Communicating Research to the General Public

At the March 5, 2010 UW-Madison Chemistry Department Colloquium, the director of the Wisconsin Initiative for Science Literacy (WISL) encouraged all Ph.D. chemistry candidates to include a chapter in their Ph.D. thesis communicating their research to non-specialists. The goal is to explain the candidate's scholarly research and its significance to a wider audience that includes family members, friends, civic groups, newspaper reporters, state legislators, and members of the U.S. Congress.

Ten Ph.D. degree recipients have successfully completed their theses and included such a chapter, less than a year after the program was first announced; each was awarded $500.

WISL will continue to encourage Ph.D. chemistry students to share the joy of their discoveries with non-specialists and also will assist in the public dissemination of these scholarly contributions. WISL is now seeking funding for additional awards.

Wisconsin Initiative for Science Literacy

The dual mission of the Wisconsin Initiative for Science Literacy is to promote literacy in science, mathematics and technology among the general public and to attract future generations to careers in research, teaching and public service.

UW-Madison Department of Chemistry
1101 University Avenue
Madison, WI 53706-1396
Contact: Prof. Bassam Z. Shakhashiri
bassam@chem.wisc.edu
www.scifun.org

January 2011

Mechanistic Investigations on

Hsp70 Chaperone-Mediated Protein Folding

and

Photo-CIDNP Enhancements in

Heteronuclear Correlation NMR Spectroscopy

by

Ashok Sekhar

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biophysics)

at the

University of Wisconsin-Madison

2011

Chapter 9

I. The Role of the Hsp70 Molecular Chaperone in Protein Folding

and

II. How Lasers and Magnets Can Be Combined to Better Understand How Proteins Work

A Thesis Summary for Non-Specialists

Submitted as part of the

Wisconsin Initiative for Science Literacy

This chapter abridges the dissertation in an attempt to communicate my research to non-specialists, as part of the Wisconsin Initiative for Science Literacy (WISL). In this chapter, I outline the fundamental science research undertaken as part of this thesis, as well as its potential application to society. I have tried to provide a flavour for the kind of questions that my research has addressed, as well as the factors motivating such questions. I have summarized several of the results stated in the thesis and hope that the reader perceives some of the excitement that drives fundamental this science research.

I have not eliminated a technical description of my research from this chapter. While the broad research areas covered in this thesis (eg. protein folding) have implications that can be conveyed in non-technical language, the results obtained as part of this dissertation are specific and require some background in order to be appreciated. Hence, instead of bypassing technical details, I have gradually developed the concepts necessary to appreciate them at the beginning of each of the two sections in this chapter, starting from a very basic level of scientific knowledge. I have provided just enough background information necessary for relating to the thesis results. The interested reader is referred to Chapters 1-8 for more comprehensive discussions of the research.

I. The Role of the Hsp70 Molecular Chaperone in Protein Folding

9.1. Background and Significance

9.1.1. Proteins and the importance of protein folding

The importance of proteins for sustaining life was summarized succinctly by Nobel Laureate Stanford Moore in his Nobel Banquet speech in 1972 as follows, “Man's health and well-being depends upon, among many things, the proper functioning of the myriad proteins that participate in the intricate synergisms of living systems.”

Proteins are linear polymers structured like beads on a string. The monomer beads are called amino acids and occur in 20 different varieties. Each amino acid has specific distinguishing features arising from its molecular structure. In particular, some amino acids preferentially associate with water (hydrophilic), while others are oily and try to avoid contact with aqueous solutions (hydrophobic). The specific sequence of amino acids in a protein chain represents its primary structure. Amino acids close by in space in a protein interact with each other to give rise to two so called secondary structural motifs, the α-helix (where a protein chain winds itself up into a helix) and the β-sheet (where many strands of the protein come together to form a sheet) (Fig. 9.1a). The three-dimensional spatial arrangement of atoms in the protein is referred to as its conformation [1].

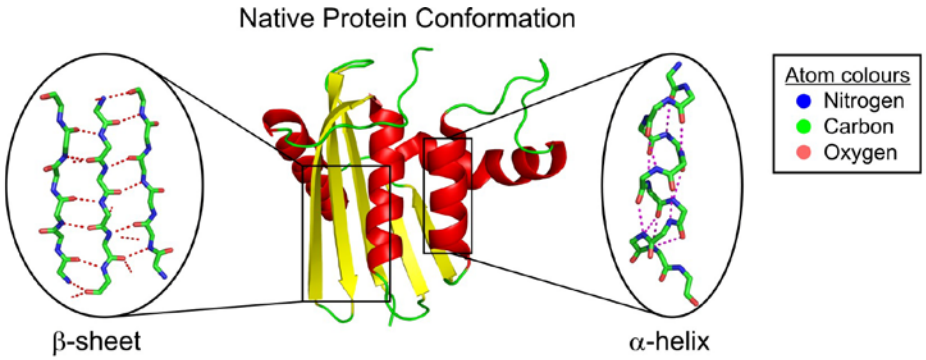
It is intriguing that most naturally occurring protein sequences adopt a unique conformation under physiological conditions, known as the native conformation or the folded state (Fig. 9.1a). Pioneering experiments by Christian Anfinsen established that the native state of the protein, and

hence the spatial location of its atoms, is completely defined by the specific sequence of amino acids constituting the protein chain [2]. Even more mysteriously, proteins attain a folded structure fast, on biological timescales (typically a few seconds) rather than the cosmological timespan (i.e., billions of years) necessary if the protein randomly searched through all its possible conformations [3].

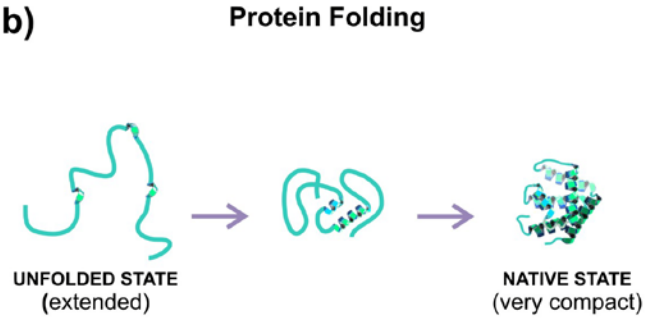
The process by which a protein reaches its native state is known as protein folding (Fig. 9.1b). The important aspects of protein folding include the rate at which the protein folds (kinetics), the stability of the native state (thermodynamics) and the pathway that a protein takes to reach the native state (mechanism). The major driving force for protein folding is the tendency of oily amino acid residues to bury themselves in the interior of the protein and avoid an aqueous environment, just like oil drops coalesce on the surface of water [4]. Protein folding is conventionally studied by unfolding the protein using physical (temperature, pressure) or chemical (acid, base, urea etc.) denaturants, and then removing the denaturant to track the protein fold. Form begets function [4], and the proper folding of proteins is essential for the healthy functioning of all life forms. Further, the disruption of protein folding results in variety of debilitating diseases such as Alzheimer's and Parkinson's disorders, diabetes, and cataract. In such diseases, several factors such as ageing and stress cause proteins to misfold and aggregate, often resulting in the formation of protein plaques called amyloids [5].

Thus, protein folding is another instance of biological wizardry, and studying it is satisfying and rewarding in itself. In addition, research in protein folding and misfolding also has the potential to provide therapeutic strategies to combat some of the most common diseases afflicting mankind.

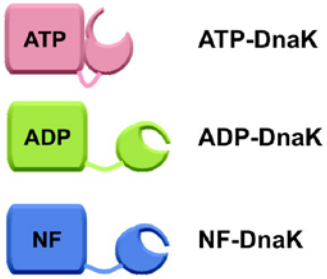
a)


Native Protein Conformation
Atom colours
Nitrogen
Carbon
Oxygen
β-sheet
α-helix

b)


Protein Folding
UNFOLDED STATE
(extended)
NATIVE STATE
(very compact)

c)


Three Key Nucleotide States of DnaK
ATP
ATP-DnaK
ADP
ADP-DnaK
NF
NF-DnaK

d)

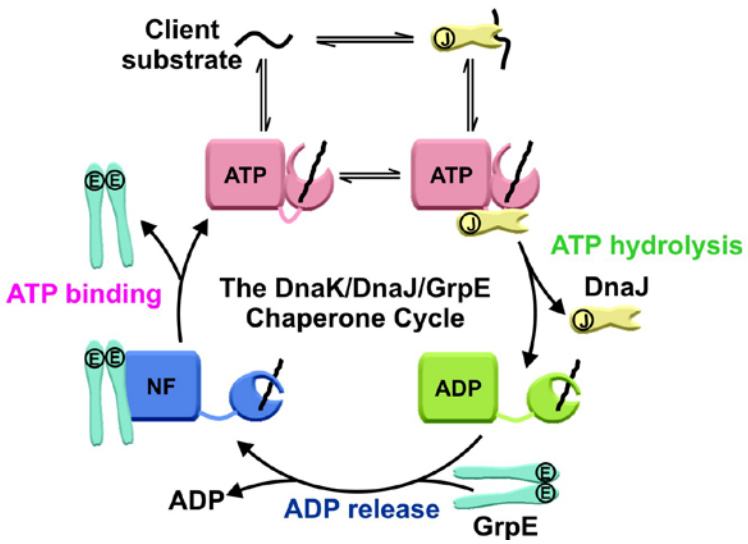

Client substrate
J
ATP
ATP
E E
ATP hydrolysis
ATP binding
The DnaK/DnaJ/GrpE Chaperone Cycle
DnaJ
NF
ADP
ADP
ADP release
GrpE

Figure 9.1. a) Ribbon diagram showing a typical native protein conformation. α-helical and β-sheet secondary structural elements are indicated in red and yellow respectively. Atom-level depictions of an α-helix and a β-sheet present in the protein structure are presented alongside. The interactions leading to the formation of these secondary structures are indicated as dotted magenta lines. b) A typical protein folding reaction, beginning from an extended unfolded state with very little structure, progressing through an intermediate with some structure and compaction and finishing up in the native state. c) Cartoon representations of the three nucleotide states of DnaK pictorially depicting the structural differences between the three forms. d) The functional chaperone cycle of *E .coli* Hsp70.

9.1.2. Molecular chaperones: the art of avoiding sticky situations [6]

While all the necessary information required for folding is encoded within the primary sequence of a protein, protein folding under cellular conditions is facilitated by a class of proteins called molecular chaperones [7]. The term 'chaperone' generally refers to an adult who supervises young unmarried men or women during social occasions, with the specific intent of preventing inappropriate or illicit interactions. Analogously, molecular chaperones typically bind to exposed hydrophobic regions in unfolded proteins (known as 'client' proteins) and prevent their premature misfolding and deleterious aggregation.

The Hsp70 (Heat Shock Protein 70) family of molecular chaperones is present in virtually all life forms [8]. Hsp70 shields client proteins concurrently with their synthesis in the cell. In addition, Hsp70 molecules participate in diverse cellular processes such as protein quality control, protein trafficking and the cellular response to heat stress. Abnormal Hsp70 activity is associated with several diseases, including neurodegenerative disorders, cancer and ischemia [9].

The Hsp70 protein in the bacterium *Escherichia coli* is called DnaK, and serves as a representative model system for the Hsp70 family. DnaK is a protein with a molecular mass of approximately 70,000 Da (the mass of the H atom is ~ 1 Da; 1 Da $\approx 10^{-27}$ kg). DnaK is equipped to utilize the energy currency in the cell, the nucleotide ATP (Adenosine triphosphate), by binding and hydrolyzing it to ADP (Adenosine diphosphate). DnaK can thus exist in three distinct states, bound to ATP (ATP-DnaK), bound to ADP (ADP-DnaK), and free of nucleotides (NF-DnaK) (Fig. 9.1c). The affinity of DnaK for its client is lower in the ATP-bound state, and higher in the ADP-bound state. Further, the three-dimensional structures of ADP- and NF-DnaK are very similar, while ATP binding causes major structural rearrangements in DnaK. In cells, DnaK collaborates with two other proteins, DnaJ and GrpE, which increase the rate at which

DnaK hydrolyzes ATP. The periodic alternation of DnaK between its three nucleotide states, mediated by DnaJ and GrpE, is known as the DnaK functional cycle (Fig. 9.1d) [8].

All the experiments reported in this thesis are performed using the *E. coli* Hsp70 chaperone system, which will be variously referred to as DnaK/DnaJ/GrpE, or K/J/E.

9.1.3. How do client-chaperone interactions influence protein folding?

Very little is known about how a client protein which is trying to fold is affected by the K/J/E chaperone system. K/J/E accelerates the folding rate of a few large proteins like firefly luciferase, which cannot fold efficiently by themselves [10,11]. However, it is unclear whether the 'folding catalysis' activity of K/J/E is a general feature, or if it is specific to the client protein. We also do not understand whether the *E. coli* Hsp70 chaperone system can speed up the folding rate of slow-folding, but nevertheless folding competent, client proteins.

We do not comprehend all the mechanisms by which K/J/E picks and chooses its clients. DnaK recognizes a 'binding motif' in proteins, which is a sequence of about seven amino acids, enriched in oily hydrophobic residues like leucine. However, this motif is common enough in naturally occurring proteins so that all proteins are expected to have at least one DnaK binding site [12]. While DnaK is thus expected to interact with all proteins, experimental measurements in living *E. coli* cells have shown that DnaK preferentially interacts with clients in the size range of 20,000-70,000 Da [13,14]. This observation is even more mysterious because it is not clear how DnaK estimates the size of its client protein.

The ATP-regulated association of DnaK with its client proteins is vital for its function as a chaperone. However, the molecular level details of how client-DnaK interactions facilitates client protein folding remains unclear. Does DnaK use the energy stored in ATP to twist a helix

here, bend a sheet there, and thereby actively fold proteins? Or does DnaK passively hold a client and prevent illicit and premature interactions?

We employed two approaches to address deficiencies in our understanding of chaperone-client interactions. First, we used computational kinetic modeling (Chapter 2) together with stopped-flow fast mixing experiments (Chapter 3) to obtain insights into the time-dependent perturbations induced by K/J/E in the folding of client proteins. Second, we used fluorescence measurements in conjunction with Nuclear Magnetic Resonance (NMR) spectroscopy to examine the conformation of client substrates bound to DnaK (Chapter 4).

9.2. Results from this thesis

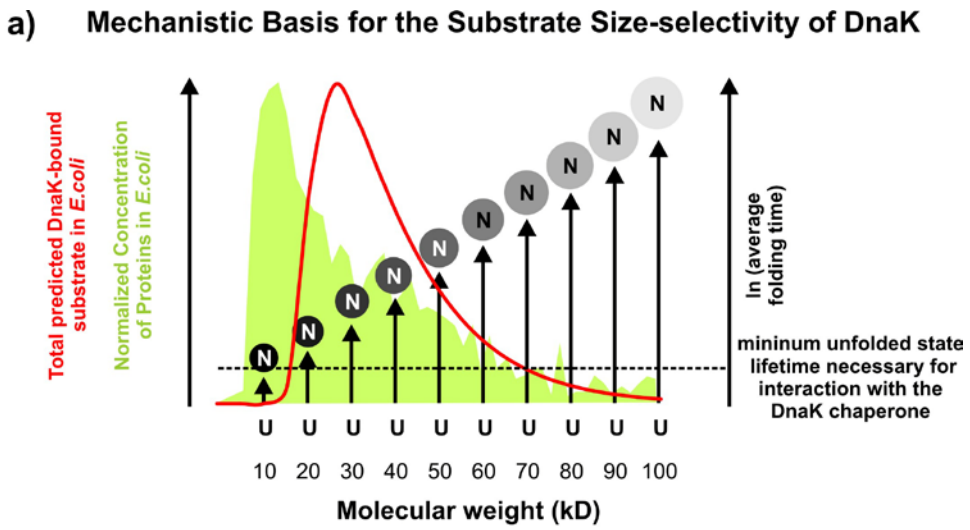
9.2.1. A summary of Chapter 2: Computational modeling of protein folding in the presence of the K/J/E chaperone system

In order to mechanistically understand chaperone-substrate interactions, we modeled them according to a ‘kinetic partitioning’ mechanism. Here, unfolded client molecules are segregated along simultaneous pathways resulting in either a chaperone-bound or a native state. The flux of molecules along each pathway is controlled by the relative rates (kinetics) of chaperone-binding and folding. We then assembled a computational model comprising the interactions of DnaK with DnaJ and GrpE, as well as with client proteins, using data available in literature. This model includes 43 reaction steps, and each step has a differential equation and a rate constant associated with it. Simulations using the model involve numerical integration of these differential equations using well-defined initial conditions. The model is a useful tool to predict the extent of interaction of a folding client protein with K/J/E, and the changes in its folding rate concomitant with such an interaction.

Using the computational model, we were able to quantify the characteristics of client proteins that would significantly interact with K/J/E. We recognized that the complex network of K/J/E interactions is ideally tailored to select clients based on the rate at which they folded to the native state, but was not as suited to interact with proteins that have already folded to their native structure. We then applied our model to computationally study the interactions of clients with K/J/E in a cell-relevant scenario. We found that the set of proteins associating with K/J/E under these conditions possessed a well-defined range of folding rates, which corresponded to the average folding rates of proteins in the size range 20,000 – 70,000 Da. Our model analysis thus provided a mechanistic basis for the experimentally observed size selectivity of the K/J/E chaperone system (Fig. 9.2a), by revealing that the chaperone is in fact sensitive to the folding rate of the client protein, and that the folding rates are size-dependent.

9.2.2. A summary of Chapter 3: Extensive but short-lived interactions of a folding client protein with K/J/E

We then experimentally studied the interaction of K/J/E with a folding client protein. We used a slow-folding variant of *E. coli* RNase H [15] as the client substrate. The time-dependent perturbations induced by K/J/E in the folding of RNase H^D were followed using circular-dichroism-detected stopped-flow fast mixing. The stopped-flow instrument is a device that can rapidly mix two solutions (with a dead time of ~ 1 millisecond) and inject them into an observation chamber for detection using methods like circular dichroism (CD). The CD signal is a measure of the amount of helical structure present in any protein. A typical protein folding process is thus expected to progress with an increase of the CD signal, because the unfolded state of proteins contains very little structure compared to the native state. Stopped-flow experiments


a)
Mechanistic Basis for the Substrate Size-selectivity of DnaK
Total predicted DnaK-bound substrate in E.coli
Normalized Concentration of Proteins in E.coli
ln (average folding time)
mininum unfolded state lifetime necessary for interaction with the DnaK chaperone
N
U
10 20 30 40 50 60 70 80 90 100
Molecular weight (kD)

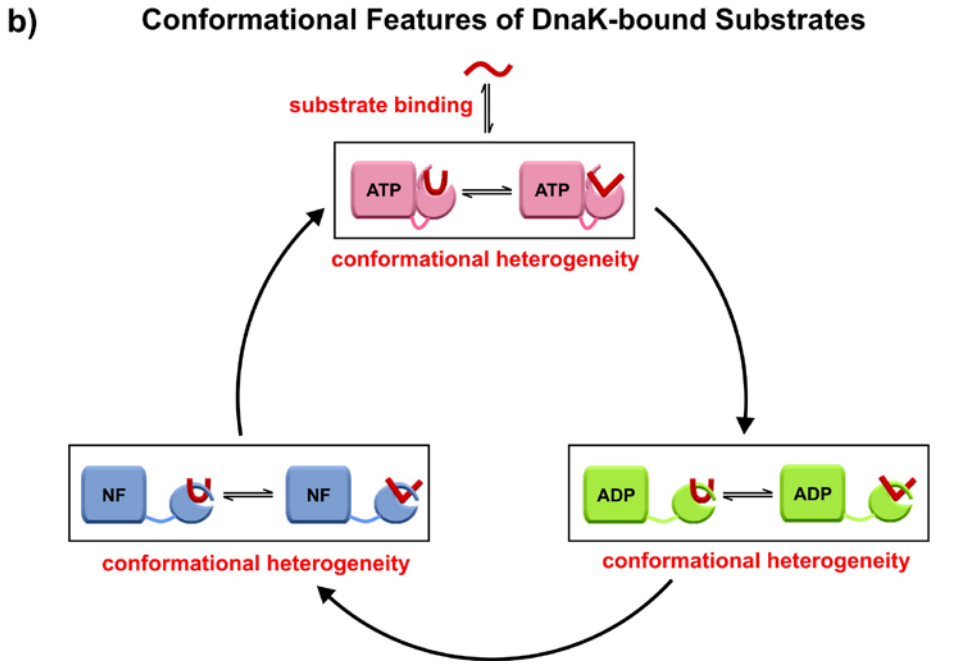

b)
Conformational Features of DnaK-bound Substrates
substrate binding
ATP
ATP
conformational heterogeneity
NF
NF
conformational heterogeneity
ADP
ADP
conformational heterogeneity

Figure 9.2. a) A mechanistic explanation for the experimentally observed size-selectivity of DnaK in *E. coli* cells. The molecular weight-dependent logarithm of the average folding time is denoted by black arrows. The horizontal dotted line indicates the minimum time the unfolded state must exist in order to interact with DnaK. Small molecular weight proteins fold too fast to associate significantly with DnaK. The green curve shows the concentration of proteins in *E. coli* as a function of molecular weight, and reveals that there are very few proteins larger than ~ 70 kD. The combinations of the above effects results in the size-selectivity of DnaK for ~ 20-70 kD substrate proteins (the red line). b) A summary of the conformational features of the substrate as it traverses the Hsp70 chaperone cycle. Binding to DnaK induces structural changes in the substrate, and the client displays multiple bound states. However, the substrate conformation is unaltered when the nucleotide state of DnaK is changed.

were carried out by unfolding RNase H^D using urea, and initiating folding in the absence or presence of K/J/E by mixing unfolded RNase H^D with a solution having no urea.

The existence of interactions between K/J/E and the client protein RNase H^D was confirmed using size-exclusion chromatography (SEC). SEC segregates molecules based on their molecular weight, and can distinguish between native RNase H^D molecules (~ 17,000 Da) and DnaK-bound RNase H^D (~ 87,000 Da).

Stopped-flow folding rates in the absence and presence of K/J/E, as well as chaperone-client interactions were quantified and analyzed independently as well as in conjunction with the previously developed computational model.

We observed that the K/J/E chaperone system slowed down the folding of its client RNase H^D. Thus, the acceleration in folding rate observed previously in literature is not a general feature of K/J/E, but is in fact client-specific. The deceleration in folding rate is very moderate, which is significant from the cellular viewpoint, because RNase H^D can fold efficiently by itself; a slow down in its folding because of interactions with K/J/E is counterproductive to the health of the cell, because the folded state is a necessary prerequisite for function.

We showed that the experimental stopped-flow and size-exclusion chromatography measurements were entirely consistent with a passive kinetic partitioning mechanism of client-chaperone interaction and could be quantitatively predicted by our computational model. Model analysis confirmed that the interactions between the client and K/J/E are extensive but short-lived. Further, the role of K/J/E in the lifetime of a protein capable of independent folding was shown to be to shield nonnative client molecules and thereby avoid premature misfolding and aggregation. The extensive but short-lived association of clients with K/J/E provides a

mechanism by which the chaperone can prevent inappropriate interactions, while at the same time circumventing large delays in the folding of the client protein.

9.2.3. A summary of Chapter 4: Conformational properties of DnaK-bound client substrates

A molecular link between chaperone-substrate interactions and Hsp70 function remains lacking. One prominent reason for this lack of understanding is that we do not know whether chaperone-binding changes the structure of the client substrate, and whether the changes in the nucleotide state of DnaK alter the conformation of the associated client. We used fluorescence and NMR measurements to probe the interactions of substrates with DnaK. We chose as client substrates, 13-residue peptides expected to have a high affinity for DnaK.

Fluorescence measurements were employed to quantify the interaction strength between ATP-, ADP- and NF-DnaK and the peptide substrates. Such measurements showed that DnaK interacted strongly with the client peptide in the NF- and ADP-bound states, but the binding affinity decreased by 5-12-fold in the ATP-bound form.

NMR spectroscopy is a technique that provides structural information on proteins at atomic resolution (also see Section 9.4.1). NMR spectra can be used to differentiate between protein segments adopting helical and sheet structures. Here, we used NMR measurements to identify the conformation of the peptide substrates, both in their free forms, as well as when bound to ATP-, ADP- and NF-DnaK.

We find that both the peptides used in the study undergo large conformational changes when they bind DnaK. Further, when associated with DnaK, the same substrate displays at least two alternative structures which are different from one another and also different from the free form

of the substrate (Fig. 9.2b). One of the peptide substrates has a helical conformation in the free form. Both the DnaK-bound forms of this peptide have reduced helicity, and one of the bound conformations has a partial β-strand character. The second peptide is unstructured in the free form. One of the DnaK-bound forms of this peptide exhibits a residual β-strand tendency, while the second conformation is twisted. The simultaneous presence of multiple conformations of the same protein, such as those we observe in our experiments, is termed as conformational heterogeneity and is known to be very important in biomolecular recognition processes.

In contrast, however, the structures of the client substrates bound to ATP-, ADP- and NF-DnaK are identical (Fig. 9.2b). Our results establish that the major role of ATP in the K/J/E functional cycle is to alter the affinity of the client for DnaK, and not to change the client structure.

9.3. Conclusions and Future Perspectives

Our computational and experimental studies examining the interaction of folding client proteins with the *E. coli* Hsp70 chaperone system have revealed intriguing features inherent to the K/J/E chaperones. The K/J/E chaperone system is tailored to select client proteins based on how slowly they fold to the native state. The sensitivity of DnaK for folding rates is indirectly reflected in the size-selectivity of K/J/E for clients observed experimentally in living *E. coli* cells. The transient association of clients with K/J/E enables the chaperone to prevent inappropriate aggregation, while concomitantly avoiding sizeable delays in the maturation of the client.

At the molecular level, we have established that the major function of the energy-rich ATP in the K/J/E functional cycle is to alter the affinity of clients for DnaK, and not to change the structure of the bound client substrate. However, DnaK-binding induces major structural changes in the client, and substrates display conformational heterogeneity in the chaperone-bound state.

The research carried out as part of this dissertation has also opened up several fascinating questions for future work. Some of the promising avenues include characterizing the structural and functional relevance of the conformational heterogeneity observed in DnaK-bound clients, examining the structure of entire client proteins associated with DnaK, and elucidating the mechanism of protein folding starting from a client-DnaK complex.

II. How Lasers and Magnets Can Be Combined to Better Understand How Proteins Work

9.4. Background and Significance

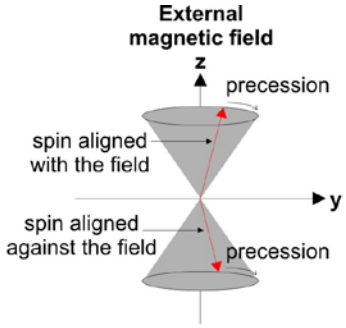
9.4.1. NMR Spectroscopy: an outstanding tool in biophysics research

A long-cherished scientific dream, voiced by the French mathematician Pierre de Laplace, has been to precisely determine the instantaneous position and velocity of every particle in this universe as a prerequisite for deciphering the past and predicting the future. We now know that this lofty goal is impossible to achieve and is inconsistent with the laws of quantum mechanics, a well-accepted theory explaining much of our physical world.

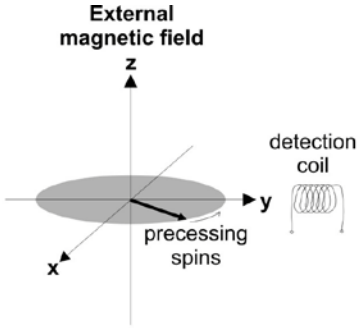
However, Nuclear Magnetic Resonance (NMR) spectroscopy comes close to realizing this goal, if not for the entire universe, at least for the sample under investigation. NMR is unique in its ability to provide structural information at atomic resolution, as well as dynamic information over a broad range of timescales (ps to s; 1 ps = 10^{-12} s) [16]. NMR spectroscopy has provided seminal insights into several frontier areas of biomolecular research, and thereby occupies a preeminent position among biophysical tools.

NMR works on the same principles as Magnetic Resonance Imaging (MRI) [17]. The nuclei of some atoms have non-zero spin, which is a complex quantum mechanical quantity that can assume integer or half-integer values. Routine NMR measurements use spin ½ nuclei like ^{1}H, ^{13}C and ^{15}N. Spin ½ nuclei behave as tiny magnets and align themselves along or against an external magnetic field. In addition, when these magnets are out of alignment with the external

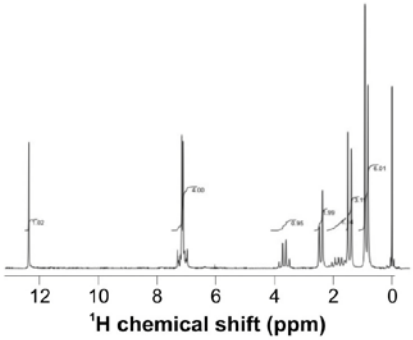
a)



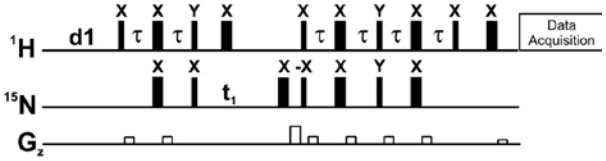
b)



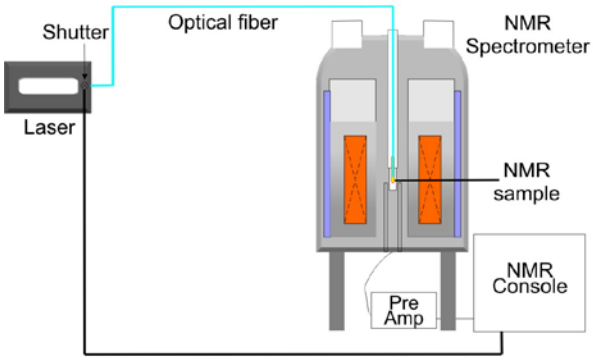
c) A Typical 1D NMR Spectrum



d) ^{1}H-^{15}N HSQC Pulse Sequence



e) Layout of the laser and the NMR spectrometer



f) Layout of the laser table

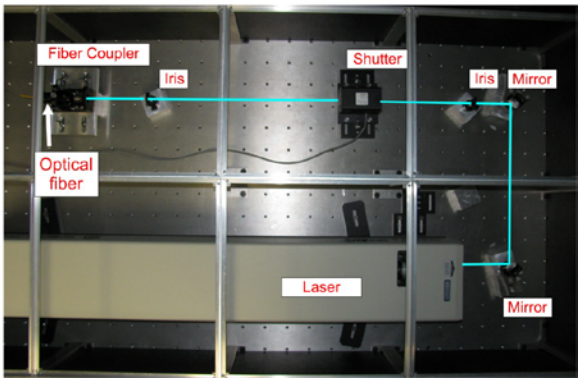


Figure 9.3. a) Precession of nuclear spins aligned with and against an external magnetic field along the z-axis. b) The detected NMR signal consists of a voltage induced in a coil placed in a plane transverse to the external magnetic field. c) A typical NMR spectrum showing several peaks as a function of their chemical shift. d) A representative heteronuclear correlation pulse sequence comprising rf pulses separated by delays. Pulse durations are generally a few μs, while τ is ~ 2.5 ms. The phase of each pulse is also indicated. In this sequence, pulses are applied either at the ^{1}H or the ^{15}N frequency. Magnetic field gradients (G_z) are also introduced at specific time points. e) The schematic layout of the Ar-ion laser and the NMR spectrometer used for photo-CIDNP experiments. A fiber optic cable is used to transport laser light into the NMR spectrometer. f) A picture of the laser table, showing the positions of the laser, the two mirrors and two irises, the shutter and the fiber coupler.

field, they undergo a precessional motion (Fig. 9.3a), in much the same way as a gyroscope does when placed at an angle to the vertical. When a coil is placed in the vicinity of the precessing nuclei, their magnetic field cuts the coil and induces a voltage in it, similar to an electricity generator (Fig. 9.3b). The frequencies present in this time-dependent voltage correspond to the nuclear precession frequencies. The chemical and electronic environment of a nucleus alters its fundamental precession frequency. An NMR spectrum thus consists of peaks arising from different NMR-active nuclei present in the molecule, as a function of their precession frequency relative to a standard (commonly known as chemical shift) (Fig. 9.3c). The NMR spectroscopist manipulates nuclear spins using a pulse sequence, which is applied to the sample before data acquisition. The pulse sequence consists of a carefully designed train of radiofrequency (rf) pulses of the appropriate frequency, phase, and amplitude, interspersed with precise time delays (Fig. 9.3d). NMR spectroscopy differs from MRI in the respect that NMR samples are homogeneous and the signal arising from different regions of the sample is averaged together, while the aim of MRI is to spatially resolve differences present in the imaged article.

NMR measurements benefit from outstanding spectral resolution. Peaks differing in frequency by as little as a few parts per billion are distinguished from each other on a routine basis [17]. Further, NMR chemical shifts are exquisitely sensitive to the physical and chemical environment of the nucleus, and serve as specific reporters of structural features such as backbone conformation, ring puckering and geometric isomerization. In addition, other NMR spectral characteristics like the broadness of the peaks (known as the linewidth) are sensitive to the fluctuations of this environment, caused by molecular motion [16].

The workhorse of biomolecular NMR spectroscopy is a class of experiments known as heteronuclear correlation NMR, in which information from several nuclei such ^{1}H, ^{13}C and ^{15}N is

integrated [18]. Heteronuclear correlation NMR was developed in concert with methods to produce proteins containing ^{13}C and ^{15}N instead of the naturally abundant ^{12}C and ^{14}N isotopes. Heteronuclear experiments are performed routinely to characterize a protein, as well as in determining the structure or the dynamics of a protein chain.

9.4.2. The low sensitivity of NMR spectroscopy and the concept of hyperpolarization

The bane of NMR spectroscopy is its low sensitivity (or a poor signal-to-noise ratio), which translates into the need for preparing concentrated samples and results in long data acquisition times (often days at a stretch). The poor NMR sensitivity arises from the weak nature of the interaction between nuclear magnets and the external magnetic field present in the NMR spectrometer. Consequently, several research groups worldwide are working towards improving the sensitivity of NMR experiments, which will broaden the repertoire of NMR spectroscopy to include larger and more complex biomolecular systems.

For the purposes of describing the results in this dissertation, it is easiest to understand NMR sensitivity in terms of the concept of spin temperature [19]. The collection of nuclear spins of a certain type (eg. the –NH proton of the 10^{th} amino acid in the protein) are associated with a temperature value (T_S), which is the same as the sample temperature (T_L) for a sample at equilibrium. Cold spins give a larger NMR signal. Therefore one way to increase sensitivity in NMR is to lower the sample temperature [17]. However, acquiring protein NMR spectra at low temperature is impractical for various reasons including the fact that aqueous samples freeze below ~ 0°C.

A prominent class of sensitivity enhancement techniques in NMR spectroscopy selectively cools the spin system (and not the sample) and observes the intense NMR signal before the spin

system has equilibrated with the sample reservoir, and the spin and sample temperatures have equalized. Such methods which rely on spin refrigeration to increase NMR sensitivity are referred to as hyperpolarization methods [20].

Photochemically induced dynamic nuclear polarization (photo-CIDNP) [21,22] is one such hyperpolarization technique with demonstrated applicability to proteins in aqueous solution.

9.4.3. Photo-CIDNP: A hyperpolarization method for solution state biomolecular NMR

Photo-CIDNP is a fascinating example of the influence of magnetic interactions on the course of a chemical reaction [23]. In photo-CIDNP, spin refrigeration occurs by means of a photochemical reaction. A photosensitive dye absorbs light and interacts with nuclear spins in the sample to hyperpolarize it. The detailed mechanism of photo-CIDNP is provided in Chapter 1, and is summarized in Hore and Broadhurst [21]. Because of the mechanistic features of photo-CIDNP, only selected nuclei in three amino acids, Trp, Tyr and His, are refrigerated. Further, since cooling and hyperpolarization involves the interaction of the dye with the sample, only amino acids on the surface of a protein are affected by photo-CIDNP. As a result, photo-CIDNP spectra have been extensively used to determine whether amino acids in a protein are buried or exposed to the solvent.

In order to carry out photo-CIDNP experiments, it is necessary to irradiate samples inside the NMR spectrometer with laser light. It must be borne in mind that photo-CIDNP relies on selectively cooling the spin system of interest, and the spins equilibrate to the sample temperature in about 1 s, so it is not feasible to irradiate samples and then load them manually into the NMR spectrometer. Consequently, a significant portion of this section of the dissertation involved building the instrumentation necessary to stably transport laser light into a sample

located inside the NMR spectrometer (Fig. 9.3 panels e and f). Measures were also taken to synchronize radiofrequency and laser pulses.

9.4.4. Exploring the potential of photo-CIDNP as a sensitivity enhancement tool

NMR spectroscopy has witnessed the development of several hyperpolarization methods such as DNP (Dynamic Nuclear Polarization) [24], PHIP (Parahydrogen Induced Polarization) [25], optical pumping [20] and photo-CIDNP. However, none of these methods except photo-CIDNP is currently applicable to enhance the sensitivity of protein peaks in solution. On the other hand, photo-CIDNP has been primarily utilized as a probe of solvent exposure, and its potential as a sensitivity enhancement tool in NMR spectroscopy remains poorly explored.

As a part of this dissertation, we examined the ability of photo-CIDNP arising from ^{1}H (Chapter 5), ^{15}N (Chapter 6) and ^{13}C (Chapter 7) nuclei to improve the sensitivity of heteronuclear correlation NMR spectroscopy. Chapter 6 also details the development of two new photo-CIDNP pulse sequences, EPIC- and CHANCE-HSQC.

9.5. Results from this thesis

9.5.1. A summary of Chapter 5: ^{1}H photo-CIDNP enhancements in ^{1}H-^{15}N correlation spectroscopy

The research described in this chapter examined the sensitivity enhancements available from the ^{1}H nucleus in the Trp amino acid sidechain, in the context of ^{1}H-^{15}N heteronuclear correlation spectra. Several routinely used ^{1}H-^{15}N correlation pulse sequences were modified to incorporate a laser irradiation module in their beginning. The protein, apoHmpH [26], was used as a model for testing the concomitant improvements in sensitivity. ApoHmpH is structurally very

similar to myoglobin, which binds oxygen in muscles and was the first protein whose structure was solved using x-ray crystallography by John Kendrew in 1958 [27]. ApoHmpH contains one Trp residue, and this Trp is exposed to the solvent in its unfolded state. Therefore, unfolded apoHmpH is an ideal candidate for testing photo-CIDNP-based pulse sequences.

We systematically quantified the sensitivity improvements originating from the inclusion of photo-CIDNP based spin refrigeration in heteronuclear correlation pulse sequences. We observed about 100% sensitivity improvements compared to currently employed pulse sequences in spectra of unfolded apoHmpH (Fig. 9.4a). We used a standard mathematical model of photo-CIDNP [28] to predict the signal improvements available from each pulse sequence, and found the predictions to agree very well with the experimentally observed values. We characterized the extent of sample damage resulting from laser irradiation, and found it to be negligible under our experimental conditions for routine acquisition of one- and two-dimensional NMR spectra (where peak intensity is a function of two distinct precession frequencies). We then demonstrated the applicability of the photo-CIDNP NMR methodology for acquiring two dimensional heteronuclear correlation NMR spectra, which are less tolerant to fluctuations in signal intensity.

9.5.2. A summary of Chapter 6: Development of novel photo-CIDNP pulse sequences EPIC- and CHANCE-HSQC

The photo-CIDNP enhancements arising from ^{15}N nuclei are known to be large [29]. However, the ^{15}N spin is a 10-fold weaker magnet than the ^{1}H nucleus, and ^{15}N spectra are much more noisy compared to ^{1}H spectra. In order to overcome the lower sensitivity of the ^{15}N nucleus and simultaneously take advantage of its large photo-CIDNP enhancement factor, we devised two

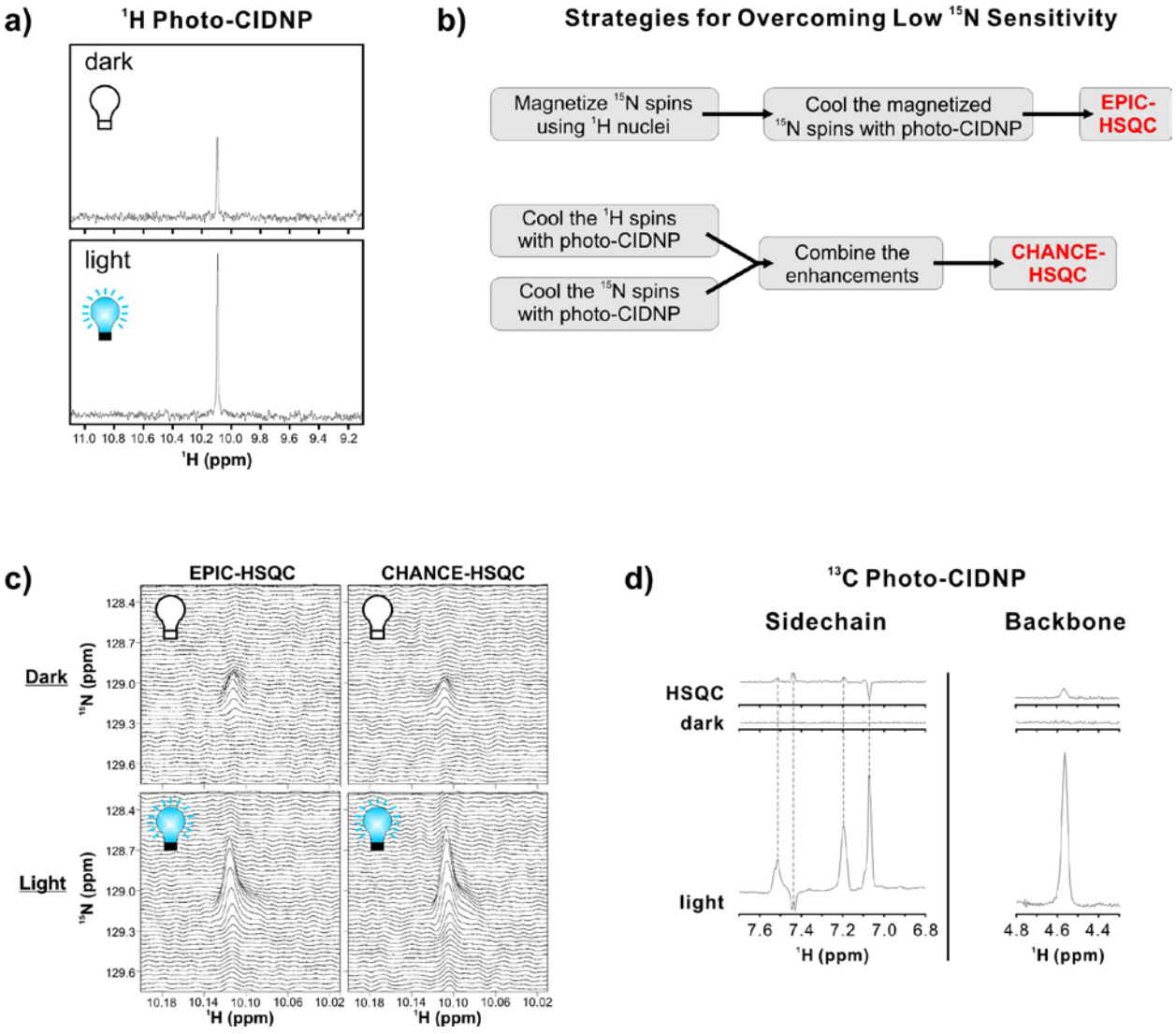

a)
^{1}H Photo-CIDNP
dark
light
11.0 10.8 10.6 10.4 10.2 10.0 9.8 9.6 9.4 9.2
^{1}H (ppm)
b)
Strategies for Overcoming Low ^{15}N Sensitivity
Magnetize ^{15}N spins using ^{1}H nuclei
Cool the magnetized ^{15}N spins with photo-CIDNP
EPIC-HSQC
Cool the ^{1}H spins with photo-CIDNP
Cool the ^{15}N spins with photo-CIDNP
Combine the enhancements
CHANCE-HSQC
c)
EPIC-HSQC
CHANCE-HSQC
Dark
Light
^{15}N (ppm)
128.4
128.7
129.0
129.3
129.6
10.18 10.14 10.10 10.06 10.02
^{1}H (ppm)
d)
^{13}C Photo-CIDNP
Sidechain
Backbone
HSQC
dark
light
7.6 7.4 7.2 7.0 6.8
4.8 4.6 4.4
^{1}H (ppm)

Figure 9.4. a) Light (laser on) and dark (laser off) NMR spectra of unfolded apoHmpH showing the sensitivity enhancement of the Trp sidechain $^{1}H^{N}$ arising from ^{1}H photo-CIDNP. b) Strategies for overcoming the low ^{15}N sensitivity, leading to the development of EPIC- and CHANCE-HSQC pulse sequences. c) 2D EPIC- and CHANCE-HSQC light and dark spectra of unfolded apoHmpH. d) Photo-CIDNP sensitivity enhancements originating from backbone and sidechain ^{13}C nuclei of the Trp amino acid in a 13-residue peptide. The reference HSQC spectrum is the prevalent way of acquiring such spectra.

strategies that led to the development of the EPIC- and CHANCE-HSQC pulse sequences (Fig. 9.4b).

In the EPIC-HSQC, the ^{15}N spins are first magnetized by the ^{1}H spins, which are more powerful magnets. The magnetized ^{15}N spins are then refrigerated by photo-CIDNP before data acquisition (Fig. 9.4b). In CHANCE-HSQC, the cooling effect of photo-CIDNP on the ^{1}H and ^{15}N spin reservoirs is combined to provide optimal sensitivity (Fig. 9.4b). A novel feature of both EPIC- and CHANCE-HSQC is that they involve laser irradiation in the middle of the radiofrequency pulse sequence.

Both the EPIC- and CHANCE-HSQC were demonstrated to be capable of robust 1D and 2D NMR data acquisition on samples of unfolded apoHmpH protein (Fig. 9.4c). The sensitivity gains from both these pulse sequences could be predicted quantitatively based on the standard mathematical model of photo-CIDNP.

9.5.3. A summary of Chapter 7: ^{13}C photo-CIDNP enhancements

Chapter 7 investigates the sensitivity enhancements arising from ^{13}C nuclei in various amino acids. Experiments in this chapter were performed with a high power (5 W) laser compared to the low power 1W laser utilized in Chapters 5 and 6. The higher available laser energy increases the output of the photochemical reaction and leads to more efficient refrigeration of the nuclear spins.

We observed several interesting features in this study. We found peaks from several ^{13}C nuclei in Trp, Tyr and His amino acids to be enhanced in sensitivity, increasing the number of photo-CIDNP-polarizable probes in proteins (Fig. 9.4d). Backbone ^{13}C spins in Trp and Tyr were distinctly cooled by photo-CIDNP, and this is important because backbone ^{13}C nuclei are robust

probes of secondary structure (helix or sheet conformation) in proteins. Finally, sensitivity enhancement factors for ^{13}C nuclei irradiated with a high power laser were very large (exceeding 200-fold) even on a 13-residue peptide sample.

9.6. Conclusions and Future Perspectives

The NMR investigation of complex biologically important molecules is often sensitivity limited. It is not possible to prepare homogeneous biological samples of arbitrarily high concentration, because of the time and monetary costs involved in making such molecules, and because of undesirable processes such as aggregation, which dominate at high concentrations. In some cases, it is not even desirable to increase the concentration indefinitely, because the nature of biomolecular interactions is concentration dependent. In addition, biological samples have limited temporal stability and it is not possible to acquire high quality NMR data for long durations. On the other hand, several such biological systems are of profound scientific interest. Research in sensitivity enhancement methodology in NMR spectroscopy acquires importance from the perspective of placing these systems within the reach of NMR investigation.

Photo-CIDNP is a promising candidate for improving the sensitivity of NMR in the case of biomolecules in solution. We have systematically explored the potential of ^{1}H, ^{15}N and ^{13}C sensitivity enhancements in the context of heteronuclear correlation NMR spectroscopy. The signal improvements, particularly from ^{13}C nuclei at high laser power, substantiate the value of photo-CIDNP in NMR spectroscopy.

The large magnitudes of the observed signal enhancements promise to enable the NMR detection of dilute biological samples. We are particularly interested in applying the photo-CIDNP methodology to study two exciting, yet poorly understood, biological systems: the

ribosome-bound nascent protein [30], which represents an intermediate stage in the production line of proteins in the cell, and client-chaperone complexes (introduced in section 9.1.). Other important avenues of future research in photo-CIDNP that have been brought to prominence by our research include extending the applicability of photo-CIDNP "refrigeration" to other types of nuclei in proteins and learning how to mitigate the damage of high power laser irradiation on delicate protein samples.

9.7. References

1. Lehninger, A. L., Nelson, D. L. and Cox, M. M. (2005). *Principles of Biochemistry*, W.H. Freeman.
2. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* (1973) **181**, 223-230.
3. Sinha, K. K. and Udgaonkar, J. B. Early events in protein folding. *Curr. Sci.* (2009) **96**, 1053-1070.
4. Dill, K. A. Dominant forces in protein folding. *Biochemistry* (1990) **29**, 7133-7155.
5. Chiti, F. and Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* (2006) **75**, 333-366.
6. Hartl, F. U., Hlodan, R. and Langer, T. Molecular chaperones in protein folding: the art of avoiding sticky situations. *Trends Biochem. Sci.* (1994) **19**, 20-25.
7. Hartl, F. U. and Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* (2009) **16**, 574-581.
8. Mayer, M. P. and Bukau, B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* (2005) **62**, 670-684.

9. Meimaridou, E., Gooljar, S. B. and Chapple, J. P. From hatching to dispatching: the multiple cellular roles of the Hsp70 molecular chaperone machinery. *J. Mol. Endocrinol.* (2009) **42**, 1-9.

10. Szabo, A., Langer, T., Schröder, H., Flanagan, J., Bukau, B. and Hartl, F. U. The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ, and GrpE. *Proc. Natl. Acad. Sci. U. S. A.* (1994) **91**, 10345-10349.

11. Sharma, S. K., De Los Rios, P., Christen, P., Lustig, A. and Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* (2010) **6**, 914-920.

12. Rüdiger, S., Germeroth, L., Schneider-Mergener, J. and Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *The EMBO journal* (1997) **16**, 1501-1507.

13. Teter, S. A., Houry, W. A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C. and Hartl, F. U. Polypeptide Flux through Bacterial Hsp70:: DnaK Cooperates with Trigger Factor in Chaperoning Nascent Chains. *Cell* (1999) **97**, 755-765.

14. Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. and Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature (London)* (1999) **400**, 693-696.

15. Spudich, G. M., Miller, E. J. and Marqusee, S. Destabilization of the *Escherichia coli* RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J. Mol. Biol.* (2004) **335**, 609-618.

16. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M. and Skelton, N. J. (2007). *Protein NMR spectroscopy*. Second edit, Elsevier Academic Press, San Diego.

17. Ernst, R. R., Bodenhausen, G. and Wokaun, A. (1989). *Principles of nuclear magnetic resonance in one and two dimensions*, Oxford University Press, New York.

18. Sattler, M., Schleucher, J. and Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* (1999) **34**, 93-158.

19. Goldman, M. Spin temperature and NMR in solids. *Mir, Moscow* (1972).

20. Reimer, J. A. Nuclear hyperpolarization in solids and the prospects for nuclear spintronics. *Solid State Nucl. Magn. Reson.* (2010) **37**, 3-12.

21. Kaptein, R., Dijkstra, K. and Nicolay, K. Laser photo-CIDNP as a surface probe for proteins in solution. *Nature* (1978) **274**, 293-294.

22. Hore, P. J. and Broadhurst, R. W. Photo-CIDNP of biopolymers. *Prog. Nucl. Magn. Reson. Spectrosc.* (1993) **25**, 345-402.

23. Turro, N. J. Influence of nuclear spin on chemical reactions: Magnetic isotope and magnetic field effects (A Review). *Proc. Natl. Acad. Sci. U. S. A.* (1983) **80**, 609.

24. Griffin, R. G. and Prisner, T. F. (2010). High field dynamic nuclear polarization—the renaissance, Vol. 12, pp. 5737-5740. Royal Society of Chemistry.

25. Natterer, J. and Bargon, J. Parahydrogen induced polarization. *Prog. Nucl. Magn. Reson. Spectrosc.* (1997) **31**, 291-314.

26. Eun, Y. J., Kurt, N., Sekhar, A. and Cavagnero, S. Thermodynamic and kinetic characterization of ApoHmpH, a fast-folding bacterial globin. *J. Mol. Biol.* (2008) **376**, 879-897.

27. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* (1958) **181**, 662-666.

28. Kuprov, I. and Hore, P. J. Chemically amplified 19F–1H nuclear Overhauser effects. *J. Magn. Reson.* (2004) **168**, 1-7.

29. Lyon, C. E., Jones, J. A., Redfield, C., Dobson, C. M. and Hore, P. J. Two-dimensional ^{15}N–^{1}H photo-CIDNP as a surface probe of native and partially structured proteins. *J. Am. Chem. Soc.* (1999) **121**, 6505–6506.

30. Cabrita, L. D., Dobson, C. M. and Christodoulou, J. Protein folding on the ribosome. *Curr. Opin. Struct. Biol.* (2010) **20**, 33-45.