution (PBS) (pH = 6.5). The result shows a decreasing trend for DA to the DA-MIP surface with increasing values of pH used in the imprinting medium [42]. Contrarily, the AA rebound to the AA-MIP increased with the increasing in pH values as shown in Figure 2(b). When the imprinting medium pH gets lower at 0.7, AA is hardly imprinted into silica–alumina gel, and consequently the rebinding process is quite low. Probably, the Lewis acid sites had a stronger repulsive force to the acidic group of AA at the lower imprinting pH. Conversely, the Lewis acid sites become weak at higher imprinting pH values, thus, it can be found that the higher rebinding amount of AA is found at an imprinting pH of 6.
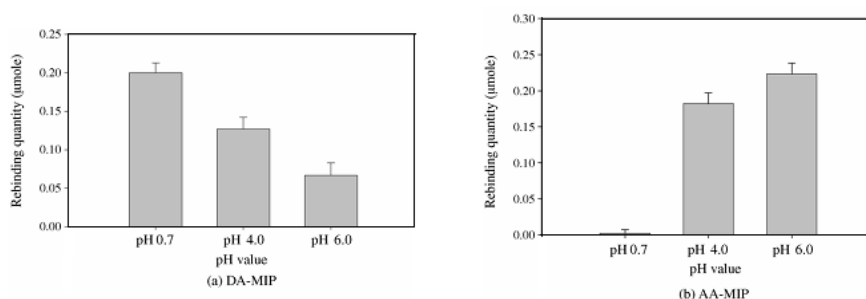
Figure 2. The imprinting pH effect on the rebinding quantity of (a) DA on the surface capped of DA-MIP, and (b) AA on the surface capped of AA-MIP. The imprinting pH values of the silica—alumina solutions were adjusted by using HCl and NaOH solutions. DA (AA) imprinted silica gel was composed of DA (AA):Si:Al = 1:10:3. Each rebinding test was based on 50 mg DA-MIP powder in 10 ml DA solution (0.2 mM) at 25 ∘ C for 30 min. The analytes solution is in 10 mM PBS, pH = 6.5.

1.4.7. PH OF THE REBINDING MIXTURE

The pH of solution from which rebinding process is conducted has great influence on the rebinding process of the template molecule to their imprinted polymers. The solution pH used

for the rebinding process affects the rebinding process in two ways. On one hand the electrostatic charge the template carries in solution is greatly controlled by the medium pH and varies with its change especially for molecules carrying functional groups susceptible for the change in pH like carboxylic groups, phenolic groups, amino groups …. etc. Consequently, the interaction between the template and the imprinted site would be affected. On the other hand, the change in the electrostatic charge of the molecular species might result in a change in the spatial configuration of the functional groups and their orientation in space with respect to the oriented functional groups in the polymer matrix. As a result of that these molecules might not be able to interact with the imprinted sites in an appropriate way and so on the efficiency of rebinding of the molecules to the molecular cavities will decline [43].

Most important to be discussed here is the effect of solution pH on rebinding process of neurotransmitter and drug like molecules which have been extensively studied during the last few years. Virtually all drug-like molecules are weak acids or bases. This means that they contain at least one site that can reversibly disassociate or associate a proton (a hydrogen ion) to form a negatively charged anion or positively charged cations. Molecules that disassociate protons are acids, and those that associate protons are bases. The reversibility means that a sample is always in equilibrium with some fraction protonated and the rest deprotonated.

$$[HA \leftrightharpoons H^+ + A^- \text{ or } HB \leftrightharpoons H^+ + B^-]$$
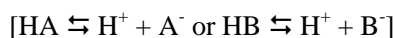
By varying the availability of protons, i.e. the acidity of the media, the balance of the equilibrium can be shifted. Alternatively, the pKa values of a site can be thought of the pH at which the protonated and deprotonated fractions are equal. Experiments were carried out from different pH solutions on dopamine imprinted surface. All pH measurements were carried out from 100_mol L^{-1} with respect to dopamine. As illustrated in Figure 3, the current response increases with the increase in pH value.

At pH-values lower than pKa1 (8.57), the protonated form of dopamine predominates than the unprotonated zwitter ion form. On the other hand, at pH-values higher than pKa2 (10.08) the deprotonated (phenoxide ion) form is higher than unprotonated zwitter ion. Between pKa1 and pKa2, the zwitter ion form predominates over the charged form [27]. This argument clarifies the fact that the protonated dopamine molecule is less interacting with the oriented functional groups in the molecular cavities while the neutral form is best fitted to the imprinted sites so by increasing the pH values current response enhances. On the other hand, there was an increase in the oxidation potential value of the adsorbed species on the imprinted surface with the increase of pH value. This result could be explained from the fact that, positively charged molecules due to protonation of dopamine at lower pH values needs higher polarization potential (overvoltage) for oxidation. This is because the average electronic charge on the molecules is lowered, therefore it is more difficult to withdraw electrons at such a lower potential, while at higher pH values the molecule is either neutral or negatively charged which increases its electro-oxidation. This study was not extended to pH values higher than 9 because the glassy silicon film could decompose under the effect of strong alkaline medium.
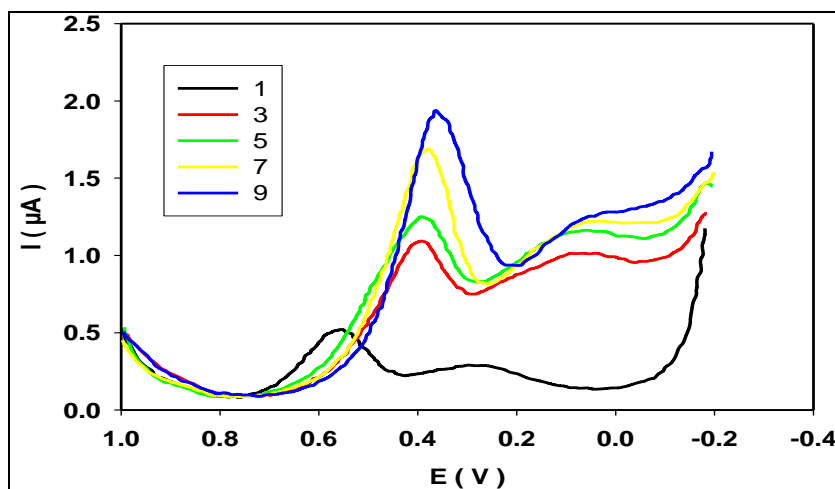
Figure 3. Square Wave Voltammetry (SWVs) of dopamine imprinted film after dipping in 100 µM of dopamine solutions at different values of pH (1.02 : ⎯ ;3.01: ⎯ ; 5.09:⎯ ; 6.99: ⎯ ; 9.1: ⎯).

1.5. IMPRINTED MATERIALS ANALYTICAL CHARACTERISTICS AND RECOGNITION AND ANALYTICAL CHALLENGES

1.5.1. REBINDING EFFICIENCY AND TEMPLATE RECOVERY

One of the most important characteristics have to be considered seriously when discussing the performance of molecularly imprinted materials is the extent to which it is able to rebind the template once again into their imprinted sites. The sensitivity of the imprinted material to recover the template to the imprinted sites depends largely on some of the factors discussed in the foregoing sections including in particular the monomer selection, progen, template concentration in particular the post-treatment of the imprinted materials during the extraction step but most important to be discussed here, in this section is the impact of the rebinding efficiency on the molecular recognition characteristics of the imprinted materials. The sensitivity of the imprinted materials is related directly to the number of imprinted sites created during the polymerization process of the molecularly imprinted material. As mentioned above the number of these active sites in the polymer matrix is greatly affected by the extraction of the template and influence of this step is governed by two important factors. On one hand, the solvent shall interact with the polymeric material in such a way that it would change the conformational structure of the molecular cavities rendering them into incompatible structures unable to interact with template in the appropriate way. On the other hand, and most important to be considered is the affinity of the polymeric materials for cross-linking after the extraction process. After the extraction of template, most polymeric materials are aged for certain period of time. The aging process is adopted to cure the polymeric material from any deformation which might occur during the post-imprinting treatments. This period of time would sufficient for extensive polymerization of the imprinted material which assists in blocking of the imprinted sites.

Moreover, the method adopted for the polymer preparation would affect the ability of the polymer to rebind its template. The use of different methods for the polymerization would absolutely vary the morphology of the polymer which results in a different reaction of these surfaces toward the template molecules. One example is the molecular imprinting of morphine as a template molecule in 3,4-ethylenedioxythiophene (EDOT) as a functional monomer using two different polymerization methods. The polymer was polymerized using traditional polymerization conditions and precipitation polymerization conditions. EDOT monomers were used to immobilize the molecularly imprinted particles in a second step. Rebinding experiments revealed that the polymer prepared using precipitation technique could recover higher concentration of morphine. Additionally the imprinted material showed higher selectivity for morphine in the presence of codeine molecules [44].

As a matter of fact, the decline in the number of imprinted sites would take place whatever the procedures being used during the extraction or aging processes but we can decrease its effect by some measures including the following:

1. Selecting a solvent for the extraction process, able to leach the template out from the matrix with a minimum interaction with the polymer.
2. Keeping the imprinted material in an isolated and dry place under low temperature to decrease the rate of polymerization process.
3. Avoiding the use of rebinding solution containing any species which would accelerate the polymerization process as much as possible.

1.5.2. MOLECULAR SELECTIVITY OF THE IMPRINTED MATERIALS

One of the challenging difficulties facing the measurement protocols involving an extraction, separation or even detection step using molecularly imprinted materials is the ability to discriminate between different molecular species, especially when they have very similar molecular structures. The selectivity of molecularly imprinted is suppressed by two factors. On one hand, the non-specific adsorption on the non-imprinted sites in the polymer matrix contributes an appreciable interference during the recognition process. Non-specific adsorption can still be considered to be acceptable within the limits which would not affect the molecular selectivity of the imprinted materials especially both the interfering moieties and template molecules which interfere almost to the same extent. Moreover, the effect of non-specific adsorption can be avoided completely by conducting a control experiment. Non-imprinted polymer, prepared using the same conditions of polymerization in the absence of the template, is tested in a solution containing the template molecule in the same concentration used in the routine rebinding experiments then the response of the non-imprinted polymer is subtracted from the basic rebinding experiments. On the other hand, the interference coming from the anchoring of molecules of similar structures partially or completely into the imprinted sties presents the most challenging problem for molecular recognition characteristics of the imprinted materials. The molecular size in conjunction with spatial configuration of the functional groups of the interfering molecules with respect to the imprinted cavities of the original template plays the most influential role on enhancing or deteriorating the molecular selectivity of these materials. In principle, the imprinted materials

are tailored so that the imprinted sites would be able to recognize only their templates having the appropriate size and structure. But some molecules may have similar structure which might help them to be partially adsorbed into the molecular cavities in the polymeric materials. Moreover, competitive molecules could be adsorbed completely into these sites when their molecular sizes are slightly smaller than the molecular size of the template. The difference in size facilitates the diffusion of these molecules more easily into the polymer matrix which decreases the selectivity toward the template molecule itself.

Carbaryl is an example depicting the highly selective molecularly imprinted polymer. The selectivity of the developed MIP-based fluorescent system for the recognition of the template was analyzed [45]. For such purpose, two other carbaryl pesticides (carbofuran and bendiocarb), structurally similar to carbamate were tested as possible interfering species, since those two compounds have similar chemical structures to that of carbaryl and show similar fluorescent spectra, they are also likely to be found in the same samples were carbaryl is to be determined. The MIP showed relatively high selectivity for the template compared to its potential interferents. One more example is the stereoselectivity of sol-gel imprinted with L-histidine. As template, L-histidine was imprinted into a hybrid sol-gel material synthesized from a mixture of functionalized organosilicon precursors (phenyltrimethoxysilane and methyltrimethoxysolane). The L-histidine imprinted and the control sol-gel films were exposed to a series of interfering substances. The sol-gel film has an excellent selectivity for L-histidine. Moreover, the imprinted sol-gel film showed a pronounced selectivity for L-histidine against its stereoisomer D-histidine [46].

On the light of these two examples, it can be concluded that the selectivity of molecularly imprinted polymers originates from the rigid three-dimensional structure created with the polymer matrix during the imprinting process and can be preserved as long as these formation were kept without any deformation of both size and configuration of their internal functional groups spatially oriented with the template molecules.

1.5.3. REPEATABILITY OF REBINDING PROCESS

Here in this section, we will discuss the repeatability of the rebinding process on the imprinted films. Repeatability of rebinding measurements means how the repeated measurements could be close to each other. In other words, rebinding process should be repeated more than one time with respect to its imprinted material with an acceptable value of standard deviation of these measurements from the mean value. Repeatability of the rebinding process is very crucial from the point of view of long term use of the imprinted materials and how trustable are the results we get. Repeatability of results is affected by all the foregoing factors described in the previous sections especially those related to the synthesis process and post-treatment of the imprinted material. It can be discussed in terms of stability of molecular cavity created within the imprinted materials. Preserving the number and structure of these active sites without any deformations can be accessed various factors involved in the imprinting and rebinding process. The stability of results obtained depends also to a large extent on the type of imprinted material and its resistance for the external factors like temperature changes, long term exposure to solvents and post-treatment of the imprinted materials. Silica based sol-gel synthesized materials is an example of an excellent material

showing good extent of results repeatability. One example is the imprinting of dopamine and tyramine molecules into a hybrid sol-gel material synthesized by acid hydrolysis of a mixture of tetraethylorthosilicate (TEOS) and phenyltriethylorthosilicate (PTEOS)[43]. Dopamine molecularly imprinted film cast on glassy carbon electrodes were tested for repeatability of their measurements. The rebinding experiments were carried out from a solution containing 50 µM dopamine prepared in PBS (pH 7.2 and 10 mM). The batch solutions from which adsorption experiments were prepared freshly for each experiment being done. The rebinding experiment was carried out five times for the same electrode with a period between each two experiments not more than 15 minutes. Results (current responses) obtained from dopamine oxidation on dopamine molecularly imprinted film as shown in table 10 indicated that the measurements are highly stable and repeatable. The standard deviation from the mean of measurements values calculated for the repeated experiments is very reasonable (7.4 %) with an accepted value 1.440×10^{-8} A, so we can write the mean value of these results as $(1.958 \pm 0.144) \times 10^{-7}$ A. The stability of measurements confirms the fact that adsorption occurs only into well defined molecular sites (molecular cavities created during the imprinting process). Measurements were carried out on three different electrodes each time.

1.6. Classification of the Imprinted Materials According to the Type of Monmer

1.6.1. Imprinting in Organic Materials

In the previous sections we have shown generally an idea about molecular imprinting. We referred also to different modes of interaction between functional monomers and template molecules being imprinted and the factors affecting the whole process. Here in this section, we shall classify the molecularly imprinted materials according to the type of monomer used in the imprinting process. We have selected a wide spectrum of functional monomers used for the molecular imprinting of various molecular templates oriented for different applications. We selected a group of representative examples as indicated in table 4. These examples cover different modes of interactions with various template molecules used in a lot of applications. The imprinted polymers tailored using these templates is almost were synthesized using only one monomer [33, 40] or a mixture of monomers [31, 32] able to interact with template molecules under study. Organic monomers have been studied intensively during the last two decades due to their ability to bear various functional groups which are able to interact with various template molecules. Tailoring of these materials has been state of the art to developing molecular imprinting of various template molecules including drugs [35], amino acids [33, 34] Viruses [40], herbicides [43].

1.6.2. Imprinting in Inorganic Materials

Undoubtedly, imprinting in inorganic materials has got wide interest from chemists and materials science groups working in this area not only because of the advantages such as durability and wide applicability of these materials, but also because of the easiness of their

preparation in laboratories. Research in this area has intensified in the last few years for different technological and scientific aspects starting from analytical purposes to nano-catalysis. The context of the following pages will survey different approaches to imprinting of different molecules into various inorganic materials. The main routes of these materials are the hydrolysis of metal alkoxides, mainly tetraalkylorthosilicates and titanium alkoxides using sol-gel process. Some examples for imprinting in silicates and titanates are shown in table 4b.

Table 4a. Organic functional monomers used for the imprinting of wide range of templates

Template Molecule	Functional Monomers and cross-linkers used
4-L-phenylalanylaminopyridine (4-LpheNHPy)	Mixture of ethylene glycol dimethacrylate (EGDMA) and methacrylic acid[47]
Tetracycline	Methacrylic acid (MAA) and ethylene glycol dimethacrylate as a cross-linker[48]
D- and L-tyrosine	Polypyrrole[49]
L-tyrptophane	Acrylamide (AM) and trimethylacrylate[50]
Ibuprofen	Methylmethacrylate (MMA) or 2-vinylpyridine in the presence of ethylene glycol dimethacrylate (EGDMA) as cross-linker[51]
Uric acid	Acrylonitile–acrylic acid[52]
Propanolol	Polymer micro-spheres [p-(divinylbenzene)-co-methacrylic acid] [53]
Morphine	Methacrylic acid and trimethylpropane trimethacrylate [54]
Tobacco mosaic virus	Polyethylene glycol hydro-gels[55]
Monocrotophos	Nylon-6 polymer [56]
Doxazosin mesylate	Methacrylic acid as a functional monomer and triallylisocyanate as a cross-linker[57]
2,4-dichlorophenoxyacetic acid	4-vinylpyridine (4-VP) and ethylene glycol dimethacrylate[58]

Table 4b. Inorganic functional monomers used for the imprinting of wide range of templates

Template Molecule	Functional Monomers and cross-linkers used
Nacfillin	Tetramethylorthosilicate (TMOS), methyltrimethoxysilame (MTMOS) and Phenyltriethylorthosilicate [59]
D- and L- 3, 4-dihydroxyphenylalanine (D- and L-Dopa) and R-and S-N,N'-dimethylferocenylethylamine (R-FC)	Tetramethylorthosilicate (TMOS) and phenyltrimethylorthosilicate (PMOS)[60]
hemoglobin (Hb)	3-aminopropyl- trimethylorthosilicateda (APTMS), and trimethylpropylorthosiilcate (TMPS)[61]
Caffeine	3-aminopropyltrimethoxysilane (APTMS), was used with tetraethylorthosilicate (TEOS) monomer [62]
Cholesterol in assembled with β-cyclodextrine)	Tetraethylorthosilicate (TEOS)[63]
Silica scaffolds	Protein (lysozyme or RNas-A)[64]
4-(4-ropyloxylazo)-benzoic acid (C3AZOCO$_2$H) and anthracene carboxylic acids (2-AnCO$_2$H and 9-AnCO$_2$)	Titanium butoxide precursor (Ti(O-n-Bu)$_4$)[65]
d-glucose	Titanium butoxide(Ti(O-n-Bu)$_4$)[66]

Most interesting in the use silicon or titanium alkoxides as precursor monomers regarding imprinting process is the potential of inserting functional organic groups on the central atom (Silicon or titanium). Consequently, we could have higher flexibility in using these materials for the imprinting of various template molecules. Moreover, the use of inorganic materials, synthesized by sol-gel technique, in molecular imprinting has showed increasing interest

during the last few years due to the following reasons i) Easiness of synthesis using sol-gel technique, ii) durability of the synthesized material compared to organic materials, iii) limited affinity for reacting with template molecules which allowed imprinting of big molecule never imprinted before like proteins[49] and iv) easiness to be applied as thin and ultra-thin films which have wide technological applications.

1.7. COMPUTATIONAL APPROACHES TO MIPs

During the planning of this thesis, part of the proposed work had been directed toward some theoretical manipulations and modeling aspects concerning the expected experimental results. Therefore, we will focus here on showing how computational work could improve research in molecular imprinting area in the last few years with some examples.

First of all, we will give a concise overview on computational chemistry which may be less common for a lot of chemists especially experimentalists. Computational chemistry is relatively a new emerging branch of chemistry and often makes use of computer software which is designed for its purposes. Computational chemistry simulates chemical processes and structures numerically, based in full or in part on the fundamental laws of physics [67].

Computational chemistry allows chemists to study chemical processes by running calculations on computers rather than by examining reactions and compounds experimentally which could largely help in understanding some experimental results and improve new experimental recipes. There are two basic methods of all calculations; the first one is molecular mechanics simulations which makes use of the laws of classical physics to predict the structures and properties of molecules. Molecular mechanics are available in many computer programs such as Hyperchem, Quanta, and Alchemy. On the other hand electronic structure methods make use of the laws of quantum mechanics rather than classical physics as the basis for their computations. The most famous and widely used software is Gaussian program. Energy and other related properties of a molecule may be obtained by solving the famous Schrödinger equation whose simple stationary form is shown in the simplest form of Schordinger Eguaiton.

$$H \psi = E \psi$$

In the following few paragraphs, we will discuss some work which had been published in the field of computational chemistry focusing on the area of molecular imprinting.

Xin Li et al have developed a computational method to screening functional monomers and polymerization solvents in rational design of molecularly imprinted polymer (MIPs). The method was based on the comparison of the binding energies of the most stable complexes formed between the template and different functional monomers. Five functional monomers, acrylamide (AAM), acrylic acid (AA), methacrylic acid (MAA), methylmethacrylate (MMA) and 2-trifluromethacrylic acid (TFMAA) were theoretically selected as possible functional monomers. Conformation optimization was performed using Hartree Fock (HF) computations at the level MP2 with 6-31++G (d) as a basis set specified for calculations [68]. According to the theoretical work done, the MIP with aniline as a template was prepared by emulsion polymerization method using acrylic acid as the best functional monomer, divinylbenzene as

cross-linker dissolved in carbontetrachloride. The synthesized MIP was then tested by equilibrium adsorption method, and the MIP demonstrated high removal efficiency for aniline.

Also, the interactions between alkylorthosilicates and a template within the scope of molecular imprinting were studied to screen the best functional monomer which is able to achieve maximum interaction energy with the template molecules [68]. Substituted tetramethylorthosilicate monomers symbolized as XTMOS were studied. TMOS, NH2TMOSm COOTMOS, SHTMOS NH2PhTMOS and PhNHTMOS were screened for the best functional monomer. PhNHTMOS was the best monomer to be used with β-damascenone which is a terpinic ketone found in the essential oils of many natural materials, as a template molecule. Computation results suggest that the use of the 3-21G basis set concomitantly with a method for basis set superimposition error(BSSE) correction represents a good compromise between the level of theory and the time consumed during computation for successful screening of functional monomers.

Also, another group dedicated their computational work to develop and apply state-of-the-art computational tools to achieve an understanding of intermolecular interactions in molecular imprinting of theophylline into complex polymeric systems [70]. Molecular dynamics (MD) simulations were carried out on different molecular systems in order to predict the interaction energies, the closest approach distances and the active site groups between the simulated molecular systems and different bio-ligands. The energy minimized structures of five ligands, theophylline and its derivatives (theobromine, theophylline-8-butanoic acid, caffeine and theophylline-7-acetic acid) have been obtained with the use of molecular mechanics approach. MD simulations at room temperature were carried out to obtain equilibrated conformations in all cases. The first simulated molecular systems consisted of a ligand and commonly used functional monomers, polymers and a substrate. The second simulated molecular system consists of a ligand and a monomer or polymer using a solvent (ethanol). During this study, it was found that electrostatic interactions play the most significant role in the formation of molecular imprinting materials. The simulated functional monomers and polymers with ligands indicate that the functional groups interacting with ligands tend to be either –COOH or CH2=CH– groups.

Another computational approach to be mentioned here was devoted to simulate the formation of possible imprinted polymers in acetonitrile solution for theophylline (THO) using combined molecular dynamics (MD), molecular mechanics (MM), docking and site mapping computational techniques [71]. Methacrylic acid (MAA) and methylmethacrylate (MMA) monomers were used to simulate possible homo and copolymer structures. The model is able to predict binding affinity and selectivity when considering THO analogues, such as caffeine, theobromine, xanthine and 3-methylxanthine. Comparison with available experimental data is also proposed.

1.8. THE POTENTIAL APPLICATIONS OF MOLECULARLY IMPRINTED MATERIALS

Due to their promising characteristics recognition elements tailored using molecularly imprinted polymers find nowadays numerous applications in a lot of fields including

medicine, science, engineering, and all fields of science and technology. MIP have been applied as trapping elements for separating and extraction of molecules, antibody analogues and receptors, molecular catalysts, as recognition elements for sensory applications, and as membranes for separation of molecules and different processes which are diffusion dependent.

Moreover, MIPs have been recently applied for two important aspects. On one hand this technique has been ameliorated for the chiral recognition and separation of stereo-isomers of the same compound. On the other hand it introduced a very reasonable solution for one of the most challenging problems in molecular recognition by imprinting big molecules like enzymes, hormones and other big protein molecules. This category of huge molecules is very difficult to be recognized by ordinary technique. We will discuss here a little bit in more details the foregoing applications.

1.8.1. SEPARATION AND EXTRACTION OF MOLECULES

Molecularly imprinted materials (MIM) could best be used as stationary phases in chromatographic separation equipments especially for enantiomeric separations. The MIP based stationary phases is more advantageous and promising than ordinary separation elements in that it does not depend only in separation process based on the polarity difference between the stationary and mobile phases or the size of the molecules relative to embedded particles but also depends on the shape size complementarities with respect to the imprinted sites.

Consequently the separation processes will more sharp than in the ordinary cases. These chromatographic materials have allowed the separation of numerous compounds such as naproxen, anti-inflammatory drug [72], timolol, a beta-blocker-A [73], and nicotine [74],. Separation performance can be expressed by a factor α, which is only dependent on the retention times of the two enantiomers, or the resolution factor R_s which takes into account the breadth of the chromatographic peaks. The higher those factors are, the better is the separation. Noticeable separations have been obtained. For example, Rs = 4.3 for a racemic mixture of phenyl-α-mannopyranoside [75].

Also, the molecularly imprinted materials (MIM) were successfully used for thin-layer chromatography [76] and capillary electrophoresis. Moreover, it has been reported on many occasions that the extraction on the surface of imprinted polymers can give better results than standard techniques such as liquid-liquid extraction or extraction on C18 phase. An example of these is the extraction of some analgesic drug, from human serum [77]. The extraction with the imprinted polymers showed high selectivity toward the extracted drug.

1.8.2. PREPARATION OF ANTIBODY ANALOGUES

One another interesting application for the MIPs is tailoring of antibody analogues. Recent studies have shown that the efficiency of antibodies and artificial receptors prepared by the molecular imprinting technique is relatively promising. Also these studies demonstrated the possibility of their application in therapeutic trials (clinical trials) [78, 79]

which have always been one of the challenging aspects in drug delivery systems. A synthesized receptor for morphine and for leu-enkephaline was found to be highly efficient not only in organic media but also in aqueous solutions [80]. Diazepam, a tranquilizer drug, synthesized receptors using molecular imprinting technology, showed also high selectivity to those of monoclonal antibodies when compared to similar drugs (benzodiazepines) [80]. From these examples we might draw a conclusion that antibody analogues (clinical trials) synthesized using molecular imprinting technique can be designed for various molecules which are of biological interest with a very selective rebinding capacity.

1.8.3. MOLECULARLY IMPRINTED CATALYSTS

One of the most interesting applications of molecular imprinting is the preparation of catalysts for molecular reactions. This area of research is very promising not only for being less expensive compared to other precious metals based catalysts but also for the availability of use to catalyzing different reactions and chemical processes. In principle, molecularly imprinted catalysts can be prepared if a substrate (a product or a transition state analog) could be successfully imprinted into a specific material. The imprinted sites would then correspond to a large extent to the substrates being formed in the transition state. As a result of this, the transition state is stabilized by adsorption on the imprinted surface. Stabilizing of the transition state underlying reaction pathway from reactants to products on the surface of the catalyst ameliorates the whole rate of reaction. The first artificial antibody used for hydrolysis of p-nitro phenyl acetate was prepared using a transition state analog, p-nitro phenyl methylphosphonate as a template. Polyvinylimidazole cross-linked by 1, 4-dibromobutane in the presence of the transition state analog exhibited 1.7 fold higher activities for catalyzing the planned hydrolysis reaction compared to that of a non-imprinted reference polymer by the addition of the template molecule during polymerization. This result suggests that imprinted cavities operated successfully as catalytic sites for the reactants undergoing certain reaction. Other reactions catalyzed by MIP, were described like Diels-Alder reactions [82,83], aldol condensations [84,85]and isomerization of benzisoxazoles [86], etc.

Finally, it was reported that enzymes could be modified using the principles of molecular imprinting technique to modulate their action [87]. Although the catalytic activity of MIP networks is still below that of real enzyme systems, they could be considered as a future perspective.

1.8.4. MOLECULARLY IMPRINTED MATERIALS FOR SENSORY APPLICATIONS

One of the most promising aspects of molecularly imprinted materials is their application as recognition elements in the design of biosensors [88] A good demonstration for the principle underlying the mechanism of sensors based on the idea of molecularly imprinted materials is given below in Figure 2. As shown in the diagram the recognition ability of the imprinted element is mainly determined by shape and size of the underlying molecules. In other words the molecule of right shape and size is almost able to rebind back to the

imprinted molecular cavities while other molecules are unable to reside into the molecular cavity of the template.

The molecularly imprinted material, as indicated in the figure is in contact with a transducer which converts the chemical or the physical signal obtained due to adsorption of the analyte into the molecular cavities to an easily quantifiable electrical signal. Various characterization techniques for the rebinding process of the detected molecules based on different transduction principles have been reported many times in the literature. Of these techniques, we can refer to ellipsometry [89] evanescent wave IR [90], fluorescence [91], amperometry [92][93], voltammetry [93] and pH based transducers [94].

1.8.5. MOLECULARLY IMPRINTED MEMBRANES

The last application to be discussed here is the molecularly imprinted membranes. Molecularly imprinted membranes were applied well for the separation process of different types of molecules. A series of enantioselective imprinted polymer membranes for amino acids and peptide derivatives were prepared using oligopeptide as functional monomers [95-98]. Also, it was reported that a surface modifications of porous membranes with molecularly imprinted polymers showed great improvement of the membrane efficiency which was indicated by better selective diffusion ability of the imprinted molecule. As an example, some membranes were photo-grafted with a functionalized monomer, 2-acrylamido-2-methylpropanesulfonic acid and a cross-linker, N,N'-methylenebis (acrylamide) in the presence of a template molecule, desmetryn, in water [99].

REFERENCES

[1]   F. H. Dickey, *Natl. Acad. Sci. USA*, 35 (1949) 227.
[2]   F. H. Dickey, *J. Phys. Chem*, 59 (1959) 695.
[3]   K. Raman, M. T. Anderson, C. J. Brinker, *Chem Mater*, 8 (1996) 1682.
[4]   S. Mann, S. L. Burkett, S. A. Davis, C. E. Fowler, N. H. Medelson, S. D. Sims, D. Walsh, N. T. Whilton, *Chem. Mater*, 9 (1997) 2300.
[5]   D. Kriz, O. Ramstrom, K. Mosbach, *Anal. Chem*, (1997) 345A.
[6]   Q. F. H. Sambe, C. Kagawa, K. K. Kunimoto, J. Haginaka, *Anal. Chem*, 75 (2003) 191.
[7]   O. Ramstrom, K. Skuder, J. Haines, P. Patel, O. Bruggeman, *J. Agric. Food Chem*. 49 (2001) 2105.
[8]   P. Spegel. Schweitz, S. Nilsson, *Anal.Chem*, 75 (2003) 6608.
[9]   G. Mayes, M. J. Whitcombe. *Advanced Drug delivery reviews*, 57 (2005) 1742.
[10] G. Wulff, A. Sarhan, Angew. *Chem. Int. Ed. Engl*, 11 (1972) 341.
[11] G. Wulff, R.Vesper, R.Grobe Einsler, A. Sarhan, Makromol. Chem, 78 (1977) 2799.
[12] G. Wulff, S. Schauhoff, *J. Org. Chem*, 56 (1991) 395.
[13] Kugimiya, J. Matsui, T. Takeuchi, K. Yano, H. Muguruma, A.V. Elgersma, J. Karube, *Anal. Lett*, 28 (1995) 2317.
[14] Kugimiya, J. Matsui, H. Abe, M. Aburatani, T. Takeuchi, *Anal. Chim. Acta*, 365 (1998) 75.

[15] G. Wulff, J. Vietmeier, *Makromol. Chem*, 190 (1989) 1727.

[16] N. Sallacan, M. Zayats, T. Bourenko, A. B. Kharitonov, I. willner, *Anal. Chem*, 74 (2002) 702.

[17] G. wulff, W. Best, A. Akelah, *React. Polymer*, 2 (1984) 167.

[18] K. J. Shea, T. K. Dougherty, *J. Am. Chem. Soc*, 108 (1986) 1091.

[19] K. J. Shea, D. Y. Sasaki, *J. Am. Chem. Soc*, 111 (1989) 3442.

[20] K. J. Shea, D. Y. Sasaki, *J. Am. Chem. Soc*, 113 (1991) 4109.

[21] J. Damenm, D. C. Neckers, *Tetrahedron.Lett,* 21 (1980) 1913.

[22] J. Damenm, D. C. Neckers, *J. Org. Chem*, 45 (1980) 1382.

[23] J. Damenm, D. C. Neckers, *J. Am. Chem. Soc*, 102 (1980) 3265.

[24] Nada F. Atta, Maher M. Hamed, Ali M. Abdel-Mageed, *Anal Chim. Acta* 667 (2010)63

[25] R. Arshady, L. Glad, K. Mosbach, *J. Chromotogr*, 299 (1981) 687.

[26] O. Norrlöw, M. Glad, K. Mosbach, *J. Chromatogr*, 299 (1984) 29.

[27] Pietrzyk, R. Wiley, D. McDaniel, *J. Org. Chem*, 22 (1957) 83.

[28] Yu, K. Mosbach, *J. Mol. Recognit*, 11 (1998) 69.

[29] O. Ramström, L. I. Andersson, K. Mosbach, *J. Org. Chem*, 58 (1993) 7562.

[30] Z. H. Meng, J. F. Wang, L. M. Zhou, Q. H. Wang, *Anal. Sci*, 15 (1999) 141.

[31] Do-Hyeon Yang, Naoki Takahara, Seung Woo Lee, Toyoki Kunitake, *Sensors and Actuators B*, 130 (2008) 379.

[32] Keith Farrington, Fiona Regan, *Biosensors and Bioelectronics*, 22 (2007) 1138.

[33] Wensheng, B. Ram, Gupta, *Separation and Purification Technology*, 35 (2004) 215.

[34] Hui-Jing Liang, Tzong-Rong Ling, John F. Rick, Jse-Chuan Choub, *Anal. Chim. Acta*, 542 (2005) 83.

[35] Feng Liu, Xiao Liu, Siu-Choon Ng, Nardy Sze-on Chan, *Sensors and Actuators*, B113 (2006) 234.

[36] Huai You Wang, Ji Gang Jiang, Li Ying Ma, Yan Ling Pang, Reactive and *Functional Polymers* 64 (2005) 119.

[37] Keith Farrington, Fiona Regan, *Biosensors and Bioelectronics*, 22 (2007) 1138.

[38] D. Silvestri, N. Barbani, C. Cristallini. P. Giusti, G. Ciardelli, *Journal of Membrane Science*, 282 (2006) 284.

[39] Y. Lu, C. X. Li, X. D. Wang, P. C. Sun, X. H. Xing, *Journal of Chromatography* B, 804 (2004) 53.

[40] David A. Spivak, *Advanced Drug delivery Reviews*, 57 (2005) 1779.

[41] Wenguo Donga, Ming Yan, Zheng Liu, Guoshi Wu, Yanmei Li, *Separation and Purification Technology* 53 (2007) 183.

[42] Hui-Jing Liang, Tzong-Rong Ling, John F. Rick, Jse-Chuan Choub, *Anal. Chim. Acta*, 542 (2005) 83.

[43] Nada F. Atta, Ali M. Abdel-Mageed, Talanta 80 (2009) 511-518.

[44] Kuo-Chuan Hoa,, Wei-Ming Yeh a, Tsai-Shih Tung , Jung-Yu Liao, *Analytica Chimica Acta* 542 (2005) 90.

[45] Israel S´anchez-Barrag´an a, Kal Karim b, Jos´e M. Costa-Fern´andez, Sergey A. Piletsky, Alfredo Sanz-Medel, *Sensors and Actuators B: Chemical*, 123 (2007)798.

[46] Zhaohui Zhang, Haiping Liao, Hui Li, Lihua Nie, Shouzhuo Yao, Analytical Biochemistry, 336 (2005)108.

[47] H. Asonuma, T. Akiyama, K. Kajiya, T. Hishiya, M. Komiyama, *Anal. Chim. Acta*, 435 (2001) 25.

[48] Yan Lu, Chenxi Li, *Anal. Chim. Acta*, 489 (2003) 33.

[49] C. Wensheng, B. Ram, Gupta, *Separation and Purification Technology*, 35 (2004) 215.

[50] Hui-Jing Liang, Tzong-Rong Ling, John F. Rick, Jse-Chuan Choub, *Anal. Chim. Acta*, 542 (2005) 83.

[51] Feng Liu, Xiao Liu, Siu-Choon Ng, Nardy Sze-on Chan, *Sensors and Actuators*, B113 (2006) 234.

[52] Keith Farrington, Fiona Regan, *Biosensors and Bioelectronics*, 22 (2007) 1138.

[53] D. Silvestri, N. Barbani, C. Cristallini. P. Giusti, G. Ciardelli, *Journal of Membrane Science*, 282 (2006) 284.

[54] Fernando Sineriza, Yasunori Ikeda, Emmanuel Petit, Laurent Bultel, Karsten Haupt, Jos eKovenskyb and Dulce papy-Garciaa, Tetrahedron, 63 (2007) 1857.

[55] Kuo-Chuan Haa, Wei-Ming Yeh, Tsai-Shih Tung, Junag-Ya Liao, *Anal. Chim. Acta*, 542 (2001) 90.

[56] Linden D. Bolisaya, James N. Culverb, Peter Kofinusa, *Biosensors and Bioelectronics*, 27 (2006) 4965.

[57] Xiaolan Zhua, Jibo Cai, Jun Yang, Qingde Sua,Yun Gaob, *Journal of Chromatography A*, 1131 (2006) 37.

[58] Hui You Wang, Ji Gang Jiang , Li Ying Ma, Ya ling pang, *Reactiveand Functional Polymers*, 64 (2005) 119.

[59] Sharon Marx, Amalya Zaltsman ,Iva Turyan and Daniel Mandler, *Anal. Chem*, 76 (2004) 120.

[60] A. Fernández-González , R. Bad´ıa La´ıño , M.E. Diaz-Garc´ıa , L. Guardia , A. Viale, *Journal of Chromatography B*, 804 (2004) 247.

[61] S. Firenman-Shoresh, Iva Turyan, Daniel Mandler, David Avir, Sharon Marx, *Lungmuir*, 22(2005)7842.

[62] Zhaohur Zhang, Yumer Lang, Lihua Nie, Shouzhuo Yao, *Biosensors and Bioactuators*, 21(2006) 1244.

[63] Raquel Gomes de Costa, Fabio Augusto, Journal of Chromatography A, 1114 (2006) 216.

[64] M. E. Brown, D. A. Puleo, *Chemical Engineering Journal*, 137 (2008) 97.

[65] Peter A. Lieberzeit, Adeel Afzal, Gerd Glanzing, Franz L. Dickert, *Anal. Bioanal. Chem*, 389 (2007) 441.

[66] Do-Hyeon Yang, Naoki Takahara, Seung-Woo Lee, Toyoki Kunitake, *Sensors and Actuators B,* 130 (2007) 385.

[67] Qing-Zhi Zhu, Karsten Haupt, Dietmar Knopp, Reinhard Niessner, *Anal. Chim. Acta*, 468 (2002) 217.

[68] James. B. Foresman, Aeleen. Frisch, Exploring chemistry with Electronic Structure Methods, Second Edition, Gaussian Inc, (1994) (Text Book).

[69] Manuel AZenh, Porkod, Kathirvel, Pedro Nogueira, Antonio Fernando-Silva, *Biosensors and Bioelecctronics*, 23 (2008) 1843.

[70] Dumitru Pavel, Jolanta Lagowski, *Polymer* 46 (2005) 7543.

[71] Susanna Monti , Chiara Cappelli , Simona Bronco , Paolo Giusti , Gianluca Ciardelli, *Biosensors and Bioelectronics* 22 (2006) 153.

[72] Katz, Davis ME, *Nature*, 403 (2000) 286.

[73] L. Fischer, R. Müller, B. Ekberg, K. Mosbach, *J. Am. Chem. Soc*, 113 (1991) 9358.

[74] A. Zander, P.Findlay, T. Renner, B. Sellenger, *Anal. Chem*, 70 (1998)3304.

[75] G. Wulff, M. Minarik, *Liq Chromatogr*, 13 (1990) 2987.

[76] D. Kriz, C. Berggren, Li. Anderson, K. Mosbach, *Anal. Chem*, 66 (1994) 2636.

[77] Li. Anderson, A. Paprica, T. Arvidsson, *Chromatographia,* 46 (1997) 57.

[78] G. Vlatakis, Li. Andersson, R. MÜller, K. Mosbach, *Makromol Chem*, 361(1993) 645.

[79] O. Ramstörm, L.Ye, K. Mosbach, *Chem. Biol.*, 3 (1996) 471.

[80] M. Kempe, K. Mosbach, *Tetrahedron L*ett, 36 (1995) 3563.

[81] D. K. Ronbinson, K. Mosbach, *J. Chem. Soc, Chem. Commun.*, 14 (1989) 969.

[82] Xc. Liu, K.Mosbach, *Macromol. Rapid Commun.*, 18 (1997) 609.

[83] J. Matsui, IA. Nicholls, I. Kerube, K.Mosbach, *Org. Chem*, 61 (1996) 414.

[84] Xc. Liu, K. Mosbach, *Macromol. Rapid Commun.*, 19 (1998) 671.

[85] Bracol, Dabulisk, AM. Klibanow, *Natl. Acad. Sci. USA*, 87 (1990) 671.

[86] FL. Dickert, R. Sikoriski, *Mater. Sci. Eng*, 10 (1999) 39.

[87] LI. Anderson, CF. Mandenius, K. Mosbach, *Tetrahedron Lett.*, 29 (1988) 437.

[88] M. Jakusch, M. Janotta, B. Mizaikoff, K. Mosbach, K.Haupt, *Anal. Chem*, 71 (1999) 4786.

[89] D. Kriz, O. Ramström, A. Svensson, K. Mosbach, *Anal. Chem*, 67 (1995) 2142.

[90] S. Al-kindy, R. Badea, M. Deaz-Garcia, *Anal Lett*, 35 (2002) 1763.

[91] D. Kriz, K. Mosbach, *Anal. Chim. Acta*, 300 (1995) 71.

[92] S. Kröger, A. Turner, K. Mosbach, K. Haupt, *Anal. Chim. Ac*ta, 71 (1999) 698.

[93] C. Malitesta, I. Losito, PG. Zambonin, *Anal. Chem*, 71 (1999) 1366.

[94] G. Chen, Z. Guan, C. Chen, L. Fu, V. Sundersan, FH. Arnold , *Nature Biotechnol*, 15 (1997) 354.

[95] J. M. Krotz, K.J.Shea, *J. Am. Chem. Soc*., 118 (1996) 8754.

[96] M. Yoshikawa, J. Izumi, T. Kitaao, *Chem. Lett.,* (1996) 611.

[97] M. Yoshikawa, J. Izumi, T. Kitao, Sakamotos, *Makromol. Rapid. Commun.*, 18 (1997) 761.

[98] M. Ykoshikawa, T. Fujisawa, J. Izumi, T. Kitao, S. Sakamoto, *Anal. Chim. Acta*, 365 (1998) 59.

[99] S. A. Piletsky, H. Matuschewski, U. Schedler, A. Wilpert, E. V. Pilteska, E, T. A. Thiele, M. Ulbricht, *Macromolecules,* 33 (2000) 3092.

Chapter 6

RECENT ADVANCES IN DNA-LIGAND MOLECULAR RECOGNITION AND ALLOSTERIC INTERACTIONS

*Jonathan T. B. Huang[1], Robin C. K. Yang[1], Wei-Kang Hung[1], Michael J. Waring[2] and Leung Sheh[1]**

[1] Department of Chemistry and Life Science Research Center,
Tunghai Christian University, Taichung 407, Taiwan, R.O.C.
[2] Department of Pharmacology, University of Cambridge,
Tennis Court Road, Cambridge CB2 1PD, England

ABSTRACT

It is generally recognized that elucidating the molecular basis for recognition of specific sequences in target DNA by proteins and small synthetic molecules vitally underpins research on the modulation of gene expression. In this review we discuss the fundamental basis of DNA sequence recognition by small molecules at the atomic bonding level based on recent X-ray diffraction results together with circular dichroism spectra and footprinting experiments on DNA-small ligand binding. Monodentate (single) interactions, intrastrand bidentate interactions and interstrand bidentate interactions are considered not only central to the capabilities of small ligands and proteins to recognize DNA sequences, but also provide the means for allosteric communication between multiple DNA binding loci. Thermodynamic, kinetic and allosteric features of molecular recognition by drugs and small ligands within the minor groove of DNA are reviewed. Allosteric interactions between small synthetic peptides and multiple DNA binding sites are discussed, and hypothetical models are proposed to interpret the complex allosteric communication process. In contrast to protein-protein interaction networks which have been extensively investigated, studies on small ligand-DNA interaction networks have only recently been commenced. In this review, three different types of novel allosteric interaction networks between peptides and DNA are considered, together with hypothetical models featuring monodentate interactions and interstrand bidentate interactions. The new concept of DNA-small ligand interaction networks illuminates

---

* Correspondence Tel: +886-4-23590248; fax: +886-4-23590426; E-mail: Lsheh@thu.edu.tw

some basic chemical rules of DNA-small ligand allostery and may find applications in future drug design as well as structural biology research.

1. INTRODUCTION

It has long been known that the heredity of living organisms is mediated by genes contained in the nuclei of cells in the form of chromosomes. Genes are long polymers of 2'-deoxyribonucleotides in double helical array, generally stabilized by nuclear proteins. The base pairing of DNA is size-complementary, that is, the large purine bases always hydrogen-bond to the small pyrimidine bases, maintaining the AT and GC pairs having almost identical dimensions. In aqueous solution, the size-complementarity of the bases and the double helical conformation of DNA in the B-form provide a wide (major) groove about 12 Å wide running in parallel with a narrow (minor) groove of about half the width.

Sequence-specific interactions between DNA and transcription factors are central to the implementation and maintenance of genomic expression. Proteins, because of their relatively large size, bind predominantly to the major groove. Well-known DNA binding motifs include the helix-turn-helix motif, zinc finger motif, homeobox domain, and bZip motif. Early studies of interactions between DNA and the amino acid side chains of proteins were pioneered by several research groups. Seeman *et al.* identified hydrogen-bonding atoms on the edges of DNA base pairs and suggested that greater specificity was more likely to arise through amino acid side chain interactions in the major groove rather than the minor groove [1]. Suzuki established early chemical rules for the recognition of DNA bases by amino acid side chains [2]. Mandel-Gutfreund *et al.* presented a systematic analysis of hydrogen bonding between regulatory proteins and DNA [3]. Later, Thornton *et al.* presented a comprehensive study of DNA-protein interactions at the atomic level, investigating hydrogen bonds, van der Waals contacts and water-mediated bonds in 129 DNA-protein complexes [4]. Cheng and Frankel computed *ab initio* interaction energies for 21 hydrogen-bonded amino acid side chain-nucleic acid interactions [5].

Interactions of DNA bases with proteins via the amino acid side chains can be divided into several main categories: monodentate interactions (single interactions), intrastrand bidentate interactions, interstrand bidentate interactions, and complex interactions [4]. In this review we shall concentrate on the role of intra- and inter-strand bidentate interactions in DNA sequence recognition by peptides and in DNA-peptide allosteric interactions.

2. MONODENTATE INTERACTIONS AND BIDENTATE INTERACTIONS

Hydrogen bonds between DNA bases and amino acid residues of proteins or small ligands are paramount for binding and molecular recognition. They are weak non-covalent interactions with bond energy around 4.5 kcal/mol. Their weak bonding character is vital for rapid bond formation and dissociation in molecular recognition processes in biological systems. Monodentate interactions arise when a single hydrogen bond forms between one hydrogen acceptor/donor atom of the DNA base and a corresponding hydrogen donor/acceptor atom of the amino acid side chain [4]. Most proteins interact with DNA in the major groove whereas many conjugates/peptides incorporating 4-amino-1-methylpyrrole-2-

carboxylic acid residues (Py) bind preferentially in the minor groove. In the major groove, the N7 of A and G, O6 of G, and O4 of T frequently act as acceptors for hydrogen bonds, whereas the C6 amino group of A and the C4 amino group of C typically operate as hydrogen bond donors (Figure 1). In the minor groove, the N3 of A and the O2 of T and C can act as acceptors for hydrogen bonding whereas the C2 amino group of G is the only available hydrogen bond donor (Figure 1).

Bidentate interactions are further sub-divided into intrastrand and interstrand modes [4,6,7]. Intrastrand bidentate interactions refer to hydrogen bonds formed between two atoms of a base and one atom of an amino acid side chain, or between one atom of a base and two atoms of an amino acid. Intrastrand bidentate interactions were identified by Cheng and Frankel from *ab initio* interaction energies of 21 hydrogen-bonded amino acid side chains with nucleic acids [5]. The most favorable interactions are Lys-G, Arg-G, Asn-A, Ser-A and Gln-A pairings in the DNA major groove.

A
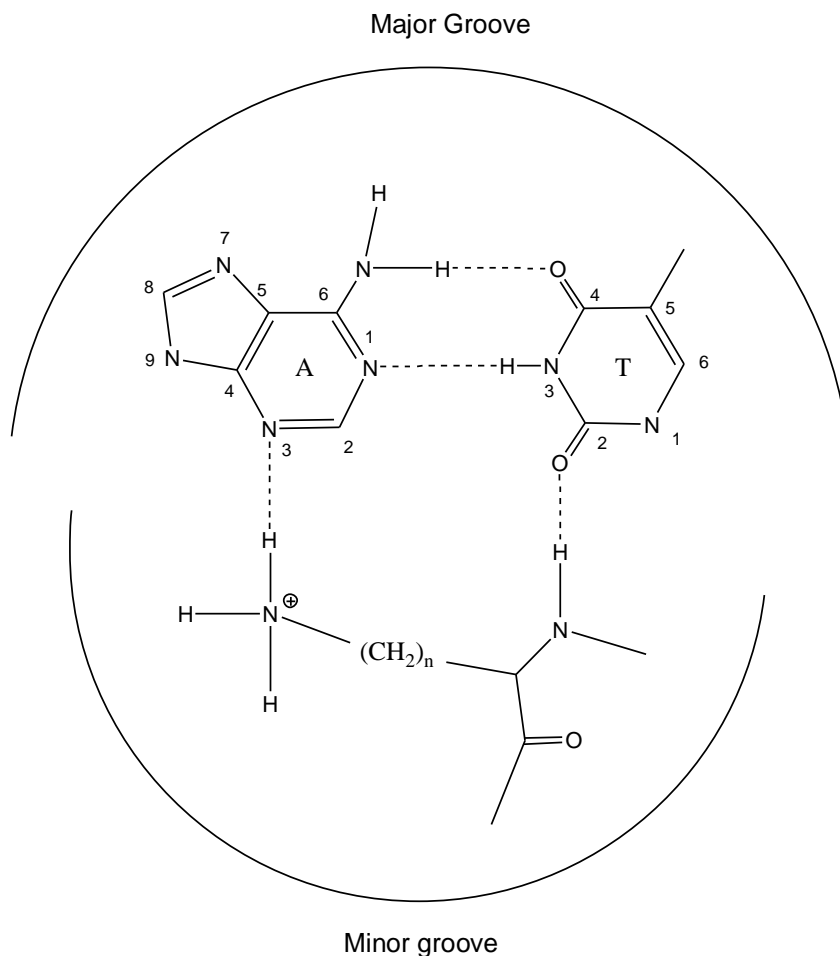
Figure 1. (Continued).

B
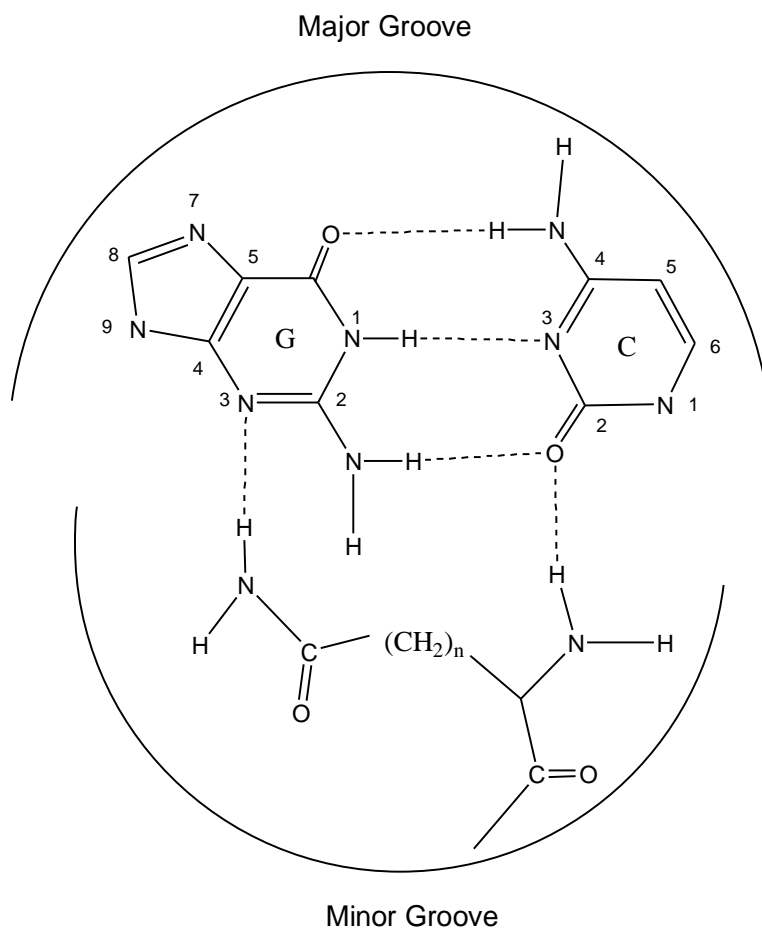
Figure 1. Interstrand bidentate interactions between peptide moieties and bases in the DNA minor groove: proposed interstrand bidentate interaction between an Orn (n = 3) or a Lys residue (n = 4) with a AT bp (panel A); proposed interstrand bidentate interactions between an Asn (n = 1) or a Gln residue (n = 2) with a GC bp (panel B) [proposed figures drawn according to X-ray data of reference [4] (Table 2)].

Interstrand bidentate interactions, on the other hand, are hydrogen bonds that form between two or three donor/acceptor atoms of an amino acid residue (including the α−amino function and side chain functional groups) with two atoms belonging to complementary bases on the two DNA strands [4] (Figure 1). Interestingly, protein residues Asn, Gln, Ser and Tyr can interact with both the major and minor grooves of DNA whereas there are several instances where Arg interacts preferentially with hydrogen bond acceptor atoms belonging to A-T pairs in the minor groove. Some years ago we reasoned that structural information deriving from DNA-protein interactions could be applied to the study of DNA-peptide interactions so far as hydrogen-bonded L-amino acid side chain-DNA interactions in the minor groove are concerned.

3. DNA-SMALL LIGAND MOLECULAR RECOGNITION IN THE MINOR GROOVE

Extensive studies employing diverse methodologies have been carried out with small ligands that are capable of sequence-specific or sequence-selective recognition of DNA. These small ligands include drugs [8-10], antitumor antibiotics [11-14], netropsin/distamycin and its analogues [15-19], and synthetic peptides [6,7, 20-24]. A deep understanding of molecular recognition of DNA by small molecules is important for disclosing molecular aspects of the modulation of gene expression *in vivo* as well as for augmenting the progress of structural biochemistry research and drug design.

An early breakthrough in molecular recognition research affecting the DNA minor groove came with the finding of the bis-intercalation of the two quinoxaline chromophores of the antitumor antibiotic echinomycin into DNA [25, 26]. DNase I footprinting and X-ray diffraction studies showed that echinomycin and its biosynthetic precursor triostin bind preferentially to CpG sequences, with hydrogen bonds forming between the 2-amino groups of guanines and the carbonyl oxygen atoms of the Ala residues, and between the N3 of guanines and the amide protons of Ala [25-27]. Recent studies have gone on to show that echinomycin specifically inhibits the DNA binding activity of hypoxia-inducible factor-1 (a transcription factor that controls genes involved in glycolysis, angiogenesis, migration, and invasion of tumors), suggesting that this antibiotic could have an important prospective in the regulation of gene expression in tumor progression and metastasis [28].

As mentioned before, the molecular basis of sequence-selective or sequence-specific interactions with DNA by drugs or other small ligands is furnished mainly by formation of adequate hydrogen bonds between functional groups of peptides or drugs and DNA bases (monodentate interactions and intrastrand/interstrand bidentate interactions). Other forms of non-covalent interaction may also occur, comprising ionic interactions between positively-charged groups of a ligand and the DNA phosphates, hydrophobic interactions, and π–interactions between intercalating aromatic groups of ligands and the DNA bases. Many synthetic minor groove-binding agents are derivatives or congeners of the antiviral antibiotics netropsin and distamycin that contain 4-amino-1-methylpyrrole-2-carboxylic acid residues (Py). X-ray diffraction studies have shown that the 4-amino group of Py residues in these compounds acts as a hydrogen bond donor to the N3 or O2 of T, while the 3-methine proton of Py is in van der Waals contact with the C2 proton of A. The netropsin molecule fits into the minor groove with guanine stacked on a pyrrole ring of the antibiotic [29]. Lexitropsins containing two thiazole rings with the sulfur atoms directed away from the minor groove bind to alternate purine-pyrimidine sequences such as 5'-TATGAC-3' and 5'-TGCATGC-3' [30]. Similar DNA binding results were obtained with the furan-containing lexitropsins [30]. It has been shown that two distamycin molecules can associate in a side-by-side and antiparallel head-to-tail orientation within the minor groove, and this finding prompted many subsequent designs for minor groove binders such as homo- and hetero-dimeric compounds [31]. Polyamides containing the imidazole residue were found to bind preferentially to GC pairs via hydrogen bonding with the 2-amino group of G; moreover, judicious juxtaposition of Py and imidazole residues allowed G · C pairs to be differentiated from C · G pairs [32].

The thermodynamic basis of molecular recognition between small ligands and DNA has been investigated by several research teams. For many studies isothermal titration calorimetry

(ITC) was the method of choice to provide the thermodynamic parameters. In 1987, Breslauer et al. reported enthalpy-entropy compensation studies based on netropsin, distamycin, ethidium bromide, and daunomycin binding to poly[d(A-T)] • poly[d(A-T)] and poly[d(A)] • poly[d(T)] using calorimetry [33]. Binding energetics of acridine-based antitumor agents were investigated by Graves et al., showing that these drugs exhibit a minor groove binding preference [34]. The enhanced binding enthalpies of C5-substituted analogues were found to be correlated with anticancer activity. Replacing various functional groups on anthracycline antibiotics resulted in a binding free energy penalty, revealing that total ligand binding energies are partitioned among various substituents in these antibiotics [35]. Enthalpy-entropy compensation has been reported in the binding of 7-amino actinomycin D to several short DNA duplexes [36]. Wilson et al. showed that a number of reversed amidine heterocyclic compounds bound to the minor groove of DNA in a mostly entropy-driven manner, with positive $T\Delta S$ values around 4.3-6.0 kcal/mol [37].

Kinetic aspects of the interaction of DNA with drugs or other small molecules were early explored by several research teams employing complementary approaches. As long ago as 1976, Davanloo and Crothers reported the kinetics of formation and dissociation of drug-dinucleotide complexes and actinomycin-nucleotide complexes [38]. In 1981, Fox, Wakelin and Waring reported that the speeding up of rate constants with an increasing level of echinomycin binding to DNA could indicate the occurrence of cooperativity associated with progressive changes in the DNA structure [39]. Stopped-flow kinetic studies on echinomycin-DNA binding showed a single exponential and suggested a molecular mechanism of binding in which both chromophores of the antibiotic become intercalated simultaneously rather than sequentially, and no transition from a mono-intercalated state to a bis-intercalated state could be detected [39]. Using stopped-flow and temperature-jump approaches, Chaires et al. proposed the mechanism of daunomycin-DNA interaction to be two-stepped: a rapid "outside" binding of daunomycin to DNA, followed by intercalation of the drug [40]. Employing a fluorescence-detected stopped-flow assay, Crothers and coworkers studied the kinetics of a series of polyamides binding to DNA and pointed out that covalent linkage of the subunits results in polyamides with dramatically enhanced affinity due to faster association rates [41]. Moreover, the basis for discrimination between matched and mismatched sites for each polyamide was found to arise mainly from differences in the rates of dissociation from these sites. Fletcher and Fox developed a convenient method for studying the dissociation kinetics of echinomycin from CpG binding sites in different sequence environments monitored by the rate of disappearance of DNase I footprints [42]. Surface plasmon resonance (SPR) appears to be another appealing approach for studying DNA-ligand interactions since equilibrium constants, association and dissociation rate constants for monomer and dimer complex formation can be determined [37]. SPR also allows binding rates to be observed in real time for complex formation and dissociation. Wilson, and Lee et al. have been using SPR successfully to analyze the kinetics of reversed amidine heterocycles and polyamides binding to various oligonucleotide duplexes [37].

Against this background we sought to investigate the molecular basis of sequence recognition as well as complex allosteric aspects of DNA-peptide interactions by designing and synthesizing a number of peptides incorporating the XPRK or XHypRK motifs together with a tract of 4-amino-1-methylpyrrole-2-carboxylic acid residues (Py) [6,7,20-24]. Synthetic peptides based upon the XPRK or XHypRK motifs possess good DNA sequence-selective binding capability toward DNA sequences incorporating three or four consecutive

W (A or T) base pairs and afford satisfactory DNA footprinting results at sub-micromolar concentrations. These peptides also possess the advantage of good aqueous solubility and obviate the frequent use of organic solvents in DNA-small molecule binding studies.

Our logic for the design of the XPRK motif stems from the fact that a naturally occurring SPXX motif was found in repeating sequences in histones, steroid hormone receptors, various segmentation gene products and some oncogene products [43,44]. It was suggested that the SPXX motif assumes a β-turn stabilized by two hydrogen bonds, and that the side chains of the two basic residues engage in salt bridges with the DNA phosphate groups [43,44].

Quantitative DNase I footprinting originally turned out to be the best methodology for DNA-peptide sequence recognition experiments since most other methods were unable to discriminate peptide binding at multiple recognition sites. In addition, to investigate conformational changes involved in the molecular recognition of DNA we resorted to circular dichroism measurements which provide spectral characterization of DNA-peptide interactions [6,7].

4. DNA-LIGAND ALLOSTERIC INTERACTIONS

Allostery is generally recognized as an indispensable process in biological control regulating biochemical efficiency and energy metabolism in nature. For almost sixty years extensive research has been devoted to understanding the allosteric behavior of functional proteins and regulatory enzymes. The complex network-based allosteric regulation of ATP/GTP and NADH/FADH$_2$ generation in glycolysis and the tricarboxylic acid cycle are representative of 'biological energy economy' that is vital for the maintenance of life in higher animals including man. The term 'allosteric' or 'allostery' was originally used to describe a macromolecule-ligand interaction that results in local conformational change of the macromolecule, thereby facilitating the binding of further ligand molecules to the binding subunits/sub-sites (positive cooperativity), or making it more difficult for subsequent ligand molecules to bind to those subunits/sub-sites (negative cooperativity). The concerted-symmetry allosteric model of Monod, Wyman, and Changeux (MWC model) proposed that all of the protein subunits change shape in a concerted manner to preserve the symmetry of the entire molecule as it is transformed from the low affinity conformation (T state) to the high affinity conformation (R state) [45]. The sequential/induced fit allosteric model of Koshland, Nemethy, and Filmer (KNF model) proposed that each subunit changes shape as ligand binds, so that changes in one subunit lead to distortions in the shape and/or interactions of other subunits within the protein [46]. The KNF model appears to provide a better explanation of the phenomenon of 'negative cooperativity' than the MWC model. Williams et al. used a thermodynamic approach to define positive cooperativity as the decreased dynamic behavior of a receptor system with a benefit in enthalpy and a cost in entropy, and negative cooperativity as the increased dynamic behavior of a receptor system with a cost in enthalpy and a benefit in entropy [47,48]. Ever since 1960, while numerous papers on protein-ligand allosteric interactions have been published, both the MWC and KNF models have been widely used to interpret complex allosteric behavior of proteins.

On the other hand, studies on possible allosteric aspects of DNA-ligand interactions are far fewer, and much less is known about their results than protein-ligand interactions.

However, some early pioneering studies did establish a significant degree of progress towards understanding how allostery might influence DNA-ligand interactions. In 1980 Crothers et al. reported that the binding of distamycin to calf thymus DNA at low levels of saturation is positively cooperative [49]. Three years later Graves and Krugh reported that adriamycin and daunomycin also exhibit positive cooperative binding to various DNAs [50]. The appearance of initial positive curvature in the binding isotherm was found to be dependent on the ionic strength, suggesting a role for DNA flexibility in positive cooperativity. In 1992 Fagan and Wemmer reported the use of ^{1}H NMR spectroscopy to study the cooperative binding of distamycin to d(CGCIIICCGGC) + d(GCCIICCCGCG), where I represents inosine [51]. In 1996 Bailly, Hamy and Waring reported cooperativity in the binding of the antitumor antibiotic echinomycin to DNA using quantitative DNase I footprinting [52]. This study demonstrated that the binding of echinomycin to the sequences ACGTACGT and TCGAACGT is highly cooperative. Also in 1996, Morii et al. reported cooperative binding to DNA of peptides conjugated to adamantyl and β-cyclodextrin groups using gel mobility shift and circular dichroism assays [53]. In 2000 Chaires et al. reported that (+)daunorubicin binds selectively to right-handed DNA and can provoke apparently allosteric conversion of left-handed polynucleotide to a right-handed intercalated form, as well as *vice versa* for a synthetic (-)analogue [54]. Shortly afterwards Laughton et al. used a molecular dynamics approach to investigate cooperativity in drug-DNA recognition [55]. Their results led them to support Cooper and Dryden's hypothesis of allosteric communication without changes [56] in the time-average structure of the macromolecule, though binding free energy can be obtained from changes in conformational flexibility alone. In 2003 Fechter and Dervan used quantitative DNase I footprinting to investigate allosteric inhibition of GCN4 bZip protein binding to the DNA major groove by polyamide-acridine conjugates that bind to the minor groove [57].

In recent years we have been studying possible allosteric features of peptide-DNA interaction using designed peptides targeted to specific recognition sites in DNA [6,7,23,24]. Our experimental protocol is based upon quantitative DNase I footprinting of designed peptides binding to a 5'-^{32}P-labeled 158-mer DNA duplex and a complementary 5'-^{32}P-labeled 135-mer DNA duplex. Binding site positions on the upper and lower strands are designated as U and L, respectively. Like proteins, monodentate interactions and interstrand bidentate interactions between the ligands and the DNA bases have been recognized as essential contributors to molecular recognition of DNA by our synthetic peptides.

To illustrate the power of quantitative DNase I footprinting as one of the best analytical techniques for studying DNA recognition and allostery, we offer the following example. A designed decapeptide His-Hyp-Arg-Lys-(Py)$_4$-Lys-Arg-NH$_2$ (HyH-10) was footprinted using DNase I and serial peptide dilutions between 10 and 3000 nM [7]. The peptide was equilibrated with the DNA for an hour before enzyme cleavage, then the reaction was stopped, the gel was electrophoresed, and subjected to autoradiography (Figure 2). In previous reports [6,7,24], we describe how the position of interstrand bidentate interactions can be assigned by inspecting differential cleavage plots (experiment versus control) so that regions of significant simultaneous blockage of DNase I cutting on the complementary strands can be cross-correlated; these we indicate by pecked lines connecting base sequences on the two strands (Figure 3).
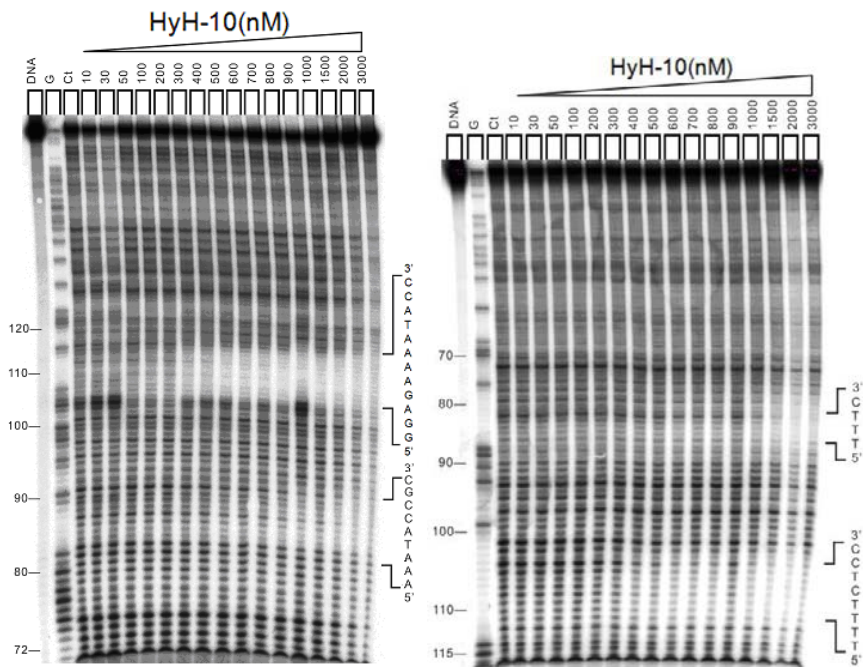
Figure 2. Example of a quantitative DNase footprinting autoradiograph showing the sequence selective binding of synthetic peptide HyH-10 on DNA duplexes labeled at either end: 5'-[^{32}P]-labeled 158-mer upper strand, left panel; and 5'-[^{32}P]-labeled 135-mer lower strand, right panel. Peptide HyH-10 was equilibrated with the DNA in 5 mM sodium cacodylate buffer, pH 6.5 at 37°C for 60 min before DNase 1 cleavage. G represents a Maxam-Gilbert guanine sequencing track and Ct shows a DNase I digestion control lane.
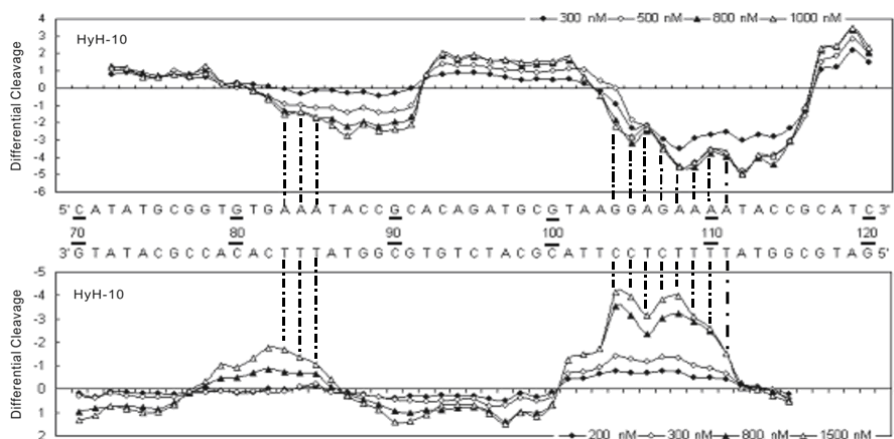
Figure 3. Differential cleavage plots comparing the susceptibility of DNA fragments to DNase I cleavage after incubation with peptide HyH-10 in cacodylate buffer at 37°C for 60 min. The upper traces represent the differential cleavage plot for a given peptide bound to the 5'-[^{32}P]-labeled upper strand (158-mer) DNA fragment; the lower traces represent the corresponding plots for each peptide bound to the 5'-[^{32}P]-labeled lower strand (135-mer) DNA fragment. The vertical dotted lines between DNA bases represent assignment of interstrand bidentate interactions where significant coincident H-bonding interactions occur involving complementary bases in both strands.
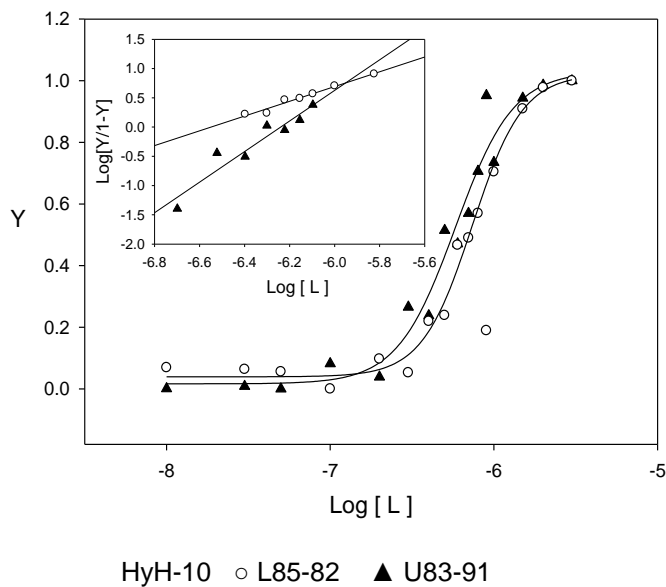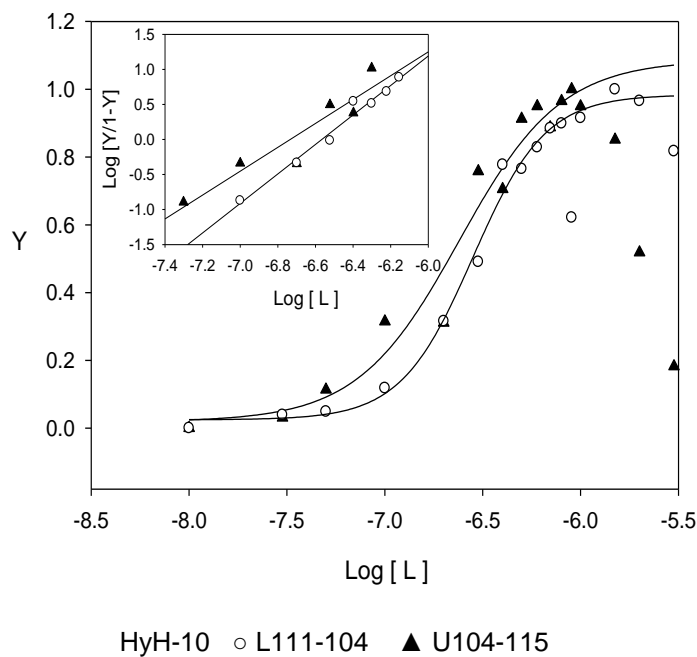
Figure 4. Binding isotherms from DNase I footprinting titrations of peptide HyH-10 (Y, the fractional saturation *versus* log molar concentration) for binding sites L111-104 and U104-115 (upper panel), and sites L85-82 and U83-91 (lower panel). Inset are the respective Hill plots of log[Y/1-Y] *versus* log[L]. The concentration of ligand (peptide) is in nM. The data points of Y at high peptide concentrations are scattered due to over-saturation.

Monodentate interactions are simply assigned by connecting regions of inhibition in the differential cleavage plot profile to bases where there is no simultaneous blockage on the complementary strand.

A Hill plot of log Y/(1-Y) *versus* log [L] yields the Hill coefficient (n_H) which records the degree of cooperativity of the binding process and allows us to compare cooperativity with respect to different binding sites as well as the relative cooperativity of different designed peptides binding to various discrete sites (see section 5, Figure 4).

The autoradiograph (Figure 2) and the differential cleavage plot (Figure 3) reveal that, on the upper strand of the 158-mer pBR322 fragment peptide HyH-10 produces a strong and very broad DNase I blockage around position U104-115, corresponding to the sequence 5'-GGAGAAAATACC-3', displaying positive cooperativity ($n_H = 1.7$), and there is also a broad DNase I blockage site around position U83-91, comprising the sequence 5'-AAATACCGC-3', where notable positive cooperativity ($n_H = 2.6$) is evident (The binding site positions on the upper and lower strands are abbreviated as U and L, respectively). On the lower strand, two regions of DNase I blockage appear around positions L85-82 and L111-104. Interstrand bidentate interactions are accordingly assigned around positions 83-85 and 104-111 (Table 1).

The wide areas of DNase cleavage inhibition around positions U104-115 and L111-104 that encompass 12 base pairs suggest that two peptide molecules bind in dimeric fashion here. Consistent with this, the binding isotherms at individual binding sites exhibit sigmoidal curvature, typical of positive cooperativity (Figure 4).

Reviewing hundreds of footprinting experiments, we have frequently observed an abrupt increase in the intensity of footprints as the peptide concentration is raised slightly. This is a clear indication of significant positive cooperativity between peptide molecules binding to adjacent sub-binding sites. Peptides incorporating the XPRK motif interact preferentially with d(AAAA)-d(TTTT) or d(AAA)-d(TTT) sites.

Footprinting studies that indicate a wide binding locus spanning over 12 base-pairs can really only be explained on the basis of closely adjacent sites that might act independently, but are most likely to accommodate two peptide molecules in a dimeric binding mode. A dimeric binding pattern is consistent with independent circular dichroism studies. Thus, in this case the apparent large single binding site appears to be composed of two adjacent sub-binding sites on each complementary strand, representing a total of four sub-sites at each locus (Figure 5).

The conformations of two adjacent sub-binding sites affect each other the most since they are linked covalently. Bonds linking opposite sub-sites are interstrand bidentate hydrogen bonding and exert less conformational influence upon one another than do those linking adjacent sub-sites. Conformational influences between neighboring binding sites also depend on the intervening distance between them. Recent footprinting experiments show that the n_H values for peptides PyHyp-12 and PyHyp-9 binding to 81-mer duplexes (S-81) containing a single d(AAAA)-d(TTTT) binding locus are significantly higher than those of 158-mer and 135-mer duplexes containing two binding loci d(AAAA)-d(TTTT) and d(AAA)-d(TTT) [6]. This was a first indication that the allosteric relay of peptide binding information between neighboring d(AAA)-d(TTT) and d(AAAA)-d(TTTT) positions spanning an intervening sequence of 12-14 base pairs on the same DNA strand might be affected by some negative cooperative effect.

Indeed, it appears that allosteric influences affecting peptide/DNA binding sites may produce either positively or negatively cooperative binding. Most of the allosteric effects found in our studies lead to positive cooperativity.

Table 1. Binding specificity, physicochemical parameters and type of allosteric interaction network for recognition of designed peptides on complementary 5'-[^{32}P]-labeled upper (158-mer) and 5'-[^{32}P]-labeled lower (135-mer) DNA strands at 37°C determined by quantitative DNase I footprinting

Ligand	Binding Position	Recognition Sequence	K_a	n_H	Position of Interstrand Bidentate interactions	Type of Interaction Network
RHyp-12	U84-90 U106-113 L110-103	5'-AATACCG-3' 5'-AGAAAATA-3' 5'-TTTCTCCT-3'	3.0×10^6 7.6×10^6 7.4×10^6	1.5 2.1 2.0	106-110	Partial Circuit
PyHK-10	U86-89 U105-115 L111-105	5'-TACC-3' 5'-GAGAAAATACC-3' 5'-TTTTCTC-3'	4.9×10^6 2.7×10^6 3.4×10^6	0.8 2.4 1.4	105-110	Partial Circuit
HyH-10	U83-91 U104-115 L85-82 L111-104	5'-AAATACCGC-3' 5'-GGAGAAAATACC-3' 5'-TTTC-3' 5'-TTTTCTCC-3'	1.8×10^6 4.7×10^6 1.4×10^6 3.6×10^6	2.6 1.7 1.3 2.0	83-85 104-111	Circuit
PyHR-9	L85-79 L112-104	5'-TTTCACA-3' 5'-ATTTTCTCC-3'	8.4×10^6 9.7×10^6	5.0 10.6	NIL	Non-circuit
PyHyp-9	U84-89 U107-113 L84-83 L111-103	5'-AATACC-3' 5'-GAAAATA-3' 5'-TT-3' 5'-TTTTCTCCT-3'	2.5×10^6 3.8×10^6 2.7×10^6 2.2×10^6	2.2 2.2 2.0 2.0	84 107-111	Circuit
PyHyp-12	U84-88 U106-114 L110-103	5'-AATAC-3' 5'-AGAAAATAC-3' 5'-TTTCTCCT-3'	2.5×10^6 3.4×10^6 3.4×10^6	1.5 1.4 1.4	106-110	Partial Circuit
PyPro-12	U84-90 U106-116 L85-83 L110-105	5'-AATACCG-3' 5'-AGAAAATACCG-3' 5'-TTT-3' 5'-TTTCTCC-3'	6.7×10^7 3.2×10^6 3.7×10^6 3.8×10^6	0.7 1.2 1.1 1.5	84-85 106-110	Circuit
HypKK-10	U83-90 U107-116 L110-104 L85-81	5'-AAATACCG-3' 5'-AGAAAATACC-3' 5'-TTTCTCC-3' 5'-TTTCA-3'	2.0×10^6 2.4×10^6 3.3×10^6 1.6×10^6	2.7 2.0 2.3 -	107-110	Partial Circuit

K_a and n_H are the apparent association constant and Hill coefficient determined from concentration-dependent DNase I footprinting studies, respectively. The binding site positions on the upper and lower strands are abbreviated as U and L, respectively. Interstrand bidentate interactions are assigned where there are coincident effects on complementary nucleotides in both strands.
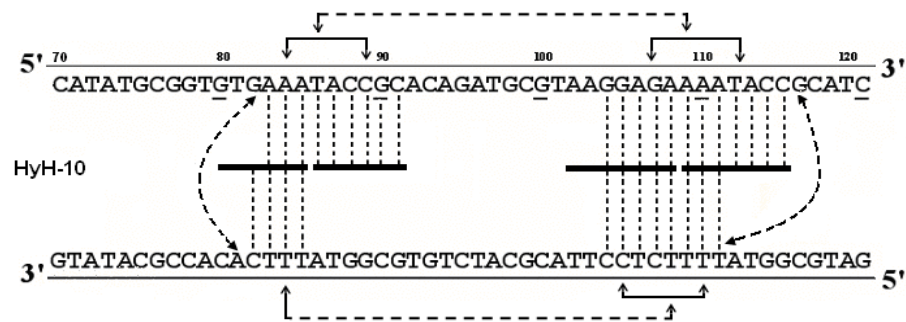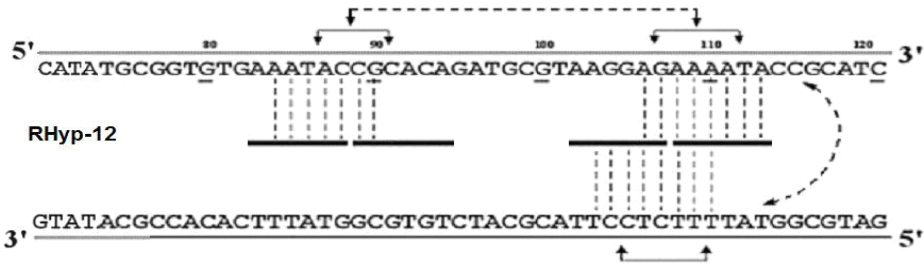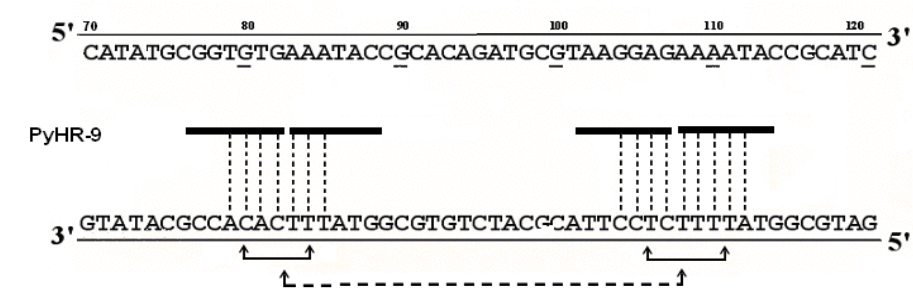
Figure 5. Three different types of allosteric interaction network models for cooperative binding of peptides to DNA fragments based on quantitative footprinting studies. The portion of the ligand binding to each DNA site/sub-site is represented by a thick horizontal line. Monodentate interactions and interstrand bidentate interactions are represented by vertical broken lines. The solid horizontal arrow lines represent communication of allosteric interaction between DNA sub-binding sites. Broken horizontal arrow lines between neighboring binding loci some 12-16 nucleotides apart are intended to represent weak negative cooperative communication.

In these cases, binding of a peptide molecule to a given sub-site generates local DNA conformational changes that enhance the binding affinity of another peptide molecule at an adjacent sub-site. In the negatively cooperative case outlined above, binding of two peptide molecules to a first site produces local DNA conformation changes that decrease the binding affinity of one or two peptide molecules at a different site about 12-18 base-pairs away.

We have also found that interstrand bidentate hydrogen bond interactions may relay conformational changes from one site or sub-site to the opposite site in the complementary strand, and may bestow negative cooperativity in this manner.

Two peptide molecules could be capable of binding in a dimeric head-to-head or head-to-tail fashion to some or all of these sites (Figure 5). Several modes of interaction with complementary or adjacent binding sites can be envisaged that might lead to cooperative binding. Interaction of one peptide molecule in a sequential manner with one sub-site (inside a binding site) on one DNA strand could facilitate the binding of a second molecule to the other sub-site on the same strand. On the other hand, binding of a peptide molecule to a first site might induce local DNA conformational change(s) mediated by interstrand bidentate interactions and facilitate sequential or concerted binding of further peptide molecules to an opposite site on the complementary strand. Of course, binding of peptide molecules to one site could invoke conformational change(s) and facilitate or hinder peptide binding to a neighboring site on the same strand. Thus, cooperative (or anti-cooperative) binding distributed over an array of binding sites may establish a network of cooperativity interconnecting several allosteric sites (Figure 5).

To gain direct insight into possible conformational changes of DNA associated with the sequence-selective binding of peptides we carried out circular dichroism measurements (CD) using a number of synthetic peptides (the amino acid sequences of some of these peptides are shown in Figure 8). As an example we present CD results for a decapeptide, HyH-10: His-Hyp-Arg-Lys-(Py)$_4$-Lys-Arg-NH$_2$ (HyH-10) [7]. A 13-mer deoxyribonucleotide duplex d(TAGGAGAAAATAC)- d(GTATTTTCTCCTA) (U4A-L4T) was synthesized to contain a sequence of 12 base-pairs that correspond to the binding site at position 103-114 of the pBR322 fragment. The CD spectra for the DNA and peptides alone are shown in Figure 6A. Titration of peptide HyH-10 versus the duplex U4A-L4T induces a negative CD band around 247 nm and a positive band around 284 nm. A strong, broad, dose-dependent positive CD enhancement appears around 330 nm (Figure 6B, C).

In the difference spectrum (Figure 6C), this CD band around 330 nm is unchanged. Another induced positive CD band is red-shifted from 284 nm to 288 nm. It has been reported that molecules which bind to the minor groove typically exhibit an induced positive CD band around 320 nm [37]. Accordingly the spectra indicate strongly that our designed peptides incorporating the XPRK/XHyPRK and polyamide motifs bind preferentially to the minor groove of the double helix.

It is evident that the peptide-induced enhancement of CD bands is correlated with peptide-DNA binding stoichiometry. Figure 7 shows a plot of [peptide]/[DNA duplex] versus Δθ at 322 nm for four synthetic peptides: RHyp-12, PyHK-10, HyH-10 and PyHR-9 [7]. The decapeptide PyHK-10 induces the greatest ellipticity change among the four peptides, suggesting that most far-reaching DNA conformational changes are associated with the binding of this peptide in the minor groove. On the other hand, significantly greater ellipticity changes are induced by peptides PyHK-10, RHyp-12 and HyH-10 compared to PyHR-9,
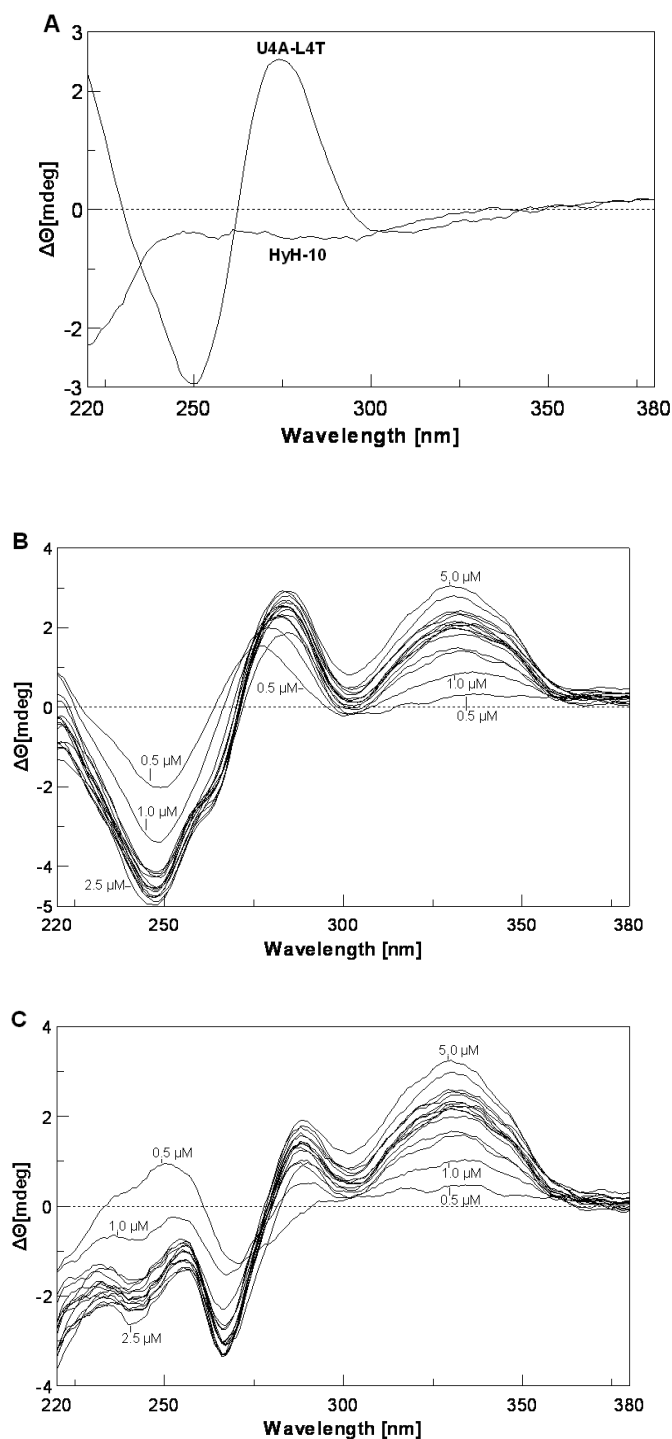
Figure 6. Panel A: CD spectra of DNA duplex d(TAGGAGAAAATAC)- d(GTATTTTCTCCTA) (U4A-L4T) alone and peptide alone. Panel B: Titration of duplex U4A-L4T versus peptide HyH-10 at peptide concentrations of 0.2, 1.0. 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 4.0, 5.0 μM. Panel C: Corresponding CD difference spectra with the contribution of duplex and peptide HyH-10 subtracted.

indicating that the nonapeptide is much less able to perturb the conformation of the helical minor groove. The ellipticity changes induced by peptides RHyp-12 and HyH-10 are of comparable magnitude.

From the $\Delta\theta$ *versus* [peptide]/[DNA duplex] plot (Figure 7), it appears that at peptide/DNA ratios of about 0.5 to 2.0, one molecule of peptide binds to the d(AAAA)-d(TTTT) locus. At peptide/DNA ratios of 2 - 3, as indicated by the small plateau region in the titration curve, two peptide molecules seem to bind ostensibly as a dimer in the minor groove. At ratios above 4, a progressive increase in $\Delta\theta$ suggests that more peptide molecules begin to engage in some sort of non-sequence-selective binding to DNA. The CD results agree well with extensive DNase I footprinting results that dimeric peptide binding to the d(AAAA)-d(TTTT) locus is favored, as shown by the very wide binding locus spanning a distance of 9-12 base pairs (Figures 3,5, Table 1).
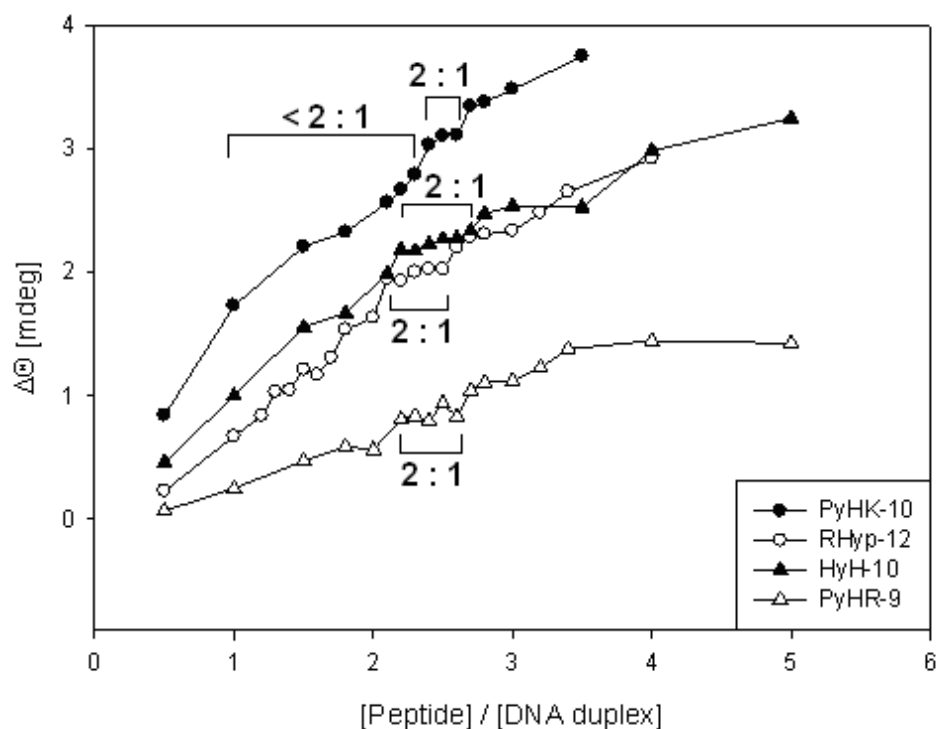
Figure 7. CD intensity at 322 nm as a function of [peptide]/[DNA duplex] The proposed stoichiometric binding ratios are as indicated, with binding below 2:1 (<2:1) considered to be predominantly 1:1.

Isothermal titration calorimetry (ITC) studies of several designed peptides incorporating the XPRK motif revealed that the binding reaction is strongly exothermic [58]. Preliminary surface plasmon resonance (SPR) studies with an oligonucleotide hairpin have shown that two peptides bind to the d(AAAA)-d(TTTT) sequence with $t_{1/2}$ values around 2-2.5 min (depending on the initial concentration of the peptide) and with k_{a1} and k_{a2} around 180-300 $M^{-1}s^{-1}$, suggesting that the binding of these designed peptides to their preferred DNA binding sites is a very slow process compared to other small molecules [58].

5. DETERMINATION OF DNA-PEPTIDE BINDING COOPERATIVITY BY THE HILL PLOT

The apparent DNA binding site saturation is determined by the equation [19].

$$Y' = 1 - (I_{tot}/I_{ref})/(I_{tot}^{o}/I_{ref}^{o})$$

where Y' is the fractional saturation, I_{ref} and I_{ref}^{o} are integrated band volumes over 5 bases of a running lane (non-binding site, with peptide) and control lane (without peptide), respectively. I_{tot} and I_{tot}^{o} are integrated band volumes of the binding site locus and corresponding control lane locus, respectively.

The value of Y is optimized by the following equation:

$$Y = (Y' - Y'_{min})/(Y'_{max} - Y'_{min})$$

where Y'_{max} and Y'_{min} are the maximum and minimum site saturation Y' values, respectively.

The Hill coefficients which indicate the magnitude of cooperativity between binding sub-sites are determined by the Hill equation:

$\log [Y/1-Y] = \log K_a + n_H \log[L]$ where Y is the fractional saturation, [L] is the peptide concentration, and n_H the Hill coefficient. A minimum of six data points within or near the linear portion of the binding isotherm (Y versus log [L]) were chosen for the Hill plot (log Y/(1-Y) versus log [L]) using a linear least-squares fitting procedure. The Hill coefficient was determined from the slope of the corresponding Hill plot (see Figure 4). The apparent binding constant K_a is determined empirically from the binding isotherm as the peptide concentration at 50% fractional saturation.

6. DNA-PEPTIDE ALLOSTERIC INTERACTION NETWORKS

Current studies of biomacromolecule-ligand allostery have progressed to the so-called "interaction network" stage. Allostery appears to interconnect different metabolic and biosynthetic pathways through interaction of various bio-molecules: namely, protein-protein interactions [59,60], protein-small ligand interactions [61], protein-nucleic acid interactions [62], and nucleic acid-small ligand interactions. In addition, many of these complex network interactions may communicate with one another, resulting in significant net biochemical and physiological phenomena. In other words, there are few biochemical processes within the cell that are independent of allosteric or regulatory influences from other metabolic pathways or from other bio-molecules. The past decade has been remarkable for an explosion of interest in network-based interactions between bio-molecules that effectively regulate biochemical and physiological processes in nature. High-throughput studies have contributed much to establishing protein-protein interaction networks, metabolic networks, and transcription regulatory networks [59].

To learn more about the chemical rules that govern DNA-ligand allostery we employed quantitative footprinting to embark upon a study of molecular aspects of cooperative interactions between peptides and multiple DNA binding sites. Thus, in 2006 we proposed an original network-based DNA-peptide allosteric model featuring intercommunicating allosteric binding sites in fragments of the latent membrane protein-1 gene from a pathogenic Epstein-Barr virus variant derived from nasopharyngeal carcinoma [23]. In recent studies we have further explored whether such network-based allosteric behavior can be detected in common DNA motifs containing consecutive A's and T's using our 158-mer and 135-mer pBR322 fragments [6,7].

His-Hyp-Arg-Lys-(Py)$_4$-Lys-Arg-Hyp-His-NH$_2$ (RHyp-12)

His-Pro-Arg-Lys-(Py)$_4$-Lys-Arg-NH$_2$ (PyHK-10)

His-Hyp-Arg-Lys-(Py)$_4$-Lys-Arg-NH$_2$ (HyH-10)

His-Pro-Arg-Lys-(Py)$_4$-Arg-NH$_2$ (PyHR-9)

His-Hyp-Arg-Lys-(Py)$_3$-His-Hyp-Lys-Arg-NH$_2$ (PyHyp-9)

His-Hyp-Arg-Lys-(Py)$_4$-His-Hyp-Arg-Lys-NH$_2$ (PyHyp-12)

His-Pro-Arg-Lys-(Py)$_4$-His-Pro-Arg-Lys-NH$_2$ (PyPro-12)

His-Hyp-Arg-Lys-(Py)$_4$-Lys-Lys-NH$_2$ (HypKK-10)

Figure 8. Sequences of eight designed peptides used in quantitative DNase 1 footprinting studies (Py = 4-amino-1-methylpyrrole-2-carbonyl-).

Based on quantitative footprinting results with eight designed peptides (Figure 8) on the complementary upper and lower DNA strands summarized in Table 1, we constructed network-based models in an effort to interpret the complex communication between DNA-peptide allosteric binding sites (Figure 5). The duplex fragments used here contain two conspicuous peptide binding loci: around position 83-89, comprising the d(AAA)-d(TTT) sequence, and position 103-116 comprising the d(AAAA)-d(TTTT) sequence. Each locus seems to afford two sites on complementary strands that consist of two to four sub-sites, depending on individual peptides. In our model, conformational changes induced at one binding site affect subsequent peptide binding to a neighboring site, and interstrand bidentate interactions are envisaged as playing an important role in relaying conformational changes between sites on the complementary strands, facilitating or inhibiting the binding of further peptide molecules to sub-sites. The different models for these eight peptides binding to DNA primarily reflect differences in the relay of conformational information between opposite sites on complementary strands, mediated via interstrand bidentate interactions. It can be concluded that peptides HyH-10, PyHyp-9 and PyPro-12 form a circuit type of a communication: that is to say, the allosteric communication between binding sites is complete. For peptides RHyp-12 , PyHK-10, PyHyp-12 and HypKK-10, the networks are referred to as incomplete-circuit type, because the allosteric communications form a partial or incomplete circuit. At the other extreme, synthetic peptides PyHR-9, (HPRK)$_3$NH$_2$ [HR-12] [23], (SPRK)$_3$NH$_2$ [SP-12] [23] whose binding to DNA fragments lacks interstrand bidentate interactions altogether, appear to engage in a non-circuit type of allosteric communication.

To summarize, the quantitative footprinting results in these studies support our hypothesis that three different types of network-based allosteric communication can be distinguished in peptide-DNA recognition: circuit type, incomplete-circuit type and non-circuit type [6,7]. The finding that interaction networks of allosteric effects exist in DNA-peptide molecular recognition serves to widen and enlarge our understanding of the fundamental nature of macromolecule-small ligand interactions. Provided that the intervening distance between binding sites lies within about 15 bp, conformational changes can readily be transmitted from one site to another. If more than two sites are sufficiently close, one can envisage that allosteric effects could be propagated over a substantial distance along the DNA duplex. This network concept provides some basic chemical rules of DNA-small ligand allosteric interaction that should find applications in future drug design and structural biology research.

ACKNOWLEDGMENTS

REFERENCES

[1]   Seeman, N.C. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* 1976, *73,* 804-808.

[2]   Suzuki, M. A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure* 1994, *2,* 317-326.

[3]   Mandel-Gutfreund, Y.; Schueler, O.; Margalit, H. Comprehensive analysis of hydrogen bonds in regulatory protein–DNA complexes: in search of common principles. *J. Mol. Biol.* 1995, *253,* 370-382 and references cited therein.

[4]   Luscombe, N.M.; Laskowski, R.A.; Thornton, J.M. Amino cid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* 2001, *29,* 2860-2869 and references cited therein. .

[5]   Cheng, A.C.; Frankel, A.D. Ab initio interaction energies of hydrogen-bonded amino acid side chain-nucleic acid base interactions. *J. Am. Chem. Soc.* 2004, *126,* 434-435 and references cited therein.

[6]   Kao, K.L.; Jonathan C.T. Huang; J.C.T.; Yang, C.K.; Jeng, K.C.G.; Chang, J.C.C.; Yao, W.C.; Hsien, S.C.; Waring, M.J.; Chen, M.H.; Ma, L.; Sheh, L. Detection of Multiple Network-based Allosteric Interactions between Peptides and Arrays of DNA Binding Sites. *Bioorg. Med. Chem.* 2010, 18, 366-376.

[7]   Huang, T.B.; Chen, Y.C.; Chang, J.C.; Jeng, K.C.; Kao, K.K.L.; Yang, R.C.K., Kan, L.S.; Wey, M.T.; Waring, M.J.; Chen, C.S.; Chien, W.J.; Sheh, L. Novel DNA-peptide interaction networks. *Bioorg. Med. Chem.* 2010, 18, 2575-2585.

[8]   Fox, K.R.; Waring, M.J. Meth. Enzymol. Use of DNA molecules substituted with unnatural nucleotides to probe specific drug-DNA interactions. 2001, *340,* 412-430.

[9]   Moravek, Z.; Neidle, S.; Schneider, B. Protein and drug interactions in the minor groove of DNA. *Nucleic Acids Res.,* 2002, *30,* 1182-1191.

[10]   Bailly, C.; Chaires, J.B. Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconjugate Chem.,* 1998, *9,* 513-538 and references cited therein.

[11]   Waring, M.J. (1979) Echinomycin, triostin and related antibiotics. In *Antibiotics,* Hahn, F.E. Eds. Springer-Verlag. Berlin, vol. 5, pp. 173-194.

[12]   Chaires, J.B.; Fox, K.R.; Herrera, J.E.; Britt, M.; Waring, M.J. Site and sequence specificity of the daunomycin-DNA interaction. *Biochemistry*, 1987, *26,* 8227-8236.

[13]   Boger, D.L.; Chen, J.H.; Saionz, K.W. (-)-Sandramycin: Total synthesis and characterization of DNA binding properties. *J. Am. Chem. Soc.* 1996, *118,* 1629-1644.

[14]   Bailly, C.; Suh, D.; Waring, M.J.; Chaires, J.B. Binding of daunomycin to diaminopurine and/or inosine-substituted DNA. *Biochemistry* 1998, *37,* 1033-1045.

[15]   Chen, Y.H.; Lown, J.W. A new DNA minor groove binding motif: cross-linked lexitropsins. *J. Am. Chem. Soc.* 1994, *116,* 6995-7005.

[16]   Walker, W.L.; Landaw, E.M.; Dickerson, R.E.; Goodsell, D.S. *Proc. Natl. Acad. Sci. U.S.A.* 1997, *94,* 5634-5639.

[17]   Satz, A.L.; Bruice, T.C. Recognition in the minor groove of double-stranded DNA by microgonotropens. *Acc. Chem. Res.* 2002, *35,* 86-95.

[18]   Bailly, C.; O'hUigin, C.; Houssin, R.; Colson, P.; Houssier, C.; Rivalle, C.; Bisagni, E.; Hénichart, J.P.; Waring, M.J. DNA binding properties of a distamycin-ellipticine hybrid molecule. *Mol. Pharmacol..* 1992, *41,* 845-855.

[19]   Mrksich, M.; Parks, M.E.; Dervan, P.B. Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* 1994, *116,* 7983-7988; Weyermann, P.; Dervan, P.B. Recognition of ten base pairs of DNA by head-to-head hairpin dimers. *J. Am. Chem. Soc.* 2002, *124,* 6872-6878 and references cited therein.

[20]   Yang, C.-H.; Chou, P.-J.; Luo, Z.-L.; Chou, I.-C.; Chang, J.-C.; Cheng, C.-C.; Martin, C.R.H.; Waring, M.J.; Sheh, L. Preferential binding to DNA sequences of peptides related to a novel XPRK motif. *Bioorg. Med. Chem*. 2003, *11,* 3279-3288.

[21]   Chang, J.-C.; Yang, C.-H.; Chou, P.-J.; Yang, W.-H.; Chou, I.-C.; Lu, C.-T.; Lin, P.-H.; Hou, R.C.-W.; Jeng, K.-C.G.; Cheng, C.-C.; Sheh, L. DNA sequence-specific recognition of peptides incorporating the HPRK and ployamide motifs. *Bioorg. Med. Chem*. 2004, *12,* 53-61.

[22]   Yang, C.-H.; Chen, W.F.; Jong, M.-C.; Jong, B.-J.; Chang, J.-C.; Waring, M.J.; Ma, L.; Sheh, L. Semiquinone footprinting. *J. Am. Chem. Soc.* 2004, *126,* 8104-8105.

[23]   Yang, C.-H.; Jeng, K.C.G.; Yang, W.H.; Chen, Y.L.; Hung, C.C.; Lin, J.W.; Chen, S.T.; Richardson, S.; Martin, C.R.H.; Waring, M.J.; Sheh, L. Unusually strong positive cooperativity in binding of peptides to latent membrane protein-1 DNA fragments of the Epstein-Barr viral gene. *ChemBioChem* 2006, *7,* 1187-1196.

[24]   Chen, Y.C.; Huang, J.T.B.; Jeng, K.C.G.; Yang, R.C.K.; Liao, M.K.; Chen, C.S.; Chien, W.J.; Wey, M.T.; Kan, L.S.; Sheh, L. Determination of Allosteric Effects and Interstrand Bidentate Interactions in DNA-peptide Molecular Recognition *J. Chin. Chem. Soc.* 2010, *57,* 266-274.

[25] Waring, M.J.; Wakelin, L.P.G. Echinomycin: a bifunctional intercalating antibiotic. *Nature* 1974, *252,* 653-657.

[26] Wakelin, L.P.G.; Waring, M.J. The binding of echinomycin to DNA. *Biochem. J.* 1976, *157,* 721-740.

[27] Waring, M.J. in molecular aspects of anticancer drug-DNA interactions (Neidle, S. and Waring, M.J. eds) 1993, Vol. 1, 213-242, MacMillan, London and references quoted therein.

[28] Kong, D.; Park, E.J.; Stephen, A.G.; Calvani, M.; Cardellina, J.H.; Monks, A.; Fisher, R.J.; Shoemaker, R.H.; Melillo, G. Echinomycin, a small-molecule inhibitor of hypoxia-inducible factor-1 DNA-binding activity. *Cancer Res.* 2005, *65,* 9047-9055.

[29] Abrescia, N.G.A.; Malinina, L.; Subirana, J.A. Stacking interaction of guanine with netropsin in the minor groove of d(CGTATATACG)$_2$. *J. Mol. Biol.* 1999, *294,* 657-666.

[30] Bailly, C.; Colson, P.; Houssier, C., Houssin, R., Mrani, D.; Gosselin, G.; Imbach, J.L.; Waring, M.J.; Lown, J.W.; Hénichart, J.P. Binding properties and DNA sequence-specific recognition of two bithiazole-linked netropsin hybrid molecules. *Biochemistry* 1992, *31,* 8349-8362.

[31] Chen, X.; Ramakrishnan, T.R.; RAo, S.T.; Sundaralingam, M. Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nature Struct. Biol.* 1994, *1,* 169-175.

[32] Dervan, P.B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* 2001, *9,* 2215-2235 and references quoted therein

[33] Breslauer, K.J.; Remeta, D.P.; Chou, W.Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J.G.; Marky, L.A. Enthalpy-entropy compensations in drug-DNA binding studies. *Proc. Natl. Acad. Sci. U.S.A.* 1987, *84,* 8922-8926.

[34] Hutchins, R.A.; Crenshaw, J.M.; Graves, D.E.; Denny, W.A. Influence of substituent modifications on DNA binding energetics of acridine-based anticancer agents. *Biochemistry* 2003, *42,* 13754-13761.

[35] Chaires, J.B.; Satyanarayana, S.; Suh, D.; Fokt, I.; Przewloka, T.; Priebe, W. Parsing the free energy of anthracycline antibiotic binding to DNA. *Biochemistry* 1996, *35,* 2047-2053.

[36] Qu, X.; Ren, J.; Riccelli, P.V.; Benight, A.S.; Chaires, J.B. Enthalpy/entropy compensation; influence of DNA flanking sequence on the binding of 7-amino actinomycin D to its primary binding site in short DNA duplexes. *Biochemistry* 2003, *42,* 11960-11967.

[37] Munde, J.; Lee. M.; Neidle, S.; Arafa, R.; Boykin, D.W.; Liu, Y.; Bailly, C.; Wilson, W.D. Induced fit conformational changes of a reversed amidine heterocycle: optimized interactions In a DNA minor groove complex. *J. Am. Chem. Soc.* 2007, *129,* 5688-5698 references cited therein.

[38] Davanloo, P.; Crothers, D.M. Kinetic studies of drug-dinucleotide complexes. *Biochemistry* 1976, *15,* 5299-5305; Davanloo, P.; Crothers, D.M. Phase partition studies of actinomycin-nucleotide complexes. *Biochemistry* 1976, *15,* 4433-4438.

[39] Fox, K.R.; Waring, M.J. Stopped-flow kinetic studies on the interaction between echinomycin and DNA. *Biochemistry* 1994, *23,* 2627-2633.

[40] Chaires, J.B.; Dattagupta, N.; Crothers, D.M. Kinetics of daunomycin-DNA interactions. *Biochemistry* 1985, *24,* 260-267.

[41] Baliga, R.; Baird, E.E.; Herman, D.M.; Melander, C.; Dervan. P.B.; Crothers, D.M. Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry* 2001, *40,* 3-8.

[42] Fletcher, M.C.; Fox, K.R. Dissociation kinetics of echinomycin from CpG binding sites in different sequence environments. *Biochemistry* 1996, *35,* 1064-1075.

[43] Churchill, M.E.A.; Suzuki, M. SPKK motifs prefer to bind to DNA at A/T –rich sites. *EMBO J*. 1989, *8,* 4189-4195.

[44] Suzuki, M. The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA. *Nature* 1990, *344,* 562-565.

[45] Monod, J.; Wyman, J.; Changeux, J.P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 1965, *12,* 88-118.

[46] Koshland, D.E. Jr. and Hamadani, K. Proteomics and models for enzyme cooperativity. *J. Biol. Chem.* 2002, *277,* 46481-46844 and references cited therein.

[47] Shiozawa, H.; Chia, B.C.S.; Davies, N.L.; Zerella, R.; Williams, D.H. Cooperative binding interactions of glycopeptide antibiotics. *J. Am. Chem. Soc.* 2002, *124,* 3914-3919.

[48] Williams, D.H.; Davies, N.L.; Zerella, R.; Bardsley, B. Noncovalent interactions: defining cooperativity. Ligand binding aided by reduced dynamic behavior of receptors. Binding of bacterial cell wall analogues to Ristocentin A. *J. Am. Chem. Soc.* 2004, *126,* 2042-2049.

[49] Dattagupta, N.; Hogan, M.; Crothers, D.M. Interaction of netropsin and distamycin with deoxyribonucleic acid: electric dichroism study. *Biochemistry* 1980, *19,* 5998-6005.

[50] Graves, D.E.; Krugh, T.R. Adriamycin and daunorubicin bind in a cooperative manner to deoxyribonucleic acid. *Biochemistry* 1983, *22,* 3941-3947.

[51] Fagan, P.; Wemmer, D.E. Cooperative binding of distamycin-A to DNA in the 2:1 mode. *J. Am. Chem. Soc.* 1992, *114,* 1080-1081.

[52] Bailly, C.; Hamy, F.; Waring, M.J. Cooperativity in the binding of echinomycin to DNA fragments containing closely spaced CpG sites. *Biochemistry* 1996, *35,* 1150-1161.

[53] Morii, T.; Yamane, J.; Aizawa, Y.; Makino, K.; Sugiura, Y. Cooperative oligomerization enhances sequence-selective DNA binding by a short peptide. *J. Am. Chem. Soc.* 1996, *118,* 10011-10017.

[54] Qu, S.; Trent, J.O.; Fokt, I.; Priebe, W.; Chaires, J.B. Allosteric, chiral-selective drug binding to DNA. *Proc. Natl. Acad. Sci. U.S.A.* 2000, *97,* 12032-12037.

[55] Harris, S.A.; Gavathiotis, E.; Searle, M.S.; Orozco, M.; Laughton, C.A. Cooperativity in drug-DNA recognition: a molecular dynamics study. *J. Am. Chem. Soc.* 2001, *123,* 12658-12663.

[56] Cooper. A.; Dryden, D.T.F. Allostery without conformational change. A plausible model. *Eur. Biophys. J.*, 1984, *11,* 103-109.

[57] Fechter, E.J.; Dervan, P.B. Allosteric inhibition of protein-DNA complexes by polyamide-intercalator conjugates. *J. Am. Chem. Soc.* 2003, *125,* 8476-8485.

[58] Yang, R.C.K. et al. paper to be submitted.

[59] Russell, R.B. and Aloy, P. Targeting and tinkering with interaction networks. *Nat. Chem. Biol.* 2008, *4,* 666-673 and references cited therein.

[60] Yu, H. et al. High-quality binary protein interaction map of the yeast interactome network. *Science,* 2008, *322,* 104-110.

[61] Vetter, D. Chemical microarrays, fragment diversity, label-free imaging by plasmon resonance-a chemical genomics approach. *J. Cell. Biochem. Suppl.* 2002, *39,* 79-84.

[62] Berger, M.F.; Bulyk, M.L. protein binding microarrays for rapid, high throughput characterization of the sequence-specificities of DNA binding proteins. *Methods Mol. Biol.* 2006, *338,* 245-260.

Chapter 7

MOLECULAR RECOGNITION OF ODORANT-BINDING PROTEINS IN INSECT OLFACTION

Xin Jin[1] and Long Zhang[2]

[1] Center for Molecular Recognition,
College of Physicians and Surgeons,
Columbia University, New York, NY 10032, USA
[2] Department of Entomology, China Agricultural University,
Beijing, 100193, China

ABSTRACT

Like other animals, insects sense chemicals evoke their behaviors, such as aggregation, mating, feeding and migration behaviors, through their olfactory and taste recognition systems. The olfactory molecules, in the air out of insect body, diffuse into sensillum lymph through the cuticle pores and at first step interact with the small soluble odorant-binding proteins (OBPs). The odor specificity of a given olfactory neuron is determined by expressed olfactory receptor (OR) genes along with other accessory proteins. The signals accepted by ORs then are sent to higher processing centers in the brain to elicit distinct behavioral outputs.

So far over 100 OBPs have been identified from more than 40 species of insects. However, only a few have been studied for their recognition with ligands during binding activity. Although OBPs certainly show binding ability to hydrophobic odorants, their physiological roles are still in controversy. There are three unclear points which people are very interested in. Firstly, whether or not insect OBPs have binding specificity and how do they perform such binding specificity? Secondly, how do the proteins recognize the odorants, and which binding sites are involved in recognition? Thirdly, how do OBPs interact with odorant receptors? To reveal mechanisms of OBPs molecular recognition is helpful in understanding their physiological functions and in designing interfering

---

[1] xj2123@columbia.edu
[2] locust@cau.edu.cn

molecules to ruin normal recognition and control insect pests. Therefore this field is attracting more and more scientists' passions and interests.

Recently some interesting experiments on insect OBPs had been conducted to partly address these questions with the combination of three dimensional structure analysis, binding experiments, site-directed mutagenesis, and simulation and docking experiments. Here we discuss insect OBPs molecules recognition and review the progress in the field. It has been found that the conformational change of OBPs can be induced either by pH value alteration or by ligand binding. Some hydrophilic amino acids at the entrance of OBPs binding pocket, as well as the hydrophobic amino acids in the pocket, are involved in ligand binding and may contribute to the recognition specificity of OBPs.

INTRODUCTION

The first insect odorant-binding protein (OBP) was identified and named pheromone binding protein (PBP) in moth *Antheraea polyphemus* in 1981, using tritium labeled pheromone (E,Z)-6,11-hexadecadienyl acetate[1]. *A. polyphemus* PBP is concentrated in sensillum lymph at around 10mM[2] and is 142 amino acids in length[3]. In the following decades, a great number of small soluble proteins with similar expression pattern and amino acid sequence to PBP of *A. polyphemus* have been identified in more than 40 insect species from 8 different orders. These proteins are subdivided as PBP, general odorant binding protein (GOBP), antennal binding protein X (ABPX), Plus-C OBP and Atypical OBP. Most of them share some common signatures, such as N-terminus signal peptide, specific expression pattern in chemosensory organs and six position-conserved cysteines (except for Plus-C and Atypical OBP, which possess two or more additional cysteines in sequence)(see review [4]). Furthermore, the large multigenic families of OBP-like proteins have been identified in some dipteran species (review in [5]), honeybee[6], silkworm[7] and pea aphid [8] by genome analysis.

The highly concentrated expression of OBPs in olfactory sensillum lymph and binding assays[4] gave people enough reasons to postulate the physiological roles of OBPs in peri-receptor events of olfactory recognition, which are still in controversy. OBPs have been suggested to act as carriers to transport odors to dendrite membrane, to deactivate stimuli, even to be a phenomenal signal trigger[4, 9]. However, direct evidences of the involvement of OBPs in insect olfaction were lacking until the late 90s of 20th century. Two well-known reports in fire ant and fruit fly respectively, showed that OBPs are required in insect olfactory behaviors.

The general protein-9 (Gp-9) gene, which encodes an OBP in fire ant *Solenopsis invicta*, regulates the polymorphism in colony social organization[10]. The colonies with a single egg-laying queen (monogyne form) include only workers bearing a class of Gp-9 alleles (B-like alleles), whereas colonies containing multiple queens (polygyne form) always additionally include workers bearing an alternate class of alleles (b-like alleles). The results implicated that the B-like and b-like protein variants may have different binding properties to the ligands, the pheromones elicited by queens. Therefore, the workers of different Gp-9

genotypes exhibit different queen recognition capabilities, leading to different colony-level phenotypes[10-12].

Drosophila melanogaster OBP LUSH is another example[13]. *lush* mutant is the first reported OBP mutant. The wild-type flies have mechanisms to avoid high concentrations of ethanol, propanol and butanol, while *lush* mutants are defective to avoid high-alcohols. Instead of deactivating or desensitizing alcohol-sensitive neurons, the authors argued that LUSH is necessary to activate neurons that mediate such avoidance mechanism[13, 14].

OBPs accumulate in the sensillum lymph, which is a hydrophilic barrier between the hydrophobic odorants and the olfactory neurons. Most likely OBPs have some kinds of interactions with odorants to process the olfactory signals. People are very curious about the molecular basis of OBPs in sensillum lymph. They want to answer whether OBPs possess ligand binding capacity, whether OBPs can recognize and discriminate different odors, and whether OBPs are involved in signal triggering?

THREE DIMENSIONAL STRUCTURES AND BINDING SITES OF OBPS

A lot of tools have been used to figure out the roles of OBPs in olfactory signal circuits. Ligand binding assay, behavior assay and specific cellular expression data provided some clues on this, but none of them can give direct evidence like structural biology does.
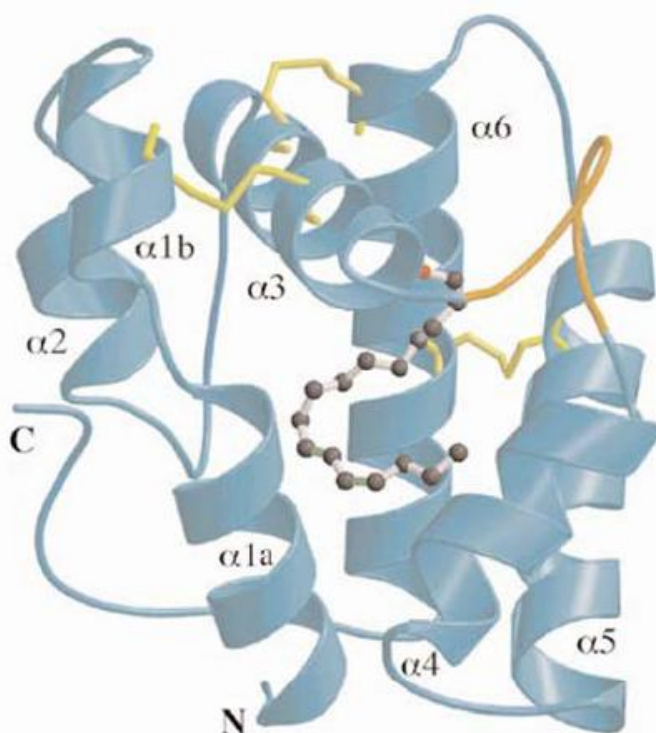
Figure 1. Three dimensional crystal structure of the *B. mori* PBP. The six α helices are stabilized by three disulfide bridges (yellow). Loop 60-69 is shown in orange. Four antiparallel helices (α 1, α 4, α 5 and α 6) form the binding pocket. Bombykol is shown in a ball-and-stick representation with double bonds in green. This figure is provided courtesy of Jon Clardy.

The first identified three dimensional structure of OBP is PBP structure from silkworm *Bombyx mori*[15-17]. The protein is mainly composed of six helices (α1-6) that are tightened together by three disulfide bridges. Two disulfide bridges formed by Cys19-Cys54 and Cys50-Cys108 stabilize helix α3, α1b and α6, meanwhile the third one by Cys97-Cys117 supports the connection between α5 and α6. Four of six helices (α1, α4, α5, α6) are anti-parallele and form a big hydrophobic cavity in the center of the PBP, which holds the moth pheromone bombykol (Figure 1). The identification of 3D structure of *B. mori* PBP stimulated the progress in the structure analysis in OBP field, and several other important OBP structures have been solved in the few years. Those OBPs all share a very similar folding pattern with *B. mori* PBP (see review [18]).

In the binding cavity of *B. mori* PBP, bombykol stays between Phe12 and Phe118, and is connected to Ser56 by hydrogen bond near the cavity entrance[15]. A serendipitous ligand *n*-butyl-benzene-sulfonamide (NBBS) interacts with Phe117 on C-terminus and Ile119 on N-terminus of *Apis mellifera* ASP1 by hydrogen bonds[19]. Within the binding pocket of *Drosophila* LUSH, the alcohol hydroxyl forms hydrogen bonds with Ser52 and Thr57 near the entrance of the pocket[20].

pH-Dependent Conformational Change and Ligand Release in OBPs

In *B. mori* PBP, it is proposed that the movement of loop 60-69 (Figure 1), or the unwinding of helix α1a might shift the cavity between "open" and "close" states, though no conformational change has been detected among the apo-PBP and PBP/bombykol complex[21]. However, a conformational change has been observed with moth PBPs when condition shifts from neutral to acidic pH value. At acidic condition, the C-terminus tail forms an extra helix and folds into the binding cavity, which is proposed to occupy the binding cavity and force the pheromone out of the cavity (Figure 2) [17, 22, 23].
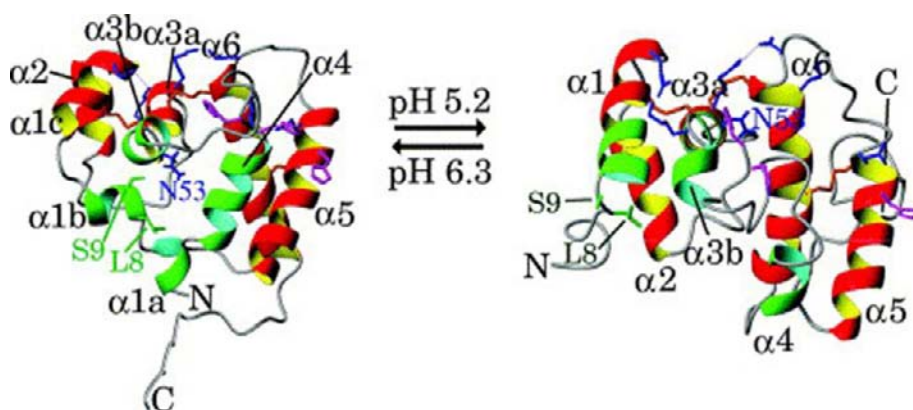
Figure 2. Ribbon view of *A. polyphemus* PBP neutral form (left) and acidic form (right). The side-chain of Asn53 and side-chains of the residues involved in salt-bridges are shown in blue. The side-chains of Leu8 and Ser9 are shown in green and the histidine side-chains of H69, H70, and H95 are shown in purple. The disulfide bonds are indicated in orange. The residues exhibiting large chemical shift changes are highlighted in green. Note that C-terminus tail stays outside in neutral form (left), while folds into the binding pocket in acidic form (right). This figure is provided courtesy of Smita Mohanty.

Such pH-dependent conformational changes have also been found in other insect orders, though they take different strategies since they are shorter in length than moth OBPs. In dipteran species, the C-terminus coil of OBP extends into the protein and covers the binding cavity by contacting with N-terminus. It is suggested that the lower pH condition may disrupt hydrogen bonds and open the cover, resulting in reduction of ligand binding affinity[20, 24, 25]. In honeybee *A. mellifera,* ASP1 is a dimer at pH7.0. The dimerization is disrupted at pH4.0 and the protein undergoes a pH-reduced domain swapping[26]. The length of PBP from cockroach *Leucophaea maderae* is even shorter than that of ASP1, and the authors suggest that it takes a different ligand binding and release mechanism[27, 28].

BINDING PREFERENCE AND INITIAL RECOGNITION OF OBPS

In some species, it has been found that OBPs showed binding preference, although to a relative small extent. Three silkmoth PBP subtypes differentially interact with the three pheromonal compounds of *Antheraea*[29-31]. In cockroach *L. maderae*, only two components of the pheromonal blend, 3-hydroxy-butan-2-one and butane-2,3-diol, showed stronger binding affinity to PBP than fluorescence probe 8-anilino-1-naphtalenesulphonic acid (ANS) by fluorescence binding studies[32]. For *Drosophila* LUSH, a series of large aromatic compounds, including some phthalates, showed better affinity than short chain alcohols, even than some compounds that elicit electrophysiological responses in the fly's antenna[33].

Locusta migratoria OBP1 is proposed to play an important role in locusts to detect semiochemicals or food odors. The highest binding affinity of ligands for *L. migratoria* OBP1 was found using C15 aliphatic alcohol pentadecanol. Aliphatic alcohols having longer carbon chain length (C16 or C17) decreased their binding affinity. So did aliphatic ketones, aliphatic aldehydes and aliphatic esters. These results suggested that chain lengths having 15 carbon atoms with the polar terminal C-O group were molecules of suitable shape and size to bind to *L. migratoria* OBP1. In other word, OBP1 has relative binding specificity to ligands. And we proposed that this binding specificity related with its binding sites. The predicted binding sites of complex of OBP1/pentadecanol model were identified through performing docking experiment with bioinformatics. The docking result showed that all of the top 20 ligand poses placed in similar position, and the polar head group facing the entrance of the protein binding pocket. The docking pose which was the most representative is shown in Figure 3a, b. All the residues with < 4.0 Å distances to pentadecanol are shown in Figure 3c. The pentadecanol neared to Leu50, Met54, Phe66, Glu71, Asn74, Val75, Pro76, Val87, Cys90, Lys91, Ile93, Ala102, Ile105, His106, Tyr109, Ler116, Tyr117, Ser118 and Leu119. Among these nineteen amino acid residues, Val87 was found at the bottom of the pentadecanol binding site and the distance from Asn74 carbonyl oxygen to the hydroxy oxygen of pentadecanol was 3.37 Å, which would be suitable for the formation of a hydrogen bond group (Figure 3c).

Three residues in the binding pocket were replaced with alanine, thus generating the three mutants S59A, N74A and V87A. The binding abilities of the three mutants were different. The mutant N74A almost lost the binding for efficient alcohols. Either alcohol could compete with 50% 1-NPN from mutant N74A even when its concentration reaches 20μM. However, mutant S59A behaved like *L. migratoria* OBP1.
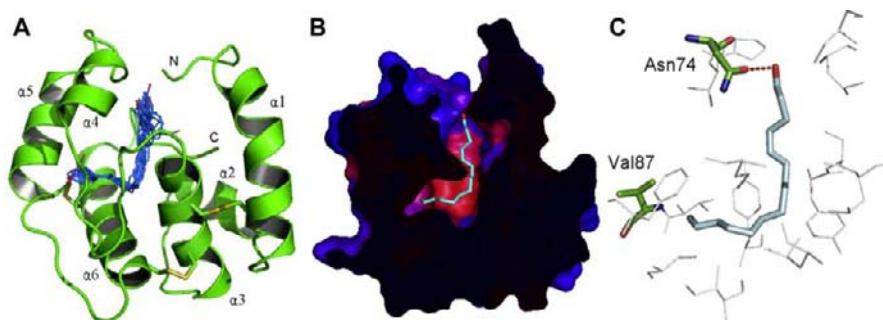
Figure 3. Docking experiment of *L. migratoria* OBP1 and ligand pentadecanol. (A) Overall structure of the *L. migratoria* OBP1 and the docking result. Disulfide bridges are shown as sticks, helices and the two termini are labeled. The top 20 docking poses are shown as stick models. Images were generated using Pymol (www.pymol.org). (B) The binding pocket of OBP1 and the docking result, with pentadecanol shown as a stick model with the hydroxyl oxygen in red. The color of the pocket surface is according to hydrophobicity, with high hydrophobicity in red and low in blue. (C) A close view of the potential binding mode of OBP1 with pentadecanol. Pentadecanol and the residues that were determined to be important for ligand binding are represented as stick models. Residues shown as line drawings have a distance to pentadecanol of less than 4 Å. Hydrogen bonds are shown as dashed lines.
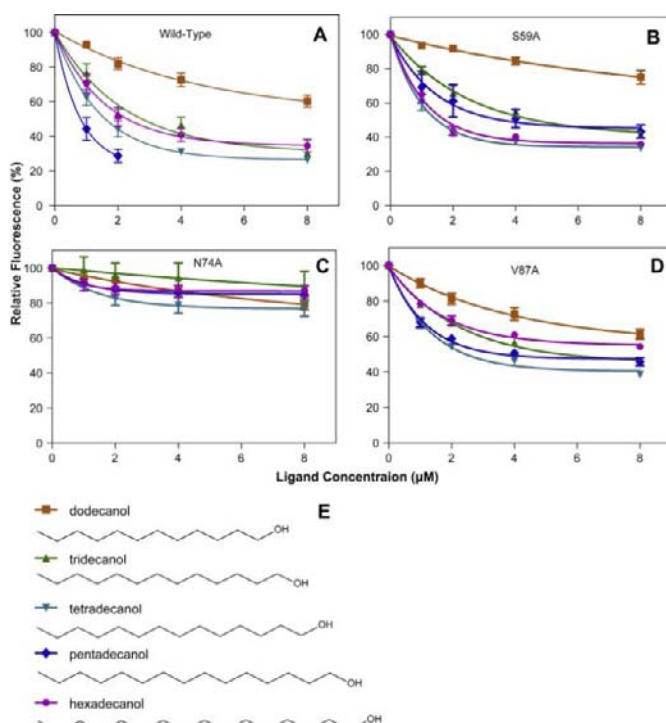
Figure 4. Binging properties of *L. migratoria* OBP1 (wild type) (A) and its three site-directed mutants S59A (B), N74A(C) and 87VA(D). Each ligand was incubated with 2 μM OBP and 2 μM 1-NPN at increasing concentrations. Fluorescence intensity values were normalized and plotted against total ligand concentration. All spectra were subjected to background subtraction. Data are an average of three independent measurements and the error bars represent the standard deviations of the mean derived from the differences between the measurements. Structures of ligands dodecanol, tridecanol, tetradecanol, pentadecanol and hexadecanol (E) were plotted by ChemSketch.

The mutant exhibited also very close Ki values for each tested ligand, except the strongest binding ligand pentadecanol with 6.0- fold decrease of Ki value. For tridecanol and tetradecanol, mutant V87A shows slightly decrease of affinities.

In contrast, the Ki value of pentadecanol and hexadecanol reduced by 3.3- and 4.2- fold respectively. Dodecanol could not compete with 50% 1-NPN from mutant N74A even when its concentration reaches 20µM (Figure 4). As the docking speculation, the experiment with site-directed mutagenesis in combination with fluorescence binding assays revealed that mutant N74A can not bind with tested ligands dodecanol, tridecanol, tetradecanol, pentadecanol or hexadecanol probably for losing the initial recognition by the hydrogen bond[34].

OBP/ODORANT COMPLEXES MAY ACT AS SIGNAL TRIGGERS

In 2005, a more surprising finding with *Drosophia* LUSH function had been unveiled[35]. *lush* mutant flies are insensitive to 11-cis vaccenyl acetate (cVA) (over 500-fold loss on sensitivity reported in more recent paper[9]) and lose cVA-evoked behaviors[35]. cVA is a pheromone mediating aggregation, aggression and sexual behaviors in *Drosophila* reportedly[35-39]. T1 neuron, which is specifically tuned to cVA[37, 40-42], has more than 400-fold reduction in spontaneous activity in *lush* mutant compared to wild type. T1 behaviors in *lush* mutant suggest that LUSH protein is not only required in cVA signal transduction, but a receptor firing activity trigger or even a signal activator as a complex with odorant. The LUSH mutations strongly support this hypothesis[9].
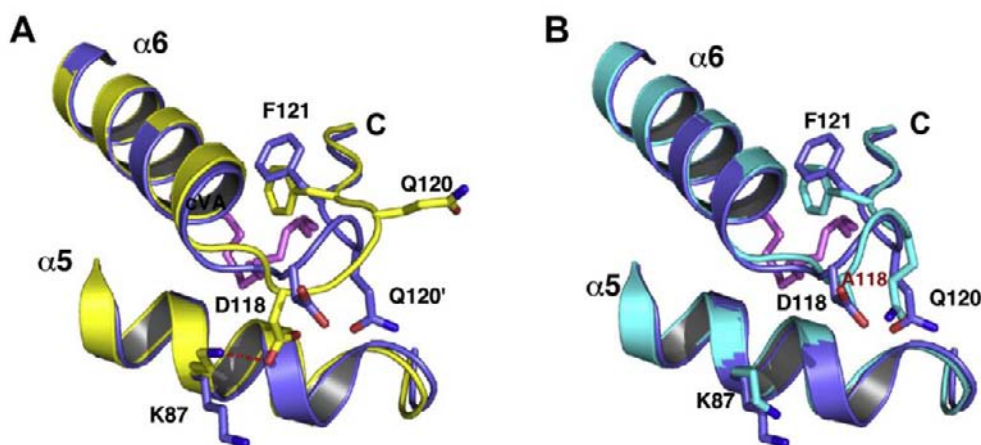
Figure 5. Structure of LUSH[D118A] mutant protein mimics cVA-bound LUSH protein. Ribbon diagram of the LUSH-cVA complex (blue) with LUSH-butanol complex (yellow) and the LUSH[D118A] (cyan). The cVA is shown as a stick model in magenta. (A) Comparison of the structures of the 117–121 loop in the LUSH-butanol (yellow) and LUSH-cVA complexes (blue). (B) Comparison of the structures of the same regions between LUSH-cVA (blue) and LUSH[D118A] (cyan). LUSH-cVA and LUSH[D118A] adopt similar conformations. This figure is provided courtesy of Dean Smith.

The acetate group of cVA makes polar interactions with The57 and Ser52 and is positioned by interaction with three aromatic residues Phe64, Phe113 and Try123.

At the opposite end of binding pocket, interaction of cVA with Phe121 induces a conformational change in the loop between residues 117 and 121, including Glu120 and Asp118, and results the disruption of a salt bridge between Asp118 and Lys87, pushing Phe121 outward.

Basing on the structural information, the authors engineered some proteins with point mutations to shape LUSH, which may mediate the cVA-induced activities in vivo. Remarkably, one of the mutations, LUSH with residue Asp118 substituted with alanine (LUSH[D118A]) even activates T1 neurons in the absence of cVA. After diffusing LUSH[D118A] into T1 sensilla lymph through recording pipette, the neuronal spontaneous activity becomes comparable to 1% cVA stimulation on wild type flies. The activity has no additional increase when cVA is applied. Infusion of LUSH[D118A] into other trichoid or basiconic sensilla has no effect on the activities of those neurons. By resolving the crystal structure of LUSH[D118A], they found D118A mutant highly mimics the conformational change induced by cVA binding to wild type LUSH, and strongly postulate that cVA/LUSH complex, not cVA alone, triggers pheromone signals on neuronal membrane physiologically (Figure 5)[9].

CONCLUSION

The odor recognition and discrimination is systematic and complicated physiological process in insect chemosensory system. Olfactory receptors (ORs), which are expressed on the membrane of olfactory receptor neurons, play a significant role during this process. The relatively small receptor repertoire in insects (62 or gens in *Drosophila* and 79 in mosquito *Anopheles gambiae*) decodes the odors by the combination of each own special odor response spectrum[43-45]. As media between the odors in the environment and ORs, OBPs may carry out some preliminary recognition with odors just before they send the stimuli to activate the olfactory neurons.

OBPs have the capacity to perform the recognition task. Firstly, OBPs are required for olfactory signal transduction. It has been found in fruit fly and fire ant[10, 13], and most recently, reports from *A. gambiae* have provided more evidence to support this[46, 47]. Secondly, the binding studies of OBPs showed the strong binding ability to some ligands. The 3D structure of OBPs even presents a binding cavity in the protein with ligand bound[4, 18]. This space of cavity is limited and can definitely not bind all sizes of ligands. Furthermore, the pH-dependent conformation of OBPs indicated that the proteins are not only capable of catching ligand, but also pushing it out. Thirdly, OBPs have some binding preference to different ligands[31-34].

By structural biology, some key residues of OBPs have been identified to help recognize the ligands. Crystal structure of *B. mori* PBP/bombykol complex suggested that bombykol is bound in the cavity through numerous hydrophobic interactions, and Ser56 is the critical residue for specific pheromone binding[15]. In *Drosophila* LUSH, Ser52 and Thr57 are located at the opening of the binding pocket and form hydrogen bonds with hydroxyl of alcohols and cVA[9, 20, 48]. Although lacking site-mutagenesis proof, three dimensional structures of *A. polyphemus* PBP1 revealed that the side-chain of Asn53 is located at the end of the hydrophobic cavity and proposed to be responsible for the recognition with the acetate group of the its pheromone[49]. In *L. migratoria* OBP1, the mutant by replacing Asn74 with

alanine can not bind tested ligands probably for losing the initial recognition by the hydrogen bond[34]. Most of the residues which may be important for ligands binding and recognition are polar and uncharged ones and are located at the entrance of the cavity. Therefore, in spite of complete metamorphosis or hemimetabolous insects, the recognition of linear aliphatic compounds by OBPs is quite conserved. The hydrophilic amino acids around the opening of the binding cavity are crucial for initial recognition and binding specificity of OBPs.

In addition, dominant LUSH mutant established a new function model for OBPs and suggested that at least in some cases, OBP/odor complex might be a signal trigger in olfactory circuit [9]. Recently, an interesting scavenger receptor, which is called sensory neuron membrane protein (SNMP), was found to be required for cVA sensitivity in T1 neurons in *Drosophila*[50, 51]. SNMP are co-localized with ORs in all types of trichoid sensilla[50]. In contrast to LUSH mutant, T1 neurons from the mutant missing SNMP show increased spontaneous activity (14-25 spikes/second compared with ~1 spike/second in wild type, while T2 neurons remain normal[51]). SNMP may function downstream of LUSH, and upstream of or in parallel with ORs. Here come some significant questions. Does LUSH trigger and SNMP inhibit the spontaneous neuron firing? Does cVA or LUSH/cVA complex interact with SNMP, or directly with ORs? Although cVA can directly activate OR in the absence of LUSH and/or SNMP [42, 50], in real physiological environment, LUSH [40] and SNMP are still required for effective and quick pheromone signal transduction in flies.

LUSH, as well as SNMP, both only affect pheromone-evoked neuronal activity in *Drosophila*. These surprising findings suggest that pheromone perception process is special and different from other general odors perception. It's very interesting to look the similar mechanisms to be identified in other insect species, even in mammals.

The interaction between odor molecules, OBPs, SNMP and ORs is a big puzzle so far. ORs were just identified as ion channels on sensory membrane[52, 53]. The functional complex of these molecules must be a research hot spot in the future. People would be very interested to understand what factors activate and close the OR channels and what else are involved in a functional channel unit. How does the channel form and whether SNMP is involved in the channel probably would be the next questions people try to answer.

REFERENCES

[1]   Vogt, R.G. and L.M. Riddiford, *Pheromone binding and inactivation by moth antennae. Nature*, 1981. 293(5828): p. 161-3.

[2]   Klein, U., *Sensillum-lymph proteins from antenna1 olfactory hairs of the moth Antheraea polyphemus (Satumiidae). Insect. Biochem.*, 1987. 17: p. 1193-204.

[3]   Raming, K., J. Krieger, and H. Breer, *Molecular cloning of an insect pheromone-binding protein. FEBS Lett*, 1989. 256(1-2): p. 215-8.

[4]   Pelosi, P., et al., *Soluble proteins in insect chemical communication. Cell Mol. Life Sci.,* 2006. 63(14): p. 1658-76.

[5]   Pelletier, J. and W.S. Leal, *Genome analysis and expression patterns of odorant-binding proteins from the Southern House mosquito Culex pipiens quinquefasciatus. PLoS One*, 2009. 4(7): p. e6237.

[6]   Foret, S. and R. Maleszka, *Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). Genome Res.*, 2006. 16(11): p. 1404-13.

[7]   Gong, D.P., et al., *The odorant binding protein gene family from the genome of silkworm, Bombyx mori. BMC Genomics*, 2009. 10: p. 332.

[8]   Zhou, J.J., et al., *Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid Acyrthosiphon pisum. Insect. Mol. Biol.,* 2010. 19 Suppl 2: p. 113-22.

[9]   Laughlin, J.D., et al., *Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein. Cell*, 2008. 133(7): p. 1255-65.

[10]   Ross, K.G., *Multilocus evolution in fire ants: effects of selection, gene flow and recombination. Genetics*, 1997. 145(4): p. 961-74.

[11]   Ross, K.G. and L. Keller, *Genetic control of social organization in an ant. Proc. Natl. Acad. Sci. USA*, 1998. 95(24): p. 14232-7.

[12]   Krieger, M.J. and K.G. Ross, *Identification of a major gene regulating complex social behavior. Science*, 2002. 295(5553): p. 328-32.

[13]   Kim, M.S., A. Repp, and D.P. Smith, *LUSH odorant-binding protein mediates chemosensory responses to alcohols in Drosophila melanogaster. Genetics*, 1998. 150(2): p. 711-21.

[14]   Kim, M.S. and D.P. Smith, *The invertebrate odorant-binding protein LUSH is required for normal olfactory behavior in Drosophila. Chem. Senses*, 2001. 26(2): p. 195-9.

[15]   Sandler, B.H., et al., *Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. Chem. Biol.,* 2000. 7(2): p. 143-51.

[16]   Damberger, F., et al., *NMR characterization of a pH-dependent equilibrium between two folded solution conformations of the pheromone-binding protein from Bombyx mori. Protein Sci.*, 2000. 9(5): p. 1038-41.

[17]   Horst, R., et al., *NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. Proc. Natl. Acad. Sci. USA*, 2001. 98(25): p. 14374-9.

[18]   Tegoni, M., V. Campanacci, and C. Cambillau, *Structural aspects of sexual attraction and chemical communication in insects. Trends Biochem Sci.*, 2004. 29(5): p. 257-64.

[19]   Lartigue, A., et al., *Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee Apis mellifera L. J. Biol. Chem.*, 2004. 279(6): p. 4459-64.

[20]   Kruse, S.W., et al., *Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster. Nat. Struct. Biol.*, 2003. 10(9): p. 694-700.

[21]   Lee, D., et al., *NMR structure of the unliganded Bombyx mori pheromone-binding protein at physiological pH. FEBS Lett.*, 2002. 531(2): p. 314-8.

[22]   Leal, W.S., A.M. Chen, and M.L. Erickson, *Selective and pH-dependent binding of a moth pheromone to a pheromone-binding protein. J. Chem. Ecol.*, 2005. 31(10): p. 2493-9.

[23]   Zubkov, S., et al., *Structural consequences of the pH-induced conformational switch in A.polyphemus pheromone-binding protein: mechanisms of ligand release. J. Mol. Biol.,* 2005. 354(5): p. 1081-90.

[24] Leite, N.R., et al., *Structure of an odorant-binding protein from the mosquito Aedes aegypti suggests a binding pocket covered by a pH-sensitive "Lid". PLoS One*, 2009. 4(11): p. e8006.

[25] Wogulis, M., et al., *The crystal structure of an odorant binding protein from Anopheles gambiae: evidence for a common ligand release mechanism. Biochem. Biophys. Res. Commun.*, 2006. 339(1): p. 157-64.

[26] Pesenti, M.E., et al., *Queen bee pheromone binding protein pH-induced domain swapping favors pheromone release. J. Mol. Biol.*, 2009. 390(5): p. 981-90.

[27] Lartigue, A., et al., *The crystal structure of a cockroach pheromone-binding protein suggests a new ligand binding and release mechanism. J. Biol. Chem.,* 2003. 278(32): p. 30213-8.

[28] Lartigue, A., et al., *Crystallization and preliminary crystallographic study of a pheromone-binding protein from the cockroach Leucophaea maderae. Acta. Crystallogr D Biol. Crystallogr,* 2003. 59(Pt 5): p. 916-8.

[29] Bette, S., H. Breer, and J. Krieger, *Probing a pheromone binding protein of the silkmoth Antheraea polyphemus by endogenous tryptophan fluorescence. Insect. Biochem. Mol. Biol.*, 2002. 32(3): p. 241-6.

[30] Mohl, C., H. Breer, and J. Krieger, *Species-specific pheromonal compounds induce distinct conformational changes of pheromone binding protein subtypes from Antheraea polyphemus. Invert. Neurosci.*, 2002. 4(4): p. 165-74.

[31] Maida, R., G. Ziegelberger, and K.E. Kaissling, *Ligand binding to six recombinant pheromone-binding proteins of Antheraea polyphemus and Antheraea pernyi. J. Comp. Physiol. B,* 2003. 173(7): p. 565-73.

[32] Riviere, S., et al., *A pheromone-binding protein from the cockroach Leucophaea maderae: cloning, expression and pheromone binding. Biochem. J.*, 2003. 371(Pt 2): p. 573-9.

[33] Zhou, J.J., et al., *Revisiting the odorant-binding protein LUSH of Drosophila melanogaster: evidence for odour recognition and discrimination. FEBS Lett.*, 2004. 558(1-3): p. 23-6.

[34] Jiang, Q.Y., et al., *Binding specificity of locust odorant binding protein and its key binding site for initial recognition of alcohols. Insect. Biochem. Mol. Biol.*, 2009. 39(7): p. 440-7.

[35] Xu, P., et al., *Drosophila OBP LUSH is required for activity of pheromone-sensitive neurons. Neuron*, 2005. 45(2): p. 193-200.

[36] Bartelt, R.J., A.M. Schaner, and L.L. Jackson, *cis-Vaccenyl acetate as an aggregation pheromone inDrosophila melanogaster. Journal of Chemical Ecology*, 1985. 11(12): p. 1747-56.

[37] Kurtovic, A., A. Widmer, and B.J. Dickson, *A single class of olfactory neurons mediates behavioural responses to a Drosophila sex pheromone. Nature*, 2007. 446(7135): p. 542-6.

[38] Ejima, A., et al., *Generalization of courtship learning in Drosophila is mediated by cis-vaccenyl acetate. Curr. Biol.*, 2007. 17(7): p. 599-605.

[39] Wang, L. and D.J. Anderson, *Identification of an aggression-promoting pheromone and its receptor neurons in Drosophila. Nature,* 2010. 463(7278): p. 227-31.

[40] Ha, T.S. and D.P. Smith, *A pheromone receptor mediates 11-cis-vaccenyl acetate-induced responses in Drosophila. J. Neurosci.*, 2006. 26(34): p. 8727-33.

[41] Clyne, P., et al., *Odorant response of individual sensilla on the Drosophila antenna. Invert. Neurosci.*, 1997. 3(2-3): p. 127-35.

[42] van der Goes van Naters, W. and J.R. Carlson, *Receptors and neurons for fly odors in Drosophila. Curr. Biol.*, 2007. 17(7): p. 606-12.

[43] Lu, T., et al., *Odor coding in the maxillary palp of the malaria vector mosquito Anopheles gambiae. Curr. Biol.*, 2007. 17(18): p. 1533-44.

[44] Hallem, E.A., A. Dahanukar, and J.R. Carlson, *Insect odor and taste receptors. Annu. Rev. Entomol.*, 2006. 51: p. 113-35.

[45] Wang, G., et al., *Molecular basis of odor coding in the malaria vector mosquito Anopheles gambiae. Proc. Natl. Acad. Sci. USA*, 2010. 107(9): p. 4418-23.

[46] Pelletier, J., et al., *Knockdown of a mosquito odorant-binding protein involved in the sensitive detection of oviposition attractants. J. Chem. Ecol.*, 2010. 36(3): p. 245-8.

[47] Biessmann, H., et al., *The Anopheles gambiae odorant binding protein 1 (AgamOBP1) mediates indole recognition in the antennae of female mosquitoes. PLoS O*ne, 2010. 5(3): p. e9471.

[48] Thode, A.B., et al., *The role of multiple hydrogen-bonding groups in specific alcohol binding sites in proteins: insights from structural studies of LUSH. J. Mol. Biol.*, 2008. 376(5): p. 1360-76.

[49] Mohanty, S., S. Zubkov, and A.M. Gronenborn, *The solution NMR structure of Antheraea polyphemus PBP provides new insight into pheromone recognition by pheromone-binding proteins. J. Mol. Biol.,* 2004. 337(2): p. 443-51.

[50] Benton, R., K.S. Vannice, and L.B. Vosshall, *An essential role for a CD36-related receptor in pheromone detection in Drosophila. Nature*, 2007. 450(7167): p. 289-93.

[51] Jin, X., T.S. Ha, and D.P. Smith, *SNMP is a signaling component required for pheromone sensitivity in Drosophila. Proc. Natl. Acad. Sci. USA*, 2008. 105(31): p. 10996-1001.

[52] Wicher, D., et al., *Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature*, 2008. 452(7190): p. 1007-11.

[53] Sato, K., et al., *Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature*, 2008. 452(7190): p. 1002-6.

Chapter 8

THE PROTEOMIC CODE: A MOLECULAR RECOGNITION CODE FOR PROTEINS

Jan C. Biro[*]

Homulus Foundation, 88 Howard, #1205, San Francisco
CA 94105, USA

ABSTRACT

The Proteomic Code is a set of rules by which information in genetic material is transferred into the physico-chemical properties of amino acids and determines how individual amino acids interact with each other during folding and in specific protein–protein interactions. The Proteomic Code is part of the redundant Genetic Code. The 25-year-old history of this concept is reviewed from the first independent suggestions by Biro and Mekler, through the works of Blalock, Root-Bernstein, Siemion, Miller and others, followed by the discovery of a Common Periodic Table of Codons and Nucleic Acids in 2003 and culminating in the recent conceptualization of partial complementary coding of interacting amino acids as well as the theory of the nucleic acid-assisted protein folding. A novel cloning method for the design and production of specific and with high affinity reacting proteins (SHARP) is presented. This method is based on the concept of proteomic codes and is suitable for large-scale, industrial production of specifically interacting peptides.

ABBREVIATIONS

Amino Acid Complementarity	The physico-chemical complementarity (size, charge, hydropathy) of amino acid pairs as well as their origin from complementary codons
Complementary Amino Acids	Amino acid pairs, coded by complementary codons.

---

[*] E-mail address: jan.biro@comcast.net

NA	Nucleic Acid
P	Protein
PC	Proteomic Code, the comprehensive rules describing the origin and nature of amino acid complementarity.
SHARP	Specific High Affinity Reacting Peptides W-C base pairs:Watson–Crick's complementary base pairs.

INTRODUCTION

Nucleic acids and proteins are the carriers of most (if not all) biological information. This information is complex, well organized in space and time. These two kinds of macromolecules have a polymer structure. Nucleic acids are built from four nucleotides and the proteins are built from 20 amino acids (as basic units). Both nucleic acids and proteins are able to interact with each other and in many cases these interactions are extremely strong (K_d ~10^9–10^{12} M) and extremely specific. The nature and origin of this specificity is well understood in the case of nucleic acid–nucleic acid (NA-NA) interactions (DNA–DNA, DNA–RNA, RNA–RNA), as is the complementarity of the Watson–Crick (W-C) base pairs. The specificity of NA-NA interactions is determined, undoubtedly, at the basic unit level where the individual bases have a prominent role.
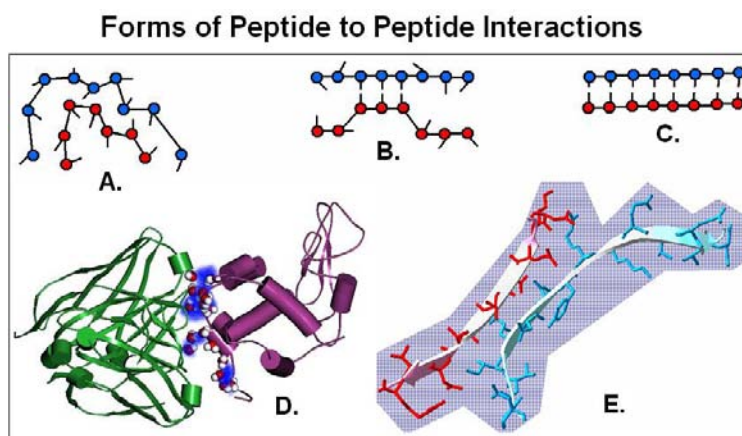
Figure 1. Forms of peptide to peptide interactions. The specificity of interactions between two peptides might be explained in two ways. First, many amino acids collectively form larger configurations (protrusions and cavities, charge and hydropathy fields) which fit each other (A and D). Second, the physico-chemical properties (size, charge, hydropathy of individual amino acids fit each other like "lock and key" (C and E). There are even intermediate forms (B).

Our most established view on the specificity of protein–protein (P-P) interactions is completely different [1]. In this case the individual amino acids in a particular protein together establish a large 3D structure. This structure has protrusions and cavities, charged and uncharged areas, hydrophobic and hydrophilic patches on its surface, which altogether form a complex 3D pattern of spatial and physico-chemical properties. Two proteins will specifically interact with each other if their complex 3D pattern of spatial and physico-chemical properties fits to each other as a mold to its template or a key to its lock. In this way

the specificity of P-P interactions is determined at a level larger than one amino acid (Figure 1).

The nature of specific nucleic acid–protein (NA-P) interactions is less understood. It is suggested that some bases together form some kind of 3D structure which fits to the 3D structure of a protein (in the case of single-stranded nucleic acids). Alternatively, double-stranded nucleic acid provides a pattern of atoms in the grooves of the double strands, with an atomic pattern that is in some way specifically recognized by nucleo-proteins [2].

Regulatory proteins are known to recognize specific DNA sequences directly through atomic contacts between protein and DNA, and/or indirectly through the conformational properties of the DNA.

There has been ongoing intellectual effort for the last 30 years to explain the nature of specific P-P interactions at the residue unit (individual amino acid) level. This view states that there are individual amino acids which preferentially co-locate with each other in specific P-P contacts and form amino acid pairs which are physico-chemically more compatible with each other than any other amino acid pairs. These physico-chemically highly compatible amino acid pairs are complementary to each other in analogy to the W-C base pair complementarity.

The comprehensive rules describing the origin and nature of amino acid complementarity is called the Proteomic Code.

THE HISTORY OF THE PROTEOMIC CODE

PEOPLE FROM THE PAST

This is my very subjective selection of scientists, for whom I have great respect and I think they contributed – in one or another way – to the development of the Proteomic Code.

Linus Pauling is regarded as "the greatest chemist who ever lived". *The Nature of the Chemical bond* is fundamental to the understanding of any biological interaction [3]. His works on the protein structure are classics [4]. His unconfirmed DNA model, in contrast to the established model, gives some theoretical ideas on how specific nucleic acid–protein interaction might happen [5,6].

Carl R Woese is famous for defining the Archaea, the third life form on Earth (in addition to bacteria and eucarya). Less known is his "RNA world" hypothesis. This is a theory which proposes that a world filled with RNA (ribonucleic acid)-based life predates current DNA (deoxyribonucleic acid)-based life. RNA, which can store information like DNA *and* catalyze reactions like proteins (enzymes), may have supported cellular or pre-cellular life. Some theories as to the origin of life present RNA-based catalysis and information storage as the first step in the evolution of cellular life.

The RNA world is proposed to have evolved into the DNA and protein world of today. DNA, through its greater chemical stability, took over the role of data storage while proteins, which are more flexible in catalysis through the great variety of amino acids, became the specialized catalytic molecules. The RNA world hypothesis suggests that messenger RNA (mRNA), the intermediate in protein production from a DNA sequence, is the evolutionary remnant of the "RNA world" [7]. His concept of a common origin of nucleic acid and protein "worlds" is very much compatible with the foundation of the Proteomic Code.

Margaret O Dayhoff is the mother of bioinformatics. She was the first who collected and edited the *Atlas of Protein Sequence and Structure* [8] and later introduced statistical methods into protein sequence analyses. Her work was a huge asset and inspiration to my first suggestion of the Proteomic Code [9,10,11].

George Gamow was a theoretical physicist and cosmologist and spent only a few years in Cambridge, UK, but he was there when the structure of DNA was discovered in 1954. He developed the first genetic code, which was not only an elegant solution for the information transfer from DNA to proteins, but at the same time explained how DNA might specifically interact with proteins [12,13,14,15,16,17]. In his mind, the codons were mirror images of the coded amino acids and they had a very intimate relationship with each other. His genetic code proved to be wrong and the nature of specific nucleic acid–protein interactions is still not known, but he remains a strong inspiration (Figure 2) [18,19].
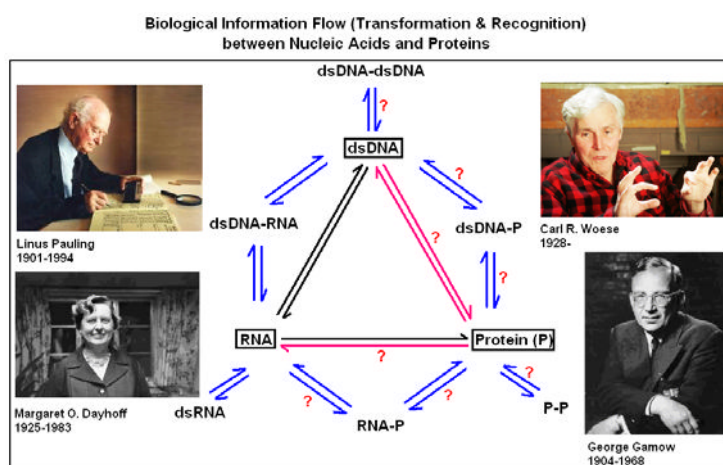
Figure 2. Biological information flow (transformation and recognition) between nucleic acids and proteins. All biological information is stored in nucleic acids (DNA/RNA) and much in proteins (P). The information transfer and interactions between nucleic acids and the formation of double-stranded (ds) forms are well known and understood. However, the exact nature of P-P and P-nucleic acid interactions is still obscure. The works of these four scientists played important roles in much that we know about these information transfers and interactions (subjectively chosen by the author of this article).

FIRST GENERATION MODELS FOR THE PROTEOMIC CODE

The first generation models (up to 2006) of the novel Proteomic Code are based on perfect codon complementarity coding of interacting amino acid pairs.

MEKLER

Mekler described an idea of sense and complementary peptides which may be able to interact specifically, mediated by specific through-space, pair wise interactions between amino acid residues [20]. He suggested that amino acids of specifically interacting proteins,

in their specifically interacting domains, are composed of two parallel sequences of amino acid pairs that are spatially complementary to each other, similar to the Watson–Crick base pairs in nucleic acids. The protein nucleic acid analogy in his theory continued and he proposed that these spatially complementary amino acids are coded by reverse-complementary codons (translational reading in the 5′→3′ direction).

It is possible to segregate 64 (the number of different codons, including three stop codons) of all the possible putative amino acid pairs ($20{\times}20/2{=}200$) into three non-overlapping groups [21].

BIRO

I was also inspired by the complementarity of nucleic acids and developed a theory of complementary coding of specifically interacting amino acids [9,10,11]. I had no knowledge of the publications of Mekler or Idlis (published in two Russian papers). I was also convinced that amino acid pairs coded by complementary codons (no matter if in the same 5′→3′/5′→3′ or opposite 5′→3′/3′→5′ orientations) are somehow special and suggested that these pairs of amino acids might be responsible for the specific intra- and intermolecular peptide interactions.

I developed a method for pair wise computer searching of protein sequences for complementary amino acids and found that these specially coded amino acid pairs are statistically overrepresented in those proteins known to interact with each other. In addition, I was able to find short complementary amino acid sequences within the same protein sequences and concluded that these might play a role in the formation or stabilization of 3D protein structures (Figure 3).

Origin of Proteomic Codes
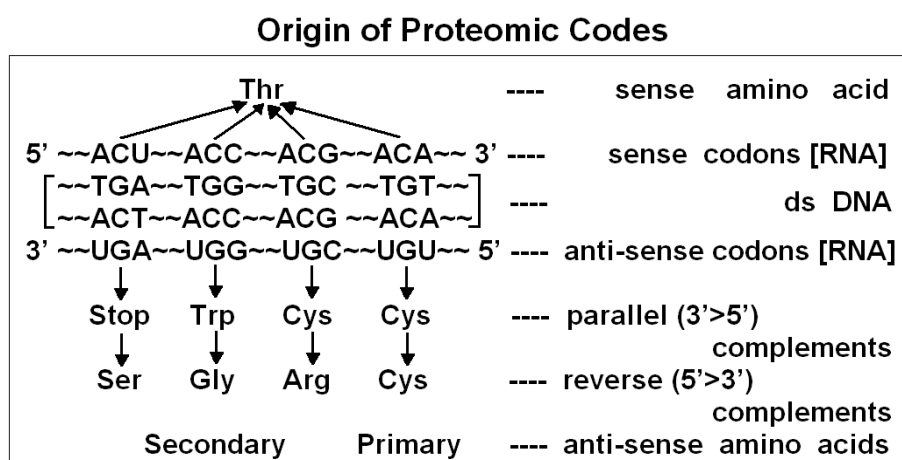
Figure 3. Origin of the Proteomic Code. Threonine (Thr) is coded by 4 different synonymous codons. Complementary triplets are coding different amino acids in parallel (3′→5′) and anti-parallel (5′→3′) readings. Amino acids coded by symmetrical codons are called "primary" and others "secondary" anti-sense amino acids (modified from [9]).

Molecular modeling showed the size compatibility of complementary amino acids and that they might form a 5–7 atoms long bridge between the alpha C atoms of amino acids. It was a rather ambitious theory at a time when the antisense DNA sequences were called nonsense, and it was an even more ambitious method when computers were programmed by punch-cards and the protein databases were based on Dayhoff's three volumes of protein sequences [8].

BLALOCK-SMITH

This theory is called the *molecular recognition theory*. Synonymous names are *hydropathy complementarity* or *anti-complementarity theory* and was based on the observation [22] that codons for hydrophilic and hydrophobic amino acids are generally complemented by codons for hydrophobic and hydrophilic amino acids, respectively. This is the case even when the complementary codons are read in the 3′→5′′ direction. Peptides, specified by complementary RNAs, bind to each other with specificity and high affinity [23,24]. The theory turned out to be very fruitful in neuro-endocrine and immune research [25,26].

A very important observation is that antibodies against complementary antibodies also specifically interact with each other. Bost and Blalock [27] synthesized two complementary oligopeptides (i.e. peptides translated from complementary mRNAs, in opposing directions). The two peptides, Leu-Glu-Arg-Ile-Leu-Leu (LERILL), and its complementary peptide, Glu-Leu-Cys-Asp-Asp-Asp (ELCDDD), specifically recognized each other in radioimmunoassay. Antibodies were produced against both peptides. Both antibodies specifically recognized their own antigen. Using radioimmunoassays, anti-ELCDDD antibodies were shown to interact with ^{125}I-labeled anti-LERILL antibodies but not with ^{125}I-labeled control antibodies. More importantly, the interaction of the two antibodies could be blocked using either peptide antigen, but not with control peptides. Furthermore, ^{125}I-labeled anti-LERILL binding to LERILL could be blocked with anti-ELCDDD antibody and vice versa. It was concluded therefore that antibody/antibody binding occurred at or near the antigen combining site, demonstrating that this was an idiotypic/anti-idiotypic interaction.

This experiment clearly showed the existence (and functioning) of an intricate network of complementary peptides and interactions. Much effort is being made to master this network and use it in protein purification, binding assays, medical diagnosis and therapy.

Recently, Blalock [28]{Blalock JE: Department of Physiology and Biophysics, The University of Alabama at Birmingham. *http://www.physiology. uab.edu/Blalock.htm# Research*} has emphasized that nucleic acids encode amino acid sequences in a binary fashion with regard to hydropathy and that the exact pattern of polar and non-polar amino acids, rather than the precise identity of particular R groups, is an important driver for protein shape and interactions. The perfect codon complementarity behind coding of interacting amino acids is no longer an absolute requirement for his theory.

It is easy to see that there is similarity between hydropathy profiles of the four possible translations of a nucleic acid (Figure 4), but the validity of hydrophobe–hydrophyl interactions remains unanswered.
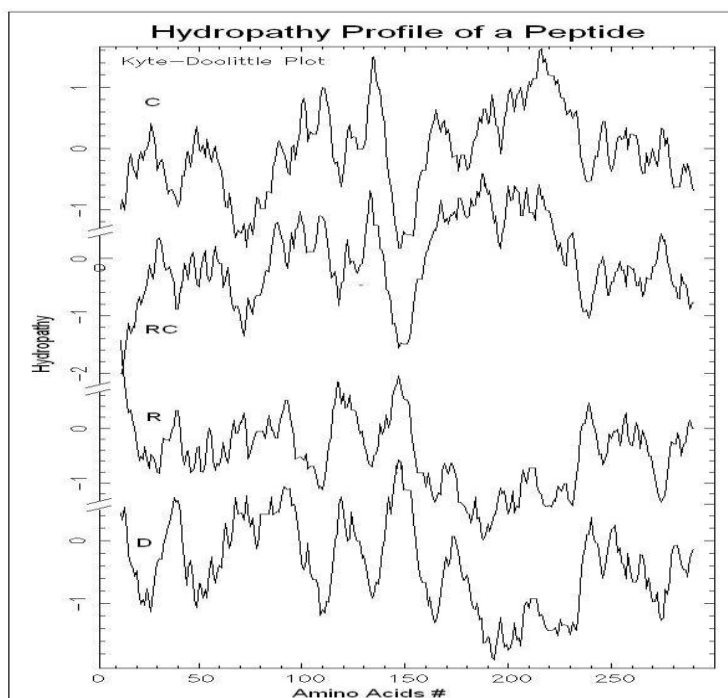
Figure 4. Hydropathy profile of a protein. An artificially constructed nucleic acid sequence was randomized and translated into the four possible directions (D, direct; RC, reverse-complement; R, reverse; C, complement) sequences. The D sequence was designed to contain equal numbers of the 20 amino acids. The hydropathy profile was determined by an online tool (http://www.tcdb.org/progs/hydropathy.php).

ROOT-BERNSTEIN

Another amino acid pairing hypothesis was presented by Root-Bernstein [29,30]]. He focused on whether it was possible to build amino acid pairs meeting standard criteria for bonding. He concluded that it was possible only in 26 cases (of 210 pairs). Of these 26, 14 were found to be genetically encoded by perfect complementary codons (read in the same orientation ($5'{\rightarrow}3'/3'{\rightarrow}5'$) while in 12 cases mismatch was found at the wobble position of pairing codons.

SIEMION

There is a regular connection between activation energies (measured as enthalpies (ΔH^{++}) and entropies (ΔS^{++}) of activation for the reaction of 18 N''-hydroxysuccinimide esters of N-protected proteinaceous amino acids with p-anisidine) and the genetic code [31,32,33].

This periodic change of amino acid reactivity within the genetic code led him to suggest a peptide–anti-peptide pairing. This is rather similar to Root-Bernstein's hypothesis.

MILLER

Practical use is the best test of a theory. Technologies based on interacting proteins have a significant market in different branches of biochemistry, as well as in medical diagnostics and therapy. The Genetic Therapies Centre (GTC) *(http://www.gtc.ch.ic.ac.uk/about_us/ overview.shtml* at the Imperial College London (*http://www.imperial.ac.uk/P2801.htm*) was founded in 2001 with major financial support from a Japanese company, the Mitsubishi Chemical Corporation, and the UK charity, the Wolfson Foundation) is one of the first academic centers that are openly investing in Proteomic Code-based technologies. With the clear intention for their science "to be used in the marketplace", Andrew Miller, the first director of GTC and co-founder of its first spin-off company, Proteom Ltd (http://www.prosarix.com/default.htm), is making major contributions to this field [34,35,36,37,38].

However, Miller and his colleagues came to realize that the amino acid pairs provided by perfectly complementary codons are not always the best pairs, and deviations from the original design sometimes significantly improved the quality of a protein–protein interaction. Therefore the current view of Miller is that there are "strategic pairs of amino acid residues that form part of a new, through-space two-dimensional amino acid interaction code (Proteomic Code).

The proteomic code and derivatives thereof could represent a new molecular recognition code relating the 1D world of genes with the 3D world of protein structure and function, a code that could shortcut and obviate the need for extensive research into the proteome to give form and function to currently available genomic information (i.e., true functional genomics)" (http://www.gtc.ch.ic.ac.uk/research_interests/code.shtml).

THE PROTEOMIC CODE AND THE 3D STRUCTURE OF PROTEINS

It is widely accepted that the 3D structure of proteins plays a significant role in their specific interactions and function. The opposite is less obvious, namely that specific and individual amino acid pairs or the sequence of these pairs might determine the foldings of proteins.

Complementarity at the amino acid level in the proteins and the corresponding internal complementarity within the coding mRNA (the Proteomic Code) raises the intriguing possibility that some protein folding information is present in the nucleic acids (in addition or within the known and redundant genetic code). A higher propensity of complementarily coded amino acids was observed in real protein sequences than in translations of corresponding random nucleotide sequences [9,10,11].

The internal amino acid complementarity allows the polypeptides coded by complementary codons to retain the secondary structure patterns of the translated strand (mRNA). Thus, genetic code redundancy could be related to evolutionary pressure towards retention of protein structural information in complementary codons and nucleic acid subsequences [39,40,41,42,43,44].

EXPERIMENTAL EVIDENCE

Experiments based on the idea of a Proteomic Code usually start with a well-known receptor-ligand type protein interaction. A short sequence is selected (often <10 amino acids long) which is known or suspected to be involved in the direct contact between the proteins in question (P-P/r). A complementary oligopeptide sequence is derived using the known mRNA sequence of the selected protein epitope, making a reverse complementation of the sequence, translating it and synthesizing it.

The flow of the experiments looks as follows:

(a) choose an interesting peptide
(b) select a short, "promising" oligo-peptide epitope (P);
(c) find the true mRNA of P;
(d) reverse-complement this mRNA;
(e) translate the reverse-complemented mRNA into complementary peptide (P/c);
(f) test P–P/c interaction (affinity, specificity);
(g) use P/c for finding P-like sequences (for histochemistry, affinity purification);
(h) use P/c to generate antibodies (P/c_ab);
(i) test P/c_ab for its interaction with the P-receptor (P/r) and use it for example labeling or affinity purification of P/r;
(j) use P_ab (as well as antibodies to P, P_ab) to find and characterize idiopathic (P_ab– P/c_ab) antibody reactions.

A nice feature of Proteomic-Code based technology is that the amino acid complementarity (information mirroring) does not stop with the P–P/c interaction but continues and involves even the antibodies generated against the original interacting domains; even P_ab–P/c_ab, i.e., antibodies against interacting proteins, will themselves contain interacting domains. They are idiotypes.

Peptides and interactions involved in Proteomic Code-based experiments are summarized in Figure 5.

A very nice example of this technology and its potential is given by Bost and Blalock [27] (described above), It is reviewed by Heal et al. [37] and McGuian [45]. Table 1 presents a number of these kinds of experiments.

Some experiments or types of experiments require further attention.

The *antisense homology box*, a new motif within proteins that encodes biologically active peptides, was defined by Baranyi an coworkers, around 1995. They used a bioinformatical method for a genome-wide search of peptides coded by complementary exon sequences. They found that amphiphilic peptides, approximately 15 amino acids in length, and their corresponding antisense peptides exist within protein molecules. These regions (termed antisense homology boxes) are separated by approximately 50 amino acids. They concluded that because many sense–antisense peptide pairs have been reported to recognize and bind to each other, antisense homology boxes may be involved in folding, chaperoning and oligomer formation of proteins. The frequency of peptides in antisense homology boxes was 4.2 times higher than expected from random sequences ($p<0.001$) [46].
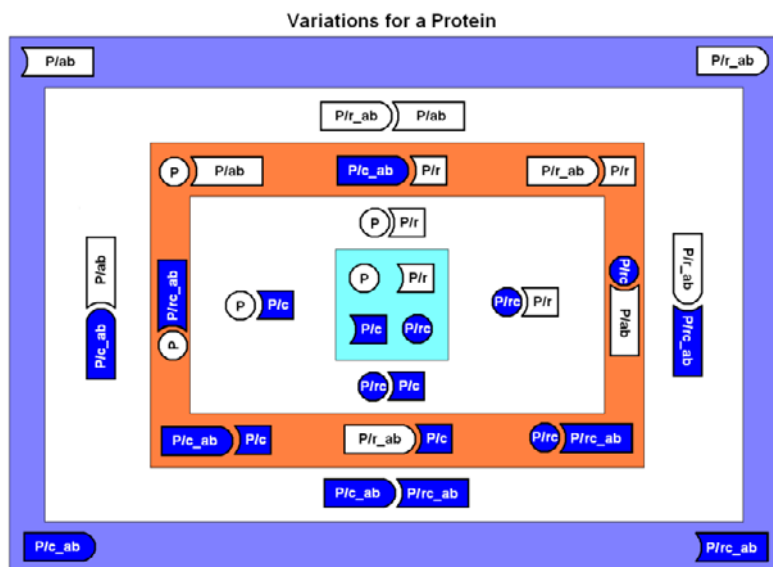
Figure 5. Variations for a protein. Experiments regarding the Proteomic Code are usually designed for the peptides and peptide interactions depicted in this figure. A peptide (P) naturally interacts with its receptor (P/r). Antibodies against this protein (P/ab) and its receptor (P/r_ab) might also be naturally present *in vivo* as part of the immune surveillance or artificially arise. The Proteomic Code gives a method to design artificial oligopeptides (P/c and P/rc) with the potency to interact with the receptor and its ligand. P and P/c as well as Pr and P/rc are expressed from complementary nucleic acid sequences. It is possible to raise antibodies against P/c (P/c_ab) and P/rc (P/rc_ab).

They successfully confirmed their suggestion by experimental results. The antisense homology box-derived peptide CALSVDRYRAVASW, a fragment of human endothelin A receptor, proved to be a specific inhibitor of endothelin peptide (ET-1) in a smooth muscle relaxation assay. The peptide was also able to block endotoxin-induced shock in rats. The finding of endothelin receptor inhibitor among antisense homology box-derived peptides indicates that searching proteins for this new motif may be useful in finding biologically active peptides [47,48,49].

A bioinformatics experiment, similar to Baranyi's, was performed by Segerstéen et al. [50]. The hypothesis that nucleic acids, which code specifically interacting receptor and ligand proteins, contain complementary sequences was tested. Human insulin mRNA (HSINSU) contained 16 sequences which were 23.8 ± 1.4 nucleotides long and were complementary to the insulin receptor mRNA (HSIRPR, $74.8\pm1.9\%$ complementary matches, $p<0.001$ compared to randomly occurring matches).

However, when examining 10 different nucleic acids (coding proteins not interacting with the insulin receptor), 81 additional sequences were found which were also complementary to HSIRPR. Although the finding of short complementary sequences was statistically highly significant, we concluded that this is not specific for nucleic acid coding of specifically interacting proteins.

There are two kinds of antisense technologies based on the complementarity of nucleic acids: (a) when the production of a protein is inhibited by an oligonucleotide sequence which is complementary to its mRNA; this is a pre-translational modification and it usually requires transfer of nucleic acids into the cells; (b) when the biological effect of an already complete

protein is inhibited by another protein which is translated from its complementary mRNA; this is a post-translational modification and does not block the synthesis of a protein.

Table I.

Experiments related to the Proteomic Code

Target Proteins				Antibodies				Lig.-Rec. typ					Ag-Ab type							Ab-Ab type				Comments	Authors	Year	Ref.
	P/r	P/c	P/rc	P/ab	P/r_ab	P/c_ab	P/rc_ab	P | P/r	P | P/c	P/r | P/c	Prc | P/r	P/rc | P/c	P | P/ab	P/r | P/r_ab	P/c_ab | P/c	P/rc | P/rc_ab	P | P/c_ab	P/r | P/c_ab	P/rc | P/ab	P/ab | P/r_ab	P/ab | P/c_ab	P/r_ab | P/rc_ab	P/c_ab | P/rc_ab				
protein (P2)		x			x			x											x						treatment of allergic neuritis	Araga et al.	1999	1
acetylcholine receptor	x	x		x		x																x			treatment of myasthenia gravis	Araga et al.	1993	2
prolactin			x						x																	Bajpai et al.	1991	3
endothelin A	x		x							x															antisense homology boxes (AHB)	Baranyi et al.	1995	4
anaphylatoxin (C5a)	x	x						x	x																antisense homology boxes (AHB)	Baranyi et al.	1996	5
interleukin-18		x							x																designe of "receptor-surrogates", lead,	Bhakoo et al.	2004	6
ACTH		x							x																	Blalock & Bost	1986	7
trypsin-TMOF	x	x						x	x																Kd~microM	Borovsky et al.	1994	8
Antibodies (Anty-idiotypic)	x			x	x								x	x							x				classical	Bost & Blalock	1989	9
ACTH		x							x																	Bost et al.	1985	10
endorphin (gamma)		x							x																	Bost et al.	1985	11
fibronectin	x	x			x			x	x				x					x							affinity purification of receptor	Brentani et al.	1988	12
substance P	x	x			x			x	x																receptor histochemy	Bret-Dibat et al.	1994	13
Somatostatin		x			x							x													receptor identification	Campbell-Thompson et al.	1993	14
enkephalin [Met]	x	x	x		x		x						x		x										opioid pep agonist	Carr et al.	1989	15
endorphin (Gamma)	x				x		x								x										affinity purification of opiate receptor	Carr et al.	1986	16
Laminin	X	X		X				x	x		x															Castronovo et al.	1991	17
fibrinogen (a_IIb_b3)	x	x							x																fibrinogen agonist	Derrick et al.	1997	18
ca-binding site	x								x																Ca-mimetic biol effect	Dillon et al.	1991	19
Angiotensin IIa	x	x			x			x	x										x						affinity purification of receptor	Elton et al.	1988	20
tumor necrosis factor		x							x																Kd 10 nM	Fassina et al.	1992	21
Interleukin-2	X	X						x	x																kd micromolar	Fassina et al.	1995	22
interleukin 1 beta		x							x																micromolar range, affinity purification	Fassina & Cassani	1992	23
Endithelin (Big)		x							x																	Fassina et al.	1992	24
Endithelin (Big)		x							x																Kd ~ microMn (monomer) 1/100x (multimer)	Fassina et al.	1992	25
c-Raf		x							x																affinity purification of c-raf	Fassina et al.	1989	26
vasopressin [Arg8]	x	x	x		x			x	x		x														10(-3) to 10(-6) M., affinity purification	Fassina et al.	1989	27
anaphylatoxin (C5a)	x				x			x	x																shock prevention	Fujita et al.	2004	28
Fibrinogen		x						x	x																	Gartner et al.	1991	29
fibrinogen		x							x																inhibirion of platelet aggregation	Gartner & Taylor	1991	30
cystatin C | C4		x							x																	Ghiso et al.	1990	31
angiogenin	x		x			x																			Kd 44 nM, inhibition of neovascularization	Gho & Chae	1997	32
angiogenin	x		x					x	x																blocks tumor-induced angiogenesis	Gho et al.	1997	33
interleukin 1beta	x	x							x																no interaction with X-ray crystallography	Heal et al.	1999	34
angiotensin II		x							x																Kd ~ 5 microM and 70 nM	Holsworth et al.	1994	35
HIV (T20)		x							x																HIV inhibition	Imai et al.	2000	36
interferon-gamma		x		x	x								x	x	x											Johnson et al.	1982	37
antibody (idiotypic,T15)		x							x																	Kang et al.	1988	38
vasopressin [Arg8]	x	x			x			o									o								negative results	Kelly et al.	1990	39
vasopressin [Arg8]	x	x						x	x																	Knigge et al.	1988	40
Insulin	x	x			x			x	x				x		x										Kd of 3 nM	Knutson	1988	41
amyloid beta		x							x																Prevention of aggregate formation	Kwak et al.	2006	42
vasopressin [Arg8]	x	x						x	x																affinity purification of receptor	Lu et al.	1991	43
HIV proteins		x			x																				HIV diagnostic & vaccine development	Ludwig et al.	2006	44
interferon-beta		x							x																Kd=1.89 x 10(-4) - 1.22 x 10(-5) M, biosensors	Luo et al.	2007	45
Prion protein		x			x										x										rceptor recognition	Martins et al.	1997	46
Gastrin		x			x			x	x						x										gastrin antagonism	McGuigan & Campbell	1992	47
Protein	x	x			x			x	x						x										receptor identification	McGuigan JE.	1994	48
angiotensin II		x			x			x	x																inhibition of biol action	Moore et al.	1989	49
LHRH	X	X			x			x	x				x												LHRH antagonists	Mulcahey et al.	1986	50
anaphylatoxin (C5a)		x							x																prevention of anaphylactic shock	Okada et al.	2007	51
La/SSB		x		x	x				x	x											x				autoimmunity	Papamattheou et al.	2004	52
substance P	x	x			x			x	x																receptor purification	Pascual et al.	1989	53
fibronectin		x			x																					Pasqualini et al.	1989	54
chemoattractant		x							x																biol. Inhibitor	Pfister et al.	1989	55
LHRH		x							x																Kd mM	Root-Bernstein & Westall	1986	56
La/SSB		x		x	x				x	x											x				autoimmunity	Routsias et al.	2003	57
La/SSB		x		x	x				x	x											x				autoimmunity	Routsias et al.	2002	58
La/SSB		x		x	x				x	x											x				autoimmunity	Sakarellos-Daitsiotis et al.	2004	59
synthase (nitric oxide)		x						x																		Sautebin et al.	2000	60
interferon-beta		x						x																	affinity purification	Scapol et al.	1992	61
endorphin (beta)	x	x			x			x	x									x								Shahabi et al.	1992	62
ribonuclease S		x						x																	10(-6) M	Shai et al.	1989	63
ribonuclease		x						x																	not saturable binding	Shai et al.	1987	64
Endothelin A	x	x	x					x	x	x															inhibitor	Wu et al.	1997	65

Many experiments (Table 1) indicate that antisense proteins inhibit the biological effects of a protein. This suggests the possibility of *antisense protein therapy*. The P–P/c reaction is in many respects similar to the antigen–antibody reaction, therefore the potential of antisense protein therapy is expected to be similar to the potential of antibody therapy (passive immunization against proteinaceous toxins, such as bacterial toxins, venoms, etc.). However, antisense peptides are much smaller than antibodies (MW as little as ~1000 Da compared to IgG ~155 kDa). This means that antisense proteins are easy to manufacture in vitro;

antibodies are produced in living animals (with non-human species characteristics). However, the small size is expected to have the disadvantage of a higher K_d and a shorter biological half-life.

Immunization with complementary peptides gains antibodies (P/c_ab) as any other protein. These antibodies contain a domain that is similar to the original protein (P) and specifically bind to the receptor of the original protein (P/r). This property is effectively used for affinity purification or immuno-staining of receptors. The P/c_ab is able to mimic or antagonize the in vivo effect of P by binding to its receptor. This property has the desired potential to treat protein-related disease, such as many pituitary gland-related diseases. A vision might be to treat, for example, pituitary dwarfism, with immunization to growth hormone complementary peptide (GH/c) or Type I diabetes with immunization to insulin/c peptide.

REVERSE BUT NOT COMPLEMENTARY SEQUENCES

The biochemical process of transcription and translation is unidirectional, $5'\rightarrow3'$ and reversion does not exist. However, there are many examples of sequences which are present in the genome (in addition to direct reading) even in reverse orientation, and if expressed (in the usual $5'\rightarrow3'$ direction) they produce mRNA and proteins that are virtually reversely transcribed and reversely translated.

An interesting observation is that direct and reverse proteins often have very similar binding properties and related biological effects even if their sequence homology is very low (<20%). For example, growth hormone-releasing hormone (GHRH) and the reverse GNRH specifically bind to the GHRH receptor on rat pituitary cells and to polyclonal anti-GHRH antibody in ELISA and RIA procedures although they share only 17% sequence similarity and they are antagonists in in vitro stimulation of GH RNA synthesis and in vitro and in vivo GH release from pituitary cells [51].

The same phenomenon is observed in complementary sequences. A peptide expressed by complementary mRNA often specifically interacts with proteins expressed by the direct mRNA and it does not matter if they are read in the same or opposite directions. A possible explanation is that many codons are actually symmetrical and have the same meaning in both directions of reading.

The physico-chemical properties of amino acids is preferentially determined by the 2nd (central) codon letter [52] and in that way the physico-chemical pattern of direct and reverse sequences remains the same. In addition, I found that protein structural information is also carried by the 2nd codon letters [53].

CONTROVERSIES REGARDING THE ORIGINAL PROTEOMIC CODES

All proteomic codes before 2006 required perfect complementarity, even if it was noticed that the "biophysical and biological properties of complementary peptides can be improved in a rational and logical manner where appropriate" [36] .

- Expression of the antisense DNA strand was simply not accepted before large scale genome sequences confirmed that genes are about equally distributed on both strands of DNA in every organisms containing dsDNA.
- Spatial complementarity is difficult to imagine between longer amino acid sequences, because the natural, internal folding of proteins will prohibit it in most cases.
- The usual is that residues with the same polarity show attraction to each other, because hydrophobes like the hydrophobe environment as well as lipohobes prefer lipohobe neighbors. Amphipathic interactions seems to be artificial for most chemists.
- Only complementary (but not reversed) sequences were also found effective. This requires $3'\rightarrow5'$ translation, which is normally prohibited.
- Inconsistency of the results; it works for some proteins but not for others; necessity to improve results, e.g., "M-I pair mutagenesis" [36].
- The protein 3D structure and interactions are thought to be arranged on a larger scale than individual amino acids.
- The number of possible amino acid pairs is $20\times20/2=200$. The number of perfect codons is 64, i.e., about a third of the number expected. This means that two-thirds of amino acid pairs are impossible to code by perfectly complementary codons.

Are these amino acid pairs not derived from complementary codons at all?
Are these amino acid pairs derived from imperfectly complementary codons?

DEVELOPMENT OF THE SECOND GENERATION PROTEOMIC CODE

What did we learn about the Proteomic Code during its first 25 years (1981–2006)? My first and most important lesson is that I realize how terribly wrong it was (and is) to believe in scientific dogmas, like sense vs nonsense DNA strands. It is almost unbelievable today that many of us were able to see difference between two perfectly symmetrical and structurally identical strands.

We were able to provide multiple independent and convincing evidence that the concept of the Proteomic Code is valid. At the same time we had to understand that the first concepts – based on perfect complementarity of codons behind interacting amino acids – are imperfect.

There is protein folding information in the nucleic acids – in addition or within the redundant genetic code – but it is unclear how is it expressed and interpreted to form the 3D protein structure.

A major physico-chemical property, the hydropathy of amino acids is coded by the codons. Proteins translated from direct and reverse as well as from complement and reverse-complement strands results in the same hydropathic profile. This is possible only if the amino acid hydropathy is related to the second, central codon letter.

There is a clear indication that some biological information exists in multiple complementary (mirror) copies: DNA–DNA/c→RNA–RNA/c→protein–protein/c→IgG–IgG/c.

Some theoretical considerations and research that led to the suggestion of the 2nd generation Proteomic Codes are now reviewed.

CONSTRUCTION OF A COMMON PERIODIC TABLE OF CODONS AND AMINO ACIDS

The Proteomic Code revitalizes a very old dilemma and dispute represented by Carl Woese and Francis Crick about the origin of the genetic code. Is there any logical connection between any properties of an amino acid on the one hand and any properties of its genetic code on the other?

Carl Woese [54] argued that there was stereochemical matching, i.e., affinity, between amino acids and certain triplet sequences. He therefore proposed that the genetic code developed in a way that was very closely connected to the development of the amino acid repertoire, and that this close biochemical connection is a fundamental of specific protein–nucleic acid interactions.

Crick [55] considered that the basis of the code might be a "frozen accident", with no underlying chemical rationale. He argued that the canonical genetic code evolved from a simpler primordial form that encoded fewer amino acids. The most influential form of this idea, "code co-evolution," proposed that the genetic code co-evolved with the invention of biosynthetic pathways for new amino acids [56].

A periodic table of codons has been designed where the codons are in regular locations. The table has four fields (16 places in each) one with each of the four nucleotides (A, U, G, C) in the central codon position. Thus, AAA (lysine), UUU (phenylalanine), GGG (glycine) and CCC (proline) are positioned in the corners of the fields as the main codons (and amino acids).

They are connected to each other by six axes. The resulting nucleic acid periodic table shows perfect axial symmetry for codons. The corresponding amino acid table also displaces periodicity regarding the biochemical properties (charge and hydropathy) of the 20 amino acids, and the position of the stop signals. Table 2 emphasizes the importance of the central nucleotide in the codons, and predicts that purines control the charge while pyrimidines determine the polarity of the amino acids.

In addition to this correlation between the codon sequence and the physico-chemical properties of the amino acids, there is a correlation between the central residue and the chemical structure of the amino acids. A central uridine correlates with the functional group $-C(C)_2-$; a central cytosine correlates with a single carbon atom, in C_1 position; a central adenine coincides with the functional groups $-CC=N$ and $-CC=O$, and finally a central guanine coincides with the functional groups $-CS$, $-C=O$, and $C=N$, and with the absence of a side chain (glycine). (Table III)

I interpret these results as a clear cut answer for the Woese vs Crick dilemma: there is a connection between the codon structure and the properties of the coded amino acids. The second (central) codon base is the most important determinant of the amino acid property. It explains why the reading orientation of translation has so little effect on the hydropathy profile of the translated peptides. Note, that 24 of 32 codons (U or C in the central position) code apolar (hydrophobe) amino acids, while only 1 of 32 codons (A or G in the central

position) codes non-apolar (non-hydrophobe, charged or hydrophilic) amino acids. It explains why complementary amino acid sequences have opposite hydropathy, even if the binary hydropathy profile is the same.

Table II. Common Periodic Table of Codons and Amino Acids

Table II. Continued

PERIODIC TABLE OF AMINO ACIDS

Table III. Effect of a Single Codon Residue on the Structure of the Amino Acids

THE PHYSICO-CHEMICAL COMPATIBILITY OF AMINO ACIDS IN THE PROTEOMIC CODE

The complementary coding of two amino acids is not a guarantee per se for the special co-location (or interaction) of these amino acids within the same or between two different peptides. Some kind of physico-chemical attraction is also necessary. The most fundamental properties to consider are, of course, the size, charge and hydropathy. Mekler and myself suggested size compatibility [9,10,11,20], obviously under the influence of the known size

complementarity of the Watson–Crick base pairs. Blalock emphasizes the importance of hydropathy, better say amphipathy (which makes some scientists immediately antipathic). Hydrophobe residues like other hydrophobe residues and hydrophilic residues like hydrophilic residues. Hydrophyl likes hydrophobe like water likes oil.

Visual studies on the 3D structures of proteins give some ideas how interacting interfaces look (Figure 6):

−   the interacting (collocating) sequences are short (1–10 amino acid long);
−   the interacting (collocating) sequences are not continuous; there are many mismatches;
−   the orientation of the collocating residues is often not the same (not parallel);
−   the contact between collocating residues might be side-to-side or top-to-top.

This is clearly a different picture than the base-pair interactions in a dsDNA spiral. Alpha-helices and beta-sheets are regular structures, which make their amino acid residues periodically ordered. Many residues are parallel to each other and W-C-like interactions are not impossible. But is it really the explanation for specific residue interactions?
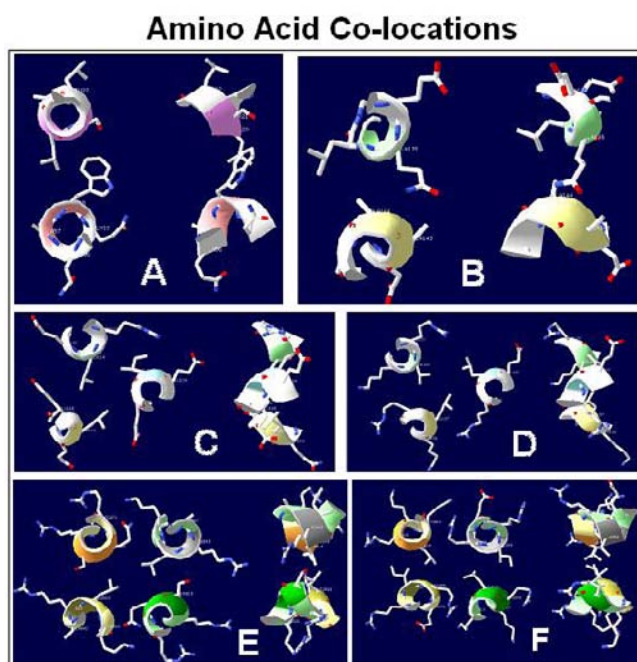
Figure 6. Amino acid co-locations. Randomly selected amino acid contacts from real proteins. The interactions between amino acid residues from 2 (A, B) 3 (C, D) and 4 (E, F) parallel alpha helices are perpendicular to the peptide backbones (helices). The orientation of residues show large variation; some are located side-by-side, others are end-to-end.

SEQX

The interacting residues of protein and nucleic acid sequences are close to each other; they are co-located. Structure databases (e.g., Protein Data Bank, PDB and Nucleic Acid Data

Bank, NDB) contain all the information about these co-locations; however it is not an easy task to penetrate this complex information. We developed a JAVA tool, called SeqX, for this purpose [57].

The SeqX tool is useful to detect, analyze and visualize residue co-locations in protein and nucleic acid structures. The user:

(a) selects a structure from PDB;
(b) chooses an atom that is commonly present in every residue of the nucleic acid and/or protein structure(s);
(c) defines a distance from these atoms (3–15 Å).

The SeqX tool then detects every residue that is located within the defined distances from the defined "backbone" atom(s); provides a dot-plot-like visualization (residues contact map), and calculates the frequency of every possible residue pair (residue contact table) in the observed structure. It is possible to exclude ±1−10 neighbor residues in the same polymeric chain from detection, which greatly improves the specificity of detections (up to 60% when tested on dsDNA). Results obtained on protein structures show highly significant correlations with results obtained from the literature ($p<0.0001$, $n=210$, four different subsets). The co-location frequency of physico-chemically compatible amino acids is significantly higher than is calculated and expected in random protein sequences ($p<0.0001$, $n=80$) (Figure 7).
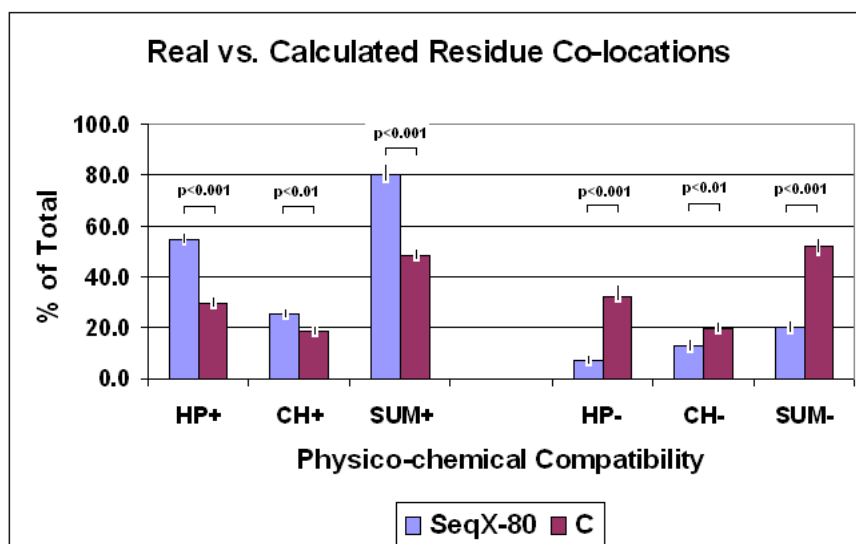
Figure 7. Real vs calculated residue co-locations (from [57] {. The relative frequency of real residue co-locations was determined by SeqX in 80 different protein structures and compared to the relative frequency of calculated co-locations in artificial, random protein sequences (C). The 200 possible residue pairs provided by the 20 amino acids were grouped into 4 subgroups regarding their physico-chemical compatibility to each other, i.e., favored (+) and un-favored (–) regarding hydrophobicity and charge. (HP+, hydrophobe–hydrophobe and lipophobe–lipophobe; HP–, hydrophobe–lipophobe; CH+, positive–negative and hydrophobe–charged; CH–: positive–positive and negative–negative and lipophobe–charged interactions). The bars represent the mean±SEM ($n=80$ for real structures and $n=10$ for artificial sequences). Student's t-test was applied to evaluate the results.

These results gave a preliminary confirmation of our expectation that physico-chemical compatibility exists between co-locating amino acid pairs. Our findings do not support any significant dominance of amphipathic residue interactions in the examined structure.

AVAILABILITY AND REQUIREMENTS

Available from http://janbiro.com/Downloads.html SeqX, Java J2SE Runtime Environment 5.0 (available from [see Additional file 1] http://www.sun.com) and at least a 1 GHz processor and with a minimum 256 Mb RAM. Source codes are available from the authors.

AMINO ACID SIZE, CHARGE, HYDROPATHY INDICES AND MATRICES FOR PROTEIN STRUCTURE ANALYSIS

It was necessary to have a closer look at the physico-chemical compatibility of co-locating amino acids [58].
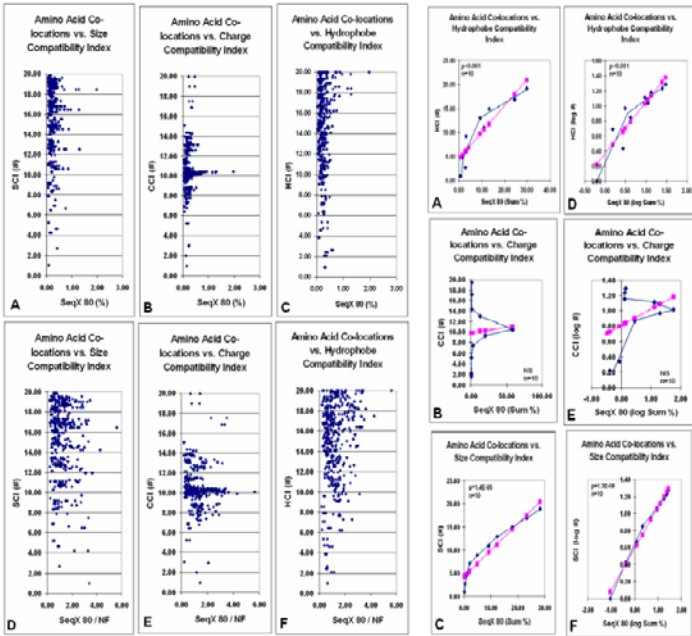
Figure 8. (a) Amino acid co-locations vs size, charge, and hydrophobe compatibility indexes (modified from [59]. Average propensity of the 400 different amino acid co-locations in 80 different protein structures (SeqX 80) are plotted against size, charge and hydrophobe compatibility indexes (SCI, CCI, HCI). The original "row" values are indicated in (A–C). The SeqX 80 values were corrected by the co-location values, which are expected only by chance in proteins where the amino acid frequency follows the natural codon frequency (NF) (D–F). (b) Amino acid co-locations vs size, charge, and hydrophobe compatibility indexes in major subgroups (modified from [59]. Data presented in Figure 1 were divided into subgroups and summed (Sum). The group averages are connected by the blue lines while the pink symbols and lines indicate the calculated linear regression.

Matrix Representation of Residue Co-location in a Protein Structure (1AP6)
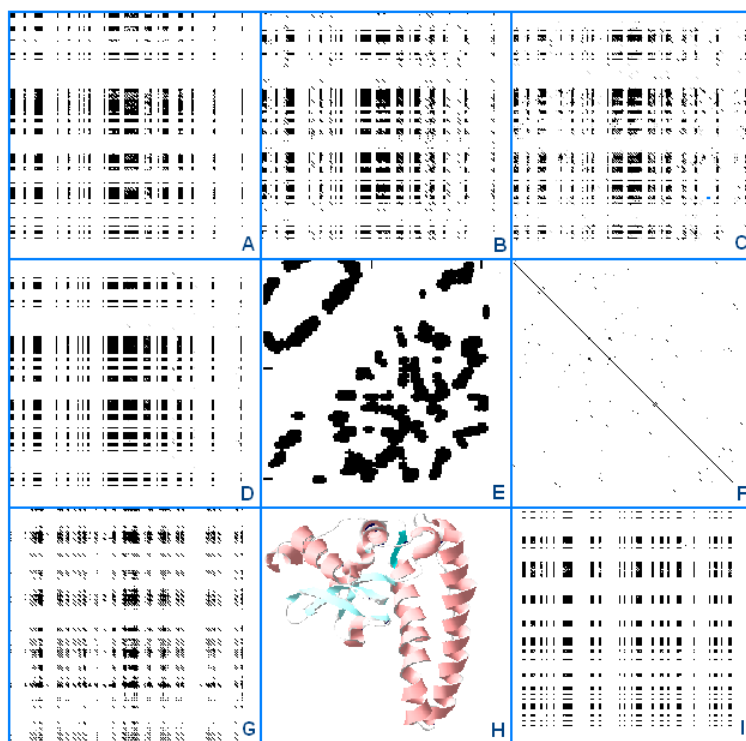
A: SCM, B: CCM, C: HCM, D:SCHM, E: SeqX, F: Blosum62, G: NFM, H: DeepView, I: Random Prot.

Figure 9. Matrix representation of residue co-locations in a protein structure (1AP6) (modified from [58.]). A protein sequence (1AP6) was compared to itself with DOTLET using different matrices, SCM (A), CCM (B), HCM (C), the combined SCHM (D) and NFM (G) and Blosum62 (F). Comparison of randomized 1AP6 using SCHM is seen in (I). The 2D (SeqX Residue Contact Map) and 3D (DeepView/Swiss-PDB Viewer) views of the structure are illustrated in (E) and (H). The black/gray parts of the dot-plot matrices indicate the respective compatible residues, except the Blosum62 comparison (F), where the diagonal line indicates the usual sequence similarity. The dot-plot parameters are otherwise the same for all matrices.

We indexed the 200 possible amino acid pairs for their compatibility regarding the three major physico-chemical properties – size, charge and hydrophobicity – and constructed size, charge and hydropathy compatibility indices (SCI, CCI, HCI) and matrices (SCM, CCM, HCM).

Each index characterized the expected strength of interaction (compatibility) of two amino acids by numbers from 1 (not compatible) to 20 (highly compatible). We found statistically significant positive correlations between these indices and the propensity for amino acid co-locations in real protein structures (a sample containing total 34,630 co-locations in 80 different protein structures): for HCI, $p<0.01$, $n=400$ in 10 subgroups; for SCI, $p<1.3E–08$, $n=400$ in 10 subgroups; for CCI, $p<0.01$, $n=175$). Size compatibility between residues (well known to exist in nucleic acids) is a novel observation for proteins (Figure 8).

We tried to predict or reconstruct simple 2D representations of 3D structures from the sequence using these matrices by applying a dot-plot-like method. The location and pattern of

the most compatible subsequences was very similar or identical when the three fundamentally different matrices were used, which indicates the consistency of physico-chemical compatibility. However, it was not sufficient to choose one preferred configuration between the many possible predicted options (Figure 9).

Indexing of amino acids for major physico-chemical properties is a powerful approach to understanding and assisting protein design. However, it is probably insufficient itself for complete *ab initio* structure prediction.

ANFINSEN'S THERMODYNAMIC PRINCIPLE AND THE PROTEOMIC CODE

The existence of physico-chemical compatibility of co-locating amino acids already on the residue level is, of course, a necessary support for the Proteomic Code. At the same time, it raises the possibility that the protein structure might be predicted from the primary amino acid sequence (de novo, ab initio prediction) and the location of physico-chemically compatible amino acid residues in the sequence. This idea is in line with a dominating statement about protein folding. Anfinsen's thermodynamic principle states that all information necessary to form a 3D protein structure is present in the protein sequence [60].

Attempts were made to use the three different matrices in a dot plot to predict the place and extent of the most likely residue co-locations. This semi-quantitative method indicated that the three very different matrices located very similar residues and subsequences as potential co-location places. No single diagonal line was seen in the dot-plot matrices, which is the expected signature of sequence similarity (or compatibility in our case). Instead, block-like areas indicated the place and extent of predicted sequence compatibilities. It was not possible to reconstruct a real map of any protein 2D structure (Figure 9) [61].

This experience with the indexes provides arguments for as well as against Anfinsen's theorem. The clear-cut action of basic physico-chemical laws at residue level is well in line with the lowest free energy requirement of the law of entropy. Furthermore, this obvious presence of physico-chemical compatibility is easy to understand, even from an evolutionary perspective. In evolution, sequence changes more rapidly than structure; however, many sequence changes are compensatory and preserve local physico-chemical characteristics. For example, if, in a given sequence, an amino acid side chain is particularly bulky with respect to the average at a given position, this might have been compensated in evolution by a particularly small side chain in a neighboring position, to preserve the general structural motif. Similar constraints might hold for other physico-chemical quantities such as amino acid charge or hydrogen bonding capacity [62].

We were not able to reconstruct any structure using our indexes. There are massive arguments against Anfinsen's principle:

1) The connection between primary, secondary and tertiary structure is not strong, i.e., in evolution, sequence changes more rapidly than structure. Structure is often conserved in proteins with similar function even when sequence similarity is already lost (low structure specificity to define a sequence). Identical or similar sequences often result in different structures (low sequence specificity to define a structure).

2) An unfolded protein has a vast number of accessible conformations, particularly in its side chains of residues. Entropy is related to the number of accessible conformations. This problem is known as the Levinthal paradox [63].

3) The energy profile characteristics of native and designed proteins are different. Native proteins usually show a unique and less stable profile, while designed proteins show lower structural specificity (many different possible structures) but high stability [64].

4) The entropy minimum is a statistical minimum. The conformation entropy change of the whole molecule is the sum of local (residue level) conformation entropy changes and it permits the co-existence of many different local conformation variations. It is doubtful whether structural variability (heterogeneity, instability) is compatible with the function (homogeneity, stability) of a biologically active molecule.

The present experiments do not decide the "fate" of Anfinsen's dogma; however, they show that the number of possible co-locating places is too large, and searching this space poses a daunting optimization problem. It is not realistic to expect the ab initio prediction of only one single structure from one primary protein sequence. The development of a prediction tool for protein structure (like an mfold for nucleic acids [65]{4 Zuker M. 2003}), which provides only a few hundred most likely (thermodynamically most optimal) structure suggestions per protein sequence, seems to be closer. It is likely that SCM, CCI and HCM (or similar matrices) will be essential elements of these tools.

Additional folding information might be necessary (in addition to that carried in the protein primary sequence) to be able to create a unique protein structure. Such information is suspected to be present in the redundant genetic code [66,67,68] and chaperons [39,70,71].

PROTEIN STRUCTURE AND THE FUNCTIONAL ASYMMETRY OF THE CODONS

The consequence of codon redundancy is that the third codon base is ambiguously defined; the same amino acid might be coded by more than one codon differing at the wobble base positions. These codons are called synonymous codons.

The Common Periodic Table of Codons and Amino Acids revealed another difference between codon bases, namely, that the second codon base is most significant in determining the physico-chemical properties of amino acids.

I wanted to find more evidence supporting the connection between amino acid (protein) and codon (nucleic acid structure). All these facts provide vital evidence for the Proteomic Code.

The protein folding problem has been one of the grand challenges in computational molecular biology. The problem is to predict the native three-dimensional structure of a protein from its amino acid sequence. It is widely believed that the amino acid sequence contains all the necessary information to make up the correct three-dimensional structure, since protein folding is apparently thermodynamically determined; i.e., given a proper environment, a protein will fold up spontaneously. This is called Anfinsen's thermodynamic principle [7].

The thermodynamic principle has been confirmed many times on many different kinds of proteins in vitro. Critics says that the in vivo chemical conditions are different from those in vitro, the correct folding is determined by interactions with other molecules (chaperons, hormones, substrate, etc.) and protein folding is much more complex than re-naturation of denatured poly amino acids. The fact that many naturally occurring proteins fold reliably and quickly to their native state, despite the astronomical number of possible configurations, has come to be known as Levinthal's Paradox [8].

Anfinsen's principle was formulated in the 1960s using purely chemical experiments and a lot of intuition. Today, we have a lot of sequences and structures available to establish a logical and understandable link between sequence, structure and function. But it is still not possible to correctly predict the structure (or a range of possible structures) purely from the sequence, ab initio and in silico [9].

There are two potential, external sources of additional and specific protein folding information: (a) the chaperons (other proteins that assist in the folding of proteins and nucleic acids [71]; and (b) the protein coding nucleic acid sequences themselves (which are templates of the protein syntheses, but are not defined as chaperons).

The idea that the nucleotide sequence itself could modulate translation and hence affect co-translational folding and assembly of proteins has been investigated in a number of studies [72,73]. Studies on the relationships between synonymous codon usage and protein secondary structural units are especially popular [68,74,75]. The genetic code is redundant (61 codons code 20 amino acids) and as many as 6 synonymous codons can code the same amino acid (Arg, Leu, Ser). The "wobble" base has no effect on the meaning of most codons but still the codon usage (wobble usage) is not randomly defined [76,77] and there are well known, stable species-specific differences in the codon usage. It seems to be logical to search for some meaning (biological purpose) of the wobble bases and try to associate them with protein folding.

Another observation concerning the code redundancy dilemma is that there is a widespread selection (preference) for local RNA secondary structure in protein coding regions [78]. A given protein can be encoded by a large number of distinct mRNA species, potentially allowing mRNAs to simultaneously optimize desirable RNA structural features in addition to their protein coding function. The immediate question is whether there is some logical connection between the possible, optimal RNA structures and the possible, optimal biologically active protein structures.

Single-stranded RNA molecules can form local secondary structures through the interactions of complementary segments. W-C base pair formation lowers the average free energy, dG, of the RNA and the magnitude of change is proportional to the number of base pair formations. Therefore the free folding energy (FFE) is used to characterize the local complementarity of nucleic acids [78]. The free folding energy is defined as FFE=(d$G_{shuffled}$ − dG_{native})/L×100, where L is the length of the nucleic acid, i.e., free energy difference between native and shuffle (randomized) nucleic acids per 100 nucleotides. Higher positive values indicate stronger bias towards secondary structure in the native mRNA, and negative values indicate bias against secondary structure in the native mRNA.

We used a nucleic acid secondary structure predicting tool, mfold [65] to obtain dG values and the lowest dG was used to calculate the FFE. mfold also provided the folding energy dot-plots, which are very useful to visualize the energetically most favored structures in a 2D matrix.

A series of JAVA tools were used: SeqX to visualize the protein structures in 2D as amino acid residue contact maps [57]; SeqForm for selection of sequence residues in predefined phases (every third in our case) [79]; SeqPlot for further visualization and statistical analyses of the dot-plot views [80]; Dotlet as a standard dot-plot viewer [81]. Structural data were downloaded from PDB [82], NDB [83], and from a wobble base oriented database called Integrated Sequence–Structure Database (ISSD) [84].

Structures were generally randomly selected regarding species and biological function (a few exceptions are mentioned below). Care was taken to avoid very similar structures in the selections. A propensity for alpha helices was monitored during selection and structures with very high and very low alpha helix content were also selected to make sure of a wide range of structural representation. Linear regression analyses and Student's *t*-tests were used for statistical analyses of the results. Observations were made on human peptide hormone structures. This group of proteins is very well defined and annotated, the intron–exon boundaries are known and even intron data are easily accessible. The coding sequences were phase separated by SeqForm into three subsequences, each containing only the 1st, 2nd or 3rd letters of the codons. Similar phase separation was made for intronic sequences immediately before and after the exon. There are, of course, no known codons in the intronic sequences, therefore we continued the same phase that we applied for the exon, assuming that this kind of selection is correct and maintained the name of the phase denotation even for non-coding regions. Subsequences corresponding to the 1st and 3rd codon letters in the coding regions had significantly higher FFEs than subsequences corresponding to the 2nd codon letters. No such difference was seen in non-coding regions (Figure 10).
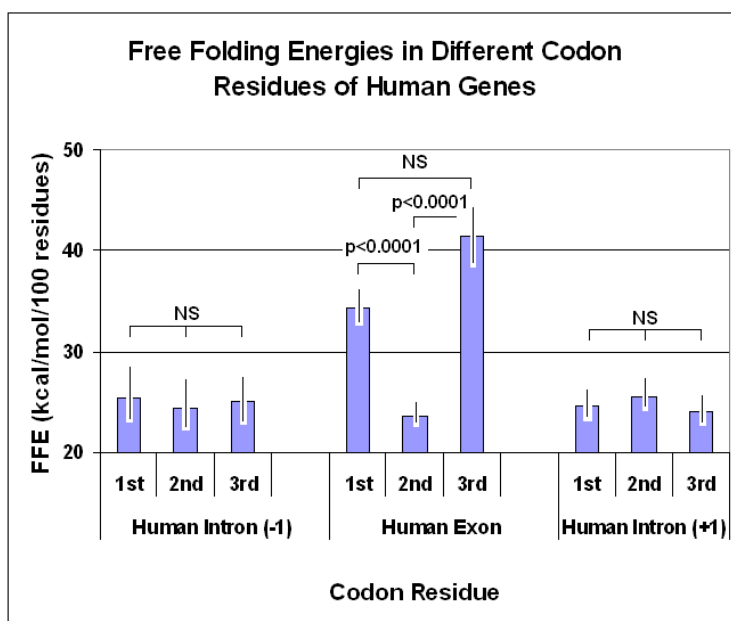
Figure 10. Free folding energies (FFE) in different codon residues of human genes. The coding sequences (exons) of 18 human hormone genes and the preceding (−1) and following (+1) sequences (introns) were phase separated into three subsequences each corresponding to the 1st, 2nd and 3rd codon positions in the coding sequence. The dG values were determined by mfold and the FFE was calculated. Each bar represents the mean±SEM, n=18.
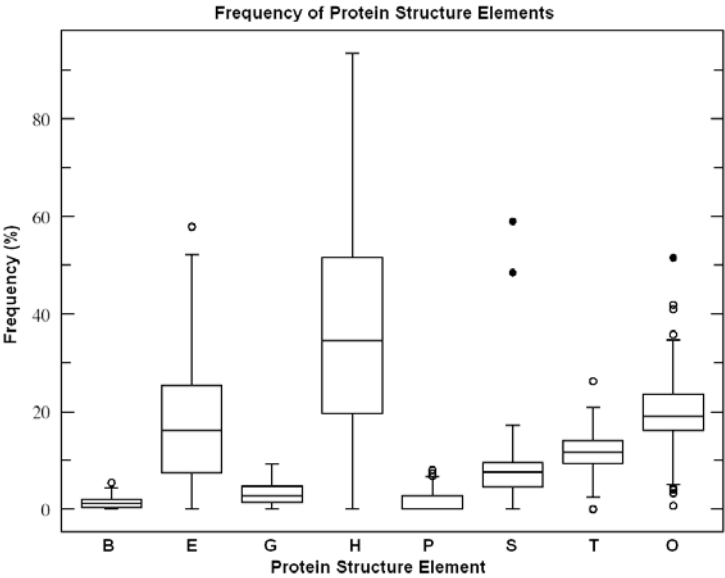
Figure 11. Frequency of protein structure elements. Box plot representation of protein secondary structure elements in 81 structures. *L*=317±20 (mean±SEM, *n*=81). Secondary structure codes: H, alpha helix; B, residue in isolated beta bridge; E, extended strand, participates in beta ladder; G, 3-helix (3/10 helix); I, 5 helix (pi helix); P, polyproline type II helix (left-handed); T, hydrogen bonded turn; S, bend.
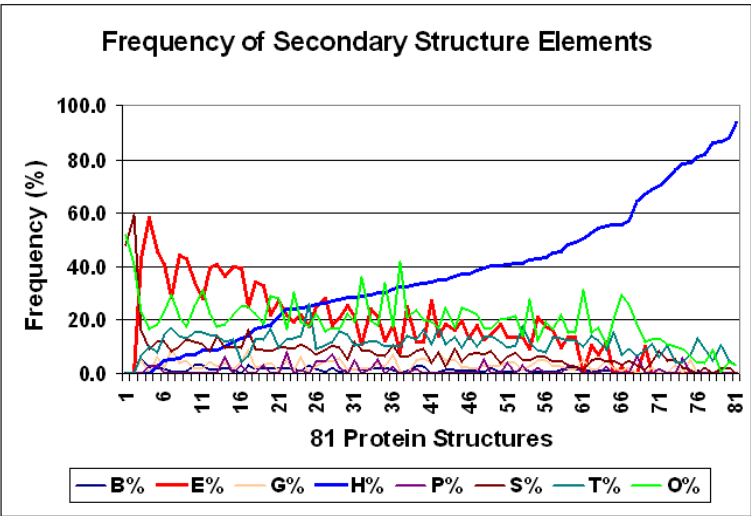
Figure 12. Frequency of secondary structure elements. The propensity of different structural elements in 81 different proteins is shown. *L*=317±20 (mean±SEM, *n*=81). Secondary structure codes: H, alpha helix; B, residue in isolated beta bridge; E, extended strand, participates in beta ladder; G, 3-helix (3/10 helix); I, 5 helix (pi helix); P, polyproline type II helix (left-handed); T, hydrogen bonded turn; S, bend.

In a larger selection of 81 different protein structures, the corresponding protein and coding sequences were used to extend the observations. These 81 proteins represented different (randomly selected) species and different (also randomly selected) protein functions and therefore the results might be regarded as more generally valid. The propensity of

different secondary structure elements was recorded (as annotated in different databases) (Figure 11). The proportion of alpha helices varied from 0 to 90% in the 81 proteins and showed a significant negative correlation to the proportion of beta sheets (Figures 12 and 13).
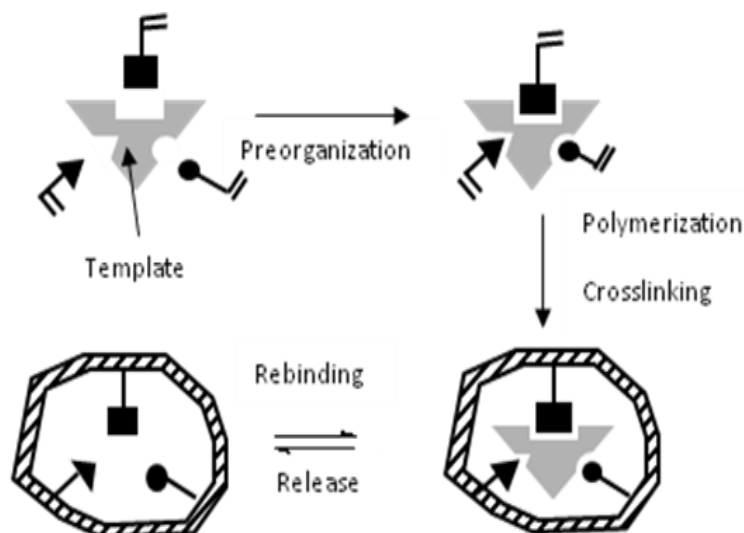
Figure 13. Correlation between two main structural elements in proteins. Data were taken from Figure 3 (H, alpha helix; E, beta sheet).

The original observation made on human hormone proteins, that significantly more free folding energy is associated with the 1st and 3rd codon residues than with the 2nd was confirmed on a larger and more heterogeneous protein selection. A significant difference showed up even between the 1st and 3rd residues in this larger selection (Figure 14).
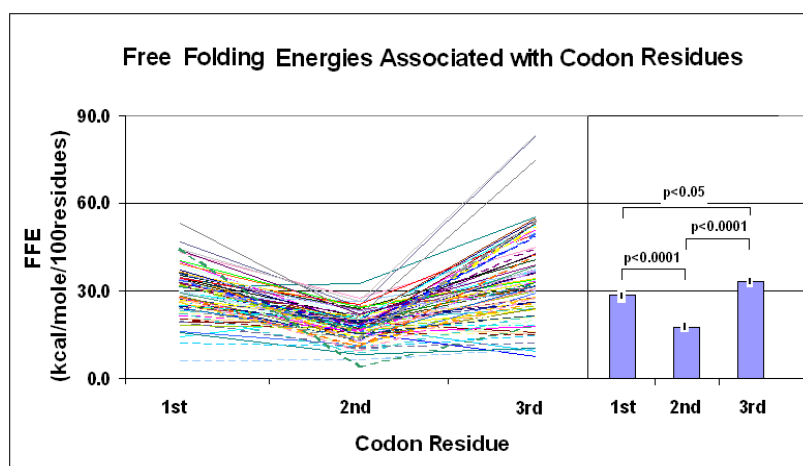
Figure 14. Free folding energies associated with codon residues (Free folding energies (FFE) were determined in phase-selected subsequences of 81 different protein coding nucleic acids. The lines indicate individual values (left part of the figure), while the bars (right part of the figure) indicate the mean±SEM (n=81).

There is a correlation between the protein structure and the FFE associated with codon residues. The correlation is negative between FFE associated with the 2nd (middle) codon residues and the alpha helix content of the protein structure. The correlation is especially significant when the FFE ratios are compared to the helix/sheet ratios (Figures 15 and 16).
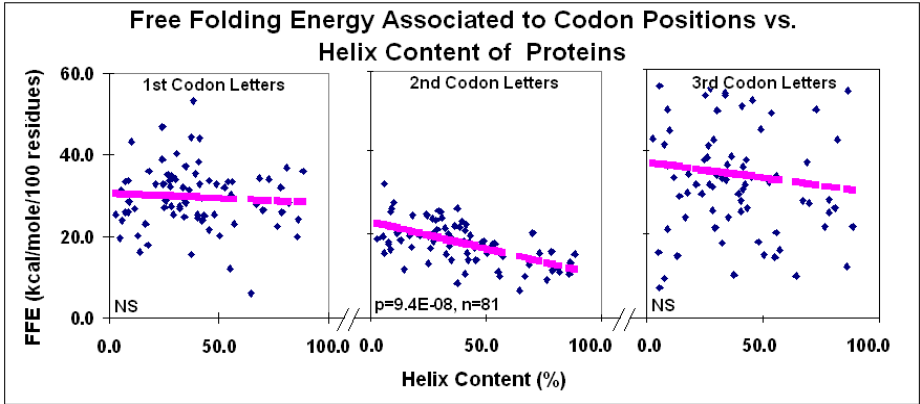
Figure 15. Free folding energy associated with codon positions vs helix content of proteins. Linear regression analyses; pink symbols represent the linear regression line.
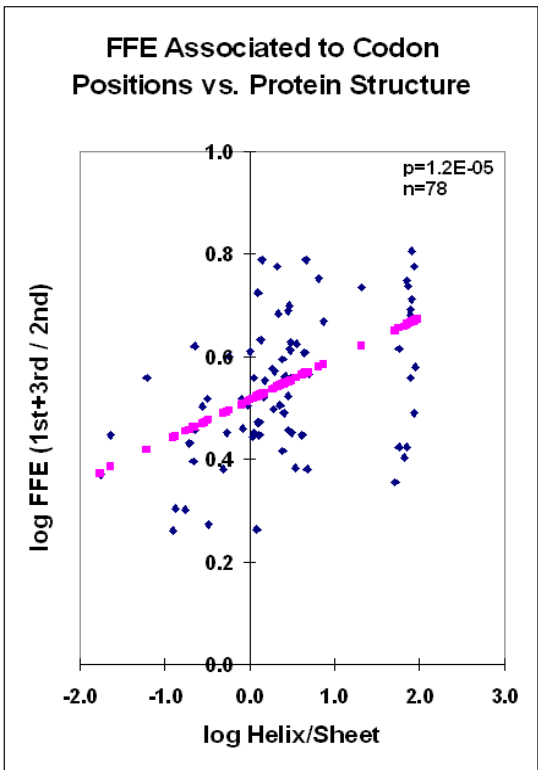
Figure 16. FFE associated with codon positions vs protein structure. Same data as in Figure 6 after calculating ratios and log transformation. Linear regression analyses; pink symbols represent the linear regression line.

Figure 17. Correlation between alpha helix content of protein structure and other protein characteristics. The alpha helix content of 80 protein structures was compared to the frequency of other major structural elements (A,B), the frequency of individual amino acids (C) and the frequency of charged and hydrophobe residues (D,E). (A) The correlation between helix (H), beta sheet (S) and turn (T); (B) the proportions between the sum of helices (SH), beta strands (SS), turns (ST) and all other structural elements (TO). (D) The proportion between the sums of apolar (S_Ap), polar (S_Pol), negatively charged (S_Neg) and positively charged (S_Poz) amino acids. (E) The linear regression analyses correlation between helix content and the percentage of polar+apolar (Polarity) and positively+negatively charged (Charge) residues.

The alpha helix is the most abundant structure element in proteins. It shows negative correlation to the frequency of the second most prominent protein structure, the beta sheet. The propensity of some amino acids and the major physico-chemical characteristics (charge and polarity) shows significant correlation (positive or negative) to this structural feature. We include statistical analyses of alpha helix content and other protein characteristics to show the complexity behind the term "alpha helix" and to show the insecurity in interpreting any correlation to this structural feature (Figures 17 and 18). Detailed analyses of these data are outwith the scope of this review.

Higher FFE in subsequences of 1st and 3rd codon residues than in the 2nd indicates the presence of a larger number of complementary bases at the right positions of these subsequences. However, this might be the case only because the first and last codons form simpler subsequences and contain longer repeats of the same nucleotide than the 2nd codons.

This would not be surprising for the 3rd (wobble) base but would not be expected for the 1st residue, even though it is known that the central codon letters are the most important to distinguish between amino acids (as shown in the in the Common Periodic Table of Codons and Amino Acids [52]. It is more significant to see that the FFEs in 1st and 3rd residues are additive and together they represent the entire FFE of the intact mRNA (Figure 19).

Correlation Between Frequency of Individual Amino Acids and the Main Secondary Structure Elements in Proteins

	A%	C%	D%	E%	F%	G%	H%	I%	K%	L%	M%	N%	P%	Q%	R%	S%	T%	V%	W%	Y%
HELIX (%)	3.E-04	1.E-04	2.E-01	1.E-05	2.E-06	2.E-06	4.E-01	2.E-01	1.E-04	9.E-04	2.E-02	1.E-01	2.E-06	1.E-04	8.E-02	2.E-01	3.E-03	4.E-03	1.E-02	6.E-02
HELIX CORRELATION	POS	NEG	NS	POS	NEG	NEG	NS	NS	POS	POS	POS	NS	NEG	POS	NS	NS	NEG	NEG	NEG	NS
SHEET (%)	2.E-03	6.E-05	6.E-02	3.E-05	5.E-06	6.E-06	5.E-01	3.E-01	1.E-03	5.E-05	4.E-02	6.E-02	9.E-02	5.E-03	4.E-02	5.E-02	8.E-04	8.E-03	2.E-01	4.E-02
SHEET CORRELATION	NEG	POS	NS	NEG	POS	POS	NS	NS	NEG	NEG	NEG	NS	NS	NEG	NEG	NS	POS	POS	NS	POS
TURN (%)	9.E-01	2.E-01	1.E-01	2.E-01	1.E-01	2.E-03	5.E-01	3.E-01	7.E-02	4.E-02	6.E-01	7.E-01	1.E-03	1.E-01	1.E-01	3.E-01	9.E-02	9.E-02	1.E-01	5.E-01
TURN CORRELATION	NS	NS	NS	NS	NS	POS	NS	NS	NS	NEG	NS	NS	POS	NS	NS	NS	NS	NS	NS	NS

A-Y: One letter code of amino acid, POS: positive correlation, NEG: negativer correlation, NS: not significant, linear regression analyses, n=79

Figure 18. Correlation between frequency of individual amino acids and the main secondary structure elements in proteins. See text for explanation.
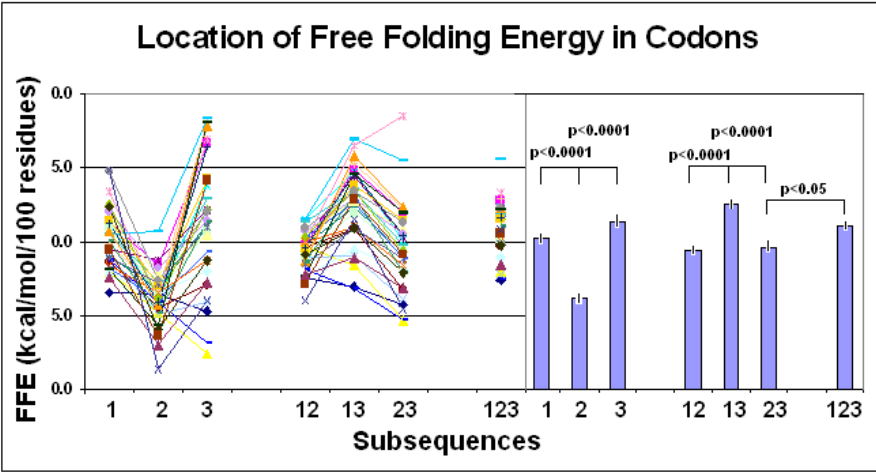
Figure 19. Location of free folding energy in codons. Free folding energies (FFE) were determined in phase-selected subsequences of 31 different protein coding nucleic acids. The original intact RNA contained the intact three-letter codons (123). Subsequences were constructed by periodical removal of one letter from the codon and maintaining the other two (12, 13, 23) or removing two letters and maintaining only one (1, 2, 3). The lines indicate individual values (left), while the bars (right) indicate the mean±SEM (n=31).

Higher FFE at the 1st and 3rd codon positions than at 2nd indicates that the number of complementary bases (a-t and g-t) is higher in the 1st and 3rd subsequences than in the second. This is possible only if more complementers are in 1-1, 1-3, 3-1, 3-3 position pairs than in 1-2, 2-1, 2-3, 3-2 position pairs. We wanted to know whether the 1-1, 3-3 (complement) or the 1-3, 3-1 (reverse-complement) pairing is more predominant.

The length of phase-separated nucleic acid subsequences (l) is a third of the original coding sequence (L). The number of different residues (a, t, g, and c) varies at different codon positions (1, 2, 3).

$$a1+u1+g1+c1=a2+t2+g2+c2=a3+t3+g3+c3=l=L/3$$

The highest number of complementary pairs might occur in the 1st subsequence if

$a1=t1$, $g1=c1$ and $a1/t1=g1/c1=1$

If, for example, $a1>t1$, $g1=c1$ an excess of unpaired a1 occurs and $a1/t1>g1/c1=1$ and the possible FFE in subsequence 1 will be less. Following the same logic for other pairs in other subsequences we can conclude that any deviation from $a/t=g/c=1$ is suboptimal regarding the FFE. Counting the different residue ratios and combinations indicates that the optima are obtained if the residues in the first position form W-C pairs with residues at the third positions (1–3) and vice versa (3–1). This is consistent with the expectation that mRNA will form local loops, in which the direction of more or less double stranded sequences is reversed and (partially) complemented (Figure 20).
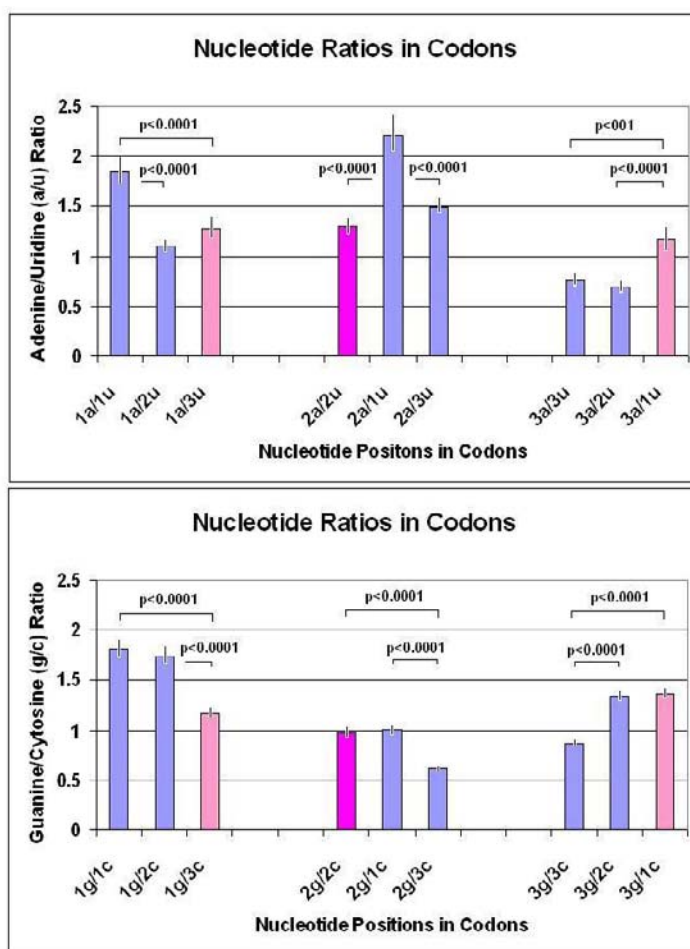
Figure 20. Nucleotide ratios in codons. The number of the 4 different nucleotide bases was counted at the 1st, 2nd and 3rd codon positions in 30 different protein coding RNA sequences. The ratio of the Watson–Crick pairs at different codon positions are indicated by bars (±SEM, n=30). Ideally, the ratio of complementary base pairs is ~1.0. This ideal situation was mostly satisfied when one of the complementary bases was located at codon position 1 with the other at codon position 3 (pink) or both complements at codon position 2 (violet).

COMPARISON OF THE PROTEIN AND MRNA SECONDARY STRUCTURES

The partial (suboptimal) reverse complementarity of codon-related positions in nucleic acids suggested some similarity between protein structures and the possible structures of the coding sequences. This possibility was examined by visual comparison of 16 randomly selected protein residue contact maps and the energy dot-plots of the corresponding RNAs. We could see similarities between the two different kinds of maps (Figure 21). However, this type of comparison is not quantitative and statistical evaluation is not directly possible.
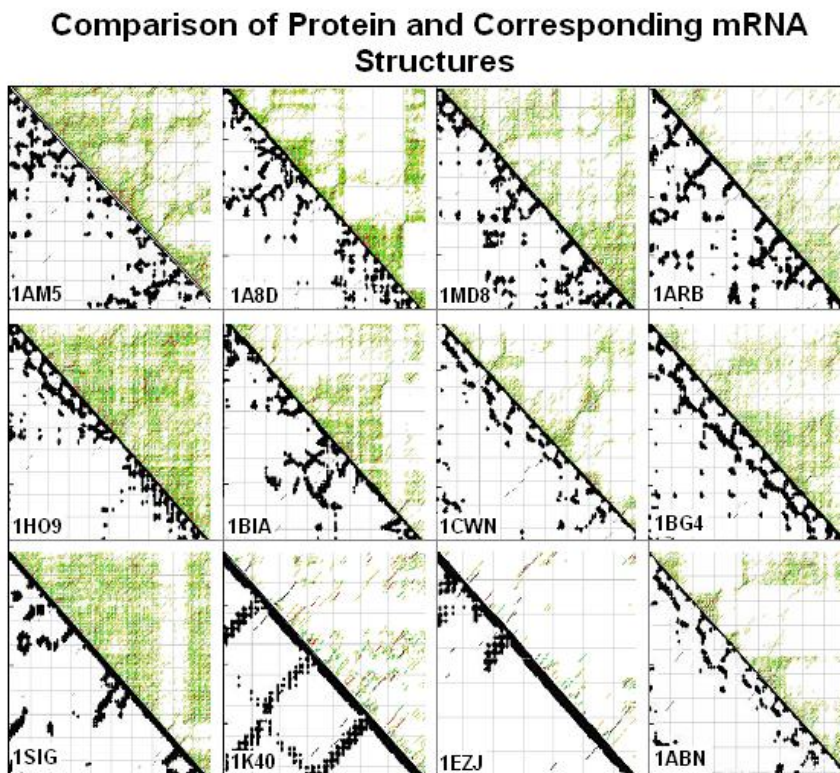
Figure 21. Comparison of protein and corresponding mRNA structures (modified from [59]). Residue contact maps (RCM) were obtained from the PBD files of protein structures using the SeqX tool (left triangles). Energy dot-plots (EDP) for the coding sequences were obtained using the mfold tool (right triangles). The two kinds of maps were aligned along a common left diagonal axis to make an easy visual comparison of the different kinds of representation possible. The black dots in the RCMs indicate amino acids that are within 6 A of each other in the protein structure. The colored (grass-like) areas in the EDPs indicate the energetically mostly likely RNA interactions (color code in increasing order: yellow, green red, black).

Another similar, but still not quantitative, comparison of protein and coding structures was performed on four proteins that are known to have very similar 3D structures but their primary structure (the sequence) is less than 30% similar, as well as the sequence of their mRNA. These four proteins are examples of the fact that the tertiary structure of proteins is much more conserved than the amino acid sequence. We asked the question whether this is true for the RNA structures and sequence. We found that there are signs of conservation even

of the RNA secondary structure (as indicated by the energy dot-plots) and there are similarities between the protein and nucleic acid structures (Figure 22).

Comparison of the Protein and mRNA Secondary Structures
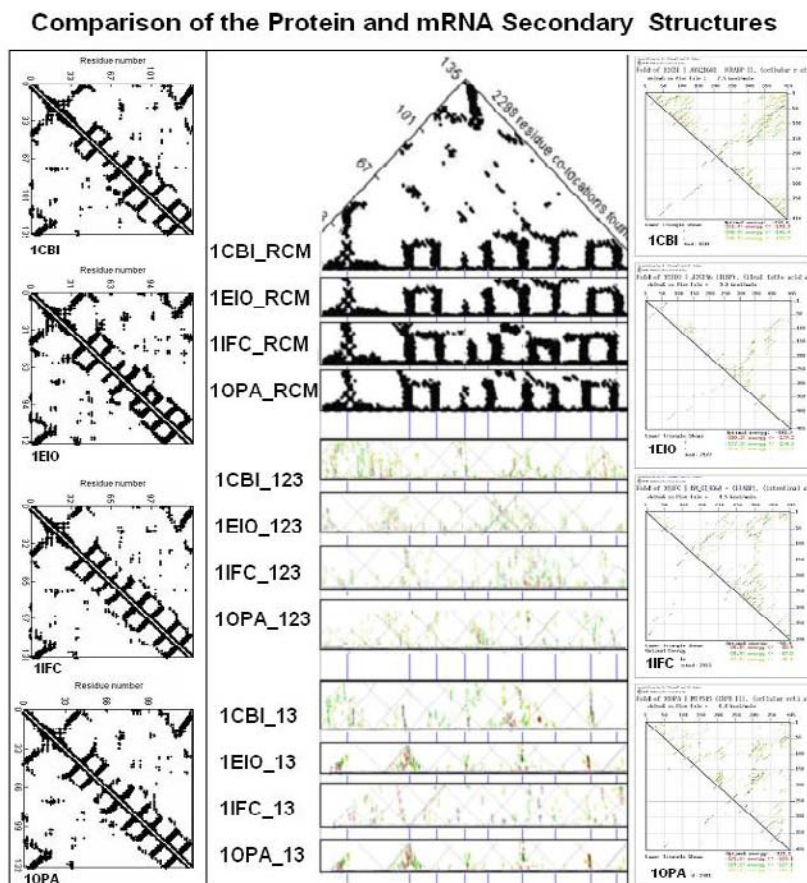
Figure 22. Comparison of the protein and mRNA secondary structures (modified from [59]). Residue contact maps (RCM) were obtained from the PBD files of 4 protein structures (1CBI, 1EIO, 1IFC, 1OPA) using the SeqX tool (left column). Energy dot-plots (EDP) for the coding sequences were obtained using the mfold tool (right column). The left diagonal portion of these two kinds of maps was compared in the central part of the figure. Blue horizontal lines in the background correspond to the main amino acid co-location sites in the RCM. Intact RNA (123) as well as subsequences containing only the 1st and 3rd codon letter (13) are compared. The black dots in the RCMs indicate amino acids that are within 6 A of each other in the protein structure. The colored (grass-like) areas in the EDPs indicate the energetically most likely RNA interactions (color code in increasing order: yellow, green red, black).

The similarity between mRNA and the coded protein secondary structure is an unexpected, novel observation. The 21/64 redundancy of the genetic code gives a 441/4.096 codon pair redundancy for every amino acid pair. It means that every amino acid pair might be coded by ~9 different codon pairs (some complementary but most not). The similarity between protein and their mRNA structures indicates extensive complementary coding of co-locating amino acids. The possible number of codon variations and possible nucleic acid structures behind a protein sequence and structure is very large (Figure 23) as well as the corresponding folding energies (dG, the stability of the mRNA).
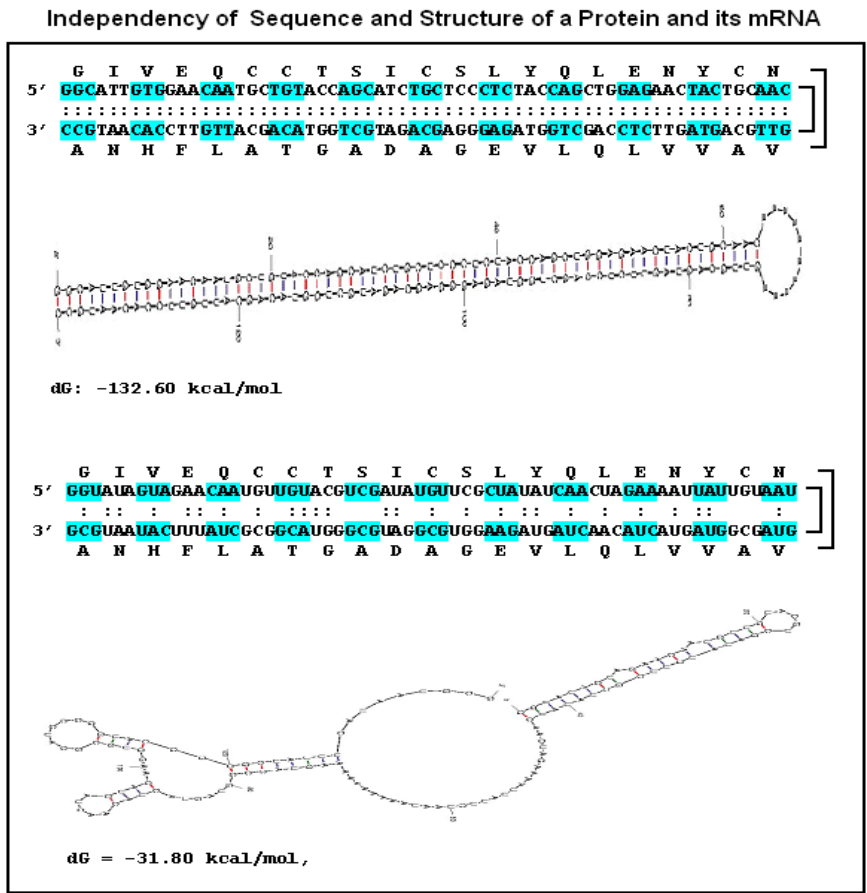
Figure 23. Independency of sequence and structure of a protein and its mRNA. The amino acids in a U-shaped protein structure are coded by complementary codons (rule PC1). The nucleic acid structure is uniform and the folding energy is –132 kcal/mol. Exactly the same amino acid sequence might be coded by non- or only partially complementary codons, which will fundamentally alter the mRNA structure and increase the folding energy to –31 kcal/mol (less stable). The nucleic acid structures were generated by mfold and dGs were calculated by the same program.

COMPLEMENTARY CODES VS AMINO ACID CO-LOCATIONS

Comparisons of the protein residue contact map with the nucleic acid folding maps suggest similarities between the 3D structures of these different kinds of molecules. However, this is a semi-quantitative method.

A more direct statistical support might be obtained by analyzing and comparing residue co-locations in these structures. Assume that the structural unit of mRNA is a tri-nucleotide (codon) and the structural unit of the protein is the amino acid. The codon may form a secondary structure by interacting with other codons according to the W-C base complementary rules, and contribute to the formation of a local double helix. The 5′-A1U2G3-3′ sequence (Met, M codon) forms a perfect double string with the 3′-U3A2C1-5′ sequence (His, H codon, reverse and complementary reading). Suboptimal complexes are 5′-

A1X2G3-3′ partially complemented by 3′-U3X2C1-5′ (AAG, Lys; AUG, Met; AGG, Arg; ACG, Pro; and CAU, His; CUU, Leu; CGU, Arg; CCU, Pro, respectively).

Our experiments with FFE indicate that local nucleic acid structures are formed under this suboptimal condition, i.e., when the 1st and 3rd codon residues are complementary but the 2nd is not. If this is the case, and there is a connection between nucleic acid and protein 3D structure, one might expect that the 4 amino acids coded by 5′-A1X2G3-3′ codons will preferentially co-locate with another 4 amino acids coded by 3′-U3X2C1-5′ codons. We have constructed 8 different complementary codon combinations and found that the codons of co-locating amino acids are often complementary at the 1st and 3rd positions and follow the D-1X3/RC-3X1 formula but not the 7 other formulas (Figures 24 and 25).
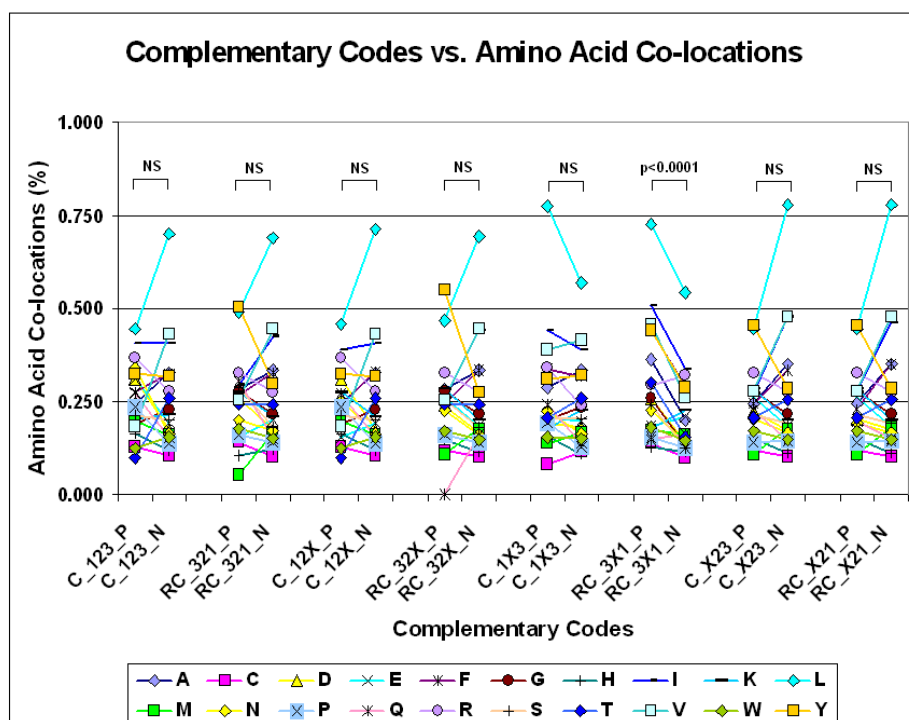
Figure 24. Complementary codes vs amino acid co-locations (modified from [59]). The propensity of the 400 possible amino acid pairs was monitored in 81 different protein structures with the SeqX tool. The tool detected co-locations when two amino acids were within 6 A of each other (neighbors on the same strand were excluded). The total number of co-locations was 34,630. Eight different complementary codes were constructed for the codons (2 optimal and 6 suboptimal). In the two optimal codes, all three codon residues (123) were complementary (C) or reverse complementary (RC) to each other. In the suboptimal codes, only two of three codon residues were C or RC to each other (12, 13, 23), while the third was not necessarily complementary (X). (For example, Complementary Code RC_1X3 means that the first and third codon letters are always complementary, but the not the second and the possible codons are read in reverse orientation. The 400 co-locations were divided into 20 subgroups corresponding to 20 amino acids (one of the co-locating pairs), each group containing the 20 amino acids (corresponding to the other amino acid in the co-locating pair). If the codons of the amino acid pairs followed the predefined complementary code the co-location was regarded as positive (P); if not, the co-location was regarded as negative (N). Each symbol represents the mean frequency of P or N co-locations corresponding to the indicated amino acid. Paired Student's t-test, $n=20$.
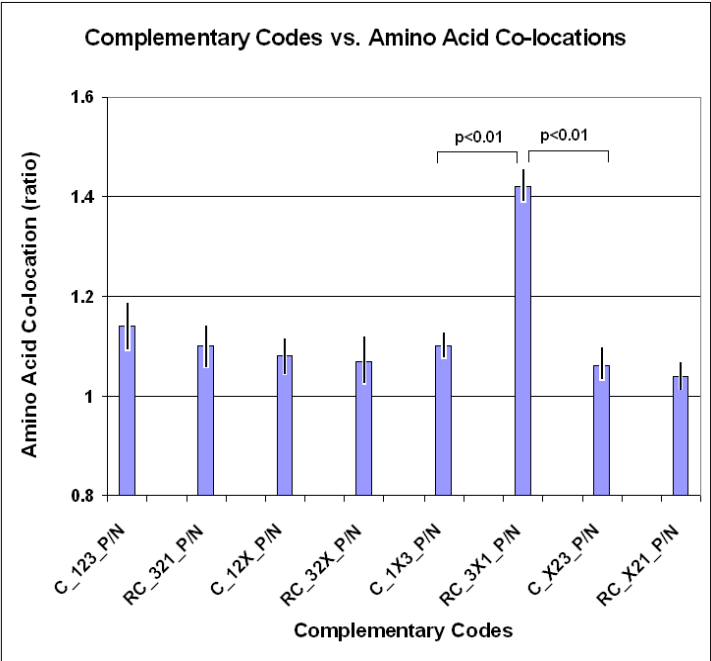
Figure 25. Complementary codes vs amino acid co-locations (ratios) (modified from [59]). The ratio of positive (P) and negative (N) co-locations was calculated on data from Figure 13. Each bar represents the mean±SEM (*n*=20).

These special amino acid pairs and their frequency are indicated and summarized in a matrix (Figure 26).

SeqX 80			1st	G	T	G	G	T	G	C	A	A	CT	A	A	C	C	AC	AT	A	G	T	T
			2nd	C	G	A	A	T	G	A	T	A	T	T	A	C	A	G	CG	C	T	G	A
			3rd	X	CT	CT	AG	CT	X	CT	ACT	AG	XAG	G	CT	X	AG	AGX	CTX	X	X	G	CT
1st	2nd	3rd	AA	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
G	C	X	A	7.5	1.4	2.6	2.9	4.1	3.9	1.5	6.3	3.3	9.9	2.8	2.6	2.0	2.5	3.3	3.7	4.5	7.5	1.6	4.7
T	G	CT	C	1.4	3.1	0.7	0.6	1.3	1.4	0.4	1.2	0.7	1.7	0.5	0.9	0.8	0.4	0.7	1.1	1.3	1.9	0.5	1.3
G	A	CT	D	2.6	0.7	1.6	1.3	1.6	2.7	1.4	2.3	3.1	3.6	0.8	1.5	1.3	3.5	2.3	2.7	3.0	1.0	2.2
G	A	AG	E	2.9	0.6	1.3	2.2	2.1	2.1	1.2	2.9	3.5	4.1	1.0	1.4	1.3	1.5	3.6	2.3	2.2	3.4	1.0	2.2
T	T	CT	F	4.1	1.3	1.6	2.1	3.3	2.8	1.3	4.8	2.4	6.5	1.9	1.8	2.2	1.7	3.3	2.4	2.4	4.9	1.6	3.0
G	G	X	G	3.9	1.4	2.7	2.1	2.8	4.2	1.3	4.2	2.3	5.2	1.4	2.3	2.1	1.7	3.0	3.0	3.3	4.4	1.3	2.7
C	A	CT	H	1.5	0.4	1.4	1.2	1.3	1.3	0.6	1.5	0.8	2.4	0.5	0.7	0.9	0.7	1.2	1.3	1.2	1.5	0.6	1.3
A	T	ACT	I	6.3	1.2	2.3	2.9	4.8	4.2	1.5	7.0	2.7	12.2	2.5	2.1	1.8	2.2	3.1	3.2	3.7	7.7	1.8	4.4
A	A	AG	K	3.3	0.7	3.1	3.5	2.4	2.3	0.8	2.7	1.5	4.2	1.3	1.6	1.3	1.7	1.7	1.9	2.3	3.4	1.0	2.5
CT	T	XAG	L	9.9	1.7	3.6	4.1	6.5	5.2	2.4	12.2	4.2	19.7	3.9	3.7	3.2	3.5	5.8	4.4	5.0	11.2	3.3	6.4
A	T	G	M	2.8	0.5	0.8	1.0	1.9	1.4	0.5	2.5	1.3	3.9	1.5	0.8	0.9	0.9	1.5	1.2	1.2	2.6	0.6	1.8
A	A	CT	N	2.6	0.9	1.5	1.4	1.8	2.3	0.7	2.1	1.6	3.7	0.8	1.6	1.5	1.4	1.8	2.2	2.1	2.1	1.0	1.9
C	C	X	P	2.0	0.8	1.3	1.3	2.2	2.1	0.9	1.8	1.3	3.2	0.9	1.5	1.5	1.0	2.2	1.4	1.6	2.8	1.2	2.1
C	A	AG	Q	2.5	0.4	1.3	1.5	1.7	1.7	0.7	2.2	1.7	3.5	0.9	1.4	1.0	0.9	1.6	1.3	1.7	2.5	0.7	1.6
AC	G	AGX	R	3.3	0.7	3.5	3.6	3.3	3.0	1.2	3.1	1.7	5.8	1.5	1.8	2.2	1.6	2.4	2.5	2.6	3.7	0.8	2.5
AT	CG	CTX	S	3.7	1.1	2.3	2.3	2.4	3.0	1.3	3.2	1.9	4.4	1.2	2.2	1.4	1.3	2.5	2.5	2.6	3.3	1.2	2.3
A	C	X	T	4.5	1.3	2.7	2.2	2.4	3.3	1.2	3.7	2.3	5.0	1.2	2.1	1.6	1.7	2.6	2.6	3.8	4.4	0.9	2.5
G	T	X	V	7.5	1.9	3.0	3.4	4.9	4.4	1.5	7.7	3.4	11.2	2.6	2.1	2.8	2.5	3.7	3.3	4.4	8.7	2.0	4.6
T	G	G	W	1.6	0.5	1.0	1.0	1.6	1.3	0.6	1.8	1.0	3.3	0.6	1.0	1.2	0.7	0.8	1.2	0.9	2.0	0.9	1.5
T	A	CT	Y	4.7	1.3	2.2	2.2	3.0	2.7	1.3	4.4	2.5	6.4	1.8	1.9	2.1	1.6	2.5	2.3	2.5	4.6	1.5	3.6

RC_3X1 code: 1st and 3rd codon letters are complementary in reverse order, indicated by complementary colors (red, blue); unfavored colocations are gray-colored.
X: any residue; AA: amino acids, one-letter code; Co-location frequency is multiplied by 10, Sum=1000

Figure 26. Complementary codes vs amino acid co-locations. See text for explanation.

It is well known that coding and non-coding DNA sequences (exon/intron) are different and this difference is somehow related to the asymmetry of the codons, i.e., that the third codon letter (wobble) is poorly defined. Many Markov models have been formulated to find this asymmetry and de novo predict coding sequences (genes). These in silico methods work rather well but not perfectly and some scientists remain unconvinced that the codon asymmetry explains the exon–intron differences satisfactorily.

Another codon-related problem is that the well-known, non-overlapping, triplet codon translation is extremely phase-dependent and there is theoretically no tolerance for any phase shift. There are famous examples of how single nucleotide deletion might destroy the meaningful translation of a sequence and which are incompatible with life. However, considering the magnitude and complexity of the eukaryotic proteome, the precision of translation is astonishingly good. Such physical precision is not possible without massive and consistent physico-chemical fundaments. Therefore, discovery of the existence of secondary structure bias (folding energy differences) in coding regions of many organisms [78] was a very welcome observation because it differentiates exons from introns on a physico-chemical basis.

Our experiments with free folding energy (FFE) confirmed that this bias exists. In addition, there is a very consistent and very significant pattern of FFE distribution along the nucleotide sequence. Comparing the FFE of phase-selected subsequences, subsequences comprised of only the 1st or only the 3rd codon letters showed significantly higher FFE than those consisting only of the 2nd letters. This FFE difference was not present in intronic sequences preceding and following the exons, but it was present in exons from different species including viruses. This is an interesting observation because this phenomenon might not only distinguish between exons and introns on a physico-chemical basis, but it might even clearly define the tri-nucleotide codons and thus the phase of the translation. This codon-related phase-specific variation in FFE may explain why mRNAs have greater negative free folding energies than shuffled or codon choice randomized sequences [85].

Free folding energy in nucleic acids is always associated with W-C base pair formation. Higher FFE indicates more W-C pairs (presence of complementarity) and lower FFE indicates fewer W-C pairs (less complementarity). The FFE in the 1st and 3rd codon positions was additive, while the 2nd letter did not contribute to the total FFE; the total FFE of the entire (intact) nucleic acid was the same as subsequences containing only the 1st and 3rd codon letters (2nd deleted). This is an indication that the local RNA secondary structure bias is caused by complementarity of the 1st and 3rd codon residues in local sequences. This partial, local complementarity is more optimal in reverse orientation of the local sequences as expected with loop formations.

It is known that single stranded RNA molecules can form local secondary structures through the interactions of complementary segments. The novel observation here is that these interactions preferentially involve the 1st and 3rd codon residues. This connection between the RNA secondary structure and codons immediately directed attention towards the question of protein folding and its long suspected connection to RNA folding [86,87].

Only about one-third (20/64) of the genetic code is used for protein coding, i.e., there is a great excess of information in the mRNA. At the same time, the information carried by amino acids seems to be insufficient (as stated by some scientists) to complete unambiguous protein folding. Therefore, it is believed that the third codon residue (wobble base) carries some additional information to that already present in the genetic code. A specialized wobble base

oriented database, the ISSD [84] was established in an effort to connect different features of protein structure to wobble bases [88] with more or less success.

We found a significant negative correlation between FFE of the 2nd codon residue and the helix content of protein structures, which was not expected even though this possibility is mentioned in the literature [74]. Our previous work on a Common Periodic Table of Codons and Nucleic Acids [52] indicated that the second codon residue is intimately coupled with the known physico-chemical properties of the amino acids. Almost all amino acids show significant positive or negative correlation to the helix content of proteins. Therefore, the real biological meaning and significance of any connection between FFE of the 2nd codon residue and the propensity of a protein structural element is highly questionable.

It was possible to make direct visual comparison of mRNA structure (as statistically predicted by the mfold energy dot-plot) and protein structures (as 2D residue contact maps). This method suggests similarity between nucleic acid and protein structures. It is known that some complex protein structures are very similar even if there is less than 30% sequence similarity. It was interesting to see that the same principle might apply for nucleic acids, and structural similarity might exist even when the sequence similarity is low. Furthermore, significant similarity between nucleic acid and protein structures might exist even without translational connection.

Structure seems to be more preserved, even in nucleic acids, than sequence.

However, even if the matrix comparisons are suggestive, they remain semi-quantitative methods. Better support was necessary.

A working hypotheses grew out of these observations, namely that (a) partial, local reverse complementarity exists in nucleic acids that form the nucleic acid structure; (b) there is some degree of similarity between the folding of nucleic acids and proteins; (c) protein structure determines the amino acid co-locations; (d) as a consequence, amino acids coded by the interacting (partially reverse complementary) codons might show preferential co-locations in the protein structures.

This seems to be the case: codons which contain complementary bases at the 1st and 3rd positions and are translated in reverse orientation result in amino acids which are preferentially co-located (interacting) in the 3D protein structure. Other complementary residue combinations or translation in the same (not reverse) direction (as much as seven combinations in total) did not result in any preferentially co-locating subset of amino acid pairs.

Construction of residue contact maps for protein structures and statistical evaluation of residue co-locations is a frequently used method for visualization and analyses of spatial connections between amino acids [89,90,91. The amino acid co-locations in real protein structures is clearly not random [92,93] and therefore residue co-location matrices are often used to assist in the prediction of novel protein structures [94,95]. We have carefully examined the physico-chemical properties of specifically interacting amino acids in and between protein structures, and we concluded that these interactions follow the well known physico-chemical rules of size, charge and hydrophobe compatibility (unpublished data) well in line with Anfinsen's prediction. A recent study supports the fact that there is a previously unknown connection between the codons of specifically interacting amino acids; those codons are complementary at the 1st and 3rd (but not the 2nd) codon positions.

The idea that sequence complementarity might explain the nature of specific protein–protein interactions is not new and was suggested already in 1981 [9,10,11].

I was never able to experimentally confirm my own original theory, which suggested a perfect complementarity between codons of interacting amino [9,10,11,50], in contrast to others [37].

The explanation is that this codon complementarity is suboptimal and does not involve the 2nd codon residue. Experimental in vitro confirmation is required to validate this recent theoretical and in silico prediction.

AVAILABILITY

http://www.janbiro.com/downloads: SeqX, SeqForm, SeqPlot, Dotlet.
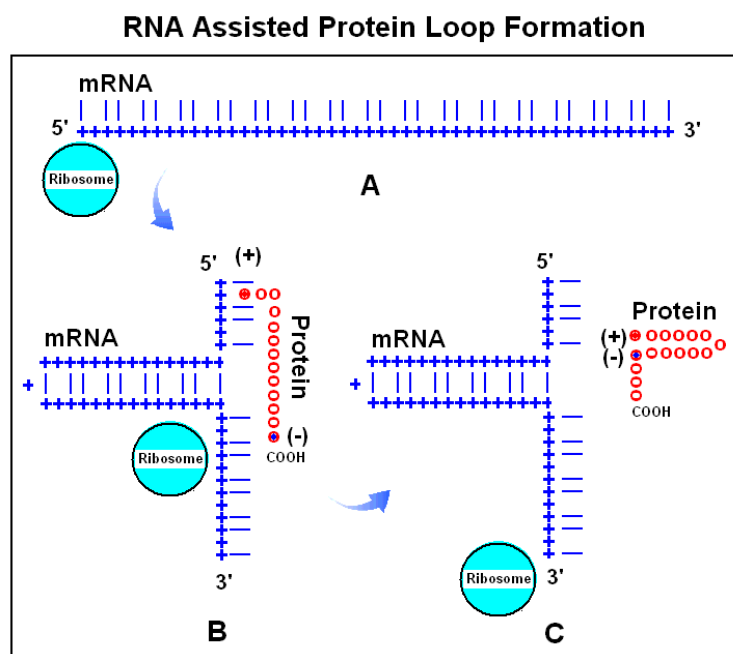
Figure 27. RNA assisted protein loop formation (from [96]. Translation begins with the attachment of the 5′ end of a mRNA to the ribosome (A). Ribonucleotides are indicated by blue "+" and the 1st and 3rd bases in the codons by blue lines; the 2nd base positions are left empty. A positively charged amino acid ((+) and red dots), for example, arginine, remains attached to its codon. The mRNA forms a loop because the 1st and 3rd bases are locally complementary to each other in reverse orientation (B). The growing protein is indicated by red circles. When translation proceeds to an amino acid with especially high affinity to the mRNA-attached arginine, for example a negatively charged Glu or Asp ((−) and blue dot), the charge attraction removes the Arg from its mRNA binding site and the entire protein is released from the mRNA and completes a protein loop (C). The protein continues to grow towards the direction of its carboxy terminal (COOH).

THEORY OF NUCLEIC ACID (CHAPERONS) ASSISTED PROTEIN FOLDING

A series of novel arguments has been presented above to support a deeper connection between nucleic acids and expressed proteins, than the traditional codon translation. The

physico-chemical properties of amino acids are clearly associated with the 2nd codon letter as was shown in the Common Periodic Table of Codons and Amino Acids. The co-locating amino acids are preferentially coded by codons that are complementary at the 1st and 3rd codon position.

The structure of proteins and their coding nucleic acid are rather similar to each other in many cases. All these observations suggest the co-evolution of codons and amino acids and that protein folding (structural) information is present in the nucleic acids in addition to the canonical genetic code.

This immediately raises the possibility of nucleic acid-assisted protein folding, i.e. the possibility of nucleic acid chaperons [96]. This is an exciting possibility, because the protein primary sequence seems not to carry the necessary information for unambiguous protein folding (in contrast to Anfinsen's theorem), while there is a two fold excess of information in the redundant genetic code. A theoretical example of how such nucleic acid-assisted protein folding may look is presented in Figures 27 and 28.
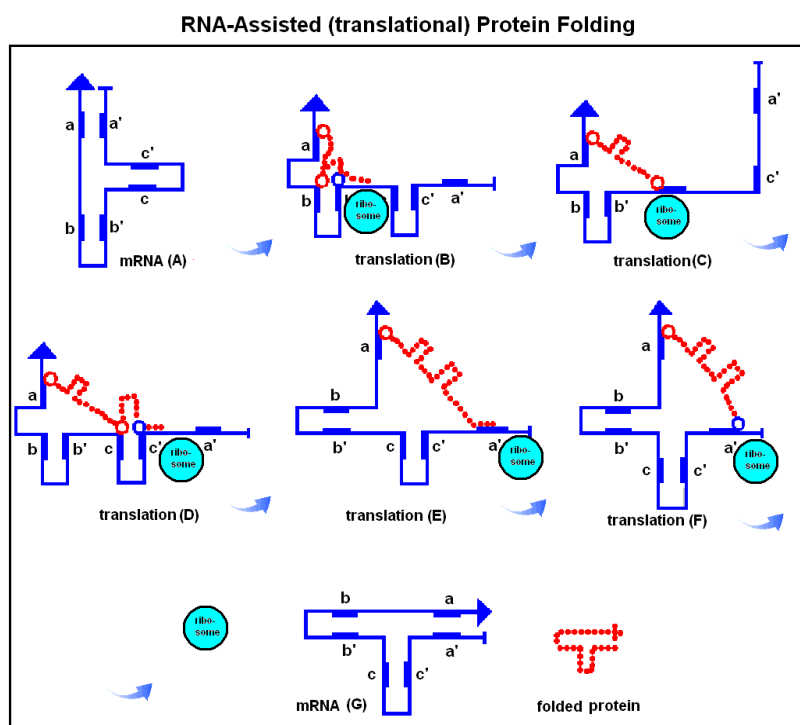
Figure 28. RNA-assisted (translational) protein folding (from [96]. There are three reverse and complementary regions in a mRNA (blue line, A): a–a′, b–b′, c–c′, which fold the mRNA into a T-like shape. During the translation process the mRNA unfolds on the surface of the ribosome, but subsequently refolds, accompanied by its translated and lengthening peptide (red dotted line, B–F). The result of translation is a temporary ribonucleotide complex, which dissociates into two T-shape-like structures: the original mRNA and the properly folded protein product (G). The red circles indicate the specific, temporary attachment points between the RNA and protein (for example a basic amino acid); the blue circles indicate amino acids with exceptionally high affinity for the attachment points (e.g., acidic amino acids); these capture the amino acids at the attachment point and dissociate the ribonucleoprotein complex. Transfer RNAs are of course important participants in translation, but they are not included in this scenario.

DEFINITION OF THE 2ND GENERATION PROTEOMIC CODE

The Proteomic Code is a set of comprehensive rules by which information in genetic material is transferred into the physico-chemical properties of amino acids and determines how individual amino acids interact with each other during folding and in specific protein–protein interactions. The Proteomic Code is part of the redundant genetic code. The theory of Proteomic Code contains the following observations:

– Co-locating (interacting) amino acids in native proteins are coded by partially (imperfect) complementary codons in reverse ($5' \rightarrow 3'/5' \rightarrow 3'$) orientation.
– Partial complementarity means that the 1st and 3rd codon bases are complementary (Watson–Crick) bases to each other, while the 2nd bases may or may not be complementary to each other.
– The physico-chemical characteristics of the coded and interacting amino acids are determined mainly by the 2nd (central) codon residues.
– The physico-chemical properties of the interacting amino acids (size, charge, and hydropathy) are compatible with each other at the individual amino acid level.
– Nucleic acids (exons) contain protein folding information in (or in addition to) the redundant genetic code.
– Nucleic acids may directly assist protein folding as chaperons.
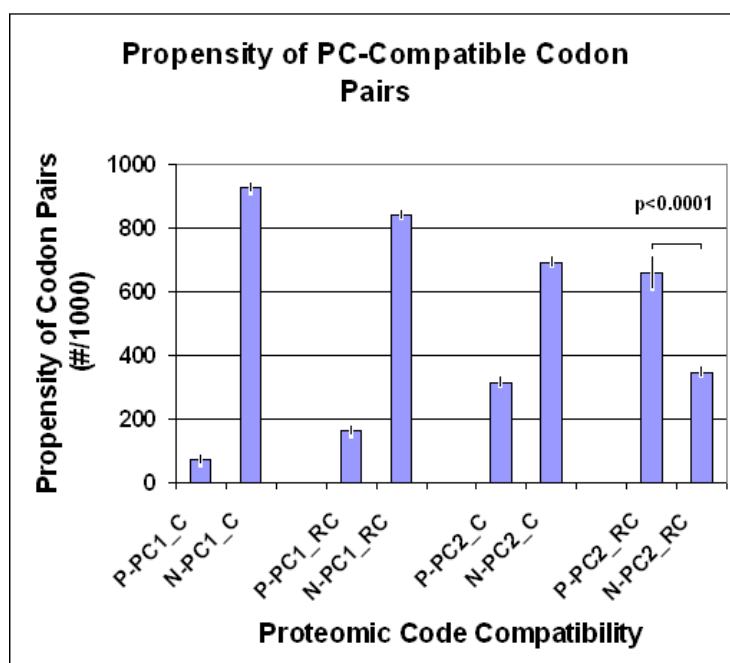
Figure 29. Propensity of PC-compatible codon pairs. There are 4096 possible codon pairs altogether (64×64, including those formed with the 3 stop codons). Codons, which are (P) or are not (N) coded by a specific complementary rule (PC) were counted. The bars represent the mean±SD of 9 independent determinations (n=9). PC1 and PC2 indicate complementarity of all 3 or only the first and third codon bases in parallel (C) or anti-parallel (RC) readings.

There are four different Proteomic Codes at this moment. PC1_C and PC1_RC are the original codes based on the perfect complementarity of all three codon bases in complementary (C, 5′→3′/3′→5′) readings; PC2_C and PC2_RC are the recent extended codes requiring base pair complementarity only at the 1st and 3rd codon position, but not necessarily at the 2nd. PC1 is part of PC2.

The PC_C variants require 3′→5′ translations, which do not exist, therefore I regard this variant as an artifact, caused by the symmetry of many codons. Only a small fraction of total codon and amino acid pairs belongs to PC1; ~50% of all amino acid pairs and >60% of all codon pairs can be classified into PC2_RC (Figures 29 and 30).
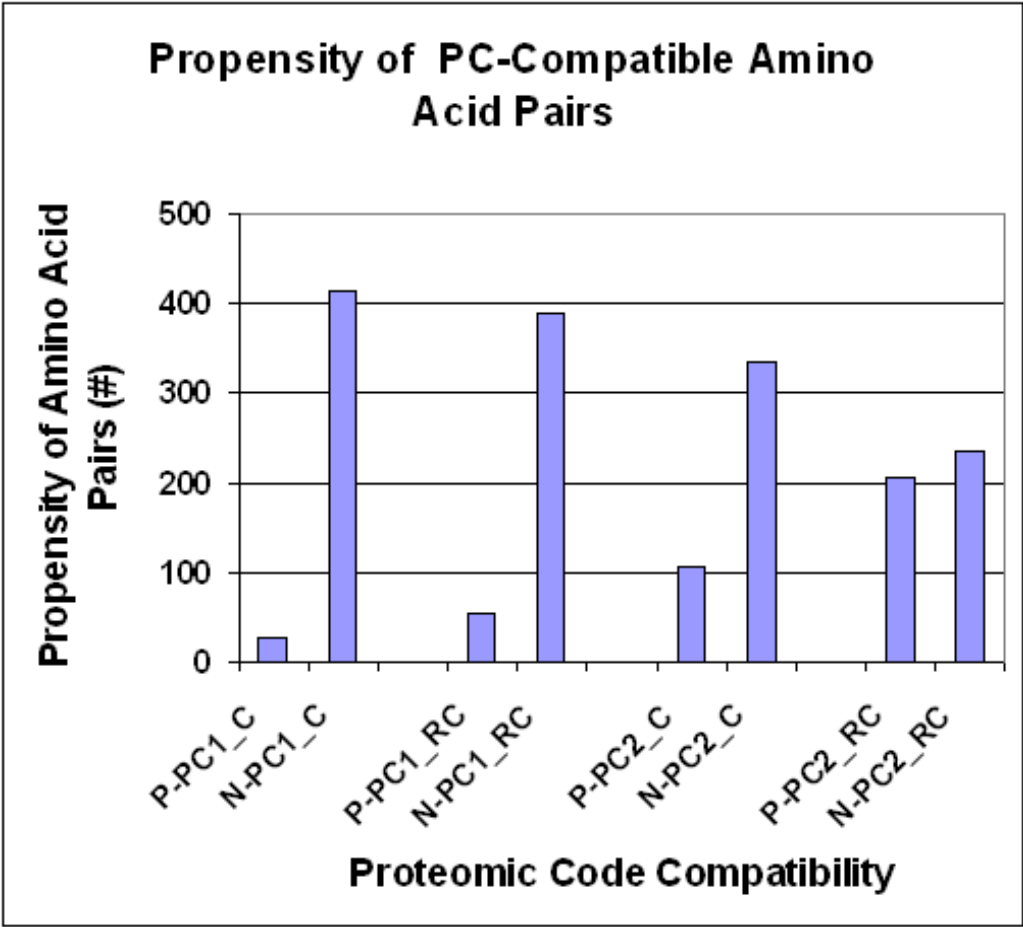
Figure 30. Propensity of PC-compatible amino acid pairs. There are 441 possible amino acid pairs altogether (21×21, including the stop place as the 21st variable). Amino acids, which are (P) or are not (N) coded by a specific complementary rule (Proteomic Code, PC) were counted. PC1 and PC2 indicate complementarity of the respective codon pairs at all 3 or only the first and third codon bases in parallel (C) or anti-parallel (RC) readings.

A list of all possible amino acid pairs (21×21=441, including the virtual pairs formed with the Stop/End signal) are listed in Table 4. The most important physico-chemical parameters (molecular weight (MW), isoelectric point (pI), hydropathy (HP)) and the derived

values of three compatibility indexes (charge (CCI), size (SCI) and hydropathy (HCI) compatibility indexes [58]) as well as the expected frequency of the amino acid pairs (natural frequency (NF), calculated from codon table 1) are included in Table 4. P (positive) and N (negative) indicate whether an amino acid pair may or may not be coded by a given Proteomic Code.

Table 4

First	Second	1st	2nd	PC1_C	PC2_C	PC1_RC	PC2_RC	mw-57_1	HP_1	pI_1	mw-57_2	HP_2	pI_2	NF*/1k	SCI	CCI	HCI
Ala	Ala	A	A	N	N	N	P	14.0	1.0	6.0	14.0	1.0	6.0	4.3	6.6	10.4	20.0
Ala	Arg	A	R	P	P	P	P	14.0	1.0	6.0	99.0	-7.5	10.8	6.4	18.6	13.1	4.8
Ala	Asn	A	N	N	N	N	P	14.0	1.0	6.0	57.0	-2.7	5.4	2.1	12.7	10.1	13.4
Ala	Asp	A	D	N	N	N	P	14.0	1.0	6.0	58.0	-3.0	2.8	2.1	12.8	8.6	12.8
Ala	Cys	A	C	N	N	P	P	14.0	1.0	6.0	46.0	0.2	5.1	2.1	11.1	9.9	18.5
Ala	Gln	A	Q	N	P	N	N	14.0	1.0	6.0	71.0	-2.9	5.7	2.1	14.7	10.2	13.0
Ala	Glu	A	E	N	N	N	N	14.0	1.0	6.0	72.0	-2.6	3.2	2.1	14.8	8.9	13.5
Ala	Gly	A	G	N	N	P	P	14.0	1.0	6.0	0.0	0.7	6.0	4.3	4.7	10.4	19.4
Ala	His	A	H	N	P	N	P	14.0	1.0	6.0	80.0	-1.7	7.6	2.1	15.9	11.3	15.2
Ala	Ile	A	I	N	N	N	P	14.0	1.0	6.0	56.0	3.1	5.9	3.2	12.5	10.4	16.2
Ala	Leu	A	L	N	P	N	P	14.0	1.0	6.0	56.0	2.2	6.0	6.4	12.5	10.4	17.8
Ala	Lys	A	K	N	N	N	N	14.0	1.0	6.0	71.0	-4.6	9.7	2.1	14.7	12.5	10.0
Ala	Met	A	M	N	N	N	N	14.0	1.0	6.0	74.0	1.1	5.7	1.1	15.1	10.3	19.8
Ala	Phe	A	F	N	N	N	P	14.0	1.0	6.0	90.0	2.5	5.5	2.1	17.3	10.2	17.3
Ala	Pro	A	P	N	P	N	N	14.0	1.0	6.0	40.0	-0.3	6.5	4.3	10.3	10.7	17.7
Ala	Ser	A	S	N	N	P	P	14.0	1.0	6.0	30.0	-1.1	5.7	6.4	8.9	10.3	16.2
Ala	Thr	A	T	N	N	N	P	14.0	1.0	6.0	44.0	-0.8	5.9	4.3	10.9	10.4	16.9
Ala	Trp	A	W	N	N	N	N	14.0	1.0	6.0	129.0	1.5	6.0	1.1	17.2	10.4	19.1
Ala	Tyr	A	Y	N	N	N	P	14.0	1.0	6.0	106.0	0.1	5.7	2.1	19.6	10.3	18.4
Ala	Val	A	V	N	N	N	P	14.0	1.0	6.0	42.0	2.3	6.0	4.3	10.6	10.4	17.7
Arg	Ala	R	A	P	P	P	P	99.0	-7.5	10.8	14.0	1.0	6.0	6.4	18.6	13.1	4.8
Arg	Arg	R	R	N	N	N	P	99.0	-7.5	10.8	99.0	-7.5	10.8	9.7	9.4	3.1	20.0
Arg	Asn	R	N	N	P	N	N	99.0	-7.5	10.8	57.0	-2.7	5.4	3.2	15.4	14.4	11.4
Arg	Asp	R	D	N	P	N	N	99.0	-7.5	10.8	58.0	-3.0	2.8	3.2	15.2	19.9	11.9
Arg	Cys	R	C	N	P	N	P	99.0	-7.5	10.8	46.0	0.2	5.1	3.2	16.9	15.1	6.3
Arg	Gln	R	Q	N	N	N	P	99.0	-7.5	10.8	71.0	-2.9	5.7	3.2	13.4	13.9	11.8
Arg	Glu	R	E	N	P	N	P	99.0	-7.5	10.8	72.0	-2.6	3.2	3.2	13.2	19.0	11.2
Arg	Gly	R	G	N	P	N	P	99.0	-7.5	10.8	0.0	0.7	6.0	6.4	16.6	13.2	5.4
Arg	His	R	H	N	N	N	P	99.0	-7.5	10.8	80.0	-1.7	7.6	3.2	12.1	9.8	9.6
Arg	Ile	R	I	N	N	N	N	99.0	-7.5	10.8	56.0	3.1	5.9	4.8	15.5	13.3	1.0
Arg	Leu	R	L	N	N	N	P	99.0	-7.5	10.8	56.0	2.2	6.0	9.7	15.5	13.2	2.6
Arg	Lys	R	K	N	N	N	P	99.0	-7.5	10.8	71.0	-4.6	9.7	3.2	13.4	5.2	14.8
Arg	Met	R	M	N	N	N	N	99.0	-7.5	10.8	74.0	1.1	5.7	1.6	13.0	13.7	4.6
Arg	Phe	R	F	N	P	N	P	99.0	-7.5	10.8	90.0	2.5	5.5	3.2	10.7	14.2	2.1
Arg	Pro	R	P	N	N	P	P	99.0	-7.5	10.8	40.0	-0.3	6.5	6.4	17.7	12.1	7.1
Arg	Ser	R	S	P	P	P	P	99.0	-7.5	10.8	30.0	-1.1	5.7	9.7	19.2	13.8	8.5
Arg	Thr	R	T	N	N	P	P	99.0	-7.5	10.8	44.0	-0.8	5.9	6.4	17.2	13.4	7.9
Arg	Trp	R	W	N	N	N	N	99.0	-7.5	10.8	129.0	1.5	6.0	1.6	5.2	13.2	3.9
Arg	Tyr	R	Y	N	P	N	P	99.0	-7.5	10.8	106.0	0.1	5.7	3.2	8.5	13.8	6.4
Arg	Val	R	V	N	P	N	P	99.0	-7.5	10.8	42.0	2.3	6.0	6.4	17.5	13.2	2.4
Asn	Ala	N	A	N	N	N	P	57.0	-2.7	5.4	14.0	1.0	6.0	2.1	12.7	10.1	13.4
Asn	Arg	N	R	N	N	N	N	57.0	-2.7	5.4	99.0	-7.5	10.8	3.2	15.4	14.4	11.4
Asn	Asn	N	N	N	N	N	P	57.0	-2.7	5.4	57.0	-2.7	5.4	1.1	18.7	9.6	20.0
Asn	Asp	N	D	N	N	N	P	57.0	-2.7	5.4	58.0	-3.0	2.8	1.1	18.9	7.2	19.5
Asn	Cys	N	C	N	N	N	N	57.0	-2.7	5.4	46.0	0.2	5.1	1.1	17.2	9.3	14.9
Asn	Gln	N	Q	N	N	N	N	57.0	-2.7	5.4	71.0	-2.9	5.7	1.1	19.3	9.8	19.6
Asn	Glu	N	E	N	N	N	N	57.0	-2.7	5.4	72.0	-2.6	3.2	1.1	19.2	7.6	19.8
Asn	Gly	N	G	N	N	N	N	57.0	-2.7	5.4	0.0	0.7	6.0	2.1	10.7	10.1	14.0
Asn	His	N	H	N	N	N	N	57.0	-2.7	5.4	80.0	-1.7	7.6	1.1	18.0	11.5	18.2
Asn	Ile	N	I	N	N	N	P	57.0	-2.7	5.4	56.0	3.1	5.9	1.6	18.6	10.0	9.6
Asn	Leu	N	L	P	P	N	N	57.0	-2.7	5.4	56.0	2.2	6.0	3.2	18.6	10.1	11.2
Asn	Lys	N	K	N	N	N	N	57.0	-2.7	5.4	71.0	-4.6	9.7	1.1	19.3	13.4	16.6
Asn	Met	N	M	N	N	N	N	57.0	-2.7	5.4	74.0	1.1	5.7	0.5	18.9	9.9	13.2
Asn	Phe	N	F	N	N	N	N	57.0	-2.7	5.4	90.0	2.5	5.5	1.1	16.6	9.6	10.7
Asn	Pro	N	P	N	N	N	N	57.0	-2.7	5.4	40.0	-0.3	6.5	2.1	16.3	10.5	15.7
Asn	Ser	N	S	N	N	P	P	57.0	-2.7	5.4	30.0	-1.1	5.7	3.2	14.9	9.8	17.1
Asn	Thr	N	T	N	N	N	P	57.0	-2.7	5.4	44.0	-0.8	5.9	2.1	16.9	10.0	16.5
Asn	Trp	N	W	N	P	N	N	57.0	-2.7	5.4	129.0	1.5	6.0	0.5	11.1	10.1	12.5
Asn	Tyr	N	Y	N	N	N	N	57.0	-2.7	5.4	106.0	0.1	5.7	1.1	14.4	9.8	15.0
Asn	Val	N	V	N	N	P	P	57.0	-2.7	5.4	42.0	2.3	6.0	2.1	16.6	10.1	11.0
Asp	Ala	D	A	N	N	N	P	58.0	-3.0	2.8	14.0	1.0	6.0	2.1	12.8	8.6	12.8
Asp	Arg	D	R	N	P	N	N	58.0	-3.0	2.8	99.0	-7.5	10.8	3.2	15.2	19.9	11.9
Asp	Asn	D	N	N	N	N	P	58.0	-3.0	2.8	57.0	-2.7	5.4	1.1	18.9	7.2	19.5
Asp	Asp	D	D	N	N	N	P	58.0	-3.0	2.8	58.0	-3.0	2.8	1.1	19.0	0.9	20.0
Asp	Cys	D	C	N	N	N	N	58.0	-3.0	2.8	46.0	0.2	5.1	1.1	17.3	6.4	14.3
Asp	Gln	D	Q	N	P	N	N	58.0	-3.0	2.8	71.0	-2.9	5.7	1.1	19.2	7.8	19.8
Asp	Glu	D	E	N	N	N	N	58.0	-3.0	2.8	72.0	-2.6	3.2	1.1	19.0	2.0	19.3
Asp	Gly	D	G	N	N	N	P	58.0	-3.0	2.8	0.0	0.7	6.0	2.1	10.9	8.6	13.4
Asp	His	D	H	N	N	N	N	58.0	-3.0	2.8	80.0	-1.7	7.6	1.1	17.9	12.4	17.7
Asp	Ile	D	I	N	N	P	P	58.0	-3.0	2.8	56.0	3.1	5.9	1.6	18.7	8.4	9.1
Asp	Leu	D	L	P	P	N	N	58.0	-3.0	2.8	56.0	2.2	6.0	3.2	18.7	8.6	10.7
Asp	Lys	D	K	N	N	N	N	58.0	-3.0	2.8	71.0	-4.6	9.7	1.1	19.2	17.5	17.1
Asp	Met	D	M	N	N	N	N	58.0	-3.0	2.8	74.0	1.1	5.7	0.5	18.7	8.0	12.7
Asp	Phe	D	F	N	N	N	N	58.0	-3.0	2.8	90.0	2.5	5.5	1.1	16.5	7.4	10.1
Asp	Pro	D	P	N	P	N	N	58.0	-3.0	2.8	40.0	-0.3	6.5	2.1	16.5	9.8	15.1
Asp	Ser	D	S	N	N	N	P	58.0	-3.0	2.8	30.0	-1.1	5.7	3.2	15.1	7.9	16.6
Asp	Thr	D	T	N	N	N	P	58.0	-3.0	2.8	44.0	-0.8	5.9	2.1	17.0	8.3	16.0
Asp	Trp	D	W	N	N	N	N	58.0	-3.0	2.8	129.0	1.5	6.0	0.5	11.0	8.6	11.9
Asp	Tyr	D	Y	N	N	N	N	58.0	-3.0	2.8	106.0	0.1	5.7	1.1	14.2	7.8	14.5
Asp	Val	D	V	N	N	P	P	58.0	-3.0	2.8	42.0	2.3	6.0	2.1	16.8	8.6	10.5

Table 4. Continued

First	Second	1st	2nd	PC1_C	PC2_C	PC1_RC	PC2_RC	mw-57_1	HP_1	pI_1	mw-57_2	HP_2	pI_2	NF°/1k	SCI	CCI	HCI
Cys	Ala	C	A	N	N	P	P	46.0	0.2	5.1	14.0	1.0	6.0	2.1	11.1	9.9	18.5
Cys	Arg	C	R	N	P	N	P	46.0	0.2	5.1	99.0	-7.5	10.8	3.2	16.9	15.1	6.3
Cys	Asn	C	N	N	N	N	N	46.0	0.2	5.1	57.0	-2.7	5.4	1.1	17.2	9.3	14.9
Cys	Asp	C	D	N	N	N	N	46.0	0.2	5.1	58.0	-3.0	2.8	1.1	17.3	6.4	14.3
Cys	Cys	C	C	N	N	N	N	46.0	0.2	5.1	46.0	0.2	5.1	1.1	15.6	8.9	20.0
Cys	Gln	C	Q	N	N	N	N	46.0	0.2	5.1	71.0	-2.9	5.7	1.1	19.2	9.5	14.5
Cys	Glu	C	E	N	N	N	P	46.0	0.2	5.1	72.0	-2.6	3.2	1.1	19.3	6.9	15.0
Cys	Gly	C	G	N	N	N	P	46.0	0.2	5.1	0.0	0.7	6.0	2.1	9.2	9.9	19.1
Cys	His	C	H	N	N	N	N	46.0	0.2	5.1	80.0	-1.7	7.6	1.1	19.6	11.6	16.6
Cys	Ile	C	I	N	P	N	P	46.0	0.2	5.1	56.0	3.1	5.9	1.6	17.0	9.8	14.7
Cys	Leu	C	L	N	N	N	N	46.0	0.2	5.1	56.0	2.2	6.0	3.2	17.0	9.9	16.4
Cys	Lys	C	K	N	P	N	P	46.0	0.2	5.1	71.0	-4.6	9.7	1.1	19.2	14.0	11.5
Cys	Met	C	M	N	P	N	N	46.0	0.2	5.1	74.0	1.1	5.7	0.5	19.6	9.6	18.3
Cys	Phe	C	F	N	N	N	N	46.0	0.2	5.1	90.0	2.5	5.5	1.1	18.2	9.4	15.8
Cys	Pro	C	P	N	N	N	N	46.0	0.2	5.1	40.0	-0.3	6.5	2.1	14.8	10.4	19.2
Cys	Ser	C	S	N	N	N	N	46.0	0.2	5.1	30.0	-1.1	5.7	3.2	13.4	9.6	17.7
Cys	Thr	C	T	P	P	P	P	46.0	0.2	5.1	44.0	-0.8	5.9	2.1	15.4	9.8	18.4
Cys	Trp	C	W	N	N	N	N	46.0	0.2	5.1	129.0	1.5	6.0	0.5	12.7	9.9	17.6
Cys	Tyr	C	Y	N	N	N	N	46.0	0.2	5.1	106.0	0.1	5.7	1.1	15.9	9.5	19.8
Cys	Val	C	V	N	N	N	P	46.0	0.2	5.1	42.0	2.3	6.0	2.1	15.1	9.9	16.2
Gln	Ala	Q	A	N	P	N	N	71.0	-2.9	5.7	14.0	1.0	6.0	2.1	14.7	10.2	13.0
Gln	Arg	Q	R	N	N	N	P	71.0	-2.9	5.7	99.0	-7.5	10.8	3.2	13.4	13.9	11.8
Gln	Asn	Q	N	N	N	N	N	71.0	-2.9	5.7	57.0	-2.7	5.4	1.1	19.3	9.8	19.6
Gln	Asp	Q	D	N	P	N	N	71.0	-2.9	5.7	58.0	-3.0	2.8	1.1	19.2	7.8	19.8
Gln	Cys	Q	C	N	N	N	N	71.0	-2.9	5.7	46.0	0.2	5.1	1.1	19.2	9.5	14.5
Gln	Gln	Q	Q	N	N	N	P	71.0	-2.9	5.7	71.0	-2.9	5.7	1.1	17.3	10.0	20.0
Gln	Glu	Q	E	N	N	N	N	71.0	-2.9	5.7	72.0	-2.6	3.2	1.1	17.2	8.1	19.5
Gln	Gly	Q	G	N	P	N	N	71.0	-2.9	5.7	0.0	0.7	6.0	2.1	12.7	10.2	13.6
Gln	His	Q	H	N	N	N	N	71.0	-2.9	5.7	80.0	-1.7	7.6	1.1	16.1	11.4	17.8
Gln	Ile	Q	I	N	N	N	N	71.0	-2.9	5.7	56.0	3.1	5.9	1.6	19.4	10.2	9.2
Gln	Leu	Q	L	N	N	P	P	71.0	-2.9	5.7	56.0	2.2	6.0	3.2	19.4	10.2	10.9
Gln	Lys	Q	K	N	N	N	N	71.0	-2.9	5.7	71.0	-4.6	9.7	1.1	17.3	13.1	17.0
Gln	Met	Q	M	N	N	N	N	71.0	-2.9	5.7	74.0	1.1	5.7	0.5	16.9	10.0	12.8
Gln	Phe	Q	F	N	N	N	N	71.0	-2.9	5.7	90.0	2.5	5.5	1.1	14.7	9.8	10.3
Gln	Pro	Q	P	N	N	N	P	71.0	-2.9	5.7	40.0	-0.3	6.5	2.1	18.3	10.6	15.3
Gln	Ser	Q	S	N	N	N	P	71.0	-2.9	5.7	30.0	-1.1	5.7	3.2	16.9	10.0	16.8
Gln	Thr	Q	T	N	N	N	N	71.0	-2.9	5.7	44.0	-0.8	5.9	2.1	18.9	10.1	16.1
Gln	Trp	Q	W	N	N	N	P	71.0	-2.9	5.7	129.0	1.5	6.0	0.5	9.2	10.2	12.1
Gln	Tyr	Q	Y	N	N	N	N	71.0	-2.9	5.7	106.0	0.1	5.7	1.1	12.4	10.0	14.7
Gln	Val	Q	V	P	P	N	N	71.0	-2.9	5.7	42.0	2.3	6.0	2.1	18.6	10.2	10.7
Glu	Ala	E	A	N	N	N	N	72.0	-2.6	3.2	14.0	1.0	6.0	2.1	14.8	8.9	13.5
Glu	Arg	E	R	N	P	N	P	72.0	-2.6	3.2	99.0	-7.5	10.8	3.2	13.2	19.0	19.3
Glu	Asn	E	N	N	N	N	N	72.0	-2.6	3.2	57.0	-2.7	5.4	1.1	19.2	7.6	19.8
Glu	Asp	E	D	N	N	N	N	72.0	-2.6	3.2	58.0	-3.0	2.8	1.1	19.0	2.0	19.3
Glu	Cys	E	C	N	N	N	P	72.0	-2.6	3.2	46.0	0.2	5.1	1.1	19.3	6.9	15.0
Glu	Gln	E	Q	N	N	N	N	72.0	-2.6	3.2	71.0	-2.9	5.7	1.1	17.2	8.1	19.5
Glu	Glu	E	E	N	N	N	N	72.0	-2.6	3.2	72.0	-2.6	3.2	1.1	17.3	3.0	20.0
Glu	Gly	E	G	N	N	N	N	72.0	-2.6	3.2	0.0	0.7	6.0	2.1	12.8	8.8	14.1
Glu	His	E	H	N	P	N	P	72.0	-2.6	3.2	80.0	-1.7	7.6	1.1	15.9	12.3	18.4
Glu	Ile	E	I	N	N	N	N	72.0	-2.6	3.2	56.0	3.1	5.9	1.6	19.3	8.7	9.8
Glu	Leu	E	L	P	P	P	P	72.0	-2.6	3.2	56.0	2.2	6.0	3.2	19.3	8.8	11.4
Glu	Lys	E	K	N	N	N	N	72.0	-2.6	3.2	71.0	-4.6	9.7	1.1	17.2	16.8	16.4
Glu	Met	E	M	N	N	N	N	72.0	-2.6	3.2	74.0	1.1	5.7	0.5	16.8	8.3	13.4
Glu	Phe	E	F	N	N	N	P	72.0	-2.6	3.2	90.0	2.5	5.5	1.1	14.5	7.8	10.9
Glu	Pro	E	P	N	P	N	N	72.0	-2.6	3.2	40.0	-0.3	6.5	2.1	18.5	9.9	15.9
Glu	Ser	E	S	N	N	N	P	72.0	-2.6	3.2	30.0	-1.1	5.7	3.2	17.0	8.2	17.3
Glu	Thr	E	T	N	N	N	N	72.0	-2.6	3.2	44.0	-0.8	5.9	2.1	19.0	8.6	16.7
Glu	Trp	E	W	N	N	N	N	72.0	-2.6	3.2	129.0	1.5	6.0	0.5	9.0	8.8	12.7
Glu	Tyr	E	Y	N	N	N	P	72.0	-2.6	3.2	106.0	0.1	5.7	1.1	12.3	8.2	15.2
Glu	Val	E	V	N	N	N	N	72.0	-2.6	3.2	42.0	2.3	6.0	2.1	18.7	8.8	11.2
Gly	Ala	G	A	N	N	P	P	0.0	0.7	6.0	14.0	1.0	6.0	4.3	4.7	10.4	19.4
Gly	Arg	G	R	N	P	N	P	0.0	0.7	6.0	99.0	-7.5	10.8	6.4	16.6	13.2	5.4
Gly	Asn	G	N	N	N	N	P	0.0	0.7	6.0	57.0	-2.7	5.4	2.1	10.7	10.1	14.0
Gly	Asp	G	D	N	N	N	P	0.0	0.7	6.0	58.0	-3.0	2.8	2.1	10.9	8.6	13.4
Gly	Cys	G	C	N	N	N	P	0.0	0.7	6.0	46.0	0.2	5.1	2.1	9.2	9.9	19.1
Gly	Gln	G	Q	N	P	N	N	0.0	0.7	6.0	71.0	-2.9	5.7	2.1	12.7	10.2	13.6
Gly	Glu	G	E	N	N	N	N	0.0	0.7	6.0	72.0	-2.6	3.2	2.1	12.8	8.8	14.1
Gly	Gly	G	G	N	N	N	N	0.0	0.7	6.0	0.0	0.7	6.0	4.3	2.7	10.4	20.0
Gly	His	G	H	N	P	N	P	0.0	0.7	6.0	80.0	-1.7	7.6	2.1	13.9	11.3	15.8
Gly	Ile	G	I	N	N	N	P	0.0	0.7	6.0	56.0	3.1	5.9	3.2	10.6	10.4	15.6
Gly	Leu	G	L	N	P	N	P	0.0	0.7	6.0	56.0	2.2	6.0	6.4	10.6	10.4	17.3
Gly	Lys	G	K	N	N	N	N	0.0	0.7	6.0	71.0	-4.6	9.7	2.1	12.7	12.6	10.6
Gly	Met	G	M	N	N	N	N	0.0	0.7	6.0	74.0	1.1	5.7	1.1	13.1	10.3	19.2
Gly	Phe	G	F	N	N	N	P	0.0	0.7	6.0	90.0	2.5	5.5	2.1	15.4	10.1	16.7
Gly	Pro	G	P	P	P	P	P	0.0	0.7	6.0	40.0	-0.3	6.5	4.3	8.3	10.7	18.3
Gly	Ser	G	S	N	N	N	P	0.0	0.7	6.0	30.0	-1.1	5.7	6.4	6.9	10.2	16.8
Gly	Thr	G	T	N	N	N	P	0.0	0.7	6.0	44.0	-0.8	5.9	4.3	8.9	10.3	17.5
Gly	Trp	G	W	N	N	N	N	0.0	0.7	6.0	129.0	1.5	6.0	1.1	19.2	10.4	18.5
Gly	Tyr	G	Y	N	N	N	P	0.0	0.7	6.0	106.0	0.1	5.7	2.1	17.6	10.2	18.9
Gly	Val	G	V	N	N	N	P	0.0	0.7	6.0	42.0	2.3	6.0	4.3	8.6	10.4	17.1

First	Second	1st	2nd	PC1_C	PC2_C	PC1_RC	PC2_RC	mw-57_1	HP_1	pI_1	mw-57_2	HP_2	pI_2	NF'/1k	SCI	CCI	HCI
His	Ala	H	A	N	P	N	P	80.0	-1.7	7.6	14.0	1.0	6.0	2.1	15.9	11.3	15.2
His	Arg	H	R	N	N	N	P	80.0	-1.7	7.6	99.0	-7.5	10.8	3.2	12.1	9.8	9.6
His	Asn	H	N	N	N	N	N	80.0	-1.7	7.6	57.0	-2.7	5.4	1.1	18.0	11.5	18.2
His	Asp	H	D	N	N	N	N	80.0	-1.7	7.6	58.0	-3.0	2.8	1.1	17.9	12.4	17.7
His	Cys	H	C	N	N	N	N	80.0	-1.7	7.6	46.0	0.2	5.1	1.1	19.6	11.6	16.6
His	Gln	H	Q	N	N	N	N	80.0	-1.7	7.6	71.0	-2.9	5.7	1.1	16.1	11.4	17.8
His	Glu	H	E	N	P	N	P	80.0	-1.7	7.6	72.0	-2.6	3.2	1.1	15.9	12.3	18.4
His	Gly	H	G	N	P	N	P	80.0	-1.7	7.6	0.0	0.7	6.0	2.1	13.9	11.3	15.8
His	His	H	H	N	N	N	N	80.0	-1.7	7.6	80.0	-1.7	7.6	1.1	14.8	10.8	20.0
His	Ile	H	I	N	N	N	N	80.0	-1.7	7.6	56.0	3.1	5.9	1.6	18.2	11.4	11.4
His	Leu	H	L	N	N	N	N	80.0	-1.7	7.6	56.0	2.2	6.0	3.2	18.2	11.3	13.0
His	Lys	H	K	N	N	N	P	80.0	-1.7	7.6	71.0	-4.6	9.7	1.1	16.1	10.1	14.8
His	Met	H	M	N	N	P	P	80.0	-1.7	7.6	74.0	1.1	5.7	0.5	15.6	11.4	15.0
His	Phe	H	F	N	N	N	N	80.0	-1.7	7.6	90.0	2.5	5.5	1.1	13.4	11.5	12.5
His	Pro	H	P	N	N	N	N	80.0	-1.7	7.6	40.0	-0.3	6.5	2.1	19.6	11.2	17.5
His	Ser	H	S	N	N	N	N	80.0	-1.7	7.6	30.0	-1.1	5.7	3.2	18.2	11.4	18.9
His	Thr	H	T	N	N	N	P	80.0	-1.7	7.6	44.0	-0.8	5.9	2.1	19.9	11.4	18.3
His	Trp	H	W	N	N	N	N	80.0	-1.7	7.6	129.0	1.5	6.0	0.5	7.9	11.3	14.3
His	Tyr	H	Y	N	N	N	N	80.0	-1.7	7.6	106.0	0.1	5.7	1.1	11.1	11.4	16.8
His	Val	H	V	P	P	P	P	80.0	-1.7	7.6	42.0	2.3	6.0	2.1	19.9	11.3	12.8
Ile	Ala	I	A	N	N	N	P	56.0	3.1	5.9	14.0	1.0	6.0	3.2	12.5	10.4	16.2
Ile	Arg	I	R	N	N	N	N	56.0	3.1	5.9	99.0	-7.5	10.8	4.8	15.5	13.3	1.0
Ile	Asn	I	N	N	N	P	P	56.0	3.1	5.9	57.0	-2.7	5.4	1.6	18.6	10.0	9.6
Ile	Asp	I	D	N	N	P	N	56.0	3.1	5.9	58.0	-3.0	2.8	1.6	18.7	8.4	9.1
Ile	Cys	I	C	N	P	N	P	56.0	3.1	5.9	46.0	0.2	5.1	1.6	17.0	9.8	14.7
Ile	Gln	I	Q	N	N	N	N	56.0	3.1	5.9	71.0	-2.9	5.7	1.6	19.4	10.2	9.2
Ile	Glu	I	E	N	N	N	N	56.0	3.1	5.9	72.0	-2.6	3.2	1.6	19.3	8.7	9.8
Ile	Gly	I	G	N	N	N	P	56.0	3.1	5.9	0.0	0.7	6.0	3.2	10.6	10.4	15.6
Ile	His	I	H	N	N	N	N	56.0	3.1	5.9	80.0	-1.7	7.6	1.6	18.2	11.4	11.4
Ile	Ile	I	I	N	N	N	P	56.0	3.1	5.9	56.0	3.1	5.9	2.4	18.5	10.3	20.0
Ile	Leu	I	L	N	P	N	N	56.0	3.1	5.9	56.0	2.2	6.0	4.8	18.5	10.4	18.4
Ile	Lys	I	K	N	N	N	N	56.0	3.1	5.9	71.0	-4.6	9.7	1.6	19.4	12.7	6.2
Ile	Met	I	M	N	N	N	N	56.0	3.1	5.9	74.0	1.1	5.7	0.8	19.0	10.2	16.4
Ile	Phe	I	F	N	P	N	P	56.0	3.1	5.9	90.0	2.5	5.5	1.6	16.8	10.1	18.9
Ile	Pro	I	P	N	N	N	N	56.0	3.1	5.9	40.0	-0.3	6.5	3.2	16.2	10.7	13.9
Ile	Ser	I	S	N	P	N	N	56.0	3.1	5.9	30.0	-1.1	5.7	4.8	14.8	10.2	12.5
Ile	Thr	I	T	N	N	N	P	56.0	3.1	5.9	44.0	-0.8	5.9	3.2	16.8	10.3	13.1
Ile	Trp	I	W	N	P	N	N	56.0	3.1	5.9	129.0	1.5	6.0	0.8	11.3	10.4	17.1
Ile	Tyr	I	Y	P	P	P	P	56.0	3.1	5.9	106.0	0.1	5.7	1.6	14.5	10.2	14.6
Ile	Val	I	V	N	N	N	P	56.0	3.1	5.9	42.0	2.3	6.0	3.2	16.5	10.4	18.6
Leu	Ala	L	A	N	P	N	P	56.0	2.2	6.0	14.0	1.0	6.0	6.4	12.5	10.4	17.8
Leu	Arg	L	R	N	N	N	P	56.0	2.2	6.0	99.0	-7.5	10.8	9.7	15.5	13.2	2.6
Leu	Asn	L	N	P	P	N	N	56.0	2.2	6.0	57.0	-2.7	5.4	3.2	18.6	10.1	11.2
Leu	Asp	L	D	P	P	N	N	56.0	2.2	6.0	58.0	-3.0	2.8	3.2	18.7	8.6	10.7
Leu	Cys	L	C	N	N	N	N	56.0	2.2	6.0	46.0	0.2	5.1	3.2	17.0	9.9	16.4
Leu	Gln	L	Q	N	N	N	P	56.0	2.2	6.0	71.0	-2.9	5.7	3.2	19.4	10.2	10.9
Leu	Glu	L	E	P	P	P	P	56.0	2.2	6.0	72.0	-2.6	3.2	3.2	19.3	8.8	11.4
Leu	Gly	L	G	N	P	N	P	56.0	2.2	6.0	0.0	0.7	6.0	6.4	10.6	10.4	17.3
Leu	His	L	H	N	N	N	N	56.0	2.2	6.0	80.0	-1.7	7.6	3.2	18.2	11.3	13.0
Leu	Ile	L	I	N	P	N	N	56.0	2.2	6.0	56.0	3.1	5.9	4.8	18.5	10.4	18.4
Leu	Leu	L	L	N	N	N	P	56.0	2.2	6.0	56.0	2.2	6.0	9.7	18.5	10.4	20.0
Leu	Lys	L	K	N	N	P	P	56.0	2.2	6.0	71.0	-4.6	9.7	3.2	19.4	12.6	7.8
Leu	Met	L	M	N	N	N	P	56.0	2.2	6.0	74.0	1.1	5.7	1.6	19.0	10.3	18.0
Leu	Phe	L	F	N	N	N	N	56.0	2.2	6.0	90.0	2.5	5.5	3.2	16.8	10.1	19.5
Leu	Pro	L	P	N	N	N	P	56.0	2.2	6.0	40.0	-0.3	6.5	6.4	16.2	10.7	15.5
Leu	Ser	L	S	N	P	N	P	56.0	2.2	6.0	30.0	-1.1	5.7	9.7	14.8	10.2	14.1
Leu	Thr	L	T	N	P	N	P	56.0	2.2	6.0	44.0	-0.8	5.9	6.4	16.8	10.4	14.7
Leu	Trp	L	W	N	N	N	P	56.0	2.2	6.0	129.0	1.5	6.0	1.6	11.3	10.4	18.7
Leu	Tyr	L	Y	N	N	N	N	56.0	2.2	6.0	106.0	0.1	5.7	3.2	14.5	10.2	16.2
Leu	Val	L	V	N	P	N	P	56.0	2.2	6.0	42.0	2.3	6.0	6.4	16.5	10.4	19.8
Lys	Ala	K	A	N	N	N	N	71.0	-4.6	9.7	14.0	1.0	6.0	2.1	14.7	12.5	10.0
Lys	Arg	K	R	N	N	N	P	71.0	-4.6	9.7	99.0	-7.5	10.8	3.2	13.4	5.2	14.8
Lys	Asn	K	N	N	N	N	N	71.0	-4.6	9.7	57.0	-2.7	5.4	1.1	19.3	13.4	16.6
Lys	Asp	K	D	N	N	N	N	71.0	-4.6	9.7	58.0	-3.0	2.8	1.1	19.2	17.5	17.1
Lys	Cys	K	C	N	P	N	P	71.0	-4.6	9.7	46.0	0.2	5.1	1.1	19.2	14.0	11.5
Lys	Gln	K	Q	N	N	N	N	71.0	-4.6	9.7	71.0	-2.9	5.7	1.1	17.3	13.1	17.0
Lys	Glu	K	E	N	N	N	N	71.0	-4.6	9.7	72.0	-2.6	3.2	1.1	17.2	16.8	16.4
Lys	Gly	K	G	N	N	N	N	71.0	-4.6	9.7	0.0	0.7	6.0	2.1	12.7	12.6	10.6
Lys	His	K	H	N	N	N	P	71.0	-4.6	9.7	80.0	-1.7	7.6	1.1	16.1	10.1	14.8
Lys	Ile	K	I	N	N	N	N	71.0	-4.6	9.7	56.0	3.1	5.9	1.6	19.4	12.7	6.2
Lys	Leu	K	L	N	N	P	N	71.0	-4.6	9.7	56.0	2.2	6.0	3.2	19.4	12.6	7.8
Lys	Lys	K	K	N	N	N	N	71.0	-4.6	9.7	71.0	-4.6	9.7	1.1	17.3	6.8	20.0
Lys	Met	K	M	N	N	N	N	71.0	-4.6	9.7	74.0	1.1	5.7	0.5	16.9	12.9	9.8
Lys	Phe	K	F	P	P	P	P	71.0	-4.6	9.7	90.0	2.5	5.5	1.1	14.7	13.3	7.3
Lys	Pro	K	P	N	N	N	N	71.0	-4.6	9.7	40.0	-0.3	6.5	2.1	18.3	11.8	12.3
Lys	Ser	K	S	N	P	N	P	71.0	-4.6	9.7	30.0	-1.1	5.7	3.2	16.9	13.0	13.7
Lys	Thr	K	T	N	N	N	N	71.0	-4.6	9.7	44.0	-0.8	5.9	2.1	18.9	12.7	13.1
Lys	Trp	K	W	N	N	N	N	71.0	-4.6	9.7	129.0	1.5	6.0	0.5	9.2	12.6	9.1
Lys	Tyr	K	Y	N	P	N	P	71.0	-4.6	9.7	106.0	0.1	5.7	1.1	12.4	13.1	11.6
Lys	Val	K	V	N	N	N	N	71.0	-4.6	9.7	42.0	2.3	6.0	2.1	18.6	12.6	7.6

Table 4. Continued

First	Second	1st	2nd	PC1_C	PC2_C	PC1_RC	PC2_RC	mw-57_1	HP_1	pI_1	mw-57_2	HP_2	pI_2	NF*/1k	SCI	CCI	HCI
Met	Ala	M	A	N	N	N	N	74.0	1.1	5.7	14.0	1.0	6.0	1.1	15.1	10.3	19.8
Met	Arg	M	R	N	N	N	P	74.0	1.1	5.7	99.0	-7.5	10.8	1.6	13.0	13.7	4.6
Met	Asn	M	N	N	N	N	N	74.0	1.1	5.7	57.0	-2.7	5.4	0.5	18.9	9.9	13.2
Met	Asp	M	D	N	N	N	N	74.0	1.1	5.7	58.0	-3.0	2.8	0.5	18.7	8.0	12.7
Met	Cys	M	C	N	P	N	N	74.0	1.1	5.7	46.0	0.2	5.1	0.5	19.6	9.6	18.3
Met	Gln	M	Q	N	N	N	N	74.0	1.1	5.7	71.0	-2.9	5.7	0.5	16.9	10.0	12.8
Met	Glu	M	E	N	N	N	N	74.0	1.1	5.7	72.0	-2.6	3.2	0.5	16.8	8.3	13.4
Met	Gly	M	G	N	N	N	N	74.0	1.1	5.7	0.0	0.7	6.0	1.1	13.1	10.3	19.2
Met	His	M	H	N	N	P	P	74.0	1.1	5.7	80.0	-1.7	7.6	0.5	15.6	11.4	15.0
Met	Ile	M	I	N	N	N	N	74.0	1.1	5.7	56.0	3.1	5.9	0.8	19.0	10.2	16.4
Met	Leu	M	L	N	N	N	P	74.0	1.1	5.7	56.0	2.2	6.0	1.6	19.0	10.3	18.0
Met	Lys	M	K	N	N	N	N	74.0	1.1	5.7	71.0	-4.6	9.7	0.5	16.9	12.9	9.8
Met	Met	M	M	N	N	N	N	74.0	1.1	5.7	74.0	1.1	5.7	0.3	16.5	10.1	20.0
Met	Phe	M	F	N	P	N	N	74.0	1.1	5.7	90.0	2.5	5.5	0.5	14.2	9.9	17.5
Met	Pro	M	P	N	N	N	P	74.0	1.1	5.7	40.0	-0.3	6.5	1.1	18.7	10.6	17.5
Met	Ser	M	S	N	P	N	N	74.0	1.1	5.7	30.0	-1.1	5.7	1.6	17.3	10.1	16.1
Met	Thr	M	T	N	N	N	N	74.0	1.1	5.7	44.0	-0.8	5.9	1.1	19.3	10.2	16.7
Met	Trp	M	W	N	N	N	N	74.0	1.1	5.7	129.0	1.5	6.0	0.3	8.7	10.3	19.3
Met	Tyr	M	Y	P	P	N	N	74.0	1.1	5.7	106.0	0.1	5.7	0.5	12.0	10.1	18.2
Met	Val	M	V	N	N	N	N	74.0	1.1	5.7	42.0	2.3	6.0	1.1	19.0	10.3	17.8
Phe	Ala	F	A	N	N	N	P	90.0	2.5	5.5	14.0	1.0	6.0	2.1	17.3	10.2	17.3
Phe	Arg	F	R	N	P	N	P	90.0	2.5	5.5	99.0	-7.5	10.8	3.2	10.7	14.2	2.1
Phe	Asn	F	N	N	N	N	N	90.0	2.5	5.5	57.0	-2.7	5.4	1.1	16.6	9.6	10.7
Phe	Asp	F	D	N	N	N	N	90.0	2.5	5.5	58.0	-3.0	2.8	1.1	16.5	7.4	10.1
Phe	Cys	F	C	N	N	N	N	90.0	2.5	5.5	46.0	0.2	5.1	1.1	18.2	9.4	15.8
Phe	Gln	F	Q	N	N	N	N	90.0	2.5	5.5	71.0	-2.9	5.7	1.1	14.7	9.8	10.3
Phe	Glu	F	E	N	N	P	P	90.0	2.5	5.5	72.0	-2.6	3.2	1.1	14.5	7.8	10.9
Phe	Gly	F	G	N	N	N	P	90.0	2.5	5.5	0.0	0.7	6.0	2.1	15.4	10.1	16.7
Phe	His	F	H	N	N	N	N	90.0	2.5	5.5	80.0	-1.7	7.6	1.1	13.4	11.5	12.5
Phe	Ile	F	I	N	P	N	P	90.0	2.5	5.5	56.0	3.1	5.9	1.6	16.8	10.1	18.9
Phe	Leu	F	L	N	N	N	N	90.0	2.5	5.5	56.0	2.2	6.0	3.2	16.8	10.1	19.5
Phe	Lys	F	K	P	P	P	P	90.0	2.5	5.5	71.0	-4.6	9.7	1.1	14.7	13.3	7.3
Phe	Met	F	M	N	P	N	N	90.0	2.5	5.5	74.0	1.1	5.7	0.5	14.2	9.9	17.5
Phe	Phe	F	F	N	N	N	N	90.0	2.5	5.5	90.0	2.5	5.5	1.1	12.0	9.7	20.0
Phe	Pro	F	P	N	N	N	N	90.0	2.5	5.5	40.0	-0.3	6.5	2.1	19.0	10.6	15.0
Phe	Ser	F	S	N	N	N	N	90.0	2.5	5.5	30.0	-1.1	5.7	3.2	19.6	9.9	13.5
Phe	Thr	F	T	N	P	N	P	90.0	2.5	5.5	44.0	-0.8	5.9	2.1	18.5	10.0	14.2
Phe	Trp	F	W	N	N	N	N	90.0	2.5	5.5	129.0	1.5	6.0	0.5	6.5	10.1	18.2
Phe	Tyr	F	Y	N	N	N	N	90.0	2.5	5.5	106.0	0.1	5.7	1.1	9.7	9.9	15.7
Phe	Val	F	V	N	N	N	P	90.0	2.5	5.5	42.0	2.3	6.0	2.1	18.7	10.1	19.6
Pro	Ala	P	A	N	P	N	P	40.0	-0.3	6.5	14.0	1.0	6.0	4.3	10.3	10.7	17.7
Pro	Arg	P	R	N	N	P	P	40.0	-0.3	6.5	99.0	-7.5	10.8	6.4	17.7	12.1	7.1
Pro	Asn	P	N	N	N	N	N	40.0	-0.3	6.5	57.0	-2.7	5.4	2.1	16.3	10.5	15.7
Pro	Asp	P	D	N	P	N	N	40.0	-0.3	6.5	58.0	-3.0	2.8	2.1	16.5	9.8	15.1
Pro	Cys	P	C	N	N	N	N	40.0	-0.3	6.5	46.0	0.2	5.1	2.1	14.8	10.4	19.2
Pro	Gln	P	Q	N	N	N	P	40.0	-0.3	6.5	71.0	-2.9	5.7	2.1	18.3	10.6	15.3
Pro	Glu	P	E	N	P	N	N	40.0	-0.3	6.5	72.0	-2.6	3.2	2.1	18.5	9.9	15.9
Pro	Gly	P	G	P	P	P	P	40.0	-0.3	6.5	0.0	0.7	6.0	4.3	8.3	10.7	18.3
Pro	His	P	H	N	N	N	N	40.0	-0.3	6.5	80.0	-1.7	7.6	2.1	19.6	11.2	17.5
Pro	Ile	P	I	N	N	N	N	40.0	-0.3	6.5	56.0	3.1	5.9	3.2	16.2	10.7	13.9
Pro	Leu	P	L	N	N	N	P	40.0	-0.3	6.5	56.0	2.2	6.0	6.4	16.2	10.7	15.5
Pro	Lys	P	K	N	N	N	P	40.0	-0.3	6.5	71.0	-4.6	9.7	2.1	18.3	11.8	12.3
Pro	Met	P	M	N	N	N	P	40.0	-0.3	6.5	74.0	1.1	5.7	1.1	18.7	10.6	17.5
Pro	Phe	P	F	N	N	N	N	40.0	-0.3	6.5	90.0	2.5	5.5	2.1	19.0	10.6	15.0
Pro	Pro	P	P	N	N	N	P	40.0	-0.3	6.5	40.0	-0.3	6.5	4.3	13.9	10.8	20.0
Pro	Ser	P	S	N	N	N	N	40.0	-0.3	6.5	30.0	-1.1	5.7	6.4	12.5	10.6	18.5
Pro	Thr	P	T	N	N	N	P	40.0	-0.3	6.5	44.0	-0.8	5.9	3.2	14.5	10.7	19.2
Pro	Trp	P	W	N	N	P	N	40.0	-0.3	6.5	129.0	1.5	6.0	1.1	13.5	10.7	16.8
Pro	Tyr	P	Y	N	N	N	N	40.0	-0.3	6.5	106.0	0.1	5.7	2.1	16.8	10.6	19.3
Pro	Val	P	V	N	P	N	P	40.0	-0.3	6.5	42.0	2.3	6.0	4.3	14.2	10.7	15.4
Ser	Ala	S	A	N	N	P	P	30.0	-1.1	5.7	14.0	1.0	6.0	6.4	8.9	10.3	16.2
Ser	Arg	S	R	P	P	P	P	30.0	-1.1	5.7	99.0	-7.5	10.8	9.7	19.2	13.8	8.5
Ser	Asn	S	N	N	P	N	P	30.0	-1.1	5.7	57.0	-2.7	5.4	3.2	14.9	9.8	17.1
Ser	Asp	S	D	N	N	N	P	30.0	-1.1	5.7	58.0	-3.0	2.8	3.2	15.1	7.9	16.6
Ser	Cys	S	C	N	N	N	N	30.0	-1.1	5.7	46.0	0.2	5.1	3.2	13.4	9.6	17.7
Ser	Gln	S	Q	N	N	N	P	30.0	-1.1	5.7	71.0	-2.9	5.7	3.2	16.9	10.0	16.8
Ser	Glu	S	E	N	N	N	P	30.0	-1.1	5.7	72.0	-2.6	3.2	3.2	17.0	8.2	17.3
Ser	Gly	S	G	N	N	P	P	30.0	-1.1	5.7	0.0	0.7	6.0	6.4	6.9	10.2	16.8
Ser	His	S	H	N	N	N	N	30.0	-1.1	5.7	80.0	-1.7	7.6	3.2	18.2	11.4	18.9
Ser	Ile	S	I	N	P	N	P	30.0	-1.1	5.7	56.0	3.1	5.9	4.8	14.8	10.2	12.5
Ser	Leu	S	L	N	P	N	P	30.0	-1.1	5.7	56.0	2.2	6.0	9.7	14.8	10.2	14.1
Ser	Lys	S	K	N	P	N	P	30.0	-1.1	5.7	71.0	-4.6	9.7	3.2	16.9	13.0	13.7
Ser	Met	S	M	N	P	N	N	30.0	-1.1	5.7	74.0	1.1	5.7	1.6	17.3	10.1	16.1
Ser	Phe	S	F	N	N	N	N	30.0	-1.1	5.7	90.0	2.5	5.5	3.2	19.6	9.9	13.5
Ser	Pro	S	P	N	N	N	P	30.0	-1.1	5.7	40.0	-0.3	6.5	6.4	12.5	10.6	18.5
Ser	Ser	S	S	P	P	N	N	30.0	-1.1	5.7	30.0	-1.1	5.7	9.7	11.1	10.0	20.0
Ser	Thr	S	T	N	P	P	P	30.0	-1.1	5.7	44.0	-0.8	5.9	6.4	13.1	10.2	19.4
Ser	Trp	S	W	N	P	N	N	30.0	-1.1	5.7	129.0	1.5	6.0	1.6	14.9	10.2	15.3
Ser	Tyr	S	Y	N	N	N	N	30.0	-1.1	5.7	106.0	0.1	5.7	3.2	18.2	10.0	17.9
Ser	Val	S	V	N	N	N	P	30.0	-1.1	5.7	42.0	2.3	6.0	6.4	12.8	10.2	13.9

First	Second	1st	2nd	PC1_C	PC2_C	PC1_RC	PC2_RC	mw-57_1	HP_1	pI_1	mw-57_2	HP_2	pI_2	NF*/1k	SCI	CCI	HCI
Thr	Ala	T	A	N	N	N	P	44.0	-0.8	5.9	14.0	1.0	6.0	4.3	10.9	10.4	16.9
Thr	Arg	T	R	N	N	P	P	44.0	-0.8	5.9	99.0	-7.5	10.8	6.4	17.2	13.4	7.9
Thr	Asn	T	N	N	N	N	P	44.0	-0.8	5.9	57.0	-2.7	5.4	2.1	16.9	10.0	16.5
Thr	Asp	T	D	N	N	N	P	44.0	-0.8	5.9	58.0	-3.0	2.8	2.1	17.0	8.3	16.0
Thr	Cys	T	C	P	P	P	P	44.0	-0.8	5.9	46.0	0.2	5.1	2.1	15.4	9.8	18.4
Thr	Gln	T	Q	N	N	N	N	44.0	-0.8	5.9	71.0	-2.9	5.7	2.1	18.9	10.1	16.1
Thr	Glu	T	E	N	N	N	N	44.0	-0.8	5.9	72.0	-2.6	3.2	2.1	19.0	8.6	16.7
Thr	Gly	T	G	N	N	P	P	44.0	-0.8	5.9	0.0	0.7	6.0	4.3	8.9	10.3	17.5
Thr	His	T	H	N	N	N	P	44.0	-0.8	5.9	80.0	-1.7	7.6	2.1	19.9	11.4	18.3
Thr	Ile	T	I	N	N	N	P	44.0	-0.8	5.9	56.0	3.1	5.9	3.2	16.8	10.3	13.1
Thr	Leu	T	L	N	P	N	P	44.0	-0.8	5.9	56.0	2.2	6.0	6.4	16.8	10.4	14.7
Thr	Lys	T	K	N	N	N	N	44.0	-0.8	5.9	71.0	-4.6	9.7	2.1	18.9	12.7	13.1
Thr	Met	T	M	N	N	N	N	44.0	-0.8	5.9	74.0	1.1	5.7	1.1	19.3	10.2	16.7
Thr	Phe	T	F	N	P	N	P	44.0	-0.8	5.9	90.0	2.5	5.5	2.1	18.5	10.0	14.2
Thr	Pro	T	P	N	N	N	P	44.0	-0.8	5.9	40.0	-0.3	6.5	4.3	14.5	10.7	19.2
Thr	Ser	T	S	N	P	P	P	44.0	-0.8	5.9	30.0	-1.1	5.7	6.4	13.1	10.2	19.4
Thr	Thr	T	T	N	N	N	P	44.0	-0.8	5.9	44.0	-0.8	5.9	4.3	15.1	10.3	20.0
Thr	Trp	T	W	P	P	N	N	44.0	-0.8	5.9	129.0	1.5	6.0	1.1	13.0	10.2	16.0
Thr	Tyr	T	Y	N	P	N	P	44.0	-0.8	5.9	106.0	0.1	5.7	2.1	16.2	10.1	18.5
Thr	Val	T	V	N	N	N	P	44.0	-0.8	5.9	42.0	2.3	6.0	4.3	14.8	10.3	14.5
Trp	Ala	W	A	N	N	N	N	44.0	-0.8	5.9	14.0	1.0	6.0	1.1	17.2	10.4	19.1
Trp	Arg	W	R	N	N	N	P	129.0	1.5	6.0	99.0	-7.5	10.8	1.6	5.2	13.2	3.9
Trp	Asn	W	N	N	P	N	N	129.0	1.5	6.0	57.0	-2.7	5.4	0.5	11.1	10.1	12.5
Trp	Asp	W	D	N	N	N	N	44.0	-0.8	5.9	58.0	-3.0	2.8	0.5	11.0	8.6	11.9
Trp	Cys	W	C	N	N	N	N	44.0	-0.8	5.9	46.0	0.2	5.1	0.5	12.7	9.9	17.6
Trp	Gln	W	Q	N	N	N	P	129.0	1.5	6.0	71.0	-2.9	5.7	0.5	9.2	10.2	12.1
Trp	Glu	W	E	N	N	N	N	44.0	-0.8	5.9	72.0	-2.6	3.2	0.5	9.0	8.8	12.7
Trp	Gly	W	G	N	N	N	N	44.0	-0.8	5.9	0.0	0.7	6.0	1.1	19.2	10.4	18.5
Trp	His	W	H	N	N	N	N	44.0	-0.8	5.9	80.0	-1.7	7.6	0.5	7.9	11.3	14.3
Trp	Ile	W	I	N	P	N	N	129.0	1.5	6.0	56.0	3.1	5.9	0.8	11.3	10.4	17.1
Trp	Leu	W	L	N	N	N	P	129.0	1.5	6.0	56.0	2.2	6.0	1.6	11.3	10.4	18.7
Trp	Lys	W	K	N	N	N	N	44.0	-0.8	5.9	71.0	-4.6	9.7	0.5	9.2	12.6	9.1
Trp	Met	W	M	N	N	N	N	44.0	-0.8	5.9	74.0	1.1	5.7	0.3	8.7	10.3	19.3
Trp	Phe	W	F	N	N	N	N	44.0	-0.8	5.9	90.0	2.5	5.5	0.5	6.5	10.1	18.2
Trp	Pro	W	P	N	N	P	P	129.0	1.5	6.0	40.0	-0.3	6.5	1.1	13.5	10.7	16.8
Trp	Ser	W	S	N	P	N	N	129.0	1.5	6.0	30.0	-1.1	5.7	1.6	14.9	10.2	15.3
Trp	Thr	W	T	P	P	N	N	129.0	1.5	6.0	44.0	-0.8	5.9	1.1	13.0	10.4	16.0
Trp	Trp	W	W	N	N	N	N	44.0	-0.8	5.9	129.0	1.5	6.0	0.3	1.0	10.4	20.0
Trp	Tyr	W	Y	N	N	N	N	44.0	-0.8	5.9	106.0	0.1	5.7	0.5	4.2	10.2	17.5
Trp	Val	W	V	N	N	N	N	44.0	-0.8	5.9	42.0	2.3	6.0	1.1	13.2	10.4	18.6
Tyr	Ala	Y	A	N	N	N	P	106.0	0.1	5.7	14.0	1.0	6.0	2.1	19.6	10.4	18.4
Tyr	Arg	Y	R	N	P	N	P	106.0	0.1	5.7	99.0	-7.5	10.8	3.2	8.5	13.8	6.4
Tyr	Asn	Y	N	N	N	N	N	106.0	0.1	5.7	57.0	-2.7	5.4	1.1	14.4	9.8	15.0
Tyr	Asp	Y	D	N	N	N	N	106.0	0.1	5.7	58.0	-3.0	2.8	1.1	14.2	7.8	14.5
Tyr	Cys	Y	C	N	N	N	N	106.0	0.1	5.7	46.0	0.2	5.1	1.1	15.9	9.5	19.8
Tyr	Gln	Y	Q	N	N	N	N	106.0	0.1	5.7	71.0	-2.9	5.7	1.1	12.4	10.0	14.7
Tyr	Glu	Y	E	N	N	N	P	106.0	0.1	5.7	72.0	-2.6	3.2	1.1	12.3	8.2	15.2
Tyr	Gly	Y	G	N	N	N	P	106.0	0.1	5.7	0.0	0.7	6.0	2.1	17.6	10.2	18.9
Tyr	His	Y	H	N	N	N	N	106.0	0.1	5.7	80.0	-1.7	7.6	1.1	11.1	11.4	16.8
Tyr	Ile	Y	I	P	P	P	P	106.0	0.1	5.7	56.0	3.1	5.9	1.6	14.5	10.2	14.6
Tyr	Leu	Y	L	N	N	N	N	106.0	0.1	5.7	56.0	2.2	6.0	3.2	14.5	10.2	16.2
Tyr	Lys	Y	K	N	P	N	P	106.0	0.1	5.7	71.0	-4.6	9.7	1.1	12.4	13.1	11.6
Tyr	Met	Y	M	P	P	N	N	106.0	0.1	5.7	74.0	1.1	5.7	0.5	12.0	10.1	18.2
Tyr	Phe	Y	F	N	N	N	N	106.0	0.1	5.7	90.0	2.5	5.5	1.1	9.7	10.1	15.7
Tyr	Pro	Y	P	N	N	N	N	106.0	0.1	5.7	40.0	-0.3	6.5	2.1	16.8	10.6	19.3
Tyr	Ser	Y	S	N	N	N	N	106.0	0.1	5.7	30.0	-1.1	5.7	3.2	18.2	10.0	17.9
Tyr	Thr	Y	T	N	P	N	P	106.0	0.1	5.7	44.0	-0.8	5.9	2.1	16.2	10.1	18.5
Tyr	Trp	Y	W	N	N	N	N	106.0	0.1	5.7	129.0	1.5	6.0	0.5	4.2	10.2	17.5
Tyr	Tyr	Y	Y	N	N	N	N	106.0	0.1	5.7	106.0	0.1	5.7	1.1	7.5	10.0	20.0
Tyr	Val	Y	V	N	N	P	P	106.0	0.1	5.7	42.0	2.3	6.0	2.1	16.5	10.2	16.0
Val	Ala	V	A	N	N	N	P	42.0	2.3	6.0	14.0	1.0	6.0	4.3	10.6	10.4	17.7
Val	Arg	V	R	N	P	N	P	42.0	2.3	6.0	99.0	-7.5	10.8	6.4	17.5	13.2	2.4
Val	Asn	V	N	N	N	P	P	42.0	2.3	6.0	57.0	-2.7	5.4	2.1	16.6	10.1	11.0
Val	Asp	V	D	N	N	P	P	42.0	2.3	6.0	58.0	-3.0	2.8	2.1	16.8	8.6	10.5
Val	Cys	V	C	N	N	N	P	42.0	2.3	6.0	46.0	0.2	5.1	2.1	15.1	9.9	16.2
Val	Gln	V	Q	P	P	N	N	42.0	2.3	6.0	71.0	-2.9	5.7	2.1	18.6	10.2	10.7
Val	Glu	V	E	N	N	N	P	42.0	2.3	6.0	72.0	-2.6	3.2	2.1	18.7	8.8	11.2
Val	Gly	V	G	N	N	N	P	42.0	2.3	6.0	0.0	0.7	6.0	4.3	8.6	10.4	17.1
Val	His	V	H	P	P	P	P	42.0	2.3	6.0	80.0	-1.7	7.6	2.1	19.9	11.3	12.8
Val	Ile	V	I	N	N	N	P	42.0	2.3	6.0	56.0	3.1	5.9	3.2	16.5	10.4	18.6
Val	Leu	V	L	N	P	N	P	42.0	2.3	6.0	56.0	2.2	6.0	6.4	16.5	10.4	19.8
Val	Lys	V	K	N	N	N	N	42.0	2.3	6.0	71.0	-4.6	9.7	2.1	18.6	12.6	7.6
Val	Met	V	M	N	N	N	N	42.0	2.3	6.0	74.0	1.1	5.7	1.1	19.0	10.3	17.8
Val	Phe	V	F	N	N	N	N	42.0	2.3	6.0	90.0	2.5	5.5	2.1	18.7	10.1	19.6
Val	Pro	V	P	N	P	N	P	42.0	2.3	6.0	40.0	-0.3	6.5	4.3	14.2	10.7	15.4
Val	Ser	V	S	N	N	N	P	42.0	2.3	6.0	30.0	-1.1	5.7	6.4	12.8	10.2	13.9
Val	Thr	V	T	N	N	N	P	42.0	2.3	6.0	44.0	-0.8	5.9	4.3	14.8	10.3	14.5
Val	Trp	V	W	N	N	N	N	42.0	2.3	6.0	129.0	1.5	6.0	1.1	13.2	10.4	18.6
Val	Tyr	V	Y	N	N	P	P	42.0	2.3	6.0	106.0	0.1	5.7	2.1	16.5	10.2	16.0
Val	Val	V	V	N	N	N	P	42.0	2.3	6.0	42.0	2.3	6.0	4.3	14.5	10.4	20.0

SYSTEM AND METHOD TO OBTAIN OLIGO-PEPTIDES WITH SPECIFIC HIGH AFFINITY TO QUERY PROTEINS

I have developed a system and method to obtain oligo-peptides with specific high affinity to query proteins [97]. The method is built on the second generation Proteomic Code which is based on the partially complementary coding of interacting amino acids.

Figures 31 and 32 show the steps for producing the target proteins of this system with a high affinity for query proteins, wherein the primary structure is known for the query proteins.
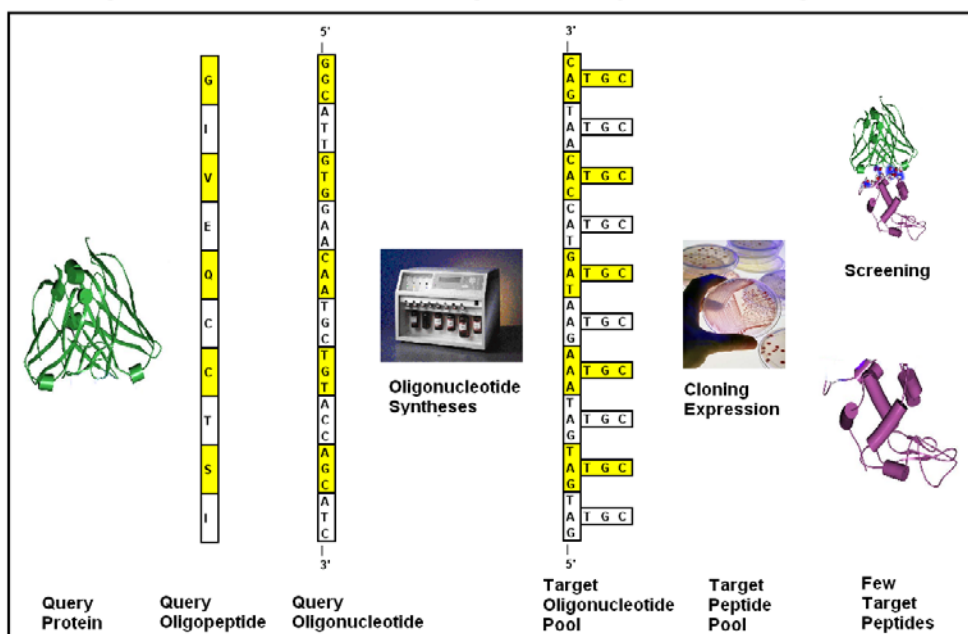
Figure 31. Design and production of specifically interacting proteins (see text for details).

There is no limitation to the size of the query; however, preferably the sequence is from about 5 amino acid residues to about 40 amino acid residues, and better from about 7 to 15 amino acid residues. Preferably, the real and natural coding sequence is known for the query protein. However, there might be some special cases when the sequence is not exactly known, for example in case of designed or artificially modified proteins. Thus, it is possible to fabricate a virtual coding sequence with back translation, using Codon Usage Frequency Tables. The present method relies on the entire information carried by the naturally occurring DNA/mRNA and not only that used for coding of the protein primary sequence.

The query sequence should be a "promising" domain of the query protein and specific domains are more important, including domains that: (a) are known to be antigenic; (b) are located on the surface of the query protein; (c) are not simple (repetitive) sequences; d) contain less frequent amino acids; and e) contain charged amino acid residues.

Once the promising area of the known amino acid sequence is chosen and the nucleotide sequence is determined, then construction of nucleic acid sequences encoding for the target

proteins is initiated. As used here, the term "nucleotide sequence" means a sequence of nucleotides connected by phosphodiester linkages.

A Novel Method to Obtain Oligo-peptides with Specific High Affinity to Query Proteins

1. Query Selection (Q1)

2. Target (T) RNA Pool Prediction
   (using partial codon-complementarity rule)

3. Syntheses of RNAs in the Target RNA Pool
   (modified method of RNA syntheses)

4. Cloning of the Target (T) RNA Pool

5. Target (T) Protein Pool Expression Library

6. Screening for Q-T interactions

7. Clone Selection
   (Few, unique, mono-clones)   >>> Large scale protein production

8. Sequencing the best monoclonal RNAs and Proteins

9. Refine Protein - Protein Interaction Code >>> Gain knowledge

10. Iterate 1-9 on another query (Q2)

Figure 32. A novel method to obtain oligo-peptides with specific high affinity to query proteins. The steps required to practice the method for obtaining oligo-peptides with specific high affinity to query proteins.

Nucleotide sequences are presented here in the direction from the 5′ to the 3′ direction and can be a deoxyribonucleic acid (DNA) molecule or ribonucleic acid (RNA) molecule. Relevant nucleotide bases are indicated here by a single letter code: adenine (A), guanine (G), thymine (T), cytosine (C), inosine (I) and uracil (U). The target nucleotide (RNA or DNA) prediction should follow a simple rule, namely that the 1st and 3rd codon letters of the target nucleotide sequences should have reverse complementary to the 1st and 3rd codon nucleotide residues of the query nucleotide sequence, but the middle, 2nd residue should be any of the four possible nucleotides. The expected number of predicted target RNAs will be 4^n, where n is the number of amino acids (=number of codons, =number of 2nd codon letters).

Synthesis of the nucleotide sequences can be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR or by excising selected DNA fragments from recombinant plasmids containing appropriate inserts and suitable restriction enzyme sites. However, synthesis of predicted (max. 4^n) sequences on a one by one basis does not seem practical. Thus, a simple mass-production is needed which will result in a mixture, containing all possible sequences in the predicted RNA/DNA pool. Fortunately, the regular nature of the nucleotides in the pool makes it possible to synthesize the entire pool of nucleotide sequences as if were only one single nucleotide sequence. For example, the usual step-by-step (base by base) protocol can be followed except at the positions for the synthesis of the 2nd codon

residue. At those points in the synthesis process, an equal mixture of the four nucleotides should be provided instead of a single nucleotide. The result of this modified oligo-nucleotide synthesis should be a mixture of the desired potential target RNAs.

The step to clone the predicted and synthesized RNAs in the pool is the regular cloning procedure which involves insertion of RNA into vector (plasmid or other carrier) and multiplying the sequences in bacteria or yeast as described in the literature. Expression vectors of the system may comprise polynucleotides operatively linked to an enhancer-promoter, such as a prokaryotic or eukaryotic promoter. Further, an enhancer may be included in the vector. A major function of an enhancer is to increase the level of transcription of a coding sequence in a cell that contains one or more transcription factors that bind to that enhancer. Unlike a promoter, an enhancer can function when located at variable distances from transcription start sites so long as a promoter is present.

Expression vectors of the present system comprise polynucleotides that encode the target peptides of the pool. Where expression of recombinant polypeptide of the present system is desired and a eukaryotic host is contemplated, it is most desirable to employ a vector, such as a plasmid, that incorporates a eukaryotic origin of replication. In addition, for the purposes of expression in eukaryotic systems, it is desired to position the peptide encoding sequence adjacent to and under the control of an effective eukaryotic promoter such as those used in combination with Chinese hamster ovary cells. To bring a coding sequence under the control of a promoter, whether it is eukaryotic or prokaryotic, what is generally needed is to position the 5′ end of the translation initiation side of the proper translational reading frame of the polypeptide between about 1 and 50 nucleotides 3′ of or downstream with respect to the promoter chosen. Furthermore, where eukaryotic expression is anticipated, one would typically desire to incorporate an appropriate polyadenylation site into the transcriptional unit which includes the different target peptides.

The pRc/CMV vector (available from Invitrogen) is an exemplary vector for expressing a peptide in mammalian cells, particularly COS and CHO cells. Target polypeptides of the present invention under the control of a CMV promoter can be efficiently expressed in mammalian cells. The pCMV plasmids are a series of mammalian expression vectors of particular utility in the present system. The vectors are designed for use in essentially all cultured cells and work extremely well in SV40-transformed simian COS cell lines. The pCMV1, 2, 3, and 5 vectors differ from each other in certain unique restriction sites in the polylinker region of each plasmid. The pCMV4 vector differs from these 4 plasmids in containing a translation enhancer in the sequence prior to the polylinker. While they are not directly derived from the pCMV1-5 series of vectors, the functionally similar pCMV6b and c vectors, available from the Chiron Corp. (Emeryville, CA), are identical except for the orientation of the polylinker region which is reversed in one relative to the other. The pCMV vectors have been successfully expressed in simian COS cells, mouse L cells, CHO cells, and HeLa cells.

Means of transforming or transfecting cells with exogenous polynucleotide such as the nucleotide molecules of the present system are well known and include techniques such as calcium-phosphate- or DEAE-dextran-mediated transfection, protoplast fusion, electroporation, liposome-mediated transfection, direct microinjection and adenovirus infection.

The most widely used method is transfection mediated by either calcium phosphate or DEAE-dextran. Although the mechanism remains obscure, it is believed that the transfected

DNA enters the cytoplasm of the cell by endocytosis and is transported to the nucleus. Depending on the cell type, up to 90% of a population of cultured cells can be transfected at any one time. Because of its high efficiency, transfection mediated by calcium phosphate or DEAE-dextran is the method of choice for experiments that require transient expression of the foreign DNA in large numbers of cells. Calcium phosphate-mediated transfection is also used to establish cell lines that integrate copies of the foreign DNA, which are usually arranged in head-to-tail tandem arrays into the host cell genome.

The application of brief, high-voltage electric pulses to a variety of mammalian and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation can be extremely efficient and can be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

Liposome transfection involves encapsulation of DNA or RNA within liposomes, followed by fusion of the liposomes with the cell membrane. The mechanism of how DNA or RNA is delivered into the cell is unclear but transfection efficiencies can be as high as 90%.

Direct microinjection of a DNA molecule into nuclei has the advantage of not exposing DNA to cellular compartments such as low pH endosomes. Microinjection is therefore used primarily as a method to establish lines of cells that carry integrated copies of the DNA of interest. A transfected cell can be prokaryotic or eukaryotic.

In addition to prokaryotes, eukaryotic microbes, such as yeast can also be used. *Saccharomyces cerevisiae* or baker's yeast is the most commonly used among eukaryotic microorganisms, although a number of other strains are also available. For expression in *Saccharomyces*, the plasmid YRp7, for example, is commonly used. This plasmid already contains the trpl gene which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1. The presence of the trpl lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Suitable promoter sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also introduced into the expression vector downstream from the sequences to be expressed to provide polyadenylation of the mRNA and termination. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter region for alcohol dehydrogenase 2, isocytochrome *c*, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector containing a yeast-compatible promoter, origin or replication and termination sequences is suitable.

In addition to microorganisms, cultures of cells derived from multicellular organisms can also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate

or invertebrate culture. However, interest has been greatest in vertebrate cells, and propagation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years. Examples of such useful host cell lines are AtT-20, VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and W138, BHK, COSM6, COS-1, COS-7, 293 and MDCK cell lines. Expression vectors for such cells ordinarily include (if necessary) an origin of replication, a promoter located upstream of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences.

For use in mammalian cells, the control functions on the expression vectors are often derived from viral material. For example, commonly used promoters are derived from polyoma, Adenovirus 2, Cytomegalovirus and most frequently Simian Virus 40 (SV40). The early and late promoters of SV40 virus are particularly useful because both are obtained easily from the virus as a fragment that also contains the SV40 viral origin of replication. Smaller or larger SV40 fragments can also be used, provided they include the approximately 250 bp sequence extending from the HindIII site towards the BglI site located in the viral origin of replication. It is also possible, and often desirable, to utilize promoter or control sequences normally associated with the desired gene sequence, provided such control sequences are compatible with the host cell systems.

Following transfection, the cell is maintained under culture conditions for a period of time sufficient for expression of the target proteins of the pool. Culture conditions are well known and include ionic composition and concentration, temperature, pH, etc. Typically, transfected cells are maintained under culture conditions in a culture medium. Suitable medium for various cell types are well known. Temperature is preferably from about 20°C to about 50°C. pH is preferably from about a value of 6.0 to a value of about 8.0, better at about 6.8–7.8 and, better again at about 7.4. Other biological conditions needed for transfection and expression of an encoded protein are well known.

Transfected cells are maintained for a period of time sufficient for expression of the target proteins. A suitable time depends inter alia upon the cell type used and is readily determinable by a skilled artisan. Typically, maintenance time is from about 2 to 14 days. Recovery of the target proteins comprises isolating and purifying the recombinant polypeptides. Isolation and purification techniques for polypeptides are well known and include such procedures as precipitation, filtration, chromatography, electrophoresis, etc.

The target proteins are preferably arranged in a library assay system for screening with samples of the query protein. Any method that detects specific, high affinity protein–protein interactions is theoretically useful to perform the screening.

Selecting the best clones with the most specific and highest affinity interacting proteins can be followed by repeated screenings, thus leading to the most desired target proteins having the highest binding affinity for the query protein. The target proteins with the highest affinity are suitable for large-scale target protein production.

These aspects and embodiments of the present system are further described in the following examples. However, the present system is not limited by the following example, and variations will be apparent to those skilled in the art without departing from the scope of the present setup.

EXAMPLE 1

Figure 33 shows the use of the present system to obtain a specific high affinity protein with binding affinity for a section of the A-peptide in human insulin.

Starting with the known protein and nucleic acid sequence of the entire Pre-pro-insulin, 1−10 residues of the A peptide and the corresponding nucleic acid sequence are selected. The selected part of the peptide, called the query, is used for screening of the target protein expression library. Therefore, this sequence should be available in pure peptide form.

Figure 33. Representative query amino acid sequences and preparation of reverse and complement sequences wherein the second nucleotide of each codon is replaced with a variable X nucleotide.

Next, a sequence is created which is complementary to the query nucleotide sequence at the 1st and 3rd codon positions but leaving the 2nd position undefined (X). The

complementary sequence is reversed and in this particular example, the bases T are changed to U. The second (central) codon position remains undefined and this undefined X position can be any one of the possible nucleotides (A, U, G, C). Therefore, this prediction method defines many different target RNA sequences. In the case of a sequence including 30 nucleotide bases, the expected number of possible target sequences will be about $4^{10}=10^6$.

The predicted pool of target RNAs is synthesized by following the usual step-by-step (base by base) protocol, known to those skilled in the art, except the syntheses of the X positions. At the X position, a mixture of nucleotide bases is provided (which contain equal amounts of A and U and G and C). The result of this modified oligo-nucleotide synthesis is a mixture of the desired potential target RNAs as shown in Figure 34. The target RNAs are cloned and transfected, via an expression vector, into a cell for expression therein of the encoded protein. An expression library of the expressed target protein is created for screening for query protein/target protein affinity binding. When binding complexes are found to meet the affinity binding levels, the target protein may be cloned for large-scale production.

Syntheses of RNAs in Target RNA Pool
(modified method of RNA syntheses)

1. Step	2. Step	5. Step	6. Step	7. Step
5′	GA	GAUGA	GAUGAU	GAUGAUG
G	GU	GUUGA	GUUGAU	GUUGAUG
	GG	GGUGA	GGUGAU	GGUGAUG
	GC	GCUGA	GCUGAU	GCUGAUG
		GAUGU	GAUGUU	GAUGUUG
	3. Step	GUUGU	GUUGUU	GUUGUUG
	GAU	GGUGU	GGUGUU	GGUGUUG
	GUU	GCUGU	GCUGUU	GCUGUUG
	GGU	GAUGG	GAUGGU	GAUGGUG
	GCU	GUUGG	GUUGGU	GUUGGUG
		GGUGG	GGUGGU	GGUGGUG
	4. Step	GCUGG	GCUGGU	GCUGGUG
	GAUG	GAUGC	GAUGCU	GAUGCUG
	GUUG	GUUGC	GUUGCU	GUUGCUG
	GGUG	GGUGC	GGUGCU	GGUGCUG
	GCUG	GCUGC	GCUGCU	GCUGCUG

Colored background: addition of nucleotide mixture

Figure 34. Synthesis pattern for construction of target sequences and the progression of the permutations depending on the number of amino acid residues.

These steps may be repeated numerous times by modifying the length of the query sequence and/or using another domain area of the query protein that may be of interest.

EXAMPLE 2. EXAMPLE FOR DESIGNING AND CHARACTERIZATION OF A SPECIFIC PROTEIN−PROTEIN INTERACTION

The BacterioMatch™ two-hybrid system (Stratagene, 11011 N. Torrey Pines Road, La Jolla, CA 92037) was used to quickly detect protein−protein interactions designed by the recent method. It is a simple alternative or complement to yeast two-hybrid systems for in

vivo detection of protein–protein interactions. Because the two-hybrid assay is performed in bacteria, the results are obtained more easily and quickly than in yeast. The system is based on transcriptional activation of a primary ampicillin-resistant reporter and a secondary β-galactosidase reporter for validation. The BacterioMatch two-hybrid system is based on a methodology developed by Dove, Joung, and Hochschild of Harvard Medical School.
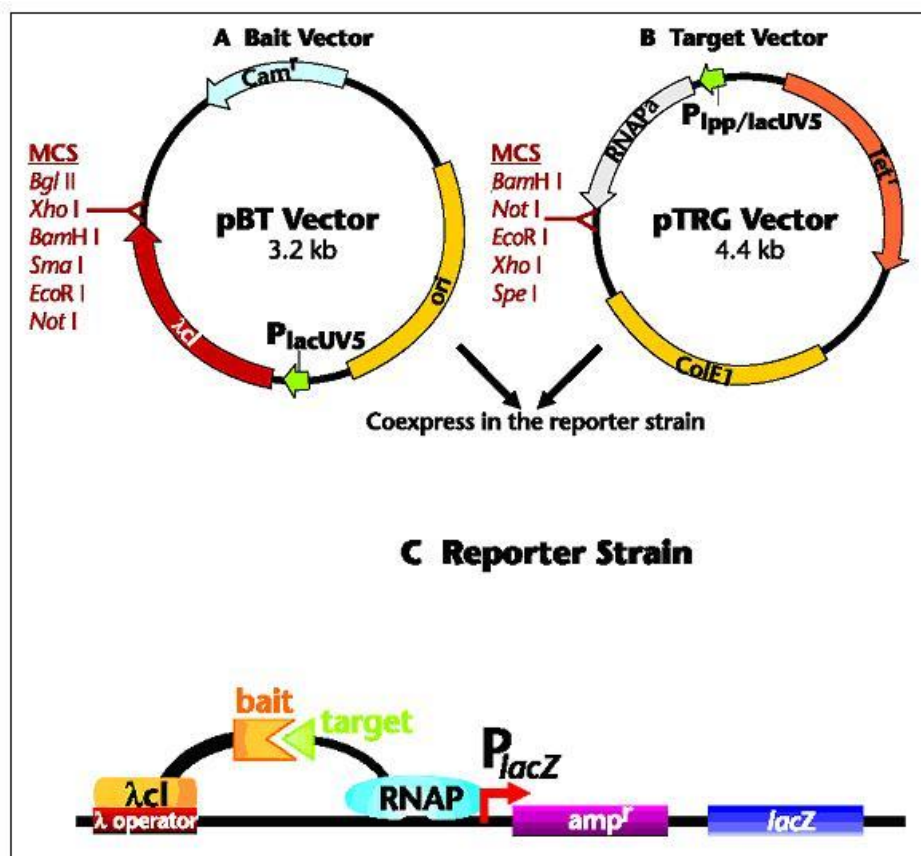
Figure 35. The BacterioMatch™. Two-hybrid system (reproduced from http://www.strategene.com).

The BacterioMatch two-hybrid system is based on transcriptional activation (Figure 35). A protein of interest – the bait – is fused to the full-length bacteriophage repressor protein (λcI). The corresponding target protein is fused to the amino-terminal domain of the α-subunit of RNA polymerase (RNAPα). The bait is tethered to the x operator sequence upstream of the reporter promoter through the DNA-binding domain of λcI. If the bait and target interact, they recruit and stabilize the binding of RNA polymerase close to the promoter and activate the transcription of the ampicillin-resistant reporter gene in the BacterioMatch two-hybrid reporter strain. The β-galactosidase reporter gene provides an additional mechanism to validate putative protein–protein interactions.

(a) *Bait vector.* The bait vector, pBT encodes the full-length bacterial phage cI protein under the control of the strong lacUV5 promoter. A protein of interest is fused to the

bacterial phage λcI protein by inserting its gene into the multiple cloning site at the 3′ end of the λcI gene. The presence of a multiple cloning site present makes it convenient to subclone a bait gene that is already present in many yeast two-hybrid bait plasmids.

(b) *Target vector*. The target plasmid, pTRG, is compatible with Stratagene's cDNA library construction kit. The target plasmid directs transcription of the amino-terminal domain of RNA polymerase α-subunit and linker region under the control of tandem promoters, lpp and lacUV5. The target gene is fused in-frame to the α-subunit NTD through a multiple cloning site at the 3′ end of the α-subunit gene.

(c) *Reporter strain*. The reporter strain is derived from XL1-Blue MRF′. The strain lacks all restriction systems in order to be compatible with current cDNA library construction methods. The lac I^q gene located on the F′ episome represses synthesis of the bait and target until induction. The reporter cassette is also located on the F′ episome in the cell. The lacZ gene serves as a secondary reporter to provide a visible phenotype for identifying positive protein–protein interactions.

DEFINITIONS

Query (or bait) is one protein sequence with which the target protein, designed and produced with the method, will specifically interact. *Target protein* is one or more protein sequence(s) designed by the method to specifically interact with the query protein sequence. The target is expected to be present in a pool of protein sequences, called the *target pool*. The target pool is designed using a *target template*, which is a nucleic acid sequence containing 2/3 defined and 1/3 undefined (any) nucleotides (X) (a target template, which contains 15 undefined nucleic acid residues, will result in $4^{15}=10^9$ different oligonucleotides which will be translated into the corresponding number of proteins). The target pool is synthesized using a *target oligo template* (TOT) which has a *constant* (C) and *variable* (V) part. The TOT-C is necessary to synthesize dsDNA of the target pool sequences and it is ~20 nucleotides long. The TOT-V (target template) is about 30–45 nucleotide long, 2/3rd of nucleotides are unambiguously defined, while 1/3rd are not (X). The X residues should be synthesized by adding a mixture of nucleotides (equal amounts of A+T+G+C) to the reaction during oligo synthesis. The results (number of highly, moderately, slightly positive clones) are evaluated by visual inspection. The positive clones are saved for further experiments. If there are no positive clones, it is necessary to validate the orientation and translation frame in the target mRNAs. This is possible by sequencing some target mRNAs. The sequence should show the residue pattern. Both TARGET TEMPLATE to ESRLERLEQLFLLIF (GAL4 09–23AA) and TARGET TEMPLATE to QLFLLIFPREDLDMI (GAL4 17–31AA) contained numerous positive bacterial clones growing on double selective medium. Sequencing of DNA from the vectors in randomly selected positive clones confirmed that:

- they contained the characteristic TOT pattern, i.e. defined 1st and 3rd codon residues;
- the nucleic acids differed only in the 2nd codon positions, while they were the same regarding the 1st and 3rd codon positions;

- The restriction endonuclease recognition sequences were present;
- the start and stop codons were present;
- the sequences were inserted into the correct, sense DNA strands;
- the codon frames were correct in relation to the start codon and were read in the correct frames.

Some positive TARGET TEMPLATE to ESRLERLEQLFLLIF (GAL4 09–23AA) clones were further processed to monoclonal colonies and proteins were extracted. Characterization of the binding properties of fluorescent labelled GAL4 peptide to the protein extract indicated the presence of saturable binding sites in the protein extracts from positive clones and the absence of saturable binding sites in the negative clones.

THE EXPERIMENT

The experiment below is specifically designed for the BacterioMatch (Stratagene) two-hybrid system. This system uses:

- a bait vector (pBT) and the manufacturer's standard are used as insert, the dimerization domain of 1HBW REGULATORY PROTEIN GAL4;
- a target vector (pTRG) and the manufacturer's standard are used as insert, and ~90 aa long mutant form of Gal11.

Figure 36. Sequences, designed by this method, were expected to produce proteins (when transcribed and translated) with the potential to specifically interact with the indicated domains of the Gal4 protein. The 1st and 3rd codon letters in these target templates are complementary to the 3rd and 1st codon letters in the Gal4 coding sequences (reverse reading direction) while the 2nd codon letter is undefined (A or T or G or C).

In the experiment below the target oligo pool is used instead of Gall 1 in the pTRG vector.

The query in this experiment is the dimerization domain of 1HBW REGULATORY PROTEIN GAL4 inserted into pBT (as provided and described by Stratagene).

The target oligo templates (TOT-V) were designed to specifically interact with K01486_SCGAL4_DIMDOM-171/9-23 and K01486_SCGAL4_DIMDOM-171/17/31 sequences.

The sequences below are sense, ssDNA sequences which means that the TOT-V in this sequence is the same as the sequence in the expected mRNAs (except T/U conversion). The TOT-C is not indicated here, BPD can decide which TOT-C to use for this purpose (Figure 36).
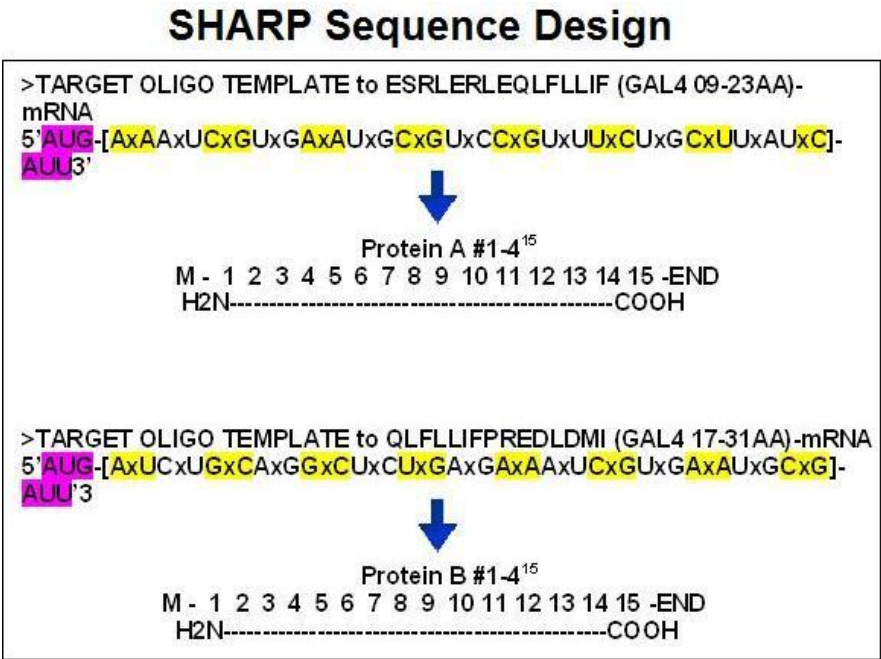
Figure 37. The transcription of TOT dsDNA will result in TOT mRNA. A 45 nucleic acid long TOT will be translated into 4^{15} different oligopeptides, each 15 amino acids long. Some of these oligopeptides are expected to specifically interact with the respective GAL4 targets.

The experiment consists of the following steps:

1) Sequence the Gal 4 DNA (provided by Stratagene) to make sure that the query sequence is as expected.
2) Synthesize the target pool using the target oligo templates. This is a single run routine oligo synthesis. Residue X is equal amounts of A+T+G+C.
3) Make dsDNAs. This is a single run PCR.
4) Make restriction enzyme cuts on the target oligo pool sequences. This is a single run RE reaction.
5) Insert the oligo pool sequences into the pTRG vector. $\sim10^9$ different vectors are expected. Make sure that the orientation of the target oligo-s is correct and the transcription will result in the following mRNA. The target oligo pool insertion is a single run ligase reaction (Figure 37). Transcription of TOT dsDNA will result in TOT mRNA. A 45 nucleic acid long TOT will be translated into 4^{15} different

oligopeptides, each 15 amino acids long. Some of these oligopeptides are expected to specifically interact with the respective GAL4 targets.

6) Insert the vectors into bacteria.
7) Perform the BacterioMatch two-hybrid assay accordingly to the Stratagene manual.

The K_d of the binding sites varied between 1 and 100 nM indicating the presence of a limited number of high affinity binding sites. Unlabelled GAL4 inhibited the binding of labeled GAL4 to the proteins from positive clones while other randomly chosen proteins (insulin, growth hormone, prolactin) were ineffective competitors even in much higher concentrations.

This experiment indicates that it is possible to design specifically interacting oligo-peptides (target) to any oligo-peptide (query) and detect the interaction in a bacterial two-hybrid system (like BacterioMatch™. The method is quick; it takes only a few days to obtain interacting monoclonal proteins. The designed protein–protein interaction is highly specific and has high affinity ($K_d \sim 1$–100 nM).

Further details can be found in the following references [58, 59, 96].

CONCLUSION

THE PROMISES OF THE SECOND GENERATION PROTEOMIC CODE

Industrial Applications

The second generation Proteomic Code and the method to develop SHARP have some potential advantages that are not obvious in the recent antibody developing methods:

−   provide quick access to interacting peptides;
−   provide direct and permanent access to monoclonal sources for large-scale production;
−   SHARP is small (MW <2000 Da) compared to antibodies (155 kDa) or affibodies (which gives therapeutic and manufacturing advantage), no need for humanization;
−   might be the key to the mass production of interactive oligopeptides (similar to the on-demand synthesis of nucleotide oligo-s;
−   a self-learning method; every single successful SHARP can contribute to a more and more exact amino acid interaction table.

Scientific Potential

The present system is a unique in silico method of identifying the most effective binding proteins to interact with reactive epitopes on a respective protein antigen. Epitopes of a protein antigen represent the sites that are recognized as binding sites by certain immunoglobulin molecules, as antibodies.

The benefits of the system are widespread and beneficial to biotechnology, and are useful, for example, in developing drugs for treatment of viral diseases such as AIDS and influenza, as well as diseases such as Alzheimer's disease and bovine spongiform

encephalopathy disease. In addition to medical research and drug development, this system has applications related to environmental health and public safety, including for example the detection of bacteria, viruses, toxins, etc. in air, water, and food supplies.

By way of further specific examples, the present system has applications in the following areas:

1) improving health care, by providing a new and easily implemented approach to the development of diagnostic kits and therapeutic drugs;
2) improving the environment, by providing new and economic approaches for detecting environmental pathogens;
3) improving working conditions by providing economic and effective ways to detect environmental pathogens; and
4) improving homeland security, by providing rapid detection of known as well as new pathogens in air, water, food, etc.

THE VISION OF A PROTEOME-SENSOR CHIP

Detection and measurement of proteins is a fundamental procedure in life sciences. Many diagnostic procedures are already based on this technique and many more will follow:

− detection of hormones and enzymes for diagnosis of organ failure;
− detection of pathogen-derived antigens or antibody responses for diagnosis of infections;
− detection of allergens for allergy diagnoses;
− tumor markers to detect and evaluate neoplasias.

Considering that the number of hormones, enzymes, pathogens, and markers is very large, it is easy to recognize that the demand for specifically interacting diagnostic proteins is large. However, there are limitations to satisfy the demand. It takes several weeks to develop one antibody with the traditional methods. Therefore it is not cheap. The traditional one protein–one kit method is simply no longer feasible. Some kind of integration is necessary. Development points to protein chips which permit the simultaneous, parallel detection of hundreds or thousands of protein signals and the computerized integrated evaluation of the results. Chip-based protein detection requires a large number of easy to produce, cheap interacting proteins.

The Proteomic Code-based method described here could significantly contribute to the industrial production of these interacting peptides.

The SHARP chip-based technology opens the way to the real possibility of monitoring the proteome, i.e. obtain detailed information on the qualitative and quantitative state of a large number of different proteins simultaneously, including even some information about the splicing and configuration changes of individual proteins (Figure 38).

The human genome contains 3×10^9 base pairs. About 2% of this is located in about 30,000 genes which are expressed, spliced into ~100,000 different proteins containing ~6 million amino acids. The $2 \times$ coverage of this size proteome with 15 amino acid long

complementary sequences requires 800,000 different oligopeptides. (This is probably the upper estimate, because the proteome consists of many similar or identical domains.) Proteome monitoring with proteomic chips is a huge challenge which is only possible if SHARP proteins are accessible in much larger number and for a much lower price than is possible today.
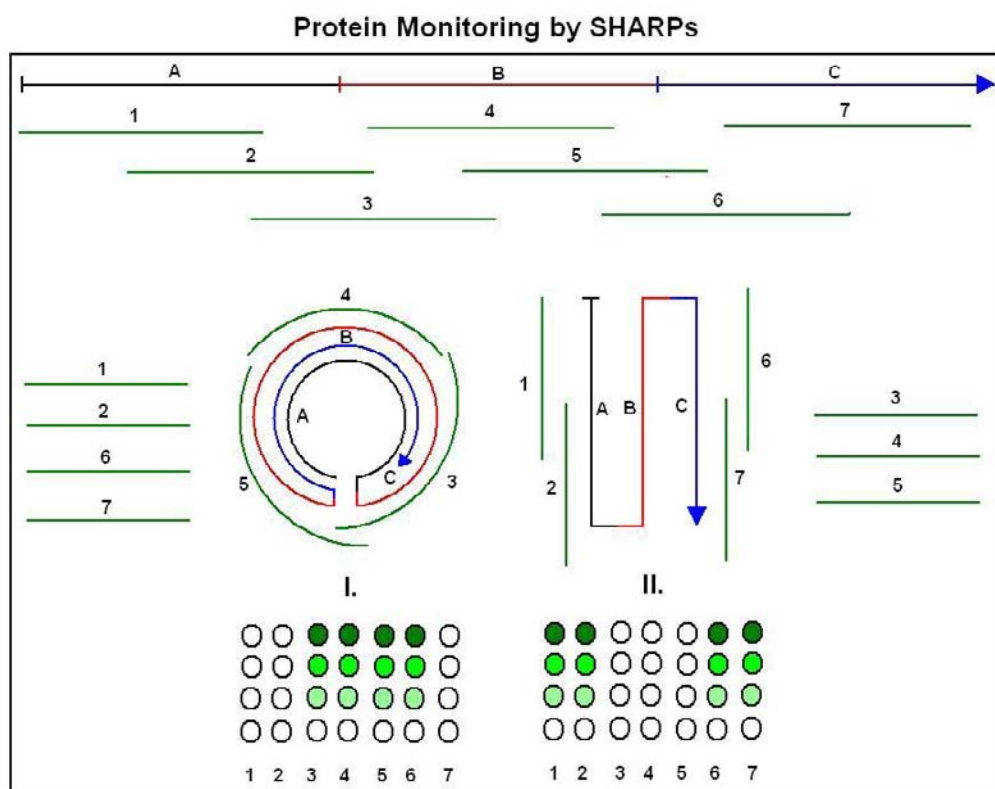
Figure 38. Protein monitoring by SHARPs. Specific and high affinity reacting proteins (SHARPs, 1–7) are designed and produced against a protein. The protein has 3 domains (A, B, C) and two main conformations (I, II). Domain B is exposed on the surface by conformation I and that form is detected by SHARPs 3–6. Domain B is sandwiched by domains A and B in configuration II and therefore this form is detected by SHARPs 1–2 and 6–7. A concentration and configuration dependent response is detected by a slot corresponding to the protein by the SHARPs CHIP (dots).

THE VISION OF A NEW PHYSIOLOGY

The receptor – ligand and antigen – antibody type or interaction has a fundamental role in the physiology of humans and animals. Peptides which are able to specifically interact with regulatory pathway have of course significant potential in manipulating these pathways. Most drugs are effective because of their interaction with regulatory pathways (e.g., GPCR or 7TM receptor system). Therefore, easy and inexpensive access to designed interacting proteins will have an impact on further development of physiological and pharmacological research and drug discovery. Some kind of industrial-scale, standardized, semi-automated physiological

research is also desirable. The traditional one-by-one approach is too slow and to expensive for the complexity of life.

THE VISION OF NEW PROTEIN-BASED THERAPEUTIC APPROACHES

Proteins/peptides are underutilized in medical therapy. Only insulin is used to treat a common disease (Type I diabetes); a few more proteins are used to treat relatively rare disease, such as growth hormone (GH) deficiency and hemophilia. Veterinary use of GH (lean meat production) or misuse in sports and cosmetology far exceeds its medical indication. Protein therapy is expensive and requires daily injections which is not attractive to most patients. Most pharmaceutical companies have large accumulated knowledge (and patent base) regarding traditional, simple molecular drug design and treatments while their knowledge in proteomics is still undergoing development. Simple molecule-based drugs are "simple"; not much specificity is provided by these structures. Protein-based, highly specific treatments are not too far off in the future and easy access to biologically active proteinaceous substances will facilitate physiological evaluation and medical application of these more complex peptide molecules.

SHARPs, complementary or not, are obvious candidates for receptor agonist or antagonist functions. However, there is an even more exciting and less expected therapeutic application of complementary peptides. Some experiments suggest that immunization with complementary peptides to receptors induces production of ligand-like antibodies, with ligand-like biological effects. Just imagine the possibility that inducing insulin-like antibodies by immunization with designed complementary peptides to insulin-receptors and obtaining insulin-like effects (regulation will of course be a problem to solve). Or treat GH deficiency by raising GH-like antibodies using GH-receptor complementary peptides for immunization. This seems like science fiction today, but for example Blalocks' experiments are already pointing in that direction [27].

Is protein therapy an alternative to gene therapy? Yes. Gene therapy is technically still difficult to perform and the effect is irreversible. The effects of protein therapy are short-lived and reversible (if immunization does not occur).

SOFTWARE

SeqForm, http://janbiro.com/Downloads.html
SeqPlot http://janbiro.com/Downloads.html
Dotlet http://janbiro.com/Downloads.html
mfold http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi
SeqX: http://janbiro.com/Downloads.html.

PATENTS

ROOT-BERNSTEIN

Blalock JE (Birmingham, AL), Bost KL (Birmingham, AL), Smith EM (Galveston, TX): Polypeptides complementary to peptides or proteins having an amino acid sequence or nucleotide coding sequence at least partially known and methods of design therefor. United States Patent 5,212,072, May 18, 1993.

Omichinski JG (Potomac, MD), Fassina G (Padova, IT), Olson AD (Silver Spring, MD), Thorgeirsson SS (Bethesda, MD): Computer-assisted design of anti-peptides based on the amino acid sequence of a target peptide. United States Patent 5,081,584, January 14, 1992.

Biro J: System and method to obtain oligo-peptides with specific high affinity to query proteins. US Patent Application Publication US 2007/0015189 A1, January 18, 2007.

COMPETING INTERESTS

The author of this article has pending US and PCT patent applications for the system and the method of obtaining oligo-peptides with specific high affinity to query proteins.

ACKNOWLEDGMENTS

REFERENCES

[1]   Pawson T, Gish GD: Protein-protein interactions: a common theme in cell biology. In *Protein-Protein Interactions*, Edited by Golemis EA, Adams PD. New York: CSHL Press; 2005.

[2]   Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R: The structure of DNA and RNA. Chapter 6. In *Molecular Biology of the Gene*, 5th edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2004:1-33.

[3]   Pauling L: *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, 2nd edition. Ithaca, NY: Cornell University Press; London: Humphrey Milford, Oxford University Press: 1940: 450 pp.

[4]   Pauling L, Corey RB, Yakel Jr HL, Marsh RE: Calculated form factors for the 18-residue 5-turn _-helix. *Acta Crystallogr* 1955, 8:853-855.

[5]   Pauling L, Corey RB: Specific hydrogen-bond formation between pyrimidines and purines in deoxyribonucleic acids. (In Linderstrom-Lang Festschrift). *Arch. Biochem. Biophys.* 1956, 56:164-181.

[6]   Biro JC: Speculations about alternative DNA structures. *Med. Hypotheses* 2003, 61:86-97.

[7]   Woese CR: Translation: in retrospect and prospect. *RNA* 2001, 7:1055-1067.

[8]   Dayhoff MO: *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, 1966.

[9]   Biro J: Comparative analysis of specificity in protein-protein interactions. Part I: A theoretical and mathematical approach to specificity in protein-protein interactions. *Med. Hypotheses* 1981, 7:969-979.

[10] Biro J: Comparative analysis of specificity in protein-protein interactions. Part II: The complementary coding of some proteins as the possible source of specificity in protein-protein interactions. . *Med. Hypotheses* 1981, 7:981-993.

[11] Biro J: Comparative analysis of specificity in protein-protein interactions. Part III: Models of the gene expression based on the sequential complementary coding of some pituitary proteins. *Med. Hypotheses* 1981, 7:995-1007.

[12] Gamow G: Possible mathematical relation between deoxyribonucleic acid and proteins. *Kgl Danske Videnskab. Selskab. Biol. Medd.* 1954, 22:1-13.

[13] Gamow G: Possible relation between deoxyribonucleic acid and protein structures. *Nature* 1954, 173:318.

[14] Gamow G: Information transfer in the living cell. *Sci. Am.* 1955, 193:70-78.

[15] Gamow G, Ycas M: Statistical correlation of protein and ribonucleic acid composition. *Proc. Natl. Acad. Sci. USA* 1955, 41:1011-1019.

[16] Gamow G, Rich A, Ycas M: The problem of information transfer from nucleic acids to proteins. In *Advances in Biological and Medical Physics*, *Volume 4*. New York: Academic Press; 1956:23-68.

[17] Gamow G, Ycas M: *Mr. Tompkins Inside Himself: Adventures in the New Biology.* New York: Viking Press; 1967.

[18] Biro JC, Biro JM: Frequent occurrence of recognition site-like sequences in the restriction endonucleases. *BMC Bioinformatics* 2004, 16:30.

[19] Biro JC, Biro JM: The BlastNP: a novel, sensitive sequence similarity searching method using overlappingly translated sequences. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2004, 4:2777-2780.

[20] Mekler LB: Specific selective interaction between amino acid groups of polypeptide chains. *Biofizika* 1969, 14:581-584.

[21] Mekler LB, Idlis RG: *VINITI Deposited Doc.* 1981, 1476-1481.

[22] Blalock JE, Smith EM: Hydropathic anti-complementarity of amino acids based on the genetic code. *Biochem. Biophys. Res. Commun.* 1984, 121:203-207.

[23] Blalock JE, Bost KL: Binding of peptides that are specified by complementary RNAs. *Biochem. J.* 1986, 234:679-683.

[24] Bost KL, Smith EM, Blalock JE: Similarity between the corticotrophin (ACTH) receptor and a peptide encoded by an RNA that is complementary to ACTH mRNA. *Proc. Natl. Acad. Sci. USA* 1985, 82:1372-1375.

[25] Blalock JE, Smith EM: A complete regulatory loop between the immune and neuro-endocrine systems. *Fed. Proc.* 1985, 44(1 Pt 1):108-111.

[26] Blalock JE, Harbour-McMenamin D, Smith EM: Peptide hormones shared by the neuroendocrine and immunologic systems. *J. Immunol.* 1985, 135(Suppl 2):858s-861s.

[27] Bost KL, Blalock JE: Production of anti-idiotypic antibodies by immunization with a pair of complementary peptides. *J. Mol. Recognit.* 1989, 1:179-183.

[28] Blalock JE: Department of Physiology and Biophysics, The University of Alabama at Birmingham. http://www.physiology.uab.edu/Blalock.htm#Research

[29] Root-Bernstein RS: Amino acid pairing. *J. Theor. Biol.* 1982, 94:885-894.

[30] Root-Bernstein RS: Protein replication by amino acid pairing. *J. Theor. Biol.* 1983, 100:99-106.

[31] Siemion IZ, Stefanowicz P: Periodical changes of amino acid reactivity within the genetic code. *Biosystems* 1992, 27:77-84.

[32] Siemion IZ, Cebrat M, Kluczyk A: The problem of amino acid complementarity and antisense peptides. *Curr. Protein Peptide Sci.* 2004, 5:507-527.

[33] Siemion IZ, Zbozień-Pacamaj R, Stefanowicz P: New hypothesis on amino acid complementarity and its evaluation on TGF-β_2-related peptides. *J. Mol. Recognit.* 2001, 14:1-12.

[34] Heal JR, Roberts GW, Christie G, Miller AD: Inhibition of beta-amyloid aggregation and neurotoxicity by complementary (antisense) peptides. *Chembiochemistry* 2002, 3:86-92.

[35] Heal JR, Bino S, Roberts GW, Raynes JG, Miller AD: Mechanistic investigation into complementary (antisense) peptide mini-receptor inhibitors of cytokine interleukin-1. *Chembiochemistry* 2002, 3:76-85.

[36] Bhakoo A, Raynes JG, Heal JR, Keller M, Miller AD: De-novo design of complementary (antisense) peptide mini-receptor inhibitor of interleukin 18 (IL-18). *Mol. Immunol.* 2004, 41:1217-1224.

[37] Heal JR, Roberts GW, Raynes JG, Bhakoo A, Miller AD: Specific interactions between sense and complementary peptides: the basis for the proteomic code. *Chembiochemistry* 2002, 3:136-151 (Review. Erratum in: *Chembiochemistry* 2002, 3:271).

[38] Heal JR, Bino S, Ray KP, Christie G, Miller AD, Raynes JG: A search within the IL-1 type I receptor reveals a peptide with hydropathic complementarity to the IL-1beta trigger loop which binds to IL-1 and inhibits in vitro responses. *Mol. Immunol.* 1999, 36:1141-1148.

[39] Zull JE, Smith SK: Is genetic code redundancy related to retention of structural information in both DNA strands? *Trends Biochem. Sci.* 1990, 15:257-261.

[40] Zull JE, Taylor RC, Michaels GS, Rushforth NB: Nucleic acid sequences coding for internal antisense peptides: are there implications for protein folding and evolution? *Nucleic Acids Res.* 1994, 22:3373-3380.

[41] Rother KI, Clay OK, Bourquin JP, Silke J, Schaffner W: Long non-stop reading frames on the antisense strand of heat shock protein 70 genes and prion protein (PrP) genes are conserved between species. *Biol. Chem.* 1997, 378:1521-1530.

[42] Duax WL, Huether R, Pletnev VZ, Langs D, Addlagatta A, Connare S, Habegger L, Gill J: Rational genomics I: antisense open reading frames and codon bias in short-chain oxido reductase enzymes and the evolution of the genetic code. *Proteins* 2005, 61:900-906.

[43] Forsdyke DR: Sense in antisense? *J. Mol. Evol.* 1995, 41:582-586.

[44] Okada H: The possible role of sense and antisense peptide interactions in the generation and maintenance of the tertiary structure of a protein. *Anticancer Res.* 1998, 18:3927-3930.

[45] McGuigan JE: Antibodies to complementary peptides as probes for receptors. *Immunomethods* 1994, 5:158-166.

[46] Baranyi L: Antisense homology boxes in proteins. http://citeseer.ist.psu.edu/16596.html (year unknown).

[47] Baranyi L, Campbell W, Ohshima K, Fujimoto S, Boros M, Okada H: The antisense homology box: a new motif within proteins that encodes biologically active peptides. *Nat. Med.* 1995, 1:894-901.

[48] Baranyi L, Campbell W, Ohshima K, Fujimoto S, Boros M, Kaszaki J, Okada H: Antisense homology box-derived peptides represent a new class of endothelin receptor inhibitors. *Peptides* 1998, 19:211-223.

[49] Baranyi L, Campbell W, Okada H: Antisense homology boxes in C5a receptor and C5a anaphylatoxin: a new method for identification of potentially active peptides. *J. Immunol.* 1996, 157:4591-4601.

[50] Segerstéen U, Nordgren H, Biro JC: Frequent occurrence of short complementary sequences in nucleic acids. *Biochem. Biophys. Res. Commun.* 1986, 139:94-101.

[51] Weigent DA, Clarke BL, Blalock JE: Peptide design using a genetically patterned binary code: growth hormone-releasing hormone as a model. *Immunomethods* 1994, 5:91-97.

[52] Biro JC, Benyo B, Sansom C, Szlavecz A, Fordos G, Micsik T, Benyo Z: A common periodic table of codons and amino acids. *Biochem. Biophys. Res. Commun.* 2003, 306:408-415.

[53] Benyo B, Biro JC, Benyo Z: Codes in the codons: construction of a codon/amino acid periodic table and a study of the nature of specific nucleic acid-protein interactions. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2004, 4:2860-2863.

[54] Woese CR: Chapters 6-7. In: *The Genetic Code. The Molecular Basis for Genetic Expression.* New York: Harper and Row: 1967:156-160.

[55] Crick FHC: The origin of the genetic code. *J. Mol. Biol.* 1968, 38:367-379.

[56] Wong JT: Evolution of the genetic code. *Microbiology* 1988, 5:174-181.

[57] Biro JC, Fordos G: SeqX: a tool to detect, analyze and visualize residue co-locations in protein and nucleic acid structures. *BMC Bioinformatics* 2005, 6:170.

[58] Biro JC: Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. *Theor Biol Med Model* 2006, 3:15.

[59] Biro JC: Indications that "codon boundaries" are physico-chemically defined and that protein-folding information is contained in the redundant exon bases. *Theor. Biol. Med. Model* 2006, 3:28.

[60] Anfinsen CB, Redfield RR, Choate WI, Page J, Carroll WR: Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J. Biol. Chem.* 1954, 207:201-210.

[61] Biro JC: A novel intra-molecular protein-protein interaction code based on partial complementary coding of co-locating amino acids, *Med. Hypotheses* 2006; 66:137-142.

[62] Neher E: How frequent are correlated changes in families of protein sequences. *Proc. Natl. Acad. Sci. USA* 1994, 91:98-102.

[63] Levinthal C: How to fold graciously in Mossbauer spectroscopy in biological systems. In *Proceedings of a Meeting held at Allerton House, Monticello, IL*. Edited by Debrunner P, Tsibris JCM, Munck E. Urbana, IL: University of Illinois Press; 1969:22-24.

[64] Isogai Y, Ota M, Ishii A, Ishida M, Nishikawa K: Identification of amino acids involved in protein structural uniqueness: implication for de novo protein design. *Protein Eng* 2002, 15:555–560 (doi: 10.1093/protein/15.7.555).

[65] Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003, 31:3406-3415.

[66] Brunak S, Engelbrecht J: Protein structure and the sequential structure of mRNA: alpha-helix and beta-sheet signals at the nucleotide level. *Proteins* 1996, 25:237-252.

[67] Gu W, Zhou T, Ma J, Sun X, Lu Z: Folding type specific secondary structure propensities of synonymous codons. *IEEE Trans. Nanobiosci.* 2003, 2:150-157.

[68] Gupta SK, Majumdar S, Bhattacharya TK, Ghosh TC: Studies on the relationships between the synonymous codon usage and protein secondary structural units. *Biochem. Biophys. Res. Commun.* 2000, 269:692-696.

[69] Gokhale KC, Newnam GP, Sherman MY, Chernoff YO: Modulation of prion-dependent polyglutamine aggregation and toxicity by chaperone proteins in the yeast model. *J. Biol. Chem.* 2005, 280:22809-22818.

[70] Fan H, Mark AE: Mimicking the action of folding chaperones in molecular dynamics simulations: application to the refinement of homology-based protein structures. *Protein Sci.* 2004, 13:992-999.

[71] Walter S, Buchner J: Molecular chaperones − cellular machines for protein folding. *Angew Chem. Int. Ed. Engl.* 2002, 41:1098-1113.

[72] Komar AA, Kommer A, Krasheninnikov IA, Spirin AS: Cotranslational folding of globin. *J. Biol. Chem.* 1997, 272:10646-10651.

[73] Thanaraj TA, Argos P: Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci.* 1996, 5:1973-1983.

[74] Chiusano ML, Alvarez-Valin F, Di Giulio M, D'Onofrio G, Ammirato G, Colonna G, Bernardi G: Second codon positions of genes and the secondary structures of proteins. Relationships and implications for the origin of the genetic code. *Gene* 2000;261:63-69.

[75] Gu W, Zhou T, Ma J, Sun X, Lu Z: The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *Biosystems* 2004, 73:89-97.

[76] Ermolaeva O: Synonymous codon usage in bacteria. *Curr. Issues Mol. Biol.* 2001, 3:91-97.

[77] Biro JC, Biro JM, Biro AM: Hidden massages in hidden sub-sequences: a study on collagens. In *30th FEBS Congress – 9th IUBMB Conference, Budapest, Hungary, 2–7 July 2005*; 2005:abstract.

[78] Katz L, Burge CB: Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.* 2003, 13:2042-2051.

[79] Biro JC: SeqForm. 2005 [http://www.janbiro.com/downloads].

[80] Biro JC: SeqPlot. 2005 [http://www.janbiro.com/downloads].

[81] Junier T, Pagni M: Dotlet: diagonal plots in a web browser. *Bioinformatics* 2000, 16:178-179 [http://www.isrec.isb-sib.ch/java/dotlet/Dotlet.html].

[82] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: The Protein Data Bank. *Nucleic Acids Res.* 2000, 28:235-242.

[83] Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B: The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 1992, 63:751–759.

[84] Adzhubei IA, Adzhubei AA: ISSD Version 2.0: taxonomic range extended. *Nucleic Acids Res.* 1999, 27:268-271.

[85] Seffens W, Digby D: mRNA have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res* 1999, 27:1578-1584.

[86] Oresic M, Dehn M, Korenblum D, Shalloway D: Tracing specific synonymous codon-secondary structure correlations through evolution. *J. Mol. Evol.* 2003, 56:473-484.

[87] D'Onofrio G, Ghosh TC, Bernardi G: The base composition of the genes is correlated with the secondary structures of the encoded proteins. *Gene* 2002, 300:179-187.

[88] Xie T, Ding D: The relationship between synonymous codon usage and protein structure. *FEBS Lett.* 1998, 434:93-96.

[89] Kumarevel TS, Gromiha MM, Ponnuswamy MN: Distribution of amino acid residues and residue-residue contacts in molecular chaperons. *Prep. Biochem. Biotechnol.* 2001, 31:163-183.

[90] Eilers M, Patel AB, Liu W, Smith SO: Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biochem. J.* 2002, 82:2720-2736.

[91] Glaser F, Steinberg DM, Vakser IA, Ben-Tal N: Residue frequencies at protein-protein interfaces. *Proteins Struct Funct Genet* 2001, 43:89-102.

[92] Naor D, Fisher D, Jernigan RL, Wolfson H, Nussinov R: Amino acid pair interchanges at spatially conserved locations. *J. Mol. Biol.* 1996, 256:924-938.

[93] Azarya-Sprinzak E, Naor D, Wolfson HJ, Nussinov R: Interchanges of spatially neighboring residues in structurally conserved environment. *Protein Eng.* 1997, 10:1109-1122.

[94] Singer MS, Vriend G, Bywater RP: Prediction of protein residue contacts with a PDB-derived likelihood matrix. *Protein Eng.* 2002, 15:721-725.

[95] Shao Y, Bystroff C: Predicting inter-residue contacts using templates and pathways. *Proteins Struct. Funct. Genet.* 2003, 53:497-502.

[96] Biro JC: Nucleic acid chaperons: a theory of an RNA-assisted protein folding. *Theor. Biol. Med. Model* 2005, 2:35.

[97] Biro JC: System and method to obtain oligo-peptides with specific high affinity to query proteins. US Patent Application Publication, US 2007/0015189 A1, Jan 18, 2007.

REFERENCES TO TABLE I

[1] Araga S, Kishimoto M, Doi S, Nakashima K: A complementary peptide vaccine that induces T cell anergy and prevents experimental allergic neuritis in Lewis rats. *J. Immunol.* 1999, 163:476-482.

[2] Araga S, LeBoeuf RD, Blalock JE: Prevention of experimental autoimmune myasthenia gravis by manipulation of the immune network with a complementary peptide for the acetylcholine receptor. *Proc. Natl. Acad. Sci. USA* 1993, 90:8747-8751.

[3] Bajpai A, Hooper KP, Ebner KE: Interactions of antisense peptides with ovine prolactin. *Biochem. Biophys. Res. Commun.* 1991, 180:1312-1317.

[4]   Baranyi L, Campbell W, Ohshima K, Fujimoto S, Boros M, Okada H: The antisense homology box: a new motif within proteins that encodes biologically active peptides. *Nat. Med.* 1995, 1:894-901.

[5]   Baranyi L, Campbell W, Okada H: Antisense homology boxes in C5a receptor and C5a anaphylatoxin: a new method for identification of potentially active peptides. *J. Immunol.* 1996, 157:4591-4601.

[6]   Bhakoo A, Raynes JG, Heal JR, Keller M, Miller AD. De-novo design of complementary (antisense) peptide mini-receptor inhibitor of interleukin 18 (IL-18). *Mol. Immunol.* 2004 Nov; 41(12):1217-24.

[7]   Blalock JE, Bost KL: Binding of peptides that are specified by complementary RNAs. *Biochem. J.* 1986, 234:679-683.

[8]   Borovsky D, Powell CA, Nayar JK, Blalock JE, Hayes TK: Characterization and localization of mosquito-gut receptors for trypsin modulating oostatic factor using a complementary peptide and immunocytochemistry. *FASEB J.* 1994, 8:350-355.

[9]   Bost KL, Blalock JE: Production of anti-idiotypic antibodies by immunization with a pair of complementary peptides. *J. Mol. Recognit.* 1989, 1:179-183.

[10]   Bost KL, Smith EM, Blalock JE: Similarity between the corticotrophin (ACTH) receptor and a peptide encoded by an RNA that is complementary to ACTH mRNA. *Proc. Natl. Acad. Sci. USA* 1985, 82:1372-1375.

[11]   Bost KL, Smith EM, Blalock JE: Similarity between the corticotrophin (ACTH) receptor and a peptide encoded by an RNA that is complementary to ACTH mRNA. *Proc. Natl. Acad. Sci. USA* 1985, 82:1372-1375.

[12]   Brentani RR, Ribeiro SF, Potocnjak P, Pasqualini R, Lopes JD, Nakaie CR: Characterization of the cellular receptor for fibronectin through a hydropathic complementarity approach. *Proc. Natl. Acad. Sci. USA* 1988, 85:364-367.

[13]   Bret-Dibat JL, Zouaoui D, Déry O, Zerari F, Grassi J, Maillet S, Conrath M, Couraud JY: Antipeptide polyclonal antibodies that recognize a substance P-binding site in mammalian tissues: a biochemical and immunocytochemical study. *J. Neurochem.* 1994, 63:333-343.

[14]   Campbell-Thompson M, McGuigan JE: Canine parietal cell binding by antibodies to the complementary peptide of somatostatin. *Am. J. Med. Sci.* 1993, 305:365-373.

[15]   Carr DJ, Blalock JE, Bost KL: Monoclonal antibody against a peptide specified by [Met]-enkephalin complementary RNA recognizes the delta-class opioid receptor. *Immunol. Lett.* 1989, 20:181-186.

[16]   Carr DJ, Bost KL, Blalock JE: An antibody to a peptide specified by an RNA that is complementary to gamma-endorphin mRNA recognizes an opiate receptor. *J. Neuroimmunol.* 1986, 12:329-337.

[17]   Castronovo V, Taraboletti G, Sobel ME: Laminin receptor complementary DNA-deduced synthetic peptide inhibits cancer cell attachment to endothelium. *Cancer Res.* 1991, 51:5672-5678.

[18]   Derrick JM, Taylor DB, Loudon RG, Gartner TK: The peptide LSARLAF causes platelet secretion and aggregation by directly activating the integrin alphaIIbbeta3. *Biochem. J.* 1997, 325(Pt 2):309-313.

[19]   Dillon J, Woods WT, Guarcello V, LeBoeuf RD, Blalock JE: A peptide mimetic of calcium. *Proc. Natl. Acad. Sci. USA* 1991, 88:9726-9729.

[20]   Elton TS, Oparil S, Blalock JE: The use of complementary peptides in the purification of an angiotensin II binding protein. *J. Hypertens* 1988, 6(Suppl):S404-407.

[21]   Fassina G, Cassani G, Corti A: Binding of human tumor necrosis factor alpha to multimeric complementary peptides. *Arch. Biochem. Biophys.* 1992, 296:137-143.

[22]   Fassina G, Cassani G, Gnocchi P, Fornasiero MC, Isetta AM: Inhibition of interleukin-2/p55 receptor subunit interaction by complementary peptides. *Arch. Biochem. Biophys.* 1995, 318:37-45.

[23]   Fassina G, Cassani G, Corti A: Binding of human tumor necrosis factor alpha to multimeric complementary peptides. *Arch. Biochem. Biophys.* 1992, 296:137-143.

[24]   Fassina G, Consonni R, Zetta L, Cassani G. Design of hydropathically complementary peptides for Big Endothelin affinity purification. *Int. J. Peptide Protein Res.* 1992, 39:540-548.

[25]   Fassina G, Corti A, Cassani G: Affinity enhancement of complementary peptide recognition. *Int. J. Peptide Protein Res.* 1992, 39:549-556.

[26]   Fassina G, Roller PP, Olson AD, Thorgeirsson SS, Omichinski JG: Recognition properties of peptides hydropathically complementary to residues 356-375 of the c-raf protein. *J. Biol. Chem.* 1989, 264:11252-11257.

[27]   Fassina G, Zamai M, Brigham-Burke M, Chaiken IM: Recognition properties of antisense peptides to Arg8-vasopressin/bovine neurophysin II biosynthetic precursor sequences. *Biochemistry* 1989, 28:8811-8818.

[28]   Fujita E, Farkas I, Campbell W, Baranyi L, Okada H, Okada N: Inactivation of C5a anaphylatoxin by a peptide that is complementary to a region of C5a. *J. Immunol.* 2004, 172:6382-6387.

[29]   Gartner TK, Loudon R, Taylor DB: The peptides APLHK, EHIPA and GAPL are hydropathically equivalent peptide mimics of a fibrinogen binding domain of glycoprotein IIb/IIIa. *Biochem. Biophys. Res. Commun.* 1991, 180:1446-1452.

[30]   Gartner TK, Taylor DB: The peptide Glu-His-Ile-Pro-Ala binds fibrinogen and inhibits platelet aggregation and adhesion to fibrinogen and vitronectin. *Proc. Soc. Exp. Biol. Med.* 1991, 198:649-655.

[31]   Ghiso J, Saball E, Leoni J, Rostagno A, Frangione B: Binding of cystatin C to C4: the importance of sense-antisense peptides in their interaction. *Proc. Natl. Acad. Sci. USA* 1990, 87:1288-1291.

[32]   Gho YS, Chae CB: Anti-angiogenin activity of the peptides complementary to the receptor-binding site of angiogenin. *J. Biol. Chem.* 1997, 272:24294-24299.

[33]   Gho YS, Lee JE, Oh KS, Bae DG, Chae CB: Development of antiangiogenin peptide using a phage-displayed peptide library. *Cancer Res.* 1997, 57:3733-3340.

[34]   Heal JR, Bino S, Ray KP, Christie G, Miller AD, Raynes JG: A search within the IL-1 type I receptor reveals a peptide with hydropathic complementarity to the IL-1beta trigger loop which binds to IL-1 and inhibits in vitro responses. *Mol. Immunol.* 1999, 36:1141-1148.

[35]   Holsworth DD, Kiely JS, Root-Bernstein RS, Overhiser RW: Antisense-designed peptides: a comparative study focusing on possible complements to angiotensin II. *Peptide Res.* 1994, 7:185-193.

[36]   Imai M, Okada N, Okada H: Inhibition of HIV-1 infection by an intramolecular antisense peptide to T20 in gp160. *Microbiol. Immunol.* 2000, 44:205-12.

[37] Johnson HM, Langford MP, Lakhchaura B, Chan TS, Stanton GJ: Neutralization of native human gamma interferon (HuIFN gamma) by antibodies to a synthetic peptide encoded by the 5' end of HuIFN gamma cDNA. *J. Immunol.* 1982, 129:2357-2359.

[38] Kang CY, Brunck TK, Kieber-Emmons T, Blalock JE, Kohler H: Inhibition of self-binding antibodies (autobodies) by a VH-derived peptide. *Science* 1988, 240:1034-1036.

[39] Kelly JM, Trinder D, Phillips PA, Casley DJ, Kemp B, Mooser V, Johnston CI: Vasopressin antisense peptide interactions with the V1 receptor. *Peptides* 1990, 11:857-862.

[40] Knigge KM, Piekut DT, Berlove D: Immunocytochemistry of a vasopressin (AVP) receptor with anti-idiotype antibody: inhibition of staining with a peptide (PVA) encoded by an RNA that is complementary to AVP mRNA. *Neurosci. Lett.* 1988, 86:269-271.

[41] Knutson VP: Insulin-binding peptide. Design and characterization. *J. Biol. Chem.* 1988, 263:14146-14151.

[42] Kwak JW, Kim HK, Chae CB: Potential lead for an Alzheimer drug: a peptide that blocks intermolecular interaction and amyloid beta protein-induced cytotoxicity. *J. Med. Chem.* 2006, 49:4813-4817.

[43] Lu FX, Aiyar N, Chaiken I: Affinity capture of [Arg8]vasopressin-receptor complex using immobilized antisense peptide. *Proc. Natl. Acad. Sci. USA* 1991, 88:3642-3646.

[44] Ludwig LB, Ambrus JL, Krawczyk KA, Sharma S, Brooks S, Hsiao CB, Schwartz SA: Human Immunodeficiency Virus-Type 1 LTR DNA contains an intrinsic gene producing antisense RNA and protein products. *Retrovirology* 2006, 3:80.

[45] Luo J, Zhang Q, Huang Y, Liu G, Zhao R: Quartz crystal microbalance biosensor for recombinant human interferon-beta detection based on antisense peptide approach. *Anal. Chim. Acta* 2007, 590:91-97.

[46] Martins VR, Graner E, Garcia-Abreu J, de Souza SJ, Mercadante AF, Veiga SS, Zanata SM, Neto VM, Brentani RR: Complementary hydropathy identifies a cellular prion protein receptor. *Nat. Med.* 1997, 3:1376-1382.

[47] McGuigan JE, Campbell-Thompson M: Complementary peptide to the carboxyl-terminal tetrapeptide of gastrin. *Gastroenterology* 1992, 103:749-758.

[48] McGuigan JE: Antibodies to complementary peptides as probes for receptors. *Immunomethods* 1994, 5:158-166.

[49] Moore GJ, Ganter RC, Franklin KJ: Angiotensin 'antipeptides': (-)messenger RNA complementary to human angiotensin II (+)messenger RNA encodes an angiotensin receptor antagonist. *Biochem. Biophys. Res. Commun.* 1989 160:1387-1391.

[50] Mulchahey JJ, Neill JD, Dion LD, Bost KL, Blalock JE: Antibodies to the binding site of the receptor for luteinizing hormone-releasing hormone (LHRH): generation with a synthetic decapeptide encoded by an RNA complementary to LHRH mRNA. *Proc. Natl. Acad. Sci. USA* 1986, 83:9714-718.

[51] Okada N, Asai S, Hotta A, Miura N, Ohno N, Farkas I, Hau L, Okada H: Increased inhibitory capacity of an anti-C5a complementary peptide following acetylation of N-terminal alanine. *Microbiol. Immunol.* 2007, 1:439-443.

[52] Papamattheou MG, Routsias JG, Karagouni EE, Sakarellos C, Sakarellos-Daitsiotis M, Moutsopoulos HM, Tzioufas AG, Dotsika EN: T cell help is required to induce idiotypic-anti-idiotypic autoantibody network after immunization with complementary

epitope 289-308aa of La/SSB autoantigen in non-autoimmune mice. *Clin. Exp. Immunol.* 2004, 135:416-426.

[53] Pascual DW, Blalock JE, Bost KL: Antipeptide antibodies that recognize a lymphocyte substance P receptor. *J. Immunol.* 1989, 143:3697-3702.

[54] Pasqualini R, Chamone DF, Brentani RR: Determination of the putative binding site for fibronectin on platelet glycoprotein IIb-IIIa complex through a hydropathic complementarity approach. *J. Biol. Chem.* 1989, 264:14566-14570.

[55] Pfister RR, Haddox JL, Blalock JE, Sommers CI, Coplan L, Villain M: *Synthetic complementary peptides inhibit a neutrophil chemoattractant found in the alkali-injured cornea. Cornea* 2000, 19:384-389.

[56] Root-Bernstein RS, Westall FC: Bovine pineal antireproductive tripeptide binds to luteinizing hormone-releasing hormone: a model for peptide modulation by sequence specific peptide interactions? *Brain Res. Bull* 1986, 17:519-528.

[57] Routsias JG, Dotsika E, Touloupi E, Papamattheou M, Sakarellos C, Sakarellos-Daitsiotis M, Moutsopoulos HM, Tzioufas AG: Idiotype-anti-idiotype circuit in non-autoimmune mice after immunization with the epitope and complementary epitope 289-308aa of La/SSB: implications for the maintenance and perpetuation of the anti-La/SSB response. *J. Autoimmun.* 2003, 21:17-26.

[58] Routsias JG, Touloupi E, Dotsika E, Moulia A, Tsikaris V, Sakarellos C, Sakarellos-Daitsiotis M, Moutsopoulos HM, Tzioufas AG: Unmasking the anti-La/SSB response in sera from patients with Sjogren's syndrome by specific blocking of anti-idiotypic antibodies to La/SSB antigenic determinants. *Mol. Med.* 2002, 8:293-305.

[59] Sakarellos-Daitsiotis M, Cung MT, Sakarellos C, El Hilali Z, Kosmopoulou A, Voitharou C: Complementary peptide epitopes and anti-idiotypic antibodies in autoimmunity. *Protein Peptide Lett.* 2004, 11:367-375.

[60] Sautebin L, Rombolà L, Di Rosa M, Caliendo G, Perissutti E, Grieco P, Severino B, Santagada V: Synthesis and structure-activity of antisense peptides corresponding to the region for CaM-binding domain of the inducible nitric oxide synthase. *Eur. J. Med. Chem.* 2000, 35:727-732.

[61] Scapol L, Rappuoli P, Viscomi GC: Purification of recombinant human interferon-beta by immobilized antisense peptides. *J. Chromatogr.* 1992, 600:235-242.

[62] Shahabi NA, Bost KL, Madhok TC, Sharp BM: Characterization of antisera to the naloxone-insensitive receptor for beta-endorphin on U937 cells generated by using the complementary peptide strategy. *J. Pharmacol. Exp. Ther.* 1992, 263:876-883.

[63] Shai Y, Brunck TK, Chaiken IM: Antisense peptide recognition of sense peptides: sequence simplification and evaluation of forces underlying the interaction. *Biochemistry* 1989, 28:8804-8811.

[64] Shai Y, Flashner M, Chaiken IM: Anti-sense peptide recognition of sense peptides: direct quantitative characterization with the ribonuclease S-peptide system using analytical high-performance affinity chromatography. *Biochemistry* 1987, 26:669-675.

[65] Wu X, Richards NT, Johns EJ, Kohsaka T, Nakamura A, Okada H: Influence of ETR-p1/f1 antisense peptide on endothelin-induced constriction in rat renal arcuate arteries. *Br. J. Pharmacol.* 1997, 122:316-320.

Chapter 9

'ON/OFF'-SWITCHED MOLECULAR RECOGNITION BY A SMART AMINOPURINE-IMPRINTED POLYMER

Songjun Li[1], Ashutosh Tiwari[2,3] and Mani Prabaharan[3]*

[1]Key Laboratory of Pesticide and Chemical Biology of Ministry of Education,
College of Chemistry, Central China Normal University,
Wuhan 430079, China
[2]Division of Engineering Materials, National Physical Laboratory,
Dr. K.S. Krishnan Marg, New Delhi 110012, India
[3]Department of Mechanical Engineering,
University of Wisconsin-Milwaukee, Milwaukee 53211, Wisconsin, USA

ABSTRACT

The 'on/off'-switched molecular recognition by a smart imprinted polymer (MIP-T) was presented in this chapter. The smart MIP-T was prepared using PNIPA as the thermosensitive element and 2-aminopurine as the template. The thermal phase transition of PNIPA induced a self-switching ability in the prepared polymer. At a relatively low temperature (such as 20 $^{\circ}$C), the MIP-T was capable of highly specifically recognizing the imprint species, i.e., 2-aminopurine. However, above the transition temperature, the MIP-T did not demonstrate significant resolution for 2-aminopurine compared with its analogue 2-amino-6-methylpurine. Such temperature-responsive recognition, in nature, was comparable to an on/off-switched process, which allows an efficient self-regulation in the molecular recognition behavior.

Keywords: Molecular imprinting; polymers; molecular recognition; modulation.

1. INTRODUCTION

Molecularly imprinted polymers (MIPs), also known as "plastic antibodies", have attracted much attention [1-3]. As the efficient molecular recognition systems, MIPs contain

usually selectively binding frameworks within their polymeric networks. This character makes them especially attractive to extraction, sensing, separation and catalysis [4]. To fabricate the MIPs, functional monomers and template are first allowed to form a self-organized architecture via molecular self-assembly [5]. Polymerization in the presence of a crosslinker is then performed to fix the self-organized architecture. The imprinted template is then removed from the matrix networks, leaving behind binding sites complementary to the template in terms of shape and structure. The arrangement of these binding sites constitutes a template's memory, which makes the polymer capable of specifically recognizing the imprint species (i.e., the template). Thus, molecular recognition by MIPs is generally comparable to a key and a lock.

On the forefront of MIPs is to develop the intelligent systems capable of self-regulating the recognition behavior. Recent advances in 'smart' materials have made this target possible [6,7]. Poly(N-isopropylacylamide) (PNIPA), an excellent smart material, has attracted special attentions due to its unique phase transition at a moderate temperature. Consisting of hydrophilic amide-chains and hydrophobic isopropyl-groups, the PNIPA polymer in aqueous media exhibits a lower critical solution temperature (LCST) at about 32 ^{o}C [8]. Below the LCST, the polymer is soluble in water because of the hydrogen bonding interactions, which are formed by the amide chains and water. Above the LCST, the relative equilibrium of hydrophilicity/hydrophobicity is broken up due to a thermodynamic transition. Water is expelled from the network interior, causing a drastic decrease in the volume of this polymer. Thus, inspired by such a unique mechanism, some researchers have developed the so-called 'smart MIPs', which are capable of self-regulating the recognition behavior.

Using PNIPA as the 'smart' element and methacrylic acid as the monomer, Alvarez-Lorenzo *et al.* have reported Ca^{2+}, Pd^{2+}-imprinted polymers [9,10]. Coinciding with the operation temperature, the prepared MIPs demonstrated the self-switching recognition for the imprinted templates. Below the transition temperature of PNIPA, the prepared polymers can specifically absorb the imprinted species. However, above the transition temperature, the prepared polymers provided a dramatically decreased affinity to their templates. Similarly, with PNIPA as the thermosensitive composition and 4-(vinylbenzyl)ethylenediamine (VBEDA) as the chelating and crosslinking agent, Tokuyama *et al.* have prepared a Cu^{2+}-imprinted polymer [11]. The result indicated that the self-switching recognition behavior by smart MIPs was potentially relevant to the deformation and/or distortion of molecularly imprinted networks, which may cause decreased recognition abilities. Although some analogous studies are also recently available [12,13], the development of smart MIPs is still in its infancy. The recognition mechanism by smart MIPs and its self-regulation are not clear yet. Furthermore, little work on the nonmetal-imprinted 'smart MIPs' has been performed.

We here presented a 2-aminopurine (AP)-imprinted smart polymer and studied its self-switching recognition behavior. AP was selected as the template because it is soluble in water and easily forms electrostatic interactions with the common hydrophilic monomer 2-acrylamide-2-methylpropane -sulfonic acid (AMPS). Meanwhile, the AP-imprinted polymers have been well studied in literature (cf. Scheme 1) [14,15]. These advantages from the AP-imprinted polymers therefore provide valuable promotion for our present study. To investigate the specificity of the prepared polymer, the structural analogue of AP, i.e., 2-amino-6-methylpurine (AMP), was selected as the control. The objective of this study is to further the general understanding on the self-regulated by smart MIPs.

Molecular self-assembly

Precursor

Remove template

Molecular recognition
in water
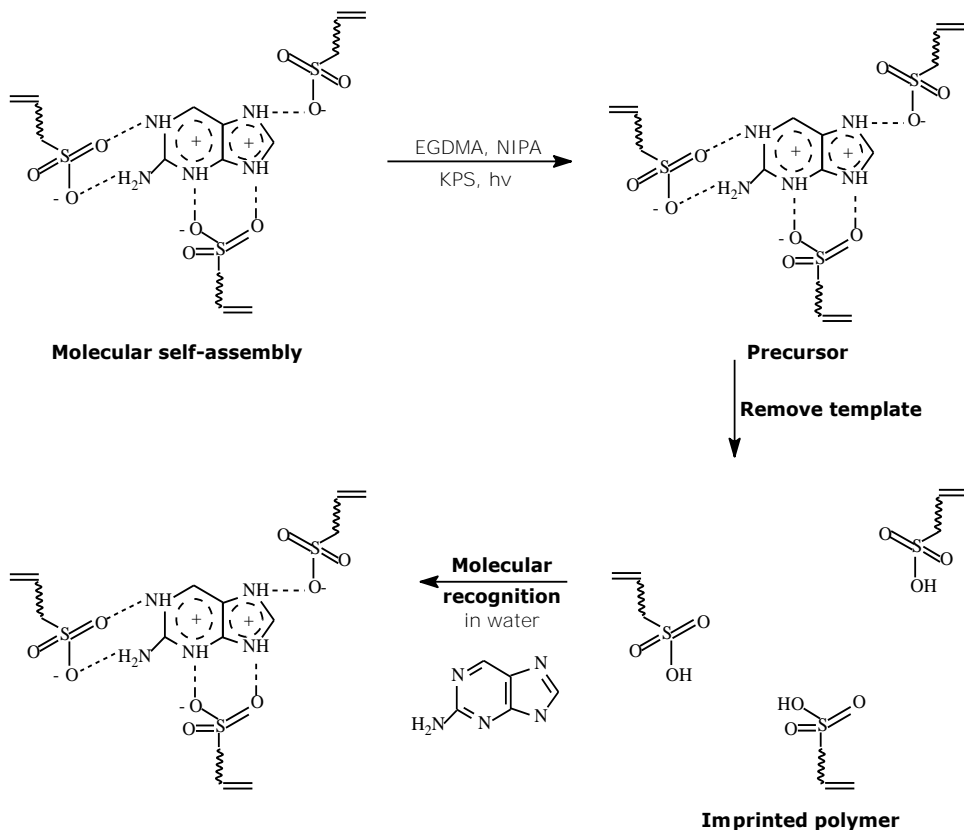
Imprinted polymer

Scheme 1. Technical outline for the preparation of MIP-T.

2. EXPERIMENTAL SECTION

2.1. PREPARATION OF IMPRINTED POLYMERS

In this study, all the imprinted polymers concerned were prepared and treated following classic molecular imprinting processes (cf. Scheme 1) [16,17]. The chemicals used were commercially available products of reagent grade and used as received. Template, monomers, crosslinker and initiator were dissolved in an acetonitrile-water solution (3:1, v/v; totally 16 mL) (cf. Table 1). After the system was dispersed and deoxygenated with sonication and nitrogen, polymerization was fully performed with the excitation of ultraviolet light (365 nm). The obtaining imprinted polymer precursor was roughly crushed and subsequently washed with a mixture of methanol and acetic acid (8:2, v/v) to remove the imprinted template and minimal unreacted monomers. The resulting imprinted polymer (i.e., MIP-T) was dried in a vacuum vessel at room temperature and then ground into a size of 50-60 mesh. For comparison, two controls named "MIP" and "NIP-T" were also prepared under comparable conditions (cf. Table 1). The MIP was prepared with the same procedures used for MIP-T except that N-isopropylacylamide (NIPA) was not used. The NIP-T is a non-imprinted polymer and no any AP was used during the preparation process.

Table 1. Synthesis composition for the imprinted and non-imprinted polymers

Composition	MIP-T	MIP	NIP-T
AP	0.35 g	0.35 g	0 g
AMPS	1.45 g	1.45 g	1.45 g
EGDMA[*]	7.0 mL	7.0 mL	7.0 mL
KPS [*]	0.1 g	0.1 g	0.1 g
NIPA	4.9 g	0 g	4.9 g

[*]EGDMA is ethylene glycol dimethacrylate and KPS is potassium peroxodisulfate.

2.2. TEMPERATURE-PROGRAMMED DESORPTION

Temperature-programmed desorption (TPD) was performed to evaluate the interaction between imprinted polymers and substrate [18]. Typically, using a device composed of a gas chromatography (TCD) and a data processing system (102G, China), the imprinted polymers (200 mg) were placed into an online U-shaped quartz tube (4 mm I.D.). After the polymers adsorbed with 10 μL substrate (1.5 μmol mL^{-1}) (i.e., AP or AMP), the polymer system was heated in a nitrogen flow (43 mL min^{-1}; 0.27 MPa) at a rate of 10 ^{o}C min^{-1} from room temperature up to the temperature at which the absorbed substrate desorbed. The desorbing signal was recorded by the data processing system.

2.3. EVALUATION OF PHASE TRANSITION

Dynamic light scattering (DLS) and swelling experiments were used to evaluate the phase transition. The DLS analysis was carried out at a scattering angle of 90° using a goniometer equipped with a He-Ne laser (Langen, Germany). To allow equilibrium, all samples were kept at specific temperatures for at least 20 min before acquiring the scattering degree. For the swelling experiments, dried samples (in triplicate) were immersed to deionized water until equilibrium was reached. These wet samples were subsequently weighed (W_t) and dried until a constant weight was achieved (W_d). The swelling ratio was calculated using the following equation and the average of the triplicate runs was reported.

$$S = \frac{W_t - W_d}{W_d} \times 100\%$$

2.4. SORPTION TEST

The sorption of the prepared polymers was performed using a batch format. In triplicate, the substrate AP was dissolved to deionized water (initial concentration, 1.2 μmol mL^{-1}; totally 20 ml). The solid content of the imprinted polymers used was 5 mg mL^{-1} in each operation. The extent to adsorption was determined at regular intervals and the change in the

concentration of solute was spectrophotometrically monitored at 274 nm. The adsorption amount of the polymer was obtained from the mass balance of solute and finally the average of the triplicate runs was presented. To study the specific recognition, the adsorption of the AP analogue, i.e., AMP, as the control, was carried out under comparable conditions.

2.5. DYNAMIC ADSORBING-DESORBING CYCLIC VOLTAMMETRY

It is known that the potential of reducing/oxidizing a binding molecule depends on the binding constant. A larger binding constant would require higher energy to overcome the binding, thereby causing the larger redox potential. Thus, the dynamic adsorbing-desorbing cyclic voltammetry (DCV) can provide valuable information on the binding behavior between molecularly imprinted polymers and their templates [19]. Typically, using an electrochemical workstation equipped with a three-electrode configuration (Pt-working and auxiliary electrodes, and Ag/Ag$^+$-ref. electrode) (CHI-600, USA), the imprinted polymers (20 mg) pre-adsorbed with 1μmol template were placed into an online cuvette equipped with a self-rotation unit (supporting electrolyte: 0.01 mmol mL^{-1} KNO$_3$; 10 mL). The absorbed template in the adsorbing/desorbing equilibrium was consecutively scanned by the workstation up to 20-cycles until the stable DCV diagram was obtained (scanning range, *ca.* 1.2~ -1.2 *V*; scanning rate, 0.1 $V \cdot s^{-1}$).

3. RESULTS AND DISCUSSION

3.1. FTIR AND SEM ANALYSIS OF MOLECULAR IMPRINTING

To show the imprinting behavior, Figure 1 shows the FTIR spectra of the prepared polymers.
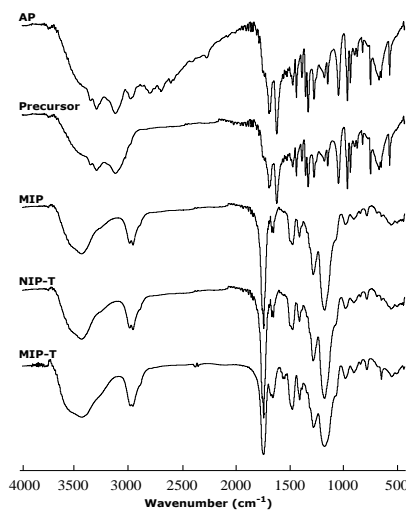
Figure 1. IR spectra of the prepared polymers.

Three main absorption bands (3250-3600, 2850-3050 and ~1750 cm^{-1}) and some fingerprints appeared in the spectra. In basic conditions [20,21], these main absorption bands can be attributed to the stretching of O-H/N-H, C-H and C=O, respectively. The fingerprints may arise from the vibration of C-N and C-C bonds and rotation of various groups. For comparison, we also included in Figure 1 the spectra of template and the MIP-T precursor (i.e., the MIP-T system in which the template AP had not been removed from the matrix yet). The MIP-T precursor exhibited the absorption band of AP. After washing, the spectrum of the resulting polymer (i.e., MIP-T) became comparable to that of the NIP-T. Figure 2 presents the SEM images of these prepared polymers. The NIP-T exhibited relatively smooth morphology. Both MIP-T and MIP appeared with speckles and cavities. Relating to the preparation process (cf. Scheme 2.1), the IR and SEM results reflect a consequence of imprinting (further information regarding imprinting was provided in Section 3.2).
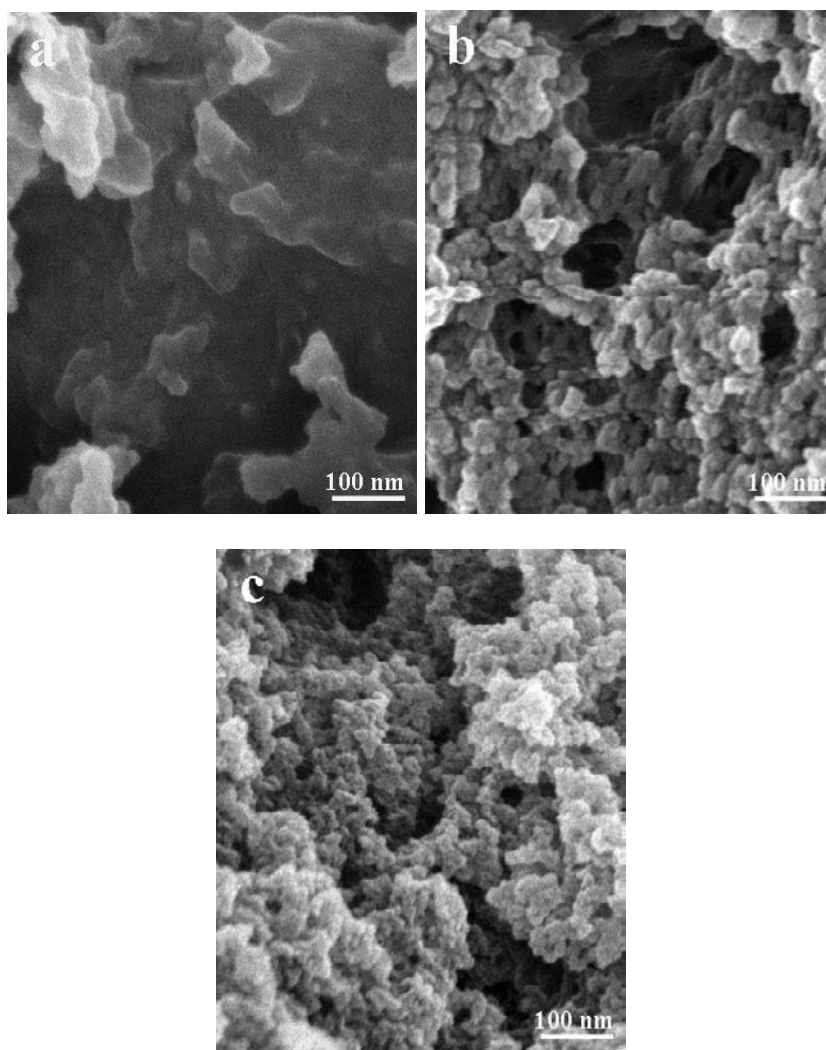
Figure 2. SEM images of the prepared polymers (a: NIP-T; b: MIP; c: MIP-T).

3.2. SPECIFIC INTERACTION BETWEEN IMPRINTED POLYMERS AND SUBSTRATE

Figure 3 presents the TPD profiles. AP, i.e., the imprint species, desorbed from MIP-T at *ca.* 199 °C, which the analogue AMP was desorbed from MIP-T at 176 °C. The MIP-T demonstrated apparently a stronger interaction with AP compared with AMP. To further study the interaction behavior, we also included the TPD profiles of two mentioned controls to Figure 3. The NIP-T did not show significant resolution between AP and AMP. The MIP showed an expected resolution for AP and its analogue. This result strongly indicates that the interaction offered by MIP-T to AP was highly AP-selective. This highly AP-selective interaction, as already explained, may be a result of the imprinting behavior. Since molecular recognition by molecularly imprinted polymers is essentially a result of the template-induced memory, the specific interaction between two AP-imprinted polymers (i.e., MIP-T and MIP) and AP is expected.
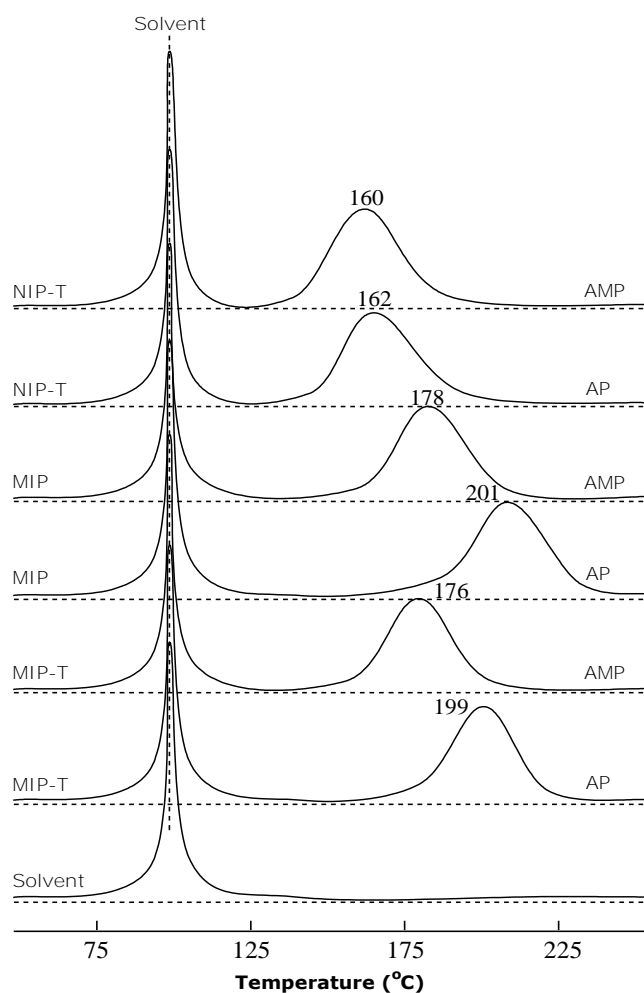
Figure 3. TPD profiles of the prepared polymers.

3.3. PHASE TRANSITION BEHAVIOR

Figure 4 shows the DLS curves of the prepared polymers. The scattering degree of these prepared polymers increased with temperature (conceptually relating to a decreased hydrodynamic radius, R_h). The MIP-T and NIP-T showed a stronger dependence on temperature, compared with MIP. A dramatic transition occurred at *ca.* 33 °C in the MIP-T and NIP-T.
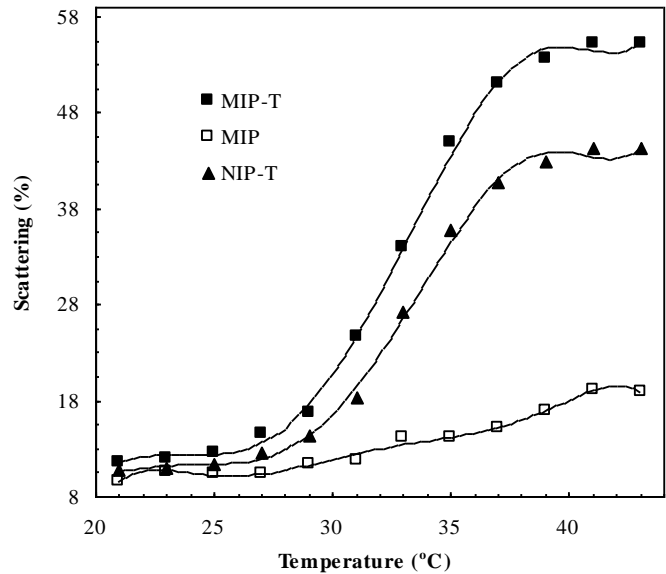
Figures 4. DLS curves of the prepared polymers.
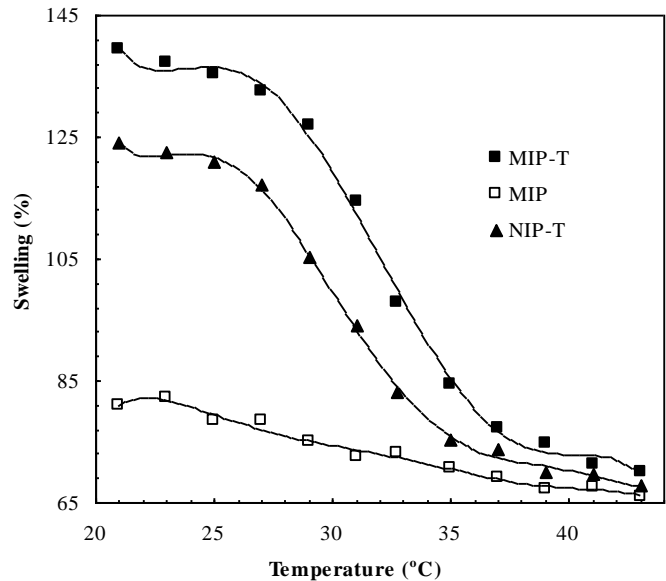
Figure 5. Swelling curves of the prepared polymers.

Below the transition temperature, the MIP-T and NIP-T demonstrated a relatively small scattering degree (i.e., relatively large R_h). Above the transition temperature, both MIP-T and NIP-T demonstrated a dramatically increased scattering. Figure 5 presents the swelling curves of these prepared polymers. Similar to the DLS curves, the swelling curves of both MIP-T and NIP-T showed also a transition temperature at ~33 °C. Below the transition temperature, the MIP-T and NIP-T demonstrated a relatively large swelling ratio. Above the transition temperature, both MIP-T and NIP-T demonstrated a dramatically decreased scattering. Clearly, the DLS and swelling curves reflect a consequence of the thermal phase transition of PNIPA. Below the transition temperature, the relatively large R_h and swelling may be due to the hydrophilic network of PNIPA, which provided access for water to the polymer interior. On the contrary, the hydrophobic network of PNIPA dispelled the water from the polymer, blocking the access for water to the gel network. As a result, the MIP-T and NIP-T demonstrated a dramatically decreased R_h and swelling degree above the transition temperature.

3.4. SWITCHED MOLECULAR RECOGNITION

Figure 6 shows the adsorption curves of the prepared polymers. To track the modulation, two typical temperatures, i.e., 20 and 40 °C, which are either lower or higher than the transition temperature 33 °C, were selected for comparison. The MIP-T at 20 °C exhibited highly specific recognition for the imprint molecule comparable to the MIP. The specificity of the MIP-S dramatically decreased at 40 °C to a level near that of the NIP-T. It appeared that molecular recognition by the MIP-T was similar to a switch-on and witch-off process. This result, as already explained, can be related to the unique phase transition behavior of MIP-T.
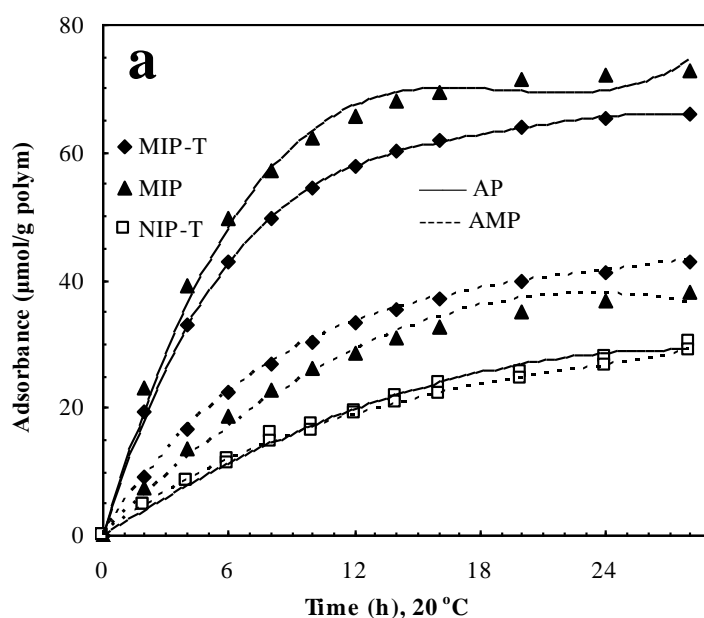
Figure 6. Continued
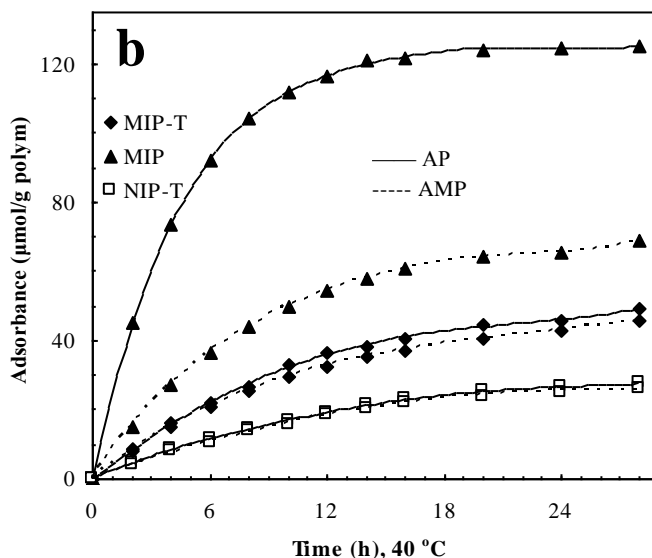
Figure 6. Adsorption curves of the prepared polymers (a: 20 oC; b: 40 oC).

The swelling of the imprinted polymer matrix of MIP-T at lower temperatures provided easy access for the substrate in water to the binding framework, thereby making efficient molecular recognition feasible. At higher temperatures, the shrinked polymer network largely blocked the diffusion of the substrate in water, which thereby led to the observed nonspecific adsorption behavior.

3.5. TRACK OF KINETICS

To track or simulate sorption, pseudo-first-order kinetics is commonly used [22-24]. This may be because such a process shows usually rapid adsorption at the beginning and then a slow approach to a limiting value, as is the case for this study. Thus, the pseudo-first-order kinetics is also tentatively used:

$$-\frac{dC}{dt} = kC \tag{1a}$$

or

$$-\frac{d(1-\theta)}{dt} = k(1-\theta) \tag{1b}$$

Here C and θ are the free concentration of substrate and the coverage degree, respectively, and $(1-\theta)$ represents the uncovered portion of a polymeric sorbent. Correlating this relationship at 20 $^{\circ}$C with 40 $^{\circ}$C (or correlating this relationship for the template with its analogue) would deduce the relative adsorption of an imprinted polymer:

$$\left.\frac{d(1-\theta_i)}{1-\theta_i}\right|_{T_1} = \frac{k_{T_1}}{k_{T_0}} \cdot \left.\frac{d(1-\theta_i)}{1-\theta_i}\right|_{T_0} \tag{2}$$

Here the subscripts 'T_0' and 'T_1' represent the relatively low and high temperatures (such as 20 and 40 $^\circ$C). Defining a constant R_a that reflects the relative adsorption will change Eqn (2) into Eqn (3):

$$\ln(1-\theta_i)_{T_1} = R_a \cdot \ln(1-\theta_i)_{T_0} \tag{3}$$

in which,

$$R_a \equiv \frac{k_{T_1}}{k_{T_0}}$$

Converting Eqn (3) to the standardized form will result in Eqn (4):

$$\ln\left(1-\frac{Q_t}{Q_m}\right)_{T_1} = R_a \cdot \ln\left(1-\frac{Q_t}{Q_m}\right)_{T_0} \tag{4}$$

Here Q_t and Q_m are the practical and the maximal adsorbed amounts, respectively. Based on Eqn (4), the relative adsorption at 20 and 40 $^\circ$C and the specificity of the template versus its analogue can be determined. Plotting $\ln(1-Q_t/Q_m)_{T1}$ versus $\ln(1-Q_t/Q_m)_{T0}$ showed a linear relationship (cf. Figures 7a and b). The relative adsorption/specificity can be revealed from the slope (i.e., A_r-value).
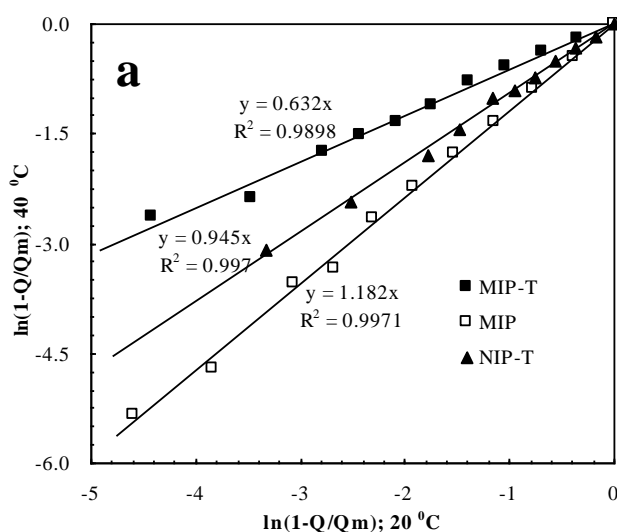
Figure 7. Continued
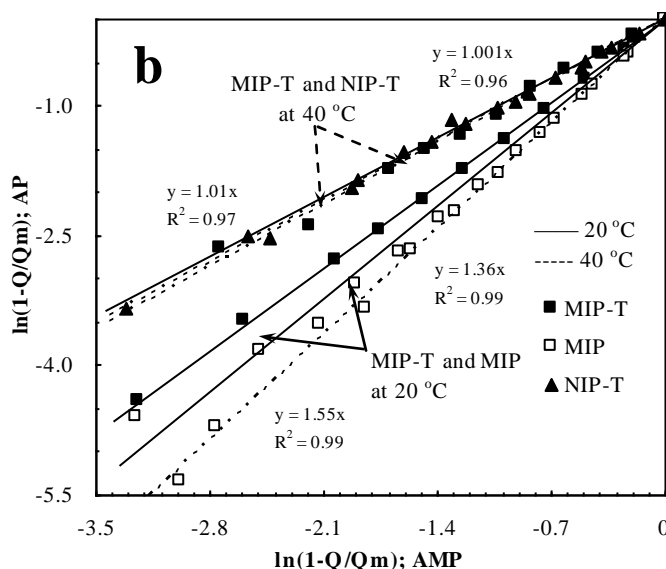
Figure 7. Kinetic simulation for the relative adsorption and specificity (a: 20 versus 40 oC; b: template versus analogue).

In the case of MIP (cf. Figure 7a), a higher temperature caused a larger adsorption (A_r >1). The increased temperature generated a limited effect on the NIP-T (R_a ≈1). However, the MIP-T demonstrated smaller adsorption at 40 ^{o}C than at 20 ^{o}C. Furthermore, the specificity of template versus analogue in MIP-T at 20 ^{o}C was comparable to that in MIP (cf. Figure 7b). The specificity of MIP-T became almost as small as that in NIP-S at 40 ^{o}C. The MIP-T apparently acted as a switch for molecular recognition.

3.6. DYNAMIC BINDING BEHAVIOR

Figures 8a and 8b present the DCV diagrams. At 20 ^{o}C, the imprint molecule binding to MIP-T showed a reduction potential at -394 mV; the reduction potential shifted to -320 mV at 40 ^{o}C. It appeared that the MIP-T have a stronger interaction with the imprint species at 20 ^{o}C than at 40 ^{o}C. For comparison, the reduction potentials of the imprint species from three controls are shown in Table 2. At 20 ^{o}C, the imprint molecule binding to MIP-T exhibited a reduction potential comparable to that of MIP (-394 versus -386 mV). The reduction potential exhibited by MIP-T became as small as that exhibited by NIP-S (-320 versus -309 mV) at 40°C.

Table 2. Reduction potentials of template binding to the prepared polymers

Polymer	Reduction potential at 20 ^{o}C (mV)	Reduction potential at 40 ^{o}C (mV)	Temperature effect (mV)
MIP-T	-394	-320	74
MIP	-386	-381	5
NIP-T	-322	-309	13

These results strongly indicate that the MIP-T presents an switchable interaction with the imprint molecule. Thus, modulated molecular recognition by the MIP-T is logically an external embodiment of this switchable interaction.
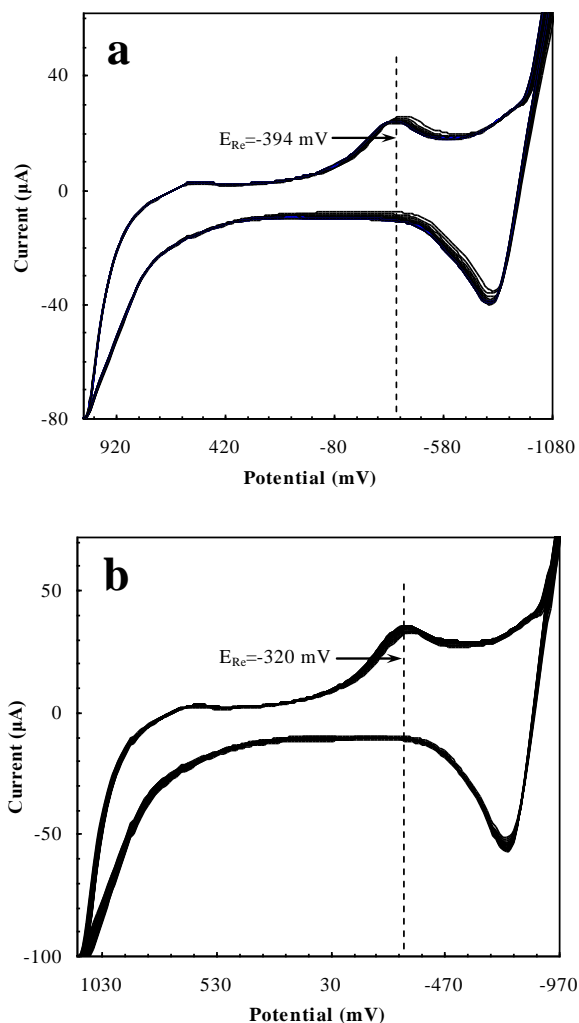
Figure 8. DCV diagrams of the imprint molecule binding to MIP-T (a: 20 oC; b: 40 oC).

CONCLUSIONS

A temperature-sensitive 2-aminopurine-imprinted polymer MIP-T was prepared using PNIPA as the thermosensitive element. The prepared MIP-T exhibited a temperature-responsive molecular recognition behavior. At a relatively low temperature (such as 20 ^{o}C), the MIP-T presented highly specific recognition for the imprint species. However, the MIP-T did not demonstrated significant resolution for the template and its analogue above the

transition temperature. Thus, the modulated molecular recognition behavior shown by the MIP-T is comparable to a switch-on and switch-off process.

ACKNOWLEDGMENTS

REFERENCES

[1]   Priego-Capote, F.; Ye, L.; Shakil, S.; Shamsi, S. A.; Nilsson, S. *Anal. Chem.* 2008, *80*, 2881-2887.
[2]   Gomy, C.; Schmitzer, A. R. *Org. Lett.* 2007, *9*, 3865-3868.
[3]   Li, J.; Kendig, C. E.; Nesterov, E. E. *J. Am. Chem. Soc.* 2007, *129*, 15911-15918.
[4]   Hoshino, Y.; Kodama, T.; Okahata, Y.; Shea, K. J. *J. Am. Chem. Soc.* 2008, *130*, 15242-15243.
[5]   Jacob, R.; Tate, M.; Banti, Y.; Rix, C.; Mainwaring, D. E. *J. Phys. Chem. A* 2008, *112*, 322-331.
[6]   Kurkuri, M. D.; Nussio, M. R.; Deslandes, A.; Voelcker, N. H. *Langmuir* 2008, *24*, 4238-4244.
[7]   Qian, J.; Wu, F. *Chem. Mater.* 2007, *19*, 5839-5841.
[8]   Arizmendi, R. C.; Del-Angel, M.; Romero-Garcia, *J. Langmuir* 2007, *23*, 8-12.
[9]   Alvarez-Lorenzo, C.; Guney, O.; Oya, T.; Sakai, Y.; Kobayashi, M.; Enoki, T.; Takeoka, Y.; Ishibashi, T.; Kuroda, K.; Tanaka, K.; Wang, G.; Grosberg. A. Y. *J. Chem. Phys.* 2001, *114*, 2812-2816.
[10] Moritani, T.; Alvarez-Lorenzo, C. *Macromolecules* 2001, *34*, 7796-7803.
[11] Tokuyama, H.; Kanazawa, R.; Sakohara, S. *Separat. Purif. Technol.* 2005, *44*, 152-159.
[12] Hua, Z.; Chen, Z.; Li, Y.; Zhao, M. *Langmuir* 2008, *24*, 5773-5780.
[13] Li, S.; Huang, X.; Zheng, M.; Li, W. *Anal. Bioanal. Chem.* 2008, *392*, 177-185.
[14] Huang, X.; Li, S. *J. Inorg. Organomet. Polym.* 2008, *18*, 277-283.
[15] Li, W.; Li, S. *Adv. Polym. Sci.* 2007, *206*, 191-210.
[16] Emgenbroich, M.; Borrelli, C.; Shinde, S.; Lazraq, I.; Vilela, F.; Hall, A. J.; Oxelbark, J.; De Lorenzi, E.; Courtois, J.; Simanova, A.; Verhage, J.; Irgum, K.; Karim, K.; Sellergren, B. *Chem. Eur. J.* 2008, *14*, 9516-9529.
[17] Beltran, A.; Marce, R. M.; Cormack, P. A. G.; Sherrington, D. C.; Borrull, F. *J. Separ. Sci.* 2008, *31*, 2868-2874.
[18] Li, S.; Tong, K.; Zhang, D.; Huang, X. *J. Inorg. Organomet. Polym.* 2008, *18*, 264-271.
[19] Tong, K.; Xiao, S.; Li, S.; Huang, J. *J. Inorg. Organomet. Polym.* 2008, *18*, 426-433.
[20] Tobita, M.; Yasuda, Y. *J. Phys. Chem. C* 2008, *112*, 13851-13855.
[21] Kakuda, H.; Okada, T.; Hasegawa, T. *J. Phys. Chem. B* 2008, *112*, 12940-12945.
[22] Fang, C.; Li, S. *J. Inorg. Organomet. Polym.* 2007, *17*, 623-629.
[23] Li, S.; Huang, X.; Zheng, M.; Li, W.; Tong, K. *Sensors* 2008, *8*, 2854-2864.
[24] Liu, Y.; Shen, L. *Langmuir* 2008, *24*, 11625-11630.

INDEX

N

O

P

W

X

Y

Z