

# **How do Proteins Work?**

A Thesis

Submitted for the Degree of Doctor of Science

to the

Faculty of Science, University of Sydney

by

Helen Jane Dyson

October 1, 2007

*Dedication*

For Nick and Kate

With love

# Acknowledgments

The work described in this thesis would have been impossible without the contributions of a very large number of collaborators and co-workers. Their names are inscribed in the author lists and acknowledgments of the published work presented here. I would particularly like to thank my colleagues at Scripps, who have provided an amazing environment for the flowering of science in so many areas.

I am particularly grateful to Maria Martinez-Yamout and Chiaki Nishimura for long and fruitful collaborations, and to John Chung and Gerard Kroon for exemplary care and feeding of the spectrometers. Many technicians have provided excellent assistance, especially Linda Tennant, who has been part of the lab since 1984. Assistance with the compilation of this thesis has been provided primarily by Cristina Mora, with contributions by Bonnie Weier and Ruby Santos.

My most important and heartfelt acknowledgment must be to my husband, Peter Wright, whose support has never wavered, even during the darkest days. The reader of this work will rapidly appreciate the extent of his contribution to “my” work.

# Contents

	Page
Synopsis	3-4
Preamble	5-68
Introduction	5-6
Chapter 1 The Protein Folding Problem	7-24
Chapter 2 Structure, Dynamics and Folding of Proteins	25-34
Chapter 3 Dynamics and Catalysis	35-42
Chapter 4 Intrinsically Unstructured Proteins	43-64
Chapter 5 Technical Advances	65-68
List of Publications (including intellectual input)	69-86
Published Work 1975-2007 in chronological order	87-2075

## SYNOPSIS

The two major long-term themes in my research are:

- the understanding of how the amino acid sequence of a protein determines its final folded structure.
- the understanding of enzyme and protein function through study of structure and dynamics.

Primary research techniques include NMR spectroscopy for study of structure and dynamics, mass spectrometry, and equilibrium and kinetic CD and fluorescence spectroscopy. Molecular cloning techniques have been an extremely powerful addition to the arsenal, allowing us to prepare labeled proteins in the amounts necessary for structural studies by NMR. A brief description of major areas of research follows:

The initial steps in protein folding are the most relevant for the exploration of the link between sequence and folded structure. They are also the most difficult to study experimentally. We initially tackled this problem by using short peptide fragments of proteins as a model system for the earliest events in the folding process. The sequence specificity for residual structure in short peptides in solution was determined, to give insights into sites that might act as folding initiation sites. Reverse turns and nascent helices were characterized, as well as propensities for the formation of ordered helical structure. In order to determine which parts of proteins potentially contained folding initiation sites, large sets of peptides were studied, corresponding to the complete sequences (in pieces) of three proteins containing widely different secondary and tertiary structures. These studies remain the definitive work in this field. More recently the emphasis has been on the use of new NMR techniques to probe residual structure in denatured forms of the full-length proteins.

NMR is the method of choice for the examination of the solution conformational ensembles of equilibrium folding intermediates and unfolded states of proteins. The primary vehicle for these studies has been the helical protein apomyoglobin in both kinetic measurements of folding and equilibrium model systems for the various stages of the folding pathway. Comparison of the folding pathways of apomyoglobin and the related but widely divergent protein apoleghemoglobin, as well as the  $\beta$ -sheet protein apoplastocyanin, have provided important insights into the role of the amino acid sequence in the initiation and propagation of protein folding, leading to the hypothesis that protein folding initiation arises in areas where there are local high densities of long (but not necessarily completely hydrophobic) side chains. This work has had a significant impact on the protein folding field.

Our most recent interest is in the interactions of chaperones and their co-chaperones and client proteins. A major new project has recently been instituted to determine the conformational states of so-called "client" proteins when bound to the chaperone Hsp90.

I and my collaborator Peter Wright are probably best known for the recognition of the role of intrinsically unstructured proteins in the metabolism of the cell. A number of protein systems are unfolded or only partly folded until they bind ligand or substrate. While this has been known for many years to apply to peptide hormones, we now realize that certain transcription factors, cyclin-dependent kinase inhibitors and other proteins are unfolded in the

absence of their natural receptors. In addition, folding processes have been implicated in a number of disease states, such as Alzheimer's disease and prion diseases

A number of protein systems have been studied to elucidate the link between chemistry, structure, dynamics and function. The system that has been studied the longest in my laboratory is *E. coli* thioredoxin, a small multi-functional thiol-disulfide oxidoreductase. The contribution from my laboratory has been particularly significant in the delineation of the mechanism of *E. coli* thioredoxin. Our hypothesis that there is a shared proton between the two active-site cysteines has received a great deal of attention.

A clear relationship between structure and function was seen for another system, a metalloprotein obtained from an acidophilic bacterium. *Thiobacillus ferrooxidans* rusticyanin has the astonishing property that it contains a highly stable Type I or "blue" copper site at pHs well below 4. In fact the protein is stable and functional in sulfuric acid at pH 0.5, in spite of a copper coordination site that includes two histidines, which would normally be protonated at such a low pH. Our solution structure of reduced rusticyanin clearly shows that the acid-stability of rusticyanin resides in the composition of the environment of the copper site, which contains a high proportion of aromatic groups, causing steric hindrance to the dissociation of the ligand histidines.

The relationship between dynamics of the polypeptide chain and enzyme catalysis is a major theme in my laboratory. The success of structure-based drug design has so far been quite modest. This may be because the design process incorporates only static structural information. Our working hypothesis is that a requirement for efficient enzyme catalysis may well be that the active site be flexible, and that enzymes have therefore evolved to incorporate this flexibility. To test this hypothesis we use mutagenesis coupled with full characterization of changes in enzymatic function, structure and dynamics of two very different enzyme systems, *E. coli* dihydrofolate reductase and a metallo- $\beta$ -lactamase. For dihydrofolate reductase, dynamics measurements have enabled us to form a detailed picture of the motions of the protein as it proceeds through its catalytic cycle.

## **Preamble: How do Proteins Work?**

### **INTRODUCTION**

Throughout the last two decades, I have been intensively involved in research at the Scripps Research Institute, California. Beginning as a postdoctoral fellow (research associate) in the Department of Molecular Biology, I have risen through the ranks to my present position as Professor in the same department. My research work shows a progression through the years, not only because I started as a bench scientist and am now mostly involved with planning, interpretation and reporting of experiments, but also because my field of study has seen astonishing developments and breakthroughs that have facilitated existing projects and suggested new and challenging ones. It is hard to remember how constrained we were during the 80s, when the only insights that we could obtain, for example into the protein folding process, were derived from studies of short peptides. Now we study full-length proteins, sometimes protein complexes, and the new insights we obtain are of a breadth commensurate with the increase in size of the molecules that we can study. It continues to be an exciting time to be a molecular biologist!

In the following discussion, I have divided the description of my work into chapters grouped by topic. However, in a somewhat surprising way, the many strands of my research, seemingly so disparate when they were started, appear to be converging. For example, the work on peptide conformation led to the protein folding projects, where we learned to deal with unfolded and partly folded proteins. In the meantime, we were frequently observing the puzzling phenomenon in studies of transcriptional activators, that unstructured domains can actually be functional. I would say that this insight, and the work that has arisen from it in the last few years, is probably the work for which I am best known in the research community.

I would like to give a short explanation for the research relationships that are apparent from the multitudes of authors on the papers that I report as “mine”. Firstly, our policy is to include as authors those people who have contributed “substantially” to the work being reported. This may be in the provision of some of the materials, or in useful experimental or theoretical input to the understanding of the subject. The topics we are researching are complex, sometimes presenting exceptional experimental challenges and taking considerable time and effort to complete. As recently reported<sup>1</sup>, it is inevitable that more than one person will be involved in such ambitious projects. I have therefore tried (see the list of annotated publications following chapter 5) to enumerate for each publication on the list an estimate of my contribution and its nature, as well as naming the “Principal Investigator” (PI) for the work, that is, the person who instigated, supervised and was responsible for the project. For approximately 55% of the publications, I was the PI, either alone (23%) or together with one or two others (32%). For about 28% of the publications, I performed experiments, either the primary work or in a secondary capacity, with someone else (frequently the first author of the paper) providing the primary experimental effort. In later years, my main role has been supervisory, assisting with interpretation and reporting of the experimental data, usually obtained by postdoctoral fellows or graduate students. I have been particularly involved in the authorship of the research papers themselves (92% of first and second drafts), frequently writing what amounts to a first draft of papers for junior workers whose first language is not English. This writing task frequently

involves considerable input into the interpretation of the results, and may result in the planning and execution of more experimental work before the paper is submitted.

Throughout my recent research career, Dr. Peter Wright has played a prominent role, reflected in the high proportion of “my” papers in which he appears as a co-author. In some cases, I have named myself as PI, as this work was independent of Dr. Wright’s projects, even when his name appears as an author; in other cases, Dr. Wright is named as PI, but I am named as an author on the paper because I provided scientific input into one of his projects. In a large number of cases, I have named both of us as PI. These projects are completely collaborative, and our contributions are equal. Particularly in the latter case, I believe that the collaborative effort has yielded much more insight, productivity and significant advances than would have occurred if we had worked totally independently. I believe that my work can be judged on its own merits, recognizing that much of it was achieved through collaborative work. I have been particularly fortunate in the collaboration with Dr. Wright, who is also my husband. During the period when the research described herein was under way, we also raised two children, both now at university.

#### Reference List: Introduction

1. Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science* **316**, 1036-1039.



## **CHAPTER 1. THE PROTEIN FOLDING PROBLEM**

Simply stated, the protein folding problem concerns the transformation of linear (one-dimensional) information (the gene sequence, coding for the amino acid sequence of the protein) into three-dimensional structures that are functional. All of the information required for the final folded state of a globular protein is contained in the amino acid sequence. Our goal for the last 20 years has been to determine the rules that govern the initiation and propagation of the folding process for a protein, utilizing the best experimental means that were available at any given time. Our primary experimental method for the study of the protein folding problem has been solution nuclear magnetic resonance spectroscopy (NMR), which is particularly appropriate since folding occurs in solution. NMR is capable of detecting quite small populations of partly-folded conformers within an ensemble. The NMR measurements are supplemented by other spectroscopic data, from circular dichroism (CD) and fluorescence spectroscopy, which can give valuable, though rather limited, corroborative information.

### **1.1 Short Peptides: A Model System for Initiation of Protein Folding**

NMR spectroscopy provides a wealth of information on almost every nucleus in a molecule: most of the protons ( $^1\text{H}$ ) are detectable, and  $^{13}\text{C}$  and  $^{31}\text{P}$  signals can also be detected at natural abundance. For complete information on the signals from carbon and nitrogen nuclei in a protein, isotopic labeling is necessary. However, for biological macromolecules, proteins and nucleic acids, the number of signals from all of the nuclei causes a problem – the signals cannot be distinguished because of resonance overlap. During the early years of our protein folding research, the state of the art in NMR spectroscopy could not be used to undertake measurements on unfolded proteins, because the resonances are badly overlapped. Only limited information was available for fully folded proteins, where the resonance dispersion is better. Full assignment of  $^1\text{H}$  resonances for folded proteins did not occur until the late 1980s (see, for example,<sup>1</sup>). It was not until the early 1990s that the spectra of unfolded proteins were even looked at<sup>2</sup>. A summary of methodological advances that allowed more information to be obtained on larger systems is shown in Table 1.1.

#### **5.3.3 Anatomy of a Peptide Reverse Turn Structure**

Our task in 1984 was therefore to design experiments that would give insights into the folding process of proteins, without being able to examine the proteins directly. We turned to protein fragments, or peptides, which were recently available through the solid-phase synthesis techniques that earned Bruce Merrifield the 1984 Nobel Prize in Chemistry. At the time, the general consensus was that peptides were uniformly random in their conformations, that is, their conformational ensemble showed no preferences for any particular structure – all possible structures would be present in equal concentrations. This was thought to be especially so for peptides dissolved in water, as the hydrogen bonding potential of the bulk water would trump any feeble attempts at intramolecular hydrogen bonding in the peptide. At about this time, Richard Lerner and his colleagues at Scripps Clinic and Research Foundation (later called the Scripps Research Institute) had made some very provocative observations when they raised antibodies in mice and other animals to peptide immunogens. If antibodies were raised against a certain peptide, coupled to a large protein carrier to stimulate the immune system, the resulting antibodies would reliably react with the full-length, folded protein from which the peptide sequence had been derived<sup>3,4</sup>. This observation suggested that peptides might sample the conformations present in the folded protein, possibly at populations that could be detected in an NMR experiment. Against all common dogma, we therefore examined one of the immunogenic

**Table 1.1. Time Line of NMR Innovations**

Year	Method	Author	Reference	$M_r(\text{max})(\text{KDa})$
1971	1D NOE	Freeman	1	5
1976	COSY	Ernst	2	10
1979	TD NOE	Wüthrich	3	
1979	NOESY	Ernst	4	
1982	RELAY	Ernst	5	12
1983	TOCSY	Ernst	6	
1983	2Q	Ernst	7	
1983	MQ filter	Ernst	8	
1985	isotope labels	Griffey	9	
1989	$^{13}\text{C}$ editing	Rance	10	
1990	2D HSQC	Campbell	11	30
1990-1994	3D, triple resonance	various	12-16	
1997	TROSY	Wüthrich	17	>100

- Freeman, R. & Hill, H. D. W. (1971). Nuclear Overhauser effect in uncoupled NMR spectra of carbon-13. *J. Magn. Reson.* **5**, 278-280.
- Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229-2246.
- Dubs, A., Wagner, G., & Wüthrich, K. (1979). Individual assignments of amide proton resonances in the proton NMR spectrum of the basic pancreatic trypsin inhibitor. *Biochim. Biophys. Acta* **577**, 177-194.
- Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546-4553.
- Eich, G., Bodenhausen, G., & Ernst, R. R. (1982). Exploring nuclear spin systems by relayed magnetization transfer. *J. Am. Chem. Soc.* **104**, 3731-3732.
- Braunschweiler, L. & Ernst, R. R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521-528.
- Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983). Analysis of networks of coupled spins by multiple quantum NMR. *Mol. Phys.* **48**, 535-560.
- Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983). Improved spectral resolution in COSY  $^1\text{H}$  NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479-485.
- Griffey, R. H., Jarema, M. A., Kunz, S., Rosevear, P. R., & Redfield, A. G. (1985). Isotopic-label-directed observation of the nuclear Overhauser effect in poorly resolved proton NMR spectra. *J. Am. Chem. Soc.* **107**, 711-712.
- Rance, M., Wright, P. E., Messerle, B. A., & Field, L. D. (1987). Site-selective observation of nuclear Overhauser effects in proteins via isotopic labeling. *J. Am. Chem. Soc.* **109**, 1591-1593.
- Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990). Comparison of techniques for  $^1\text{H}$ -detected heteronuclear  $^1\text{H}$ - $^{15}\text{N}$  spectroscopy. *J. Magn. Reson.* **87**, 488-501.
- Bax, A., Clore, G. M., & Gronenborn, A. M. (1990).  $^1\text{H}$ - $^1\text{H}$  correlation via isotropic mixing of  $^{13}\text{C}$  magnetization, a new three-dimensional approach for assigning  $^1\text{H}$  and  $^{13}\text{C}$  spectra of  $^{13}\text{C}$ -enriched proteins. *J. Magn. Reson.* **88**, 425-431.
- Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496-514.
- Ikura, M., Kay, L. E., & Bax, A. (1991). Improved three-dimensional  $^1\text{H}$ - $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy of a  $^{13}\text{C}$ -labeled protein using constant-time evolution. *J. Biomol. NMR* **1**, 299-304.
- Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432-440.
- Zhang, O., Kay, L. E., Olivier, J. P., & Forman-Kay, J. D. (1994). Backbone  $^1\text{H}$  and  $^{15}\text{N}$  resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* **4**, 845-858.
- Pervushin, K., Riek, R., Wider, G., & Wüthrich, K. (1997). Attenuated  $T_2$  relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366-12371.

Peptides (sequence Tyr-Pro-Tyr-Asp-Val-Pro-Asp-Tyr-Ala) by NMR, utilizing the then-new 2-dimensional  $^1\text{H}$  methods. We found unequivocal evidence for a conformational preference for a reverse turn conformation in the Tyr-Pro-Tyr-Asp sequence, which was published in 1985<sup>5</sup>. This seminal paper showed firstly that conformational preferences could be present in short linear peptides in water solution, and secondly that these conformational preferences could be detected and characterized by NMR. It provided us with a basis for further studies to begin to map out the possible ways that protein folding can be initiated in a linear polypeptide sequence.

Since the reverse turn conformation appeared to be quite sequence-specific – it was observed in the Tyr-Pro-Tyr-Asp sequence, but not, for example, in the Val-Pro-Asp-Tyr sequence of the same peptide – our next set of experiments was designed to probe the sequence specificity of the reverse turn conformation. Starting with the known reverse-turn sequence, we made several series of one-position variants on the sequence Tyr-Pro-Tyr-Asp-Val (YPYDV in single-letter notation), giving sets of peptides YPXDV and YPYXV, where the third and fourth amino acids in the sequence were replaced by all 19 other amino acids<sup>6</sup>. The highest population of the reverse turn conformation was found for the sequence YPGDV. Several years later, we finally completed this study, for the peptide sets XPGDV and AXGDV<sup>7</sup>: although we were able to show that the sequence APGD was overall the best at reverse turn formation, the absence of proline at position 2 of the turn was surprisingly not as deleterious as we had predicted, as long as the A, G and D were present.

Our original observations<sup>5</sup> indicated that the 9-residue peptide YPYDVPDYA contained two populations of molecules, which differed in the conformation of the proline at position 2. The major population contained the *trans* isomer of Pro2, but a significant minor population of the *cis* isomer was present in this peptide. The population of *cis* isomer depends mainly on the nature of the amino acids at positions 1 and 3 of the peptide, with aromatic residues (Tyr, Phe, Trp) in both positions being the most favorable for formation of the *cis* peptide bond<sup>6,7</sup>. In the course of these studies, we made the astonishing observation that the addition of an extra residue (or even an acetyl protecting group) at the N-terminus of the peptide resulted in a very significant increase in the population of the *cis* isomer, from 35% *cis* for YPYDV and 23% *cis* for YPGDV to 57% *cis* for AYPYDV<sup>6</sup>. This increase was attributed to the formation of a stable Type VI reverse turn conformation in the sequence Ala-Tyr-(*cis*)Pro-Tyr. The sequence dependence of this turn conformation was extensively investigated<sup>8</sup> and, since the population of the turn was so high (> 70% of the *cis* peptide molecules) we performed a structure calculation for the turn, based on observed NMR parameters (nuclear Overhauser effects [NOE] and coupling constants), which gave a structural basis for the preference for aromatic amino acids flanking the proline: the three rings are neatly stacked together in the structure, consistent with the observed extensive shifts of the *cis*-proline ring resonances from random-coil values<sup>9</sup>. We were also able to show that solvent water was excluded from this cluster, a likely factor contributing to its stability<sup>10</sup>.

### 1.1.2 Links between Secondary Structure Propensities and Protein Folding

Among those who thought about protein folding at that time<sup>11</sup>, there was a general consensus that the initiation of folding would have to occur associated with secondary structure formation, either helix or reverse turn, although there were countervailing opinions that hydrophobic collapse might precede secondary structure formation<sup>12</sup>. The latter suggestion seemed counter-intuitive to us (how would the “collapsed” chain sort itself into a proper structure?). We therefore set out to look for secondary structures in peptides derived from proteins where the three-dimensional structure was known. The first such peptide that we

examined was the C-helix segment of the all-helix protein myohemerythrin. This helix is regular and well formed in the intact protein. As an isolated peptide, this sequence does not form helical structure in solution. However, we noticed some distinctive features in the NMR spectrum that indicated that the C-helix peptide populated states that could lead to helix. Specifically, there was evidence for the formation of multiple, interlinked reverse turn conformations in part of the peptide. When the helix-promoting solvent trifluoroethanol was added to the solution, helical structure was formed, but only in the region of the peptide where we had observed the interlinked reverse turns in pure water. We named this state “nascent helix”<sup>13</sup>, a term that has now entered the general literature to describe such states.

Acting on the assumption that the presence of nascent secondary structure in peptide fragments identified potential initiation sites for folding of the intact protein, the myohemerythrin study was later extended to encompass peptides corresponding to the entire sequence of myohemerythrin<sup>14</sup>, with a second set of peptides representing the entire sequence of an all- $\beta$  protein, plastocyanin<sup>15</sup>. We found that the peptides corresponding to helices frequently populated nascent or folded helix in solution, while the peptides representing  $\beta$ -strands appeared to have no propensity for helix or turn formation, but remained extended for the vast majority of the population. Although this result might seem unsurprising, this was the first demonstration that, in general, the isolated sequences corresponding to secondary structure elements have a propensity for formation of the same type of secondary structure as in the intact protein. This is not always the case, as demonstrated in an X-ray study at about the same time<sup>16</sup>. This project was further extended following the publication of extensive evidence that helical structure could be stabilized by the provision of the correct capping sequences at the ends<sup>17</sup>: the presence of greater length and a capping sequence stabilizes the helical structure in a peptide representing the C-terminal sequence of myohemerythrin<sup>18</sup>.

The search for folding initiation sites by using peptides was extended to a third protein, apomyoglobin. Initial work focused on the two C-terminal helices of apomyoglobin, which were studied as isolated peptides corresponding to the G and H helices<sup>19</sup>, as a series of short peptides spanning the G-H linker region<sup>20</sup> and as a 51-residue peptide encompassing both helices<sup>21</sup>. Again, detectable propensities for helix or nascent helix were observed only for the regions that were helical in the intact protein, while the linker region contained a highly populated reverse turn (sequence HPGDF, reminiscent of the 1988 peptide work) that appeared to act as a “helix stop” signal. These results were all consistent with our hypothesis that incipient secondary structure was the key to protein folding initiation. Interestingly, the long peptide encompassing both G and H helices and the linker between them showed some unexpected properties: although carefully engineered by “mutations” at sites that would contact other parts of the protein in the fully folded form, the 51-residue peptide formed a very stable  $\square$ immer, giving a 4-helix bundle structure quite unlike that found in the intact protein. Nevertheless, the G and H helices appeared to have formed a helical hairpin structure as expected, consistent once again with the “secondary structure” hypothesis. The apomyoglobin peptide story was later amplified to include the entire sequence of apomyoglobin<sup>22</sup>. The A helix peptide proved quite insoluble, and could not be studied by itself, but a longer peptide containing both the A and B helix sequences was soluble, and showed significant propensity for helix throughout its length. The implications for these results for the folding mechanism of apomyoglobin will be discussed in a later section, together with results obtained for the intact protein.

### 1.1.3 Digression: Immunogenicity of Peptides

Although our interest in peptide structure became more and more applied to the protein folding problem, we remained fascinated by the phenomenon of peptide immunogenicity. We had found and thoroughly characterized a conformational propensity for a recognizable structure in a short immunogenic peptide, but it was not at all clear that this structured form would necessarily also be the primary immunogen. Indeed, it appeared that the most immunogenic peptides were those that corresponded to the most mobile parts of the parent protein<sup>4</sup>. On one level, this makes perfect sense, as long as peptide immunogens were completely unfolded. What does a conformational preference in a peptide immunogen mean? We explored this question in several ways, firstly by pooling our biophysical insights with the immunological insights of the Lerner lab in a series of review articles<sup>23,24</sup> and secondly by examining a large number of different immunogenic peptides for patterns<sup>25-31</sup> and for the effects of conformational restriction<sup>32,33</sup>. At this stage, the jury is still out, but it appears that most of these phenomena can be explained by the extreme versatility of the immune system.

### 1.2 Folding of Proteins Studied by NMR

During the period 1984-1993, a number of innovations, both in NMR (Table 1.1) and in protein expression techniques in bacteria and other systems, enabled the examination of the folding of proteins by NMR. The technique, pioneered in the Baldwin<sup>34</sup> and Englander<sup>35</sup> labs, takes advantage of the differential rates of deuterium exchange of amide protons in proteins, depending on the pH of the solution and on whether the amide proton is hydrogen bonded. If a protein is placed in a solution of <sup>2</sup>H<sub>2</sub>O (D<sub>2</sub>O), the majority of the protons, which are attached to carbon atoms, will not be exchanged, while the protons attached to nitrogen and oxygen will be exchanged for deuterium. Since D is invisible in the NMR experiments used to detect protons, the signals of the exchanged protons will disappear from the spectrum. If an unfolded protein is placed in D<sub>2</sub>O buffer, all of the amide protons, as well as –OH and –NH<sub>2</sub>, will be exchanged for D within a few seconds. However, if a folded protein is placed in D<sub>2</sub>O, a subset of the backbone amide protons may be protected from exchange due to stable hydrogen bonding, for example in helical or β-sheet secondary structure. Some of these protected amides may persist for days or longer; these hyper-protected amides are the “probes” that allow site-specific estimates of folding rates to be made, using the “quench-flow” experiment.

In order for classical quench-flow measurements to be used to give information on the folding of a protein, several conditions must be met: there must be sufficient probe amides, the protein must be reversibly refoldable and tolerant of a variety of pH conditions without aggregation, and the NMR spectrum of the fully-folded protein must be good enough that signals can be distinguished and resonance assignments made. The first quench-flow experiments done in the Wright/Dyson lab reported on the folding pathway of apomyoglobin<sup>36</sup>. This report generated much excitement<sup>37,38</sup>, as it showed for the first time that a folding intermediate could form part of the folding pathway, rather than being an off-pathway dead end. The results showed unequivocally that the folding intermediate formed most rapidly (within ~ 6 ms) in the quench-flow experiments (termed the “burst phase” intermediate) contained much of the A, G, and H helices and part of the B helix already folded and stabilized sufficiently for amide protons to be protected from exchange.

### 1.3 Kinetic and Equilibrium Folding Studies of Apomyoglobin

The direction for the next stage of our inquiry into the protein folding process appeared quite clear in 1993: the peptide studies showed that the isolated H helix contained a significant

population of helical structures in water solution. While the G helix appeared to contain very little helix in water, the population of helix for both G and H helices could be increased by addition of trifluoroethanol or by the proximity of other helical segments. The quench-flow experiments indicated that the H helix formed part of the burst phase intermediate of folding, that is, it was among the secondary structure elements that were folded first. Was the H helix the major initiation site for the folding process for apomyoglobin?

### 1.3.1 Site-Directed Mutagenesis Unlocks the Secret

To test this hypothesis, we turned to mutagenesis, an option that became available to us only in the mid-1990s. The myoglobin holoprotein (containing the heme prosthetic group) had been the focus of one of the major research efforts in the Wright lab during the 1980s<sup>39-41</sup>. All of these studies used sperm whale myoglobin (actually from sperm whales, which became a problem at about this time, due to bans on whaling). During this period, cloning and expression technology was beginning to appear, and in 1987, the gene for sperm whale myoglobin was reported<sup>42</sup>; the expression of this gene was optimized in *E. coli* for efficient isotope labeling for NMR<sup>43</sup>. With this methodology in hand, we were able to design mutant proteins to assess the role of various parts of the protein in initiation and propagation of folding. To date, the Wright/Dyson lab has prepared and characterized 16 mutants (Table 1.2), with more to come. This is an enormous amount of work, which has served to refine our ideas on the factors that promote the initiation of protein folding. In short, these experiments and their interpretation have identified the key interactions that occur to initiate folding, a major step forward in the understanding of the folding process. Although our results have been obtained for only one system, we believe that the fundamental nature of the interactions will make our conclusions widely applicable to proteins in general. I will summarize these results in the following paragraphs.

To test the hypothesis that helical structure in the H helix was “the” initiation site for apomyoglobin folding, the double mutant N132G, E136G was prepared<sup>44</sup>. These two substitutions were designed to disrupt the intrinsic helical propensity in the H helix, as observed in the corresponding peptides<sup>19-21</sup> and in acid-denatured apomyoglobin<sup>45</sup> (see later section). If helical structure in the H helix is essential for the formation of the burst phase intermediate in the folding process, then we should observe changes in the burst phase, in the best possible scenario, a slowing of this process into the observable time regime. Although the mutant protein contains less helical structure in the acid denatured state and in the equilibrium molten globule state, the kinetic folding process for the mutant, measured by optically-detected stopped-flow methods, is very similar to that of the wild-type protein, both in the amplitude of the burst phase and in the rate of the observable phase. Upon close examination of the quench flow data, it is clear that the H helix has been considerably destabilized in the burst phase, yet the overall effect on the folding process is virtually undetectable. Clearly our hypothesis that the H helix is the primary initiation site is not correct. This study demonstrated that the folding mechanism of apomyoglobin is extremely robust to sequence changes, a conclusion that is consistent with the similarity of myoglobin structures with widely differing amino acid sequences. Another important revelation from this study was that accurate quench flow-NMR experiments were capable of detecting subtle changes in folding behavior at individual sites in the protein, giving important insights into the folding process at the level of individual amino acids, and even individual nuclei.

**Table 1.2. Apomyoglobin Mutants**

Mutation(s)	Location	reference
(N132G,E136G)	H helix	1
H64F	E helix	2
I28A, L29A, I30A, L32A	B helix	3
(W14G,G73W)	A, E helix	4
V10A, V17A	A helix	5
L69A, L72A	E helix	5
L104A, I111A, L115A	G helix	5
L135A, I142A	H helix	5

1. Cavagnero, S., Dyson, H. J., & Wright, P. E. (1999). Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* **285**, 269-282.
2. Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J., & Wright, P. E. (2000). Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. *Biochemistry* **39**, 11227-11237.
3. Nishimura, C., Wright, P. E., & Dyson, H. J. (2003). Role of the B helix in early folding events in apomyoglobin: evidence from site-directed mutagenesis for native-like long range interactions. *J. Mol. Biol.* **334**, 293-307.
4. Nishimura, C., Lietzow, M. A., Dyson, H. J., & Wright, P. E. (2005). Sequence determinants of a protein folding pathway. *J. Mol. Biol.* **351**, 383-392.
5. Nishimura, C., Dyson, H. J., & Wright, P. E. (2006). Identification of Native and Non-native Structure in Kinetic Folding Intermediates of Apomyoglobin. *J. Mol. Biol.* **355**, 139-156.

The next phase of the mutagenesis work (apart from the H64F mutant<sup>46</sup>, which was designed to probe a different question) involved an intensive collaboration between Dr. Chiaki Nishimura, an exceptionally capable senior scientist in the group, as well as Dr. Wright and me. A suite of 14 mutant proteins was prepared, where hydrophobic residues in the A, B, E, G and H helices were substituted with alanine<sup>47-49</sup>. Each mutant protein was thoroughly characterized, both under equilibrium conditions at pH 2, pH 4 and pH 6 (see later section), and in quench flow experiments, to gauge the effects of the mutations on the folding mechanism at various sites in the protein. Firstly, as expected, there were detectable, usually major, changes in the protection of amides in the vicinity of the mutation sites. This is consistent with the type of change made, from a bulky hydrophobic group to alanine. However, interestingly, each mutant protein showed a different set of long-range effects on the constitution of the kinetic burst phase intermediate. For example, the L32A mutant protein showed a very significant destabilization of the G helix in the kinetic intermediate, relative to the wild-type protein, whereas the L29A mutant showed additional stabilization of the E helix<sup>47</sup>. Both of the mutated residues are in the B helix, within 3 positions in the sequence, yet when mutated they show detectably different folding behavior. Examination of the structure of fully-folded (wild-type) myoglobin shows that the side chain of L32 makes contact with residues in the G helix, while the side chain of L29 approaches the E helix. Thus, the structure of the kinetic burst phase intermediate, formed within 6 ms of the initiation of the folding process, must already have a native-like topology, with helical segments packed more or less in the correct geometry for the formation of the final structure. This is of course consistent with all of the previous evidence that the apomyoglobin burst phase intermediate is on the folding pathway, and indeed, is obligatory<sup>50</sup>, that is, all molecules in the ensemble of the folding protein must pass through this state. However, it begs the question as to the nature of the barriers that prevent the entire molecule from folding in the burst phase.

Further studies refined our understanding of the folding process. In a folding experiment, the protein begins as an ensemble of unfolded conformers. Thus, every molecule in the ensemble starts to fold from a different starting point, and there will be as many “folding pathways” that

occur as there are molecules in the ensemble. Nevertheless, there are observable trends in the process, which we and many others have observed by various spectroscopic means. This implies that there are preferred routes that are taken by a majority of the molecules as they fold. This has been pictured as a “folding funnel”, a rough landscape with high-energy ridges and lower-energy valleys leading to the final lowest-energy, fully folded state<sup>51</sup>. For some proteins, such as lysozyme, there are several possible pathways of similar low energy, so that multiple phases are observed in the folding process<sup>52</sup>. For apomyoglobin, all of the molecules appear to pass through the burst phase kinetic intermediate<sup>50</sup>. The examination of the folding behavior of mutants of the A, E, G and H helices<sup>49</sup> illustrates the mechanism by which the folding of part of the molecule is “stalled” in the kinetic intermediate. Because we are able to dissect the folding behavior of apomyoglobin at almost atomic-level detail, we are able to observe slight anomalies in the positions of residues affected by the mutations in the A, E, G and H helices. Like the B helix mutants described in the previous paragraph, this set of mutants showed evidence of native-like topology of the burst phase intermediate. However, there were subtle differences in the residues that were affected by the mutations, that is, the affected residues were close to, but not exactly the same as, the residues that would be expected from the three-dimensional structure of the fully folded protein. By carefully localizing the long-range effects for each of the mutants, we were able to show that the burst phase intermediate of apomyoglobin contains a translocation of about one helical turn in the position of the H helix, relative to its position in the fully folded protein. This is illustrated in Figure 1.1.

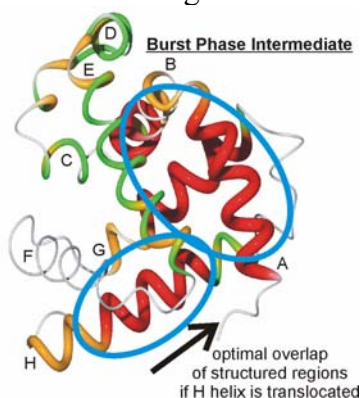


Figure 1.1 Structured regions in the burst phase intermediate, mapped onto the X-ray structure of fully folded apomyoglobin. Optimal contact between the protected areas would place the H helix in a non-native translocated position as indicated. Red: amides are highly protected within the burst phase of the quench flow experiment (~ 6.4 ms). Yellow: slight protection; green: protected only in the detectable phase of folding (~ 150 ms).

We infer that correction of this minor structural aberration is the cause of the slowing of the folding process for the residues involved in the observable phase of apomyoglobin folding. The subtle nature of this change is also consistent with the on-pathway nature of the intermediate – if a major structural change were required to correct a seriously mis-folded intermediate, an additional kinetic phase should be observed. In addition, we might expect that the pattern of amide proton protection throughout the folding process would not be so uniformly monotonic: amides protected early in the process as the misfolded intermediate was formed might be observed to become solvent exposed and exchanged at later stages of folding.

#### **1.4 So How is Protein Folding Initiated?**

The characteristic features of the amino acid sequence that govern the initiation of protein folding continued to intrigue us. In order to describe the progression of ideas that led to our present understanding of the mechanism of protein folding initiation, it is necessary first to describe the results of an extensive series of experiments on the unfolded and partly folded forms of apomyoglobin. This work also has considerable bearing on the issue of intrinsically disordered proteins (see Chapter 4).



### **1.5 Unfolded and Partly Folded Apomyoglobin at Equilibrium**

It has long been known that apomyoglobin forms an intermediate structure under equilibrium conditions, either at pH ~ 4 or in intermediate concentrations of denaturants such as urea. At low pH (~ 2) and in the presence of high concentrations of denaturant, the protein is unfolded, but can be reversibly refolded by removal of the denaturant or raising the pH (this is the basis for the quench-flow kinetic experiment described above). As NMR techniques became more powerful, and as the spectrometers themselves became larger and more sensitive, we were able to utilize multi-dimensional techniques and isotopic labeling to begin to explore the conformational ensembles of these unfolded and partly folded forms of apomyoglobin. The first step in a quantitative evaluation of residual structure in unfolded proteins was to define NMR parameters that could be used as diagnostics of, for example, helical structure. The primary datum that is obtained from the NMR experiment is the chemical shift. In itself, it contains a great deal of information about the local environment of the nucleus, but there are many influences on the chemical shift, only some of which are understood, with the result that the calculation of chemical shifts a priori from structures (and conversely, the prediction of structures from chemical shifts alone) are still in their infancy. Other NMR parameters can be used to give information on the composition of the ensemble, for example, coupling constants, NOEs and relaxation times. Of these, the coupling constant and NOE are not usually very useful, particularly in highly heterogeneous ensembles, due to the effects of ensemble averaging. For both of these quantities, a rather significant proportion of the ensemble must show a conformational preference for a similar local structure before they can be interpreted with any validity. On the other hand, while chemical shifts and relaxation time measurements are also subject to ensemble averaging, the populations of molecules with local residual structure can be considerably smaller for a detectable effect to be observed, provided that the system is correctly calibrated.

#### **1.5.1 Calibration and Referencing of Protein NMR Spectra**

The need for correct calibration and referencing, particularly of chemical shift data, has been a cause for concern as long as protein NMR spectra have been acquired. Since the resonance frequency of each spectrometer is slightly different, and indeed varies slightly over time, it is necessary to utilize independent references so that data obtained at different times and in different places can be compared. Proton chemical shifts were traditionally referenced to the resonance of the trimethylsilane molecule (TMS), which was set to 0 parts-per-million (ppm). This proved to be a problem for samples in water solution, as TMS is insoluble in water. Several water-soluble compounds have been used over the years for referencing spectra, including dioxane (set to 3.75 ppm) and 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), set to 0 ppm. In 1995, these procedures were standardized, and through a consortium of several different NMR groups, a consensus set of guidelines was published<sup>53</sup>. Additionally for proteins and other biological macromolecules there is the problem of so-called “random coil” chemical shifts. Since a protein is a polymer of defined peptide units, with variations in the chemical composition of side chains, the chemical shifts of the resonances of each of the types of nuclei are to a certain extent characteristic. For example, the protons of the CH<sub>3</sub> group of an alanine side chain resonate at a frequency around 1.4 ppm in the absence of other influences such as might occur in a folded protein, whereas the “random coil” chemical shift of the methyl group of a methionine is 2.1 ppm, and of a leucine is 0.9 ppm. Deviations of the chemical shifts of Ala, Met or Leu methyl groups in a protein give a sensitive measure of the presence and extent of a propensity for residual non-random structure in the conformational ensemble. The measurement of random coil

chemical shifts has been an ongoing process, started in the Wüthrich group in the 1970s<sup>54</sup>, measuring chemical shifts observed of amino acids in a set of peptides. Other sets of random coil shifts were prepared from our lab and others<sup>55-57</sup>. More recently, following the initiation of the studies to be described in the next few paragraphs, we have measured a new set of random coil shifts for <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N, in urea solution<sup>58</sup>. In addition, some of the chemical shifts in proteins are highly sequence-dependent<sup>59</sup>, an advantage for assignment of resonances in the spectra of unfolded proteins, but problematic for the evaluation of small chemical shift differences from random coil values. We therefore evaluated the sequence dependence of a number of reference chemical shifts in a set of random coil peptides, with an algorithm (implemented in the commonly-used program NMRView) for correction of random coil chemical shifts for the influence of local sequence<sup>60</sup>. The use of sequence-corrected random coil chemical shifts to evaluate the so-called “secondary chemical shift” (ie, observed chemical shift minus random coil chemical shift) has proven invaluable in the evaluation of local propensities for structure in unfolded proteins, and remains the cornerstone of our experimental strategy for these systems.

### 1.5.2 Equilibrium Intermediates as Models for Stages on the Folding Pathway

Since it appeared that the kinetic and equilibrium intermediates observed for apomyoglobin were very similar<sup>36</sup>, we reasoned that the unfolded and partly folded forms of the protein could be used as models for stages in the kinetic folding pathway. These forms have the advantage for NMR study that they are stable in solution over the (lengthy) period required for acquisition of 3-dimensional spectra for resonance assignment. Four equilibrium states were chosen: the unfolded state formed in acidic 8M urea<sup>61</sup>, the acid-unfolded state at pH 2 in the absence of urea<sup>62</sup>, the equilibrium intermediate state at pH 4<sup>63</sup> and the folded apoprotein at pH 6<sup>64</sup>. A summary of these data appeared in 1998<sup>45</sup> following its presentation at a conference in Europe; this material was highlighted in a report in *Science*<sup>65</sup>. The findings from the apomyoglobin equilibrium studies are summarized in Figure 1.2.

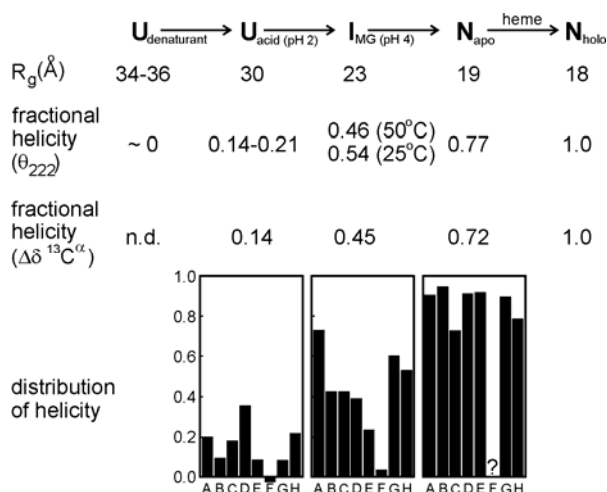


Figure 1.2 Summary of equilibrium data for the various forms of apomyoglobin, representing models for various stages in the folding pathway. R<sub>g</sub> is the radius of gyration, measured by small angle X-ray scattering. θ<sub>222</sub> is the ellipticity at 222 nm in a CD spectrum, Δδ is the secondary chemical shift (observed – random coil), averaged over the entire molecule. The bottom panels show the same data, classified according to individual helical segments. (Figure derived from <sup>45</sup>)

The urea-denatured state contains negligible propensity for helical structure, as measured by the sequence-corrected secondary chemical shifts, although it does contain some intriguing features, which are described in a later paragraph. The acid-denatured state in the absence of urea contains some residual helical structure, a small amount in the A and H helices, but also an equally significant population of non-native helical backbone dihedral angles in the D-E helix linker sequence<sup>62</sup>. Extensive relaxation data were measured for the acid-denatured state, in order to determine which parts of the polypeptide chain might be less flexible as a result of the

formation of residual structure (see Chapter 3 for a fuller explanation of the use of relaxation data to obtain dynamic information). Consistent with the results from the chemical shift analysis, the regions of the sequence where residual helical structure was observed (Figure 1.2) corresponded to regions of slight motional restriction on a ps-ns time scale in the polypeptide backbone. However, the analysis of the relaxation data showed an additional very intriguing result: two regions of the polypeptide were exchanging on a  $\mu$ s-ms time scale, which begins to approach the time scale of the folding process. Even more intriguingly, these two regions formed parts of the A and G helices, which are present in the kinetic and equilibrium intermediate states. Are we observing the first signs of a long-range interaction that presages collapse of the molecule to the folded state? The same evidence for  $\mu$ s-ms time scale motion appears to be present in solutions of varying concentrations, indicating that it probably does not arise from intermolecular interactions<sup>62</sup>. We were able to make a definitive identification of the source of this behavior by utilizing apomyoglobin mutants that had been site-specifically spin-labeled<sup>66</sup>. A spin-label such as a nitroxide group (or a paramagnetic metal ion) contains an unpaired electron, which has the effect of broadening the resonances of nearby protons, with a  $1/r^6$  distance dependence. In a folded protein, the presence of a spin label causes most of the resonances to be broadened, an observation that can be used to obtain structural information<sup>67</sup>. In the conformational ensemble of an unfolded or partly folded protein, the presence of the spin label allows an estimate of the internuclear distance together with the population of a particular conformation. The two quantities cannot be separated without other sources of information. However, the observation of resonance broadening at sites distant in the sequence from the site of the spin label is good evidence that there is at least a threshold population of conformers where the two sites are in close proximity. This is illustrated in Figure 1.3. At pH 2 in the presence of 8 M urea, only the resonances of protons in the immediate sequence vicinity of the spin label are broadened (Figure 1.3A), whereas at pH 2 in the absence of urea there are many more areas of at least partial broadening (Figure 1.3B).

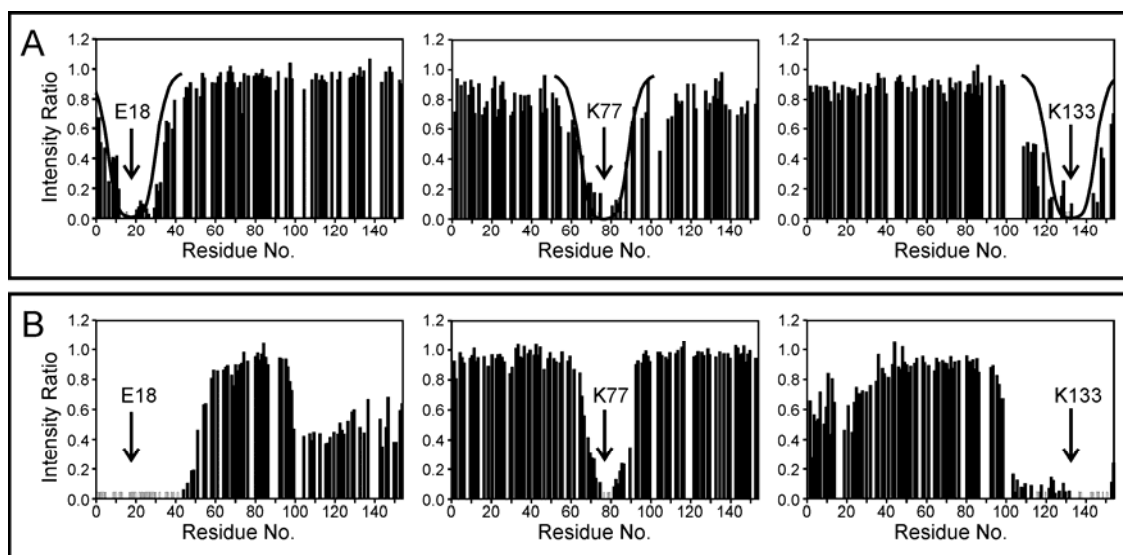


Figure 1.3 Intensity ratio of cross peaks in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of spin-labeled and spin-label-reduced apomyoglobin. A ratio of 1 indicates no broadening at this position by the spin label. A ratio of 0 (and the short unfilled bars) indicates that the cross peak is invisible due to broadening. Vertical arrows in each panel indicate the site of the covalently-attached spin label. Inverted bell-shaped curves are the expected intensity ratios for a completely random chain, calculated according to a Gaussian distribution. Intensity ratios for A. apomyoglobin in the presence of 8 M urea, pH 2. B. apomyoglobin at pH 2 in the absence of urea. (redrawn from <sup>66</sup>)

The most significant aspect of the results shown in Figure 1.3B is that the effect is sequence-specific: a spin label in the A helix will broaden resonances in the G and H helices, while a spin label in the H helix will broaden resonances in the A helix. However, the spin label in the E helix does not show any differences in broadening between the solutions with and without urea, indicating that this region of the sequence does not participate in the interactions that are apparent for the A, G and H helices. We conclude that even in the acid-unfolded state of apomyoglobin, the polypeptide chain samples long-range interactions, and further, that this sampling is not random, but rather that native-like contacts between the ends of the polypeptide are preferentially sampled. If we can identify the characteristics of the amino acids in the A, G and H helices that are promoting these interactions, we will have answered the question as to the nature of the impetus for initiation of protein folding.

### **1.6 Cracking the Sequence Code for Folding Initiation**

The unfolded forms of apomyoglobin, even in 8 M urea at pH 2, show slight sequence-specific differences in behavior<sup>61,62</sup>. In particular, the analysis of the dynamics of the backbone, through <sup>15</sup>N relaxation analysis, shows variation along the sequence that appears to be correlated with an intrinsic property of the individual amino acids. This property, which has been called “average area buried upon folding” (AABUF)<sup>68</sup> or “hydrophobicity”<sup>69</sup> recognizes that amino acid side chains are not chemically homogeneous, but may exhibit different properties in different parts of the side chain. For example, a lysine side chain is traditionally classified as a “positively charged” side chain, since the terminal amine group is usually present as  $-\text{NH}_3^+$  at neutral pH where most proteins are stable. Yet this amine group is only a small part of the lysine side chain, which contains four methylene  $-\text{CH}_2$  groups, which can behave like a hydrophobic side chain, especially if the charged group is neutralized, for example by the formation of a salt bridge in a folded protein. Thus, although a charged side chain would be expected to prefer the surface of a folded protein, where it can interact with solvent water, lysine side chains are frequently found buried in the interior. The AABUF quantity thus takes into account not only the chemical nature of the amino acid side chain, but also its length and bulk, recognizing that bulky side chains, even if they contain + or – charges, can make interactions through their hydrophobic groups<sup>70</sup>. The dynamic behavior of the pH 2 form of apomyoglobin as a function of residue number is correlated with sequence-dependent changes in the AABUF<sup>62</sup>: areas of slight motional restriction occur at the same positions in the sequence as peaks in the AABUF. That is, local motion of the backbone is being restricted because of the presence of clumps of bulky amino acid side chains. The converse effect is seen in the dynamics of the protein in 8 M urea<sup>61</sup>: here the polypeptide chain shows noticeable dips in motion in areas of the sequence where there are clumps of small amino acids such as glycine, alanine and serine. These results suggest that the intrinsic chemical properties of the amino acid side chains, and their locations in clumps along the amino acid sequence are causing recognizable motional restriction of the backbone, the beginnings of the organized collapse that is the folding of the protein. Our hypothesis is that the local (sequence-specific) “hydrophobic” interactions between the areas containing a high proportion of contiguous bulky side chains causes highly specific local chain collapse, leading to the initiation of folding. Put another way, it is the propensities for formation of transient local hydrophobic interactions (NOT the formation of secondary structures, as we had originally thought) that initiate the orderly folding of the protein. Clearly, secondary structure formation will be an important component of the early folding process, since the backbone peptide groups must be organized into hydrogen bonding networks in order that they can be buried in the hydrophobic core of the folding protein<sup>71,72</sup>.

If this hypothesis is correct, then we ought to be able to design changes in the amino acid sequence specifically to alter the local AABUF, say, in the A helix, then measure the folding of the protein to see whether the A helix remains part of the burst phase intermediate. Since mutations may affect all of the states of the protein, unfolded, intermediate and folded, we specifically designed a double mutant that would change the AABUF of the A and E helices while preserving as well as possible the interactions in the final folded structure of the protein<sup>48</sup>. This mutant protein is illustrated in Figure 1.4.

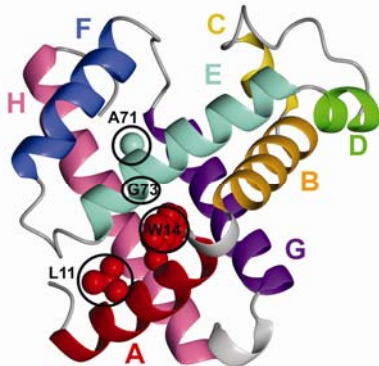


Figure 1.4 X-ray crystal structure of myoglobin, showing the positions of the residues that are changed in the GGLW mutant (L11G, W14G, A71L, G73W). (from<sup>48</sup>).

The tryptophan at position 14 in the sperm whale apomyoglobin sequence was replaced by a glycine, while the residue opposite Trp14 in the 3-dimensional structure, the glycine at position 73, was replaced by a tryptophan (the W14G, G73W or GW mutant). The effect of these two changes on the 3-dimensional structure of the protein ought to be minimal, since the bulky tryptophan side chain can occupy the same position in the mutant and wild type proteins, but the effects on the AABUF of the sequences of the A and E helices is profound. Additional minor mutations enhance these effects, as illustrated in Figure 1.5A: the intrinsic hydrophobicity or AABUF of the A helix sequence is significantly lowered, while that of the E helix is significantly enhanced. The results of quench flow experiments on the wild type and GGLW mutant proteins are also shown in Figure 1.5B (solid points)<sup>48</sup>.

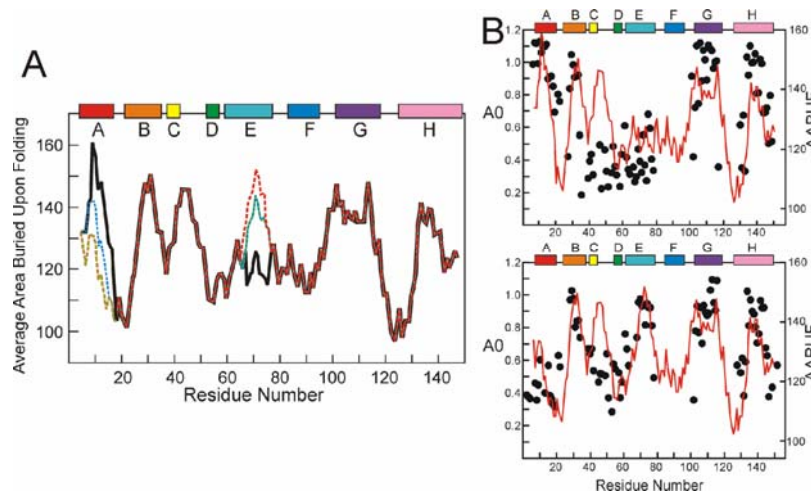


Figure 1.5 A. Plot of AABUF as a function of residue number. Black: wild type protein, blue, GW mutant, green, GGW mutant, red, GGLW mutant. B. Plots of A0 (a measure of proton occupancy in the burst phase intermediate) (black points) and AABUF (same data as part A) (red lines) for wild-type apomyoglobin (top) and for the GGLW mutant protein (bottom).

Clearly the mutations, which were designed to alter the intrinsic hydrophobicity of the sequence, have resulted in a significant change in the composition of the burst phase intermediate, consistent with our hypothesis: where the AABUF is high, either in the wild type or mutant protein, there is greater likelihood of that region forming part of the initially-folded core of the protein.

*These experiments thus identify the primary driving force for the initiation of protein folding: the local interactions between groups of bulky side chains that are close together in the amino acid sequence.*

### **1.7 New Methods, New Experiments, New Insights**

During the course of the experiments described in the previous paragraph, it became clear that the classic quench flow method could not give the information required to answer the questions we were posing. In particular, as the overall stability of the protein was reduced by mutations, the number of data points decreased, due to the requirement for very slowly-exchanging amide protons in the final folded state of the protein. To overcome this difficulty, we devised a method that was identical in all respects to the “classic” experiment, except in the final step, where instead of forming the folded protein and measuring amide protection in the final folded state, we rapidly froze and lyophilized the samples, and redissolved them in an aprotic organic solvent DMSO<sup>73</sup>. The pattern of protected amides produced in the quench flow experiment would be unchanged, as there would be no amide exchange in the aprotic solvent. Folding data would thus be available from a much larger set of amides, for example, the classic quench flow experiment gives data on 50 amides (of the 145 in apomyoglobin), while the DMSO experiment yields data on 95.

### **1.8 Folding Studies of Other Proteins**

Folding experiments on the small 3-helix protein Protein A (B domain)<sup>74,75</sup> were made primarily for comparison with theoretical calculations<sup>76,77</sup> that had implicated different helices in a folding intermediate of Protein A. We were unable to find evidence for either intermediate in our folding experiments, possibly saving us from quarrels with either of our theoretical colleagues.

To broaden the spectrum beyond helical proteins, we undertook folding experiments on a  $\beta$ -sheet protein, plastocyanin<sup>78</sup>. The folding pathway of apoplastocyanin are completely different from those of apomyoglobin: the majority of the molecule folds extremely rapidly, but the final folding stage is extremely slow due to the isomerization of two proline residues (from *trans* to *cis*). The slow folding of apoplastocyanin enabled us to undertake real-time NMR experiments to characterize the folding on a site-specific basis<sup>79</sup>. Apoplastocyanin also populates an unfolded state under non-denaturing conditions (in the presence of low salt), which has been characterized by NMR<sup>80</sup>.

## Reference List: Chapter 1

1. Dyson, H. J., Holmgren, A., & Wright, P. E. (1989). Assignment of the proton NMR spectrum of reduced and oxidized thioredoxin: sequence-specific assignments, secondary structure and global fold. *Biochemistry* **28**, 7074-7087.
2. Neri, D., Wider, G., & Wüthrich, K. (1992). Complete <sup>15</sup>N and <sup>1</sup>H NMR assignments for the amino-terminal domain of the phage 434 repressor in the urea-unfolded form. *Proc. Natl. Acad. Sci. USA* **89**, 4397-4401.
3. Lerner, R. A. (1982). Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature* **299**, 592-596.
4. Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A., & Hendrickson, W. A. (1984). The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature* **312**, 127-133.
5. Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., & Lerner, R. A. (1985). The immunodominant site of a synthetic immunogen has a conformational preference in water for a Type-II reverse turn. *Nature* **318**, 480-483.
6. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988). Folding of immunogenic peptide fragments of proteins in water solution. I Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161-200.
7. Dyson, H. J., Bolinger, L., Feher, V. A., Osterhout, J. J., Jr., Yao, J., & Wright, P. E. (1998). Sequence requirements for stabilization of a peptide reverse turn in water solution - Proline is not essential for stability. *Eur. J. Biochem.* **255**, 462-471.
8. Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., & Dyson, H. J. (1994). Stabilization of a Type VI turn in a family of linear peptides in water solution. *J. Mol. Biol.* **243**, 736-753.
9. Yao, J., Dyson, H. J., & Wright, P. E. (1994). Three-dimensional structure of a Type VI turn in a linear peptide in water solution: evidence for stacking of aromatic rings as a major stabilizing factor. *J. Mol. Biol.* **243**, 754-766.
10. Yao, J., Brüschweiler, R., Dyson, H. J., & Wright, P. E. (1994). Differential side chain hydration in a linear peptide containing a Type VI turn. *J. Am. Chem. Soc.* **116**, 12051-12052.
11. Karplus, M. & Weaver, D. L. (1979). Diffusion-collision model for protein folding. *Biopolymers* **18**, 1421-1437.
12. Chan, H. S. & Dill, K. A. (1990). Origins of structure in globular proteins. *Proc. Natl. Acad. Sci. USA* **87**, 6388-6392.
13. Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E., & Lerner, R. A. (1988). Folding of immunogenic peptide fragments of proteins in water solution. II The nascent helix. *J. Mol. Biol.* **201**, 201-217.
14. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992). Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin. *J. Mol. Biol.* **226**, 795-817.
15. Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. E. (1992). Folding of peptide fragments comprising the complete sequence of proteins: models for the initiation of protein folding. II. Plastocyanin. *J. Mol. Biol.* **226**, 819-835.
16. Wilson, I. A., Haft, D. H., Getzoff, E. D., Tainer, J. A., Lerner, R. A., & Brenner, S. (1985). Identical short peptide sequences in unrelated proteins can have different conformations: a testing ground for theories of immune recognition. *Proc. Natl. Acad. Sci. USA* **82**, 5255-5259.
17. Serrano, L. & Fersht, A. R. (1989). Capping and  $\alpha$ -helix stability. *Nature* **342**, 296-299.
18. Reymond, M. T., Huo, S. Q., Duggan, B., Wright, P. E., & Dyson, H. J. (1997). Contribution of increased length and intact capping sequences to the conformational preference for helix in a 31-residue peptide from the C terminus of myohemerythrin. *Biochemistry* **36**, 5234-5244.
19. Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry* **32**, 6337-6347.

20. Shin, H.-C., Merutka, G., Waltho, J. P., Wright, P. E., & Dyson, H. J. (1993). Peptide models of protein folding initiation sites. 2. The G-H turn region of myoglobin acts as a helix stop signal. *Biochemistry* **32**, 6348-6355.
21. Shin, H.-C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J., & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry* **32**, 6356-6364.
22. Reymond, M. T., Merutka, G., Dyson, H. J., & Wright, P. E. (1997). Folding propensities of peptide fragments of myoglobin. *Protein Sci.* **6**, 706-716.
23. Dyson, H. J., Lerner, R. A., & Wright, P. E. (1988). The physical basis for induction of protein-reactive antipeptide antibodies. *Ann. Rev. Biophys. Biophys. Chem.* **17**, 305-324.
24. Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988). Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* **27**, 7167-7175.
25. Dyson, H. J., Satterthwait, A. C., Lerner, R. A., & Wright, P. E. (1990). Conformational preferences of synthetic peptides derived from the immunodominant site of the circumsporozoite protein of *Plasmodium falciparum* by <sup>1</sup>H NMR. *Biochemistry* **29**, 7828-7837.
26. Oldstone, M. B. A., Tishon, A., Lewicki, H., Dyson, H. J., Feher, V. A., Assa-Munt, N., & Wright, P. E. (1991). Mapping the anatomy of the immunodominant domain of the human immunodeficiency virus gp41 transmembrane protein: peptide conformation analysis using monoclonal antibodies and proton nuclear magnetic resonance spectroscopy. *J. Virol.* **65**, 1727-1734.
27. Chandrasekhar, K., Profy, A. T., & Dyson, H. J. (1991). Solution conformational preferences of immunogenic peptides derived from the principal neutralizing determinant of the HIV-1 envelope glycoprotein gp120. *Biochemistry* **30**, 9187-9194.
28. Dyson, H. J., Norrby, E., Hoey, K., Parks, D. E., Lerner, R. A., & Wright, P. E. (1992). Immunogenic peptides corresponding to the dominant antigenic region Ala597 to Cys619 in the transmembrane protein of simian immunodeficiency virus have a high folding propensity. *Biochemistry* **31**, 1458-1463.
29. Dyson, H. J. & Wright, P. E. (1995). Antigenic peptides. *FASEB J.* **9**, 37-42.
30. Campbell, A. P., Sykes, B. D., Norrby, E., Assa-Munt, N., & Dyson, H. J. (1996). Solution conformation of an immunogenic peptide derived from the principal neutralizing determinant of the HIV2 envelope glycoprotein gp125. *Fold. Design* **1**, 157-165.
31. Lehmann, T. E., Kroon, G., Dyson, H. J., Lorenzo, M. A., Bermudez, H., & Perez, H. (2003). Plasmodium vivax CS peptides display conformational preferences for folded forms in solution. *J. Pept. Res.* **61**, 252-262.
32. Satterthwait, A. C., Chiang, L.-C., Arrhenius, T., Cabezas, E., Zavala, F., Dyson, H. J., & Wright, P. E. (1990). The conformational restriction of synthetic vaccines for malaria. *Bull. WHO* **68**, 17-25.
33. Ghiara, J. B., Ferguson, D. C., Satterthwait, A. C., Dyson, H. J., & Wilson, I. A. (1997). Structure-based design of a constrained peptide mimic of the HIV-1 V3 loop neutralization site. *J. Mol. Biol.* **266**, 31-39.
34. Udgaonkar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694-699.
35. Roder, H., Elöve, G. A., & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* **335**, 700-704.
36. Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-896.
37. Englander, S. W. (1993). In pursuit of protein folding. *Science* **262**, 848-849.
38. Baldwin, R. L. (1995). On-pathway versus off-pathway folding intermediates. *Fold. Des* **1**, R1-R8.
39. Dalvit, C. & Wright, P. E. (1987). Assignment of resonances in the <sup>1</sup>H nuclear magnetic resonance spectrum of the carbon monoxide complex of sperm whale myoglobin by phase-sensitive two-dimensional techniques. *J. Mol. Biol.* **194**, 313-327.
40. Mabbutt, B. C. & Wright, P. E. (1985). Assignment of heme and distal amino acid resonances in the <sup>1</sup>H-NMR spectra of the carbon monoxide and oxygen complexes of sperm whale myoglobin. *Biochim. Biophys. Acta* **832**, 175-185.
41. Wright, P. E., Cooke, R. M., Cross, K. J., Mabbutt, B. C., Messerle, A., & Wellington, J. E. (1985). *NMR studies of the structure and dynamics of monomeric hemoglobins and myoglobins*. In *Magnetic Resonance in Biology and Medicine* (Govil, G., Khetrpal, & Saran, eds), pp. 131-150, Tata-McGraw-Hill.
42. Springer, B. A. & Sligar, S. G. (1987). High-level expression of sperm whale myoglobin in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84**, 8961-8965.



43. Jennings, P. A., Stone, M. J., & Wright, P. E. (1995). Overexpression of myoglobin and assignment of the amide, C $\alpha$  and C $\beta$  resonances. *J. Biomol. NMR* **6**, 271-276.
44. Cavagnero, S., Dyson, H. J., & Wright, P. E. (1999). Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* **285**, 269-282.
45. Eliezer, D., Yao, J., Dyson, H. J., & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of myoglobin and implications for protein folding. *Nature Struct. Biol.* **5**, 148-155.
46. Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J., & Wright, P. E. (2000). Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. *Biochemistry* **39**, 11227-11237.
47. Nishimura, C., Wright, P. E., & Dyson, H. J. (2003). Role of the B helix in early folding events in apomyoglobin: evidence from site-directed mutagenesis for native-like long range interactions. *J. Mol. Biol.* **334**, 293-307.
48. Nishimura, C., Lietzow, M. A., Dyson, H. J., & Wright, P. E. (2005). Sequence determinants of a protein folding pathway. *J. Mol. Biol.* **351**, 383-392.
49. Nishimura, C., Dyson, H. J., & Wright, P. E. (2006). Identification of Native and Non-native Structure in Kinetic Folding Intermediates of Apomyoglobin. *J. Mol. Biol.* **355**, 139-156.
50. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., & Wright, P. E. (1999). Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* **8**, 45-49.
51. Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA* **89**, 8721-8725.
52. Radford, S. E., Dobson, C. M., & Evans, P. A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302-307.
53. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., & Sykes, B. D. (1995). <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135-140.
54. Bundi, A. & Wüthrich, K. (1979). <sup>1</sup>H-NMR parameters of the common amino acid residues measured in aqueous solution of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* **18**, 285-297.
55. Merutka, G., Dyson, H. J., & Wright, P. E. (1995). 'Random coil' <sup>1</sup>H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR* **5**, 14-24.
56. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995). <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N random coil NMR chemical shifts of the common amino acids: I. Investigations of nearest neighbor effects. *J. Biomol. NMR* **5**, 67-81.
57. Braun, D., Wider, G., & Wüthrich, K. (1994). Sequence-corrected <sup>15</sup>N "random coil" chemical shifts. *J. Am. Chem. Soc.* **116**, 8466-8469.
58. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., & Dyson, H. J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil chemical shift data in NMRView. *J. Biomol. NMR* **18**, 43-48.
59. Yao, J., Dyson, H. J., & Wright, P. E. (1997). Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* **419**, 285-289.
60. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., & Dyson, H. J. (2001). Sequence dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**, 2970-2978.
61. Schwarzinger, S., Wright, P. E., & Dyson, H. J. (2002). Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry* **41**, 12681-12686.
62. Yao, J., Chung, J., Eliezer, D., Wright, P. E., & Dyson, H. J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* **40**, 3561-3571.
63. Eliezer, D., Chung, J., Dyson, H. J., & Wright, P. E. (2000). Native and non-native structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* **39**, 2894-2901.
64. Eliezer, D. & Wright, P. E. (1996). Is apomyoglobin a molten globule? Structural characterization by NMR. *J. Mol. Biol.* **263**, 531-538.
65. Balter, M. (1997). NMR maps giant molecules as they fold and flutter. *Science* **278**, 1014-1015.
66. Lietzow, M. A., Jamin, M., Dyson, H. J., & Wright, P. E. (2002). Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **322**, 655-662.

67. Kosen, P. A., Scheek, R. M., Naderi, H., Basus, V. J., Manogaran, S., Schmidt, P. G., Oppenheimer, N. J., & Kuntz, I. D. (1986). Two-dimensional <sup>1</sup>H NMR of three spin-labeled derivatives of bovine pancreatic trypsin inhibitor. *Biochemistry* **25**, 2356-2364.
68. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-838.
69. Matheson, R. R., Jr. & Scheraga, H. A. (1978). A method for predicting nucleation sites for protein folding based on hydrophobic contacts. *Macromolecules* **11**, 819-829.
70. Dyson, H. J., Wright, P. E., & Scheraga, H. A. (2006). The role of hydrophobic interactions in initiation and propagation of protein folding. *PNAS* **103**, 13057-13061.
71. Yang, A. S. & Honig, B. (1995). Free energy determinants of secondary structure formation .2. Antiparallel  $\beta$ -sheets. *J. Mol. Biol.* **252**, 366-376.
72. Yang, A.-S., Sharp, K. A., & Honig, B. (1992). Analysis of the heat capacity dependence of protein folding. *J. Mol. Biol.* **227**, 889-900.
73. Nishimura, C., Dyson, H. J., & Wright, P. E. (2005). Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc. Natl. Acad. Sci. U. S. A* **102**, 4765-4770.
74. Bai, Y., Karimi, A., Dyson, H. J., & Wright, P. E. (1997). Absence of a stable intermediate on the folding pathway of protein A. *Protein Sci.* **6**, 1449-1457.
75. Karimi, A., Matsumura, M., Wright, P. E., & Dyson, H. J. (1999). Characterization of monomeric and dimeric B domain of Staphylococcal protein A. *J Pept. Res.* **54**, 344-352.
76. Skolnick, J., Kolinski, A., Brooks, C. L., Godzik, A., & Rey, A. (1993). A method for predicting protein structure from sequence. *Curr. Biol.* **3**, 414-423.
77. Boczko, E. M. & Brooks, C. L. (1995). First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269**, 393-396.
78. Koide, S., Dyson, H. J., & Wright, P. E. (1993). Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry* **32**, 12299-12310.
79. Mizuguchi, M., Kroon, G. J., Wright, P. E., & Dyson, H. J. (2003). Folding of a  $\beta$ -sheet protein monitored by real-time NMR spectroscopy. *J. Mol. Biol.* **328**, 1161-1171.
80. Bai, Y., Chung, J., Dyson, H. J., & Wright, P. E. (2001). Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions. *Protein Sci.* **10**, 1056-1066.

## **CHAPTER 2. STRUCTURE, DYNAMICS AND FUNCTION OF PROTEINS**

The majority of the work-horse molecules in a cell are proteins. Stable proteins provide scaffolding for internal structures such as microtubules, and membrane-associated proteins ensure the integrity of the cell and regulate the ingress and egress of substrates and waste products. In particular, most enzymes are proteins; an understanding of the way the cell works necessarily requires an understanding of the mechanisms of enzymes, and this has been a fruitful area of study for well over a century. As biophysical techniques have developed, we have been able to study proteins and enzymes on an increasingly more detailed level. We are now beginning to understand some of the factors that enable enzymes to act as catalysts with such efficiency (described in Chapter 3). This is clearly a huge field. In this brief chapter I will introduce some of the protein systems that I have studied, most of which were initiated as part of a collaborative effort. In most cases, I have been the Principal Investigator in the structural studies described below.

### **2.1 Disulfide-dithiol Chemistry in the Cell: Thioredoxin**

The correct functioning of cellular metabolism requires that solution conditions within the cell be kept in homeostasis within certain strict limits in terms of pH, salt concentration and redox potential. Redox homeostasis is mediated by a large number of cellular systems, most of which rely upon the balance between (reduced) thiol groups and (oxidized) disulfides. A large pool of each of these forms is generally present, so that challenges to the cell, such as by the formation of reactive oxygen species (ROS) or under hypoxic conditions, can be met efficiently and rapidly. Two of the best characterized disulfide-dithiol proteins are thioredoxin and glutaredoxin, which were initially described in the laboratory of Arne Holmgren in Sweden<sup>1-3</sup>.

#### **2.1.1 Solution NMR Studies of *Escherichia coli* Thioredoxin**

Our initial interest in thioredoxin was because it presented a structural problem that apparently could not be solved by X-ray crystallography. Thioredoxin from *E. coli* is present in two forms, oxidized (with a disulfide bond between Cys32 and Cys35) and reduced (where the two cysteines contain thiol –SH groups). The two forms differ slightly in function, for example, the reduced form (Trx-(SH)<sub>2</sub>) is capable of acting as a subunit in the DNA polymerase of the bacteriophage T7<sup>4</sup> and is required for the assembly of filamentous bacteriophages<sup>5</sup>; the oxidized form (Trx-S<sub>2</sub>) does not function in either of these systems. These functional differences implied that there was a fundamental difference between Trx-(SH)<sub>2</sub> and TrxS<sub>2</sub>, presumably a structural difference. However, at the time that we began work on thioredoxin, it was impossible to compare the structures of the two forms.

The oxidized form of thioredoxin was crystallized only with difficulty, in the presence of Cu<sup>2+</sup> ions. The structure of Trx-S<sub>2</sub> was solved<sup>6</sup> and refined<sup>7</sup>, but no crystals were available for Trx-(SH)<sub>2</sub>. We undertook a comparison of the two forms of *E. coli* thioredoxin by proton NMR<sup>8,9</sup> and calculated the solution structure of Trx-(SH)<sub>2</sub><sup>10</sup>. This NMR analysis and structure calculation was the largest attempted at that time (thioredoxin has 108 residues, molecular weight 11,200 Da), and remains the largest protein structure reported using only proton data. Virtually all subsequent studies made use of stable isotope labeling, which we also employed<sup>11,12</sup> to obtain highly-refined structures of both Trx-(SH)<sub>2</sub> and Trx-S<sub>2</sub> in solution<sup>13</sup>. Comparison of the two proteins showed that there were only very small structural differences between them (Figure 2.1), yet the functional differences mentioned in the previous paragraph are unequivocal. We concluded that there must be other factors at work governing the ability of Trx-(SH)<sub>2</sub> to operate in the bacteriophage systems.

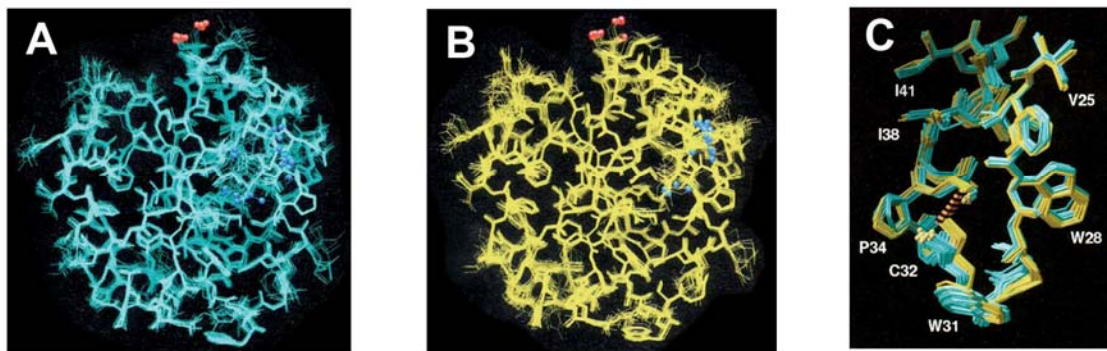


Figure 2.1 A. Family of 20 NMR structures of reduced *E. coli* thioredoxin [Trx-(SH)<sub>2</sub>]. B. Family of 20 NMR structures of oxidized *E. coli* thioredoxin [Trx-S<sub>2</sub>]. Red and blue balls indicate C- and N-termini respectively. C. Superposition of the families in A and B in the vicinity of the active site. The disulfide bond is shown as a hatched red bar.

### 2.1.2 A Dynamic Difference, not a Structural Difference

Backbone and tryptophan side chain dynamics were measured for the two forms of thioredoxin<sup>14</sup>. Analysis of these measurements revealed a localized difference in the flexibility of the backbone in the vicinity of the active-site dithiol/disulfide. In particular we observed slow time scale ( $\mu$ s-ms) motions for residues 73-75, which form part of a loop that contacts the active site sulfur atoms. These results imply that the additional flexibility in the active site region, imparted by the presence of two thiols instead of a rigid disulfide link, allows Trx-(SH)<sub>2</sub> to function in the bacteriophage systems. This is an important insight for our future studies on the role of flexibility in function (see Chapter 3): for thioredoxin, only if the region surrounding the active site was sufficiently flexible could it be functional. If this region is too rigid, as when the disulfide is present, it can no longer perform the same function in the phage systems. A further clue that the answer lies neither in structural differences nor in the chemical difference (perhaps a thiol group is required for the interaction?) comes from mutagenesis studies, where the active site cysteines were replaced with serine<sup>15,16</sup>: all of these mutants were active in the bacteriophage systems, and their NMR spectra resemble those of wild-type Trx(SH)<sub>2</sub><sup>17</sup>.

### 2.1.3 Insights into the Mechanism of Thiol-Disulfide Exchange in Thioredoxin

The conversion of two thiol groups into a disulfide (or the reverse reaction) requires the transfer of two protons and two electrons from some external source. In the cell, one of the normal functions of thioredoxin is to mediate redox homeostasis, which most often requires that the molecule act as a disulfide reductase, converting disulfide bonds in other proteins to thiol groups. Regeneration of the dithiol from the disulfide formed occurs via an enzyme, thioredoxin reductase, which utilizes reducing equivalents from NADPH. Since the oxidation/reduction reaction of thioredoxin involves the transfer of protons as well as electrons, its activity ought to be pH dependent; this idea has received a great deal of attention, some of it controversial, from a number of different research groups.

We initially performed a simple pH titration of the proton NMR spectrum of thioredoxin<sup>18</sup>. This study showed simple two-state pH-dependent behavior for Trx-S<sub>2</sub> in the pH range 5.5-10, but that there are several titrating groups that affect the local environment of the cysteine thiol groups in Trx-(SH)<sub>2</sub> in the same range (Figure 2.2).

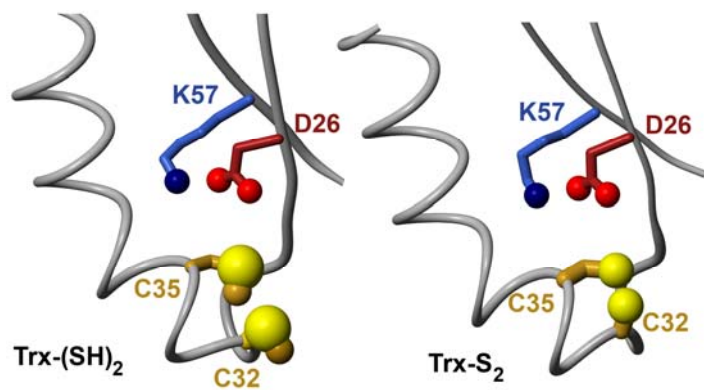


Figure 2.2 Close up view of representative structures of Trx-(SH)<sub>2</sub> and Trx-S<sub>2</sub> showing the proximity of titrating groups.

The buried carboxylate group of Asp26 is most likely responsible for the behavior of Trx-S<sub>2</sub>, while the complex behavior of Trx-(SH)<sub>2</sub> is probably due to the presence not only of the two thiols but of the Asp26 carboxylate as well. A more comprehensive study was later performed using isotopically-labeled thioredoxin<sup>19</sup>. Because this study obtained a complete set of titration data for all nuclei that were affected by pH changes, we were able to dissect the influence of the three possible titrating groups on all sites in the vicinity of the active site. Both Asp26 and Cys32 titrate with a pK<sub>a</sub> of ~7.5, while the Cys35 thiol titrates with pK<sub>a</sub> ~ 9.5. Our conclusion from this study was that the three titrating groups acted in a synergistic manner, with the titration of one group affecting those of the other two. We postulated that the thiol proton that remains at pHs above 7, nominally attached to the Cys35 thiol, is actually shared between the Cys32 and Cys35 sulfur atoms, thus facilitating the transfer of electrons and protons and the formation of mixed disulfide intermediates in the enzymatic reactions of thioredoxin. We later confirmed the pK<sub>a</sub> of Asp26 by direct titration<sup>20</sup>, and showed by examination of mutant proteins that this residue and its hydrogen-bonded, buried partner Lys57 are essential for the correct functioning of thioredoxin<sup>21</sup>. A number of other mutant proteins, with specific changes in the vicinity of the active site, have also been characterized<sup>22,23</sup>, including a covalent complex between a single-cysteine mutant and a cysteine-containing peptide, which provides a model for the mixed-disulfide intermediate postulated in the reductive mechanism of thioredoxin<sup>24</sup>.

These studies and the conclusions we derived from them proved controversial, with a number of groups weighing in with different types of measurements and various theories to account for them<sup>25-27</sup>. Our final study in this series was a theoretical treatment of the electrostatics of the active-site region under differing pH conditions<sup>28</sup>, which showed unequivocally that the pK<sub>a</sub>s of the various groups were likely to be close to the ones that we (but not the other groups) had reported. This study also provided support for our original idea that the two cysteine sulfur atoms share a proton (one hesitates to call it a hydrogen bond!) at intermediate pH values.

## **2.2 Structure of an Unusual Glutaredoxin**

Glutaredoxins are generally smaller than thioredoxins, and are principally involved in interactions with ribonucleotide reductase in the production of deoxyribonucleotides, using the small molecule glutathione as a source of reducing equivalents. The importance of the ribonucleotide reductase system for viability of the cell results in the provision of multiple redundant systems for the supply of reducing equivalents for this enzyme. Multiple thioredoxins and glutaredoxins can substitute for each other in the ribonucleotide reductase system<sup>2</sup>. There are three glutaredoxins in *E. coli*, Grx1, Grx2 and Grx3<sup>29</sup>. Grx1 and Grx3 are typical glutaredoxins, with molecular weights ~ 10,000 Da, and both are capable of reducing ribonucleotide reductase.

By contrast, Grx2, which is highly abundant in the *E. coli* cell, is significantly larger (24,300 Da) and is incapable of reducing ribonucleotide reductase, although highly efficient in general disulfide-dithiol reactions. The solution structure of *E. coli* Grx2 (Figure 2.3) shows a complex helical molecule that appears to contain as a subsidiary domain the structure of a classic glutaredoxin<sup>30</sup>. The general similarity of the Grx2 fold to a number of glutathione-S-transferase structures provides a clue as to the major function of Grx2 in the *E. coli* cell: it is likely to act as a detoxifying enzyme. This would be consistent with its observed role in electron donation to arsenate reductase<sup>31</sup>.

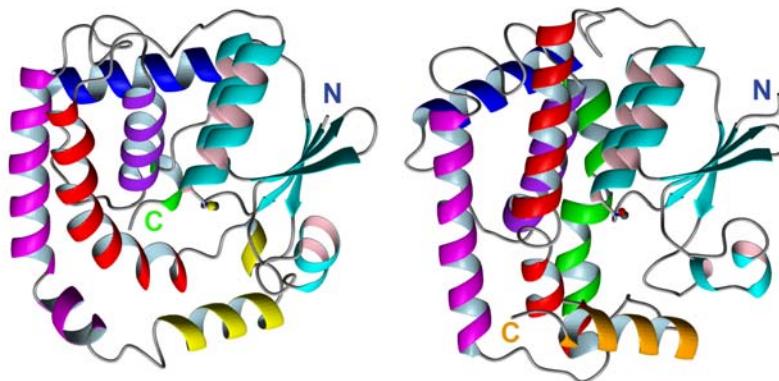


Figure 2.3 Solution structure of reduced Grx2 (left), and human  $\theta$  class GST (figure adapted from ref. <sup>30</sup>).

### **2.3 Anatomy of an Extremophile Redox Protein**

We are accustomed to thinking about protein structure and function under “normal” conditions: temperatures of  $\sim 37^\circ$  or lower, reasonably low salt concentrations and near-neutral pH. Yet many organisms live and thrive under conditions that are wildly different. One of the intriguing questions about such organisms is how their proteins and other macromolecules are constituted so as to be stable under extreme conditions. One study that addressed this was a structural and dynamic comparison of two copper proteins, one from green plants and the other from an acidophilic bacterium *Thiobacillus ferrooxidans*, a principal component of acid mine drainage.

The blue copper protein plastocyanin has been a staple system in the lab for many years. Plastocyanin is a component of the photosynthetic pathway in green plants, and consists of a well-defined  $\beta$ -sandwich containing two  $\beta$ -sheets<sup>32</sup>, with a tetrahedral binding site for a single copper ion (Cu(I) or Cu(II)) that is defined by the polypeptide chain<sup>33</sup>. That is, unlike a number of other metal-containing proteins such as zinc fingers, the metal ion is not necessary for the protein to be folded. Plastocyanin can be crystallized at a number of pH values, but at low pH, the copper ion is trigonally coordinated, resulting in a preference for the Cu(I) state that renders plastocyanin redox-inactive at low pH<sup>34</sup>. By contrast, the blue copper protein rusticyanin, from the acidophilic *T. ferrooxidans* is maximally active at pH  $\sim 2$ , the normal pH of growth of the bacterium, which obtains reducing equivalents from the oxidation of ferrous ores in mine tailings, principally ferrous sulfides, producing the characteristic copious deposits of rust-like hydrated ferric oxides and an equivalent amount of sulfuric acid, which serves to keep the local environment acidic to the point that atmospheric oxidation of the ferrous compounds is disfavored. We undertook a structure determination of Cu(I) rusticyanin by NMR, in order to determine what characteristics of the protein allowed it to operate in the Cu(I)/Cu(II) redox system at pH 2 [studies of Cu(II) proteins are difficult by NMR, due to relaxation by the unpaired electron]. Resonance assignments for rusticyanin were completed using <sup>15</sup>N-labeled

protein alone<sup>35</sup>, a significant feat for a protein of this size (16,500 Da). The protein was produced by expression in *T. ferrooxidans* by our collaborator Robert Blake. Although the <sup>15</sup>N-labeled protein could be produced with ease, the bacteria simply would not grow on <sup>13</sup>C media, an interesting insight into a possible fundamental difference in the metabolism of these bacteria from those encountered more commonly. In order to produce <sup>13</sup>C-labeled protein, we developed a simple new technique for synthesis of a gene coding for rusticyanin that was optimized for expression in *E. coli*<sup>36,37</sup>. This method has been successfully used in a number of other systems in the lab<sup>38,39</sup>. The solution structure of Cu(I) rusticyanin<sup>40</sup> [an X-ray crystal structure of Cu(II) rusticyanin<sup>41</sup> was published simultaneously] revealed a tightly-packed  $\beta$ -sandwich of two  $\beta$ -sheets, with a short helix (Figure 2.4). The copper site is a distorted tetrahedron of four standard type-I ligands (1 cysteine, 1 methionine, 2 histidine side chains) and resembles that of the neutral-pH form of plastocyanin. The copper site is placed deep inside a hydrophobic region, which probably accounts for the stability of the tetrahedral conformation at pH 2, where plastocyanin becomes 3-coordinate. This paper included a theoretical electrostatic analysis of the structural differences between plastocyanin and rusticyanin, concluding that the disposition of charged side chains and peptide dipoles probably accounted for much of the observed difference in redox potential between the two proteins, as well as promoting the acid-stability of the protein by inhibition of protonation of one of the copper ligand histidines. A follow-up study of the electronic structure of the rusticyanin copper site was also published<sup>42</sup>.

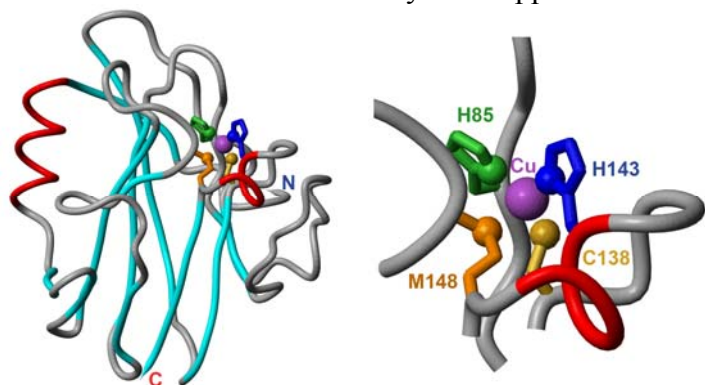


Figure 2.4 Representative structure of *T. ferrooxidans* rusticyanin (left), showing helices red and B-strands blue. An enlarged view of the copper site is shown at right.

## **2.4 Structural and Dynamic Studies on the Prion Protein**

Because we are interested in the intersection between structure, dynamics, folding and function of proteins, we became involved in structural studies on the prion protein, which is apparently the causative agent for a number of neurodegenerative diseases, including scrapie of sheep, mad cow disease (bovine spongiform encephalopathy) and numerous mostly rare inherited human diseases, including Creutzfeldt-Jakob disease. Prion diseases appear to be caused by a misfolded form of a regular cellular protein, the cellular prion protein (PrP<sup>C</sup>), which forms amyloid-like aggregates in the brains of infected animals and humans<sup>43</sup>. The aggregates contain protein that is identical to PrP<sup>C</sup>, but is folded differently forming the aggregate. This form of the protein is termed the scrapie form and is denoted PrP<sup>S</sup>. We defined and characterized the prion protein from Syrian hamster in the “full-length” form (residues 29-231) and in the minimally infective form (residues 90-231)<sup>44</sup>, both of which have a C-terminal folded domain and a long N-terminal sequence that is unstructured in solution<sup>45</sup>. The normal function of the prion protein is unknown, but it appears to have a role in extracellular copper metabolism, as it binds specifically and tightly to Cu(II). We investigated the Cu(II) binding of a portion of the N-terminal domain of PrP, a region termed the “octapept” region for its repetition of the sequence PHGGGWGQ<sup>46</sup>.

Finally, we have calculated a solution structure for a molecule that appears to be able to substitute for the prion protein in some mammalian systems. Termed the Doppel protein, it shows some structural similarities to the prion protein (Figure 2.5).

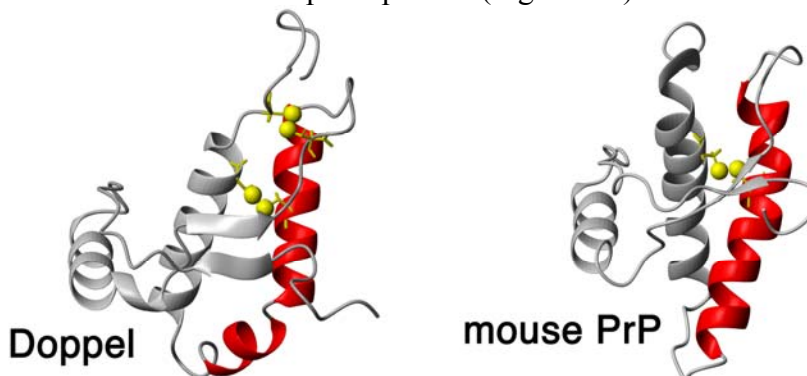


Figure 2.5 A comparison of representative structures from the solution structure families of mouse Doppel protein<sup>47</sup> and mouse prion protein<sup>48</sup>.

### **2.5 Four Proteins – Four Questions**

The following paragraphs describe structural studies of rather more limited scope, which were initiated to answer questions posed by collaborators. These studies have frequently provided interesting sidelights to our main research focus.

#### **2.5.1 How do Adhesion Molecules Work?**

Our collaborator, Bruce Cunningham, was puzzled by the very low affinity of the neural cell adhesion molecule (NCAM) for other adhesion molecules. We initiated a project to use NMR measurements to detect small changes in the local structure of a purified immunoglobulin domain of the protein upon treatment with other domains from the NCAM molecule<sup>49</sup>. These studies demonstrated that the first two immunoglobulin domains (of the five present in NCAM) participated in a weak though specific interaction that was consistent with the multiple weak interactions involved in cell-cell adhesion. A later study investigated the hypothesis that the third immunoglobulin domain self-associates in the homophilic binding interaction<sup>50</sup>. The major conclusion from these studies is that the interactions observed in solution may not reflect those that occur on the cell surface.

#### **2.5.2 How does an Anticoagulant Work?**

Our collaborators in the Edgington lab were unable to determine a crystal structure of an anticoagulant protein, nematode anticoagulant protein c2 (NAPc2)<sup>51</sup>. We were able to show that this small protein is inherently flexible (consistent with its inability to crystallize) – this flexibility likely has a functional application in the wide substrate specificity of this protein.

#### **2.5.3 What is the Structure of the Inserted Domain of the LFA1?**

Our colleagues at the pharmaceutical company Novartis were interested in the inserted (I) domain of the leukocyte function associated antigen-1 (LFA1) as a possible drug target. We agreed to provide resonance assignments<sup>52</sup> and undertook a structure determination<sup>53</sup> that provided the major impetus for the development of the program SANE<sup>54</sup>.

#### **2.5.4 A 44-Residue Protein Won't be a Problem, Will It?**

We were asked by our collaborator David Loskutoff to calculate a structure of a small domain from the blood protein vitronectin. At only 44 residues, the recombinant somatomedin-B domain appeared to be a relatively easy short project. How wrong we were. This protein contains



8 cysteines and 4 disulfides, which define the structure<sup>55</sup>. There is very little interaction other than the disulfides that keeps the molecule together, with the result that data for defining the three-dimensional structure are extremely sparse. The results of the structure calculation remained ambiguous in one critical point, despite our best efforts: the Loskutoff group maintained that their mass spectrometry measurements defined a particular pattern of disulfides in the molecule<sup>56</sup>, while other studies, including an X-ray crystal structure that was published while our study was under way<sup>57</sup>, suggested different disulfide patterns. Exhaustive NMR experiments and energy calculations were unable to distinguish definitively between the possible patterns, although there was a slight preference for the pattern seen in the X-ray structure. Interestingly, any one of three disulfide bonding patterns were consistent with the NMR data, and each of these three patterns delineated a hydrophobic surface of the domain that included residues that were implicated from mutagenesis studies in the binding of SMB to its physiological partner, the serpin PAI-1<sup>55</sup>. To add to the confusion, a third group published a structure of the SMB domain derived from plasma, with a completely different structure<sup>58</sup>, suggesting that the Loskutoff recombinant material used in our structure determination was incorrectly folded<sup>59</sup> and necessitating a whole new set of experiments to disprove it<sup>60</sup>. It has been rightly said that size is no predictor of difficulty.

### Reference List: Chapter 2

1. Holmgren, A. (1985). Thioredoxin. *Ann. Rev. Biochem.* **54**, 237-271.
2. Holmgren, A. & Åslund, F. (1995). Glutaredoxin. *Methods Enzymol.* **252**, 283-292.
3. Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. *J. Biol. Chem.* **264**, 13963-13966.
4. Adler, S. & Modrich, P. (1983). T7-induced DNA polymerase. Requirement for thioredoxin sulfhydryl groups. *J. Biol. Chem.* **258**, 6956-6962.
5. Russel, M. & Model, P. (1985). Thioredoxin is required for filamentous phage assembly. *Proc. Natl. Acad. Sci. USA* **82**, 29-33.
6. Holmgren, A., Söderberg, B.-O., Eklund, H., & Brändén, C.-I. (1975). Three-dimensional structure of *Escherichia coli* thioredoxin-S<sub>2</sub> to 2.8 Å resolution. *Proc. Natl. Acad. Sci. USA* **72**, 2305-2309.
7. Katti, S. K., LeMaster, D. M., & Eklund, H. (1990). Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J. Mol. Biol.* **212**, 167-184.
8. Dyson, H. J., Holmgren, A., & Wright, P. E. (1988). Structural differences between oxidized and reduced thioredoxin monitored by two-dimensional <sup>1</sup>H NMR spectroscopy. *FEBS Lett.* **228**, 254-258.
9. Dyson, H. J., Holmgren, A., & Wright, P. E. (1989). Assignment of the proton NMR spectrum of reduced and oxidized thioredoxin: sequence-specific assignments, secondary structure and global fold. *Biochemistry* **28**, 7074-7087.
10. Dyson, H. J., Gippert, G. P., Case, D. A., Holmgren, A., & Wright, P. E. (1990). Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**, 4129-4136.
11. Chandrasekhar, K., Krause, G., Holmgren, A., & Dyson, H. J. (1991). Assignment of the <sup>15</sup>N NMR spectra of reduced and oxidized *Escherichia coli* thioredoxin. *FEBS Lett.* **284**, 178-183.
12. Chandrasekhar, K., Campbell, A. P., Jeng, M.-F., Holmgren, A., & Dyson, H. J. (1994). Effect of disulfide bridge formation on the NMR spectrum of a protein: studies on oxidized and reduced *Escherichia coli* thioredoxin. *J. Biomol. NMR* **4**, 411-432.
13. Jeng, M.-F., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E., & Dyson, H. J. (1994). High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. *Structure* **2**, 853-868.

14. Stone, M. J., Chandrasekhar, K., Holmgren, A., Wright, P. E., & Dyson, H. J. (1993). Comparison of backbone and tryptophan side-chain dynamics of reduced and oxidized *Escherichia coli* thioredoxin using  $^{15}\text{N}$  NMR relaxation measurements. *Biochemistry* **32**, 426-435.
15. Huber, H. E., Russel, M., Model, P., & Richardson, C. C. (1986). Interaction of mutant thioredoxins of *Escherichia coli* with the gene 5 protein of phage T7. The redox capacity of thioredoxin is not required for stimulation of DNA polymerase activity. *J. Biol. Chem.* **261**, 15006-15012.
16. Russel, M. & Model, P. (1986). The role of thioredoxin in filamentous phage assembly. Construction, isolation and characterization of mutant thioredoxins. *J. Biol. Chem.* **261**, 14997-15005.
17. Dyson, H. J., Jeng, M.-F., Model, P., & Holmgren, A. (1994). Characterization by  $^1\text{H}$  NMR of a C32S,C35S double mutant of *Escherichia coli* thioredoxin confirms its resemblance to the reduced wild-type protein. *FEBS Lett.* **339**, 11-17.
18. Dyson, H. J., Tennant, L. L., & Holmgren, A. (1991). Proton-transfer effects in the active-site region of *Escherichia coli* thioredoxin using two-dimensional  $^1\text{H}$  NMR. *Biochemistry* **30**, 4262-4268.
19. Jeng, M.-F., Holmgren, A., & Dyson, H. J. (1995). Proton sharing between cysteine thiols in *Escherichia coli* thioredoxin: Implications for the mechanism of protein disulfide reduction. *Biochemistry* **34**, 10101-10105.
20. Jeng, M. F. & Dyson, H. J. (1996). Direct measurement of the aspartic acid 26  $\text{pK}_a$  for reduced *Escherichia coli* thioredoxin by  $^{13}\text{C}$  NMR. *Biochemistry* **35**, 1-6.
21. Dyson, H. J., Jeng, M. F., Tennant, L. L., Slaby, I., Lindell, M., Cui, D. S., Kuprin, S., & Holmgren, A. (1997). Effects of buried charged groups on cysteine thiol ionization and reactivity in *Escherichia coli* thioredoxin: Structural and functional characterization of mutants of Asp 26 and Lys 57. *Biochemistry* **36**, 2622-2636.
22. Jeng, M.-F., Reymond, M.T., Tennant, L.L., Holmgren, A. and Dyson, H.J. NMR characterization of a single-cysteine mutant of *Escherichia coli* thioredoxin and a covalent thioredoxin-peptide complex. (1998). *Eur. J. Biochem.*
23. Slaby, I., Cerna, V., Jeng, M.-F., Dyson, H. J., & Holmgren, A. (1996). Replacement of Trp-28 in *Escherichia coli* thioredoxin by site-directed mutagenesis affects thermodynamic stability but not function. *J. Biol. Chem.* **271**, 3091-3096.
24. Jeng, M. F., Reymond, M. T., Tennant, L. L., Holmgren, A., & Dyson, H. J. (1998). NMR characterization of a single-cysteine mutant of *Escherichia coli* thioredoxin and a covalent thioredoxin-peptide complex. *Eur. J. Biochem.* **257**, 299-308.
25. Li, H., Hanson, C., Fuchs, J. A., Woodward, C., & Thomas, G. J., Jr. (1993). Determination of the  $\text{pK}_a$  values of active-center cysteines, cysteines-32 and -35, in *Escherichia coli* thioredoxin by Raman spectroscopy. *Biochemistry* **32**, 5800-5808.
26. Wilson, N. A., Barbar, E., Fuchs, J. A., & Woodward, C. (1995). Aspartic acid 26 in reduced *Escherichia coli* thioredoxin has a  $\text{pK}_a > 9$ . *Biochemistry* **34**, 8931-8939.
27. Chivers, P. T., Prehoda, K. E., Volkman, B. F., Kim, B. M., Markley, J. L., & Raines, R. T. (1997). Microscopic  $\text{pK}_a$  values of *Escherichia coli* thioredoxin. *Biochemistry* **36**, 14985-14991.
28. Dillet, V., Dyson, H. J., & Bashford, D. (1998). Calculations of electrostatic interactions and  $\text{pK}_a$ s in the active site of *Escherichia coli* thioredoxin. *Biochemistry* **37**, 10298-10306.
29. Åslund, F., Ehn, B., Miranda-Vizuete, A., Pueyo, C., & Holmgren, A. (1994). Two additional glutaredoxins exist in *Escherichia coli*: Glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin 1 double mutant. *Proc. Natl. Acad. Sci. USA* **91**, 9813-9817.
30. Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, P. E., & Dyson, H. J. (2001). Solution structure of *Escherichia coli* glutaredoxin 2 shows similarity to mammalian glutathione-S-transferases. *J. Mol. Biol.* **310**, 907-918.
31. Shi, J., Vlamis-Gardikas, A., Åslund, F., Holmgren, A., & Rosen, B. P. (1999). Reactivity of glutaredoxins 1, 2, and 3 from *Escherichia coli* shows that glutaredoxin 2 is the primary hydrogen donor to ArsC-catalyzed arsenate reduction. *J. Biol. Chem.* **274**, 36039-36042.
32. Guss, J. M. & Freeman, H. C. (1983). Structure of oxidized poplar plastocyanin at 1.6 Å resolution. *J. Mol. Biol.* **169**, 521-563.
33. Garrett, T. P. J., Clingeffer, D. J., Guss, J. M., Rogers, S. J., & Freeman, H. C. (1984). The crystal structure of poplar apo-plastocyanin at 1.8 Å resolution: The geometry of the copper-binding site is created by the polypeptide. *J. Biol. Chem.* **259**, 2822-2825.
34. Guss, J. M., Harrowell, P. R., Murata, M., Norris, V. A., & Freeman, H. C. (1986). Crystal structure analyses of reduced ( $\text{Cu}^{\text{I}}$ ) poplar plastocyanin at six pH values. *J. Mol. Biol.* **192**, 361-387.

35. Hunt, A. H., Toy-Palmer, A., Assa-Munt, N., Cavanagh, J., Blake, R. C., II, & Dyson, H. J. (1994). Nuclear magnetic resonance  $^{15}\text{N}$  and  $^1\text{H}$  resonance assignments and global fold of rusticyanin: Insights into the ligation and acid stability of the blue copper site. *J. Mol. Biol.* **244**, 370-384.
36. Casimiro, D. R., Toy-Palmer, A., Blake, R. C., II, & Dyson, H. J. (1995). Gene synthesis, high-level expression and mutagenesis of *Thiobacillus ferrooxidans* rusticyanin. His85 is a ligand to the blue copper center. *Biochemistry* **34**, 6640-6648.
37. Toy-Palmer, A., Prytulla, S., & Dyson, H. J. (1995). Complete  $^{13}\text{C}$  assignments for recombinant Cu(I) rusticyanin: Prediction of secondary structure from patterns of chemical shifts. *FEBS Lett.* **365**, 35-41.
38. Casimiro, D. R., Wright, P. E., & Dyson, H. J. (1997). PCR-based gene synthesis and protein NMR spectroscopy. *Structure* **5**, 1407-1412.
39. Prytulla, S., Dyson, H. J., & Wright, P. E. (1996). Gene synthesis, high-level expression and assignment of backbone  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances of soybean leghemoglobin. *FEBS Lett.* **399**, 283-289.
40. Botuyan, M. V., Toy-Palmer, A., Chung, J., Blake, R. C., II, Beroza, P., Case, D. A., & Dyson, H. J. (1996). NMR solution structure of Cu(I) rusticyanin from *Thiobacillus ferrooxidans*: Structural basis for the extreme acid stability and redox potential. *J. Mol. Biol.* **263**, 752-767.
41. Walter, R. L., Ealick, S. E., Friedman, A. M., Blake, R. C., II, Proctor, P., & Shoham, M. (1996). Crystal structure of Cu(II) rusticyanin: a cupredoxin with extreme redox potential and acid stability. *J. Mol. Biol.*
42. Bender, C. J., Casimiro, D. R., & Dyson, H. J. (1997). Electron spin echo modulation study of the Type I copper protein rusticyanin and its mutant variant His85Ala. *J. Chem. Soc. (Faraday)* **93**, 3967-3980.
43. Prusiner, S. B. & DeArmond, S. J. (1994). Prion diseases and neurodegeneration. [Review]. *Annu. Rev. Neurosci.* **17**, 311-339.
44. Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., Prusiner, S. B., Wright, P. E., & Dyson, H. J. (1997). Structure of the recombinant full-length hamster prion protein PrP(29-231): The N terminus is highly flexible. *Proc. Natl. Acad. Sci. USA* **94**, 13452-13457.
45. Viles, J. H., Donne, D. G., Kroon, G. J. A., Prusiner, S. B., Cohen, F. E., Dyson, H. J., & Wright, P. E. (2001). Local structural plasticity of the prion protein. Analysis of NMR relaxation dynamics. *Biochemistry* **40**, 2743-2753.
46. Viles, J. H., Cohen, F. E., Prusiner, S. B., Goodin, D. B., Wright, P. E., & Dyson, H. J. (1999). Copper binding to the prion protein: Structural implications of four identical cooperative binding sites. *Proc. Natl. Acad. Sci. USA* **96**, 2042-2047.
47. Mo, H., Moore, R. C., Cohen, F. E., Westaway, D., Prusiner, S. B., Wright, P. E., & Dyson, H. J. (2001). Two different neurodegenerative diseases caused by proteins with similar structures. *Proc. Natl. Acad. Sci. USA* **98**, 2352-2357.
48. Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R., & Wüthrich, K. (1996). NMR structure of the mouse prion protein domain PrP(121-231). *Nature* **382**, 180-182.
49. Atkins, A. R., Osborne, M. J., Lashuel, H. A., Edelman, G. M., Wright, P. E., Cunningham, B. A., & Dyson, H. J. (1999). Association between the first two immunoglobulin-like domains of the neural cell adhesion molecule N-CAM. *FEBS Letters* **451**, 162-168.
50. Atkins, A. R., Chung, J., Deechongkit, S. P., Little, E. B., Edelman, G. M., Wright, P. E., Cunningham, B. A., & Dyson, H. J. (2001). Solution structure of the third immunoglobulin domain of the neural cell adhesion molecule N-CAM: can solution studies define the mechanism of homophilic binding? *J. Mol. Biol.* **311**, 161-172.
51. Duggan, B. M., Dyson, H. J., & Wright, P. E. (1999). Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. *Eur. J. Biochem.* **265**, 539-548.
52. Kriwacki, R. W., Legge, G. B., Hommel, U., Ramage, P., Chung, J., Tennant, L. L., Wright, P. E., & Dyson, H. J. (2000). Assignment of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  resonances of the I-domain of human leukocyte function associated antigen-1. *J. Biomol. NMR* **16**, 271-272.
53. Legge, G. B., Kriwacki, R. W., Chung, J., Hommel, U., Ramage, P., Case, D. A., Dyson, H. J., & Wright, P. E. (1999). NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. *J. Mol. Biol.* **295**, 1251-1264.
54. Duggan, B. M., Legge, G. B., Dyson, H. J., & Wright, P. E. (2001). SANE (Structure assisted NOE evaluation): an automated model-based approach for NOE assignment. *J. Biomol. NMR* **19**, 321-329.

55. Kamikubo, Y., De Guzman, R. N., Kroon, G., Curriden, S. A., Neels, J. G., Churchill, M. J., Dawson, P. E., Oldziej, S., Jagielska, A., Scheraga, H. A., Loskutoff, D. J., & Dyson, H. J. (2004). Disulfide bonding arrangements in active forms of the somatomedin B domain of human vitronectin. *Biochemistry* **43**, 6519-6534.
56. Kamikubo, Y., Okumura, Y., & Loskutoff, D. J. (2002). Identification of the disulfide bonds in the recombinant somatomedin B domain of human vitronectin. *J. Biol. Chem.* **277**, 27109-27119.
57. Zhou, A., Huntington, J. A., Pannu, N. S., Carrell, R. W., & Read, R. J. (2003). How vitronectin binds PAI-1 to modulate fibrinolysis and cell migration. *Nat. Struct. Biol.* **10**, 541-544.
58. Mayasundari, A., Whittemore, N. A., Serpersu, E. H., & Peterson, C. B. (2004). The solution structure of the N-terminal domain of human vitronectin: proximal sites that regulate fibrinolysis and cell migration. *J Biol Chem.* **279**, 29359-29366.
59. Horn, N. A., Hurst, G. B., Mayasundari, A., Whittemore, N. A., Serpersu, E. H., & Peterson, C. B. (2004). Assignment of the four disulfides in the N-terminal somatomedin B domain of native vitronectin isolated from human plasma. *J Biol Chem.* **279**, 35867-35878.
60. Kamikubo, Y., Kroon, G., Curriden, S. A., Dyson, H. J., & Loskutoff, D. J. (2006). The reduced, denatured somatomedin B domain of vitronectin refolds into a stable, biologically active molecule. *Biochemistry* **45**, 3297-3306.

## **CHAPTER 3. DYNAMICS AND CATALYSIS**

Most studies of enzymes focus on the immediate vicinity of the active site, and all but ignore the remainder of the protein molecule. Yet many of the amino acids distant from the active sites of enzymes are highly conserved, and do not appear to be always involved with the stabilization of the three-dimensional structure of the enzyme. The working hypothesis of our work on catalysis is that the characteristics of the entire protein, not just the local active-site environment, play an important role.

Dynamic processes play a role in enzyme catalysis<sup>1-3</sup>. Protein motions are implicated in events such as binding of substrate or cofactor, product release, or allosteric regulation. These processes can potentially be mediated by conformational fluctuations of active site loops, by hinge-bending motions, or by reorientation of protein domains or entire protein subunits. In addition, the catalyzed reaction itself involves an inherently dynamic process, with changes in atomic coordinates required along the reaction coordinate<sup>4</sup>. Our exploration of the role of motion of the polypeptide chain in catalysis has been focused on three systems: catalytic antibodies, a metallo- $\beta$ -lactamase and the *E. coli* enzyme dihydrofolate reductase (DHFR).

### **3.1 Crude Man-Made Enzymes: Catalytic Antibodies**

Our early connection with the Lerner group ensured our familiarity with anti-peptide antibodies (see Chapter 1), but direct NMR studies of antibodies and antibody domains were not really possible until quite recently, with the development of labeling techniques and the application of triple-resonance and TROSY experiments, together with high-field spectrometers. We had early on devised specific-labeling methods for use with antibodies<sup>5</sup>, but these had proved impractical except with the best-behaved antibody domains. The Fab domains of antibodies, which were the systems of choice for crystallography, were too large for NMR studies, while the Fv fragments, at approximately half the size, were amenable but frequently difficult to study in solution due to aggregation problems. In our antibody studies, we have found a significant variability in the solution behavior of Fv fragments, for no apparent reason. For example, the Fv that we have reported most results for, the esterase/amidase NPN43C9<sup>6</sup> is well-behaved in solution, whereas a series of aldolase antibodies<sup>7</sup> proved recalcitrant, both to soluble expression in *E. coli* and refolding of inclusion body protein, despite many years of effort.

Initial studies on the catalytic antibody NPN43C9 involved elucidation of mechanistic details, primarily using mass spectrometry techniques<sup>8,9</sup>. This antibody was also characterized by NMR<sup>10</sup> and the influence of substrate (hapten) binding on the backbone dynamics of the polypeptide chain measured<sup>11</sup>. Catalytic antibodies remain of great interest, as a model for early non-optimized biological catalysts: most enzymes, for example, are very highly evolved, precision machines. Catalytic antibodies are by contrast extremely crude, and may give important insights into basic catalytic mechanisms, not least because site-directed mutagenesis has a chance, if correctly designed, of improving catalysis. Improvements are much less likely for normal enzymes. This work is being continued in a collaboration with Dr. Floyd Romesberg.

### **3.2 A Slightly-Evolved Enzyme, the Metallo- $\beta$ -Lactamase from *Bacteroides fragilis***

Bacterial antibiotic resistance is becoming increasingly important in the treatment of many infectious diseases. Resistance usually occurs as a result of the acquisition of genes for enzymes that inactivate the antibiotic or block its action. For example, penicillin-related antibiotics, which act on bacterial polysaccharide cell walls, contain a  $\beta$ -lactam moiety that is essential for activity. One of the major routes to antibiotic resistance is for an enzyme to cleave

the lactam ring, inactivating the antibiotic. One of the approaches to treatment for antibiotic-resistant infections is to treat with a cocktail of an antibiotic plus an inhibitor of the lactamase enzyme, for example in the drug Augmentin. This drug contains clavulanic acid, a potent inhibitor of class A lactamase enzymes. However, there are many other classes of lactamases, most of which are insensitive to clavulanic acid. One of these is the metallo- $\beta$ -lactamase from the opportunistic pathogen *Bacteroides fragilis*. This enzyme is particularly worrying, as it appears that it has evolved into a lactamase encoded from the bacterial chromosome, but has since been observed incorporated into plasmids, which can then be spread to other pathogens.

The study of the metallo- $\beta$ -lactamase has obvious pharmacological utility for the design of new agents to combat this form of antibiotic resistance, but it is also of fundamental interest in the elucidation of catalytic mechanism. The metallo- $\beta$ -lactamase is most probably adapted on a short evolutionary time scale to allow bacteria with this capability to survive the strong selection pressure of the presence of the antibiotic. It may therefore be possible to design mutations that will enhance catalysis, unlike the DHFR system (see following section) where the enzyme has been highly evolved over eons, and any mutation that we design is likely to be deleterious.

Under the auspices of a multi-project Program Project grant from the National Institutes of Health (Dyson, P.I.), we undertook a structural and dynamic analysis of two enzymes. The individual projects consisted of: NMR structural and dynamic analysis of *E. coli* DHFR (Wright), NMR structural and dynamic analysis of the *B. fragilis* metallo- $\beta$ -lactamase (Dyson), mechanistic enzymology of the metallo- $\beta$ -lactamase (Benkovic) and computational analysis of DHFR catalysis (Brooks) and lactamase catalysis (Case). This focused consortium has been extremely productive, publishing landmark papers during the course of the grant<sup>4,12-39</sup>.

The metallo- $\beta$ -lactamase enzyme was thoroughly characterized by NMR<sup>19-23</sup>. This enzyme has an extremely broad range of substrate specificity, which renders it extremely dangerous in the medical sphere. We were able to elucidate the structural basis for the broad specificity: the active site of the enzyme contains a binuclear zinc catalytic site, which is contacted by an unusual  $\beta$ -hairpin flap structure, which is extremely flexible in the free state of the enzyme, but is bound stably to the enzyme when inhibitor (and presumably substrate) is present, sequestering the active site from solvent and allowing the hydrolysis reaction to occur with great efficiency<sup>20</sup> (Figure 3.1). The dynamics of the enzyme are thus extremely important to its mechanism, but the nature of the interactions and motions are extremely simple, and its effects appear to be understandable. We therefore decided that further work on the lactamase system was probably unwarranted, and went on to study other enzyme systems.

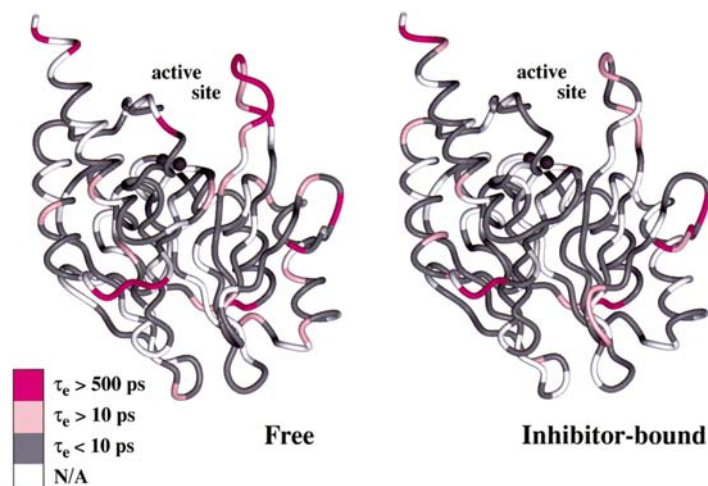


Figure 3.1 Effective correlation times for internal motion mapped onto the X-ray crystal structure for the free and inhibitor-bound forms of metallo- $\beta$ -lactamase. Long correlation times ( $\tau_e > 500$  ps) are shown in hot pink (model 5), while intermediate correlation times ( $10 > \tau_e > 500$  ps) are shown in light pink (model 2). Residues with fast internal motions ( $\tau_e < 10$  ps) are shown in gray. Residues that did not fit any model or residues for which relaxation data was unobtainable due to spectral overlap are shown in white.

### **3.3 An Extremely Sophisticated Enzyme: DHFR**

If the metallo- $\beta$ -lactamase can be regarded as a blunt instrument for indiscriminately destroying any antibiotic, dihydrofolate reductase can be regarded as a precision machine akin to a Swiss watch. It catalyzes the extremely specific and vital function of hydride transfer from a cofactor, NADPH, to a substrate, dihydrofolate (DHF), to form tetrahydrofolate (THF), which is absolutely required for the production of deoxyribonucleotides from ribonucleotides. Inactivation of bacterial DHFR is a commonly used and very effective mechanism for a number of anti-infectives, and human DHFR is the target for a number of anti-cancer drugs. These agents are effective because they target an extremely important enzyme, and also because the mechanism of the enzyme is complex, allowing a number of avenues for disruption of catalysis. As a result of the work that we have reported over the last 10 years or so, new avenues, concerned with disruption of DHFR function in a more subtle way, may be opening, with the possibility of new therapies with fewer side effects than the current ones.

Work on DHFR began in the Wright lab, as a collaboration with Dr. Stephen Benkovic of Pennsylvania State University. Initial characterization of *E. coli* DHFR by NMR was achieved in the 1990s<sup>40-44</sup>. The backbone dynamics of DHFR showed several very unusual characteristics. Unlike, for example, the lactamase enzyme, where the molecule as a whole was rather uniformly rigid except in the vicinity of the active site, certain parts of the polypeptide chain of DHFR, well removed from the active site, were much more mobile than others<sup>45</sup>. The rich diversity of dynamical features reported in this paper provided the major impetus for the further study of DHFR dynamics and the application for the Program Project grant mentioned above.

Estimation of polypeptide chain dynamics by NMR involves measurement of relaxation behavior on a per-residue basis. Most commonly, measurements are made for the polypeptide backbone amide groups of <sup>15</sup>N-labeled proteins.  $T_1$  and  $T_2$  relaxation times and heteronuclear [<sup>1</sup>H]-<sup>15</sup>N nuclear Overhauser enhancements (NOE) are measured, and the results analyzed using the so-called “model-free” approach<sup>46,47</sup>. The parameters derived from this approach are the generalized order parameter  $S^2$ , the effective internal correlation time  $\tau_e$  and the exchange rate

$R_{ex}$ , each of which informs on local motions on different time scales from picosecond ( $S^2$ ) and nanosecond ( $\tau_e$ ) to microsecond-millisecond ( $R_{ex}$ ).

The effort on DHFR in the 2000s has been an ambitious attempt to define the motions of the polypeptide chain for all of the stages in the catalytic cycle. Single-turnover kinetic measurements<sup>48</sup> showed that DHFR operates via a cyclic ping-pong mechanism, with the beginning of the cycle consisting of a complex between the enzyme and the NADPH cofactor. Upon binding of substrate DHF, the Michaelis complex is formed. Hydride transfer from NADPH to DHF results in a ternary DHFR:NADP<sup>+</sup>:THF complex. The next step is dissociation of the oxidized cofactor to give the product binary complex, and NADPH rebinding is necessary, forming the DHFR:NADPH:THF complex, before dissociation of the product THF in the rate-determining step, to yield the DHFR:NADPH complex again. There are thus 5 major complexes that participate in the catalytic cycle. All of these complexes can be prepared with the native substrate, product and cofactor, and with analogs, which are necessary, for example, when studying the Michaelis complex, due to enzyme turnover. Over 60 X-ray crystal structures of *E. coli* DHFR in various substrate/product and cofactor complexes are available<sup>49</sup>. Our backbone dynamics measurements showed that the loops of DHFR, which were thought to affect the catalytic mechanism, were in fact undergoing characteristic motion in the various complexes of the catalytic cycle, not only in the backbone<sup>14,18,50</sup> but in the side chains as well<sup>16</sup>. A full relaxation treatment of one of the DHFR complexes, including relaxation dispersion measurements that directly inform on the  $\mu$ s-ms motions of the backbone, demonstrated that the rates of interconversion between the ground state and the lowest-energy excited state were of the same order of magnitude as the rate of interconversion between this complex and the next one in the catalytic cycle<sup>13</sup>. This meant that we were possibly observing the marriage between the motions of the protein and the steps of the catalytic cycle. The ultimate demonstration that this is the case for DHFR was made using all of the members of the catalytic cycle was reported in 2006 in the journal *Science*<sup>12</sup>. This highly cited paper was the subject of a number of prominent commentaries (ref. <sup>51</sup>, <http://www.f1000biology.com/article/id/1044073/evaluation>). The major conclusion of this paper is that the motions of a given state of the protein sample the conformations of the next state in the catalytic cycle, and in some cases, the previous state as well. These higher-energy states can be discriminated by the characteristics of their NMR spectra, and the rates of interconversion, as well as the populations of these excited states, can be estimated directly from the relaxation dispersion measurements. A summary of the results reported so far on DHFR is shown in Figure 3.2.



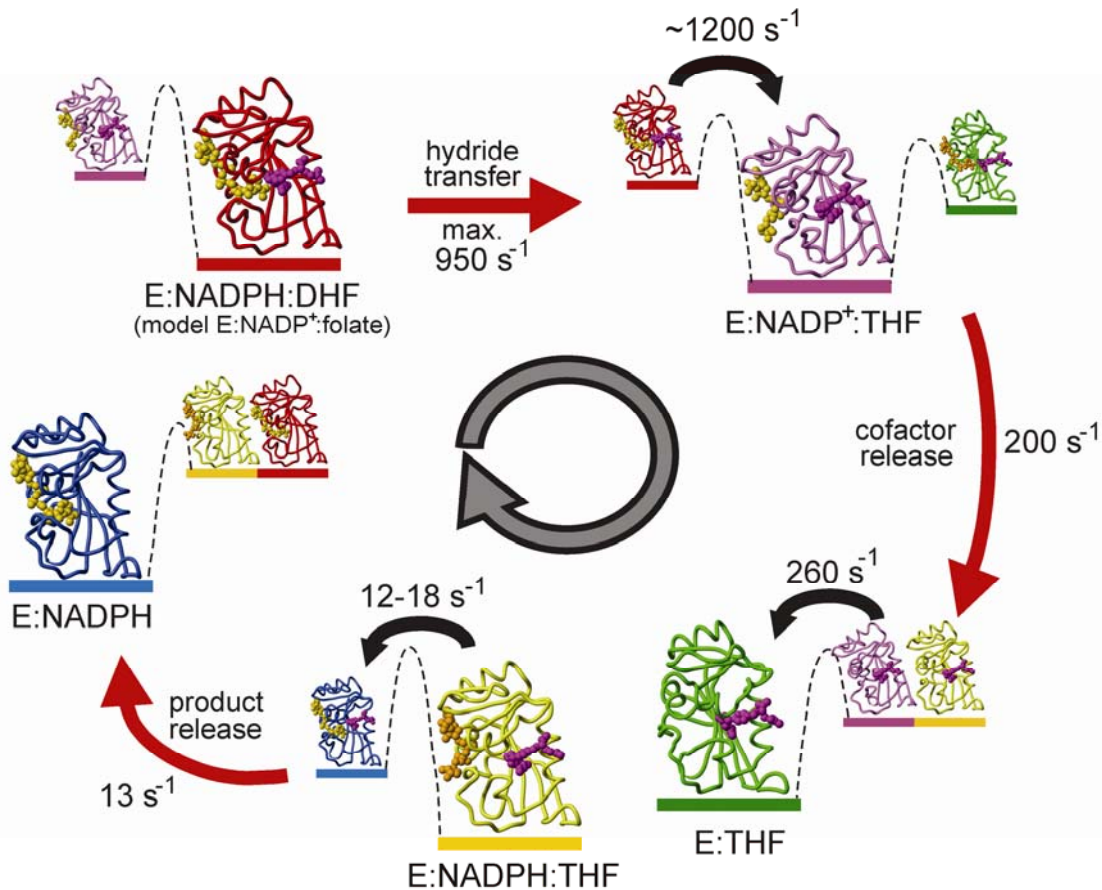


Figure 3.2 The dynamic energy landscape of DHFR catalysis<sup>12</sup>. Ground state (larger) and higher energy (smaller) structures of each intermediate in the cycle, modeled on the published x-ray structures<sup>49</sup> are shown color-coded, with NADPH and NADP<sup>+</sup> shown in gold and substrate, product, and analogs shown in magenta. For each intermediate in the catalytic cycle, the higher energy conformations detected in the relaxation dispersion experiments resemble the ground-state conformations of adjacent intermediates; their interconversion rates, also obtained from the relaxation dispersion experiments, are shown with black arrows. Rate constants for the interconversion between the complexes, measured by pre-steady state enzyme kinetics at 298 K, pH 6<sup>48</sup> are indicated with red arrows. R<sub>2</sub> relaxation dispersion measurements were made at pH 6.8 (E:NADP<sup>+</sup>:folate) or pH 7.6 (E:NADPH:THF, E:NADP<sup>+</sup>:THF, E:NADPH, and E:THF), at 281K (E:NADPH), 300K (E:NADPH:THF, E:NADP<sup>+</sup>:THF and E:THF), or 303K (E:NADP<sup>+</sup>:folate).

### Reference List: Chapter 3

1. Somogyi, B., Welch, G. R., & Damjanovich, S. (1984). The dynamic basis of energy transduction in enzymes. *Biochim. Biophys. Acta* **768**, 81-112.
2. Hammes, G. G. (1964). Mechanism of enzyme catalysis. *Nature* **204**, 342-343.
3. Karplus, M. (2000). Aspects of protein reaction dynamics: deviations from simple behavior. *J. Phys. Chem. B* **104**, 11-27.
4. Cannon, W. R., Singleton, S. F., & Benkovic, S. J. (1996). A perspective on biological catalysis. *Nature Struct. Biol.* **3**, 821-833.
5. Wright, P. E., Dyson, H. J., Lerner, R. A., Riechmann, L., & Tsang, P. (1990). Antigen-antibody interactions: an NMR approach. *Biochem. Pharm.* **40**, 83-88.
6. Janda, K. D., Schloeder, D., Benkovic, S. J., & Lerner, R. A. (1988). Induction of an antibody that catalyzes the hydrolysis of an amide bond. *Science* **241**, 1188-1191.
7. Hoffmann, T., Zhong, G., List, B., Shabat, D., Anderson, J., Gramatikova, S., Lerner, R. A., & Barbas, C. F., III (1999). Aldolase antibodies of remarkable scope. *J. Am. Chem. Soc.* **120**, 2768-2779.
8. Krebs, J. F., Siuzdak, G., Dyson, H. J., Stewart, J. D., & Benkovic, S. J. (1995). Detection of a catalytic antibody species acylated at the active site by electrospray mass spectrometry. *Biochemistry* **34**, 720-723.
9. Siuzdak, G., Krebs, J. F., Benkovic, S. J., & Dyson, H. J. (1994). Binding of hapten to a single-chain catalytic antibody demonstrated by electrospray mass spectrometry. *J. Am. Chem. Soc.* **116**, 7937-7938.
10. Kroon, G. J. A., Martinez-Yamout, M. A., Krebs, J. F., Chung, J., Dyson, H. J., & Wright, P. E. (1999). Backbone resonance assignments for the Fv fragment of the catalytic antibody NPN43C9 with bound *p*-nitrophenol. *J. Biomol. NMR* **15**, 83-84.
11. Kroon, G. J. A., Mo, H., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2003). Changes in structure and dynamics of the Fv fragment of a catalytic antibody upon binding of inhibitor. *Protein Sci.* **12**, 1386.
12. Boehr, D. D., McElheny, D., Dyson, H. J., & Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638-1642.
13. McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J., & Wright, P. E. (2005). Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. USA* **102**, 5032-5037.
14. Osborne, M. J., Schnell, J., Benkovic, S. J., Dyson, H. J., & Wright, P. E. (2001). Backbone dynamics in dihydrofolate reductase complexes: Role of loop flexibility in the catalytic mechanism. *Biochemistry* **40**, 9846-9859.
15. Osborne, M. J., Venkitakrishnan, R. P., Dyson, H. J., & Wright, P. E. (2003). Diagnostic chemical shift markers for loop conformation and cofactor binding in dihydrofolate reductase complexes. *Protein Sci.* **12**, 2230-2238.
16. Schnell, J. R., Dyson, H. J., & Wright, P. E. (2004). Effect of cofactor binding and loop conformation on side chain methyl dynamics in dihydrofolate reductase. *Biochemistry* **43**, 374-383.
17. Schnell, J. R., Dyson, H. J., & Wright, P. E. (2004). Structure, dynamics and catalytic function of dihydrofolate reductase. *Ann. Rev. Biophys. Biomol. Struct.* **33**, 119-140.
18. Venkitakrishnan, R. P., Zaborowski, E., McElheny, D., Benkovic, S. J., Dyson, H. J., & Wright, P. E. (2004). Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry* **43**, 16046-16055.
19. Huntley, J. J., Fast, W., Benkovic, S. J., Wright, P. E., & Dyson, H. J. (2003). Role of a solvent-exposed tryptophan in the recognition and binding of antibiotic substrates for a metallo-beta-lactamase. *Protein Sci.* **12**, 1368-1375.
20. Huntley, J. J. A., Scrofani, S. D. B., Osborne, M. J., Wright, P. E., & Dyson, H. J. (2000). Dynamics of the metallo- $\beta$ -lactamase from *Bacteroides fragilis* in the presence and absence of a tight-binding inhibitor? *Biochemistry* **39**, 13356-13364.
21. Scrofani, S. D., Wright, P. E., & Dyson, H. J. (1998).  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  NMR backbone assignments of 25.5 kDa metallo- $\beta$ -lactamase from *Bacteroides fragilis*. *J. Biomol. NMR* **12**, 201-202.
22. Scrofani, S. D. B., Chung, J., Huntley, J. J. A., Benkovic, S. J., Wright, P. E., & Dyson, H. J. (1999). NMR characterization of the metallo- $\beta$ -lactamase from *Bacteroides fragilis* and its interaction with a tight-binding inhibitor: Role of an active-site loop. *Biochemistry* **38**, 14507-14514.

23. Scrofani, S. D., Wright, P. E., & Dyson, H. J. (1998). The identification of metal-binding ligand residues in metalloproteins using nuclear magnetic resonance spectroscopy. *Protein Sci.* **7**, 2476-2479.
24. Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T. R., Benkovic, S. J., & Hammes-Schiffer, S. (2002). Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **99**, 2794-2799.
25. Antikainen, N. M., Smiley, R. D., Benkovic, S. J., & Hammes, G. G. (2005). Conformation coupled enzyme catalysis: single-molecule and transient kinetics investigation of dihydrofolate reductase. *Biochemistry* **44**, 16835-16843.
26. Benkovic, S. J. & Hammes-Schiffer, S. (2003). A perspective on enzyme catalysis. *Science* **301**, 1196-1202.
27. Benkovic, S. J. & Hammes-Schiffer, S. (2006). Enzyme motions inside and out. *Science* **312**, 208-209.
28. Rajagopalan, P. T., Lutz, S., & Benkovic, S. J. (2002). Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: mutational effects on hydride transfer rates. *Biochemistry* **41**, 12618-12628.
29. Rajagopalan, P. T. & Benkovic, S. J. (2002). Preorganization and protein dynamics in enzyme catalysis. *Chem. Rec.* **2**, 24-36.
30. Wang, L., Tharp, S., Selzer, T., Benkovic, S. J., & Kohen, A. (2006). Effects of a distal mutation on active site chemistry. *Biochemistry* **45**, 1383-1392.
31. Wang, L., Goodey, N. M., Benkovic, S. J., & Kohen, A. (2006). Coordinated effects of distal mutations on environmentally coupled tunneling in dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.*
32. Wang, L., Goodey, N. M., Benkovic, S. J., & Kohen, A. (2006). The role of enzyme dynamics and tunnelling in catalysing hydride transfer: studies of distal mutants of dihydrofolate reductase. *Philos. Trans. R. Soc. Lond B Biol Sci.* **361**, 1307-1315.
33. Zhang, Z., Rajagopalan, P. T., Selzer, T., Benkovic, S. J., & Hammes, G. G. (2004). Single-molecule and transient kinetics investigation of the interaction of dihydrofolate reductase with NADPH and dihydrofolate. *Proc. Natl. Acad. Sci. U. S. A* **101**, 2764-2769.
34. Radkiewicz, J. L. & Brooks, C. L. (2000). Protein dynamics in enzymatic catalysis: Exploration of dihydrofolate reductase. *J. Am. Chem. Soc.* **122**, 225-231.
35. Rod, T. H., Radkiewicz, J. L., & Brooks, C. L. (2003). Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA* **100**, 6980-6985.
36. Rod, T. H. & Brooks, C. L. (2003). How dihydrofolate reductase facilitates protonation of dihydrofolate. *J. Am. Chem. Soc.* **125**, 8718-8719.
37. Thorpe, I. F. & Brooks, C. L. (2003). Barriers to hydride transfer in wild type and mutant dihydrofolate reductase from *E. coli*. *J. Phys. Chem. B* **107**, 14042-14051.
38. Thorpe, I. F. & Brooks, C. L. (2004). The coupling of structural fluctuations to hydride transfer in dihydrofolate reductase. *Proteins in press*.
39. Thorpe, I. F. & Brooks, C. L., III (2005). Conformational substates modulate hydride transfer in dihydrofolate reductase. *J. Am. Chem. Soc.* **127**, 12997-13006.
40. Falzone, C. J., Wright, P. E., & Benkovic, S. J. (1994). Dynamics of a flexible loop in dihydrofolate reductase from *Escherichia coli* and its implication for catalysis. *Biochemistry* **33**, 439-442.
41. Falzone, C. J., Cavanagh, J., Cowart, M., Palmer, A. G., Matthews, C. R., Benkovic, S. J., & Wright, P. E. (1994). <sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C resonance assignments, secondary structure, and the conformation of substrate in the binary folate complex of *Escherichia coli* dihydrofolate reductase. *J. Biomol. NMR* **4**, 349-366.
42. Falzone, C. J., Benkovic, S. J., & Wright, P. E. (1990). Partial <sup>1</sup>H assignments of the *Escherichia coli* dihydrofolate reductase complex with folate: Evidence for a unique conformation of bound ligand. *Biochemistry* **29**, 9667-9677.
43. Falzone, C. J., Wright, P. E., & Benkovic, S. J. (1991). Evidence for two interconverting protein isomers in the methotrexate complex of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **30**, 2184-2191.
44. Li, L., Falzone, C. J., Wright, P. E., & Benkovic, S. J. (1992). Functional role of a mobile loop of *Escherichia coli* dihydrofolate reductase in transition-state stabilization. *Biochemistry* **31**, 7826-7833.
45. Epstein, D. M., Benkovic, S. J., & Wright, P. E. (1995). Dynamics of the dihydrofolate reductase folate complex: Catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features. *Biochemistry* **34**, 11037-11048.
46. 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.
47. 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.

48. Fierke, C. A., Johnson, K. A., & Benkovic, S. J. (1987). Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **26**, 4085-4092.
49. Sawaya, M. R. & Kraut, J. (1997). Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586-603.
50. Osborne, M. J. & Wright, P. E. (2001). Anisotropic rotational diffusion in model-free analysis for a ternary-DHFR complex. *J. Biomol. NMR* **19**, 209-230.
51. Vendruscolo, M. & Dobson, C. M. (2006). Structural biology. Dynamic visions of enzymatic reactions. *Science* **313**, 1586-1587.

## **CHAPTER 4. INTRINSICALLY UNSTRUCTURED PROTEINS**

As advances were made in molecular genetics techniques in the 1980s and 1990s, there was a gradual change in the basic philosophy and methodology of biochemical studies of proteins. Previously, the sequence of events would involve the identification of some observable function in an organism or tissue, followed by the isolation of the molecules responsible for this function and subsequent structure determination using the purified protein. Nowadays we are increasingly using information derived from genetic studies to identify regions of genes and the proteins for which they code that are responsible for observable cellular functions. The specific functions of the proteins involved may be quite difficult to determine – in these cases, the determination of a structure may occur before the specific protein function is known. This latter sequence of events allows the detection of important proteins that may not be fully folded in their active states. We are only now beginning to understand that such proteins can be functional because we have only recently become aware of their existence. In addition, there was a general prejudice for studies of folded proteins – if a protein construct proved to be intractably unfolded, it was frequently discarded or ignored as “too difficult”. Certainly an unfolded protein is impossible to study by X-ray techniques, but we have increasingly been able to use spectroscopic techniques in solution, especially NMR, to give sometimes quite startling insights into the presence and function of unstructured domains of proteins.

### **4.1 A Brief Historical Perspective**

Because of our interest in protein folding and peptide structure (Chapter 1), in protein structure and its relation to function (Chapter 2) and in protein motion and its relation to function (Chapter 3), we were particularly open to the idea that a functional protein need not be rigidly structured. Nevertheless, we took a great deal of time persuading ourselves that it was so. Examples were known of functional unstructured proteins: peptide hormones are well-known to be unfolded in the absence of their binding partners<sup>1</sup>. Early NMR work by Peter Wright had even identified unstructured storage proteins in living cells<sup>2</sup>. The p21 system<sup>3,4</sup> finally persuaded us not only that unstructured proteins were functional, but that their functions were important in regulation of cellular metabolism, and indeed, relied upon their unstructured nature for carrying out their function. Extensive collaborations with the labs of Ron Evans and Marc Montminy at the Salk Institute gave us access to genetics-derived information on the regions of a number of multi-domain proteins involved in the process of transcriptional activation. Our studies of these proteins and their interactions have formed the core group of observations that have enabled us to formulate a picture of the nature and function of functional, intrinsically unstructured proteins in cellular metabolism. This field is receiving a great deal of attention from a number of quarters, and we have written several influential recent review articles that have in some senses defined the field<sup>5-8</sup>. In the following sections, I will describe some of our seminal studies on CREB-binding protein, zinc finger and other DNA-binding proteins, and conclude with a brief description of a new research direction on the interactions of chaperones with client proteins.

### **4.2 CREB-Binding Protein – the Transcriptional Coactivation Machine**

When a signal is received by a cell, for example by the binding of a hormone at the surface of the cell, the signal is translated into intracellular instructions for the activation and expression of genes by one of a number of specific signaling pathways. One prominent pathway includes the second messenger cyclic AMP, and one of the targets of cyclic AMP is the so-called “cyclic AMP Response-Element binding protein” (CREB), which binds to upstream elements of gene promoters prior to their transcription. Transcription requires the presence of a large number

of protein components, including the RNA polymerase, various transcription factors and the TATA binding protein. The initiation of transcription requires the recruitment of a transcriptional coactivator such as CREB-binding protein (CBP) or its close relative p300. Multiple overlapping routes of signal transduction are used in cells<sup>9</sup>; a simplified scheme showing the CREB-CBP interaction is shown in Figure 4.1.

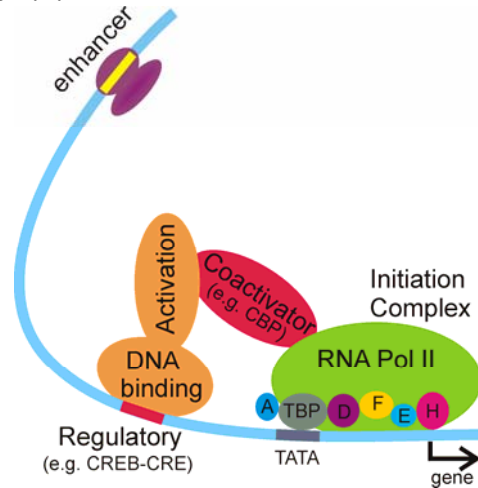


Figure 4.1 Schematic diagram showing the interactions of several of the proteins involved in transcription. The RNA polymerase complex with its attendant transcription factors A, D, F, E and H are connected to the TATA element upstream of the gene by the TATA binding protein. The regulatory element is bound by the response-element binding protein (e.g. CREB) via its DNA-binding domain. The activation domain interacts with a recruited coactivator (e.g. CBP) which also interacts with the polymerase and with the histones attached to the DNA. Enhancer elements distant from the transcriptional initiation sites promote the interactions between these components for efficient transcriptional initiation.

CBP is a multi-domain protein with 2441 amino acids. Several of the domains appear to be well folded, but there is a considerable portion of the sequence that appears to be unable to fold into a compact 3-dimensional structure on its own. This is illustrated in Figure 4.2, which shows the amino acid sequence of CBP, annotated according to the amino acid type. Certain parts of the sequence show a mixture of these amino acid types, while other parts are highly biased towards the amino acids in the “green” classification, that is, they possess small, usually hydrophilic side chains. This lack of sequence complexity is the datum that is used to predict the presence of intrinsic disorder in proteins<sup>10</sup>.

CBP is also a multi-functional binding partner for a large number of proteins in the transcriptional machinery. In many cases, the binding sites have been mapped to individual domains of CBP, as indicated schematically in Figure 4.3.

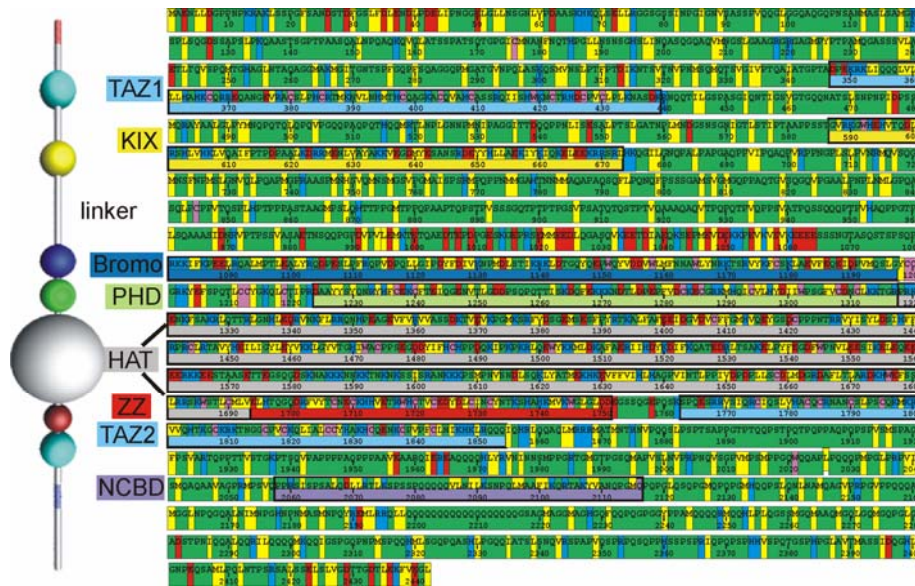


Figure 4.2 Amino acid sequence of mouse CBP. Amino acids have been colored according to the following broad groupings: red, acidic (Asp, Glu); blue, basic (Arg, Lys, His); yellow, bulky hydrophobic (Val, Leu, Ile, Phe, Tyr, Met), green, small hydrophilic (Gly, Ala, Asn, Gln, Ser, Thr, Pro) and pink, unusual (Cys, Trp). In addition, the locations of the structured domains are indicated by a superimposed block of color as indicated in the key to the left. It is noticeable that many of the regions outside the structured domains are “green”.

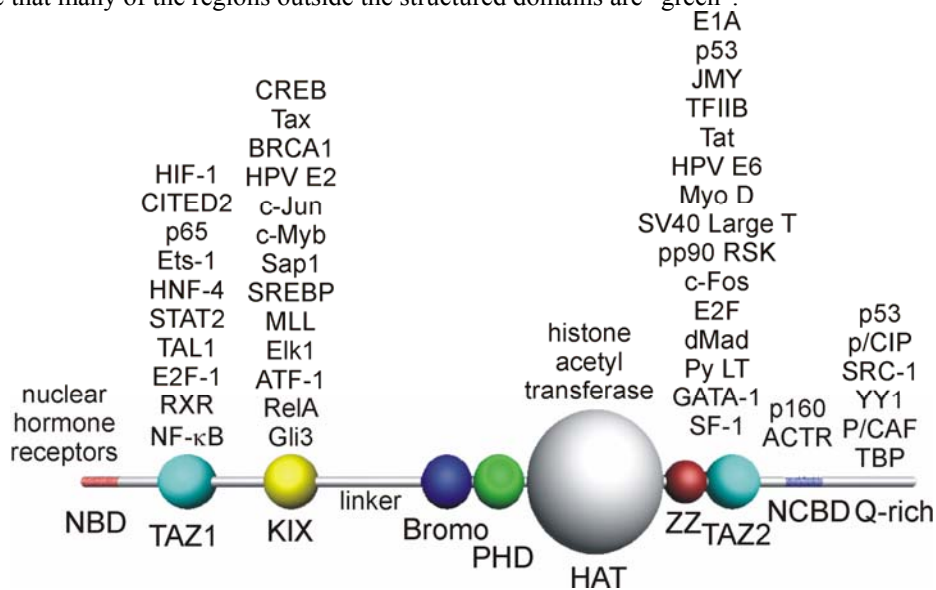


Figure 4.3 Binding partners for the various domains of CBP and its close relative p300.

### 4.3 Dissection of CBP into Domains and Structural Basis for Binding of Partners

The extensive published mapping experiments summarized in Figure 4.3 form the basis for the design of short constructs that contain one or more of the domains of CBP. This project represents years of work on the part of a series of excellent postdoctoral fellows, but primarily the expertise of a senior member of the lab, Dr. Maria Martinez-Yamout, who has been primarily responsible for the design, cloning and expression of the CBP domains and their partners, frequently overcoming multiple problems in her pursuit of an NMR-suitable protein for this project.

#### 4.3.1 The KIX domain and its Interaction Partners

Our first CBP project was a collaboration with the lab of Marc Montminy, then at the Salk Institute. He had identified a region of the CREB protein that interacted with the KIX domain of CBP, a region called the kinase inducible domain (KID), where a crucial phosphorylation step was required to promote the interaction<sup>11</sup>. If Ser133 is not phosphorylated, the interaction between the KID and KIX domains is not detectable in biological assays (we can in fact detect a weak interaction between unphosphorylated KID and KIX by NMR). The NMR solution structure of the complex between pKID and KIX<sup>12</sup> shows that the pKID domain binds to the KIX domain as a pair of bent helical structures, with the phosphoserine at the intersection of the two helices, forming hydrogen bonds with tyrosine and lysine side chains of KIX. The NMR spectra show that the KIX domain is stably folded in solution in the absence of binding partners, whereas the pKID is completely unstructured. Upon binding to KIX, the pKID folds: this is one of the earliest well-documented examples of an intrinsically unstructured protein that folds into a well-defined structure upon binding to its partner.

The pKID-KIX system has provided fascinating insights in a number of follow-up studies, both on the mechanics and thermodynamics of folding-coupled binding, but on the structural and chemical basis for the switch that allows the all-important discrimination between the phosphorylated CREB (switch ON) and the unphosphorylated CREB (switch OFF). Initial studies focused on defining the conformational differences between KID and pKID free in solution<sup>13</sup>. There are no detectable differences in the conformational ensemble of free KID and free pKID, indicating that the structural discrimination occurs upon binding to KIX and that the phosphoryl group does not influence the intrinsic structural disorder of this domain. The KIX-pKID system was also used as a paradigm to report methods of designing constructs for structural studies in solution<sup>14</sup>.

The requirement for the presence of the highly charged phosphate group before pKID could bind to KIX, plus the obvious thermodynamic penalty that would have to be paid for folding of the peptide upon binding intrigued us and prompted a series of new studies that have served to define not only the CREB-CBP interaction, but provide a basis for thinking about folding upon binding in general, as described in our seminal review article in 1999<sup>8</sup>. Firstly, we examined the influence of residual secondary structure formation in the free KIX ligand by comparing the thermodynamics of binding of pKID (and KID) and another KIX partner c-Myb<sup>15</sup>. Unlike pKID, an inducible ligand, c-Myb is a “constitutive” ligand, that is, KIX binding by c-Myb does not require post-translational modification (phosphorylation for example) in order to bind KIX. C-Myb contains a significant amount of residual helical structure in the free state, which should by all accounts make the binding stronger to KIX<sup>16</sup>. The dissociation constants for the two ligands are comparable, but isothermal titration calorimetry reveals that the thermodynamic basis for binding of the two ligands is completely different. As expected, the entropy loss in folding of pKID from an unstructured state provides a large unfavorable component in the thermodynamics of binding, but this is more than compensated by the large favorable enthalpy component that arises due to the presence of the phosphoryl group in pKID. Many of the hydrophobic contacts between the two ligands and KIX are similar<sup>17</sup>, but the difference in the conformational ensemble of the free state renders the c-Myb capable of binding to KIX, while the unphosphorylated KID cannot overcome the entropic barrier to give a significant binding free energy. This comparison thus gives a neat explanation, on the basis of the underlying structure and chemistry, of the inducible nature of the pKID-KIX interaction, and



points to an important aspect of protein-protein interactions in this case. The switch only works correctly if pKID and KID are unstructured in the free state, so that KIX binding can only occur when the phosphate is present to overcome the entropic folding barrier.

Although its interaction with CREB served as the original impetus for naming CBP, it turns out that the KIX domain is extremely multi-faceted, and we are continually finding new functions. It is clear from Figure 4.3 that KIX binds a multitude of protein factors. We have recently begun to understand that KIX (and other CBP domains) binds many of these factors in different sites and may bind to more than one factor, with synergy between them. This was shown initially by the observation that another protein, the activation domain of the mixed-linear leukemia protein (MLL) binds to the KIX domain in the presence of c-Myb or pKID, and further, that this binding is synergistic<sup>18</sup>. Chemical shift perturbations showed that the binding site for MLL was different from that of pKID or c-Myb<sup>18</sup>, and structure determination of the ternary complex between KIX, MLL and c-Myb shows the structural basis for the cooperativity – additional interactions are formed when both ligands are bound to KIX<sup>19</sup>. The KIX NMR structures are illustrated in Figure 4.4.

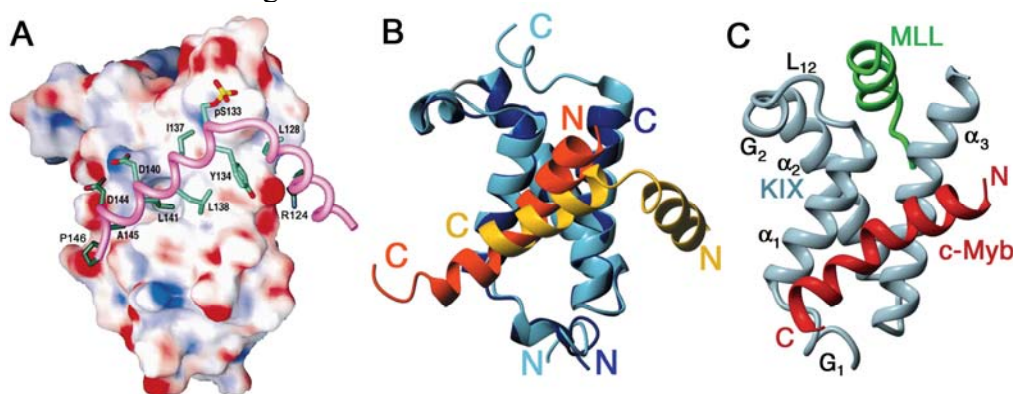


Figure 4.4 A. Representative NMR structure of the KIX domain of CBP (surface colored according to charge: red, negative, blue, positive) in complex with pKID (pink ribbon), showing hydrophobic residues buried in the interface, together with the position of the phosphoserine residue (pSer)<sup>12</sup>. B. Superposition of the structures of the KIX-pKID complex (dark blue and yellow ribbons) and the KIX-c-Myb complexes (light blue and red ribbons)<sup>17</sup>. C. NMR structure of the ternary complex of KIX (grey), c-Myb (red) and MLL (green)<sup>19</sup>.

#### 4.3.2 The KIX-pKID Interaction as a Prototype Coupled Folding and Binding Interaction

Most recently, we have achieved an interesting milestone in our quest to understand the role of unstructured protein domains. The pKID-KIX interaction was subjected to a meticulous analysis by NMR chemical shift titration and relaxation dispersion measurements<sup>20</sup>. These experiments were able to show that, for this system, the interaction occurs first by the formation of non-specific encounter complexes, primarily mediated by hydrophobic interactions. The next stage is the formation of an intermediate state, where the pKID is bound at the final binding site, but is not folded correctly. The final step is the conversion of this intermediate to the fully folded state. This study represents one of our proudest achievements, and has been recognized (<http://www.fl1000biology.com/article/id/1086851/evaluation>).

#### 4.3.3 Solution Structures of Other CBP Domains

Following intensive work by Dr. Yamout and her associates, we were able in a very short period of time to determine solution structures of a number of the domains of CBP. Some of the CBP (or p300) domains proved impossible to clone, express or purify, and so, for some cases, we resorted to using homologous domains from other species. All told, we have published a type

structure for all of the domains of CBP except the histone acetyl transferase domain (HAT), which has up till now eluded us due to its large size and (most likely) the presence of significant unstructured regions within the sequence of the CBP HAT (see Figure 4.2). The structures of the free TAZ2 domain<sup>21</sup>, TAZ1 domain<sup>22</sup> and ZZ domain<sup>23</sup> are all derived from the CBP sequence. The bromodomain structure<sup>24</sup> was derived from the GCN5 sequence and the PHD domain<sup>25</sup> from the Williams-Beuren syndrome transcription factor. Of these, all except the bromodomain contain structural zinc ions, and are unfolded in the absence of zinc. The structures of these domains are illustrated in Figure 4.5.

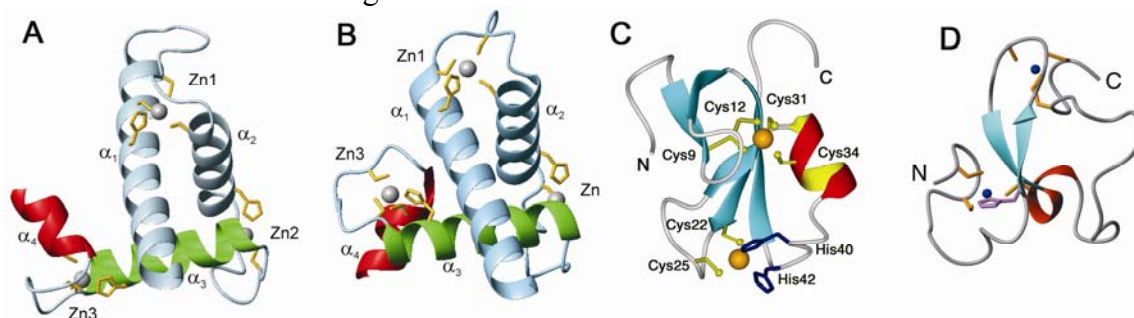


Figure 4.5 Representative solution structures of CBP domains. A. TAZ1<sup>22</sup>. B. TAZ2<sup>21</sup>. C. ZZ<sup>23</sup>. D. PHD from Williams-Beuren syndrome transcription factor<sup>25</sup>.

#### **4.4 Variations on “Coupled Folding and Binding” to CBP Domains**

The structure determinations described in the previous paragraph excited a great deal of interest: the TAZ domains provided the first example of an entirely new zinc-binding structural motif, and the ZZ and PHD domains also bear very little resemblance to other known motifs (PHD has some small resemblance to RING and FYVE domains). Nevertheless, our major focus in studies with these domains was to define the structural basis for their interactions with transcription factor ligands. Following our experience with pKID and KIX (see previous section), we made sure to design the constructs carefully so as not to exclude important elements of the final structure. Even with our best skill, it sometimes takes several attempts before a successful design is realized.

##### **4.4.1 Structural Basis for the Action of Hypoxia-Inducible Factor**

The series of interactions that result in the expression of genes as a response to hypoxia (low oxygen tension) in the cell are of great interest to a number of medical fields. This system is the target, for example, of drug discovery efforts in cancer therapy, since the growth of tumors can be halted if the response to hypoxia (which includes the expression of genes for inducing vascularization) can be knocked out. The hypoxic response provides a fascinating example of the direct effects of chemical changes at the molecular level on the response of organisms to environmental changes by the expression of genes.

Under normal oxygen conditions in the cell, the hypoxia response is turned off, by direct chemical modification of the hypoxia inducible factors, the transcription factors responsible for turning on the hypoxia genes. This is illustrated in Figure 4.6. For these proteins, chemical modification, degradation and chaperone interactions work in concert to control the response (reviewed in<sup>26</sup>). The HIF proteins, of which HIF-1 is the prototype, are heterodimers, consisting of  $\alpha$  and  $\beta$  domains. The HIF-1 $\beta$  subunit, also known as aryl hydrocarbon receptor nuclear translocator (ARNT), is constitutively expressed and is not sensitive to oxygen like the HIF-1 $\alpha$  domain. Under normoxic conditions, the HIF-1 $\alpha$  subunit is post-translationally modified by

hydroxylation at several positions. Hydroxylation at two proline residues in the oxygen-dependent degradation (ODD) domain allows binding of the von Hippel-Lindau tumor suppressor protein, which acts as a ubiquitin ligase, targeting these domains to the proteasome for degradation. Hydroxylation at an asparagine residue (Asn803) in the C-terminal activation domain inhibits the interaction of HIF-1 $\alpha$  with the transcriptional coactivators CBP and p300. Under hypoxic conditions, these residues are no longer hydroxylated: HIF-1 $\alpha$  is not now degraded, but dimerizes with HIF-1 $\beta$  and is translocated to the nucleus, followed by transactivation and gene transcription. Thus, remarkably, genes are turned on and off by a direct chemical reaction of the signal molecule itself with the proteins involved in the signaling pathway.

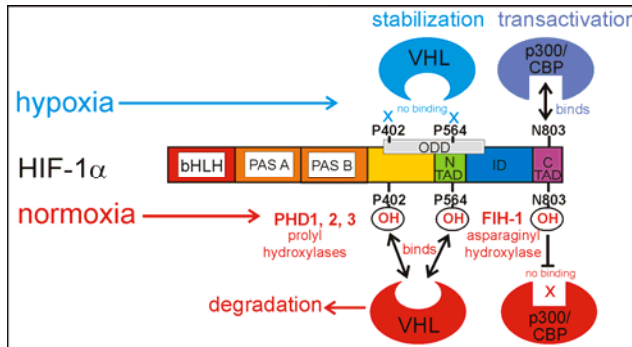


Figure 4.6 Schematic diagram showing the domains of HIF-1 $\alpha$  and the locations of the hydroxylated residues. VHL: von Hippel-Lindau tumor suppressor factor; bHLH: helix-loop-helix domain; PAS A, B: Per-Arnt-Sim domains; ID, inhibitory domain; N, CTAD: N-terminal and C-terminal activation domains; ODD: oxygen-dependent degradation domain; FIH-1: “factor inhibiting HIF”; PHD: prolyl hydroxylase domain-containing protein.

The C-terminal activation domain (CTAD) of HIF-1 $\alpha$  binds to the TAZ1 domain of CBP, as long as Asn803 is not hydroxylated. This characteristic was exploited in the expression of the CTAD in *E. coli*. Both the HIF-1 $\alpha$  CTAD and the TAZ1 domain were co-expressed in a bicistronic (two-gene) expression vector, thus avoiding degradation of the CTAD, which is unstructured in solution in the absence of TAZ1. The structure of the TAZ1- HIF-1 $\alpha$  CTAD complex<sup>27,28</sup> (Figure 4.7, 4.8) illustrates several of the reasons why a binding partner might be unstructured. The CTAD wraps almost completely around the folded TAZ2, forming three short helices with interconnecting loops. Such a structure could not arise if the CTAD was stably folded before the interaction. In addition, it has been pointed out that the formation of complexes by the folding of an unstructured region gives a much greater interaction surface, providing both specificity and affinity, than if both components of the complex were folded before the interaction<sup>29</sup>. Indeed, to form an interaction surface of comparable size to that seen for the TAZ1- HIF-1 $\alpha$  CTAD complex if both proteins were fully folded would require that the size of each protein be much greater, which would prove a burden to the cell. Another reason why the CTAD might be unstructured is so that it can bind to different partners with different structures. This is illustrated in Figure 4.7: the enzyme that catalyzes the post-translational modification of Asn803 of the HIF-1 $\alpha$  CTAD is named “factor inhibiting HIF” or FIH. In the X-ray crystal structure of the complex of FIH with the HIF-1 $\alpha$  CTAD<sup>30</sup>, the sequence containing Asn803 is in an extended conformation (Figure 4.7B), whereas the same sequence in the TAZ1 complex is in a helical conformation (Figure 4.7A). Versatility in the formation of complexes with multiple partners is a hallmark of the intrinsically unstructured sequence.

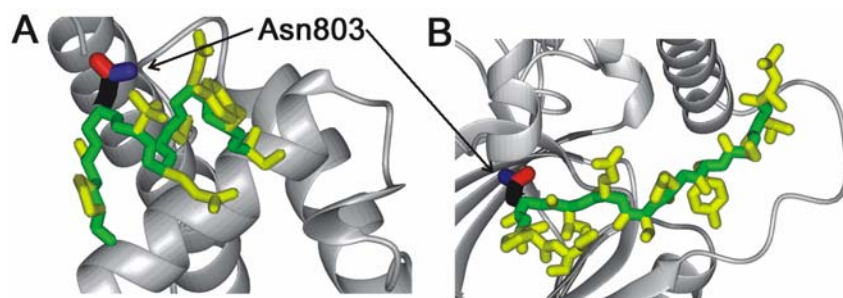


Figure 4.7 A. Portion of a representative NMR structure of TAZ1 (grey) in complex with the HIF-1 $\alpha$  CTAD (green backbone, yellow side chains), showing the helical backbone conformation and the position of Asn803<sup>27</sup>. B. Portion of the X-ray structure of FIH (grey) in complex with the same portion of the HIF-1 $\alpha$  CTAD shown in part A, showing the extended backbone conformation<sup>30</sup>.

#### 4.4.2 Competition for CBP – Alternative TAZ1 Complexes and their Interconversion

If a complex is formed between a ligand and a CBP domain in response to a signal, then there must also be a mechanism for turning off the switch after the need represented by the signal is passed. In view of the high affinity of, for example, the HIF-1 $\alpha$  CTAD for the TAZ1 domain (7 nM<sup>27</sup>), and the burial of Asn803 in the complex, it is difficult to envision how the complex is to be dissociated. A clue to the mechanism comes from the structures of TAZ1 complexes with other ligands, CITED2<sup>31,32</sup> and STAT2 (manuscript submitted). These ligands bind to TAZ1 quite differently from the HIF-1 $\alpha$  CTAD. For CITED2, the binding site overlaps one of the interaction sites of the short helices of the HIF-1 $\alpha$  CTAD, but the rest of the domain binds in a different site. One could envision that CITED2 could replace the HIF-1 $\alpha$  CTAD by binding to the free TAZ1 site and “peeling off” the HIF-1 $\alpha$  CTAD by competition. The STAT2 complex is even more striking: the STAT2 activation domain actually binds in the opposite sense to the HIF-1 $\alpha$  CTAD and CITED2. These structures are illustrated in Figure 4.8.

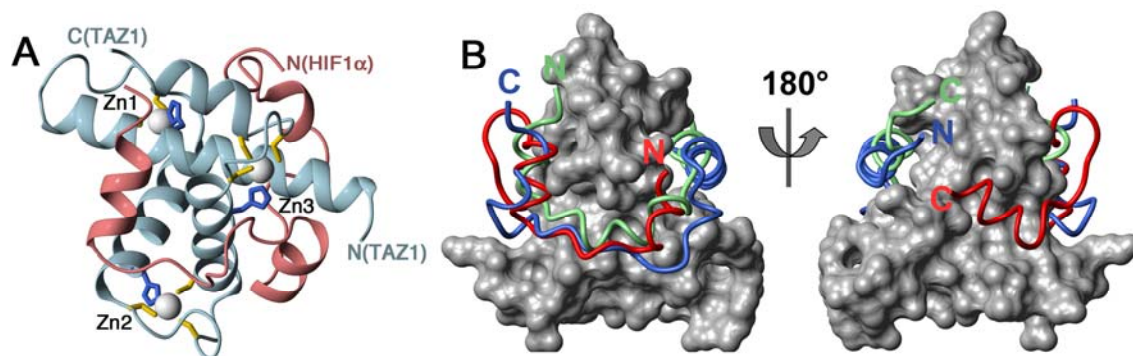


Figure 4.8 A. Representative NMR solution structure of the complex between the TAZ1 domain of CBP (blue) and the C-terminal activation domain of HIF-1 $\alpha$  (pink)<sup>27</sup>. Zinc ions are shown as grey spheres and zinc ligands are colored yellow for Cys and blue for His. B. Superposition of the TAZ1 structure (grey surface) with the ligands HIF-1 $\alpha$  CTAD (red)<sup>27</sup>, CITED2 (blue)<sup>31</sup> and STAT2 (green) (manuscript submitted). The two structures in part B differ by a 180° rotation about the axis lying in the page.

#### 4.4.3 The Ultimate Coupled Folding and Binding Interaction: Mutual Synergistic Folding

Parts of CBP participate in ligand binding even though they are themselves unstructured. An example of such a region is the nuclear coactivator binding domain (NCBD) towards the C-terminus of CBP (see Figure 4.2). Interaction of the NCBD sequence with one of its partners, the ACTR domain of p160, results in the formation of a folded complex from two unfolded sequences, which we have called “mutual synergistic folding”<sup>33</sup>. The NCBD has a small

propensity for the formation of helical secondary structure according to the CD spectrum, while the ACTR domain has no residual structure. When bound together, the CD spectrum shows the presence of a significant amount of helical structure, and, more importantly, shows a cooperative unfolding transition in the presence of denaturant, which occurs for neither of the free components (Figure 4.9).

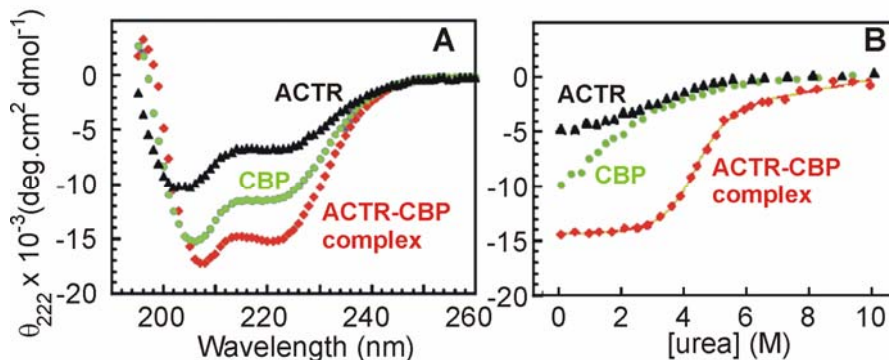


Figure 4.9 A. Circular dichroism spectra of free ACTR (black), free NCB of CBP (green) and the 1:1 complex between the two components (red). B. Urea titration of ACTR, CBP and the complex.

Once again, the structures calculated for the complex reveal that the disordered nature of the component parts of the complex is probably a necessary circumstance for its formation. In the complex, the ACTR polypeptide wraps around the CBP-NCBD portion. The NCBD consists of three helices; one helical segment of ACTR interacts with these 3 helices to form a 4-helix bundle structure (Figure 4.10), with the hydrophobic surface formed by the helical disposition of the diagnostic LXXLL motif fitting snugly into the groove between helices 2 and 3 of the NCBD (Figure 4.10B).

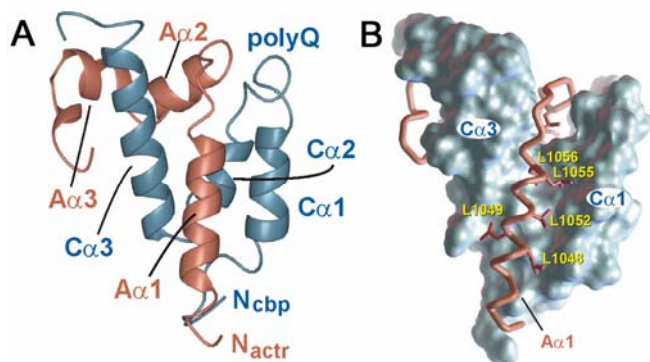


Figure 4.10 A. Representative structure from the ensemble of the complex of the NCBD of CBP (blue) and the ACTR interaction domain (pink). B. Surface representation of CBP-NCBD (blue) with ACTR bound (pink) showing the packing of the side chains in the LXXLL motif.

#### 4.4.4 An Example where No Binding Occurs (despite Literature Reports!)

CBP is an important component of the machinery that regulates the metabolism of the tumor suppressor protein p53. This tetrameric protein has been the subject of a vast array of studies, a tribute to its central role in cellular control mechanisms, and the apparent role of p53 mutants in cancers of many types. The regulation of p53 is complex, and it appears to interact with several of the domains of CBP, including KIX, TAZ1, TAZ2 and the NCBD. A major effort in our lab at present is to characterize these interactions and relate them to the overall function of p53 and its relationship with CBP. Another major regulatory mechanism for p53 is the promotion of its proteasomal degradation by the protein MDM2 (“mouse double-minute protein 2”). An X-ray crystal structure of the complex between p53 and MDM2 was solved several years ago<sup>34</sup>. A more recent report apparently saw evidence for interactions between MDM2 and the

TAZ1 domain of CBP<sup>35</sup>, which we found intriguing and proceeded to investigate by expression of the purified components of the putative complex. However, we observed no affinity of the TAZ1 domain of CBP for the human MDM2 protein. When we re-investigated the original observations<sup>36</sup>, it became clear that they were incorrect. Grossman et al.<sup>35</sup> had included EDTA in the buffer used to observe the interaction. Under these conditions, the TAZ1 domain loses its zinc and becomes irreversibly unfolded. The original observations reported only on a non-specific interaction of MDM2 with an unfolded protein! We have since come across other examples where reported interactions turn out to be erroneous or artifactual. A recent example is described in the later section on nucleic acid-binding zinc finger proteins.

#### 4.4.5 The HDM2 RING Finger Protein

MDM2 (or the human analogue HDM2) contain a number of domains; the functions of all of these domains have not yet been completely defined. The protein contains two zinc-binding regions, a zinc finger domain that was structurally characterized in another lab<sup>37</sup> and a RING finger domain that functions as a ubiquitin E3 ligase, for which we recently calculated solution structures<sup>38</sup>. Interestingly, this protein is a highly symmetric homodimer in solution (Figure 4.11A), a rather unusual structure. RING domains frequently form  $\square$ immer $\square$ , but the HDM2 RING is the only example where the dimerization occurs completely within the domain itself, without the participation of outside elements of structure. By analogy with other published structures, we were able to construct a model of the complex between the RING domain (an E3 ubiquitin ligase) with a UbcH5 ubiquitin-conjugating enzyme (Figure 4.11B).

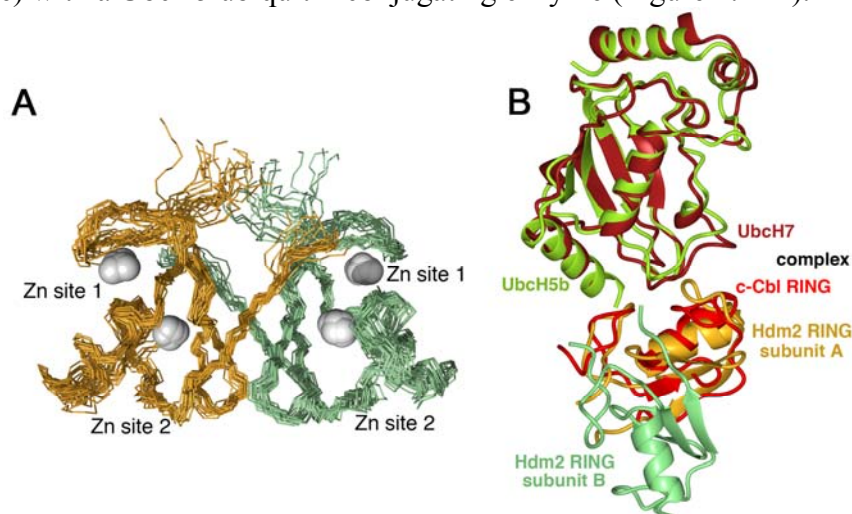


Figure 4.11 A. Family of 20 structures of the HDM2 RING domain homodimer. The two polypeptide chains are colored green and yellow. B. Structural overlay of the UbchH7/c-Cbl complex<sup>39</sup> with the UbchH5b structure<sup>40</sup> and the lowest-energy structure of the Hdm2 RING homodimer<sup>38</sup>.

### 4.5 Interactions of Zinc Finger Proteins with DNA and RNA

Zinc finger proteins are a long-standing interest of the Wright laboratory, which was the first to describe the structure of a zinc finger domain<sup>41</sup>. My own interest in zinc finger proteins dates from my realization that their mode of action in binding to nucleic acids is another example where proteins that are incompletely folded have functional advantages.

#### 4.5.1 Interaction of Hormone Receptors with DNA Requires Protein Flexibility

The retinoid X receptor (RXR) is a central member of the nuclear hormone receptor superfamily of ligand-controlled transcription factors<sup>42</sup>, forming DNA-binding homodimers and

heterodimers with other hormone receptors such as thyroid hormone receptor and retinoic acid receptor. The structure of RXR is fairly typical of nuclear hormone receptors<sup>43</sup>, containing two zinc ions as part of a rather compact structure. From NMR measurements in the presence of DNA<sup>44</sup>, it appears that conformational changes in each monomer, induced by binding DNA, increase the affinity of the monomer units for each other, providing an example of the utility of protein flexibility in the performance of, in this case, a binding function. Flexibility in certain parts of an otherwise well-structured domain also appears to mediate the function of the estrogen-related receptor ERR in binding to DNA<sup>45</sup>. ERR binds to its cognate DNA sequence as a monomer, but contains a long C-terminal tail that is unstructured in the free protein, but which wraps around the DNA in the complex, providing additional affinity and specificity to the monomer binding.

#### 4.5.2 Flexibility Mediates DNA Binding in C<sub>2</sub>H<sub>2</sub> Zinc Fingers

The Cys<sub>2</sub>His<sub>2</sub> zinc finger motif is one of the most common motifs in the human genome, and is frequently associated with binding to nucleic acids. The Wright lab has a long-standing interest in the zinc finger protein TFIIIA, a collaboration with the lab of Joel Gottesfeld at Scripps. Much of the knowledge about this group of proteins has been published from this lab<sup>46-51</sup> and it has also been used as a model system for assessing new measurement techniques<sup>52,53</sup>. My own interest in the zinc finger system arose through comparative dynamic studies of the first three zinc fingers of TFIIIA free<sup>52</sup> and in complex with cognate double-stranded DNA<sup>48</sup>. It appeared that, like the RXR and ERR systems, there was a distinct difference in the dynamics of the protein upon binding to DNA. For TFIIIA and for the WT1 zinc fingers (see next section) this difference takes the form of a very specific change in the linker regions between the zinc fingers. This combination of flexibility in the free state and rigidity in the bound state can be used to explain the mechanism of sequence-specific DNA binding by these proteins.

#### 4.5.3 The Wilms Tumor Suppressor Protein – Zinc Fingers with a Difference

The Wilms tumor suppressor protein gene codes for a protein WT1 that contains 4 C<sub>2</sub>H<sub>2</sub> zinc fingers. Mutations in WT1 are associated with many cancer phenotypes, including the titular Wilms tumor, the most common pediatric solid tumor. The WT1 zinc fingers were of interest as a variant of the most common zinc finger paradigm, that of TFIIIA, as introduced above, or the EGR-1 system, for which a canonical X-ray crystal structure was reported<sup>54</sup>. Alternate RNA splicing gives rise to four splice variants of WT1. One of the splice variants occurs between the third and fourth fingers of the zinc finger domain, with an insertion into the linker sequence of three amino acids, Lys-Thr-Ser (KTS). The form without this insertion (–KTS) has the greater affinity for double-stranded DNA, while the form with the insertion (+KTS) has a greater affinity for RNA. The NMR spectra of the two alternative splice forms are almost identical in the free state, but show specific chemical shift differences in the vicinity of the fourth finger in the bound state<sup>55</sup>. Even more significant are changes in the polypeptide chain dynamics<sup>56</sup> (Figure 4.12), which show unequivocally that the insertion of the KTS sequence causes the fourth finger to be unable to contact the DNA as the other three fingers do, and as all four fingers are able to do in the –KTS splice form. This provides a reasonable rationale for the difference in the affinity of the two splice forms for DNA.

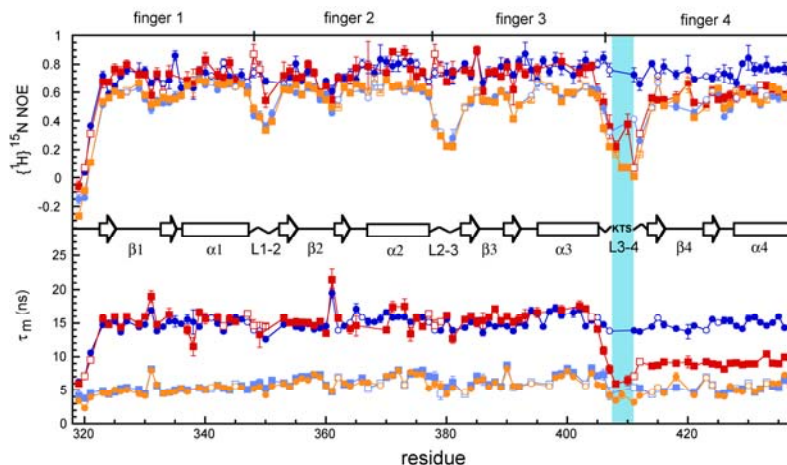


Figure 4.12 Relaxation parameters  $[^1\text{H}]-^{15}\text{N}$  NOE and  $\tau_m$  for the WT1 zf1-4 splice variants -KTS free (light blue) and bound to DNA (dark blue) and +KTS free (orange) and bound to DNA (red)<sup>56</sup>. The position of the KTS insertion is shown by the vertical blue bar. The NOE data indicate that the motions of the fourth finger in the +KTS DNA complex are equivalent to those of the same finger in the free proteins, whereas the fourth finger in the -KTS protein is motionally equivalent to the first three fingers in both complexes. These results show that the presence of the KTS insertion disallows the binding of the fourth finger to the DNA duplex, accounting for the lower affinity of this isoform.

#### 4.5.4 Functional Uses of Motion and Flexibility: the “Snap-Lock”

The effects on the affinity of zf4 of WT1 when the canonical linker sequence is disrupted by the insertion of the KTS sequence prompted a closer look at the structure and function of the zinc finger linker sequences. These sequences are among the most highly conserved in zinc fingers, yet they appear to be almost completely unstructured in the free form<sup>52</sup>. This provides an important clue as to the function of these linkers in DNA binding. A comparison of the sequences of a large number of DNA-binding zinc fingers shows that a consensus 5-amino acid sequence, with small variations, is present throughout this family. This sequence, Thr-Gly-Glu-Lys-Pro (TGEKP), appears well-structured in X-ray and NMR structures of the DNA complexes of zinc finger proteins; the structures of the linker regions can be superimposed, and show a very specific structure that provides a C-terminal cap to the helix of the preceding finger. That is, upon binding to the correct DNA sequence, the linker between the zinc fingers undergoes a disorder-order transition, increasing affinity for the DNA and snapping the zinc fingers into place in the complex. We have termed this process a “snap-lock”<sup>57</sup>; this concept provides a satisfying rationale for specific binding of zinc finger proteins to DNA: the fingers have residual affinity for the DNA, and can search along the sequence for the correct cognate DNA sequence. Once this sequence is found, the affinity for the DNA increases radically due to the additional protein-protein interactions of the “snap-locked” linker.

#### 4.5.5 Interaction of Zinc Fingers and RNA – Even More Mobility

Zinc finger proteins are also heavily involved in binding to RNA. Since RNA is frequently of widely variable structure, compared to the DNA double helix, the proteins that interact with RNA are frequently of unusual form, and frequently less than fully structured in the free state. One example of such a protein is the TIS11 family of zinc finger proteins, which are responsible for regulation of messenger-RNA (mRNA) stability by binding to the 3' untranslated regions of the message. The 3' UTR is rich in uracil and adenine bases (termed AU-rich) and the TIS11d protein specifically binds to the sequence UUAUUUAUU. The TIS11d protein comprises two zinc fingers that are unstructured in the absence of zinc, but which fold into two



independent domains in the presence of zinc. Upon binding to the RNA sequence, the protein takes up a highly specific structure embracing the RNA and providing stabilization to the complex by intercalation of aromatic groups between the RNA bases<sup>58</sup>. Clearly both RNA and protein undergo disorder-order transitions upon complex formation. This structure is illustrated in Figure 4.13A.

The zinc finger domains of the transcription factor TFIIIA interact with RNA as well as DNA; part of the function of TFIIIA in *Xenopus* oocytes is to store and transport 5S ribosomal RNA. While the first three of the 9 zinc fingers have primary function in sequence-specific DNA binding, as described briefly above, the second three fingers have the primary affinity for 5S ribosomal RNA. While the interactions of TFIIIA and similar zinc finger proteins with DNA is by now well-established, the interactions with RNA have only recently been elucidated. The TFIIIA zf4-6 complex with 5S RNA proved to be a very difficult project by NMR, due to the large size of all of the component molecules and the presence of exchange processes that considerably degraded the quality of the spectra. This structure was finally completed in 2006<sup>59</sup>; a 2003 crystal structure that was published in the middle of the study showed overall similarity, but some crucial differences in one of the binding regions, which we attribute to less-than-optimal solution conditions that were required for the formation of crystals. Once again, it is clear from the NMR spectra that both RNA and protein have undergone disorder-order transitions in the formation of the complex. The zf4-6/5S RNA structure is shown in Figure 4.13B.

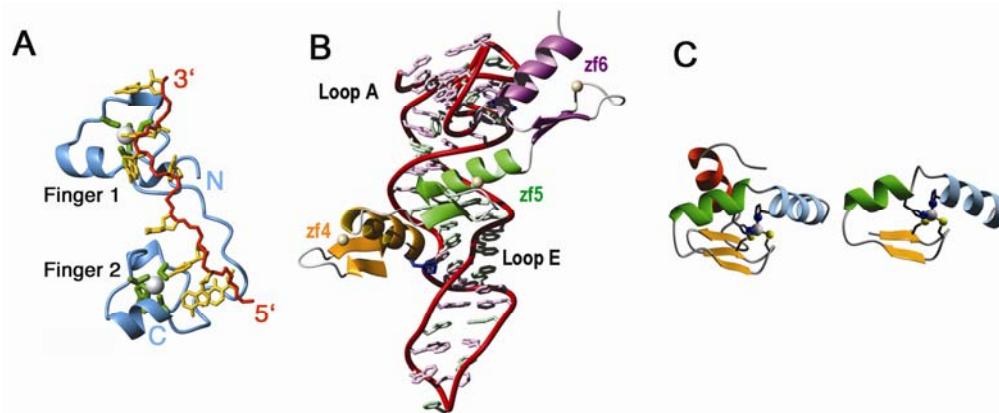


Figure 4.13 A. Representative solution structure of the TIS11D protein (blue ribbon) in complex with the UUAUUUAUU RNA oligonucleotide from the 3' untranslated region of mRNA (backbone red, bases yellow)<sup>58</sup>. Zinc is shown as a grey sphere, with green ligands. B. Complex between zinc fingers 4-6 of TFIIIA and a 55-nucleotide RNA sequence from 5S ribosomal RNA<sup>59</sup>. C. Separate structures of the two zinc finger domains of the RNA-binding protein ZFA, which are joined by a long, unstructured linker in solution<sup>60</sup>.

We have recently begun work on a number of other RNA-binding “zinc finger” proteins. Not all of these proteins have been identified as to function, although they clearly bind RNA, some double-stranded and some single-stranded. It is frequently a challenge to determine the optimal RNA sequence for binding to these proteins. For most of these proteins, there are long and variable linkers between the folded zinc modules. This remains a fascinating field that we are only just beginning to tackle. One example of such a zinc finger is the two-finger construct ZFA<sup>60</sup>, which is shown in Figure 4.13C.

#### 4.5.6 Another Literature Mistake – the Zinc Finger that Wasn't

A report of a zinc finger protein that appeared to be involved in embryonic neural development caught our eye. This protein was identified on the basis of sequence homology as

containing two C<sub>4</sub>-type zinc finger domains, and on this basis was named “Churchill”, a reference to the famous “V-for-Victory” hand signal of the British Prime Minister during World War II<sup>61</sup>. The authors inferred from an immunoprecipitation DNA selection assay utilizing an N-terminal GST-ChCh fusion protein that ChCh specifically binds DNA. Unfortunately, our structure determination and subsequent exhaustive search for a DNA target revealed that ChCh is neither a zinc finger protein nor a DNA-binding protein<sup>62</sup>. The structure of ChCh is shown in Figure 4.14. It contains a highly unusual single-layer  $\beta$ -sheet, with three zinc ions bound, not two. Two of the zinc ions participate in a cysteine-bridged binuclear cluster, while the third connects a long loop to the  $\beta$ -sheet. The N-terminal residue is a cysteine that forms part of the zinc binuclear cluster, which may provide an explanation for the apparent observation of DNA binding by Sheng et al<sup>61</sup>: we observe that the protein is unfolded when additional residues are added at the N-terminus, probably due to the disruption of the zinc binuclear cluster. Although it doubtless functions as a transcriptional activator in neural development, the function of ChCh cannot include DNA binding; its exact function remains unknown at present.

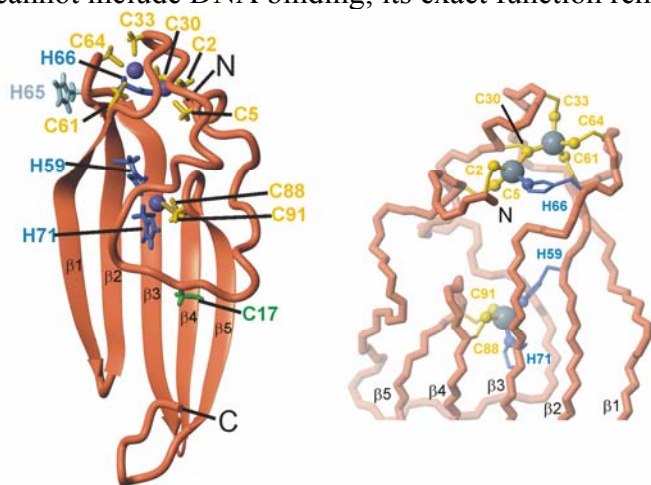


Figure 4.14 (left) Representative structure of ChCh showing the single-layer  $\beta$ -sheet and loop, and the three zinc sites. (right) close-up view of the three zinc sites, showing the shared Cys30 thiolate<sup>62</sup>.

#### 4.5.7 Miscellaneous Non-Zinc Finger Transcriptional Activator Domains

We have published the results of several studies of transcriptional activator domains<sup>63-65</sup>. All of these illustrate that conformational restriction is an important part of complex formation in the interactions between nucleic acids and proteins. Perhaps the most unusual of these is the HMG domain of LEF-1. The structure of LEF-1 in complex with DNA was published some years ago<sup>66</sup>, and it was noted at the time firstly that free LEF-1 was highly unstructured, and secondly that, upon complex formation, the DNA duplex was bent through an unusually large angle, implying that both DNA and protein must undergo a mutual conformational change. This conformational change was recently investigated by NMR<sup>63</sup>. The free protein contains a substantial population of helical structure, but no evidence of stable tertiary structure. The observations argue that, prior to binding, bending and distorting DNA, the HMG domain of LEF-1 exists in a segmentally disordered or partially folded state. Upon complex formation, the protein domain undergoes a cooperative folding transition together with the DNA to a highly ordered and well folded state.

#### 4.6 The Role of Disorder in Protein Function – A Synthesis to Date

It appears that there are many degrees of disorder that can be utilized in the functions of proteins – and nucleic acids. We recently suggested a synthesis of these ideas, referring to published studies from our own and other labs. This synthesis is illustrated in Figure 4.15.

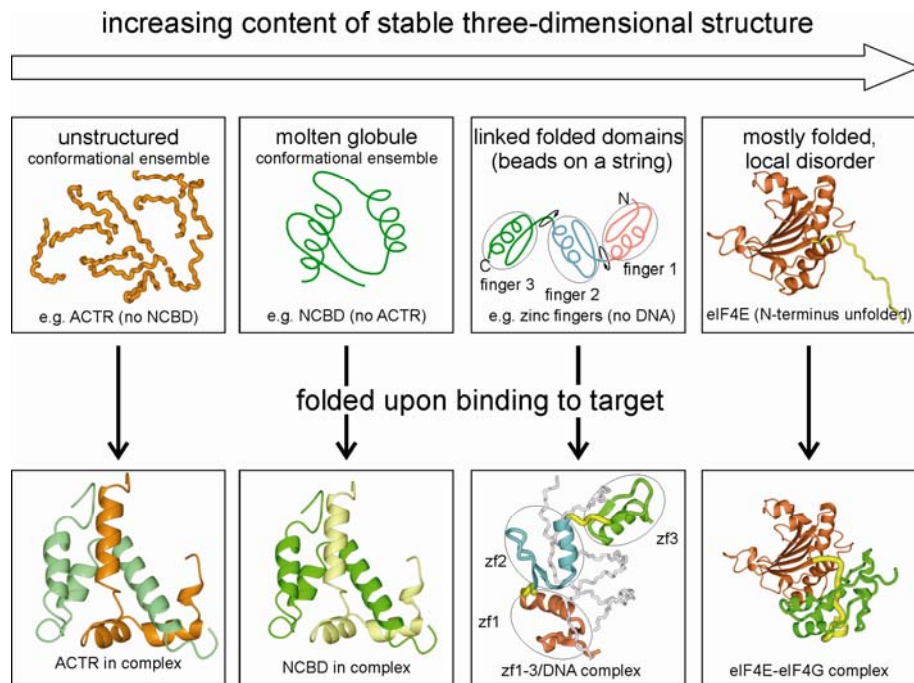


Figure 4.15 The upper panels show schematic diagrams of examples on the continuum of protein structure: an unstructured conformational ensemble, the interaction domain of activator for thyroid hormone and retinoid receptors (ACTR)<sup>33</sup> (left panel); a molten globule-like domain such as the nuclear-coactivator-binding domain (NCBD) of cyclic-AMP-response-element-binding protein (CREB)-binding protein (CBP)<sup>33,67</sup> (second panel); linked folded domains such as a construct that contains the first three zinc fingers of transcription factor-IIA (TFIIIA)<sup>68</sup> (third panel) and free eukaryotic initiation factor-4E (eIF4E)<sup>69</sup>. The lower panels show the structures of the domains in the upper panels folded onto their biological target domains or sequences. The first two lower panels show the mutually folded structure of a complex<sup>33</sup> (PDB code: 1KBH) between the ACTR domain of the p160 coactivator (orange) and the NCBD domain of CBP (green) (in the first panel, the ACTR domain is highlighted; in the second, the NCBD domain is highlighted). The third panel shows the well-ordered structure of the first three zinc-fingers of TFIIIA bound to an oligonucleotide that contains its cognate DNA sequence<sup>51</sup> (PDB code: 1TF3). The fourth panel shows the complex between eIF-4E and eukaryotic initiation factor-4G (eIF4G), which highlights the mutual folding of the N-terminal tail of eIF4E (thick yellow line) and eIF4G (green)<sup>69</sup> (PDB code: 1RF8). All of the three-dimensional structure figures were drawn using MOLMOL<sup>70</sup>. (from <sup>7</sup>.)

#### **4.7 The Next Thought – How Does it Work In Vivo?**

It is clear from the topology of the complexes that are reported in the preceding paragraphs that the components of the complexes have to be unfolded immediately before the formation of the complex. Does this mean that they are present in the cell as unfolded proteins? Or that they are induced to fold, perhaps by the crowded milieu of the cytoplasm? The latter case appears counter-intuitive, as there would have to be an additional mechanism in place to unfold the proteins before they could make their complexes, and that would require the expenditure of energy. In addition, it has recently been shown that molecular crowding has only a minor effect, if any, on the conformational ensemble of unfolded or partly folded proteins<sup>71-74</sup>. A clue to the possible in vivo mechanism of employment of unstructured proteins comes from the observation that molecular chaperones account for 1-2% of the total protein in the cell. It is our current hypothesis that the interactions of molecular chaperones may account for the prevalence of facile interactions of unfolded protein domains, by stabilizing these domains in interaction-competent states ready for complex formation when the partner is available. This mechanism has already

been invoked to explain the role of the chaperone Hsp90 in the stabilization of the ligand-binding domains of nuclear hormone receptors in the absence of the hormone signal<sup>75</sup>. It turns out that almost nothing is known about the structures of these and other so-called “client proteins” when bound to chaperones. This question forms the basis for a great deal of ongoing work in my lab, to be described in the following paragraphs.

#### 4.7.1 Domains of Chaperones and Co-Chaperones

As a beginning to our studies of the interactions of client proteins with chaperone domains, we began by selecting from literature reports the peptide- or protein-binding domains of chaperone proteins. Only recently has structural information become available on important chaperone proteins such as Hsp90, in part because of its large size and probable fluxional nature. A structure of full-length yeast Hsp90 with its co-chaperone Sba1 bound was recently published<sup>76</sup>, as well as an image-reconstruction effort was made to elucidate the structure of Hsp90 bound to the client protein Cdc37<sup>77</sup>. Our lab began our investigations into chaperone-client protein interactions by dissecting several chaperone molecules into domains. One of these chaperones is the *E. coli* chaperone DnaJ, which contains a cysteine-rich domain of unusual structure<sup>78</sup> which was thought to be involved in the binding of unfolded proteins. It is a measure of the difficult nature of these studies that we were unable to find any protein, unfolded or folded, to which this domain would bind. We endeavoured to lengthen the construct to include the domain C-terminal to the Cys-rich domain, but this construct was insoluble. The answer to this question came from a crystal structure published in another lab, illustrated in Figure 4.16A: Part of the C-terminal domain consists of a sequence located N-terminal to the Cys-rich domain. Once this sequence was included in the construct, it was able to fold correctly.

Another chaperone domain that appeared to be the operative peptide binding domain was the C-terminal domain of the ribosome-bound chaperone trigger factor. Interestingly, like the C-terminal domain of DnaJ, the structure of the C-terminal domain included a sequence that was located N-terminal to an intercalated folded domain, in this case, a peptidyl-prolyl  $\alpha$ -immerse. For this domain, two quite different crystal structures were published. We used NMR measurements to distinguish between the two structures, and to determine which, if either, was the form present in solution. A paper has been submitted on this work, illustrated in Figure 4.16B.

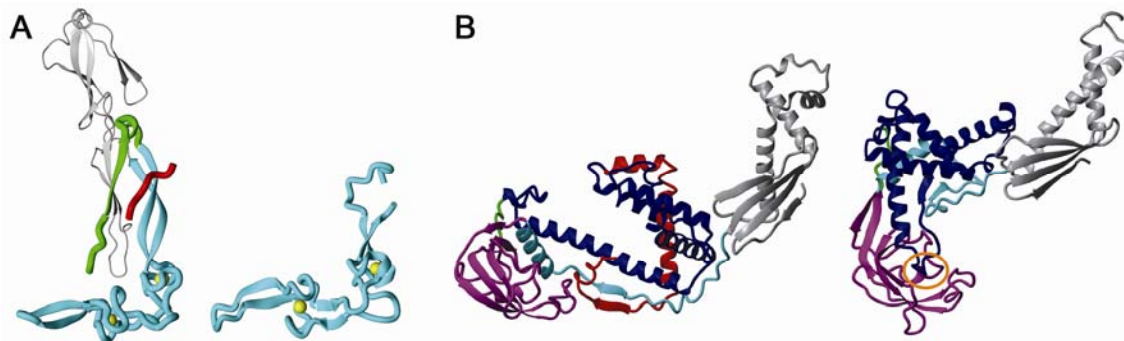


Figure 4.16 A. (left) Crystal structure of the *Saccharomyces cerevisiae* Hsp40 protein Ydj1, in complex with a 7-residue peptide<sup>79</sup>. (right) Average NMR structure of the cysteine-rich domain of *E. coli* Hsp40 protein DnaJ<sup>78</sup>. The zinc atoms are shown in yellow, the backbone of the cysteine-rich domain in blue, and the C-terminal sequence (residues 220-360) in gray. The sequence N-terminal to the cysteine-rich domain that participates in the structure is shown in green. The bound peptide is shown in red. B. (left) *E. coli* trigger factor from the structure of a complex with the ribosome<sup>80</sup> (right) *Vibrio cholerae* trigger factor<sup>81</sup> The ribosomal binding domain (residues 1-111) is in

grey, the PPIase domain (residues 151-242) is in magenta. The apparent linker sequences between the C-terminal and PPIase domains are in green. The sequence between residues 112-148, which connects the ribosomal binding domain and the PPIase domain, and which is folded as part of the C-terminal domain, is shown in cyan. The C-terminal domain (residues 250-432) is shown in blue. The final 53 residues of the C-terminal domain, missing in the *V. cholerae* structure, are shown in red.

#### 4.7.2 A Redox-Regulated Chaperone

We became intrigued by a report of a chaperone protein that responded to oxidative stress. The bacterial heat-shock protein Hsp33 is regulated both at the transcriptional level and post-translationally, through the operation of an unusual redox switch<sup>82</sup>. Under normal oxygen conditions in the cell, the Hsp33 is reduced and monomeric, and binds one equivalent of zinc to form a compact, folded domain<sup>83</sup>. Under oxidative stress conditions, the protein is activated to form a dimer. The crystal structures of Hsp33<sup>84,85</sup> did not show density for the C-terminal sequence, which is the site of redox regulation. We undertook the NMR characterization of the zinc-dependent structure of the C-terminal domain, and were able to show that it forms a compact domain with a novel fold<sup>86</sup> (Figure 4.17A). The mechanism of activation thus occurs in several stages, firstly loss of zinc, then disulfide formation between the cysteines of the C-terminal domain, followed by dimerization<sup>86</sup>.

#### 4.7.3 A Folding Switch – Quorum Sensing in Bacteria

Many species of bacteria sense the presence of members of their own or other bacterial species through the secretion and detection of small molecule communication chemicals known as quorum sensing factors. This communication system was first described for the luminescent bacteria of the genus *Vibrio* that exist in the deep ocean<sup>87</sup>, and consists of a recognizable group of genes, one of which codes for an enzyme that synthesizes the quorum-sensing factor or autoinducer (frequently an acyl homoserine lactone) and one that codes for the protein that detects the quorum-sensing factor. Work on the quorum sensing system of the plant pathogen *Agrobacterium tumefaciens* revealed that the quorum sensing works through the operation of a folding switch<sup>88</sup>: in the presence of the autoinducer, the detector protein is able to fold correctly and thereafter proceed to act as a transcriptional activator for a series of genes. In the absence of autoinducer, the protein is unstable and is degraded. We became interested in this system primarily because of the intriguing notion of a folding switch, but also because the common bacterium *E. coli* appears to be anomalous: its genome contains a gene homologous to the detection proteins of *Vibrio* and *Agrobacterium*, but is missing the autoinducer synthetase. That is, if it is at all analogous to those previously described, the *E. coli* quorum sensing system must be geared towards the sensing of autoinducers from other species of bacteria. The NMR solution structure of *E. coli* SdiA, the quorum sensing protein<sup>89</sup>, shows a single-domain protein that is highly homologous to one of the monomer units of the *Agrobacterium* TraR structure<sup>90,91</sup>. However, it is clear from the NMR spectrum as well as the structure that the binding site for the acyl homoserine lactone autoinducer is quite large for SdiA, and as a result, the structure of the autoinducer within the protein is highly heterogeneous, consistent with the idea that SdiA may represent a quorum sensing protein that responds to many different acyl homoserine lactone autoinducers. The structure of SdiA is shown in Figure 4.17B.

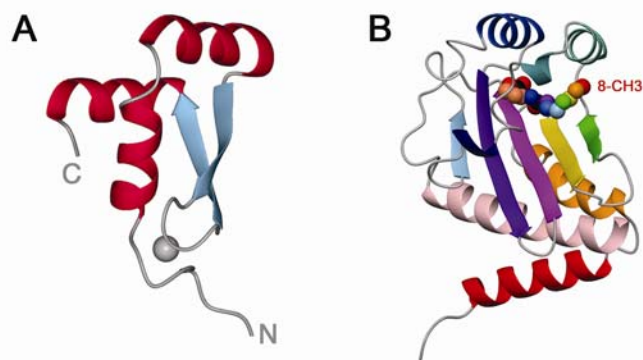


Figure 4.17 A. Representative solution structure of the structured portion of the C-terminal domain of Hsp33, bound to zinc (grey sphere)<sup>86</sup>. B. Representative solution structure of SdiA complexed with the autoinducer N-octanoyl-homoserine lactone (heavy atoms shown in space-filling representation<sup>89</sup>).

#### 4.7.4 Towards the Elucidation of Client Protein Structure

It is a challenge to work with chaperones and their client proteins, not least because we have little idea of the structural states of client proteins in their chaperone complexes. Are they folded, unfolded, partly folded, misfolded? Also, we have very little idea what effect the binding of a client protein has on the chaperone, and what is the function of the co-chaperones, which appear to be required for client specificity. We have made some small steps in the direction of an NMR characterization of the client-protein interaction. So far, we have mainly established solution conditions and explored labeling schemes and types of NMR experiments that could be used. We were able to demonstrate the location of the Hsp90 binding site on the co-chaperone p23, using strategies that we are hopeful will also work in larger systems<sup>92</sup>; it was gratifying that the conclusions that we came to in this paper were consistent with the X-ray structure that was subsequently published<sup>76</sup>. This work is ongoing.

#### Reference List: Chapter 4

1. Boesch, C., Bundi, A., Oppliger, M., & Wüthrich, K. (1978). <sup>1</sup>H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur. J. Biochem.* **91**, 209-214.
2. Daniels, A. J., Williams, R. J. P., & Wright, P. E. (1978). The character of the stored molecules in chromaffin granules of the adrenal medulla: a nuclear magnetic resonance study. *Neuroscience* **3**, 573-585.
3. Kriwacki, R. W., Wu, J., Siuzdak, G., & Wright, P. E. (1996). Probing protein/protein interactions with mass spectrometry and isotopic labeling: analysis of the p21/Cdk2 complex. *J. Am. Chem. Soc.* **118**, 5320-5321.
4. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., & Wright, P. E. (1996). Structural studies of p21<sup>Waf1/Cip1/Sdi1</sup> in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. USA* **93**, 11504-11509.
5. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54-60.
6. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607-3622.
7. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol* **6**, 197-208.

8. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321-331.
9. Rosenfeld, M. G., Lunyak, V. V., & Glass, C. K. (2006). Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes and Development* **20**, 1405-1428.
10. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., & Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **11**, 161-171.
11. Chrivia, J. C., Kwok, R. P., Lamb, N., Hagiwara, M., Montminy, M., & Goodman, R. H. (1993). Phosphorylated CREB binds specifically to nuclear protein CBP. *Nature* **365**, 855-859.
12. Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., & Wright, P. E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:Coactivator interactions. *Cell* **91**, 741-752.
13. Radhakrishnan, I., Pérez-Alvarado, G. C., Dyson, H. J., & Wright, P. E. (1998). Conformational preferences in the Ser<sup>133</sup>-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Letters* **430**, 317-322.
14. Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., & Wright, P. E. (1999). Structural analyses of CREB-CBP transcriptional activator-coactivator complexes by NMR spectroscopy: Implications for mapping the boundaries of structural domains. *J. Mol. Biol.* **287**, 859-865.
15. Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R., & Wright, P. E. (2002). Roles of Phosphorylation and Helix Propensity in the Binding of the KIX Domain of CREB-binding Protein by Constitutive (c-Myb) and Inducible (CREB) Activators. *J. Biol. Chem.* **277**, 42241-42248.
16. Parker, D., Rivera, M., Zor, T., Henrion-Caude, A., Radhakrishnan, I., Kumar, A., Shapiro, L. H., Wright, P. E., Montminy, M., & Brindle, P. K. (1999). Role of secondary structure in discrimination between constitutive and inducible activators. *Mol. Cell Biol.* **19**, 5601-5607.
17. Zor, T., De Guzman, R. N., Dyson, H. J., & Wright, P. E. (2004). Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* **337**, 521-534.
18. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2002). Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the Kix domain. *J. Biol. Chem.* **277**, 43168-43174.
19. De Guzman, R. N., Goto, N. K., Dyson, H. J., & Wright, P. E. (2006). Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J. Mol. Biol.* **355**, 1005-1013.
20. Sugase, K., Dyson, H. J., & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* (in press).
21. De Guzman, R. N., Liu, H. Y., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2000). Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. *J. Mol. Biol.* **303**, 243-253.
22. De Guzman, R. N., Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2005). CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry* **44**, 490-497.
23. Legge, G. B., Martinez-Yamout, M. A., Hambly, D. M., Trinh, T., Lee, B. M., Dyson, H. J., & Wright, P. E. (2004). ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. *J. Mol. Biol.* **343**, 1081-1093.
24. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2000). Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J. Mol. Biol.* **304**, 355-370.
25. Pascual, J., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2000). Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J. Mol. Biol.* **304**, 723-729.
26. Hirota, K. & Semenza, G. L. (2005). Regulation of hypoxia-inducible factor 1 by prolyl and asparaginyl hydroxylases. *Biochem. Biophys. Res. Commun.*
27. Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J., & Wright, P. E. (2002). Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc. Natl. Acad. Sci. U. S. A* **99**, 5271-5276.
28. Freedman, S. J., Sun, Z. Y., Poy, F., Kung, A. L., Livingston, D. M., Wagner, G., & Eck, M. J. (2002). Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1alpha. *Proc. Natl. Acad. Sci. USA* **99**, 5367-5372.

29. Gunasekaran, K., Tsai, C. J., Kumar, S., Zanuy, D., & Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.* **28**, 81-85.
30. Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M., & Ryu, S. E. (2003). Structure of Human FIH-1 Reveals a Unique Active Site Pocket and Interaction Sites for HIF-1 and von Hippel-Lindau. *J. Biol. Chem.* **278**, 7558-7563.
31. De Guzman, R. N., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2004). Interaction of the TAZ1 Domain of CREB-Binding Protein with the Activation Domain of CITED2: Regulation by Competition between Intrinsically Unstructured Ligands for Non-Identical Binding Sites. *J. Biol. Chem.* **279**, 3042-3049.
32. Freedman, S. J., Sun, Z. Y., Kung, A. L., France, D. S., Wagner, G., & Eck, M. J. (2003). Structural basis for negative regulation of hypoxia-inducible factor-1alpha by CITED2. *Nat. Struct. Biol.* **10**, 504-512.
33. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., & Wright, P. E. (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549-553.
34. Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., & Pavletich, N. P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948-953.
35. Grossman, S. R., Perez, M., Kung, A. L., Joseph, M., Mansur, C., Xiao, Z. X., Kumar, S., Howley, P. M., & Livingston, D. M. (1998). p300/MDM2 complexes participate in MDM2-mediated p53 degradation. *Mol. Cell* **2**, 405-415.
36. Matt, T., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2004). The CBP/p300 TAZ1 domain in its native state is not a binding partner of MDM2. *Biochem. J.* **381**, 685-691.
37. Yu, G. W., Allen, M. D., Andreeva, A., Fersht, A. R., & Bycroft, M. (2006). Solution structure of the C4 zinc finger domain of HDM2. *Protein Sci.* **15**, 384-389.
38. Kostic, M., Matt, T., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2006). Solution Structure of the Hdm2 C2H2C4 RING, a Domain Critical for Ubiquitination of p53. *J. Mol. Biol.* **363**, 433-450.
39. Zheng, N., Wang, P., Jeffrey, P. D., & Pavletich, N. P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin- protein ligases. *Cell* **102**, 533-539.
40. Özkan, E., Yu, H., & Deisenhofer, J. (2005). Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl. Acad. Sci. USA* **102**, 18890-18895.
41. Lee, M. S., Gippert, G., Soman, K. Y., Case, D. A., & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger binding domain. *Science* **245**, 635-637.
42. Mangelsdorf, D. J. & Evans, R. M. (1995). The RXR heterodimers and orphan receptors. *Cell* **83**, 841-850.
43. Holmbeck, S. M. A., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J., & Wright, P. E. (1998). High resolution solution structure of the retinoid X receptor DNA binding domain. *J. Mol. Biol.* **281**, 271-284.
44. Holmbeck, S. M. A., Dyson, H. J., & Wright, P. E. (1998). DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J. Mol. Biol.* **284**, 533-539.
45. Gearhart, M. D., Holmbeck, S. M. A., Evans, R. M., Dyson, H. J., & Wright, P. E. (2003). Monomeric complex of human orphan estrogen related receptor-2 with DNA: A pseudo-dimer interface mediates extended half-site recognition. *J. Mol. Biol.* **327**, 819-832.
46. Clemens, K. R., Liao, X., Wolf, V., Wright, P. E., & Gottesfeld, J. M. (1992). Definition of the binding sites of individual zinc fingers in the TFIIIA-5S RNA gene complex. *Proc. Natl. Acad. Sci. USA* **89**, 10822-10826.
47. Clemens, K. R., Zhang, P., Liao, X., McBryant, S. J., Wright, P. E., & Gottesfeld, J. M. (1994). Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J. Mol. Biol.* **244**, 23-35.
48. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., & Wright, P. E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nature Struct. Biol.* **4**, 605-608.
49. Liao, X., Clemens, K. R., Tennant, L., Wright, P. E., & Gottesfeld, J. M. (1992). Specific interaction of the first three zinc fingers of TFIIIA with the internal control region of the *Xenopus* 5 S RNA gene. *J. Mol. Biol.* **223**, 857-871.
50. Neely, L. S., Lee, B. M., Xu, J., Wright, P. E., & Gottesfeld, J. M. (1999). Identification of a minimal domain of 5 S ribosomal RNA sufficient for high affinity interactions with the RNA-specific zinc fingers of transcription factor IIIA. *J. Mol. Biol.* **291**, 549-560.



51. Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M., & Wright, P. E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: Determinants of affinity and sequence specificity. *J. Mol. Biol.* **273**, 183-206.
52. Bruschiweiler, R., Liao, X., & Wright, P. E. (1995). Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886-889.
53. Tsui, V., Zhu, L., Huang, T. H., Wright, P. E., & Case, D. A. (2000). Assessment of zinc finger orientations by residual dipolar coupling constants. *J. Biomol. NMR* **16**, 9-21.
54. Pavletich, N. P. & Pabo, C. O. (1991). Zinc-finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809-817.
55. Laity, J. H., Chung, J., Dyson, H. J., & Wright, P. E. (2000). Alternative splicing of Wilms' tumor suppressor protein modulates DNA binding activity through isoform-specific DNA-induced conformational changes. *Biochemistry* **39**, 5341-5348.
56. Laity, J. H., Dyson, H. J., & Wright, P. E. (2000). Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proc. Natl. Acad. Sci. USA* **97**, 11932-11935.
57. Laity, J. H., Dyson, H. J., & Wright, P. E. (2000). DNA-induced  $\alpha$ -helix capping in conserved linker sequences is a determinant of binding affinity in Cys<sub>2</sub>-His<sub>2</sub> zinc fingers. *J. Mol. Biol.* **295**, 719-727.
58. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2004). Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.* **11**, 257-264.
59. Lee, B. M., Xu, J., Clarkson, B. K., Martinez-Yamout, M. A., Dyson, H. J., Case, D. A., Gottesfeld, J. M., & Wright, P. E. (2006). Induced Fit and "Lock and Key" Recognition of 5 S RNA by Zinc Fingers of Transcription Factor IIIA. *J. Mol. Biol.* **357**, 275-291.
60. Möller, H. M., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2005). Solution structure of the N-terminal zinc fingers of the *Xenopus laevis* double-stranded RNA-binding protein ZFa. *J. Mol. Biol.* **351**, 718-730.
61. Sheng, G., dos Reis, M., & Stern, C. D. (2003). Churchill, a zinc finger transcriptional activator, regulates the transition between gastrulation and neurulation. *Cell* **115**, 603-613.
62. Lee, B. M., Buck-Koehntop, B. A., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2007). Embryonic neural inducing factor Churchill is not a DNA-binding zinc finger protein: solution structure reveals a solvent-exposed  $\beta$ -sheet and zinc binuclear cluster. *J. Mol. Biol.* (in press).
63. Love, J. J., Li, X., Chung, J., Dyson, H. J., & Wright, P. E. (2004). The LEF-1 HMG domain undergoes a disorder-to-order transition upon complex formation with cognate DNA. *Biochemistry* **43**, 8725-8734.
64. Perez-Alvarado, G. C., Martinez-Yamout, M., Allen, M. M., Grosschedl, R., Dyson, H. J., & Wright, P. E. (2003). Structure of the nuclear factor ALY: insights into post-transcriptional regulatory and mRNA nuclear export processes. *Biochemistry* **42**, 7348-7357.
65. Pérez-Alvarado, G. C., Munnerlyn, A., Dyson, H. J., Grosschedl, R., & Wright, P. E. (2000). Identification of the regions involved in DNA binding by the mouse PEBP2 $\alpha$  protein. *FEBS Letters* **470**, 125-130.
66. Love, J. J., Li, X. A., Case, D. A., Giese, K., Grosschedl, R., & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-795.
67. Lin, C. H., Hare, B. J., Wagner, G., Harrison, S. C., Maniatis, T., & Fraenkel, E. (2001). A small domain of cbp/p300 binds diverse proteins. solution structure and functional studies. *Mol. Cell* **8**, 581-590.
68. Brüschweiler, R., Liao, X., & Wright, P. E. (1995). Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886-889.
69. Gross, J. D., Moerke, N. J., von der, H. T., Lugovskoy, A. A., Sachs, A. B., McCarthy, J. E., & Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* **115**, 739-750.
70. Koradi, R., Billeter, M., & Wüthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51-55.
71. Flaugh, S. L. & Lumb, K. J. (2001). Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules* **2**, 538-540.
72. McPhie, P., Ni, Y. s., & Minton, A. P. (2006). Macromolecular Crowding Stabilizes the Molten Globule Form of Apomyoglobin with Respect to Both Cold and Heat Unfolding. *J. Mol. Biol.* **361**, 7-10.
73. Qu, Y. & Bolen, D. W. (2002). Efficacy of macromolecular crowding in forcing proteins to fold. *Biophys. Chem.* **101-102**, 155-165.

74. van den, B. B., Ellis, R. J., & Dobson, C. M. (1999). Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* **18**, 6927-6933.
75. Pratt, W. B. & Toft, D. O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol Med. (Maywood.)* **228**, 111-133.
76. Ali, M. M., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C., & Pearl, L. H. (2006). Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013-1017.
77. Vaughan, C. K., Gohlke, U., Sobott, F., Good, V. M., Ali, M. M., Prodromou, C., Robinson, C. V., Saibil, H. R., & Pearl, L. H. (2006). Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol Cell* **23**, 697-707.
78. Martinez-Yamout, M., Legge, G. B., Zhang, O. W., Wright, P. E., & Dyson, H. J. (2000). Solution structure of the cysteine-rich domain of the *Escherichia coli* chaperone protein DnaJ. *J. Mol. Biol.* **300**, 805-818.
79. Li, J., Qian, X., & Sha, B. (2003). The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate. *Structure. (Camb.)* **11**, 1475-1483.
80. Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., & Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590-596.
81. Ludlam, A. V., Moore, B. A., & Xu, Z. (2004). The crystal structure of ribosomal chaperone trigger factor from *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A* **101**, 13436-13441.
82. Jakob, U., Muse, W., Eser, M., & Bardwell, J. C. A. (1999). Chaperone activity with a redox switch. *Cell* **96**, 341-352.
83. Graf, P. C., Martinez-Yamout, M., VanHaerents, S., Lilie, H., Dyson, H. J., & Jakob, U. (2004). Activation of the redox-regulated chaperone Hsp33 by domain unfolding. *J. Biol. Chem.* **279**, 20529-20538.
84. Vijayalakshmi, J., Mukherjee, M. K., Graumann, J., Jakob, U., & Saper, M. A. (2001). The 2.2 Å crystal structure of Hsp33. A heat shock protein with redox-regulated chaperone activity. *Structure* **9**, 367-375.
85. Kim, S. J., Jeong, D. G., Chi, S. W., Lee, J. S., & Ryu, S. E. (2001). Crystal structure of proteolytic fragments of the redox-sensitive Hsp33 with constitutive chaperone activity. *Nat. Struct. Biol.* **8**, 459-466.
86. Won, H. S., Low, L. Y., Guzman, R. D., Martinez-Yamout, M., Jakob, U., & Dyson, H. J. (2004). The zinc-dependent redox switch domain of the chaperone Hsp33 has a novel fold. *J Mol Biol* **341**, 893-899.
87. Nealson, K. H. & Hastings, J. W. (1979). Bacterial bioluminescence: its control and ecological significance. *Microbiol. Rev.* **43**, 496-518.
88. Zhu, J. & Winans, S. C. (2001). The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proc. Natl. Acad. Sci. USA* **98**, 1507-1512.
89. Yao, Y., Martinez-Yamout, M. A., Dickerson, T. J., Brogan, A. P., Wright, P. E., & Dyson, H. J. (2006). Structure of the *Escherichia coli* quorum sensing protein SdiA: activation of the folding switch by acyl homoserine lactones. *J. Mol. Biol.* **355**, 262-273.
90. Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R., Neddermann, P., & Marco, S. D. (2002). The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* **21**, 4393-4401.
91. Zhang, R. G., Pappas, T., Brace, J. L., Miller, P. C., Oulmassov, T., Molyneaux, J. M., Anderson, J. C., Bashkin, J. K., Winans, S. C., & Joachimiak, A. (2002). Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. *Nature* **417**, 971-974.
92. Martinez-Yamout, M. A., Venkitakrishnan, R. P., Preece, N. E., Kroon, G., Wright, P. E., & Dyson, H. J. (2006). Localization of sites of interaction between p23 and Hsp90 in solution. *J Biol. Chem.* **281**, 14457-14464.

## **CHAPTER 5. TECHNICAL ADVANCES**

Our lab has traditionally focused on the application of new technology to a series of biological problems (Chapters 1-4), and we have not generally made a point of undertaking technical innovations for their own sake. However, in several instances, the presence of particular problems were solved by technical innovations within the lab, and these are briefly described here.

### **5.1 Simulation of 2D and 3D NOESY Spectra**

One of the major absences in the determination of 3-dimensional structures by NMR methods is the lack of a calculation of the fit of the calculated model to the experimental data. X-ray crystallographers readily calculate the “R-factor”, an index of the reflections expected from a given structural model compared with the observed reflections. Although NMR methods of structure calculation use experimental data to derive structural models, it has been difficult to formulate a valid “NMR R-factor”, not least because of the wide variety of types of experimental data that are used. The most important, as well as the most numerous, data used for NMR structure calculations are the NOEs, the matrix of short distances that arise as a result of the three-dimensional structure and which are measured in NOESY spectra. The NOE is highly distance dependent (proportional to  $r^{-6}$ ).

The first step in the derivation of an NMR R-factor is the calculation of a theoretical NOESY spectrum from the calculated structures. The program SPIRIT<sup>1</sup> was written in our lab by a postdoc, Leiming Zhu, and the program was tested on the thioredoxin structures (see chapter 2), for which we had excellent NMR data and highly refined structures of two highly similar forms. The program works well, and in fact was able to distinguish on the basis of calculations for a structural ensemble rather than for single structures, that there were slight differences in the solution structures of thioredoxin, reported by the experimental NOESY spectrum, which were inconsistent with the published X-ray crystal structure of oxidized thioredoxin.

Although the SPIRIT program was published several years ago, we have not succeeded in attracting anyone to work on the fascinating problem of coding for an NMR R-factor, or, even more exciting, for a direct refinement method for solution structures, utilizing the experimental NOESY spectrum. Such advances will surely occur in future.

### **5.2 Development of an Orienting Medium for Use at Low pH**

The NOE distance restraints and coupling-constant-derived dihedral angle restraints that are the common data used to calculate solution structures are all very localized: because of its  $r^{-6}$  distance dependence, the NOE does not report on inter-atom distances longer than about 6 Å, while coupling constant data are only available over at maximum 4-5 connected atoms. A new method developed in the late 90s in the Bax laboratory utilizes one of the major parameters measured in solid state NMR experiments, dipolar coupling. In a solid, all of the atoms are coupled, through dipolar coupling. For a macromolecule, this means that, without sophisticated spinning techniques and pulse selection, the resonances are extremely broad and are all present at nearly the same chemical shift – no structural data can be obtained. In solution, the dipolar coupling does not interfere, because its effect is averaged out by the isotropic tumbling of the molecules. However, in a liquid crystal or other anisotropic medium, the environment is somewhere between a solid and a liquid, and a small “residual dipolar coupling” can be measured<sup>2</sup>. In practice, the RDC takes the form of a vector calculated for each of the relevant bonds (e.g. N-H bonds in a <sup>15</sup>N experiment) relative to the overall diffusion tensor of the

molecule. This parameter has advantages, because it is non-local. That is, the vector orientation of a given bond is measured independently of all the other bond orientations. The analysis of all of the bond orientation vectors can give useful structural information, though it has not proved possible to calculate structures from RDCs alone. The RDC is especially useful for determining the relative orientations of independent domains of a single molecule, which may not be well-defined by NOEs.

We were (and remain) very interested in defining the structural characteristics of the equilibrium molten globule intermediate of apomyoglobin (see Chapter 1). Despite much work, neither we nor any other lab has been able to observe long-range NOEs in intermediates and unfolded states. This is apparently due on the one hand to the heterogeneity of the conformational ensemble in these systems and on the other hand to the broadening of the resonances that occurs even in the most favorable cases. We viewed the RDC as a promising means of obtaining long-range structural information on the molten globule state of apomyoglobin. However, at the time, the only liquid crystal orienting medium involved the use of so-called “bicelles”, which are formed using two different sizes of detergent molecules<sup>3</sup>. The detergents that were used were unstable in acidic solution, which is the medium where the apomyoglobin molten globule is formed. We therefore set about developing a new bicelle medium for use in acidic media, employing non-ionic detergents<sup>4</sup>. These bicelles proved to be excellent for RDC measurements for the acidophilic protein rusticyanin (see Chapter 2), but were no good at all for the apomyoglobin molten globule. The RDC measurement relies on transient contacts between the sample molecules and the orienting medium, giving a very slight orientation or anisotropy in the solution. If instead of transient contact there is more substantial contact or binding to the bicelles, the NMR spectrum of the sample molecules disappears, broadened beyond detection by the large size of the bicelles. It turned out that the apomyoglobin molten globule had an excellent affinity for the low-pH bicelles, which is perhaps not surprising since they contain long-chain fatty acids and the molten globule has a non-globular structure with exposed hydrophobic groups. In the meantime, other groups were pursuing different types of orientational media, including filamentous bacteriophages<sup>5</sup> and stressed polyacrylamide gels<sup>6</sup>. In the end, we were able to use the polyacrylamide gel technique to obtain information on the unfolded states of apomyoglobin<sup>7</sup>, in the process providing a new polymer-based analysis of the RDC data obtained for unfolded proteins. Unfortunately, the molten globule state continues to present difficulties, although with the use of spin labels in conjunction with RDC and mutagenesis data (see Chapter 1), we are now beginning to obtain a picture of the structural features of this intermediate.

### **5.3 Structure Determination Techniques**

Our lab has been heavily involved in the determination of NMR structures, and we have pioneered a number of techniques, including the use of the powerful molecular dynamics program AMBER<sup>8</sup>, with the collaboration of our colleague David Case. In general, the advances that we have reported in structure determination techniques have formed part of reports of the structures themselves. This is particularly true of our earliest structures, which in many ways were reports of the methods used to obtain them as much as of the structures themselves<sup>9-11</sup>. Nevertheless, we have reported several techniques that have aided in structure determination in a more general way.

### 5.3.1 Structure-Assisted NOE Evaluation (SANE)

During the course of the structure determination of the LEF-1 I domain<sup>12</sup>, it became clear that a published structure of a similar, though not identical, protein could be utilized during the tedious NOE assignment phase of the NMR structure calculation to speed up the process. The technique, SANE, employs a number of selection criteria and a protocol of gradually increasing restraint to iterate into the correct structure<sup>13</sup>. The process is illustrated in Figure 5.1.

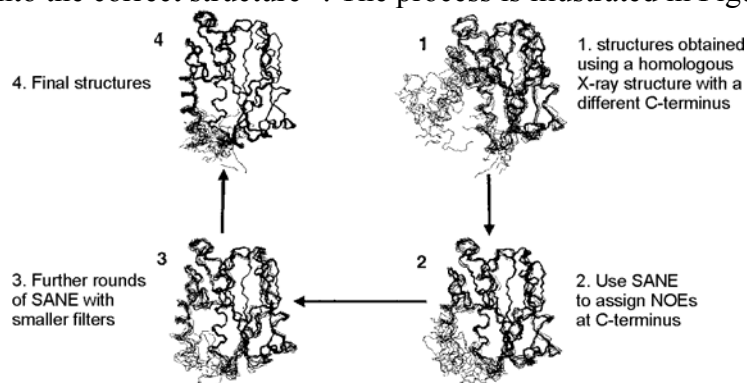


Figure 5.1 Sets of structures of the LFA-1 I-domain showing the intermediate steps in the usage of SANE for structure determination<sup>13</sup>.

### 5.3.2 Inclusion of Solvent Water in NMR Structure Calculations

Because of the large amount of data that is accumulated during a structure calculation, it has been impossible, within the limits of the available computing power, to calculate structures with explicit water solvent. We evaluated a Born continuum method developed in the lab of David Case to account for solvent water without the necessity for a full explicit solvent calculation<sup>14</sup>. Our comparison showed that the quality of the structures obtained using the Born solvent model were equal in every way to those obtained with the much more lengthy and expensive explicit solvent model, and we have since incorporated a final Born solvent calculation as part of our refinement protocol. For well-determined and exhaustively refined structures such as those of Grx2<sup>15</sup> (see Chapter 2), the improvements can be quite subtle, but in other cases, where the NMR data are not so complete, anomalous backbone conformations such as  $\gamma$ -turns can occur in the calculated structures. We observe that these backbone distortions are removed by refinement with the Born solvent model (for example<sup>16</sup>), which clearly increases structure accuracy as well as accelerating the process of structure refinement.

### 5.3.3 Provision of Good Starting Models

The use of SANE allows the structure calculation process to proceed more rapidly than a “bootstrap” approach, but it does require the presence of a relatively realistic starting model, otherwise the calculations may never converge, or may converge to an incorrect structure. The provision of a good starting model from a small subset of the NMR data was the subject of a recent collaboration with the laboratory of Charles Brooks<sup>17</sup>. In this case, a limited set of data from the spectra of the zinc-bound C-terminal domain of Hsp33<sup>18</sup> was used “blind” to generate starting structures for SANE refinement using the full data set. The structures calculated using the “limited-set” starting structures refined to the same structure as that calculated by the more laborious classical methods.

## Reference List: Chapter 5

1. Zhu, L., Dyson, H. J., & Wright, P. E. (1998). A NOESY-HSQC simulation program, SPIRIT. *J. Biomol. NMR* **11**, 17-29.
2. Bax, A. & Tjandra, N. (1997). High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *J. Biomol. NMR* **10**, 289-292.
3. Sanders, C. R., II & Schwonek, J. P. (1992). Characterization of Magnetically Orientable Bilayers in Mixtures of Dihexanoylphosphatidylcholine and Dimyristoylphosphatidylcholine by Solid-State NMR. *Biochemistry* **31**, 8898-8905.
4. Cavagnero, S., Dyson, H. J., & Wright, P. E. (1999). Improved low pH bicelle system for orienting macromolecules over a wide temperature range. *J. Biomol. NMR* **13**, 387-391.
5. Hansen, M. R., Mueller, L., & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5**, 1065-1074.
6. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T., & Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* **18**, 303-309.
7. Mohana-Borges, R., Goto, N. K., Kroon, G. J. A., Dyson, H. J., & Wright, P. E. (2004). Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* **340**, 1131-1142.
8. Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M. J., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688.
9. Moore, J. M., Case, D. A., Chazin, W. J., Gippert, G. P., Havel, T. F., Powls, R., & Wright, P. E. (1988). Three-dimensional solution structure of plastocyanin from the green alga *Scenedesmus obliquus*. *Science* **240**, 314-317.
10. Lee, M. S., Gippert, G., Soman, K. Y., Case, D. A., & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger binding domain. *Science* **245**, 635-637.
11. Dyson, H. J., Gippert, G. P., Case, D. A., Holmgren, A., & Wright, P. E. (1990). Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**, 4129-4136.
12. Legge, G. B., Kriwacki, R. W., Chung, J., Hommel, U., Ramage, P., Case, D. A., Dyson, H. J., & Wright, P. E. (1999). NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. *J. Mol. Biol.* **295**, 1251-1264.
13. Duggan, B. M., Legge, G. B., Dyson, H. J., & Wright, P. E. (2001). SANE (Structure assisted NOE evaluation): an automated model-based approach for NOE assignment. *J. Biomol. NMR* **19**, 321-329.
14. Xia, B., Tsui, V., Case, D. A., Dyson, H. J., & Wright, P. E. (2002). Comparison of solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model and with explicit water. *J. Biomol. NMR* **22**, 317-331.
15. Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, P. E., & Dyson, H. J. (2001). Solution structure of *Escherichia coli* glutaredoxin 2 shows similarity to mammalian glutathione-S-transferases. *J. Mol. Biol.* **310**, 907-918.
16. Perez-Alvarado, G. C., Martinez-Yamout, M., Allen, M. M., Grosschedl, R., Dyson, H. J., & Wright, P. E. (2003). Structure of the nuclear factor ALY: insights into post-transcriptional regulatory and mRNA nuclear export processes. *Biochemistry* **42**, 7348-7357.
17. Chen, J., Won, H. S., Im, W., Dyson, H. J., & Brooks, C. L., III (2005). Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *J. Biomol. NMR* **31**, 59-64.
18. Won, H. S., Low, L. Y., Guzman, R. D., Martinez-Yamout, M., Jakob, U., & Dyson, H. J. (2004). The zinc-dependent redox switch domain of the chaperone Hsp33 has a novel fold. *J. Mol. Biol.* **341**, 893-899.

## List of Publications

[annotated with a description of my contribution]

(citations from Science Citation Index as of August 28 2007)

1. Characteristics of a *Bacillus subtilis* W23 mutant temperature sensitive for initiation of chromosome replication. P. Upcroft, H.J. Dyson, & R.G. Wake (1975) *J. Bacteriol.* **121**, 121-127. [This was my Honors project. I performed some of the experimental work.] (9 citations)
2. Spin state and unfolding equilibria of ferricytochrome c in acidic solutions. H.J. Dyson & J.K. Beattie (1982) *J. Biol. Chem.* **257**, 2267-2273. [This was my Ph.D. project. I designed and performed the experiments and wrote the first draft of the paper.] (82 citations)
3. The immunodominant site of a synthetic immunogen has a conformational preference in water for a Type-II reverse turn. H.J. Dyson, K.J. Cross, R.A. Houghten, I.A. Wilson, P.E. Wright, & R.A. Lerner (1985) *Nature* **318**, 480-483. [This work marks the beginning of my major research. I designed and performed most of the experiments and wrote the paper. Cross performed some of the early experiments; the peptides were synthesized by Houghten and Wilson; Wright and Lerner helped design the experiments and edited the paper.] (226 citations)
4. Antipeptide antibodies and the disorder-order phenomenon. P.E. Wright, H.J. Dyson, M. Rance, J. Ostresh, R.A. Houghten, I.A. Wilson, & R.A. Lerner (1985) In: Modern Approaches to Vaccines (R.A. Lerner, R.M. Chanock, & F. Brown Eds.) Cold Spring Harbor Laboratory, pp 15-19. [A book chapter, written by me, recapitulating the material of ref. 3.]
5. Selection by site-directed antibodies of small regions of peptides which are ordered in water. H.J. Dyson, K.J. Cross, J. Ostresh, R.A. Houghten, I.A. Wilson, P.E. Wright, & R.A. Lerner (1986) In: Synthetic Peptides as Antigens, Ciba Foundation Symposium 119 (R. Porter & J. Whelan Eds.) John Wiley & Sons, Chichester. pp 58-75. [A book chapter, written by me, recapitulating the material of ref. 3.] (11 citations)
6. The order-disorder paradox in antigen-antibody union: anti-peptide antibodies as a probe for structured regions of small peptides. H.J. Dyson, M. Rance, R.A. Houghten, P.E. Wright, & R.A. Lerner (1987) In: Biological Organization: Macromolecular Interactions at High Resolution (R.M. Burnett & H.J. Vogel Eds.) Academic Press Inc., London.. pp 227-234. [A book chapter, written by me, recapitulating the material of ref. 3.] (7 citations)
7. Identification of folded structures in immunogenic peptides by 2D NMR Spectroscopy. P.E. Wright, H.J. Dyson, M. Rance, R.A. Houghten, & R.A. Lerner (1987) In: Protides of the Biological Fluids, 35th Colloquium (H. Peeters Ed.) Pergamon, pp 477-480. [A book chapter, written by me, recapitulating the material of ref. 3.]
8. The physical basis for induction of protein-reactive antipeptide antibodies. H.J. Dyson, R.A. Lerner, & P.E. Wright (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 305-324. [An invited review, written by me, following ref. 3.] (118 citations)
9. Folding of immunogenic peptide fragments of proteins in water solution. I Sequence requirements for the formation of a reverse turn. H.J. Dyson, M. Rance, R.A. Houghten, R.A. Lerner, & P.E. Wright (1988) *J. Mol. Biol.* **201**, 161-200. [This paper and the following, ref. 10, mark the next phase of the peptide project. For this paper, I was responsible for virtually all of the experimental work, and most of the data analysis and writing of the paper was done by me. Rance was responsible for technical aspects of NMR. Houghten was responsible for peptide synthesis. Lerner suggested the original project and helped edit the paper. As for most subsequent papers where his name appears, this work was done by NMR in Wright's lab, and he provided major input into all of the experimental design, analysis and write-up.] (597 citations)

10. Folding of immunogenic peptide fragments of proteins in water solution. II The nascent helix. H.J. Dyson, M. Rance, R.A. Houghten, P.E. Wright, & R.A. Lerner (1988) *J. Mol. Biol.* **201**, 201-217. [The nascent helix, a term invented by us to explain the observations reported in this paper, has become generally used to describe the phenomenon of a fluctuating ensemble that samples helix-like conformations. Contributions to this paper were similar to those of ref 9.] (448 citations)
11. Structural differences between oxidized and reduced thioredoxin monitored by two-dimensional <sup>1</sup>H NMR spectroscopy. H.J. Dyson, A. Holmgren, & P.E. Wright (1988) *FEBS Lett.* **228**, 254-258. [This paper marks the beginning of a major new line of research, into the structure and dynamic differences between the oxidized (disulfide) and reduced (dithiol) forms of thioredoxin. For this paper, I did all of the experimental measurements and data analysis, and wrote the paper. Holmgren provided the protein.] (39 citations)
12. Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. P.E. Wright, H.J. Dyson, & R.A. Lerner (1988) *Biochemistry* **27**, 7167-7175. [This major invited review synthesized our thoughts on the role of residual structure in peptides as a paradigm for initiation of protein folding. I was responsible for writing the paper, together with Wright. Lerner was responsible for the immunological aspects.] (455 citations)
13. Assignment of the proton NMR spectrum of reduced and oxidized thioredoxin: sequence-specific assignments, secondary structure and global fold. H.J. Dyson, A. Holmgren, & P.E. Wright (1989) *Biochemistry* **28**, 7074-7087. [This paper represented the “state of the art” in the assignment of protein NMR spectra at the time, and was a major achievement, done without the use of stable isotope labeling. This paper is frequently mentioned as the largest (and last) protein that was assigned by proton 2D methods. For this paper, I did all of the experimental measurements and data analysis, and wrote the paper.] (87 citations)
14. <sup>1</sup>H NMR studies of the solution conformations of an analogue of the C-peptide of ribonuclease A. J.J. Osterhout, Jr., R.L. Baldwin, E.J. York, J.M. Stewart, H.J. Dyson, & P.E. Wright (1989) *Biochemistry* **28**, 7059-7064. [This work was a collaboration between the labs of Baldwin (Stanford University) and Wright. My role was primarily in data analysis and writing of the paper, although the first draft was prepared by Osterhout.] (134 citations)
15. Folding of peptide fragments of proteins in water solution: implications for initiation of protein folding. P.E. Wright, R.A. Lerner, & H.J. Dyson (1989) In: Advances in Protein Design (H. Blöcker, J. Collins, R.D. Schmid, & D. Schomburg Eds.) VCH Publishing, pp 13-19. [This chapter in a conference proceedings recapitulates much of the solution peptide work. I was responsible for the entire paper.]
16. Folding of peptide fragments of proteins in aqueous solution. P.E. Wright, H.J. Dyson, V.A. Feher, L.L. Tennant, J.P. Waltho, R.A. Lerner, & D.A. Case (1990) In: Frontiers of NMR in Molecular Biology, UCLA Symposia in Molecular and Cellular Biology, New Series vol.109 (D. Live, I. Armitage, & D. Patel Eds.) Alan R. Liss, Inc., New York. pp 1-13. [Also a chapter in a conference proceedings. Experimental work was done by Dyson, Feher, Tennant, and Waltho. I was responsible for most of the paper writing.]
17. Folding of peptide fragments of proteins in water solution. P.E. Wright, H.J. Dyson, J.P. Waltho, & R.A. Lerner (1990) In: Protein Folding (L.M. Gierasch & J. King Eds.) AAAS, Washington. pp 95-102. [Book chapter much like ref. 16.]
18. Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. H.J. Dyson, G.P. Gippert, D.A. Case, A. Holmgren, & P.E. Wright (1990) *Biochemistry* **29**, 4129-4136. [I was responsible for all of the experimental work and structure calculations. Gippert and Case were responsible for the technical aspects of the structure calculations, as many of the methods had to be developed at this time. I wrote the paper, Wright edited.] (152 citations)



19. Antigen-antibody interactions: an NMR approach. P.E. Wright, H.J. Dyson, R.A. Lerner, L. Riechmann, & P. Tsang (1990) *Biochem. Pharm.* **40**, 83-88. [Chapter in conference proceedings. Further development of NMR methods for peptides and their interactions with antibody domains. For this paper, I was mostly responsible for the paper writing.] (23 citations)
20. Conformational preferences of synthetic peptides derived from the immunodominant site of the circumsporozoite protein of *Plasmodium falciparum* by <sup>1</sup>H NMR. H.J. Dyson, A.C. Satterthwait, R.A. Lerner, & P.E. Wright (1990) *Biochemistry* **29**, 7828-7837. [Another example of nascent structure in an immunodominant site, in this case from the malaria parasite. Satterthwait and Lerner suggested the protein for study. I did all of the NMR experiments, analyzed the data and wrote the paper.] (56 citations)
21. The conformational restriction of synthetic vaccines for malaria. A.C. Satterthwait, L.-C. Chiang, T. Arrhenius, E. Cabezas, F. Zavala, H.J. Dyson, & P.E. Wright (1990) *Bull. WHO* **68**, 17-25. [An advance on ref. 20 where the peptides were covalently restricted. I had only minor contribution to this paper, making some of the NMR measurements.] (11 citations)
22. Defining solution conformations of small linear peptides. H.J. Dyson & P.E. Wright (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 519-538. [This invited review synthesizes the methods we use to detect residual structure in small peptides in aqueous solution. I wrote the first draft. Wright edited the paper.] (437 citations)
23. Active conformation of an insect neuropeptide family. R.J. Nachman, V.A. Roberts, H.J. Dyson, G.M. Holman, & J.A. Tainer (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4518-4522. [My contribution to this work was fairly minor – I measured and interpreted the NMR spectra of these short peptides.] (62 citations)
24. Proton-transfer effects in the active-site region of *Escherichia coli* thioredoxin using two-dimensional <sup>1</sup>H NMR. H.J. Dyson, L.L. Tennant, & A. Holmgren (1991) *Biochemistry* **30**, 4262-4268. [I did all of the experimental measurements for this paper, analyzed the data and wrote the paper. Tennant and Holmgren were involved in the production of protein. Since I had been promoted to Assistant Professor, I now began publishing independent papers on my own projects, while continuing to collaborate with Wright on other projects.] (53 citations)
25. Mapping the anatomy of the immunodominant domain of the human immunodeficiency virus gp41 transmembrane protein: peptide conformation analysis using monoclonal antibodies and proton nuclear magnetic resonance spectroscopy. M.B.A. Oldstone, A. Tishon, H. Lewicki, H.J. Dyson, V.A. Feher, N. Assa-Munt, & P.E. Wright (1991) *J. Virol.* **65**, 1727-1734. [Another collaborative effort on immunogenic peptide conformations. My contribution was mainly in the measurement and interpretation of the peptide NMR spectra.] (49 citations)
26. Assignment of the <sup>15</sup>N NMR spectrum of reduced and oxidized *Escherichia coli* thioredoxin. K. Chandrasekhar, G. Krause, A. Holmgren, & H.J. Dyson (1991) *FEBS Lett.* **284**, 178-183. [I designed the experiments and wrote the paper. Chandrasekhar was my postdoctoral fellow and did the NMR measurements and analysis. Krause and Holmgren prepared the labeled protein.] (24 citations)
27. Polypeptide backbone resonance assignments and secondary structure of *Bacillus subtilis* Enzyme IIIg<sup>lc</sup> determined by two-dimensional and three-dimensional heteronuclear NMR spectroscopy. W.J. Fairbrother, J. Cavanagh, H.J. Dyson, A.G. Palmer, III, S.L. Sutrina, J. Reizer, M.H. Saier, Jr, & P.E. Wright (1991) *Biochemistry* **30**, 6896-6907. [A new collaborative project between Saier (UCSD) and Wright. I was only marginally involved in the early NMR spectroscopy of this protein.] (46 citations)
28. Solution conformational preferences of immunogenic peptides derived from the principal neutralizing determinant of the HIV-1 envelope glycoprotein gp120. K. Chandrasekhar, A.T. Profy, & H.J. Dyson (1991) *Biochemistry* **30**, 9187-9194. [A collaborative project with the

Repligen company (Profy), who provided the peptides. Experimental work was done by Chandrasekhar; I analyzed the data and wrote the paper.] (141 citations)

29. Immunogenic peptides corresponding to the dominant antigenic region Alanine-597 to Cysteine-619 in the transmembrane protein of simian immunodeficiency virus have a propensity to fold in aqueous solution. H.J. Dyson, E. Norrby, K. Hoey, D.E. Parks, R.A. Lerner, & P.E. Wright (1992) *Biochemistry* **31**, 1458-1463. [Another collaboration on immunogenic peptides. For this work, I did all of the NMR experiments, analyzed the data and wrote the paper. Norrby suggested the project and edited the paper. Hoey and Parks made the peptides.] (19 citations)
30. A comparison of the requirements for pre-formed secondary structure in proteins with different structures in the folded state. H.J. Dyson & P.E. Wright (1992) *Structure and Function* **2**, 113-120. [A short review on peptide conformation. I did most of the writing for this paper.]
31. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding I. Myohemerythrin. H.J. Dyson, G. Merutka, J.P. Waltho, R.A. Lerner, & P.E. Wright (1992) *J. Mol. Biol.* **226**, 795-817. [This paper and the one following represent the next major step in the understanding of peptide conformation in solution, based on the hypothesis that conformational preferences observed in short peptide sequences ought to be related to the initiation sites for folding of the entire molecule. The sequences of two proteins were “dissected” by synthesis of short peptides corresponding to fragments of the entire polypeptide, and these peptides were examined for residual structure in solution. This paper examines a helical protein, the following paper looks at a  $\beta$ -sheet protein. I made most of the NMR measurements, analyzed the data and wrote the paper. Wright and I designed the peptides. Merutka and Waltho did some of the NMR experiments.] (344 citations)
32. Folding of peptide fragments comprising the complete sequence of proteins: models for the initiation of protein folding II Plastocyanin. H.J. Dyson, J.R. Sayre, G. Merutka, H.-C. Shin, R.A. Lerner, & P.E. Wright (1992) *J. Mol. Biol.* **226**, 819-835. [For this paper, I made most of the NMR measurements, analyzed the data and wrote the paper. Wright and I designed the peptides. Sayre, Merutka and Shin did some of the NMR experiments.] (242 citations)
33. Peptide conformation and protein folding. H.J. Dyson & P.E. Wright (1993) *Curr. Opin. Struct. Biol.* **3**, 60-65. [An invited review that summarizes our thoughts on the interpretation of peptide data for the initiation of protein folding. I wrote the first draft, edited by Wright.] (163 citations)
34. Comparison of backbone and tryptophan side-chain dynamics of reduced and oxidized *Escherichia coli* thioredoxin using  $^{15}\text{N}$  NMR relaxation measurements. M.J. Stone, K. Chandrasekhar, A. Holmgren, P.E. Wright, & H.J. Dyson (1993) *Biochemistry* **32**, 426-435. [This paper was one of the first to describe the dynamics of a protein and relate the results to subtle structural and functional differences between two forms of a protein. Stone and Chandrasekhar made the NMR measurements and analyzed the data. Holmgren prepared the protein. Wright was involved with the data interpretation. I was responsible for the interpretation of the relaxation and dynamic data in terms of the structure and function of thioredoxin.] (107 citations)
35. Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. J.P. Waltho, V.A. Feher, G. Merutka, H.J. Dyson, & P.E. Wright (1993) *Biochemistry* **32**, 6337-6347. [This and the next two papers describe a systematic study of the conformational preferences of two helical segments of myoglobin. The majority of the NMR measurements were made by Waltho, with assistance of Feher and Merutka. My role was primarily one of data analysis and interpretation. The papers were entirely written by me.] (186 citations)
36. Peptide models of protein folding initiation sites. 2. The G-H turn region of myoglobin acts as a helix stop signal. H.-C. Shin, G. Merutka, J.P. Waltho, P.E. Wright, & H.J. Dyson (1993) *Biochemistry* **32**, 6348-6355. [The majority of the NMR measurements were made by Shin, with assistance of Merutka and Waltho.] (88 citations)

37. Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. H.-C. Shin, G. Merutka, J.P. Waltho, L.L. Tennant, H.J. Dyson, & P.E. Wright (1993) *Biochemistry* **32**, 6356-6364. [The majority of the NMR measurements were made by Shin, with assistance of Merutka and Waltho.] (100 citations)
38. Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. S. Koide, H.J. Dyson, & P.E. Wright (1993) *Biochemistry* **32**, 12299-12310. [My role here was data interpretation and editing of the paper.] (69 citations)
39. Protein structure calculation using NMR restraints. H.J. Dyson & P.E. Wright (1994) In: Two-Dimensional NMR Spectroscopy: Applications for Chemists and Biochemists (W.R. Croasmun & R. Carlson Eds.) VCH Publishers, Inc., New York. pp 655-698. [I wrote this book chapter, with some editing by Wright.] (13 citations)
40. Peptide structure in solution. H.J. Dyson (1994) In: Synthetic Vaccines (B.H. Nicholson Ed.) Blackwell Scientific Publications Ltd, Oxford, U.K.. pp 246-267. (Book chapter written by me.)
41. Characterization by <sup>1</sup>H NMR of a C32S,C35S double mutant of *Escherichia coli* thioredoxin confirms its resemblance to the reduced wild-type protein. H.J. Dyson, M.-F. Jeng, P. Model, & A. Holmgren (1994) *FEBS Lett.* **339**, 11-17. [I planned the experiments following a conversation with Peter Model in New York. I made most of the measurements, with some input from Jeng. Protein was provided by Model and Holmgren. I analyzed and interpreted the data and wrote the paper]. (20 citations)
42. Effect of disulfide bridge formation on the NMR spectrum of a protein: studies on oxidized and reduced *Escherichia coli* thioredoxin. K. Chandrasekhar, A.P. Campbell, M.-F. Jeng, A. Holmgren, & H.J. Dyson (1994) *J. Biomol. NMR* **4**, 411-432. [Chandrasekhar made the NMR measurements with some help from Campbell and Jeng. Holmgren made the protein. I interpreted the results and wrote the paper.] (26 citations)
43. Use of chemical shifts and coupling constants in NMR structural studies on peptides and proteins. D.A. Case, H.J. Dyson, & P.E. Wright (1994) *Methods Enzymol.* **239**, 392-416. [For this collaborative review, Case wrote the section on chemical shifts and I wrote the coupling constant section, edited by Wright.] (80 citations)
44. Detecting nascent structures in solution using NMR. H.J. Dyson, J. Yao, & P.E. Wright (1994) In: Peptides: Chemistry, Structure and Biology (R.S. Hodges & J.A. Smith Eds.) ESCOM, Leiden. pp 1093-1095. [book chapter written by me.]
45. Binding of hapten to a single-chain catalytic antibody demonstrated by electrospray mass spectrometry. G. Siuzdak, J.F. Krebs, S.J. Benkovic, & H.J. Dyson (1994) *J. Am. Chem. Soc.* **116**, 7937-7938. [A new collaboration with Benkovic of Penn State University on structure and dynamics of catalytic antibodies. The paper describes a new MS method, was written by Krebs and Siuzdak and edited by me.] (24 citations)
46. High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. M.-F. Jeng, A.P. Campbell, T. Begley, A. Holmgren, D.A. Case, P.E. Wright, & H.J. Dyson (1994) *Structure* **2**, 853-868. [This paper provides the definitive solution structure of both the oxidized and reduced forms of thioredoxin. Experiments and calculations were planned and supervised by me, with input from Wright. My postdoc Jeng made all of the measurements with input from Begley and Campbell, and calculated the structures, with input from Case. I interpreted the results and wrote the paper.] (175 citations)
47. Stabilization of a Type VI turn in a family of linear peptides in water solution. J. Yao, V.A. Feher, B.F. Espejo, M.T. Reymond, P.E. Wright, & H.J. Dyson (1994) *J. Mol. Biol.* **243**, 736-753. [The next major phase of the peptide conformation project. Wright and I planned the experiments. Yao made the measurements, with input from Feher, Espejo and Reymond. I interpreted the data and wrote the paper.] (122 citations)

48. Three-dimensional structure of a Type VI turn in a linear peptide in water solution: evidence for stacking of aromatic rings as a major stabilizing factor. J. Yao, H.J. Dyson, & P.E. Wright (1994) *J. Mol. Biol.* **243**, 754-766. [This was a pioneering study that showed that the structure of a highly-populated conformation in the ensemble of a peptide could be defined by NMR. Yao made the measurements and calculated the structures, with input from Wright and me. I interpreted the data and wrote the paper.] (90 citations)
49. Nuclear magnetic resonance <sup>15</sup>N and <sup>1</sup>H resonance assignments and global fold of rusticyanin: Insights into the ligation and acid stability of the blue copper site. A.H. Hunt, A. Toy-Palmer, N. Assa-Munt, J. Cavanagh, R.C. Blake, II, & H.J. Dyson (1994) *J. Mol. Biol.* **244**, 370-384. [A new project. I initiated the collaboration with the bacteriologist Blake, who supplied protein for these experiments. Hunt and Toy-Palmer made the NMR measurements with some input from Assa-Munt and Cavanagh. I interpreted the results and wrote the paper.] (26 citations)
50. The folding pathway of apomyoglobin. P.A. Jennings, H.J. Dyson, & P.E. Wright (1994) In: Statistical Mechanics, Protein Structure and Protein Substrate Interactions (S. Doniach Ed.) Plenum Press, New York. pp 7-18. [Book chapter written by Jennings, edited by Wright and me.]
51. Analysis of peptide and protein structure. H.J. Dyson (1994) In: Immunological recognition of peptides in medicine and biology (N. Zegers, W. Boersma, & E. Claassen Eds.) CRC Press, London. pp 133-145. [Book chapter written by me.] (1 citation)
52. Differential side chain hydration in a linear peptide containing a type VI turn. J. Yao, R. Brüschweiler, H.J. Dyson, & P.E. Wright (1994) *J. Am. Chem. Soc.* **116**, 12051-12052. [further work on the system of ref. 48. Yao made the measurements, with input from Brüschweiler. I interpreted the results and wrote the paper, edited by Wright.] (32 citations)
53. Antigenic peptides. H.J. Dyson & P.E. Wright (1995) *FASEB J.* **9**, 37-42. [Invited review written by me with some editing by Wright.] (54 citations)
54. Comparison of the hydrogen exchange behavior of reduced and oxidized *Escherichia coli* thioredoxin. M.-F. Jeng & H.J. Dyson (1995) *Biochemistry* **34**, 611-619. [This very thorough study of HX behavior has been widely cited in the mass spectrometry field as a standard work. Jeng made the measurements, I planned the experiments, interpreted the results and wrote the paper.] (51 citations)
55. Detection of a catalytic antibody species acylated at the active site by electrospray mass spectrometry. J.F. Krebs, G. Siuzdak, H.J. Dyson, J.D. Stewart, & S.J. Benkovic (1995) *Biochemistry* **34**, 720-723. [Application of the method described in ref. 45. Krebs and Siuzdak made the measurements. My role was primarily in writing and editing the manuscript.] (33 citations)
56. "Random coil" <sup>1</sup>H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. G. Merutka, H.J. Dyson, & P.E. Wright (1995) *J. Biomol. NMR* **5**, 14-24. [This work establishes chemical shift standards for "random coil" values that are used in protein NMR. Wright and I planned the experiments. Merutka made the measurements and analyzed the data. I wrote the paper, Wright edited.] (314 citations)
57. NMR of thioredoxin and glutaredoxin. H.J. Dyson (1995) *Methods Enzymol.* **252**, 293-306. [Invited review article written by me.] (3 citations)
58. Gene synthesis, high-level expression and mutagenesis of *Thiobacillus ferrooxidans* rusticyanin. His85 is a ligand to the blue copper center. D.R. Casimiro, A. Toy-Palmer, R.C. Blake, II, & H.J. Dyson (1995) *Biochemistry* **34**, 6640-6648. [This report describes an efficient method for synthesis of a gene in order to express a particular protein for NMR. I designed the experiments, Casimiro designed the experimental system and made the measurements, assisted by Toy-Palmer. I wrote the paper.] (45 citations)

59. Complete  $^{13}\text{C}$  assignment for recombinant rusticyanin: Prediction of secondary structure from patterns of chemical shifts. A. Toy-Palmer, S. Prytulla and H.J. Dyson (1995) *FEBS Lett.* **365**, 35-41. [I designed the experiments and wrote the paper.] (9 citations)
60. Proton sharing between cysteine thiols in *Escherichia coli* thioredoxin: implications for the mechanism of protein disulfide reduction. M.-F. Jeng, A. Holmgren, & H.J. Dyson (1995) *Biochemistry* **34**, 10101-10105. [This study was somewhat controversial at the time. Later work (refs. 62, 84) showed that our conclusion in this paper was probably correct. Jeng made the measurements, I interpreted the results and wrote the paper.] (60 citations)
61.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift references in biomolecular NMR. D.S. Wishart, C.G. Bigam, J. Yao, F. Abildgaard, H.J. Dyson, E. Oldfield, J.L. Markley, & B.D. Sykes (1995) *J. Biomol. NMR* **6**, 135-140. [This is a very highly cited reference paper. The authors formed a consortium to measure chemical shift references for use in biomolecular NMR. Yao and I were part of this consortium and made several of the measurements.] (928 citations)
62. Direct measurement of the Asp-26  $\text{pK}_a$  for reduced *Escherichia coli* thioredoxin by  $^{13}\text{C}$  NMR. M.-F. Jeng and H. J. Dyson (1996) *Biochemistry* **35**, 1-6. [Corroboration for the interpretation made in ref. 60] (35 citations)
63. Folding of proteins and protein fragments. H.J. Dyson & P.E. Wright (1996) In: Encyclopedia of Nuclear Magnetic Resonance (ed. D.M. Grant and R.K. Harris) John Wiley, Chichester. pp 3811-3820. [Contribution to a major reference work. I wrote, Wright edited.] (1 citation)
64. Thioredoxin and glutaredoxin. H.J. Dyson (1996) In: Encyclopedia of Nuclear Magnetic Resonance (ed. D.M. Grant and R.K. Harris) John Wiley, Chichester. pp 4727-4733. [Contribution to a reference work.]
65. NMR structural studies of flexible molecules. P.E. Wright & H.J. Dyson (1996) In: NMR as a Structural Tool for Macromolecules (B.D. Nageswara Rao and M.D. Kemple, Eds) Plenum Publishing Corp, New York, pp 245-249. [Book chapter. I wrote, Wright edited.]
66. Replacement of Trp-28 in *Escherichia coli* thioredoxin by site-directed mutagenesis affects thermodynamic stability but not function. I. Slaby, V. Cerna, M.-F. Jeng, H.J. Dyson and A. Holmgren (1996) *J. Biol. Chem.* **271**, 3091-3096. [Slaby and Jeng did the experimental work. Holmgren and I wrote the paper.] (20 citations)
67. Solution conformation of an immunogenic peptide derived from the principal neutralizing determinant of the HIV-2 envelope glycoprotein gp125. A.P. Campbell, B.D. Sykes, E. Norrby, N. Assa-Munt and H. J. Dyson (1996) *Folding and Design* **1**, 157-165. [Campbell did the experimental work, partly in Sykes's lab. Assa-Munt did some of the early NMR work. Norrby suggested the project. I interpreted the results and wrote the paper.] (10 citations)
68. Insights into protein folding from NMR. H.J. Dyson and P.E. Wright (1996) *Ann Rev. Phys. Chem.* **47**, 369-395. [Invited review. I wrote, Wright edited.] (78 citations)
69. NMR solution structure of Cu(I) rusticyanin from *Thiobacillus ferrooxidans*: Structural basis for the the extreme acid stability and redox potential. M. V. Botuyan, A. Toy-Palmer, J. Chung, R. C. Blake II, P. Beroza, D.A. Case & H. J. Dyson (1996) *J. Mol. Biol.* **263**, 752-767. [A difficult structure calculation. Botuyan and Toy-Palmer made the measurements with help from Chung. Beroza and Case assisted with the structure calculations. I planned the experiments, interpreted the results and wrote the paper.] (54 citations)
70. Gene synthesis, high level expression and assignment of backbone  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances of soybean leghemoglobin. S. Prytulla, H.J. Dyson and P.E. Wright (1996) *FEBS Lett.* **399**, 283-289. [My contribution to this work was the assembly of information and writing the paper.] (9 citations)
71. Structure-based design of a constrained-peptide mimic of the HIV-1 V3 loop neutralization site. J.B. Ghiara, D. Ferguson, A.C. Satterthwait, H.J. Dyson & I.A. Wilson (1997) *J. Mol. Biol.* **266**, 31-39. [Collaborative work with Wilson. I provided some NMR data for this paper.] (54 citations)

72. Folding propensities of peptide fragments of myoglobin. M.T. Reymond, G. Merutka, P.E. Wright and H.J. Dyson (1997) *Protein Sci.* **6**, 706-716. [Further work on myoglobin fragments. See refs. 35-37. Reymond synthesized the peptides and made the measurements. Some preliminary work had been done by Merutka. I interpreted the data and wrote the paper.] (63 citations)
73. Effects of buried charged groups on cysteine thiol ionization and reactivity in *Escherichia coli* thioredoxin: Structural and functional characterization of mutants of Asp 26 and Lys 57. H.J. Dyson, M.-F. Jeng, I. Slaby, M. Lindell, D.-S. Cui & A. Holmgren (1997) *Biochemistry* **36**, 2622-2636. [Jeng in my lab, Slaby, Lindell and Cui in Holmgren's lab made the measurements. I interpreted the data and wrote the paper.] (73 citations)
74. Absence of a stable intermediate on the folding pathway of Protein A. Y. Bai, A Karimi, H.J. Dyson and P.E. Wright (1997). *Protein Sci.* **6**, 1449-1457. [Bai and Karimi made the measurements. Experiments planned by Wright. I interpreted the data and wrote the paper.] (68 citations)
75. Contribution of increased length and intact capping sequences to the conformational preference for helix in a 31-residue peptide from the C-terminus of myohemerythrin. M.T. Reymond, S. Huo, B. Duggan, P.E. Wright and H.J. Dyson (1997) *Biochemistry* **36**, 5234-5244. [A follow-up to ref. 9. Reymond made the peptide and the experimental measurements, with assistance from Huo and Duggan. Wright and I planned the experiments. I interpreted the results and wrote the paper.] (39 citations)
76. Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR. D. Eliezer, P.A. Jennings, H.J. Dyson and P.E. Wright (1997) *FEBS Lett.* **417**, 92-96. [My main contribution was to write the paper.] (25 citations)
77. PCR-based gene synthesis for protein over-production. D. Casimiro, P.E. Wright and H.J. Dyson (1997) *Structure* **5**, 1407-1412. [Review of the gene synthesis protocol. I edited the paper.] (14 citations)
78. Electron spin echo modulation study of the Type I protein rusticyanin and its mutant variant His85Ala: Implications for the general analysis of weak <sup>14</sup>N superhyperfine interactions. C.J. Bender, D. Casimiro, J. Peisach, H.J. Dyson (1997) *J. Chem. Soc. (Faraday)* **93**, 3967-3980. [Bender (Albert Einstein University) made the measurements, Casimiro and I supplied the protein and helped interpret the data, and edited the paper.] (11 citations)
79. Structure of the recombinant full-length hamster prion protein PrP(29-231): the N-terminus is highly flexible. D.G. Donne, J.H. Viles, J. Chung, D. Groth, I. Mehlhorn, F.E. Cohen, S.B. Prusiner, P.E. Wright & H.J. Dyson (1997) *Proc. Natl. Acad. Sci. USA* **94**, 13452-13457. [A major new collaboration with Prusiner and Cohen (UCSF). Donne and Viles did the experimental work, with input from Chung. Groth and Mehlhorn synthesized the protein. I wrote the paper, edited by Wright.] (366 citations)
80. Chemical shift dispersion and secondary structure prediction in unfolded and partly-folded proteins. J. Yao, H.J. Dyson and P.E. Wright (1997) *FEBS Lett.* **419**, 285-289. [Yao made the measurements. I interpreted the data and wrote the paper, Wright edited.] (48 citations)
81. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions I. Radhakrishnan, G. Perez-Alvarado, D. Parker, H.J. Dyson, M. Montminy and P.E. Wright (1997) *Cell* **91**, 741-752. [My role was in interpretation of the data and writing of the paper.] (274 citations)
82. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. D. Eliezer, J. Yao, H.J. Dyson and P.E. Wright (1998) *Nature Struct. Biol.* **5**, 148-155. [This is a major synthesis of our thoughts on the folding process of apomyoglobin. Eliezer and Yao made the measurements. Wright and I interpreted the data and wrote the paper. This work was featured in a commentary in *Science* magazine 7 November 1997: Vol. 278. no. 5340, pp. 1014 – 1015.] (225 citations)

83. A NOESY-HSQC simulation program SPIRIT. L. Zhu, H.J. Dyson and P.E. Wright (1998) *J. Biomol. NMR* **11**, 17-29. [Zhu wrote the program. My thioredoxin NMR data were used. Wright and I wrote the paper.] (13 citations)
84. Calculations of electrostatic interactions and pKas in the active site of *Escherichia coli* thioredoxin. V. Dillet, H.J. Dyson and D. Bashford (1998) *Biochemistry* **37**, 10298-10306. [A definitive theoretical treatment of the thioredoxin pKa controversy (see refs. 60, 62). Dillet and Bashford did this work. My contribution was mainly in the extensive discussions on the interpretation of the results of the calculations.] (47 citations)
85. Equilibrium NMR studies of unfolded and partially folded proteins. H.J. Dyson and P.E. Wright (1998) *Nature Struct. Biol.* **5**, 499-503. [Invited review written and edited by Wright and me.] (119 citations)
86. High resolution solution structure of the retinoid X receptor DNA binding domain. S.M.A. Holmbeck, M. P. Foster, D. R. Casimiro, D. Sem, H. J. Dyson and P. E. Wright (1998) *J. Mol. Biol.* **281**, 271-284. [My contribution was in the analysis and interpretation of data and assisting in writing the paper.] (30 citations)
87. Sequence requirements for stabilization of a peptide reverse turn in water solution: proline is not essential for stability. H.J. Dyson, L. Bolinger, V.A. Feher, J.J. Osterhout, Jr., J. Yao, & P.E. Wright (1998) *Eur. J. Biochem.* **255**, 462-471. [A further visit to the short peptide field. I synthesized a large amount of experimental data and wrote the paper.] (19 citations)
88. Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB I. Radhakrishnan, G.C. Perez-Alvarado, H.J. Dyson and P.E. Wright (1998) *FEBS Lett.* **430**, 317-322. [I was involved with data interpretation and paper writing.] (35 citations)
89. <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N NMR backbone assignments of 25.5 kDa metallo-β-lactamase from *Bacteroides fragilis*. S.D.B. Scrofani, P.E. Wright and H.J. Dyson (1998) *J. Biomol. NMR* **12**, 201-202. [The first publication from a new project aimed at elucidation of the effects of protein dynamics on catalysis, the subject of an NIH program project grant in my name, and including 4 sub-projects headed by Wright, Dyson, Benkovic and Case. For this paper, I planned the experiments and wrote the paper.] (6 citations)
90. The identification of metal-binding ligand residues in metalloproteins using nuclear magnetic resonance spectroscopy. S.D.B. Scrofani, P.E. Wright and H.J. Dyson (1998) *Protein Sci.* **7**, 2476-2479. [I interpreted the results and edited the paper.] (4 citations)
91. Glycosylation of threonine of the repeating unit of RNA polymerase II with β-linked N-acetylglucosamine leads to a turnlike structure. E. E. Simanek, D.-H. Huang, L. Pasternack, O. Seitz, D.S. Millar, H.J. Dyson and C.-H. Wong (1998) *J. Am. Chem. Soc.* **120**, 11567-11575. [For this paper from Wong's lab, I provided some NMR data.] (27 citations)
92. NMR characterization of a single-cysteine mutant of *Escherichia coli* thioredoxin and a covalent thioredoxin-peptide complex. M.-F. Jeng, M.T. Reymond, L.L. Tennant, A. Holmgren and H.J. Dyson (1998). *Eur. J. Biochem.* **257**, 299-308. [I planned the experiments, interpreted the results and wrote the paper.] (10 citations)
93. DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. S.M.A. Holmbeck, H.J. Dyson and P.E. Wright (1998) *J. Mol. Biol.*, **284**, 533-539. [I had major input in the structure calculation, interpretation of the results and writing of the paper.] (40 citations)
94. Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. V. Tsui, C. Garcia-Gonzalez, S. Cavagnero, G. Siuzdak, H.J. Dyson and P.E. Wright (1999) *Protein Sci.* **8**, 45-49. [My contribution was synthesis and interpretation of results and assistance with paper writing.] (65 citations)

95. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. S. Cavagnero, H.J. Dyson and P.E. Wright. (1999) *J. Mol. Biol.* **285**, 269-282. [I had major input in the interpretation of data and writing the paper.] (46 citations)
96. Copper binding to the prion protein. Structural implications of four identical cooperative binding sites. J.H. Viles, S.B. Prusiner, F.E. Cohen, D.D. Goodin, P.E. Wright and H.J. Dyson (1999) *Proc. Natl Acad. Sci. USA* **96**, 2042-2047. [This project was introduced by Cohen and Prusiner under the auspices of a shared NIH grant. Viles did the experimental work. I did a significant proportion of the data interpretation and paper writing.] (290 citations)
97. Improved low pH bicelle system for orienting macromolecules over a wide temperature range S. Cavagnero, H. J. Dyson and P.E. Wright (1999) *J. Biomol. NMR* **13**, 387-391. [My input was mainly in data interpretation and paper writing.] (40 citations)
98. Structural analyses of CREB-CBP transcriptional activator-coactivator complexes by NMR spectroscopy: implications for mapping the boundaries of structural domains. I. Radhakrishnan, G.C. Perez-Alvarado, D. Parker, H.J. Dyson, M.R. Montminy and P.E. Wright (1999) *J. Mol. Biol.* **287**, 859-865. [My input was in the interpretation of data and writing of the paper.] (31 citations)
99. Association between the first two immunoglobulin-like domains of the neural cell adhesion molecule N-CAM. A. Atkins, M.J. Osborne, H.A. Lashuel, G.M. Edelman, P.E. Wright, B. A. Cunningham and H.J. Dyson (1999) *FEBS Lett.* **451**, 162-168. [A new collaboration with Cunningham of Scripps Department of Neurobiology. Atkins made the experimental measurements under my supervision and we both wrote the paper.] (19 citations)
100. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. P.E. Wright and H.J. Dyson (1999) *J. Mol. Biol.* **293**, 321-331. [This paper provides a major synthesis of our thinking on intrinsically unstructured proteins. It is frequently cited as the beginning of this new field, which is rapidly gaining recognition, including its own section of the Biophysical Society. It was written as a close collaboration between Wright and me.] (604 citations)
101. Characterization of monomeric and dimeric B domain of staphylococcal protein A: Sources of stabilization of a 3-helix bundle protein. A. Karimi, M. Matsumura, P.E. Wright & H.J. Dyson. (1999) *J. Pept. Res.* **54**, 344-352. [I synthesized a large amount of experimental data and wrote the paper.] (3 citations)
102. Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. B.M. Duggan, H.J. Dyson & P.E. Wright (1999) *Eur. J. Biochem.* **265**, 539-548. [I played a major role in structure calculation and data interpretation and in writing the paper.] (27 citations)
103. Assignment of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  resonances of reduced *Escherichia coli* glutaredoxin 2. B. Xia, J. Chung, A. Vlamis-Gardikas, A. Holmgren, P.E. Wright and H.J. Dyson (1999) *J. Biomol. NMR* **14**, 197-198. [I wrote this short paper.] (1 citation)
104. NMR characterization of the metallo- $\beta$ -lactamase from *Bacteroides fragilis* and its interaction with a tight-binding inhibitor: Role of a flexible loop S.D.B. Scrofani, J. Chung, J.J.A. Huntley, S. J. Benkovic, P.E. Wright and H.J. Dyson (1999) *Biochemistry* **38**, 14507-14514. [I planned the experiments and supervised the data interpretation and writing of the paper.] (44 citations)
105. Backbone resonance assignments for the Fv fragment of the catalytic antibody NPN43C9 with bound *p*-nitrophenol. G. Kroon, M. Martinez-Yamout, J.F. Krebs, John Chung, H.J. Dyson & P.E. Wright (1999) *J. Biomol. NMR* **15**, 83-84. [I supervised writing this short paper.] (4 citations)
106. Amide proton hydrogen exchange rates for sperm whale myoglobin obtained from  $^{15}\text{N}$ - $^1\text{H}$  NMR. S. Cavagnero, Y. Thériault, S.S. Narula, H.J. Dyson & P.E. Wright. (2000) *Protein Sci.* **9**, 186-193. [I synthesized experimental results dating back a number of years, and wrote the paper.] (10 citations)



107. NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. G.B. Legge, R.W. Kriwacki, J. Chung, U. Hommel, P. Ramage, D.A. Case, H.J. Dyson and P.E. Wright (2000) *J. Mol. Biol.* **295**, 1251-1264. [I supervised data collection and interpretation and structure determination, and largely wrote the paper.] (35 citations)
108. DNA-induced  $\alpha$ -helix capping in conserved linker sequences is a determinant of binding affinity in Cys<sub>2</sub>-His<sub>2</sub> zinc fingers. J.H. Laity, H. J. Dyson and P. E. Wright (2000) *J. Mol. Biol.* **295**, 719-727. [I made a major contribution to the interpretation of data and writing the paper.] (50 citations)
109. Spectroscopic techniques. H.J. Dyson (2001) In: *Encyclopedia of Life Sciences* John Wiley and Sons, Ltd. Chichester <http://www.els.net/10.1038/npg/els.0002738> [short review article]
110. Backbone HN, N, C $\alpha$ , C' and C $\beta$  assignments of the 19 kDa DHFR/NADPH complex at 9°C and pH 7.6. E. Zaborowski, J. Chung, G. Kroon, H.J. Dyson and P.E. Wright (2000) *J. Biomol. NMR* **16**, 349-350. [My contribution was in interpretation of data and writing of the paper.] (3 citations)
111. Assignment of <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N resonances of the I-domain of human leukocyte function associated antigen-1. R.W. Kriwacki, G.B. Legge, U. Hommel, P. Ramage, J. Chung, L.L. Tennant, P.E. Wright and H.J. Dyson (2000) *J. Biomol. NMR* **16**, 271-272. [I assembled the data and wrote the paper.] (2 citations)
112. Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. D. Eliezer, J. Chung, H.J. Dyson and P.E. Wright (2000) *Biochemistry* **39**, 2894-2901. [I played a role in data interpretation and a major role in paper writing.] (67 citations)
113. Identification of the regions involved in DNA binding by the mouse PEBP2 $\alpha$  protein. G.C. Perez-Alvarado, A. Munnerlyn, H.J. Dyson, R. Grosschedl and P.E. Wright (2000) *FEBS Lett.* **470**, 125-130. [I was involved in data interpretation and paper writing.] (11 citations)
114. Alternative splicing of Wilms' tumor suppressor protein modulates DNA binding activity through isoform-specific DNA-induced conformational changes. J.H. Laity, J. Chung, H.J. Dyson, P.E. Wright (2000) *Biochemistry* **39**, 5341-5348. [I was involved in data interpretation and helped write the paper.] (22 citations)
115. Conservation of folding pathways in evolutionarily distant globin sequences. C. Nishimura, S. Prytulla, H.J. Dyson & P.E. Wright (2000) *Nature Struct. Biol.* **7**, 679-686. [I was heavily involved in data interpretation and wrote most of the paper.] (53 citations)
116. Solution structure of the cysteine-rich domain of the *Escherichia coli* chaperone protein DnaJ. M.A. Martinez-Yamout, G.B. Legge, O. Zhang, P.E. Wright and H.J. Dyson. (2000) *J. Mol. Biol.* **300**, 805-818. [I initiated the project and played a major role in data interpretation and paper writing.] (48 citations)
117. Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. C. Garcia, C. Nishimura, S. Cavagnero, H.J. Dyson and P.E. Wright (2000) *Biochemistry* **39**, 11227-11237. [I played a major role in supervision of data collection, data interpretation and paper writing.] (38 citations)
118. Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. R.N. De Guzman, H.Y. Liu, M. Martinez-Yamout, H.J. Dyson and P.E. Wright (2000) *J. Mol. Biol.* **303**, 243-253. [I had significant input into data interpretation and played a major role in writing the paper.] (32 citations)
119. Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView. S.Schwarzinger, G.J.A. Kroon, T.R. Foss, P.E. Wright and H.J. Dyson (2000) *J. Biomol. NMR* **18**, 43-48. [A well-cited reference work. I helped plan the experiments, supervised the experimental work and wrote the paper.] (81 citations)
120. Dynamics of the metallo- $\beta$ -lactamase from *Bacteroides fragilis* in the presence and absence of a tight-binding inhibitor. J.J.A. Huntley, S.D.B. Scrofani, M.J. Osborne, P.E. Wright and H.J.

- Dyson (2000) *Biochemistry* **39**, 13356-13364. [This paper represents the major findings from the lactamase project (see ref. 89). I planned the experiments, supervised the work and wrote the paper.] (30 citations)
121. Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. J.H. Laity, H.J. Dyson and P.E. Wright (2000) *Proc. Natl. Acad. Sci. USA* **97**, 11932-11935. [I played a major role in data interpretation and paper writing.] (39 citations)
  122. Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. B.P. Hudson, M.A. Martinez-Yamout, H.J. Dyson and P.E. Wright (2000). *J. Mol. Biol.* **304**, 355-370. [I played a significant role in data interpretation and paper writing.] (68 citations)
  123. Structure of the PHD zinc finger from human Williams-Beuren Syndrome transcription factor. J. Pascual, M. Martinez-Yamout, H.J. Dyson and P.E. Wright (2000). *J. Mol. Biol.* **304**, 723-729. [I played a significant role in structure determination and paper writing.] (67 citations)
  124. NMR methods for the elucidation of the structure and dynamics in disordered states. H.J. Dyson and P.E. Wright (2001) *Methods Enzymol.* **339**, 258-270. [Invited review. I wrote, Wright edited.] (43 citations)
  125. Local structural plasticity of the prion protein. Analysis of NMR relaxation dynamics. J.H. Viles, D. Donne, G.J.A. Kroon, S.B. Prusiner, F.E. Cohen, H.J. Dyson and P.E. Wright (2001) *Biochemistry* **40** (9), 2743-2753. [Both Wright and I had major input into the design and interpretation of results in this difficult project. I had major input in the writing of the paper.] (59 citations)
  126. Two different neurodegenerative diseases caused by proteins with similar structures. H. Mo, R.C. Moore, F.E. Cohen, D. Westaway, S.B. Prusiner, P.E. Wright and H.J. Dyson (2001) *Proc. Natl. Acad. Sci. USA* **98**, 2352-2357. [This paper describes the structure of a prion-like protein called Doppel. I supervised the experimental work and structure calculations and wrote the paper.] (76 citations)
  127. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin: provides insights into the early events in protein folding. J. Yao, J. Chung, D. Eliezer, P.E. Wright and H.J. Dyson (2001) *Biochemistry* **40**, 3561-3571. [I supervised the experimental work and synthesized a large amount of data, and wrote most of the paper.] (102 citations)
  128. SANE (Structure assisted NOE evaluation): An automated model-based approach for NOE assignment. B.M. Duggan, G.B. Legge, H.J. Dyson and P.E. Wright (2001). *J. Biomol. NMR* **19**, 321-329. [A technical paper. My role was mostly in evaluating the results and assisting with writing the paper.] (41 citations)
  129. Genomic-scale comparison of sequence- and structure-based methods of function prediction: Does structure provide additional insight? J.S. Fetrow, N. Siew, J.A. Di Gennaro, M. Martinez-Yamout, H.J. Dyson and J. Skolnick (2001) *Protein Sci.* **10**, 1005-1014. [I was involved with the design of experiments to test the hypothesis of Skolnick that sequence comparisons could be used to predict function.] (34 citations)
  130. Sequence-dependent correction of random coil NMR chemical shifts. S. Schwarzingler, G.J.A. Kroon, T.R. Foss, J. Chung, P.E. Wright and H.J. Dyson (2001). *J. Am. Chem. Soc.* **123**, 2970-2978. [This is a widely used method for sequence correction of random coil chemical shifts, used in protein structure evaluation in NMR. I was involved in designing and interpreting the experiments, and played a major role in writing the paper.] (127 citations)
  131. Potential bias in NMR relaxation data introduced by peak intensity analysis and curve fitting methods. J.H. Viles, B.M. Duggan, E. Zaborowski, S. Schwarzingler, J.J.A. Huntley, G.J.A. Kroon, H.J. Dyson and P.E. Wright (2001). *J. Biomol. NMR* **21**, 1-9. [I was heavily involved in the interpretation of data for this large collaborative effort among members of the Wright and Dyson labs. I was also involved in collating the data and in writing the paper.] (23 citations)

132. Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under non-denaturing conditions. Y. Bai, J. Chung, H.J. Dyson and P.E. Wright (2001) *Protein Sci.* **10**, 1056-1066. [I helped plan the experiments and interpret the data, and was mainly responsible for writing the paper.] (43 citations)
133. Solution structure of *Escherichia coli* glutaredoxin-2 shows similarity to mammalian glutathione-S-transferases. B. Xia, A. Vlamis-Gardikas, A. Holmgren, P.E. Wright and H.J. Dyson (2001) *J. Mol. Biol.* **310**, 907-918. [I planned the experiments and supervised data collection and structure calculation, and was largely responsible for interpretation of the results and writing of the paper.] (35 citations)
134. Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism. M.J. Osborne, J. Schnell, S.J. Benkovic, H.J. Dyson and P.E. Wright (2001). *Biochemistry* **40**, 9846-9859. [I was involved in data interpretation and extensively in paper writing.] (100 citations)
135. Solution structure of the third immunoglobulin domain of the neural cell adhesion molecule N-CAM: Can solution studies define the mechanism of homophilic binding? A. Atkins, J. Chung, S. Deechongkit, E. B. Little, G. M. Edelman, P. E. Wright, B. A. Cunningham and H. J. Dyson (2001) *J. Mol. Biol.* **311**, 161-172. [I planned the experiments, supervised data acquisition and structure calculation, and led the interpretation of the results and writing of the paper.] (11 citations)
136. Conformational and dynamic characterization of the molten globule state of an apomyoglobin mutant with an altered folding pathway. S. Cavagnero, C. Nishimura, S. Schwarzingler, H.J. Dyson and P.E. Wright (2001). *Biochemistry* **40**, 14459-14467. [I was responsible for data collation and interpretation, and heavily involved in paper writing.] (22 citations)
137. Structure and dynamics of disordered proteins. H.J. Dyson and P.E. Wright (2002) *Encyclopedia of NMR* **9**, 449-457. [reference work. Written by me, edited by Wright.]
138. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. S.J. Demarest, M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H.J. Dyson, R.M. Evans and P.E. Wright (2002) *Nature* **415**, 549-553. [I was heavily involved in the interpretation of the results, which provide the first example of structural characterization of the folded complex formed between two intrinsically unstructured proteins.] (86 citations)
139. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model and with explicit water. B. Xia, V. Tsui, D.A. Case, H.J. Dyson and P.E. Wright (2002) *J. Biomol. NMR* **22**, 317-331. [A technical paper. I was mainly involved with data interpretation and paper writing.] (36 citations)
140. Coupling of folding and binding for unstructured proteins. H.J. Dyson and P.E. Wright (2002) *Curr. Opin. Struct. Biol.* **12**, 54-60. [Invited review. Written by me, edited by Wright.] (318 citations)
141. Assignment of a 15 kDa protein complex formed between the p160 coactivator ACTR and CREB binding protein. S.J. Demarest, J. Chung, H.J. Dyson and P.E. Wright (2002) *J. Biomol. NMR* **22**, 377-378. [See ref. 138.]
142. Structural basis for Hif-1 $\alpha$ /CBP recognition in the cellular hypoxic response. S.A. Dames, M. Martinez-Yamout, R.N. De Guzman, H.J. Dyson and P.E. Wright (2002) *Proc. Natl Acad. Sci. USA*, **99**, 5271-5276. [I had major input into the interpretation of data and paper writing.] (102 citations)
143. Insights into the structure and dynamics of unfolded proteins from NMR. H. Jane Dyson and P.E. Wright (2002) *Adv. Protein Chem.*, **62**, 311-340. [Invited review. Written by me, edited by Wright.] (68 citations)

144. The apomyoglobin folding pathway revisited: Structural heterogeneity in the kinetic burst phase intermediate. C. Nishimura, H.J. Dyson and P.E. Wright (2002) *J. Mol. Biol.*, **322**, 483-489. [I was closely involved with all aspects of this paper, planning experiments, interpreting data and writing for publication.] (38 citations)
145. Mapping long-range contacts in a highly unfolded protein. M.A. Lietzow, M. Jamin, H.J. Dyson and P.E. Wright (2002) *J. Mol. Biol.*, **322**, 655-662. [I was closely involved with all aspects of this paper, planning experiments, interpreting data and writing for publication.] (49 citations)
146. Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. S. Schwarzhinger, P.E. Wright and H.J. Dyson. (2002) *Biochemistry* **41**, 12681-12686. [I was responsible for planning the experiments, interpreting the data and writing the paper.] (49 citations)
147. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. T. Zor, B.M. Mayr, H.J. Dyson, M.R. Montminy and P.E. Wright (2002). *J. Biol. Chem.* **277**, 42241-42248. [I was closely involved in the planning of experiments, interpretation of data and paper writing.] (24 citations)
148. Cooperativity in transcription factor binding to the coactivator CREB-binding Protein (CBP). N.K. Goto, T. Zor, M. Martinez-Yamout, H.J. Dyson and P.E. Wright (2002). *J. Biol. Chem.* **277**, 43168-43174. [Companion to ref. 147. I was closely involved in the planning of experiments, interpretation of data and paper writing.] (24 citations)
149. Folding of a  $\beta$ -sheet protein monitored by real-time NMR spectroscopy. M. Mizuguchi, G. Kroon, P.E. Wright and H.J. Dyson (2003) *J. Mol. Biol.*, **328**, 1161-1171. [Wright and I were responsible for planning experiments. I was heavily involved in interpretation of the data and wrote the paper.] (12 citations)
150. *Plasmodium vivax* CS peptides display conformational preferences for folded forms in solution. T.E. Lehmann, G. Kroon, M.A. Lorenzo, H. Bermúdez, H. Perez and H.J. Dyson (2003) *J. Peptide Research* **61**, 252-262. [Lehmann visited from Venezuela to perform these experiments. Together we planned the experiments and interpreted the data. I edited the paper.]
151. Role of a solvent-exposed tryptophan in the recognition and binding of antibiotic substrates for a metallo- $\beta$ -lactamase. J.J.A. Huntley, W. Fast, S.J. Benkovic, P.E. Wright and H.J. Dyson (2003). *Protein Sci.* **12**, 1368-1375. [Together with postdoctoral fellow Huntley, I designed the experiments, interpreted the data and wrote the paper.] (13 citations)
152. Monomeric complex of human orphan estrogen related receptor-2 with DNA: A pseudodimer interface mediates extended half-site recognition. M.D. Gearhart, S.M.A. Holmbeck, R.M. Evans, H.J. Dyson and P.E. Wright (2003) *J. Mol. Biol.* **327**, 819-832. [I was heavily involved in this structure calculation and in the writing of the paper.] (16 citations)
153. Changes in structure and dynamics of the Fv fragment of a catalytic antibody upon binding of inhibitor. G. J.A. Kroon, H. Mo, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright (2003) *Protein Sci.* **12**, 1386-1394. [I was responsible for much of the supervision of this project, and for the interpretation of results and writing of the paper.] (5 citations)
154. Structure of the nuclear factor ALY: insights into post-transcriptional regulatory and mRNA nuclear export processes. G.C. Perez-Alvarado, M. Martinez-Yamout, M. Allen, R. Grosschedl, H.J. Dyson and P.E. Wright (2003) *Biochemistry* **42**, 7348-7357. [I assisted with interpretation of the data and writing the paper.] (6 citations)
155. Diagnostic chemical shift markers for loop conformation and substrate and cofactor binding in dihydrofolate reductase complexes. M. J. Osborne, R. P. Venkitakrishnan, H. J. Dyson and P. E. Wright (2003). *Protein Science* **12**, 2230-2238. [My input was in data interpretation and paper writing.] (10 citations)
156. Role of the B helix in early folding events in apomyoglobin: Evidence from site-directed mutagenesis for native-like long range interactions. C. Nishimura, P.E. Wright and H.J. Dyson

- (2003) *J. Mol. Biol.* **334**, 293-307. [Wright and I were responsible for planning the experiments. I did a major amount of work on data interpretation and paper writing.] (14 citations)
157. Packing, specificity and mutability at the binding interface between the p160 coactivator and CREB-binding protein. S.J. Demarest, S. Deechongkit, H.J. Dyson, R.M. Evans and P.E. Wright (2004). *Protein Science* **13**, 203-210. [Follow-up to ref. 138.] (7 citations)
  158. Effects of cofactor binding and loop conformation on side chain methyl dynamics in dihydrofolate reductase. J.R. Schnell, H.J. Dyson, P.E. Wright (2004) *Biochemistry* **43**, 374-383. [I was involved with paper writing and data interpretation.] (27 citations)
  159. Unfolded proteins and protein folding studied by NMR. H.J. Dyson, and P.E. Wright (2004) *Chemical Reviews* **104**, 3607-3622. [Wright and I were invited to act as guest editors for this issue of *Chemical Reviews*. This paper is our contribution to that issue, which also includes a commentary by the two of us. Introduction: Biological Nuclear Magnetic Resonance. H. Jane Dyson and Peter E. Wright, pp 3517 - 3518; (Editorial) Written by me, edited by Wright.] (75 citations)
  160. Structure, dynamics and catalytic function of dihydrofolate reductase. J.R. Schnell, H.J. Dyson, and P.E. Wright (2004) *Ann. Rev. Biophys. Biomol. Struct.* **33**, 119-140. [Invited review. Schnell wrote the first draft, Wright and I extensively revised.] (56 citations)
  161. Interaction of the TAZ1 domain of CREB-binding protein with the activation domain of CITED2: Regulation by competition between intrinsically unstructured ligands for non-identical binding sites. R.N. De Guzman, M.A. Martinez-Yamout, H.J. Dyson, and P.E. Wright (2004) *J. Biol. Chem.* **279**, 3042-3049. [I was heavily involved with the structure determination, data interpretation and paper writing.] (15 citations)
  162. Structure and function of the CBP/p300 TAZ domain. R.N. De Guzman, M.A. Martinez-Yamout, H.J. Dyson, and P.E. Wright (2004) In: Zinc Finger Proteins: From Atomic Contact to Cellular Function (S. Iuchi and N. Kuldell, eds), Landes Biosciences. Chapter 17. [Invited book chapter. I assisted with writing and editing.]
  163. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. B.P. Hudson, M.A. Martinez-Yamout, H.J. Dyson, P.E. Wright (2004) *Nature Structural Biology* **11**, 257-264. [I was involved in the structure calculation, data interpretation and paper writing.] (34 citations)
  164. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. T. Zor, R.N. De Guzman, H.J. Dyson, P.E. Wright (2004) *J. Mol. Biol.* **337**, 521-534. [I was involved in the structure determination and data interpretation and in the paper writing.] (16 citations)
  165. Activation of the redox-regulated chaperone Hsp33 by domain unfolding. P.C.F. Graf, M.A. Martinez-Yamout, S. VanHaerents, H. Lilie, H.J. Dyson and U. Jakob (2004). *J. Biol. Chem.* **279**, 20529-20538. [A new collaboration initiated by me. For this paper, we supplied some NMR data.] (17 citations)
  166. Disulfide bonding arrangements in active forms of the somatomedin B domain of human vitronectin. Y. Kamikubo, R. De Guzman, G. Kroon, S. Curriden, J. Neels, M.J. Churchill, P. Dawson, S. Oldziej, A. Jagielska, H.A. Scheraga, D.J. Loskutoff and H.J. Dyson (2004). *Biochemistry* **43** 6519-6534. [An extensive collaboration between members of several Scripps departments, as well as with Scheraga at Cornell University. This structure determination proved to be exceedingly difficult. I was responsible for all of the data interpretation, structure calculations and paper writing.] (11 citations)
  167. The CBP/p300 TAZ1 domain in its native state is not a binding partner of MDM2. T. Matt, M.A. Martinez-Yamout, H.J. Dyson and P.E. Wright (2004). *Biochem. J.* **381**, 685-691. [I was involved in data interpretation and writing the paper.] (11 citations)
  168. The LEF-1 HMG domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. J.J. Love, X. Li, J. Chung, H.J. Dyson and P.E. Wright (2004). *Biochemistry*

- 43**, 8725-8734. [I was involved in synthesizing a large amount of experimental data over a number of years to write the paper.] (11 citations)
169. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. R.M. Borges, N.K. Goto, G.J.A. Kroon, H.J. Dyson and P.E. Wright (2004) *J. Mol. Biol.* **340**, 1131-1142. [I was involved at all stages of this project, and mostly wrote the paper.] (37 citations)
  170. The zinc-dependent redox switch domain of the chaperone Hsp33 has a novel fold. H.-S. Won, L. Y. Low, R. N. De Guzman, M.A. Martinez-Yamout, U. Jakob and H.J. Dyson (2004). *J. Mol. Biol.* **341**, 893-899. [See ref. 165. I was responsible for all aspects of this project.] (9 citations)
  171. ZZ Domain of CBP – an unusual zinc finger fold in a protein interaction module. G.B. Legge, M.A. Martinez-Yamout, D. Hambly, T. Trinh, B.M. Lee, H. J. Dyson and P.E. Wright (2004). *J. Mol. Biol.* **343**, 1081-1093. [I was heavily involved in the structure determination, data interpretation and paper writing.] (12 citations)
  172. Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. R.P. Venkitakrishnan, E. Zaborowski, D. McElheny, S.J. Benkovic, H.J. Dyson and P.E. Wright (2004). *Biochemistry* **43**, 16046-16055. [I was responsible for collating the experimental data and writing the paper.] (23 citations)
  173. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. R.N. De Guzman, J.M. Wojciak, M.A. Martinez-Yamout, H.J. Dyson and P.E. Wright (2005). *Biochemistry*, **44**, 490-497. [I assisted in writing the paper] (2 citations)
  174. Generation of native-like models from limited NMR data, modern force fields and advanced conformational sampling. J. Chen, H.-S. Won, W. Im, H.J. Dyson and C.L. Brooks, III (2005). *J. Biomol. NMR* **31**, 59-64. [For this paper I and my postdoctoral fellow Won provided experimental input for this theoretical study.] (7 citations)
  175. Intrinsically unstructured proteins and their functions. H.J. Dyson and P.E. Wright (2005). *Nature Reviews* **6**, 197-208. [An extremely influential invited review written by me, edited by Wright.] (222 citations)
  176. Elucidation of the protein folding landscape by NMR. H.J. Dyson and P.E. Wright (2005). *Methods Enzymol.* **394**, 299-322. [Invited review, written by me, edited by Wright.] (20 citations)
  177. Backbone and side chain <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N assignments for *Escherichia coli* SdiA1-171, the autoinducer binding domain of a quorum sensing protein. Y. Yao, M.A. Martinez-Yamout and H.J. Dyson (2005). *J. Biomol. NMR* **31**, 373-374. [I originated the project, planned the experiments, supervised the work and wrote it up.] (2 citations)
  178. Enhanced picture of protein folding intermediates using organic solvents in HD exchange and quench flow experiments. C. Nishimura, H.J. Dyson and P.E. Wright (2005). *Proc. Natl Acad. Sci. USA* **102**, 4765-4770. [Wright and I initiated this project, I helped analyze the data and wrote the paper.] (10 citations)
  179. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. D. McElheny, J.R. Schnell, J. Lansing, H.J. Dyson and P.E. Wright (2005). *Proc. Natl. Acad. Sci. USA* **102**, 5032-5037. [I interpreted data and wrote the paper.] (29 citations)
  180. Sequence determinants of a protein folding pathway. C. Nishimura, M.A. Lietzow, H.J. Dyson and P. E. Wright (2005). *J. Mol. Biol.* **351**, 383-392. [Wright and I initiated this project. I mad major contributions in the synthesis of the large amount of experimental data and in writing the paper.] (7 citations)
  181. Solution structure of the N-terminal zinc fingers of the *Xenopus laevis* double-stranded RNA-binding protein ZFa. H. M. Möller, Maria A. Martinez-Yamout, H. J. Dyson and P. E. Wright (2005). *J. Mol. Biol.* **351**, 718-730. [I supervised much of the structure determination, synthesized the results and co-wrote the paper.] (2 citations)

182. Structural basis for cooperative transcription factor binding to the CBP coactivator. R.N. De Guzman, N. Goto, H.J. Dyson and P.E. Wright (2006). *J. Mol. Biol.* **355**, 1005-1013. [I helped interpret the data and write the paper.] (5 citations)
183. NMR solution structure of the peptide fragment 1-30, derived from unprocessed mouse doppel protein, in DHPC micelles. E. Papadopoulos, K. Oglęcka, L. Mäler, J. Jarvet, P.E. Wright, H.J. Dyson, A. Gräslund (2006). *Biochemistry* **45**, 159-166. [A collaborative effort with the Gräslund lab (Stockholm University). My contribution was in planning the experiments and in the interpretation of data.] (5 citations)
184. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. C. Nishimura, H.J. Dyson and P.E. Wright (2006). *J. Mol. Biol.* **355**, 139-156. [Wright and I planned the experiments and interpreted the data. I wrote the paper.] (11 citations)
185. Structure of the *Escherichia coli* quorum sensing protein SdiA: activation of the folding switch by acyl homoserine lactones. Y. Yao, M.A. Martinez-Yamout, T.J. Dickerson, A.P. Brogan, P.E. Wright and H.J. Dyson (2006). *J. Mol. Biol.* **355**, 262-273. [I was responsible for all aspects of this project. I originated the idea, planned the experiments, interpreted the data and wrote the paper.] (8 citations)
186. Induced fit and “lock-and-key” recognition of 5S RNA by zinc fingers of transcription factor IIIA. B.M. Lee, J. Xu, M.D. Gearhart, B.K. Clarkson, M.A. Martinez-Yamout, H.J. Dyson, D.A. Case, J.M. Gottesfeld, and P.E. Wright (2006). *J. Mol. Biol.* **357**, 275-291. [I was heavily involved in the structure determination and in the interpretation of results, as well as in writing the paper.] (5 citations)
187. The reduced, denatured somatomedin B domain of vitronectin refolds into a stable, biologically active molecule. Y. Kamikubo, G. Kroon, S.A. Curriden, H.J. Dyson and D.J. Loskutoff (2006). *Biochemistry* **45**, 3297-3306. [I and my lab (Kroon) supplied extensive NMR data to this project, and played a significant role in drafting the paper.] (2 citations)
188. According to current textbooks, a well-defined three-dimensional structure is a prerequisite for the function of the protein. Is this correct? H.J. Dyson and P.E. Wright (2006) *IUBMB Life* **58**, 107-109. [Invited review, written by me, edited by Wright.] (3 citations)
189. Localization of sites of interaction between p23 with Hsp90 in solution. M.A. Martinez-Yamout, R.P. Venkitakrishnan, N.E. Preece, G. Kroon, P.E. Wright and H.J. Dyson (2006). *J. Biol. Chem.* **281**, 14457-14464. [I originated the idea, planned the experiments, interpreted the data and wrote the paper.]
190. An NMR perspective on enzyme dynamics. D.D. Boehr, H.J. Dyson and P.E. Wright (2006). *Chem. Rev.* **106**, 3055-3079. [Invited review. Boehr wrote the paper, Wright and I edited.] (6 citations)
191. The dynamic energy landscape of dihydrofolate reductase catalysis. D.D. Boehr, D. McElheny, H. Jane Dyson, P.E. Wright (2006). *Science* **313**, 1638-1642. [This prominent paper provides a synthesis of a large number of experiments and a great deal of analysis, providing an overall view of the role of dynamics in the catalytic cycle of dihydrofolate reductase. A commentary also appeared in *Science* **313**, 15 September 2006: 1586-1587. “Dynamic visions of enzymatic reactions”. M. Vendruscolo and C.M. Dobson. For this paper I was heavily involved in data interpretation and in writing the paper.] (16 citations)
192. The role of hydrophobic interactions in initiation and propagation of protein folding. H.J. Dyson, P.E. Wright, H.A. Scheraga (2006) *Proc. Natl. Acad. Sci. USA* **103**, 13057-13061. [For this synthesis of experimental and theoretical results, Scheraga and I each wrote part of the paper. Wright edited.] (1 citation)
193. Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination of p53. M. Kostic, T. Matt, H.J. Dyson and P.E. Wright (2006). *J. Mol. Biol.* **363**, 433-450. [I played a major role in the structure determination and in writing the paper.] (5 citations)

194. Mechanism of coupled folding and binding of an intrinsically unstructured protein. K. Sugase, H.J. Dyson and P.E. Wright. (2007) *Nature* **447**, 1021-1025. [For this prominent paper, I was involved in data synthesis and interpretation and in the writing of the paper. A commentary appears in *Nature* (2007) **447**, 920-921, "Proteins hunt and gather". D. Eliezer and A.G. Palmer III.] (1 citation)
195. Embryonic neural inducing factor Churchill is not a DNA-binding zinc finger protein: solution structure reveals a solvent-exposed  $\beta$ -sheet and zinc binuclear cluster. B.M. Lee, B.A. Buck-Koehntop, M.A. Martinez-Yamout, H.J. Dyson and P.E. Wright (2007). *J. Mol. Biol.* **371**, 1274-1289. [I was heavily involved in interpretation of data and in writing the paper.]
196. Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. R. Stoll, B.M. Lee, E.W. Debler, J.H. Laity, I.A. Wilson, H.J. Dyson and P.E. Wright (2007). *J. Mol. Biol.* (in press). [I was heavily involved in data interpretation and in writing the paper.]
197. NMR detection of adventitious xylose binding to the quorum-sensing protein SdiA of *Escherichia coli*. Y. Yao, T.J. Dickerson, M.S. Hixon, and H.J. Dyson (2007). *Bioorg. Med. Chem. Lett.* (in press). [I designed many of the experiments, interpreted the data and wrote the paper.]