NMR studies of protein dynamics and structure

Jörgen Ådén

Umeå Universitet
Department of Chemistry
901 87 Umeå, Sweden

ISBN: 978-91-7459-092-0
Printed by: VMC, KBC, Umeå University
Umeå, Sweden 2010

Table of Contents

1 List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals I – IV.

I: **Ådén J**, Wolf-Watz M. (2007)
NMR Identification of Transient Complexes Critical to Adenylate Kinase Catalysis.
J. Am. Chem. Soc. Nov 14;129(45):14003-12

II: Rundqvist L, **Ådén J**, Sparrman T, Wallgren M, Olofsson U, Wolf-Watz M. (2009)
Non-cooperative Folding of Subdomains in Adenylate Kinase.
Biochemistry, 48 (9), pp 1911–1927

III: **Ådén J**, Wallgren M, Storm P, Weise C, Christiansen A, Schröder W, Funk C, Wolf-Watz M. (2010)
Arabidopsis thaliana Peroxiredoxin Q is extraordinary dynamic on the µs-ms timescale.
Submitted

IV: Wallgren M[†], **Ådén J**[†], Pylypenko O, Mikaelsson T, Johansson L. B, Rak A, Wolf-Watz M. (2008)
Extreme Temperature Tolerance of a Hyperthermophilic Protein Governed by Residual Structure in the Unfolded State.
J. Mol. Biol. Jun 13;379(4):845-58

[†] The authors contributed equally to this work.

2 Abbreviations

Aadk	Adenylate kinase isolated from *Aquifex aeolicus*
Adk	Adenylate kinase
Ap5A	(P^1, P^5-di(adenosine-5´-)penta-phosphate)
CD	Circular Dichroism
CPMG	Carr–Purcell–Meiboom–Gill
CSA	Chemical shift anisotropy
Eadk	Adenylate kinase isolated from *Escherichia coli*
HSQC	Heteronuclear Single Quantum Coherence
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
PrxQ	Peroxiredoxin Q isolated from *Arabidopsis thaliana*
RDC	Residual Dipolar Couplings
TCEP	tris(2-carboxyethyl)phosphine
Trx	Thioredoxin

3 Abstract

Enzymes are extraordinary molecules that can accelerate chemical reactions by several orders of magnitude. With recent advancements in structural biology together with classical enzymology the mechanism of many enzymes has become understood at the molecular level. During the last ten years significant efforts have been invested to understand the structure and dynamics of the actual catalyst (i. e. the enzyme). There has been a tremendous development in NMR spectroscopy (both hardware and pulse programs) that have enabled detailed studies of protein dynamics. In many cases there exists a strong coupling between enzyme dynamics and function. Here I have studied the conformational dynamics and thermodynamics of three model systems: adenylate kinase (Adk), Peroxiredoxin Q (PrxQ) and the structural protein S16. By developing a novel chemical shift-based method we show that Adk binds its two substrates AMP and ATP with an extraordinarily dynamic mechanism. For both substrate-saturated states the nucleotide-binding subdomains exchange between open and closed states, with the populations of these states being approximately equal. This finding contrasts with the traditional view of enzyme-substrate complexes as static low entropy states. We are also able to show that the individual subdomains in Adk fold and unfold in a non-cooperative manner. This finding is relevant from a functional perspective, since it allows a change in hydrogen bonding pattern upon substrate-binding without provoking global unfolding of the entire enzyme (as would be expected from a two-state folding mechanism). We also studied the structure and dynamics of the plant enzyme PrxQ in both reduced and oxidized states. Experimentally validated structural models were generated for both oxidation states. The reduced state displays unprecedented μs-ms conformational dynamics and we propose that this dynamics reflects local and functional unfolding of an α-helix in the active site. Finally, we solved the structure of S16 from *Aquifex aeolicus* and propose a model suggesting a link between thermostability and structure for a mesophilic and hyperthermophilic protein pair. A connection between the increased thermostability in the thermophilic S16 and residual structure in its unfolded state was discovered, persistent at high denaturant concentrations, thereby affecting the difference in heat capacity difference between the folded and unfolded state. In summary, we have contributed to the understanding of protein dynamics and to the coupling between dynamics and catalytic activity in enzymes.

4 Introduction

4.1 Aim of this thesis

In the last decades, the area of NMR spectroscopy has proven extremely useful for studies of proteins since it provides residue-specific information about dynamic processes ranging from picoseconds motions to hydrogen exchange processes with lifetimes of hours or more. In the work presented in this thesis, solution state NMR together with classical biophysical techniques have been applied to small proteins (~ 20 kDa) to analyze their structure, thermostability and dynamics. In order to study the coupling between protein dynamics and function, suitable model systems must be identified. For NMR studies, this means the protein must be relatively small, soluble at high concentrations and stable for a long time. I have worked with three different model systems matching these criteria to study general aspects of protein structure and dynamics. The goal has been to improve our knowledge about underlying mechanisms dictating protein structure and function, an area very important for protein folding and dynamics in general, but also how proteins (e.g. thermostable variants) can be engineered for industrial applications and commercial use.

4.2 Adenylate kinase

Adenylate kinase (Adk) catalyses the reversible phosphoryl transfer reaction (Fig. 1) in the presence of magnesium and has the primary role of maintaining the energy balance in cells. Adk consists of three subdomains which are defined as the AMPbd, the ATPlid and the CORE (Fig. 2) [1]. Adk also shows residual activity without magnesium, which are the experimental conditions used in our studies. During catalysis, Adk utilizes a random bi-bi mechanism [2; 3] and the phosphoryl transfer chemistry is likely to follow an associative mechanism [4]. During its catalytic cycle, Adk undergoes large conformational perturbations when AMP or ATP is bound, directly linked to enzymatic catalysis [5]. The exchange between the open and closed Adk conformations occurs in the µs-ms timescale and is rate-limiting for its catalytic function [6]. Adk from *E.coli* has previously been solved in the open [5] and Ap5A [7] bound closed conformation [8] using X-ray crystallography. Throughout the years, Adk has proved to be a very successful model system for dynamic and functional studies [6; 9; 10], and its biological importance is reflected in its ubiquitous presence in many different organisms and tissues [11].

$$AMP + ATP \underset{Mg^{2+}}{\overset{Mg^{2+}}{\rightleftharpoons}} 2\,ADP$$

Fig. 1. The catalytic mechanism of adenylate kinase.

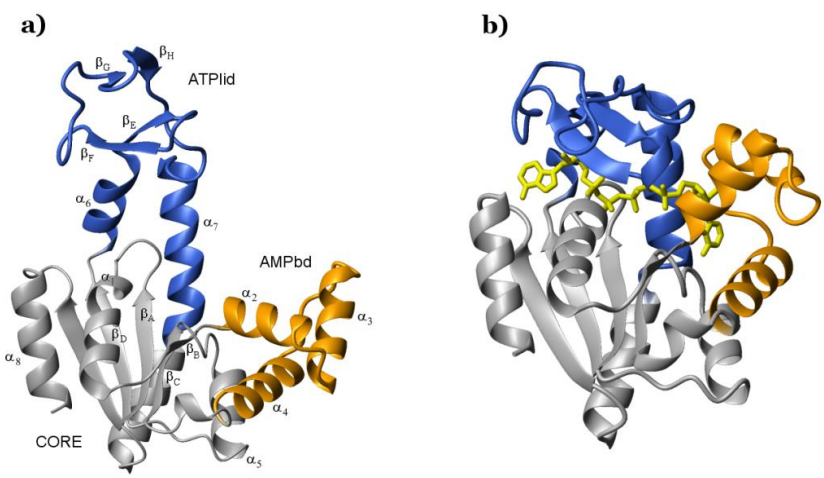

Fig. 2. Crystal structures of *E. coli* adenylate kinase in the open and closed conformation. Adk consists of three subdomains: the CORE, ATPlid (defined as residues 113-176) and AMPbd (residues 31-72) a) The 2.2 Å structure of Adk displayed in the substrate-free state (4AKE.pdb), and b) the 1.9 Å structure of Adk bound to the inhibitor Ap5A colored in yellow (1AKE.pdb).

4.3 Peroxiredoxin Q

Peroxiredoxin Q is a protein found in plants and bacteria and belongs to a group of antioxidant enzymes denoted as thioredoxin- or glutaredoxin-dependent peroxidases catalyzing the reduction of peroxides [12; 13]. *Arabidopsis thaliana* PrxQ is a sequence homolog to the bacterioferritin co-migratory proteins, Bcps [14], as well as Prx from *Aeropyrum pernix* [15], which belongs to a group of atypical Prxs due to the two intermolecular disulfides and its monomeric state. So far 10 Prxs have been found in *A. thaliana*, which can be divided into four subgroups: 1-Cys Prx, 2-Cys Prx, Prx II and PrxQ. These proteins all contain two catalytic cysteines with the exception of 1-Cys Prxs. All Prx enzymes however include a conserved N-terminal cysteine required for catalysis, often referred as the peroxidatic cysteine (C_P) [16; 17]. In 2-Cys Prx enzymes, activation also involves an additional cysteine, the resolving cysteine, or C_R, forming a disulfide bond during their catalytic cycle [17]. The exact cellular location of PrxQ is still debated. It has previously been proposed by studies in *A. thaliana* that the enzyme is positioned close to photo-system II on the stromal side of the grana thylakoids [18]. Other reports suggest that the subcellular location of the enzyme is the thylakoid lumen [19].

The catalytic step starts with a nucleophilic attack of the peroxidatic cysteine, forming a sulfenic acid intermediate with a subsequent release of water. An intramolecular disulfide bond is then created by interaction of the resolving cysteine. This disulfide is finally reduced by thioredoxin, and the reaction cycle starts over [12; 20]. An illustration of the catalytic cycle in 2-Cys Prx from poplar PrxQ is shown in Fig. 3, based on the model proposed by Rouhier and co-workers in 2004 [12].

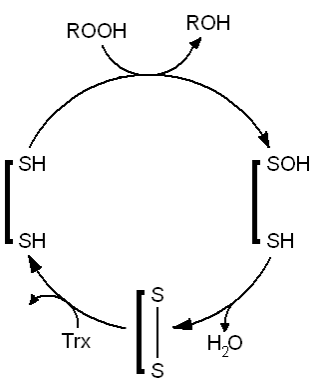

Fig. 3. The catalytic mechanism of PrxQ. This model is based on the mechanism of poplar peroxiredoxin Q described previously by Rouhier and co-workers [12]. In the presence of peroxides, catalysis is initiated by oxidation of the peroxidatic cysteine, forming a sulfenic acid intermediate followed by a release of water. The reaction continues with the resolving cysteine, forming an intramolecular disulfide bridge. The cysteines are then reduced by thioredoxin, and the reaction can start over.

In paper III, solution state NMR was used to study the link between structure and function of PrxQ isolated from *A. thaliana*. To construct a working model of PrxQ in both redox states, structures were generated with Swiss-MODEL [21; 22; 23] based on the crystal data of the Bcp homolog *A. pernix* (2CX3.pdb and 2CX4.pdb). The oxidized and reduced structures of PrxQ are displayed in Fig. 4. These models were validated with NMR using the C_α chemical index and circular dichroism (CD) spectroscopy, serving as a reliable model of *A. thaliana* PrxQ. As judged from these structures, it appears that the active part of the enzyme is involved in a large conformational rearrangement upon catalysis. This is an interesting observation since it can be accessed with solution state NMR and relaxation experiments. The reduced state is indeed populating a more compact conformation which promotes formation of helix 1, harboring the catalytic cysteins. In this state, both the C_P and C_R cysteines are buried into the hydrophobic part of the molecule. In the oxidized state the cysteines are solvent-accessible forming a disulfide bridge.

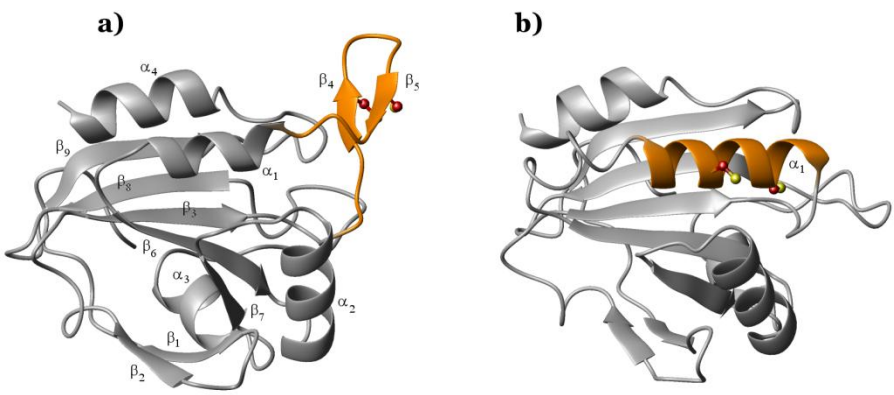

Fig. 4. Structural models of *A. thaliana* PrxQ in the oxidized and reduced state. a) In the oxidized state, the catalytic cysteines form a disulfide bond (illustrated with a ball-and-stick representation) exposed towards the solvent. b) In the reduced state, these residues are protected in an α-helical arrangement. Residues converting between an open and closed conformation upon catalysis are colored in orange. These models were made using the Swiss-MODEL protein structure homology-modeling server with the crystal structures of the Prx homolog *A. pernix* K1 peroxiredoxin (Mizohata *et. al*, unpublished data) used as templates.

4.4 The ribosomal protein S16

The ribosome is a protein-RNA complex divided into two subunits (16S and 30S) binding mRNA upon protein translation. S16 is one of the small proteins involved in the ribosomal assembly, belonging to the small ribosomal 16S subunit. There exist more than 20 proteins building up the small subunit, all of relatively small size and therefore well suitable for biophysical characterization using NMR techniques. The structure of S16 from *Thermus thermophilus* was previously solved using solution NMR [24], and is also similar to the crystal structure of the *A. aeolicus* S16 reported by us in paper IV, here presented in Fig. 5. The difference is two additional β-strands at the C-terminus in S16 from *A. aeolicus*. The exact function of S16 is still under investigation. Studies of the *Escherichia coli* 30S ribosome have shown that S16 induces large conformational changes in the 16S rRNA, crucial for ribosomal assembly [25; 26]. Also, binding of S5 and S12 is strongly dependent on S16 [25]. For the work presented in paper IV, a hyperthermophilic and mesophilic homologous model system was established using a recombinant expression system based on S16 from *Aquifex aeolicus* and *Chlamydia pneumoniae*. Using this model system, we show that the increased thermostability of *A. aeolicus* S16 is linked to residual structure in its unfolded state.

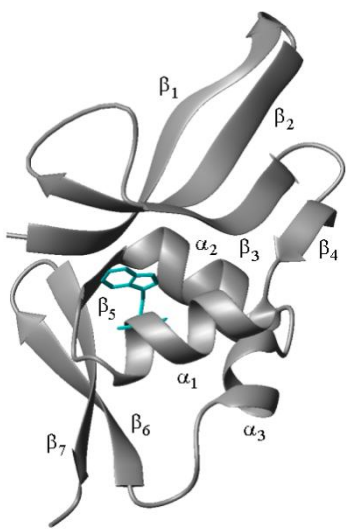

Fig. 5. The 2.0 Å crystal structure of S16 from *A. aeolicus*. S16 is a monomeric 112 residue globular protein consisting of three α-helices and seven β-strands. The side chain of Trp 58 (important for fluorescence experiments) is highlighted in cyan, and is completely buried in the structure.

5 Protein stability

5.1 Protein folding and the two-state model

The living cell is built up by thousands of proteins responsible for very different tasks. In order to perform their functions, proteins must fold into their three-dimensional structure determined by the primary sequence [27]. This process leaves very little room for error, since misfolded proteins can be associated with severe diseases, such as Alzheimers or Creutzfeldt-Jakob disease [28; 29]. Due to this connection, the area of protein folding has gained a lot of interest in the last decades, although it is still one of the biggest unsolved questions in chemistry. We still do not fully understand how proteins fold.

In the end of the 1960's, Cyrus Levinthal stated that folding of a small protein would require billions of years to complete if the process was completely random due to the high number of conformers available in an unfolded polypeptide [30]. However, it has been shown that some proteins can fold within milliseconds [31; 32], so protein folding seems to follow a pre-determined trajectory. In its unfolded state, the polypeptide chain is unstructured and contains both hydrophobic and polar amino acids. This is an energetically unfavored state since non-polar residues want to escape from the solvent-exposed atmosphere. During folding, interactions between amino acids are formed, burying these non-polar residues, known as the hydrophobic effect [33; 34], which is very important for protein folding [33]. The general concept of protein folding can be described by a two-state mechanism where the protein is either folded (N) or unfolded (U) in an "all-or-nothing" process (Eq. 1 and Fig. 6).

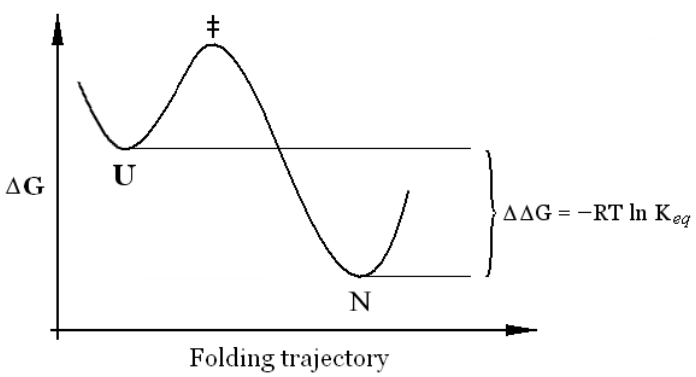

Fig. 6. The two-state folding pathway. The relationship between free energy and ground state energy levels is the energy difference between U and N calculated according to Eq. 2. This model assumes no stable intermediate present through the folding pathway.

Also, formation of intermolecular hydrogen bonds in the folded state contributes significantly to the overall stability [35]. One way to energetically describe protein folding is by the energy landscape theory where the unstructured polypeptide chain can adopt a variety of conformations of higher energy states [36]. As the protein folds, fewer possibilities of conformational freedom are available and the free energy of the system is decreased.

$$U \underset{k_u}{\overset{k_f}{\rightleftarrows}} N \quad (1)$$

The free energy, ΔG, for a system under equilibrium can be calculated according to the following relationship:

$$\Delta G = -RTlnK_{eq} \quad (2)$$

In Eq. 2, R is the gas constant, T the temperature in Kelvin, and K_{eq} the equilibrium constant, which can be calculated from the fractions of the native and unfolded populations according to:

$$K_{eq} = \frac{[U]}{[N]} \quad (3)$$

The free energy in a protein is also a balance between entropy and enthalpy using the relationship in Eq. 4.

$$\Delta G = \Delta H - T\Delta S \quad (4)$$

ΔH is the difference in enthalpy and ΔS the change in entropy. The enthalpy can be described as the energy gained by chemical bonding within the structure, whereas entropy is the level of conformational freedom of the protein. The free energy in a protein is a delicate balance between entropy and enthalpy acting against each other. Due to this relationship, a protein is only marginally stable [37].

5.2 Protein denaturation

Under native conditions the folded conformation is strongly favored, i.e. $k_f >> k_u$. In order to observe and measure kinetics and biophysical parameters we need to control the folding process. There are many substances that can be used to destabilize protein molecules, but the most commonly used are urea or guanidinium hydrochloride. The mechanism behind denaturation by these compounds has been investigated for decades [38]. It is suggested that the main reason for denaturant-induced protein unfolding is the difference in accessible surface area between the folded and unfolded state [39]. One way of retrieving the free energy of a system under native conditions is

to use the Linear Extrapolation Method (LEM) [40] which assumes a linear relationship between the free energy of unfolding and the denaturant concentration. By linear regression ΔG_{H_2O} can be extracted, also yielding the slope or m-value corresponding to the denaturant dependence of ΔG [41] (Eq. 5). It has been experimentally shown that the m-value is strongly correlated with the difference in accessible surface area between the folded and unfolded state, ΔASA [42; 43].

$$\Delta G = \Delta G_{H_2O} - m[urea] \qquad (5)$$

5.3 The unfolded state

Since the unfolded state in most cases is difficult to study and was for a long time assumed to only consist of a complete random coil structure, it gained less attention than the folded state. However, observations that even small proteins can contain residual structure in their unfolded state [44] clearly show the importance of unfolded conformations. The difference in heat capacity between the folded and unfolded state, ΔCp, can be obtained by calorimetry or fitting a protein stability curve to the Gibbs-Helmholtz equation (Eq. 6) [45]. The curvature of a fitted stability curve is determined by ΔCp. Both the denaturant dependence and the m-value, as well as ΔCp reflect the burial of non-polar groups [46]. A protein can adopt a high melting temperature (T_m) in several ways [47; 48] (Fig. 7). Lowering the ΔCp will make the T_m higher by flattening the curve. Shifting the curve to higher temperatures will make the maximum stability higher, as well as the T_m. Another approach is to shift the stability curve towards a higher temperature maximum with a maintained ΔCp value. Thermophilic proteins apparently use a variety of strategies to increase their T_m, in many cases using all these mentioned methods [49].

$$\Delta G_{N-D}(T) = \Delta H_m(1 - T/T_m) - \Delta Cp[(T_m - T) + T\ ln(T/T_m)] \qquad (6)$$

In Eq. 6 [45] $\Delta G_{N-D}(T)$ is the Gibbs energy change at zero denaturant. T_m is the temperature at the denaturation midpoint and ΔH_m the enthalpy change at T_m. ΔCp is the difference in heat capacity between the folded and unfolded state.

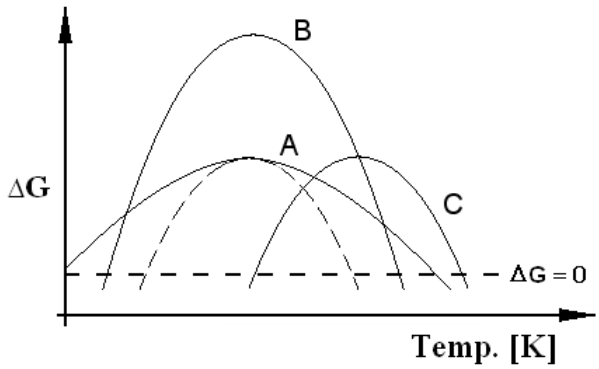

Fig. 7. Thermodynamic mechanisms to obtain a high T_m. The melting temperature is determined by the curvature of the stability curve, ΔCp, and the maximum stability (dotted line). T_m is reached when $\Delta G = 0$. A) Lowering ΔCp will make T_m higher, but the maximum stability remains the same. B) The maximum stability is increased but ΔCp remains the same, resulting in a higher T_m. C) Here the maximum stability is maintained, but T_m is increased due to a shift of the entire temperature interval.

Since the backbone topology and predicted fold between homologous proteins are very similar [50], a reduced ΔCp and ΔASA value is a strong indication of residual structure in the unfolded state. This is an intriguing observation since homologous proteins can differ remarkably in their resistance to thermal unfolding (see Paper IV). Today there exists no exact answer on how thermophilic proteins gain their amazing thermostability in contrast to their mesophilic counterparts, but many ideas have been proposed [49; 51; 52; 53], such as optimization of hydrogen bonding [54] or ion networks [55; 56], all an important denominator for protein stability.

5.4 Enzyme dynamics

Enzymes are remarkable biocatalysts that can enhance the rate of chemical reactions several orders of magnitude by lowering the activation barrier associated with the chemical reaction. For instance, ortodine 5'-phosphate decarboxylate enhances the rate of 1-methylortoic acid decarboxylation with a factor of 10^{17} compared to the uncatalyzed rate in aqueous solution [57]. The barrier is lowered through high binding affinity (K_d) between the enzyme and the transition state configuration of the altered substrate. For example, ortodine 5´-phosphate decarboxylate has been estimated to bind the transition state compound with a K_d of $5 \cdot 10^{-24}$ M [57]. Detailed chemical mechanisms for many enzymes have been elucidated by combining information from three-dimensional structures with classical enzymology. One of the most well studied enzyme groups are the serine proteases, and it is established that the peptide bond cleavage is catalyzed with a catalytic triad containing one aspartic acid residue, one histidine residue and one serine residue located in close proximity in the structure.

Enzymatic catalysis is a dynamic process. The catalytic step, during which the chemical reaction occurs, involves motion along the reaction coordinate in conjunction with bonds being broken and formed. Binding and release of substrates often occur with conformational changes that can be of varying amplitude. During the last few years there have been increasing efforts to probe motions and fluctuations of the actual catalyst (i.e. the enzyme). Protein motions occur on many timescales (see Fig. 8) ranging from fast (ps-ns) motions associated with bond vibrations and side chain rotations to slow motions in the μs-ms timescale. Catalysis, ligand binding and also protein folding often occur in the μs-ms timescale [39]. Therefore understanding of enzyme motions on this timescale is important in order to seek for possible couplings and correlations between dynamics and catalysis. There exist different biophysical methods that can be applied to probe motions on the μs-ms timescale, for instance single molecule methods and NMR spectroscopy. NMR has proven to be particularly useful, especially in light of the development of relaxation dispersion experiments [58; 59]. These experiments can in favorable cases provide residue specific information on kinetics, thermodynamics (populations) and structure (chemical shifts) of low populated states present during enzymatic reaction cycles.

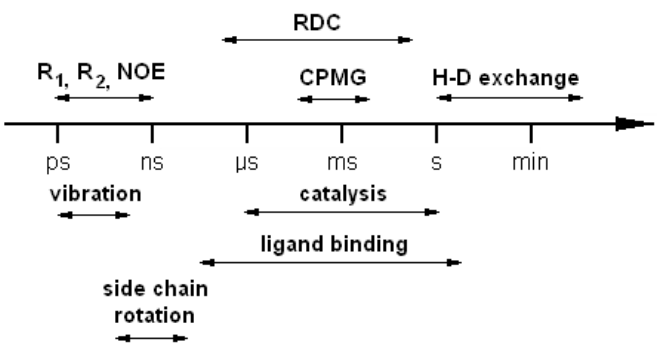

Fig. 8. Different molecular timescales detectable with solution state NMR spectroscopy [60]. Virtually all molecular processes can be observed, covering picoseconds dynamics to backbone proton-exchange mechanisms with rates of minutes or even days.

Here below some enzymes are described where a clear link between enzyme dynamics and catalysis has been established. One must, however, be careful with terminology. If a correlation between slow dynamics and the rate of catalysis is found, this does not necessarily mean that dynamics is affecting the rate of the actual chemical step. In a recent theoretical publication it was explicitly stated that slow conformational dynamics can not affect the rate enhancement of the chemistry catalyzed [61].

Dihydrofolate reductase (DHFR): DHFR catalyses the reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate (THF) by using NADPH as a cofactor (electron donor). THF is required for synthesis of purine that is used as a building-block in RNA and DNA. The enzyme cycles through five sequential major intermediates during the catalytic cycle. By using ^{15}N and ^{1}H relaxation dispersion measurements it was shown that each intermediate samples high energy states that resemble the ground states of the next intermediate in the reaction cycle [62]. This "conformational sampling" occurs on the μs-ms timescale. The substrate-free enzyme samples a high energy state that resembles the conformation of the ligand-bound first intermediate. This observation is consistent with "conformational selection" as the mechanism for ligand binding by DHFR. It was also found that the rates for hydride transfer and release of product are rate-limited by the conformational dynamics. Clearly, there exists a strong linkage between the catalytic efficiency and dynamics in DHFR.

RNase A: RNase A is an enzyme that catalyses the cleavage of the P-O5 bond in single-stranded RNA. The rate-limiting step for RNase A catalysis is the product release that occurs with a rate of around 2000 s^{-1} [63]. In an extensive relaxation dispersion study it was observed that RNase A display conformational fluctuations for residues lining the active site on the same timescale both in the apo- and substrate-bound state [64]. It was suggested that the conformational fluctuations in this case are dictating the rate of product release. Interestingly, the conformational dynamics that is required for catalysis is present already in the substrate-free state. It therefore appears that conformational dynamics is an intrinsic property of RNase A and that these motions are "harvested" for catalysis once the substrate is bound.

Cyclophilin A (Cyp A): CypA belongs to the family of prolyl-isomerases that catalyze the cis-trans isomerization of prolyl peptide bonds. Again by utilizing relaxation dispersion experiments it was shown that Cyp A residues lining the active site display μs-ms motions with frequencies close to the catalytic turn-over rate in both the apo and substrate-bound states [65]. In analogy to RNase A it was proposed that the catalytic motions are present already in the absence of substrate and that these motions are utilized for catalysis.

Adenylate kinase: The reversibility of the reaction (Fig. 1) enables quantification of dynamics on an active enzyme. It was shown with NMR relaxation dispersion experiments that the rate-limiting step for Adk catalysis is the opening of the substrate-binding domains ATPlid and AMPbd [6]. This observation also explains why a hyperthermophilic Adk variant displays low activity at ambient temperatures compared to mesophilic Adk. In hyperthermophilic Adk the rate of subdomain opening is significantly slower compared to mesophilic Adk. Single-molecule fluorescence resonance energy transfer experiments have shown that the closed Adk conformation is sampled in the absence of substrate [66].

Conclusions from these studies: A general theme in the studies mentioned above is that conformational rearrangements can be rate-limiting for catalysis. Covalent chemistry has been speeded up by evolution to the point that conformational fluctuations have become rate-limiting. A second general observation is that enzymes may sample high energy structural states that resemble ligand-bound conformations. This feature is consistent with a "conformational selection" mechanism for ligand binding. In the conformational selection model a native protein samples many different conformations (i.e. rugged energy landscape) and the conformation that is closest to the "bound conformation" will interact with a specific binding partner. This mechanism has been suggested for ubiquitin through advanced NMR approaches [67]. Using the same model system it was recently proposed that conformational selection is followed by residual "induced fit" rearrangements for residues lining the binding site [68]. It must, however, be pointed out that interconversion between different protein conformations in absence of ligand was proposed already in 1965 in the MWC model for cooperativity [69]. The third observation is that the actual timescale of conformational dynamics in ligand-free states can be preserved in substrate-bound states. These motions are utilized by the enzyme during substrate turn-over to perform the conformational rearrangements required for catalysis.

6 Biophysical techniques

6.1 CD spectroscopy

CD spectroscopy can be used to probe secondary structure content in macromolecules by measuring the difference in absorption between left- and right-handed circular polarized light, caused by asymmetry in the molecule [70]. In an unfolded protein this means no (or weak) signal; whereas a structured protein yields a CD signal deviating from zero. CD spectroscopy is a versatile technique which can be utilized for many purposes such as studies of protein folding, kinetics, ligand binding, and to retrieve structural information. It is also an alternative to fluorescence methods when a fluorescence probe (tryptophan) is not available, which otherwise would provide insufficient fluorescence signals. For protein studies, two main wavelength windows are normally studied to measure secondary (far-UV) and tertiary structure (near-UV), where far-UV is the most commonly used. See Fig. 9 for typical far-UV CD spectra.

Far-UV: The CD signal (between ~ 190-250 nm) is affected by the peptide bond. α-helices, β-turn and random coil conformations give characteristic CD bands.

Near-UV: The CD signal (between ~ 250-350 nm) is affected by aromatic residues and disulfide bonds in the polypeptide, and thus, is sensitive to the tertiary structure.

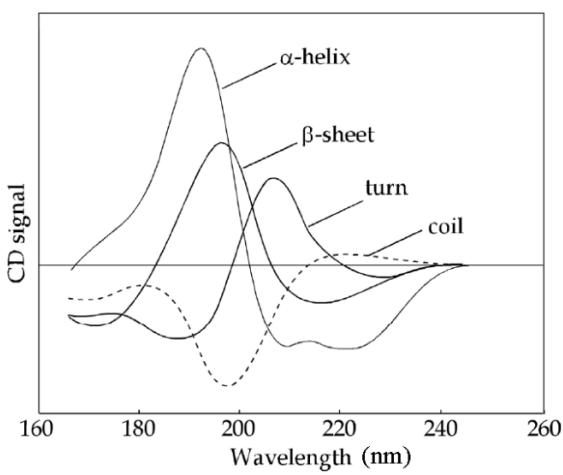

Fig. 9. An example of different secondary structures detectable in the far-UV region [71]. The CD signal yields characteristic patterns for each type of secondary structure. The signal amplitude is proportional to the sample concentration according to Lambert-Beers law.

Far-UV CD spectroscopy can also be used to measure thermal unfolding, giving the unfolding enthalpy and T_m of a system. A fit of the CD signal response as a function of temperature can be performed using Eq. 7 and 8 [72]. Here CD_{mdeg} is the CD signal at a specific temperature, and S_f and S_u is the CD signal for the folded and unfolded state, respectively. a is the slope of the folded baseline, and b the slope of the unfolded baseline. In Eq. 8 ΔH_m is the enthalpy obtained at T_m.

$$CD_{mdeg}(T) = \frac{\left(S_f + aT + K_{obs}(S_u + bT)\right)}{1 + K_{obs}} \quad (7)$$

$$K_{obs}(T) = e^{\left(\frac{\Delta H_m}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right)} \quad (8)$$

6.2 Fluorescence spectroscopy

Fluorescence is a phenomenon which occurs when a chromophore is excited at a specific wavelength and as the system relaxes back to equilibrium, light is emitted at a longer wavelength. In the folded state, the fluorescent probe (such as an aromatic residue) is normally positioned in a hydrophobic environment. As the micro-environment of the chromophore changes due to protein unfolding (increased polarity), the fluorescence emission shifts to a higher wavelength (red shift) and also decreases in amplitude as the signal is quenched by the solvent. Due to the relationship between the chemical environment and protein denaturation, this enables studies of protein folding and kinetics. The majority of the fluorescence signal arises from emission of aromatic residues such as tryptophan, tyrosine or phenylalanine, dominated by tryptophan fluorescence which makes Trp residues very useful for protein studies [73; 74]. Protein concentrations of around 1 μM can be analyzed, making fluorescence spectroscopy a very sensitive method.

Fluorescence spectroscopy can also be used for distance measurements and single molecular studies using FRET (Förster Resonance Energy Transfer). This technique can be utilized for intramolecular distance measurements [75; 76] when a donor and acceptor molecule are present in the structure. The use of naturally occurring chromophores is in this case a big advantage since the protein is not modified. However, mutagenic approaches can be used to introduce chromophores of interest, which makes fluorescence spectroscopy a versatile method.

6.3 Stopped-flow spectroscopy

The use of stopped-flow spectroscopy enables quantification of folding kinetics by measuring the fluorescence emission in a protein (discussed in the previous section) as a function of time at various denaturant concentrations. As many proteins contain tryptophan residues they can be used as fluorescence probes for global folding studies. The most used denaturants for the stopped-flow experiment are urea or guanidinium hydrochloride, both shifting the equilibrium towards the unfolded state (see Eq. 1).

The typical stopped-flow experiment consists of two steps:

1) The protein is kept in a native buffer and is allowed to unfold at increasing denaturant concentrations.

2) The protein is kept unfolded and is allowed to refold in buffer with decreasing denaturant concentrations (requires a reversible system).

In a stopped-flow experiment fluorescence can be used as a detection method. The observable is the fluorescence emission of an aromatic residue present in the polypeptide. Folding kinetics can be obtained by plotting the difference in amplitude as an exponential increase or decrease as the chemical environment of the sample changes. For a two-state folding protein, the rate constants should be linearly dependent on the denaturant concentration [77; 78] (Eq. 9 and 10).

$$\ln k_u = ln{k_u}^{H_2O} + m_u \times [D] \quad (9)$$

$$\ln k_f = ln{k_f}^{H_2O} - m_f \times [D] \quad (10)$$

Here [D] is the denaturant concentration and m_f and m_u the denaturant dependence of the rate constants [40]. ln $k_{u}{}^{H_2O}$ and ln $k_{f}{}^{H_2O}$ are the extrapolated rate constants in water. For a two-state system, the observable kinetic rate constant, ln k_{obs} ($k_{obs} = k_f + k_u$) can be fitted to a V-shaped chevron plot using the linear relationship between the denaturant concentration and kinetic rates [39].

$$lnk_{obs} = ln\left(e^{(lnk_f - m_f[D])} + e^{(lnk_u + m_u[D])}\right) \quad (11)$$

The *m*-value of the folding reaction is related to its denaturant dependence. Eq. 12 is therefore valid under the assumption of a two-state mechanism where a linear correlation exists between the rate constants and the denaturant concentration.

$$m_{D-N} = m_u - m_f \quad (12)$$

Given the *m*-value of a system in the folded and unfolded conformations, the compactness of the transition state can be calculated using the Tanford value (Eq. 13) [39], which is used for phi-value analysis [39; 79].

$$\beta = \frac{m_f}{m_{D-N}} \tag{13}$$

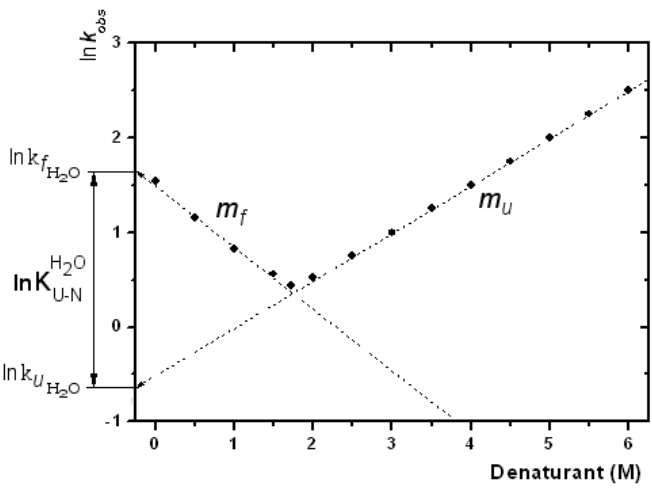

Fig. 10. The chevron plot. By unfolding a folded protein and vice versa, the kinetic rate constants can be calculated based on the change in fluorescence amplitude as a function of time. In a two-state system, the rate constants and denaturant concentrations are linearly dependent on each other. Extrapolation to native conditions yields the folding and unfolding rates and thermodynamic stability of the system.

Since the thermodynamic stability of a two-state system can be calculated with Eq. 2 and 3, extrapolation of rate constants also yields the free energy according to:

$$\Delta G_{D-N} = -RT\ ln\left(k_f{}^{H_2O}/k_u{}^{H_2O}\right) \tag{14}$$

The two-state model is a fundamental description of protein folding, but the true picture can be much more complicated. Especially for bigger proteins, the protein can fold via a more complex pathway, causing the chevron plot to deviate from the true linear two-state relationship. As the *m*-value is found to be connected to the accessible surface area between the native and denatured state [42], the origin of curvature in a kinetic experiment might be linked to residual structure that are preserved in these states.

7 NMR theory

A major advantage of NMR is that it provides residue-specific information and can monitor events on many timescales from picoseconds up to days (Fig. 8) depending on the experiment. The drawbacks are that it requires highly concentrated samples and that the possibility to study large protein molecules is fairly limited. When only looking at small systems such as organic compounds, the information provided by the simplest 1D experiment is normally sufficient. However, when studying bigger systems such as macromolecules, the high number of resonances complicates the spectra. The size of the molecule also affects the tumbling rate which contributes negatively to the signal amplitude. In the last decades, methods to study larger systems such as proteins and protein complexes have been greatly refined [80]. Rapid developments of new pulse programs and achievements in computer technology have extended the methods to analyze proteins. The invention of multi-dimensional NMR spectroscopy and cryogenic systems where the probe and preamplifier are cooled down to minimize electronic noise enable NMR studies at lower protein concentrations. Today, proteins of around 200 residues can be fairly easily studied using routine experiments with two or more dimensions.

7.1 The NMR signal

Many nuclei have a property called spin, defined by the quantum number (I) of the nucleus. When a sample is inserted into an NMR magnet, the energy states of the nuclei are split (Fig. 11), resulting in a very small net energy difference, ΔE, between the states. It is the small energy difference that makes NMR a relatively insensitive technique. ΔE is proportional to the magnetic field B_0 and increases with a stronger field. The sensitivity of NMR increases with ΔE, making higher fields desirable for NMR studies. The ability to detect a certain nucleus is determined by the gyromagnetic constant and its natural abundance if $I \neq 0$. Since ΔE is proportional to the gyromagnetic constant γ, nuclei with larger γ result in larger NMR signals. A large number of NMR active nuclei can be observed, but the most commonly used in biological NMR are ^{1}H, ^{13}C, ^{15}N and ^{31}P, all with $I = ½$. Modern NMR magnets can reach field strengths of around 24 T, corresponding to a proton frequency of 1 GHz.

When positioned in the magnet, the spins tend to align parallel to the magnetic field, B_0, along the z-axis. If a spin system of $I = ½$ is positioned along B_0, two energy states are formed ($\alpha = ½$ and $\beta = -½$) with opposite magnetic moments, resulting in a small net energy difference (Fig. 11). Since slightly more spins populate the lower energy state, this results in a bulk magnetization (M_z). To create transverse magnetization, the M_z component is tilted from the z-plane by radio-frequency (RF) pulses. The

magnetization starts to rotate about the z-axis in the x-, y-plane, generating a current in the detection coil positioned around the sample. The precession frequency ω (known as the Larmor frequency) can be expressed as $\omega = -\gamma B_o$. By Fourier transformation, the data can then be converted from time domain data to interpretable NMR signals in a frequency spectrum.

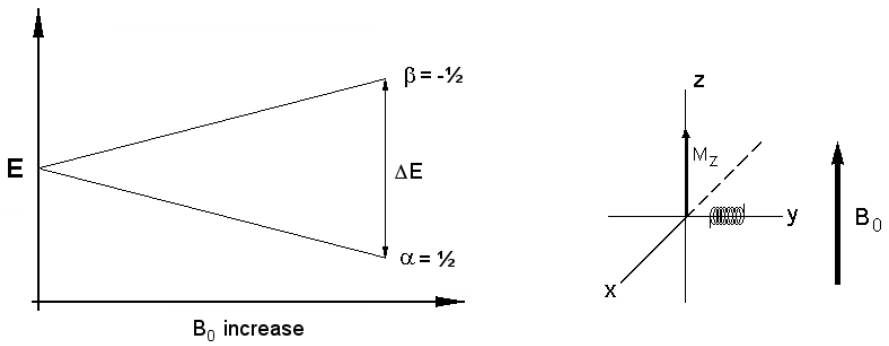

Fig. 11. Energy splitting in a nucleus with spin ½. Left) The magnetic moment is split into two energy states (α and β) when positioned in a static field B_o. A small difference in energy, ΔE, is formed, favoring the lower energy state. The obtained energy difference increase as B_o gets stronger. Right) A vector model representing the M_z magnetization resulting from ΔE and its alignment in the z-plane. Transverse magnetization is created by applying RF pulses at the resonance frequency, tilting the bulk magnetization from z. The rotation induces a current in the detection coil which generates the NMR signal.

7.2 The rotating frame

The RF pulse that tilts the bulk magnetization away from B_0 is short, usually a few μs in duration. Tilted away from z, the nuclear spins now rotate about the z-axis with the Larmor frequency, ω. We can simplify the description of the precession around the z-axis by introducing a laboratory frame that rotates with the frequency ω_0. If the Larmor frequency is equal to ω_0 ($\omega_0 = \omega$), the spin magnetization appears to be static in the rotating frame since they move with the same speed (Fig. 12).

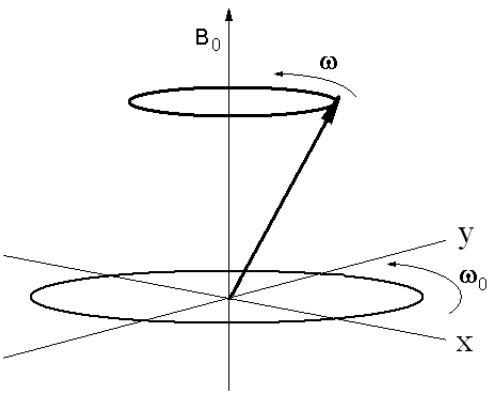

Fig. 12. Larmor precession along the rotating frame. A short RF pulse that rotates perpendicular to B_0 tilts the bulk magnetization from the z-axis. This causes the magnetization to rotate around the z-axis with the frequency ω. If we introduce a laboratory frame rotating with the frequency ω_0, its precession around the z-axis appears to be static when $\omega_0 = \omega$.

7.3 The chemical shift

The exact resonance frequency of a nucleus is dependent of its electronic environment and serves as a marker of it. Chemical shifts can serve as a fingerprint for a compound since each nucleus is shielded differently in the structure. When a nucleus is positioned in the magnet, its electrons generate local magnetic fields that interfere with the externally applied field. The effect of the electron shielding is very small, and the divergence is only a few ppm from the reference frequency. As B_0 differs between NMR spectrometers and so does the chemical shift when measured in Hz, a standardized scale was introduced to normalize the chemical shift frequency from Hz to parts per million, ppm. This causes the chemical shifts to be independent of the spectrometer frequency, and NMR data can be directly compared regardless of the field strength. For chemical shift calibration, an internal standard can be introduced together with the sample. For this purpose, different compounds can be

used, such as tetramethylsilane (TMS) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

Chemical shifts can also directly predict secondary structure in proteins, which is the basis behind the ^{13}C chemical shift index [81]. In this experiment, the carbon shifts in a protein sequence are recorded and compared to chemical shifts expected for residues in a random coil structure.

7.4 Chemical shift anisotropy

Chemical shift anisotropy (CSA) is caused by an uneven electron density around the nuclei, causing the resonance frequency to vary according to the orientation of the molecule in an external magnetic field, which also affects the relaxation rate. In solution state NMR the orientation dependence is averaged out by rapid tumbling of the nucleus in the sample in contrast to solid state NMR. For stronger fields the CSA effect contributes increasingly to the relaxation.

7.5 J couplings

J couplings or scalar couplings are mediated through chemical bonds and are essential for magnetization transfer from one nucleus to another in many NMR experiments. J couplings cause the NMR signals to split into two or more lines depending on the number of neighboring atoms close to the observed nucleus. This is a very useful tool in 1D NMR to identify unknown chemical compounds. If required, the effect of J couplings can also be removed which is often the case in heteronuclear experiments when the resonances otherwise would overlap. This can be achieved with a broadband decoupling sequence such as GARP or WALTZ-16, involving continuous irradiation of the nucleus with a designed set of pulses.

In protein NMR, J couplings can also be used to retrieve secondary structure information. For example, the $^3J_{HNHA}$ coupling is dependent on the torsion angles of the backbone conformation, Φ. In the backbone of an amino acid, Φ describes the rotation about the NH - H_α bond. The connection between torsion angles and J coupling is defined by the Karplus relationship [82] (Eq. 15), where A, B and C are experimentally derived constants. Calculations of $^3J_{HNHA}$ couplings can be accomplished with the HNHA experiment [83].

$$^3J(\Phi) = A\cos^2(\Phi) + B\cos(\Phi) + C \qquad (15)$$

7.6 Dipolar couplings

Dipolar couplings are caused by the interaction of two magnetic dipoles and depend on the distance (r) between two interacting spin systems in an r^{-3} fashion, and the angle relative to the B_0 field (Fig. 13). In solution NMR, dipolar couplings are normally averaged out to zero by rapid isotropic tumbling of the sample whereas their contribution in solid state NMR is highly significant. There exists however methods to introduce partial alignment of the NMR sample and thereby re-introduce dipolar couplings. This can be achieved with different methods discussed below.

7.7 Residual dipolar couplings

The obtained dipolar coupling between two nuclei is dependent on the bond angle of the intermolecular axis relative to the external magnetic field (Fig. 13), a feature which can be used for structural analyses [84; 85].

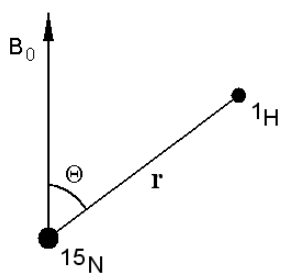

Fig. 13. The dipolar interaction is distance and alignment dependent. The obtained dipolar coupling is dependent on the distance such as r^{-3} between two spins, and the angle θ in respect to the static magnetic field along the z-axis. In isotropic solution θ varies due to molecular tumbling.

By introducing partial alignment in a sample, dipolar couplings are not averaged out completely, instead they become dependent on the three-dimensional structure (see Fig. 14). There are several ways to introduce partial anisotropic alignment in a sample [86]. Two methods are use of stretched polyacrylamide gels [87; 88] or bacteriophages [89]. The RDC can be calculated according to the following relationship:

$$D_{(\theta,\varphi)} = D_a\left[(3cos^2\theta - 1) + \frac{3}{2}Rsin^2\theta cos2\varphi\right] \qquad (16)$$

Here D_a is the axial vector component and R describes its rhombicity [90; 91]. The parameters θ and φ are dependent on the direction of intermolecular vectors in the alignment frame. The observable RDC is the average over internal motion of the intermolecular vectors, and therefore θ and φ are averaged by motions occurring on

the ns-ms timescale [92]. RDC measurements can be performed using the IPAP ^{1}H-^{15}N HSQC pulse sequence [93].

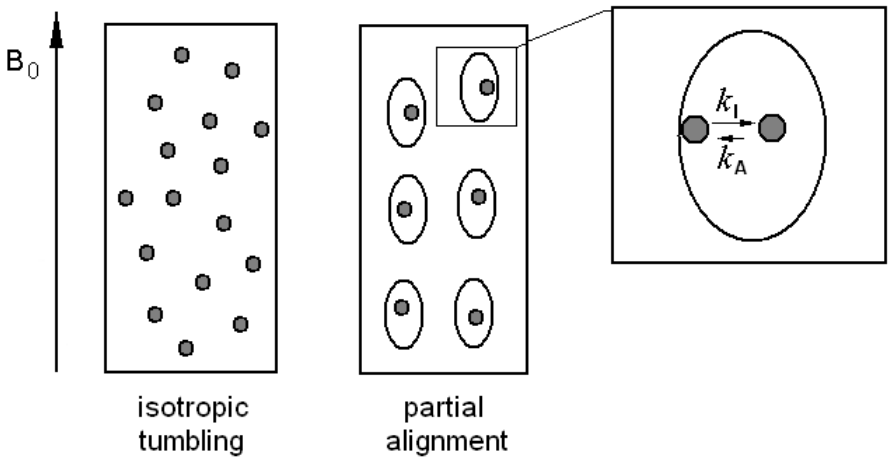

Fig. 14. Partial anisotropic alignment achieved in a stretched polyacrylamide gel. Left) In a polyacrylamide gel there exist spherical cavities which can be filled with protein molecules. The protein in these cavities will sense an isotropic (i.e. buffer-like) environment and RDCs will be averaged to zero. Under these conditions, the sample rotates randomly in solution. Right) To introduce anisotropy, the gel can be stretched. This causes the cavities to assume an elongated shape. A protein can interact with the surface of the gel, and under anisotropic conditions the observed RDC will depend on the fraction of protein molecules at the interface (given by the rate constants k_I and k_A in the expansion).

7.8 Relaxation

Relaxation is the process whereby the magnetization returns to a thermodynamic equilibrium distribution. The relaxation process is very dependent on the interaction between a spin and other surrounding spin systems and is an important probe for dynamic processes covering time spans from picoseconds to milliseconds. In an NMR experiment, the signal intensity is strongly dependent on the tumbling rate in solution, or the rotational correlation time, τ_c. The smaller the molecule, the faster it rotates, corresponding to a short τ_c and a sharper signal. In general τ_c is in the order of 4-12 ns for small proteins. There are many contributing factors for relaxation, such as fluctuating fields caused by molecular motion, dipolar interactions, CSA and paramagnetic species. For a long time, intermolecular motions that were slower than τ_c could not be observed in NMR since the overall tumbling of the molecule causes averaging of the bond orientation. The use of residual dipolar couplings has filled this gap. Spin relaxation can be divided into two categories which are dependent on each other and occur at the same time: T_1 and T_2 relaxation.

7.9 T_1 relaxation

T_1 or longitudinal relaxation is the process by which the z-component of the magnetization along the z-axis (M_z) returns to equilibrium, and is sensitive to very fast molecular motions on the ps-ns timescale. T_1 relaxation can be measured using the inversion recovery experiment. In this experiment a 180° pulse is applied followed by a delay τ, a 90° pulse and acquisition. By alteration of τ, the T_1 relaxation time can be determined. In practice, T_1 relaxation also sets the time limit for how fast an experiment can be repeated. If the pulsing is too fast, the z-magnetization does not have time to reach equilibrium, resulting in a reduced NMR signal. A delay of approximately 5 times T_1 is required for the magnetization to relax back to z before next pulse is applied to ensure that equilibrium is reached.

7.10 T_2 relaxation

T_2 or transverse relaxation is affected by ps-ns and μs-ms motions and describes how the magnetization in the x-, y-plane (M_x, M_y) returns to equilibrium, and is an important property for studies of protein dynamics. For a folded protein of around 150 residues, T_1 relaxation occurs in the order of one second, whereas T_2 relaxation lasts for about 50-80 milliseconds. The relaxation rate can be quantified by measuring the signal intensity which decays exponentially as the transverse relaxation dephase (Eq. 17). Here I is the initial intensity, t the relaxation delay and I_0 is the intensity at a relaxation delay t.

$$I = I_0 \, e^{-R_x \cdot t} \tag{17}$$

The contribution to R_2 ($R_2 = 1/T_2$) is mainly the sum of the contribution from chemical exchange (R_{ex}) and the intrinsic R_2 rate of the molecule itself, R_2^0 (Eq. 18) [94]:

$$R_2 = R_2^0 + R_{ex} \tag{18}$$

For a two-state system under fast exchange (see Fig. 15) transversal relaxation can be used to retrieve dynamic information under equilibrium conditions (Eq. 19). P_A and P_B are the populations of the folded and unfolded states, and Δω is the difference in chemical shift between those. k_{ex} is the exchange rate between P_A and P_B.

$$R_{ex} = \frac{P_A P_B (\Delta\omega)^2}{k_{ex}} \tag{19}$$

7.11 CPMG relaxation dispersion

In general, conformational exchange on the µs-ms timescale causes the NMR signal to dephase or weaken. ^{1}H-^{15}N CPMG corrected relaxation dispersion can be used to recover resonance-broadened transverse magnetization useful for structural studies on the µs-ms timescale [58; 95; 96]. In this experiment a series of 180° CPMG spin-echo pulses are applied during a fixed time τ_{cp}, enhancing the signal-to-noise by refocusing the magnetization dephasing [94]. The measured relaxation rate constant R_2 is dependent on the number of spin-echo pulses during τ_{cp}. In the intermediate exchange regime, the fitted parameters from this experiment yield the population distribution, the difference in chemical shift between those, and the exchange rate, $k_{ex,}$ of the populations [94]. The individual exchanging rates for a protein with relaxation dispersion can be achieved with a detection limit of around ≥ 0.5 % of the low-populated species, making it very sensitive for low-populated states [97]. CPMG dispersion has been shown to be an extremely useful approach for studies of low-populated states and protein dynamics.

7.12 Hydrogen-deuterium exchange

As non-hydrogen bonded protons are constantly exchanging with the solvent, their exchange rates are dependent on their protection level and bond strength. Since ^{2}H has a different gyromagnetic ratio than ^{1}H, it is invisible in a ^{1}H-^{15}N HSQC experiment due to the difference in resonance frequency. This means that this setup is suitable for monitoring amide proton exchange as the signal decay over time, which can vary between seconds to even days. The apparent exchange rate is heavily dependent on the pH in the solution (Eq. 21), [98] and can be altered by adjusting buffer conditions to fit a convenient laboratory timescale. The hydrogen exchange rate of a certain amino acid in the three-dimensional structure is dependent on the opening and closing rates according to Eq. 20 [99]:

$$NH_{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} NH_{open} \xrightarrow{k_{ch}} NH_{exchanged} \qquad (20)$$

k_{op} and k_{cl} refer to the opening and closing rate of protected hydrogen groups, respectively. k_{ch} is the intrinsic exchange rate in an unfolded polypeptide chain, which is affected by neighbouring residues, pH and buffer conditions [98; 100]. The hydrogen exchange mechanism can either follow an EX1 ($k_{cl} << k_{ch}$) or EX2 ($k_{cl} >> k_{ch}$) mechanism [101]. By plotting the NMR signal intensity in the EX2 regime as it decays over time, it is possible to quantify the thermodynamic stability and determine kinetics of intermediate folding species on a residue-specific level [102; 103]. A method to obtain EX1 conditions in a hydrogen exchange system is to increase the pH since k_{ch} is base-catalyzed (Eq. 21) [101]. The amide proton of an amino acid can also be

exchanged through local fluctuations [100]. These fluctuations are insensitive to the denaturant at low concentrations since the difference in accessible surface area for these residues is very small, but they still contribute to the free energy of the protein. Local fluctuations are still poorly understood but can be defined as "local breathing" of secondary structure elements [104], and observed by plotting ΔG_{HX} [105] at low denaturant concentrations.

$$k_{ch} = k_{int}\,[OH^-] \tag{21}$$

In most cases, $k_{cl} >> k_{ch}$ is true, i.e. EX2 conditions [106], where the exchange rate is pH dependent. Here the exchange rate can be described as in Eq. 22 [101]:

$$k_{ex} = \frac{k_{op} k_{ch}}{k_{cl}} \tag{22}$$

This means that in the limit of EX2, the apparent exchange rate for an amino acid is determined by the equilibrium constant K_{op} for disruption of hydrogen bonds multiplied with the intrinsic rate [102]. In the EX2 regime, Eq. 22 can be reduced to:

$$k_{ex} = \frac{k_{op}}{k_{cl}} k_{ch} = K_{op} k_{ch} \tag{23}$$

The free energy of a particular amino acid in a protein is then directly given by the relationship between Eq. 2 and 23 as shown below [103; 106]:

$$\Delta G_{HX} = -RT ln K_{op} = -RT ln\left(\frac{k_{ex}}{k_{ch}}\right) \tag{24}$$

7.13 The NOESY experiment

As seen throughout this introduction, NMR is a versatile tool covering a broad time span from intramolecular motions to reactions involving chemical exchange of hours or more (see Fig. 8). However, until now one of the most important tools in protein NMR has been left out: the NOESY. It is an experiment to probe intramolecular distances and can also be used for structural assignment since an α-helix and β-strand have different bonding patterns. The signal intensity of the NOE depends on through-space interactions between nuclei due to cross relaxation. The relaxation effect strongly vary depending of the distance between the nuclei as the NOE $\propto r^{-6}$, where r is the distance between two spin systems and NOE is the observable signal intensity. The observed peak volume is related to the distance between two nuclei, i.e. the lower the intensity, the larger is the distance. The sensitivity of the experiment can be used to probe distances of around 5 - 6 Å or shorter as the dipole-dipole interaction rapidly decreases with increased distance.

The 2D NOESY pulse program consists of three 90° pulses. First, a 90° pulse is applied followed by t1, which is the evolution time for the chemical shift labeling in the first dimension. A second pulse is then applied to transfer magnetization to the z-axis followed by the mixing period. During the mixing period magnetization is exchanged by cross relaxation. In the final pulse, magnetization is flipped back to the transverse plane and finally detected. The mixing time is normally set to around 100 milliseconds. Too long mixing times can cause spin diffusion which distorts the relaxation due to energy exchange with distant spin systems.

7.14 The NMR timescale

It is not always possible to observe protein dynamics as the exchange rate can interfere with what is technically possible to detect with NMR. Conformational exchange occurs at different timescales and can either be slow, intermediate or fast on the NMR timescale. Fig. 15 illustrates the term "NMR timescale" based on the line shape of NMR signals here represented by two states, A and B.

Slow on the NMR timescale: A sample with two exchanging conformations, A and B, is separated according to their resonance frequencies as ν_A and ν_B, both visible in the NMR spectrum. In this case, the difference in frequency between these is much larger than the exchange rate.

Fast on the NMR timescale: The rate of interconversion between the two states is much faster than the difference in frequency between A and B, leading to an unresolved NMR spectrum. The NMR spectrum represents the signal average between A and B, which is the population weighted average between ν_A and ν_B (Eq. 25). P_X here represents the population of state x, and x the chemical shift of that state.

$$\Delta\nu_{obs} = P_A\nu_A + P_B\nu_B \qquad (25)$$

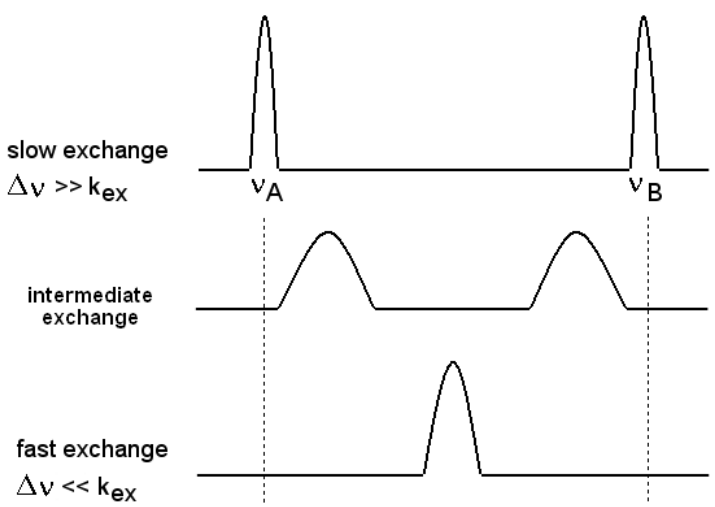

Fig. 15. The NMR timescale. In the slow exchange regime k_{ex} is much slower than the difference in frequency between state A and B, separating the signals as sharp peaks in the NMR spectrum. As the exchange rate increases, the signals dephase and drop in intensity. In the limit of fast exchange (~ 10 kHz) the observable signal is given by the population average between A and B (in this example $P_A = P_B$).

7.15 Assignment procedures

The simplest NMR experiment is the 1D, where only one 90° pulse is applied, followed by detection and Fourier transformation. Since proteins are large molecules, the 1D NMR experiment cannot provide very detailed information due to large signal overlap. The approach for working with protein NMR instead requires multidimensional techniques separating the chemical shifts. This process requires 2D and 3D experiments and the introduction of other nuclei such as carbon 13 and nitrogen 15. Since the natural abundance of ^{1}H is 99.98 %, whereas carbon and nitrogen contain very little of the NMR-active isotopes ^{13}C (1.11 %) and ^{15}N (0.37 %), isotope enrichment is required which can be done by recombinant over-expression in a medium supplemented with the desired isotopes.

In biological NMR, the most used experiment is the ^{1}H-^{15}N HSQC, which yields the backbone chemical shift correlation between a proton and its attached nitrogen for each amino acid in the sequence, except for prolines and the first N-terminal amide. In a fully decoupled HSQC one single peak is generated by each amino acid in the primary sequence with its corresponding chemical shift, serving as a fingerprint of a protein structure. To identify the observable resonances, the HSQC spectrum needs to be assigned. There is a number of NMR experiments designed for this purpose, where the most used are the 3D experiments HNCA [107], HNCACB [108] and CBCA(CO)NH [109].

Looking at the backbone structure in a polypeptide, the chemical shifts of the C_α and C_β atoms can together be used as a marker for the amino acids in a polypeptide [81]. HNCACB provides shift information for $C_{\alpha\, i,i-1}$ and $C_{\beta\, i,i-1}$, where i is an amino acid in the primary sequence. In the CBCA(CO)NH experiment, the magnetization is transferred only via the carbonyl, and is selective only for $C_{\alpha\ i-1}$ and $C_{\beta\ i-1}$, distinguishing amino acids that connect to each other through the peptide bond. With its greater sensitivity due to the less complex magnetization transfer, HNCA [107] is used as complementary experiment together with the others, yielding the chemical shift of $C_{\alpha\ i,\ i-1}$. This approach is a very valuable strategy for assignments of an unknown protein structure. Even the redox state of a cysteine can be distinguished with its characteristic C_β shift.

8 Summary of papers

8.1 **Paper I** - NMR Identification of Transient Complexes Critical to Adenylate Kinase Catalysis

Aim of the study

Since molecular motions are a prerequisite for enzymatic functions, conformational dynamics and catalysis are intimately linked together. In adenylate kinase catalysis is occurring with large conformational movements where both nucleotide-binding domains undergo large subdomain motions. Here we use a solution state NMR approach to understand the underlying mechanism for substrate-binding in Adk.

Methods and results

Solution structures of apo- and ADP-bound states: In this paper and paper II, the subdomains binding ATP and AMP are denoted ATPlid and AMPbd (Fig. 2). The crystal structures of both the open (apo) state and Ap5A (closed) state were previously solved by crystallography to 2.2 Å and 1.9 Å, respectively [5; 8]. We were able to show with residual dipolar couplings that ligand-free Adk mainly populates the crystallographic open state. Moreover, we show that Adk in complex with both Ap5A and ADP mainly populate the crystallographic closed state. These results are significant since chemical shifts in apo- and ADP-saturated states can be used as probes of the fully open and closed conformations.

Binding of ATP or AMP does not resemble the fully closed state: Addition of the natural substrates ATP and AMP to Adk from *E. coli* (Eadk) causes chemical shift perturbations at the expected binding sites. The amplitude of the chemical shifts is however smaller than expected from fully closed ATPlid and AMPbd (fully closed chemical shifts are derived from the ADP-saturated state as explained above).

Chemical shifts suggest population averaging: We analyzed the reduced chemical shift perturbation amplitude in detail for both the AMP- and ATP-saturated state. Interestingly, for residues in the ATPlid the chemical shifts in ^{1}H-^{15}N spectra of the ATP-saturated state falls onto a straight line between the open (apo) and closed (ADP-saturated) states. This behavior is indicative of population-weighted averaged chemical shifts (i.e. the shifts in the ATP state is a linear combination of the shifts in the open and closed states). A related scenario is observed for residues in the AMPbd in response to AMP-saturation, where the chemical shifts in the AMP-saturated state fall onto a straight line between open and closed states. To analyze all residues

simultaneously we developed a global method to fit all data in the respective subdomain. In conclusion, all residues in the ATPlid are interconverting between the open and closed conformation in presence of saturating ATP concentrations (Fig. 16). Notably, the populations of the open and closed states are almost equal. Likewise, all residues in the AMPbd are interconverting between open and closed conformations when saturated with AMP (Fig. 16). Based on measurements of the dissociation constants (K_d values) for ATP binding we were able to show that there exists intramolecular crosstalk between the ATPlid and AMPbd.

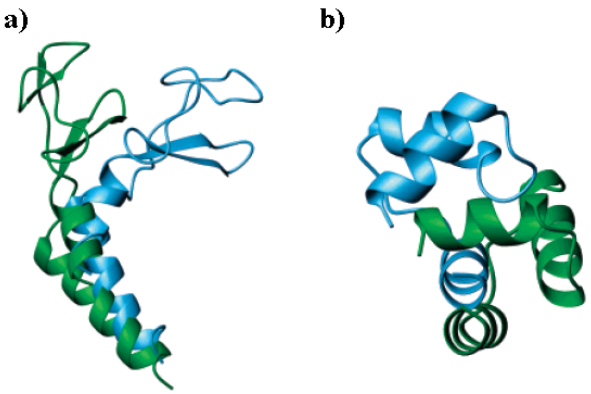

Fig. 16. Transient structures formed during Adenylate kinase catalysis. a) Addition of ATP causes the ATPlid to fluctuate between the open crystallographic (green) and closed (blue) conformations with almost equal populations. b) Structural fluctuations in response to AMP binding.

Conclusions

We have discovered that Adk binds the substrates AMP and ATP with a highly dynamic mechanism. In the ATP-saturated state the ATPlid is interconverting between open and closed conformations with an almost equal population distribution. A very similar scenario is observed for AMP binding where the AMPbd also is converting between open and closed conformations (again with almost equal weights). This highly flexible ligand binding mode contrast the commonly accepted inference that enzyme-ligand complexes are static low-entropy states. Based on the (close to) zero difference in free energy between the open and closed single nucleotide-bound states, we propose that Adk has the innate property to sample the closed conformation already in the absence of substrate. Only in the presence of both ATP and AMP will both nucleotide-binding domains populate the fully closed conformation. This cooperative nature of subdomain closure is explained structurally with domain-domain interactions, but also with interaction of R156 in the ATPlid to both the ATP- and AMP substrates.

8.2 **Paper II** - Non-cooperative Folding of Subdomains in Adenylate Kinase

Aim of the study

In paper I we discovered a highly dynamic ligand binding mechanism in adenylate kinase. In paper II, we studied the protein folding mechanism of Adk from mesophilic (*E. coli*) and hyperthermophilic (*A. aeolicus*) organisms primarily using NMR methods. Here we gain information about the cooperativity and thermodynamics involved in the folding of Adk, and take a first step to probe the potential coupling between protein folding and functional energy landscapes.

Methods and results

Structural distributions of thermodynamic stability: The local thermodynamic stabilities (ΔG_{HX}) of Eadk and Aadk were quantified using hydrogen-deuterium exchange experiments. The two enzymes display a striking similarity in the spatial distribution of thermodynamic stabilities (Fig. 17), with the only difference being that Aadk is overall more stable. In both enzymes the flexible nucleotide-binding motifs display significantly reduced ΔG_{HX} values compared to the CORE subdomain, and in Eadk many residues in ATPlid and AMPbd exchange completely within the dead-time of the experiment. We were able to quantify ΔG_{HX} values for many fast exchanging residues by stabilizing the enzymes with Ap5A or ADP. The results are an indication that Adk folds with a mechanism that is more complex than two-state.

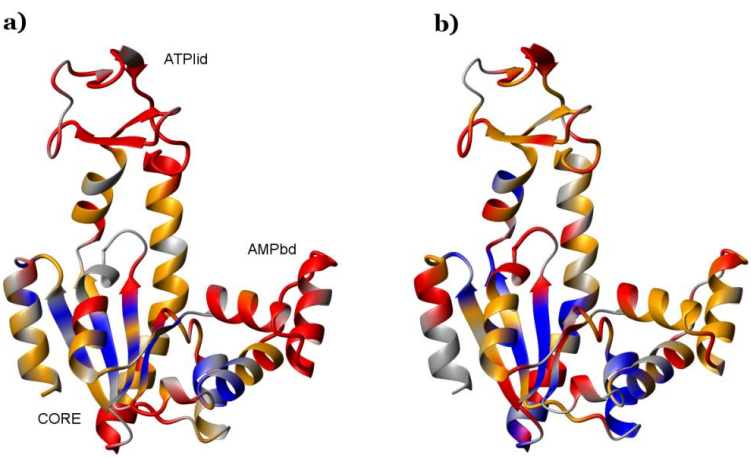

Fig. 17. Stability values for adenylate kinase obtained with hydrogen-deuterium exchange. Local stabilities are shown on the open crystal structure (4AKE.pdb) for a) Eadk, and b) Aadk. Blue) ΔG_{HX} > 29 kJ mol^{-1}, orange 29 > ΔG_{HX} > 11 kJ mol^{-1}, and red) residues exchanging within the dead-time of the experiment. Overlapping, unassigned or proline residues are colored in grey.

Subdomain folding in Eadk: It has been shown that some proteins consist of structural units that fold cooperatively (foldons) [110]. Foldons can be identified with hydrogen-deuterium exchange experiments as a function of denaturant concentration [105]. We used this approach to investigate the foldon substructure in Eadk. The denaturant dependencies of ΔG_{HX} were used to identify two foldons in Eadk corresponding to: 1) CORE and parts of helix 4 in the AMPbd, 2) helices α_7 and α_8 in the ATPlid. Together with the hydrogen exchange data acquired under native buffer conditions an emerging picture is that Adk folds in a non-cooperative manner with the CORE representing the most stable substructure.

The ATPlid or AMPbd subdomains can fold independently of the CORE domain: The observation that both the nucleotide-binding domains show lower ΔG_{HX} values than the CORE, together with the foldon substructure of Eadk is consistent with a model where ATPlid or AMPbd can fold and unfold on an otherwise folded CORE domain. To test this idea experimentally we substituted key non-polar clusters in the ATPlid (Ile116, Val117, Val164 and Leu168) or AMPbd (Val39, Ala49 and Met53) with glycine residues to induce local unfolding in these sections. If correct, selective unfolding of either the ATPlid or AMPbd should be accommodated without provoking unfolding of the remaining parts of the enzyme. By analyzing $^{13}C_\alpha$ chemical shifts and transverse relaxation lifetimes (T_2) we were able to verify our hypothesis. In summary, we showed that Adk folds in a non-cooperative manner, and importantly the ATPlid and AMPbd can fold/unfold while the CORE remains stably folded.

Conclusions

We have shown that the individual subdomains in Adk fold and unfold in a non-cooperative manner. The non-cooperative nature of Adk subdomain folding may have an important functional consequence. Notably, the functional open-to-closed transition in Adk is accompanied by significant rearrangements in backbone hydrogen bonding patterns in both the ATPlid and AMPbd. Non-cooperative subdomain folding allows these rearrangements without provoking global unfolding of the entire molecule which would be expected if Adk would follow a two-state ("all-or-nothing") folding mechanism. Our results are an indication that the protein folding landscape may be important for, or overlapping with the functional energy landscape.

8.3 **Paper III** - *Arabidopsis thaliana* Peroxiredoxin Q is extraordinary dynamic on the µs-ms timescale

Aim of the study

We have used NMR to study the structure and dynamics of the plant enzyme Peroxiredoxin Q from *A. thaliana*. The aim was to understand the molecular details of its substrate-binding and enzymatic function, and to generate a model which can be applied for Prx enzymes in general.

Methods and results

PrxQ is verified as an atypical 2-Cys enzyme: Our homology models based on the crystal structures of the PrxQ homolog *A. pernix* reveal that the enzyme undergoes a significant conformational change during catalysis (Fig. 4). An initial fluorescence study was performed together with NMR spectroscopy and kinetic experiments, providing us with a detailed view of the enzyme. PrxQ was recombinantly expressed and purified using routine protocols in our lab, and elutes as a monomeric protein as confirmed with NMR experiments and size exclusion chromatography. The monomeric state supports the idea that PrxQ is an atypical 2-Cys Prx enzyme. To obtain the reduced state, a stoichiometric excess of the reducing agent TCEP was added under an argon atmosphere prior to the NMR experiments. In the experimental conditions used throughout the study, 1 mM of DTT was used to keep the thioredoxin reduced. It has previously been shown that DTT itself cannot work as an electron donor in the absence of Trx [18] which also was confirmed by our control experiments.

The structural models of PrxQ are valid representations of the enzyme: Since no structure of PrxQ from *A. thaliana* currently is available, a sequence modeling approach was used to construct the structures of PrxQ by protein modeling using the PrxQ homolog *A. pernix* as a reference (Mizohata *et. al*, unpublished data) for both redox states. The convergence between our models and the experimental data was verified by the C_α chemical shift index and NOE connectivities together with CD spectroscopy. In addition, biophysical characterizations using CD spectroscopy show that the melting temperature of the reduced state is five degrees higher compared to the oxidized state, which is fully consistent with our constructed models. As shown from the PrxQ models as well as for *A. pernix* and other Prx homologs [15; 20], the reduced state harboring the catalytic cysteins is folded into a helical structure which creates a more compact state [15]. This is also reflected by the higher enthalpy of the reduced state (95 kJ mol^{-1}), that can be explained by additional binding contacts.

PrxQ undergoes excessive conformational exchange in both oxidation states: By comparing the ^{1}H-^{15}N HSQC spectra of the oxidized and reduced enzyme, conformational dynamics on the μs-ms timescale was detected due to intense line-broadening. CPMG relaxation dispersion experiments of $PrxQ^{red}$ show that the conformational dynamics occur on the μs-ms timescale. 132 of a total of 140 residues in the oxidized state could be assigned using triple-resonance NMR experiments, whereas the number decreased to only 63 for the reduced state. The amino acid residues undergoing conformational exchange are mainly localized in the catalytic site of the protein in both redox states. The intense dynamics were further supported using R_1 and R_2 relaxation, where nearly all residues in the catalytic site are line-broadened beyond detection. In the oxidized state almost all residues are detectable, but many show increased relaxation rates which is an indication of conformational exchange. We further studied the relaxation parameters R_1, R_2 and steady-state NOEs using the Lipari-Szabo modelfree formalism [111; 112; 113], where the R_2/R_1 distributions correspond to a correlation time of 9.3 and 10.2 ns for $PrxQ^{ox}$ and $PrxQ^{red}$, respectively. Estimations of correlation values for PrxQ using HYDRONMR [114] yielded τ_c values as expected for a monomeric protein of the size of PrxQ. These findings support the previous conclusions using size exclusion chromatography that PrxQ is monomeric.

Catalysis may be linked to local unfolding of the enzyme: We noticed that PrxQ is active even at temperatures exceeding its melting temperature. One intriguing explanation for this observation is that helix 1 in PrxQ undergoes local unwinding to expose its cysteine residues to the substrate, coupled to its function. This idea is further supported by looking at the structural model (Fig. 4), where the cysteine residues are partially buried in helix 1.

Conclusions

PrxQ has been studied in both oxidation states using a variety of NMR experiments together with classical biophysical techniques. The protein folds with an apparent two-state transition as judged from kinetic and equilibrium experiments. The R_1, R_2 and relaxation dispersion data provided with this report clearly show that the protein undergoes significant dynamics and conformational exchange on the μs-ms timescale in both oxidation states. Modelfree [115; 116] analyses show that PrxQ is monomeric under our experimental conditions. We suggest that the conformational exchange in PrxQ depends on structural transitions to one or more states, enabling local unwinding of helix 1, coupled to its enzymatic function.

8.4 **Paper IV** - Extreme Temperature Tolerance of a Hyperthermophilic Protein Governed by Residual Structure in the Unfolded State

Aim of the study

For many applications, understanding the underlying mechanism between protein stability and function is required to construct enzymes surviving high temperatures with maintained activity. To gain knowledge about the molecular details behind extreme thermostability, the folding kinetics and thermodynamic parameters of two structural homologs were compared: the ribosomal protein S16 from a mesophilic and hyperthermophilic organism.

Methods and results

High difference in thermostability between the S16 homologues: The crystal structure of S16 from *A. aeolicus* was solved to a resolution of 2.0 Å using X-ray crystallography. The ^{1}H-^{15}N HSQC spectra of both homologs were assigned using a combination of triple-resonance NMR experiments. The C-terminal residues 103 – 112 of the mesophilic protein are concluded to be unstructured in solution due to the lack of NOE magnetization transfer and intense peaks in a ^{1}H-^{15}N HSQC spectrum. Equilibrium fluorescence experiments were used to determine the stability of both proteins which was 24.8 kJ mol^{-1} and 13.8 kJ mol^{-1} for the hyperthermophilic and mesophilic variant, respectively. We determined the T_m of the mesophilic S16 to 59°C using CD spectroscopy. Due to the extreme stability of the thermophilic protein, this approach was not feasible. In this case the melting temperature was estimated by fitting the stability curves to the Gibbs-Helmholtz equation [45] (Eq. 6), yielding a T_m of 111°C. Residue-specific stabilities of the thermophilic protein were also obtained using hydrogen-deuterium exchange experiments, showing that the hydrophobic core is the most stable part of the structure.

The folding kinetics of S16 from *A. aeolicus* deviates from a two-state mechanism: By using stopped-flow spectroscopy experiments we confirmed that the extreme thermostability of the thermophilic S16 can be explained by its much slower unfolding rate. Kinetic studies show that the mesophilic S16 follows a two-state mechanism whereas the folding of the thermophilic protein is more complex. To gain insight in the difference in thermostability in more detail, we analyzed the thermodynamics and linked it to the relationship between ΔCp and ΔASA.

High thermostability linked to residual structure in the unfolded state: ΔCp can be measured by fitting the free energy of unfolding as a function of temperature to Gibbs-Helmholtz equation [45]. We used this approach to compare the ΔCp values between the two homologs. It was observed that the ΔCp is almost two-fold larger for

the mesophilic protein; this discrepancy can be linked to residual structure in the unfolded state since ΔCp reflects the burial of non-polar groups [46]. To test the hypothesis that residual structure is present in the thermophilic protein, two mutants were constructed: a stabilizing V73L meso and a destabilizing I20V thermo mutant. The free energy and ΔCp values between these were measured using fluorescence spectroscopy. We observed that the mesophilic mutant V73L is stabilized by 1.7 kJ mol^{-1} whereas the stability of I20V is reduced by 6.3 kJ mol^{-1}. An interesting observation was that the chevron plot of V73L shows the same type of curvature as detected for the thermophilic wild-type protein, and the same observation is true for the I20V mutant. In addition, the calculated ΔCp values are also lower for V73L and higher for I20V compared to the wild-type. Our results clearly show a link between mutant thermostability and differences in heat capacity when comparing the folded and unfolded state. Due to the link between ΔASA and ΔCp, we have shown that a viable strategy to manipulate protein thermostability is to identify parameters altering ΔCp.

Intramolecular measurements confirm a compact unfolded state: The fluorescence emission from a Trp residue can be used to excite a BODIPY probe for intramolecular distance measurements [75]. This approach was used to study the unfolded state of the thermophilic S16 by Förster [76] resonance energy transfer from Trp58 to BODIPY linked to residue 10. It was observed that the distances between these probes are shorter in the thermophilic protein under unfolded conditions, and the distance distribution is wider in the mesophilic S16. These results support our view where the thermostable S16 is more compact in its denatured state.

Conclusions

We have shown by kinetic experiments together with thermodynamics that the increased thermostability of *A. aeolicus* S16 is linked to residual structure in its unfolded state, and also that the unfolding rate between these proteins are significantly different. By constructing mutants from both thermophilic and mesophilic S16, we were able to alter the thermostability by affecting the ΔCp value. The relationship between intramolecular distances and compactness of the unfolded state investigated with FRET experiments further supports the idea where compactness in the unfolded state affects ΔCp. Taken together, we hope that our observations can be a viable starting point when constructing thermostable proteins and enzymes that are used for various applications and for commercial use.

9 Summary and outlook

We have made contributions at the forefront of the field of enzyme dynamics and the coupling between dynamics and enzymatic activity. Specifically we have shown that the ligand binding mechanism of Adk follows a highly flexible, previously not observed, mechanism. We have also made initial experiments to probe if the protein folding and functional energy landscapes in the enzymes Adk and PrxQ are overlapping. Our data is pointing towards such overlap. It is established that enzyme dynamics on the slow μs-ms timescale is intimately linked to catalysis in a small subset of enzymes that have been studied with NMR spectroscopy. A remaining open question is the actual molecular mechanism underlying conformational exchange processes. In order to fully comprehend dynamics it is necessary to understand the underlying molecular details. Adk provides an excellent system to study the mechanisms of nucleotide-binding, domain-opening and closure. It has been suggested from simulations that the open/closed transition in Adk is accompanied with local unfolding in well-defined regions of the enzyme [117]. By producing selected Adk mutants this mechanism could be tested with NMR spectroscopy and other biophysical techniques.

The first event in enzymatic catalysis is the formation of transient substrate-enzyme encounter complexes, which are inherently difficult to study using conventional techniques. Complete understanding of enzymatic reaction cycles demand knowledge about the three-dimensional structures of encounter complexes. Adk again seems to be a suitable model system where strategic mutations that stabilize an ATP-bound encounter complex probably can be designed. In the long-term perspective, scientists can use the knowledge gathered from studies of dynamics to design enzymes with novel properties.

10 Acknowledgements

Först av allt vill jag naturligtvis tacka min handledare **Magnus** för det samarbete vi haft genom åren.

Gerhard Gröbner, min biträdande handledare, för trevliga diskussioner och umgänge.

Marcus Wallgren, för allt intensivt arbete med proteinrening, sena kvällar/nätter då vi for in på labbet och tog hand om prover som ”bara måste” in i kylen.

Christoph Weise, for both fun and serious chatting about everything, och för trevligt arbete tillsammans.

Ulrika Olsson, för två roliga och trevliga år i gruppen.

Tobias Sparrman, för underhåll av NMR-spektrometrarna samt intressanta fikadiskussioner.

Alexander Christiansen, för arbete med PrxQ-projektet.

“Tips-gänget” **Michael Hall** och **Patrik Storm**, för en uppsjö ”tior” (utan utdelning).

Katarina Wallgren och **Viola Tegman**, för att ni tog hand om allt inför kursstart och såg till att reagenser alltid fanns till hands.

Anders Öhman, för att du tog dig tid och handledde mitt examensarbete så jag kom in på NMR-spåret som mer eller mindre bidrog till denna avhandling.

Göran Karlsson på Swedish NMR Centre för NMR-tid på ”600:an och 800:an” i Göteborg.

Alla ni som inte är med på denna lista men som jag utelämnade och som tycker ni borde blivit nämnda.

11 References

1. Shapiro, Y. E. & Meirovitch, E. (2006). Activation energy of catalysis-related domain motion in E. coli adenylate kinase. *J Phys Chem B* **110**, 11519-24.
2. Rhoads, D. G. & Lowenstein, J. M. (1968). Initial velocity and equilibrium kinetics of myokinase. *J. Biol. Chem.* **243**, 3963-3972.
3. Sheng, X. R., Li, X. & Pan, X. M. (1999). An iso-random Bi Bi mechanism for adenylate kinase. *J. Biol. Chem.* **274**, 22238-22242.
4. Abele, U. & Schulz, G. E. (1995). High-resolution structures of adenylate kinase from yeast ligated with inhibitor Ap5A, showing the pathway of phosphoryl transfer. *Protein Sci.* **4**, 1262-1271.
5. Muller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **4**, 147-56.
6. Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z. & Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* **11**, 945-949.
7. Lienhard, G. E. & Secemski, II. (1973). P 1 ,P 5 -Di(adenosine-5')pentaphosphate, a potent multisubstrate inhibitor of adenylate kinase. *J Biol Chem* **248**, 1121-3.
8. Muller, C. W. & Schulz, G. E. (1992). Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state. *J Mol Biol* **224**, 159-77.
9. Shapiro, Y. E., Sinev, M. A., Sineva, E. V., Tugarinov, V. & Meirovitch, E. (2000). Backbone dynamics of escherichia coli adenylate kinase at the extreme stages of the catalytic cycle studied by (15)N NMR relaxation. *Biochemistry* **39**, 6634-44.
10. Hanson, J. A., Duderstadt, K., Watkins, L. P., Bhattacharyya, S., Brokaw, J., Chu, J. W. & Yang, H. (2007). Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc Natl Acad Sci U S A* **104**, 18055-60.
11. Noda, L. (1973). Adenylate kinase. *The enzymes* **In Boyer, P. D. (ed.) VIII, Academic Press, New York, NY**, 279-305.
12. Rouhier, N., Gelhaye, E., Gualberto, J. M., Jordy, M. N., De Fay, E., Hirasawa, M., Duplessis, S., Lemaire, S. D., Frey, P., Martin, F., Manieri, W., Knaff, D. B. & Jacquot, J. P. (2004). Poplar peroxiredoxin Q. A thioredoxin-linked chloroplast antioxidant functional in pathogen defense. *Plant Physiol* **134**, 1027-38.
13. Motohashi, K., Kondoh, A., Stumpp, M. T. & Hisabori, T. (2001). Comprehensive survey of proteins targeted by chloroplast thioredoxin. *Proc Natl Acad Sci U S A* **98**, 11224-9.
14. Kong, W., Shiota, S., Shi, Y., Nakayama, H. & Nakayama, K. (2000). A novel peroxiredoxin of the plant Sedum lineare is a homologue of Escherichia coli bacterioferritin co-migratory protein (Bcp). *Biochem J* **351**, 107-14.

15. Mizohata, E., Sakai, H., Fusatomi, E., Terada, T., Murayama, K., Shirouzu, M. & Yokoyama, S. (2005). Crystal structure of an archaeal peroxiredoxin from the aerobic hyperthermophilic crenarchaeon Aeropyrum pernix K1. *J Mol Biol* **354**, 317-29.
16. Ellis, H. R. & Poole, L. B. (1997). Roles for the two cysteine residues of AhpC in catalysis of peroxide reduction by alkyl hydroperoxide reductase from Salmonella typhimurium. *Biochemistry* **36**, 13349-56.
17. Wood, Z. A., Schroder, E., Robin Harris, J. & Poole, L. B. (2003). Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci* **28**, 32-40.
18. Lamkemeyer, P., Laxa, M., Collin, V., Li, W., Finkemeier, I., Schottler, M. A., Holtkamp, V., Tognetti, V. B., Issakidis-Bourguet, E., Kandlbinder, A., Weis, E., Miginiac-Maslow, M. & Dietz, K. J. (2006). Peroxiredoxin Q of Arabidopsis thaliana is attached to the thylakoids and functions in context of photosynthesis. *Plant J* **45**, 968-81.
19. Petersson, U. A., Kieselbach, T., Garcia-Cerdan, J. G. & Schroder, W. P. (2006). The Prx Q protein of Arabidopsis thaliana is a member of the luminal chloroplast proteome. *FEBS Lett* **580**, 6055-61.
20. D'Ambrosio, K., Limauro, D., Pedone, E., Galdi, I., Pedone, C., Bartolucci, S. & De Simone, G. (2009). Insights into the catalytic mechanism of the Bcp family: functional and structural analysis of Bcp1 from Sulfolobus solfataricus. *Proteins* **76**, 995-1006.
21. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* **37**, D387-92.
22. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201.
23. Peitsch, M. C. (1995). Protein Modeling by E-Mail. *Bio-Technology* **13**, 658-660.
24. Allard, P., Rak, A. V., Wimberly, B. T., Clemons, W. M., Jr., Kalinin, A., Helgstrand, M., Garber, M. B., Ramakrishnan, V. & Hard, T. (2000). Another piece of the ribosome: solution structure of S16 and its location in the 30S subunit. *Structure* **8**, 875-82.
25. Stern, S., Powers, T., Changchien, L. M. & Noller, H. F. (1989). RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* **244**, 783-90.
26. Ramaswamy, P. & Woodson, S. A. (2009). S16 throws a conformational switch during assembly of 30S 5' domain. *Nat Struct Mol Biol* **16**, 438-45.
27. Anfinsen, C. B. (1972). The formation and stabilization of protein structure. *Biochem J* **128**, 737-49.
28. Glatzel, M., Stoeck, K., Seeger, H., Luhrs, T. & Aguzzi, A. (2005). Human prion diseases: molecular and clinical aspects. *Arch Neurol* **62**, 545-52.
29. Hardy, J. & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353-6.

30. Levintha.C. (1968). Are There Pathways for Protein Folding. *Journal De Chimie Physique Et De Physico-Chimie Biologique* **65**, 44-&.
31. Schechter, A. N., Chen, R. F. & Anfinsen, C. B. (1970). Kinetics of folding of staphylococcal nuclease. *Science* **167**, 886-7.
32. Spector, S. & Raleigh, D. P. (1999). Submillisecond folding of the peripheral subunit-binding domain. *J Mol Biol* **293**, 763-8.
33. Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640-7.
34. Dyson, H. J., Wright, P. E. & Scheraga, H. A. (2006). The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc Natl Acad Sci U S A* **103**, 13057-61.
35. Rose, G. D., Fleming, P. J., Banavar, J. R. & Maritan, A. (2006). A backbone-based theory of protein folding. *Proc Natl Acad Sci U S A* **103**, 16623-33.
36. Plotkin, S. S. & Onuchic, J. N. (2002). Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys* **35**, 111-67.
37. Pace, C. N., Trevino, S., Prabhakaran, E. & Scholtz, J. M. (2004). Protein structure, stability and solubility in water and other solvents. *Philos Trans R Soc Lond B Biol Sci* **359**, 1225-34; discussion 1234-5.
38. Schellman, J. A. (2002). Fifty years of solvent denaturation. *Biophys Chem* **96**, 91-101.
39. Fersht, A. (2000). Structure and mechanism in protein science. *W.H. Freeman and Company, New York*.
40. Greene, R. F., Jr. & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* **249**, 5388-93.
41. Pace, C. N. & Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **Suppl 4**, 1-7.
42. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138-48.
43. Livingstone, J. R., Spolar, R. S. & Record, M. T., Jr. (1991). Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry* **30**, 4237-44.
44. Li, Y., Horng, J. C. & Raleigh, D. P. (2006). pH dependent thermodynamic and amide exchange studies of the C-terminal domain of the ribosomal protein L9: implications for unfolded state structure. *Biochemistry* **45**, 8499-506.
45. Pace, C. N. & Laurents, D. V. (1989). A new method for determining the heat capacity change for protein folding. *Biochemistry* **28**, 2520-5.
46. Otzen, D. E. & Oliveberg, M. (2004). Correspondence between anomalous m- and DeltaCp-values in protein folding. *Protein Sci* **13**, 3253-63.
47. Razvi, A. & Scholtz, J. M. (2006). Lessons in stability from thermophilic proteins. *Protein Sci* **15**, 1569-78.
48. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M. & Marqusee, S. (2003). Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci U S A* **100**, 11345-9.

49. Petsko, G. A. (2001). Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol* **334**, 469-78.
50. Robic, S., Berger, J. M. & Marqusee, S. (2002). Contributions of folding cores to the thermostabilities of two ribonucleases H. *Protein Sci* **11**, 381-9.
51. Sterner, R. & Liebl, W. (2001). Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* **36**, 39-106.
52. Hilser, V. J., Gomez, J. & Freire, E. (1996). The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins* **26**, 123-33.
53. Chen, J. & Stites, W. E. (2004). Replacement of staphylococcal nuclease hydrophobic core residues with those from thermophilic homologues indicates packing is improved in some thermostable proteins. *J. Mol. Biol.* **344**, 271-280.
54. Bougault, C. M., Eidsness, M. K. & Prestegard, J. H. (2003). Hydrogen bonds in rubredoxins from mesophilic and hyperthermophilic organisms. *Biochemistry* **42**, 4357-4372.
55. Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H. & Robb, F. T. (1998). Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc Natl Acad Sci U S A* **95**, 12300-5.
56. Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R. & Consalvi, V. (1995). The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147-1158.
57. Radzicka, A. & Wolfenden, R. (1995). A Proficient Enzyme. *Science* **267**, 90-93.
58. Loria, J. P., Rance, M. & Palmer, A. G. (1999). A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for charactherizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121**, 2331-2332.
59. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341-11352.
60. Boehr, D. D., Dyson, H. J. & Wright, P. E. (2006). An NMR perspective on enzyme dynamics. *Chem Rev* **106**, 3055-79.
61. Pisliakov, A. V., Cao, J., Kamerlin, S. C. & Warshel, A. (2009). Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc. Natl. Acad. Sci. U S A* **106**, 17359-17364.
62. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638-1642.
63. Lee, G. C. Y. & Chan, S. I. (1971). P-31 Nmr Study of Association of Uridine-3'-Monophosphate to Ribonuclease-A. *Biochemical and Biophysical Research Communications* **43**, 142.

64. Beach, H., Cole, R., Gill, M. L. & Loria, J. P. (2005). Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state. *J. Am. Chem. Soc.* **127**, 9167-9176.
65. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. & Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-121.
66. Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G. & Kern, D. (2007). Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838-844.
67. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C. & de Groot, B. L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471-1475.
68. Wlodarski, T. & Zagrovic, B. (2009). Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. U S A* **106**, 19346-19351.
69. Monod, J., Wyman, J. & Changeux, J. P. (1965). On the Nature of Allosteric Transitions: a Plausible Model. *J. Mol. Biol.* **12**, 88-118.
70. Johnson, W. C., Jr. (1990). Protein secondary structure and circular dichroism: a practical guide. *Proteins* **7**, 205-14.
71. Lindberg, M. (2005). Protein Folding Studies on the Ribosomal Protein S6: the Role of Entropy in Nucleation, Umeå.
72. Consalvi, V., Chiaraluce, R., Giangiacomo, L., Scandurra, R., Christova, P., Karshikoff, A., Knapp, S. & Ladenstein, R. (2000). Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima. *Protein Engineering* **13**, 501-507.
73. Vivian, J. T. & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* **80**, 2093-109.
74. Chen, Y. & Barkley, M. D. (1998). Toward understanding tryptophan fluorescence in proteins. *Biochemistry* **37**, 9976-82.
75. Olofsson, M., Kalinin, S., Zdunek, J., Oliveberg, M. & Johansson, L. B. (2006). Tryptophan-BODIPY: a versatile donor-acceptor pair for probing generic changes of intraprotein distances. *Phys Chem Chem Phys* **8**, 3130-40.
76. Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluorescenz. *Ann. Phys.* **2**.
77. Tanford, C. (1968). Protein denaturation. *Adv Protein Chem* **23**, 121-282.
78. Tanford, C. (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* **24**, 1-95.
79. Fersht, A. R. & Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci U S A* **101**, 7976-81.
80. Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. (2002). Solution NMR techniques for large molecular and supramolecular structures. *J Am Chem Soc* **124**, 12144-53.

81. Wishart, D. S. & Sykes, B. D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**, 171-80.
82. Cung, M. T., Marraud, M. & Neel, J. (1974). Experimental calibration of a Karplus relationship in order to study the conformations of peptides by nuclear magnetic resonance. *Macromolecules* **7**, 606-13.
83. Vuister, G. W. & Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear 3-bond J(H(N)H(alpha) coupling-constants in N-15-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772-7777.
84. Prestegard, J. H., al-Hashimi, H. M. & Tolman, J. R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* **33**, 371-424.
85. Wiesner, S., Stier, G., Sattler, M. & Macias, M. J. (2002). Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J Mol Biol* **324**, 807-22.
86. Ruan, K. & Tolman, J. R. (2005). Composite alignment media for the measurement of independent sets of NMR residual dipolar couplings. *J Am Chem Soc* **127**, 15032-3.
87. Trempe, J. F., Morin, F. G., Xia, Z., Marchessault, R. H. & Gehring, K. (2002). Characterization of polyacrylamide-stabilized Pfl phage liquid crystals for protein NMR spectroscopy. *J Biomol NMR* **22**, 83-7.
88. Tycko, R., Blanco, F. J. & Ishii, Y. (2000). Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J. Am. Chem. Soc.* **122**, 9340-9341.
89. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**, 1065-74.
90. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-4.
91. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A* **92**, 9279-83.
92. Tolman, J. R. & Ruan, K. (2006). NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem Rev* **106**, 1720-36.
93. Ottiger, M., Delaglio, F., Marquardt, J. L., Tjandra, N. & Bax, A. (1998). Measurement of dipolar couplings for methylene and methyl sites in weakly oriented macromolecules and their use in structure determination. *J Magn Reson* **134**, 365-9.
94. Kovrigin, E. L., Kempf, J. G., Grey, M. J. & Loria, J. P. (2006). Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *J Magn Reson* **180**, 93-104.
95. Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y. & Kay, L. E. (2005). Multiple-site exchange in proteins studied with a suite of six NMR

relaxation dispersion experiments: an application to the folding of a Fyn SH3 domain mutant. *J Am Chem Soc* **127**, 15602-11.

96. Korzhnev, D. M. & Kay, L. E. (2008). Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts of Chemical Research* **41**, 442-451.
97. Vallurupalli, P., Hansen, D. F., Lundstrom, P. & Kay, L. E. (2009). CPMG relaxation dispersion NMR experiments measuring glycine 1H alpha and 13C alpha chemical shifts in the 'invisible' excited states of proteins. *J Biomol NMR* **45**, 45-55.
98. Molday, R. S., Englander, S. W. & Kallen, R. G. (1972). Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150-8.
99. Berger, A. & Linderstrom-Lang, K. (1957). Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* **69**, 106-18.
100. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. (2003). Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci* **12**, 153-60.
101. Bai, Y. (1999). Equilibrium amide hydrogen exchange and protein folding kinetics. *J Biomol NMR* **15**, 65-70.
102. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86.
103. Bai, Y. & Englander, S. W. (1996). Future directions in folding: the multi-state nature of protein structure. *Proteins* **24**, 145-51.
104. Englander, S. W. & Kallenbach, N. R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **16**, 521-655.
105. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192-197.
106. Krishna, M. M., Hoang, L., Lin, Y. & Englander, S. W. (2004). Hydrogen exchange methods to study protein folding. *Methods* **34**, 51-64.
107. Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31-Kda protein. *J. Magn. Reson.* **96**, 432-440.
108. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J. Magn. Res. Series B* **101**, 201-205.
109. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291-6293.
110. Bedard, S., Mayne, L. C., Peterson, R. W., Wand, A. J. & Englander, S. W. (2008). The foldon substructure of staphylococcal nuclease. *J Mol Biol* **376**, 1142-54.
111. Lipari, G. & Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J. Am. Chem. Soc.* **104**, 4546-4559.
112. Lipari, G. & Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **104**, 4559-4570.

113. Clore, G. M., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990). Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy. *Biochemistry* **29**, 7387-7401.

114. Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138-146.

115. Mandel, A. M., Akke, M. & Palmer, A. G. (1995). Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **246**, 144-163.

116. Palmer, A. G., Rance, M. & Wright, P. E. (1991). Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy. *J. Am. Chem. Soc.* **113**, 4371-4380.

117. Whitford, P. C., Miyashita, O., Levy, Y. & Onuchic, J. N. (2007). Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.* **366**, 1661-1671.