# E 8 Protein Crystallography and Protein Dynamics

R. Biehl and T. E. Schrader
Jülich Centre for Neutron Science & Institute of Complex Systems
Forschungszentrum Jülich GmbH

## Contents



Lecture Notes of the 50th IFF Spring School "Scattering! Soft, Functional and Quantum Materials" (Forschungszentrum Jülich, 2019). 

# 1 Introduction

Apart from water proteins are the most abundant molecules in living cells. They are constantly synthesized in the cell by the transcription and translation mechanism where first the DNA is read out to produce a messenger RNA which encodes the proteins and in the subsequent translation process the protein is synthesized by the ribosomes which itself is a protein/RNA-complex. Proteins fulfil numerous functions in the cell, for example enzymatic catalysis which enhances the speed with which molecules like fatty acids are synthesized or transport and storage of important molecules like oxygen or they are involved in immunology, just to name a few [1]. In order to perform these functions, proteins adopt a unique three-dimensional structure with a carefully controlled mixture of flexibility and stiffness. For an understanding of their function knowledge of this three-dimensional structure is a prerequisite.

Today it is realized that proteins are quite flexible objects, which show configurational changes on all length- and timescales. To reach a buried active site it is often necessary to open a cleft that the substrate can enter and to release the product. Binding of substrates in specific pockets allows to bring them close together in a specific configuration for e.g. phosphate or hydrogen transfer in a protected environment. Necessary conformational changes can be the rate-limiting step in catalysis. In other cases as for kinesin walk on myesin the configurational change is the aim of a chemical process to allow transport of cargo [2]. Therefore, protein dynamics on all length scales is a key to understand how conformational changes are related to function and which mechanisms are involved to allow the rich functionality of proteins.

Ideally one would like to produce a movie where one can follow the functioning protein in action in slow motion with atomic resolution. In practice, there are techniques available which have a sufficient time resolution but do only provide very limited structural information as for example time resolved infrared spectroscopy. On the other hand, there are methods which provide full atomic resolution but with limited time resolution. With those methods one often stops the enzymatic process of a protein in an intermediate state by using inhibitor molecules which trap the catalytic process of the protein in a certain intermediate state.

The first part of the article will focus on the latter static, elastic scattering techniques among which x-ray protein crystallography is the most widely used one. Here, there are some promising attempts to do time resolved x-ray scattering to have both structural resolution at the atomic level and time resolution [3]. But this technique of time resolved x-ray scattering is often limited to certain systems which undergo a photo-initiated process. This part will also introduce the method of neutron protein crystallography since there are some similarities but also some differences to x-rays as probes. Finally, an example case study is discussed where these both techniques give complimentary information.

The second part of the article illustrates the power of inelastic neutron scattering for the elucidation of equilibrium motions of proteins in solution. In recent years the biological community became more and more aware of the importance of motions and dynamics in proteins that can play an important role in understanding function. The importance of protein dynamics may be highlighted in the frame of drug design. While in the past in general the development of drugs was done using static crystallographic structures during the last 10 years the state-of-the-art involves ensemble docking. This part shows examples how protein internal dynamics can be measured.

# 2 Some Basics about Proteins

## 2.1 Amino acids as the building blocks of proteins

Proteins consist of a chain of amino acids. In that sense they are biopolymers and since they in general have charged side chains they can also be called polyelectrolytes. The first information about a protein is therefore the number of amino acid residues it contains. This number can span quite a wide range between 10 and 25 000. Typically, a protein contains 100 amino acids. The average molecular weight per amino acid is around 100 g/mol. Despite there are many different amino acids (or to be more exact: 2-amino carboxylic acids) present in living organisms not all of them are used to build up proteins. The proteinogenic amino acids are shown in Figure 1. Covalently attached to the central C-atom, the so-called $C_\alpha$ atom, are an amino functional group, a carboxylic group, the side chain atoms and finally one hydrogen atom. Since the $C_\alpha$ atom has a $sp^3$ hybridization sterically all four constituents point into the corners of a tetrahedron. Since (apart from the amino acid glycine) this $C_\alpha$ atom has four different constituents it forms a chiral center. So, two different arrangements of these four constituents are possible which lead to the L- and D-enantiomers of the corresponding amino acid, according to Fischer's convention. But in nature only the L-enantiomers are found as building blocks of proteins.

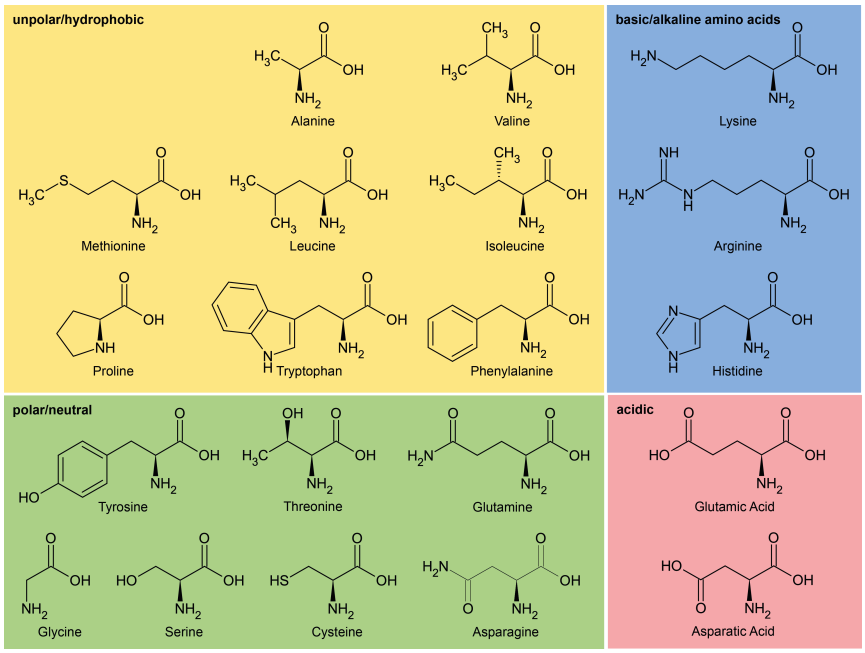



Figure 1 A compilation of all 20 amino acids found in natural proteins. The N-terminal amino group is shown in its neutral charge state pointing to the bottom of the page. Covalently attached to it is the Cα atom which carries the corresponding side chain group (Adapted from http://upload.wikimedia.org/wikipedia/commons/7/7d/Overview_proteinogenic_amino_acids-DE.sv).

The sequence of amino acids is the **primary structure** of the protein. It can be displayed as a line of text with a one letter code representing one amino acid. It is a common convention that the line of text starts at the N-terminal end of the amino acid chain i. e. with the amino acid with side chain $R_1$ in the example given in Figure 2 on the left.

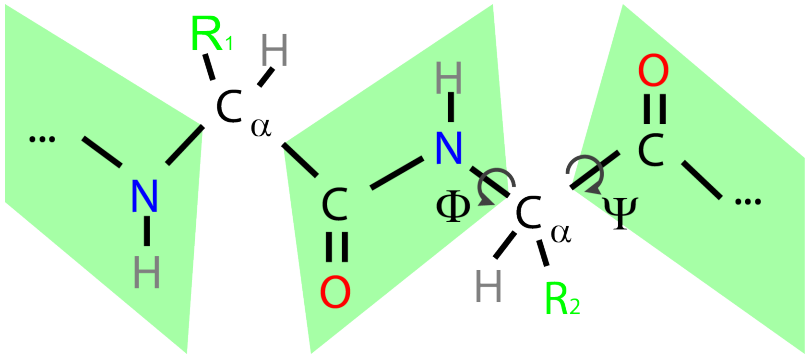



Figure 2 Due to a partial double bond character of the C-N peptide bond the rotation around it is hindered by a steep potential. This causes the four atoms OCNH to form a planar structure (marked in green).

A special feature of the peptide bond is its partial double bond character of the chemical bond between the carbon and nitrogen atom. The lone electron pair at the nitrogen atom is delocalized and has some existence probability between the atoms forming the peptide bond. This causes the bond length to shrink below the value of a single bond. As a consequence, the rotation around the CN-bond is hindered and the four atoms OCNH form a planar geometry denoted by a green polygon in Figure 2. This is why one amino acid only contributes two degrees of freedom to the amino acid backbone which are denoted by the dihedral angles $\Phi$ and $\Psi$, for their definition see Figure 2.

### 2.2 The three-dimensional structure of proteins

This structure is stabilized by four different interactions. First of all, there is the possibility of establishing **hydrogen bonds** between two parts of the backbone, but also between side chains or between a side chain and a part of the backbone. Another stabilizing mechanism is a formed **salt bridge** between a negatively and a positively charged side chain, e. g. aspartic acid and lysine. The third interaction is the formation of a **hydrophobic cluster** or core. Hereby the surrounding water plays a major role which makes it a mostly entropic effect. It is more favorable for the water molecules to form hydrogen bonds with each other than to stick between some hydrophobic side chains. This is why those side chains tend to be packed together in the folding process resulting in van der Waals interactions among them. The fourth stabilizing moment of a three-dimensional fold of a protein is a **disulfide bridge** between two cysteine residues.

When the primary structure only gives the linear sequence of amino acids, the **secondary structure** of a protein denotes all arrangements of the protein backbone stabilized by a regular hydrogen bonding pattern. These hydrogen bonds are solely between different parts of the backbone. There are several structural motifs of that kind which occur frequently in proteins. Some of these motifs have been given a name e.g. $\alpha$-helix or parallel $\beta$-sheet. The **tertiary structure** of a protein denotes the three-dimensional arrangement of all atoms of the protein in space, including the side chains. This information can be obtained by structural techniques as for example protein crystallography (will be discussed below), or solution NMR. Some proteins need more than one amino acid chain to be functional. The arrangement of the different amino acid chains is then referred to as the protein's **quaternary structure**.

# 3 Crystallography

### 3.1 The Physics of Protein Crystallography

The following chapter will show how most of this structural information has been obtained. For both techniques x-ray and neutron protein crystallography a single crystal of the protein of

interest is required since the scattering of one protein molecule is very weak. So, in general a crystal has to be grown, especially large ones in case of neutron crystallography since the neutron luminosity of modern sources is much smaller than the x-ray flux reached by synchrotron sources.

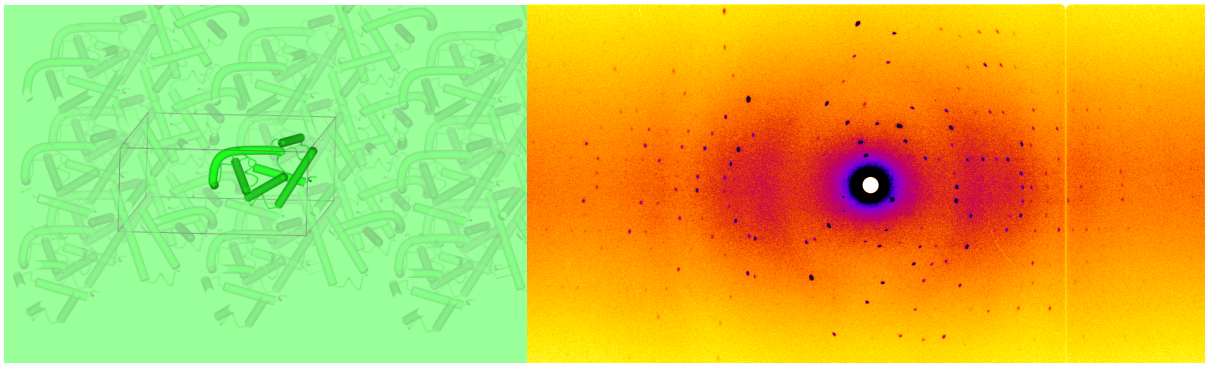


Figure 3 Real space arrangement of myoglobin molecules in a crystal of space group $P2_1$ (on the left) versus diffraction pattern (right) of a myoglobin crystal.

The crystal then serves as a noiseless amplifier of the diffraction signal. But the arrangement of proteins in a crystal brings in another advantage, since the orientational averaging can be avoided, which is always present in the solution phase. Figure 3, left hand side shows the regular arrangement of myoglobin molecules in a crystal lattice. The unit cell of the monoclinic lattice (space group $P2_1$) is indicated by black lines. It bears two myoglobin molecules in one unit cell. The picture on the right shows a diffraction pattern recorded with the instrument BioDiff on a myoglobin crystal.

**An X-ray macromolecular crystallography beamline**

Synchrotron beamlines provide extremely high photon flux for x-ray crystallography. Due to the high demand from the structural biology community, often more than one macromolecular crystallography beamline is operated at a synchrotron. Those beamlines are optimized for special wavelengths and focal diameters. As an example, the beamline BL14.2 has a beam size at the sample position of 150 µm x 100 µm (FWHM). Its Rayonix MX-225 detector has a pixel size of 37 µm. Without on chip binning one frame amounts to 6144 x 6144 pixels. The exposure time per frame is typically between 3 to 10 seconds. Typically, the sample crystals are kept at temperature as low as 100 Kto avoid radiation damages. To record a full data set takes about 10-30 minutes. The largest diagonal of a typical protein crystal ranges between 10 to 500 µm

**A neutron macromolecular crystallography instrument**

Since x-rays are scattered from the electrons in the crystal and neutrons from the nuclei, hydrogen atoms are hardly seen in x-ray crystallography experiments. Only at very high resolutions of 1 Å or less there is a chance to observe hydrogen atom positions. This resolution is often not within reach because of the quality of protein crystals. Here neutron protein crystallography must be employed to retrieve the hydrogen atom positions. Moreover, neutron scattering can distinguish between different isotopes, especially between hydrogen and deuterium. Whereas from x-ray crystallography the electron density in the unit cell of the crystal can be calculated, neutron protein crystallography yields the nuclear scattering length density, which is a signed quantity.

A major drawback of the method neutron protein crystallography is the required crystal size. Due to the much smaller neutron flux as compared to typical x-ray fluxes the crystals required for a neutron crystallography study must be much larger as compared to x-ray crystallography. Here, often crystal diagonals of 1 mm and more have to be reached.

As an example of a neutron diffraction instrument optimized for protein crystallography the instrument BioDiff at the FRM II shall be introduced. It is a collaboration between the

Forschungszentrum Jülich (FZJ) and the Forschungs-Neutronenquelle Heinz Maier-Leibnitz (FRM II).

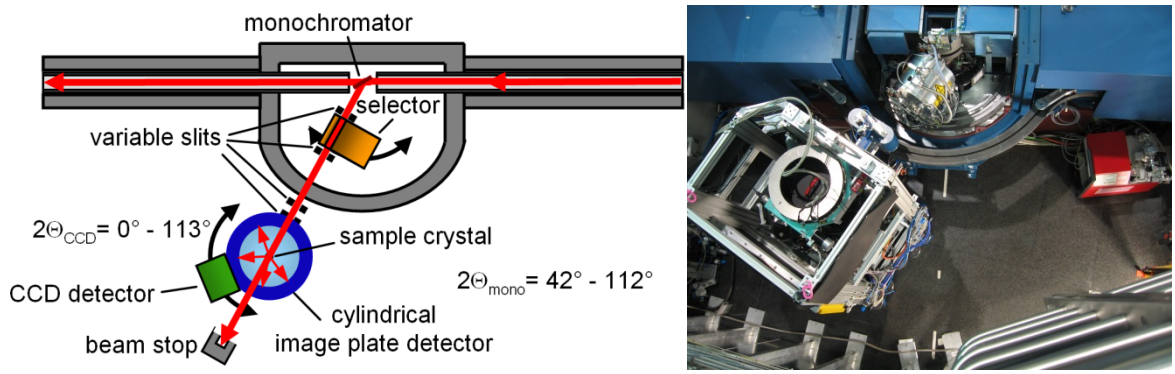



Figure 4 Schematic view of the BioDiff instrument (left) and a picture taken from a similar view point with the biological shielding removed (right).

Figure 4 shows a schematic top view of the instrument and a corresponding picture when the biological shielding has been removed. The neutron beam from the cold neutron source of the FRM II reactor enters from the right. By Bragg reflection from a pyrolytic graphite crystal (002-reflex) neutrons are taken out from the white neutron spectrum and pass a variable slit and a velocity selector. The velocity selector acts as a $\lambda/2$ filter to remove higher orders. Together with the pyrolytic graphite crystal it forms a monochromator with a $\Delta\lambda/\lambda$ of ca. 2.5 %. Behind the velocity selector the beam passes a second variable slit and the main instrument shutter. Before entering the detector drum of the image plate detector through a Zirconium window a collimator made out of boron carbide apertures with fixed diameters between 3 mm and 5 mm shape the beam to fit to the sample size. At present the sample is usually contained in a quartz glass tube. It is fixed to a standard goniometer which is mounted upside-down from the sample stage on top of the instrument. After passing the sample the main neutron beam exits the detector drum through a second Zirconium window and hits finally the beam stop. The cylindrical image plate detector can be read out with three different resolutions of 125 µm, 250 µm and 500 µm. As an alternative, one can lower the image plate detector and swing in a neutron sensitive scintillator which is imaged onto a CCD-chip. This CCD-camera set up is used for a fast alignment of the sample crystal with respect to the neutron beam.

The diffraction pattern shown in Figure 3, right hand side has been recorded with the image plate detector. In fact, a complete crystallographic data set on a typical protein crystal takes about 10 days to record and involves about 300 detector images. In between two detector images the crystal is rotated with the goniometer by 0.3° typically. Of course, the required time to record a data set is much longer as in case of x-ray diffraction.

**Some mathematical aspects of diffraction by a protein crystal**

Having recorded a complete data set on a crystal some data treatment is necessary in order to calculate meaningful atom positions. Here, only a brief introduction can be given. More details can be found in text books [4,5].

Assuming a number of n atoms per unit cell the structure factor of a single unit cell can be written as:

$$F(\mathbf{S}) = \sum_{j=1}^{n} f_j \exp(2\pi i \mathbf{r}_j \mathbf{S}) \qquad (1)$$

Here $\mathbf{r}_j$ denote the atom position of atom j and $\mathbf{S}$ is the scattering vector perpendicular to the plane which reflects the incident beam. In the previous lectures the scattering vector $\mathbf{Q}$ was defined. It relates to $\mathbf{S}$ with the following relation: $\mathbf{Q} = 2\pi\mathbf{S}$ . In crystallography it is just more

convenient to use $\mathbf{S}$ instead of $\mathbf{Q}$. $f_j$ can be seen here either as the scattering length of atom j in the neutron scattering case or the atomic scattering factor in case of x-ray diffraction. One can generalize this approach by switching form the summation to an integration to yield:

$$F(\mathbf{S}) = \int_{unitcell} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r}\mathbf{S}) d^3\mathbf{r} \tag{2}$$

where $\rho(\mathbf{r})$ is the electron density distribution or the scattering length density respectively. Since a crystal consists of AxBxC unit cells, the structure factor of the crystal can be composed as

$$F_{cryst.}(\mathbf{S}) = F(\mathbf{S}) \sum_{u=0}^{A} \exp(2\pi i u \mathbf{a}\mathbf{S}) \sum_{v=0}^{B} \exp(2\pi i v \mathbf{b}\mathbf{S}) \sum_{w=0}^{C} \exp(2\pi i w \mathbf{c}\mathbf{S}) \tag{3}$$

The vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ denote basis vectors of the unit cell. For an increasing number of unit cells, the sums can be represented by delta functions leading to the Laue conditions for the structure factor being non-zero:

$$\mathbf{a}\mathbf{S} = h, \mathbf{b}\mathbf{S} = k, \mathbf{c}\mathbf{S} = l \tag{4}$$

This means that one only gets constructive interference, when the scattering vector is perpendicular to planes in the crystal which can be denoted by the index vector $\mathbf{h} = hkl$. For this reason, the diffraction pattern of a single crystals shows distinct peaks, the so-called Bragg peaks. The Bragg law can be easily derived from equation 4. Figure 5 shows the Ewald sphere construction. It is a tool to construct the direction of the diffracted beam. The Ewald sphere has its origin at the position of the crystal. Its radius is the reciprocal wavelength used in the scattering experiment. The origin of the reciprocal lattice is placed at the intersection of the sphere with the incident beam direction. Whenever the orientation of the reciprocal lattice is such that another point of the reciprocal lattice lies on the Ewald sphere a diffracted beam results in the direction of line running from the center of the Ewald sphere through that point.

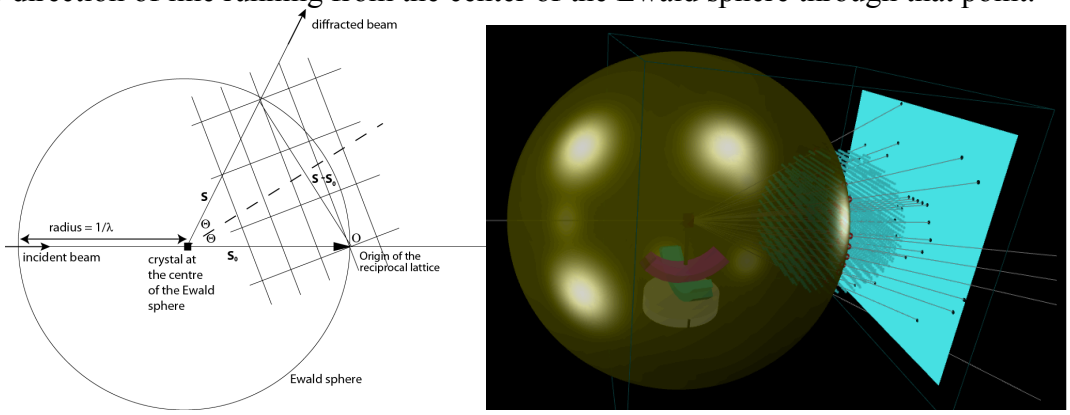



Figure 5 Ewald sphere: On the left the schematic shows how to construct the Ewald sphere. On the right an Ewald sphere (golden colour) construction is shown in three dimensions. The blue square represents a flat two-dimensional detector.

When the crystal is rotated the reciprocal space rotates with it resulting in other lattice points to cause diffracted beams. In practice the incident beam is not strictly monochromatic but has a wavelength distribution which causes the Ewald sphere to be elongated to form a spherical shell of a certain thickness. This increases the number of diffracted beams observed. The beam divergence adds also to its thickness.
So, the positions of the diffracted beams on the detector only depend on the reciprocal lattice. The structure of the protein inside the unit cell is encoded in the amplitude and phase of the structure factor.

To obtain the electron density or the nuclear scattering length density one has to perform the inverse Fourier transformation:

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h}) \exp(-2\pi i \mathbf{rh}) \qquad (5)$$

Here V is the volume of the unit cell. Unfortunately, only the modulus squared of the structure factor is measured as intensity on the detector. The phase information is lost which is known as the phase problem of crystallography.

There are several solutions to the phase problem which are predominantly applicable for the x-ray diffraction case:

a. isomorphous replacement: Several crystals of the same crystal structure have to be available for this method. First a crystallographic data set is recorded on an untreated crystal. Then crystals are soaked in at least two different heavy atom salt solutions. In the best case, the different heavy atom ions occupy different regular positions in the unit cell. From these (at least) two crystallographic data sets recorded on the heavy atom treated crystals phase information can be retrieved which is then used to determine the phases of the data set of the untreated crystal.
b. anomalous dispersion: Often it is possible to replace one distinct methionine amino acid with an artificial selenomethionine one. The selenium atom has a suitable absorption edge on which anomalous scattering can be performed by tuning the wavelength of the beamline to the anomalous regime. Crystallographic data sets are then recorded at different wavelengths from which the phase information can be calculated. In some cases this approach can also be adopted for sulfur atoms present in naturally occuring cysteine residues.
c. molecular replacement: From the primary structure one can search the protein data bank (pdb) for proteins with a similar amino acid sequence. If one finds enough fragments which seem to be sufficiently homologous to the unknown structure one can use those fragments for the calculation of initial phases. In further refinement steps these phases can be improved. Since the number of unique structures entered in the protein data bank is growing this method is increasingly favored over other methods.

The phase problem of the neutron data sets is solved by using the x-ray structure and the molecular replacement technique.

**Model building and refinement**

With the data treatment one has now arrived at a contour map $\rho(\mathbf{r})$ be it either a nuclear scattering length density or an electron density. Now the information on the primary structure of the protein is used and the backbone is coarsely fitted into the contour map either manually or employing software. From this model new amplitudes and phases of the structure factor are calculated using eq. 1. The modulus squared of the structure factor is again compared with the data. One could now think of a least square based fitting procedure to find the optimum arrangement of the protein atoms in the unit cell. In practice however, maximum likelihood and simulated annealing molecular dynamics simulations are used because those are superior to the least square approach in terms of overcoming local minima. In these molecular dynamics simulations, a lot of stereochemical information is used as restraints for example known bond lengths of single bonds between two carbon atoms or bond angles. The agreement between model and observed contour map is often measured by calculating a so-called R-factor. A good agreement between model and measured data leads to an R-factor of about 0.2. R-factors of 0.5 and above are indicative for a random or no agreement between model and data.

## 3.2 An example case study: Fighting Meticillin resistant bacteria

This example also shows the nice interplay between x-ray and neutron crystallography. It is chosen because of the relevance it has to our everyday life. The protein involved here is the β-lactamase. It is produced by bacteria and partly secreted to their surrounding environment in order to split a certain bond in the four-membered ring (=β-lactam motif) of a series of very common antibiotics. This is one of the mechanisms which render such bacteria resistant against this type of antibiotics. The antibiotic drug called meticillin also bears such a four membered ring. Meticillin resistant *Staphylococcus aureus* is assumed to possess this mechanism of resistance against β-lactam antibiotics. This is why it would be interesting to investigate the catalytic cylce of this protein in order to block it. This would result in the bacteria being affected by the antibiotics again.

The experiment was designed in a clever way. The catalytic cycle of the β-lactamase consists of an acylation step where a covalent adduct between protein and substrate (here the antibiotics molecule) is formed (Figure 6 top part). Here, the CO double bond is split into a single bond and a tetrahedral intermediate between substrate and protein is produced. The second part is a deacylation step where the four-membered ring (the so-called β-lactam motif) is split and the product is released. The product cannot act as an antibiotic any more due to its different chemical structure.

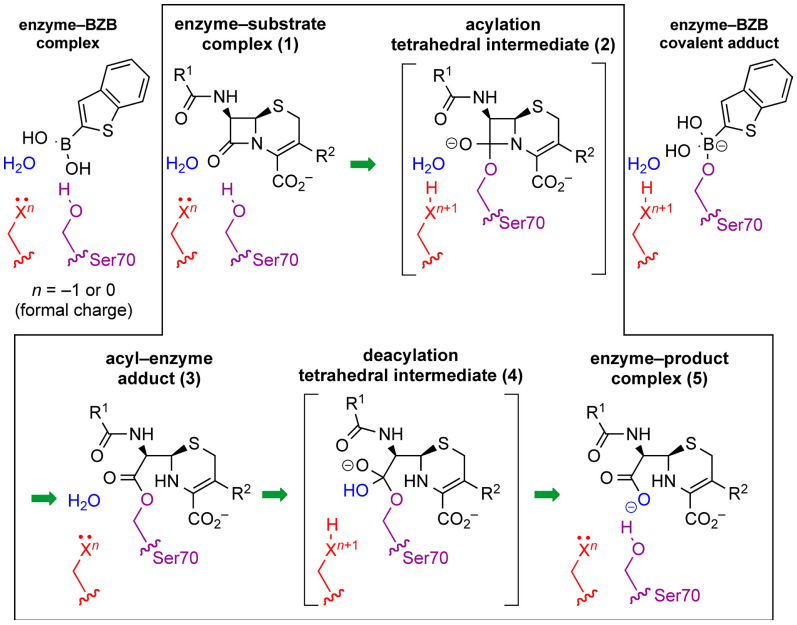



Figure 6 Catalytic cycle of a β -lactamase enzyme. (The Figure is taken from ref. [6]. This research was originally published in Journal of Biological Chemistry, Stephen J. Tomanicek et al. "Neutron and X-ray Crystal Structures of a Perdeuterated Enzyme Inhibitor Complex Reveal the Catalytic Proton Network of the Toho-1 β-Lactamase for the Acylation Reaction." © the American Society for Biochemistry and Molecular Biology.)

Due to its relevance to clinics the enzyme family of β-lactamases has been subjected to countless studies mostly by x-ray crystallography but the main question yet not addressed was the nature of the active base which takes the excess proton in the acylation step. It was not clear which part, which side chain of the protein takes over this role. But this was the key information for improving drugs to block this enzyme. Since a base is best detected when it takes a proton in the acylation step, the problem consisted of stopping the catalytic cycle in the acylation step and hunting for a proton which was not there in the ground state of the protein when no substrate was bound.

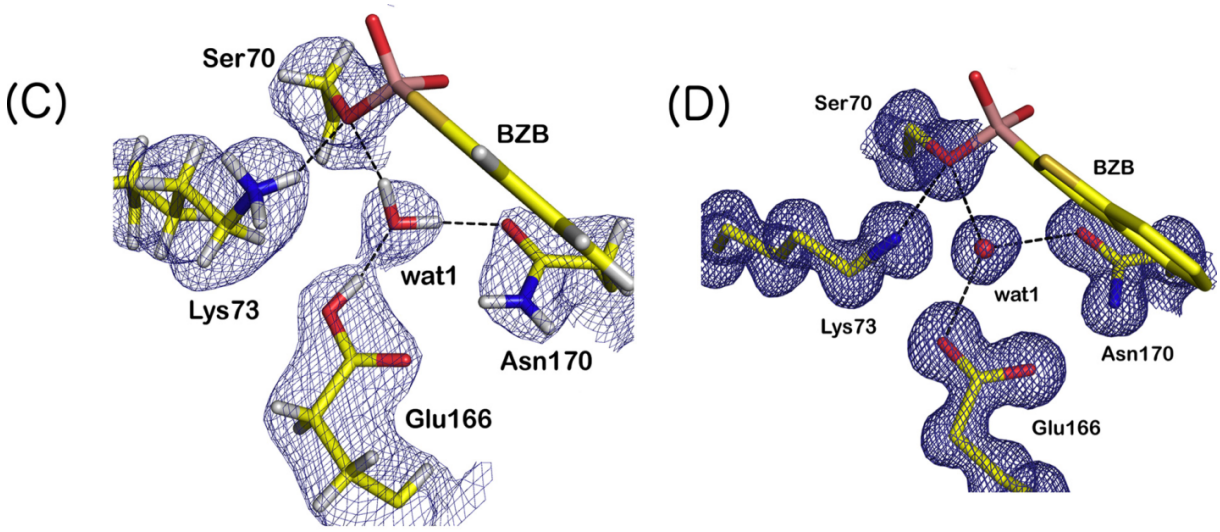



Figure 7 X-ray (right) and neutron (left) data on the BZB adduct in the catalytic cycle of a β-lactamase enzyme. (The Figure is taken from ref. [6]. This research was originally published in Journal of Biological Chemistry, Stephen J. Tomanicek et al. "Neutron and X-ray Crystal Structures of a Perdeuterated Enzyme Inhibitor Complex Reveal the Catalytic Proton Network of the Toho-1 β-Lactamase for the Acylation Reaction." © the American Society for Biochemistry and Molecular Biology.)

Here, a different ligand was used as the natural antibiotics: benzothiophene-2-boronic Acid (BZB). This ligand was known to stop the protein in the acylation step. In addition to that, this ligand was also known to mimic this covalent intermediate between substrate and protein. In addition to that, neutron protein crystallography was used to detect the proton unambiguously. For the latter technique the protein was expressed in deuterated media such that a fully deuterated protein resulted. The BZB was also synthesized using deuterium instead of hydrogen atoms. In fact, even a special Boron isotope was used which shows less neutron absorption than the natural abundant boron which is known as a good neutron absorber. Figure 7, left side shows the x-ray structure and Figure 7 right side shows the neutron structure. The proton is clearly seen in the neutron structure and therefore the side chain Glutamate 166 could be identified as the base in the acylation step. With this knowledge better drugs can be designed which bind this sidechain firmly blocking the protein's catalytic cycle in its first step. These additional drugs will break the resistance of those bacteria against common antibiotics.

# 4 Protein dynamics

An early model about protein activity and specificity was the "lock and key" model, which assumes an exact fit of the protein active site to the substrate due to complementary geometrical shapes but with a rigid conformation as found in crystal structures [7]. To explain also the stabilization of the transition state with bound substrate in a different configuration compared to the unbound state the later "induced fit" model [8] allows a reshaping of the binding site to the substrate including local configurational changes of amino acids or large structural changes as for allosteric transitions. Still the protein is viewed as a rigid structure in liganded and unliganded case. Including the dynamics of proteins as transitions in a configurational ensemble the model of "conformational selection" allows substrate binding in a specific configuration as reached due to the thermal motions.

There are two different but linked types of dynamic motions. In terms of an energy landscape view thermal motions are motions that cover the configurational space at thermal energy kT in equilibrium [9]. The accessible configurational space can spread over a single deep minimum or a broader rugged valley where in both the depth defines the occupancy of a configuration in the valley within Boltzmann distribution. Kinetic motions try to find the equilibrium from a

higher energy level and are directed towards equilibrium. The higher energy level can be due to an excitation (e.g. photolysis) or binding of a ligand that changes the local energy landscape. Nevertheless, thermal motions occur also in the excited state and may help to overcome energy barriers on the kinetic pathway.
The fastest motions in proteins are bond vibrations, side chain rotation at the protein surface or torsion of buried methyl groups with sub-angstrom amplitudes on picosecond timescales. Rearrangements of amino acids to adjust the orientation of functional groups may require local flexibility of neighboring amino acids in a cooperative manner that slows down the process. Movements of secondary structure elements or rearrangement of groups of amino acids on nanosecond timescale allow the adaption of the protein structure to bind specific ligands. Slower motions with larger angstrom amplitudes are relative motions of complete domains as hinge bending movements or swapping of domains depends strongly on the local environment and can be fast as several nanoseconds or slow as up to seconds dependent on the needed rearrangements and the involved interactions. Allosteric transitions, functional conformational changes, folding and unfolding will happen on microsecond timescale and nanometer length scales. In general, all these motions are dependent on each other and are coupled. The local atomic fluctuations lubricate the domain motions on larger length scales and domain motions change the shape of the protein.

## 4.1 Local movements

Local movements comprise movements of single atoms or small atom groups. Due to thermal excitation, bonded atoms show vibrating movements as eigenmodes of their specific configuration. Each sidechain of an amino acid can have different configurations with respect to the backbone dependent on the space required for a change e.g. in orientation. Amino acids at the surface have more configurational freedom compared to completely buried amino acids. Configurational changes can be due to thermal movements or due to specific processes as binding of a substrate.

**Atomic vibration**

The fastest movements with highest energy are atomic vibrations as they are common for any molecule. Figure 8a shows as an example the geometry of torsion around the C-C bond of a methyl group, a symmetric stretching of the C-H bonds and a symmetric bending of the c-H bonds of the sidechain of alanine. These are only a few possible vibrations of alanine sidechain and each amino acid has different vibrational frequencies dependent on the atomic structure. For an overview see Barth et al [10]. The exact frequency depends not only on the geometry as for free molecules of the same architecture, but also on the direct environment and neighboring amino acids. Hydrogen bonds or polar interactions alter the vibration frequency. Figure 8c shows an example spectrum of $(PPG)_n$ an synthetic polypeptide with a similar structure to natural collagen [11]. The frequency range reaches from 100 $cm^{-1}$ to about 2000 $cm^{-1}$ (12.4 meV – 250 meV) describing motions on a timescale of 0.01ps to 0.3 ps. The difference in the spectra is related to a partial deuteration that changes the frequency and the amplitude of specific vibrations due to the change of the hydrogen mass. In this way the specific exchange allows to separate some of the vibrational modes present in the sample. Typical instruments with the necessary large energy transfer are neutron time of flight instruments.
For proteins the Amide I absorption band between 1600 cm $^{-1}$ and 1700 $cm^{-1}$ is of importance because it allows the determination of the relative content of secondary structure in the protein. It can be measured by infrared spectroscopy (e.g. FTIR). In this absorption band the C=O stretch vibration of the peptide backbone is dominant [12]. As shown in Figure 8b the hydrogen bond between the C=O and the N-H of a different amino acid stabilize the secondary structure elements. A specific hydrogen bond modifies the stretching vibration in a way that is

characteristic for the local geometry defined by the secondary structure. Measuring the different contributions allows extracting the fractional content of secondary structure elements.

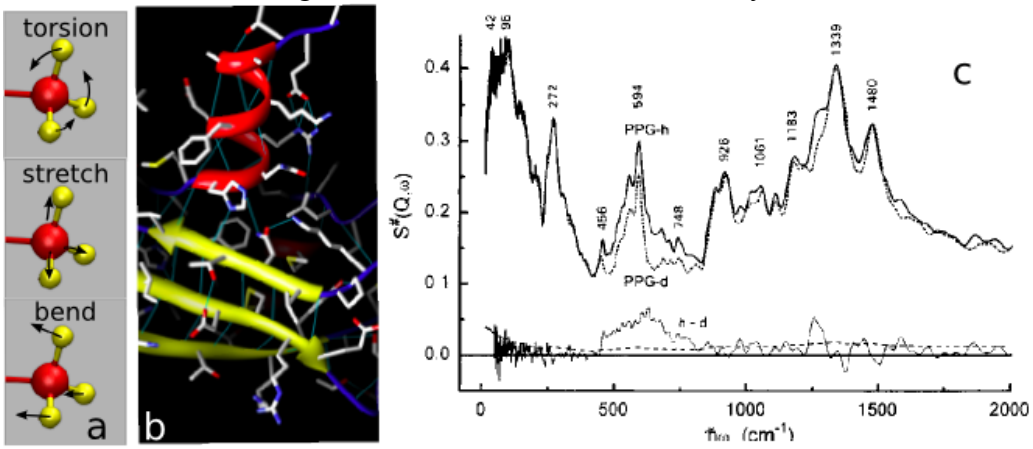



Figure 8 Example geometries of torsion, symmetric stretching or bending vibrations of a methyl group as the side group of alanine. b) Secondary structure elements β-sheets and α-helices (in yellow, red) with side chain elements. The secondary structure is stabilized by the hydrogen bonds (thin blue lines). c) Vibrational spectrum of synthetic polypeptide (PPG)n associated in a right-handed supercoiled triple helical arrangement as found in natural collagen measured at the time-focused crystal analyzer spectrometer (TFXA) at ISIS, UK by Middendorf et al. [11]

**Sidechain movements with Atomic Resolution: Time Resolved X-ray Crystallography.**

Conventional x-ray crystallography measures the diffraction patterns in different orientations of a single crystal with hundreds to thousands of Bragg-reflexes from which the three-dimensional electron density is calculated. The crystal structure of atom positions is generated as found e. g in the Brookhaven Protein Databank (PDB). Configurational changes due to activity can only be examined if it is possible to grow a crystal in both states as e.g. liganded with the substrate or an inactive replacement of the substrate and without the substrate. Diffusion trapping can be used to find the binding site of a substrate that is able to diffuse into a substrate free crystal. A larger configurational change due to substrate binding is difficult to access in this way and the timescale of a process cannot be accessed.

Time resolved crystallography tries to measure diffraction patterns with a delay time to observe the time evolution of a process. To access a dynamical process the protein must be active in the crystalline state and the process needs to be triggered inside the crystal with some reasonable amount of concentration. One way is to use a pump-probe method with measurements after a trigger event. Measurements with defined time delay relative to the trigger event e.g. photolysis of a process by a laser pulse allows to follow the process with different delays. On the other hand, a complete series with defined delays can be acquired if the repetition rate for a measurement is high enough. To overcome the need of different orientations for sub-second resolution a polychromatic Laue X-ray diffraction technique can be used [13]. Larger motions as domain movements cannot be observed as the crystal structure limits the configurational freedom to move over larger distances.

Schotte et al. have reported 100 picosecond time resolution x-ray crystallography of a myoglobin mutant after photolysis by an orange laser flash [14]. Figure 9A shows the electron density map of the unphotolyzed myoglobin and 100ps after the flash showing 3 larger configurational changes (arrows) and distributed smaller changes in the protein.

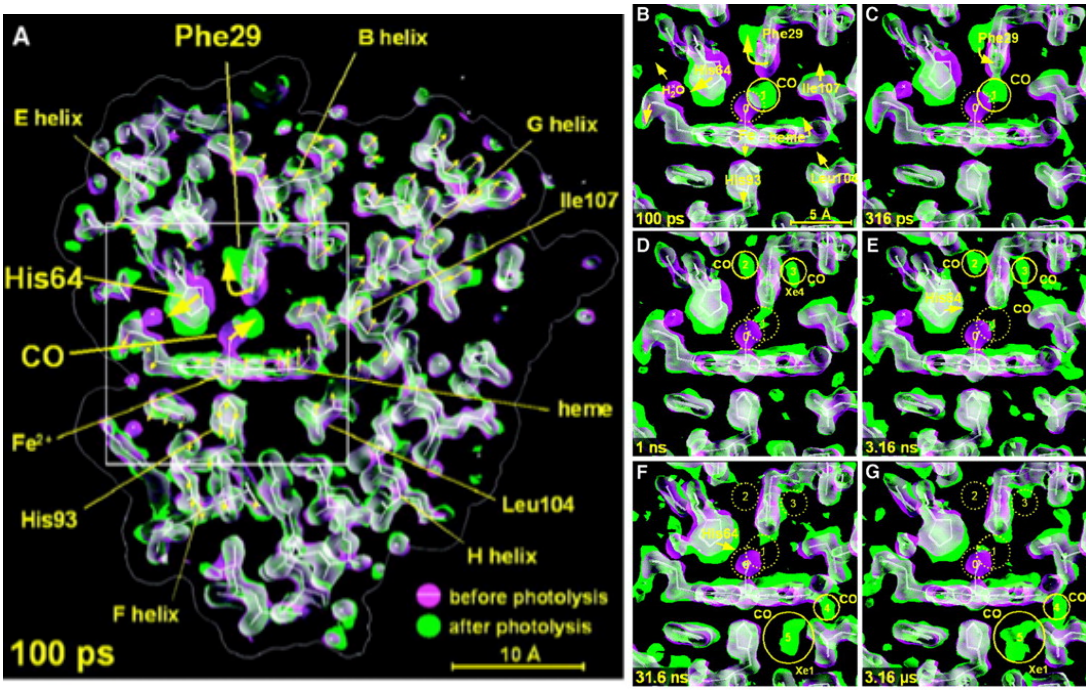



Figure 9: Electron density map of myoglobin before (magenta) and after photolysis (green). Overlapping densities are shown in white. The stick model included shows the unphotolyzed state. Large arrows indicate 3 large changes while small arrows indicate small rearrangements in the whole protein. Sequence of an enlarged view of A. B-E) Times are 100 ps, 316 ps, 1 ns, 3,16 ns, after the laser flash. F and G are 31.6ns and 3.16 µs after a longer ns laser flash. Circles indicate location of CO molecules. Figure from [14].

Figure 9 B-E show a sequence of delay times after the laser flash. The CO molecule dissociates and is trapped 2Å apart from the original binding site (positions 2 + 3). Phe29 is displaced but relaxes back to its original position within 316 ps. After several nanoseconds His64 relax to their deoxy position and the CO molecule has migrated to position 4 and 5 where it is trapped for several microseconds.
Here the conformational change due to photolysis is demonstrated with a correlated motion of sidechains rapidly sweeping away the CO molecule from its preferred binding site.
The experiment demonstrates that fast sub nanosecond reorientations are possible even in a restricted Crystal structure.

## 4.2 Domain motions

Domain motions are correlated motions between different domains or inside of domains like bending and torsion of the domain. The mechanisms are described as shear motions along an interface or hinge motions where such an interface is missing, but unclassifiable motions are allowed. The connection between domains can be a broad soft hinge as in the case of phosphoglycerate kinase, a single α-helix as in the case of lactoferrin or a disordered amino acid sequence of the protein chain as in the case of immunoglobulin G1 or mercury ion reductase.
Configurational changes changing the shape of a protein by rearranging complete domains can be observed by methods, which allow the detection of correlated motions over larger distances. SANS, SAXS allow examination of the low-resolution structure and the ability to compare structural changes in solution due to changes of pH, salt concentrations, temperature or substrate addition. Time resolved SAXS can reach sub-millisecond resolution and can be combined with stopped flow experiments or trigger events as in time resolved x-ray crystallography [15]. Neutron spin echo spectroscopy (NSE) is a technique that is able to access the timescales from 0.1 up to several hundred nanoseconds and simultaneously covers the

length scale relevant for protein domain movements as in SAXS or SANS of several nanometers distance between domains [16]. NSE measures the temporal correlation of configurational changes under utilization of the neutron spin to detect tiny velocity changes during the scattering process. The measured intermediate scattering function can be interpreted as a time correlation between small angle scattering patterns with nanosecond resolution. Main contributions to the correlation come from translational and rotational diffusion due to the spatial correlation of different proteins diffusing in the solution and internal domain dynamics on nanosecond timescale. In the following we present exemplary results to demonstrate a small variety of possible motional patterns.

**Phosphoglycerate kinase as a classical hinge**

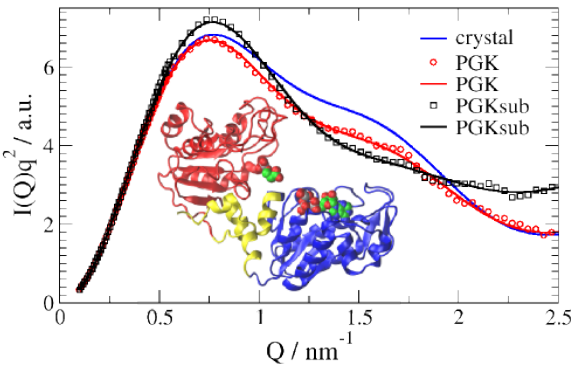



Figure 10 Form factor measured by SANS from PGK and PGKsub (Kratky plot: Q versus $Q^2I(Q)$). Lines show the calculated form factor from the crystal structure (blue) and from structures deformed along the softest normal modes to fit the experimental data. The inset shows the protein with the hinge in yellow, the main domains in blue and red and the substrates as spheres. Figure from [20].

Phosphoglycerate kinase (PGK) is an enzyme that is involved in glycolysis. It relocates a phosphate group from 1,3-biphosphoglycerate, an intermediate product in glycolysis, to ADP to synthesize ATP [17]. PGK is composed of two separated domains connected by a hinge region as shown in Figure 8 [18]. 1,3-biphosphoglycerate and ADP are bound at opposite positions in the cleft at the two domains. The active site is located at the hinge between the C- terminal and N-terminal domains. Evidence for a hinge bending motion induced by substrate binding was found by Bernstein at al. by comparing crystallographic structures of different species with and without bound substrates bringing the substrates closer together [19]. The crystal structure without substrate has an open cleft configuration with key residues Arg-39 and Gly-376 in the active center separated by about 1.18 nm. The proposed mechanism of induced fit due to substrate binding closes the cleft by a 32° hinge closure to the active configuration with bound substrate. This cleft closing motion mainly brings key residues Arg-39 and Gly-376 of the active center together with the substrates to a distance of 0.35 nm as found in the closed cleft crystal structure.

Figure 8 shows SANS measurements of PGK in solution with and without substrate demonstrating that the solution structures are different from the substrate bound crystal structure [20]. Modeling the structure by deformations along softest elastic normal modes (torsion and 2 perpendicular bends of the hinge) allowed modeling of the deformation due to substrate binding in solution. The distance between the active residues was reduced from 1.14 nm without substrate to 0.82 nm with bound substrate, but still to far to allow activity within a static structure.

NSE measurements show the relaxation of the intermediate scattering function dependent on diffusion and internal dynamics as shown in Figure 11 left. At low scattering vectors Q (observing large length scales) the protein looks point like and a single exponential relaxation is observed. t larger Q (observing length scale of protein size) additional contribution arise from internal dynamics. At long times diffusion is observed, that can be described within a single model for all Q values by rotational and translational diffusion constants. The additional

component at short times is described by a Q dependent amplitude and a relaxation time τ, which both depend on the construction of the specific hinge and influence function.

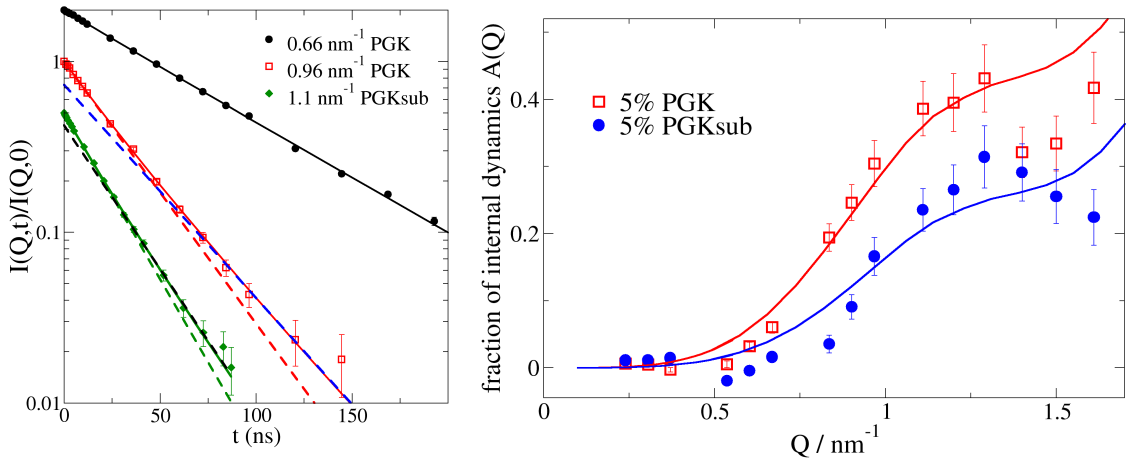
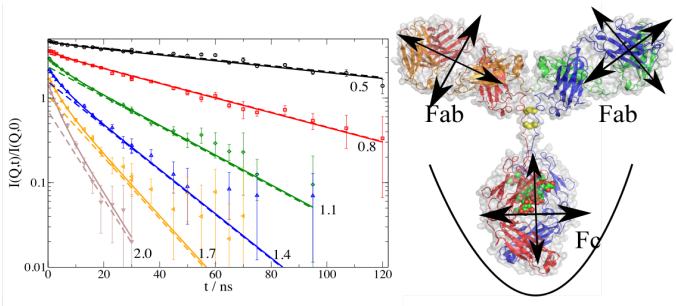



Figure 11 Left: Semi logarithmic plot of I(Q,t)/I(Q,t=0) for selected Q values (PGK, black and red; PGKsub, green; data are shifted for clarity and are equal 1 for t=0). Red and green dashed lines represent the initial slope extrapolated to long times. The blue and black dashed lines represent the long-time limit extrapolated to t=0. The long-time limit corresponds to rigid-body diffusion including inter-particle effects. The difference between extrapolated long time diffusion at t=0 to the initial slope amplitude is the internal dynamics contribution A(Q)exp(-t/τ). Right: Q dependence of A(Q) compared to model calculations allowing . Figure from [20].

The observed relaxation times are 60 ns for the substrate free and 45 ns for the substrate bound PGK. Figure 11 shows at the right the amplitude A(Q) with model calculations based on the softest normal mode deformations. From the deformation of the hinge the distance of the active residues can be calculated. It was found that the thermal driven deformation of the protein at the hinge is strong enough to reach configurations that allow activity as the active residue come close enough together.

In summary, the NSE investigation has demonstrated that the approach to a functional configuration of PGK needs to be attributed to the dynamic fluctuations of the main domains instead of an earlier proposed induced fit by substrate binding. Thus, in the case of PGK, hinge dynamics enables function.

**Fast antibody fragment motion: flexible linkers act as entropic spring**




**Intermediate scattering functions *I*(*Q*,*t*)/*I*(*Q*,0) of IgG.** Dashed lines show translational and rotational diffusion and are close to a single exponential for times t > 15 ns. Solid lines include additional internal dynamics of each domain in its own harmonic potential with a 7ns relaxation time.

The fragment motion of IgG was examined by Stingaciu et al by NSE [21]. In the experiment, aside from strong diffusion contributions, a clear signature of the internal dynamics on a timescale of 6 - 7 ns was observed. The combination of translational and rotational diffusion of the rigid protein describes the observed long-time dynamics on an absolute scale, despite the fact that we have a mixed population of monomers and dimers, as found in serum.

The fragment motion shows itself in terms of a strong decay of $S(Q,t)$ at short times, well separated from the overall diffusional relaxation. The data were analysed in terms of an Ornstein-Uhlenbeck like relaxation within a harmonic potential. Such an approach neglects the details of the complex linker region interaction on the residue level. Nevertheless, the dynamics

is described in an excellent way. Even for a short flexible linker the ensemble average seems to be sufficient to produce the characteristics of a harmonic spring. The obtained spring constant ($\approx 10\ pN/nm$) appears to be realistic compared to an entropic spring of similar length. The resulting forces stay below the limits that would be necessary to unfold the secondary structure of the attached fragments. The observed effective friction appears to be close to the friction of a free unbound fragment in the solvent. None of the directions seems to be suppressed.
The pre-existing equilibrium hypothesis [22], with multiple local minima in the configurational energy landscape, will have long transition times between the minima but is compatible with fast 6-7 ns motions in the local minima. Consequently, NSE observes the fast dynamics in pre-existing equilibrium states. The conformational flexibility in a pre-existing equilibrium configuration might be needed to adapt faster to a specific antigen when the antigen approaches the binding site of a fragment.

**Intrinsically Disordered Proteins (IDP)**

About 40 % of all proteins in the human body are intrinsically disordered. These IDP's do not exhibit any well-defined folded structure that could be crystallized. Their structural and dynamic properties reach from very soft structures over folded elements connected by extended and flexible loops to fully disordered polypeptide chains. The biological role of the IDP is founded in their high conformational adaptivity, enabling them to respond rapidly to environmental changes or other macromolecules allowing them to fold into different states or

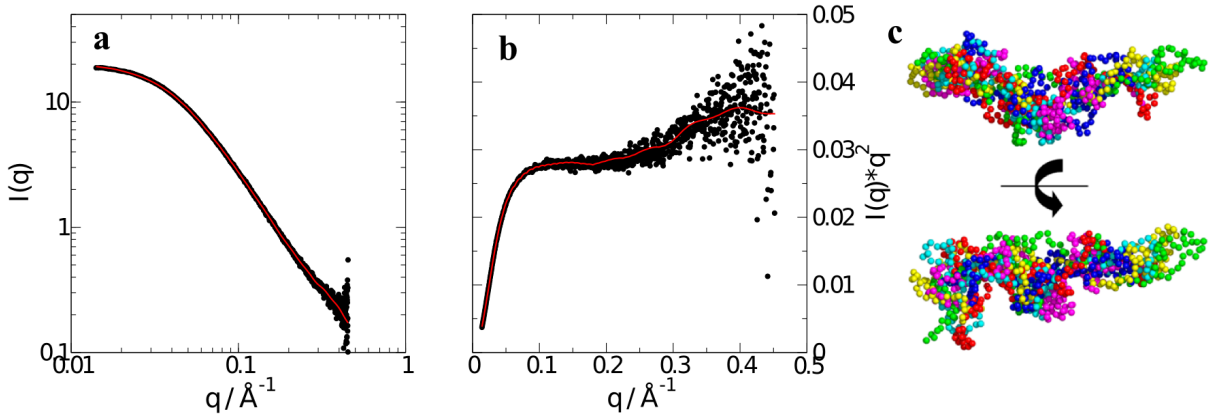



Figure 12 SAXS data from MBP at 4.5 mg/ml. a) The red solid lines is a fit with the Debye equation for a Gaussian chain. (b) Kratky plot. The line is a result of the scattering from the most probable conformational ensembles shown in c. c) Representative coarse-grained conformations of MBP as determined by inverse Monte Carlo. The structures are rotated by 90 °C in the lower part of the figure. The color code relates to six different realizations of the ensemble, which represent the data best. Figure from [23].

even a rigid 3D structure. For these properties dynamics is essential. Specifically the sampling of the energy landscape and the exploration of the large conformational space are driven by conformational motions of the unfolded peptide chain. On the other hand IDPs show the same dynamics as proteins during the early stage of folding. Since the intrinsic disorder prevents crystallographic structure determination, only low resolution SANS and SAXS information about the average structure in the disordered state exists.
Myelin basic protein (MBP) is a major component of the Myelin sheaths in the central nervous system [24]. In the human body MBP is of significant importance as there are many neurological disorders, as e.g. multiple sclerosis, that are related to MBP mal function. Lipid free MBP is not completely unfolded but retains some elements of the alpha helix and beta sheet (about 60 % of the protein is unfolded) [25].
Figure 12 a+b display X-ray form factors of MBP that display strong similarities to polymer form factors as the high Q data show a power law close to $Q^{-2}$ that is characteristic for Gaussian chain polymers in theta solvents. The small increase visible in the Kratky plot at high Q

indicates a length scale where the random oriented character vanishes and the linear character of short chain segments becomes visible. A Monte Carlo simulation was used to generate a coarse-grained ensemble representing the structural characteristics of and a

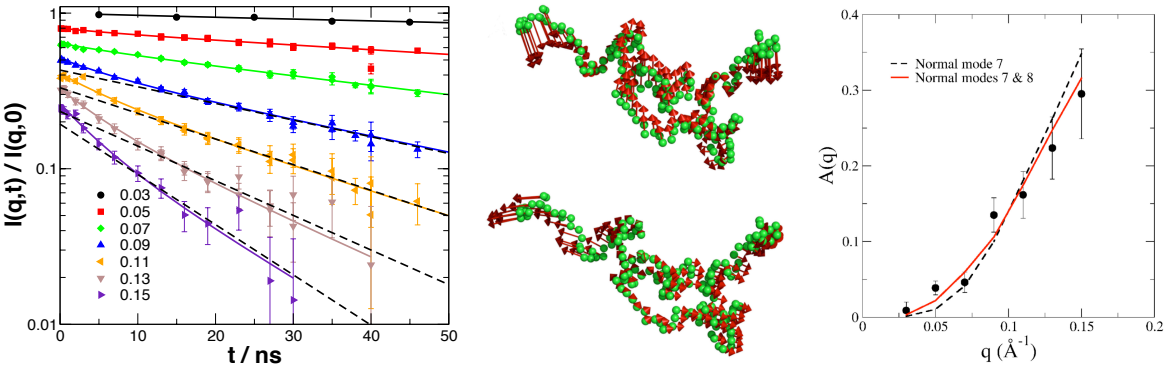



Figure 13 Left) NSE Data for MBP: All spectra start at unity but are shifted consecutively by a factor of 0.8 for clarity shown up to 50 ns. Solid lines are fits to the NSE data with the structural model. The dashed lines are exponential fits for t>20 ns to extrapolate the long time rigid body dynamics. A clear separation between the internal and the global dynamics is obvious. Middle) Displacement pattern of the normal modes 7 (upper part, bending) and 8 (lower part, stretching) from the structural model according to normal mode analysis. The lengths of the vectors are increased for better visibility. Right) Amplitude of the internal protein dynamics as obtained from the fit. The solid and dashed lines are the calculated mode amplitude according to equation 4. Figures from [23].

ensemble was selected that represent the SAXS data [26]. The resulting characteristic molecular shapes are displayed in Figure 12c. The model conformations indicate an elongated structure with a relatively compact core and flexible ends on both sites.

Figure 13 at left displays NSE spectra from 54 mg/ml solutions (times up to 140 ns could be accessed). Inspecting this figure, the two-component structure of the NSE spectra at Q-values above 0.9 nm$^{-1}$ is visible. Thus, we deal with long time rigid body motion augmented by internal dynamics with relaxation times below 10 ns.

The structural models based on SAXS analysis were used to describe the long time translational and rotational diffusion combined with a Q-dependent motional amplitude A(Q) and an internal mode relaxation time. The characteristic internal relaxation time $\tau_{int}$=8.4±2.0 ns is found for the whole structural ensemble and the corresponding amplitude A(Q) is displayed in Figure 13 at the right. Normal mode analysis was used to describe the deformation of the structural models. Figure 13 middle shows the first two normal modes as a bending of the structure, which already give a satisfactory description of the observed amplitudes. Comparing the normal modes with the structural models in Figure 12c it can be concluded that the normal modes describe approximately the motion from one structural model to the next. Here it is shown that even for very flexible structure as the amino acid chain still low frequency collective stretching and bending motions of the outer part of the structure describe the essential features of the large-scale dynamics.

# Acknowledgement

The author would like to thank Christian Felder for software support. The authors are especially grateful to Andreas Ostermann for supplying some of the Figures.

# References

[1] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemie* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

[2] R. D. Vale, Cell **112**, 467 (2003).

[3] P. Fromme and J. C. Spence, Curr. Opin. Struct. Biol. **21**, 509 (2011).

[4] J. Drenth, *Principles of Protein X-Ray Crystallography* (Springer-Verlag GmbH, 1999).

[5] N. Niimura and A. Podjarny, *Neutron Protein Crystallography: Hydrogen, Protons, and Hydration in Bio-Macromolecules* (Oxford University Press, 2011).

[6] S. J. Tomanicek, R. F. Standaert, K. L. Weiss, A. Ostermann, T. E. Schrader, J. D. Ng, and L. Coates, J. Biol. Chem. **288**, 4715 (2013).

[7] E. Fischer, Berichte Der Dtsch. Chem. Gesellschaft **27**, 2985 (1894).

[8] D. E. Koshland, Proc. Natl. Acad. Sci. U. S. A. **44**, 98 (1958).

[9] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).

[10] A. Barth, Prog. Biophys. Mol. Biol. **74**, 141 (2000).

[11] H. D. Middendorf, R. L. Hayward, S. F. Parker, J. Bradshaw, and a Miller, Biophys. J. **69**, 660 (1995).

[12] H. Susi and D. M. Byler, *Enzyme Structure Part K* (Elsevier, 1986).

[13] Z. Ren, D. Bourgeois, J. R. Helliwell, K. Moffat, V. Šrajer, and B. L. Stoddard, J. Synchrotron Radiat. **6**, 891 (1999).

[14] F. Schotte, M. Lim, T. A. Jackson, A. V Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, and P. A. Anfinrud, Science (80-. ). **300**, 1944 (2003).

[15] R. Graceffa, R. P. Nobrega, R. A. Barrea, S. V Kathuria, S. Chakravarthy, O. Bilsel, and T. C. Irving, J. Synchrotron Radiat. **20**, 820 (2013).

[16] R. Biehl and D. Richter, J. Phys. Condens. Matter **26**, 503103 (2014).

[17] R. K. Scopes, Enzym. **8**, 335 (1973).

[18] T. N. Bryant, H. C. Watson, and P. L. Wendell, Nature **247**, 14 (1974).

[19] B. E. Bernstein, P. A. Michels, W. G. Hol, P. A. M. Micheis, and W. G. Hol, Lett. to Nat. **385**, 275 (1997).

[20] R. Inoue, R. Biehl, T. Rosenkranz, J. Fitter, M. Monkenbusch, A. Radulescu, B. Farago, and D. Richter, Biophys. J. **99**, 2309 (2010).

[21] L. R. Stingaciu, O. Ivanova, M. Ohl, R. Biehl, and D. Richter, Sci. Rep. **6**, 22148 (2016).

[22] L. Pauling, J. Am. Chem. Soc. **62**, 2643 (1940).

[23] A. M. Stadler, L. Stingaciu, A. Radulescu, O. Holderer, M. Monkenbusch, R. Biehl, and D. Richter, J. Am. Chem. Soc. **136**, 6987 (2014).

[24] G. Harauz, N. Ishiyama, C. M. . Hill, I. R. Bates, D. S. Libich, and C. Farès, Micron **35**, 503 (2004).

[25] E. Polverini, A. Fasano, F. Zito, P. Riccio, and P. Cavatorta, Eur. Biophys. J. **28**, 351 (1999).

[26] P. Bernado, E. Mylonas, M. V Petoukhov, M. Blackledge, and D. I. Svergun, J Am Chem Soc **129**, 5656 (2007).

[27] O. Korb, T. S. G. Olsson, S. J. Bowden, R. J. Hall, M. L. Verdonk, J. W. Liebeschuetz, and J. C. Cole, J. Chem. Inf. Model. **52**, 1262 (2012).

[28] N. Andrusier, E. Mashiach, R. Nussinov, and H. J. Wolfson, Proteins **73**, 271 (2008).

[29] S.-Y. Huang and X. Zou, Proteins **66**, 399 (2007).