International Journal of Terraspace Science and Engineering 6(1)2014 7-16

Protein structure and dynamics

Gábor Náray-Szabó[a,*], András Perczel [b]

[a] *Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary*

[b] *Protein Modeling Group HAS-ELTE, Institute of Chemistry, Eötvös Loránd University, H–1538, Budapest, P.O.B. 32, Hungary*

Abstract

Proteins are essential components of biological processes, this explains why understanding their structure, function and dynamics is so important. In the following, we give an overview on various methods for the determination of three-dimensional structure and dynamics of proteins. We discuss the most important experimental methods, X-ray diffraction and NMR spectroscopy, as well as computer modelling techniques and their application to the construction of graphics models, which can be inspected visually. We also treat prediction as well as molecular graphics representation of protein structures. We devote special attention to dynamics, where time scales of protein movement, structures and interactions are discussed. We wish to demonstrate that protein structure determination and computer representation is now at a very high degree of sophistication and reliability.

Keywords**:** *Protein structure;Dynamics; X-ray crystallography;Nuclear magnetic resonance (NMR);Molecular mechanics*

1. Introduction

The primary structure (amino acid sequence) information content of a typical protein is coupled to the encoding genes, which are part of the chromosomes. *Via* various ribonucleic acids, by using the complex machinery of the ribosome, the sequential information comes alive in the form of various polypeptides and proteins. Most probably among biomolecules, proteins have the most versatile nature, fulfilling a very broad role in cellular life. Recent advances in X-ray and neutron diffraction as well as nuclear magnetic resonance (NMR) spectroscopy methods have opened the way to describe and understand structures and events at the molecular level. Furthermore, molecules can now be described by means of their shape, internal motions and interactions. Beside their three-dimensional structures, their dynamics and interaction profiles with additional features are required to truly understand the way how the coupled biological information (bioactivity) is handled. The present review is not an original research paper, not even a thorough review, rather an ensemble of selected ideas and concepts related to molecular modeling, bioNMR spectroscopic techniques, X-ray crystallography, to pin down some mechanical and dynamic aspects of proteins. These macromolecules present a broad time scale of internal dynamics (motion), from picoseconds to minutes or even hours and years. In the following, a brief introduction of protein motion will be followed by the description of the different experimental techniques used to capture the required information. For a more detailed review of protein structure and dynamics see [1].

.

2. Structure Determination

Various *in vitro*, *in vivo* and *in silico* techniques are available for the determination of the 3D structure of proteins, which differ in their fundamentals, speed resolution and performance. However, none of the available techniques offers a unique tool to determine the structure and dynamics simultaneously. On the contrary, a modern approach to structural biology utilizes as many methods as possible to decipher a convergent molecular picture. 3D-structures are determined commonly by X-ray and neutron-diffraction methods [2] or NMR spectroscopy [3]. Low resolution structures can be obtained by cryo-electron microscopy (Cryo-EM) [4 and references therein] or small-angle X-ray scattering (SAXS) [5]. A global picture on shape and fold of proteins may also be determined by electronic circular dichroism (ECD) [6] and fluorescence spectroscopy [7]. In all cases, computer assisted data manipulation is required but, in addition, computer modeling and bioinformatics methods help research. In order to get a reliable, broad and good enough picture on structure and dynamics of proteins, a careful evaluation of the experimental data is needed by typically using a variety of techniques. The first protein structures were determined by X-ray diffraction in the early 1950s and until now over 80,000 structures were deposited in the

*Corresponding author: Gábor Náray-Szabó
E-mail address: narayszabo@yahoo.com

Protein Data Bank [8]. Structure determination by NMR spectroscopy is also possible, but the number of deposited structures to date does not exceed 10,000 entries.

2.1 X-ray crystallography

The major drawback of protein crystallography is that it needs relatively large single crystals of the target protein, presenting a task often tedious and hard to fulfill. Once a suitable crystal appropriate in size and quality is found, it is irradiated by X-rays and the obtained diffraction pattern is detected and subsequently analyzed. The resulting electron density map has peaks, indicating where nuclei are positioned. By using the primary sequence information of proteins *i.e.* their chemical constitution and molecular topology, an electron density map can be evaluated and refined. At the end of the iterative procedure, a 3D representation of the protein molecule is obtained and visualized by various techniques (see Fig. 1). The following steps have to be followed [2].

i) *Sample purification and crystallization.* This step consists of purification and homogenization of proteins to select conditions to grow a well-ordered crystal [9]. Nowadays, different robotics help the scientist to find the optimum conditions (e.g. salt concentration, pH optimum, precipitants) as fast as possible [10].

ii) *Data collection.* With a suitable crystal available, the symmetry, unit cell and resolution limits obtained from the X-ray diffraction pattern are to be recorded, either in-house or at a suitable synchrotron radiation source. Time-resolved crystallography is also available if critical experimental and reaction conditions are met [11].

iii) *Structure solution.* Bragg's law, the basis of X-ray diffraction, provides the relationship between the scattering angle and the distance between planes passing through the atoms in the crystal. The inverse Fourier transform of the diffraction pattern determined from the amplitude and the phase of the diffracted waves, provides the electron density of the protein.

iv) *Model building.* In an ideal case (perfectly ordered crystal), the electron density map has peaks at each of the atomic locations, thus Cartesian coordinates of the atoms are obtained. In a non-idealistic case, a model is needed, which can be fitted to the electron density map. Typically, the backbone atoms of the protein are fitted first (see the pink ribbon of Fig 1), this is followed by those of the side-chains of the residues. X-ray methods are called blind for highly mobile regions of the structures, such as loops or terminal chains, since the corresponding electron density is smeared and therefore does not provide peaks in the diffraction pattern.

v) *Refinement and validation.* The resulting atomic resolution structure is improved in an iterative manner when the quality of the atomic model is judged on the basis of the *R*-factor. This is the average fractional error of the calculated amplitudes in relation to their experimental counterparts. Final structures are validated by various bioinformatics methods.

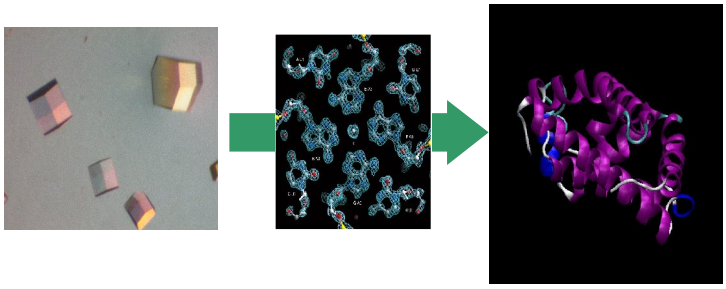

Fig. 1 The single crystal (left) provides an atomic resolution electron density map (center), which can be turned into the molecular modeling representation of the 3D-structure of the protein (right).

Neutron diffraction is a method requires high thermal-neutron fluxes obtained from nuclear reactors and provides special information on proteins [12]. Hydrogen atoms can be precisely located, which is almost impossible by X-ray diffraction. A diffraction experiment can be performed on a crystal; the results can be evaluated similarly, as done for the X-ray technique.

2.2 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is an important method of modern structural biology, allowing to determine three-dimensional protein structures in solution, and even in case of some proteins for which X-ray diffraction does not provide enough result. It became an almost routine method for the structure determination of proteins up to about 30 kDa molecular weight. Structure determination by NMR is based on the following consecutive steps (for a simplified scheme see Fig. 2).

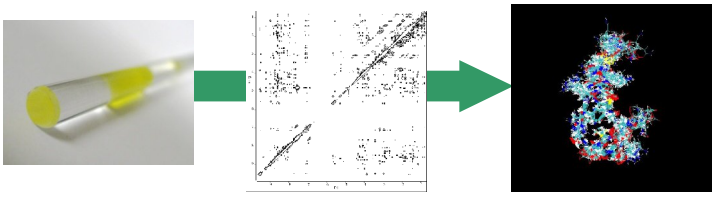

Fig. 2. The small protein in solution (left) provides a correlation NMR spectrum (center), which is analyzed to obtain a graphical representation of the ensemble of its atoms (right).

i) Sample preparation. For an NMR measurement, a protein solution of at least 95 % purity is needed, which is stable over a week and has an appropriate concentration (0.1–1 mM). The total sample volume

should vary between 350 and 550 μl, the total mass between 3 and 30 mg. 2,2-Dimethyl-2-Silapentane-5-Sulfonic acid (DSS) is used as a reference compound. Samples are expressed by an appropriate recombinant technique, which eliminates toxic effects due to the overproduction of recombinant proteins.

ii) Data acquisition [13]. In a magnetic field of at least 10 T some nuclei (e.g. ^{1}H, ^{13}C, or ^{15}N) absorb energy, which is manifested by the appearance of a characteristic resonance frequency, detected as a signal in the NMR spectrum. The resonance frequency is between 500 and 950 MHz for protons. It refers to the energy of absorption and it varies linearly with the strength of the static magnetic field. Due to the variability of the local chemical environment, nuclei of the same type in a molecule resonate at slightly different frequencies. The deviation from the reference resonance frequency is the chemical shift. Proteins from natural sources, the utilized resonances are mostly from protons. Further NMR active nuclei (typically ^{13}C and ^{15}N) may be inserted in the molecule in order to collect more information.

In the ideal case, each magnetically active atom has its characteristic chemical shift, allowing to be recognized and assigned. Assignment of the experimental chemical shifts to atoms or groups, like a methyl group, is the next step of structure determination. This is a very complicated procedure but it provides useful information on protein structure and dynamics. Beside the most common dimension typically associated with protons, additional dimensions can be generated to minimize overlap of the spectra. For structure determination, two methods are in everyday use. One is based on coherence transfer between chemically linked atoms (through bond coupling), another on magnetization transfer between atoms lying close in space (through space coupling).

iii) Resonance frequency assignment. The basis of the analysis of spectra is to distinguish between real and artificial (background) signals. Then, chemical shifts should be assigned to various nuclei of the macromolecule. Several protocols are available to achieve this goal, mainly depending on the method of isotope labeling of proteins. Assignment for proteins of molecular weights between 5 and 15 kDa relies on ^{15}N-labeled samples. Larger proteins require labeling of two or three types of magnetically active nuclei, which allows resonance assignment for both backbone and side-chain atoms. In principle, the full assignment procedure can be automated, by *e.g.* the software DYNASSIGN [14].

iv) Restraint collection. Structure determination is based on various structural restraints. Distance restraints may be typically obtained from Nuclear Overhauser Effect (NOE) experiments where cross peaks related to magnetization transfer between spatially close nuclei are detected. Each of the assigned cross peaks can be converted into a corresponding internuclear distance, which cannot be determined precisely. They are classified as short, medium or large distances. Restraints on the backbone and side-chain dihedral torsion angles are obtained by detecting conformation sensitive indirect coupling constants. Various methods are available for the accurate measurement of these, which can be converted into restraints of the dihedral angles. The coupling is used in solid state NMR and provides information on the orientation of bonds. Hydrogen bonds can also be detected by a hydrogen/deuterium (H/D) exchange experiment. The exchange can be monitored on the basis of the deuterium isotope effect on the chemical bond strength. The slower the proton of the NH bond is replaced by deuterium, the more the amide group is buried within the protein. Additionally, for smaller proteins, the temperature dependence of NH chemical shifts can be used to determine the involvement of the NH bond in a hydrogen bond. Order parameters, introduced as restraints, provide information on the dynamic structure of proteins.

v) Structure determination and refinement. Experimentally obtained restraints are used as input for the elucidation of the structure. Computer programs, like XPLOR [15], use as many restraints as possible and combine them with structural properties of proteins. The different structural restraints are converted into target functions describing energy terms, which have to be minimized simultaneously. A manifold of structures is thus obtained reflecting a molecular fold. This procedure is at present a time consuming iterative process requiring well trained specialists although the lengthiest procedures are the chemical shift and the NOE assignments. Efforts are made to fully automate assignment and integrate it with structure calculations [16].

vi) Structure validation. This indispensable step can be done by using a special software (*e.g.* CheckShift [17]) to check both the quality of the experimental data and structural information. Validation refers to both geometry and restraint violations of a given protein structure. The output includes a statistics and visualizations on restraint violations, furthermore the agreement with deposited models and the quality of geometrical parameters.

NMR is a versatile tool for protein structure determination, however, in some cases it has severe limitations, among others spectral crowdedness or too fast relaxation. Traditionally, it has been used only for small or medium size proteins. Problems originate from the limited spectral resolution as well as signal overlap. Multidimensional NMR spectroscopy combined with different isotope labeling schemes as well as increase of the applied static magnetic field

helped considerably in enhancing its power. Another limitation is connected to spin relaxation since magnetization decays rapidly for larger macromolecules. This means that there is less time to measure NMR signals leading to information loss. Techniques have been introduced to overcome this problem and technical innovations allow to determine structures of larger water soluble and membrane proteins, the latter being studied very difficultly by other methods [18].

2.3 Computer modeling

Modern computer technology allows to model proteins at the atomic resolution. Several methods are known for structure determination, interpretation of structure-function relationships, as well as construction of protein models from smaller fragments.

2.3.1 Molecular mechanics

Since proteins contain several thousand atoms, explicit calculations on them is very laborious. Instead of using *a priori* quantum mechanical approaches, the so-called molecular mechanics (MM) methods are often applied, where a molecular system is handled as a collection of atoms, which are connected by bonds. Forces acting between atoms are described by simple quadratic or trigonometric terms, the former taken over from Hooke's law of vibrations. The approximate energy is a sum of interaction terms, describing different types of strain within the molecule. The total energy is calculated as the sum of the following terms:

$$E_{MM} = V_{stretch} + V_{bend} + V_{torsion} + V_{outofplane} + V_{nonbonded}$$

where subscripts refer to stretching, bending, torsion, out-of-plane bending, and non-bonded terms, respectively. For a review and comparison of parametrization schemes see *e.g.* [19].

Parameters in the energy expression are fitted either to experimental or calculated data, their total number may be quite large. Since the number of atom pairs in $V_{nonbonded}$ increases quadratically, most methods use a cutoff to reduce computation efforts, beyond which the interactions are set to zero. A certain set of parameters is valid only for a given force field therefore a parameter cannot be transferred from one method to the other.

Two types of parametrization are used. Class I force fields *e.g.* [20] work with a simpler energy expression and use experimental data. They are applied to proteins, nucleic acids, carbohydrates and their associations. Class II force fields *e.g.* [21] include higher order and cross terms, too. They are calibrated to quantum mechanically calculated quantities, which increases their transferability and reliability. Several freely or commercially available software packages are available, like INSIGHT II (http://accelrys.com/), SYBYL (http://www.tripos.com/), CHARMM (http://www.charmm.org) or SCHRODINGER (http://www.schrodinger.com/).

2.3.2 Structure prediction

Protein structures are very complicated, but fortunately, several methods are available for their prediction, see *e.g.* [22, 23].

Homology modeling. Proteins are called homologous if their primary sequence overlaps to some extent. It has been observed that such an overlap means also the partial conservation of three-dimensional structures. Based on homologue sequences, which therefore possess some similarity, a model of an unknown protein can be constructed. The modeling procedure is described in the following.

Identification of templates is based on sequence alignments. After the template was selected on the basis of the similarity with the target protein, it should be improved by placing missing residues and correcting improper bond lengths. Then, it has to be mutated to fit the corresponding residues of the target protein. Furthermore, the gaps in the target sequence have to be excised; loops should be built and inserted by selecting anchoring residues. At last, a molecular mechanics method is used to optimize the generated model.

Assessment of homology models can be performed either with statistical potentials or energy calculations. Both methods calculate an energy estimate for the model. Unfortunately, neither of the two methods correlates very well with the accuracy of the structure. Statistical potentials have computational advantages as they can be constructed using proteins of known structure. For the energy calculations see Sec. 2.3.1. The above methods are based on the supposition that the native structure of a protein refers to the minimum energy, *i.e.* the lower is the energy the closer is the structure to the true one. The quality of homology models becomes lower with decreasing sequence identity. Several computer programs are available for homology modeling, see *e.g.* [24].

Protein threading (fold recognition) makes use of the observation that the number of different protein folds is about one thousand. 90% of the new structures, recently submitted to the Protein Data Bank, have similar folds to those already stored. The method uses sequences with the same fold as a known structure, which are, however, not homologous with any of them. Prediction is based on a statistical relationship between structure and sequence. The prediction uses alignment of each amino acid in the target sequence to a respective position in the template. The target fit to the template is evaluated and the best-fit one is selected. The structural model of the sequence is

constructed on the basis of the alignment with the selected template.

Ab initio methods. If structural analogs are not available, other information is needed for the prediction. Various potentials and other information may be used in constructing spatial restraints, as well as identifying local structural building blocks. A well-known *ab initio* method is ROSETTA [25]. It uses conformations determined by the protein backbone and C^{β} atoms of side-chains and are found for each short sequence segment in known protein structures. These are used to approximate the set of local conformers in the modeled protein having the same sequence. The combination of local conformers with the lowest overall energy is considered as the best model. Despite some spectacular successes, the very large computer time required is still prohibitive for a routine use.

3. Structure representation

Protein structures contain typically thousands of atoms, thus their graphical representation is a key factor for the understanding of their structure and function. Modern computer technology allows to get insight into structural and functional details.

3.1 Molecular graphics

Three-dimensional structure of a protein may be displayed on the computer screen in various representations [26]. In the simplest, the molecule is considered as a chemical graph, in which vertices and edges represent atoms and bonds, respectively. This representation is not very useful for proteins, since the picture becomes too crowded. However, it is often used for displaying structures derived from X-ray data. In most cases, evaluation of NMR spectroscopic data leads to several structures, which may be superimposed. Ten to twenty structures fitting best to restraints may be extracted and, for simplification, a single representation adopting the characteristic conformation is shown.

The ball-and-stick model displays atoms as balls, bonds as sticks connecting them. In the space-filling representation, atoms are represented by spheres with van der Waals radii in order to give an impression of the amount of space they occupy. For proteins, ribbon diagrams offer a versatile means to visualize special features of the structure. Here, α-helices are shown as ribbons, β-strands as arrows, random coils as thin tubes (see Fig. 3). The ribbon representation allows recognition of structural motifs while the space-filling model provides information on *e.g.* location and shape of crevices. By molecular graphics, it is possible to generate surfaces around the molecule, on which its properties can be displayed. For proteins the electrostatic potential (see Sec. 3.2), water accessibility and hydrophobicity patterns are the most important.

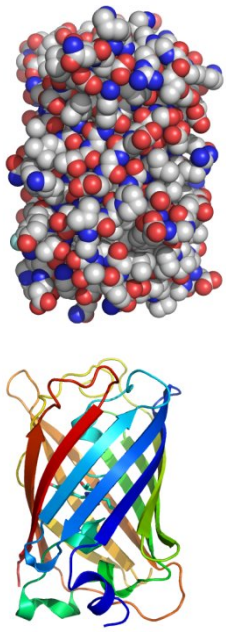

Fig. 3. Space-filling (left) and ribbon (right) representation of the green fluorescent protein (PDB code: 1RRX).

3.2 Electrostatics

Proteins have numerous polar or charged side-chains on their surface, therefore electrostatics is a simple, yet reliable tool to study their properties [27]. Electrostatic interactions play a role in protonation, ligand binding, enzyme catalysis, redox processes, even in photosynthesis.

The electrostatic potential generated by a protein molecule in a given point, $\boldsymbol{r}_i$, is given as follows

$$V(\boldsymbol{r}_i) = \int \rho(\boldsymbol{r}) / | \boldsymbol{r} - \boldsymbol{r}_i | \mathrm{d}\boldsymbol{r} + \Sigma \boldsymbol{Z}_a / | \boldsymbol{r}_i - \boldsymbol{R}_{ia} |$$

where ρ is the electron density generated by the electrons and nuclei of the protein, Z_a is the nuclear charge of an atom located at $\boldsymbol{R}_{ia}$. Calculation of the electron density needs the quantum mechanically determined total wave function. Its exact determination for proteins would require astronomical computer time. Fortunately, the density can be approximated by a sum of transferable molecular fragments, atomic monopoles or multipoles.

The above equation refers to a molecule *in vacuo*, a realistic model should include some treatment of the surrounding biophase. This goal is fulfilled in the Poisson-Boltzmann approach, where the ions near the protein are represented by an average field. A linearized equation can be derived, combining the electron density of the protein and the charge represented by the ions

$$\nabla[\varepsilon(\boldsymbol{r})\nabla V(\boldsymbol{r})] + 4\pi[-2IV(\boldsymbol{r})/kT + \rho(\boldsymbol{r})] = 0$$

$V(\boldsymbol{r})$, $\varepsilon(\boldsymbol{r})$ and $\rho(\boldsymbol{r})$ are the electrostatic potential, the dielectric constant and the protein electron density in point $\boldsymbol{r}$, respectively. I denotes the ionic strength. DelPhi, a software package based on the solution of the Poisson-Boltzmann equation, is available for calculating and visualizing the protein electrostatic potential [28].

The electrostatic potential is useful in describing the binding of small ligands to proteins. Molecular recognition by proteins is also partly determined by electrostatics.

4. Dynamics

It may be misleading to use a rigid representation of a protein, which is in most cases flexible to a given extent. However, the need to provide simple three-dimensional models of proteins often leads to the neglect of their inherent dynamical properties. The static picture fails to explain several phenomena, *e.g.* it is not fully appropriate to treat ligand binding. Below we give an account on protein dynamics occurring on different timescales, methods of including them in protein representation and on the relevance of dynamics to molecular biology. A graphical representation shows a correlation of various molecular events with different time scales (Fig. 4.).

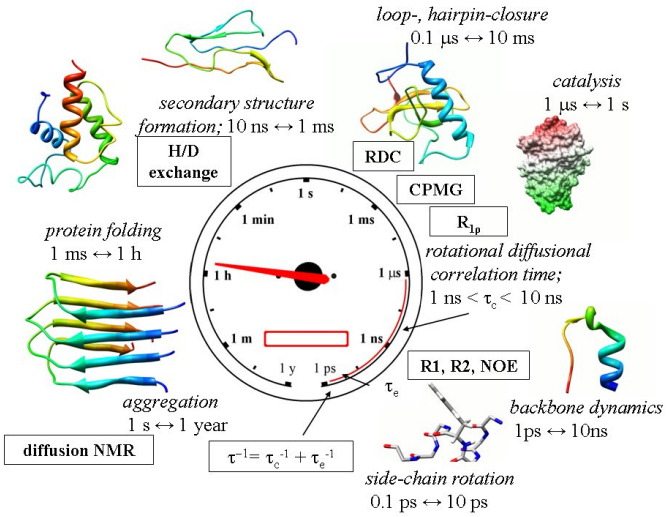

Fig. 4. The time scale for events related to proteins. NMR-based techniques appropriate for studying dynamics at different time scales are shown in boxes.

4.1 Time scales

Proteins undergo several types of time-dependent events. For example, folding occurs on the millisecond time scale, the half-life time of a protein varies from a few minutes to several weeks. In general, regarding protein dynamics, we consider much faster events, like secondary structure formation, which needs 10 ns to 1 ms, or opening and closing of loops, being on the 10 μs to 1 ms time scale.

The inverse of the rotational diffusion correlation time, τ_C^{-1}, is the average frequency of molecular tumbling in solution. The *correlation time* of a protein is the time that is needed to reorient by 1 radian through tumbling in solution. For proteins in aqueous solution at normal conditions, τ_C varies between 1 and 10 ns. Further types of motions are protein backbone (time span: 1 ps to 10 ns) and protein side-chain (time span: 0.1 to 10 ps) dynamics. Although even faster events (*e.g.* bond vibrations) may occur in proteins, they provide less information on the molecule as a whole.

NMR spectroscopy is an appropriate tool to study the dynamics of proteins since relaxation times and other properties are sensitive to side-chain motions. In case of proteins in solution, relaxation of the magnetically exited states takes place as a result of various segmental motions. Since more than a single relaxation process characterizes the correlation function $G(\tau)$, which often spans a large time scale, it is advised to observe several different relaxation parameters at various magnetic fields and temperatures.

Routine work with ^{15}N and ^{13}C labeled proteins allows estimation of the degree of main-chain mobility. Coupled primarily to fast motion of the protein, heteronuclear NMR relaxation of backbone ^{15}N atoms is interpreted by supposing that global and local motions are fully uncorrelated. Local N–H motions are quantified by the generalized order parameter, S^2, ranging between zero and unity, which reflects the amplitude of N–H fluctuations [29]. A graphical representation of the backbone dynamics is offered by mapping S^2 values on the protein surface (*cf.* Fig. 5). It may be understood that movement of the central Trp residue, within the hydrophobic cluster of the protein, is most restricted.

Recent molecular dynamics techniques allow consideration of mobility and NOE restraints jointly [30]. As a result, a more realistic representation is possible in solution than by conventional NMR methods. The dynamically restrained ensembles occupy a much larger conformational space than the conventionally calculated ones, allowing to reproduce independent NMR parameters much better.

The model-free approach is not applicable for highly asymmetric and partially folded or fully unfolded proteins. These do not meet the criteria of the separation of internal and global motions. The reconstruction of spectral density components by specific measurements is possible but more experiments are required. IUPs (intrinsically

unstructured proteins) can also be analyzed in terms of raw data (see *e.g.* [31]).

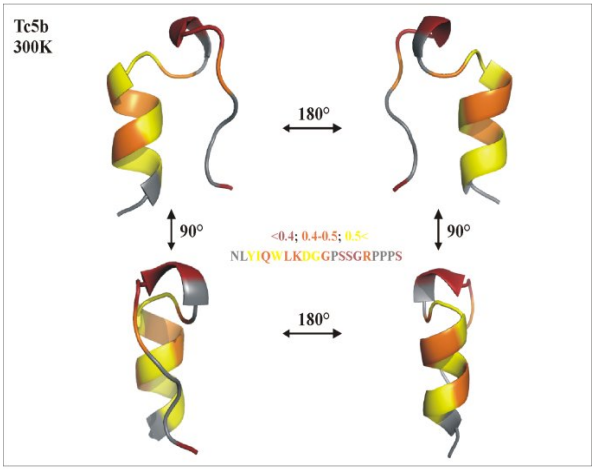

Fig. 5. S^2, mapped on the Tc5b miniprotein ribbon representation. Brown stands for high (<0.4), orange for medium (0.4-0.5), yellow for low (>0.5) backbone NH mobility (Pohl et al., unpublished).

Above, we discussed molecular motions ten or hundred times faster than the rotational diffusion correlation time. Slower motions (*e.g.* loop fluctuations, secondary structure reorientations) are also important. These are accessible to spin relaxation due to modulation of the isotropic chemical shift of ^{1}H, ^{13}C and ^{15}N nuclei coupled to chemical exchange. NMR spectroscopy can also provide information on this much slower motion. Appropriate techniques are sensitive to protein motions or chemical kinetic processes occurring at microsecond time scales [32]. Beside conformational events like protein folding, other biochemical processes (*e.g.* enzyme catalysis) also appear on this time scale. It seems that information on slower dynamics could be of great importance, especially if cooperative conformational transitions or enzyme dynamics are modeled.

4.2 Dynamical structures

Partially folded and molten globule states are of considerable interest and can be characterized by NMR and other spectroscopic techniques. The molten globule state of a protein is partially folded and has secondary structural elements forming a 3D-structure, which is similar to the native state. The main difference from the latter is the absence of some side-chain/side-chain interactions. This intermediate structure can be stabilized under selected conditions allowing to obtain structural information. Proteins in their molten globule form are characterized as ensembles of a set of various conformers [33]. The study of the molten globule states is often impossible by X-ray diffraction and may also be difficult if using NMR techniques. However, from simple line-shape and H/D exchange analysis by the latter technique, measurement of the hydrodynamic radius, deciphering stable secondary structural subunits by *e.g.* circular dichroism or fluorescence spectroscopy, these can be appropriately characterized [34].

5. Protein-protein interactions

Protein function is often exerted through interaction with each other, forming complexes *e.g.* in the ribosome, in cell membranes or during enzyme catalysis. Complex structures are crucial in understanding action, therefore beyond experimental techniques, *in silico* methods are also of great importance. Due to a number of practical difficulties, it seems unlikely that solving the structures of protein complexes will become routine in the near future. Hence, computational techniques play an important role [35].

The simplest methods are based on the complementarity principle focusing on the properties of the contact surface. Spatial, electrostatic and hydrophobic matching between interaction partners should be ensured to allow maximum interaction in the biophase [36]. A special problem is to identify the binding sites on the interacting proteins. In several cases (*e.g.* antibody-antigen or most enzyme-inhibitor interactions), these are known, however in others, they may be located on the basis of mutagenic or phylogenetic evidence.

It is a significant challenge to predict the binding sites of protein-protein interaction surfaces using computational techniques alone [37]. Machine learning techniques are used to develop automated protein-protein interface prediction methods. These are trained using *e.g.* buried surface areas, desolvation and electrostatic interaction energies, hydrophobicity and residue conservation scores.

For protein-protein docking two alternatives are at hand [38]. In the rigid-body approach interacting partners do not relax, their structures remain unchanged during complex formation. Quite often a conformational change within the partners accompanies the interaction. In such cases, a soft docking technique has to be applied, which means that conformational relaxation is allowed.

Docking involves two main steps. First, a set of reliable configurations has to be generated. Second, right configurations have to be distinguished from those that are not appropriate for complex formation. Docking algorithms start with a rigid body projection of the interacting proteins onto a three-dimensional grid. Then, grid cells will be distinguished according to whether they are near the protein surface or buried within the core. A search is performed by scoring the degree of overlap between pairs of grids in various

relative orientations. After excluding interactions on the basis of experimental evidence or steric conflicts, the structures will be sampled by scoring each configuration. In the geometric hashing approach, each protein surface is first pre-processed and a list of a few hundred critical points is obtained. These are then compared through a special algorithm to generate a relatively small number of trial orientations, which can be used for scoring [39].

An automated algorithm was developed for the identification of molecular surface complementarity [40]. It involves a digital representation of the molecules distinguishing between surface and interior. After calculating a correlation function, which assesses the degree of surface overlap and penetration resulting from the displacement, it scans relative orientations in three dimensions. The algorithm estimates the degree of spatial match between the surfaces of the interacting molecules. The procedure is the same as a six-dimensional search but it is much faster as the computation time does not increase rapidly with molecular size. Such methods are used to evaluate very many of configurations. This advantage is lost in case if conformational changes are also considered. It is possible to construct reasonable scoring functions combining requirements for steric and electrostatic fit.

Conformational changes can also be considered in Monte Carlo methods. An initial configuration is refined on the basis of random steps, which are accepted or rejected on the basis of the improvement in score [41]. It is assumed that convergence to the best structure may be reached from a large number of initial configurations, of which only one will be finally accepted. Because of the difficulty of finding an appropriate scoring function, which is highly discriminating and converges to the correct configuration, refinements have been proposed. Monte Carlo methods do not necessarily lead to an exhaustive search, thus the best configuration may be missed even using an appropriate scoring function.

To find the proper scoring function, which allows selecting the right configuration, a benchmark of protein-protein interactions is needed. Such a benchmark with several dozens of known protein-protein complexes was developed for testing docking methods [42]. The set covers a wide range of enzyme-inhibitor, antigen-antibody and other interaction types in order to avoid repetition. The best scoring functions are selected, and scores are defined on the basis of residue contacts, shape complementarity, interaction energies or other considerations. Hybrid scores may also be created by combining one or more categories. Although solvation effects are very important in complex formation, most docking algorithms do not take these into account.

Interactions between proteins have two contradicting aspects. First, partners should have complementary shapes, which are considerably rigid over time. Second, they must possess an inherent flexibility and plasticity otherwise several phenomena cannot be understood. Thus, the protein should remain rigid with a well-defined shape to enhance selectivity but it should be able to change its shape and conformation dynamically upon binding.

Although the classical lock-and-key model has been refined, further modifications are needed in case of canonical protease inhibitors. Proteases cut other proteins into pieces in order to help digestion, activation, degradation and other processes. Accordingly, they should be under severe control often ensured by fine tuned canonical inhibitors. Upon binding, these may not influence conformational changes near the binding site since they bind in a substrate-like manner to inhibit activity. Recently, backbone dynamics of some inhibitors was measured and it was found that, in contrast to the expectations, they are quite flexible. In fact, these disulfide bonds stabilized proteins are quite mobile, like their protease binding region. Thus, in this case, the classical lock-and-key model does not work perfectly. It was shown that the interaction cannot be properly described without taking into account the internal mobility of the partner proteins. The presence of flexible rather than rigid interacting partners allows them to avoid having an entirely complementary shape. In fact, partners can gently adapt to each other and they slightly differ from case to case [43]. Flexibility of partners was found to be synchronized and also encoded [44]. This might mean that the lock-and-key concept widely used earlier should be reinterpreted and replaced by the dynamical hand-and-a-glove analogy [45].

6. Conclusions

Protein dynamics involves very broad time scales falling within the ps and ks range. These scales refer to phenomena, which, in most cases, should be treated separately; however, there exist techniques, like Monte Carlo or some special statistical mechanics or molecular dynamics methods, which allow a simultaneous treatment. Accordingly, protein dynamics involves a way of thinking, which is similar to that followed in mesomechanics, considering the simultaneous interaction of different scale levels and attempts to bridge the gap in the time scale. Thus, the introduction of principles of physical mesomechanics to protein dynamics may provide insight into the mechanism interrelating structure and function. On the other hand, techniques, which are appropriate for the treatment of dynamic effects in proteins, may also be applicable in mesomechanics [46].

References

[1] G. Náray-Szabó, A. Perczel, and A. Láng, Protein modeling in, J. Leszczynski (ed.), Handbook of Computational Chemistry, Berlin-Heidelberg: Springer-Verlag (2011), p. 1096.

[2] C. Branden, and J. Tooze, *Introduction to Protein Structure*. New York: Garland (1994).

[3] J. Keeler, Understanding NMR spectroscopy, New York: Wiley (2002).

[4] X. Zhang, Z.H. Zhou, Limiting Factors In Atomic Resolution Cryo Electron Microscopy: No Simple Tricks. *J. Struct. Biol.* 175 (2011) 253.

[5] H.D. Mertens, D.I. Svergun, Structural Characterization Of Proteins And Complexes Using Small-Angle X-Ray Solution Scattering. *J. Struct. Biol.* 172 (2010) 128.

[6] S.M. Kelly, T.J., Jess, and N.C. Price, How To Study Proteins By Circular Dichroism. *Biochim Biophys Acta.* 1751 (2005) 119.

[7] D.M. Luykx, S.S. Goerdayal, P.J. Dingemanse, W. Jiskoot, P.M. Jongen, Hplc And Tandem Detection To Monitor Conformational Properties Of Biopharmaceuticals. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 821 (2005) 45.

[8] Protein Data Bank, PDB Current Holdings Breakdown, http://www.rcsb.org/pdb/statistics/holdings.do [retrieved on April 20, 2012.

[9] R. Horsefield, and R. Neutze, Crystallization of Lysozyme by the Hanging Drop Method, http://www.csb.gu.se/rob/PDFs/Crystallisation_Course_2006_Part-I.pdf, [retrieved on April 20, 2009

[10] J.R. Luft, R.J. Collins, N.A. Fehrman, A.M. Lauricella, C.K. Veatch, and G.T. DeTitta, A Deliberate Approach to Screening for Initial Crystallization Conditions of Biological Macromolecules. *J. Struct. Biol.* 142 (2003) 170.

[11] D. Bourgeois, and A. Royant, Advances in Kinetic Protein Crystallography. *Curr. Opin. Struct. Biol*. 15 (2005) 538.

[12] J. Fitter, T. Gutberlet, and J. Katsaras, (eds.), Neutron Scattering in Biology, Techniques and Applications, Berlin: Springer (2006).

[13] C. Schorn and B.J. Taylor, NMR-Spectroscopy: Data Acquisition, 2nd ed., New York: Wiley-VCH (2004).

[14] R. Schmucki, S. Yokoyama, and P. Güntert, Automated Assignment of NMR Chemical Shifts Using Peak-Particle Dynamics Simulation with the DYNASSIGN Algorithm. *J. Biomol. NMR* 43 (2009) 97.

[15] C.D. Schwieters, J.J. Kuszewski, and G.M. Clore, Using Xplor-NIH for NMR Molecular Structure Determination. *Prog. Nucl. Magnet. Reson. Spectr.* 48 (2006) 47.

[16] J. Volk, T. Herrmann, and K. Wüthrich, Automated Sequence-Specific Protein NMR Assignment Using the Memetic Algorithm MATCH. *J. Biomol. NMR* 41 (2008) 127.

[17] S.W. Ginzinger, F. Gerick, M. Coles, and V. Heun, CheckShift: Automatic Correction of Inconsistent Chemical Shift Referencing. *J. Biomol. NMR* 39 (2007) 223.

[18] K. Wüthrich, NMR Studies of Structure and Function of Biological Macromolecules. http://nobelprize.org/nobel_prizes/chemistry/laureates/2002/wutrich-lecture.pdf, [retrieved on August 6, 2009].

[19] Z. Hu, and J. Jiang, Assessment of Biomolecular Force Fields for Molecular Dynamics Simulations in A Protein Crystal. *J. Comput. Chem.* 31 (2010) 371.

[20] GROMOS (2011) Dynamic Modelling of Molecular Systems, https://www1.ethz.ch/igc/ GROMOS/ [retrieved on March 10, 2011].

[21] T.A. Halgren, Merck Molecular Force Field. I-V. *J. Comp. Chem*. 17 (1996) 490.

[22] J.M. Bujnicki, Prediction of Protein Structures, Functions, and Interactions, Chichester: Wiley-Blackwell (2009).

[23] E. Krieger, S.B. Nabuurs, and G. Vriend, Homology Modeling, *Methods Biochem. Anal.* 44 (2003) 509.

[24] Y. Xu, D. Xu, and J. Liang, Computational Methods for Protein Structure Prediction and Modeling, Vols. 1-2, New York: Springer (2007).

[25] R. Das, B. Qian, S. Raman, R. Vernon, J. Thompson, P. Bradley, S. Khare, M.D. Tyka, D. Bhat, D. Chivian, D.E. Kim, W.H. Sheffler, L. Malmström, A.M. Wollacott, C. Wang, I. Andre, and D. Baker, Structure Prediction for CASP7 Targets Using Extensive All-Atom Refinement with Rosetta@home. *Proteins: Struct. Funct. Bioinf.* 69, (Suppl 8) (2007) 118.

[26] A.M. Lesk, H.J. Bernstein, and F.C. Bernstein, Molecular Graphics in Structural Biology, in, Computational Structural Biology, Methods and Applications. M. Peitsch, and T. Schwede, (eds.), Singapore: World Scientific Publishing (2008) p. 729.

[27] N.A. Baker, and J.A. McCammon, Electrostatic Interactions, In Structural Bioinformatics, 2nd ed., J. Gu and P.E. Bourne (eds.), Chichester: Wiley-Blackwell (2009) p. 575.

[28]wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:DelPhi [retrieved on July 20, 2009].

[29] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, M. Rance, and N.J. Skelton, Protein NMR Spectroscopy, Principles and Practice, 2nd ed., Amsterdam: Elsevier (2007).

[30] B. Richter, J. Gsponer, P. Várnai, X. Salvatella, and M. Vendruscolo, The MUMO (Minimal Under-Restraining Minimal Over-Restraining) Method for the Determination of Native State Ensembles of Proteins. *J. Biomol. NMR* 37 (2007) 117.

[31] R. Kiss, D. Kovács, P. Tompa, and A. Perczel, Local Structural Preferences of Calpastatin, the Intrinsically Unstructured Protein Inhibitor of Calpain, *Biochemistry* 47 (2008) 6936.

[32] A.G. Palmer III, C.D. Kroenke, and J.P. Loria, Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. *Meth. Enzymol.* 339 (2001) 204.

[33] J. Gsponer, H. Hopearuoho, S.B.M. Whittaker, G.R. Spence, G.R. Moore, E. Paci, S.E. Radford, and M. Vendruscolo, Determination of an Ensemble of Structures Representing the Intermediate State of the Bacterial Immunity Protein Im7. *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 99.

[34] C. Redfield, NMR Studies of Partially Folded Molten Globule States, in, Protein NMR Techniques, 2nd ed., A.K. Downing, (ed.), New Jersey: Humana Press (2004).

[35] D.W. Ritchie,_Recent Progress and Future Directions in Protein-Protein Docking. *Curr. Prot. Pep. Sci.* 9 (2008) 1.

[36] V. Harmat, and G. Náray-Szabó, Theoretical Aspects of Molecular Recognition. *Croat. Chem. Acta,* 82 (2009) 277.

[37] A. Porollo, and J. Meller, Prediction-based Fingerprints of Protein Interactions. *Proteins: Struct. Funct. Bioinf.* 66 (2007) 630.

[38] L.P. Ehrlich, and R.C. Wade, Protein-Protein Docking. *Rev. Comp. Chem.* 17 (2001) 61.

[39] D. Fischer, S. Lin, H.L. Wolfson, and R. Nussinov, A Geometry-Based Suite of Molecular Docking Processes. *J. Mol. Biol.* 248 (1995) 459.

[40] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, and I.A. Vakser, Molecular Surface Recognition: Determination of Geometric Fit between Proteins and Their Ligands by Correlation Techniques. *Proc. Natl. Acad. Sci. U.S.A.* 89 (1992) 2195.

[41] J.J. Gray, S.E. Moughan, C. Wang, O. Schueler-Furman, B. Kuhlman, C.A. Rohl, and D. Baker, Protein-Protein Docking with Simultaneous Optimization of Rigid-Body Displacement and Side-Chain Conformations. *J. Mol. Biol.* 331 (2003) 281.

[42] J. Mintseris, K. Wiehe, B. Pierce, R. Anderson, R. Chen, J. Janin, and Z. Weng, Protein-Protein Docking Benchmark 2.0: An Update. *Proteins: Struct. Funct. Bioinf.* 60 (2005) 214.

[43] K. Fodor, V. Harmat, J. Kardos, J. Antal, C. Hetényi, A. Perczel, B. Szenthe, Z. Gáspári, G. Katona, and L. Gráf, Conformational Adaptation of a Canonical Protease Inhibitor upon its Binding to the Target Protease Increases Specificity. *FEBS J.* 272 (2005) 167.

[44] B. Szenthe, Z. Gáspári, A. Nagy, A. Perczel, and L. Gráf, Same Fold with Different Mobility: Backbone Dynamics of Small Protease Inhibitors from the Desert Locust, *Schistocerca Gregaria*, Biochemistry 43 (2004) 3376.

[45] C.S. Goh, D. Milburn, and M. Gerstein, Conformational Changes Associated with Protein-Protein Interactions. *Curr. Opin. Struct. Biol.* 14 (2004) 104.

[46] M.F. Horstemeyer, Multiscale Modeling: A Review in, Practical Aspects of Computational Chemistry, Methods, Concepts and Applications, J. Leszczynski and M.K. Shukla (eds.), Berlin: Springer, 2011, p. 87-135。