Charles University in Prague
Faculty of Science
Department of Physical and Macromolecular Chemistry

Doctoral Thesis

Intermolecular Interactions in Proteins

Mgr. Jiří Kysilka

Supervisor:
RNDr. Jiří Vondrášek, Ph.D.
Consultant:
RNDr. Ota Bludský, Ph.D.

Institute of Organic Chemistry and Biochemistry AS CR
Center for Biomolecules and Complex Molecular Systems

Universita Karlova v Praze
Přírodovědecká fakulta
Katedra fyzikální a makromolekulární chemie

Disertační práce

Mezimolekulové interakce v proteinech

Mgr. Jiří Kysilka

Školitel:
RNDr. Jiří Vondrášek, Ph.D.
Konzultant:
RNDr. Ota Bludský, Ph.D.

Ústav organické chemie a biochemie AV ČR
Centrum biomolekul a komplexních molekulových systémů

I hereby declare that I have written the presented thesis solely by myself and all the literature is properly cited. Neither the thesis nor its parts have been used for obtaining any academic degree.

Prague, 2nd January 2013 Jiří Kysilka

Acknowledgement

I would like to thank Dr. Jiří Vondrášek and Dr. Ota Bludský for their guidance and support. I would also like to thank to all members of the Center for Biomolecules and Complex Molecular Systems for advice and assistance. Special thanks belong to Miroslav Rubeš for his kind introduction into the Matlab programming. I would also like to thank Boris Fačkovec, Jiří Vymětal, Jan Heyda, Jindřich Famfrlík, Erik Wernersson and Ondřej Maršálek for their inspiration, advice and support when dealing with computational procedures.

Contents

1 Preface

Proteins make up the most diverse group of macromolecules in the cell. They perform various functions – they are incredible catalysts and they are able to selectively recognize each other during the signaling and transport processes. All the wide range of their functions share one common denominator: molecular recognition – proteins excel in their ability to selectively recognize other small molecules, ions and other biomolecules, including other proteins or nucleic acids.

The function of proteins is directly related to their structure. Considering the complexity of protein chemistry and topology, it is incredible to realize that the covalent structure of proteins is given by the sequence of twenty basic amino acids. Each amino acid possesses wide possibilities of interactions. It is the non-covalent interactions that determine the actual structure of the protein and that dictate the interaction of the protein with other molecules. Protein folding is quite complex process, during which the protein acquires its native structure. It is controlled by the intermolecular interactions among the protein amino acids and also by the interaction of the protein molecule with a solvent. In terms of thermodynamics, the process can be described by the free energy decrease of the whole system. While the entropic contribution is quite difficult to be dealt with, enthalpic contribution can be described quite satisfactorily, using various computational strategies.

In order to understand the behavior of proteins, it is crucial to have reliable model of intermolecular interactions. While it is difficult to investigate the amino acid interactions experimentally, the theoretical methods are more suitable for this task. Yet, the accurate description of intermolecular interaction is still a challenge for the contemporary computational chemistry. As the most sophisticated wave function based methods are prohibitively expensive for the extended biological systems, one has to search for a compromise that is computationally feasible and yet provides reliable results.

This thesis explores the role of intermolecular interactions in proteins. First, it examines the physical nature of intermolecular interactions and summarizes their basic types. It outlines the available computational methods and discusses their suitability for the description of intermolecular interactions in proteins. Finally, it shows how intermolecular interactions influence the structure of proteins and which actual intermolecular interactions appear in proteins.

The motivation of this research is to understand the role of the intermolecular interaction of proteins. The first study answers the request for an accurate description of the non-covalent interaction of the protein functional groups. The DFT/CC method has been utilized for the analysis of the interaction of twelve small molecules with the graphitic surface, being the model of hydrophobic environment. It has been shown that the results can serve as a good benchmark for the estimation of the order of the interaction energies of the non-covalent interactions in proteins. In the following part of the work, the tools of bioinformatics were used to handle the experimental crystallographic data from the protein data bank. In the consequent study, the principles of the protein-protein interactions were studied on a set of protein dimers. The study focused on the role of side-chain interactions during the interaction process. The last study presents a systematic method for the analysis of the protein hydration structure. This method spots the sites of the protein molecule that play an important role for its interaction with water. It also enables the overall topology analysis of the hydration shell.

2 Intermolecular interactions

This chapter gives a basic overview of the intermolecular interaction from the physical point of view. Intermolecular interactions determine the physico-chemical properties of solids, liquids and gases[1].

2.1 Nature of intermolecular interaction

Although intermolecular interactions can be classified into several groups, it is important to note that they all share the same electromagnetic nature. In the figure 1 there is a typical potential curve of the interaction energy.

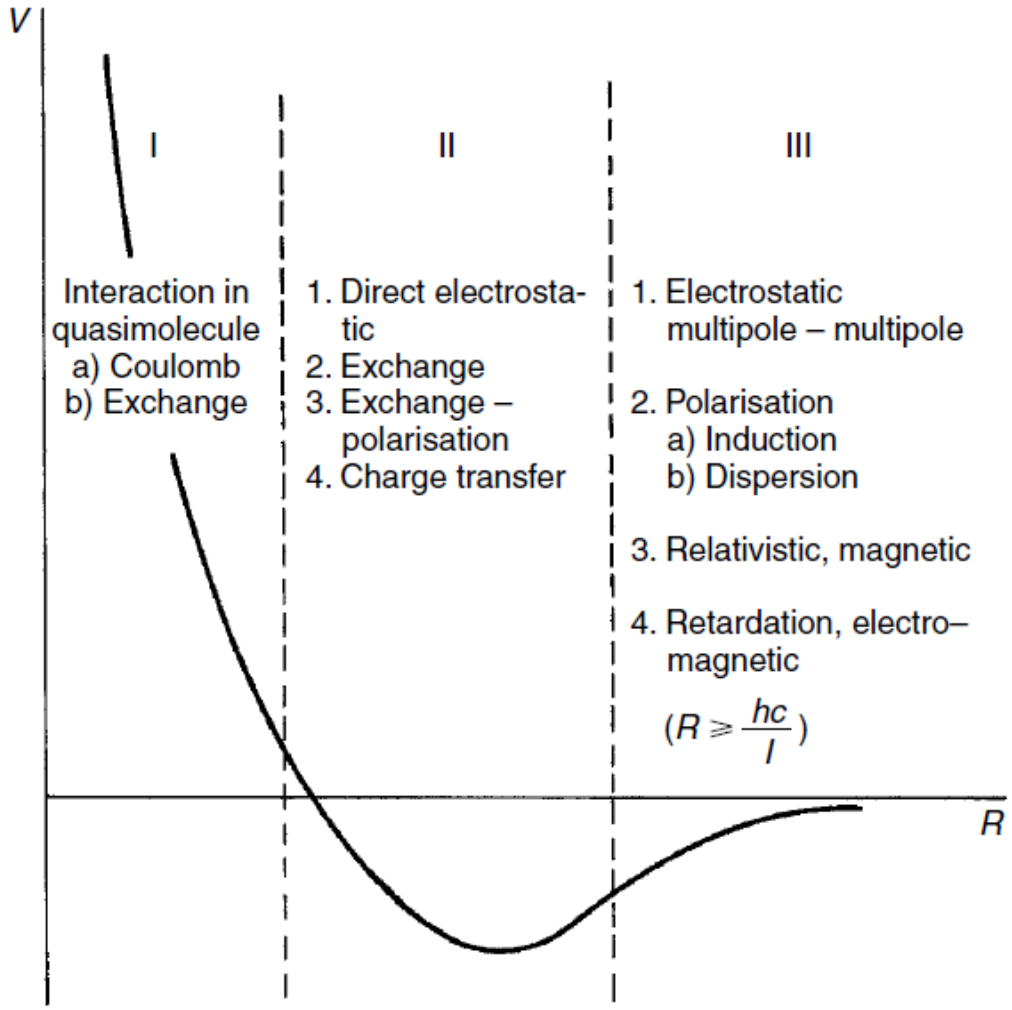

Fig. 1 – Classification of intermolecular interactions[1]

In each region of the intermolecular potential curve, different types of intermolecular forces play an important role.

If the distance of the two molecules is short enough, the intermolecular potential is repulsive and the dominating force is the electronic exchange that is the result of the overlapping molecular electronic shells.

In the medium range of intermolecular distances, there is a minimum of the potential energy. This is the result of balancing the attractive and repulsive forces. Interaction energy at this region is smaller than the energies of the individual molecules and can be therefore treated as a perturbation.

In the region of the long range of intermolecular distances exchange effects can be neglected. In this region, multipole expansion of the electrostatic potential can be used. If the distance is great enough, the first term of this expansion is sufficient for a good description of the interaction energy (the dipole term in case of the polar molecules).

2.2 Types of intermolecular interactions

2.2.1 Direct electrostatic interaction

For the system of two interacting molecules, the total Hamiltonian can be expressed as a sum of the Hamiltonian of the isolated molecules, H_0, and the operator describing the intermolecular interaction, V:

$$H = H_0 + V \,. \tag{1}$$

The operator of the intermolecular interaction, V, is defined as

$$V = -\sum_{a=1}^{n_A}\sum_{j=1}^{N_B}\frac{Z_a e}{r_{aj}} - \sum_{b=1}^{n_B}\sum_{i=1}^{NA}\frac{Z_b e}{r_{bi}} + \sum_{i=1}^{N_A}\sum_{j=1}^{N_B}\frac{e^2}{r_{ij}} + \sum_{a=1}^{n_A}\sum_{b=1}^{n_B}\frac{Z_a Z_b e^2}{R_{ab}} \,, \tag{2}$$

where the indices *a*, *b* correspond to the nuclei, and the indices *i*, *j* correspond to the electrons of molecules *A* and *B*.

When the distance of the two interacting molecules is large enough, the electron exchange can be neglected and the operator *V* can be treated as a perturbation. From perturbation theory, the energy of direct electrostatic interaction can be expressed as

$$E^{(1)} = \left\langle \Psi_n^A \Psi_m^B \middle| V \middle| \Psi_n^A \Psi_m^B \right\rangle, \tag{3}$$

where indices *n* and *m* correspond to the sets of quantum numbers of the isolated molecules.

2.2.1.1 Multipole moment

If the distance between the interacting molecules is greater than the dimensions of the individual molecules, the overlap of electronic densities can be neglected and therefore it is possible to represent the electrostatic energy with good accuracy as an expansion of powers of 1/*R*. This introduces the model of the multipole moments.

The system of charges then creates a potential, which can be described by the following Taylor expansion:

$$\varphi(R) \approx \frac{q}{r} + \frac{\mu}{r^2} + \frac{1}{2}\frac{Q}{r^3} + \ldots, \tag{4}$$

the origin of the Cartesian system lies inside the system of charges.

in the following equations, e_i is the charge *i*, r_i is the position of the ith charge, $x_{i\alpha}$ is the αth Cartesian component of r_i and δ is the Kronecker symbol.

The monopole (total charge) *q* is defined as:

$$q = \sum_i e_i \,. \tag{5}$$

The dipole moment, μ, is defined as:

$$\mu = \sum_i e_i r_i \,. \tag{6}$$

The quadrupole moment, *Q*, is defined as:

$$Q_{\alpha\beta} = \frac{1}{2}\sum_i e_i\left(3x_{i\alpha}x_{i\beta} - r_i^2\delta_{\alpha\beta}\right) . \quad (7)$$

The expansion goes further; other terms are called octopole and hexadecapole moment tensors. As you can see, the earlier terms in the expansion have prevailing effect over the succeeding ones. Therefore, for the system that possesses an overall charge, the first term will have the most influence on the overall potential.

2.2.1.2 Multipole-multipole interaction

The electrostatic energy of two interacting molecules *A* and *B* can be understood as the potential energy of one molecule in an external field of the other:

$$V_{AB} = \sum_{i\in A} e_i\varphi_B(r_i) , \quad (8)$$

where the potential φ_B of the charge system of the molecule *B* operates on the points of the charge system *A*.

The resulting formula for the multipole-multipole interaction depends on the nature of the interacting multipoles. This dictates the dependence of the interaction energy on the distance. For example, in case of two interacting charges, the dependence is $1/R$, in case of dipole-dipole interaction, the dependence is $1/R^3$, in case of dipole-quadrupole interaction, the dependence is $1/R^4$. The electrostatic interaction can be either attractive or repulsive, depending on the signs of monopoles and directions of multipoles.

2.2.2 Polarization interaction

The electronic structure of one molecule adapts to the field of another molecule. This effect is called polarization and is described by the second- and higher-order terms of the perturbation theory. There are two types of polarization interaction – induction interaction, which is the interaction of induced dipole with the electric field of another polar molecule, and dispersion interaction, which is purely quantum correlation effect.

2.2.2.1 Induction interaction

Induction interaction is the result of the adaptation of one molecule to the electric field of the other. In the electric field *E*, a molecule with the polarizability α gaines an induced dipole μ_{ind}, according to the formula:

$$\mu_{ind} = \alpha \cdot E \tag{9}$$

Whilst the molecular polarizability is an isotropic property, α has a form of a tensor.

The induction energy for the molecules in their ground state has always attractive nature. The interaction of induced dipole is similar as the interaction of a permanent dipole. Thus, the approximate induction energy can be expressed as

$$E_{ind} = -\frac{\mu_1^2 \alpha_2}{r^6} - \frac{\mu_2^2 \alpha_1}{r^6} . \tag{10}$$

2.2.2.2 Dispersion interaction

Dispersion is a non-classical effect, arising from the correlation between the movements of electrons. We can imagine it as an interaction between a dipole of one molecule, induced by the instantaneous dipole of the other molecule, arising from the fluctuation of the electronic shell.

The dispersion energy can be expressed as:

$$E_{disp} = -\frac{\alpha_2 \langle \mu_1^2 \rangle}{r^6} - \frac{\alpha_1 \langle \mu_2^2 \rangle}{r^6} \tag{11}$$

where $\langle \mu \rangle$ is the mean induced dipole, compared to the previous case, it is a quantum variable.

Similarly to the induction interaction, dispersion interaction is always attractive. This interaction is always present, even in case of neutral atoms. It is responsible for the fact that rare gases can be liquefied.

2.2.3 Other types of interaction

The other types of interaction are resonance interaction, exchange interaction and magnetic interaction.

Resonance interaction is the interaction between the ground state of one molecule and the excited state of another molecule. The transition energy of both molecules has to be equal. This interaction always happens in case of two identical molecules.

The exchange interaction is the consequence of the Pauli principle – the many electron wave function must be antisymmetric with respect to the electron permutation. Electrons with the like spin therefore cannot share the same space.

Magnetic interaction is the result of the fact that any system of moving charges is characterized by the magnetic multipole moments that exhibit a magnetic field.

3 Modeling of intermolecular interactions

Computational chemistry provides a wide range of methods for modeling intermolecular interactions. These methods differ in cost and accuracy. The hardest task is dealing with dispersion, which is the effect of the correlation of electron motion and demands sophisticated mathematical apparatus for a proper description. This chapter aims to present a brief overview of the methods available and discuss their suitability for the description of the intermolecular interactions.

3.1 Wave function based methods

Wave function based methods rely on solving the time-independent Schrödinger equation for the system:

$$(\hat{T}+\hat{V})\psi = E\psi .$$

The Schrödinger equation of complex systems has to be solved approximatively. There are two approaches for obtaining the approximate solution of Schrödinger equation – variational and perturbational. Variational methods utilize the variational theorem, that states that the energy of the exact ground state E_0 is always lower or equal than energy E of a trial wave function. The examples of variational methods are Hartree-Fock method and configuration interaction. Perturbational methods consider exact Hamiltonian H as a sum of two contributions – hamiltonian of a unperturbed model system with a known solution, H_0, and a small perturbation V. The example of perturbational methods is Moller-Plesset MP2 method.

3.1.1 Hartree-Fock Method

Hartree-Fock (HF) method is the simplest *ab initio* method. The principle of the HF method is the idea of an electron moving in an averaged field caused by other electrons. This method therefore neglects the electron correlation and thus is not able to describe dispersion interaction.

3.1.2 Møller-Plesset Method

Møller-Plesset (MP) method is based on the Rayleigh-Schrödinger perturbation theory, where the exact Hamiltonian $\hat{H}$ is expressed as the sum of Hamiltonian of unperturbed problem $\hat{H}_0$ with known solution, and a small perturbation $\hat{V}$:

$$\hat{H} = \hat{H}_0 + \lambda\hat{V},$$

where λ is an arbitrary parameter. The perturbation in MP theory corresponds to electron correlation.

Since the zeroth order, MP0, energy is the HF energy and the first order does not bring any improvement (first order correction is zero), the most popular of the MP methods is MP2. Higher-order terms are computationally very demanding. The MP4 method gives very good results, but is much more expensive than MP2 method.

MP2 and derived methods are relatively inexpensive and at the same time covering the correlation energy to a great degree. This combination makes them suitable for the description of intermolecular interactions.

3.1.3 Coupled Clusters Method

Coupled Clusters method[2,3] elegantly deals with the higher-order terms of the correlation energy. CCSD(T), using the single and double excitations, augmented by the perturbative triples correction, is probably the most reliable yet feasible method for the description of the weakly-bound systems, where the dispersion plays a major role. It has been shown that the results of the CCSD(T) method are close to the results of the more accurate CCSDT method, which deals with the triples rigorously. As the CCSDT method is very close to the ideal case of the full configuration interaction limit, there is a common agreement that CCSD(T) method can be considered being a reliable benchmark[4].

3.2 DFT methods

Density functional theory provides different approach to modeling chemical systems. Instead of dealing with the wave function itself, it utilizes the electron density. This

frees us from dealing with many-dimensional wave functions and therefore makes the computation much simpler. All the properties of the system are then expressed as a functional of the electron density.

The approach of the density functional theory leads to the Kohn-Sham equations[5]. These equations, similarly to the Hartree-Fock equations, are solved iteratively.

The energy of the system in the DFT is expressed as a functional of the electron density. The total functional form can be expressed as a sum of several terms:

$$E[\rho] = T_{ni}[\rho] + V_{ne}[\rho] + V_{ee}[\rho] + \Delta T[\rho] + \Delta V_{ee}[\rho]. \quad (12)$$

The first term, T_{ni}, is the kinetic energy of the non-interacting system, the second term, V_{ne}, is the interaction energy between electrons and nuclei, the third term, V_{ee}, is the classical electron-electron repulsion. All these terms can be expressed exactly. The last two terms are the problematic part, for which the exact functional form is not known. It is the difference of the exact kinetic energy from the kinetic energy of the model of non-interacting electrons, and the difference of the potential energy due to the electron correlation. These two terms together form so called exchange-correlation energy.

This term is the holy grail of the density functional theory. If the exact functional form for this term was known, we could find an exact solution of the Schrödinger equation. Unfortunately, this is not the case. Various DFT methods differ in the approximation used for this term.

3.2.1 Local functionals

Although DFT itself has, at least in principle, the capability to provide the exact solution of the Schrödinger equation including the long-range dispersion effects, most widely used functionals do not perform very well when trying to describe dispersion. When evaluating the exchange-correlation energy term, these functionals use only the local properties of the electron density. Therefore, the resulting energy is also local and does not give a good description of any long-range

effect. This is well known, as the most commonly used DFT functionals usually underestimate the stability of dispersion complexes.

3.2.1.1 LDA functionals

Local Density Approximation (LDA)[6] uses the model of the homogeneous electron gas for the calculation of the exchange-correlation energy. Exchange and correlation terms are expressed as a function of the electron density.

3.2.1.2 GGA functionals

In the Generalised Gradient Approximation (GGA)[7], the exchange correlation term is dependent not only on the electron density itself, but also on its gradient. It leads to a significant improvement of the performance of the potentials. The most used from this group are PBE[8,9], PW91[9] and BLYP[10] functionals.

In order to still improve the performance, other semi-local information, such as higher-order gradients or density of the kinetic energy, is inserted into the expression for the exchange-correlation term. These functionals are called meta-GGA functionals and the widely used example is the TPSS functional[11].

3.2.1.3 Hybrid functionals

As the Hartree-Fock method is able to describe the exchange energy accurately, hybrid functionals utilize the idea of incorporating a fixed part of this energy to the usual DFT exchange energy. The performance of hybrid functionals is generally better than the pure GGA ones. The most popular hybrid functional is B3LYP[9,10,12]; other examples are PBE0 and PBE1 functionals[9,13].

3.2.2 Non-local functionals

As mentioned above, the conventional DFT functionals are not able to describe dispersion properly. In orde to handle the dispersion correctly in DFT, more complex methods have to be utilized.

Dispersion can be introduced into DFT in several ways. The simplest way involves adding some correction – whether empirical correction[14] or correction to the results

of the *ab initio* methods. More sophisticated ways utilize pseudopotentials[15] or reparametrization[16,17] of the DFT functionals. The most sophisticated approaches generate truly non-local functionals[18].

3.2.2.1 DFT-D

DFT-D[14,19] is a general term used for several generations of methods based on including an empirical correction in the DFT scheme. The form of this correction is an asymptotic $1/R^6$ formula, which best describes the long-range interactions. Because DFT provides a fairly good description of the short-range correlation, it is necessary to include a damping function in the DFT-D scheme. The task of the damping function is to continuously switch off the dispersion correction from the intermediate to the short-range region. The DFT-D methods are very successful in the description of the weakly bound complexes.

3.2.2.2 DFT/CC

DFT/CC[20] is a methodology that utilizes theoretical correction. The coupled clusters method is used as a benchmark that can describe dispersion interaction properly. The costly coupled clusters computations are performed on a simple system and pairwise correction functions are constructed. These correction functions are transferred to larger systems with similar atom types. DFT/CC is a promising method for the description of systems where dispersion interaction is important.

3.3 Empirical force fields

Empirical force fields are based on a simple molecular mechanics. They are utilized to compute the potential energy of a system of particles – atoms and molecules. Every force field consists of the functional form and the set of parameters. They are obtained by fitting the parameters to both the experimental data and the results of accurate quantum chemistry computations.

3.3.1 Functional form

The functional form of a force field basically consists of the bonded and nonbonded term. The bonded term describes the atoms linked by the covalent bonds, while the

nonbonded term examines the long range electrostatic and van der Waals interactions.

$$E = E_{bonded} + E_{nonbonded} \quad (13)$$

The bonded term consists of bond, angle and dihedral terms:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}. \quad (14)$$

The bond and angle terms are usually modeled as harmonic oscillators in the models that do not allow bond breaking. If a more realistic description of a covalent bond is needed, the more extensive Morse potential is usually employed. The functional form of the dihedral term depends on the particular force field. It always includes proper dihedral potentials and additionally improper torsions that ensure the proper planar behavior of aromatic and conjugated systems.

The non-bonded term includes the electrostatic and van der Waals terms:

$$E_{nonbonded} = E_{electrostatic} + E_{vanderWaals}. \quad (15)$$

The electrostatic term is usually computed with the Coulomb's law, while the van der Waals term utilizes the popular Lennard-Jones potential. Some force fields also include terms that account for the electronic polarizability. In most force fields, the interactions are limited to pairwise energies.

3.3.2 Parameterization

Each force field is defined by the set of parameters. All-atom force fields possess the parameters for every atom, while the coarse-grained force fields[21] use cruder description, in order to be used for long-time simulations of proteins.

The typical set of parameters includes the atomic masses, van der Waals radii and partial charges for individual atoms. It also includes equilibrium bond lengths, bond angles, dihedral angles and effective spring constants for pairs, triplets and quadruplets of bonded atoms. Most of the force fields use the fixed charge model, where the partial charge of atom is not influenced by its electrostatic environment.

The next generation of force fields[22] accounts for polarizability, where the partial charges of atoms change according to their environment.

The parameters for a given atom type are usually gained by the experimental and quantum chemistry observation of small organic molecules. Various experimental data, such as enthalpy of vaporization, enthalpy of sublimation, dipole moments or spectroscopy are utilized.

3.3.3 Performance

The force fields are several order of magnitude faster than quantum chemical methods. For extended biomolecular systems and molecular dynamics simulations they often provide the only feasible solution. The results of the force field calculations can be surprisingly reliable, as long as the parameters were fitted to the same type of the system. As Berka et. al. showed, the non-covalent interactions are described quite satisfactorily by the most widely used force fields[23].

Nevertheless, all the force fields are based on numerous approximations. As most of the contemporary force fields do not account with the polarization effect of the environment, they can significantly overestimate the electrostatic interactions of partial charges. Van der Waals forces, which originate from the interaction of induced and instantaneous dipoles, are also strongly dependent on the environment.

3.3.4 Examples

AMBER, CHARMM and GROMOS are force fields utilized primarily for molecular dynamics simulations, although they are also widely used for energy minimization. AMBER (Assisted Model Building and Energy Refinement)[24] is a force field used for proteins and DNA. CHARMM (Chemistry and HARvard Molecular Mechanics)[25] was developed for both macromolecules and small molecules. GROMOS[26] (GROningen MOlecular Simulation package) was originally developed for simulations of aqueous and apolar solutions of proteins. OPLS[27] (Optimized Potentials for Liquid Simulations) also belongs to popular force fields.

4 Intermolecular interactions of proteins

4.1 Protein structure

Proteins are polymeric biomolecules. Their monomeric units are 20 amino acids which are connected into a polypeptide chain in a protein molecule. Amino acids in a polypeptide chain are connected with an amide bond between carboxyl and amino group of each amino acid. Alternating carbonyl groups, amide groups and alpha carbons make up a main chain of a protein molecule; the various amino acid residues that are attached to alpha carbons are called side chains.

It is believed that the specific amino acid sequence, called primary structure, dictates all the structural properties of the protein[28]. Each of the 20 amino acids possesses different affinity to water molecules, ions and other amino acids. Interplay of these forces drives the process of protein folding, where some of the side-chains tend to avoid the contact with water molecules, thus forming the hydrophobic core[29], while the others remain at the surface, interacting with water. During the folding process, the main-chain groups of the protein form hydrogen bonds[30-32] and the protein chain folds into the elements of secondary structure – α-helices and β-sheets. The final shape of the protein, its tertiary structure, is thus influenced by the interaction within its main-chain, interaction between the main-chain and side-chains, side-chain side-chain interaction and interactions of the water molecules.

4.2 Types of amino acids

Amino acid side chains vary in the physico-chemical properties. They have different size, flexibility and polarity. Amino acids can be divided into groups according to their properties. There is a number of ways in which amino acids can be classified, depending on the properties of interest. In this thesis, a simple division into charged, polar and non-polar amino acids has been chosen.

4.2.1 Charged amino acids

Charged amino acids contain a charged group in their side chain. This group can be negatively charged carboxyl group, as in aspartate (D) and glutamate (E), or positively charged group – guanidine group in case of arginine (R) and amino group in case of lysin (K). Charged amino acids generate salt bridges and can also participate in hydrogen bonds. In acidic solutions, the amino acid histidine (H) is in its protonated state, also carrying a positive charge. Therefore, in some cases, it can also qualify for this group.

4.2.2 Polar amino acids

Amino acids, whose side chains contain any polar group, are called polar. It can be hydroxyl group, as in case of serine (S), threonine (T) and tyrosine (Y) or amide group, as in case of asparagine (N) and glutamine (Q). Amino acid histidin (H), that contains imidazole ring, can be also understood as a part of this family. Polar side chains can form hydrogen bonds with other electronegative groups and also with solvent molecules.

4.2.3 Hydrophobic amino acids

Amino acids, whose side chains are not polar, are called hydrophobic. This group consists of glycine (G), which does not possess any side chain, proline (P) with cyclic side chain. These two amino acids specifically influence the structure of the main chain and therefore and are mainly found in bends and hinges, or in the specific secondary structures as in collagen. Other amino acids from this group have aliphatic hydrocarbon chain – alanine (A), valine (V), leucine (L) and isoleucine (I). Yet other amino acids posses aromatic side chain – phenylalanine (F) and tryptophane (W) (polar amino acid tyrosine can also posses an aromatic side chain). Aromatic amino acids can enter specific interactions utilizing their π-electrons. Amino acids containing sulphur – cystein (C) and methionine (M) can be also considered as being a part of this group. These amino acids, and especially cystein, can create covalent bonds, called disulfide bridges.

Hydrophobic amino acids tend to congregate together in their endeavor to minimize their contact with water[29,33]. This process is governed by the increase of the entropy of the hydration shell, as the water molecules around the hydrophobic residues loose their degrees of freedom. It was shown that the water near the large hydrophobic surfaces is more mobile than the bulk water[34]. This process is called hydrophobic effect[35] and drives the protein folding and also the protein-protein interaction.

4.2.4 Occurrence of the particular amino acids in the protein

The occurrence of each of the 20 standard amino acids differs. Figure 2 shows the typical amino acid composition averaged through all the species. The profile of amino acid composition will differ for the particular species.

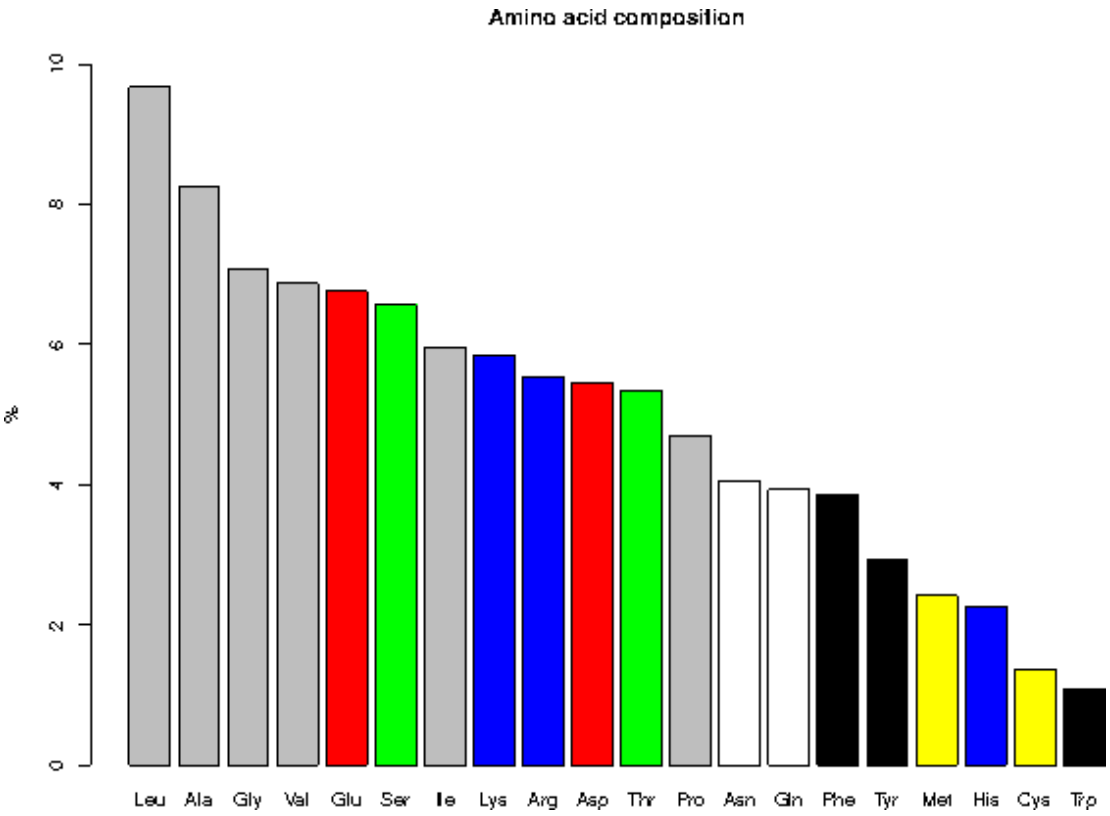

Fig. 2 – Average amino acid composition of proteins (source: UniProt)

Each of the groups of the amino acids tends to occupy different places in the protein structure. Charged and polar amino acids tend to be exposed to the water and therefore occur at the surface, while hydrophobic amino acids tend to avoid the water and therefore occur mostly inside the protein core.

4.3 Protein and intermolecular interactions

Proteins participate in numerous interactions. They interact with solvent molecules, ions[36], ligands or enzyme substrates[37]. They also form complexes with other proteins and DNA[38] and RNA[39,40] molecules. All these interactions are governed by the interaction of particular amino acids. Moreover, the amino acids themselves interact with each other. They can form salt bridges, hydrogen bonds or dispersive interactions in an isolated protein molecule. The intention of this chapter is to briefly describe these particular types of interactions.

4.3.1 Salt bridges

Salt bridge is an electrostatic interaction between two oppositely charged amino acid residues – the positively charged residues (aspartate and glutamate) and the negatively charged residues (lysine, arginine or histidine). The distance of the charged atoms is lower than 4 Å and the interaction energy in the gas phase can reach 600 kJ/mol. The interaction energy of the salt bridge is comparable with the energy of a covalent bond; nevertheless, salt bridges can be easily distracted by the presence of water. Their importance for protein stability is thus highly dependent on their location within a protein structure.

4.3.2 Hydrogen bonds

Hydrogen bond is a bond in which two electronegative atoms share one hydrogen atom. One of the atoms acts as a donor of hydrogen and its bond with hydrogen weakens and lengthens. In proteins, hydrogen bonds can be formed by main-chain carbonyl and amide groups, and by oxygen and nitrogen atoms of side-chains of charged and polar amino acids. Hydrogen bonds between main chain atoms stabilize the secondary structures of a protein, but also side chain atoms can be a part of a hydrogen bond. Protein atoms also form the hydrogen bonds with water molecules. The interaction energy of a hydrogen bond ranges between 5 and 30 kJ/mol.

4.3.3 Dispersion interactions

Dispersion interactions are always present in proteins. Although they are not so strong, they are numerous. In case of the non-polar residues, this kind of interaction

is the only interaction possible. The energy of the dispersion interaction is the most important part of the overall interaction energy of aromatic-aromatic, aromatic-aliphatic and aliphatic-aliphatic side-chains. The energy of the dispersion interaction is proportional to the size of the interacting side-chains and is generally smaller than other types of interactions, ranging typically between 1 and 20 kJ/mol.

5 Aims of the thesis

This thesis explores the role of intermolecular interactions in proteins from two different points of view – quantum chemistry and bioinformatics.

1. The DFT/CC method[20,41] is employed for the modeling the interaction of protein functional groups (represented by the set of small molecules) with a hydrophobic graphitic surface[42], which serves as a model for the hydrophobic surface. The reliability of the method is discussed, comparing its results with other theoretical and experimental approaches. The data could serve as a benchmark for the intermolecular interactions in proteins.

In the following part of the thesis, bioinformatic and computational chemistry tools are used for the analysis of the protein-protein interactions based on crystallographic data from the protein data bank. The data should be represented in a way that enabled a better understanding of the protein-protein interaction and of the hydration structure of proteins.

2. The theme of the second study is the analysis of the set of protein dimers. The aim is to localize and characterize the protein interfaces. The main focus is to explore the role of amino acid side-chains for the interaction, as they are expected to determine the selectivity of the whole process. Do the residues at the interface differ from the non-interacting surface? Are there similarities their intramolecular and intermolecular pairing tendencies? What about the energy content of the interaction? And the final question – is it possible to utilize these findings for the algorithmization of the interface prediction?

3. The third study deals with the hydration structure of the protein. T4 lysozyme has been chosen as a case study protein because of its multiple structures in the protein databank. The question is – is it possible to superimpose all the relevant structure and localize distinct spots with high water occupancy? Do these spots prefer definite parts of the protein structure? How are these spots interrelated? What can we tell

about the overall topology of the hydration structure? And finally – can we utilize the proposed methodology for other proteins?

The next chapters present a detailed introduction into the field of each study and describe the computational strategy utilized.

5.1 DFT/CC as a benchmark method for interactions in proteins

For the quality description of the protein behavior, an accurate description of the intermolecular interaction is needed. As the post Hartree-Fock methods are computationally prohibitive for the complex molecular systems and the DFT method utilizing LDA- and GGA-based functionals do not handle the dispersion well, the possibilities of the DFT/CC method were investigated in this thesis. DFT/CC has proved to be a quality correction scheme for the DFT[20,41]. It strives to correct the DFT calculations to the coupled clusters accuracy.

The physical adsorption of various molecules on a graphite surface was studied by means of DFT/CC and the results were compared with experimental data and also with other theoretical approaches. The graphitic surfaces serves as a model of the hydrophobic surface and the spectrum of model molecules (C_2H_2, C_2H_4, C_2H_6, C_6H_6, CH_4, H_2, H_2O, N_2, NH_3, CO, CO_2, Ar) covers the main functional groups in proteins.

5.2 Protein-protein interactions

Protein molecules interact together and form higher molecular complexes of varying stability and duration. Protein-protein interactions (PPI) are important during many transport, catalytic, regulation and signaling processes. All of these processes require high selectivity. Great theoretical and experimental effort has been devoted to understanding the PPI.[43-52] A very important task in exploring the PPI is the prediction of interaction interfaces[43-46,49,51,52]. Some of the methods for interface prediction use empirical scoring functions, others utilize energy data.

There are two essential features of the PPI – affinity and selectivity. Affinity describes the stability of the protein complex and can be characterized by the free

energy of dissociation, ΔG_d. Specificity, however, is much more difficult for the thermodynamic description. It corresponds to the fact that two particular proteins (and not others) can prefer one specific interaction arrangement (and not other).

Because proteins always occur in the water environment, the dissociation energy of the protein complex consist not only of the of the protein-protein dissociation, but includes also the reorganization of the solvent layer and therefore changes in the protein-solvent and solvent-solvent interaction energies. The enthalpy of a protein-protein interaction is considered to be the main stabilizing contribution forming a protein complex and can overpower the possible destabilizing entropic effects[48,52]. However, the overall change of the entropy can be also favorable for the complex formation, in case there are hydrophobic patches at the interaction interfaces. When hydrophobic residues on both interacting protein surfaces collapse together to produce hydrophobic patches, the organized water molecules are excluded from the interface, which is accompanied by an increase of entropy[52-54]. This effect is in principle the same as the hydrophobic driving forces of protein folding[55].

From the structural point of view, molecular recognition is enabled by the shape and chemical complementarity[48,54,56]. Amino-acid side chains at the binding interface form preferentially stabilizing interactions of an electrostatic nature. It becomes evident that particular amino-acid side-chains play different roles during the formation of a protein complex[52,57]. The study of protein-protein interactions utilizes statistical tools for the determination of the key residues and their pairing preference. The interaction energy is investigated on the level of empirical force field amber ff03.

5.3 Hydration structure of proteins

Water is the natural environment for most of the protein molecules. It is therefore essential to understand the structure and function of the proteins in the context of their ubiquitous interactions with water molecules[58-60]. Water evidently affects the majority of biological processes – it enables the folding process[61] or stabilizes the active conformation of enzymes[62,63]. Water molecules also interfere with the protein-protein interaction process[61,64,65] and mediate the interaction between the

protein and the ligand[66-68]. Understanding the structure and dynamics of protein hydration is therefore of a great importance in order to explain biological phenomena.

Water exhibits its characteristic physico-chemical properties and dynamics due to the formation of hydrogen-bonds[69,70]. The polar groups of the protein also have a strong tendency to participate in hydrogen bonding. The water molecules therefore bind to the protein in the whole range of stability and time persistence. Due to this fact, water and protein mutually influence each other. The water layer covering the protein surface has clearly different properties from the bulk water[60,71]. Reversely, the modes of the motion of this hydration shell can trigger large-scale motions of the protein or they have a great influence on the protein dynamics.[72,73].

The hydration structure of proteins has been studied via numerous experimental[74-83] and theoretical[68,84-90] approaches. The x-ray crystallography reveals the hydration water molecules[75], however, if the studies are carried out at the ambient temperature, they reveal only the most stable water molecules, called buried waters[86]. These waters contribute significantly to the protein stabilization and form an integral part of it[86,91-93]. The positions of these hydration sites tend to be conserved to a great degree in similar structures[63,75,77,94]. It has been shown that these buried water molecules occupy preferably those protein main-chain polar groups that are not part of any secondary structure[86].

Cryogenic x-ray crystal structure analysis revealed the large-scale network of hydrogen bonds[95-97], that are also visible in molecular dynamics simulations trajectories[98-100]. These hydration networks link secondary structures of the protein. Under normal conditions, this hydrogen bond network supposedly undergoes reorganization, accompanying the motion of the protein[101]. As a molecular dynamics simulation revealed, a lot of coherent patterns, such as fair currents, vortices and divergent flows can be observed in the first solvation layer of the protein. This collective nondiffusive behavior happens on the time scales shorter than 10 ps and a length scales smaller than 12 Å[89]. In contrast to the cryogenic x-ray measurements,

molecular dynamics simulations provide also the temporal development of the hydration structure of the protein[98-100].

The study of hydration structure uses the multiple crystallographic data of the lysozyme protein to reveal water clusters – hot spots in the hydration shell. Statistical and topological tools are used for the characterization of these clusters and their overall structure.

6 Methods

6.1 DFT/CC

6.1.1 DFT/CC methodology

DFT/CC correction scheme is based on two assumptions:

(i) the DFT error can be represented in a pairwise (atom-atom) manner

(ii) transferability – the corrections can be transferred from the model system to the target system if the chemical nature of the systems is similar enough

The DFT error, ΔE, is defined as the difference between the CCSD(T) and DFT energies and can be expressed as a sum of correction functions for each atom pair:

$$\Delta E = E_{CCSD(T)} - E_{DFT} = \sum_{ij} \varepsilon_{ij}\left(R_{ij}\right) \quad . \tag{16}$$

In order to obtain correction functions, a one-dimensional scan of the potential energy surface for the reference set of the molecules (the adsorbate molecules Ar, H_2, CH_4, C_2H_6, C_2H_4, C_6H_6, C_2H_2, CO, CO_2, H_2O, N_2, NH_3 in the interaction with H_2 and benzene) has to be performed on the DFT and CCSD(T) level. The interaction energies are calculated as a difference between the energy of the complex and energies of the individual molecules. The correction functions are the result of applying Reciprocal Power Reproducing Kernel Hilbert Space Interpolation (RP-RKHS).

Ab initio calculation were performed with the augmented Dunning's correlation-consistent valence-X-ζ basis sets with polarization function (X = D, T, Q, 5 for the double, triple, quadruple and pentuple) – AVDZ, AVTZ, AVQZ and AV5Z. The results of these computations were used for the CBS estimate.

For the DFT interaction energies of the reference system, the AVQZ basis set was used. It has been shown that this basis set is consistent with the saturated plane-

wave basis set used in periodic DFT calculations. In all DFT calculations, the Perdew-Burke-Ernzerhof (PBE) exchange correlation functional has been used. The plane-wave basis set with an energy cutoff of 800 eV and the PAW pseudopotential with ENMAX = 700 eV were used. Γ-point calculations of the interaction energies shown themselves to be accurate enough.

The *ab initio* and DFT cluster calculations were performed with Molpro2010 and Gaussian09 program suites. The periodic DFT calculations were carried out with the Viena *ab initio* simulation package (VASP).

For the details of the methodology, see the references[20,41].

6.1.2 Model systems

The studied model system consisted of twelve small molecules (Ar, H_2, CH_4, C_2H_6, C_2H_4, C_6H_6, C_2H_2, CO, CO_2, H_2O, N_2, NH_3) adsorbed on the graphitic surface. The spectrum of the model molecules covers the most important functional groups in proteins. As there are accurate experimental data available for the argon ⋯ graphite system, argon serves as a reliability check here. The graphitic surface models at some extent a hydrophobic surface.

6.2 Protein-protein interactions

6.2.1 Model set

Protein-protein interactions were studied on the defined set of protein-protein complexes. The structures of these complexes were obtained from the PDB database from March 16, 2010. Selection was narrowed to the x-ray structures of proteins with a resolution better than 1.7 Å. The number of oligomers and the number of entities was set to 2. The sequence length of individual interacting proteins ranged between 50 and 500 amino acids. Structures with sequence identity higher than 50% were removed. The resulting set included 69 protein complexes.

6.2.2 Definition of the protein compartments

6.2.2.1 Surface and interior of the protein

The solvent-accessible surface (SAS) area of all the amino acid was computed by the Protein Dossier module of the STING Millennium web application[102]. Amino acids with a relative solvent-accessible surface area greater than 10% were considered as surface amino acids.

6.2.2.2 Protein interface localization

In order to map contacts between particular amino acids of both proteins, a contact matrix was constructed. Contact matrix[103] is the matrix that for every two amino acids of the proteins gives the shortest distance between two heavy atoms of the amino-acids side-chains. In this analysis, the distance was expressed relatively, in terms of the multiple of van der Waals radii of both interacting atoms.

Two amino acids were defined to be in contact if the distance between any pair of their side-chain atoms (the backbone atoms were excluded) was below the threshold of 1.25 multiple of van der Waals radii of these atoms. This threshold was found to be a good compromise between the extensive omission and large number of amino acids that further formed interfaces.

An amino acid was defined to be an interface amino acid when it has at least one contact with any amino acid of the second protein partner. All the interface amino acids then created the interface of this particular protein dimmer.

6.2.3 Processing of the protein interfaces

Protein-protein interfaces were stored as xyz coordinates of the component amino acids in a pdb file. Hydrogen atoms were added to the interfaces, using the Gromacs procedure pdb2gmx. The positions of the hydrogen atoms were optimized using the ff03 potential[104] with the weight of the heavy atoms being strongly restrained by the constant 50000000 to avoid their movement during the hydrogen atom optimization.

6.2.4 χ^2 test

For the evaluation of the statistical significance of our results, we used the standard χ^2 test:

$$\chi^2 = \frac{(n - n_{teor})^2}{n_{teor}} + \frac{(n^* - n^*_{teor})^2}{n^*_{teor}}, \quad (17)$$

where n is the actual number of positive results while n_{teor} is the theoretical number of positive results. The values with asterisk correspond to negative results. If the value of χ^2 exceeds 3.841, it corresponds to a probability of 0.05 that the result belongs to the theoretical distribution; if the value exceeds 10.828, the corresponding probability is 0.001. The higher the value of χ^2, the more statistically significant the result.

The χ^2 test was used to evaluate the statistical significance of the amino-acid frequencies' differences between distinct sets of amino acids.

6.2.5 Pair statistics

In order to evaluate the significance of the tendency that two particular amino acids will create a pair, we expressed the difference

$$\Delta p = p_{actual} - p_{theory}, \quad (18)$$

where p_{actual} is the actual frequency of amino-acid pairs (related to the total number of pairs) and p_{theory} is the theoretical frequency of amino-acid pairs, resulting from independent frequencies of amino acids 1 and 2,

$$p_{theory} = p_1 p_2 . \quad (19)$$

The difference $_{\Delta p}$ corresponds to the pairing tendency of these two amino acids.

6.2.6 Interaction energies

Interaction energies between amino acid side-chains were evaluated using an empirical force field. The interaction energies consisted of Coulomb and Lennard-Jones terms.

The electrostatic (Coulomb) energy was evaluated using the formula

$$E_{COUL} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}, \quad (20)$$

where indexes *i* and *j* correspond to the side-chain atoms of the first and second amino acids, q_i is the partial charge on the amino-acid atom, ε_0 is the permittivity of the vacuum, ε_r is the relative permittivity and r_{ij} is the distance between the two atoms.

The Lennard-Jones energy was computed according to the formula below:

$$E_{LJ} = \sum_{i,j} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right), \quad (21)$$

where $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ is the geometric mean of the ε constants, which correspond to the depth of the potential well on the Lennard-Jones curve, while $\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$ is the arithmetic mean of the σ constants, which correspond to the finite equilibrium distance of two atoms, where the potential energy equals zero. Parameters *q*, ε and σ for the amino-acid atoms were taken from the ff03 force field. All the energies were calculated in the gas phase (ε_r = 1). The interaction energies were calculated using a homemade MATLAB script (MATLAB version 7.4.0., 2007).

Amino acids were represented in a C_α representation, i.e. no backbone atom except the C_α. C_α was retained in its original crystallographic position and hydrogens were added along the direction of the removed backbone atoms[105]. The Lennard-Jones parameters of the added hydrogens are the same as for the original C_α hydrogen, and their partial charges are set to preserve the overall charge of the amino acid (zero for most amino acids, +1 for arginine, lysine and protonated histidine, -1 for glutamate and aspartate).

The total interaction energy of the amino-acid pair is constructed as the sum of the Lennard-Jones and the Coulomb contributions. The corresponding data were stored

in an interaction energy matrix (IEM), defined according to Bendova et al.[106]. It is a matrix of interaction energies for each pair of residues.

6.2.7 Residue interaction energies (RIE) and their distribution functions

The residue interaction energy (RIE) is defined as the sum of the pairwise energy contributions of a particular residue in contact with all the other protein or complex residues. Technically, it is the sum of one row or column in the IEM.

In order to evaluate the distribution of the interaction energies, we used a cumulative distribution function. The distribution function describes the probability that the values of energy are lower than or equal to a given value of energy. The distribution function allows us to see how the interaction energies are distributed and to compare different distributions in different sets.

6.3 Hydration structure

6.3.1 Lysozyme as a case study model

One of the most populated structures in the PDB database is lysozyme. Therefore it was chosen as a case study protein for the analysis of the hydration shell. PDB database was searched for the following criteria: lysozyme, x-ray structures, one-chain protein, x-ray resolution < 2.5 Å. This query yielded 837 structures. The aim of the following procedure was to detach a set of corresponding structures in terms of symmetry and sequence, perform the superposition algorithm on these structures and represent the positions of all the water molecules present in all the structures with a 3D grid, suitable for the further analysis.

The whole set was divided into subsets different in a symmetry group. The most populated symmetry P 32 2 1, with 427 representatives, was chosen for the further study. Next, the sequence length for the structures within this group was determined. Structures with the most populated sequence lengths and its minimum variation were chosen. Thus, only the structures with a sequence length of 162, 163 and 164 amino acids qualified for the final set. In the group with the shortest

sequence length, the consensus sequence was determined and aligned with all the other sequences. The final set consisted of 391 structures of T4 phage lysozyme.

In the next step, the structures were superimposed utilizing the Kabsch superposition algorithm[107] as the reference method. It was implemented on the backbone atoms corresponding to the consensus sequence. This structure superposition yielded in total 56,386 superimposed water molecules coming from 391 structures of the T4 phage lysozyme.

6.3.2 Water density grid

In order to represent the hydration structure of the lysozyme in a more suitable form, a 3D grid (centered around the common center of mass of the protein molecules) was constructed with the density of the nodes being 1 grid point every 0.2 Å. The extremes of the 3D grid were determined up to the point where the number of the eliminated waters exceeded 0.5 ‰ of the total number of waters.

Every grid point was assigned with a certain number corresponding to the number of waters in a 1-Å cube with its center located at that grid point. This value defines water density at the grid point (*wdgp*) function. Thanks to this approach, the water density could be represented as a three-dimensional numeric scalar function.

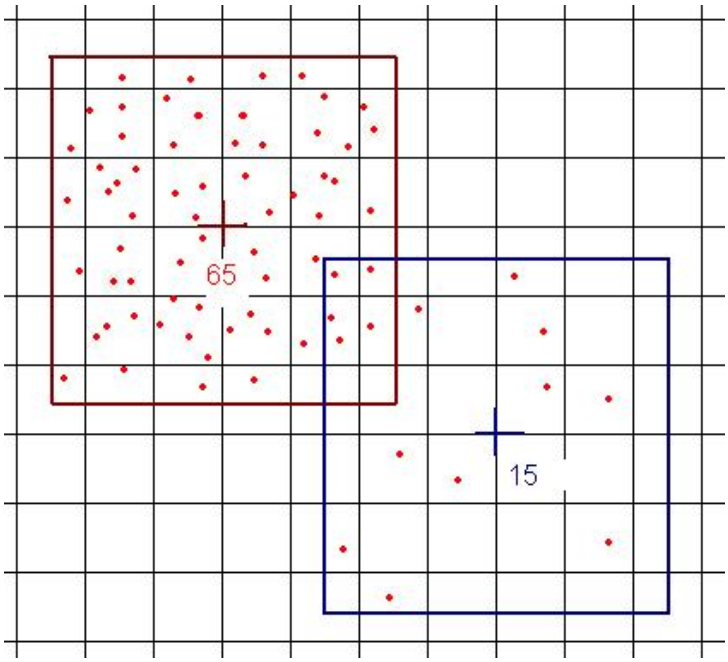

Fig. 3 – An illustration of the definition of the water density grid point

6.3.3 Water clusters

The water density grid was subjected to a simple clustering algorithm. As the first step, a threshold for the water density function was set and only the *wdgp*s exceeding this threshold value were classified for the following clustering algorithm. In this algorithm, two *wdgp*s belonged to the same cluster if they shared at least the body diagonal of the grid cube. The number and size of the acquired cluster is dependent on the threshold, which was expressed in percent of the maximum possible number of waters, given by the total number of structures in the set.

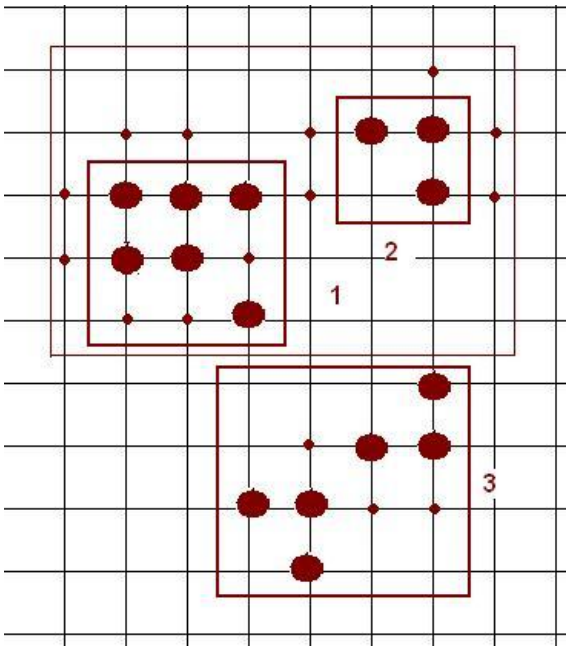

Fig. 4 – An illustration of the clustering algorithm. The big spots symbolize the *wdgp*s satisfying the higher threshold, the small spots symbolize the *wdgp* satisfying only the lower threshold. For the lower threshold, clusters 1 and 2 merge.

Each cluster has a *wdgp* with the highest value. This value defines the occupancy of the cluster, and the coordinates of this *wdgp* represent the position of the cluster. The volume of the cluster can be easily calculated as the number of *wdgp* of this cluster times 0.008, which is a volume corresponding to one *wdgp*. An approximate diameter of the cluster is defined as the third root of this volume. As the 10% threshold ensures that the clusters do not merge yet, while preserving most of the information, this threshold has been chosen for the most of the further analysis.

6.3.3.1 Position of the water clusters

The relative positions of the clusters and the protein were examined. Each protein structure was searched for the cluster-representative water molecules. A cluster-representative water molecule is the water molecule closest to the cluster center, whose distance must not exceed the approximate diameter of the cluster.

Consequently, each of the protein structure was analyzed to determine heavy atoms closer than 3.1 Å to the representative water – the upper limit of the consensus length of the hydrogen bond[108]. The result is a list of the interacting atom types and the list of the structures in which this interaction takes place. For the purpose of the simple statistical analysis, only the atom types which occurred in at least 50 % of the structures containing the water molecule were taken into account.

For each localized water cluster, all the other clusters within the 3.1 Å distance were determined. Each cluster was thus assigned with the number of the protein partners and the number of the water cluster partners. The clusters were further connected into higher organizational groups – superclusters – based on the criteria that two clusters are part of the same supercluster if their distance is less than 3.1 Å.

6.3.3.2 Solvent-accessible surface area

The solvent-accessible surface area is determined as a part of the overall van der Waals surface of a molecule which is in contact with the solvent approximated by a sphere of a certain diameter. This value can be calculated for residues or even for particular atoms. The solvent-accessible surface area per atom was computed utilising the web application GETAREA[109]; the radius of the water probe was set to 1.4 Å.

6.3.3.3 Interaction energy of the water-cluster representatives

The cluster-representative water molecules were identified as follows. The structures with a resolution better than 1.5 Å and with no missing atoms were chosen as a representative set for interaction energy evaluation. Hydrogen atoms were added using the amber tool LEaP and their positions were optimised in AMBER10 using amber99sb force field parameterisation[104]. The structures of 224

water molecule representatives and their closest amino-acid neighbours from the protein were selected for their interaction energy calculations. The pairwise interaction energies between the water molecule and a particular amino acid were determined as the single point values of the experimentally determined positions of an amino acid with optimised hydrogen atoms and the water molecules.

7 Results

7.1 DFT/CC benchmark energies

7.1.1 Coronene ··· A complexes

Coronene is a frequently used model system for the graphitic surface. For this system, both the density-functional-based and wave-function-based methods are available. It therefore allows the comparison of DFT/CC results with the MP2/AVTZ interaction energies.

The structures of all the complexes were optimized using the MP2/AVDZ level of theory. The global minimum structures are presented in the figure 5.

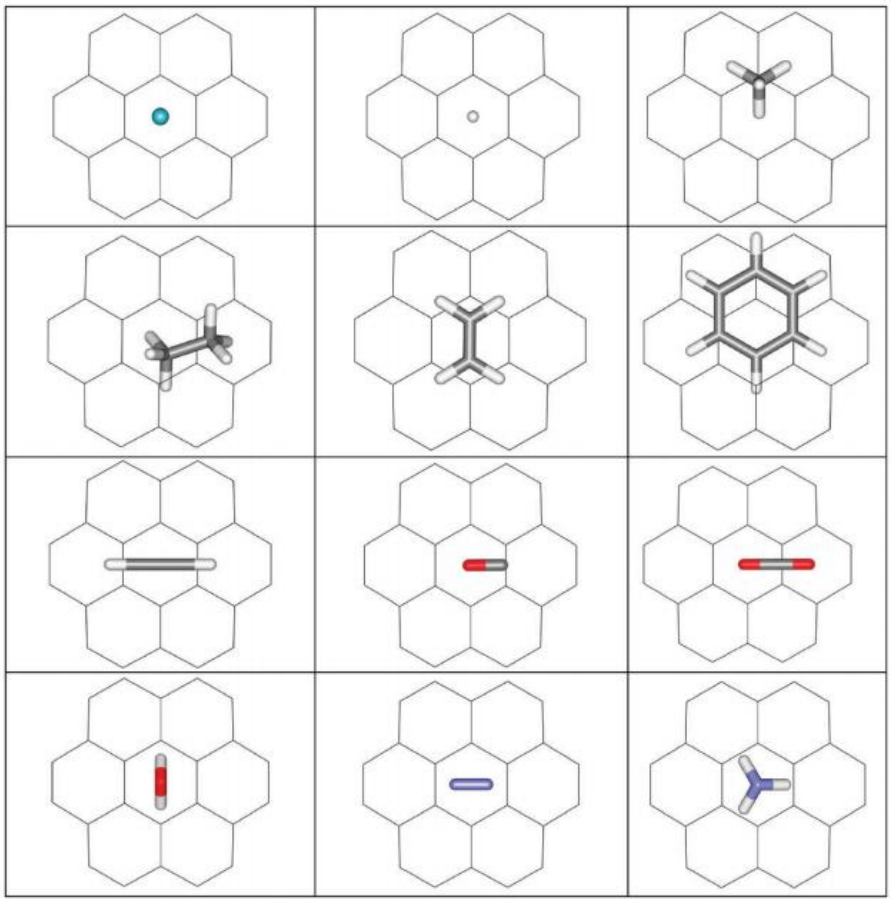

Fig. 5– The global minimum structures of coronene ··· A complexes (A = Ar, H_2, CH_4, C_2H_6, C_2H_4, C_6H_6, C_2H_2, CO, CO_2, H_2O, N_2, NH_3) calculated at the MP2/AVDZ level.

Table 1. MP2/AVTZ and DFT/CC/AVQZ interaction energies E_{int} and equilibrium distances R_e of coronene ⋯ A complexes

Method	MP2/AVTZ		DFT/CC (AVQZ)		PBE (AVQZ)
System	R_e (Å)	E_{int} (kJ/mol)	R_e (Å)	E_{int} (kJ/mol)	E_{int} (kJ/mol)
Ar	3.27	-11.24	3.31	-9.31	1.08
H_2	3.03	-6.05	3.07	-4.98	-0.28
N_2	3.12	-15.41	3.22	-9.40	1.49
CH_4	3.24	-15.69	3.31	-11.71	2.01
C_2H_6	3.40	-24.44	3.50	-17.49	2.30
C_2H_4	3.13	-26.19	3.25	-17.36	3.61
C_2H_2	3.15	-23.11	3.28	-14.67	2.73
C_6H_6	3.19	-60.90	3.37	-32.82	9.97
CO	3.17	-14.17	3.23	-10.73	0.85
CO_2	3.09	-20.79	3.14	-16.47	2.38
H_2O	3.20	-15.87	3.20	-14.87	-4.32
NH_3	3.30	-15.14	3.31	-13.96	-0.92

The interaction energies and equilibrium distances (i.e. the distance of the molecular center of mass from the coronene plane) were obtained with the one-dimensional scan with the MP2/AVTZ and DFT/CC methods. As the table 1 shows, the MP2 interaction energies are in all the studied cases larger than the DFT/CC ones. The discrepancy of both methods is smaller (7-8 %) for the molecules possessing a large dipole moment (water and ammonia). For the other molecules, the difference is higher, with the coronene ⋯ benzene complex having the largest difference of 28 kJ/mol. This difference can be compared with the sandwich structure of the benzene dimer, where the interaction energy predicted by the MP2 method is higher than the reliable CCSD(T) benchmark calculation[41] by a factor of 1.84. It is very close to the ratio 1.86 between the results of the MP2 and the DFT/CC methods for the coronene ⋯ benzene system.

The tendency of the MP2 method to overestimate the interaction energy of weak intermolecular complexes is now well documented[41,110]. The MP2 results should therefore be understood as an upper limit for the interaction energy and not considered as a benchmark for the verification of the DFT/CC results.

7.1.2 Adsorption of single molecules on graphene

DFT/CC optimization of adsorption complexes on graphene was performed with the graphene PBE equilibrium geometry kept frozen. The MP2/AVDZ global minima shown in the figure 5 were used as a starting point for the optimization. The resulting structures are basically the same as in the case of coronene ··· A complexes, especially for the structures with the adsorbed molecule located above the center of the six-member ring. The minor lateral shifts were observed for methane, benzene and carbon dioxide molecules; it reflects the higher symmetry of the physisorbed complexes on the graphene.

The interaction energies for the graphene ··· A complexes are shown in the table 2. For the physisorbed complexes at the graphene surface, the dispersion contribution is generally higher than for the complexes with coronene. It can be attributed to the increased number of interacting carbon atoms. The exceptions are water and ammonia. They can be explained by the electrostatic effects, because these molecules are the only molecules that possess large dipole moment. The interaction of water with aromatic molecules was studied by the density-fitting density-functional-theory symmetry-adapted perturbation-theory (DF-DFT-SAPT)[111]. It was shown that the electrostatic contribution to the interaction energy drops in the row of benzene, coronene and dibenzocoronene. Extrapolation to the graphene plane explains the lower interaction energy for the physisorbed water compared to the complex with coronene.

Because the evaluation of the properties of molecules adsorbed on the extended substrate is beyond the applicability of reliable wave-function-based methods, the results can be compared only with the density-functional-based methods or with empirical potentials. The interaction energy of graphene ··· benzene adsorption complex was calculated with the vdW-DF method[112] (-47.8 kJ/mol). It is somewhat larger than the DFT/CC result (-43.1 kJ/mol). However, the vdW-DF result for the sandwich structure of benzene dimer is overestimated, compared to the best theoretical estimates[113,114]. It is therefore reasonable to assume that the interaction energy of the graphene ··· benzene complex is also overestimated.

Table 2. Physisorption of the small molecules on a graphene surface – confrontation of the DFT/CC results with experimental values. R_e – equilibrium distance, D – potential well depth, $\Delta_{ad}H$ – adsorption enthalpy (corrected for the zero point vibrational energy), Q_1 – heat of adsorption.

	DFT/CC			Experiment		
System	**R_e (Å)**	**D (kJ/mol)**	**$\Delta_{ad}H$ (kJ/mol)**	**R_e (Å)**	**D (kJ/mol)**	**Q_1 (kJ/mol)**
Ar	3.29	9.9	-9.6	3.2 ± 0.1[115]	9.6 ± 0.4[116]	
H_2	3.06	5.4	-4.4	2.87[116]	5.0 ± 0.05[117]	
N_2	3.23	10.9	-10.6	3.34[116]	10 ± 0.3[116]	
CH_4	3.31	13.5	-13.0	3.45[116]	12.5 ± 1.0[116]	
C_2H_6	3.44	20.8	-20.3			18[118]
C_2H_4	3.24	20.2	-19.7		18.9[116]	
C_2H_2	3.26	17.1	-16.7		17.2[116]	
C_6H_6	3.30	43.1	-42.6			41[118]
CO	3.23	12.3	-11.8		10.6[116]	
CO_2	3.10	19.1	-18.7	3.2[116]	17.2[116]	
H_2O	3.19	13.5	-13.0			19[118]
NH_3	3.31	13.5	-13.0			19[118]

7.1.3 Comparison with experimental results

Because the accurate theoretical benchmarks are lacking for the molecules adsorbed on extended substrates, the results need to be confronted with the experimental data. Several experimental techniques provide data for the determination of the parameters of the laterally averaged molecule-surface potential energy functions. The most accurate experimental data come from the bound-state resonances in the elastic scattering of light atoms and molecule from a substrate surface gained by selective adsorption measurements. The more abundant source of data are thermodynamic experiments, such as measurement of isosteric heat of adsorption. Another common source of information about adsorption potentials are the experiments measuring the desorption rate. It is important to add that especially the last two experimental techniques are very susceptible to the heterogeneity of the substrate surface. The experimental determination of the interaction energies of single molecules demand a low coverage regime, where the molecules bind preferably to the sites with the largest interaction energies and do not interact laterally. The situation is further complicated by the fact that the heats of adsorption

cannot be experimentally determined at 0 K, which is the temperature of the theoretic calculations. The most straightforward way to discuss the reliability of the results is to compare them with the parameters of the laterally averaged potentials obtained from the experimental data.

Table 2 summarizes the DFT/CC results and the experimental data. The experimental heats of adsorption were used in cases where the potentials of physical adsorption were not available. It is important to note that the values of differential heats of adsorption, Q_1, cannot be directly compared with the DFT/CC values. Nevertheless, as the Q_1 values on carbon blacks decrease very little with the increasing temperature, the calculated *D*s are usually very close to the experimentally measured heats of adsorption.

The first four systems – argon, molecular hydrogen, molecular nitrogen and methane – are experimentally well defined. The agreement between the DFT/CC and experimental values is clearly satisfactory. The DFT/CC interaction energy is only slightly overestimated, which can be partly attributed to the fact that the experimental potential is laterally averaged. In case of other hydrocarbons, carbon monooxide and carbon dioxide, the experimental data are rare or unreliable. Although the absolute difference of DFT/CC and experimental results is somewhat larger in this group of molecules, the both data still correlate. Moreover, it is not clear whether the error comes from the experimental or theoretical side. For the water and ammonia, the reliable experimental data basically do not exist. The heats of adsorption reported for these molecules are clearly too high for a single molecule interaction with a graphitic surface. The binding between the physisorbed molecules is actually stronger for the water and ammonia than the binding with the substrate[112,119,120]. The interaction energy of the water-graphite system has been recently computed by the DF-DFT-SAPT and DFT-D methods[111,121]. The results are in a good agreement with the proposed DFT/CC values.

The results of DFT/CC calculation can also be compared with the reliable experimental data from scattering experiments for molecular hydrogen or from measurements of the graphite exfoliation energy. The interaction energy for the

physisorbed hydrogen gained from the scattering experiment[117] (-5 kJ/mol) is in a good agreement with the DFT/CC value[122] (-5.4 kJ/mol). The DFT/CC structural parameters for the graphite (a = 2.46 Å, c = 6.60 Å) are in good agreement with experiment (a = 2.46 Å, c = 6.67 Å) and the calculated exfoliation energy[123] (5.2 kJ/mol per atom) agrees quite well with the most recent experimental value[124] (5.0 kJ/mol).

7.1.4 Discussion

The DFT/CC equilibrium distances and interaction energies are in reasonable agreement with available experimental and theoretical data. The agreement between the DFT/CC and the best experimental estimates (Ar, H_2, N_2, CH_4) is within a few tenths of kJ/mol, and within 1-2 kJ/mol for other systems where the values derived from thermodynamic experiments are less reliable[116]. The correlation between experimental and theoretical data is poor only in case of water and ammonia; this can be however attributed to the fact that it is experimentally demanding to determine the values relevant to the zero-coverage limit because of the strong interaction between the adsorbate molecules[119,120]. The calculated results are thus relevant guess for interaction strength of the functional groups of proteins with the hydrophobic surface.

7.2 Protein-protein interactions

Protein-protein interactions were studied on a set of 69 protein complexes, i.e. 138 monomeric protein units. 52 of the monomers were animal (36 human), 5 plant, 13 fungi, 64 procaryotic and 4 viral proteins. 31 complexes were obligate complex, 38 were non-obligate. Obligate protein complexes (OPC) are complexes of the proteins that cannot exist as individual units, while non-obligate protein complexes (NPC) are proteins that can exist also independently. Most of the obligate protein complexes were enzyme complexes, while the group of non-obligate complexes consisted of transport proteins, structural proteins, signaling proteins and enzyme-inhibitor complexes. All the studied structures were biologically relevant complexes which selectively and meaningfully bind to each other.

Obligate protein complexes are formed on average by 435 amino acids. Approximately ninety-nine amino acids create their interfaces which represent 22.7 % of the average protein complex size. Non-obligate protein complexes on the other hand consist of 286 amino acids on average, of which approximately 36 amino acids belong to the interface that is 12.5 %. Non-obligate interface makes up on average 14.1 % of the total surface (6.9–26.4 %) while obligate interfaces make up on average 24.1 % of the total surface (7.7–43.6 %).

Tables 3 and 4 present the overview of results and summarize the features of obligate and non-obligate proteins and their interfaces. The columns represent the pdb ID, number of amino acids of chains A and B, number of amino acids at interfaces A and B, buried interface area, total interaction energy and Coulomb and Lennard-Jones contributions.

Table 3. Obligate protein complexes

pdb	protein A/B [amino acids]	interface [amino acids]	interface area [A^2]	E int [kJ/mol]	Coulomb contribution [kJ/mol]	Lennard-Jones contribution [kJ/mol]
1AJS	412/412	63/60	7106	-913	-291	-622
1F2T	149/143	33/39	4669	-1160	-709	-450
1FM0	81/149	19/15	1944	-683	-461	-222
1GK9	208/557	123/135	14396	-3743	-1838	-1905
1GM7	207/557	122/131	14123	-3356	-1505	-1852
1GVE	324/324	22/23	2040	-606	-355	-251
1H32	261/137	22/19	2615	731	983	-253
1HFE	396/88	76/53	7680	-4265	-3294	-971
1IRD	141/146	18/18	1894	-328	-119	-209
1LUC	355/320	40/36	4246	-721	-191	-530
1MTP	320/35	61/32	5348	-3077	-2434	-644
1N1J	87/78	43/37	4580	-1724	-1129	-595
1NME	146/92	39/41	4494	-1706	-1175	-531
1PHN	174/174	26/33	3167	-1669	-1298	-371
1Q7L	192/88	66/61	8363	-1032	-111	-922
1UGP	203/226	73/78	8747	-4203	-3190	-1012
1W6N	134/134	9/10	1343	-433	-395	-39
1WUI	534/267	60/72	7429	-3080	-2419	-662
2FOM	54/150	23/37	3393	-3476	-3068	-408
2FP7	40/152	27/50	4324	-1989	-1504	-486
2G2S	64/164	36/46	4664	-1830	-1321	-509
2HP0	447/447	39/39	4151	-1324	-989	-335
2JBA	125/121	11/11	1185	-509	-435	-75
2OPL	181/182	100/102	10805	-5595	-4212	-1382
2QDY	197/211	69/73	7512	-2780	-1907	-873
2QM6	342/186	120/91	12015	-2665	-1293	-1373
2VB7	404/406	54/54	6850	531	557	-26
3BZY	17/83	13/25	2202	-902	-800	-102
3CLS	262/318	66/69	7405	-2956	-2440	-515
3D0Y	89/90	26/28	2803	420	734	-314
3IDB	343/157	28/22	2855	-2165	-1884	-281

Table 4. Non-obligate protein complexes

pdb	protein A/B [amino acids]	interface [amino acids]	interface area [A^2]	E int [kJ/mol]	Coulomb contribution [kJ/mol]	Lennard-Jones contribution [kJ/mol]
1CSE	275/63	22/11	1535	-169	-21	-148
1EUV	221/77	21/19	2404	-2079	-1907	-172
1F60	440/90	32/30	3682	-1359	-1027	-333
1JAT	152/136	11/15	1535	-2584	-2455	-129
1R8S	160/187	26/33	3124	-1741	-1374	-367
1TO2	281/63	22/12	1869	-281	-140	-142
1TX4	196/177	23/19	2428	-2534	-2285	-250
1V5I	275/76	20/17	2059	-337	-190	-147
1WMH	83/82	12/11	1266	-3299	-3228	-71
1WXC	273/83	23/18	2051	-1532	-1318	-214
1XD3	229/75	24/15	2750	-1222	-1013	-209
1Z0J	169/51	15/12	1434	-722	-564	-158
2BCG	442/194	31/19	2975	-1218	-903	-315
2DRK	59/10	12/6	996	-1678	-1542	-136
2FHZ	106/93	24/19	2612	-2373	-2138	-235
2GRR	157/157	11/7	1185	-314	-222	-92
2HQS	415/108	30/21	3356	-659	-398	-261
2OMZ	465/104	25/23	3054	-1468	-1179	-289
2OXG	112/108	20/22	3251	-781	-592	-189
2OZN	133/131	16/16	1685	-285	-115	-170
2UYZ	156/78	10/13	1392	-2662	-2543	-120
2VLQ	84/134	16/16	1561	-2341	-2222	-119
2VN6	151/64	16/15	1570	-84	85	-169
2VPB	57/35	15/11	1327	-158	-17	-141
2ZA4	108/90	14/13	1801	-1261	-1105	-156
2ZFD	183/119	31/27	3305	-439	-104	-336
3BY4	172/75	23/16	2034	-2001	-1870	-131
3CIP	371/128	27/23	2682	685	846	-161
3CJS	59/72	12/12	1804	-496	-330	-166
3D3B	139/87	16/16	1778	-2180	-2103	-76
3F1P	114/111	17/19	1996	-1172	-965	-206
3F6Q	169/72	20/18	1989	-400	-149	-250
3FIL	55/56	8/8	1003	28	124	-96
3H7H	118/95	29/27	3061	-1185	-831	-354
3KF6	139/105	18/18	2057	-1703	-1500	-203
3KNB	98/96	15/13	1417	-734	-600	-134
3KYJ	129/135	12/8	1175	-166	-47	-118
3L51	155/166	12/11	1484	-143	-20	-123

7.2.1 Amino acid composition

The first task was to examine characteristic composition of the interacting interfaces. Amino acid composition was analyzed in the whole set and its distinct subsets – surface, interior and interface. The significance of the proposed results was checked with the standard χ^2 test. Figure 6 shows the typical difference between the protein interior and its surface, where the hydrophobic amino acids tend to occupy the interior of the protein (most abundant are Leu, Val and Ala), while the charged and polar amino acid tend to prevail at the protein surface (with the most frequent being Glu, Lys and Asp). Figure 7 presents the comparison of the interface composition with the average amino acid composition. The most striking is the decrease of the occurrence of Gly and Ala at the interfaces, while the occurrence of aromatic amino acids Phe and Tyr is significantly higher. The profile of charged amino acids is very interesting – Arg shows an important increase of its occurrence at the interfaces, while the other charged amino acids exhibit opposite behavior. Figure 8 compares an average surface of the protein with the composition of the interaction interface. It is clearly visible that protein interfaces are more hydrophobic compared to the typical surface – particularly containing the hydrophobic amino acids Ile, Leu, Phe and Val and polar aromatic amino acid Tyr. Charged amino acids unexpectedly present a significant drop of occurrences at the interface. The only exception is the aforementioned Arg, whose frequency at the interface is similar to that of the average surface.

As follows from this simple statistical analysis, the interfaces have characteristic amino acid composition. This fact enables us to distinguish possible interface regions at the surface. Interfaces are more hydrophobic than the rest of the surface, while it somehow excludes the smallest hydrophobic amino acids Gly and Ala. Branched and aromatic residues, such as Ile, Leu, Val, Phe and Tyr are preferred here. The charged amino acids Lys, Asp and Glu show significantly lower occurrence at the interface, whereas Arg is the only charged amino acid to be distributed similarly at the interacting and non-interacting surface.

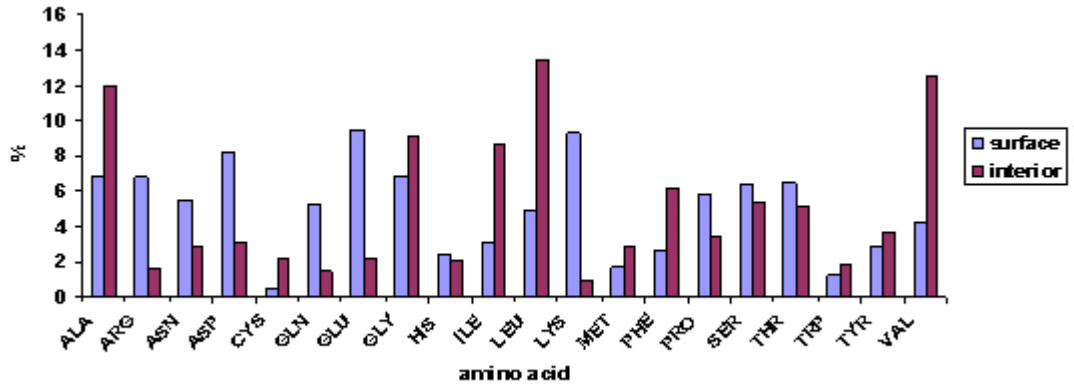

Fig. 6 – The chemical composition of the protein surface and interior. The columns are measures of the populations of amino acids at the surface (blue) and in the interior (red)

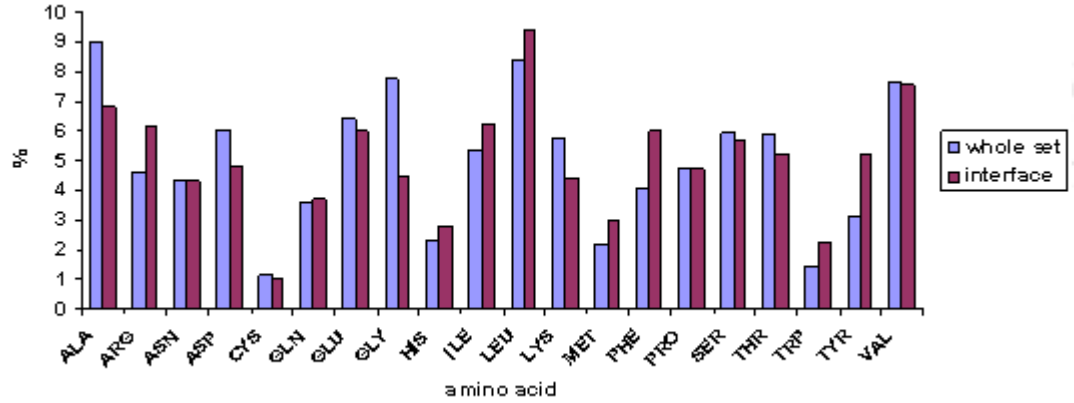

Fig. 7– The chemical composition of the proteins and protein interfaces. The columns are measures of the populations of amino acids in the whole set (blue) and at the interfaces (red)

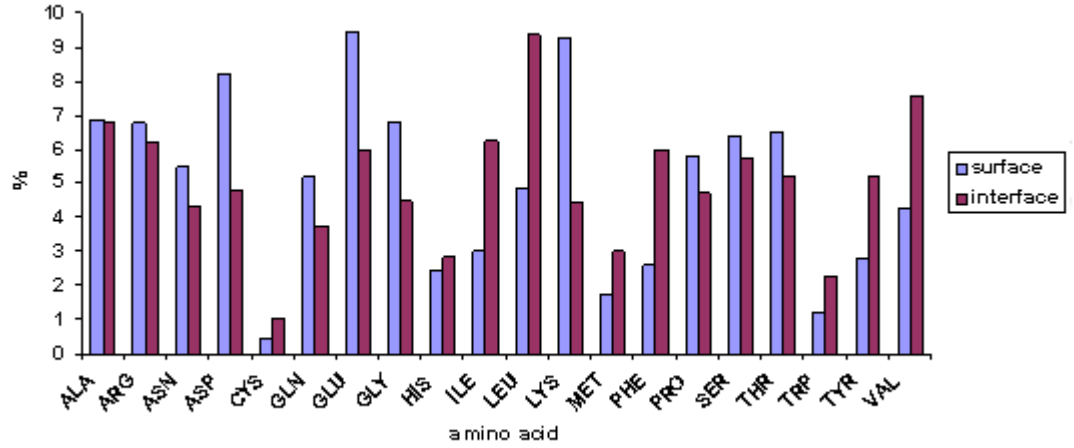

Fig. 8 – The chemical composition of the protein surface and interface. The columns are the measures of the populations of amino acids at the surface (blue) and at the interface (red)

7.2.2 Statistics of the amino-acid pairs

In the following analysis, the particular interactions of amino acid side-chains were our focus. The question was, whether some of the side-chain side-chain interactions are significantly pronounced, and whether the interface amino acid pairing exhibits a similar behavior as in the protein interior. It is the key for understanding the selectivity of the process, which is governed mainly by the specific interactions of the side-chains.

Preferences for the forming of particular amino acid pairs, presented in the figure 9, are independent of the actual amino acid composition. It is noticeable that amino acid side-chains selectively prefer some interacting partners over the others. This behavior is rather independent of the actual structural context – amino acids select the same partners at the interfaces as well as in the protein interior. This comparison is depicted on the panels A and B. In contrast, the pairing tendencies between amino-acid side chains at the protein surface shown in panel C differ from those of intramolecular and intermolecular pairing tendencies significantly. The pairing of hydrophobic amino acids is more pronounced while the pairing of charged amino acids clearly diminishes. Hydrophobic amino acids have a tendency to create hydrophobic patches on the surface of a protein. Surprisingly, charged amino acids usually do not create salt bridges on the surface but rather control intramolecular stabilization and the interaction with solvent.

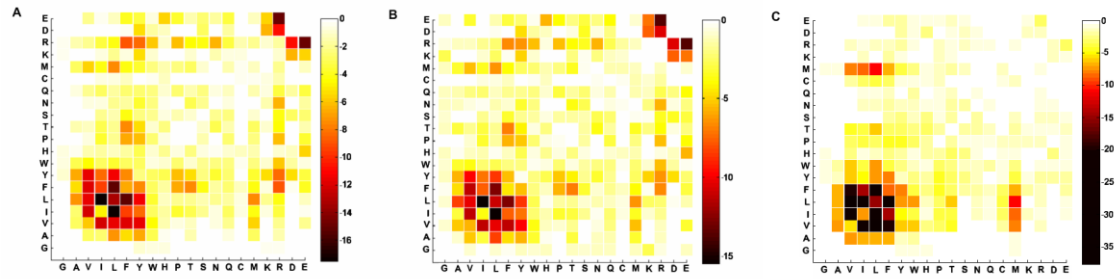

Fig. 9 – The preferences of amino-acid pairing. A – intramolecular, B – intermolecular, C – surface

7.2.3 Interaction energy analysis

Interaction energy between amino acids[125] could be a good measure characterizing not only intramolecular protein stability[126], but also the protein-protein interaction.

The energy behavior of a particular amino acid is therefore an important attribute that has to be examined in order to understand the nature of protein-protein affinity, specificity and selectivity. This analysis aimed to find amino acids with significant energetic contribution to the overall interaction energy and spot possible differences between the protein interior and interface.

This work concentrates solely on the impact of amino-acid side chains and their energy contribution to the protein-protein interaction. Thus, the amino acids were presented in C alpha representation[127], where backbone atoms are replaced with a methyl group that contains a C alpha atom. It evaluates the importance of the side-chain contribution to the overall energy and separates it from the influence of the backbone. The interaction energies were computed for all intramolecular pairs of the amino acids and for all the pairs at interfaces.

Residue interaction energy (RIE) is defined as a sum of the pairwise interaction-energy contributions of one particular amino acid with all the others in a protein. For every kind of amino acid, the RIEs were evaluated for the whole set and in three different subsets – the surface, interior and interface – and were calculated in the gas phase. The RIEs of the amino acids, which belong to an interaction interface, were evaluated for two different arrangements – the intramolecular and the intermolecular. They should represent the energy content of a particular amino acid in two different structural contexts.

Generally, the more interaction partners an amino acid has, the lower the value of the RIE is. For one particular amino acid, the RIE in the interior is always lower than at the surface. It is therefore interesting to compare the RIEs of surface amino acids with those composing the interaction interfaces. The interior amino acids can be seen as fully saturated in their interacting capabilities, whereas the surface amino acids remain unsaturated. It is therefore important to note whether the interaction capability of interface amino acids is saturated after a contribution from a second protein is added.

In order to cover the general features, the RIE distribution functions were constructed for every amino acid in each set. For all cases, median energies were

computed to obtain a characteristic – the value of the energy at the point where the distribution function equals 0.5.

Analysis of the results suggested that the interacting amino acids can be divided into three distinct groups, according to the physico-chemical and interaction properties. These groups are – charged, polar and hydrophobic amino acids. Figures 10, 12 and 14 show the median RIEs for all the groups and Figures 11, 13 and 15 show the example curves of the distribution functions for the RIEs.

7.2.3.1 Charged amino acids

The RIEs of charged amino acids have the lowest values between the studied types of amino acids. The average RIEs of a fully saturated positively charged amino acid ranges between -2500 and -3300 kJ/mol for the protein interior, whereas the average RIEs for fully-saturated negatively-charged amino acids in the same environment range between -1000 and -1500 kJ/mol.

The contributions of the other protein partner to the RIEs of the interface residues range between -800 and -1200 kJ/mol. There is a remarkable difference between the RIEs of positively charged amino acids at the surface (cyan) and at the interface (red). The intramolecular RIEs of interface residues are always significantly higher; in the case of Arg and Asp, the increase is by about 330 kJ/mol.

Charged amino acids seem to be responsible for a significant part of both intramolecular and intermolecular interaction energies. Because they carry a charge, their contributions are long-ranged. Although the RIEs of charged amino acids at an interface, including the interaction partner contribution (orange) do not reach the level of fully saturated interior RIEs, it is important to note that the contribution from the other partner comes from the interface residues only and including the rest of the protein would further increase the overall interface RIE.

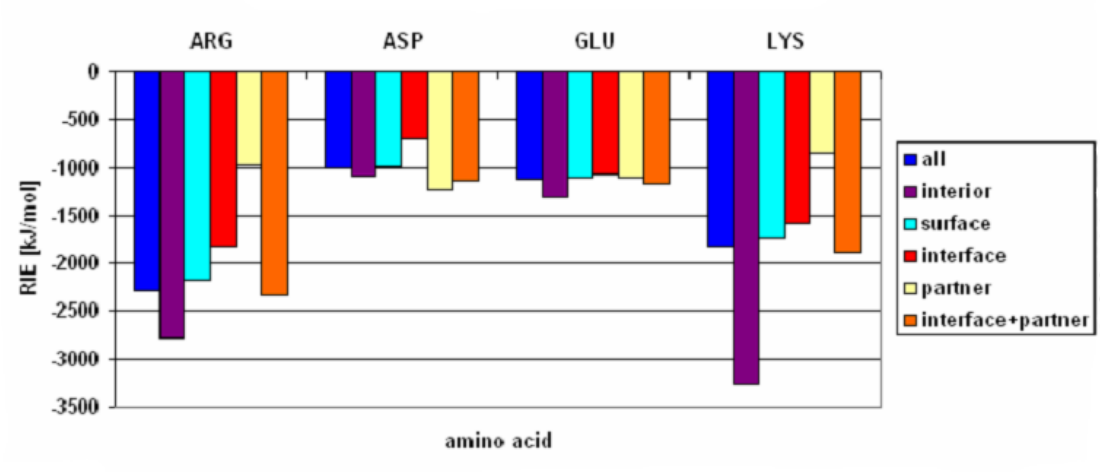

Fig. 10 – The medians of the RIE for charged amino acids

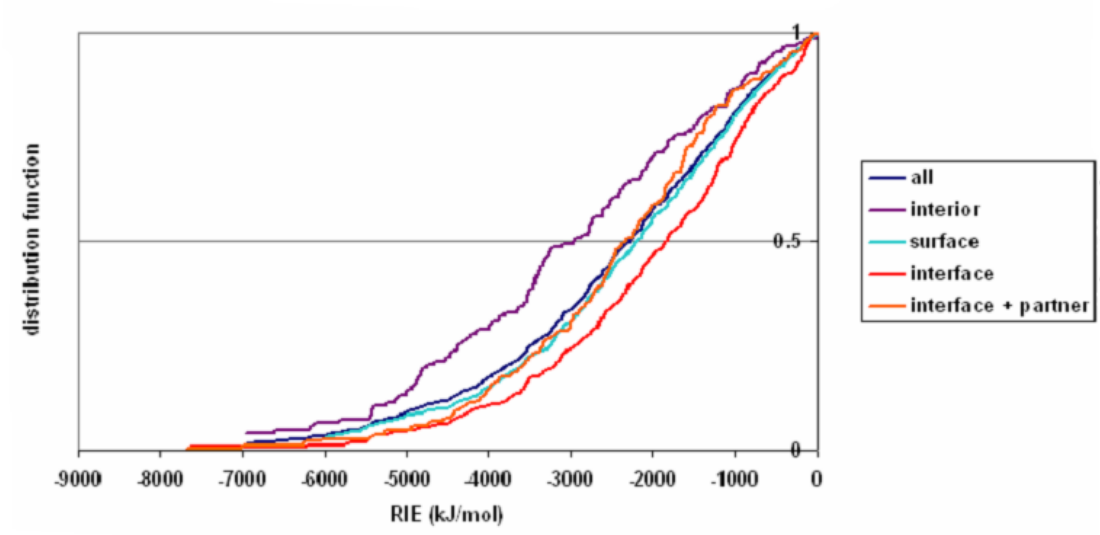

Fig. 11 – Arginine – the RIE distribution functions

7.2.3.2 Polar amino acids

As the value of the median for the RIEs for polar amino acids shows, they are between -150 and -300 kJ/mol when fully saturated. The largest contributions come from Gln, His and Tyr and RIEs range between -40 and -150 kJ/mol. It is interesting to notice that the RIEs of the interface amino acids themselves do not differ much from those of the surface amino acids and sometimes are even larger. That would lead us to the conclusion that interface polar amino acids are already well stabilized intramolecularly and that their arrangement at interface is quite compact. It is also important to note that interface residues are saturated to a similar degree than intramolecular ones.

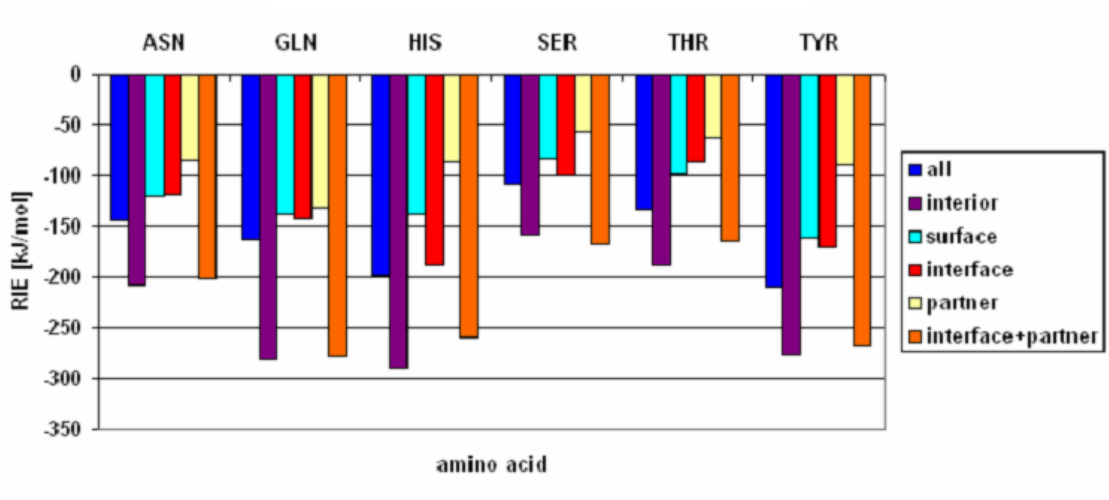

Fig. 12 – The medians of the RIE for polar amino acids

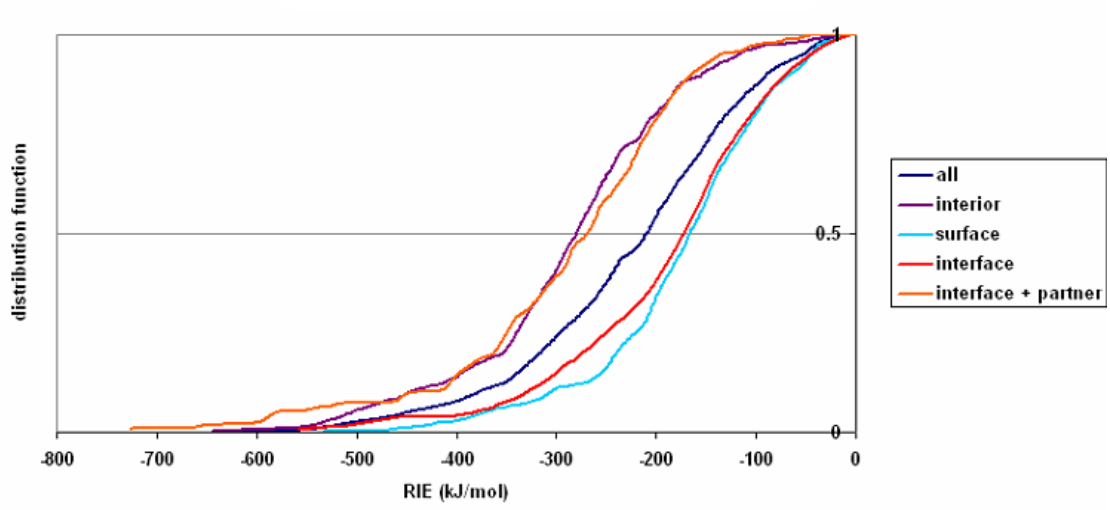

Fig. 13 – Tyrosine – the RIE distribution functions

All these features are schematically captured in Figure 13, showing the RIEs' distribution functions for Tyr. The interface amino acids (red), as compared to those at the average surface (cyan), are similarly saturated; the RIEs of interface amino acids when the interaction partner is taken into account (orange) show a dramatic drop, which leads to a curve almost identical with the RIEs' distribution function for Tyr in the protein interior (violet).

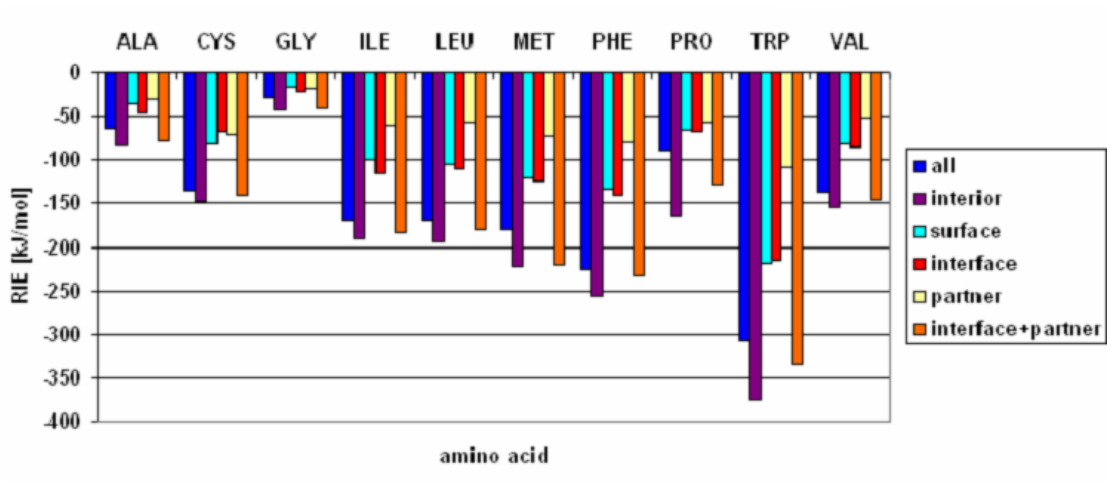

Fig. 14 – The medians of the RIE for hydrophobic amino acids

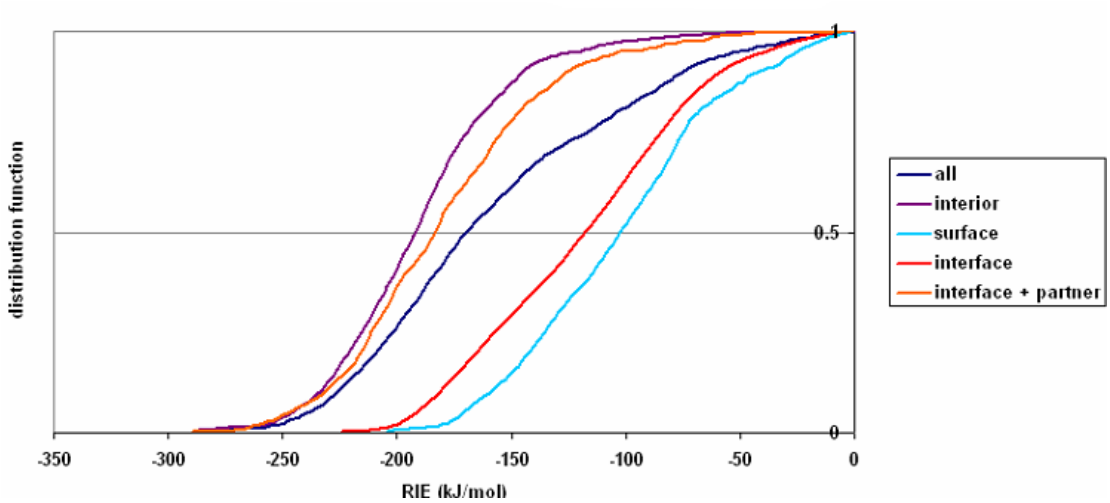

Fig. 15 – Isoleucine – the RIE distribution functions

7.2.3.3 Hydrophobic amino acids

The behavior of hydrophobic amino acids is similar to that of polar amino acids. Their interior RIEs, when fully saturated, correlate with their molecular weight and span a range from -40 kJ/mol for Gly to -380 kJ/mol for Trp. The contributions to the RIEs coming from the second protein (interface + partner) attain a -100 kJ/mol value for tryptophan. The RIEs of the interface amino acids without interacting partner are usually very similar to those at the surface of the protein. Upon the interaction, the interface amino acids RIEs are nearly saturated, but not to the same extent as the polar amino acids are.

Figure 15 reveals the most typical situation for the RIE distribution functions in the case of Ile. The intramolecular RIE of interface residues (red) is systematically lower than the RIE of the surface residues (cyan). When the contribution from other partner is included (orange), the RIE is closer to the fully saturated interior residues (violet).

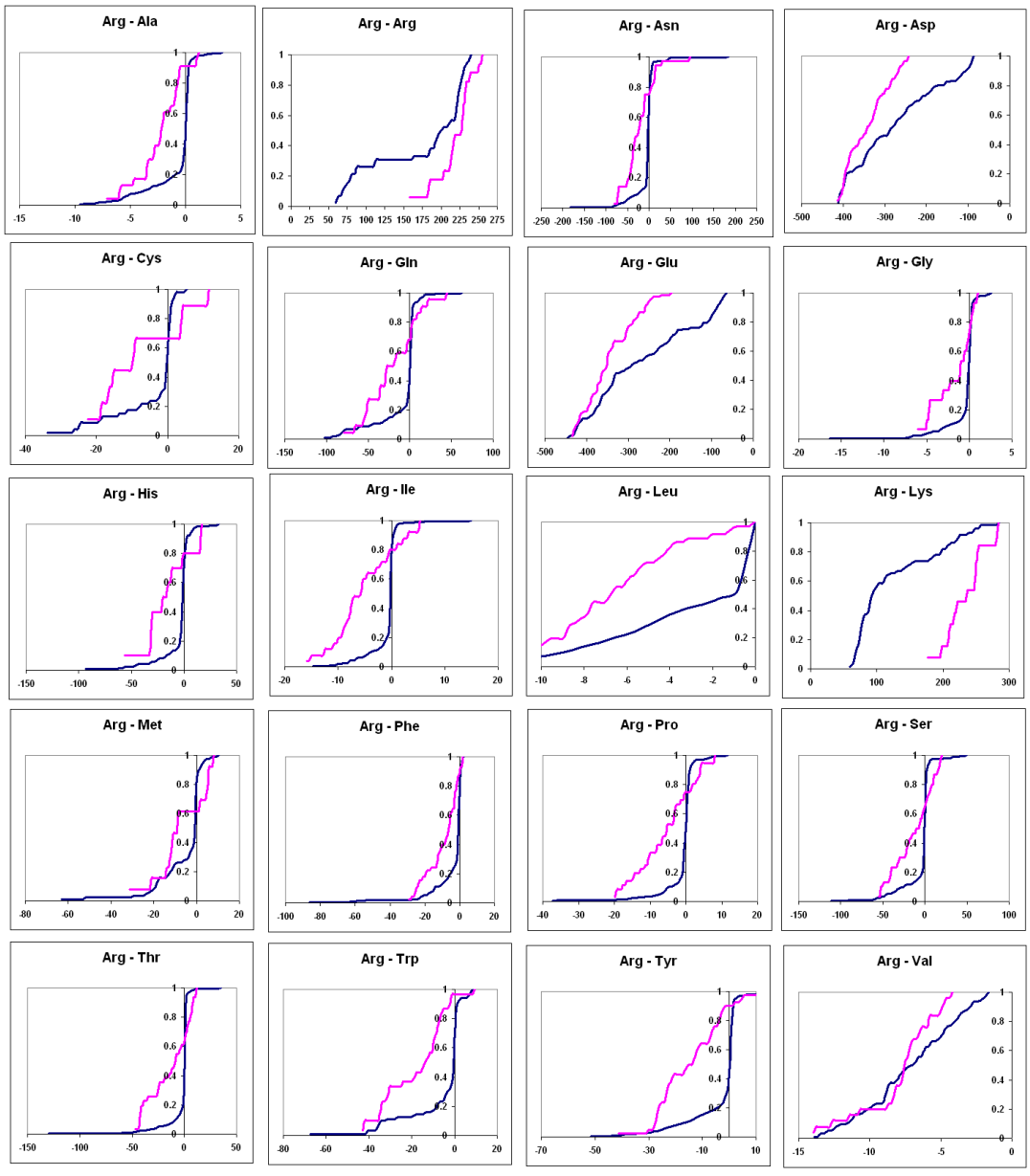

Fig. 16 – The distribution functions of the pairs of all 20 amino acids with arginine – a comparison of intramolecular (blue) and intermolecular (pink) pairs. The horizontal axis corresponds to the energies of the pairs in kJ/mol.

7.2.3.4 Comparison of intra- and intermolecular pairwise interaction energies

Interaction energies of particular amino acid pairs were computed for intramolecular and intermolecular pairs and analyzed statistically. The comparison of the distribution functions for the interaction energies of intramolecular and intermolecular pairs for Arg with all twenty amino acids are shown in the figure 16. The comparison of the curves reveals that the Arg interaction energies at the interfaces are generally substantially higher than its intramolecular interaction energies. It could be explained by the fact that the amino acids at the interfaces have the ability to find energetically more favorable positions than in the protein interior, which means that interfaces are more stabilized during protein-protein interaction than in the protein interior.

7.2.4 Analysis of the protein surface

In order to reveal characteristic clustering of the amino acid side-chains at the interfaces, the protein surface was dissected to the interacting and non-interacting parts. Every amino acid was analyzed in terms of its contacts with other side-chains, which were closer than a certain limit – which was set as the double of the corresponding van der Waals radii of the side-chain atom. The frequencies of these neighboring pairs are shown in Figure 17, panels A and B. The main conclusion is that hydrophobic neighbors (hydrophobic patches) are much more pronounced at the interfaces and hardly occur at the non-interacting surface. In contrast, the number of charged pairs at the interfaces is even lower than at the non-interacting surface.

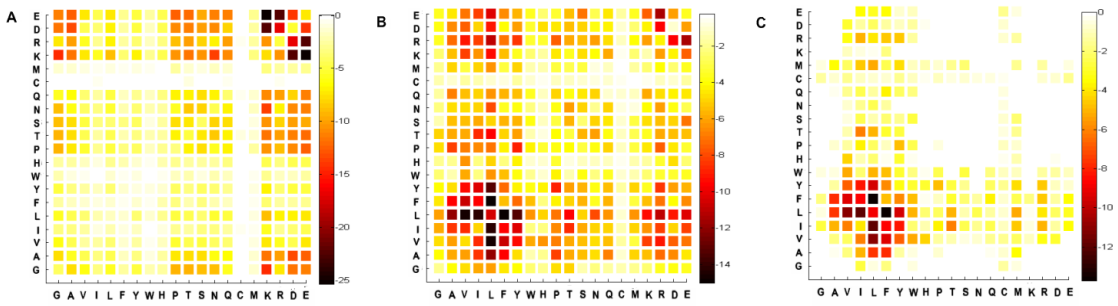

Fig. 17 – Neighboring amino acids at the non-interacting surface (A) and interface (B). C shows the difference between the two – the neighbors preferred at the interface.

7.2.5 Discussion

An analysis of protein-protein interfaces was carried out on a set of x-ray structures with high resolution. Characteristic composition of interaction interfaces was revealed. Interfaces prefer branched hydrophobic and aromatic amino acids, namely Leu, Ile, Val, Phe, Tyr, Met and Trp. On the other hand, interfaces lack the small hydrophobic amino acids Ala and Gly and charged amino acids Asp, Glu and Lys, when compared with the non-interacting surfaces. Arg makes an exception among the charged amino acids, occurring at the interfaces with the similar frequency as at the non-interacting surface. It was also shown that hydrophobic amino acids tend to group together at the interfaces, while the charged amino acids tend to be more separated. The enhanced hydrophobic character of the interfaces and lower occurrence of charged amino acids can destabilize the hydration layer of the interface area. Thus, forming the protein-protein complex can be favorable because of the favorable change of hydration structure.

The analysis of amino acid pairs at the interfaces revealed that pair forming at the interfaces obeys the same rules as in the protein interior. However, the comparison of the energetics of intramolecular and intermolecular pairs revealed that interface pairs generally adopt much more favorable mutual orientation and thus possess lower interaction energies. It suggests very high geometric and chemical complementarity at the interacting interfaces.

Analysis of residue interaction energies (RIE) showed that charged amino acids at the interfaces, especially Arg, are less saturated by the intramolecular interactions than at the non-interacting protein surface. Their tendency to find proper interacting partners can be the key to the selectivity of the process of protein-protein interaction.

These findings not only help us understand the process of PPI, but they can also serve as a tool for searching the interface area on a protein surface. We propose that the marks that specify the possible interaction interface are: amino acids Arg, Ile, Leu, Met, Phe, Trp, Tyr, Val, in the vicinity of which other amino acids from this group are present (especially other hydrophobic amino acids) and surface charged

amino acids, whose RIEs are lower than the typical value for an ordinary surface. These features are a clear result of our analysis and are suitable for further algorithmization of the process of interface prediction. The basis of the proposed algorithm is an iterative scoring function. In the first round, all the residues would be scored according to their chemical nature, accessible surface area and in case of charged amino acids also residue interaction energy. Then, in the other rounds, the scores would be updated according to the vicinity of other residues with high scores. After few rounds, patches of highly scored residues should emerge. These patches would be candidates for interface. The first step would be adjusting this proposed algorithm on the set of protein dimers utilized in this work. Then its predictive power could be tested on other structures for which their interface areas are known. If it proved itself useful, it could be used for prediction of possible interface areas of proteins with unknown interaction properties.

7.3 Protein hydration structure

The PDB database was searched for the sets of redundant structures. These sets are potential candidates for implementing this analysis. 131 sets containing at least 100 structures were found – 78 monomeric proteins, 20 dimers, 8 trimers and 25 higher oligomers. Our model case was the T4 phage lysozyme. The x-ray resolution of the pdb structures ranged between 0.15 and 0.25 nm. The set of structures consisted of 391 structures with the symmetry P 32 2 1.

7.3.1 Water density grid

The water density grid (wdg) contains all the information about the protein hydration structure that can be obtained by the superposition of all the x-ray structures in a set. It is a map of the water occupancy. The wdg of lysozyme consisted of 12,269,796 grid points, thus covering the overall volume of 98.16 nm^3. Figure 18 shows how the increasing threshold of water density determines the number of wdgps whose water density exceeds this threshold. Only about 4.5 % of all the wdgps possess nonzero water density.

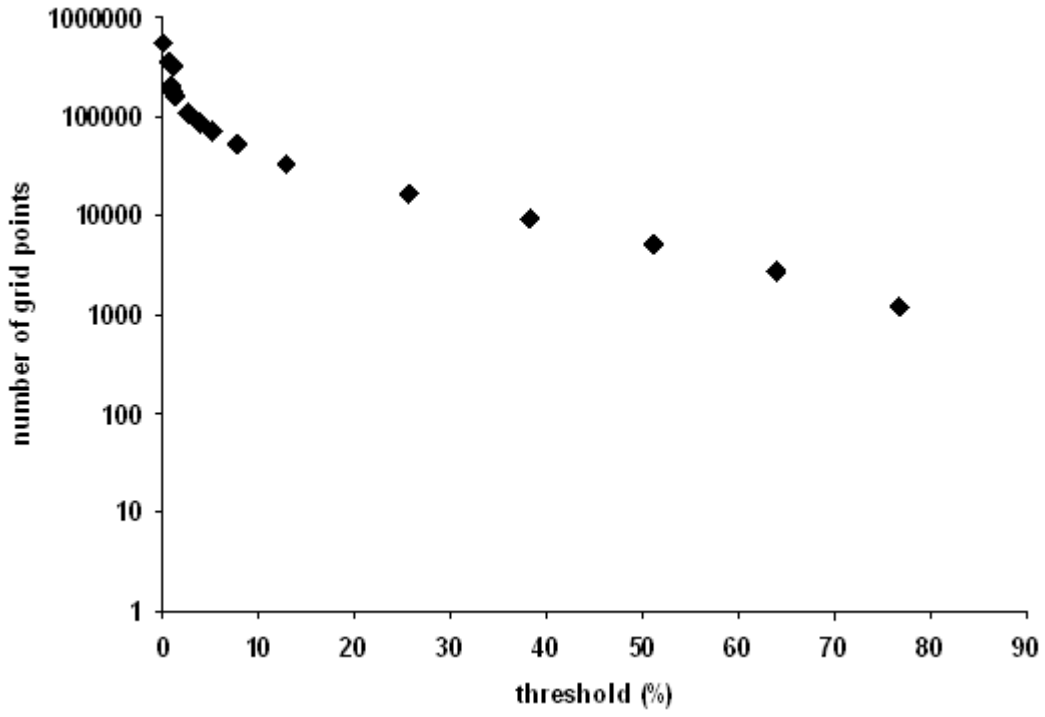

Fig. 18 – The number of grid points exceeding the threshold water density

The wdg was processed by a simple clustering algorithm. In order to identify distinct clusters, a threshold for the water density must be set. The resulting number of the clusters and their size depends on the threshold value, as shown in the figure 19. Higher threshold yield rather low number of distinct clusters whose maximum and mean size is similar; when approaching the lower thresholds, the size of the cluster is growing and some of the clusters start to associate. The dependence of the maximum and mean size of the clusters is illustrated in the figure 20.

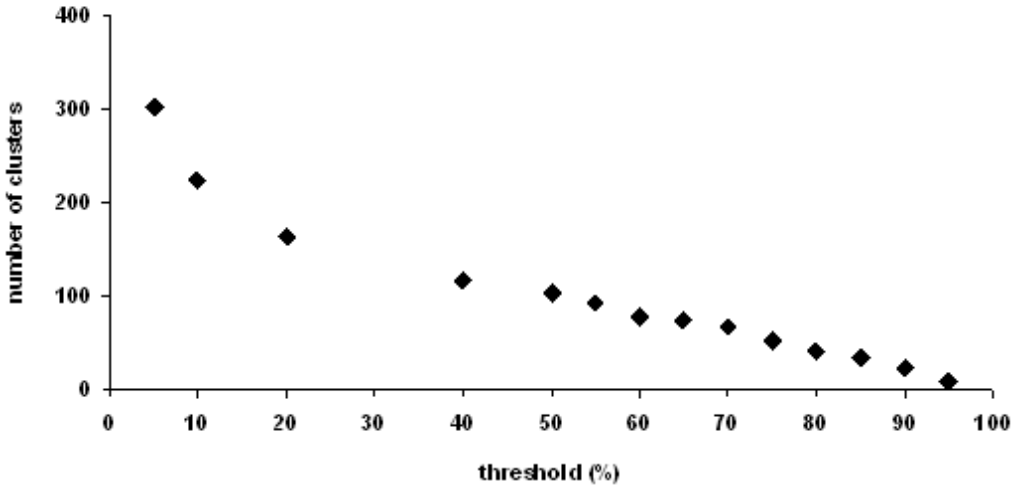

Fig. 19 – The dependence of the number of clusters on the threshold

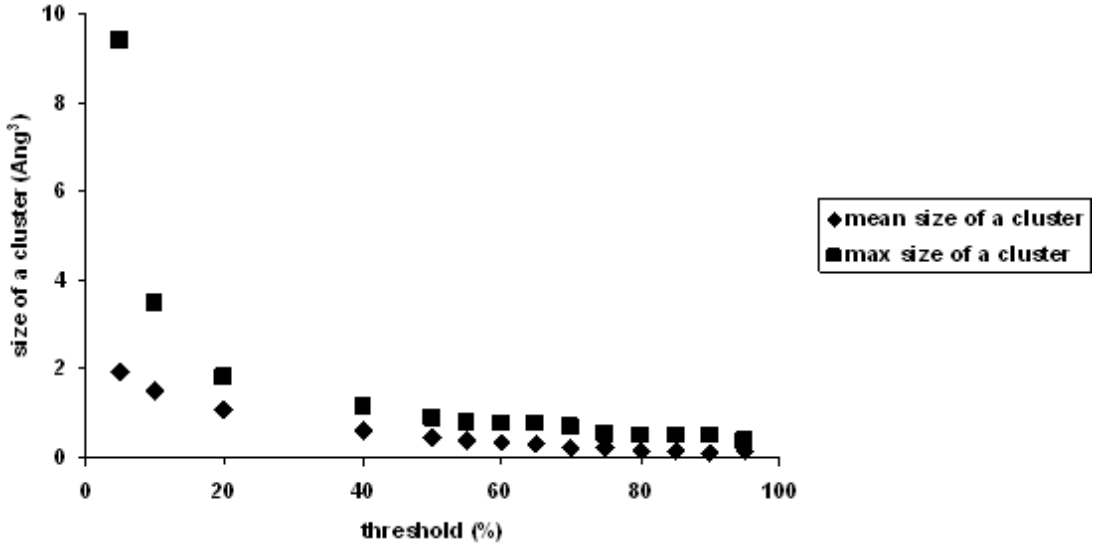

Fig. 20 – The dependence of the size of clusters on the threshold

In further analysis, the interaction of the clusters with the protein was examined. Waters form hydrogen bonds with the electronegative protein atoms. The preference of the water clusters for the particular protein atoms is relatively independent of the selected threshold, as shown in the figure 21. Figure 22 completes the whole picture, showing the percentage of the particular atom type that participates in the hydrogen bonds with water clusters. Figure 21 therefore shows the absolute numbers while Figure 9 shows the relative numbers.

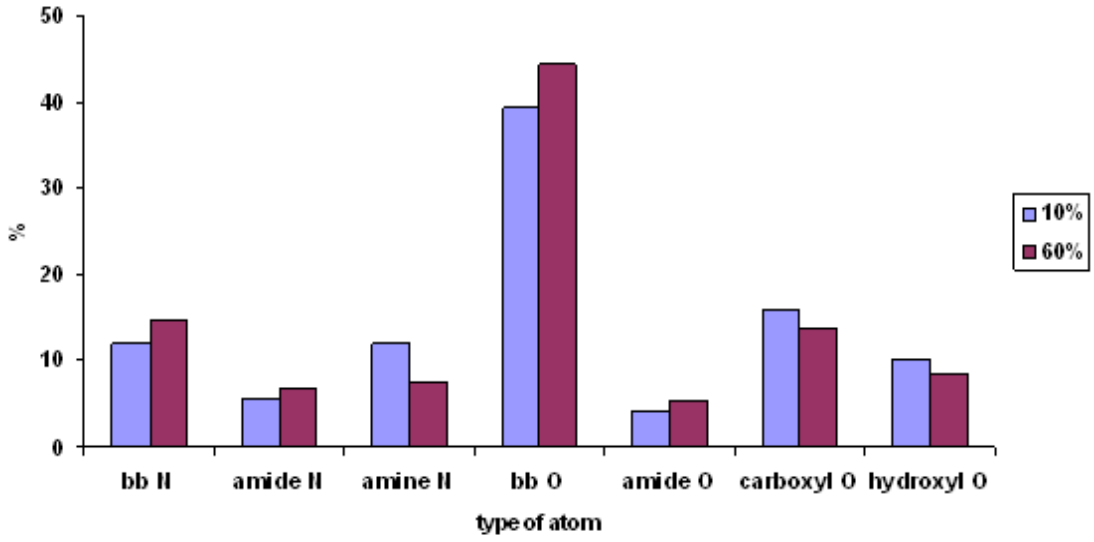

Fig. 21 – The absolute number of the protein atoms interacting with water clusters: 10% threshold (blue) and 60% threshold (purple).

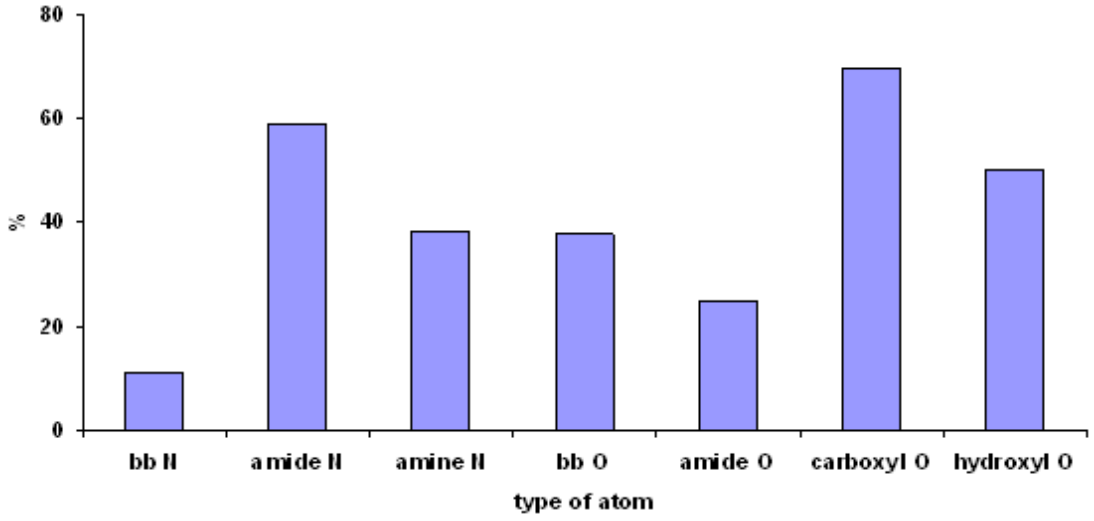

Fig. 22 – The relative number of the protein atoms forming contacts with waters.

54.5 % of all the protein atoms forming hydrogen bonds with waters are the backbone atoms, 45.5 % are the side-chain atoms. Most of the water clusters interacting with the backbone bind to the backbone oxygen atoms. The population of the water molecules bound to them is more than three times higher than the population of the waters interacting with the backbone nitrogens. Among the side-chain atoms, the water population around the carboxyl oxygens is clearly the highest – it is nearly two times higher than the water population around the amine nitrogens. Amide nitrogens, although not so numerous in the case of lysozyme, are also frequent binding partners of water molecules. Within the amidic group, the binding frequency of the nitrogen atoms is more than two times higher than that of the oxygen atoms. The water population around the hydroxyl atoms is also high. It can be concluded that the population of waters is the lowest around the nitrogen backbones and amide oxygens while the highest values are reached for carboxyl oxygens and amide nitrogens.

All the oxygen and nitrogen atoms from the protein were divided into two groups: those having a contact with the clusters within a 3.1 Å range, and those having no contact. The profile of atomistic solvent accessible surface area and corresponding beta factor was examined. Figure 23 shows clearly that the fact that some atoms lack the contact with waters is clearly connected with their solvent inaccessibility. The beta factors of the two groups, depicted in the figure 24, do not differ so significantly.

Furthemore, the relation of the number of intramolecular hydrogen bonds and interaction with water clusters was examined. Each oxygen and nitrogen atom of the protein was assigned by an index describing the number of other oxygen and nitrogen atoms of the protein within the range of 2.5 to 3.1 Å – apparently being a candidate for a hydrogen bond. The figure 25 shows the difference of the atoms having contact with water clusters and atoms without contact. While the group of atoms having no contact shows a quite even distribution, the group of atoms having contact with the water clusters consists primarily of atoms without internal hydrogen bonds or with only one hydrogen bond (it is more than 80 % of all the atoms of this group). This confirms the observation that waters preferably bind to

the nitrogen and oxygen atoms that are not saturated with the intermolecular hydrogen bonds[86]. These results can support the the endeavor to localize possible water binding spots in the protein structure. The most pronounced indicator is the low number of internal hydrogen bonds, but the combination with the monitoring of the atomic SAS area and atom type can lead to quite robust results.

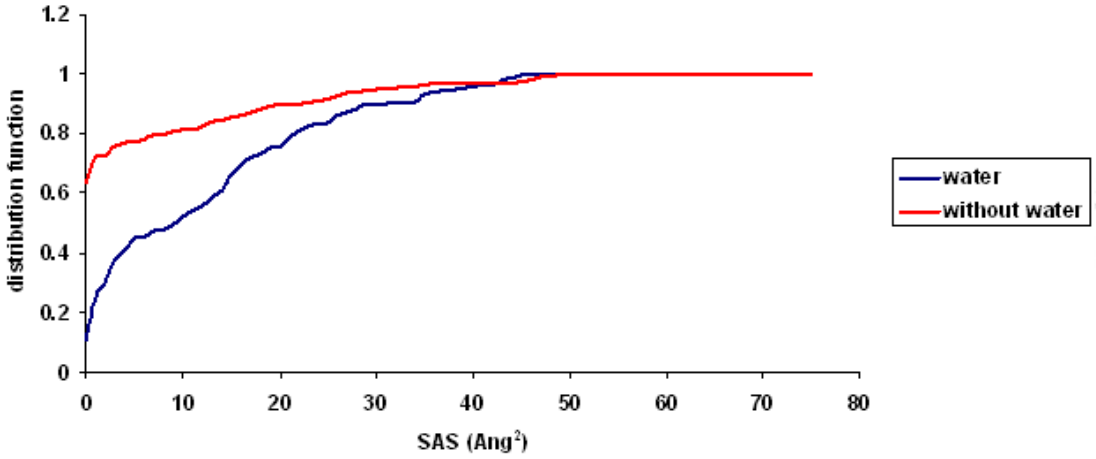

Fig. 23 – The SAS of the oxygen and nitrogen atoms having contact with water clusters (blue) and atoms with no contact (red).

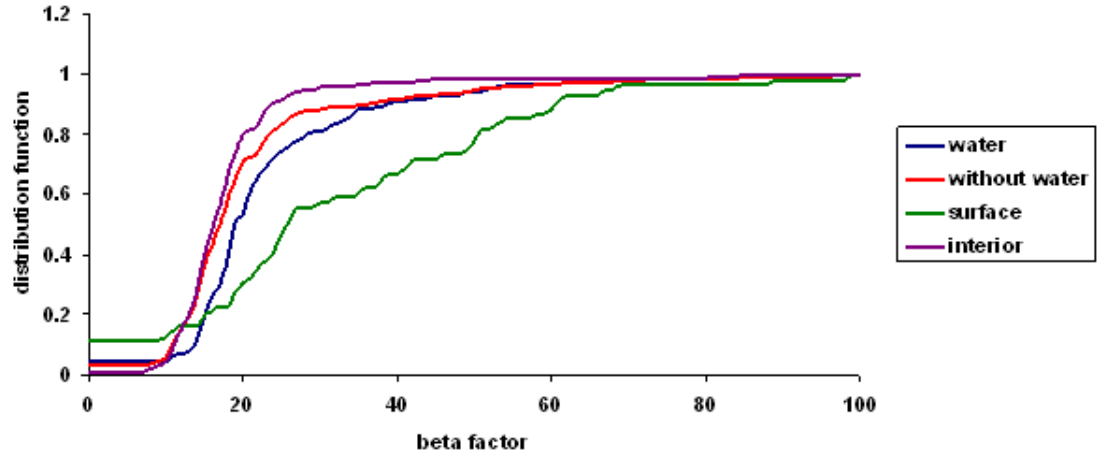

Fig. 24 – The beta factor of the oxygen and nitrogen atoms having contact with clusters (blue, water) and the atoms with no contact (red, without waters). The beta factor of the surface atoms without contact (green, surface) and the interior atom with no contact (violet, interior)

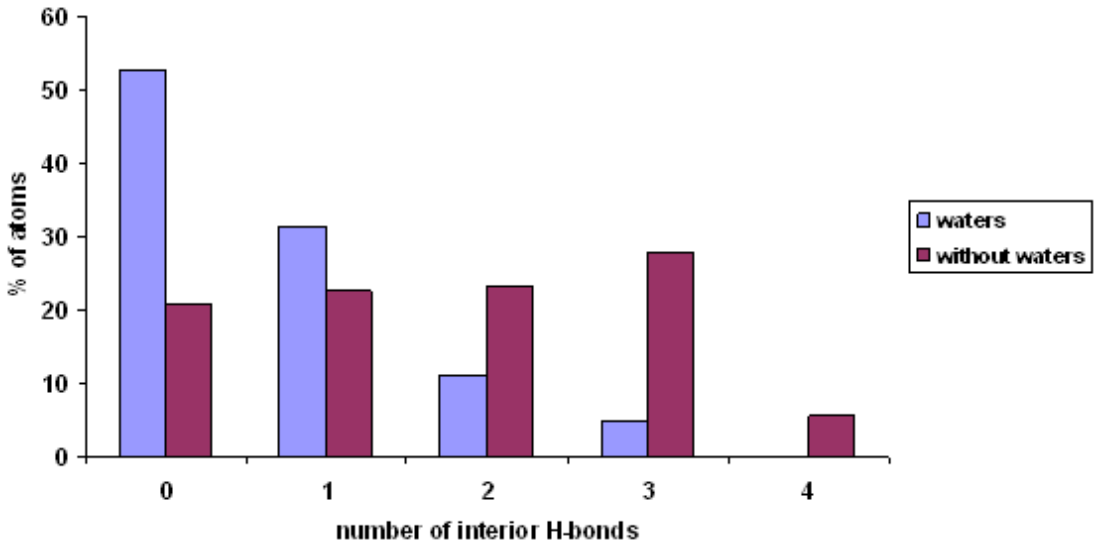

Fig. 25 – The percentage of the electronegative atoms with various number of interior hydrogen bonds – a comparison of the atoms having contact with water clusters (blue, waters) and the atoms with no contact (red, without waters).

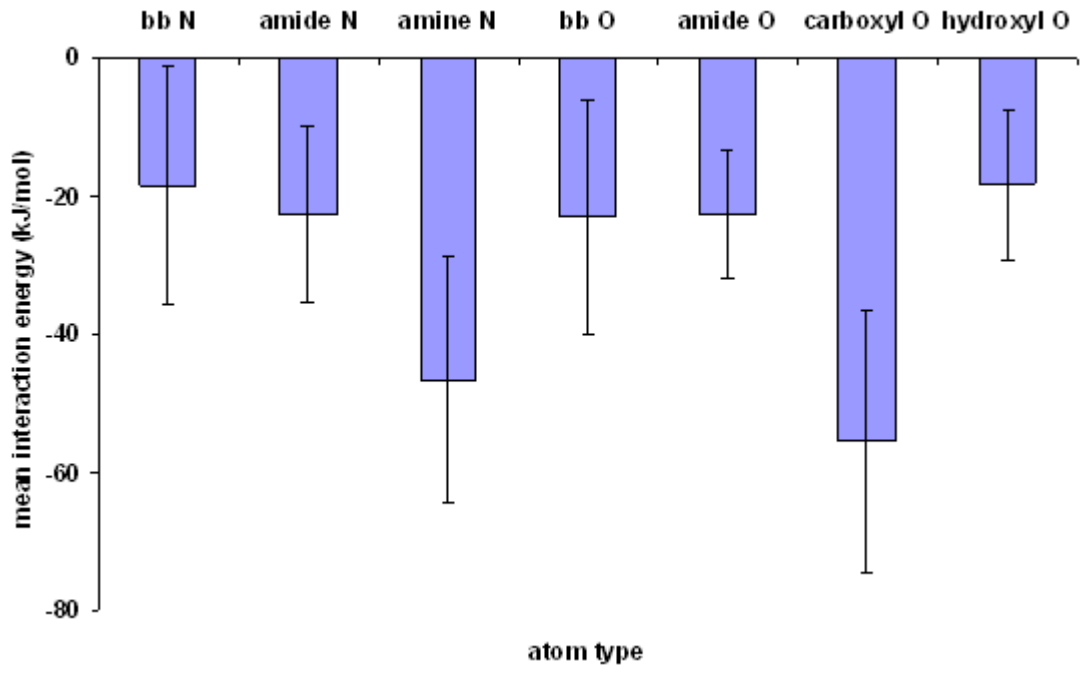

Fig. 26 – The mean pairwise interaction energies of water clusters depending on the interacting atom type.

7.3.2 Interaction energies

The mean pairwise interaction energies of a water molecule with one amino acid partner are shown in the figure 26. The interaction energies of charged amino acids

are reasonably higher (between 30 and 40 kJ/mol) than the interaction energies of other neutral side-chain and backbone atoms (between 10 and 20 kJ/mol). We cannot directly compare the energetic contribution of all amino acids, since they differ in their characteristics and their values are calculated in the gas phase approximation. The energy of the hydrogen bond with carboxyl oxygen is slightly higher than the energy of the hydrogen bond with amine nitrogen. Interaction energies with other atom types are nearly similar.

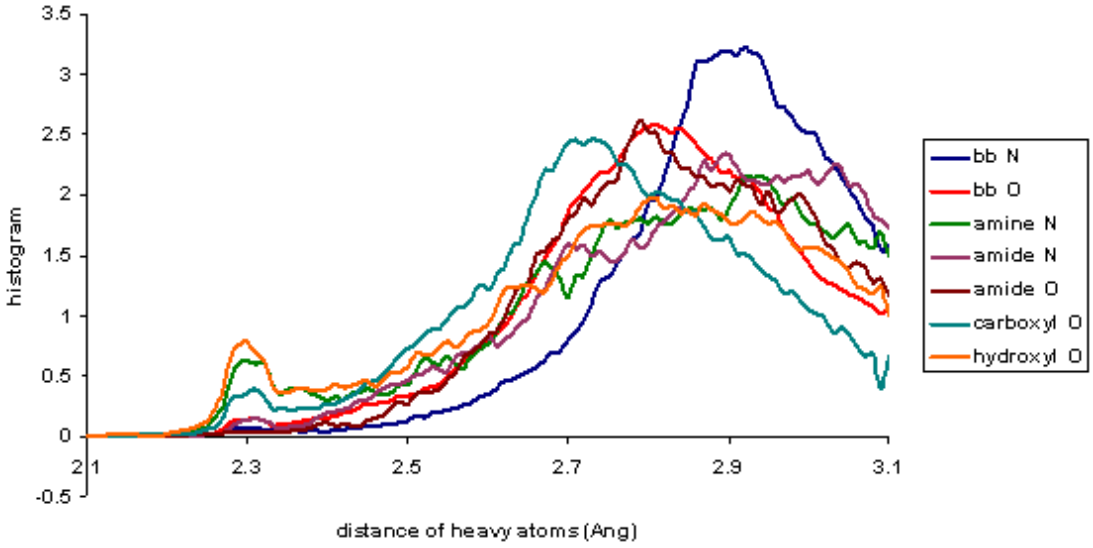

Fig. 27 – The histograms of hydrogen-bond lengths according to atom types. The backbone nitrogen (blue, bb N), backbone oxygen (red, bb O), amine nitrogen (green, amine N), amide nitrogen (purple, amide N), amide oxygen (brown, amide O), carboxyl oxygen (turquois, carboxyl O), hydroxyl oxygen (orange, hydroxyl O)

7.3.2.1 Geometry of the hydrogen bonds

The energetic profile of protein-water interaction is completed with the geometry analysis of hydrogen bonds. Figure 27 shows the histograms of hydrogen-bond lengths, according to the amino-acid atom types. The hydrogen bonds with the backbone nitrogen are longer (2.9 Å) than the corresponding hydrogen bonds with the backbone oxygen (2.8 Å). The geometry profile of the hydrogen bonds with the amide oxygens is similar to that of the backbone oxygen, as the chemical nature of this function is similar. The hydrogen bonds with carboxyl oxygens are the shortest – their peak lies at 2.7 Å. It corresponds the high energy of this hydrogen bond. The lengths of the hydrogen bonds with the other side-chain atoms are generally longer

and more variable. There is a small but distinguishable peak around 2.3 Å – evidently it doesn't correspond to a standard hydrogen bond. We assume that it corresponds to the length of non-covalent bond of sodium or potassium ions, whose electronic densities were probably wrongly attributed to the water molecules.

7.3.2.2 Total interaction energies of water clusters

The total interaction energies of cluster-representative water molecules were calculated summing the individual pairwise interaction energies for each cluster. Each particular interaction was characterized according to the interacting atom type as charged (C), polar (P), non-polar (N) or backbone (B). The interacting clusters can have one, two or three protein partners. Each cluster was denoted according to the type of its interactions with the above mentioned abbreviations. E.g. CP denotes the clusters that interact with one charged and one polar amino acid. This division enables the comparison of the interaction energies within each group. The results are shown in the figures 28, 29 and 30.

There is not enough data for the proper analysis of the clusters interacting with three partners, as we can see in the figure 28. As only the BCP group contains more than one cluster, the interpretation of this data is not possible. The situation is slightly better, but still not convincing, in the case of the clusters interacting with two partners, depicted in the figure 29. The number of clusters in each group is still rather small, but the cluster energy of the members of a particular group tends to fall into the similar range. This tendency is even more pronounced in the case of the clusters interacting with one amino acid partner in the figure 30. Unexpectedly, there is no clear connection between the interaction energy of the water-cluster representative and the actual cluster occupancy. An overall look at the results reveals that the interaction energies are roughly proportional to the number and type of the interacting atoms.

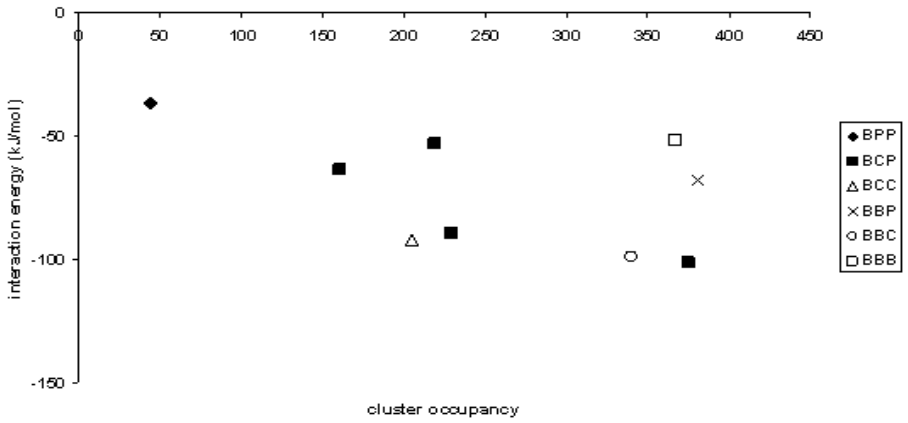

Fig. 28 – The clusters interacting with three amino-acid partners. B – backbone, C – charged, P – polar

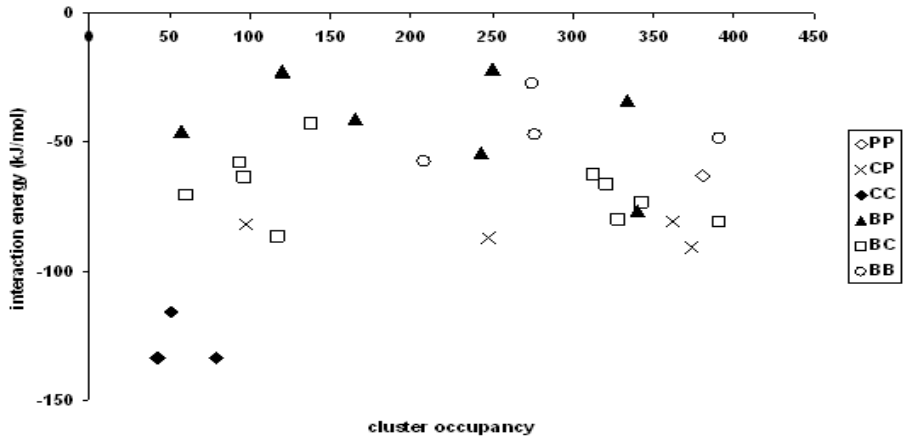

Fig. 29 – The clusters interacting with two amino-acid partners. B – backbone, C – charged, P – polar

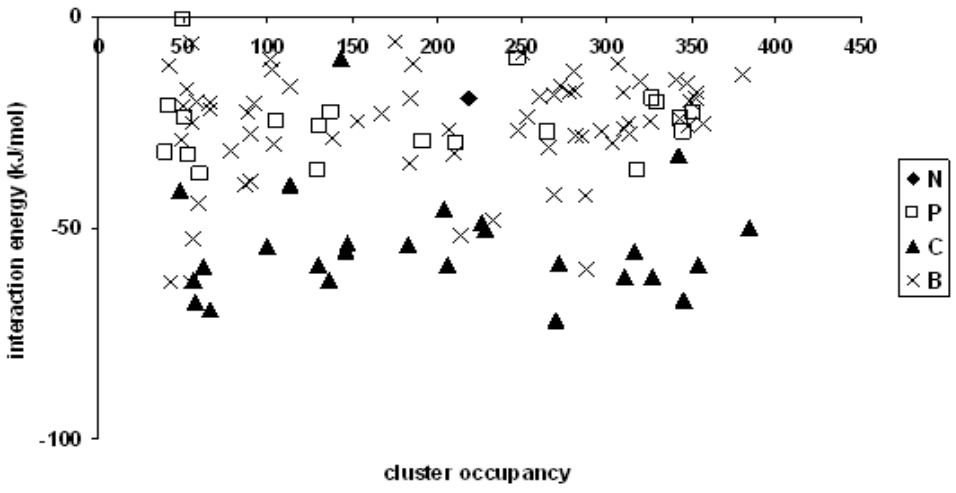

Fig. 30 – The clusters interacting with one amino-acid partner. B – backbone, C – charged, P – polar, N – non-polar

7.3.3 Topological analysis of the hydration shell

The water clusters form a net structure of hydrogen bonds around the whole protein. The clusters were analyzed to find protein-atom partners and other water-cluster partners within a distance range of 2.6 Å and 3.1 Å, which corresponds to a conventional hydrogen bond. Each cluster can form internal and external hydrogen bonds – their overall number defines the order of the cluster, which describes the topological position of a cluster in the whole net. Figure 31 shows the number of clusters of a distinct order, which undergo the Gauss distribution with its peak around the order of two.

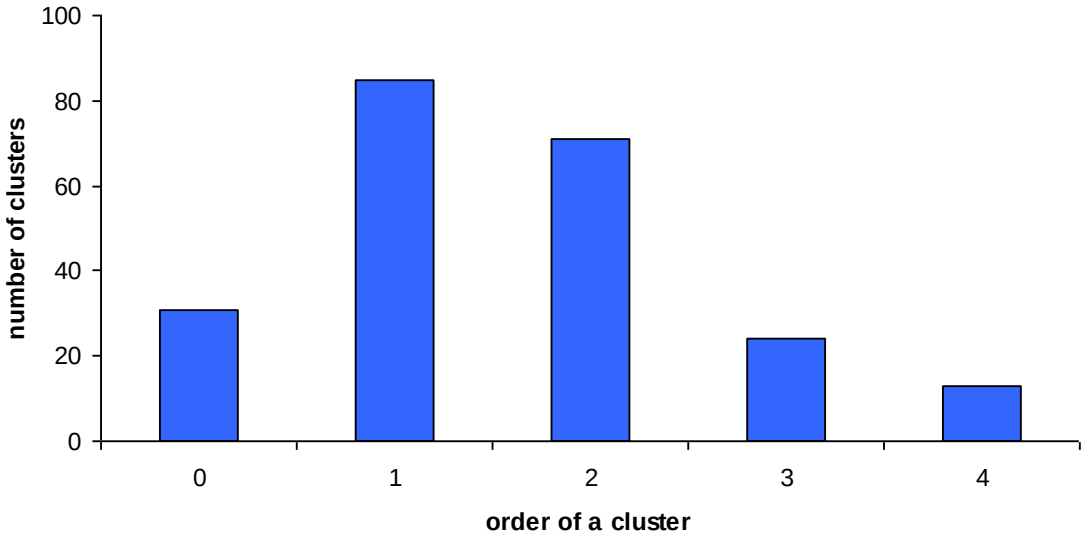

Fig. 31 – The number of clusters of a distinct order

Figure 32 shows that the mean occupancy of the water cluster grows with the cluster order. That means that the more hydrogen bonds a cluster forms with its partners, the higher the probability that water can be found in the x-ray structure. Table 5 shows more detailed information for each type of cluster.

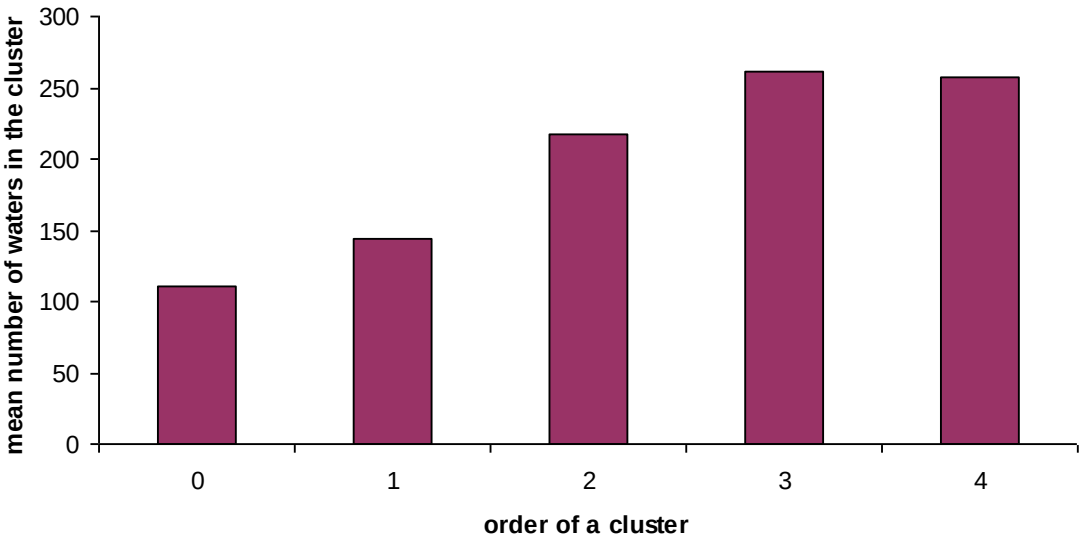

Fig. 32 – The mean occupancy of clusters of a distinct order

Table 5. The mean occupancy and the occurrence of clusters depending on their interaction. The first number in a cell is the mean occupancy of the clusters and the value in parentheses represents the number of clusters.

	number of cluster partners				
number of protein partners	**0**	**1**	**2**	**3**	**mean**
0	110.5 (31)	115.1 (23)	167.6 (13)	194 (1)	**124.2**
1	155.0 (62)	238.5 (43)	291.8 (9)	333.0 (2)	**199.6**
2	202.0 (15)	251.4 (9)	240.4 (5)		**224.0**
3	237.8 (5)	283.0¨ (4)			**257.9**
4	169.0 (2)				**169.0**
mean	**153.0**	**206.3**	**222.5**	**286.7**	

The clusters in the first column of the table do not interact with other clusters. 31 clusters that do not interact with the protein can arise due to the symmetry of the crystallographic cell. The rest of the clusters that do not form hydrogen bonds with other water clusters interact only with the protein molecule. The other clusters form

hydrogen bonds also with each other and thus are part of a higher organisational structure. They make up distinct superclusters that consist of 2 to 8. Figure 33 shows the number of each kind of supercluster (including the isolated clusters) and number of waters present in these superclusters. As the purple columns of this figure reveal, the total number of the water clusters participating in any supercluster is higher than the number of isolated water clusters. Figure 34 shows an example of a structure of such a supercluster.

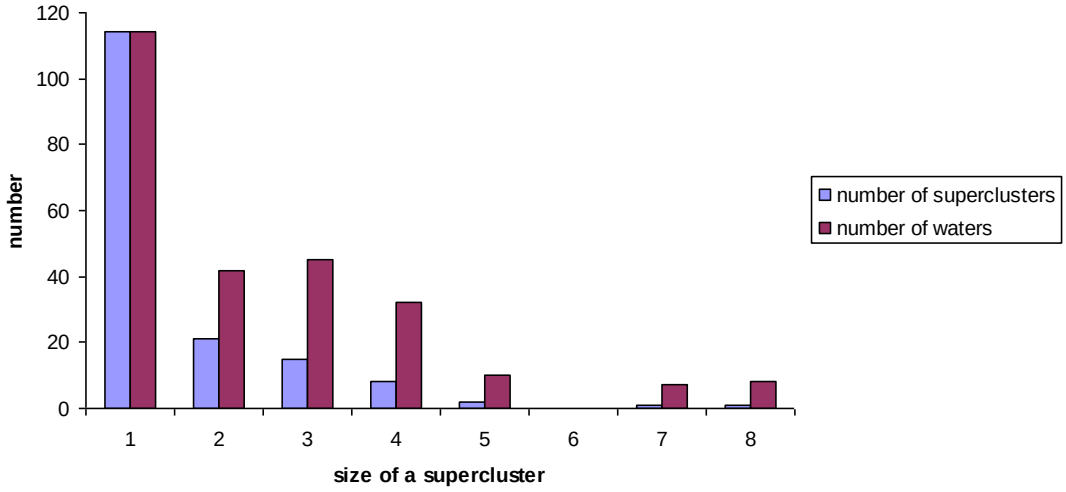

Fig. 33 – Higher organizational structure of the hydration shell

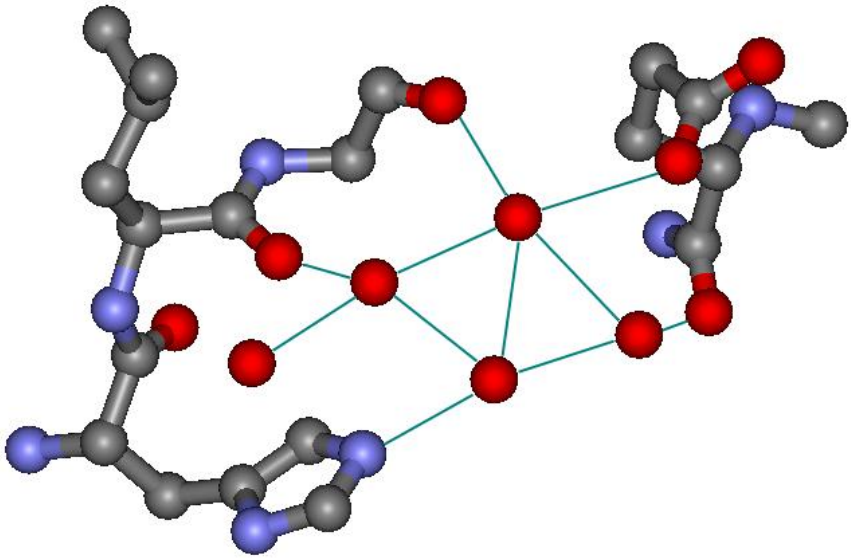

Fig. 34 – An example of the supercluster of water clusters and its connection to the protein molecule.

7.3.4 Discussion

A systematic approach to study the protein hydration structure utilizing the multiple crystallographic data was presented using the example of T4 phage lysozyme. This method enables to localize important sites in the protein hydration structure.

Hydration waters interact with the main-chain and the side-chain atoms of the protein to a similar extent. However, the water molecules tend to reside near the oxygen atoms much more than near the nitrogen atoms – this difference is most pronounced in the case of the backbone atoms. Moreover, it was shown that the hydration waters seek the oxygen or nitrogen atoms of the protein that are not saturated with the interior hydrogen bonds – more than 80 % of the localized water clusters interacted with the electronegative atom with maximum one interior hydrogen bond. This fact is clearly illustrated in the figure 25.

The interaction energies of the water molecules with charged amino acids range between 30 and 80 kJ/mol, while the interaction energies with the other amino acids is generally smaller – usually between 5 and 40 kJ/mol. The interaction energies of the water clusters are roughly proportional to the number and type of the interacting protein atoms. However, the clear connection between the interaction energy and the occupancy of the water cluster could not be found in this approximation.

The results of this work is that the structure of the hydration layer is quite complex – the fact that some places are regularly occupied by the water molecules, while others are not, cannot be unambiguously explained solely by the affinity to the protein environment, but the interaction with other water molecules need to be taken into account. It was shown that water occupancy is higher at the spots with high overall number of hydrogen bonds – with protein atoms and with other water clusters.

These results suggest that the structure of the hydration shell of the protein is an important complement of the structure of the protein itself. The methodology used can be extended to dozens of other structures in the protein data bank suitable for

this kind of analysis. This work defines features for searching the sites where interaction with water is possible – particularly the electronegative atoms that are unsaturated in terms of the internal hydrogen bonds and that are accessible for the solvent. The water molecules generally prefer the oxygen atoms over the nitrogen atoms. On the other hand, the results also suggest that protein hydration is a complex phenomenon which needs more sophisticated modeling and larger set to validate presented results.

8 Conclusions

This doctoral thesis presents three studies of the non-covalent interactions of proteins which explore this subject from three different points of view.

The first study uses the DFT/CC methodology to obtain reliable benchmark interaction energies of the most common protein functional groups with the hydrophobic surface. It has been shown that the results are in a good agreement with the most accurate experimental data and with the results of other theoretical methods that can account for the dispersion interaction. It is the first study of this methodology with such a vast spectrum of interacting atom types and groups. These results open a possible use of this methodology for the description of biological systems. In the next step, it would be needful to choose good representatives for each protein functional group and model all the possible kinds of the interactions that occur in proteins, not only hydrophobic, as presented in here. The possible candidate molecules for the reference systems modeling the protein functional groups could be ammonia, guanidine, water, methanol, formaldehyde, formic acid, hydrocarbons, benzene, phenol, pyridine, pyrrole and hydrogen sulphide. Although the journey towards the DFT/CC description of interactions of amino acids is still far away, the results of this study are promising.

The two studies of bioinformatic character were performed to crystallographic data in the protein data bank, and therefore present more empirical attempt that is closer to the experimental conditions. The study of protein-protein interaction examined chemical and energetic properties of protein interfaces. Key features are the presence of hydrophobic patches and isolated charged amino acids that are not saturated in their interaction with intramolecular amino acids. These results propose an iterative algorithm for interface localization. In the first step, this algorithm would score every amino acid according to its SAS area, RIE and chemical character. The scoring function would be based on this analysis and adjusted for the best performance during this algorithm. In the next step, the scores would be update according to the vicinity of other amino acids with high scores. After few iterative

steps, possible candidates for the interfaces would be localized. In the first phase, the ability of this algorithm to predict interfaces could be tested on this set of 69 protein complexes with known interfaces.

Moreover, the presence of hydrophobic patches revealed that the reorganization of the hydration structure of both proteins could be one of the important driving forces during the formation of a protein complex. The hydration structure of the protein is therefore an integral part of the protein, which can influence the whole process of molecular recognition. This fact raised the need to examine the hydration structure of the proteins.

T4 phage lysozyme was chosen as a case study protein, because having multiple records in the PDB. The superposition and clustering algorithms revealed the net of water clusters – spots where the water molecules occurred in a substantial portion of structures. This method enables the study of the overall topology of the protein hydration shell, as well as the study of the interaction of particular water clusters with the protein. This thesis presents the development of the methodology and shows its possible use. However, PDB contains tens of other structures suitable for this kind of analysis, including monomeric, dimeric and oligomeric proteins. It therefore allows a study of the hydration structure of various proteins and their complexes.

Reference List

1. Kaplan, I. G. *Intermolecular interactions;* Wiley: 2006.
2. Forner, W.; Cizek, J.; Otto, P.; Ladik, J.; Steinborn, E. O. Coupled-Cluster Studies .1. Application to Small Molecules, Basis Set Dependences. *Chemical Physics* **1985,** *97* (2-3), 235-249.
3. Forner, W.; Ladik, J.; Otto, P.; Cizek, J. Coupled-Cluster Studies .2. the Role of Localization in Correlation Calculations on Extended Systems. *Chemical Physics* **1985,** *97* (2-3), 251-262.
4. Burda, J. V.; Zahradnik, R.; Hobza, P.; Urban, M. Dimers of rare gas atoms: CCSD(T), CCSDT and FCI calculations on the (He)(2) dimer, CCSD(T) and CCSDT calculations on the (Ne)(2) dimer, and CCSD(T) all-electron and pseudopotential calculations on the dimers from (Ne)(2) through (Xe)(2). *Molecular Physics* **1996,** *89* (2), 425-432.
5. Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** *140* (4A), 1133-&.
6. Zupan, A.; Causa, M. Density-Functional Lcao Calculations for Solids - Comparison Among Hartree-Fock, Dft Local-Density Approximation, and Dft Generalized Gradient Approximation Structural-Properties. *International Journal of Quantum Chemistry* **1995,** *56* (4), 337-344.
7. Hua, X. L.; Chen, X. J.; Goddard, W. A. Generalized generalized gradient approximation: An improved density-functional theory for accurate orbital eigenvalues. *Physical Review B* **1997,** *55* (24), 16103-16109.
8. Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999,** *110* (11), 5029-5036.
9. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996,** *77* (18), 3865-3868.
10. Becke, A. D. Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98* (7), 5648-5652.
11. Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Physical Review Letters* **2003,** *91* (14).
12. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *Journal of Physical Chemistry* **1994,** *98* (45), 11623-11627.
13. Burke, K.; Ernzerhof, M.; Perdew, J. P. The adiabatic connection method: A non-empirical hybrid. *Chem. Phys. Lett.* **1997,** *265* (1-2), 115-120.
14. Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004,** *25* (12), 1463-1473.

15. von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Performance of optimized atom-centered potentials for weakly bonded systems using density functional theory. *Physical Review B* **2005,** *71* (19).
16. Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
17. Zhao, Y.; Truhlar, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2008,** *4* (11), 1849-1868.
18. Dion, M.; Rydberg, H.; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals density functional for general geometries. *Physical Review Letters* **2004,** *92* (24).
19. Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006,** *124* (3).
20. Bludsky, O.; Rubes, M.; Soldan, P.; Nachtigall, P. Investigation of the benzene-dimer potential energy surface: DFT/CCSD(T) correction scheme. *J. Chem. Phys.* **2008,** *128* (11).
21. Mashayak, S. Y.; Aluru, N. R. Coarse-Grained Potential Model for Structural Prediction of Confined Water. *Journal of Chemical Theory and Computation* **2012,** *8* (5), 1828-1840.
22. Xie, W. S.; Gao, J. L. Design of a next generation force field: The X-POL potential. *Journal of Chemical Theory and Computation* **2007,** *3* (6), 1890-1900.
23. Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrasek, J. Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated ab Initio Quantum Chemical and Empirical Potential Procedures. *Journal of Chemical Theory and Computation* **2009,** *5* (4).
24. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197.
25. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.
26. Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *Journal of Physical Chemistry A* **1999,** *103* (19), 3596-3607.

27. Jorgensen, W. L.; Tiradorives, J. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *Journal of the American Chemical Society* **1988,** *110* (6), 1657-1666.
28. Anfinsen, C. B. Principles That Govern Folding of Protein Chains. *Science* **1973,** *181* (4096), 223-230.
29. Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959,** *14*, 1-63.
30. Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins - 2 Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proceedings of the National Academy of Sciences of the United States of America* **1951,** *37* (4), 205-211.
31. Pauling, L.; Corey, R. B. Atomic Coordinates and Structure Factors for 2 Helical Configurations of Polypeptide Chains. *Proceedings of the National Academy of Sciences of the United States of America* **1951,** *37* (5), 235-240.
32. Pauling, L.; Corey, R. B. Configuration of Polypeptide Chains. *Nature* **1951,** *168* (4274), 550-551.
33. Chandler, D. The role of solvation dynamics in hydrophobic collapse. *Abstracts of Papers of the American Chemical Society* **2002,** *224*, U473.
34. Abseher, R.; Schreiber, H.; Steinhauser, O. The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins* **1996,** *25* (3), 366-378.
35. Chandler, D. Hydrophobicity: Two faces of water. *Nature* **2002,** *417* (6888), 491.
36. Steiner, S. A.; Hill, K. A.; Castellino, F. J. The Interaction Between Activated Bovine Protein-C and Metal-Ions. *Federation Proceedings* **1983,** *42* (7), 1860.
37. Williams, D. H.; Zhou, M.; Stephens, E. Ligand binding energy and enzyme efficiency from reductions in protein dynamics. *Journal of Molecular Biology* **2006,** *355* (4), 760-767.
38. Ding, X. M.; Pan, X. Y.; Xu, C.; Shen, H. B. Computational Prediction of DNA-Protein Interactions: A Review. *Current Computer-Aided Drug Design* **2010,** *6* (3), 197-206.
39. Mattaj, I. W. A Selective Review of Rna-Protein Interactions in Eukaryotes. *Molecular Biology Reports* **1990,** *14* (2-3), 151-155.
40. Derrigo, M.; Cestelli, A.; Savettieri, G.; Di Liegro, I. RNA-protein interactions in the control of stability and localization of messenger RNA (review). *International Journal of Molecular Medicine* **2000,** *5* (2), 111-123.
41. Rubes, M.; Bludsky, O.; Nachtigall, P. Investigation of the benzene-naphthalene and naphthalene-naphthalene potential energy surfaces: DFT/CCSD(T) correction scheme. *Chemphyschem* **2008,** *9* (12), 1702-1708.
42. Rubes, M.; Bludsky, O. DFT/CCSD(T) Investigation of the Interaction of Molecular Hydrogen with Carbon Nanostructures. *Chemphyschem* **2009,** *10* (11), 1868-1873.

43. Berggard, T.; Linse, S.; James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **2007,** *7* (16).
44. Burgoyne, N. J.; Jackson, R. M. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006,** *22* (11).
45. Ezkurdia, L.; Bartoli, L.; Fariselli, P.; Casadio, R.; Valencia, A.; Tress, M. L. Progress and challenges in predicting protein-protein interaction sites. *Briefings in Bioinformatics* **2009,** *10* (3).
46. Jones, S.; Thornton, J. M. Prediction of protein-protein interaction sites using patch analysis. *Journal of Molecular Biology* **1997,** *272* (1).
47. Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* **1999,** *285* (5).
48. Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic aspects of protein-protein interactions. *Current Opinion in Structural Biology* **2000,** *10* (2).
49. Shi, T. L.; Li, Y. X.; Cai, Y. D.; Chou, K. C. Computational methods for protein-protein interaction and their application. *Current Protein & Peptide Science* **2005,** *6* (5).
50. Shoemaker, B. A.; Panchenko, A. R. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput. Biol.* **2007,** *3* (3).
51. Shoemaker, B. A.; Panchenko, A. R. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput. Biol.* **2007,** *3* (4).
52. Zhou, H. X.; Qin, S. B. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics* **2007,** *23* (17).
53. Lijnzaad, P.; Argos, P. Hydrophobic patches on protein subunit interfaces: Characteristics and prediction. *Proteins* **1997,** *28* (3).
54. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins-Structure Function and Bioinformatics* **2007,** *68* (4).
55. Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990,** *29* (31).
56. Lee, L. P.; Tidor, B. Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. *Protein Science* **2001,** *10* (2).
57. Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Science* **1997,** *6* (1).
58. Levitt, M.; Park, B. H. Water - Now You See It, Now You Dont. *Structure* **1993,** *1* (4), 223-226.
59. Rupley, J. A.; Careri, G. Protein Hydration and Function. *Adv. Protein Chem.* **1991,** *41*, 37-172.
60. Purkiss, A.; Skoulaikis, S.; Goodfellow, J. M. The protein-solvent interface: a big splash. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* **2001,** *359* (1785), 1515-1527.
61. Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. Water in protein structure prediction. *Proceedings of the National*

Academy of Sciences of the United States of America **2004,** *101* (10), 3352-3357.

62. Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M.; Mansuy, D.; Stout, C. D. Structure of mammalian cytochrome P4502C5 complexed with diclofenac at 2.1 angstrom resolution: Evidence for an induced fit model of substrate binding. *Biochemistry* **2003,** *42* (31), 9335-9345.
63. Shaltiel, S.; Cox, S.; Taylor, S. S. Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (2), 484-491.
64. Janin, J. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure with Folding & Design* **1999,** *7* (12), R277-R279.
65. Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of protein-protein interfaces. *Proteins-Structure Function and Bioinformatics* **2005,** *60* (1), 36-45.
66. Rejto, P. A.; Verkhivker, G. M. Mean field analysis of FKBP12 complexes with FK506 and rapamycin: Implications for a role of crystallographic water molecules in molecular recognition and specificity. *Proteins* **1997,** *28* (3), 313-324.
67. Palomer, A.; Perez, J. J.; Navea, S.; Llorens, O.; Pascual, J.; Garcia, L.; Mauleon, D. Modeling cyclooxygenase inhibition. Implication of active site hydration on the selectivity of ketoprofen analogues. *Journal of Medicinal Chemistry* **2000,** *43* (11), 2280-2284.
68. Ni, H. H.; Sotriffer, C. A.; McCammon, J. A. Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: A molecular dynamics study. *Journal of Medicinal Chemistry* **2001,** *44* (19), 3043-3047.
69. Daniel, R. M.; Finney, J. L.; Stoneham, M. The molecular basis of life: is life possible without water? *Philosophical Transactions of the Royal Society B-Biological Sciences* **2004,** *359* (1448), 1143.
70. Finney, J. L. Water? What's so special about it? *Philosophical Transactions of the Royal Society B-Biological Sciences* **2004,** *359* (1448), 1145-1163.
71. Despa, F.; Fernandez, A.; Berry, R. S. Dielectric modulation of biological water. *Physical Review Letters* **2004,** *93* (22).
72. Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. Protein States and Protein Quakes. *Proceedings of the National Academy of Sciences of the United States of America* **1985,** *82* (15), 5000-5004.
73. Hayward, S.; Kitao, A.; Hirata, F.; Go, N. Effect of Solvent on Collective Motions in Globular Protein. *Journal of Molecular Biology* **1993,** *234* (4), 1207-1217.
74. Billeter, M. Hydration water molecules seen by NMR and by X-ray crystallography. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1995,** *27*, 635-645.

75. Blake, C. C. F.; Pulford, W. C. A.; Artymiuk, P. J. X-Ray Studies of Water in Crystals of Lysozyme. *Journal of Molecular Biology* **1983,** *167* (3), 693-723.
76. Ferrand, M.; Dianoux, A. J.; Petry, W.; Zaccai, G. Thermal Motions and Function of Bacteriorhodopsin in Purple Membranes - Effects of Temperature and Hydration Studied by Neutron-Scattering. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** *90* (20), 9668-9672.
77. Loris, R.; Stas, P. P. G.; Wyns, L. Conserved Waters in Legume Lectin Crystal-Structures - the Importance of Bound Water for the Sequence-Structure Relationship Within the Legume Lectin Family. *Journal of Biological Chemistry* **1994,** *269* (43), 26722-26733.
78. Nakasako, M. Structural characteristics in protein hydration investigated by cryogenic X-ray crystal structure analyses. *Journal of Biological Physics* **2002,** *28* (2), 129-137.
79. Otting, G. NMR studies of water bound to biological molecules. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1997,** *31*, 259-285.
80. Otting, G. NMR studies of water bound to biological molecules (vol 31, pg 259, 1997). *Progress in Nuclear Magnetic Resonance Spectroscopy* **1998,** *32*, 191.
81. Savage, H.; Wlodawer, A. Determination of Water-Structure Around Biomolecules Using X-Ray and Neutron-Diffraction Methods. *Methods in Enzymology* **1986,** *127*, 162-183.
82. Shou, J. J.; Wang, F.; Zeng, G. A.; Zhang, Y. H. Adsorption and Desorption Kinetics of Water in Lysozyme Crystal Investigated by Confocal Raman Spectroscopy. *J. Phys. Chem. B 115* (13), 3708-3712.
83. Syvitski, R. T.; Li, Y. M.; Auclair, K.; de Montellano, P. R. O.; La Mar, G. N. H-1 NMR detection of immobilized water molecules within a strong distal hydrogen-bonding network of substrate-bound human heme oxygenase-1. *Journal of the American Chemical Society* **2002,** *124* (48), 14296-14297.
84. Bui, H. H.; Schiewe, A. J.; Haworth, I. S. WATGEN: An algorithm for modeling water networks at protein-protein interfaces. *J. Comput. Chem.* **2007,** *28* (14), 2241-2251.
85. Friedman, R.; Nachliel, E.; Gutman, M. Molecular dynamics of a protein surface: Ion-residues interactions. *Biophys. J.* **2005,** *89* (2), 768-781.
86. Park, S.; Saven, J. G. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins-Structure Function and Bioinformatics* **2005,** *60* (3), 450-463.
87. Steinbach, P. J.; Brooks, B. R. Protein Hydration Elucidated by Molecular-Dynamics Simulation. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** *90* (19), 9135-9139.
88. Tsui, V.; Radhakrishnan, I.; Wright, P. E.; Case, D. A. NMR and molecular dynamics studies of the hydration of a zinc finger-DNA complex. *Journal of Molecular Biology* **2000,** *302* (5), 1101-1117.

89. Umezawa, K.; Higo, J.; Shimotakahara, S.; Shindo, H. Collective solvent flows around a protein investigated by molecular dynamics simulation. *J. Chem. Phys.* **2007,** *127* (4).
90. Virtanen, J. J.; Makowski, L.; Sosnick, T. R.; Freed, K. F. Modeling the Hydration Layer around Proteins: HyPred. *Biophys. J. 99* (5), 1611-1619.
91. Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical hydration shell around proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (52), 20749-20752.
92. Lu, Y. P.; Wang, R. X.; Yang, C. Y.; Wang, S. M. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem Inf. Model.* **2007,** *47* (2), 668-675.
93. Takano, K.; Yamagata, Y.; Funahashi, J.; Hioki, Y.; Kuramitsu, S.; Yutani, K. Contribution of intra- and intermolecular hydrogen bonds to the conformational stability of human lysozyme. *Biochemistry* **1999,** *38* (39), 12698-12708.
94. Sreenivasan, U.; Axelsen, P. H. Buried Water in Homologous Serine Proteases. *Biochemistry* **1992,** *31* (51), 12785-12791.
95. Nakasako, M. Large-scale networks of hydration water molecules around bovine beta-trypsin revealed by cryogenic X-ray crystal structure analysis. *Journal of Molecular Biology* **1999,** *289* (3), 547-564.
96. Higo, J.; Nakasako, M. Hydration structure of human lysozyme investigated by molecular dynamics simulation and cryogenic X-ray crystal structure analyses: On the correlation between crystal water sites, solvent density, and solvent dipole. *J. Comput. Chem.* **2002,** *23* (14), 1323-1336.
97. Yokomizo, T.; Higo, J.; Nakasako, M. Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme. *Chem. Phys. Lett.* **2005,** *410* (1-3), 31-35.
98. Komeiji, Y.; Uebayasi, M.; Someya, J.; Yamato, I. A Molecular-Dynamics Study of Solvent Behavior around a Protein. *Proteins* **1993,** *16* (3), 268-277.
99. Levitt, M.; Sharon, R. Accurate Simulation of Protein Dynamics in Solution. *Abstracts of Papers of the American Chemical Society* **1989,** *197*, 19-HYS.
100. Pettitt, B. M.; Makarov, V. A.; Andrews, B. K. Protein hydration density: theory, simulations and crystallography. *Current Opinion in Structural Biology* **1998,** *8* (2), 218-221.
101. Nakasako, M.; Fujisawa, T.; Adachi, S.; Kudo, T.; Higuchi, S. Large-scale domain movements and hydration structure changes in the active-site cleft of unligated glutamate dehydrogenase from Thermococcus profundus studied by cryogenic X-ray crystal structure analysis and small-angle X-ray scattering. *Biochemistry* **2001,** *40* (10), 3069-3079.
102. Neshich, G.; Togawa, R. C.; Mancini, A. L.; Kuser, P. R.; Yamagishi, M. E. B.; Pappas, G.; Torres, W. V.; Campos, T. F. E.; Ferreira, L. L.; Luna, F. M.; Oliveira, A. G.; Miura, R. T.; Inoue, M. K.; Horita, L. G.; de Souza, D. F.; Dominiquini, F.; Alvaro, A.; Lima, C. S.; Ogawa, F. O.; Gomes, G. B.;

Palandrani, J. F.; dos Santos, G. F.; de Freitas, E. M.; Mattiuz, A. R.; Costa, I. C.; de Almeida, C. L.; Souza, S.; Baudet, C.; Higa, R. H. STING Millennium: a web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Research* **2003,** *31* (13).

103. Rodionov, M. A.; Galaktionov, S. G. Analysis of the 3-Dimensional Structure of Proteins in Terms of Residue Residue Contact Matrices .1. the Contact Criterion. *Molecular Biology* **1992,** *26* (5), 773-776.
104. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26* (16), 1668-1688.
105. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16).
106. Bendova-Biedermannova, L.; Hobza, P.; Vondrasek, J. Identifying stabilizing key residues in proteins using interresidue interaction energy matrix. *Proteins-Structure Function and Bioinformatics* **2008,** *72* (1).
107. Kabsch, W. Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallographica Section A* **1976,** *32* (SEP1), 922-923.
108. Wallwork, S. C. Hydrogen-Bond Radii. *Acta Crystallographica* **1962,** *15* (JUL), 758-&.
109. Fraczkiewicz, R.; Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **1998,** *19* (3), 319-333.
110. Tkatchenko, A.; DiStasio, R. A.; Head-Gordon, M.; Scheffler, M. Dispersion-corrected Moller-Plesset second-order perturbation theory. *J. Chem. Phys.* **2009,** *131* (9).
111. Jenness, G. R.; Jordan, K. D. DF-DFT-SAPT Investigation of the Interaction of a Water Molecule to Coronene and Dodecabenzocoronene: Implications for the Water-Graphite Interaction. *Journal of Physical Chemistry C* **2009,** *113* (23), 10242-10248.
112. Chakarova-Kack, S. D.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. Application of van der Waals density functional to an extended system: Adsorption of benzene and naphthalene on graphite. *Physical Review Letters* **2006,** *96* (14).
113. Puzder, A.; Dion, M.; Langreth, D. C. Binding energies in benzene dimers: Nonlocal density functional calculations. *J. Chem. Phys.* **2006,** *124* (16).
114. Janowski, T.; Pulay, P. High accuracy benchmark calculations on the benzene dimer potential energy surface. *Chem. Phys. Lett.* **2007,** *447* (1-3), 27-32.
115. Shaw, C. G.; Fain, S. C.; Chinn, M. D.; Toney, M. F. Overlayer-Substrate Spacing for Argon and Krypton on Graphite Determined by Leed Intensity Analysis. *Surface Science* **1980,** *97* (1), 128-136.
116. Vidali, G.; Ihm, G.; Kim, H. Y.; Cole, M. W. Potentials of Physical Adsorption. *Surface Science Reports* **1991,** *12* (4), 133-181.

117. Mattera, L.; Rosatelli, F.; Salvo, C.; Tommasini, F.; Valbusa, U.; Vidali, G. Selective Adsorption of H-1(2) and H-2(2) on the (0001) Graphite Surface. *Surface Science* **1980,** *93* (2-3), 515-525.
118. Avgul, N. N.; Kiselev, A. V. Chemistry and Physics of Carbon 6, 1-124. 1970.
Ref Type: Generic
119. Bolina, A. S.; Brown, W. A. Studies of physisorbed ammonia overlayers adsorbed on graphite. *Surface Science* **2005,** *598* (1-3), 45-56.
120. Bolina, A. S.; Wolff, A. J.; Brown, W. A. Reflection absorption infrared spectroscopy and temperature-programmed desorption studies of the adsorption and desorption of amorphous and crystalline water on a graphite surface. *J. Phys. Chem. B* **2005,** *109* (35), 16836-16845.
121. Cabaleiro-Lago, E. M.; Carrazana-Garcia, J. A.; Rodriguez-Otero, J. Study of the interaction between water and hydrogen sulfide with polycyclic aromatic hydrocarbons. *J. Chem. Phys.* **2009,** *130* (23).
122. Rubes, M.; Nachtigall, P.; Vondrasek, J.; Bludsky, O. Structure and Stability of the Water-Graphite Complexes. *Journal of Physical Chemistry C* **2009,** *113* (19), 8412-8419.
123. Rubes, M.; Bludsky, O. Intermolecular pi-pi interactions in solids. *Physical Chemistry Chemical Physics* **2008,** *10* (19), 2611-2615.
124. Baskin, Y.; Meyer, L. Lattice Constants of Graphite at Low Temperatures. *Physical Review* **1955,** *100* (2), 544.
125. Riley, K. E.; Vondrasek, J.; Hobza, P. Performance of the DFT-D method, paired with the PCM implicit solvation model, for the computation of interaction energies of solvated complexes of biological interest. *Physical Chemistry Chemical Physics* **2007,** *9* (41), 5555-5560.
126. Vondrasek, J.; Kubar, T.; Jenney, F. E.; Adams, M. W. W.; Kozisek, M.; Cerny, J.; Sklenar, V.; Hobza, P. Dispersion interactions govern the strong thermal stability of a protein. *Chemistry-A European Journal* **2007,** *13* (32), 9022-9027.
127. Berka, K.; Laskowski, R. A.; Hobza, P.; Vondrasek, J. Energy Matrix of Structurally Important Side-Chain/Side-Chain Interactions in Proteins. *Journal of Chemical Theory and Computation 6* (7).

List of abbreviations

AMBER – Assisted Model Building and Energy Refinement
amide N – amide nitrogen
amide O – amide oxygen
amine N – amine nitrogen
AV5Z – Augmented Valence Pentuple Zeta basis set
AVDZ – Augmented Valence Double Zeta basis set
AVQZ – Augmented Valence Quadruple Zeta basis set
AVTZ – Augmented Valence Triple Zeta basis set
B3LYP – Becke 3-parameter Lee Yang Parr functional
B – Backbone cluster
BB – Backbone Backbone cluster
bb N – backbone nitrogen
bb O – backbone oxygen
BBB – Backbone Backbone Backbone cluster
BBC – Backbone Backbone Charged cluster
BBP – Backbone Backbone Polar cluster
BC – Backbone Charged cluster
BCC – Backbone Charged Charged cluster
BCP – Backbone Charged Polar cluster
BLYP – Becke Lee Yang Parr functional
BP – Backbone Polar cluster
BPP – Backbone Polar Polar cluster
BSSE – Basis Set Superposition Error
C – Charged cluster
carboxyl O – carboxyl oxygen
CBS – Complete Basis Set
CC – Charged Charged cluster
CC – Coupled Clusters
CCSD(T) – Coupled Clusters with Single, Double and Perturbative Triple Excitations
CCSDT – Coupled Clusters with Single, Double and Triple Excitations
CP – Charged Polar cluster
DF-DFT-SAPT – density-fitting density-functional-theory symmetry-adapted perturbation-theory
DFT – Density Functional Theory
DFT/CC – Density Functional Theory/Coupled Clusters method
DFT-D – Density Functional Theory with Empirical Dispersion
DNA – Deoxyribonucleic Acid
eV - electronvolt
ff03 – amber 03 force field
GGA – General Gradient Approximation
GROMOS – GROningen MOlecular Simulation package
HF – Hartree Fock method
hydroxyl O – hydroxyl oxygen

CHARMM – Chemistry and HARvard Molecular Mechanics
IEM – Interaction Energy Matrix
LDA – Local Density Functional
MP – Møller-Plesset Method
MP2 – Second-order Møller-Plesset Method
MP4 – Fourth-order Møller-Plesset Method
N – Nonpolar cluster
NOPC – Non-obligate Protein Complexes
OPC – Obligate Protein Complexes
OPLS – Optimized Potential for Liquid Simulations
P – Polar cluster
PAW – Projector Augmented Wave
PBE – Perdew Burke Ernzerhoff potential
PDB – Protein Data Bank
PP – Polar Polar cluster
PPI – Protein-Protein Interactions
PW91 – Perdew Wang functional from 1991
RIE – Residue Interaction Energy
RNA – Ribonucleic acid
RP-RKHS – Reciprocal Power Reproducing Kernel Hilbert Space Interpolation
SAS – Solvent Accessible Surface
TPSS – Tao Perdew Staroverov Scuseria functional
VASP – Viena *Ab initio* Simulation Package
wdg – water density grid
wdgp – water density grid point
vdW-DF – van der Waals density functional

List of figures

List of tables