NMR STUDIES OF FOLDING AND INTERNAL MOTIONS OF PROTEINS

A Thesis
Submitted for the Degree of

DOCTOR OF PHILOSOPHY
by

D. KRISHNA RAO

SCHOOL OF CHEMISTRY
UNIVERSITY OF HYDERABAD
HYDERABAD 500 046
INDIA
December 2007

నాన్నమ్మ కు అంకితం

CONTENTS

Chapter 3

Chapter 4

Complexity of Aromatic Ring-flip Motions in Proteins: Y97 Ring Dynamics in Cytochrome *c* Observed by Cross-relaxation Suppressed Exchange NMR **67 -96**

Chapter 5

School of Chemistry
University of Hyderabad
Central University P. O.,
Hyderabad 500 046
India

STATEMENT

I hereby declare that the work embodied in this dissertation is the result of the investigation carried out by me in the School of Chemistry, University of Hyderabad, Hyderabad, under the supervision of **Prof. Abani K. Bhuyan.**

D. KRISHNA RAO

December, 2007

STATEMENT VERIFIED

(Prof. ABANI K. BHUYAN)

PROJECT SUPERVISOR

School of Chemistry
University of Hyderabad
Central University P. O.,
Hyderabad 500 046
India

CERTIFICATE

Certified that the work embodied in this thesis entitled **"NMR STUDIES OF FOLDING AND INTERNAL MOTIONS OF PROTEINS"** has been carried out by Mr. D. Krishna Rao under my supervision, and the work has not been submitted elsewhere for a Degree.

Prof. ABANI K. BHUYAN
(THESIS SUPERVISOR)

Dean
School of Chemistry

ACKNOWLEDGEMENTS

I express my deep sense of gratitude and profound thanks to **Prof. Abani K. Bhuyan** for all of his support and guidance over the years. It has been an honor to work with and learn from him.

I wish to express my sincere gratitude to Prof. G. R. Desiraju, Prof. E. D. Jemmis (Former Deans) and Prof. M. Periasamy (Dean, School of Chemistry) for providing all necessary facilities to work in this School, and to other faculty members for direct and indirect help.

It has been a pleasure to work with such a fine group of people over the years: Rajesh, Prakash, Yadaiah, Nageswar, Harish, and Umakar. I am extremely thankful to all of them for stimulating discussions, timely help, and morale support. I thank Mr. S. Satyanarayana for his constant encouragement and technical support in the NMR operation. I thank to all my M. Sc classmates for their moral support. I am very thankful to all members of our school: research scholars, M.Sc., students and non-teaching staffs. I like to thank the University authorities for providing necessary facilities. Research fellowships and financial support from CSIR, DST, and DBT Government of India, are gratefully acknowledged.

I thank all my teachers who taught me and impressed me at all stages of my career. In particular, Mr. Shankaraiah, Mr. Seshigiri Rao, Mr. G. V. Linga Reddy, Dr. Subhash Chander Bose and Dr. Srinivasa Chowdary for their guidance and encouragement even in my failures.

I thank my listless number of friends love and support me for what I am. Particularly M. Harish, Vijay, Haneef, Amar, Seetharamulu, S. Ramababu, Sunitha, Laxmi, Prasad, Subhash, Ganesh, Sunil, MS, D. Rambabu, Upender, Srinivas M.

I acknowledge the timely helps of Mr. Manjunath and Chandu from Bruker, India. I thank all the staffs in CMSD, UoH for their kind help, especially Rajender and Fasi.

I am very grateful to my parents for their blessings, love and concern about my career and success. I thank my brother for his affection and care. I thank my wife Dr. Nagaraja for her love and support during my research.

There is no word to thank people who dedicated their life to research. I salute them, without their intense work and inherent wisdom I could not indeed make this thesis.

D. Krishna Rao

ABBREVIATIONS

1. NMR	nuclear magnetic resonance
2. CD,	circular dichroism
3. BPTI	bovine pancreatic trypsin inhibitor
4. GdnHCl	guanidinium hydrochloride
5. GdnDCl	guanidinium deutrochloride
6. Cyt *c*	cytochrome *c*
7. Ferricyt *c*	ferricytochrom *c*
8. Ferrocyt *c*	ferrocytochrome *c*
9. Carbonmonoxycyt *c*	the carbonmonoxide complex of ferrocytochrome *c*
10. ΔG_D	Gibbs free energy of denaturation
11. $\Delta G°$	Gibbs free energy of denaturation in the absence of denaturant (unfolding free energy)
12. NCO	native state of carbonmonoxycytochrome *c*
13. TOF	time of flight
14. COSY	J- correlated spectroscopy
15. NOESY	nuclear Overhausser enhancement spectroscopy
16. EXSY	cross-relaxation suppressed exchange Spectrscopy
17. $\Delta G^{\#}$	barrier free energy

Symbols for different parameters are explained fully in the text

CHAPTER 1

Introduction

The mechanism by which a protein folds to its biologically active state after biosynthesis remains one of the central unresolved puzzle in natural science. How does a protein make the transition from a diverse ensemble of unfolded structures into a unique conformation in the native form? One possibility would be that all possible conformations are tried out to find the energetically most favorable one. How long would such a random search take?

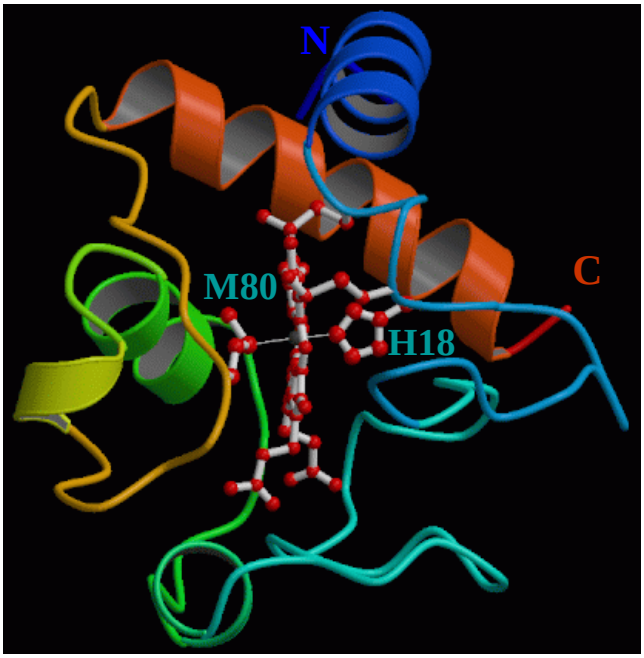


Figure 1 Structure of Horse cytochrome *c*.

Introduction

For a hypothetical small protein with 100 residues, Cyrus Levinthal calculated that if each residue can assume three different conformations, the total number of structures would be 3^{100}, which is equal to 5×10^{47}. If it takes 10^{-13} s to convert one structure into another, the total search time would be $5 \times 10^{47} \times 10^{-13}$ s, which is equal to 5×10^{34} s, or 1.6×10^{27} years. Clearly, it would take much too long for even a small protein to fold properly by randomly trying out all possible conformations. The enormous difference between calculated and actual folding times is called Levinthal's paradox (1). In the early 1960s, Anfinsen and co-workers observed that small protein molecules could fold spontaneously to the native state. This observation led to a thermodynamic hypothesis: the native state has the lowest free energy (2).

Proteins fold and unfold depending on the physical and chemical conditions of the environment. Protein folding can be studied in the test tube by dissolving a purified protein in aqueous solution and denaturing it by using chaotropic agents which are polar or charged molecules. Diluting this chaotropic molecules induces a spontaneous refolding of the polypeptide into its original, native fold. There are several spectroscopic methods available, including NMR, fluorescence labeling, and circular dichroism (CD), to study the structural changes associated with the folding reaction.

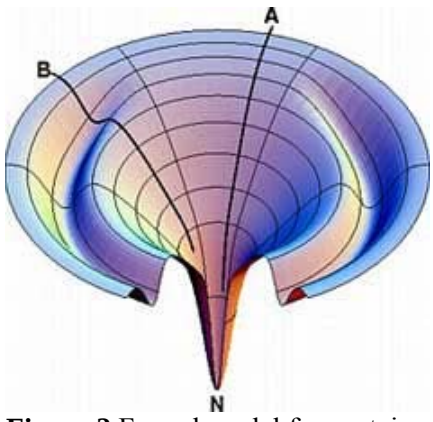


Figure 2 Funnel model for protein folding

Many models have been proposed to explain the protein folding problem. The framework model describes a stepwise mechanism to greatly narrow down the conformational search (3-5). The nucleation model suggests that tertiary structure forms through the formation of native secondary structure by only a few residues (e.g. a beta-turn, or the first turn of an alpha-helix), and structure propagates out from this nucleus (6, 7). The hydrophobic collapse model proposes that the native protein conformation forms by rearrangement of a compact collapsed structure (8, 9). The framework and hydrophobic collapse models suggest the formation of kinetic intermediates, whereas the nucleation model does not (10). Modifications of the nucleation model, whereby a diffused folding nucleus is formed and consolidated through the transition state concomitant with tertiary structure formation, leads to the nucleation-condensation model proposed by Fersht (10, 11).

A more general protein folding theory is the funnel theory (12, 13), which describes the energy profile of protein folding as a funnel with unfolded confirmations at the rim (Figure 2, 14). According to this view, the existence of intermediates is not essential for unfolded molecules to reach the native state; rather, these intermediates populate as a result of the ruggedness of the energy landscape (15). In earlier experimental studies, intermediates were commonly found, supporting the classical view (4). The neutral and alkaline pH forms of horse cytochrome *c*, which are greatly different in thermodynamic stability, have been used to demonstrate that folding is under thermodynamic control and that an initial search regime limits the folding speed. Biologically relevant folding time of small two-state proteins can be approximately equated to the time needed for

the nascent polypeptide to find a stable but less native like rate limiting barrier (16).

1.1 Molten globule in protein folding

Molten globule (MG) is a compact denatured state with a significant amount of native-like secondary structure, but largely disordered tertiary structure, which is known to be shaped during the highly cooperative process of folding. Understanding of structures and thermodynamics of such states should facilitate resolving the mechanism of folding (17). Molten globule can be obtained by titrating pH or heat denatured states of a protein with salt. Salt-induced formation of MG is explained in terms of preferential binding of counter ions to molten globule compared with the unfolded protein (18). Ptitsyn and coworkers represents (19, 20) MG was earlier taken to be a common early intermediate in the folding pathway, but later studies suggest that they correspond to late folding intermediates (21, 22). A highly ordered alkali molten globule state of ferrocyt *c* represents an off-pathway intermediate (chapter 3).

1.2 NMR Investigations of protein folding

Starting from its discovery in the laboratories of Purcell and Bloch in 1945-46, NMR has become the single most powerful technique in both chemistry and biology. The first NMR spectrum of a protein was recorded about 10 years after the discovery of NMR signals, but it was another 25 years to assign the resonances to individual atoms and the prospect of a complete structure determination (23). NMR spectroscopy began to make major contributions to

biology only when it was developed to the extent to solve macromolecular structures. In molecular biology and pharmaceutical research, 3D structures of biological macromolecules provide necessary information for detailed understanding of molecular functions and intermolecular interactions, and a foundation for protein engineering and drug design. NMR not only provides structural data, but also explores information on dynamics, conformation, and folding of a protein (24, 25). Along with many other applications, NMR became a powerful technique for drug discovery in pharmaceutical industry (26).

Most of the protein folding reactions are too fast to be monitored directly by NMR spectroscopy (27, 28), because the time required to record one 1D ^{1}H spectrum of a protein is at least 100 ms. Therefore, fast protein folding reactions have been measured by lineshape analysis of spectra recorded under equilibrium conditions. However progress has been made in recent years, and NMR is emerging as a major technique for real time observation of structure and dynamics of polypeptides in the course of protein folding (for example, 29).

The first application of NMR in protein folding involved site specific H/D exchange experiments (30), and the calculation of protection factors, P (the ratio of exchange rates in an unfolded model peptide to the interested protein) revealed the thermodynamics of local and global unfolding of the native state as well as intermediates (31). Protection factors in the hydrophobic core of native proteins can exceed 10^4, $P<5$ suggests that site is either extremely fluctuating between secondary structure and random coil or is entirely unfolded (32). Amide protons exchange with solvent by proton transfer reactions (33) that depend on a number of factors such as pH, temperature, neighboring side chains, and the involving

isotopes. After calibration of all these factors, H/D exchange experiments can define the presence or absence of hydrogen bonding at a large number of identifiable amide sites in a protein (34). Linderstrom-Lang and colleagues (35, 36) described the kinetic and thermodynamic relationships for a comprehensive analysis of structural unfolding reactions by using the measured H/D exchange rates.

Protein folding intermediates can be also described by pulse labeled hydrogen exchange. Baldwin and coworkers (37) have developed the pulse labeled hydrogen exchange experiments to describe the folding of many proteins. Kinetic H/D labeling has been applied to a number of proteins with similar and non-controversial results (38-45). In the case of two-state folding, all amides gain protection at the folding rate. In three-state folding, one generally sees behavior of showing the population of a discrete native-like intermediate (34). Sometimes intermediates revisit the unfolded state in order to reach its native confirmation, as seen in the refolding of alkaline ferrocyt *c* (46). Carbonmonoxycyt *c* refolds to native state via an off-pathway intermediate.

NMR line shape analysis and transverse relaxation measurements allow protein folding studies on a microsecond to millisecond time scale (47). Dynamic NMR opens up a large number of experimental opportunities to define the mechanism of fast protein folding reactions (48, 49). When a protein folds in a two state manner, any NMR resonance in a well dispersed spectrum and assigned in both folded and unfolded confirmations can be used to find the folding and unfolding rates. NMR can detect the specific contacts between residues as unfolded protein convert to native protein by monitoring the NOE enhancement

in a 2D experiment (50).The idea here is to allow the reaction to occur during the accumulation of a single 2D experiment. By recording the appropriate NMR spectrum, the kinetic history of the events occurring during the reaction can be deduced. Defining the structural and energetic properties of protein intermediates is important to understand the nature of protein folding pathways, and NMR is the only method that can provide the structural details, unless the intermediates can be trapped and crystallized for X-ray analysis. Recent advances in NMR relaxation dispersion methods along with the stopped-flow kinetic measurements can probe the folding and unfolding processes (51).

1.3 Side-chain and main-chain motions studied by NMR

All Biological processes that involve motions find their origin in protein dynamics (52). The mechanism by which a protein folds to its native confirmation is a highly dynamic process (53). Proteins without motion are biologically inactive. Enzyme catalysis is accompanied by the movement and rearrangement of protein segments (54). Biophysical experimental techniques, such as enzyme kinetics (55), Hydrogen exchange (56), kinetic studies of folding (53), NMR analysis of spin relaxation (57) and residual couplings (58), and single molecule spectroscopy (59) have provided evidences for different types of motions in proteins. Depending on their amplitude, internal motions are classified as small amplitude collective motions, large amplitude breathing modes, structural fluctuations, and deformations at sub-global level (60).

A special class of side-chain motions involving aromatic ring flips in proteins has been addressed by the use of NMR. These are 180^{o} flips between two

equivalent confirmations, crossing an energy barrier due to the interactions between ring and protein matrix. A free aromatic ring in solution is expected to show only two symmetric lines in NMR spectrum; this is because flipping process is too fast on NMR timescale. However side chain interactions and hydrogen bonding decelerate the aromatic ring-flip in protein. The transition from slow to fast rotation can be observed by changing the temperature. Aromatic side chains of BPTI (figure 7 of 64), shows two peaks in 1D NMR spectrum at low temperature and averaged peak appeared at high temperature. This suggests that the ring spent more time at equilibrium than the time spent to cross the barrier. From many years ring flip is considered as two site exchange. The rates of these two site exchange process can be determined quantitatively by line shape analysis (65), saturation transfer techniques (66) and EXSY (chapter 4, 67) Aromatic ring flip introduces conformational exchange-mediated magnetization transfer cross peaks into NOESY spectra (68). That suggests that if the ring motion is fast enough to transfer the longitudinal magnetization during NOESY mixing time, it can lead to a pure exchange cross peak. In such cases, NOESY spectrum is a combination of both NOE and exchange peaks. The rate process of such aromatic ring flip can be determined by introducing a pulse train during mixing time to suppress the cross-relaxation effects in NOESY. The complexity of these ring motions in proteins is elaborately

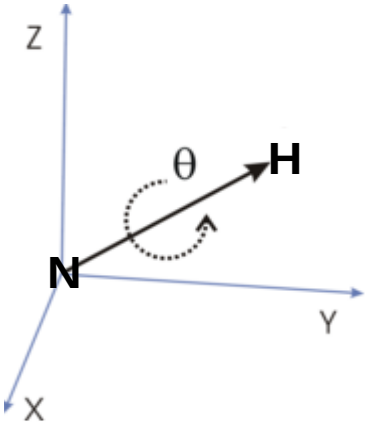


Figure 3 N-H vector in 3D making an angle θ with reference plane

explained in chapter 4.

Spin relaxation rate measurements provide information about site-specific fast internal dynamics that occur on subnanosecond time scale. NMR relaxation measurements are highly informative in finding which part of protein is flexible or rigid during protein-protein (69) or protein-ligand (70) interactions, or in the native state (71). ^{15}N relaxation studies on barstar (72) suggest rigid body movement of helix and twisting of C-terminal end of the β sheet. These large amplitude motions are perhaps relevant to making the complex with barnase.

Model free analysis (76) of relaxation rates (R1 & R2) and the heteronuclear NOE provides the order parameter (S^2) and effective correlation time (τ_e). S^2 and τ_e characterize the internal motions of bond vectors (N-H, C-H and C-C). Protein internal dynamics often changes with ligand binding that suggests that ligand binding involves a change in conformational entropy. Several Biological process including macromolecules interaction and protein folding involves change in entropy and enthalpy (77). Several independent studies suggest that changes in conformational entropy facilitate the ligand binding process. Stone and co-workers (79, 80) suggest that backbone conformational entropy may increase upon binding of a ligand. At the same time, there can be a change in the conformational entropy of side chains. Wand and co-workers (81) observed the increase in conformational entropy of Met 72 and Met 124 when a substrate binds to calcium-loaded calmodulin. This reduced entropy is perhaps compensated by enthalpy due to enhanced interactions with substrate. These resources in NMR with improved pulse sequences and labeling strategies make it practical to monitor the full range of time scales of protein motions.

1.4 References

1. Levinthal, C. *J. Chim. Phys.* **1968,** 65, 44–45
2. Anfinsen, C. B. *Science* **1973,** *181*, 223–230.
3. Ptitsyn, O. B. *Doklady Akademii Nauk SSSR*. **1973,** *210*, 1213-1215.
4. Kim, P. S and Baldwin, R. L. *Annu. Rev. Biochem.* **1990,** *59*, 631-660.
5. Dyson, H. J., and Wright, P. E. *Curr. Opin. Struc. Biol.* **1993,** *3,* 60-65.
6. Abkevich, V. I., Gutin, A. M., and Shakhnovich, E. I. *Biochemistry* **1994,** *33*, 10026-10036.
7. Wetlaufer, D. B. *Proc. Natl Acad. Sci. USA*, **1973,** *70*, 697-701.
8. Dill, K. A., Bromberg, S., Yue, K. Z., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. *Protein Sci.* **1995,** *4,* 561-602.
9. Ptitsyn, O. B. *Nat. Struct. Biol.*, **1996,** *3,* 488-490.
10. Fersht, A. R. *Curr. Opin. Struc.Biol.* **1997,** *7,* 3-9.
11. Fersht, A. R. *Proc. Natl Acad. Sci. USA*, **1995,** *92,* 10869-10873.
12. Leopold, P. E., M. M. J. N, and Onuchic. *Proc. Natl Acad. Sci. USA*, **1992,** *89,* 8721- 8725.
13. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. *Proteins: Struct. Funct. Genet.* **1995,** *21,* 167-195.
14. Dill, K. A., and Chan, H. S. *Nat. Struct. Biol.* **1997,** *4,* 10-19.
15. Bai, Y. *Biochemical and Biophysical Research Communications* **2003,** *305,* 785–788.
16. Bhuyan, A. K., Rao, D. K., Prabhu, N. P. *Biochemistry* **2005,** *44,* 3034-3040.
17. Goto, Y. and Fink, A. L. *Biochemistry* **1989,** *28,* 945-952.

18. Hamada, D., Kidokoro, S., Fukada, H., Tahahashi, K. and Goto, Y. *Proc. Natl. Acad. Sci.* USA, **1994,** *91,* 10325-10329.
19. Ptitsyn, O. B. *Ad*v. *Protein Chem.* **1995,** *47,* 83-229.
20. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. *FEBS Lett.* **1990**, *262*, 20-24.
21. Kuwajima, K., and Arai, M. *Mechanism of Protein Folding* (Pain, R. H., Ed.) **2000**, *2nd ed.,* pp 138-174, Oxford University Press, New York.
22. Mok, K. H., Nagashima, T., Day, I. J., Hore, P. J., and Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8899-8904.
23. Wuthrich, K., Wider, G., Wagner, G., and Braun, W. *J. Mol. Biol.* **1982,** *155,* 311-319.
24. Dyson, H. J., and Wright, P. E. *Annu.Rev. Phys. Chem.* **2000,** *47,* 369-395.
25. Kay, L. E. *Biochem. Cell Biol.***1998,** *76,* 145-152.
26. Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. *Science* **1996,** *274,* 1531-1534.
27. Balbach, J., Forge, V., Lau, W. S., van Nuland, N. A. J., Brew, K. and Dobson, C. M. *Science* **1995,** *274,* 1161-1163.
28. Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J. and Dobson, C. M. *Nat. Struct. Biol.* **1995,** *2,* 865-870.
29. Bhuyan , A. K. and Udgaonkar, J. B. *current science* **1999,** *77,* 942-950.
30. Wagner, G., and Wuthrich, K. *J. Mol. Biol.* **1982,** *160,* 343-361.
31. Englander, S. W., Sosnick, T. R., Englander, J. J. and Mayne, L. *Curr. Opin. Struc. Biol.* **1996,** *6,* 18-23.
32. Gruebele, M. *Annu. Rev. Phys. Chem.* **1999,** *50,* 485–516.

33. Eigen M. *Angew.Chem. Intl. Ed. English* **1964,** 3, 1–19.
34. Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000,** *29,* 213–38.
35. Linderstrom-Lang, K. U. *Symposium on Protein Structure.* **1958,** Edited by Neuberger A. London, Methuen 23-34.
36. Hvidt A, Nielsen S. O. *Adv. Protein Chem.* **1966,** *21,* 287-386.
37. Kim, P. S., Baldwin, R. L. *Biochemistry* **1980,** *19,* 6124–6129.
38. Bycroft, M., Matouschek, A., Kellis, J. T. Jr., Serrano, L., Fersht, A. R. *Nature* **1990,** 346, 488–490.
39. Jacobs, M. D., Fox, R. O. *Proc. Natl. Acad. Sci. USA* **1994,** *91,* 449–453.
40. Jennings, P. A., Wright, P. E. *Science* **1993,** *262,* 892–896.
41. Lu, J., Dahlquist, F. W. *Biochemistry* **1992,** *31,* 4749–4756.
42. Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M. and Fersht, A. R. *J. Mol. Biol.* **1992,** *224,* 837–845.
43. Mullins, L. S., Pace, C. N., Raushel, F. M. *Biochemistry* **1993,** *32,* 6152–6156.
44. Radford, S. E., Dobson, C. M., Evans, P. A. *Nature* **1992,** *358,* 302–307.
45. Roder, H., Wuthrich, K. *Proteins: Struc. Funct. Genet.* **1986,** *1,* 34–42.
46. Rao, D. K., Prabhu, N. P., Bhuyan, A. K *Biochemistry* **2006,** *45,* 8393-8401.
47. Zeeb, M. and Balbach, J. *Protein and Peptide Letters,* **2005,** 12, 139-146.
48. Huang, G. S. and Oas, T. G. *Proc. Natl. Acad. Sci.* USA, **1995,** *92,* 6878-6882.
49. Wang, M., Tang, Y., Sato, S., vugmeyster, L., Mcknight, C. J. and raleigh, D. P. *J. Am. Chem. Soc.* **2003,** *125,* 6032-6033.

50. Balbach, J., Forge, V., Lau, W. S., Jones, J. A., van Nuland, N. A. J. and Dobson, C. M. *Proc. Natl. Acad. Sci.* USA, **1997,** *94,* 7182-7185.
51. Teilum, K., Poulsen, F. M. and Akke, M. *Proc. Natl. Acad. Sci.* USA, **2006,** *103,* 6877-6882.
52. Akasaka, K. *Biochemistry* **2003,** 42, 10875-10884.
53. Roder, H., and Colon, W. *Curr. Opin. Struc. Biol.* **1997,** *7,* 15-28.
54. Ishima, R., and Torchia, D. A. *Nat. struct. Biol.* **2000,** *7,* 740-743.
55. Hammes, G. G. *Biochemistry* **2002,** *41,* 8221-8228.
56. Englander, S. W., and Mayne, L. *Annu. Rev. Biophys. Biomol. Struct.* **1992,** *21,* 243-265.
57. Kay, L. E., *J. Magn. Reson.* **2005,** *173,* 193-207.
58. Meiler, J., Peti, W., and Griesinger, C. *J. Am. Chem. Soc.* **2003,** *125,* 8072-8073.
59. Wazawa, T., Ishii, Y., Funatsu, T., and Yanagida, T. *Biophs. J.* **2000,** *78,* 1561-1569.
60. Karplus, M. *Methods Enzymol.* **1986,** *131,* 283-307.
61. Denisov, V. P., and Halle, B. *Faraday Discuss* **1996,** *103,* 227-244.
62. Halle, B., Denisov, V. P., and Venu, K. *Biological magnetic resonance* **1999,** *17*, 410-484.
63. Binsch, G. *J. Am. Chem.* **1969,** *91,* 1304-1309.
64. Wagner, G., and Wuthrich, K. *Methods in Enzymology* **1986,** *131,* 307-326.
65. Hattori, M., Hua Li, H., Yamada, H., Akasaka, K., Hengstenberg, W., Gronwald, W., and Kalbitzer, H. R. *Protein Science* **2004,** *13,* 3104-3114.

66. Campbell, I. D., Dobson, C. M., Moore, G. R., Perkins, S. J., and Williams, R. J. P. *FEBS Lett.* **1976,** *70,* 96-100.
67. Skalicky, J. J., Mills, J. L., Sharma, S., & Szyperski, T. *J. Am. Chem. Soc.* **2001,** *123*, 388-397.
68. Ernst, R. R., Bodenhausen, G., & Wokaun, A. *Principles of Nuclear Nagnetic Resonance in One and Two Dimensions* **1988** (Clarendon Press, Oxford).
69. Sahu, S. C., Bhuyan, A. K., Udgaonkar, J. B., and Houser, R. V. *J. Bio. NMR* **2000,** *18,* 107-118.
70. Yun, S., Jang, D. S., Kim, D. H., Choi, K. Y., and Lee, H. C. *Biochemistry* **2001,** *40,* 3967-3973.
71. Barbato, G., ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. *Biochemistry* **1992,** *31,* 5269-5278.
72. Wong, K-B., Fersht, A. R., and Freund, S. M. V. *J. Mol. Biol.* **1997,** *268,* 494-511.
73. Brutscher, B., bruschweiler, R., and Ernst, R. R. *Biochemistry* **1997,** *36,* 13043-13053.
74. Beeser, S. A., Oas, T. G., and Goldenberg, D. P. *J. Mol. Biol.* **1998,** *284,* 1581-1596.
75. Meekhof, A. E., and Freund, S. M. V. *J. Mol. Biol.* **1999,** *286,* 579-592.
76. Lipari, G., and Szabo, A. 1. *J. Am. Chem. Soc.* **1982,** *104,* 4546-4570.
77. Cavanagh, J., and Akke, M. *Nature Stru. Biol.* **2000,** *7,* 11-13.
78. Yang, D. and Kay, L.E. *J. Mol. Biol.*, **1996,** *263,* 369–382.
79. Zidek, L., Novotny, M. V., and Stone, M. J. *Nature Struct. Biol.* **1999,** *6,* 1118-1121.

80. Forman-kay, J. D. *Nature Struct. Biol.* **1999,** *6,* 1086-1087.

81. Lee, A. L., Kinnear, S. A., Wand, A. J. *Nature Struct. Biol.* **2000,** *7,* 72-77.

CHAPTER 2

Dad-end Intermediate(s) in the Refolding of Alkaline Ferrocytochrome *c*: a Pulse-labeling HX NMR Study

2.1 Abstract

Proteins meet with the stipulations of Levinthal. The classical mechanism involving a pre-transition search topology step can explain adequately how the biologically relevant time scale is achieved. Association of the classical mechanism of chemical kinetics naturally allows depiction of folding pathways with dead-end intermediates. By using the approach of hydrogen exchange (HX) pulse labeling in conjunction with NMR spectroscopy, it is shown here that carbonmonoxycytochrome *c* when allowed to refold from a strongly alkaline medium to neutral pH folds via a highly misfolded state. The misfolded state placed at the dead end of the folding pathway must then melt to the initial unfolded state to have a chance to refold to the correct native state.

2.2 Introduction

Emerging ideas of landscape perspective or the funnel model (1-6) were first used to interpret both two-state folding kinetics and accumulation of misfolded structures on the basis of size, slope, and roughness of energy landscapes (7, 8). Corroboration between experimental and simulation results has also been shown regarding the fractional φ values for CI2 (9). The landscape view holds that the funneled organization of the energy landscape dominates the folding kinetics, because the great variety of detailed mechanisms existing on the

funnel influences the folding trajectories (6). The emerging ideas, however, often appear to undermine some of the strongly established empirical patterns of folding kinetics. These patterns, including the energy and the location of the folding barrier, that form the key elements of the classical view of folding, are not reproduced in the new paradigm. Overlooking empirical results is difficult. A number of recent studies have demonstrated the involvement of sequential intermediates in the folding pathway of cyt *c* (10-13). It has also been shown that the misfold organization of the folding intermediate is associated with an optional event, and it does not negate the classical description of folding (1). Application of the classical paradigm to a number of proteins exhibiting diverse folding behavior has been discussed (14).

Supported by empirical observations on ferrocytochrome *c* (ferrocyt *c*), this work shows extreme misfolding when alkaline ferrocytochrome *c* liganded with CO is allowed to refold to a final pH of 7. Earlier reports of misfold organization of a kinetic intermediate of ferricytochrome *c* (ferricyt *c*) also raised the question of how conducive intermediates are to folding (15, 16). Whether it is conducive or not, polypeptides have the option to make mistake and misfold. If a misfolded state is blocked from further folding, it has to go back to the initial unfolded state to start afresh. Results presented here elucidate this situation.

2.3 Materials and Methods

Cyt *c* was from Sigma (Type VI). GdnHCl and sodium dithionite were obtained from USB and Merck, respectively. Experiments were done in strictly anaerobic atmosphere at 22°C using 0.1 M sodium phosphate buffer for pH 7

conditions, and NaOH or NaOD (with or without 1 mM CAPS) for pH 12.7 conditions. Solutions contained 0.5-3 mM freshly prepared sodium dithionite, and experiments were completed within 2 hours of exposing the protein to high pH. High-pH fluorescence changes were corrected by using NATA fluorescence.

2.3.1 Equilibrium unfolding

Cyt *c* solutions with or without 1mM CO, prepared in the 0-7 M range of GdnHCl or 7-13.25 range of pH, were deaerated and reduced under nitrogen with ~2 mM sodium dithionite, and incubated in tightly capped quartz cuvettes or rubber-capped small glass tubes for ~30 minutes. Tryptophan fluorescence emission spectra (ex: 280 nm) were taken in a FluoroMax-3 instrument (Jobin-Yvon, Horiba). Optical absorption spectra were recorded in a Cary 100 (Varian) spectrophotometer. Data were analyzed using the standard two-state equations (17).

2.3.2 Stopped-flow kinetics

Cyt *c* initially dissolved at pH 13 in the presence or absence of 4 M GdnHCl, was reduced under nitrogen by adding a concentrated solution of sodium dithionate to a final concentration of ~3mM. The solution was then saturated with CO under 1 atm of the gas. The resulting carbonmonoxycyt *c* is unfolded under the conditions used. Refolding was initiated by eight-fold dilution of the protein solution with the refolding buffer. The final protein concentration in the refolding mixture was in 5-50 μM range. For unfolding experiments, the native state of carbonmonoxycyt *c* (NCO) was prepared by manual dilution of the alkali-unfolded protein with the refolding buffer in 1:40 ratio. NCO was then unfolded by eight-fold dilution into the unfolding buffer (0.1 M phosphate, 3 mM

sodium dithionite, pH 7) containing different concentrations of GdnHCl. The final protein concentration in these experiments was ~4 μM. A Bio-Logic SFM 400 instrument was used for these experiments.

2.3.3 Hydrogen exchange pulse labeling and NMR spectroscopy

Cyt *c* with fully deuterated backbone amide sites was dissolved in NaOD at pD 12.9 with or without 3 M GdnDCl, reduced with sodium dithionite (3 mM), and unfolded by adding 1 atm CO. Initial protein concentration was 6 mM. Refolding was initiated by 6-fold dilution into a H_2O buffer containing 75 mM phosphate, pH 6.4. The final pH of the refolding medium was 7.3. After variable times of refolding, the solution was mixed with an equal volume of the H_2O pulse buffer containing 50 mM CAPS, pH 10.4. The pH of the mixture at this stage was 9.6. After 50 ms of pulse time, the solution was combined with 0.7 volume of an H_2O quench buffer containing 0.5 M sodium acetate and 0.35 M ascorbate, pH 5. The pH at this stage was 5.05. The quenched solution was washed and concentrated at 5°C using a D_2O buffer consisting of 7 mM citrate and 7 mM ascorbate, pH 5. The samples were then subjected to NMR analysis by recording magnitude COSY and TOCSY spectra in 400 MHz (Bruker) or 600 MHz (Varian) spectrometers. Proton occupancies at individual amide sites were calculated by using the cross peak intensities of a control unfolded sample whose amide sites were labeled maximally. Rates were extracted from plots of proton occupancies as a function of refolding time. Pulse-labeled samples were prepared using the Bio-Logic instrument in the quench-flow mode.

For samples not subjected to pulse labeling, D_2O solutions of 1 mM cyt *c*, the amides of which were preexchanged, were adjusted to pH 7 (0.1 M

phosphate) or pH 12.7 using NaOD. Unfolded protein was prepared in GdnDCl. For reduction solid sodium dithionite was added under nitrogen to obtain a final concentration of 20 mM. The tubes were sealed with sleeved rubber stoppers. Spectra were taken at 22°C in a 400 MHz Bruker spectrometer (AV400).

2.4 Results

Ferrocytochrome *c* (cyt *c*) unfolds in alkaline aqueous medium when carbonmonoxide is allowed to bind to its ferrous heme iron (Figure 1).

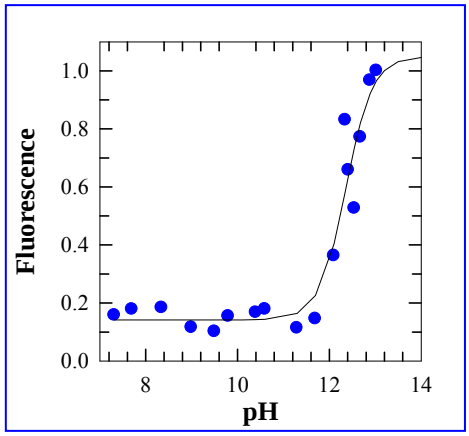


Figure 1 The alkaline pH-induced unfolding of ferrocyt *c* in the presence CO. The pH-midpoint for the transition is ~12.5.

The unfolding of the resultant CO-bound protein, called carbonmonoxycytochrome *c* (Cyt-CO) is substantial even in the aqueous medium, because the addition of the denaturant guanidinium hydrochloride (GdnHCl) to the protein solution at pH 13 or higher does not produce any considerable change in the fluorescence emission by lone tryptophan (W59). This however suggests little that the alkaline Cyt-CO is random coil-like. Completeness of unfolding is a controversial issue and cannot be shown unequivocally with a number of techniques.

2.4.1 Neutral and alkaline forms of ferrocyt *c*

On the other hand, ferrocyt *c* at pH 12.7 (alkaline ferrocyt *c*) is substantially structured. The NMR spectrum in Figure 2a shows chemical shift dispersion characteristic of a secondary- and tertiary-structured protein. Minor changes in both intensities and chemical shifts of resonances, relative to those seen for the native protein at pH 7 (neutral ferrocyt *c*), indicate some readjustments of side chains. The spectrum does not hint at molecular aggregation. In fact, TOF mass spectra and extensive ligand- binding experiments (data not shown) indicate no deamidation or aggregation at alkaline pH. A large number of NOE peaks seen at neutral pH appear at alkaline pH also, albeit somewhat shifted (Figure 2b, c). Figure 2a also compares the GdnDCl-unfolded NMR spectra of neutral- and alkaline ferrocyt *c*. Chemical shift dispersion is lost in both, but differences between the two are apparent, indicating differences in unfolded chain configuration. The latter also shows base-catalyzed fast exchange of residual hydrogens of GdnDCl with residual water protons coalesce as the exchange rate approaches the difference of the two chemical shifts.

Dead-end intermediate

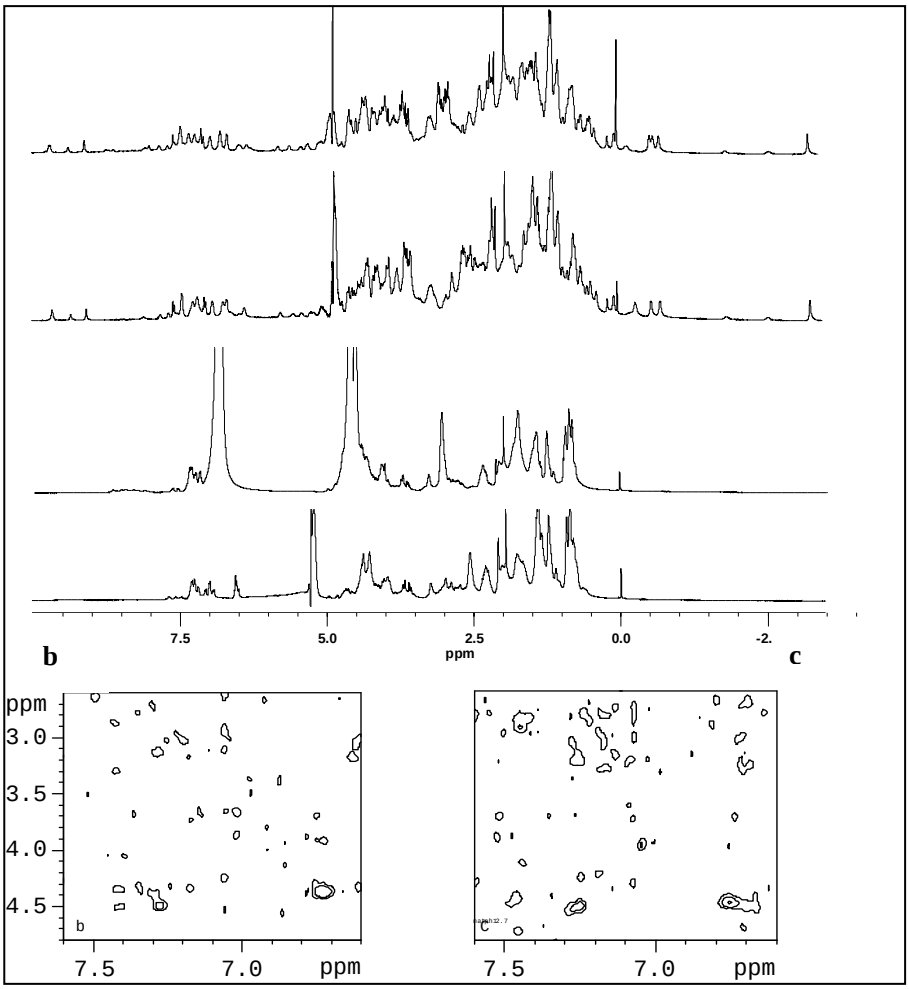


Figure 2 NMR spectra of ferrocytochrome *c*. (a) Native states at pH 7 (N) and pH 12.7 (N12.7), and GdnDCl-unfolded states at pH 7 (U 7) and pH 12.7 (U112.7). (b and c) Regions native-state NOESY spectra of neutral and alkaline ferrocyt *c*.

2.4.2 Stability of neutral and alkaline ferrocyt *c*

Figure 3a compares GdnHCl unfolding transitions of the two forms. Two-state analysis of data yields a $\Delta G_D°$ value of 18(±0.3) kcal mol^{-1} and 3(±0.5) kcal mol^{-1} for neutral and alkaline ferrocyt *c* respectively. Such a large difference in $\Delta G_D°$ for the neutral and alkaline ferrocyt *c* indicates drastic reduction of aqueous stability of the latter. The equilibrium *m*-values, m_g, for the neutral and alkaline forms are 3.6(±0.1) and 1.2(±0.1) kcal mol^{-1} M^{-1}, respectively. Since the native-state NMR spectra indicate that significant portions of secondary and tertiary structures are similar in neutral and alkaline ferrocyt *c*, the observed disparity in the m_g value is likely due to relatively small solvent accessible surface area in the denatured state of the alkaline form.

2.4.3 Folding and unfolding kinetics

In general, kinetic traces were fitted to 1-exponential. In some cases 2-exponentials were needed to improve the overall quality of fits. The slow minor phase, the amplitude of which averages to ~10% of the observed signal, represents most likely the fraction of the oxidized protein. Figure 3b shows denaturant dependences of the logarithm of the fast refolding (unfolding) phase for the two proteins. In each case, the relaxation minimum closely matches the midpoint of the individual equilibrium unfolding transition (Figure 3a).

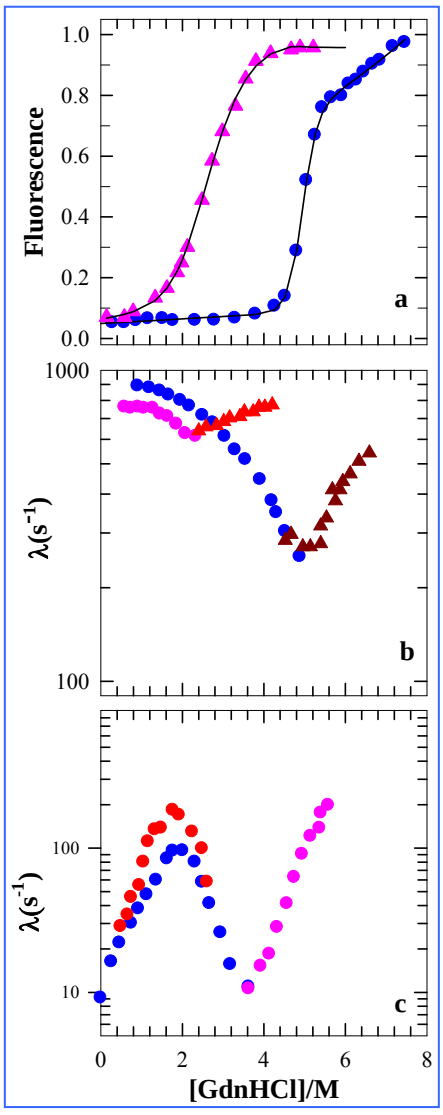


Figure 3 (a) GdnHCl unfolding of ferrocytochrome c at pH 7 (●), and pH 12.7 (▲). (b) GdnHCl dependence of rates for the pH 7→pH 7 (●, folding and ▲, unfolding) and pH 12.7→pH12.7 (●, folding and ▲, unfolding) equilibria. (c) Observed relaxation rate constants versus GdnHCl. The symbols (●) and (●) represent refolding rates when the alkali-unfolded CO-bound protein solution contained 0 M and 4 M GdnHCl, respectively. (●) represents the unfolding rates.

2.4.4 Chevron inversion in strongly native like condition

Curiously enough, when Cyt-CO is allowed to refold from pH 13 to pH 7 with or without 4 M GdnHCl in the intial unfolded solution, the measured millisecond refolding rate constant under strongly native like conditions (<2 M GdnHCl) decreases linearly with decreasing concentrations of denaturant, producing a sharp inversion in the folding limp of the chevron (Figure 3c). The accentuated inversion of the rate-denaturant profile is taken as an evidence for extensive misfolding: the misfolded species is increasingly stabilized

and hence accumulates as the denaturant concentration in the refolding medium falls lower. The misfolded product cannot, however, achieve forward folding because of the preponderance of non-native interactions. In this sense it is a dead end intermediate, and it needs to unfold extensively in order to find a native like topology which can facilitate correct folding.

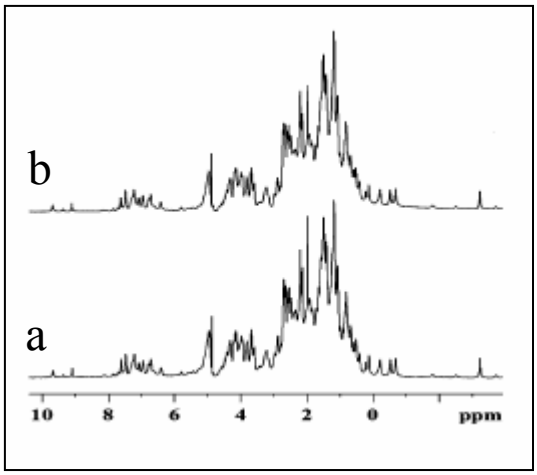


Figure 4 The protein does not aggregate or dimerize at different stages of folding. The nmr spectrum of ferrocyt *c* at pH 12.9 soon after sample preparation (a), and 38 h later (b).

Here, it is necessary to show that the observed misfolding is not related to protein dimerization or aggregation at any stage of folding, since short-lived aggregates may cause a downward curvature in the folding limb of the chevron plot (18). Protein aggregation at extreme pH was checked by several methods, including NMR and optical absorption (data not shown). Since, the sharp and well resolved resonances of the NMR spectrum of alkaline ferrocyt *c* (pH 12.9) do not change at all even after tens of hours of sample preparation (figure 4), the evidence is compelling that the protein is monomeric all the time.

Dead-end intermediate

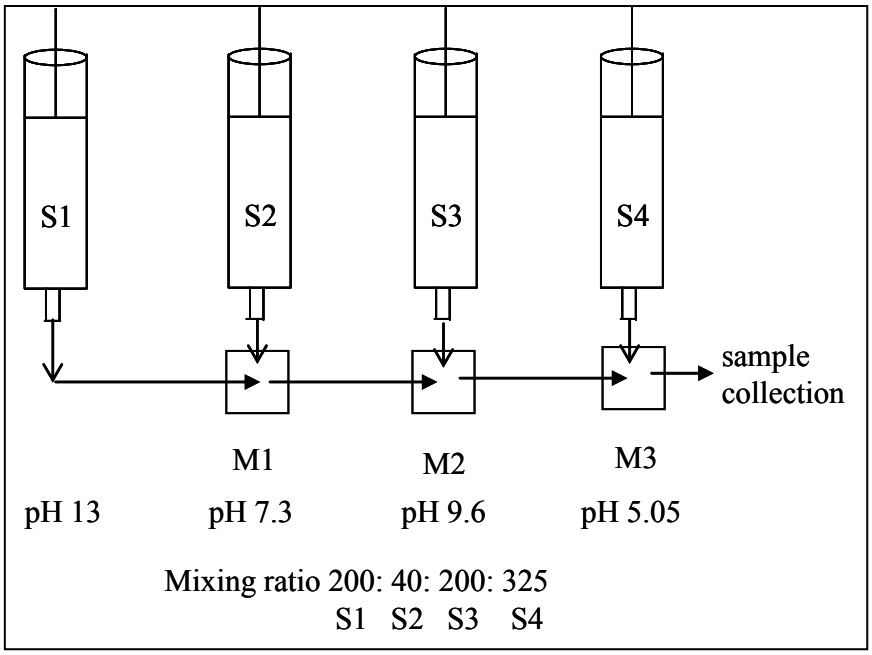


Figure 5 Schematic representation of the mixing sequence for the hydrogen exchange pulse labeling experiment. The mixing ratio was designed to achieve the pH values at different stages indicated.

2.4.5 Pulse labeled hydrogen exchange kinetics

The misfolding is better shown by using the pulse labeled hydrogen exchange strategy for probing structure formation during refolding. A Cyt-CO solution unfolded in the alkaline medium in the presence of GdnHCl was allowed to refold to a final condition of pH 7 and 0.5 M GdnHCl using the quench-flow protocol as shown in Figure 5. The pH of the solutions at different stages of refolding and the delay-times (aging time) were standardized to maximize labeling during refolding but to minimize exchange in the final refolded state.

Control samples to check for pulse leakage were prepared by manual mixing instead of using the quench-flow mixing. Both TOCSY and COSY spectra were run on the concentrated and washed samples. Figure 6 shows the NH fingerprint regions of representative COSY spectra depicting the H/D exchange as folding progresses.

For all NMR-resolved NH sites, only moderate protection against proton labeling was seen at ~14 ms of refolding time, although at 55ms the proton occupancy dropped to about 60-70% in a lone kinetic phase corresponding to the $I \rightarrow N$ transition via U.

The rate extracted from the time course of protection against H-exchange labeling at four representative NH sites is $30 \pm 2\ s^{-1}$ (figure 6D). This rate is also seen in stopped-flow interrupted refolding experiments (data not shown). The single rate for protection suggests that no transient structure accumulates during the folding.

Dead-end intermediate

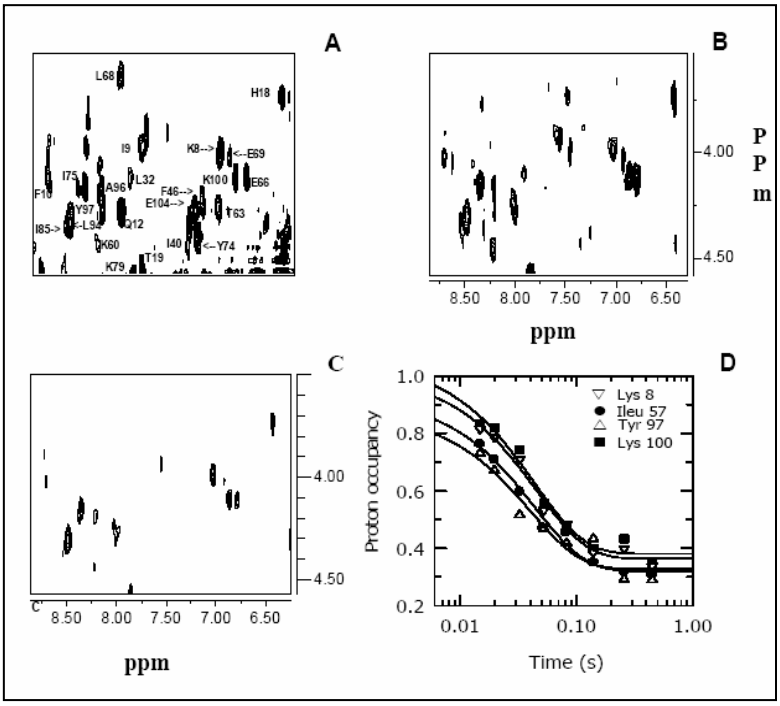


Figure 6 NMR analysis of representative samples prepared by hydrogen exchange pulse labeling. (*A*) NH-C_αH region of the magnitude COSY (*J*-correlated) spectrum of ferrocyt *c* at pH 5.1. (*B*) As in (*A*) for a pulse-labeled sample at 14 ms of refolding time. (*C*) As in (*B*) at 55 ms of refolding time. (*D*) Time course of protection against hydrogen exchange for a set of four representative amide sites. Single-exponential fits to the data give 30 ± 2 s^{-1} for the refolding rate constant.

2.5 Discussion

2.5.1 Folding models and the reality

The principles and the relaxation mechanisms by which the unfolded polypeptide acquires the native state are not understood. There are models, classical and contemporary, for the description of the folding phenomenon. While the classicals are deeply rooted into the principles of chemical kinetics and the law of mass action, the contemporaries are fundamentally based on the principles of statistical mechanics. It must be kept in mind though that a model, whatever is the basis of that, is simply a representation of the event. It could be considered a principle provided it facilitates rational interpretation of experimental results overwhelmingly. Thus, experimental results of folding kinetics for numerous proteins are viewed in model perspectives in order to examine the success of a model. The task here is to examine which model best explains the observation of the misfolded protein elucidated in this work.

2.5.2 Theory-based models and protein misfolding

These models subtly elude the misfolding phenomenon. Although the funnel model held right from the beginning that proteins in the folding run may become trapped in shallow minima at later stages because of certain non-native contacts (19, 20), occurrence of large scale misfolding is not explained. In the case of cytochrome *c* for example, the NCO species (see results) is actually a trapped misfold, because structurally NCO and N are virtually identical except for the heme ligand. The ligand is the extrinsic CO for the former and M80 for the latter. The NCO→N folding transition whose rate is limited by the thermal

dissociation of CO has actually been studied by laser photolysis and interpreted in terms of glassy trapped dynamics as held by the funnel model (21). However, the large scale misfolding in the refolding of alkaline cyt-CO shown here is difficult to interpret in the funnel perspective.

Quick and large-scale misfolding is difficult to envisage in the theoretical models. Also, these models refrain from using on-pathway and off-pathway terminologies. Nonetheless, an intermediate may be called off-pathway if the structure does not increasingly resemble the native state while on the folding run (2). For example, numerical results of Monte Carlo simulations of protein-like polymers on a lattice (22-27), and molecualrt dynamics simulations of heteropolymers in continuum (28, 29) show that a heteropolymer/protein quickly acquires a compact structure carrying native as well as non-native contacts. In the course of further evolution, some native-like structural elements are dissolved while others are formed. Experimental results for folding of β-lactoglobulin for which a non-native α-helical conformation formed during the burst folding is transformed to the native β-sheet conformation have been interpreted in this scenario (30). Similarly, transient formation and disappearance of non-native β-sheet during the folding of α-lactalbumin has been reported (31). Thus, overwhelming unfolding of a misfolded state appears not considered by theoretical models.

2.5.3 Classical models and protein misfolding

Being based on the principle of microscopic reversibility and the law of mass action, all of such models allow misfolding of various degrees. In the

classical realm, an off-pathway intermediate can be operationally defined as the one that has to undergo, because of its arrival at the dead-end, a large scale unfolding to resemble the initial unfolded state both structurally and energetically. In the present case, a comparison of the chevron gradients for the unfolding and pretransition regions of GdnHCl concentration (Figure 3c) illustrates this point. An on-pathway intermediate, on the other hand, does not have to unfold structurally except for occasional rupture of a few long range interactions or rearrangements of folding chain segments such that it can continue to fold to the native state. Both off- and on-pathway intermediates can result from the initial chain collapse and the formation of local interactions such that both states are lower in energy relative to the unfolded state, and both of them would appear to reduce the conformational search problem.

It should be mentioned though that the phenomenon of the deviation from the predicted linear dependence of logarithm of folding rate constant on denaturant can be both qualitatively interpreted and quantitatively modeled in terms of on-pathway as well as off-pathway placement of an intermediate (32, 33). Whether intermediates are on the pathway or are merely side precuts is one of the longest-running controversies (34). While the on-pathway intermediates would appear conducive to folding because they are expected to speed up folding by guiding the polypeptide chain to the native state, the role of off-pathway intermediates becomes less obvious. Here, experimental record of a fundamental kinetic criterion, namely, inversion of the denaturant dependence of the observed folding rate constant reveals the off-pathway dead-end mechanism.

Dead-end intermediate

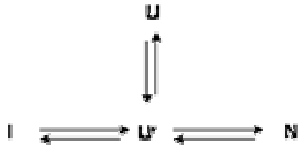
2.5.4 How could a polypeptide fold to the dead end?

The arrival of a protein chain at the dead-end is possibly a consequence of unsuccessful search for the correct transition state for folding. This possibility is best illustrated by the topology search model of folding advanced by several laboratories in recent years (for example, 14-16, 35, 36). As soon as a refolding condition is created, the unfolded chains (U) initially at low-energy equilibrium states are driven instantaneously to new higher energy level(s) (37). The excited chains, now in a nonequilibrium state, respond to the new solvent condition where only a few, if any, denaturant molecules can affect the nascent polypeptides (U′). This ultrafast relaxation process that may involve collapse or chain contraction, desolvation, and/or redistribution of backbone dihedral φ,ψ angles (16, 35, 38, 39) are missed out in millisecond stopped-flow measurements, and is termed burst folding phase. Because the U→U′ is a submillisecond process and the experimentally observed folding time falls in millisecond to seconds range, it is the U′→N transition that sets the folding time. Further, for apparently two-state fast-folding proteins like ferrocyt *c* (40, 35) and Csp B (41), where post-barrier kinetic intermediates do not accumulate, folding from the transition state to the native state will be rapid and barrier-free. These considerations imply that almost the entire length of the observed folding time is the time needed for the U′ state to find a transition state. It consumes time because of the search-find nature of the process (11, 14, 16, 35). It is thermodynamically constrained because a stable transition state that can support forward folding is to be found. Search continues until a relatively low-energy transition state stabilized by a few correctly formed native-like tertiary contacts is achieved. This is the essence of the topology search

model (14, 16, 35, 42) that draws support from observations that folding rates and mechanisms often appear to be largely determined by the native-state topology or contact order (43, 44). Native-like chain topology stipulates certain long-range interactions, not local contacts as viewed by theory (37, 45). In addition, being supported by only a few tertiary contacts the transition state barrier must be less organized and therefore more unfolded-like, consistent with $\alpha<0.3$. Finally, because the topology search is an unsystematic process, the possibility of error commitment is not excluded. A wrong transition state should lead to an off-pathway structure, a convincing experimental evidence of which is provided in this work.

The abortive structure then unfolds, and the latter searches for the correct transition state. When found, folding to the native state is rapid. Achieving the transition state is analogous to the assumption of transition state theory that the systems that have surmounted the col in the direction of products do not turn back. Kinetic intermediates, I_n, of sufficient lifetime, when present, populate the post-barrier pathway.

Finally, inadequacy of data at hand does not allow a structural description of the abortive intermediate. It is also clear that a single-exponential decay of proton occupancy as a function of time (Fig. 6) means accumulation of no other structural intermediate. Thus, the refolding of alkaline cyt-CO can be schematized classically as the following:



Dead-end intermediate

2.6 Conclusion

The folding time of cytochrome *c* is the time taken for nascent unfolded poly peptide chain to search the correct transition state topology. The search is unsystematic, and in this regime the polypeptide may achieve a wrong transition state leading to a dead-end structure. The polypeptide can nearly or completely misfold in the test tube as some do *in vivo.*

2.7 References

1. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z., and Socci, N. D. *Proc. Natl. Acad. Sci. USA* **1995,** *92,* 3626-3630.
2. Dill, K. A., and Chan, H. S. *Nat.Struct. Biol.* **1997,** *4*, 10-19.
3. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997,** *48*, 545-600.
4. Socci, N. D., Onuchic, J. N., and Wolynes, P. G. *Proteins* **1998,** *32*, 136-158.
5. Schonbrun, J. & Dill, K. A. *Proc. Natl. Acad. Sci USA* **2003,** *100*, 12678-12682.
6. Onuchic, J. N., and Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004,** *14*, 70-75.
7. Baldwin, R. L. *J. Biomol. NMR* **1995,** *5*, 103-109.
8. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. *Proteins: Struct. Funct. Genet.* **1995,** *21,* 167-195.
9. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. *Fold. Des.* **1996,** *1*, 441-450.
10. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. *Science* **1995,** *269*, 192-197.

11. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. *Proc. Natl. acad. Sci. USA*, **2001,** *98*, 105-112.
12. Hoang, L., Bédard, S., Krishna, M. M. G., Lin, Y., and Englander, S. W. *Proc. Natl. Acad. Sci. USA*, **2002,** *99*, 12173-12178.
13. Krishna, M. M. G., Lin, Y., Rumbley, J. N., and Englander, S. W. *J. Mol. Biol.* **2003,** *331*, 29-36.
14. Makarov, D. E., and Plaxco, K. W. *Protein Sci.* **2003,** *12*, 17-26.
15. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. *Nat. Struct. Biol.* **1994,** *1*, 149-156.
16. Sosnick, T. R., Mayne, L., and Englander, S. W. *Proteins* **1996,** *24*, 413-426.
17. Santoro, M. M., and Bolen, D. W. *Biochemistry* **1988,** *27*, 8063-8068.
18. Silow, M., and Oliveberg, M. *Biochemistry,* **1997,** *36*, 7633-7637.
19. Bryngelson, J. D., and Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **1987,** *84,* 7524-7528.
20. Wolynes, P. G., Onuchic, J. N., and Thirumalai, D. *Science* **1995,** *267*, 1619-1620.
21. Yadaiah, M., Kumar, R., and Bhuyan, A. K. *Biochemistry* **2007,** *46,* 2545-2551.
22. Camacho, C. J., and Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **1993,** *90,* 6369-6372.
23. Gutin, A. M., Abkevich, V. I., and Shakhnovich, E. I. *Biochemistry* **1995,** *34,* 3066-3076.
24. Miller, R., Danko, C. A., Fasolka, M. J., Balazs, A. C., Chan, H. S., and Dill, K. A. *J. Chem. Phys.* **1992,** *96,* 768-780.

Dead-end intermediate

25. Sali, A., Shakhnovich, E., and Karplus, M. *Nature* **1994,** *369,* 248-251.
26. Sali, A., Shakhnovich, E., and Karplus, M. *J. Mol. Biol.* **1994,** *235,* 1614-1636.
27. Skolnick, J., and Kilinski, A *Science* **1990,** *250,* 1121-1125.
28. Guo, Z., Thirumalai, D., and Honeycutt, J. D. *J. Chm. Phys.* **1992,** *97,* 525-535.
29. Honeycutt, J. D., and Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **1990,** *87,* 3526-3529.
30. Hamada, D., Segawa, S, I., and Goto, Y. *Nat. Struct. Biol*. **1996,** *3,* 868-873.
31. Troullier, A., Reinstadler, D., Dupont, Y., Naumann, D., and Forge, V. *Nat. Struct. Biol*. **2000,** *7,* 78-86.
32. Wildegger, G., and Kiefhaber, T. *J. Mol. Biol.* **1997,** *270,* 294-304.
33. Kumar, R., and Bhuyan, A. K. *Biochemistry* **2005,** *44*, 3024-3033.
34. Udgaonker, J. B., and Baldwin, R. L. *Nature* **1988,** *335,* 694-699.
35. Prabhu, N. P., Kumar, R., and Bhuyan, A. K. *J. Mol. Biol.* **2004,** *337*, 195-208.
36. Bhuyan, A. K., Rao, D. K., and Prabhu, N. P. *Biochemistry* **2005,** *44,* 3034-3040.
37. Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997,** *7*, 29-40.
38. Sosnick, T. R., Mayne, L., and Englander, S. W. *Proteins* **1997,** *24*, 413-426.
39. Jacob, J., Krantz, B., Dothager, R. S., Thiyagarajan, P., and Sosnick, T. R. *J. Mol. Biol.* **2004,** *338*, 369-382.
40. Bhuyan, A.K., and Udgaonkar, J.B. *J. Mol. Biol.* **2001,** *312,* 1135-1160.

41. Schindler, T., Herrler, M., Marahiel, M. A., and Schmid, F. X. *Nature Struct. Biol.* **1995,** *2,* 663-673.

42. Bai, Y., Zhou, H., and Zhou, Y. *Protein Sci.* **2004,** *13*, 1173-1181.

43. Abkevich, A. I., Gutin, A. M., and Shakhnovich, E. I. *J. Mol. Biol.* **1995,** *252*, 460-471.

44. Bhuyan, A.K., and Udgaonkar, J.B. *Proteins* **1998,** *32*, 241-247.

45. Xu, Y., Mayne, L. C., and Englander, S. W. *Nat. Struct. Biol*. **1998,** *5*, 774-778.

CHAPTER 3

The Alkali Molten Globule State of Ferrocytochrome *c*: Extraordinary Stability, Persistent Structure, and Constrained Overall Dynamics

3.1 Abstract

This chapter describes the structural and dynamic properties of a hitherto uncovered alkali molten globule state of horse "ferrocytochrome *c*". Several experimental difficulties due mainly to heme autooxidation and extraordinary stability of ferrocytochrome *c* have been overcome by working with the carbonmonoxide-bound molecule under extremely basic condition (pH 13) in a strictly anaerobic atmosphere. Structural and molecular properties extracted from basic spectroscopic experiments suggest that cations drive the base-denatured CO-liganded protein to the molten globule state. The stability of this state is ~5.2 kcal mol^{-1}, and the guanidinium-induced unfolding transition is sharp (m_g~2.3 kcal mol^{-1} M^{-1}), suggesting persistent rigid tertiary structure. NMR chemical shift dispersion and NOE also supports the long range interactions and rigid tertiary structure. Strategic experiments involving the measurement of CO association rate to the base-denatured protein indicate substantially restricted overall motion and stiffness of the polypeptide chain in the molten globule state. Possible placement of the state in the folding coordinate of ferrocytochrome *c* is discussed.

3.2 Introduction

During the past thirty years or so a large volume of literature has most definitely indicated the existence of the molten globule state, which, viewing from the classical perspective of protein folding, has been referred to as the third thermodynamic state of proteins; the native and the unfolded state are the other

two (1). The concept of the molten globule (MG) state emerged at a time when the view of the two-state nature of folding of small globular proteins (2-4) had appeared to break down. ‘Compact molecular states containing native-like secondary structure’ were temporally resolved in the early stages of folding of an increasingly large set of proteins (5-7). These intermediate folding structures appeared to fit nearly flawlessly into the molecular organizational definition of the classic MG state (1), and thus provided the basis to propose that the MG state correspond to early kinetic intermediates in protein folding. Several concerns arise, however. How can the unfolded chain achieve so native-like structural organization so early, in perhaps the submillisecond regime? Where along the reaction coordinate the transition state positioned? Does the transition from the unfolded to the MG state involve a rate-limiting barrier crossing event? About the same time, theoretical and computer simulation studies of protein folding detected ensembles of MG-like structures that occur very late in the folding course (8). It is now generally accepted that the MG state characterized by the presence of native-like secondary structure and molecular compactness but lack both native-like side chain packing and solvation of the partially hydrophobic core, corresponds to late folding intermediates (9-11).

While the classic MG state was detected in anion-containing acid medium (12), stabilization under diverse solvent conditions of equilibrium protein states that broadly meet with the definition of the MG state has been reported in due course (1, 13-16, for example). These developments throw open the issue of the importance of the MG-like states in the protein folding problem. It is thus

necessary to not only look at different aspect of their structure, dynamics, and molecular organization, but also locate their positions in the folding coordinate.

The acid MG state is known for numerous proteins, particularly well known for ferricytochrome *c* (12). The complementary alkali state has however been reported for only a few proteins, including β-lactamase (17), and barstar (18). Many reports on the nature of the alkaline transition and the properties of the alkali-unfolded state of ferricytochrome *c* (ferricyt *c*) have appeared (19-25), but the alkali MG state remained elusive.

Being aware of the extraordinary native-state stability and apparent two-state fast folding of the reduced form of horse cytochrome *c* (26-28), it is of interest to know if a MG-like state exists for this oxidation state of cyt *c*. There are experimental difficulties though, particularly because of enhanced autooxidation rate of the ferrous heme in acid solutions. To counter this problem and to suppress the excessive stability of the protein, the protein was test-tube engineered by liganding the ferrous heme iron with extrinsically added CO at pH 13. Under these conditions the reduced state of cytochorme *c* (to be called ferrocyt *c* henceforth) is fully denatured. Upon addition of cations to the medium, the protein undergoes transition to the MG state as judged from spectroscopic signatures. Specifically, the aspects of tertiary structural organization, folding stability, and the dynamic nature of the alkali MG state were examined by employing a set of strategic experiments. Very interestingly, the properties observed are atypical of the MG state, although the basic requirements of the definition are met with. This alkali molten globule state has significant relevance in explaining ferrocyt *c* folding.

3.3 Materials and Methods

Cyt *c* from Sigma (type VI) was used without further purification. Sodium dithionate was from Merck. All experiments were done in controlled anaerobic atmosphere at 22^{o} C.

3.3.1 Equilibrium unfolding and NaCl titration studies

For pH unfolding studies, an 8 μM solution of cyt *c*, prepared in an aqueous medium containing 10 mM each of Tris, disodium hydrogen phosphate, and CAPS (3-[Cyclohexylamino]-1-propanesulfonic acid), was titrated to different pH values in the 7-13.25 range by the use of NaOH. The titration did not upset the uniformity of the protein concentration in the samples. Samples were deaerated by using nitrogen or argon gas, and reduced by adding a small volume of freshly prepared dithionite to obtain a final concentration of 2-3 mM. The samples were then incubated for ~30 minutes under 1 atm CO gas pressure in tightly capped cuvettes or glass tubes. Fluorescence emission spectra (ex: 280 nm) were taken in a FluoroMax-3 instrument (Jobin-Yvon, Horiba). Optical absorption spectra were recorded in a Cary 100 instrument. The pH titration curves were analyzed using the following transformed Henderson-Hasselbalch equation

$$y = \frac{c_u + c_f\left[10^{n(pH-Cm)}\right]}{1+10^{n(pH-Cm)}} \qquad (1)$$

where, c_u and c_f are normalized fluorescence signals for the unfolded and the refolded state, respectively, n is the number of OH^- titrated, and c_m is the pH-midpoint for the transition. The procedure for GdnHCl-induced unfolding experiments, where the pH was held constant at 12.9(±0.1) by using NaOH, was

same as described. The data were analyzed using the two-state equation (29). NaCl titration studies were carried out with protein solutions (6-8 μM) prepared in NaOH containing the salt in 0-1.5 M range. pH of the solutions were adjusted to12.9(±0.1). Protein reduction and CO ligation were carried out using the protocol detailed above. CD spectra were taken in a JASCO J-710 spectropolarimeter using 1 mm light path length cuvettes.

3.3.2 NMR spectroscopy

NaCl-containing D_2O solutions of 2 mM cyt *c*, the backbone amides of which were preexchanged, were adjusted to the pH meter reading of 12.9(±0.1) by adding NaOD. Protein solutions contained in NMR tubes were reduced by adding solid sodium dithionite (20 mM) under nitrogen. A gentle stream of CO was bubbled into the solutions. The tubes were then sealed with sleeved rubber stoppers and equilibrated at 22°C for ~30 minutes. Spectra were taken at 22°C in a 400 MHz Bruker spectrometer (AV400). 90°-pulse length was calibrated for each concentration of salt used. For phase sensitive NOESY spectra (τ_m=150 ms) 512 t_1 were used. Spectral width was 7184 Hz.

3.3.3 Kinetic measurements of association of CO

Cyt *c* (1 mM) dissolved in aqueous NaOH, pH 12.9(±0.1) is deaerated and reduced by adding sodium dithionite (2 mM). 25 μl of this ferrocyt *c* solution is mixed rapidly with 2 ml of NaOH solution containing 2 mM dithionite, a given amount of NaCl, and 1 mM CO. Association kinetics of CO were followed at 550 nm (dead time ~5 s) at a peltier-controlled temperature of 22°C in a Cary 100 spectrophotometer.

3.4 Results

3.4.1 NaCl-induced resistance toward alkali denaturation of carbonmonoxide-bound ferrocytochrome *c* (Cyt-CO)

Because low pH enhances autooxidation of the ferrous heme, experiments with ferrocyt *c* are best performed at neutral to alkaline range of pH. Therefore the choice was to work under highly basic pH conditions in which a MG-like state of ferrocyt *c* could possibly be found. But, due to its extraordinary stability, ferrocyt *c* does not denature even in extreme alkaline conditions unless CO is allowed to bind to the ferrous heme (32). Figure1a shows fluorescence-monitored base titrations of cyt-CO in the absence and the presence of 0.03, 0.1, and 1 M NaCl. In saltless medium the titration is nearly complete at pH 13. The fit to the data by the use of eq 1 indicates that three OH^- are titrated. The titration curve clearly shifts toward higher pH values as NaCl is included incrementally; in the presence of 1 M NaCl hardly any sign of denaturation appears within the limit of achievable pH. This suggests that cations increase the stability of the protein under conditions fully denaturing otherwise.

3.4.2 Alkali-denatured cyt-CO and its molecular compaction in the presence of NaCl

Fluorescence emission intensity due to the lone tryptophan (W59) provides a reliable marker for the molecular compactness of cyt *c*. Native cyt *c* is fluorescence-silent due to excitation energy transfer from W59 to the heme (33). Unfolding results in an increase in the heme-tryptophan distance due to molecular

expansion, and hence a dramatic increase in the fluorescence quantum yield (33, 34). Figure 1b shows fluorescence emission spectra of alkali-denatured cyt-CO with increments of NaCl concentration. Clearly, the denatured polypeptide becomes increasingly compact as the salt concentration increases. Figure 1c shows this result even more clearly- the molecular compactness of alkali-denatured cyt-CO in the presence of >0.5 M NaCl is very close to that of native cyt *c*, indicating salt-induced transition of the alkali-unfolded cyt-CO to the MG state.

3.4.3 Far-UV CD-monitored secondary structure in the presence of NaCl

Typically, far-UV CD spectra of proteins in the MG state are similar to those of the native state (5). Figure 1(d) shows changes in the far-UV CD spectrum of alkali-unfolded cyt-CO induced by added NaCl. Figure 1(e) shows the variation of the 222-nm ellipticity with salt in the 0-1.1 M range. Clearly, cyt-CO at pH 13 acquires substantial secondary structure in the presence of >0.5 M NaCl, suggesting transformation to the MG state.

.

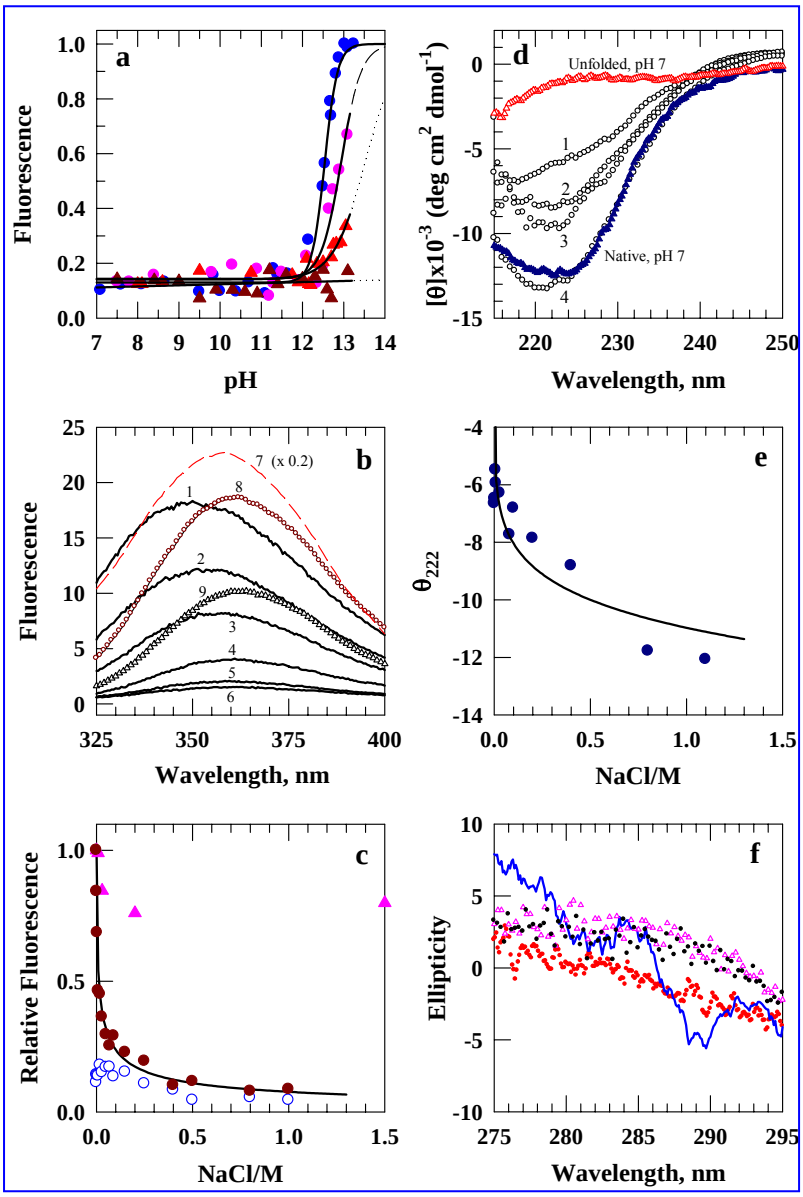


Figure 1 Basic results indicating cation-induced transformation of base-unfolded cyt-CO into the molten globule state. (a) Alkaline pH-induced unfolding of cyt-CO at 22°C: (●)

no salt; (●) in 30 mM NaCl; (▲) in 0.1 M NaCl; (▲) in 1 M NaCl. (b) Fluorescence emission spectra of cyt-CO at 22°C. the NaCl concentrations and pH values are 1, 0 M at pH 13; 2, 0.005 Mat pH 13; 3, 0.01 M at pH13; 4, 0.15 M pH 13; 5, 0.5 M at pH 13; 6, 1M at pH13; 7, 0 M at pH6.3 and 6.8 M GdnHCl; 8, 0 M NaCl at pH 13and 3.7 M GdnHCl; 9, 1M at pH13 and 3.7 M GdnHCl.The native state at pH 7 is fluroscence-silent. (c) Relative quenching of W59 fluorescence as a function of salt (●). In control experiments, fluorescence of ferrocyt *c* without CO ligation (○) and NATA (▲) were recorded. (d) Far-UV CD spectra of cyt-CO at 22 °C. NaCl concentrations and pH values are 1, 0 M NaCl at pH 13 ; 2, 0.2 M NaCl at pH 13; 3, 0.4 M NaCl at pH 13 and 4, 1.1 M NaCl at pH 13. The CD spectra in the presence of high pH and GdnHCl is obscured and is not shown. (e) Variation of CD absorption at 222 nm with salt. (f) Near-UV CD spectra: cyt-CO at pH 13, 1 M NaCl, 22°C (○); native ferrocyt c at pH 7, 1 M NaCl, 22°C (—); unfolded ferrocyt *c* liganded with CO at pH 7 and 7M GdnHCl (●).

3.4.4 Absence of near-UV CD signal

Molten globules are generally distinguished by a dramatic loss of near-UV CD signal (1, 35). Figure 1f compares the aromatic CD signals of native ferrocyt *c* and base-denatured cyt-CO in the presence of 1 M NaCl. The 282- and 290-nm bands of native ferrocyt *c* (pH 7, 1 M NaCl) that arise from the coupling of the electric transition dipole moments of the W59 indole and the heme are not traceable in the salt-containing alkaline medium (pH 13, 1 M NaCl), apparently indicating some loosening of tertiary structure. It does not however necessarily mean an increased fluctuation and conformational averaging for the W59 side chain (see Discussion).

3.4.5 NMR spectral features of alkali-denatured cyt-CO in the presence of NaCl

Increased internal mobility and side-chain environmental averaging, a documented characteristic of the MG state, is better checked by ^{1}H NMR (12, 36). Ferrocyt *c* as such is native at pH 13, and the NMR lines are narrow and well-dispersed (Figure 2a). Upon addition of CO to the same solution under identical conditions, the protein unfolds as indicated by the loss of both chemical shift dispersion and lineshape (Figure 2b). When NaCl is included in the unfolded protein solution, the spectrum appears to regain partly both dispersion and sharpness of resonances all over the spectral width (parts c-e of Figure 2). This result is inconsistent with the generic observation that NMR spectra of the MG state and the unfolded state are similar (37-39). Line broadening of the resonances, associated characteristically with the MG state (40), is not seen either. The spectra presented show that the MG being interrogated here carries reasonable traces of rigid tertiary structure involving both aliphatic and aromatic side-chains. Sections of phase-sensitive NOESY spectra presented at the bottom of figure 2 (pH 13, 1 M NaCl) substantiate the claim. A good number of long-range NOE interactions involving aliphatic (Figure 2, bottom right), and aliphatic and aromatic side chains (Figure 2, bottom left) indicates the presence of considerable tertiary structure.

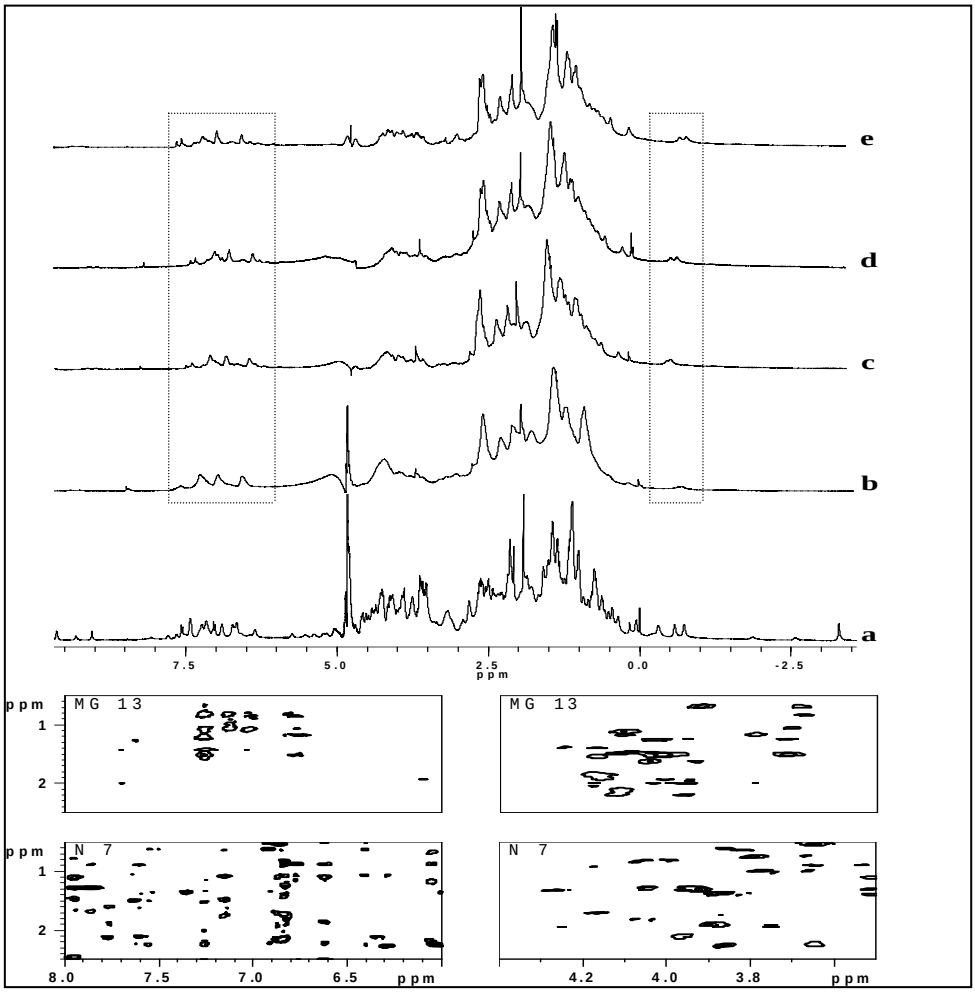


Figure 2 400 MHz NMR spectra. (a) Native state of ferrocyt *c*, pH 13, no salt. (b) Unfolded cyt-CO, pH 13, no salt. (c) Cyt-CO, pH 13, 0.02 M NaCl. (d) Cyt-CO, pH 13, 0.5 M NaCl. (e) Cyt-CO, pH 13, 1 M NaCl. The boxed panels at the bottom present sections of phase-sensitive NOESY spectra showing interactions involving aliphatics and aliphatic-aromatic side chains in native ferrocyt *c*, pH 7 (N 7), and in the molten globule state of cyt-CO, pH 13, 1 M NaCl (MG 13). All spectra were taken at 22°C.

3.4.6 Charge screening by Na^+ ions also leads to dynamic constraints: A general effect

In terms of internal mobility, the MG exhibits increased fluctuations of the side chains relative to that native state (37, 41). Although the overall motional freedom is restrained when compared with the dynamics of the denatured state held in the absence of the stabilizing counterions. To show how the charge-screening effect of counterions generally reduces the fluctuations of groups of atoms, experiments involving kinetics of association of CO with ferrocyt *c* were conducted at pH 13 in the presence of variable concentration of NaCl. The rationale of the experiment is the following. Destabilized ferrocyt *c* binds CO when the latter is used in saturating concentration (≈1 mM) (42, 43). Since intramolecular thermal collisions provide the energy for barrier crossing in the association reaction, cyt + CO $\rightarrow$ cyt-CO, the rate coefficient for the reaction (k_{ass}) is expected to decrease if the amplitudes of thermal fluctuations are reduced as a result of constraints on the collective modes of intramolecular motion (30, 31). The association kinetics are slow, and can be conveniently measured by monitoring the absorbance at 550 nm following the addition of a small volume of the protein solution to a CO-saturated aqueous alkali solution containing the salt. The two representative traces for single-phase CO association kinetics shown in Figure 3a indicate that the association rate is much slower in the presence of 1 M NaCl relative to that in the absence of the salt. Figure 3b shows the variation of the association rate coefficient, k_{ass}, with molar concentration of NaCl. Also plotted is NaCl-dependent variation of k_{ass} for native ferrocyt *c* in the presence of 0.5 M GdnHCl, pH 7. Clearly, the internal dynamics of alkaline ferrocyt *c* is

substantially constrained relative to that of the initial state in the absence of salt. It is important to note that the experiment described here monitors the salt-induced constraints in the internal motion of the alkaline ferrocyt *c* only, not the molten globule state, since it is not until CO binds that the molten globule forms at pH 13. From the temperature dependence of a series of association reactions (Figure 3c), the activation enthalpy ($H \approx E_a$) for CO association at different concentrations of NaCl in the 0-1 M range (Figure 3d) were determined. Expectedly, E_a for CO association to the alkali denatured state of cyt *c* increases with salt concentration, and approaches the value measured for the native-like protein at pH 7. This is due to reduced motional freedom of the charge-shielded alkaline ferrocyt *c*. In terms of internal dynamics, the charge-screened state closely matches the native state. Thus, the reduced mobility is a consequence of reduced charge-charge repulsion.

3.4.7 Stability of the molten globule state to unfolding by GdnHCl

Encouraged by the observations of the compactness, and dynamic control in the MG state, tryptophan fluorescence was used to monitor GdnHCl-induced unfolding of cyt-CO at pH 13, 1 M NaCl. Figure 4 shows the highly cooperative transition. Since an all-or-none type transition operates between the native and the MG state of small proteins (48-50), a two-state analysis was carried out to extract the thermodynamic parameters. The fit yields $\Delta G° = 5.2(\pm 1)$ kcal mol^{-1}.

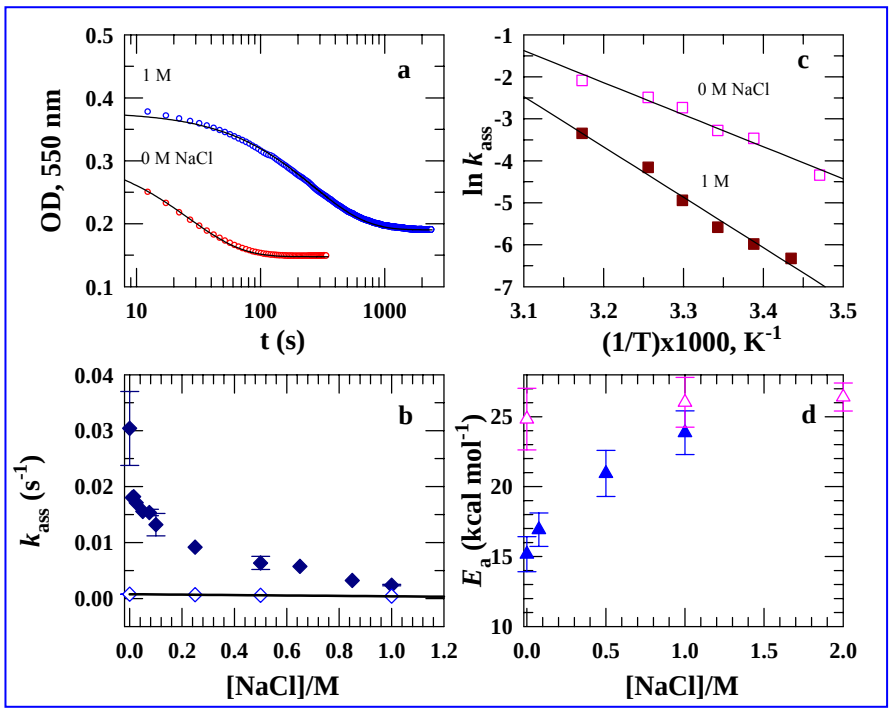


Figure 3 Results for the CO association reaction, cyt *c*+CO → cyt–CO monitored by 550-nm heme absorbance. (a) Representative kinetic traces showing the influence of NaCl on the CO association rate at pH 13. (b) NaCl dependence of the association rate constant, k_{ass}, at pH 13 (♦). In the control experiment, dependence of k_{ass} on NaCl was measured for native ferrocyt *c* at pH 7, 0.5 M GdnHCl (◇). (c) Arrhenius plots for the CO association reaction, pH 13. (d) Dependence of the activation energy, E_a, for CO association reactions on the concentration of NaCl at pH 13 (▲), and pH 7, 0.5 M GdnHCl (△).

This is enormous stability, comparable to the unfolding free energy of many small proteins (51). Equally interesting is the magnitude of the surface area associated with the unfolding transition. The fit yields m_g=2.3(±0.6) kcal mol^{-1} M^{-1}. This value, by comparison with 2.95(±0.28) kcal $mol^{-1}M^{-1}$ for GdnHCl unfolding of ferrocyt c in the presence of CO at pH 7, indicates a large amount of solvent exposure of amino acid residues accompanying the unfolding of the molten globule, suggestive of a significant hydrophobic core rather than patches of exposed and buried hydrophobic surfaces.

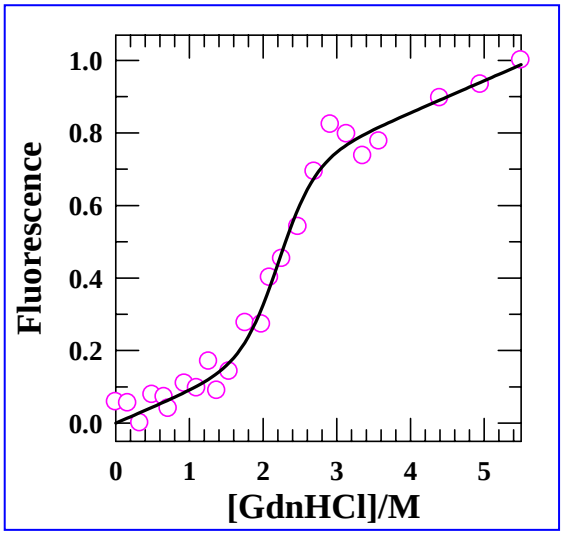


Figure 4. GdnHCl-induced equilibrium unfolding of cyt-CO at pH 13, 1 M NaCl, 22°C. The solid line is the standard two-state fit to data (santoro 1988, *29*). The fit yields $\Delta G°$=5.2(±1) kcal mol^{-1} and m_g=2.3(±0.6) kcal mol^{-1} M^{-1}), respectively.

3.5 Discussion

3.5.1 The alkali- and acid-denatured forms of cytochrome *c*

pH-denatured states are structurally not comparable with urea- or GdnHCl-unfolded states (*52*). But how different are the acid- and base-denatured forms (U_A and U_B states, respectively)? Since these conformations represent starting points for the transitions to the corresponding molten globule states, a comparison of their structural features is essential. Unfortunately, a detailed comparison is not feasible with the limited results available for the U_B state. Table 1 lists a few properties based on the available data for the U_A state of ferricyt *c* (12, 53-56), and results from this work (Figures 1, 2). Both states are maximally expanded, but are different in terms of environmental polarity and water exposure of the fluorophore. The secondary structural content is similar. The NMR resonances for both U_A and U_B states are exchanged-broadened due to motions in the 100-1000 s^{-1} range, the latter being relatively much slower (compare Figure 2 with Figure 1 of ref. 56). This qualitative sketch suggests relatively constrained dynamics for alkali-denatured cyt-CO where the W59 indole and certain other side-chains perhaps, form apolar clusters. It is also note that several control experiments, including NMR spectroscopy (chapter 2), have indicated that the U_B state is monomeric and is stable for at least 9 hours.

3.5.2 The alkali MG (B state) of ferrocyt *c*

Results presented show that a MG state of ferrocyt *c* populates in the highly alkaline salt-containing aqueous medium. Transformation of the destabilized states of many proteins to MG-like forms by a wide variety of solution conditions, solvent additives, and cosolvents have been documented (1,

13-16, 50, 53). All of these forms exhibit generally similar overall properties (1, 57), although the mechanism of stabilization depends on the specific action of the additive employed. Since ions fundamentally influence protein electrostatics, the mechanism of ion-induced stabilization of denatured proteins at both extremes of pH is expected to be the same. Judging by the pK_a values of ionizable groups of amino acids, the net charges of cyt *c* at pH<2 and pH>12.5 are +24 and −17, respectively. Under these conditions, the extrinsically added counterions can exert effect on the protein by either reducing the electrostatic repulsion or direct binding to form ion pairs (53, 54, 58), or possibly by both. Looking from this perspective, the acid MG state, extensively studied for a large number of proteins, including the paradigmatic "ferricyt *c*" (53-56), and the alkali MG state of "ferrocyt *c*" are expected to share at least qualitatively similar structural and dynamic properties. The molecular compactness and secondary structure content of the alkali state observed here are qualitatively comparable with those reported for the acid MG state of ferricyt *c* (12, 56, 59, 60; see Table 1)). Certain other properties observed for the alkali MG form of ferrocyt *c*, however, do not concur with the generic prescription for the acid state. In the following, the customary labels 'A' and 'B' states will be used for the acid and alkali molten globule states, respectively.

Table 1. Properties of U_B and U_A states of horse cytochrome *c*

	U_B	U_A
Fluorescence emission, λ_{max}	350 nm	356 nm
Molecular expansion	Equivalent to the U state	simulates the U state
Peptide CD (MRE)	~5900	~3800
Near-UV CD	weak	weak
^{1}H NMR resonances	Relatively more dispersed and sharp	Undispersed and broad

3.5.3 Moderately rigid tertiary structure in the B state of ferrocyt c.

Disordered nature or apparent absence of tertiary structure is a hallmark of the MG state for all the proteins studied (1, 5). The B state of ferrocyt *c* however contains definite traces of tertiary interactions. For cyt *c*, the absence of the near-UV CD signal, as seen in the present study also (Figure 1f), is tacitly assumed to be an indication of the loss of tertiary structure (1, 12), while it does not need always be so. The near-UV CD signal of cyt *c* originates from the rotational strength of W59 side chain due to the coupling of its electric transition dipole moment with that of the heme group (61), and since the dipole interaction potential depends on the geometric relationship between the two participating groups, the absence of the near-UV CD signal is a sure indication of a perturbation in the relative orientation and the distance between the two chromophores. It does not necessarily mean a general loss of the tertiary structure leading to an increased fluctuation and conformational averaging of the side chain. The NMR spectra (Figure 2) provide the most convincing evidence for the

presence of some defined tertiary interactions. NaCl encourages the growth of a number of both aromatic and aliphatic resonances associated with tertiary interactions, and the observation of long-range NOE interactions strengthens the fact. The cooperativity observed for the GdnHCl-induced unfolding transition for the B state (Figure 5) also points to the presence of ordered tertiary structure, because the steepness of denaturant-induced unfolding curves depends strongly on the content of rigid tertiary structure (49, 62). Moderately rigid tertiary interactions, when present, as in the case of B states of ferrocyt *c* and β-lactamase (17), can not only contribute to molecular compaction, but also results in restricted environmental averaging for the side chains as indicated by the fairly narrow NMR lines (Figure 2e).

3.5.4 Constrained overall dynamics in the U_B state of ferrocyt *c* induced by NaCl.

Spatial displacement of thermal fluctuations, and hence collisions between different groups of atoms each exhibiting collective motions is dramatically reduced in the U_B state of ferrocyt *c*. The inference is drawn from the results of a set of specific experiments designed to observe the rate constant for association of CO with the alkali-denatured ferrocyt *c* (k_{ass}). This reaction is not diffusion or encounter controlled. It rather involves substantial energy barrier or steric requirements, and hence many collisions between the CO and protein groups, particularly the heme ring, are required before the reaction ensues. The value of k_{ass} under a given solution condition then depends on the frequency of collisions involving the CO and the heme side chains that exhibit highly collective motions (63). The rate constant for the CO association reaction, k_{ass}, decreases dramatically as the base-denatured polypeptide is held in the presence increments

of salt (Figure 3b). Accordingly, the activation energy for the reaction, E_a, also goes up (Figure 3d). In fact, in the presence of NaCl concentration sufficient to fully populated B state, values for both k_{ass} and E_a match those observed for the native state (Figure 3b, d). Such hindered collective motions are not commonly associated with the U_A state. Because of dynamic disorder in the tertiary structure, the A state is known to be highly mobile (64) with low energetic barriers for conformational fluctuations in the millisecond timescale (65). Absence of non-helical hydrogen bonds due to a high degree of disorder in the side chain of majority of residues has been observed for the A state of ferricyt *c* (55, 56). On the other hand, restricted mobility of aromatic side chains has also been reported for the A state of other proteins (37, 41). The conclusion that the collective motions involving the side chains of the B state of ferrrocyt *c* are relatively restrained is consistent with the indications of organized tertiary structure discussed above.

To summarize, the B state of ferrocyt *c* is distinguished by an appreciable level of ordered tertiary interactions, restrained collective motions that cover larger length scales, and a stiff backbone, and it shares with other molten globules the common property of being compact and having similar secondary structure content to the native state.

3.5.5 Stability of the B state of ferrocyt *c*.

The B state is expected to be sufficiently stable, since it is braced by a native-like structural order, modest dynamics, and a compact molecular organization. The measured stability to unfolding by GdnHCl is 5.2(±1) kcal mol^{-1} (Figure 5), substantially higher than the reported free energy of unfolding of

molten globules in water (1, 51, 68-70). Several non-structural auxiliary factors can determine the energetic stability of molten globules. For example, time-averaged internalization of any uncompensated charge can reduce the stability, since charge burial in the low dielectric protein interior is energetically not preferred. The charge on the heme group of cytochrome *c* itself provides an illustrative example. Ferricyt *c* A state has a mixed-spin ferric iron (12), meaning that a fraction of the A state molecules carries a +ve charge, whereas the charge on the persistent low spin ferrous heme is zero due to the pairing of all six *d*-orbital electrons. Thus the A state of ferricyt *c* is less stable than the B state of ferrocyt *c*. Charge density in relation to protein motions can also play a role. High frequency atomic fluctuations at the site of a protein charge can decrease the charge density by distorting the Debye-Huckel ion sphere leading to relatively weaker ion-protein interaction. The variable strength of interaction of ions with the molten globule states of different proteins can explain the observed differences in stability.

In the context of stability it is worthwhile considering the relative cooperativity of GdnHCl unfolding transitions of A and B states. For GdnHCl unfolding of the A state of ferricyt *c* and several acetylated forms of the protein, Goto and coworkers report a m_g value of 1.85 kcal mol^{-1} M^{-1} (54). Apparently, the cooperativity of the unfolding transition is independent of the extent of acetylation. The corresponding value for the B state is 2.3(±0.6) kcal mol^{-1} M^{-1} (Figure 5) which is fairly comparable. These comparisons suggest that the amount of solvent exposure of amino acid residues accompanying the unfolding of A and

B states is similar. Thus the two states appear equivalent in terms of the content of buried hydrophobic surfaces.

3.5.6 How ordered is the B state?

Molten globules of different proteins are structurally different. Furthermore, even for a given protein molten globules stabilized by different conditions could be very different. For example, at least three structurally distinct A states each separated by significant energy barriers has been shown for staphylococcal nuclease (71, 72). The two A states for ferricyt c have similar secondary structure (35, 53) but different molecular compactness (24). In general, molten globules could be divided into three major classes: highly ordered, classical, and disordered. Taking specific tertiary structure as the orderly mark (73), the classical molten globules that exhibit reduced tertiary structure and increased fluctuations of side chains are in between the highly ordered and disordered categories. In this perspective, the B state of cyt c could be described as an ordered molten globule, since it exhibits fairly ordered tertiary interactions. By considering secondary structure content for the description of structural order (71), the B state can be categorized as a highly ordered molten globule, since it retains near-native secondary structure, and nearly half of the folding stability of the native state.

3.5.7 B state simulates the dead end intermediate in folding of alkaline ferrocyt *c*.

The classical view proposes that the molten globule state represents a common early intermediate in the folding pathway of all proteins (1, 74). Later advances in the field indicate that in some cases they correspond to late folding

intermediates (10, 11). Horse cyt c is an apparent two-state folder with no detectable accumulation of intermediates in both equilibrium and kinetic pathways (28). Kinetic studies have indicated that due to a downhill-biased run in the post-barrier course of folding, the reaction coordinate shows no minima appreciably low in energy (27, 28). However, real-time NMR measurement of amide hydrogen exchange in the presence of subdenaturing concentrations of GdnHCl (75), and the native-state equilibrium hydrogen exchange experiments developed by Englander and colleagues (76, 77) have demonstrated the existence of short-lived native-like intermediates, I_n, in the late or post-transition stages of folding. They are transient because no appreciable energy barrier separates them from the globally folded native state, and hence they do not populate the kinetic pathway unless fortuitous barriers due to bad intramolecular contacts frustrate folding in the post-transition stages (78). But studied in this work may not appear to match such a late short-lived intermediate. Stopped-flow kinetics of B state suggests this molten globule simulates the off-path way intermediate (79) in refolding of alkaline ferrocyt *c*.

3.6 Conclusions

The molten globule of alkaline ferrocyt *c* is distinguished by an appreciable level of ordered tertiary interactions, restrained collective motions that cover larger length scales, and stiff backbone, and it shares with other MGs the common property of being compact and having similar secondary structure. Kinetic data suggests that highly ordered alkaline molten globule simulates the dead end intermediate in folding of alkaline ferrocyt *c*.

3.7 References

1. Ptitsyn, O. B. *Adv. Protein Chem.* **1995,** *47*, 83- 229.
2. Privalov, P. L. *Adv. Protein Chem.* **1979,** *33*, 167-241.
3. Privalov, P. L., and Khechinashvili, N. N. *J. Mol. Biol.***1974,** *86*, 665-684.
4. Jackson, S. E., and Fersht, A. R. *Biochemistry* **1991,** *30*, 10428-10435.
5. Kuwajima, K. *Proteins* **1989,** *6*, 87-103.
6. Matthews, C. R. *Annu. Rev. Biochem.* **1993,** *62*, 653-683.
7. Evans, P. A., and Radford, S. E. *Curr. Opin. Struct. Biol.***1994,** *4*, 100-106.
8. Wolynes, P. G., Onuchic, J. N., and Thirumalai, D. *Science,* **1995,** *267*, 1619-1620.
9. Colón, W., and Roder, H. *Nat. Struct. Biol.* **1996,** *3*, 1019-1025.
10. Kuwajima, K., and Arai, M. *Mechanism of Protein Folding* **2000,** (Pain, R. H., Ed.) 2nd ed., pp 138-174, Oxford University Press, New York.
11. Mok, K. H., Nagashima, T., Day, I. J., Hore, P. J., and Dobson, C. M. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 8899-8904.
12. Ohgushi, M. & Wada, A. *FEBS Lett.* **1983,** *164*, 21-24.
13. Bychkova, V. E., Dujsekina, A. E., Klenin, S. I., Tiktopulo, E. I., Uversky, V. N., and Ptitsyn, O. B. *Biochemistry* **1996,** *35*, 6058-6063.
14. Davis-Searles, P. R., Morar, A. S., Saunders, A. J., Erie, D. A., and Pielak, G. J. *Biochemistry* **1998,** *37*, 17048-17053.
15. Moosavi-Movahedi, A. A., Chamani, J., Goto, Y., and Hakimelahi, G. H. *J. Biochem.* **2003,** *133*, 93-102.
16. Uversky, V. N. *Protein Sci.* **2002,** *11*, 739-756.
17. Goto, Y., and Fink, A. L. *Biochemistry* **1989,** *28*, 945-952.

18. Rami, B. R., and Udgaonkar, J. B. *Biochemistry* **2002,** *41*, 1710-1716.
19. Lambeth, D. O., Campbell, K. L., Zand, R., and Palmer, G. *J. Biol. Chem.* **1973,** *248*, 8130-8136.
20. Chalikian, T. V., Gindikin, V. S., and Breslauer, K. J. *FASEB J.* **1996,** *10*, 164-170.
21. Wison, M. T., and Greenwood, C. *Cytochrome c: A multidisciplinary Approach* **1996**(Scott, R. A., and Mauk, A. G., Eds.) pp 611-634, University Science Books, Sausilito, CA.
22. Battistuzzi, G., Borsari, M., Loschi, L., Martinelli, A., and Sola, M. *Biochemistry* **1999,** *38*, 7900-7907.
23. Nelson, C. J., and Bowler, B. E. *Biochemistry* **2000,** *39*, 13584-13594.
24. Cinelli, S., Spinozzi, F., Itri, R., Finet, S., Carsughi, F., Onori, G., and Mariani, P. *Biophys. J.* **2001,** *81*, 3522-3533.
25. Hoang, L., Bédard, S., Krishna, M. M. G., Lin, Y., and Englander, S. W. *Proc. Natl. Acad. Sci. USA*, **2002,** *99*, 12173-12178.
26. Bhuyan, A. K., and Udgaonkar, J. B. *J. Mol. Biol.* **2001,** *312*, 1135-1160.
27. Prabhu, N. P., Kumar, R., and Bhuyan, A. K. *J. Mol. Biol.* **2004,** *337*, 195-208.
28. Kumar, R., and Bhuyan, A. K. *Biochemistry* **2005,** *44*, 3024-3033.
29. Santoro, M. M., and Bolen, D. W. *Biochemistry* **1988,** *27*, 8063-8068.
30. Bhuyan, A. K. *Biochemistry* **2002,** *41*, 13386-13394.
31. Kumar, R., Prabhu, N. P., Yadaiah, M., and Bhuyan, A. K. *Biophys. J.* **2004,** *87*, 2656-2662.
32. Kumar, R., Prabhu, N. P., and Bhuyan, A. K. *Biochemistry* **2005,** *44*, 9359-

9367.

33. Vanderkooi, J. M., and Erecinska, M. *Eur. J. Biochem.* **1975,** *60*, 199-207.
34. Tsong, T, Y. *J. Biol. Chem.* **1974,** *249*, 1988-1990.
35. Goto, Y., Calciano, L. J., and Fink, A. L. *Proc. Natl. Acad. Sci. USA* **1990,** *87*, 573-577.
36. Ptitsyn, O. B. *J. Protein. Chem.***1987,** *6*, 273-293.
37. Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Gilmanshin, R. I., Lebedev, O. Y., Semisotnov, G. V., Tiktopulo, E. I., and Ptitsyn, O. B. *Eur. Biophys. J.* **1985,** *13*, 109-121.
38. Kuwajima, K., Harushima, Y., and Sugai, S. *Int. J. Peptide Protein Res.* **1986,** *27*, 18-27.
39. Baum, J., Dobson, C. M., Evans, P. A., and Hanley, C. *Biochemistry* **1989,** *28*, 7-13.
40. Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., and Dobson, C. M. *Biochemistry 32*, **1993,** 1707-1718.
41. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y., and Ptitsyn, O. B. *FEBS Lett.***1981,** *136*, 311-315.
42. Bhuyan, A. K., and Kumar, R. *Biochemistry* **2002,** *41*, 12821-12834.
43. Theorell, H., and Akesson, A. *J. Am. Chem. Soc.* **1941,** *63*, 1812-1827.
44. Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W. A. *Proc. Natl. Acad. Sci. USA* **1993,** *90*, 11860-11864.
45. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. *Proc. Natl. acad.*

Sci. USA **1996,** *93*, 11615-11617.

46. Hagen S. J., Latypov, R. F., Dolgikh, D. A., and Roder, H. *Biochemistry* **2002,** *41*, 1372-1380.
47. Jun, S., Bechhoefer, J., and Ha, B. Y. *Europhys. Lett.* **2003,** *64*, 20-426.
48. Uversky, V. N., Semisotnov, G. V., Pain, R. H., and Ptitsyn, O. B. *FEBS Lett.* **1992,** *314,* 89-92.
49. Ptitsyn, O. B., Uversky, V. N. *FEBS Lett*. **1994,** *341,* 15-18.
50. Qureshi, S. H., Moza, B., Yadav, S., and Ahmad, F. *Biochemistry* **2003,** *42*, 1684-1695.
51. Pfeil, W. Springer-Verlag, Berlin, Heidelberg **1998,** pp 17-343.
52. Tanford, C. *Adv.Protein chem.* **1968,** *23,* 121-282.
53. Goto, Y., Takahashi, N., and Fink, A. L. *Biochemistry* **1990,** *29*, 3480-3488.
54. Hagihara, Y. Tan, Y., and Goto, Y. *J. Mol. Biol.* **1994,** *237*, 336-348.
55. Jeng, M-F., Englander, S. W., Elöve, G. A., Wand, A. J., and Roder, H. *Biochemistry* **1990,** *29*, 10433-10437.
56. Jeng, M-F., and Englander, S. W. *J. Mol. Biol.***1991,** *221*, 1045-1061.
57. Arai, M., and Kuwajima, K. *Adv. Protein Chem.* **2000,** *53*, 209-271.
58. Goto, Y., and Nishikori, S. *J. Mol. Biol.* **1991,** *222*, 679-686.
59. Kataoka, M., Hagihara, Y., Mihara, K., and Goto, Y. *J. Mol. Biol.* **1993,** *229*, 591-596.
60. Chalikian, T. V., Gindikin, V. S., and Breslauer, K. J. *J. Mol. Biol.* **1995,** *250*, 291-306.
61. Urry, D. W. *Proc. Natl Acad. Sci. USA* **1965,** *54*, 640-648.
62. Uversky, V. N., and Ptitsyn, O. B. *Fold. Design* **1996,** *1*, 117-122.

63. Morgan, J. D., and McCammon, J. A. *Biopolymers* **1983,** *22*, 1579-1593.

64. Chakraborty, S., Ittah, V., Bai, P., Luo, L., Haas, E., and Peng, Z. *Biochemistry* **2001,** *40*, 7228-7238.

65. Kim, S., Bracken, C., and Baum, J. *J. Mol. Biol.***1999,** *294*, 551-560.

66. Redfield, C., Smith, R. A. G., and Dobson, C. M. *Nat. Struct. Biol.* **1994,** *1*, 23-29.

67. Daggett, V., and Levitt, M. *Proc. Natl. Acad. Sci. USA* **1992,** *89*, 5142-5146.

68. Ikeguchi, M., Kuwajima, K., Mitani, M., and Sugai, S. *Biochemistry* **1986,** *25*, 6965-6972.

69. Barrick, D., and Baldwin, R. L. *Protein Sci.* **1993,** *2*, 869-876.

70. Hagihara, Y., Aimoto, S., Fink, A. L., and Goto, Y. *J. Mol. Biol.* **1993,** *231*, 180-184.

71. Uversky, V. N., karnoup, A. S., Segel, D. L., Seshadri, S., Doniach, S., and Fink, A. L. *J. Mol. Biol.* **1998,** *278,* 879-894.

72. Nishimura, C., Riley, R., Easman, P., and Fink, A. L. *J. Mol. Biol.* **2000,** *299,* 1133-1146.

73. Fink, A. L. *Curr. Opin. Stru. Biol.* **2005,** *15,* 35-41.

74. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I *FEBS Lett.* **1990,** *262*, 20-24.

75. Bhuyan, A. K., and Udgaonkar, J. B. *Proteins* 1998, *32*, 241-247.

76. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. *Science* **1995,** *269*, 192-197.

77. Xu, Y., Mayne, L. C., and Englander, S. W. *Nat. Struct. Biol* **1998,** *5*, 774-778.

78. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. *Proc. Natl. acad. Sci. USA*, **2001,** *98*, 105-112.

79. Rao, D. K., Kumar, R., Yadaiah, M., and Bhuyan, A. K. *Biochemistry* **2006,** *45,* 3412-3420.

CHAPTER 4

Complexity of Aromatic Ring-flip Motions in Proteins: Y97 Ring Dynamics in Cytochrome *c* Observed by Cross-relaxation Suppressed Exchange NMR Spectroscopy

4.1 Abstract

Dynamics of large-amplitude conformational motions in proteins are complex and less understood, although these processes are intimately associated with structure, folding, stability, and function of proteins. A large set of spectra obtained by cross-relaxation suppressed exchange NMR spectroscopy (EXSY) is used to study the 180° flipping motion of the Y97 ring of horse ferricytochrome *c* as a function of near-physiological temperature in the 288-308 K range. With rising temperature, the ring-flip rate constant makes a continuous transition from Arrhenius to anti-Arrhenius behavior through a narrow Arrhenius-like zone. This behavior is seen not only for the native state of the protein, but also for native-like states generated by adding subdenaturing amounts of guanidine deuterochloride (GdnDCl). Moderately destabilizing concentrations of the denaturant (1.5 M GdnDCl) completely removes the Arrhenius-like feature from the temperature window employed. The Arrhenius to anti-Arrhenius transition can be explained by the heat capacity model where temperature strengthens ground state interactions, perhaps hydrophobic in nature. The effect of the denaturant may appear to arise from direct protein-denaturant interactions that are structure-stabilizing under subdenaturing conditions. The temperature distribution of rate constants under different stability conditions also suggests that the prefactor in

Arrhenius-like relations is temperature dependent. Although the use of the transition state theory (TST) offers several challenges associated with data interpretation, the present results and a consideration of others published earlier provide evidence for complexity of ring-flip dynamics in proteins.

4.2 Introduction

In recent years, structural studies by X-ray techniques (1) and NMR spectroscopy (2, 3), and mutational analyses (4, 5) have greatly facilitated relating structure, conformational changes, function, and stability of proteins. Investigations of protein dynamics are still in its infancy though (6). Internal dynamics that include small amplitude collective motions (7) and large amplitude cooperative breathing modes (8), and other structural fluctuations and deformations at the subglobal level (9) are relatively less understood. The complexity of certain of these processes has been recognized for quite some time in both experimental (6, 10, 11) and theoretical studies (12, 13). For analytical solutions of experimental data, they often appear complex because of either unusual time evolution or complex temperature dependence or both. In theoretical studies, the complexity surfaces when a suitable reaction coordinate and an appropriate transition state along the coordinate is difficult to define. There, of course, is the problem of current atomistic simulations not reaching times longer than microseconds. However, a comprehensive description of protein motions is important not only for connecting structure and function to dynamics, but also for understanding their role in folding and stability of native and natively unfolded states (14-16).

This chapter presents the temperature and denaturant related complexity of the ring motion of Y97 side chain (180° rotational jump of the ring about the C_β–C_γ bond axis) of horse ferricytochrome *c* observed by cross-relaxation suppressed exchange NMR spectroscopy (EXSY). The rate of the process does not follow Arrhenius temperature dependence under native and native-like conditions. The changes in the rate-temperature trend under different conditions of protein stability appear to indicate temperature dependence of the Arrhenius-type prefactor. Data are analyzed invoking a heat capacity model that assumes a considerable change in the ground-state heat capacity with temperature.

Starting with the seminal work of Wüthrich and Wagner (1975) on BPTI, the rate-temperature relationship for aromatic ring flip has been reported for just about four proteins to date: BPTI (17-20), ferrocytochrome *c* (21), iso-2-cytochrome *c* (22), and HPr protein (23). Although the activation parameters from all data sets, generated largely by simulation of variable temperature NMR spectra, were extracted by assuming temperature independence of enthalpy and entropy in the Eyring equation, it is believed that they are generally temperature dependent to different extents, and hence could be modeled better by using temperature dependent functions. Indeed, this was indirectly observed in a revisit to the BPTI Y35 ring-flip dynamics (18). Also, rate-temperature curves calculated for several cases (20) using activation parameters published earlier indicate visibly non-Arrhenius behavior. Together with the observations of the present study of ferricytochrome *c*, ring-flip dynamics appear generally complex.

4.3 Materials and Methods

All experiments were performed using cytochrome *c* (Type VI from Sigma) and GdnHCl (from USB, GE Healthcare). The labile hydrogens of both of which were preexchanged for deuterium. Exposure of D_2O solutions of cytochrome *c* to pH 10 at 60°C for an hour followed by lyophilization of the neutralized solution caused complete deuteration of the labile proton sites. GdnDCl was prepared by repeated lyophilization of D_2O solutions of the denaturant.

4.3.1 NMR Spectroscopy

A 2-mM solution of cytochrome *c* containing a desired concentration of GdnDCl was prepared in 50 mM phosphate buffer in D_2O, pH 6.5. The same sample was used to take variable-temperature NOESY and EXSY spectra. Each 2D spectrum was of 400 t_1 increments and 2048 t_2 data points over a spectral width of 18 ppm. Quadrature detection along the indirectly detected dimension was achieved by the States-TPPI method. For exchange rate measurement at a given concentration of GdnDCl and temperature, EXSY spectra (24) were recorded using five mixing times (10, 20, 30, 40, and 50 ms). The details of the mixing pulse scheme are shown in Figure 1. With 128 scans averaged for each t_1, the time required for a 2D spectrum was over 16 hours at 400 MHz. Spectra were recorded using a Bruker spectrometer, and processed and analyzed with XWIN NMR (Bruker) and Felix (Accelrys) softwares.

4.4 Results and Discussion

4.4.1 Y97 ring-flip rate constant as a function of protein stability coordinate measured by NMR spectroscopy

The earliest NMR investigations of protein dynamics that employed techniques of 1D time-resolved saturation transfer, isotope-edited spectroscopy in conjunction with spectra simulation, and analysis of temperature dependence of resonances have made several fundamental contributions to the understanding of aromatic ring rotational motions in proteins (17, 20, 22, 25). The advent of 2D NMR spectroscopy has greatly facilitated studies of such dynamic processes that are too slow to affect the lineshapes (26, 27). Subsequent introduction of pulse sequences designed to produce chemical exchange cross-peaks by elimination of cross-relaxation and spin coupling-induced coherence transfer effects in 2D spectra (24, 28) has further augmented the NMR approach to studies of exchange dynamics in proteins. Indeed, numerous recent work have employed EXSY to investigate dynamic effects; particularly, the ring inversion phenomenon in biology and chemistry (20, 29-32).

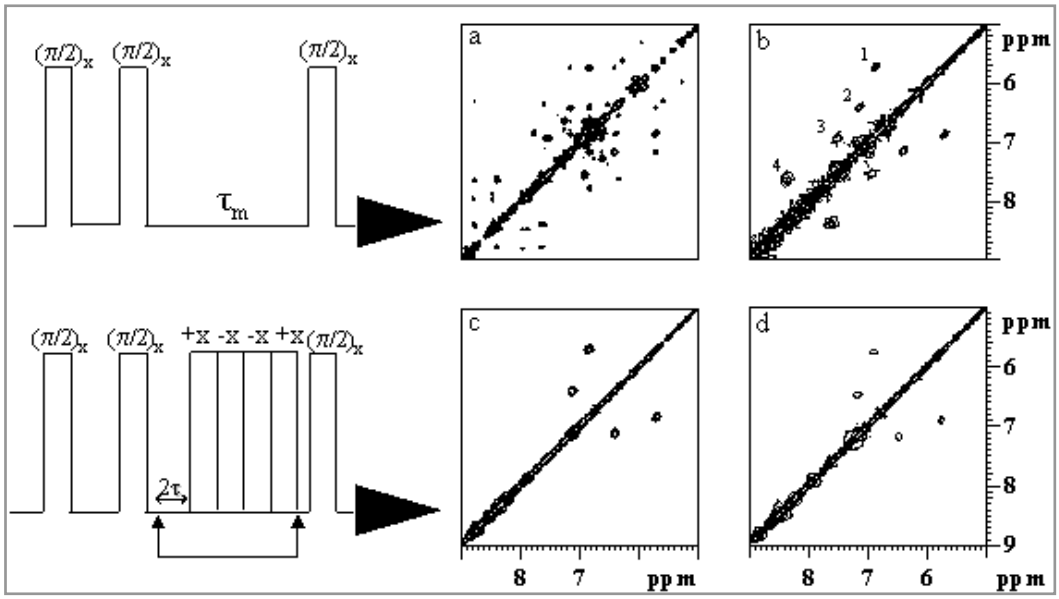


Figure 1 Relaxation suppressed EXSY largely eliminates NOESY peaks. (a) The aromatic region of a NOESY spectrum of ferricyt *c* at 22°C, pH 6.3. The same region of EXSY spectra recorded with the same sample at 20°C (b), 30°C (c), and 35°C (d). Cross peak labels in (b) are 1: $Y97_{3,5}$; 2: $Y97_{2,6}$; 3: $F10_{2,6}$; and 4: $F10_{3,5}$. The pulse sequences shown at the left are for NOESY (top) and EXSY (bottom).

With the objective of characterizing the NMR-defined slow dynamic processes in cytochrome *c* under native to mild denaturing conditions, It has collected a large set of exchange spectra recorded at several low concentrations of GdnDCl, each as a function of temperature, and in turn as a function of mixing time in the 0-50 ms range. This approach allows mapping the activation parameters for the rate processes, as the protein is progressively destabilized. Figure 1 presents excerpts from the data set showing the extraction and temperature dependence of cross peaks due to rotational motions of Y97 and F10

rings in native cytochrome *c*. Another ten cross peaks assigned to P30 $C_\beta H$, L68 $C_\delta H_3$, Y48/T42 $C_\alpha H$, T28 $C_\gamma H$, F36/T47 $C_\beta H$, K22/G23 $C_\alpha H$, C14/Q16 $C_\gamma H$, $F82_{2,6}$, F46 $C_3 H$, and H26 Cβ appear in the aliphatic region of the native-state spectrum at 22°C.

To calculate the rate coefficients and activation parameters of the dynamic processes along the protein stability coordinate, it is necessary that the set of exchange-connected cross peaks in the NMR spectrum consistently appear at all concentrations of the denaturant in a suitable range of temperature. However, since both temperature and denaturant affect the dynamics, those processes the interconversion times of which are shorter or longer than the mixing time of the NMR experiment fail to produce cross peaks; shorter interconversion times lead to exchange broadening, and longer interconversion times make only a minute fraction of exchanging population available for interrogation. Such conditions do not produce quantifiable cross peaks. The requirement of the suitable exchange time severely constrains the number of cross peaks, and hence the dynamic processes that can be examined in detail. Two cross peaks, $Y97_{3,5}$ and $Y97_{2,6}$, (peaks 1 and 2, respectively, in Figure 1b) corresponding to 180° rotational motion of the Y97 ring are consistently observed in the 288-313 K temperature range at 0-1.5 M GdnDCl concentration, and were used for further analysis. The ring-flipping rate constant, k_{flip}, was calculated according to Ernst et al. (26).

$$\ln(-2\Gamma + 1) = -2 k_{\text{flip}} \tau_m, \qquad (1)$$

where $\Gamma = V_c/(V_c + V_d)$ with V_d and V_c the diagonal and cross-peak volumes, respectively, and τ_m is the mixing time.

Figure 2 shows an example of the plot of $\ln(-2\Gamma+1)$ vs τ_m.

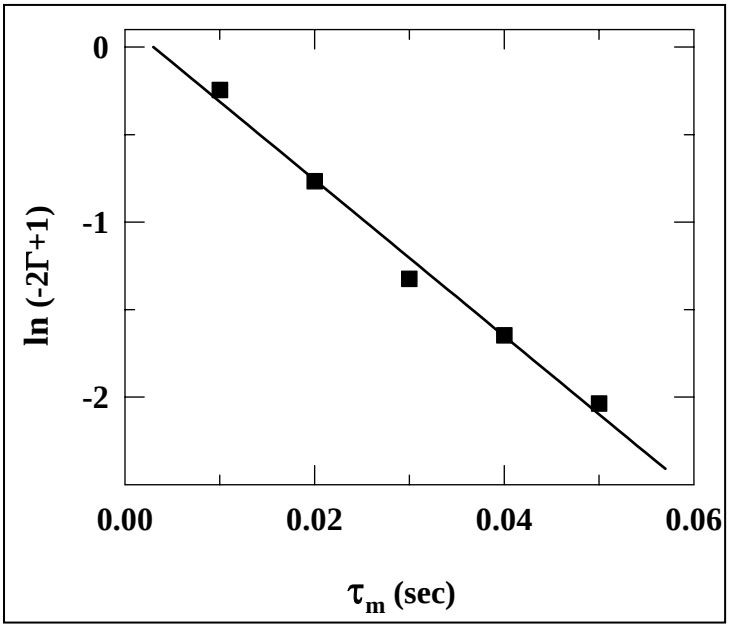


Figure 2 Extraction of the flip rate constant for the Y97 ring in the native protein at 20°C, pH 6.3. The experimental procedure and data analyses are detailed in the text. Results are in table 1.

Values of k_{flip} thus obtained should preferably be corrected for the viscosity effect due to GdnDCl. Although the relative viscosity of the medium increases only marginally in the range of denaturant concentrations used in this study (0-1.5 M), a correction can be introduced according to $k_{\text{flip}} \propto 1/\eta$, where η is the bulk viscosity. To eliminate the effect of viscosity due to GdnDCl on the ring-flip rate constant, it has simply used relative viscosity η/η_o, where η_o is the viscosity in the absence of GdnDCl. The corrected k_{flip} is then given by the relation

$$\ln k_{\text{flip}}^{\text{corr}} = \ln k_{\text{flip}} + \ln\left(\frac{\eta}{\eta_o}\right). \quad (2)$$

To obtain the values of η/η_o, viscosity values of aqueous solutions of GdnDCl from literature (33) were used to find the following empirical relation

$$\frac{\eta}{\eta_o} = 1 + 0.005[\text{D}]^{2.6144} + 0.018[\text{D}]^{0.6594} + 0.01213[\text{D}]^{0.6636}, \quad (3)$$

where [D] represents molar concentration of GdnDCl.

Table 1. Values of k_{flip} (s^{-1}) as a function of temperature and GdnDCl concentration

Temperature (K)	GdnDCl (M)			
	0	0.5	1.0	1.5
289	4.60	7.56	9.47	10.66
293	13.25	11.85	10.41	9.39
298	22.30	14.56	15.75	2.10
303	33.40	13.91	3.36	0.89
306	26.48	–	–	–
308	–	3.38	–	–
313	–	0.68	–	–

Error in determination of rates: 2-20%

4.4.2 Non-Arrhenius behavior for the ring-flip rate constant

The temperature dependence of k_{flip} for Y97 at four concentrations of GdnDCl is presented in Figure 3. The plots of $\ln(k_{flip}/T)$ against 1000/T are not at all linear. The gradients for the curves at 0, 0.5, and 1 M GdnDCl make continuous transition from negative at low temperatures to positive at high temperatures, more accentuated for the latter two concentrations of the denaturant. In general, it has identified the curved temperature dependence of rate constants with non-Arrhenius behavior, which could be operationally termed Arrhenius-like (negative slope) or anti-Arrhenius (positive slope) for a small segment of the temperature axis. Indeed, one notices a gradual turn from Arrhenius-like to anti-Arrhenius behavior as the temperature is raised from 15 to 40 ºC. At 1.5 M GdnDCl, where the protein is still native-like but denaturing conditions are approached (34), the flip rate constant shows only anti-Arrhenius temperature dependence.

To explain the diminishing enthalpic but growing entropic contribution to the ring isomerization barrier with increasing temperature, it is chose to use the transition state theory (TST) rate expression. For justification of this choice, a detailed consideration of reaction rate theory more appropriate for protein reactions is required, a general understanding of which is still not available. It is also uncertain if the ring motion is fully in the diffusive regime, so a Kramers-like rate description becomes essential. The magnitude of the internal friction, and the extent of influence of system-solvent collisions on the ring-flip reactive trajectories across the barrier are other uncertainties. Further, the ring motion is unlikely to be associated with a low and flat barrier for which one might expect a

diffusive rate. On the other hand, within the limitations of presently available reaction rate theories and microscopic models, a full justification for the adequacy of TST is difficult to achieve.

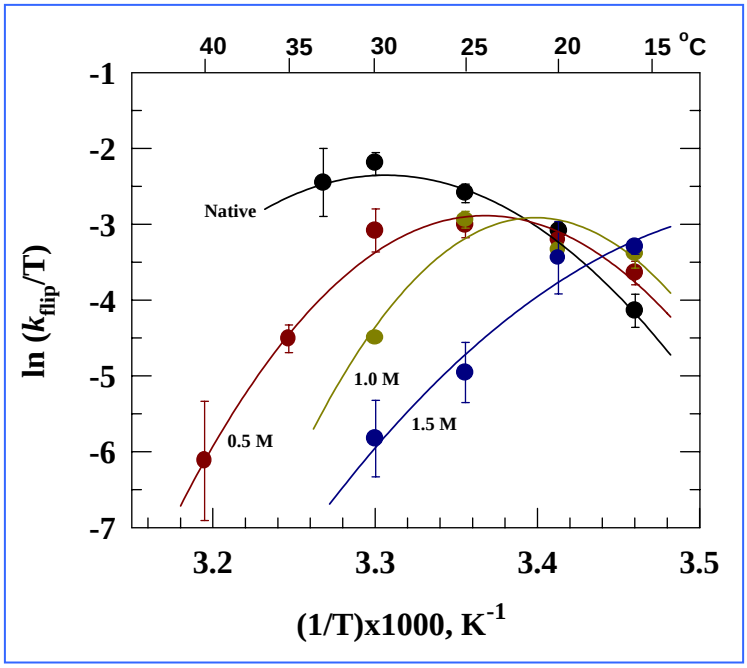


Figure 3 Temperature dependence of the Y97 ring-flip rate constant at four concentrations of GdnDCl: 0 M (black), 0.5 M (red), 1.0 M (yellow), and 1.5 M (blue). Solid lines are fits to data by equation 6. Values of $\Delta H^{\ddagger}$, $\Delta C_p^{\ddagger}$, and $\Delta S^{\ddagger}$ obtained from the fits are listed in Table 2.

To make progress with the data at hand, the analysis was continued with the Eyring-type expression

$$\ln k_{\text{flip}} = \ln A - \frac{\Delta G^{\ddagger}}{RT}, \qquad (4)$$

where $\Delta G^{\ddagger}=\Delta H^{\ddagger}-T\Delta S^{\ddagger}$ is the barrier free-energy. The exact nature of the prefactor A is not known; it is perhaps determined by microscopic dynamics entailed in

barrier crossing (14). $\Delta G^{\ddagger}$ is expanded by casting the enthalpy and entropy changes in their basic forms

$$\Delta H_T^{\ddagger} = \Delta H_{T_0}^{\ddagger} + \int_{T_0}^{T} \Delta C_p^{\ddagger} dT, \quad \text{and} \quad \Delta S_T^{\ddagger} = \Delta S_{T_0}^{\ddagger} + \int_{T_0}^{T} \Delta C_p^{\ddagger} \frac{dT}{T}, \tag{5}$$

which allow evaluation of enthalpy and entropy changes at temperatures T with respect to a reference temperature T_0. $\Delta C_p^{\ddagger}$ is the difference in heat capacity of the transition state and the initial state.

Table 2. Activation parameter for flipping motion of the Y97 ring

	Thermodynamic Parameters		
GdnDCl/M	$\Delta H^{\ddagger}$ (kcal mol^{-1})	$\Delta C_p^{\ddagger}$ (kcal mol^{-1} 1000K^{-1})	$\Delta S^{\ddagger}$ (kcal mol^{-1} 1000K^{-1})
0.0	15.3	-3.448	0.019
0.5	-4.9	-4.697	-0.049
1.0	-25	-6.645	-0.117
1.5	-38	-2.196	-0.165

Estimated error range for these values is 3-5%

T_0 is set at 25°C (see below), although for accurate calculations of heat-induced conformational transitions, the midpoint temperature should be used. Thus, the equation

$$\ln k_{flip} = \ln A - \frac{1}{RT}\left[\Delta H_{T_0}^{\ddagger} + \Delta Cp^{\ddagger}(T - T_0) - T\left(\Delta S_{T_0}^{\ddagger} + \Delta C_p^{\ddagger} \ln\left(\frac{T}{T_0}\right)\right)\right] \tag{6}$$

was used to fit the data shown in Figure 3. Values of $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, and $\Delta C_p^{\ddagger}$ for all four curves are listed in Table 2.

4.4.3 Variation of the activation parameters along the protein stability coordinate

The details of the ring-flip activation parameters in the presence of low concentrations of GdnDCl could potentially provide important information regarding collective motion, stability, and interactions of the relevant structural elements under subdenaturing to denaturing conditions. For scaling the activation parameter, it is reasonable to assume that the changes of the differences in thermodynamic properties between the transition state and the ground state under different conditions of temperature and stability are largely due to effects at the ground state level (35), because the state variables are expected to cause relatively less changes in the thermodynamic properties of a changing transition state which is structurally diminished and distorted.

The temperature gradient of $\Delta H^{\ddagger}$ (Figure 4a) is negative for all concentrations of GdnDCl. The curves precisely indicate the turn from Arrhenius to anti-Arrhenius through a relatively smaller Arrhenius-like zone of temperature. For all conditions, the value of $\Delta H^{\ddagger}$ passes from positive to negative with increasing temperature registering 'zero' at 298 K, a consequence of choosing T_o=298 K. Obviously, T_o cannot be 298 K all along the stability coordinate. Even for the native state, the chosen T_o may be way off. It is hard to determine the melting temperature for the specific subglobal part of the protein that is involved in the ring isomerization process. Nonetheless, the arbitrary choice of T_o can still

provide qualitative information and relative values of activation parameters, since the temperature gradient of $\Delta H^{\ddagger}$ will depend little on T_o.

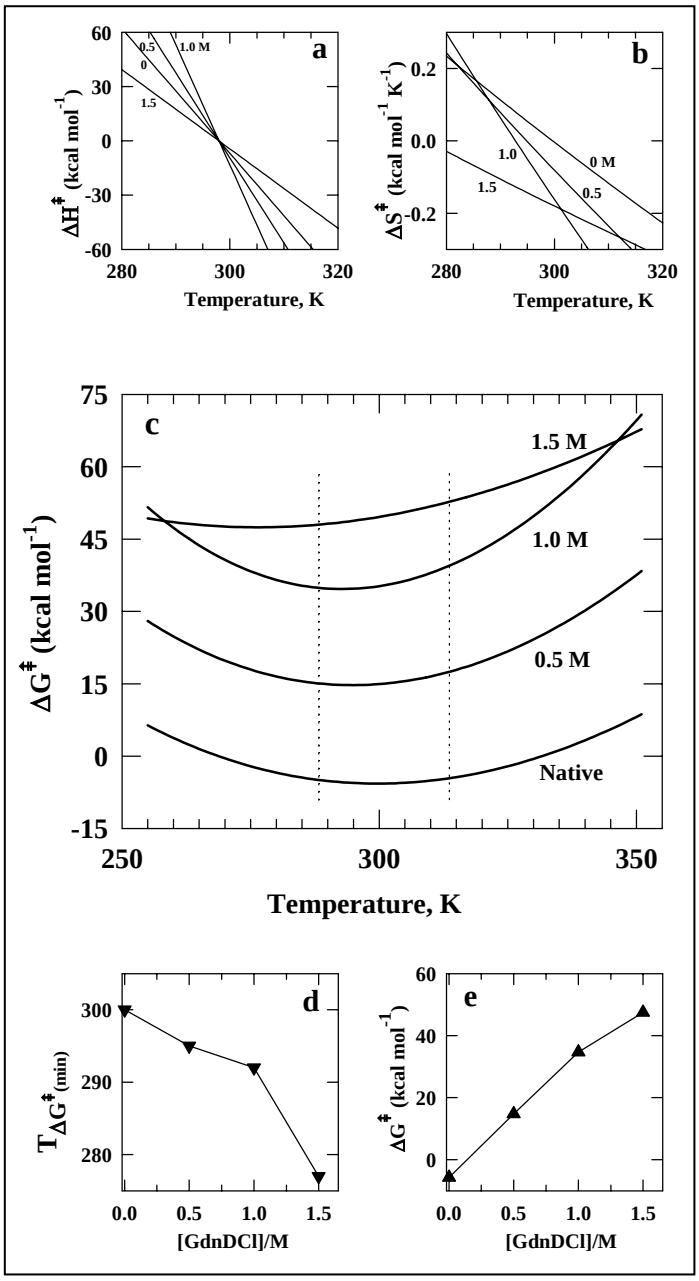


Figure 4 Thermodynamic parameters for the Y97 ring-flip dynamics calculated from the data shown in Fig. 3. Temperature distribution of (a) $\Delta H^{\ddagger}$, (b) $\Delta S^{\ddagger}$, and (c) $\Delta G^{\ddagger}$ for the four concentrations of GdnDCl indicated. The vertical dotted lines in panel (c) show the temperature range used in this study. (d) GdnDCl distribution of the temperature corresponding to minimal $\Delta G^{\ddagger}$. (e) GdnDCl distribution of the minimal $\Delta G^{\ddagger}$.

The increase in the slope of the enthalpy curves with increments of GdnDCl up to 1 M (Figure 4a) suggests that in the subdenaturing limit of its concentration, the denaturant acts to increase the difference in heat capacity between the transition state and the ground state ($\Delta C_p^{\ddagger}$). At 1.5 M GdnDCl, denaturing conditions are approached, and $\Delta C_p^{\ddagger}$ begins to shrink. This is also seen from the values listed in Table 2, where the –ve sign for $\Delta C_p^{\ddagger}$ indicates larger values for the ground state C_p. An interpretation of this result could be that within the subdenaturing limit of its concentration, GdnDCl somehow modulates the solvent properties and introduces additional interactions by virtue of its mechanism of action on proteins. This aspect of GdnDCl-protein interaction and the consequences thereof is being discussed in detail in a later section.

Like the temperature dependence of $\Delta H^{\ddagger}$, the value of $\Delta S^{\ddagger}$ passes from positive to negative with increasing temperature for all concentrations of GdnDCl (Figure 4b). Clearly, the entropic contribution to the barrier energy dominates in the respective anti-Arrhenius temperature zones (Figure 3). The slope of the entropy-temperature curve increases in going from 0 to 1 M denaturant, suggesting that in the presence of increasing subdenaturing concentrations of GdnDCl, relatively large-scale conformational constraint must be negotiated to make to the transition state. The constraint is generated by the binding interaction of the denaturant with the protein. The decrease of the slope at 1.5 M denaturant is due to the onset of denaturation, where entropy-lowering restraints in the ground state begin to weaken because of increasing structure-breaking action of the denaturant in at least that part of the protein molecule which harbors the flipping ring.

Temperature variation of $\Delta G^{\ddagger}$ for different concentrations of GdnDCl, plotted in figure 4c, shows vertical shift of energy and horizontal shift of the temperature corresponding to the energy minimum. The magnitude and the sign of these shifts are specific to the reference temperature chosen (T_o=298 K). For the same reason, $\Delta G^{\ddagger}$ value for the native-state protein comes out as negative, and therefore, the shifts obtained from Figure 4c must be scaled relatively. Figure 4d shows how the temperature corresponding to minimal $\Delta G^{\ddagger}$ decreases with increasing amount of GdnDCl; this is due to the general destabilizing action of the denaturant. On the other hand, the increase in the minimal $\Delta G^{\ddagger}$ with denaturant (Figure 4e), implying a lowering of the free energy of the ground state relative to the transition state, is due to the stabilizing action of the denaturant. The increase is nonlinear though, suggesting that the initial protein stabilization caused by subdenaturing amounts of GdnDCl withers as denaturation begins to set in (see below). By magnitude, the increase of $\Delta G^{\ddagger}$ as a function of GdnDCl is nearly the same within the temperature range employed in our experiment (marked by dotted line in Figure 4c). The observed variation of rates in this temperature range, however, appears complex. For example, the k_{flip} value for 0 M GdnDCl relative to that for 1 M GdnDCl is an order of magnitude larger at 30°C, is matching at ~20°C, and is smaller at lower temperature. These differences are not due to error of measurement. They rather originate from a temperature dependence of the prefactor A, possibly compounded by the effect of GdnDCl on the internal friction of the protein.

4.4.4 Temperature dependence of $\Delta G^{\ddagger}$

In Arrhenius-type rate expressions, $\Delta G^{\ddagger}$ should be virtually independent of temperature. However, variable degrees of temperature dependence of $\Delta G^{\ddagger}$ have been described for different systems, including simple metathesis reactions (36, 37) enzyme mechanisms (38), DNA fluctuations (39), protein-ligand reactions (10), and protein refolding kinetics in general (35, 40). The unusual temperature dependence of certain gas-phase reactions has been explained by a modified transition state theory (37). For protein reactions, folding processes in particular, two explanations are provided to account for temperature dependence of $\Delta G^{\ddagger}$. The first, explicable by statistical mechanical principles, is derived for a random walk of an activation within a Gaussian distribution of conformations (41), and is often referred to as super-Arrhenius temperature dependence where $exp–(\Delta G^{\ddagger}/RT)^2$ is used instead of $exp–(\Delta G^{\ddagger}/RT)$ in the rate-temperature expression (42, 43). The second, based on foolproof experimental evidence for strong temperature dependence of hydrophobic interactions (44, 45), proposes that temperature dependence of $\Delta G^{\ddagger}$ is due to the disruption of large buried apolar surfaces in going from the ground state to the transition state; the associated heat capacity change $\Delta C_p^{\ddagger}$, and hence $\Delta G^{\ddagger}$, is strongly temperature dependent (35, 40). Another explanation, more appropriate for bimolecular protein reactions, accords complex temperature dependence of $\Delta G^{\ddagger}$ when the binding rate competes with the rates of conformational fluctuations or relaxations of the target site. A well-documented

example is rebinding of CO after flash photolysis of carbonmonoxymyoglobin at low temperature (10, 46).

Of these, the second explanation corresponds to the heat capacity model used here for analyzing the ring-flip data. The heat capacity model draws support from strong connections recognized for large amplitude breathing motions of the protein and the ring isomerization event (17, 18, 20, 23, 47). Since deformations of structure and interactions are involved in large amplitude breathing modes (8, 48), a difference in heat capacity between the transition state and the ground state ($\Delta C_p^{\ddagger}$) is expected. It is also generally accepted that $\Delta C_p^{\ddagger}$ contributes at least in part to non-Arrhenius temperature dependence of protein reactions. The less preference for the use of a super Arrhenius-type model for ring-flip analysis rests on two considerations. First, unlike the folding of protein chains, ring-flip dynamics is not quite a diffusive process in configurational space. Second, the super-Arrhenius relation applies better for an extended temperature range. At temperatures well above the glass transition, the rate coefficient is smaller than the relation predicts (46).

4.4.5 Challenges associated with data interpretation

The analyses and discussions presented above also illustrate some of the major difficulties associated with data interpretation and calculation of heat capacity differences between ground and excited states. Fundamentally, the use of a kinetic theory and an Eyring-like expression modeled on simple chemical reaction may not necessarily lead to uncontestable insight. Even when granted, the use of an arbitrarily chosen constant reference temperature (T_o=298 K) provides at best a qualitative picture, because the value of T_o must change

according to the protein stability influenced by the denaturant. Accurate values for T_o relevant for the subglobal part harboring the ring are often difficult to obtain. Another concern is that the ground and excited states vary greatly as a function of denaturant concentration and temperature. Clearly, a detailed knowledge of the landscape topography under denaturing to native-like conditions (i.e., near the bottom of the funnel) is essential for a more quantitative description of the problem. It should mention that the strong temperature dependence of exchange rate of a buried water molecule in BPTI has been previously associated with internal motions involving multiple conformational substate interconversions in a rugged energy landscape (49).

4.4.6 Structural basis for non-Arrhenius ring-flip dynamics in cytochrome *c*

As mentioned already, fairly good evidence exists for the concerted control of large-scale protein motions on the ring isomerization process. Assuming generality of this, numerous influential work on structure and folding of cytochrome *c* facilitates surmising the motional mode that could drive the Y97 ring-flip process. Y97 is a resident of the C-terminal helix, and it is near this residue that the N- and C-terminal helices form a tight contact to produce an orthogonal geometry of helix docking (Figure 5; see also ref. 50, 51).

This contact region is made of predominantly hydrophobic residues that are highly conserved, and the majority of the very slowly exchanging amide protons, including residues 94-99, are found on the N-and C-terminal helices (49). Sufficient exposure of the hydrophobic surface may be required to allow flipping motion of the Y97 ring.

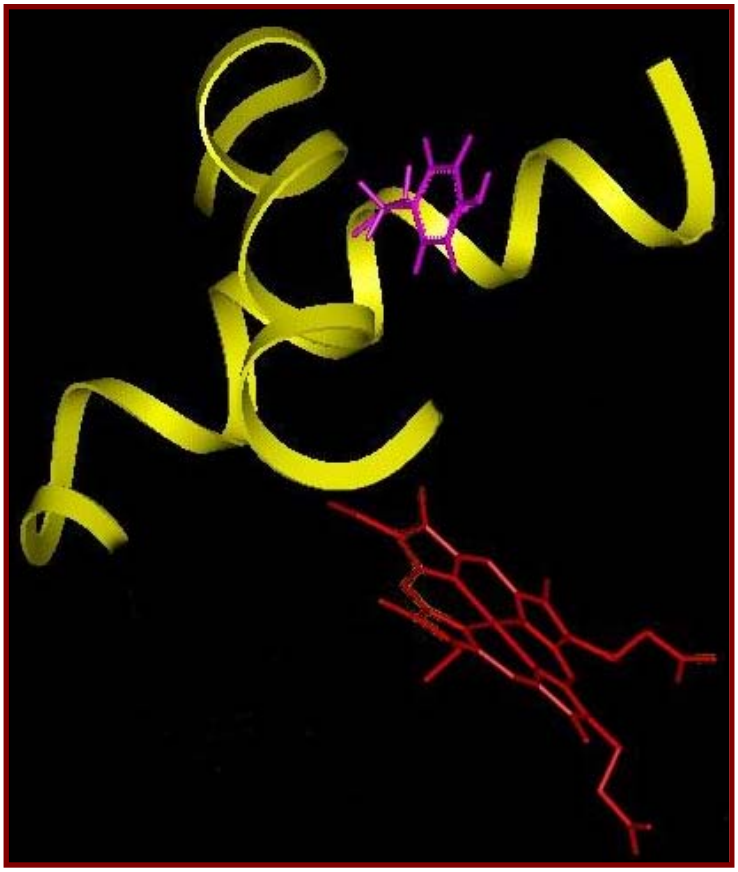


Figure 5 The N- and C-terminal helices of cytochrome *c* with the disposition of the Y97 ring as seen in the crystal structure (PDB file: 1HRC). Also shown is the heme ring.

It does not have a direct evidence to show how the motion of the helices arranges for the ring-flip. Reports that an early kinetic intermediate of cytochrome *c* comprises the structures of these two helices (50, 52), the corresponding peptides in solution self-associate (53), and that the associated structure of these helices in the native protein forms a low-energy cooperative unfolding unit simulating the

global unfolding behavior (54), all provide a seeming basis to assume that intrinsic association and dissociation or relative sliding motions of the docked helices render Y97 ring flipping. In thermodynamic terms, the relative motions of the helices that expose the hydrophobic surface in the vicinity of Y97 involve a sizable heat capacity C_p, the magnitude of which increases with strengthening of the hydrophobic interactions by temperature. This motional mechanism must be treated as a possibility. The $\Delta C_p^{\ddagger}$ value is negative for all curves (Table 2), implying that an extensive hydrophobic surface characterizes the ground state, and a large heat is involved in the dissolution of the surface.

4.4.7 Complex action of GdnDCl and temperature on protein stability and ring-flip rate

One of the major objectives of this study was to learn how the rates of slow dynamic processes, Y97 ring-flip in the context, change as the protein is placed in incrementally destabilizing conditions. A gradual decrease of k_{flip} occurs with increments of GdnDCl, clearly noticeable at temperatures higher than ~20°C. This general decrease can be explained by constrained dynamics of the protein as a result of its interaction with the denaturant. An explanation and the illustration of this effect in the case of cytochrome *c* have been provided from our lab with different set of experiments (55, 56). Briefly, GdnDCl can interact directly with protein backbone and side-chains by variable-length hydrogen bonding and van der Waals interactions (57-60). Such interactions produce cross-links or non-specific network of intraprotein interactions (57), leading to protein stiffening, reduction of motional freedom, increased internal friction, and thus entropic stabilization. In the present case also, GdnDCl-induced constraints on the motions

of N- and C-terminal helices of cytochrome *c* is likely to retard the ring-flip motion. The protein stabilizing effect of subdenaturing amount of GdnDCl is overrun by its structure unfolding action when used at higher concentrations. The present results indicate that the denaturing effect begins to appear at ~1.5 M GdnDCl.

But the observed k_{flip} dependence on GdnDCl is more complex when the temperature variable is brought into picture. The k_{flip}-temperature curves appear to meet near 20°C, and diverge on either side showing a reversal of temperature dependence (Figure 3), suggesting that the ring flip rate constant near this temperature does not change with GdnDCl as long as the latter is present in subdenaturing concentrations. The slight difference in the behavior of the curve for 1.5 M GdnDCl is likely due to approach of denaturing conditions. This behavior of k_{flip} as a function of denaturant and temperature (Figure 3) is inconsistent with the temperature distribution of $\Delta G^{\ddagger}$ at the four GdnDCl concentrations (Figure 4c). The vertical dotted lines in Figure 4c show the temperature range used in our experiments, and within this range the variation of $\Delta G^{\ddagger}$ with denaturant is very similar for all temperatures, suggesting that the k_{flip} curves for different denaturant concentrations should not converge at any temperature within this range. This observation should mean that the prefactor *A* is acting in a temperature dependent manner leading to the coalescence of k_{flip} values at ~20°C. But why this happens near 20°C is not clear at present. In Eyring-type relations, *A* generally shows only weak temperature dependence. However, the fits of the data to the heat capacity model with three floating parameters ($\Delta H^{\ddagger}$, $\Delta C_p^{\ddagger}$, $\Delta S^{\ddagger}$) cannot rule out a significant nonlinear dependence of

A on temperature in a denaturant dependent manner. Studies on the effect of GdnDCl on *A* in Eyring-type expressions will shed light on this issue.

4.4.8 Aromatic ring-flip motions in other proteins

Following the earliest study with BPTI (25), ring-flip dynamics of only four proteins, namely, ferrocytochrome *c* (21), fd bacteriophase (60), yeast iso-2-cytochrome *c* (22), and *S. carnosus* HPr (23) have been studied in some detail, although the problem in BPTI has been revisited on several occasions (17, 18, 20, 47). In general, the rate strongly varies with packing and dynamics of the ring environment. This is the likely reason why the rates for different aromatic rings within a protein, and for one protein from another could differ, often very significantly, under identical experimental conditions. For example, the rates for Y35 and Y23 of BPTI near room temperature may vary by an order of magnitude, from ~10^2 to ~10^3 s^{-1} (17), and the aromatic rings of fd bacteriophage flip at ~10^6 s^{-1} (61).

Even more interesting is the complex temperature dependence, which the present study exposes in some detail. A review of earlier rate-temperature data for ring dynamics in different proteins also reveals some deviation from simple Arrhenius behavior. The temperature dependence of Y35 flip rate constant in BPTI (17), for example, would appear nonlinear unless large errors in data are allowed. As Karplus points out (13), in the reinvestigation of Y35 flip rate by the use of high resolution EXSY, Otting et al. (18) tie the activation parameters for the flipping process with those for disulfide conformational exchange motion at high temperature. The flip rate at low temperature (<20°C) was not measured. However, since the temperature graph for the activation energy of the disulfide

exchange motion shows different dependences at low and high temperatures registering a sharp inflection centered around 20°C, and because the conformational motion and flip motion are linked, the latter should also follow the trend for the former process and hence a nonlinear function of temperature. In a more recent paper, Skalicky et al. (20) used activation parameters published earlier to calculate rate-temperature curves for several residues of BPTI, iso-2-cytochrome *c*, and horse cytochrome *c*. The curves in their semilogarithmic plot of k_{flip} versus T, calculated using $1/k_{flip}=h/(k_BT)\exp[(\Delta H^{\ddagger}/RT-\Delta S^{\ddagger}/R)]$, are clearly not linear. The flip dynamics of Y6 in *S. carnosus* HPr deduced from spectra simulation (23) has been analyzed assuming temperature independence of the activation parameters. Here also, the rate-temperature behavior, especially at 200 Mpa, may not simply be linear; an even closer look is perhaps worth considering. Thus, evidences are rather strong that aromatic ring dynamics in proteins are complex, and they do not follow an Arrhenius-like equation.

4.5 Conclusions

Unusual temperature dependence for dynamic processes like ring rotation within a temperature range where spectroscopic measurements do not reveal any change in the time-averaged native structure of the protein is a complex phenomenon by itself. The complexity is compounded when subdenaturing amounts of denaturant act on the protein. Aromatic ring isomerization in dense systems like proteins and polymers involves large amplitude cooperative motion of the ring and certain structural elements, often engaging the chain backbone (18, 62). It has used the heat capacity model to explain the cytochrome *c* results.

Further studies may reveal alternative, perhaps even better models for such phenomena. Irrespective of the merit of the model, the data and evidences presented are compelling that aromatic ring motions are complex.

4.6 References

1. Pellicena, P., Kuriyan, J. *Curr Opin Struct Biol* **2006,** *16,* 702-709.
2. Grishaev, A., Wu, J., Trewhella, J., Bax, A. *J Am Chem Soc* **2005,** *127,* 16621-16628.
3. Stangler, T., Hartmann, R., Willbold, D., Koenig, B. W. *Z .Phys Chem* **2006,** *220,*567-613.
4. Dill, K. A., Bromberg, S., Yue, K. Z., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. *Protein Sci.* **1995,** 4, 561-602.
5. Shaw, B. F., Valentine, J. S. *Trends Biochem Sci* **2007,** *32,* 78-85.
6. Fenimore, P. W., Frauenfelder, H., McMahon, B. H., Young, R. D. *Proc Natl Acad Sci USA* **2004,** *101,* 14408-14413.
7. Karplus, M. *Methods Enzymol.* **1986,** *131*,283-307.
8. Englander, S. W., Downer, N. W., Teitelbaum, H. *Annu Rev Biochem* **1972,** *41*, 903-924.
9. Ansari, A., Berendzen, J., Bowne, S. F., frauenfelder, H., Iben, I. E. T., Sauke, T. B., Shyamsunder, E., young, R. D. *Proc. Natl. Acad. Sci. USA* **1985,** *82*, 5000-5004.
10. Austin, R. H., Beeson, K. W., Eisenstein, L. *Biochemistry* **1975,** *14,* 5355-5373.
11. Steinbach, P. J., Ansari, A., Berendzen, J. et al *Biochemistry* **1991,** *30*, 3988-

4001.

12. McCammon, J. A., Wolynes, P. G., Karplus, M. *Biochemistry* **1979,** *18*, 927-942.
13. Karplus, M. *J Phys Chem B* **2000,** *104,* 11-27.
14. Portman, J. J., Takada, S., Wolynes, P. G. *J Chem Phys* **2001,** *114*,:5082-5096.
15. Okazaki, K-i., Koga, N., Takada, S., Onuchic, J. N., et al *Proc Natl Acad Sci USA* **2006,** *103,* 11844-11849.
16. Doan-Nguyen, V., Loria, J. P. *Protein Sci.* **2007,** *16,* 20-29.
17. Wagner, G., DeMarco, A., Wüthrich, K. *Biohys Struct Mechanism* **1976,** *2,* 139-158.
18. Otting, G., Liepinsh, E., Wüthrich, K. *Biochemistry* **1993,** *32*, 3571-3582.
19. Li, H., Yamada, H., Akasaka, K. *Biochemistry* **1998,** *37,* 1167-1173.
20. Skalicky, J. J., Mills, J. L., Sharma, S., & Szyperski, T. *J. Am. Chem. Soc.* **2001,** *123*, 388-397.
21. Campbell I. D, Dobson C. M, Moore G. R, Perkins S.J, Williams R. J. P *FEBS Lett*. **1976**, *70,* 96-100..
22. Nall, B. T., Zuniga, E. H. *Biochemistry* **1990,** *29,* 7576-7584.
23. Hattori, M., Li, H., Yamada, H., Akasaka, K., Hengstenberg, W., Gronwald, W., and Kalbitzer, H. R. *Protein Science* **2004,** *13,* 3104-3114.
24. Fejzo, J., Zolnai, Z., Macura, S., Markley, J. L. *J Magn Reson* **1990,** *88*, 93-110.
25. Wüthrich, K., Wagner, G. *FEBS Lett* **1975,** *50,* 265-268.

26. Ernst, R. R., Bodenhausen, G., & Wokaun, A. *Principles of Nuclear Nagnetic Resonance in One and Two Dimensions* **1988** (Clarendon Press, Oxford).
27. Roder, H. *Methods Enzymol.* **1989,** *176,* 446-473.
28. Fejzo, J., Westler, W. M., Macura, S., Markley, J. L. *J Magn Reson.* **1991,** *92,* 20-29.
29. Shokhirev, N. V., Shokhireva, T. K., Polam, J. R., Watson, C. T. et al. *J Phys Chem A* **1997,** *101,* 2778-2786
30. Whittaker, S. B-M., Boetzel, R., MacDonald, C., Lian L-Y et al. *J Biomol NMR* **1998,** *12*, 145-159.
31. Lam, P. C-H., Carlier, P. R. *J Org Chem* **2005,** *70,* 1530-1538.
32. Gallego, J. *Nucleic Acid Res* **2004,** *32,* 3607-3614.
33. Kawahara, K., and Tanford, C. *J. Biol. Chem* **1966,** *241,* 3228-3232.
34. Kumar, R., Bhuyan, A. K. *Biochemistry* **2005,** *44,* 3024-3033.
35. Oliveberg, M., Tan, Y-J., Fersht, A. R. *Proc Natl Acad Sci USA* **1995,** *92,* 8926-8929.
36. Benson, S. W., Dobis, O. *J Phys Chem A* **1998,** *102,* 5175-5181.
37. Krasnoperov, L. N., Peng, J., Marshall, P. *J Phys Chem A* **2006,** *110,* 3110-3120.
38. Truhlar, D. G., Kohen, A. *Proc Natl Acad Sci USA* **2001,** *98,* 848-851.
39. Wallace, M. I., Ying, L., Balasubramanian, S., Klenerman, D. *Proc Natl Acad Sci USA* **2001,** *98,* 5584-5589.
40. Chan, B-l., Baase, W. A., Schellman, J. A. *Biochemistry* **1989,** *28,* 691-699.
41. Bryngelson, J. D., Wolynes P. G. *J Phys Chem.* **1989,** *93*:6902-69151995

42. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., Wolynes, P. G. *Proteins* **1995,** *21,* 167-195.
43. Scalley, M. L., Baker, D. *Proc Natl Acad Sci USA* **1997,** *94,*10636-10640.
44. Baldwin, R. L. *Proc Natl Acad Sci USA* **1986,** *83,* 8069-8072.
45. Dill, K. A. *Biochemistry* **1990,** *29,* 7133-7155.
46. Steinbach PJ, Ansari A, Berendzen J et al.*Biochemistry***1991,***30,*3988-4001
47. Wagner, G., Wüthrich, K. *Nature* **1978,** *275,* 247-248.
48. Englander SW, Mayne L. *Annu. Rev Biophys Biomol Struct.* **1992,** *21*:243-265
49. Denisov, V. P., Peters, J., Hörlein, H. D., Halle, B. *Nat. Struct. Biol.* **1996,** *3,* 505-509.
50. Roder. H., Elöve G. A., Englander S. W. *Nature* **1988,** *335*, 700-704
51. Bushnell, G. W., Louie, G. V., Brayer, G. D. *J Mol Biol* **1990,***214,* 585-595
52. Elöve, G. A., Bhuyan, A. K, Roder, H. *Biochemistry* **1994,** *33,* 6925-6935.
53. Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J. et al *Biochemistry* **1993,** *32,* 10271-19276.
54. Maity, H., Rumbley, J. N., Englander, S. W. *Proteins* **2006,** *63,* 349-355.
55. Bhuyan, A. K. *Biochemistry* **2002,** 41:13386-13394.
56. Kumar, R., Prabhu, N. P., Yadaiah, M., Bhuyan, A. K. *Biophys J* **2004,** *87,* 2656-2662.
57. Dunbar, J., Yennawar, H. P., Banerjee, S., Luo, J., et al. *Protein Sci* **1997,** *6,* 1272-1733.
58. Makhatadze, G. I., Privalov, P. L. *J Mol Biol* **1992,** *226,* 491-505.
59. Pike, A. C. W., Acharya, R. *Protein Sci.* **1994,** *3,* 706-710.

60. Zarrine-Afsar, A., Mittermaier, A., Kay, L. E., Davidson, A. R. *Protein Sci* **2006,** *15,*162-170.

61. Gall, C. M., Cross, T. A., DiVerdi, J. A., Opella, S. J. *Proc Natl Acad Sci USA* **1982,** *7,* 101-105.

62. Khare R, Paulaitis M. E. *Macromolecules* **1995,** *28,* 4495-4504.

CHAPTER 5

Dependence of ^{15}N NMR Backbone Dynamic Parameters on Secondary Structure and Amino acid Type

5.1 Abstract

Delineating the determinants of backbone motions of small proteins is essential to decode the connections amongst structure, motions, and function. Analyses carried out using literature data for backbone ^{15}N NMR relaxation of 20 proteins that differ widely in their sequence, structure, and function show that the segments with defined secondary structural elements are generally stiff relative to less structured or unstructured regions. A fairly strong positive correlation between steady-state $\{^1H\}$-^{15}N NOE (Nuclear Overhauser Enhancement) and S^2 is observed for all regions of proteins, suggesting that values for the latter should be predictable from measured NOEs. Both size and polarity of amino acid correlate very weakly with backbone dynamics parameters, indicating that amino acid side chains do not influence backbone motions.

5.2 Introduction

That NMR could provide the best picture of protein motions was recognized as early as 1971 when Allerhand's laboratory showed the utility of ^{13}C and ^{15}N nuclear relaxation parameters to study dynamic processes in macromolecules (1, 2). NMR dynamics studies of these nuclei is facilitated by the fact that they relax principally by virtue of dipolar interactions with the directly bonded protons; the chemical shift anisotropy (CSA) relaxation mechanism making only marginal contribution (3). Indeed, ^{15}N and ^{13}C relaxation

measurements by one-dimensional NMR at natural abundance were reported for a large number of biopolymers and peptides in the following years. Obviously, limited resolution provided by 1D NMR and lower sensitivity of these heteronuclei stood as real problems in the endeavor of protein NMR dynamics studies. The advent of 2D pulse schemes specifically designed for indirect measurement of relaxations of insensitive nuclei (4-7) made it possible to overcome the problems that earlier workers struggled with.

The first comprehensive NMR backbone dynamics study was reported from Ad Bax's laboratory in 1989. The authors employed ^{15}N inverse detected heteronuclear spectroscopy to study the dynamics of Staphylococcal nuclease (8). Since then the backbone dynamics of a large set of proteins, a partial list of which is found in Table 1, has been worked out. Experimentally, recombinant proteins are isotopically labeled with ^{15}N, and longitudinal relaxation ($R1$) and transverse relaxation ($R2$) rate constants and the nuclear Overhauser enhancement (NOE) of the amide ^{15}N spin relaxed by dipolar coupling to the bonded proton is measured by proton detected 2D $\{^1H\}$-^{15}N NMR spectroscopy. The measurements are performed by using inversion recovery for $R1$, Carr-Purcell-Meiboom-Gill (CPMG) sequence for $R2$, and steady-state $\{^1H\}$-^{15}N NOE by using appropriate pulse sequences (9, for example). Data analyses and parameter calculations involve the following general considerations.

Intensities for the amide ^{15}N-^{1}H cross peaks are fitted to the experimental time series employed for $R1$ and $R2$ by using the following equation

$$I(t) = A + Be^{-R_{1,2}t} \quad (1)$$

where t refers to the inversion recovery delays for $R1$ measurement, and CPMG delays for $R2$ measurement. $A+B$ is the intensity at time $t = 0$, and A is the steady state value which is the intensity at $t = \infty$. The $\{^1H\}$-^{15}N heteronuclear NOE is calculated from the equation

$$NOE = \frac{I_{sat}}{I_{eq}} \tag{2}$$

where I_{sat} and I_{eq} are the intensities of a peak in the spectra collected with and without proton saturation, respectively.

The movement of the NH bond axis is characterized by the spectral density function J(ω), which is related to three parameters that describe the $R1$, $R2$, and NOE enhancement for the ^{15}N spin (10)

$$R1 = \frac{1}{4}d^2\{J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\} + c^2 J(\omega_N) \tag{3}$$

$$R2 = \frac{1}{8}d^2\{4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\} + \frac{c^2}{6}\{4J(0) + 3J(\omega_N)\} + R_{ex} \tag{4}$$

$$NOE = \frac{d^2}{4R_1}\frac{\gamma_H}{\gamma_N}\{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\} + 1 \tag{5}$$

where

$$d = \frac{\mu_0}{4\pi}\gamma_H\gamma_N\frac{h}{2\pi}(r_{NH}^{-3}) \tag{6}$$

$$c = \omega_N(\sigma_{\xi\xi} - \sigma_\perp)/\sqrt{3} \tag{7}$$

where μ_0 is the permeability of the free space, γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N (2.6752×10^8 and -2.712×10^7 rad s^{-1} T^{-1}, respectively); ω_H and

ω_N are the Larmor frequencies of ^1H and ^{15}N respectively, r_{NH} is the N-H bond length (taken here to be 1.02 Å), and $J(\omega_i)$ are the spectral densities at the angular frequencies ω_i. Generally, an axially symmetric chemical shift tensor is assumed for ^{15}N with $\sigma_{||}$ - $\sigma_{\perp}$ = -160 ppm (11). R_{ex} is included in equation 4 to accommodate chemical exchange and other pseudo-first-order processes that contribute to the decay of transverse magnetization (12). The R_{ex} term in equation 4 represents line broadening due to chemical exchange and/or conformational averaging on a time scale slower than the overall rotational correlation time, τ_m.

The amplitudes and effective correlation times of the internal motions of protein are determined from the relaxation data by using the model-free formalism pioneered by Lipari & Szabo (13, 14) and extended by Clore et al. (15, 16). In this analysis, the spectral density function, $J(\omega)$, is modeled differently depending upon whether the rotational diffusion tensor is isotropic or anisotropic. In the former case, when the internal motions of the NH bond occur on two fast but significantly different time scales so that they are characterized by two effective correlation times, τ_f and τ_S, with $\tau_f << \tau_S << \tau_m$ (15)

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S_f^2)\tau_f'}{1+(\omega\tau_f')^2} + \frac{(S_f^2-S^2)\tau_s'}{1+(\omega\tau_s')^2}\right] \quad (8)$$

in which,

$$\frac{1}{\tau_f'} = \frac{1}{\tau_f} + \frac{1}{\tau_m} \quad (9)$$

$$\frac{1}{\tau_s'} = \frac{1}{\tau_s} + \frac{1}{\tau_m} \tag{10}$$

$S^2 = S^2_f S^2_s$ is the square of the generalized order parameter characterizing the amplitude of internal motions of each NH bond, and S^2_f and S^2_s are the squares of the order parameters for the internal motions on the fast and slower time scales, respectively. The model-free spectral density function in equation 8 assumes that the overall tumbling motion of the molecule is isotropic. The order parameter specifies the degree of spatial restriction of the NH bond; S^2=1 for completely restricted motion and S^2=0 for completely free motion. S^2 can also have a value of zero when the NH bond vector is static and points along the magic angle with respect to the principal diffusion axis.

In the case of axially symmetric tensor, the spectral density function is approximated for the situations where the internal motions are much faster than overall tumbling rate as (17)

$$J(\omega) = \frac{2}{5}\left[S^2 \sum_{k=1}^{3} \frac{A_k \tau_k}{1+(\omega\tau_k)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2} \right] \tag{11}$$

where, $A_1 = (1.5\cos^2\alpha - 0.5)^2$, $A_2 = 3\sin^2\alpha\cos^2\alpha$ and $A_3 = 0.75\sin^4\alpha$. α is the angle between the NH bond vector and the unique axis of the principal frame of the diffusion tensor, $\tau_1 = (6D_\perp)^{-1}$, $\tau_2 = (D_\parallel + 5D_\perp)^{-1}$, $\tau_3 = (4D_\parallel + 2D_\perp)^{-1}$, and $\tau^{-1} = 6D + \tau_e^{-1}$. D is the isotropic diffusion constant, $D_\parallel$ and $D_\perp$are the components of the diffusional tensor parallel and perpendicular to the principal axis of the axial symmetry, respectively. The isotropic correlation time, τ_m, is related to D by the relation $\tau_m = (6D)^{-1}$.

The availability of backbone dynamics data for a sizable set of proteins provides an opportunity to search for general patterns and structural determinants of protein dynamics. The strategy of database search and correlation amongst NMR parameters, including chemical shifts and scalar couplings, has been widely adopted to delineate the determinants of protein phenomena such as structure fold (18) backbone angle restraints (19), and sequence homology (20). To determine the common patterns of backbone motions and interdependence of dynamic parameters a database of 20 of those proteins whose backbone dynamics have been published was searched (Table 1, Figure 1, and ref 21-38). This chapter briefly presents the analyses and observations.

5.3 Materials and Methods

Protein ^{15}N dynamics parameters and their uncertainties for 20 different proteins were collected from corresponding articles and supplied supporting information along with conditions and pulse sequences they have applied. Backbone dihedral angles were obtained from PDB (www.rcsb.org) using corresponding pdb ids. The dynamic parameters for all 20 amino acids were separated, and listed according to their location in different structural regions like α-helix, β-strand, H-bonded turns, and unstructured parts of proteins. Each amino acid in different structural parts is averaged to 99% level of confidence. Simple graphical analyses were performed using SigmaPlot.

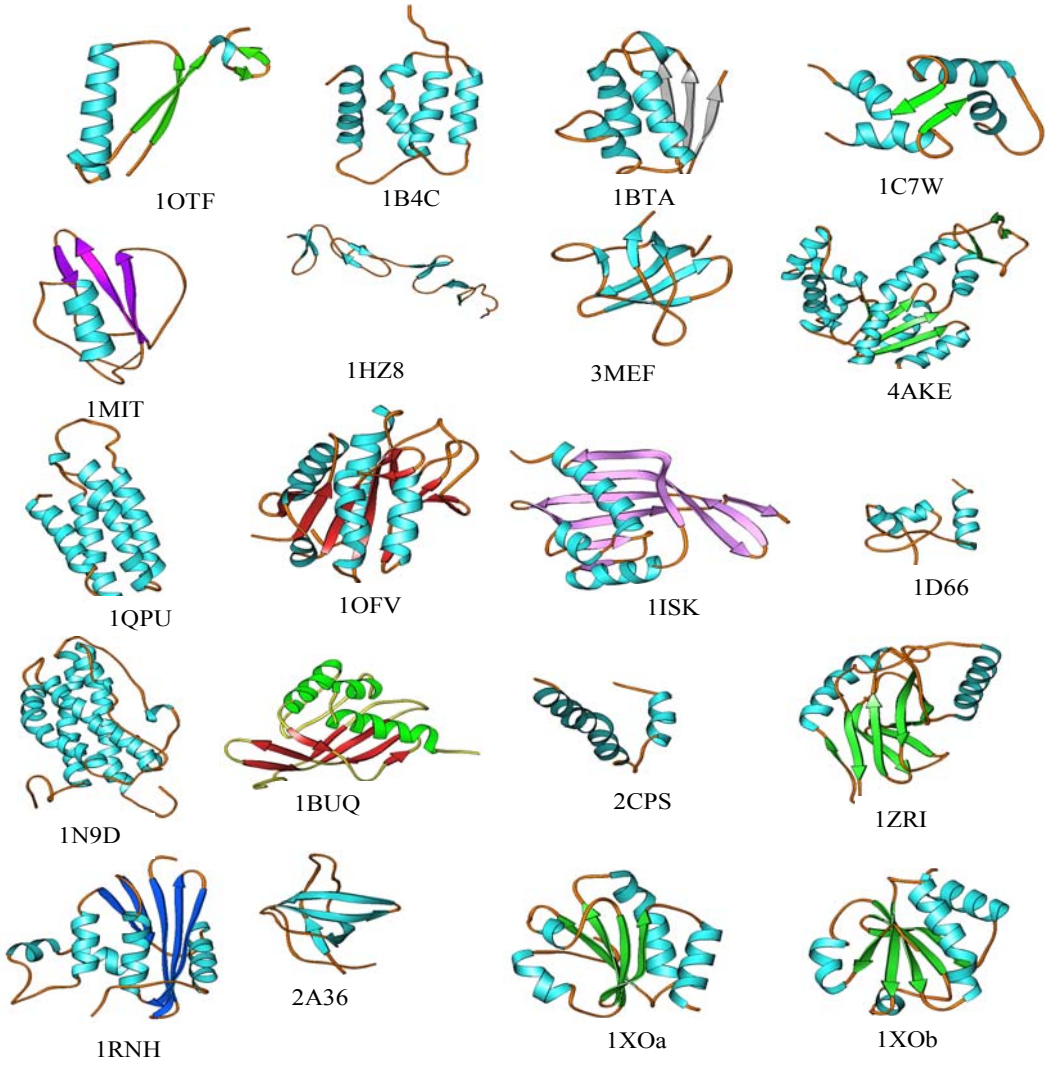


Figure 1. Structural representation of the proteins used for analysis in this work. The protein names corresponding to the labels, the pdb IDs, and some minor details are given in Table 1.

Table 1. Minor details of proteins used in this database

Protein	Pdb	Number of residues	Type of Structure	Experimental Method
1. Oxalocrotonate Tautomerase[21]	1OTF	62	α & β	Crystal
2. Apo S100B[22]	1B4C	92	α	NMR
3. Barstar[23]	1BTA	89	α & β	NMR
4. Calcium Vector Protein[24]	1C7W	81	α & β	NMR
5. Curcurbita Maxima Trypsin Inhibitor V[25]	1MIT	69	α & β	NMR
6. Human low density lipo protein receptor[26]	1HZ8	82	α & β	NMR
7. Major cold shock protein[27]	3MEF	69	α & β	NMR
8. Adenylate kinase[28]	4AKE	214	α & β	Crystal
9. Cytochrome b_{562} [29]	1QPU	106	α & β	NMR
10. Flavodoxin[30]	1OFV	169	α & β	Crystal
11. 3-Oxo-Delta 5 Steroid isomerase[31]	ISK	125	α & β	NMR
12. GAL4[32]	1D66	66	α	NMR
13. Human prolactin[33]	1N9D	199	α	NMR
14. δ-5-3- Ketosteroid isomerase Complex[31]	1BUQ	125	α & β	NMR
15. Major coat protein of filamentous Bacteriophage M13[34]	2CPS	50	α	NMR
16. Rat Olfactory maker protein[35]	1ZRI	163	α & β	NMR
17. Ribonuclease H[36]	1RNH	155	α & β	Crystal
18. N-terminal SH3 domain of DRK[37]	2A36	59	β	NMR
19. Thioredoxin oxidized[38]	1XOa	108	α & β	NMR
20. Thioredoxin reduced[38]	1XOb	108	α & β	NMR

5.4 Results and Discussion

Twenty proteins (Figure 1) whose NMR or X-ray structures as well as backbone dynamics have been worked out are considered. The proteins used in the database have 50-200 amino acid residues, and all are folded and conformationally ordered. Table 1 provides their minor details.

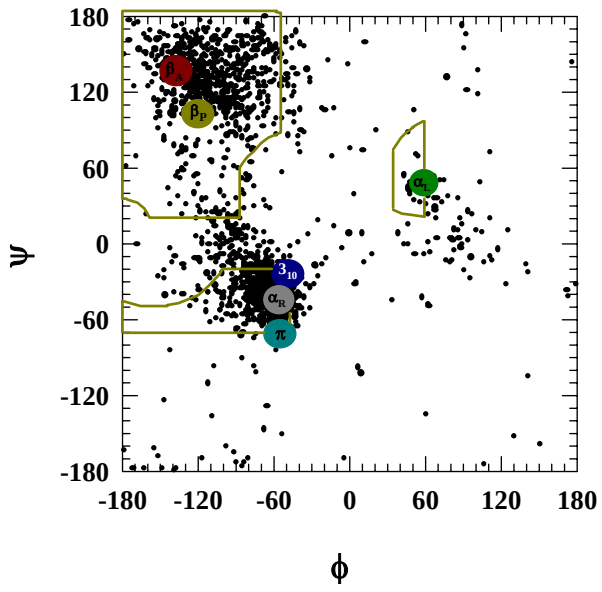


Figure 2. The backbone torsion angles obtained from the dataset of 20 proteins, showing that the dihedral angles may lie outside the allowed regions. The outliers need not always be glycines and prolines.

5.4.1 Distortions of secondary structures in proteins

To correlate the internal dynamic parameters with backbone torsion angles, the $\varphi\psi$ angles corresponding to the individual NH vectors were noted from NMR and crystal structures. This also provides an opportunity to examine how concentrated the dihedral angles are in the allowed regions of the $\varphi\psi$ map (39). Figure 2 simply shows a plot of φ *vs* ψ for all pairs of dihedral angles available in the database. Clearly, a considerable set of $\varphi\psi$ pairs does not lie in the strictly

canonical regions corresponding to the secondary structure types. Some $\varphi\psi$ pairs are even outliers with respect to the allowed regions, indicating a general distortion of structures in proteins, in addition to the fact that structures different from either α-helix and β-sheet exist.

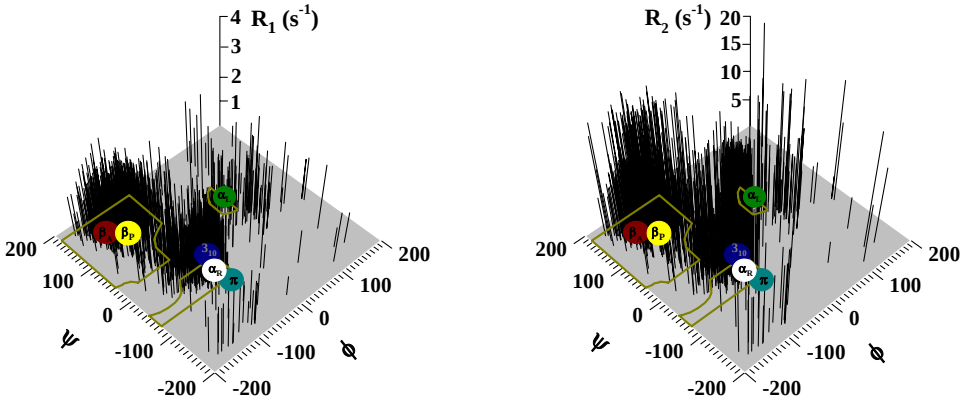


Figure 3. R_1 and R_2 relaxation rate constants projected from the $\phi\psi$ space to show that structured regions generally have higher values of both R_1 and R_2.

5.4.2 Relaxation parameters and dihedral angles

The database provides a total of 1908, 1906, and 1832 values for R1, R2, and NOE, respectively. The R1 and R2 values cannot be used directly to sense the motional properties of protein chains; the reason being the sensitivity of these two relaxation parameters to different motional frequencies. The ^{15}N R1 values across individual protein sequences generally appear uniform (for example, ref. 32, 40). This is often not the case with R2 though, since slow conformational exchange motions and other pseudo first-order processes (faster than the CPMG repetition

rate) contribute to the decay of transverse magnetization. Thus, it cannot apriori be assumed that residues in the segments of secondary structures and turns would have larger relaxation rate constants. However, as Figure 3 indicates, the survey using the database of 20 proteins indicates that the backbone nitrogens of those segments that are relatively more structured as defined by the backbone torsion angles show higher values for both *R*1 and *R*2.

The interpretation of NOE data is also not straightforward, because the long relaxation time (small R1 value) for water provides sufficient duration for chemical exchanges between the water and backbone amide protons producing NOEs, and thus effectively decreasing the measured $\{^1H\}$-^{15}N NOEs. They can still be used for qualitatively sensing fast internal motions. The measured NOEs for all 20 proteins as a function of torsion angles for the individual amino acid residues (Figure 4) shows some intense negative NOEs mostly concentrated in the α-helical region of the $\varphi\psi$ space. Negative NOEs in the regions of β-structures are relatively fewer, apparently suggesting that rapid internal motions are present in the structured regions also, and are relatively more frequent within helical segments. The data also indicates that the NOEs in the less structured segments of proteins are generally lower.

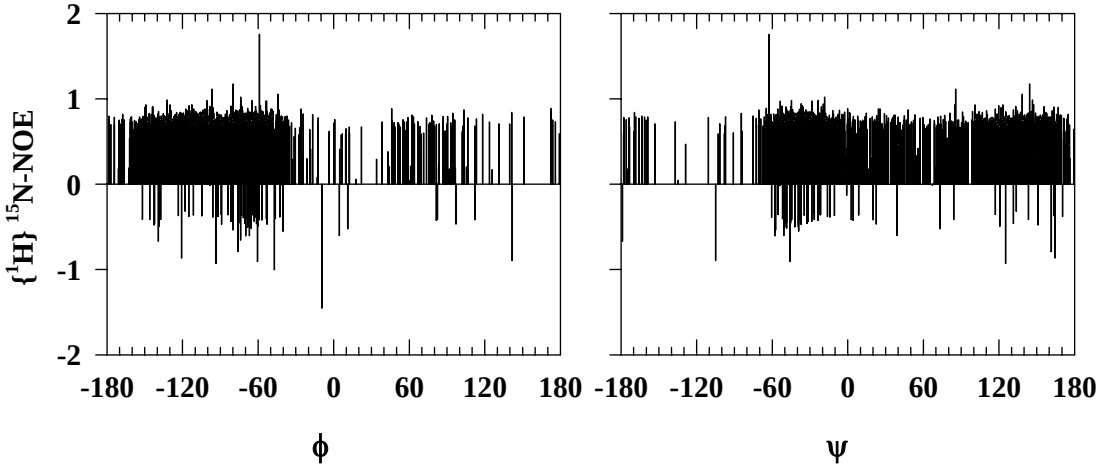


Figure 4. NOEs in the less structured regions are lower. However, negative NOEs are also observed in segments of secondary structure, suggesting that rapid internal motions may be present in these regions too.

5.4.3 Order parameter and dihedral angles

S^2 characterizes the amplitude of internal motions of the NH bond specifying the degree of its spatial restriction. Studies with individual proteins have indicated a general independence of S^2 on secondary structures (for example, 41, 42). The database of the 20 proteins now does indicate a discernible dependence of S^2 on dihedral angles (Figure 5). NH bonds in the α-helical segments on an average are more rigid than in the β-sheets. In less structured regions, the motions are not as constrained (Figure 5a, b).

The dependences of dynamic parameters on the secondary structure type have been examined by projecting the parameter values in the backbone torsion angle space. Attempt was also made to distinguish the secondary structure type quantitatively.

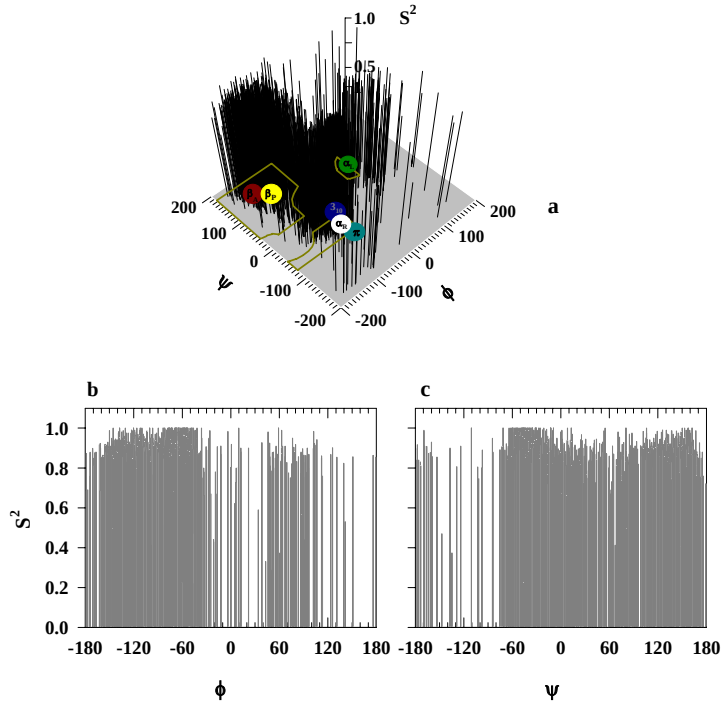


Figure 5. (a) Values of S^2 projected from the $\phi\psi$ space. Two-dimensional plots of ϕ *vs* S^2 and ψ *vs* S^2 are also shown for clearer comparison (b, c).

For example, the frequencies of occurrence of different values of all four dynamic parameters were determined separately for α-helical and β-sheet regions of all 20 proteins (data not shown). However, more rigorous statistical analyses will be needed in order to establish the dependences quantitatively.

5.4.4 S^2 *vs* NOE

Both NOE and S^2 describe internal motions. While the $\{^1H\}$-^{15}N heteronuclear NOE is obtained directly from experiments as a relaxation parameter ($NOE=I_{sat}/I_{eq}$, where I_{sat} and I_{eq} are the intensities of a peak with and without proton saturation, respectively), the latter is determined from the

relaxation data by using the model-free formalism of Lipari and Szabo (13-16, 42). A positive correlation of these two parameters should therefore provide a strong basis for using the model free expressions for calculating the spectral density function $J(\omega)$. Figure 7 shows the plot of $\{^1H\}$-^{15}N NOE *vs* S^2 for 1359 data points (correlation coefficient r=0.42). A close scrutiny of the data showed nearly 100 outliers near 0.3 NOE (gray colored in Figure 7), the exclusion of which improved the correlation coefficient to 0.71. Thus, it should be possible to use experimental values of steady-state NOE to extract the order parameter, and hence the semiangle for angular rotation of the NH vectors, $S^2=[0.5\ \cos\theta(1+\cos\theta)]^2$.

5.4.5 Backbone dynamic parameters do not provide a basis for classification of amino acids

In the next stage of analyses, the averaged values of dynamic parameters for all amino acids in different structural regions of proteins were examined. The rationale was to find out if individual amino acids have any dominating role in the resident parts of protein structures. The results shown in Figure 6 do not provide any striking contrast amongst the amino acid types in terms of their dynamic behavior, suggesting that the size of the side chain does not influence the dipolar relaxation of ^{15}N. Additionally, the following indications are noted.

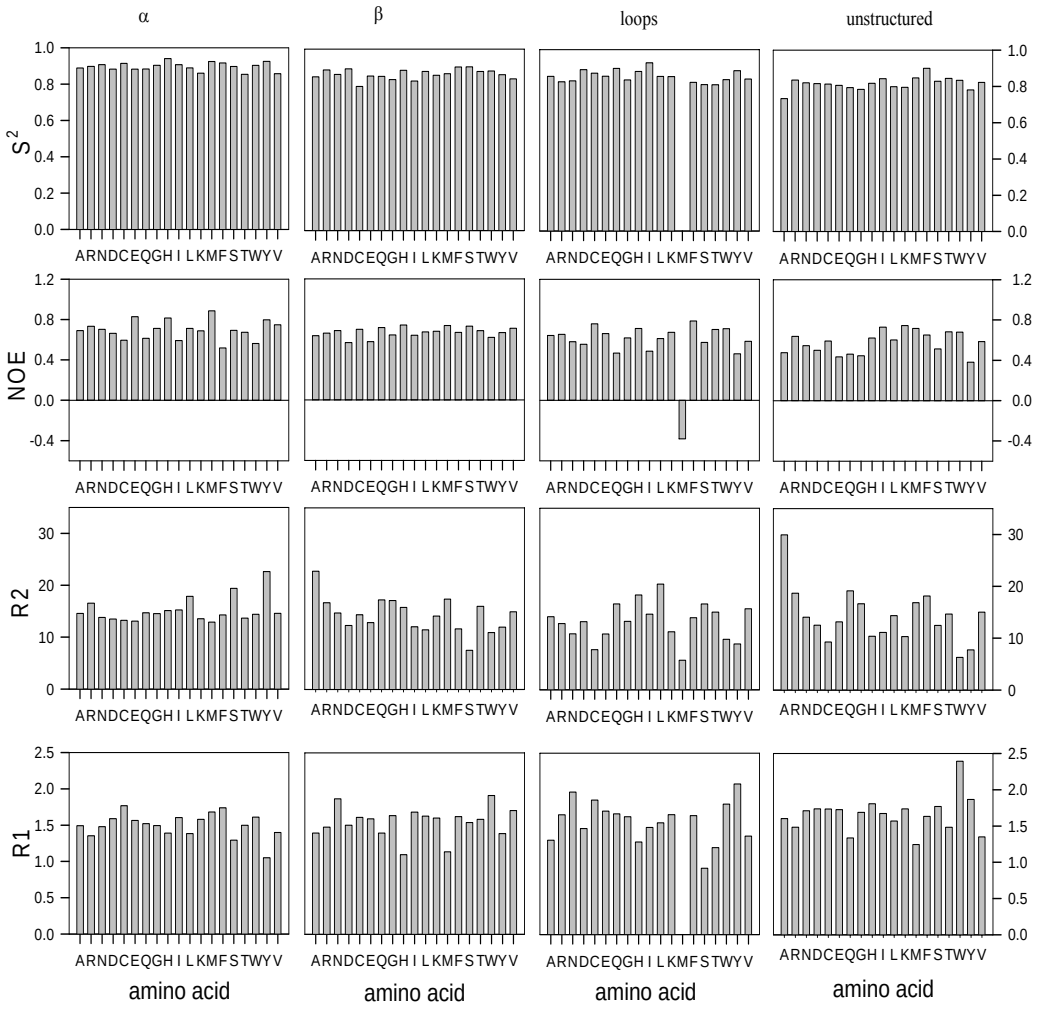


Figure 6. Averaged backbone dynamics parameters for all amino acids in different structural parts of proteins. Individual amino acids have do not have dominating roles.

1. In general, the S^2 values for all amino acids display the order: α-helix > β-sheet > loop region > unstructured or less structured segments.
2. Steady-state $\{^1H\}$-^{15}N NOEs for the amino acids in α-helix and β-sheet regions are similar, but tend to be lower in other regions. Of all, Gln, Tyr, and Val show smaller NOEs in the loop and less structured regions. Averaged NOE value for Met in loop is found to be negative, the significance of which is hard to tell even if the bar is not an outlier. Otherwise, none of the residues register negative NOE value in apparent contrast with the results shown in Figure 4. The averaging process has yielded positive values for all amino acids, suggesting that rather fewer number of a residue type exhibit the mobility of the magnitude observed in Figure 4. It is concluded that the type of residue and the structure type for its residence do not determine the magnitude and sign of NOE.
3. R2 values are fairly uniform for all types of amino acids in the α-helical segments, and the uniformity is less pronounced for the other parts. In β-sheets, loops, and less structured regions, Trp and Tyr exhibit consistently low values of R2.
4. R1 values are also nearly uniform for all types of residues in the α-helical region.

These results do not allow a classification of amino acid types on the basis of dynamic parameters. The observation that all amino acids consistently display higher S^2 value in the α-helical region merely reflects the fact that the rotational motion of amide vectors is generally more constrained in these regions, consistent with the inference made earlier (see Figure 5).

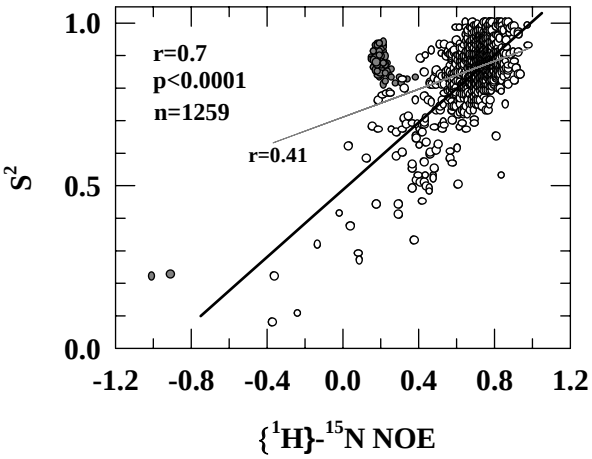


Figure 7. Correlation between S^2 and NOE for 20 proteins. The solid gray line is the best fit for the entire dataset, and the solid black line is obtained after exclusion of the apparent outliers.

5.4.6 Correlations amongst backbone dynamic parameters

Of the five dynamic parameters, R_1, $R2$, NOE, S^2, and τ_e, the first three are obtained directly from experiments, and the latter two are obtained in a model-independent way, although interpretable within the formalism of physically plausible models (13, 14). The dependence of the three relaxation parameters (R_1, $R2$, and NOE) on the two internal mobility parameters (S^2 and τ_e) is seen from the equations

$$R1 = S^2 R1_{(isot)} \left[1 + \frac{(10+\delta)}{(3+\delta)} \left(\frac{1-S^2}{S^2} \right) \left(\frac{\tau_e}{\tau_m} \right) (\omega_x \tau_m)^2 \right] \tag{12}$$

$$R2 = S^2 R2_{(isot)} \left[1 + \frac{(10+1.16\delta)}{(2+0.67\delta)} \left(\frac{1-S^2}{S^2} \right) \left(\frac{\tau_e}{\tau_m} \right) \right] \tag{13}$$

$$NOE = NOE_{(isot)} - \frac{50}{3+\delta} \left[\left(\frac{1-S^2}{S^2} \right) \left(\frac{\tau_e}{\tau_m} \right) (\omega_x \tau_m)^2 \right] \tag{14}$$

derived by Kay et al. (8) assuming that $[(\omega_A \pm \omega_x)\tau_e]^2 \ll 1$, $(\omega_x \tau_m)^2 \gg 1$, and

$[(\omega_A \pm \omega_x)\tau_m]^2 \gg (\omega_x \tau_m)^2$. Here, ω_A and ω_x are Larmor frequencies for ^1H and ^{15}N, respectively, τ_m is the correlation time for the overall motion of the protein, and $\delta=(c/d)^2$, where the expressions for c and d are as defined earlier in the context of equations 3-5. The subscript 'isot' (isotropic) refers to the value when there is no internal motion.

At megahertz frequencies, and with typical values of nano- and picoseconds for τ_m and τ_e, respectively, the second term for all three Equations makes little contribution to the ^{15}N relaxation parameters, and therefore τ_e need not be considered as an influencing factor. Hence, analyses were carried out to check for $R1$ *vs* S^2, $R1$ *vs* NOE, $R2$ *vs* S^2, $R2$ *vs* NOE, S^2 *vs* NOE, and $R1$ *vs* $R2$ correlations for all amino acid types irrespective of their presence in different structural regions of proteins. The correlation plots provided in Figure 8, where the coefficient '*r*' (outside the parentheses) in the ordinate label refers to correlation coefficient, suggest the following.

1. The $R1$ and S^2 values for none of the amino acids correlate significantly. Further, consideration of *r*-values for Trp and Tyr, and Asn and Gln, for example, shows that the bulk of the side chain is not a factor for the observed correlations.
2. The $r(R1\text{-NOE})$ values for Asn, Asp, Glu, Ile, Met, Phe, and Tyr are considerable (>0.5). It is possible that the side chain size influences these two relaxation parameters to a similar extent, but more work will be needed to substantiate the claim.
3. $R2$ and S^2 are poorly correlated for all amino acids.

4. Similarly, *R*2 and NOE are hardly correlated.
5. Better correlation of NOE and S^2 is observed for Ala, Gly, His, and Tyr.

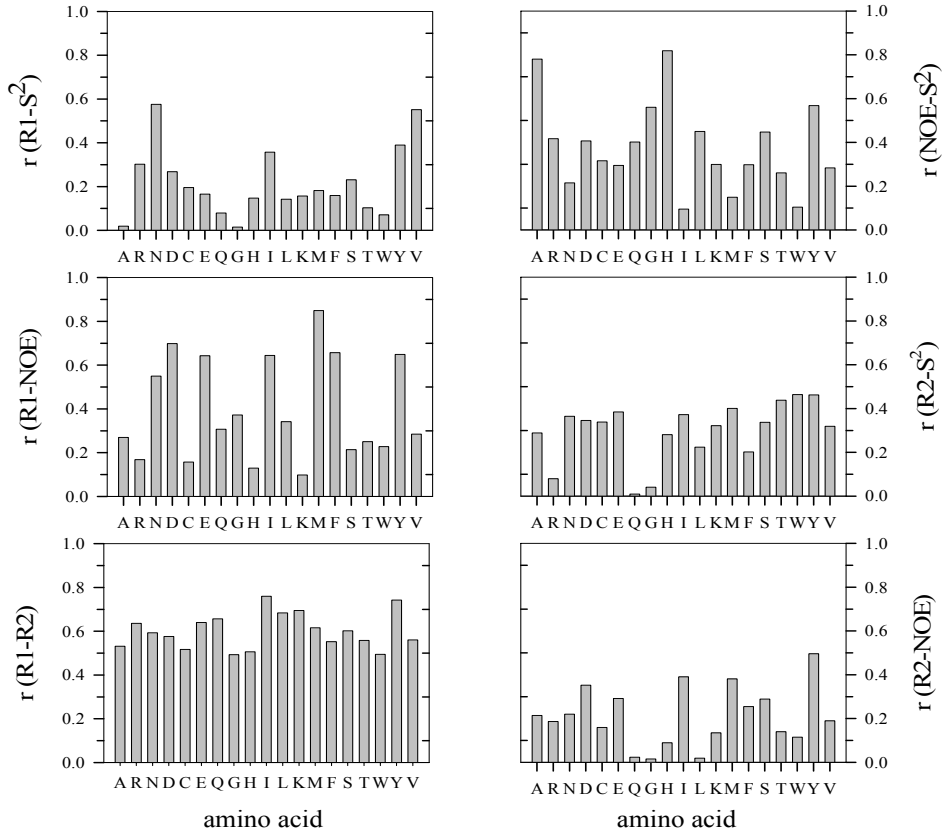


Figure 8. Linear correlations (*r* values) for (R1-R2), (R1-NOE), (R1-S^2), (R2-NOE), (R2-S^2), and (NOE-S^2) for individual amino acid types. The type and size of side chains are not determinants of any of the observed correlations between the parameters.

6. $r(R1\text{-}R2) \geq 0.5$ for all amino acids. The overall improved correlation in this case relative to the other correlations considered could arise from the fact that under extreme narrowing conditions for fast internal motions and in the absence of conformational exchanging motions, the correlation frequencies for internal motions affect $R1$ and $R2$ to the same extent.

The results in general do not indicate the influence of side chains on dynamic parameters. In particular, the type and size of side chains are not determinants of any of the observed correlations between the parameters.

5.5 Conclusions

The observed pattern of influence of backbone torsion angles on ^{15}N relaxation and ^{15}NH motions for 20 proteins suggest that protein parts consisting of secondary structural elements are generally stiff relative to less structured or unstructured regions. There is a considerable positive correlation between $\{^1H\}$-^{15}N NOE and S^2, suggesting that both are functions of rapid internal motions, and values for the latter should be predictable from measured NOEs. The size and polarity of amino acid side chains do not appear to influence backbone dynamics directly; rather their ability to support a particular structure type may dictate the terms of protein backbone dynamics.

5.6 References

1. Allerhand, A., Doddrell, D., Glushko, V., Cochran, D. W., Wenkert, E., Lawson, P. J., Gurd, F. R. N. *J. Am. Chem. Soc.* **1971,** *93*, 544-546.
2. Glushko, V., Lawson, P. J., Gurd, F. R. N. *J. Biol. Chem.* **1972,** *247*, 3176-3185.
3. Allerhand, A., Doddrell, D., Komoroski, R. *J. Chem. Phys.* **1971,** *55*, 189-198.
4. Kay, L.E., Marion, D. & Bax, A. *J. Magn. Reson.* **1989,** *84,* 72−84.
5. Skelnar, V.; Torchia, D.; Bax, A. *J. Magn. Reson.* **1987,** *73*, 375-379.
6. Nirmala., N. R.; Wagner, G. *J. Am. Chem. Soc.* **1988,** *110*, 7557-7558.
7. Nirmala., N. R.; Wagner, G. *J. Magn. Reson.* **1989,** *82*, 659-661.
8. Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972-8979.
9. Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G. S. E.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L.E. *Biochemistry* **1994**, *33*, 5984-6003.
10. Abragam, A. *Principles of Nuclear Magnetism, Clarendon Press, Oxford*, **1961**.
11. Hiyama, Y.; Niu, C.-H.; Silverton, J. V.; Bavoso, A.; Torchia, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 2378-2383.
12. Bloom, M., Reeves, L. W., Wells, E. J. *J. Chem. Phys.* **1965**, *42*, 1615-1624.
13. Lipari, G., Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.
14. Lipari, G., Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559-4570.
15. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989-4991.

16. Clore, G. M., Driscoll, P. C., Wingfield, P. T., Gronenborn, A. M. *Biochemistry* **1990**, *29*, 7387-7401.
17. Tjandra, N., Feller, S. E., Pastor, R. W., Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562-12566.
18. Bonvin, A. M., Houben, K., Guenneugues, M., Kaptein, R., Boelens, R. *J. Biomol. NMR.* **2001**, *21*, 221-233.
19. Cornilescu, G., Delaglio, F., Bax, A. *J. Biomol. NMR* **1999**, *13*, 289-302.
20. Shen, Y., Bax, A. *J. Biomol. NMR* **2007**, *38*, 289-302.
21. Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Whitman, C. P. *Biochemistry*, **1996,** *35,* 16036-16047.
22. Inman, K. G., Baldisseri, D. M., Miller, K. E., Weber, D. J. *Biochemistry,* **2001,** *40,* 3439-3448.
23. Sahu. S. C., Bhuyan, A.K., Udgoankar, J. B., Hosur, R. V. *J. Biomol. NMR* **2000,** *18,* 107-118.
24. Théret, I., Cox, J. A., Misplter. J., Craescu, C. T. *Protein science*. **2001,** *10,* 1393-1402.
25. Liu, J., Prakash, O., Cai, M., Gong, Y., Huang, Y., Wen, L., Wen, J. J., Huang, J.-K., Krishnamoorthi, R. *Biochemistry,* **1996,** *35,* 1516-1524.
26. Kurnia, N. D., Aliabadizadeh, K., Brereton, I. M., Kroon, P. A., Smith, R. *J Mol Biol* **2001,** *311*, 341
27. Feng, W., Tejero, R., Zimmerman, D. E., Inouye, M., Montelione, G. T. *Biochem*istry, **1998,** *37,* 10881-10896.
28. Shapiro, Y. E., Sinev, M. A., Sineva, E. V., Tugarinov, V., Meirovitch, E. *Biochem*istry, **2000,** *39,* 6634

29. Assfalg, M., Banci, L., Bertini, I., Ciofi-Baffoni, S., Barker, P. D. *Biochem*istry. **2001,** *40,* 12761-12771.
30. Zhang, P., Dayie, K. T., Wagner, G. *J. Mol. Biol.* **1997,** *272,* 443-455.
31. Yun, S., Jang, D. S.; Kim, D.-H., Choi, K. Y., Lee, H. C. *Biochem*istry, **2001,** *40,* 3967-3973.
32. Lefevre, J.-F., Dayie, K. T., Peng, J. W., Wagner, G. *Biochem*istry, **1996,** *35,* 2674-2686.
33. Keeler, C., Dannies, P. S., Hodsdon, M. E., *J Mol Biol* **2003,** *328,* 1105-1121.
34. Papavoine, C. H. M., Remerowski, M. L., Horstink, L. M., Konings, R. N. H., Hilbers, C. W., van de Ven, F. J. M. *Biochem*istry, **1997**, *36,* 4015-4026.
35. Gitti, R. K., Wright, N. T., Margolis, J. W., Varney, K. M., Weber, D. J., Margolis, F. L. *Biochemistry.* **2005,** *44,* 9673-9679.
36. Mandel, A. M., Akke, M., Palmer, A. G., III *Biochem*istry, **1996,** *35,* 16009-16023.
37. Farrow, N. A., Zhang, O., Forman-Kay, J. D., Kay, L. E. *Biochem*istry, **1995,** *34,* 868-878.
38. Stone, M. J., Chandersekhar, K., Holmgren, A., Wright, P. E., Dyson, J. *Biochem*istry, **1993,** *32,* 426-435.
39. Ramachandran GN, Sasisekharan V. *Adv Protein Chem.* **1968,** *23,* 283–438.
40. Penkett, C. J., Redfield, C., Jones, J. A., Dodd, I., Hubbard, J., Smith, R. A. G., Smith,L. J., Dobson, C. M. *Biochemistry* **1998,** *37,* 17054-17067.
41. Kay, L. E., Torchia, D. A., and Bax, A. *Biochemistry,* **1989,** *28,* 8972-8979.
42. Sahu, S. C., Bhuyan, A. K., Majumdar, A., Udgaonkar, J. B. *Proteins: Structure, Function, and Genetics,* **2000,** *41,* 460-474.

List of Publications

1. Bhuyan, A. K., **Rao, D. K.,** and Prabhu, N. P.(2005) Protein folding in classical perspective: folding of horse cytochrome *c*, *Biochemistry 44,* 3034-3040.
2. **Rao, D. K.,** Kumar, R., Yadaiah, M., and Bhuyan, A. K. (2006) The alkali molten globule state of ferrocyochrome *c*: extraordinary stability, persistent structure, and constrained overall dynamics, *Biochemistry 45*, 3412-3420.
3. **Rao, D. K.,** Prabhu, N. P., and Bhuyan, A. K. (2006) Extensive misfolding in the refolding reaction of alkaline ferrocytochrome *c*, *Biochemistry 45*, 8393-8401.
4. Kumar, R., Prabhu, N. P., **Rao, D. K.,** and Bhuyan, A. K. (2006) The alkali molten globule state of horse ferricytochrome *c*: Observation of cold denaturation, *J. Mol. Biol. 364*, 483-495.
5. **Rao, D. K.** and Bhuyan, A. K. (2007) Complexity of aromatic ring-flip motions in proteins: Y97 ring dynamics in cytochrome *c* observed by cross-relaxation suppressed exchange NMR spectroscopy, *J. Biomol. NMR 39,* 187-196.
6. Vijayalakshmi, A., **Rao, D. K.,** Bhuyan, A. K., Rao, N. M. (2007) Ester hydrolysis in triton X-100 micelles in the presence of gaunidinium chloride and urea, *Langmuir* (submitted).
7. **Rao, D. K.,** Rao, M.T., Bhuyan, A. K., Venu, K., and Sastry, V. V. S. (2007) Denaturant induced Stability in Proteins: A PMRD Study on Lysozyme and BSA. (manuscript under prepartion).
8. **Rao, D. K.,** and Bhuyan, A. K. Correlation between 15N NMR backbone dynamics and protein structure. (manuscript under preparation).