

TOWARDS UNDERSTANDING
HYPERTHERMOSTABILITY OF RUBREDOXIN
FROM *PYROCOCCUS FURIOSUS*

Thesis by
Silvia Cavagnero

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1997

(Submitted June 21, 1996)

© 1997

Silvia Cavagnero

All Rights Reserved

To my family

ACKNOWLEDGMENTS

My years at Caltech would not have been so intense and positive without the contributions of many people who interacted with me and helped in my efforts.

First and foremost, I would like to thank my advisor, Sunney Chan, for his constant encouragement, patience and great faith in me, regardless of what the results of my experiments were. Thanks to him, I have learned and improved so much over the past years, and I truly benefited from his push for independence and demand for original thinking. I am also glad our relationship has matured over the years. Sunney has a great mind and a gentle spirit, as I was able to appreciate during our many conversations on scientific matters and otherwise.

This project would not have been possible without the collaboration of Michael Adams. I was fortunate to interact with him and I truly appreciated the scientific freedom he allowed, as well as his enthusiasm when good results came forward. Michael Day deserves my thanks for being the first person to introduce me to *Pyrococcus furiosus* rubredoxin and hyperthermophiles. Doug Rees has also contributed substantially to this project, which actually started, with his blessing, in part as a result of his personal interest in proteins from hyperthermophiles. I had a few productive conversations with Michael Chan, whom I thank for sharing his views and thoughts on aldehyde ferredoxin oxidoreductase (AOR) even before they went to press. I appreciated collaborating with Phoebe Dea on some microcalorimetry experiments on rubredoxin.

The members of my Ph.D. committee, Jack Roberts, Doug Rees, Jack Richards and Steve Mayo, have always been supportive, encouraging and interested in my progress.

I found it extremely easy and productive to interact with all the past and present members of the Chan group. In particular, Scott Ross was great with me in my early days. He introduced me to the lab, taught me several theoretical and experimental aspects of NMR, and patiently answered my questions. John Evans has also been very generous in sharing his scientific expertise. I learned a lot from discussions with Randy Larsen, Ted DiMagno, Reggie Waldeck and Brian Schultz. I am also indebted to the members of the protein folding subgroup, Ron Rock, Ruth Ann Bertsch, Gary Mines, Derek Debe and Brian Jenkins for many stimulating discussions. Jian Lin, Dennis Anderson, Sean Elliott, Siegfried Musser, Qizhi He, Dorothy Zolandz, Michael Stowell, Mei Zhu, John Yip, Hung-Kay Lee, Zhuyin Li, Guangyang Wang, Shao-Ching: it was great to have you around. Over the years, Hoa Nguyen and Gary Mines have been dear friends (I am sorry we never had enough time to talk!).

Tom Dunn's exceptional abilities have been invaluable in fixing equipment, finding water baths for me to break, and providing witty suggestions and technical assistance with the NMR spectrometer.

Brian Schultz has been an excellent reviewer and proofreader of this thesis. I also thank Reggie Waldeck for some pertinent discussions on the research presented in chapter 3. Conversations with Siddarth Dasgupta and Komala Krishnaswamy straightened me up on the issue of protein electrostatic free energy.

Finally, I would like to thank my parents and parents-in-law, especially my mom, for truly believing in me all the way. I will leave the last, most deeply felt, thank you for Sandro, who has been a fantastic and caring soulmate, and for my sweet Cecilia.

ABSTRACT

Hyperthermophiles are phylogenetically primitive organisms, whose ability to thrive at temperatures up to about 110°C is unmatched by other life forms. This study focuses on the small iron protein rubredoxin (RdPf) from the hyperthermophile *Pyrococcus furiosus*. Understanding the physical bases for the extraordinary *in vivo* and *in vitro* properties of this model protein is complicated by the fact that its structural characteristics are extremely similar to those of its mesophilic counterparts. The first part of our work has been concerned with the controversial role of ion pairs in the thermostabilization. RdPf undergoes minimal structural perturbations upon disruption of its salt bridges. We identified an ion pair depleted RdPf "soft" form, which is able to reversibly bind the hydrophobic probe anilinonaphthalene-8-sulfonate (ANS) with micromolar affinity and 1:1 stoichiometry, while maintaining all of its other structural properties. The location of the putative binding site has been identified by electrostatic surface potential calculations. The retained low-pH structural features, coupled with an observed decrease in apparent melting temperature by about 40°C upon ion pair disruption, suggest that pH dependent intramolecular forces appreciably modulate RdPf hyperthermostability, but there must be other noncovalent forces involved in causing it.

Our next investigations compared unfolding rates and activation parameters of the hyperthermostable RdPf with those of the mesophilic rubredoxin from *Clostridium pasteurianum* (RdCp), and of other typical mesophilic proteins. Arrhenius plots unequivocally show that, under all experimentally accessible temperatures, the

unfolding of RdPf is several orders of magnitude slower than that of all of the examined mesophilic proteins. Furthermore, a comparative study of the pH dependence of Arrhenius plots of RdPf and RdCp suggests a quantifiable role of ion pairs in slowing down the unfolding kinetics. This novel kinetic, as opposed to thermodynamic, role of ion pairs may be of key importance in determining folding/unfolding kinetic barriers in proteins. Another intriguing implication of our data is that, in the specific case of RdPf, projected unfolding rates at 100°C and pH 7 yield an expected lifetime of thousands of years, vs only submilliseconds to minutes for all examined mesophilic proteins. This suggests that the hyperthermostability of RdPf may be ascribed to its being kinetically trapped at the physiological temperature of *Pyrococcus furiosus*, rather than to traditional thermodynamic factors. However, our kinetic data should be complemented by thermodynamic studies, in order to put our hypothesis on more solid ground.

We also asked whether hyperthermostability has any bearing on the features of the energy landscapes of RdPf, which ultimately dictate its folding and unfolding mechanism. This was achieved by studying the temperature dependence of the unfolding kinetics of RdPf by a variety of complementary spectroscopic techniques. Our data show that at least three kinetic intermediates are present along the unfolding pathways. The transition state takes place late during unfolding, which suggests that it occurs early in folding. The implications of these results for the protein folding problem have been discussed.

TABLE OF CONTENTS

Acknowledgments	iv
Abstract	vii
List of Figures and Tables	xii
Abbreviations	xvii

CHAPTER 1.

Understanding the Novel Properties of Proteins from Hyperthermophiles: Basic Concepts	1
Abstract	2
Introduction	3
The World of Hyperthermophilic Bacteria	9
The Thermodynamic Hypothesis and Its Role in Explaining Protein Stability	22
Proteins from Hyperthermophiles and Their Mesophilic Counterparts: Comparisons Based on Structural Analysis	34
Concluding Remarks	53
References	54

CHAPTER 2.

Response of Rubredoxin from <i>Pyrococcus Furiosus</i> to Environmental Changes: Implications for the Origins of Hyperthermostability	59
Abstract	60
Introduction	60
Materials and Methods	63
Results	68
Discussion	92
Concluding Remarks	104

References	106
------------------	-----

CHAPTER 3.

Arrhenius Analysis of the Unfolding of Rubredoxins from *Pyrococcus Furiosus* and *Clostridium Pasteurianum*: A New Role for

Ion Pairs in Proteins	111
Abstract	112
Introduction	112
Materials and Methods	114
Results	116
Discussion	133
Concluding Remarks	142
Appendix - The Electrostatic Free Energy of a Protein	143
References	156

CHAPTER 4.

The Unfolding Mechanism of Rubredoxin from <i>Pyrococcus Furiosus</i>	159
Abstract	160
Introduction	160
Materials and Methods	163
Results	168
Discussion	185
Concluding Remarks	192
References	194

CHAPTER 5.

Unfolding Kinetics and the Protein Folding Problem	196
Abstract	197
Hypercomplex Potential Energy Surfaces and Folding Funnels	197
What Unfolding Experiments Reveal (or Do Not Reveal)	
about Folding	206
Protein Unfolding Viewed According to Unimolecular	

Reaction Rate Theory	210
Concluding Remarks	216
References	217

LIST OF FIGURES AND TABLES

FIGURES

FIGURE 1-1: Universal phylogenetic tree	14
FIGURE 1-2: Influence of incubation temperature on growth of <i>Pyrococcus furiosus</i>	19
FIGURE 1-3: Representative stability curves for proteins	26
FIGURE 1-4: Sequence alignment for the rubredoxins from anaerobic bacteria ...	40
FIGURE 1-5: Ribbon diagram of the structure of rubredoxin from <i>Pyrococcus furiosus</i>	43
FIGURE 1-6: Surface potential images of rubredoxins from <i>Pyrococcus</i> <i>furiosus</i> , <i>Clostridium pasteurianum</i> , and <i>Desulfovibrio desulfuricans</i>	48
FIGURE 1-7: Surface potential images of rubredoxin from <i>Pyrococcus</i> <i>furiosus</i> , <i>Desulfovibrio gigas</i> and <i>Desulfovibrio vulgaris</i>	50
FIGURE 2-1: Circular dichroism spectra of rubredoxin from <i>Pyrococcus</i> <i>furiosus</i> (RdPf) in the far-UV region. The plot shows a superposition of the spectra of RdPf at pH 2.0 and pH 7.2, in the presence and absence of 1M NaCl	69

FIGURE 2-2: Circular dichroism spectra of rubredoxin from <i>Pyrococcus furiosus</i> (RdPf) in the near-UV/visible region. The plot shows a superposition of the spectra of RdPf at pH 2.0 and pH 7.2, in the presence and absence of 1M NaCl	71
FIGURE 2-3: Absorption spectra of rubredoxin from <i>Pyrococcus furiosus</i> at room temperature in water at variable pH	73
FIGURE 2-4: Thermal denaturation of rubredoxin from <i>Pyrococcus furiosus</i> in H ₂ O at pH 2.0, as followed by absorption spectroscopy	76
FIGURE 2-5: Denaturation curve for thermal unfolding of rubredoxin from <i>Pyrococcus furiosus</i> in H ₂ O at pH 2.0, as followed by absorption spectroscopy	78
FIGURE 2-6: Fluorescence emission spectra of rubredoxin from <i>Pyrococcus furiosus</i> upon excitation at 280 nm, at variable pH and ionic strength	80
FIGURE 2-7: ANS fluorescence emission spectra in the absence and presence of rubredoxin from <i>Pyrococcus furiosus</i> , at variable pH and ionic strength	82
FIGURE 2-8: Saturation curve (a) of rubredoxin from <i>Pyrococcus furiosus</i> at pH 2.0. Klotz plot (b) of ANS binding to rubredoxin from <i>Pyrococcus furiosus</i> . at pH 2.0	85
FIGURE 2-9: Electrometric pH titration curve of rubredoxin	

from <i>Pyrococcus furiosus</i>	89
FIGURE 2-10: Surface potential images of rubredoxin from <i>Pyrococcus</i> <i>furiosus</i> at pH 7 and pH 2	93
FIGURE 2-11: Surface potential images of rubredoxin from <i>Clostridium</i> <i>pasteurianum</i> at pH 7. All the ionizable amino acids have been labeled	95
FIGURE 3-1: Temperature dependence of <i>Pyrococcus furiosus</i> rubredoxin unfolding kinetics as followed by UV-vis absorption spectroscopy (380 nm) at a) 62.4°C, b) 66.6°C, c) 71.0°C, d) 75.5°C, and e) 76.7°C, pH 2.0	117
FIGURE 3-2: Eyring plot for rubredoxin from <i>Pyrococcus</i> <i>furiosus</i> unfolding kinetics, as followed by UV-vis absorption spectroscopy at 380 nm	121
FIGURE 3-3: Arrhenius plots for the unfolding of rubredoxin from <i>Pyrococcus furiosus</i> and a few mesophilic proteins	123
FIGURE 3-4: Arrhenius plots for the unfolding of rubredoxin from <i>Pyrococcus furiosus</i> and rubredoxin from <i>Clostridium</i> <i>pasteurianum</i> at different pH values	126
FIGURE 3-5: pH dependence of the apparent melting temperature of rubredoxin from <i>Pyrococcus furiosus</i>	138
FIGURE 4-1: Unfolding kinetics of rubredoxin from <i>Pyrococcus furiosus</i>	

as followed by Trp fluorescence emission spectroscopy.

The experimental pH is 2.0. The excitation wavelength

is 280 nm. Temperatures are: a) 63.3°C, b) 68.3°C,

c) 72.3°C, d) 74.6°C, and e) 77.8°C..... 169

FIGURE 4-2: Eyring plot for the two *Pyrococcus furiosus*

rubredoxin unfolding kinetic phases (k_1 , faster kinetic phase;

k_2 , slower kinetic phase) as detected by fluorescence

experiments, at pH 2.0 176

FIGURE 4-3: Unfolding kinetics of rubredoxin from *Pyrococcus furiosus*

as followed by circular dichroism spectroscopy. The experimental

pH is 2.0. Temperatures are: a) 63.6°C, b) 66.3°C,

c) 70.1, d) 75.7°C, and e) 78.5°C 178

FIGURE 4-4: Comparative Arrhenius plots for unfolding of

rubredoxin from *Pyrococcus furiosus* , monitored

by different spectroscopic techniques 183

FIGURE 5-1: Folding energy landscape, better known as

“folding funnel” 202

FIGURE 5-2: Diagram of the energies for the RRKM theory 212

TABLES

TABLE 1-1: Outline of the main characteristics of high temperature

	environments	11
TABLE 1-2:	Exposed and buried surface areas for rubredoxins from different species	46
TABLE 3-1:	Estimated unfolding half-lives ($t_{1/2}$) of rubredoxin from <i>Pyrococcus furiosus</i> and some mesophilic proteins	129
TABLE 3-2:	Activation parameters for rubredoxin unfolding, according to both Arrhenius and transition-state theory	131
TABLE 4-1:	First order kinetic rate constants of rubredoxin from <i>Pyrococcus furiosus</i> unfolding	174
TABLE 4-2:	Activation parameters for the unfolding of rubredoxin from <i>Pyrococcus furiosus</i> , as followed by UV-vis absorption, fluorescence and circular dichroism spectroscopies	187

ABBREVIATIONS

ANS:	1-Anilinonaphthalene-8-sulfonate
Brom:	Stem bromelain
CD:	Circular dichroism
CI2:	Chymotrypsin inhibitor 2
GnCl:	Guanidine hydrochloride
Lyz:	Lysozyme
NMR:	Nuclear magnetic resonance
RdCbt	Rubredoxin from <i>Chlorobium thiosulfatophilum</i>
RdCp:	Rubredoxin from <i>Clostridium pasteurianum</i> .
RdCpf:	Rubredoxin from <i>Clostridium perfringens</i>
RdCts:	Rubredoxin from <i>Clostridium thermosaccharolyticum</i>
RdDd:	Rubredoxin from <i>Desulfovibrio desulfuricans</i>
RdDv:	Rubredoxin from <i>Desulfovibrio vulgaris</i>
RdDvH:	Rubredoxin from <i>Desulfovibrio vulgaris</i> , strain Hildenborough
RdDvm:	Rubredoxin from <i>Desulfovibrio vulgaris</i> , strain Miyazaki
RdMe:	Rubredoxin from <i>Megasphaera elsdenii</i>
RdPa:	Rubredoxin from <i>Peptococcus aerogenes</i>
RdPf:	Rubredoxin from <i>Pyrococcus furiosus</i> .
Trp Syn:	α -Subunit of Trp synthase
Tryp:	Trypsin

RRKM: Rice-Ramsperger-Kassel-Marcus (referred to the well known unimolecular reaction rate theory)

CHAPTER 1.

**Understanding the novel properties of proteins from
hyperthermophiles : basic concepts**

ABSTRACT

There are, in nature, many primitive life forms that live under very harsh conditions. These include pH extremes (i.e., $\text{pH} > 10$ or $\text{pH} < 2$), very high ionic strengths (i.e., salt concentration about 4 molar), temperatures up to 110°C , and pressures up to about 1000 atmospheres. These life forms are mainly present in salty sediments, underwater vents in the vicinity of volcanoes, and a variety of geothermal emissions, such as geysers and solfataric fields. Collectively known as extremophiles, these organisms produce enzymes and nucleic acids with extraordinary chemical and physical characteristics.

We have focused on the study of the origins of the unusual properties of the small iron protein rubredoxin (RdPf) from the hyperthermophile *Pyrococcus furiosus*. We are especially interested in identifying the determinants of the hyperthermostability of RdPf, and in investigating its folding pathways. As explained in the course of the chapter, RdPf, believed to be involved in electron transfer processes of *Pyrococcus furiosus*, is an ideal model system for biophysical studies.

This chapter is intended to provide basic concepts and preliminary results, necessary to understand the hyperthermophilic world and the main concepts related to protein stability. A brief introduction to the world of extremophiles will be followed by a more specific description of the realm of hyperthermophilic microorganisms. An outline of old and new ideas about protein stability will then be presented. Finally, the

available information on structural comparisons between RdPf and its mesophilic counterparts will be reviewed.

INTRODUCTION

Nature has perfected, over the years, its own way of enabling species that live in different environments to take the greatest advantage of what the surroundings can offer. In some cases, living organisms have built-in mechanisms to defend themselves from potential damaging effects caused by the environment. This has become especially important during the course of evolution, since it has allowed entire species to survive sudden climatic or, more generally, environmental changes. For complex organisms, it is usually straightforward to identify, at a macroscopic level, what features give them advantage in a habitat over another. At a microscopic level, however, there are specific, sometimes subtle, metabolic, biochemical, and structural changes that take effect, depending on environmental conditions. These are still in part unidentified and, in general, poorly understood.

A very interesting case is that of some primitive microorganisms, mostly belonging to the *Archaea* domain (Woese et al., 1978), whose living environment is particularly harsh, being characterized by extreme pH values, high pressures (up to 1000 atmospheres), elevated temperatures (up to 110°C), or high salt concentration (ionic strengths of about 4 molar and higher). These organisms, collectively termed extremophiles, live in places like geothermal and hydrothermal vents in the vicinity of

volcanoes, highly saline water reservoirs, and so on. The living conditions of extremophiles are in striking contrast with the distribution of life in the biosphere, as it is currently understood, which is strongly biased to moderate temperatures (between 20 and 40°C), neutral pH, and moderate ionic strength. Extremophiles are interesting because they are relatively simple life forms, generally amenable to *in vitro* studies. In some cases, their metabolic pathways are different from those of their mesophilic counterparts, but, even more interesting, their proteins and nucleic acids exhibit very unusual physico-chemical properties. The term mesophilic literally means "that likes to be in the middle". It usually refers to molecules which are stable in moderate conditions (i.e., pH about neutral, $20 < T < 50^{\circ}\text{C}$, etc.).

There are two possible strategies by which nature enables extremophiles to survive the aforementioned harsh conditions. These organisms either adopt defensive or adaptive strategies. According to the former, they build physical barriers to prevent the cytoplasm from becoming in contact with the undesired conditions. This is the case, for example, of some extremophiles that live at extremes of pH. They are able to maintain an intracellular environment that is close to neutral. As a consequence, the intracellular components do not have any unusual features in these organisms.

Adaptive strategies are much more common in extremophiles. They comprise all the chemical, biochemical and biophysical changes that take place in the microorganism in order to enable to cope with its unusual environment, rather than getting shielded from it. Life forms that live in high salt concentrations (normally ranging from 3.4 and 5.1 molar NaCl) are a typical example. They resist osmotic

stress by maintaining an intracellular concentration of KCl that is close to saturation. As a consequence, the enzymes from these extreme halophiles function only at very high salt concentration. It is believed that high salt induces some structural changes in the proteins, possibly partial denaturation, in regions important for catalytic activity. The extracellular halophilic protease recently purified by Dordick and coworkers (Ryu et al., 1994) has an irreversibly decreased catalytic activity when salt concentration in the medium is decreased from 4 to 1 molar. In this case, though, it has been found that the drop in activity is mainly due to a decreased affinity of the enzyme for its substrate, rather than to denaturation. A different class of life forms that adopt a similar strategy is composed of the cryophilic microorganisms, also termed psychrophiles. They live in cold waters in Antarctica and other cold places around the world. Their enzymes usually spontaneously denature just above room temperature. Related to these are the anti-freeze proteins, that, because of their particular way to hydrogen-bonding to water, prevent the internal fluids of fishes that live in freezing-cold waters from condensing. Moving to a different temperature range, hyperthermophilic organisms are another relevant example. These life forms live at temperatures that range between 80 and about 110°C. Their enzymes only function at temperatures close to the optimal growth temperature of the parent organism. More will be said about these life-forms in the next section.

Sometimes extremophiles adopt both defensive and adaptive strategies. This happens, for instance, in the case of the extreme pH microorganisms. We have already seen how these life forms defend themselves from the environment by being able to

maintain a neutral intracellular environment. At the same time, they also produce some extracellular enzymes, mainly hydrolases, amylases and proteases, that must be competent to function under the prevailing acid or alkaline environment that is present extracellularly. These enzymes typically have optimum activity at the pH and temperature of the organism's native environment.

We have seen that extremophiles often adopt adaptive strategies. These generally enable them to produce intracellular components, such as proteins and nucleic acids, that have extremely different chemical and physical properties from those of the corresponding mesophilic species. Preliminary studies in the field indicate that thermal adaptation of polypeptides is achieved by minimal amino acid replacements, with respect to the primary sequence of analogous mesophilic proteins. Structural homology between mesophilic proteins and their thermophilic counterparts is usually greater than 60%. Unlike the case of nucleic acids, it seems that no unusual building blocks (i.e., amino acids for proteins, nucleotides for nucleic acids) need to be employed. We are interested in studying macromolecules that display unusual physical properties. Proteins from the hyperthermophilic realm are extremely interesting from this point of view. Furthermore, quite a few of them have been purified and some have also been structurally characterized and, in some cases, overexpressed. This makes them suitable candidates for thermostability studies.

Aside from its intellectual value, understanding the origin of protein hyperthermostability is an issue of considerable practical interest, as well. It is likely that these studies will ultimately lead to the synthesis of any arbitrary enzymes, for

which an efficient expression system is already available, in a thermostabilized version. This is obviously of enormous potential utility in biotechnology, since it will enable desired processes involving thermostabilized enzymes to be carried out at temperatures where impurities or any other interfering species are completely denatured. Based on the same principles, useful medical applications can also be envisioned. Incidentally, a number of applications along these lines are already being pursued by the food-processing industry with wild type hyperthermostable sugar isomerases used for the production of syrup, and by biotechnology labs with hyperthermostable DNA polymerases, which are becoming very popular for PCR applications, since they are about ten times more efficient than mesophilic DNA polymerases.

The physical bases of thermal adaptation, i.e., the structural features that cause melting temperatures to be higher than in analogous mesophilic proteins, have not been elucidated yet. There are, however, useful starting points, since several primary sequences of hyperthermophilic proteins, and those of their mesophilic counterparts, are available in the literature. One can then start by analyzing homology maps for these molecules. In fortunate cases, i.e., if the crystal structures of both mesophilic and hyperthermophilic proteins are available, changes in primary sequence can be related to structural variations. As we will see in the last section of this chapter, some studies along these lines were already available in the literature at the time we started this project and others have been published since then. One point that emerges from this approach, though, is that there is no consensus on what the structural determinants of

hyperthermostability are. This is because the three-dimensional features of mesophilic and hyperthermophilic proteins are often nearly identical. In many cases, the backbones are virtually superimposable, and the overall properties are very similar. More will be said on this point later. What we would like to emphasize now is that a comparison of native structures may not provide all the answers to the problem, since either the unfolded form or the transition state may play an important role in the thermostability. In fact, the surprising results presented in chapter 3 offer evidence in support of this point. The third section of this chapter also reviews and challenges the main concepts related to a thermodynamic view of protein stability.

Our studies have focused on the small iron protein rubredoxin from the marine hyperthermophile *Pyrococcus furiosus* (RdPf). Research efforts have concentrated in two directions:

- a) Study of the causes of protein thermal adaptation on a molecular basis. We have tried to learn the structural/dynamic consequences that thermal adaptation has on protein from hyperthermophilic microorganisms. Studies performed in solution have analyzed both ground state and transition state properties of RdPf.
- b) Understanding the folding pathways of proteins from extremophiles. In particular, we wanted to learn if and how extreme environmental conditions affect folding mechanisms. We pursued some of these ideas by investigating the unfolding mechanism of RdPf at low pH.

The present work attempts to give a modest contribution to the above issues by providing some initial steps in this virtually unexplored area of study. It is hoped that our efforts will serve as a stimulus for further research in this field.

THE WORLD OF HYPERTHERMOPHILIC BACTERIA

Generalities. The exploration of extreme natural environments has recently led to the discovery of unusual microorganisms that thrive at elevated temperatures (Stetter, 1982) but, on the other hand, do not exhibit any growth, and do not appreciably reproduce, at temperatures compatible with most life forms on earth, i.e., between 5 and 55-60°C (Stetter et al., 1983). These life forms, collectively known as hyperthermophiles, are organisms which have an optimal growth temperature of 80°C or higher (Adams, 1993). Interestingly, recent studies confirmed that, although these organisms do not grow at temperatures lower than about 60°C, they can survive under these conditions for years (Stetter et al., 1990).

To date, the highest temperature which is known to be compatible with hyperthermophilic life is 113°C. The existence of life in the so called black smokers has been reported by John Baross and coworkers in 1983 (Baross & Deming, 1983). Black smokers are dark gaseous underwater hot emissions found along the East Pacific Rise. Their temperature is 250°C or higher. However, it is now believed that the life forms found in their vicinity are, in fact, contaminants that have merged into the black smoker environment from environments immediately peripheral to the

smoker sampling point (Jannasch & Mottl, 1985; Karl et al., 1984). Therefore, the existence of life at 250°C and beyond has not been confirmed to date. As a consequence, the current upper temperature limit for the existence of life on earth remains 113°C.

Hyperthermophilic microorganisms were first isolated in 1982 by Stetter (Stetter, 1982) from shallow marine volcanic vents off the coast of the Italian island Vulcano. Since then, about twenty different genera of organisms have been discovered, many of which grow near 100°C (Adams, 1990; Adams & Kelly, 1992; Kelly & Deming, 1988; Kristjansson & Stetter, 1991; Stetter, 1986; Stetter et al., 1990).

Habitat. These organisms have generally been isolated from volcanic habitats, which include mainly submarine hydrothermal areas and continental solfataras. The marine locations include anaerobic hot sediments, submarine fumaroles, hydrothermal vents and hot springs. Some of the marine locations are oceanic deep-sea vents. The terrestrial locations, within solfataras, include sulfur-rich boiling springs, mud holes and heated soils. Thermophilic organisms have also been isolated from terrestrial geyser emissions in the United States.

Table 1-1 summarizes some of the properties of terrestrial and submarine hot springs, showing that the main differences are the higher hydrostatic pressures and temperatures in deep-sea hydrothermal systems, and the diversity of geothermal hot springs in terrestrial environments. An important difference between terrestrial and marine environments is also the chemical composition of the environment. The

TABLE 1-1: Outline of the main characteristics of high temperature environments.

Modified after (Kelly et al., 1994).

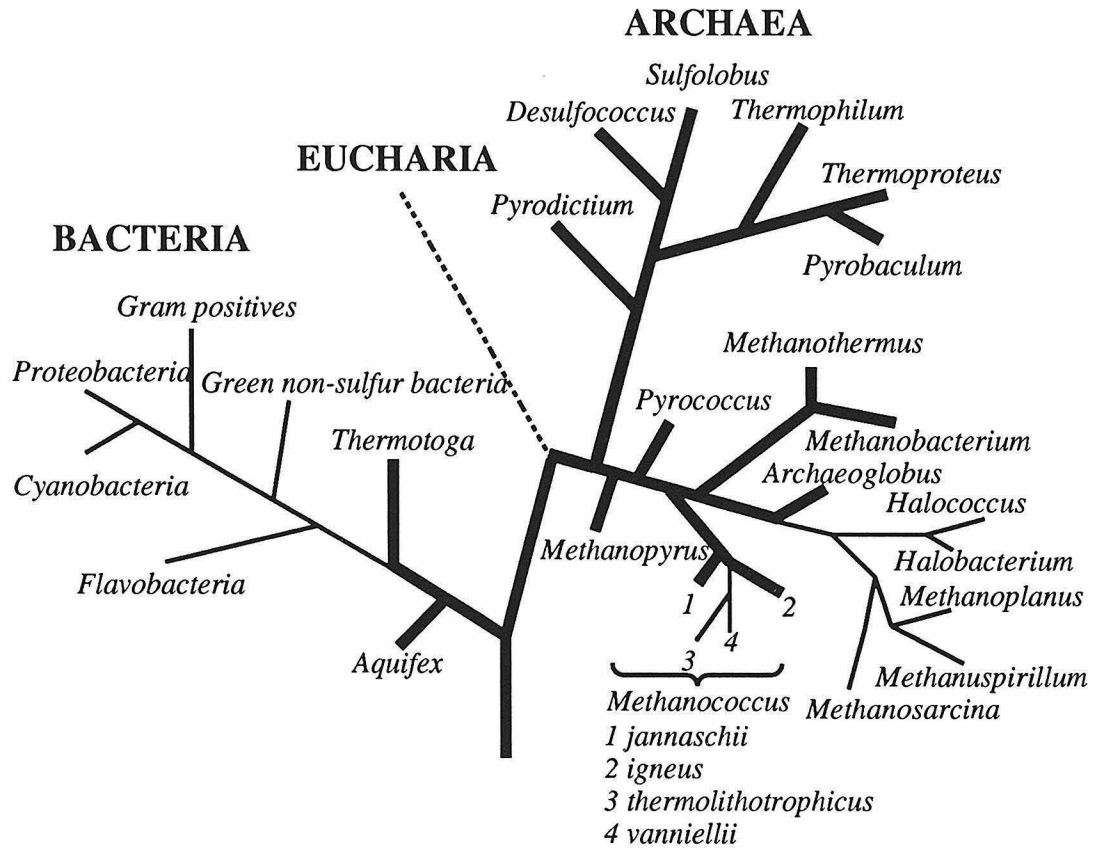
Characteristic	Marine	Terrestrial
Temperature range	A few degrees above ambient (2°C) to > 400°C, depending on mixing with sea water	Ambient to > 100°C for some fumaroles; usually <100°C at high elevations
Hydrostatic pressure	200 to 350 atm at depths of most ocean ridges; higher pressures for subsurface environments ^{a)}	Depends on depth of thermal environment
Habitats	Shallow and deep hydrothermal vents (most have high sulfide content)	Geothermal wells, thermal streams, seeps, mud holes, fumaroles, including solfataras and non-sulfur containing environments
pH	Between 3 and 8 (most hot fluids between 4 and 6)	0.5 to 9
Oxygen Contents	Generally anaerobic at temperatures exceeding 20-30°C because of high H ₂ S content	Variable
Water Chemistry	Generally high levels of H ₂ S, CH ₄ , H ₂ , Mn, Fe, Cu, W and variable concentrations of CO, NH ₄ ⁺ , chloride and organic compounds	Variable concentrations of CO ₂ , H ₂ , H ₂ S, CO, CH ₄ , NH ₄ ⁺ , SO ₄ ⁻ , metals and organic compounds
Locations	All sea and oceans spreading centers	Yellowstone, Mount St. Helens, Iceland, New Zealand, Vulcano (Italy)

a) A useful rule of thumb to estimate the expected underwater pressure is that the pressure increases by 1 atm every ten meters of underwater depth. For example, a 1 km deep underwater site is subject to a pressure of 101 atm.

concentrations of some important biological chemicals are much higher in hydrothermal vents. The concentration of H_2S , for example, ranges between 3.5 and 8.4 mM in hot hydrothermal fluids, whereas H_2S rarely exceeds 1 mM concentrations in terrestrial hot springs. Similarly, methane is much more concentrated in marine than in terrestrial springs. High levels of methane generally correlate with low levels of H_2 in both terrestrial and submarine hot springs. There are other important substances found in these environments. These are: nitrogen, phosphorus, organic compounds and transition metals such as iron, molybdenum, manganese and, notably, tungsten. While metal concentrations are generally high in all terrestrial and submarine hot springs, the concentration of nitrogen compounds varies considerably, depending on the concentration of organic molecules. Although there are marked differences between marine and terrestrial hot springs from which hyperthermophiles are isolated, the geophysical processes driving all volcanic systems are all fundamentally the same. However, presumably due to their different chemical composition, marine and terrestrial high temperature biotopes are characterized by distinct communities of hyperthermophiles.

Occasionally, hyperthermophiles can also be found in anthropogenic locations. These include mainly heated overflow waters from geothermal power plants (Huber et al., 1987; Stetter, 1985), acidic hot water drainages (Darland et al., 1970), and a uranium mine containing graphite and pyrite in Eastern Germany (A.H. Segerer & K. O. Stetter, unpublished observations).

FIGURE 1-1: Universal phylogenetic tree (Woese, 1987; Woese et al., 1990; Woese et al., 1978). Thermophilic and hyperthermophilic genera are indicated by a bold line.



Classification. As Fig. 1-1 shows, most thermophilic and hyperthermophilic genera belong to the domain of *Archaea* (Woese et al., 1978). The only exception known to date are the two genera *Aquifex* and *Thermotoga*, which belong to the bacterial domain (Woese, 1987; Woese et al., 1990). As Fig. 1-1 implies, the *Archaea* and *Eucharia* likely had a common ancestor, not shared by the *Bacteria*. This has the interesting implication that *Archaea* are evolutionarily closer to *Eucharia* (i.e. to complex living organisms) than *Bacteria* are. Fig. 1-1 also shows that the first organisms to have diverged from the *Eucharia/Archaea* branch were hyperthermophiles. Moreover, phylogenetic studies indicate that the most ancient of all microorganisms known to date are hyperthermophiles (Burggraph et al., 1992; Woese et al., 1990). This suggests that these are more closely related to the ancestor of all life than all known organisms. Life probably evolved when earth was much hotter than it is at present (Woese, 1987; Woese et al., 1990). In support of this idea, studies suggests that evolution has gradually produced organisms accustomed to lower and lower temperatures. The members of the *Archaea* domain are also believed to be among the most slowly evolving species.

Earlier in this section we have mentioned thermophiles. It is now a good time to define them, since they can be easily confused with hyperthermophiles. Thermophiles are organisms whose optimal growth temperature ranges between about 60 and 80°C. Hyperthermophiles, as already seen, have an optimal growth temperature of 80°C and beyond.

Biochemical Diversity. Most hyperthermophiles are strict anaerobes. The ability of these organisms to use O_2 is very limited, and the reduction of elemental sulfur, rather than O_2 , appears to be the predominant source of energy conservation by hyperthermophiles. Many of hyperthermophiles are also obligate heterotrophs, i.e., they utilize complex carbohydrates or proteins as their carbon source. A smaller number are autotrophs and thus fix carbon dioxide.

Most hyperthermophiles depend on elemental sulfur for growth and produce high levels of H_2S . This is why the environment where they live, or are artificially grown, is usually infested by unpleasant odors and corrosive vapors. This is also the reason why enzyme purification is not trivial. The production of large amounts of H_2S is also a problem because of corrosion: conventional stainless steel fermentors cannot be used and the required high salt concentration usually makes things worse. Despite these practical obstacles, recent years have seen an explosion of interest in the isolation and characterization of proteins from hyperthermophiles (Adams, 1993; Adams & Kelly, 1996; Flam, 1994).

The main classes of hyperthermostable proteins, mostly enzymes, that have been purified so far include redox proteins (Aono et al., 1989; Blake et al., 1991), proteases (Cowan et al., 1987; Eggen et al., 1990), hydrogenases (Bryant & Adams, 1989; Juszczak et al., 1991; Pihl & Maier, 1991), dehydrogenases (Robb et al., 1992), ferredoxin-dependent oxidoreductases (Mukund & Adams, 1990), DNA polymerases (Lundberg et al., 1991) and enzymes involved in carbohydrate metabolism (Constantino et al., 1990; Koch et al., 1990). Despite its length, the above list is far

from being complete; other classes of proteins, including transcription factors and membrane proteins, are currently being purified in different labs.

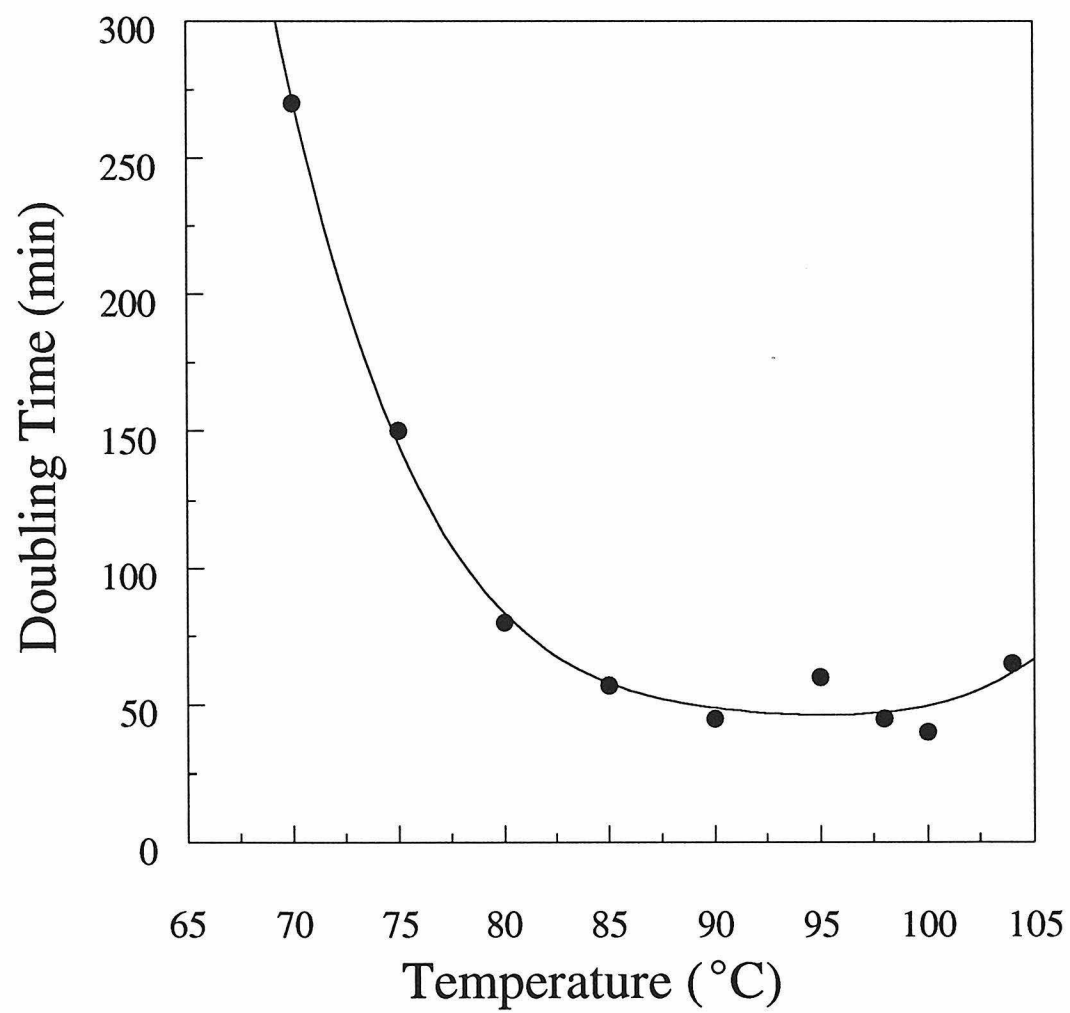
The metabolic pathways of hyperthermophilic organisms have many interesting peculiarities which, however, are not very relevant to the present work. Therefore, they will not be discussed here. The interested reader should consult the primary literature in the field (Adams, 1993).

Pyrococcus Furiosus. The hyperthermophilic organism *Pyrococcus furiosus* has been discovered by Fiala and Stetter in 1986 (Fiala & Stetter, 1986). *Pyrococcus* has been isolated from geothermally heated marine sediments at the beach of Porto di Levante, in the Vulcano island (Italy). It is a spherical organism (the average diameter is 1.5 μm) with polytrichous flagella. Its name literally means “furious fireball.” Surprisingly, the DNA G/C content of *Pyrococcus furiosus* is only about 50% (Fiala & Stetter, 1986).

Pyrococcus furiosus is a strictly anaerobic heterotroph, which grows on starch, maltose, peptone and complex organic substrates. Its metabolic products are H_2 and CO_2 . The former inhibits the growth at high concentrations. To prevent this from happening, elemental sulfur is usually added to the media, thus recruiting hydrogen gas while producing H_2S . Sulfur is therefore not strictly required for the growth of *Pyrococcus furiosus*, although it is usually added to the medium to increase the yields. The growth of this microorganism is severely temperature-dependent (Fig. 1-2). Its optimal growth temperature is 100°C . Optimal growth conditions also include high salt concentration (about 2.5 %) and pH close to neutral. *Pyrococcus* grows rapidly to

FIGURE 1-2: Influence of incubation temperature on growth of *Pyrococcus furiosus*.

Modified after (Fiala & Stetter, 1986).



relatively high cell densities. Its culture does not necessarily require stainless steel fermentors, since elemental sulfur is not strictly needed for the growth. This practical advantage makes *Pyrococcus* a precious and convenient source of interesting hyperthermostable proteins, including the one utilized in the present work, rubredoxin.

Structural Studies on Proteins from Hyperthermophiles. Given the relatively young age of the field, only a few proteins from thermophilic and hyperthermophilic microorganisms have been completely structurally characterized to date. On the other hand, the primary sequence of several of these proteins is known. Sequence analysis reveals that proteins from hyperthermophiles contain no unnatural or modified amino acids. This is quite remarkable, since it shows that the 20 amino acid code is sufficiently flexible to accommodate the thermostabilizing features that must be present in these proteins. On the other hand, it is interesting to notice that initial studies on hyperthermostable RNAs (Kowalak et al., 1994) have revealed that they contain a number of unnatural nucleotides, which are believed to be important for the thermostabilization. In the mesophilic world, it is generally accepted that nucleic acid thermostability is related to the G/C content. However, nature decided not to choose this strategy for the RNA from hyperthermophiles. We propose the following simple explanation for this phenomenon. A hypothetical hyperthermostabilized G/C enriched nucleic acid requires quite a high G/C percentage. This molecule would melt at very high temperatures but, at the same time, its free energy difference (between the hybridized and the annealed forms) would be too high for performing its normal biological functions. These, as known, require flexibility, which is, in turn, inherently

linked to marginal stability. Nature modulated the stability of nucleic acids in a much more subtle way than a brute force "free energy difference increment" approach. Future experimental studies in this area are needed to test these ideas.

The three-dimensional structures of proteins from hyperthermophiles that have been solved so far include rubredoxin from *Pyrococcus furiosus* (Day et al., 1992), aldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus* (Chan et al., 1995), glutamate dehydrogenase from *Pyrococcus furiosus* (Yip et al., 1995), indole-3-glycerolphosphate syntase from *Sulfolobus solfataricus* (Henning et al., 1995), and glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima* (Korndörfer et al., 1995). Several people have made comparisons between these structures and those of their mesophilic counterparts. The results will be presented in a later section of this chapter.

THE THERMODYNAMIC HYPOTHESIS AND ITS ROLE IN EXPLAINING PROTEIN STABILITY

Proteins, as we usually refer to them, are structured macromolecules composed of a polypeptide chain with well-defined spatial arrangement. Their specific three-dimensional features are what enables them to interact with other cell components and, in general, to have a well defined biological function. However, when subject to a harsher chemical environment than their physiological conditions (i.e., pH extremes, temperature extremes, presence of interfering chemicals commonly referred to as

denaturants), proteins undergo a pseudo-first order phase transition called unfolding or denaturation. Although poorly studied to date, the denatured state (Shortle, 1996) is, generally speaking, highly disordered, and it has a smaller amount of residual secondary and tertiary structure.

In principle, there are two potential reasons why a protein is successfully able to maintain its native state under physiological conditions: a) either the equilibrium between native and unfolded form is shifted in favor of the folded form (thermodynamic hypothesis), or b) the solvent bath does not provide sufficient energy to escape from the deep potential energy well in which the protein lies (kinetic hypothesis).

Carl Anfinsen conducted a series of brilliant experiments in 1973 (Anfinsen, 1973), where he showed that proteins fold spontaneously and reversibly into their functional conformation, which he called the native state. Since then, the validity of the thermodynamic hypothesis has widely been accepted. The fact that Anfinsen's experiments cannot be repeated on all proteins is usually only taken as an indication that most proteins unfold irreversibly due to aggregation or other irreversible processes that take place subsequent to the unfolding event. Therefore, the following treatment will assume thermodynamic equilibrium between native and unfolded species.

The free energy for protein unfolding (ΔG) is defined according to the known relation

$$\Delta G = \Delta H - T\Delta S , \quad (1-1)$$

where ΔH and ΔS represent the unfolding enthalpy difference and the unfolding entropy difference, respectively.

Proteins have the very peculiar property of increasing their heat capacity upon unfolding (Privalov, 1979). This property makes proteins special, and it is not shared by other important biomolecules. Lipids, for example undergo phase transitions that have a null ΔC_p . A positive ΔC_p is mainly a consequence of the fact that proteins possess a hydrophobic core, which is differentially solvated in the folded and unfolded form (Privalov & Gill, 1988). At constant pressure, the heat capacity difference between unfolded and native states (ΔC_p) is defined as

$$\Delta C_p = \frac{d\Delta H}{dT} , \quad (1-2)$$

where T is the absolute temperature. Equation 1-2, as all the following expressions, assume a two-state process between unfolded and native forms, unless otherwise stated. This is rarely the case in protein folding. However it is a useful working approximation to derive meaningful and simple expressions. Formula 1-2 shows that ΔC_p is positive in proteins because their ΔH increases with temperature. We can now rewrite the definition of ΔC_p , to emphasize its relation with the unfolding entropy difference (ΔS):

$$\Delta C_p = \frac{d\Delta S}{(dT)/T} = \frac{d\Delta S}{d(\ln T)} . \quad (1-3)$$

Therefore, a ΔC_p different from zero is manifested in a temperature dependent ΔH as well ΔS , since $\Delta H = T\Delta S$ at the transition temperature.

We can now consider a general definition of ΔH and ΔS as a function of temperature, where, for convenience, the difference is computed between the protein melting temperature T_m and a generic temperature T . We have that

$$\Delta H(T) = \Delta H_{T_m} + \Delta C_p (T - T_m) , \quad (1-4)$$

and

$$\Delta S(T) = \Delta S_{T_m} + \Delta C_p \ln(T/T_m) . \quad (1-5)$$

Expressions 1-4 and 1-5 are valid for a temperature independent ΔC_p . This is generally the case for a protein, but exceptions have been recently found, as in the case of bovine pancreas trypsin inhibitor (Makhatadze et al., 1993). Equations 1-1, 1-4 and 1-5 can now be combined to give the simple expression for the temperature dependence of ΔG

$$\Delta G(T) = \Delta H_{T_m} - T\Delta S_{T_m} + \Delta C_p [T - T_m - T \ln(T/T_m)] . \quad (1-6)$$

FIGURE 1-3. Representative stability curves for proteins.

a) (A): Stability curve for a typical mesophilic protein with $\Delta H_{Tm} = 100$ kcal/mol, $\Delta S_{Tm} = 294$ cal mol⁻¹ deg⁻¹ and $\Delta C_p = 1$ kcal mol⁻¹ deg⁻¹. The arrow indicates the temperature value (T^*) corresponding to the maximum thermodynamic stability of the protein, i.e., ΔG_{max} .

(B): Same stability curve as in (A) but with an added weak temperature dependence to ΔC_p according to the arbitrary function $\Delta C_p(T) = 228 \times \Delta C_p / T$; the asymmetry of this stability curve is noticeable.

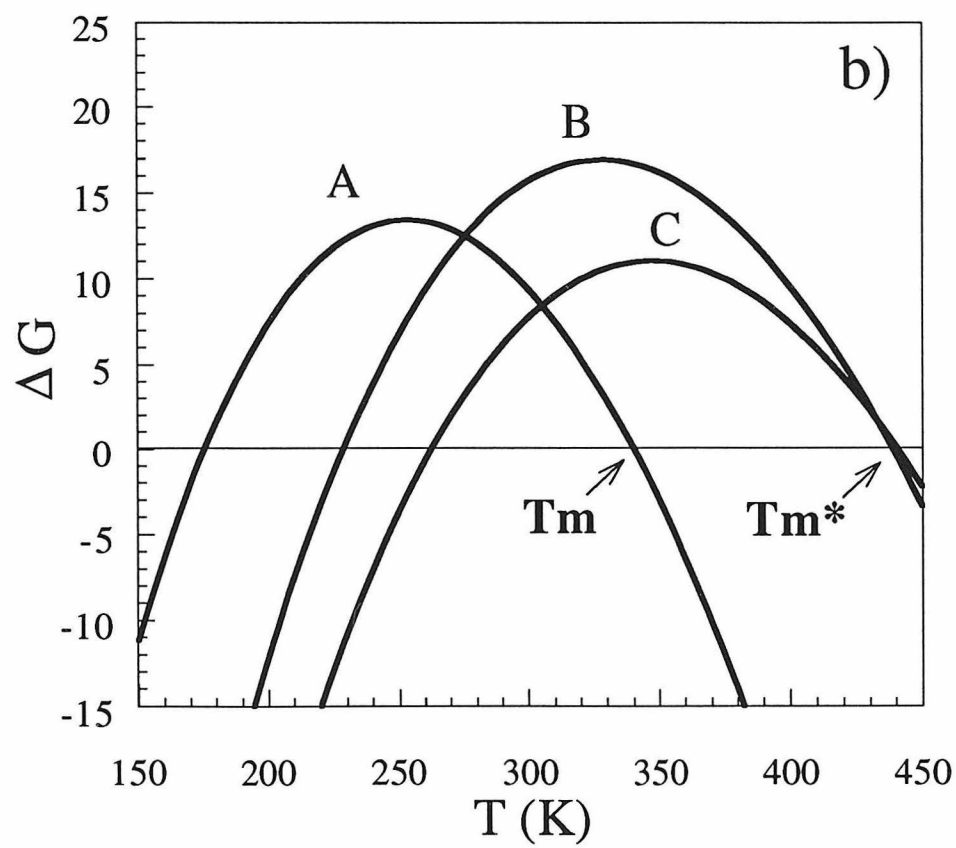
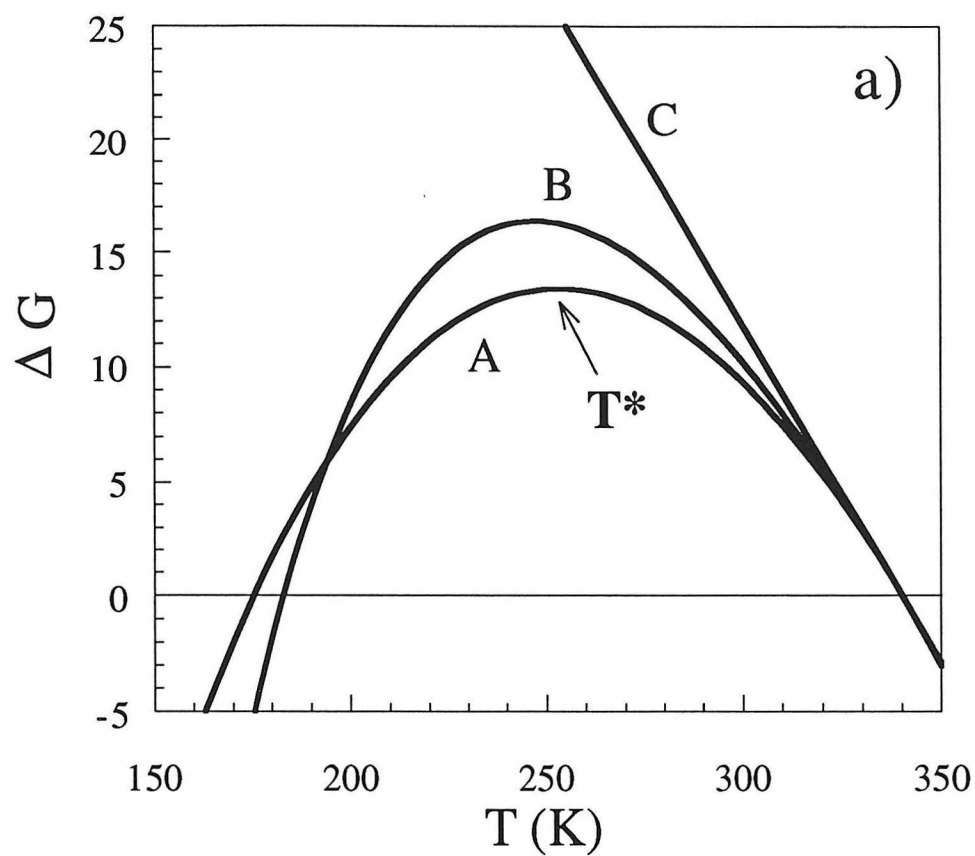
(C): Stability curve for an idealized protein with $\Delta H_{Tm} = 100$ kcal/mol, $\Delta S_{Tm} = 294$ cal mol⁻¹ deg⁻¹ and $\Delta C_p = 0$ kcal mol⁻¹ deg⁻¹.

b) (A) Stability curve for a typical mesophilic protein with $\Delta H_{Tm} = 100$ kcal/mol, $\Delta S_{Tm} = 294$ cal mol⁻¹ deg⁻¹ and $\Delta C_p = 1$ kcal mol⁻¹ deg⁻¹.

(B): Limiting case of the stability curve for a hyperthermophilic protein whose

thermostability is completely enthalpy driven. The enhanced melting temperature T_m^* has been constrained to be 440K, which gives $\Delta H_{T_m} = 129 \text{ kcal/mol}$, considering that $\Delta H_{T_m} = T_m \Delta S_{T_m}$. The other thermodynamic parameters are $\Delta S_{T_m} = 294 \text{ cal mol}^{-1} \text{ deg}^{-1}$ and $\Delta C_p = 1 \text{ kcal mol}^{-1} \text{ deg}^{-1}$.

(C): Limiting case of the stability curve for a hyperthermophilic protein whose thermostability is completely entropy driven. The enhanced melting temperature T_m^* has been constrained to be 440K, which gives $\Delta S_{T_m} = 227 \text{ kcal mol}^{-1} \text{ deg}^{-1}$, considering that $\Delta H_{T_m} = T_m \Delta S_{T_m}$. The other thermodynamic parameters are $\Delta H_{T_m} = 100 \text{ kcal mol}^{-1}$ and $\Delta C_p = 1 \text{ kcal mol}^{-1} \text{ deg}^{-1}$.



Equation 1-6 is, as implied by the above discussion, valid at constant pressure, for a two-state process and for a temperature independent ΔC_p . A plot of $\Delta G(T)$ vs T yields a bell-shaped curve, known as the **stability curve** of a protein (Becktel & Schellman, 1987; Privalov, 1979). A typical stability curve is represented in Fig. 1-3a, curve A. The thermodynamic parameters used to construct the curve are as follows: $\Delta H_{T_m} = 100$ kcal/mol, $\Delta S_{T_m} = 294$ cal mol⁻¹ deg⁻¹ and $\Delta C_p = 1$ kcal mol⁻¹ deg⁻¹.

Let's now take a quick look at the validity of the assumptions that we made in deriving eq. 1-6.

- (i) Constant pressure is usually observed both in *in vivo* and in *in vitro* situations for hyperthermostable proteins; so this is not a problem. It is well known that many hyperthermophiles, especially those found in deep hydrothermal vents, live at high pressures. However, we have neglected to consider pressure effects in our thermodynamic treatments because the pressures experienced by the hyperthermophiles are not sufficient to induce any phase transition (see Table 1-1). Incidentally, pressures required to unfold a mesophilic protein are on the order of 3000 atm; they are therefore much higher than the pressure experienced by typical hyperthermophiles in their living environments.
- (ii) The assumption that ΔC_p is temperature independent may well be unjustified, as already discussed (Makhatadze et al., 1993). To take this into account, we have estimated the behavior of the stability curve for a typical protein with a ΔC_p decreasing with temperature. The resulting curve is presented in Fig. 1-3a, curve B. The thermodynamic parameters used to construct the curve are the same as for

curve A, except that ΔC_p has been defined according to the arbitrary temperature dependent function $\Delta C_p(T) = 228 \times \Delta C_p / T$. It is apparent that temperature dependence of ΔC_p simply introduces asymmetry in the stability curve. Nonetheless, qualitative considerations still remain entirely acceptable.

(iii) Finally, the assumption that folding/unfolding is a two-state process is probably not accurate, since it is established that many known proteins have folding intermediates. However, referring to a simplified two-state situation does not affect the validity of our qualitative observations.

It is useful, in the first place, to compare curve A and C of Fig. 1-3a. The former has a ΔC_p of 1 kcal mol⁻¹ deg⁻¹, while the latter has ΔC_p equal to zero. In both cases, ΔC_p has been considered as temperature independent. It is apparent that a positive ΔC_p is responsible for introducing curvature in the stability curve of a protein. Accordingly, the curvature of a stability protein is equal to

$$\left(\frac{\partial^2 \Delta G}{\partial T^2} \right)_p = - \left(\frac{\partial \Delta S}{\partial T} \right)_p = - \frac{\Delta C_p}{T}, \quad (1-7)$$

where all the parameters have been defined before. The presence of cold denaturation, induced by this ΔC_p effect will not be discussed here.

Another feature of the stability curve is that its slope, at each temperature value, gives the stability entropy difference

$$\left(\frac{\partial \Delta G}{\partial T} \right)_p = -\Delta S. \quad (1-8)$$

Incidentally, it is to be noted that all the quantities we have talked about so far are to be considered **standard state values** (e.g., ΔH should be replaced by ΔH° , etc.). Ordinate values are thus ΔG° and not ΔG . For reasons not clear to us, nobody in the literature has pointed this out, and there is some potential for confusion. Nonetheless, to be consistent with the previous notation and with the literature in the field, we will omit the superscripts. We will leave the reader with a reminder that they should be there.

Another relevant feature of the curve is its maximum amplitude, i.e., the ΔG of maximum stability of the folded protein. This is easily obtained by equating the derivative of eq. 1-6, with respect to T , to zero. It is then necessary to solve for T . The value obtained is the temperature of maximum stability (T^*), which needs to be substituted back into eq. 1-6. The result is

$$\Delta G(T^*) = \frac{\Delta H_{T_m}^2}{2T_m \Delta C_p} . \quad (1-9)$$

It follows from the above discussion that the goal of **increasing T_m values**, that is, the job accomplished by nature with hyperthermostable proteins, can be accomplished by a number of routes:

- 1) by having a stability curve with shallow slopes (i.e., small ΔS values);
- 2) by making ΔH larger, i.e., increasing the area covered by the $\Delta G(T)$ curve;

- 3) by reducing the curvature of the stability curve (i.e. having smaller ΔC_p) and, at the same time, having larger absolute values for ΔH .

Notice that, according to the aforementioned criteria, only scenario 2 predicts that an increase of T_m is necessarily accompanied by an increased $\Delta G(T^*)$. On the contrary, in case 1, T_m will be larger at the expenses of a lower $\Delta G(T^*)$. This is not a problem, since all that is needed is to increase the melting temperature value of the protein, regardless of its maximum stability. This is even better illustrated by simply considering eq. 1-6 at the melting temperature ($T=T_m$) of a generic protein, whose stability curve is given by curve A of Fig. 1-3b. We have

$$\Delta H_{T_m} = T_m \Delta S_{T_m} . \quad (1-10)$$

Expression 1-10 simply indicates that T_m can be increased to a thermophilic value T_m^* by three simple means:

- a) increase ΔH_{T_m} (enthalpy driven thermostabilization);
- b) decrease ΔS_{T_m} (entropy driven thermostabilization);
- c) a combination of 1) and 2), such that $T_m^* > T_m$.

The limiting cases a and b are illustrated by curves B and C of Fig. 1-3b, respectively. The numerical values used to build the curves are those of a typical

situation, and they are detailed in the legend to Fig. 1-3b. As the figure shows, a limiting entropy driven process conceives that a hyperthermophilic protein may even be less stable than a mesophilic one at room temperature (i.e. 300K). Of course this is not the case for an enthalpy driven process.

One may speculate that the addition of a few extra non covalent intramolecular interactions (hydrogen bonds and ion pairs) to a mesophilic protein is an enthalpy driven process. If this were the case, sample calculations (based on the numerical values of Fig. 1-3) show that an increase in T_m by 100°C would require the gain of about 29 kcal/mol. This would, indeed, require several new interactions to be established. An entropy-driven process would need a decrease in entropy by about 67 entropic units. The above arguments are only valid to a first approximation. They are suitable, however, to introduce a word of caution, in that it may be extremely risky to draw quick conclusions upon simple examination and comparison of three-dimensional structures of mesophilic and thermophilic proteins.

A real life situation, unfortunately, cannot be easily described by the aforementioned limiting cases. One needs to remember that the stability of proteins is given by the difference of two large numbers, i.e., the free energy value of the denatured and that of the native states, respectively. Usually, the resulting ΔG is a remarkably small value, oscillating between 8 and 15 kcal/mol (see Fig. 1-3a, curve A). Additionally, the introduction of a new interaction in a protein introduces an enthalpic, as well as an entropic component. The two terms give rise to some enthalpy-entropy compensation, so that the resulting ΔG may turn out to be a

relatively small value. In conclusion, we believe that thermodynamic data on hyperthermophilic proteins should be collected in order to build a stability curve or get equivalent data. This is important in order to obtain a clearer picture of what the origin of the hyperthermostabilization is. To date, these studies have never been performed on a reversibly folding/unfolding system. Although we made several attempts, we did not succeed in collecting thermodynamic data, under reversible conditions, on rubredoxin from *Pyrococcus furiosus*, the reason being that this protein gives rise to an irreversible phase transition under all conditions that we have tested.

All of the above treatment has assumed the validity of the thermodynamic hypothesis of protein stability. This concept is currently well established, except for a few notable exceptions (see Discussion section of chapter 3) in the protein folding field. However, our kinetic data, presented in chapter three, challenge this idea by suggesting that, for these proteins, the thermodynamic hypothesis is not a requirement. We truly hope that future thermodynamic experiments on proteins from hyperthermophiles will integrate and corroborate our findings.

PROTEINS FROM HYPERTHERMOPHILES AND THEIR MESOPHILIC COUNTERPARTS: COMPARISONS BASED ON STRUCTURAL ANALYSIS

This section is intended to provide a brief outline of the most important published studies that address the origin of protein hyperthermostability. Since the studies by Perutz in 1975 on bacterial ferredoxins (Perutz & Raidt, 1975), several groups have been interested in investigating the origin of protein thermostability. Before hyperthermophiles were isolated and purified, proteins from moderate thermophiles were popular subjects of study. We will not discuss these studies here since we are more concerned with issues peculiar to hyperthermophiles in particular. The general discussion below will then be followed by a brief description of the background literature for rubredoxin from *Pyrococcus furiosus* (RdPf), the protein we have utilized for the present studies.

Proteins from hyperthermophiles. No high quality thermodynamic studies have been published, to date, on hyperthermophilic proteins. The main underlying reasons are that a) it is not trivial to identify the right protein and suitable experimental conditions under which thermal transitions are reversible; and b) hyperthermophilic proteins pose the additional challenge of requiring equipment that can be operated efficiently and reproducibly at high temperatures. In the present work, we have been able to overcome the latter problem by lowering the pH.

Most of the literature concentrates on the structural comparison between crystal structures of hyperthermophilic proteins and their mesophilic counterparts (Day et al., 1992; Goldman, 1995; Henning et al., 1995; Korndörfer et al., 1995; Perutz & Raidt, 1975; Rees & Adams, 1995; Yip et al., 1995). Additionally, some

interesting solution experiments have also been performed in the lab of Jaenicke and coworkers (Jaenicke, 1996).

A first amazing trend that emerges from these studies is that hyperthermostability is achieved by minimal structural variations between mesophilic and hyperthermophilic versions of the same protein. Both versions have virtually superimposable backbones, and the same can be said about their secondary and tertiary folds. Both sequence and structure of the hydrophobic cores are usually highly conserved. As we have seen, typical sequence homology among the two classes amounts to about 50-60%. Differences can essentially be identified in some details about the side chain interactions. One important thing to be noticed is that no generalizations can be safely made concerning this problem. Overall, and with the exclusion of RdPf, which will be described separately, essentially three classes of structural features can be identified. We will describe the first two classes below. The third class, which comprises proteins where different small effects have been invoked, will not be presented here, since it has been already effectively summarized in the review by Adrian Goldman (Goldman, 1995).

The only representative of the first class, so far, is aldehyde ferredoxin oxidoreductase (AOR) from *Pyrococcus furiosus*, whose crystal structure and discussion of the three-dimensional features have recently appeared in the literature (Chan et al., 1995; Rees & Adams, 1995). Perhaps not surprisingly, after what we said before, this protein has all the features of a typical mesophilic protein as far as amino acid composition, and secondary and tertiary structures, number of hydrogen bonds

and ion pairs. However, compared with a number of other enzymes of known structure, AOR has a reduced surface area and an increased fraction of buried atoms. This led the authors of the crystal structure paper (Chan et al., 1995) to formulate the interesting hypothesis that thermostability may be achieved by minimizing the ratio of surface area to volume. A simple analogy that has been suggested to make this more clear is that of droplets or soap bubbles (Rees & Adams, 1995). Incidentally, it is worth noticing that AOR has been a particularly difficult protein to study, since it does not have any direct mesophilic counterpart, unlike all the other thermophilic and hyperthermophilic proteins whose crystal and/or NMR structures have been solved.

The second class comprises glyceraldehyde 3-phosphate dehydrogenase from *Thermotoga maritima* (Korndörfer et al., 1995), glutamate dehydrogenase from *Pyrococcus furiosus* (Yip et al., 1995), indole-3-glycerolphosphate synthase from *Sulfolobus sulfataricus* (Henning et al., 1995) and a conspicuous number of other proteins from thermophilic species. The common feature that has been identified therein is an increased number of surface ion pairs. The average number of ion pairs per residue in mesophilic proteins is 0.04 (Barlow & Thornton, 1983). This number, for instance, is increased to 0.11 in glutamate dehydrogenase from *Pyrococcus furiosus*, which compares favorably with 0.06 found in its mesophilic counterpart from *Clostridium symbiosum*.

The problem of ion pairs in proteins has been studied extensively and it has always been somewhat controversial. The most extreme view is that proposed by Linus Pauling in the early days of the protein structure and folding field. He thought

that salt bridges are the glue that keeps proteins in their native compact state, i.e., the dominant force in protein folding. This concept is now considered obsolete and it has been replaced by a completely opposite view, based on a wealth of experimental data, that electrostatic interactions between exposed amino acids give little contribution to protein stability (Matthews, 1993). Theoretical studies explain that the favorable coulombic interaction, established by a new ion pair in water, is counterbalanced by the desolvation cost that the two solvent screened ions have to pay upon desolvation. Furthermore, the resulting favorable enthalpy, if any, resulting from this effect is effectively compensated by the entropy decrease which takes place upon spatially constraining two floppy amino acid side chains in the newly formed bond.

As justly noted by Goldman (Goldman, 1995), the current view about ion pairs is in net contrast with the structural comparison data presented above, which emphasize the importance of surface salt bridges in hyperthermophilic proteins. According to Goldman, the contradiction may be resolved by noting that what makes hyperthermophilic proteins so stable are not just ion pairs *per se*, but the fact that these salt bridges are arranged in a network on the protein surface. This then implies that the desolvation cost for introducing a new member of the network to a preexisting ion pair is approximately halved, since only one residue, as opposed to two, must be desolvated and immobilized. This point is extremely well taken. However, we feel that an important factor has been totally neglected by these studies. Ion pairs can, if strategically arranged on the protein surface, constrain it to the point that a kinetic barrier to unfolding is generated. This may, in turn, provide a “kinetic” explanation to

hyperthermostability. The interested reader is referred to chapter 3 of this work for experimental evidence (on rubredoxin from *Pyrococcus furiosus*) in support of this possibility. Finally, it is also interesting to notice that the kinetic hypothesis has the potential to provide a more general justification of hyperthermostability. The reason for this is that only a few ion pairs placed in key-positions can be sufficient to create a kinetic barrier, whereas an extensive number of salt bridges is required to satisfy the ion pair network hypothesis. Now, a slight excess of ion pairs has generally been found for all hyperthermophilic proteins belonging to this class, while a well defined ion pair network has only been identified in the case of *Pyrococcus furiosus* glutamate dehydrogenase.

Rubredoxin from *Pyrococcus furiosus*. The protein we have focused on for our studied is rubredoxin from *Pyrococcus furiosus* (RdPf). Aside from being very well characterized from both the biochemical (Blake et al., 1991), and structural (Blake et al., 1992; Day et al., 1992) points of view, RdPf is of an ideal size (only 53 amino acids) for model studies. Furthermore, three-dimensional structural data of several mesophilic rubredoxins are available in the literature. (see legend to Fig. 1-4).

Sequence comparison between RdPf and a variety of rubredoxins from different mesophilic species (Fig 1-4) reveal striking similarities. There is about 60 % primary sequence homology between hyperthermophilic and mesophilic species. Moreover, the cysteines involved in Fe^{3+} ligation are strictly conserved. The overall three-dimensional features of RdPf (Fig. 1-5) reveal the dominance of β -sheet secondary structure elements. Several turns are also present. Rubredoxin is an iron

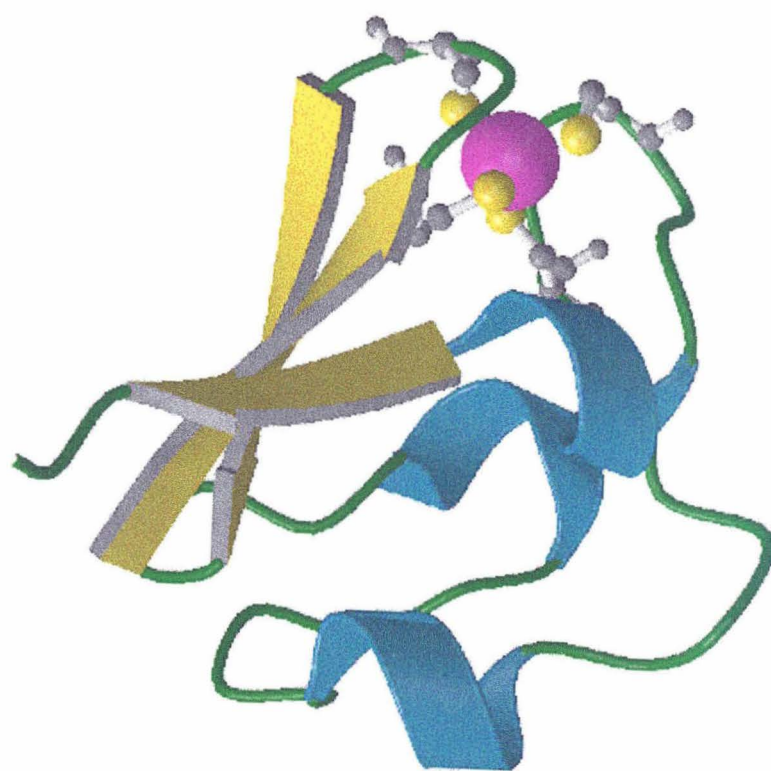
FIGURE 1-4: Sequence alignment for the rubredoxins from anaerobic bacteria. Conserved amino acids among different species are indicated in boldface. Residues that are unique to rubredoxin from *Pyrococcus furiosus* are underlined. Modified after Day *et al.* (Day et al., 1992).

Abbreviations for rubredoxins from different organisms are as follows: RdPf, rubredoxin from *Pyrococcus furiosus* (Blake et al., 1991); RdCp, rubredoxin from *Clostridium pasteurianum* (Watenpaugh et al., 1979; Yasunobo & Tanaka, 1983); RdDd, rubredoxin from *Desulfovibrio desulfuricans* (Hormel et al., 1983); RdDg, rubredoxin from *Desulfovibrio gigas* (Bruschi, 1976b); RdDv, rubredoxin from *Desulfovibrio vulgaris* (Bruschi, 1976a); RdCpf rubredoxin from *Clostridium perfringens* (Seki et al., 1989); RdCts rubredoxin from *Clostridium thermosaccharolyticum* (Meyer et al., 1990); RdCbt rubredoxin from *Chlorobium thiosulfatophilum* (Wooley & Meyer, 1987); RdDvm rubredoxin from *Desulfovibrio v*

vulgaris strain Miyazaki (Shimizu et al., 1989); RdDvH rubredoxin from *Desulfovibrio vulgaris* strain Hildenborough (Bruschi, 1976a; Voordouw, 1988); RdMe rubredoxin from *Megasphaera elsdenii* (Bachmeyer et al., 1968b); RdPa rubredoxin from *Peptococcus aerogenes* (Bachmeyer et al., 1968a); RdBm rubredoxin from *Butyrubacterium methylophilum* (Saeki et al., 1989).

	5	10	15	20	25	30	35	40	45	50
RdPf	AKWVCKICGYIYDEDAAGDPDNGISPGTKFEELPDDWVCPICGAPKSEFEKLED									
RdCp	MKKYTCTVCGYIYDPEDGDPDDGVNPGTDFKDI PDDWVCPLCGVGKDEFEEVEE									
RdDd	MQKYVCNVCGYEYDPAEHD-----NVPFDQLPSSWCCPVCVSKDQFSPA									
RdDg	MDIYVCTVCGYEYDPAKGDPDSGIKPGTKFEDLPDDWACPVCGASKDAFEKQ									
RdDv	MKKYVCTVCGYEYDPAEGDPDNGVKPGTSFDDL PADWVCPVCGAPKSEFEAA									
RdCpf	MKKFICDVCGYIYDPAVGDPDNGVEPGTEFKDI PDDWVCPLCGVDKSQFSETEE									
RdCts	MEKWQCTVCGYIYDPEVGDPTQNIPPGTKFEDLPDDWVCPDCGVDKDQFEKI									
RdCbt	MQKYVCSVCGYVYDPAEDGE PDDPIDPGTG FEDLPDEWVCPVCGVDKDLFEPES									
RdDvm	MKKYVCTVCGYEYDPAEGDPDNGVKPGTAFEDVPADWACPVCGAPKSEFEPA									
RdDvH	MKKYVCTVCGYEYDPQEGDPDNGVKPGTSFDDL PADWVCPVCGAPKSEFEAA									
RdMe	MDKYECSICGYIYDEAEGD-DGNVAAGTKFADLPADWVCPTCGADKDAFVKMD									
RdPa	MQKFECTLCGYIYDPALVGPDPDPDQDG-AFEDVSENWVCPLCGAGKEDFEVYED									
RdBm	MQKYVCDICGYVYDPAVGDPDNGVAPGTAFADLPEDWVCPECVSKDEFSPEA									

FIGURE 1-5: Ribbon diagram of the structure of rubredoxin from *Pyrococcus furiosus*. The image has been prepared by the program MOLSCRIPT (Kraulis, 1991). The β -sheet residues are colored yellow, the short α -helix portions are shown in light blue, the iron atom is shown in pink, the side chain atoms of the cysteines ligands are colored gray, and the remaining residues are colored green.



protein believed to be involved in redox processes in its host cell. However, this has never been unequivocally proven. Therefore, the biological role of RdPf is currently unknown. The most stable form has a single Fe^{3+} in a tetrahedral ligand coordination environment. Both the Fe^{3+} and Fe^{2+} forms have been isolated and even crystallized (Day et al., 1992). The structures of the protein in both oxidation states are virtually superimposable.

No surprises arise upon comparing the three-dimensional structure of RdPf to that of several mesophilic rubredoxins (Day et al., 1992): secondary structure elements are virtually identical to the point that backbones are virtually superimposable, exposed and buried surface areas do not differ significantly (Table 1-2), packing densities are similar (Rees, D. C., personal communication), and many residues in the hydrophobic core are conserved. We have also calculated the surface charge potential distribution of RdPf and of the mesophilic rubredoxins for which the crystal structure is available. The calculations were performed with the program Grasp (Nicholls et al., 1991). The rubredoxin coordinates were taken from the Brookhaven data bank. These coordinate files have been generated by the authors who solved the crystal structures. These authors are referenced in the legend to Fig. 1-4. More experimental details on how the calculations were performed can be found in the Materials and Methods section of chapter 2. The results of our calculations are presented in Figs 1-5 and 1-6. Although an initial comparison of RdPf with rubredoxin from *Clostridium pasteurianum* seems to suggest a different overall charge distribution (see also Discussion section of chapter 2), no striking variations were observed when RdPf was

TABLE 1-2: Exposed and buried surface areas for rubredoxins from different species.

Abbreviations of rubredoxins are as follows: RdPf, rubredoxin from *Pyrococcus furiosus*; RdCp, rubredoxin from *Clostridium pasteurianum*; RdDg, rubredoxin from *Desulfovibrio gigas*; RdDv, rubredoxin from *Desulfovibrio vulgaris*; RdDd, rubredoxin from *Desulfovibrio desulfuricans*. Modified after (Day et al., 1992).

PROTEIN	$A_{\text{exp}}^{\text{a)}$ ($\text{\AA}^2$)		$A_{\text{bur}}^{\text{b)}$ ($\text{\AA}^2$)			
	A_{TOT}	A_{NP}	A_{C}	$A_{\text{N/O}}$	$A_{\text{O-}/\text{N+}}$	A_{S}
RdPf	3398	1884	3637	1223	197	106
RdCp	3572	1917	3580	1178	195	129
RdDg	3325	1809	3438	1186	175	126
RdDv	2887	1902	3335	1167	206	141
RdDd	3029	1520	2997	1042	71	125

a) Exposed surface area (A_{exp}). Contributions from total exposed surface area (A_{TOT}) and non polar (carbon) exposed surface area (A_{NP}).

b) Buried surface area (A_{bur}) for carbon atoms (A_{C}), uncharged N/O atoms ($A_{\text{N/O}}$), charged N/O atoms ($A_{\text{N+}/\text{O-}}$) and sulfur atoms (A_{S}).

FIGURE 1-6: Surface potential images of rubredoxin from *Pyrococcus furiosus*, rubredoxin from *Clostridium pasteurianum*, and rubredoxin from *Desulfovibrio desulfuricans*. Surfaces are color-coded as follows: red, white and blue represent region of negative, zero and positive charge density, respectively. The images have been generated by the program GRASP (Nicholls et al., 1991).

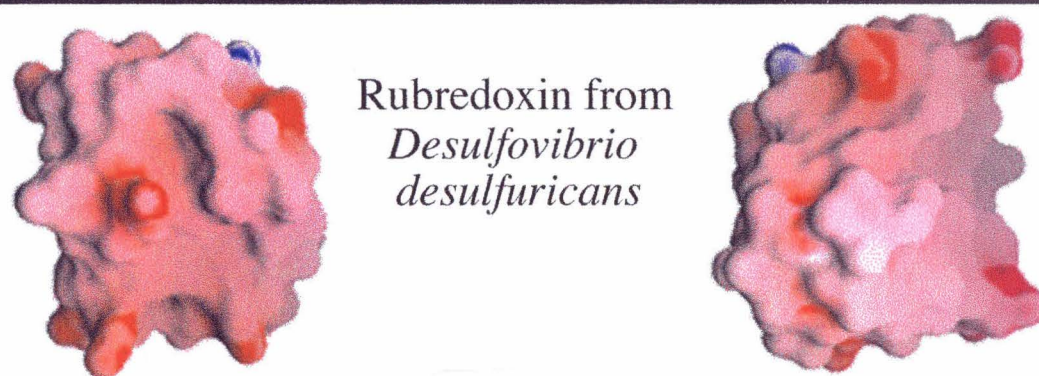
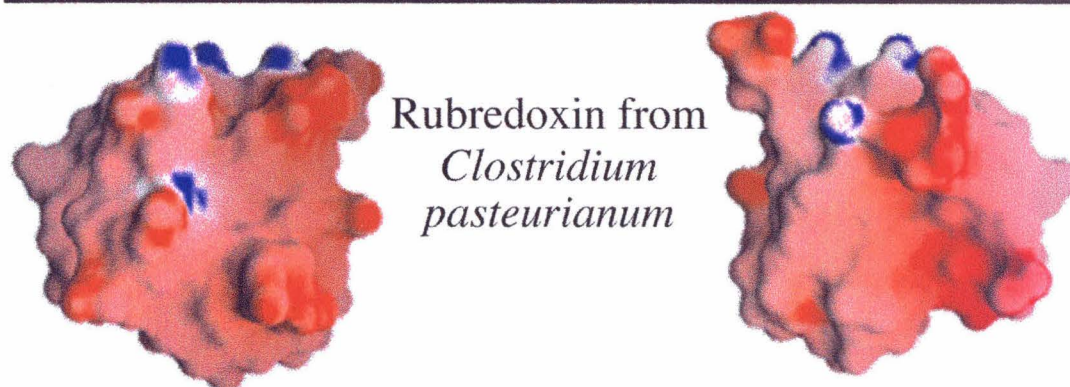
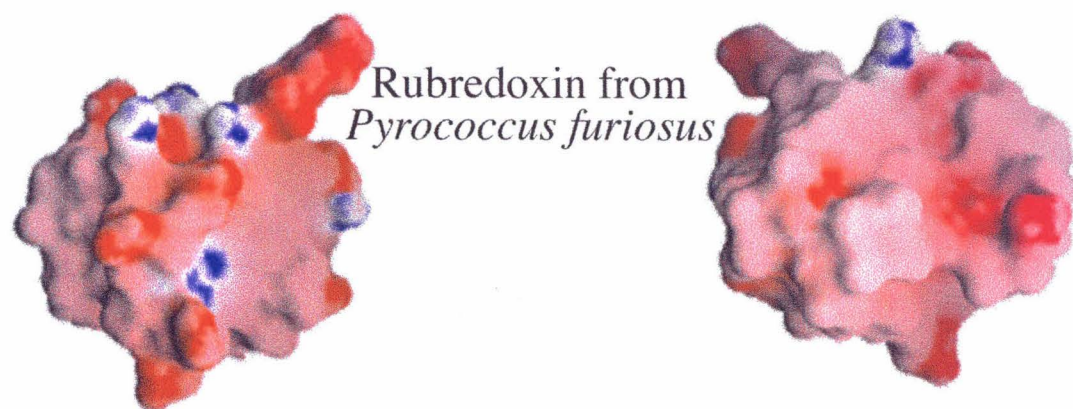
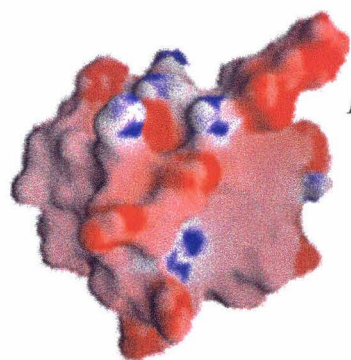
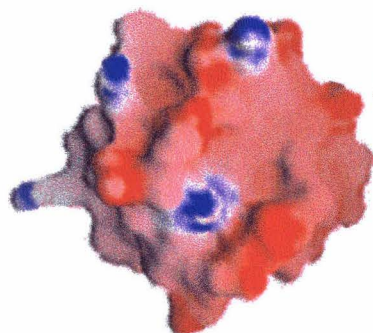
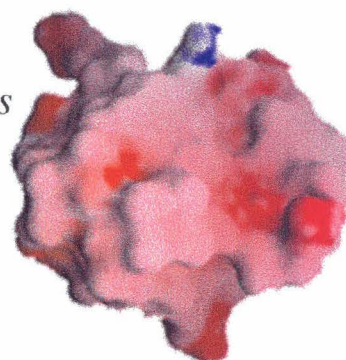


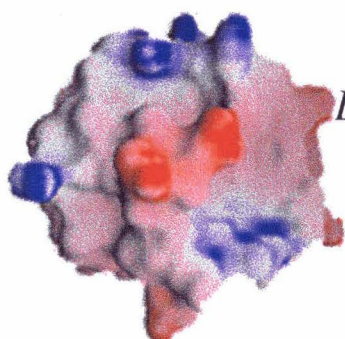
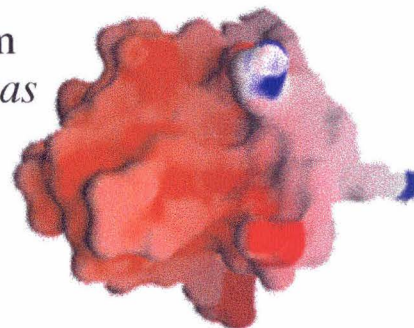
FIGURE 1-7: Surface potential images of rubredoxin from *Pyrococcus furiosus*, rubredoxin from *Desulfovibrio gigas* and rubredoxin from *Desulfovibrio vulgaris*. Surfaces are color-coded as follows: red, white and blue represent region of negative, zero and positive charge density, respectively. The images have been generated by the program GRASP (Nicholls et al., 1991).



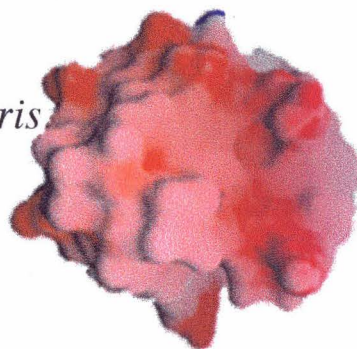
Rubredoxin from
Pyrococcus furiosus



Rubredoxin from
Desulfovibrio gigas



Rubredoxin from
Desulfovibrio vulgaris



compared to all the other species. No ion pairs networks in RdPf have been identified by this method, either. A rigorous assessment of the contribution of the charged residues to the overall free energy of the protein, should involve the explicit calculation of total electrostatic free energy (see appendix to chapter 3). However, the qualitative GRASP images are the most effective way to visualize the surface charge density distribution of a protein and look for qualitative differences among species.

This brings us to the issue of ion pairs. RdPf only has about four ion pairs *versus* two ion pairs found, for instance, in rubredoxin from *Clostridium pasteurianum*. A comparable number of ion pairs have been found in other mesophilic rubredoxins. The salt bridges of RdPf are: Glu 14 to Ala 1 (amino terminus), Glu 30 (and/or Asp 34) to Lys 28 and Lys 45, Glu 47 (and/or Glu 49) to Lys 6. Additionally, Glu 14 is also thought to form a hydrogen bond with the indole nitrogen of Trp 3. It has been proposed that these salt bridges, together with a small number of extra hydrogen bonds peculiar to RdPf, may be responsible for the hyperthermostability of the protein (Day et al., 1992). Other investigators have also suggested that the ion pair between Glu 14 and Ala 1 is crucial (Blake et al., 1992), in that it caps the N-terminus of the protein and it is therefore responsible for its thermostability. Our studies (Cavagnero et al., 1995) integrate these proposals and, to some extent, deviate from them, while providing a new perspective to the problem. We will not anticipate our results here, and we refer the interested reader to chapters 2 and 3 of this work.

CONCLUDING REMARKS

Throughout this introductory chapter, we have presented the main characteristics of the microorganisms which live in the harshest conditions presently known to be compatible with life: the extremophiles. The attention has then been focused on a subclass of these organisms, termed hyperthermophiles, which thrive at temperatures ranging between 80 and about 110°C. An outline of the thermodynamic view of protein stability, together with some preliminary results and background information on structural comparisons, have then been given with the hope that they can serve as a starting point to appreciate our contributions towards understanding the amazing properties of proteins from hyperthermophiles.

REFERENCES

- Adams, M. W. W. (1990) *FEMS Microbiol. Rev.* 75, 219.
- Adams, M. W. W. (1993) *Ann. Rev. Microbiol.* 47, 627.
- Adams, M. W. W. & Kelly, R. M. (1996) *Chem. & Eng. News Dec 18*, 32.
- Anfinsen, C. B. (1973) *Science* 181, 223.
- Aono, S., Bryant, F. O. & Adams, M. W. W. (1989) *J. Bacteriol.* 171, 3433.
- Bachmeyer, H., Benson, A. M., Yasunobo, K. T., Garrad, W. T. & Whitely, H. R. (1968a) *Biochemistry* 7, 986.
- Bachmeyer, H., Benson, A. M., Yasunobo, K. T., Peel, J. L. & Mayhew, S. (1968b) *J. Biol. Chem.* 243, 1022.
- Barlow, D. J. & Thornton, J. M. (1983) *J. Mol. Biol.* 168, 857.
- Baross, J. A. & Deming, J. W. (1983) *Nature* 303, 423.
- Becktel, W. J. & Schellman, J. A. (1987) *Biopolymers* 26, 1859.
- Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. & Adams, M. W. W. (1991) *Biochemistry* 30, 10885.
- Blake, P. R., Park, J.-B., Zhou, Z. H., Hare, D., Adams, M. W. W. & Summers, M. F. (1992) *Protein Sci.* 1, 1508.
- Bruschi, M. (1976a) *Biochim. Biophys. Acta* 434, 4.
- Bruschi, M. (1976b) *Biochem Biophys. Res. Comm.* 70, 615.
- Bryant, F. O. & Adams, M. W. W. (1989) *J. Biol. Chem.* 264, 5070.

- Burggraph, S. O., G. J., Stetter, K. O. & Woese, C. R. (1992) *Arch. Microbiol.* 15, 352.
- Cavagnero, S., Zhou, Z. H., Adams, M. W. W. & Chan, S. I. (1995) *Biochemistry* 34, 9865.
- Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995) *Science* 267, 1463.
- Constantino, H. R., Brown, S. H. & Kelly, R. M. (1990) *J. Bacteriol.* 172, 3654.
- Cowan, D. A., Smolenski, K. A., Daniel, R. M. & Morgan, H. W. (1987) *Biochem. J.* 247, 121.
- Darland, G., Brock, T. D., Samsonoff, W. & Conti, S. F. (1970) *Science* 170, 1416.
- Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J.-B., Zhou, Z. H., Adams, M. W. W. & Rees, D. C. (1992) *Protein Sci.* 1, 1494.
- Eggen, R., Geerling, A. & Watts, J. (1990) *FEMS Microbiol.* 71, 17.
- Fiala, G. & Stetter, K. O. (1986) *Arch. Microbiol.* 145, 56.
- Flam, F. (1994) *Science* 265, 471.
- Goldman, A. (1995) *Structure* 3, 1277.
- Henning, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995) *Structure* 3, 1295.
- Hormel, S., Walsh, K. A., Prickril, B. C., Titani, K. & Le Gall, J. (1983) *FEBS Lett.* 201, 147.
- Huber, R., Kristjansson, J. K. & Stetter, K. O. (1987) *Arch. Microbiol.* 149, 95.
- Jaenicke, R. (1996) *FASEB J.* 10, 84.

- Jannasch, H. W. & Mottl, M. (1985) *Science* 229, 717.
- Juszczak, A., Aono, S. & Adams, M. W. W. (1991) *J. Biol. Chem.* 266, 13834.
- Karl, D. M., Burns, D. J., Orrett, K. & W., J. H. (1984) *Mar. Biol. Lett.* 5, 227.
- Kelly, R. M., Baross, J. A. & Adams, M. W. W. (1994) *Chem. in Britain July issue*, 555.
- Kelly, R. M. & Deming, J. W. (1988) *Biotech. Prog.* 4, 47.
- Koch, R., Zabłowski, P., Spreinat, A., Lemke, K. & Antranikian, G. (1990) *FEMS Microbiol. Lett.* 71, 21.
- Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995) *J. Mol. Biol.* 246, 511.
- Kowalak, J. A., Dalluge, J. J. & McCloskey, J. A. (1994) *Biochemistry* 33, 7869.
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946.
- Lundberg, K. S., Shoemaker, D. D., Short, J. M., Sorge, J. A., Adams, M. W. W. & Mathur, E. (1991) *Gene* 108, 1.
- Makhatadze, G. I., Kim, K.-S., Woodward, C. & Privalov, P. L. (1993) *Protein Sci.* 2, 2028.
- Matthews, B. W. (1993) *Curr. Opin. Struct. Biol.* 3, 589.
- Meyer, J., Gagnon, J., Sieker, L. C., van Dorsselaer, A. & Moulis, J.-M. (1990) *Biochem J.* 271, 839.
- Mukund, S. & Adams, M. W. W. (1990) *J. Biol. Chem.* 265, 11508.
- Nicholls, A., Sharp, K. & Honig, B. (1991) *Proteins: Struct., Funct., Genet.* 11, 281.
- Perutz, M. & Raidt, H. (1975) *Nature* 255, 256.

- Pihl, T. D. & Maier, R. J. (1991) *J. Bacteriol.* 173, 1839.
- Privalov, P. L. (1979) *Adv. Prot. Chem.* 33, 167.
- Privalov, P. L. & Gill, S. J. (1988) 39, 191.
- Rees, D. C. & Adams, M. W. W. (1995) *Structure* 3, 251.
- Robb, F. T., Park, J.-B. & Adams, M. W. W. (1992) *Biochim. Biophys. Acta* 1120, 267.
- Ryu, K. G., Kim, J. B. & Dordick, J. S. (1994) *Enzyme Micr. Tech.* 16, 266.
- Saeki, K., Yao, Y., Wakabayashi, S., Shen, G. J., Zeikus, J. G. & Matsubara, H. (1989) *J. Biochemistry* 106, 656.
- Seki, Y., Seki, S., Satoh, M., Ikeda, A. & Ishimoto, M. (1989) *J. Biochem.* 106, 336.
- Shimizu, F., Ogata, M., Yagi, T., Wakabayashi, S. & Matsubara, H. (1989) *Biochimie* 71, 1171.
- Shortle, D. (1996) *FASEB J.* 10, 27.
- Stetter, K. O. (1982) *Nature* 300, 258.
- Stetter, K. O. (1985) *Naturwissenschaften* 72, 291.
- Stetter, K. O. (1986) in (Brock, T. D., Eds.) 39, Wiley & Sons, New York.
- Stetter, K. O., Fiala, G., Huber, G., Huber, R. & Segerer, A. (1990) *FEMS Microbiol. Rev.* 75, 117.
- Stetter, K. O., König, H. & Stackebrandt, E. (1983) *Syst. Appl. Microbiol.* 4, 535.
- Voordouw, G. (1988) *Gene* 67, 75.
- Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1979) *J. Mol. Biol.* 131, 509.
- Woese, C. R. (1987) *Microbiol. rev.* 51, 221.

- Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* 87, 4576.
- Woese, C. R., Magrum, L. J. & Fox, G. E. (1978) *J. Mol. Evolution* 11, 245.
- Wooley, K. J. & Meyer, T. E. (1987) *Eur. J. Biochem.* 163, 161.
- Yasunobo, K. T. & Tanaka, M. (1983) in *Iron-sulfur Proteins* (Lovenberg, W., Eds.) 27, Academic Press, New York.
- Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995) *Structure* 3, 1147.

CHAPTER 2.

Response of Rubredoxin from *Pyrococcus* *Furiosus* to Environmental Changes: Implications for the Origin of Hyperthermostability

ABSTRACT

The bases of the hyperthermostability of rubredoxin from *Pyrococcus furiosus* (RdPf) have been probed by structural perturbations induced by solution pH and ionic strength changes. Comparison of the solution behavior at pH 7 and pH 2, as probed by far- and near-UV circular dichroism, Trp fluorescence emission, 1-anilinonaphthalene-8-sulfonate (ANS) binding and NMR spectroscopy, reveals the presence of only minimal structural variations at room temperature. At pH 2, the protein displays a surprising nearly native behavior at high ionic strength while, at low ionic strength, it is capable of strongly binding the hydrophobic probe ANS. All of the secondary and tertiary structural features, including the environment of the hydrophobic core, appear to be intact regardless of pH and ionic strength. The apparent "melting" or denaturation temperature at pH 2, however, is 42°C lower than at pH 7. This is attributed to the perturbation of many electrostatic interactions, including the disruption of all the ion pairs, which is complete at pH 2, as indicated by electrometric pH titrations. The implications of these findings for the origins of the hyperthermostability of rubredoxin are discussed.

INTRODUCTION

The existence of life forms that live at elevated temperature and pressure is a manifestation of nature's ingenuity in managing to support life even under the most

adverse conditions. A variety of so-called hyperthermophilic microorganisms have been isolated in recent years (Stetter et al., 1990; Adams, 1990; Stetter, 1986) in the vicinity of both terrestrial and submarine hydrothermal vents. The majority of them are classified as *Archaea* (Woese et al., 1990) and thrive at temperatures up to 110°C, the highest temperature known to be compatible with life. They are also the most slowly evolving or "primitive" of known organisms. Little is known of the novel biochemistry that must be required to sustain life above 100°C. However, the quest for biomolecules displaying unusual and interesting physico-chemical properties has recently led to the discovery and isolation of specific enzymes from these hyperthermophilic organisms. Among those that have been purified so far are proteases (Eggen et al., 1990; Cowan et al., 1987), redox proteins (Aono et al., 1989; Blake et al., 1991), hydrogenases (Bryant & Adams, 1989; Juszczak et al., 1991; Pihl & Maier, 1991), dehydrogenases (Robb et al., 1992), ferredoxin-dependent oxidoreductases (Mukund & Adams, 1990), DNA polymerases (Lundberg et al., 1991) and enzymes involved in carbohydrate metabolism (Koch & Zabłowski, 1990; Constantino et al., 1990). These enzymes are composed of the regular 20 amino acids and display a thermal stability unknown to their mesophilic counterparts. Yet, sequence comparisons reveal no apparent striking differences. At this juncture, the origin of protein hyperthermostability (Flam, 1994; Adams & Kelly, 1992) is still an unsolved problem. Its understanding presents a great intellectual challenge to the chemist and the biochemist, not to mention its potentially enormous biotechnological impact (Kelly et al., 1994; Flam, 1994).

Several studies based on site-directed mutagenesis of non-thermophilic proteins have been able to show that even minimal structural variations are sufficient to cause some change in stability. T4 lysozyme, for instance, is a typical protein on which mutational analysis has been carried out extensively (Matsumura et al., 1989a,b; Nicholson et al., 1988; Matthews et al., 1987). However, relatively small variations have not, as yet, been sufficient to generate a hyperthermostable protein starting from a non-thermostable one. This immediately prompts the question of whether the structural/dynamic signatures of hyperthermostability are to be ascribed to a small number of specific non-covalent interactions, or more general rules exist that go beyond the identification of a few specific contributions.

We have started to address the problem by focusing our studies on the small (53 amino acids) hyperthermostable non-heme iron protein rubredoxin produced by the hyperthermophilic archaeon *Pyrococcus furiosus*. This microorganism, isolated by Fiala and Stetter (1986) in the vicinity of a mediterranean volcanic vent, grows under strictly anaerobic conditions and has an optimal growth temperature of 100°C. Despite the fact that rubredoxin from *P. furiosus* (RdPf) was isolated only recently (Blake et al., 1991), it is a remarkably well characterized protein: its sequence is known (Blake et al., 1991); the crystal structures of both the oxidized and the reduced forms have been solved (Day et al., 1992); and an NMR solution structure is available for the Zn-substituted protein (Blake et al., 1992a). The metal environment has been specifically probed by heteronuclear NMR experiments on various metal substituted forms (Blake et al., 1994; Blake et al., 1992b,c, 1994) and by X-ray absorption spectroscopy (George et al.,

1992). The thermodynamics of stability has also been studied by differential scanning calorimetry (Klump et al., 1994).

The present study is designed to elucidate some of the factors that might be important in stabilizing a hyperthermophilic protein based on the structural perturbations induced by changes in solution pH, ionic strength and temperature. The effects produced have been followed by Trp fluorescence emission, NMR, far- and near-UV circular dichroism (CD), 1-anilinonaphthalene-8-sulfonate (ANS) binding and electrometric pH titrations. The surface charge density distribution has also been probed by surface potential calculations. All the experiments have been carried out on the high spin Fe(III) form of RdPf. The implications of the results to the understanding of the hyperthermostability of RdPf and, possibly, of other proteins belonging to the same class, are discussed.

MATERIALS AND METHODS

RdPf was isolated and purified as previously described (Blake et al., 1991). Protein concentrations were determined by using the previously published extinction coefficient at $\lambda=280$ nm. High purity ANS was purchased from Molecular Probes. 99.9% D₂O was purchased from Cambridge Isotopes.

Circular Dichroism spectra were acquired with a JASCO 600 spectrophotometer using a water-jacketed 0.1 cm path quartz cell from Hellma in 10 mM phosphate buffer at pH 2.0 or 7.2. When appropriate, the buffers also contained 1M NaCl; after salt

addition the pH was readjusted to the original value. RdPf concentration was 10 to 50 μM (far UV CD) and 100 μM to 1 mM (near UV CD). Temperature control for the CD experiment was achieved by connecting the cuvette to a water bath appropriately set to the desired temperature.

Fluorescence spectra were recorded with an SLM 4800 fluorimeter. The Trp fluorescence emission spectra of RdPf were recorded at room temperature in a stirred fluorescence quartz cell. The sample was excited at 280 nm, and the emission maximum was observed at 335 nm. The excitation bandwidth was set to 4 nm, and the emission bandwidth was set to 8 nm. The spectra were recorded in 20 or 40 mM phosphate buffers at variable pH, containing, when appropriate, 1M NaCl. The excitation wavelength for the ANS emission spectra was 350 nm. The emission maximum varied between 510 nm and 480 nm depending on the fraction of protein-bound ANS. Binding constants were determined by adding increasing amounts of ANS to RdPf solutions (3 to 4 μM). The fraction of bound ANS (α) was obtained by comparing the observed ANS fluorescence emission areas with the emission areas of quinine disulfate solutions (F_{qds}) having the same optical density (at 350 nm) as the ANS solutions (Kolb & Weber, 1975), according to the formula:

$$\alpha = \frac{(F_{\text{obs}} - F_0)}{[(F_{\text{qds}} / A) - F_0]}, \quad (2-1)$$

where F_{obs} represents the observed ANS fluorescence in the presence of protein, F_0 is the fluorescence of ANS solutions of the same concentration but containing no protein.

The constant A was determined via the relation

$$A = \frac{F_{\text{qds}}}{F_{\text{b}}} , \quad (2-2)$$

where F_{b} represents the fluorescence of an ANS solution in the presence of protein at such a low ANS/protein ratio that virtually all the ANS is in bound form. The above procedure also has the advantage of internally correcting for inner filter effects due to the ANS. Apparent dissociation constants ($K_{\text{d,app}}$) were determined via non-linear regression analysis. A plot of the average number of ANS molecules bound per protein molecule

($\langle n \rangle$) vs concentration of free ANS ($[\text{ANS}]$) was fit to the expression

$$\langle n \rangle = \frac{N [\text{ANS}]}{([\text{ANS}] + K_{\text{d,app}})} . \quad (2-3)$$

The number of binding sites per protein molecule (N) was determined via linear regression analysis of Klotz plots (Klotz & Hunston, 1971).

UV-VIS absorption spectra were recorded with a Hewlett-Packard 8452A single-beam diode array spectrophotometer. Spectra were obtained in water or aqueous buffer with a 1 cm quartz cuvette under constant stirring at 5 μM RdPf concentrations. Published values for the extinction coefficients of RdPf at the wavelengths of maximum absorption were used (Blake et al., 1991). The extinction coefficients of native RdPf in

the UV-VIS region are not temperature-dependent over the temperature range relevant to the present study (20 to 95°C). However, they differ from the extinction coefficients of the unfolded form. In thermal denaturation experiments, unfolding was followed by monitoring the absorbance changes at 494 nm. The observed absorbance (F_{obs}) reflected the different populations of folded and unfolded forms at each individual temperature. The fraction of unfolded protein (X_u) was determined by using the formula:

$$X_u = \frac{F_f - F_{\text{obs}}}{F_f - F_u} . \quad (2-4)$$

F_f and F_u refer to the absorbance of the fully folded and fully unfolded protein, respectively. Apparent melting or denaturation temperature (T_m) values were derived as the temperature corresponding to $X_u=0.5$. No curve fitting was employed, given the relatively high experimental error associated with the instrumentation used for temperature measurements. The temperature was controlled via a Haake D1 thermostated water bath connected to the compartment surrounding the cuvette housing. In thermal denaturation experiments, the temperature was increased at a rate of 2°C/min. Temperature was monitored using a thermocouple probe connected to a Fluke 80TK module plugged into a multimeter. Readings were accurate to $\pm 1^\circ\text{C}$. The spectra were baseline corrected with the software Lab Calc.

Proton NMR spectra were acquired with a Bruker AMX-500 spectrometer. Protein concentrations for the NMR spectra were 1 mM; the experiments were

performed either in 90:10 H₂O/D₂O or in D₂O. In this latter case, RdPf was concentrated in a Centricon 3 device (Amicon), diluted in D₂O, incubated at 37°C for several hours, concentrated and rediluted in D₂O; the procedure was repeated as necessary. The SUPERWEFT (Inubushi & Becker, 1983) experiment was performed at varying intervals of temperature up to 85°C, using an interpulse delay of 50 msec, a recycle time of 0.5 sec, and a sweep width of 250 ppm.

Electrometric pH titrations were performed with a pH meter from Radiometer Copenhagen model PHM 93 equipped with a general purpose pHC2406 electrode. Titrations were performed under N₂ following previously described procedures (Tanford, 1962). KOH and HCl solutions were carefully standardized in order to obtain reliable results. RdPf solutions were concentrated with a Centricon 3 device (Amicon), desalted using a PD10 column (Pharmacia), and brought to a total ionic strength of 0.01 with KCl. The solution was gently stirred through the duration of the whole experiment and acid/base additions were made while keeping conditions anaerobic. Data from calculations based on blank titrations have been subtracted from the experimental points for the protein-containing solutions. RdPf concentration was typically 0.3 mM.

The program GRASP (Nicholls et al., 1991) was used for surface potential calculations based on the Poisson-Boltzmann equation. The ionizable amino acids were assigned full or zero charge depending on the simulated pH; backbone carbons, nitrogens and oxygens were assigned partial charges according to the CHARMM (Brooks et al., 1983) charge set. The crystal structures of RdPf (Day et al., 1992) and

RdCp (Watenpaugh et al., 1980; Watenpaugh et al., 1979) have been used as the starting templates.

RESULTS

Circular Dichroism. Figures 2-1 and 2-2 show a comparison of the CD spectra of RdPf in the far, near UV and visible regions at pH 7 and 2 at variable ionic strengths. The far UV CD (Fig. 2-1) probes the three-stranded antiparallel β -sheet environment that comprises about 30% of the RdPf primary sequence (Blake et al., 1991), while the CD in the near UV and visible (Fig. 2-2), recorded between 240 and 700 nm, reflects the environment of the aromatic amino acids and the Fe(III) site. In all pH and ionic strength ranges examined, the spectra appear essentially unchanged, indicating that there are no significant changes in the secondary and tertiary structural elements under these conditions.

UV-VIS Absorption. The most informative feature of the absorption spectrum of oxidized RdPf comes from the signals with maxima centered at 380, 494 and 570 nm (Fig. 2-3), which are attributed to ligand to metal charge transfer transitions by analogy with the corresponding visible absorption bands of the non-thermophilic rubredoxin (Eaton & Lovenberg, 1973). At 25°C, no spectral changes were observed between pH 6.8 and 1.8 (Fig. 2-3), in agreement with the conclusions from the CD data.

FIGURE 2-1: CD spectra of RdPf in the far-UV region. The plot shows a superposition of the spectra of RdPf at pH 2.0 and pH 7.2, in the presence and absence of 1M NaCl. The spectra have been recorded at room temperature.

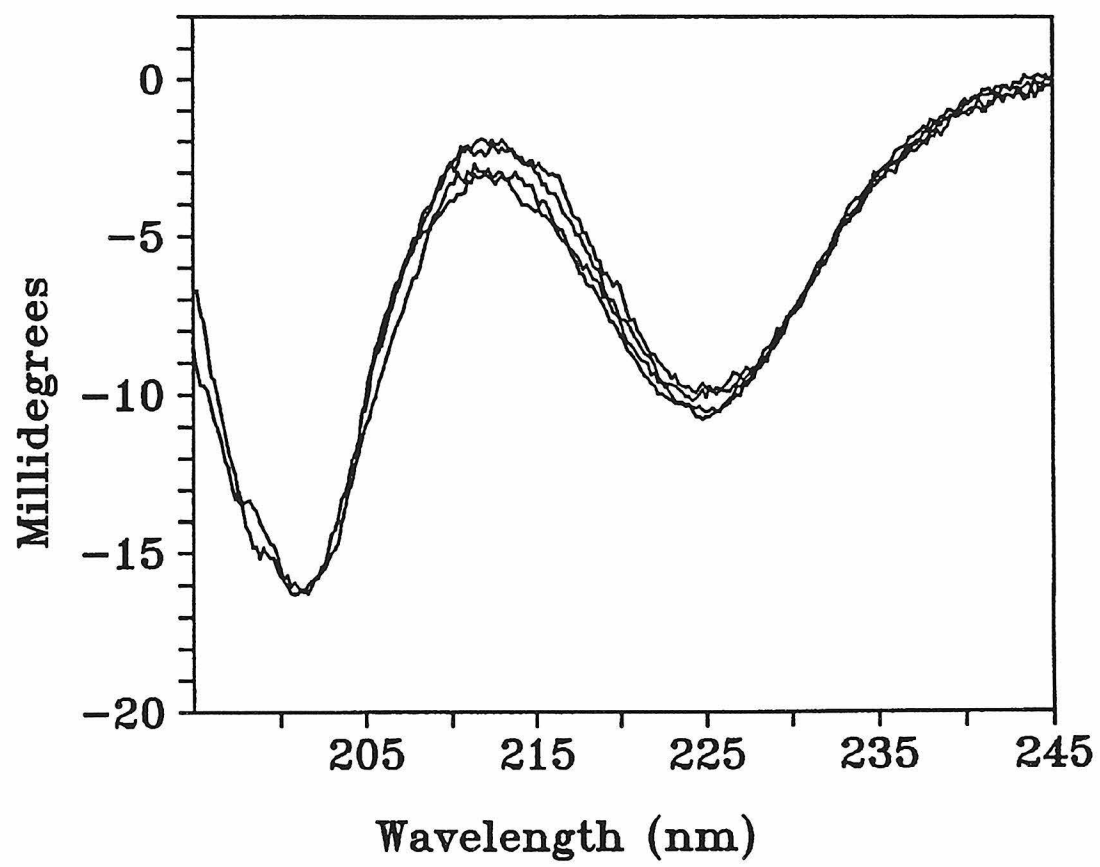


FIGURE 2-2: CD spectra of RdPf in the near-UV/visible region. The plot shows a superposition of the spectra of RdPf at pH 2.0 and pH 7.2, in the presence and absence of 1M NaCl. The spectra have been recorded at room temperature.

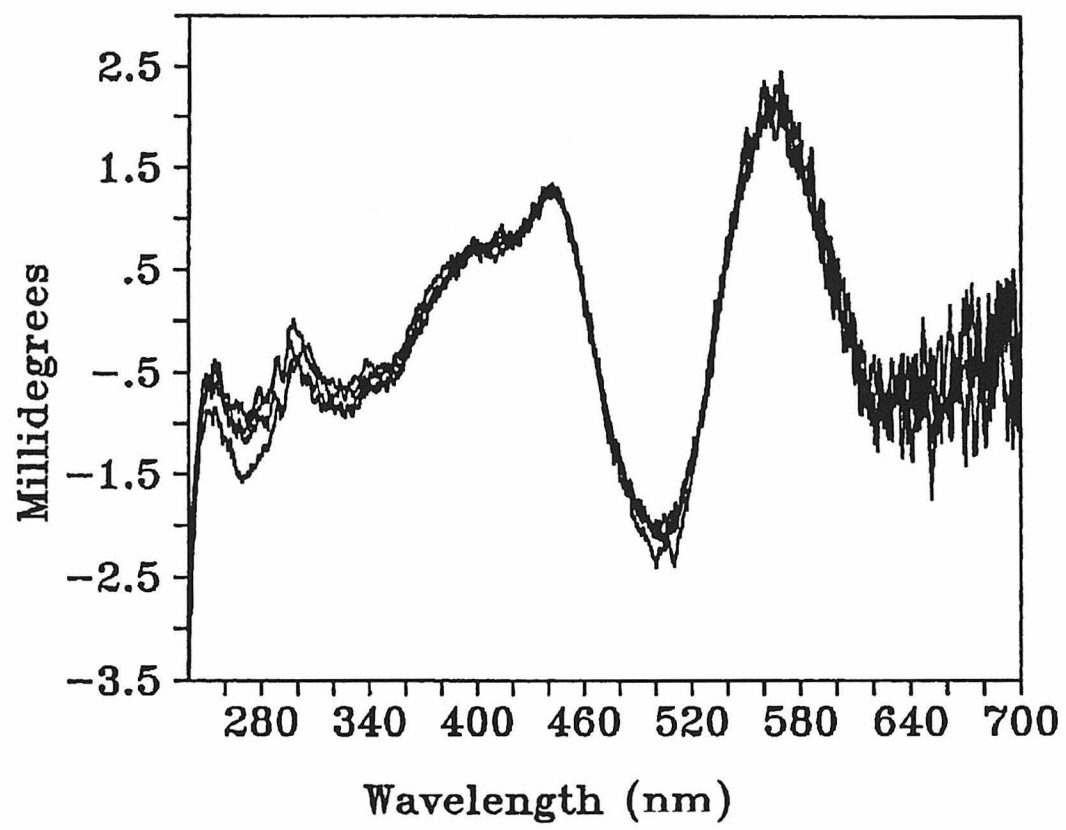
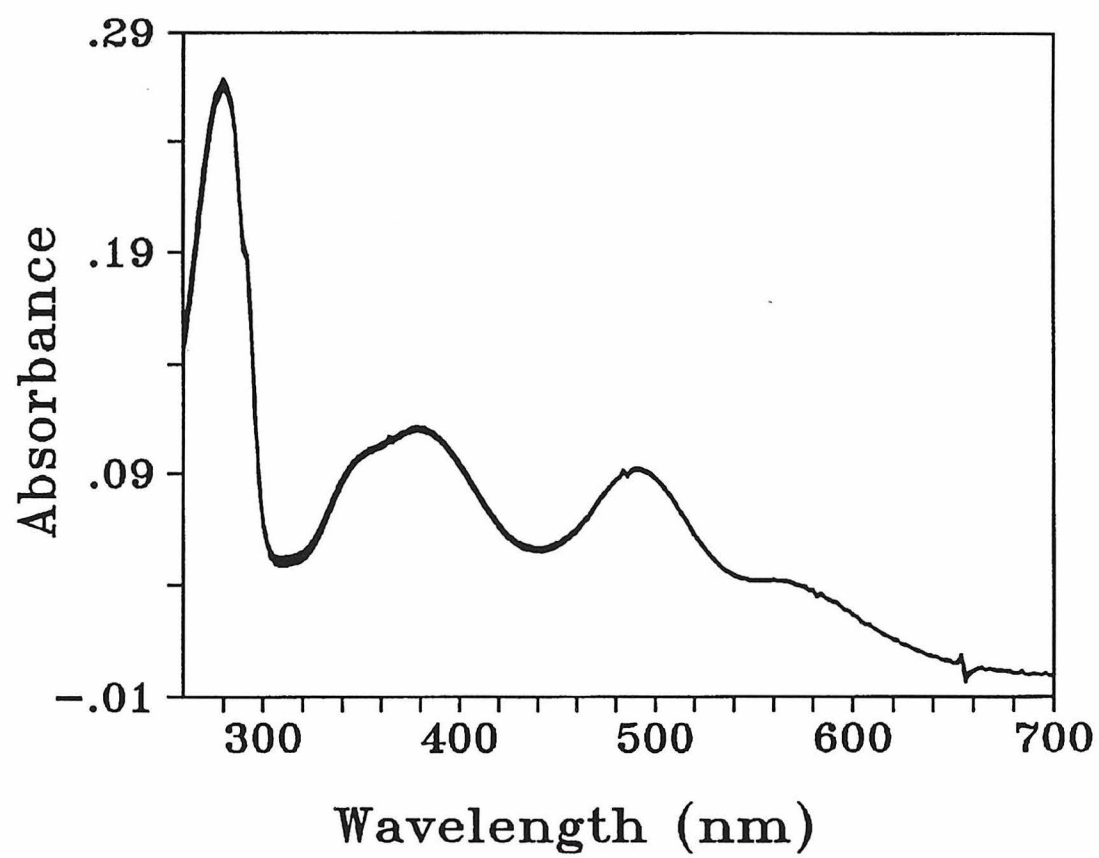


FIGURE 2-3: Absorption spectra of RdPf at room temperature in water at variable pH
(pH values are: 6.8, 5.4, 4.2, 3.8, 3.6, 3.3, 2.9, 2.4, 2.1, 1.8).



RdPf undergoes thermal denaturation around 70°C at pH 2. The protein absorption spectra as a function of temperature, and the corresponding thermal unfolding profile as followed by the decrease in absorbance at 494 nm, are shown in Figures 2-4 and 2-5, respectively. At pH 2 in H₂O, T_m was 71°C. As expected, the unfolding was irreversible, i.e. upon cooling there was no increase in the visible absorption. In the presence of 1 M NaCl at pH 2 the melting temperature was higher ($T_m = 80^\circ\text{C}$).

Some unfolding experiments at low ionic strength were performed by first bringing the temperature of the solution to 71°C and then lowering the pH stepwise. In this case it was found that the denaturation was 50% complete at pH 1.85.

Trp Fluorescence Emission. The intensity of the steady state fluorescence emission of RdPf upon excitation at 280 nm (Trp) is rather weak due to the sulfur-coordinated iron center which, because of charge-transfer transitions, functions as an inner filter, thereby decreasing the effective Trp fluorescence quantum yield. Although this decreases sensitivity, Trp fluorescence emission data of RdPf show that the Trp environment is relatively insensitive to pH and ionic strength. For example, Figure 2-6 compares the Trp emission at pH 2 and 7.2, room temperature, in various buffers and ionic strengths, demonstrating the exceptional resistance of RdPf to pH extremes when compared to the typical behavior of non-thermophilic proteins.

ANS Binding. ANS is a polycyclic aromatic fluorescent probe that interacts with proteins whose hydrophobic sites are, to some extent, solvent-exposed (Semisotnov et al., 1991). It has been used extensively to probe the molten globular

FIGURE 2-4: Thermal denaturation of RdPf in H₂O at pH 2.0 as followed by absorption spectroscopy.

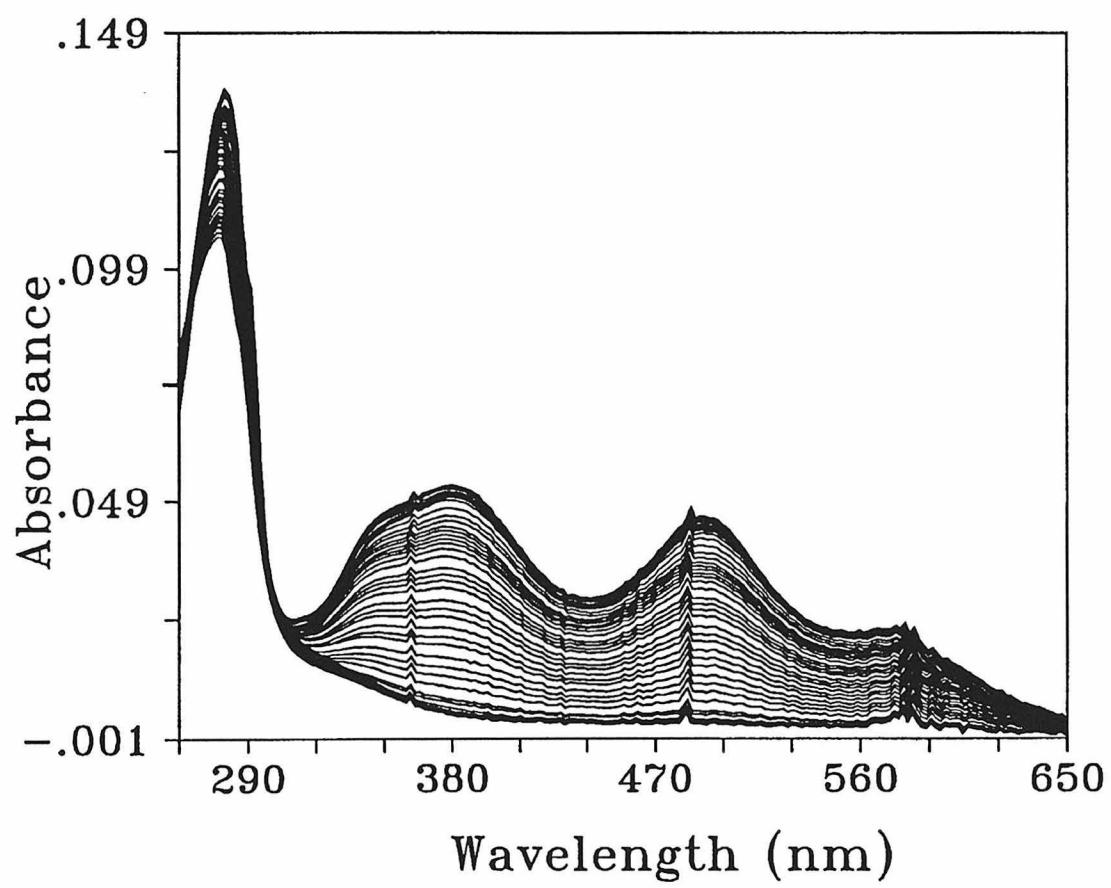


FIGURE 2-5: Denaturation curve for the thermal unfolding of RdPf in H₂O at pH 2.0 as followed by absorption spectroscopy.

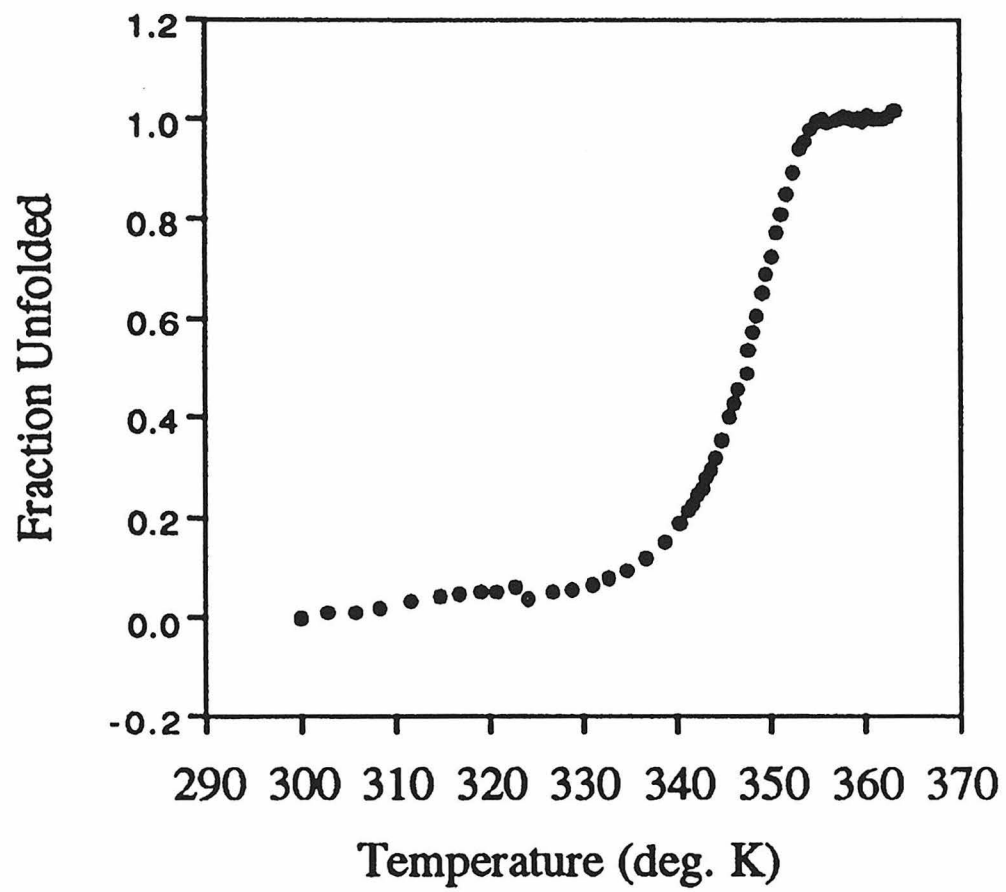


FIGURE 2-6: Fluorescence emission spectra of RdPf upon excitation at 280 nm. Buffers are as follows: 20 mM phosphate 1M NaCl pH 2.0 (A), 20 mM phosphate pH 2.0 (B), 20 mM phosphate 1M NaCl pH 7.0 (C), 20 mM phosphate pH 7.0 (D). The spectra have been recorded at 24°C.

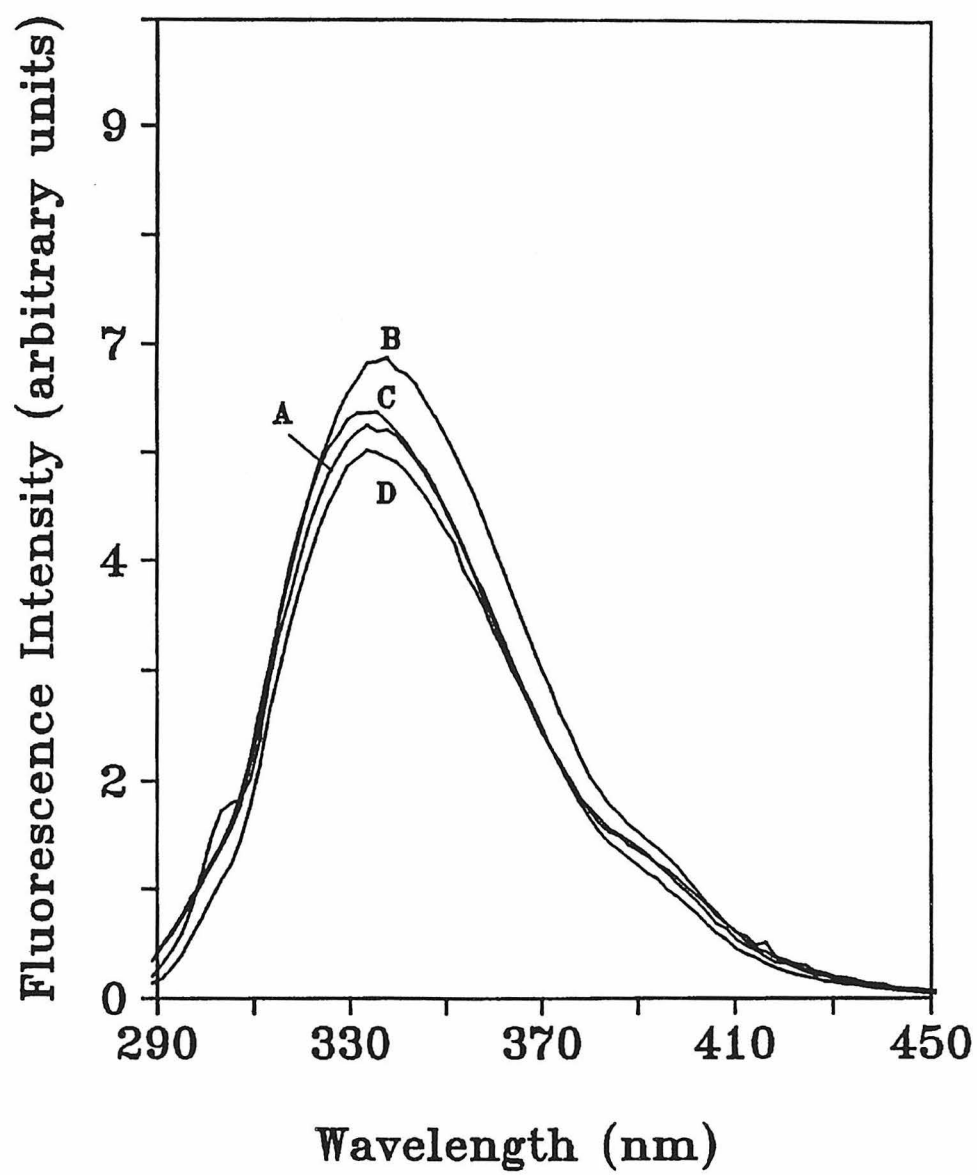
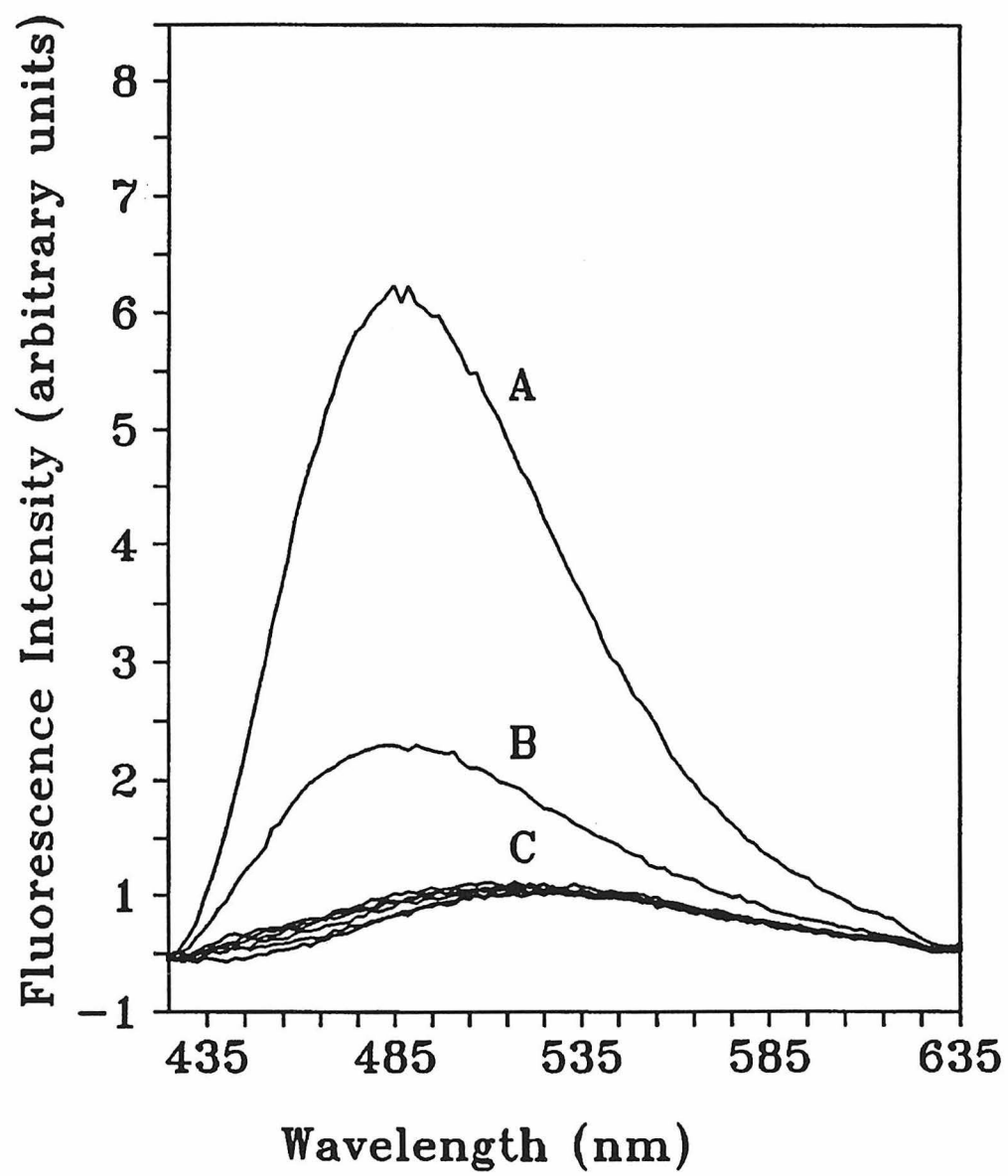


FIGURE 2-7: ANS fluorescence emission spectra in the absence and presence of RdPf at variable pH and ionic strength values: ANS+RdPf, pH 2.0 (A); ANS+RdPf, 1M NaCl, pH 2.0 (B); ANS+RdPf, pH7.0 (C); ANS+RDPf, 1M NaCl, pH 7.0 (C); ANS, pH 2.0 (C); ANS, pH 7.0 (C). Phosphate buffer (40 mM) has been used. ANS concentration is 0.8 μ M in all spectra. Buffer blanks have been subtracted. All the spectra have been recorded at 24°C.

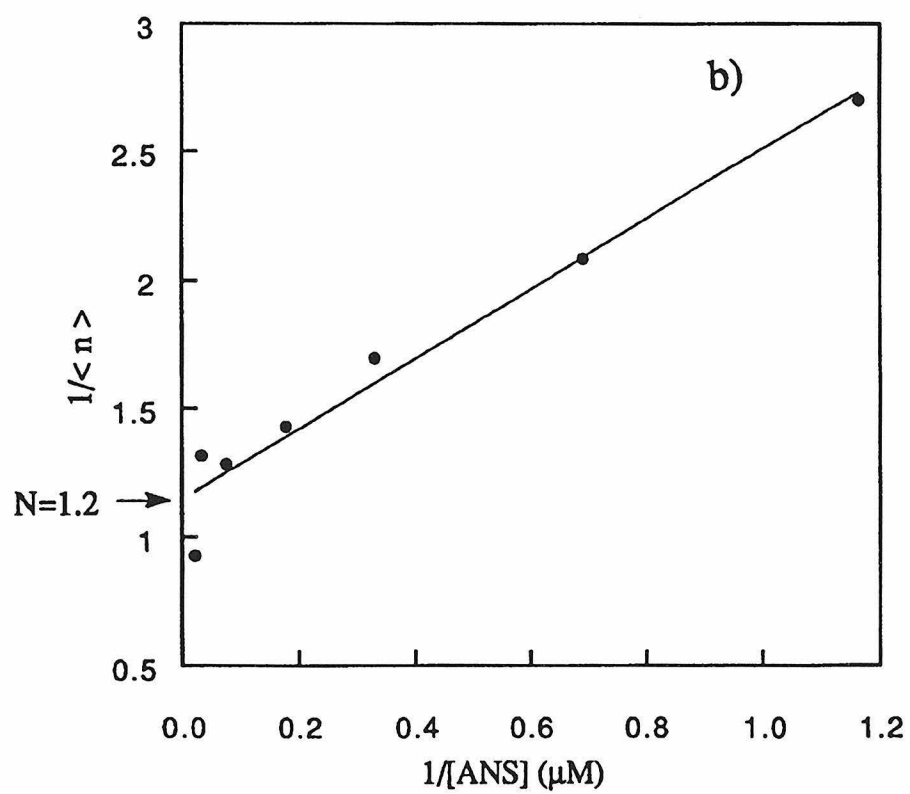
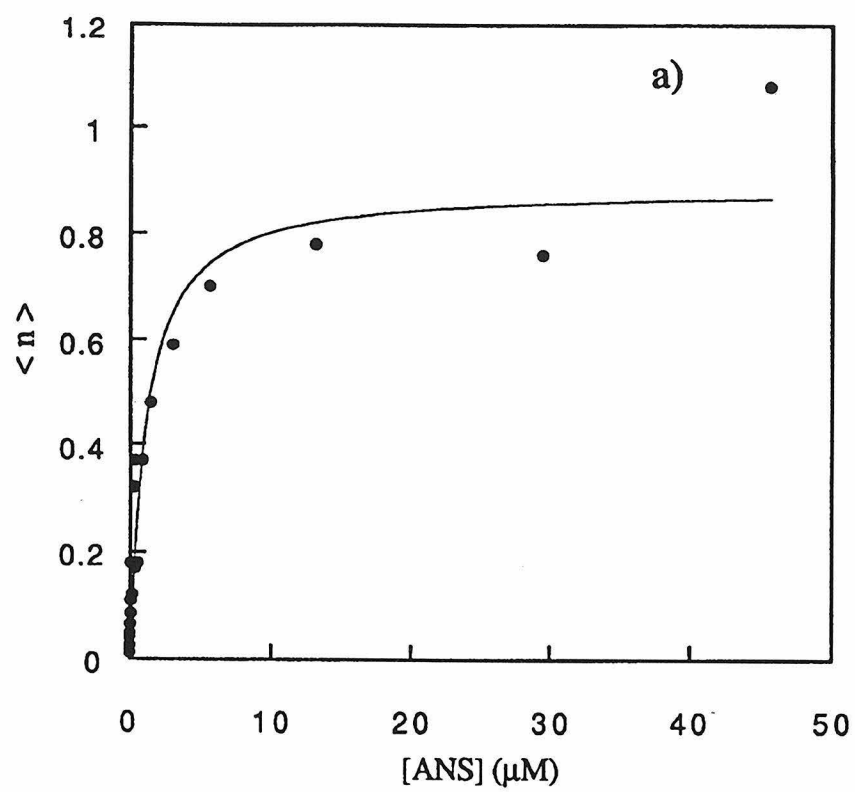


states (Kuwajima, 1989) of proteins; we have applied this approach to RdPf to probe the overall folding of the protein at low pH. Figure 2-7 shows that, at pH 2, room temperature, and low ionic strength, the quantum yield for the ANS fluorescence emission increases dramatically upon addition of RdPf to the solution in nearly equimolar amounts. Analysis of the concentration dependence of ANS binding at pH 2 yields a $K_{d,app}$ of 1.1 μM (Figure 2-8a). The Klotz plot (Klotz & Hunston, 1971) shown in Figure 2-8b is consistent with a 1:1 ANS to protein binding stoichiometry ($N=1.2$). However, the binding takes several hours to reach equilibrium. At pH 2, the ANS binding becomes weaker at high ionic strength ($K_{d,app} = 37 \mu\text{M}$), and it is completely suppressed at pH 7, regardless of the ionic strength.

NMR Spectroscopy. Proton NMR spectra of RdPf were recorded at pH ranging from 8 to 2. The 1D ^1H -NMR spectrum changes only slightly upon lowering the pH. Although the spectral changes due to individual peaks are hard to follow, careful inspection indicates that some lines undergo extensive broadening without chemical shift changes. We emphasize that the 1D NMR spectrum has revealed no overall variations in chemical shift dispersion over the pH range 8 - 2. Furthermore, the observed spectral changes were completely reversible.

The presence of high spin Fe(III) in the oxidized RdPf introduces a hyperfine and pseudohyperfine chemical shift contribution to the proton resonances (Bertini & Luchinat, 1986). The former only affects protons four or five bonds removed from the iron site; these protons, due to the high magnetic moment of high spin Fe(III), undergo extensive line-broadening. The pseudohyperfine contribution to the chemical shift is

FIGURE 2-8: Saturation curve (a) of RdPf at pH 2.0 upon addition of variable amounts of ANS. RdPf concentration is $3.8 \mu\text{M}$. $\langle n \rangle$ is the average number of ANS molecules bound per protein molecule; $[\text{ANS}]$ represents the free ANS concentration. Klotz plot (b) of ANS binding to RdPf at pH 2.0. The total number of binding sites per protein molecule (N) derived from curve fitting is indicated on the Figure.



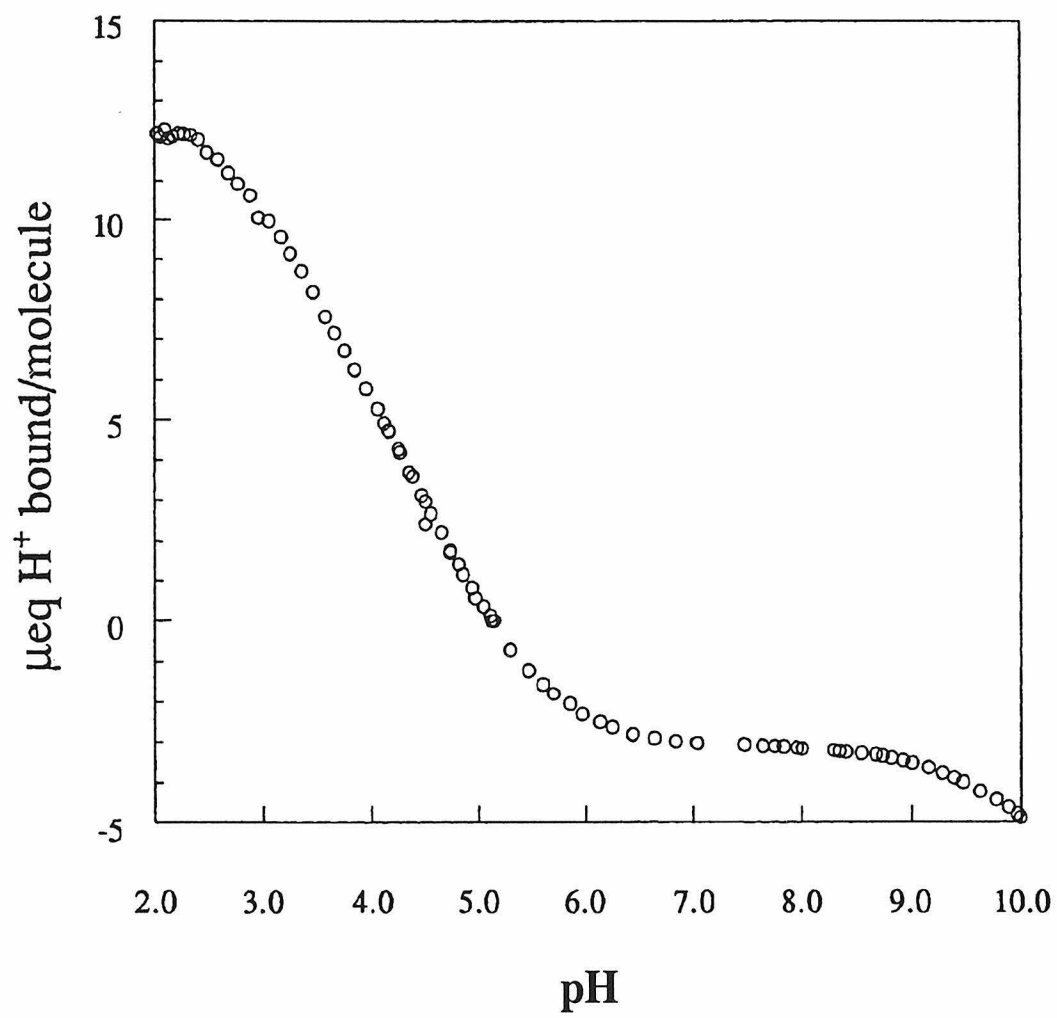
dipolar in nature and, in the case of axially symmetric systems, it is proportional to ($g_{\parallel} - g_{\perp}$), which vanishes in a perfectly tetrahedral system. According to the crystal structure of RdPf (Day et al., 1992), the S-Fe-S bond angles deviate from perfect tetrahedral symmetry by less than 6° and the Fe-S bonds are all of nearly equal length, ranging from 2.27 to 2.31 Å. Low temperature EPR studies on rubredoxin from *Pseudomonas oleovorans* revealed dispersion arising from zero-field splitting, with $D \approx 1.51 \text{ cm}^{-1}$ and $E/D = 0.28$ (Peisach et al., 1971). This result has been interpreted by Peisach and coworkers in terms of "substantial" rhombic distortion. However, at ambient temperatures, we expect rapid relaxation among the zero-field states, in which case any subtle effects of the zero-field splitting on the proton chemical shifts are expected to be small, if not negligible. This should be particularly relevant here, since D is small. If so, the pseudo-contact shifts should be governed by any intrinsic anisotropy in the electron Zeeman interaction. Judging from the low temperature EPR data, where signals from at least two pairs of the zero field levels are observed, we have concluded that $\Delta g \leq 0.01$. This implies that the pseudocontact contributions to the observed proton NMR chemical shifts of oxidized RdPf should be minimal. In support of this, proton NMR analysis of oxidized rubredoxin from *Clostridium pasteurianum* (RdCp) did not reveal any resonance out of the diamagnetic envelope at 200 MHz (Phillips et al., 1970) and some very broad peaks only slightly downfield of the diamagnetic envelope were noted when the analysis was performed at 470 MHz (Krishnamoorthi et al., 1986); we have found similar results for RdPf after extensive accumulation with the SUPERWEFT pulse sequence (data not shown). As a consequence, most of the chemical shift values of Zn-

substituted RdPf (Blake et al., 1991) should be very similar to those of the oxidized native RdPf.

In the case of RdPf, the ^1H -NMR resonance most diagnostic of the tertiary folding of the protein is the peak at -1.61 ppm (at 27°C) which has been assigned to Leu 32 δCH_3 in the Zn-substituted RdPf. This methyl group has an average distance of 12.5 Å from the metal. This, together with the above discussion, implies negligible pseudocontact contributions to the observed proton chemical shift of this resonance in the native RdPf; the assignment is therefore extended to the native oxidized RdPf. This signal is an NMR probe of the hydrophobic core. Careful scrutiny of the spectral data shows that this resonance does not undergo any chemical shift change when the pH is lowered to 2.

Electrometric pH titrations. Electrometric pH titrations are the most direct and unequivocal technique to detect the number of groups that become ionized as the pH changes. The pH titration curve of RdPf at low ionic strength is shown in Figure 2-9. RdPf has several ionizable groups: 7 Asp, 6 Glu and 5 Lys out of 53 residues. The titration end-point at low pH is clearly marked as the curve flattens towards pH 2. Proton consumption and group counting shows that, upon going from pH 7 to pH 2, all of the Asp, Glu and the carboxyl termini have been protonated. The titration end points of individual residues are not resolved, but this is typical of the pH titration curves of charged polymers. The curve does not differ, in its general features, from the corresponding curves of common proteins.

FIGURE 2-9: Electrometric pH titration curve of RdPf.



Surface Potential Calculations. The surface potential of RdPf has been calculated with the software GRASP (Nicholls et al., 1991). The charge files used are intended to simulate the observed protonation state of the protein at pH 7 and pH 2. The calculations illustrate the surface charge density distribution and its changes as a function of pH. Figure 2-10a simulates the protonation state of RdPf at pH 7 and shows that all the charge density due to the positively- and negatively-charged amino acids resides on the surface of the protein. Furthermore, most of the negatively-charged functionalities (Asp, Glu and carboxyl terminus) are distributed on one side of the molecule (see image on the left of Figure 2-10a). Interestingly, the 5 positively charged amino acids (Lys) and the amino terminus are also clustered on the same side of the molecule allowing for a number of ion pairs to be formed, including those between Glu 14 and the amino terminus and between Glu 47/Glu 49 and Lys 6. These are the ion pairs that have been proposed to be unique to RdPf (Day et al., 1992).

Figure 2-11 illustrates the surface charge density distribution of rubredoxin from *Clostridium pasteurianum*, a mesophilic organism. All of the ionizable amino acids have been labeled in the Figure. It is noted that the negatively-charged functionalities are distributed much more evenly than in the case of RdPf. The 4 Lys and the amino terminus, on the other hand, are clustered on one side of the molecule (see upper region of the RdCp images). This results, as in the case of RdPf, in a greater concentration of ion pairs on one side of the molecule.

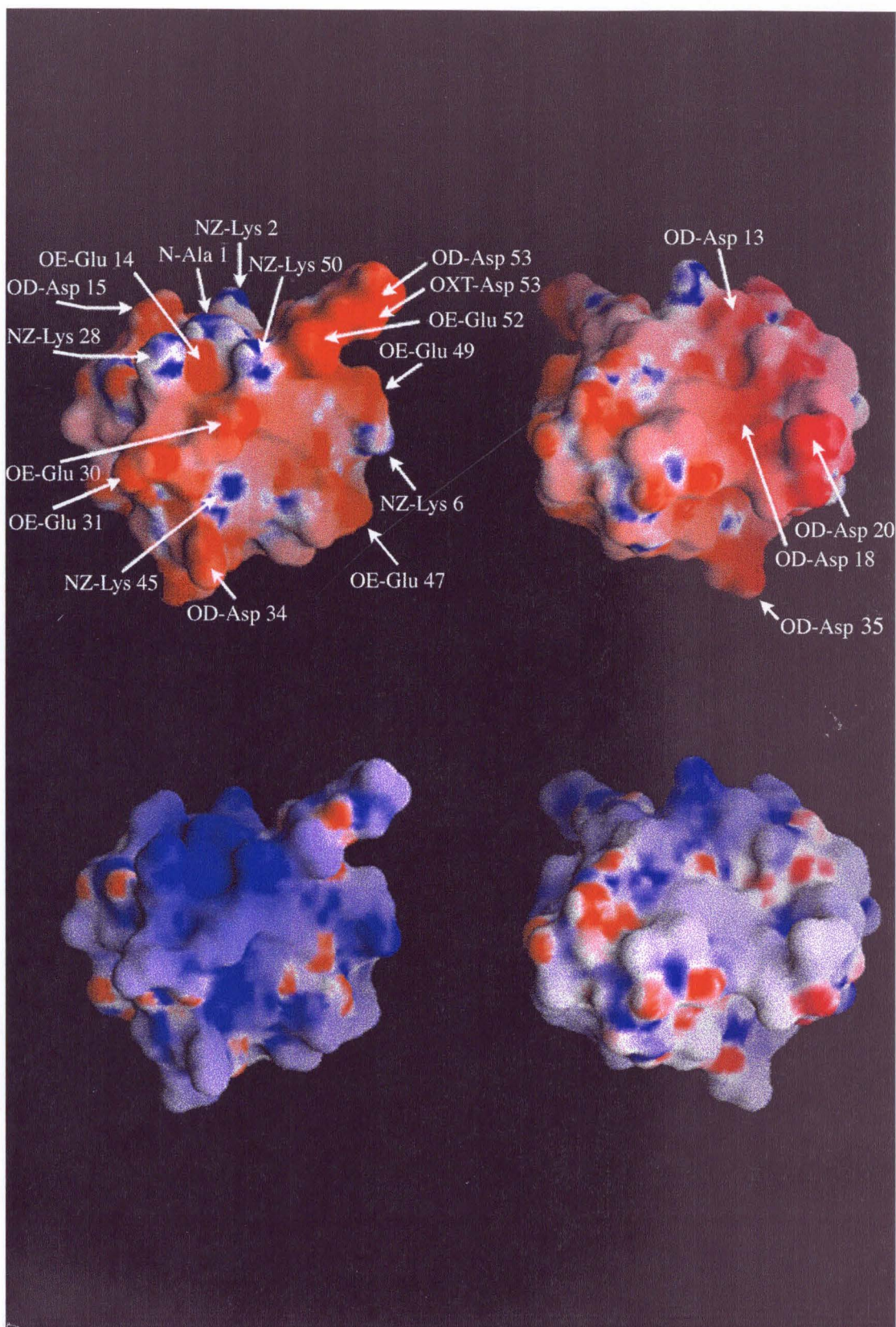
The surface charge density distribution of RdPf at pH 2 is illustrated in Figure 10b. Since the simulated pH is considerably lower than the isoelectric point, the protein

appears, as expected, mostly positively charged. Furthermore, there appears to be an area of particularly highly positive surface potential in the region comprised among the amino terminus and the amino groups of Lys 28 and Lys 50. This region is a strong candidate for the site of ANS binding at pH 2, as further discussed in the next section.

DISCUSSION

The phenomenon of protein hyperthermostability, i.e. the ability of proteins from hyperthermophilic organisms to be in their folded state at temperatures near 100°C, is not well understood at present, and its study is made more challenging by the experimental problems associated with working at temperatures near and above 100°C in aqueous buffer. In addition, the phenomenon may also depend on contributions from unfolded forms of the proteins which are hard to examine experimentally (Dill & Shortle, 1991). The origins of thermal stability may be addressed by the comparison of the native and perturbed states of a protein at ambient temperature, and more quantitative information may be gained by evaluating all the parameters in the relevant thermodynamic cycle. This is feasible provided that all the steps involved are reversible. Destabilization of a protein at a given temperature by a chemical or physical perturbation, though, does not necessarily imply a decrease in T_m ; conversely, a lower T_m is not a sufficient condition to guarantee decreased stability at room temperature. This arises from the intrinsic shape of the stability curve which is determined, in turn, by the enthalpy difference between the folded and unfolded states at $T=T_m$ (ΔH_m), the

FIGURE 2-10: Surface potential images of RdPf at pH 7 (top images) and pH 2 (bottom images). For each pH value, the protein is represented by two images that correspond to opposite sides of the molecule. Surfaces are color-coded as follows: red, white and blue represent region of negative, zero and positive charge density, respectively. All the negatively and positively charged groups have been labeled in part (a). The prefixes of the amino acid labels represent the following: OD and OE, oxygens of the side chain carboxylates of Asp and Glu, respectively; NZ, nitrogen of the side chain amino group of Lys; OXT, oxygens of the carboxyl terminus; N, nitrogen of the amino terminus.



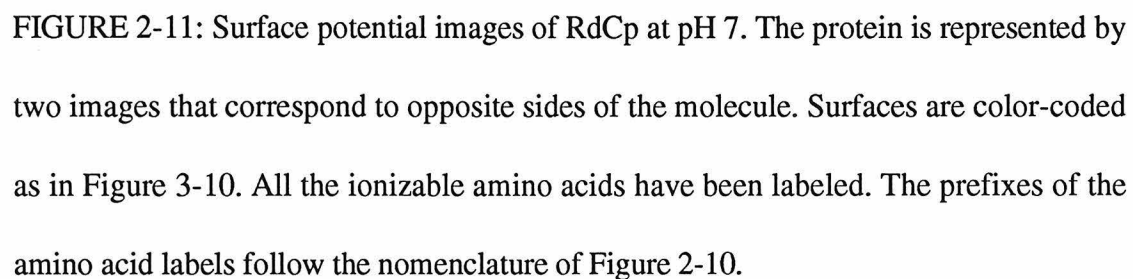
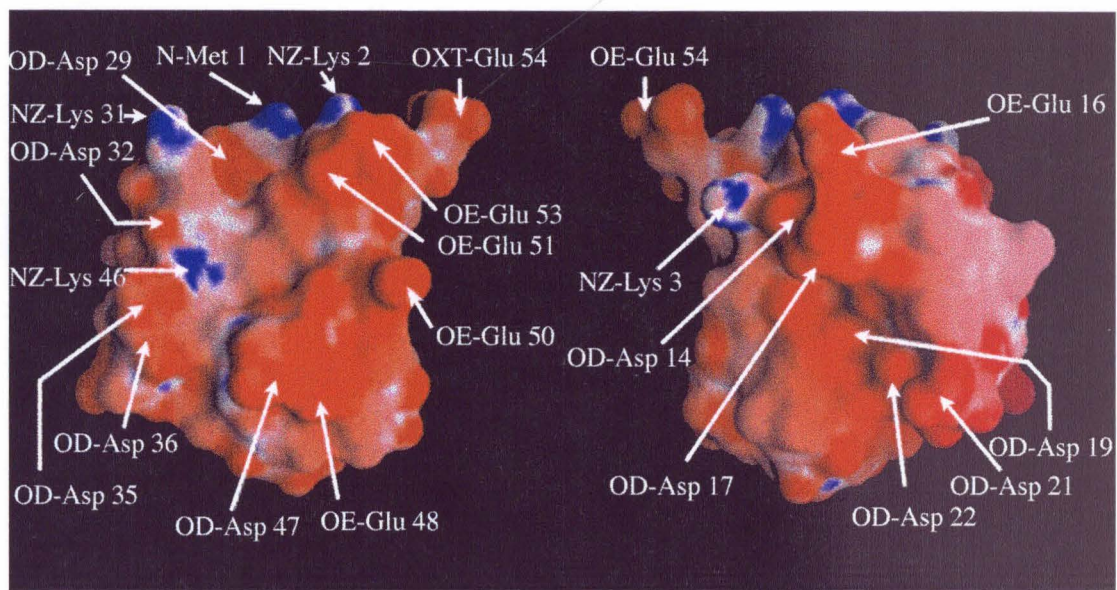


FIGURE 2-11: Surface potential images of RdCp at pH 7. The protein is represented by two images that correspond to opposite sides of the molecule. Surfaces are color-coded as in Figure 3-10. All the ionizable amino acids have been labeled. The prefixes of the amino acid labels follow the nomenclature of Figure 2-10.



heat capacity difference between the two states (ΔC_p), and T_m (Becktel & Schellman, 1987). For a two-state process and for ΔC_p independent of temperature, the stability curve is represented by:

$$\Delta G(T) = \Delta H_m (1 - T / T_m) - \Delta C_p (T_m - T + T \ln (T / T_m)) . \quad (2-5)$$

However, when a protein is examined under different experimental conditions, such as those arising from pH variations, one can assume that the free energy of stabilization at ambient temperature ($\Delta G(25^\circ\text{C})$) is proportional to T_m , provided some conditions are met. The fact that the enthalpy of thermal unfolding (ΔH_m) and the melting temperature (T_m) at low pH are generally lower than that at neutral pH, as well as the relative independence of ΔC_p on pH usually lead to a stability curve at pH close to neutral which lies above the curve at lower pH. As long as one deals with proteins which satisfy the aforementioned conditions, one can make comparisons between melting temperatures and qualitatively relate them to structural features at room temperature. However, it is necessary to keep in mind that the validity of these assumptions in the case of hyperthermophilic proteins would have to be experimentally verified by accurate calorimetric studies. In particular, it has been seen that, for most of the non-thermophilic proteins that have been studied by calorimetry, ΔC_p is, to a first approximation, temperature independent. However, recent evidence coming from the study of BPTI (Makhatadze et al., 1993) shows that ΔC_p decreases considerably at higher temperatures, introducing some asymmetry in the stability curve. This may also

turn out to be true for hyperthermophilic proteins. Under the experimental conditions of this study, the thermal denaturation of RdPf turned out to be irreversible. We postulate that this is due to processes that take place at high temperature upon denaturation. However, as long as the heating rates are not too slow compared with the rate of the irreversible step, a qualitative comparison of experimental melting temperatures (T_{ms}) that are sufficiently far apart is meaningful.

The totality of the experimental evidence presented herein from absorption spectroscopy, CD, NMR, and Trp fluorescence emission indicates that RdPf is still folded at pH 2. Both the secondary structure of the molecule as well as the environments around the metal and the aromatic amino acids including Trp 3, Trp 36, and Leu 32 remain unchanged between pH 7 and pH 2. These observations document the exceptional resistance of the protein to very low pH conditions. Although experiments at even lower pH were undertaken as part of this study, these results are not reported here as the observations obtained at pH lower than 2 may be compromised by peptide bond hydrolysis.

The ANS binding data have allowed us to refine our picture of the structure of RdPf by revealing that, at pH 2 and low ionic strength, a folded form with enhanced accessibility to hydrophobic groups is generated. This structure differs from what has traditionally been called a molten globule (Kuwajima, 1989), since there is no substantial disruption of tertiary interactions as probed by near UV circular dichroism and absorption spectroscopy. Moreover, the chemical shift dispersion in the proton NMR spectrum remains fully preserved. This low-pH pseudo-native state of RdPf is

unique in that it is even more structured than the recently discovered highly ordered molten globular forms of interleukin-4 (Redfield et al., 1994). In this case, the ANS binding was accompanied by changes in the near UV CD spectrum, even though a small change in the proton chemical shift dispersion was not observed.

In the present case of RdPf at pH 2, the ANS binding was found to be greatly diminished at 1M NaCl. This result suggests that the structural change that causes ANS binding is, at least partially, of electrostatic origin. In order to study whether the surface charge density distribution might have any relationships to hyperthermostability and the observed features of ANS binding, we have calculated the surface potential of RdPf according to the Poisson-Boltzmann equation. The amino acid charges at either pH 7 or pH 2, consistent with the results from the pH titration data (see below), have been incorporated into the calculations. The overall increased density of positive charge at low pH, a mere consequence of the protonation state of RdPf and its low isoelectric point, is surprisingly accompanied by the creation of a well defined region of extremely high and positive charge density (see upper region of left image of Figure 10b). We propose that this is the site responsible for the initial interaction of the negatively charged ANS with the protein. This idea is supported by the observed 1:1 binding stoichiometry and the ionic strength dependence of the binding. We propose that this event is followed by the penetration of ANS into the hydrophobic core of the protein. The accessibility of the hydrophobic core might be explained in terms of a slight "opening" of the protein surface driven by the elevated charge repulsion at the putative binding site. The high surface charge density at pH 7 makes RdPf very hydrophilic and

therefore explains its high water solubility. The elevated positive surface charge density at low pH is also consistent with the increased thermal stability at higher ionic strengths which was observed under those conditions. In the absence of salt bridges, the dominant contribution to the electrostatic free energy comes from the positive surface charge, which can be effectively shielded in the presence of salt. This may, in fact, be a simple trick played by Nature in order to increase the thermal stability of soluble proteins whose habitat has very high intracellular salt concentrations. It has been found, for instance, that elevated potassium salt concentrations (about 0.5 M) are present in the cells of the archaeon *Pyrococcus woesei* (Scholz et al., 1992; Zwickl et al., 1990).

The pH titration data at low ionic strength indicate that at pH 2 all the titratable groups have been protonated. This implies that the low pH state of RdPf no longer has any ion pairs at pH 2. This is an important observation, since it shows that ion pairs do not play a significant role in the remarkable thermal stability of the pH 2 form of RdPf. Thermal denaturation data do indicate that the pH 2 form of RdPf has a lower apparent T_m than a pH 8.25 state (Klump et al., 1994; Klump, 1994) by 42°C. The observed apparent decreased thermostability at low pH may well be accounted for by the disruption of all the ion pairs observed in RdPf at the higher pH: Glu 14 to Ala 1 (amino terminus), Glu 30 (and/or Asp 34) to Lys 28 and Lys 45, Glu 47 (and/or Glu 49) to Lys 6. Additionally, Glu 14 is also thought to form a hydrogen bond with the indole nitrogen of Trp 3. However, protonation of the carboxylate of Glu 14, which has definitely occurred at pH 2, does not result in sizable changes in the Trp fluorescence emission profile, which is believed to be a sensitive probe of the Trp environment.

The correlation between decrease in thermostability and disruption of ion pairs suggests that these ion pairs play some role in making a protein hyperthermostable. However, our experiments have shown that RdPf is folded at room temperature and pH 2. This is a remarkable observation, as typical proteins at pH 2 (Privalov, 1979), including metalloproteins such as cytochrome *c* (Goto et al., 1990; Stellwagen & Babul, 1975; Babul & Stellwagen, 1972), zinc finger peptides (Krizek et al., 1991), lactoferrin and transferrin (Mazurier & Spik, 1980) are either denatured or in a molten globular state. A known exception is given by some disulfide-containing proteins such as lysozyme (Privalov & Khechinashvili, 1974) and bovine pancreatic trypsin inhibitor (Makhatadze et al., 1993). However, it is generally believed that their high thermal and pH stability is mainly due to the presence of the disulfides, which provide a covalent structural constraint. Upon reduction, these proteins become fully unfolded at pH 2 and room temperature (Privalov et al., 1989). RdPf, as all the rubredoxins produced by non-hyperthermophilic organisms, has no disulfides. On the other hand, here the four conserved cysteine sulfurs are engaged in tetrahedral iron coordination. This is a common feature of all the rubredoxins and, to a first approximation, one would expect the free energy contribution given by the tetracoordinate iron to be fairly similar for different species. It is possible that the high resistance to low pH is a common characteristic of rubredoxins and it is related to the presence of the iron-sulfur core, rather than being a peculiar property of RdPf. Indirect evidence about mesophilic rubredoxins comes from a report that RdCp is present as a distribution of holo- and apo-forms when analyzed by electrospray ionization mass spectrometry under conditions

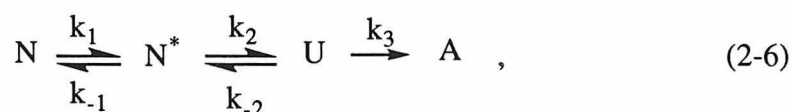
when the protein exists as a positive ion (pH=2.4; Petillot et al., 1993). The fact that the apparent "melting" temperature of RdPf is lower at pH 2 than at pH 8.25 suggests that ion pairs play a role in the thermal stability. However, their contribution cannot be dominant over other forces since the protein still has an unusually high T_m at pH 2. The observed apparent T_m is, in fact, higher than that of typical non-disulfide containing mesophilic proteins, including metalloproteins, under low pH conditions.

Analysis of Figure 10a shows that, at neutral pH, both the negative and positive charges of RdPf are concentrated on one side of the molecule, allowing all the above cited ion pairs to be clustered in close proximity of each other. The non-thermophilic rubredoxins that have been analyzed also have a number of ion pairs relatively close to each other (Figures 1-6 and 1-7). Furthermore, there are no striking qualitative differences that allow definite conclusions to be made on the charge density distribution of the two classes of proteins.

The above conjectures are consistent with the proposal, based on mutation studies on lysozyme from bacteriophage T4 (Daopin et al., 1991), that the enthalpic gain realized by the addition of an ion pair to a mesophilic protein is, at least partially, counterbalanced by the entropic loss due to decreased degrees of freedom of the side chain, with resulting small free energy advantage. More generally, the fact that electrostatic interactions are not the dominant force of protein folding has been discussed previously (Dill, 1990). It could well be that the importance of ion pairing in hyperthermophilic protein stability has been overemphasized in the past, at least in the sense that these interactions are not the dominant factors affecting hyperthermostability.

The present study represents one of the first attempts to address this issue experimentally. However, these thermodynamic considerations may not pertain to the situation under study here.

In the course of this work, we have noted that the denaturation of RdPf is irreversible. The denaturation process may be described by the following general kinetic model:



where N represents the native state, N* is an intermediate, U is the unfolded form, and A is a possible stable unfolded-like state to account for additional processes (aggregation or chemical modification) if the overall unfolding event is irreversible. A hyperthermophilic metalloprotein such as RdPf might undergo disruption of all the noncovalent interactions not involving the iron in the first step, and disruption of the iron-sulfur interactions in the second step. If these steps are sufficiently rapid at temperatures approaching T_m , the concentration of N* may be sufficiently negligible that $N \rightleftharpoons U$ approximates a two-state process, provided that the unfolding is reversible (i.e. k_3 is small). The difference between mesophilic and thermophilic rubredoxins could merely reflect the magnitudes of the rate constants k_1 and k_2 . For instance, k_1 could be smaller due to a much higher kinetic barrier for the first step in the case of RdPf. In this case the thermal transition would be under kinetic control, i.e. it

would be characterized by a relatively high kinetic or activation barrier that could only be "crossed" to a significant extent at temperatures close to the apparent T_m . It is possible that lowering the solution pH causes the height of this barrier to go down. This would result in a lower apparent T_m . Such a scenario could arise from the disruption of the structural constraints exerted by the ion pairs. Nothing can be said at this point about the height of the kinetic barrier associated with the second step, except that it is not exceedingly high for RdPf at low pH. The kinetic experiments presented in chapter 3 will shed more light on these interesting issues.

The unfolding experiments undertaken in the present study were typically performed by bringing the pH to 2 and then heating the protein solution. On the other hand, when we performed the experiments by first bringing the temperature up to T_m and then lowering the pH stepwise, we noticed some hysteresis (see Results section). These observations are consistent with the nature of the denaturation mentioned above.

CONCLUDING REMARKS

The data obtained here provide the first example of a detailed characterization of the highly ordered low-pH state of a hyperthermophilic protein. The results obtained should have implications for the understanding of the origins of the hyperthermostability of RdPf. Ion pairing seems to play a role but it does not appear to be the main contributing factor to the hyperthermostability. This conclusion may not be true for RdPf alone and may apply to other thermophilic proteins as well. Contributions due to

other forces need to be examined and should shed further light on this intriguing problem.

REFERENCES

- Adams M.W.W. (1990) *FEMS Microbiol. Rev.* 75, 219.
- Adams, M.W.W. (1993) *Ann. Rev. Microbiol.* 47, 627.
- Adams, M.W.W., & Kelly, R.M. (1992) in *Biocatalysis at Extreme Temperatures, Enzyme Systems Near and Above 100°C* (Adams, M.W.W. & Kelly, R.M., Ed.), 1-3, ACS Symposium Series 498.
- Adams, M.W.W. (1990) *FEMS Microbiol. Rev.* 75, 219.
- Aono, S., Bryant, F.O., & Adams, M.W.W. (1989) *J. Bacteriol.* 171, 3433.
- Babul, J. & Stellwagen, E. (1972) *Biochemistry* 11, 1195.
- Becktel, W.J., & Schellman, J.A. (1987) *Biopolymers* 26, 1859.
- Bertini, I., & Luchinat, C. (1986) in *NMR of Paramagnetic Molecules in Biological Systems*, pp 19-46, The Benjamin/Cummings Publishing Company, Inc., Menlo Park.
- Blake, P.R., Lee, B., Summers, M.F., Park, J.-B., Zhou, Z.H., & Adams, M.W.W. (1994) *New J. Chem.* 18, 387.
- Blake, P.R., Park, J.-B., Zhou, Z.H., Hare, D., Adams, M.W.W., & Summers M.F. (1992a) *Protein Science* 1, 1508.
- Blake, P.R., Lee, B., Summers, M.F., Adams, M.W.W., Park, J.-B., Zhou, Z.H., & Bax, A. (1992b) *J. Biomol. NMR* 2, 527.
- Blake, P.R., Park, J.-B., Adams, M.W.W., & Summers, M.F. (1992c) *J. Am. Chem. Soc.* 114, 4931.

- Blake, P.R., Park, J.-B., Bryant, F.O., Aono, S., Magnuson, J.K., Eccleston, E., Howard, J.B., Summers, M.F., & Adams, M.W.W. (1991) *Biochemistry* 30, 10885.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem.* 4, 187.
- Bryant, F.O., & Adams, M.W.W. (1989) *J. Biol. Chem.* 264, 5070.
- Constantino, H.R., Brown, S.H., & Kelly, R.M. (1990) *J. Bacteriol.* 172, 3654.
- Cowan, D.A., Smolenski, K.A., Daniel, R.M., & Morgan, H.W. (1987) *Biochem. J.* 247, 121.
- Daopin, S., Nicholson, H., Baase, W.A., Zhang, X.J., Wozniak, J.A., & Matthews, B.W. (1991) *Ciba Found. Symp.* 161, 52.
- Day, M.W., Hsu, B.T., Joshua-Tor, L., Park, J.-B., Zhou, Z.H., Adams, M.W.W., & Rees, D.C. (1992) *Protein Science* 1, 1494.
- Dill, K.A. (1990) *Biochemistry* 29, 7133.
- Dill, K., & Shortle, D. (1991) *Annu. Rev. Biochem.* 60, 795.
- Eaton, W.A. & Lovensberg, W. (1973) in *Iron-Sulfur Proteins* (Lovenberg W., Ed.), pp 131-162, Academic Press, New York and London.
- Eggen, R., Geerling, A., Watts, J., & de Vos, W.M. (1990) *FEMS Microbiol. Lett.* 71, 17.
- Fiala, G., & Stetter K.O. (1986) *Arch. Microbiol.* 145, 56.
- Flam, F. (1994) *Science* 265, 471.

- George, S.J., van Elp, J., Chen, J., Ma, Y., Chen, C.T., Park, J.-B., Adams, M.W.W., Searle, B.G., de Groot, F.M.F., Fuggle, J.C., & Cramer, S.P. (1992) *J. Am. Chem. Soc.* 114, 4426.
- Goto, Y., Calciano, L.J., & Fink, A.L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 573.
- Inubushi, T., & Becker, E.D. (1983) *J. Magn. Res.* 51, 128.
- Juszczak, A., Aono, S., & Adams, M.W.W. (1991) *J. Biol. Chem.* 266, 13834.
- Kelly, R.M., Baross, J.A., & Adams, M.W.W. (1994) *Chem. in Brit.* 30 (7), 555.
- Klotz, I.M., & Hunston, D.L. (1971) *Biochemistry* 10, 3065.
- Klump, H.H., Adams, M.W.W., & Robb F.T. (1994) *Pure & Appl. Chem.* 66 (3), 485.
- Klump, H.H. (1994) unpublished results.
- Koch, R., & Zabłowski, P. (1990) *FEMS Microbiol. Lett.* 71, 21.
- Kolb, D.A., & Weber, G. (1975) *Biochemistry* 14, 4471.
- Krishnamoorthi, R., Markley, J.L., Cusanovich, M.A., & Przysiecki, C.T. (1986) *Biochemistry* 25, 50.
- Krizek, B.A., Amann, B.T., Kilfoil, V.J., Merkle, D.L., & Berg, J.M. (1991) *J. Am. Chem. Soc.* 113, 4518.
- Kuwajima, K. (1989) in *Proteins: Structure, Function and Genetics* 6, 87.
- Lundberg, K.S., Shoemaker, D.D., Short, J.M., Sorge, J.A., Adams, M.W.W., & Mathur, E. (1991) *Gene* 108, 1.
- Makhatadze, G.I., Kim, K.-S., Woodward, C., & Privalov, P.L. (1993) *Protein Science* 2, 2028.

- Matsumura, M., Becktel, W.J., Levitt, M., & Matthews, B.W. (1989a) *Proc. Natl. Acad. Sci. U.S.A.* 86, 6562.
- Matsumura, M., Signor, G., & Matthews, B.W. (1989b) *Nature* 342, 291.
- Matthews, B.W., Nicholson, H., & Becktel, W.J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 6663.
- Mazurier, J., & Spik, G. (1980) *Biochim. Biophys. Acta* 629, 399.
- Mukund, S., & Adams, M.W.W. (1990) *J. Biol. Chem.* 265, 11508.
- Nicholls, A., Sharp, K., & Honig, B. (1991) in *Proteins: Structure, Function and Genetics II*, 281.
- Nicholson, H., Becktel, W.J., & Matthews, B. W. (1988) *Nature* 336, 651.
- Peisach, J., Blumberg, W.E., Lode, E.T., & Coon, M.J. (1971) *J. Biol. Chem.* 246, 5877.
- Petillot, Y., Forest, E., Mathieu, I., Meyer, J., & Moulis, J.-M. (1993) *Biochem. J.* 296, 657.
- Phillips, W.D., Poe, M., Weiher, J.F., Mc Donald, C.C., & Lovenberg, W. (1970) *Nature* 227, 574.
- Pihl, T.D., & Maier, R.J. (1991) *J. Bacteriol.* 173, 1839.
- Privalov, P.L., Tiktopulo, E.I., Venyaminov, S.N., Griko, Y.V., Makhatadze, G.I., & Khechinashvili, N.N. (1989) *J. Mol. Biol.* 205, 737.
- Privalov, P.L. (1979) *Adv. Prot. Chem.* 33, 167.
- Privalov, P.L., & Khechinashvili, N.N. (1974) *J. Mol. Biol.* 86, 665.
- Redfield, C., Smith, R.A.G., & Dobson, C.M. (1994) *Nature Str. Biol.* 1, 23.

- Robb, F.T., Park, J.-B., & Adams, M.W.W. (1992) *Biochim. Biophys. Acta* 1120, 267.
- Scholz, S., Sonnenbichler, J., Schäfer, W., & Hensel R. (1992) *FEBS Lett.* 306, 239.
- Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky, V.N., Gripas, A.F., & Gilmanshin, R.I. (1991) *Biopolymers* 31, 119.
- Stellwagen, E., & Babul, J. (1975) *Biochemistry* 14, 5135.
- Stetter, K.O., Fiala, G., Huber, G., Huber, R., & Segerer, G. (1990) *FEMS Microbiol. Rev.* 75,117.
- Stetter, K.O. (1986) in *The Thermophiles: General, Molecular, and Applied Microbiology* (Brock, T.D., Ed.), pp 39-74, John Wiley, New York.
- Tanford, C. (1962) *Adv. Prot. Chem.* 17, 69.
- Watenpaugh, K.D., Sieker, L.C., & Jensen, L.H. (1980) *J. Mol. Biol.* 138, 615.
- Watenpaugh, K.D., Sieker, L.C., & Jensen, L.H. (1979) *J. Mol. Biol.* 131, 509.
- Woese, C.R., Kandler, O., & Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576.
- Zwickl, P., Fabry, S., Bogedain, C., Haas, A., & Hensel, R. (1990) *J. Bacteriol.* 172, 4329.

CHAPTER 3.

**Arrhenius Analysis of the Unfolding of Rubredoxins
from *Pyrococcus Furiosus* and *Clostridium*
Pasteurianum: A New Role for Ion Pairs in Proteins**

ABSTRACT

The temperature dependence of the unfolding kinetics of rubredoxins from the hyperthermophile *Pyrococcus furiosus* (RdPf) and the mesophile *Clostridium pasteurianum* (RdCp) has been studied. Results show that RdPf unfolds much more slowly, under all experimentally accessible temperature regimes, than typical mesophilic proteins, including RdCp. This introduces the possibility that the RdPf is kinetically trapped at physiological temperatures. The most direct implication of this hypothesis is the possibility that protein hyperthermostability can be explained based on kinetic, rather than thermodynamic, arguments.

The transition state for the step involving Fe^{3+} loss has been characterized as being "loose" and with a hydrophobic character similar to that of the native state. Ion pairs, whenever present, participate to the unfolding kinetics by further slowing down the unfolding rates. This is presumably due to a surface-clamping effect exerted by the salt bridges on the native structure. Ion pairs give a contribution to the unfolding kinetics of both mesophilic and thermophilic rubredoxins; therefore, these findings may be of general significance for proteins.

INTRODUCTION

Nature has equipped proteins with an organized three-dimensional structure that enables them to perform important biological functions, which include signalling,

transport and catalysis. These require a machinery with a considerable degree of conformational flexibility, that can only be achieved by reversible breakage/formation of relatively weak non-covalent interactions, triggered by small energetic inputs. It is not surprising, then, that the thermodynamic stability of proteins is only moderate, and it does not usually exceed 15 kcal/mol, with melting temperatures typically ranging between 40 and 60°C (chapter 1, part 1.2). A less well understood situation occurs in the case of a selected number of microorganisms, commonly referred to as hyperthermophiles, that have the peculiar ability to thrive at temperatures up to 110°C. While it is not clear, at present, what causes the exceptional *in vivo* and *in vitro* thermostability displayed by the proteins produced by hyperthermophiles, it is known that heat resistance does not minimally compromise their ability to efficiently perform catalytic and transport functions (Adams, 1993). Therefore, the thermodynamic stability of these proteins at their physiological temperatures is not necessarily higher than that of their mesophilic counterparts. Most hyperthermophiles are classified as *Archaea* (Woese et al., 1990), and they live in the vicinity of submarine hydrothermal vents and continental hot springs or solfataric fields. A more extensive description of the living environment and biodiversity of hyperthermophiles is presented in chapter 1.

An excellent model system for the study of the unusual properties of these proteins is the iron-protein rubredoxin from *Pyrococcus furiosus* (RdPf), a microorganism with an optimal growth temperature of 100°C (Blake et al., 1991; Woese et al., 1990). The overall structural features of RdPf in the crystalline form are

very similar to those of rubredoxins from mesophilic organisms, except for a small number of extra ion pairs and hydrogen bonds (Day et al., 1992, chapter 1). Complete disruption of the ion pairs, at low pH, generates a fully folded structure with a somewhat lower apparent melting temperature (Cavagnero et al., 1995, chapter 2). Since there is not, at present, a satisfactory explanation of the features responsible for the thermostability, we have decided to investigate if, and to what extent, kinetic arguments could play any role in this matter.

MATERIALS AND METHODS

RdPf has been isolated and purified as previously described (Blake et al., 1991). Rubredoxin from *Clostridium pasteurianum* (Watenpaugh et al., 1973; Yasunobo & Tanaka, 1983) was purchased from SIGMA. Concentrated stock solutions of the two proteins were used for the experiments described below. The kinetics of unfolding were monitored by UV-vis absorption spectroscopy, with a Hewlett-Packard 8452A single-beam diode array spectrophotometer. Full spectra were recorded per each time point. Spectra were typically analyzed at 380 nm. A 1 cm quartz cuvette was used. The temperature was controlled via a Haake D1 thermostatted water bath connected to the compartment surrounding the cuvette housing. The experiments were performed in aqueous buffer under vigorous stirring, at different temperatures. Buffer solutions were placed in the spectrophotometer cuvette and allowed to reach thermal equilibrium at the desired temperature. The

temperature of the solution inside the cuvette was carefully measured with a resistor sensor. After the buffer solution had reached thermal equilibrium, a concentrated RdPf solution was rapidly added and spectral acquisition started. The dead time of this apparatus, inclusive of the time required for mixing and, supposedly, temperature equilibration, is 1 to 2 seconds. No denaturing agents were employed. Buffers consisted of 40 mM glycine solutions at pH 2.0 or 2.7, and 40 mM either tris (hydroxymethyl) aminomethane or 3-(N-morpholino) propanesulfonate solutions at pH 7.0. All of the spectroscopic data were baseline-corrected. This was done for each time point by subtracting experimental absorbances from the corresponding absorbances at 700 nm, where no contribution is expected from RdPf transitions. The absorption spectra were analyzed on an IBM-compatible computer, with the software Lab Calc (Galactic Industries Corporation). Non linear least squares curve fitting was done on a Macintosh computer with the program Kaleidagraph (Synergy Software). The following expression was used to fit the data recorded at 380 nm:

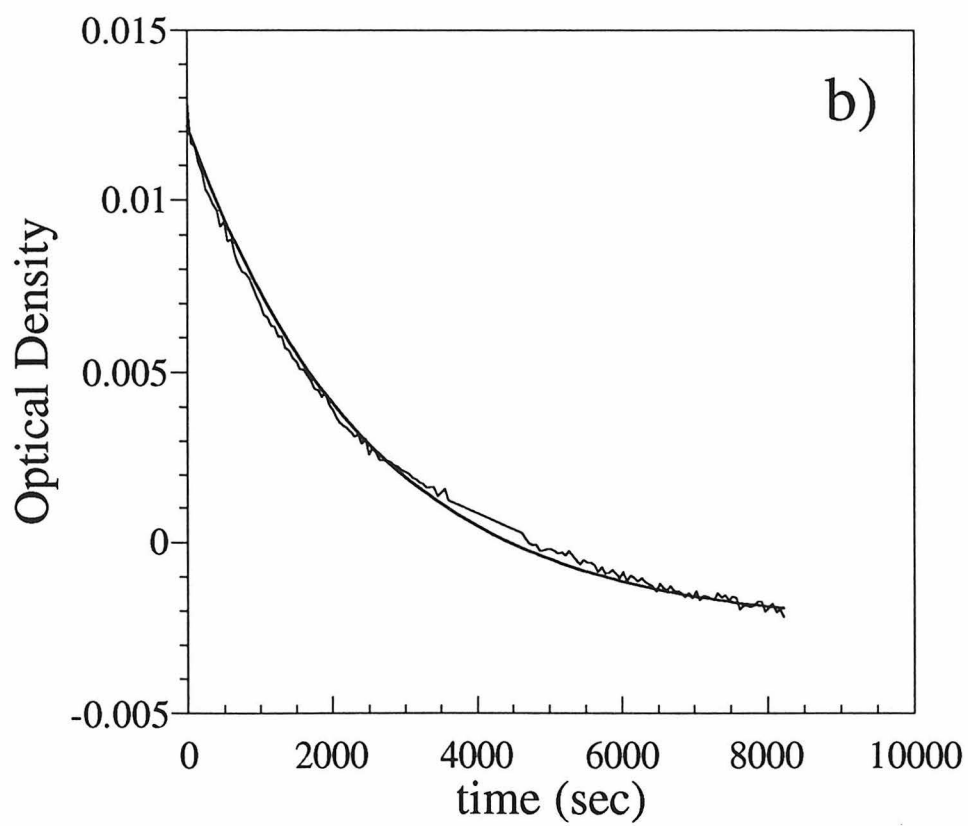
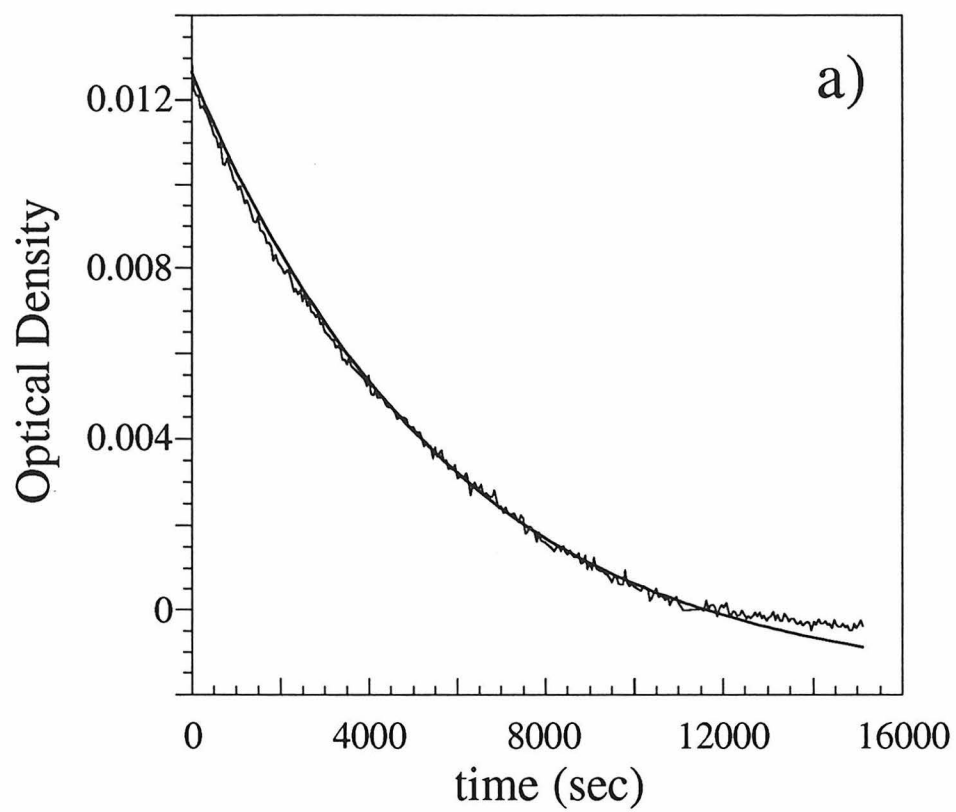
$$OD_{OBS} = OD_0 \exp[-kT] + a \quad (3-1)$$

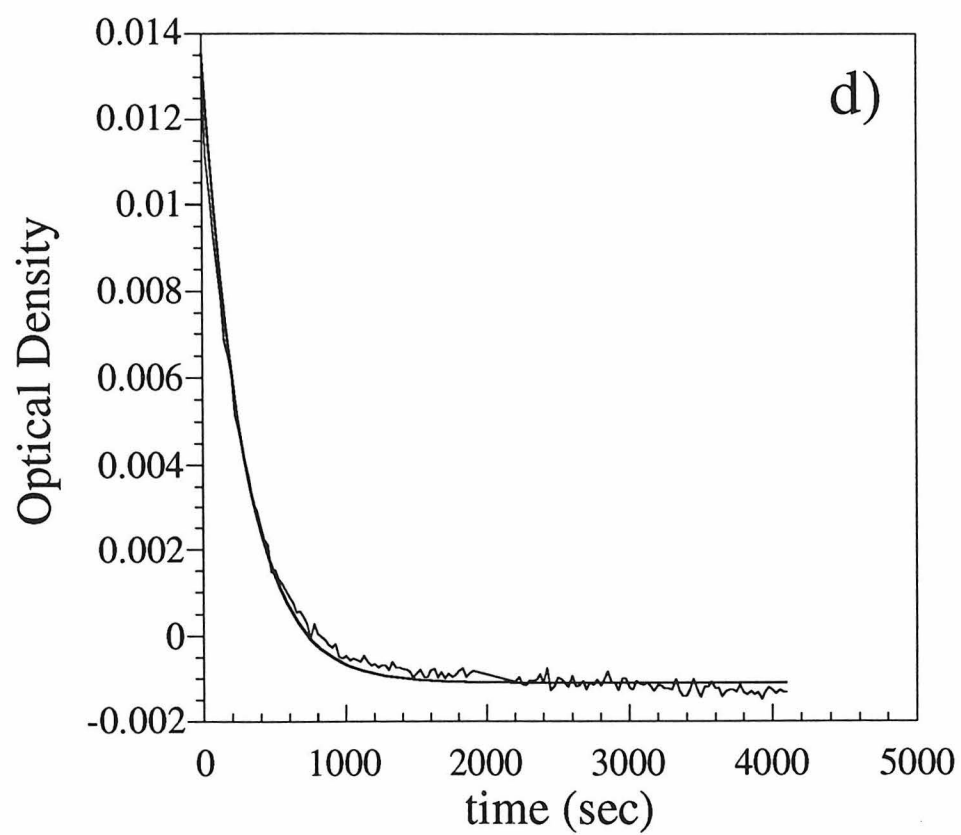
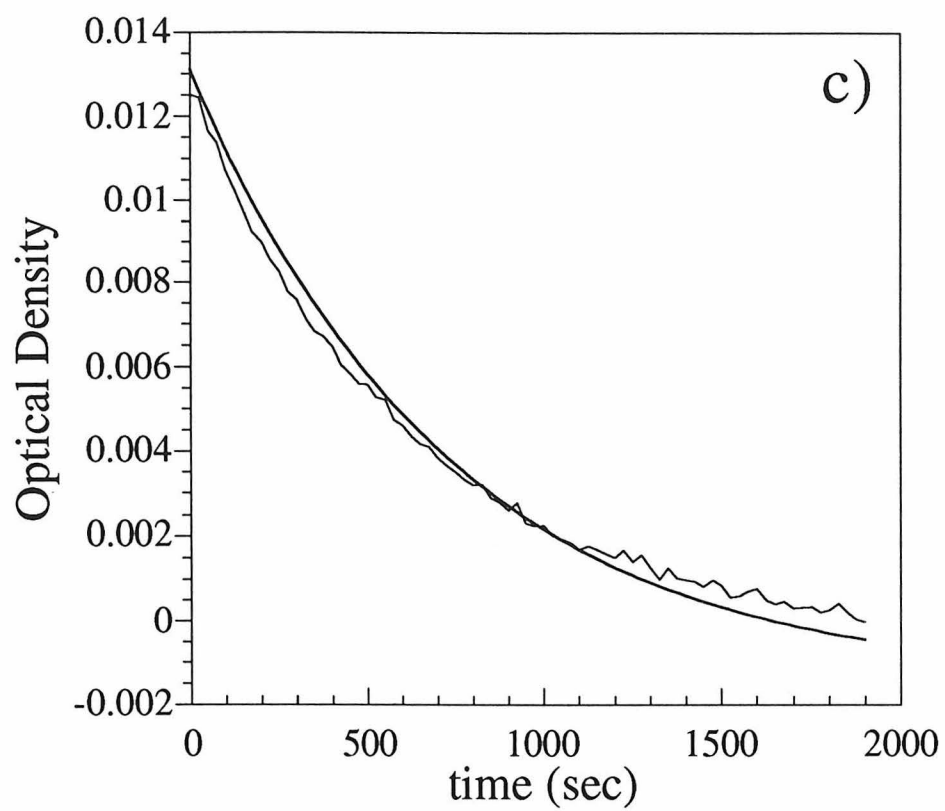
where OD_{OBS} represents the observed absorption signal, OD_0 is the absorption signal at time zero, and T is the absolute temperature. The unfolding rate constant k and the constant a , which has been introduced to account for small baseline errors, have been treated as adjustable parameters.

RESULTS

In the temperature range that has been analyzed (62-96°C), the rates of unfolding of RdPf relative to the step involving Fe^{3+} , are extremely slow. Fig. 3-1 shows the temperature dependence of the time-decay of the Cys $\rightarrow$ M charge-transfer absorption bands of RdPf at pH 2. The curves follow monoexponential kinetics, which suggest that Fe^{3+} is lost in a single step. First-order rate constants are readily obtained (Table 4-1), and the Eyring plot (Eyring, 1935) gives an excellent fit (Fig. 3-2). A comparison between RdPf and the available published data for other proteins is presented in Fig. 3-3. After extrapolating the RdPf curve to the temperature range relevant to mesophilic species, it becomes apparent that the unfolding rates of RdPf are considerably slower (about three orders of magnitude) than those relative to the non-thermophilic proteins for which data are available: tryptophan synthase (Chen & Matthews, 1994), lysozyme (Segawa & Sugihara, 1984), stem bromelain (Arroyo-Reyna & Hernandez-Arana, 1995), chymotrypsin inhibitor 2 (Jackson & Fersht, 1991) and trypsin (Pohl, 1968). In order to ensure a fair comparison among different proteins, some of the original data from the literature have been extrapolated to zero guanidine hydrochloride concentration. This was only possible in cases where the authors had studied the guanidine hydrochloride dependence of the kinetic parameters (Chen & Matthews, 1994; Jackson & Fersht, 1991). Additionally, it is important to notice that our kinetic data derive only from measurements of the unfolding kinetic step involving iron loss. However, data presented in the next chapter strongly suggest

FIGURE 3-1: Temperature dependence of RdPf unfolding kinetics as followed by UV-vis absorption spectroscopy at 380 nm, pH 2.0. Kinetic traces are shown with the corresponding curve fits. The temperatures for the different traces are: a) 62.4°C, b) 66.6°C, c) 71.0°C, d) 75.5°C, and e) 76.7°C. First-order rate constants were determined by non-linear regression analysis, assuming a two-state model.





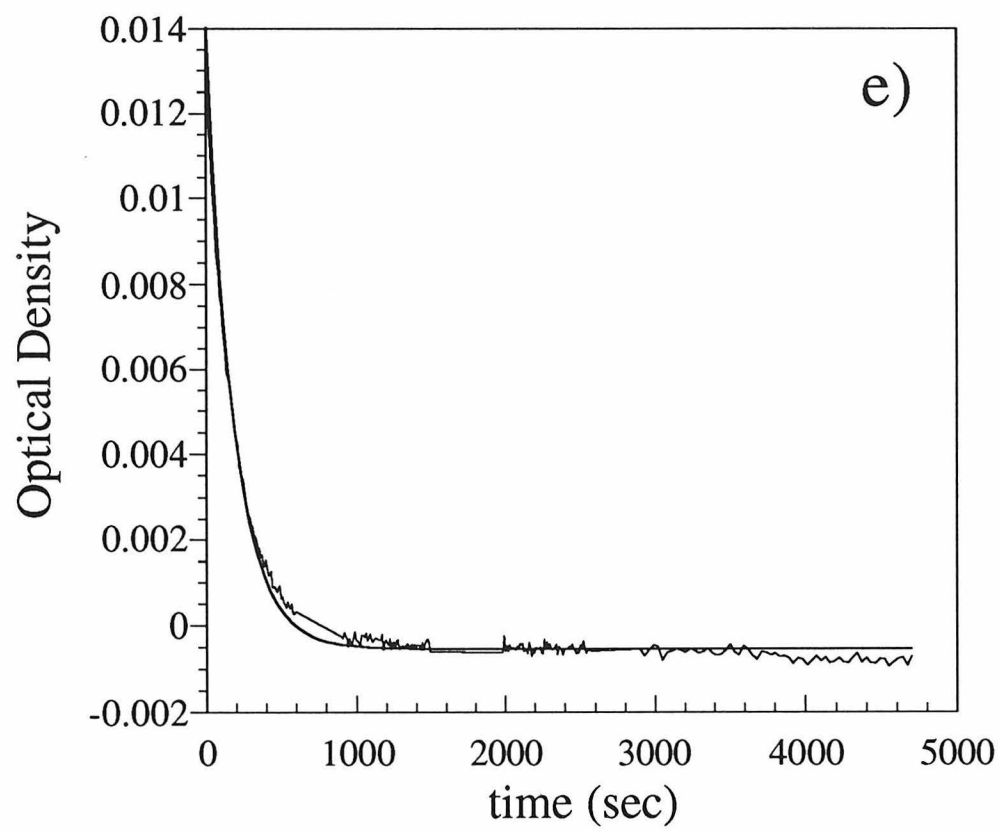


FIGURE 3-2: Eyring plot of the RdPf unfolding kinetics, as followed by UV-vis absorption spectroscopy at 380 nm.

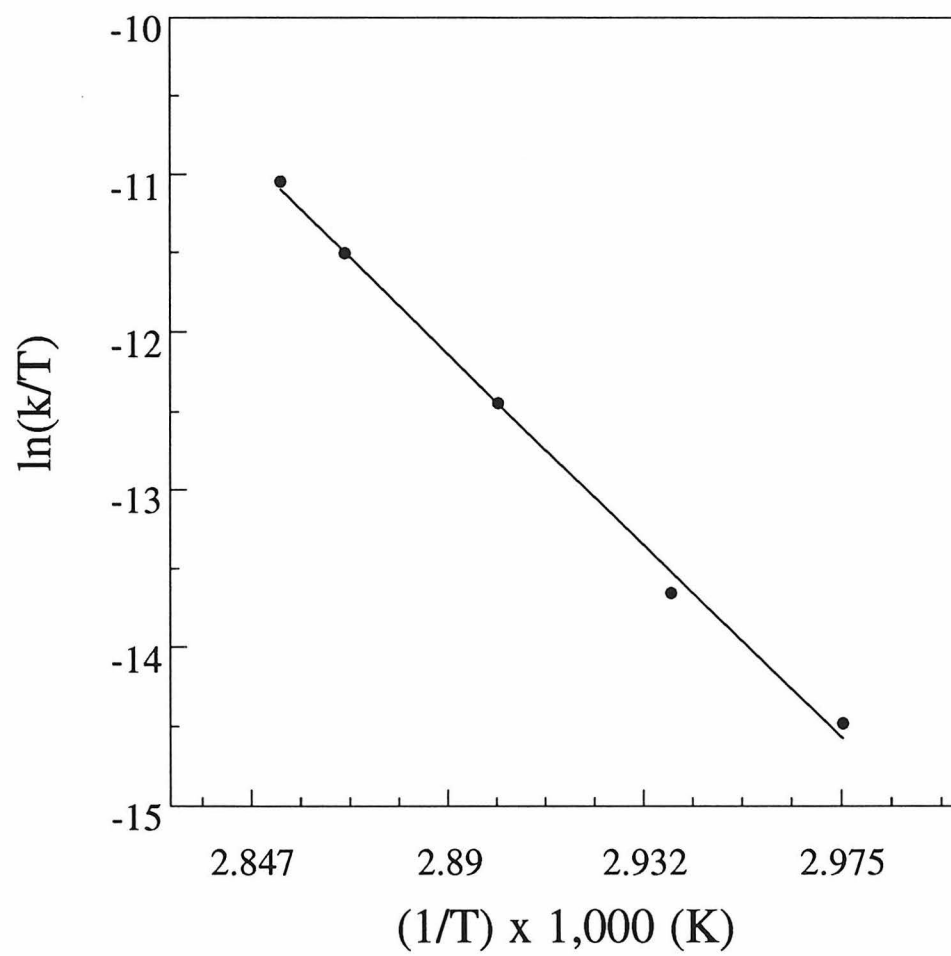
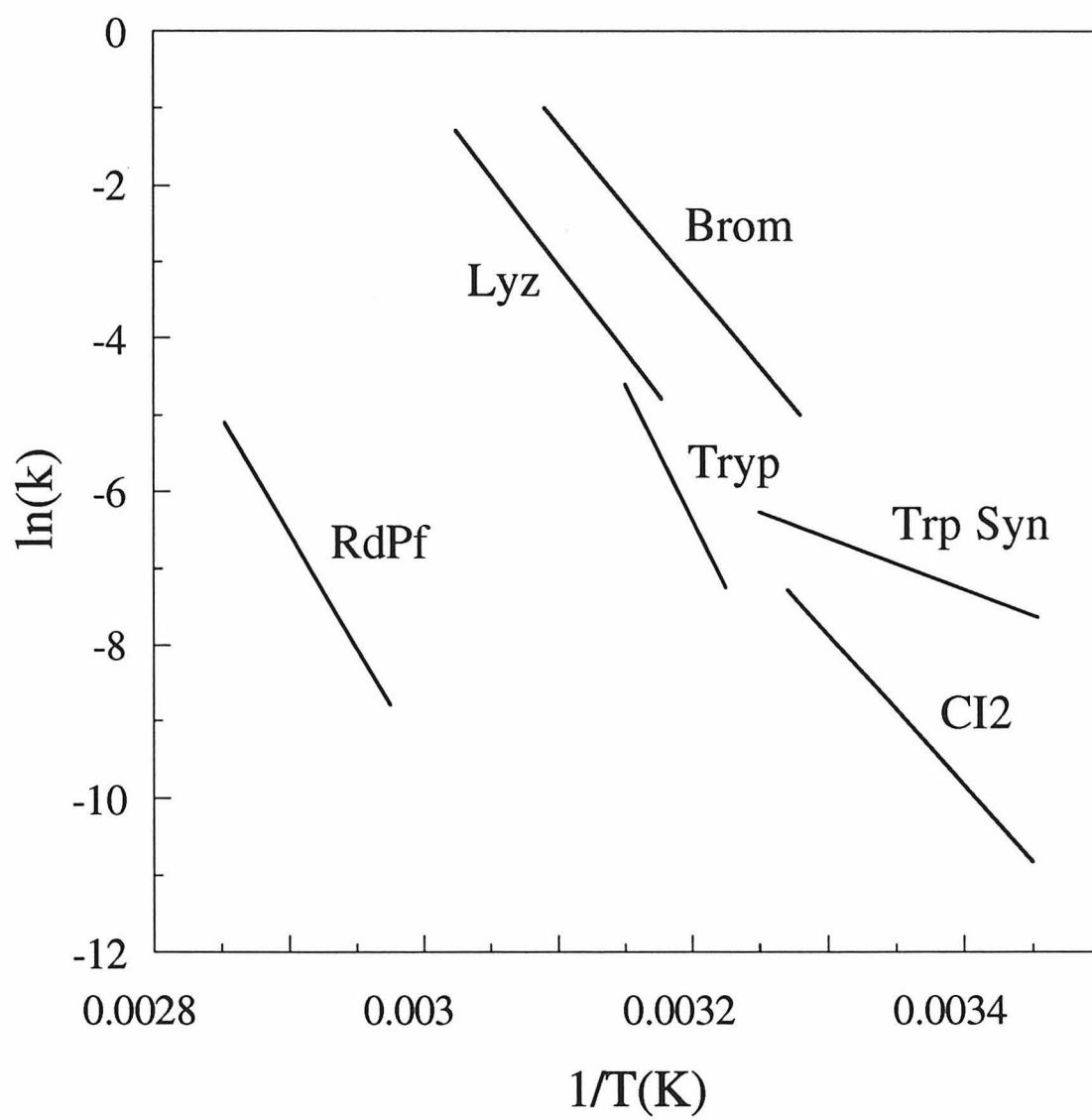


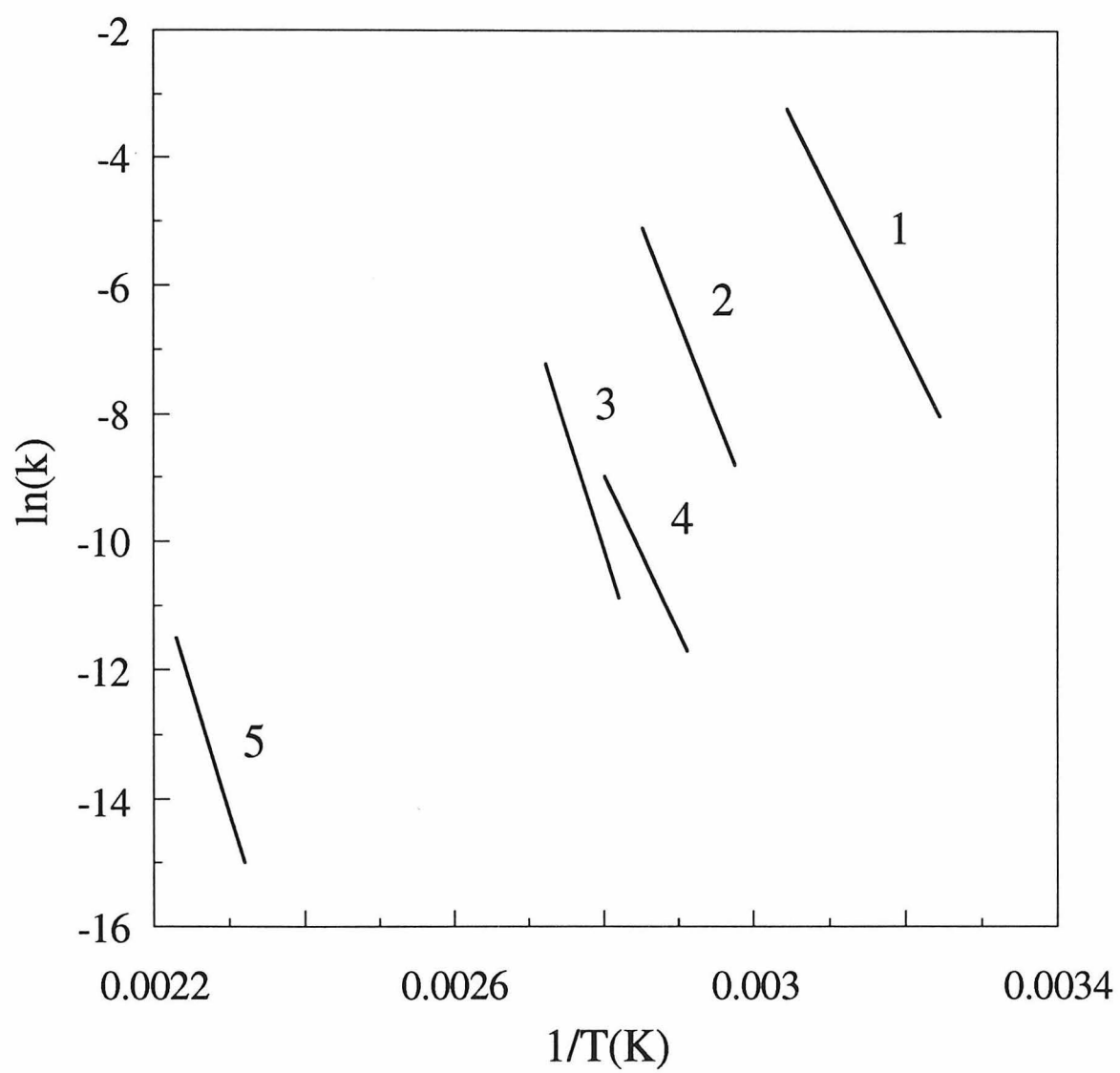
FIGURE 3-3: Arrhenius plots for the unfolding of RdPf and a few mesophilic proteins. RdPf, pH 2.0, as probed by absorption spectroscopy at 380 nm; Lyz, lysozyme, 0.7 M GnCl; Brom, stem bromelain, pH 3.4; Tryp, trypsin, pH 2; Trp Syn, α -subunit of Trp synthase, data extrapolated to zero M GnCl; CI2, chymotrypsin inhibitor 2, data extrapolated to zero M GnCl.



that this step, although intrinsically very slow, is not the rate-determining step of the unfolding process. The rate-determining step is even slower and it occurs later during unfolding. This further supports our results, and suggests that the rate of iron loss defines an upper limit to the intrinsic unfolding rates. RdPf unfolding was irreversible, under our experimental conditions. The unfolding kinetics were wavelength independent, over the wavelength range of absorption by the charge-transfer bands.

The pH dependence of the unfolding rates of RdPf and rubredoxin from the mesophilic microorganism *Clostridium pasteurianum* (RdCp) as a function of temperature are illustrated by the Arrhenius plots (Arrhenius, 1889) in Fig. 3-4. As pH is increased, the Arrhenius curves keep approximately the same slope but shift to the left, implying slower unfolding rates at a given temperature. This applies both to hyperthermophilic and mesophilic rubredoxins. In the case of RdCp, which possesses 2 ion pairs (Spasov et al., 1995) at neutral pH, the shift in going from pH 2 to pH 7 is quite pronounced, and it is ascribed to the ion pairs which are formed as the pH is raised. Similarly, the Arrhenius plot of RdPf shifts to the left upon varying the pH from 2 to 2.7. This pH change accounts for the protonation of a single negatively-charged amino acid, based on the electrometric pH titration curve of RdPf (Cavagnero et al., 1995, chapter 2 Figure 2-9). The RdPf curve shifts to the left, upon varying the pH from 2.0 to 2.7, by about half of what the RdCp curve shifts when the pH is changed from 2.0 to 7.0. This strongly supports the idea that the RdPf amino acid which gets protonated in going from pH 2 to pH 2.7 is part of a salt bridge. Furthermore, given the extent of curve shift per ion pair and assuming that the

FIGURE 3-4: Arrhenius plots for the unfolding of RdPf and RdCp at different pH values. 1, RdCp, pH 2.0; 2, RdPf, pH 2.0; 3, RdPf, pH 2.7; 4, RdCp, pH 7.0; 5, RdPf, data extrapolated to pH 7. The experimental curves are the result of the study of the temperature dependence of the unfolding kinetics. by absorption spectroscopy, at 380 nm.



Arrhenius curve shifts linearly with the number of salt bridges, it is possible to predict the approximate position of the Arrhenius unfolding curve for RdPf at pH 7 (Fig. 3-4). This curve has been generated by taking into account that RdPf has four ion pairs, according to the crystal structure (Day et al., 1992; Spassov et al., 1995). All of the salt bridges are located on the protein surface.

One can estimate the unfolding half-lives of RdCp and RdPf, at different pH values, at 100°C (Table 3-1), which is the optimal growth temperature of *Pyrococcus furiosus*. Similar data can be derived for other mesophilic proteins for which Arrhenius studies have been published (Table 3-1). As Table 3-1 shows, at 100°C the half-life of all the proteins, except for RdPf at pH 7, is very short, ranging from tens of microseconds to seconds. In striking contrast, RdPf has an approximate half-life of 1582 years, at this temperature. This has important implications for the origins of hyperthermostability, as will be further discussed in the next section.

In order to confirm the validity of our results, we have repeated the unfolding experiments a number of times and derived the error relative to the data collected under similar experimental conditions. In general, data collected at higher pH were affected by greater uncertainties, presumably because unfolding gets exceedingly slow (up to several days at the temperatures that we have studied) as the pH is raised. Very slow processes are less likely not to be highly reproducible. A compilation of the unfolding activation parameters, derived according to both Arrhenius and transition state theory, and their experimental errors is presented in Table 3-2.

TABLE 3-1: Estimated unfolding half-lives ($t_{1/2}$) of RdPf and some mesophilic proteins.

Protein	Experimental Conditions	$t_{1/2}$ extrapolated to 100°C
Tryptophan Synthase ^{a)} (α -subunit)	0 M GnCl ^{b)}	2 sec
Stem Bromelain ^{c)}	pH 3.4	6×10^{-4} sec
Trypsin ^{d)}	pH 2	5×10^{-5} sec
Lysozyme ^{e)}	0.7 M GnCl	6×10^{-3} sec
Chymotrypsin Inhibitor 2 ^{f)}	0 M GnCl ^{b)}	10^{-2} sec
RdCp	pH 2.0	2×10^{-3} sec
RdCp	pH 7.0	3 min
RdPf	pH 2.0	0.3 sec
RdPf	pH 2.7	2 min
RdPf	pH 7 ^{g)}	1582 years

a) Chen & Matthews, 1994.

b) Extrapolated from GnCl concentration dependence studies.

c) Arroyo-Reyna & Hernandez-Arana, 1995.

d) Pohl, 1968.

e) Segawa & Sugihara, 1984.

f) Jackson & Fersht, 1991.

g) Extrapolated from pH dependence studies (see Fig. 3-4).

TABLE 3-2: Activation parameters for rubredoxin unfolding, according to both Arrhenius and transition-state theory. Experimental errors are expressed as standard deviation of the mean.

Protein	E_a [kcal mol ⁻¹] (exp. error)	A [sec ⁻¹] (exp. error)	$\Delta H^\ddagger$ [kcal mol ⁻¹] (exp. error)	$\Delta S^\ddagger$ [cal mol ⁻¹ deg ⁻¹] (exp. error)	$\Delta G^\ddagger$ [kcal mol ⁻¹]
RdPf - pH 2.0	61 (3)	4×10^{36} (4×10^{36})	60 (3)	101 (7)	25
RdPf - pH 2.7	74	5×10^{40}	73	125	28
RdCp - pH 2.0	51 (2)	4×10^{33} (4×10^{33})	50 (2)	88 (7)	23
RdCp - pH 7.0	49 (5)	2×10^{32} (2×10^{32})	49 (5)	60 (12)	28

DISCUSSION

We have seen in the previous section that the unfolding kinetics of RdPf is considerably slower than in the case of several mesophilic proteins (Fig. 3-3), including RdCp (Fig. 3-4, Table 3-1). More specifically, the unfolding half-lives of RdPf and RdCp differ by two orders of magnitude at pH 2. This difference is dramatically magnified at pH 7, where the unfolding half-lives of RdCp and RdPf are 3 min and 1582 years, respectively (Table 3-1). These results introduce the possibility that the hyperthermostability of RdPf is due to kinetic, rather than thermodynamic, factors. In fact, given its exceedingly long half-life, RdPf may be kinetically trapped at temperatures close to 100°C, the optimal growth temperature of *Pyrococcus furiosus*. Unfolding may then take place only at higher temperatures, which would allow the solvent bath to transfer the required amount of energy to the protein. This proposal represents a viable alternative to the traditional thermodynamic considerations, which have not been able to convincingly explain the origins of the hyperthermostability of RdPf so far.

It is interesting to notice that, since the experiments performed by Anfinsen in the sixties and seventies (Anfinsen, 1973), protein folding has generally been assumed to be under thermodynamic control (Dill, 1989; Dill, 1990; Kim & Baldwin, 1990). However, there have been a number of instances where the existence of kinetic traps has been unequivocally identified, either along the folding pathways, as in the case of lysozyme (Kiefhaber, 1995), or in pseudo-native states, such as the pro-region-

depleted form of α -lytic protease (Baker et al., 1992). The occurrence of kinetically trapped misfolded species is also well documented during folding *in vivo*. Nature has solved this problem by having chaperones interact with the misfolded species and, in so doing, lower the kinetic barriers for refolding. This then allows proteins to "escape" from the traps (Ellis & Hartl, 1996). The same phenomenon is also frequently encountered in RNA folding, where specific RNA chaperones help resolve misfolded forms (Coetzee et al., 1994; Herschlag, 1995; Herschlag et al., 1994). To our knowledge, however, no one has ever shown before that the native state itself could be trapped to prevent the protein from unfolding. A direct implication of our hypothesis is that either the denatured state of RdPf is much more unstable than for mesophilic proteins, or refolding is nearly as slow as unfolding. This immediately prompts the intriguing question of how folding could then take place *in vivo*, as RdPf gets synthesized by the ribosomal machine. While a few answers are possible, further kinetic and thermodynamic studies will be needed in order to settle this issue. The present studies suggest that RdPf may be trapped in its native state at pH 7, at temperatures at least up to 100°C. Furthermore, our findings definitely show that kinetics ought to be considered in order to fully understand protein hyperthermostability.

Closely related to the above issues, is the observed pH dependence of the Arrhenius plots of both RdPf and RdCp (Fig. 3-4). In the interpretation of our results, we have taken advantage of the fact that changes in pH can be correlated to the protonation state of the ionizable surface residues through electrometric pH titration

curves (Tanford, 1961; Tanford, 1962). Additionally, we have considered that protonation of any negatively charged residue involved in an ion pair destroys it.

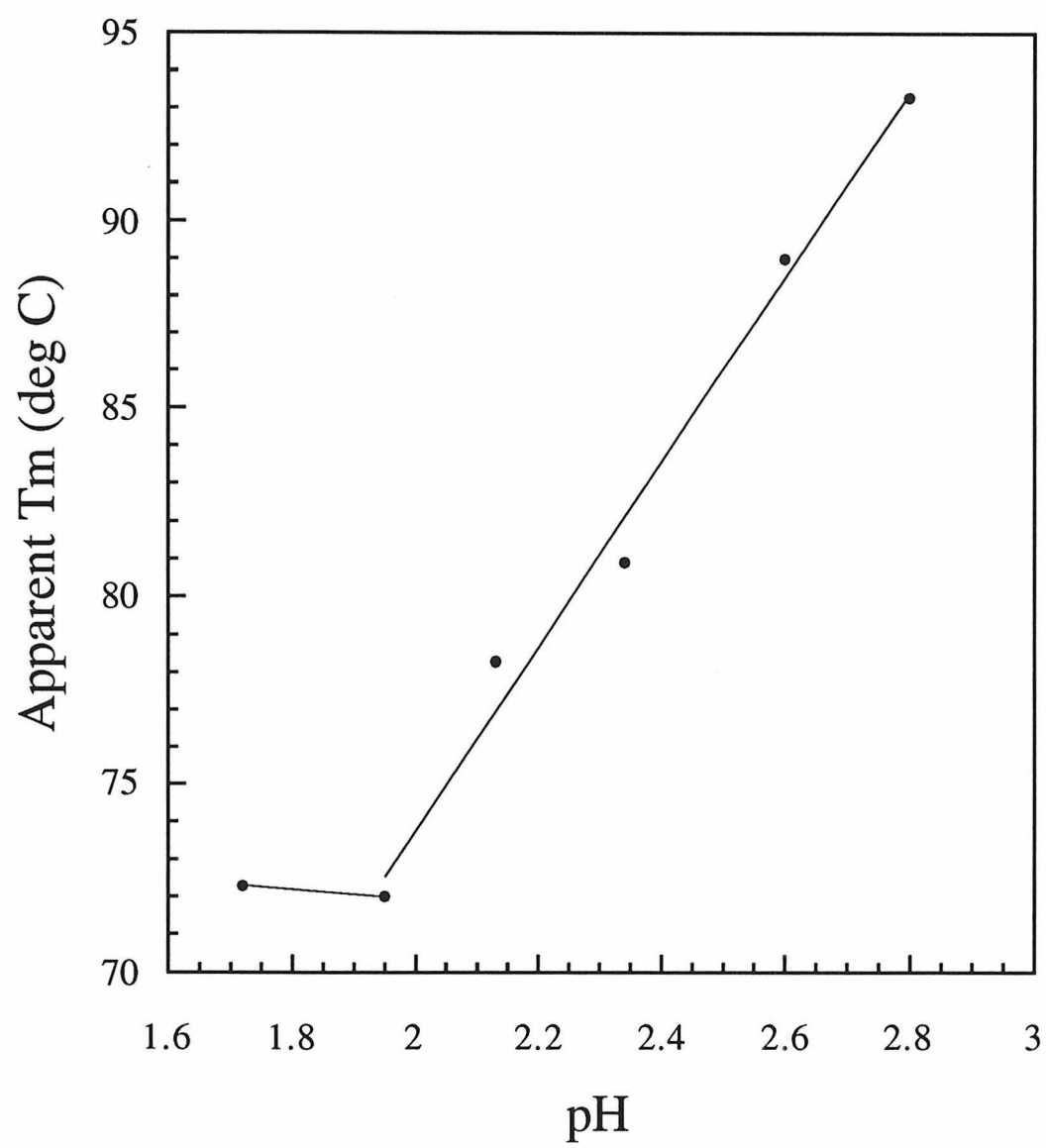
The electrostatic free energy of a protein (see appendix to this chapter) is essentially composed of a nonspecific term and a specific term. The nonspecific term (Honig & Nicholls, 1995a; Kauzmann, 1954; Linderstrom-Lang, 1924; Tanford, 1961) comprises the electrostatic energy contributions which arise when a protein has a net charge density; this term is proportional, to a first approximation, to the square of the net charge. The specific term explicitly takes into account interactions such as localized salt bridges, and their interactions with other charged and polar groups in the rest of the protein. We have not performed detailed electrostatic free energy calculations. However, based on the following considerations, we estimate that the nonspecific electrostatic energy components of RdPf and RdCp do not affect the validity of our results. Consequently, we have interpreted our data only in terms of the contributions given by the salt bridges. From the primary sequence (Blake et al., 1991) and electrometric pH titrations (Cavagnero et al., 1995) of RdPf we know that the protein is positively charged at pH 2 (net charge = + 6) and 2.7 (net charge = +5), and that it is negatively charged at pH 7 (net charge = -8). Therefore, the nonspecific electrostatic free energy component will be somewhat higher at pH 7 than at lower pH values. This implies that there will be a greater extent of intramolecular electrostatic repulsion at pH 7. It follows that this effect alone would tend to speed up the unfolding kinetics, at higher pH. Since our data indicate an opposite trend, we conclude that the nonspecific electrostatic free energy component, if appreciable,

would partially mask the effects due to the ion pairs. Our conclusions about the ion pairs would then be only reinforced in the presence of this effect. Similar considerations apply to RdCp. The above also applies when stability is analyzed in thermodynamic terms, in that a protein is expected to be more stable at a pH where the net charge is lower.

We have also worked under the assumption that forces of nonelectrostatic nature are affected by pH changes by a much smaller extent than electrostatic forces. This view is corroborated by the fact that the experimental apparent melting temperature of RdPf decreases linearly with pH up until protonation of all the negatively-charged amino acid side chains (i.e., at pH 2; see Fig. 2-9), and then it flattens out at lower pH values (Fig. 3-5). It is important to notice that, as ionic strength is increased, salt bridges are expected to get weaker even at high pH, making it very hard to draw any conclusions. We kept the ionic strength down to 40 mM in all of our pH dependence studies. A final point to be noted is that forces such as ion-dipole, dipole-dipole, dipole-induced dipole, induced dipole-induced dipole, are all essentially electrostatic in origin (Maitland et al., 1981). Additionally, hydrogen bonds are also believed to be of partially electrostatic nature (Maitland et al., 1981). Therefore, all of these noncovalent forces are expected to display some pH dependence. Nonetheless, we predict that this is going to be much weaker than in the case of ion pairs and classical electrostatic interactions. Therefore, we have neglected to consider those forces, to a first approximation.

Our studies on the unfolding pH dependence can be summarized as follows: a) Arrhenius plots relative to the thermal denaturation of both RdPf and RdCp have similar slopes; however, the curves are shifted with respect to each other. b) Given the same pH value, the Arrhenius plot of RdCp is always displaced to the right of the RdPf curve, indicating faster unfolding rates for the mesophilic protein. c) Examination of the RdCp curves at pH 2 and pH 7 strongly suggests that each added salt bridge contributes a given left-shift to the Arrhenius curve. This shift can be readily evaluated, knowing that RdCp contains four salt bridges at pH 7 (Spasov et al., 1995). d) The Arrhenius left-shift per salt bridge approximately corresponds to the RdPf curve shift in going from pH 2 to pH 2.7. Knowing that this increase in pH accounts for the deprotonation of a single carboxylate (Cavagnero et al., 1995), we deduce that this carboxylate is likely to belong to a salt bridge. e) The observations in part c) and d) allow to estimate the expected Arrhenius plot for RdPf at pH 7. This curve is remarkably shifted to the left, compared to any mesophilic protein, including RdCp, regardless of the pH value of the mesophilic protein. This indicates that RdPf is expected to have much slower unfolding kinetics than its mesophilic counterparts, at neutral pH. The estimated half-lives of mesophilic and hyperthermophilic proteins at 100°C (Table 3-1) give a more quantitative illustration of this concept. f) RdPf is expected to have a half-life greater than 1000 years at 100°C, at pH 7. Therefore, we propose that RdPf is kinetically trapped in physiological conditions. The origin of the hyperthermostability of RdPf may well be of kinetic nature, since our arguments do not require thermodynamics to play a role.

FIGURE 3-5: pH dependence of the apparent melting temperature (T_m) of RdPf.



The pH-dependent differences in kinetic behavior can be ascribed to a surface "clamping" effect exerted by the salt bridges on the native structure. By providing localized structural constraints on the protein surface, ion pairs restrict the number of rotational and vibrational normal modes of the protein. This results in decreased unfolding rates.

There has been much discussion in the literature about the general role of ion pairs in a protein (Daopin et al., 1991a; Daopin et al., 1991b; Hendsch & Tidor, 1994; Sali et al., 1991), and it is currently believed that salt bridges contribute to the thermodynamic stability by only a small extent, due to enthalpy-entropy compensation effects. In the specific case of RdPf, breakage of the salt bridges is known to lower the apparent melting temperature by several degrees (Cavagnero et al., 1995). Within the context of protein folding, however, ion pairs evidently play an important kinetic role that had not been pointed out before. Since the above applies to both hyperthermophilic and mesophilic rubredoxins alike, this effect may be of general significance for proteins. The fact that the unfolding rates of RdCp and RdPf are different at pH 2 indicates that these rates are determined to a certain extent by structural elements other than the ion pairs, since these are no longer present at this pH value. Conversely, the observation that the unfolding rates for the two proteins are expected to diverge so dramatically at pH 7, where RdCp and RdPf have approximately two and four ion pairs, respectively, indicates that the two additional ion pairs of RdPf have a surprisingly pronounced effect on its kinetic properties at high pH.

Within the temperature range analyzed in this study, the Arrhenius plots of both RdPf and RdCp are linear, suggesting that there is no detectable heat capacity change in going from the native state to the activated complex (Oliveberg et al., 1995). This supports the notion that the activated complex does not have a greater number of exposed hydrophobic groups than native RdPf (Privalov & Makhatadze, 1990), a concept which is fairly in line with what has generally been found for other proteins (Pohl, 1968; Segawa & Sugihara, 1984).

The curves in Fig 3-3 have similar slopes, which suggests that the low rate constants of RdPf are governed by low Arrhenius preexponential factors (i.e. low entropies of unfolding), compared to mesophilic proteins. However, given the high experimental error associated with the Arrhenius preexponential factors (Table 3-2), one can only say that the slower unfolding rates are due to a combination of lower preexponential factors and/or higher activation energies, or, alternatively, lower activation entropies and higher activation enthalpies. Analysis according to transition state theory yields a $\Delta H^\ddagger$ of 60 kcal mol⁻¹ and a $\Delta S^\ddagger$ of 101 cal mol⁻¹ K⁻¹. The calculated $\Delta G^\ddagger$ value is 25 kcal mol⁻¹ for the unfolding step involving iron loss at pH 2 (Table 3-2). Both activation enthalpy and entropy are very large, being about four times as high as typical unimolecular gas-phase dissociations and isomerizations of small molecules, that take place at even higher temperatures than those relative to the present study.

The recent implementation of an overexpression system for RdPf in *E. coli* (M.W.W. Adams, unpublished results) will soon allow mutational analysis to be

performed on this system. It will then be possible to ask more detailed questions about the kinetic role of individual salt bridges, in relationship to their ability to modify the protein topology and the character of its vibrational normal modes. While this study has focused on kinetic analysis, it is hoped that future experimental efforts will also be addressed to investigate whether thermodynamics plays in the hyperthermostabilization. This can be done by calorimetry, which depends on the large-scale availability of a hyperthermostable protein whose folding/unfolding transition is reversible.

CONCLUDING REMARKS

These studies have definitely shown that the thermal unfolding of RdPf at neutral pH is several orders of magnitude slower than in the case of mesophilic proteins. It is thus viable that the hyperthermostability of RdPf is kinetically controlled. Ion pairs play a special role in the unfolding kinetics by slowing down the unfolding of both RdPf and RdCp. This effect arises from the introduction of non-covalent structural constraints on the native protein surface.

APPENDIX

The Electrostatic Free Energy of a Protein

We will now briefly describe how the electrostatic component of the free energy of a protein is calculated. In the present treatment, we will only be concerned with proteins that have the charged amino acid side chains pointing towards the surface, since this case is more pertinent to the study of RdPf. However, it is important to remember that several proteins have amino acid residues with charged side chains which are buried in the hydrophobic core. These are usually associated with oppositely charged groups to form an ion pair. The contribution of electrostatic interactions in the interior of the protein to stability has been studied both from the theoretical (Hendsch & Tidor, 1994) and experimental (Waldburger et al., 1995) points of view. The general treatment of buried electrostatic interactions is similar to that of surface salt bridges. However, especially when crucial approximations or assumptions have to be made, one model may dramatically differ from another in its ability to correctly describe buried electrostatic interactions.

A protein can be visualized as a low-dielectric self-contained medium some of whose atoms bear a given residual charge. At constant temperature and pressure, the free energy consists of the work done on a system. Therefore, the electrostatic contribution to the free energy of a protein is simply the work done in placing charges on it. It should also be understood that the electrostatic free energy difference (ΔG_{el}) is the difference between the work done to put charges on the native state of the protein

and the corresponding work done on the unfolded state. If an infinitesimal amount of charge dq is to be placed at a position of electrostatic potential ψ , the amount of work done is

$$dW_{el} = \psi dq. \quad (3-1A)$$

The electrostatic free energy is then equal to

$$W_{el} = G_{el} = \int \psi dq, \quad (3-2A)$$

where the integral extends to the whole volume of the protein. Since we are just concerned with the surface charges here, we will consider the integral as extended to the protein surface only. The center of the coordinate system is located in the center of mass of the protein. The protein surface charges are considered fixed in space, i.e., their positions are indicated by a fixed set of coordinates. Since the protein surface potential ψ is a function of the precise location on the surface, we will represent it by $\psi(r)$ from now on. The problem is then reduced to finding a suitable expression for $\psi(r)$.

We will assume, for now, that the entire medium has a uniform susceptibility, and therefore, that there is a single dielectric constant, called ϵ . In the case where the charges on the surface are a discrete set of point charges, we have the simple expression derived from Coulomb's Law:

$$\psi(r) = \frac{1}{\epsilon} \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}, \quad (3-3A)$$

where $\mathbf{r}$ is the vector from the center of the coordinate system to any point in space, $\mathbf{r}_i$ is the vector to the location of the surface charge, and ϵ is the dielectric constant of the medium. Sometimes " $4\pi\epsilon_0$ " is used instead of " ϵ ". Because Coulomb's law ignores the finite sizes of ions, it is valid only at distances that are significantly greater than atomic dimensions. The charge of an atom is separated between the nucleus and the diffuse electron cloud, so it cannot be treated as a point charge at short distances. The problem of charge separation is even more severe with proteins, in which the charges of the ionized groups of the side chains of Lys, Arg, His, Asp, and Glu residues, plus the α -amino and α -carboxyl groups, are distributed over two or more hydrogen or oxygen atoms. Therefore, expression 3-3A is not adequate to describe charges in a protein. In order to describe the potential for a continuous distribution of charges on a closed surface, we start from the expression for the surface integral of the normal component of an electric field ($\mathbf{E}$) over a closed surface (S). This expression is the well known Gauss' theorem:

$$\int \mathbf{E} \cdot \mathbf{n} \, da = \sum_i \frac{4\pi q_i}{\epsilon}, \quad (3-4A)$$

where $\mathbf{E}$ is the electric field vector, $\mathbf{n}$ is the unit vector along the outer normal, and "da" is an infinitesimal element of the area of the closed surface S. In the case of a continuum charge distribution on a surface, the Gauss' theorem becomes:

$$\int_S \mathbf{E} \cdot \mathbf{n} \, da = \frac{4\pi}{\epsilon} \int_V \rho \, dv, \quad (3-5A)$$

where the surface charge density ρ indicates the charge per unit surface (in coulombs per square meter in the mks system). The volume integral is over the volume V enclosed by the surface S. After applying the divergence theorem of vector analysis, we have:

$$\int_V \nabla \cdot \mathbf{E} \, dv = \frac{4\pi}{\epsilon} \int_V \rho \, dv, \quad (3-6A)$$

where ∇ is defined as:

$$\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}. \quad (3-7A)$$

This implies that:

$$\nabla \cdot \mathbf{E} = \frac{4\pi \rho}{\epsilon}. \quad (3-8A)$$

Remembering that $\mathbf{E} = -\nabla \psi$, and substituting this expression in eq. 3-8A, we obtain the Poisson equation:

$$\nabla^2 \psi = -\frac{4\pi \rho}{\epsilon} . \quad (3-9A)$$

Eq. 3-9A is of general validity. Essentially, all electrostatic models used in studying macromolecules are based on the Poisson equation. The application of eq. 3-9A to proteins requires two important modifications to be made to take into account a) the fact that the dielectric constant varies with the spatial coordinates, and b) the presence of salts, i.e. of a variable ionic strength, in the solvent. Many different models have been proposed to include these two effects. We will only present the most successful treatments here. When applying eq. 3-9A, it is usually necessary to average an environmental property or response over some region of space. Depending on the spatial resolution at which the electrical response is modeled, a particular treatment is said to be microscopic or macroscopic. Microscopic treatments are most successful in describing electrostatic potentials in the interior of proteins, and we will not discuss them here.

Spatially dependent dielectric constants. A simplified macroscopic treatment of surface electrostatic potentials is due to Tanford and Kirkwood. According to their model, the protein is treated as an idealized spherical medium of low dielectric constant (ϵ_{in}) surrounded by water. Water has a large dielectric constant called ϵ_{out} (Tanford & Kirkwood, 1957). This treatment is very sensitive to the value chosen for

ϵ_{in} , and gross mistakes can easily be made in the description of internal charges. However, acceptable results are usually obtained for surface charges.

The best model that takes into account the variable susceptibility of the medium is commonly referred to as “discretized continuum approach” (Warwicker & Watson, 1982). The solvent and the protein are divided into a three-dimensional grid with a local value of ϵ (i.e., either ϵ_{in} or ϵ_{out}) associated with each grid point. This grid is exploited by solving the Poisson equation in the following form:

$$\nabla^2 \psi(\mathbf{r}) = -\frac{4\pi \rho_f(\mathbf{r})}{\epsilon(\mathbf{r})} \quad (3-10A)$$

by a finite difference method. ρ_f in eq. 3-10A represents the free surface charge density. An improved algorithm based on this method, created by Anthony Nicholls and Barry Honig (Nicholls & Honig, 1991), has proven to be extremely fast and it has recently given very good results (Honig & Nicholls, 1995b). In order to model internal charges as successfully as microscopic methods do, it is necessary to explicitly include the internal polarity of the protein in the treatment.

Dependence on ionic strength. The most widely used treatment that takes the presence of salts into account is Debye-Hückel theory. The univalent mobile salt ions are treated as point charges and enter into the calculation only in that they determine the charge density of the solution around the protein. At distances far removed from the protein the time-averaged charge density and the potential will be zero, since the concentration of positive and negative ions will be equal. In close proximity of the

protein, however, the concentration of positively charged mobile ions will differ from the concentration of negatively charged mobile ions. The concentration of the positively charged ions will be greater when the protein has an overall negative charge, and vice versa. Hückel theory postulates that this ion gradient may be computed by the Boltzmann distribution law, i.e., that the ratio of the concentration of one kind of ion near the protein to its concentration far away from it is given by $\exp[-W_i/(kT)]$, where W_i is the work that must be done in order to move the ion from infinity ($\psi=0$) to the potential ψ prevailing at the point where the concentration is to be computed. We have that $W_i = z\psi$ for a positive ion of charge $+z$ and $W_i = -z\psi$ for a negative ion of charge $-z$, in proximity of the protein surface. The Boltzmann distribution law takes the form $N_+ = N e^{-z\psi/kT}$ and $N_- = N e^{+z\psi/kT}$. N_+ and N_- are the number of negatively charged ions, respectively, N indicates the total number of salt molecules, assuming monovalent salts, and k is the Boltzmann constant. Far removed from the protein surface, where $\psi=0$, we have that $N_+ = N_- = N$. The charge density at any point is given by

$$\rho = N_{+z} - N_{-z} = Nz \left(e^{-z\psi/kT} - e^{+z\psi/kT} \right). \quad (3-11A)$$

We can now do a series expansion of the exponentials in eq. 3-11A:

$$\rho = -2Nz \left[\frac{z\psi}{kT} + \frac{1}{3!} \left(\frac{z\psi}{kT} \right)^3 + \frac{1}{5!} \left(\frac{z\psi}{kT} \right)^5 + \dots \right]. \quad (3-12A)$$

Another important assumption of the Debye-Hückel theory is that the potential is sufficiently small at all points where Poisson's equation is to be used. This implies that

$$\frac{z\psi}{kT} \ll 1 . \quad (3-13A)$$

As a consequence, all but the first term in eq. 3-12A can be dropped, which leaves us with:

$$\rho = -\frac{2Nz^2\psi}{kT} . \quad (3-14A)$$

Relation 3-13A is justified for a central ion with a rather low charge density, and for a sufficiently low concentration of mobile ions. Since charges are relatively dilute in proteins, the protein charge density is usually not a problem. However, it is generally assumed that the univalent salt concentration must be equal or smaller than 0.1 M for 3-13A to be valid. The Poisson equation becomes

$$\nabla^2\psi = \frac{8\pi Nz^2}{\epsilon kT} \psi = \kappa^2 \psi ; \quad (3-15A)$$

κ is the well-known Debye-Hückel parameter, defined as

$$\kappa = \sqrt{\frac{8\pi z^2 N}{\epsilon kT}} . \quad (3-16A)$$

W_{el} is proportional to the square of the protein net charge. We now want to show that the electrostatic free energy of a protein is proportional to the square of its net charge. In order to do this, we consider the solution of eq. 3-14A in a simplified case where the protein surface is approximated by a sphere of radius R . The solution of eq. 3-15A on the surface of the sphere is:

$$\psi = \frac{q}{\epsilon R} \left(1 - \frac{\kappa R}{1 + \kappa a} \right), \quad (3-17A)$$

where a is the distance between the center of the sphere and the center of an ion, at its closest approach distance to the protein. Remembering 3-2A, we can now write an expression for the electrostatic free energy of a protein of net charge nZ :

$$W_{el} = \int_{q=0}^{q=nZ} \frac{q}{\epsilon R} \left(1 - \frac{\kappa R}{1 + \kappa a} \right) dq = \frac{(nZ)^2}{2\epsilon R} \left(1 - \frac{\kappa R}{1 + \kappa a} \right). \quad (3-18A)$$

This is the result we sought. As a consequence, one should expect a protein to have no electrostatic free energy contribution at its isoelectric point, i.e., when its net charge is zero. The above simplified treatment considers the protein as being uncharged in its interior. However, if buried charges are present, this will no longer be a valid assumption. In this case, W_{el} is expected to be different from zero at a pH corresponding to the isoelectric point of the protein.

Modern Treatments. The treatment by Nicholls and Honig (Nicholls & Honig, 1991), which is particularly successful and commonly used, uses the following modified form of the Poisson-Boltzmann equation:

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \psi(\mathbf{r})] - \kappa(\mathbf{r})^2 \sinh(\psi(\mathbf{r})) = -4\pi\rho(\mathbf{r}) , \quad (3-19A)$$

where both a term containing the Debye-Hückel constant and a term containing the charge distribution function $\rho(\mathbf{r})$ are present. According to this treatment, the definition of κ has been slightly modified, to make it dielectric independent:

$$\kappa = \sqrt{\frac{8\pi z^2 N_A I}{1000kT}} , \quad (3-20A)$$

where I is the bulk salt concentration in moles per liter, N_A is the Avogadro number, and z is the charge of an electron in SI units. In fact, the algorithm that computes $\psi(\mathbf{r})$ solves a linearized, simplified version of 3-19A:

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \psi(\mathbf{r})] - \kappa(\mathbf{r})^2 \psi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) . \quad (3-21A)$$

Although the dielectric constant appears to be a general function of distance in eqns 3-19A and 3-21A, calculations are usually performed assuming that ϵ has only two values, one for the solvent medium and another one for the "protein interior" medium.

An additional advantage of the approach by Nicholls and Honig is that it uses a realistic protein surface, as calculated via computer. This surface is generated starting from the three-dimensional coordinates of the protein. This is a big improvement over older calculations based on a spherical surface approximation, since the electrostatic energy is a steep function of the distance between a probe charge and the protein surface. As a consequence, distance errors which arise from significant approximations on the surface shape, will be magnified in electrostatic energy computations.

In order to obtain the electrostatic free energy difference (ΔG_{el}) of a protein, calculated as described above, it is necessary to compute the difference between the folded and the unfolded states. Most treatments use simplified models for unfolded proteins. These models are always somewhat arbitrary, since the structural features of the denatured state are not well known at present.

Ion pairs and other specific electrostatic interactions. The electrostatic free energy difference obtained by the methods outlined above is acceptable only if the non specific effects, simply due to the existence of an asymmetric surface charge density, are to be considered. A more realistic free energy difference also takes into account the specific electrostatic interactions, such as localized salt bridges and polar interactions with the rest of the protein (Hendsch & Tidor, 1994). The calculations to be performed in this case are a little more complicated. They are based on the fragmentation of the electrostatic free energy difference (ΔG_{el}) into different contributions:

$$\Delta G_{\text{el}} = \Delta G_{\text{sol}} + \Delta G_{\text{bridge}} + \Delta G_{\text{protein}} . \quad (3-22A)$$

ΔG_{sol} is the difference between the solvation of the folded and unfolded states. Solvation free energies are calculated by subtracting the electrostatic free energy (calculated, as shown above, by first solving 3-21A, and then applying 3-2A) in water (i.e., the water external dielectric constant is used) from the electrostatic free energy *in vacuo*, in the gas phase. ΔG_{bridge} is due to the specific ion pair electrostatic bonding energy. Here the unfolded state is assumed to give no contribution. Finally, $\Delta G_{\text{protein}}$ takes into account the specific interaction of the surface charges with other polar groups in the protein. Again, the unfolded state is expected to give no contribution here. If one wants, for example, calculate the contribution of ion pairs to the electrostatic free energy of the protein, it is sufficient to calculate $\Delta\Delta G_{\text{el}}$.

$\Delta\Delta G_{\text{el}}$ is defined as follows:

$$\Delta\Delta G_{\text{el}} = \Delta G_{\text{el WT}} - \Delta G_{\text{el HP}} , \quad (3-23A)$$

where $\Delta\Delta G_{\text{el}}$ is the free energy difference between ΔG_{el} for the wild-type protein ($\Delta G_{\text{el WT}}$) and ΔG_{el} for a mutant in which the amino acids involved in the salt bridge(s), whose contribution(s) need to be assessed, have been replaced by their hydrophobic isosteres ($\Delta G_{\text{el HP}}$). Amino acids that only differ by their charged natural

amino acid counterparts by the fact that their charges have been turned down to zero, are called hydrophobic isosteres.

REFERENCES

- Adams, M. W. W. (1993) *Ann. rev. Microbiol.* 47, 627.
- Anfinsen, C. B. (1973) *Science* 181, 223.
- Arrhenius, S. (1889) *Z. Phys. Chem.* 4, 226.
- Arroyo-Reyna, A. & Hernandez-Arana, A. (1995) *Biochim. Biophys. Acta* 1248, 123.
- Baker, D., Sohl, J. L. & Agard, D. A. (1992) *Nature* 356, 263.
- Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. & Adams, M. W. W. (1991) *Biochemistry* 30, 10885.
- Cavagnero, S., Zhou, Z. H., Adams, M. W. W. & Chan, S. I. (1995) *Biochemistry* 34, 9865.
- Chen, X. & Matthews, C. R. (1994) *Biochemistry* 33, 6356.
- Coetzee, T., Herschlag, D. & Belfort, M. (1994) *Genes Dev.* 8, 1575.
- Daopin, S., Nicholson, H., Baase, W. A., Zhang, X. J., Wozniak, B. W. & Matthews, B. W. (1991a) *Ciba Found. Symp.* 161, 52.
- Daopin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991b) *Biochemistry* 30, 7142.
- Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J.-B., Zhou, Z. H., Adams, M. W. W. & Rees, D. C. (1992) *Protein Science* 1, 1494.
- Dill, K. A. (1989) 29, 133.
- Dill, K. A. (1990) *Biochemistry* 29, 7133.

- Ellis, R. J. & Hartl, F.-U. (1996) *FASEB J.* 10, 20.
- Eyring, H. (1935) *Chem. Rev.* 17, 226.
- Hendsch, Z. S. & Tidor, B. (1994) *Protein Sci.* 3, 211.
- Herschlag, D. (1995) *J. Biol. Chem.* 270, 20871.
- Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R. L. (1994) *EMBO J.* 13, 2913.
- Honig, B. & Nicholls, A. (1995a) *Science* 268, 1144.
- Honig, B. & Nicholls, A. (1995b) *Science* 268, 1144.
- Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* 30, 10436.
- Kauzmann, W. (1954) in *The Mechanism of Enzyme Action* (McElroy, W. D. & Glass, B., Eds.) 70, Johns Hopkins Press, Baltimore, MD.
- Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. USA* 92, 9029.
- Kim, P. & Baldwin, R. L. (1990) *Ann. Rev. Biochem.* 59, 631.
- Linderstrom-Lang, K. U. (1924) *C. R. Trav. Lab. Carlsberg* 15, 70.
- Nicholls, A. & Honig, B. (1991) *J. Comp. Chem.* 12, 435.
- Oliveberg, M., Tan, Y.-J. & Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* 92, 8926.
- Pohl, F. M. (1968) *Eur. J. Biochem.* 7, 146.
- Privalov, P. L. & Makhatadze, G. I. (1990) *J. Mol. Biol.* 213, 385.
- Sali, D., Bycroft, M. & Fersht, A. R. (1991) *J. Mol. Biol.* 220, 779.
- Segawa, S.-I. & Sugihara, M. (1984) *Biopolymers* 23, 2473.
- Spassov, V. Z., Karshikoff, A. D. & Ladenstein, R. (1995) *Protein Science* 4, 1516.
- Tanford, C. (1962) *Adv. Protein Chem.* 17, 69.
- Tanford, C. & Kirkwood, J. G. (1957) *J. Am. Chem. Soc.* 79, 5333.

- Waldburger, C. D., Schildbach, J. F. & T., S. R. (1995) *Nature St. Biol.* 2, 122.
- Warwicker, J. & Watson, H. C. (1982) *J. Mol. Biol.* 157, 671.
- Watenpaugh, K. D., Sieker, L. C., Herriot, J. R. & Jensen, L. H. (1973) *Acta Crystallogr. B* 29, 943.
- Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576.
- Yasunobo, K. T. & Tanaka, M. (1983) in *Iron-Sulfur Proteins* (Lovenberg, W., Eds.) 27, Academic Press, New York.

CHAPTER 4.

The Unfolding Mechanism of Rubredoxin from *Pyrococcus Furiosus*

ABSTRACT

As part of our studies on the structural and dynamic properties of hyperthermostable proteins, we have investigated the unfolding mechanism of the small iron-sulfur protein rubredoxin from *Pyrococcus furiosus* (RdPf) at pH 2. Unfolding was initiated by a temperature jump, triggered by manual mixing of a concentrated protein solution into a thermally preequilibrated buffer. The process was then followed in real time by absorption, fluorescence, and circular dichroism spectroscopies. A complex mechanism emerges, which points to the formation of at least three kinetic intermediates. All of the steps are very slow, the loss of hydrophobic environment of the protein core being rate-determining. According to a proposed sequential mechanism, partial release of secondary structure elements precedes iron loss, which is followed by further loss of β -sheet content and, finally, by hydrophobic relaxation. The implications of this mechanism for the folding pathways and energy landscapes of RdPf have been discussed.

INTRODUCTION

The understanding of macromolecular folding mechanisms (Creighton, 1990; Lecomte & Matthews, 1993; Matthews, 1993) is a fascinating and, at the same time, formidably challenging problem. This is testified by the countless number of scientists that have worked on this topic, as well as by the fact that many questions about the

primary issues involved in this field still wait to be answered. The results of the kinetic experiments presented in chapter 3 stimulated our interest in the folding pathways of proteins from hyperthermophiles. In particular, an important issue that needs to be settled is whether there is anything peculiar about the folding pathways of proteins from hyperthermophilic organisms, that differentiates them from proteins from mesophilic species. Until recently, very little has been known on the subject (Jaenicke, 1996; Rehaber & Jaenicke, 1993; Schultes & Jaenicke, 1991).

As discussed more extensively in Chapter 1, hyperthermophilic microorganisms, collectively known as *Archaea* (Woese et al., 1978), are able to withstand high temperature and pressure. They grow optimally at temperatures greater than 80°C. Not surprisingly, the fact that proteins from hyperthermophilic life forms are underrepresented in the folding/unfolding literature is, in part, due to the fact that most of these proteins have been isolated and characterized only within the last few years (Adams, 1993; Adams & Kelly, 1995; Flam, 1994; Stetter et al., 1990). The whole field actually started in the 1980s, after Karl O. Stetter and his colleagues from the University of Regensburg, Germany, discovered a number of life forms in the geothermal sediments and hot waters of Italy's "Vulcano" Island (Stetter et al., 1983). We have seen in the previous chapters that rubredoxin from *Pyrococcus furiosus* (RdPf) is an excellent model system for the study of the properties of hyperthermostable proteins. In particular, the surprising result that electrostatic interactions slow down the unfolding kinetic step involving iron loss (see chapter 3) has prompted us to learn more about the details of the unfolding mechanism of this

protein. The results presented in chapter 3 leave, among others, two fundamental and interesting questions unanswered:

- 1) Does RdPf lose its secondary and tertiary fold (i.e., it unfolds) at the same time as iron comes off?
- 2) Do ion pairs affect the kinetic behavior of any of the unfolding steps other than the one involving iron loss?

We have started to address the aforementioned issues by investigating the denaturation of RdPf in real time by different spectroscopic techniques. Each of the techniques employed is suitable to probe different events such as changes in iron coordination environment (absorption spectroscopy), variations in overall secondary structure (far-UV circular dichroism spectroscopy), and exposure of the tryptophan (Trp) sensitive portion of the hydrophobic core to the solvent (fluorescence emission spectroscopy). As discussed below, unfolding turns out to be surprisingly more complex than for the majority of mesophilic proteins that have been studied so far. Due to the unusually slow observed rates, manual mixing was sufficient to ensure that no kinetic processes were missed during the dead time of the experiments. Our interests are not limited to mechanistic studies at a fixed temperature, and we intend to gain some insight into the contribution of the Arrhenius preexponential factors and activation energies to the observed rates for each individual unfolding step. As a result of the study of the temperature dependence of the unfolding kinetics, we determined that, over the experimentally accessible temperature range, no significant changes in the observed proposed mechanism take place. It is important to remember that the

results presented in the next sections have been derived at a pH where all the salt bridges of RdPf are no longer present. However, previous studies have shown that RdPf assumes pseudo-native structural features under those conditions (Cavagnero et al., 1995). Additional results and insights will be discussed in the next sections of this chapter.

MATERIALS AND METHODS

RdPf was isolated and purified according to known procedures (Blake et al., 1991). The kinetics of unfolding was monitored by UV-vis absorption, Trp fluorescence emission and far-UV circular dichroism spectroscopies. Experimental details for the absorption measurements have been described in chapter 3. Unfolding was irreversible, under our experimental conditions. No denaturing agents were employed. Protein concentrations were in the micromolar range. Buffers consisted of 40 mM glycine solutions at pH 2.0. No salt was added to the buffers (ion strength 40 mM).

Fluorescence experiments were performed with an SLM 4800 fluorometer, in a thermostatted cuvette, under vigorous stirring. Samples were excited at 280 nm and the timecourse of fluorescence emission monitored at 340 nm. Unfolding was initiated as described in chapter 3 for the absorption spectroscopy experiments. The dead time of the kinetic experiments ranged between 1 and 2 seconds.

Far-UV circular dichroism experiments were performed on a Jasco-600 instrument, in a water-jacketed thermostatted 0.1 cm path cuvette. The timecourse of the spectroscopic signal was monitored at 225 nm. Unfolding was initiated as described in chapter 3 for the absorption spectroscopy experiments; in this case, though, manual mixing rapidly followed the addition of protein stock solution to the buffer. The dead time for the CD experiments was about 60 sec.

Kinetic parameters were derived by fitting the experimental kinetic traces to an expression for the observed spectroscopic signal of the form:

$$f_{\text{obs}} = f_{\text{N}} x_{\text{N}} + f_{\text{I}} x_{\text{I}} + f_{\text{U}} x_{\text{U}} , \quad (4-1)$$

where f_{OBS} represents the observed spectroscopic signal, f_{N} is the spectroscopic signal of the native state (N) at time zero, f_{I} is the hypothetical spectroscopic signal to be expected in a case where the whole protein is present as the intermediate (I), and f_{U} is the expected spectroscopic signal at the end of the reaction, when all the protein is in its unfolded state (U). Eq. 4-1 simply reflects the fact that signals detected by optical spectroscopy are a linear combination of the signals due to each of species that absorbs at the wavelength, or wavelength range, under examination. The symbols x_{N} , x_{I} , and x_{U} represent the mole fractions of the species N, I and U, respectively. These mole fractions are defined as follows:

$$x_{\text{N}} = \frac{[\text{N}]}{[\text{N}_0]} , \quad x_{\text{I}} = \frac{[\text{I}]}{[\text{N}_0]} , \quad x_{\text{U}} = \frac{[\text{U}]}{[\text{N}_0]} , \quad (4-2)$$

where $[N_0]$ is the initial concentration of native state, $[N]$ is the concentration of native state at time t , and $[I]$ is the concentration of the intermediate at time t .

Two different reaction schemes were considered when fitting the data. Kinetic analysis of the first scheme,



yielded an expression for the timecourse of the N and U concentrations. These were inserted into eq. 4-1, leading to the following relation:

$$f_{\text{obs}} = f_N e^{-kt} + f_U (1 - e^{-kt}) + a, \quad (4-4)$$

where t is the time and a is a generic constant. The unfolding rate constant k and the constant a , which has been introduced to account for small baseline errors, have been treated as adjustable parameters. As mentioned, eq. 4-1 has been utilized when deriving eq. 4-4. It is important to notice that, during the derivation, the second term of eq. 4-1 has been neglected, since the pertinent kinetic model does not predict the formation of any intermediate. In the specific case of the kinetic traces from absorption spectroscopy in the visible region, the optical density of the unfolded form is zero. Therefore, eq. 4-4 reduces to

$$f_{\text{obs}} = f_N e^{-kt} + a. \quad (4-5)$$

The second reaction scheme is



Conventional kinetic analysis of reaction scheme 4-6 yielded an expression for the timecourse of the N, I and U concentrations. No approximations were made. Substitution of these expressions into 4-1 gives rise to:

$$f_{\text{obs}} = f_N e^{-k_1 t} + f_I \frac{k_1}{(k_2 - k_1)} \left[e^{-k_1 t} - e^{-k_2 t} \right] + f_U \left\{ 1 + \frac{1}{(k_1 - k_2)} \left[k_2 e^{-k_1 t} - k_1 e^{-k_2 t} \right] \right\}, \quad (4-7)$$

where k_1 and k_2 are the kinetic constants for the first and second step, respectively, of reaction scheme 4-6. The constants f_N and f_U were derived from the initial and final conditions, respectively, whereas f_I was regarded as an adjustable parameter in the fit. f_I should be, to a first approximation, temperature independent. In cases where f_I was observed to vary with temperature by more than about 20%, the value obtained from lowest temperature data in the series was substituted into eq. 4-7. This modified equation, which no longer contains f_I as an adjustable parameter, was used for the fits at the higher temperatures. The rationale for selecting the lowest

temperature f_I as the “good” value, is that, in a variable temperature series, it was found that the two kinetic phases are more clearly discriminated at the lowest temperature. As temperature increases, often the first phase becomes too fast to be observed. Therefore, data at lower temperature are expected to more accurately fit eq. 4-7.

Equations 4-4, 4-5 and 4-7 were used to fit the kinetic traces. Typically, curve fitting was done by first testing the experimental traces against eq. 4-4 or 4-5. If a satisfactory fit was obtained, the calculated first-order rate constant k was regarded as the true value. Otherwise, fitting to the more complicated expression 4-7 was performed. In the latter case, the process was considered to take place according to scheme 4-6, i.e., via formation of an intermediate. As discussed in the next section, some of the low temperature fluorescence data were also fit to an equation which was derived starting from the following kinetic model:



where the forward rate constants for the first and the second step are k_1 and k_2 , respectively, and the rate constant for the back-reaction is k_{-1} . The usual procedure yields the appropriate kinetic expression for the scheme 4-8:

$$f_{\text{obs}} = \frac{f_N}{(1+K)} e^{-\frac{k_2 K}{(1+K)} t} + \frac{f_I K}{(1+K)} e^{-\frac{k_2 K}{(1+K)} t} + f_U \left[1 - e^{-\frac{k_2 K}{(1+K)} t} \right], \quad (4-9)$$

where K is the equilibrium constant for the first step (i.e., $K=I/N$). In order to derive eq. 4-9, it was not necessary to make the steady-state approximation. It is interesting to notice that eq. 4-9 contains only a single exponential, although it describes a two-step process. The software Kaleidagraph (Synergy Software) was used to implement curve fitting routines.

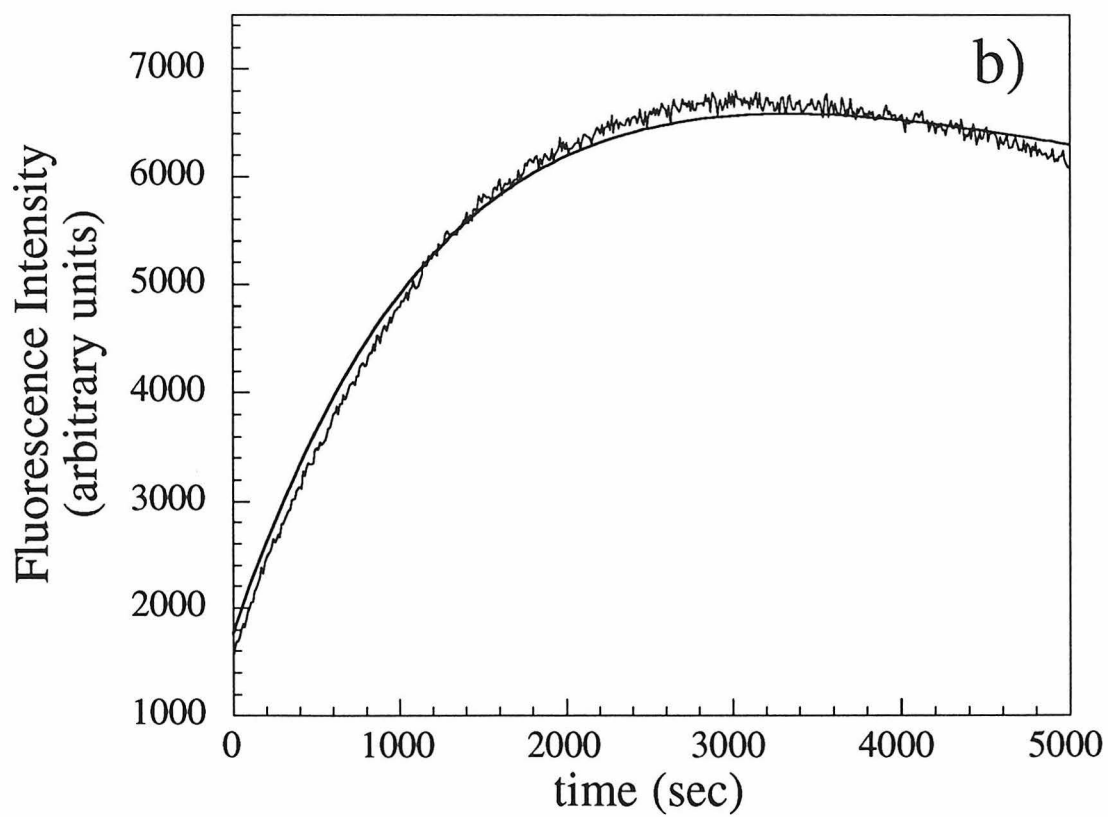
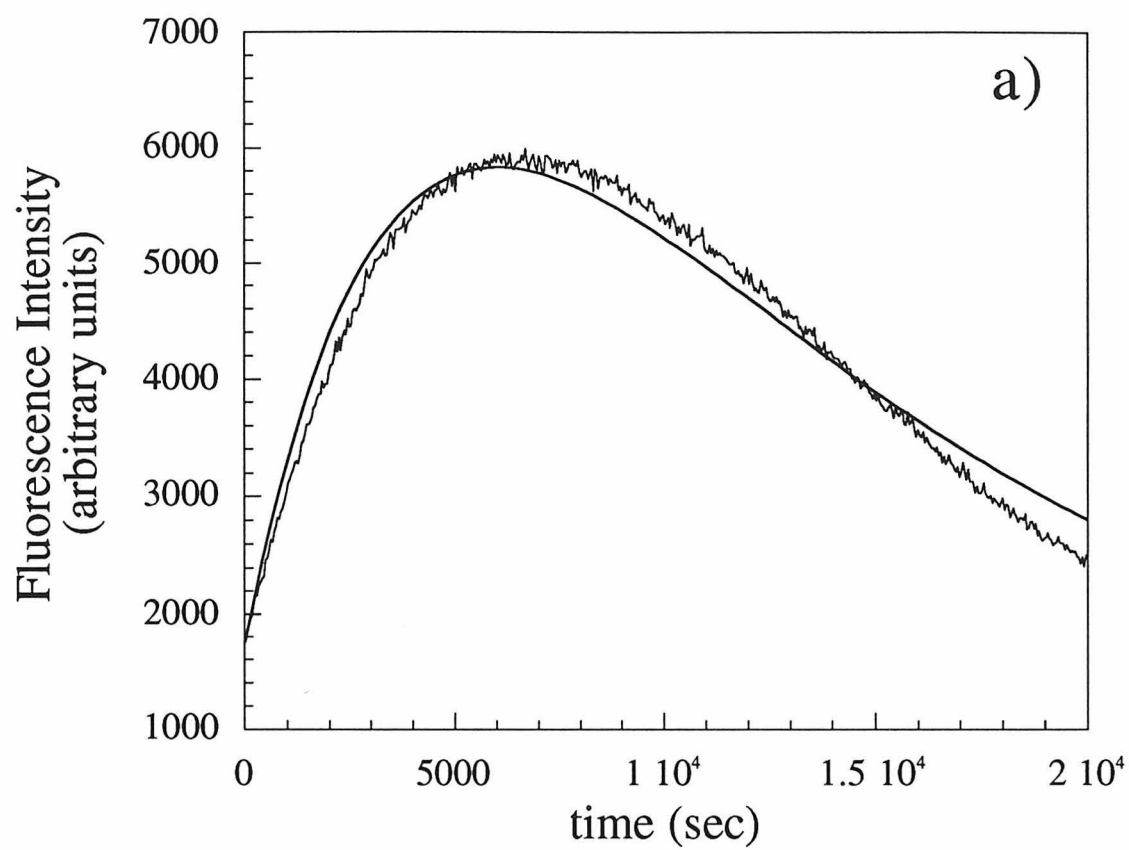
RESULTS

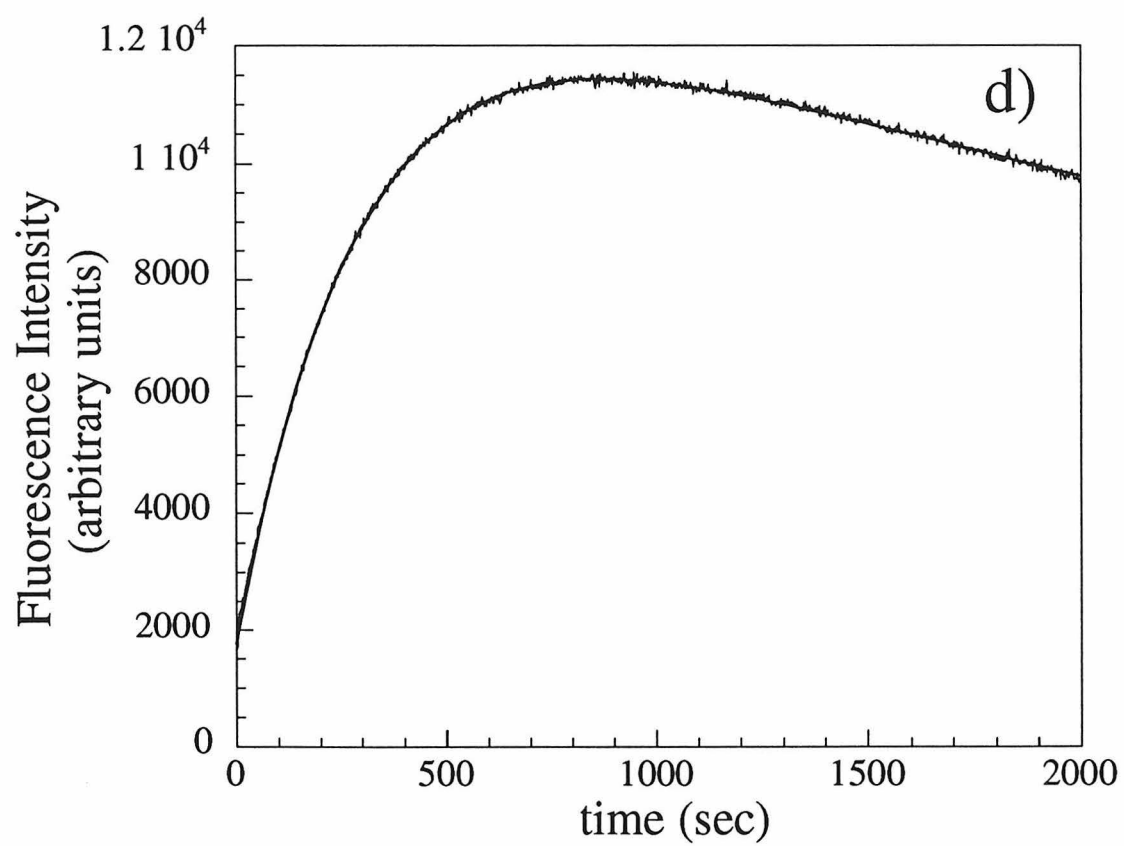
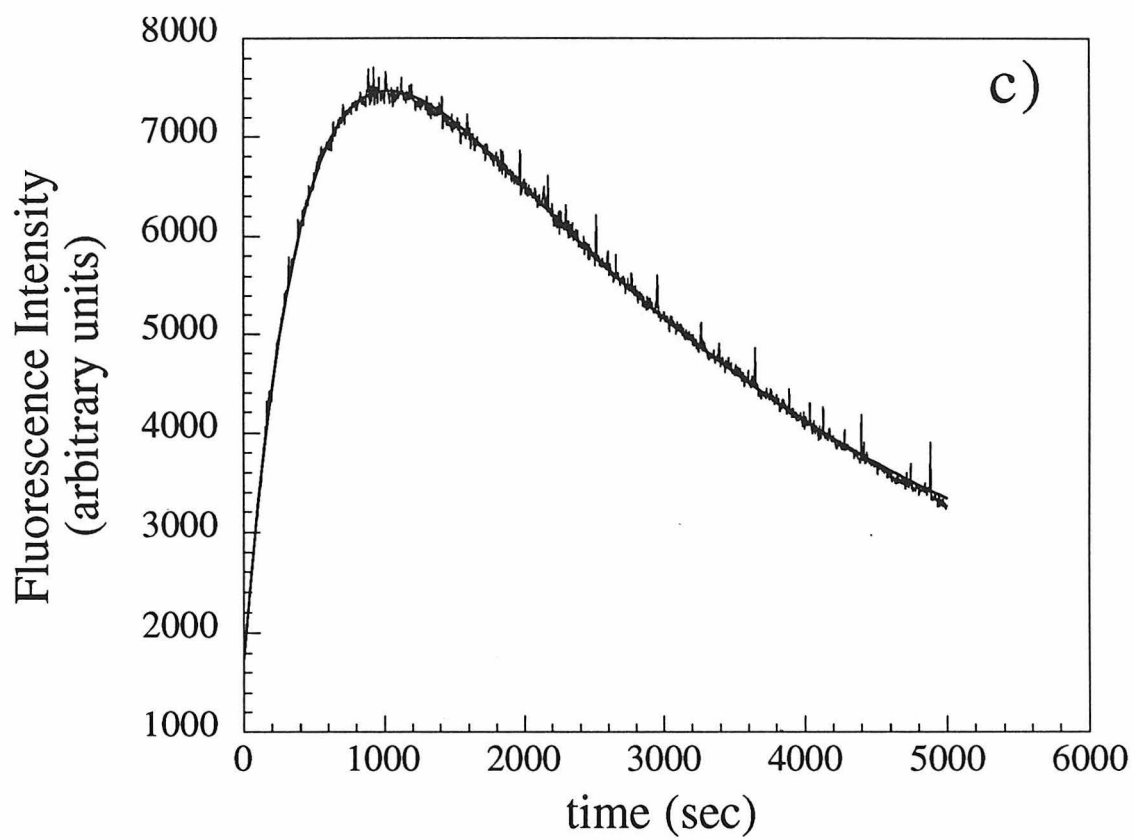
All of the unfolding studies described in this chapter have been performed at pH 2. Regardless of the spectroscopic probe used, unfolding rates are extremely slow in the temperature range that has been analyzed.

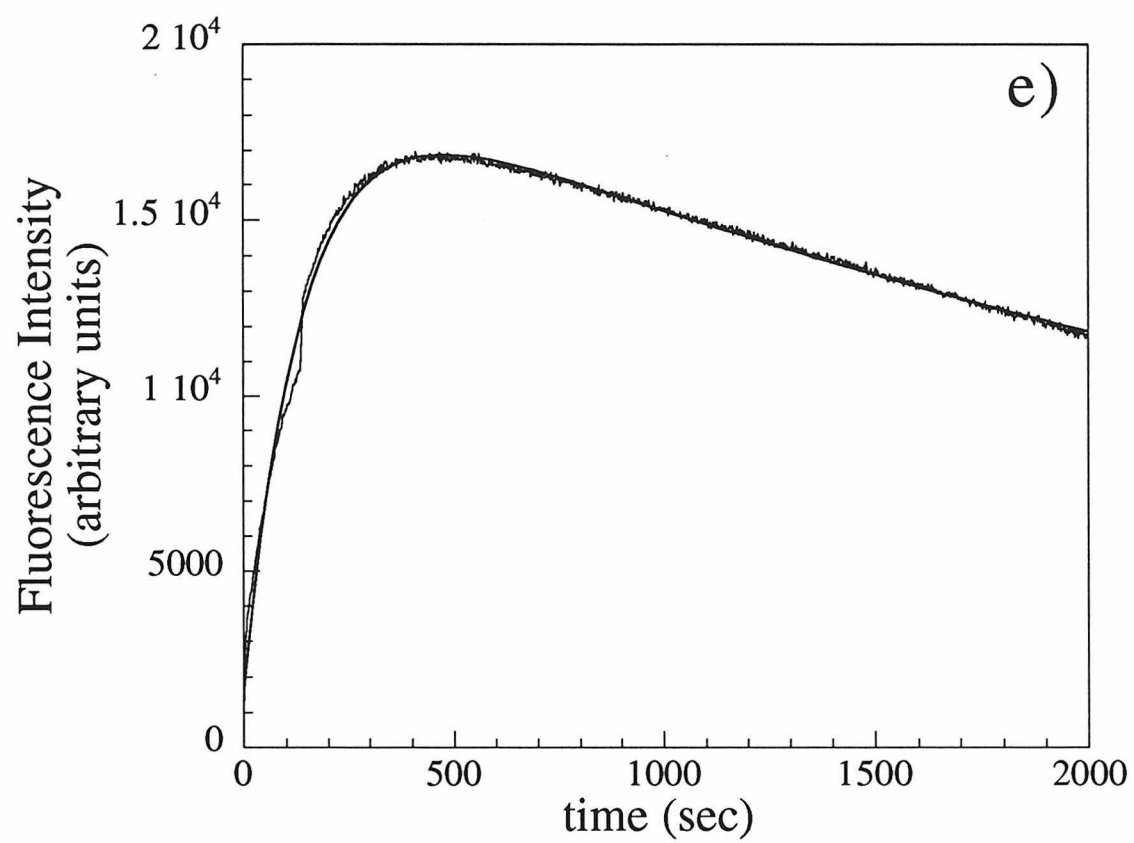
Absorption Spectroscopy. Fig 3-1 shows the temperature dependence of the time-decay of the charge-transfer absorption bands of RdPf at pH 2. As discussed in the previous chapter, the curves follow monoexponential kinetics. This suggests that Fe^{3+} is lost in a single step, during the unfolding process. First-order rate constants are readily obtained (Table 4-1), and the Eyring plot (Fig. 3-2) gives an excellent fit.

Fluorescence Spectroscopy. Results from Trp fluorescence emission spectroscopy are more interesting, and they add complexity to the problem. The kinetic traces displayed in Fig 4-1 clearly indicate the formation of an unfolding intermediate on a slow time scale. This is unusual since unfolding of virtually all mesophilic proteins studied so far follows monophasic kinetics (Creighton, 1993). We will have further comments on this point in the next section. All the kinetic

FIGURE 4-1: RdPf unfolding kinetics as followed by Trp fluorescence emission spectroscopy. The experimental pH is 2.0. The excitation wavelength is 280 nm. The experimental temperatures are: a) 63.3°C, b) 68.3°C, c) 72.3°C, d) 74.6°C, and e) 77.8°C. Non linear least squares curve fitting was performed. The kinetic model and the equation used for curve fitting are described in the "Materials and Methods" section.







traces fit well to a biexponential expression, arising from deriving the expected observed fluorescence trace for the simple two-step mechanism $N \rightarrow I \rightarrow U$, where N is the native state, I is a generic intermediate, and U is the unfolded state. Fig 4-1 displays the curve fits to the fluorescence emission traces, superimposed on the original data. The fits at 72-78°C are quite satisfactory. However, slightly poorer results have been obtained for the two traces recorded at 63.3°C and 68.3°C. Simple analysis of Fig. 4-1, graphs a) and b), indicates that the fits apparently overestimate the rate of the first kinetic phase, and they underestimate the rate of the second kinetic phase. This suggests that the backrate for the first step of mechanism 4-6 should not be neglected at the lower temperature values, where it is possible that $k_{-1} [\text{Fe}^{3+}] \approx k_1$ and k_2 . We fitted the low temperature traces to equation 4-9, which refers to kinetic scheme 4-8, but, unfortunately, we did not get better results. The reasons for this are not entirely clear to us at this point. In any case, we consider the fits that we got from eq. 4-7 to be satisfactory. The kinetic rate constants for the fluorescence data are presented in Table 4-1. Fig. 4-2 illustrates the Eyring plot for the first and second fluorescence kinetic phases. The fit for the first step (k_1) is very good, whereas the fit for the second phase (k_2) is affected by a larger error. This may be explained in terms of our inability to detect some mechanistic subtleties or, as discussed in the next section, in terms of a temperature dependent heat capacity of activation.

Circular Dichroism spectroscopy. Experimental far UV circular dichroism (CD) spectra are shown in Fig 4-3. The data are best fit by curves derived from

TABLE 4-1: First-order kinetic rate constants for rubredoxin from *Pyrococcus furiosus* unfolding.

Spectroscopic Technique	Temperature (°C)	Unfolding Rate Constant (sec ⁻¹)	
		k ₁	k ₂
Absorption	62.4	$0.000173 \pm 2 \cdot 10^{-6}$	--
	66.6	$0.000399 \pm 6 \cdot 10^{-6}$	--
	71.0	$0.00138 \pm 5 \cdot 10^{-5}$	--
	75.0	$0.00352 \pm 4 \cdot 10^{-5}$	--
	76.7	$0.00558 \pm 5 \cdot 10^{-5}$	--
Trp Fluorescence Emission	63.3	0.000158 ± 0.0041	0.000158 ± 0.0041
	68.3	$0.000729 \pm 7 \cdot 10^{-6}$	$7.90 \cdot 10^{-5} \pm 2 \cdot 10^{-6}$
	72.3	$0.00215 \pm 4 \cdot 10^{-6}$	$0.000305 \pm 4 \cdot 10^{-7}$
	74.6	$0.00334 \pm 4 \cdot 10^{-6}$	$0.000210 \pm 5 \cdot 10^{-7}$
	77.8	$0.00703 \pm 2 \cdot 10^{-5}$	$0.000278 \pm 1 \cdot 10^{-6}$
Circular Dichroism	63.6	$0.00109 \pm 1 \cdot 10^{-5}$	$4.75 \cdot 10^{-5} \pm 5 \cdot 10^{-8}$
	66.3	$0.00197 \pm 3 \cdot 10^{-5}$	$0.000109 \pm 2 \cdot 10^{-7}$
	70.1	0.382 ± 0.95	$0.000347 \pm 1 \cdot 10^{-6}$
	75.7	$1.24 \cdot 10^{47} \pm 1 \cdot 10^{57}$	$0.000884 \pm 4 \cdot 10^{-6}$
	78.5	0.0442 ± 0.0009	$0.00331 \pm 9 \cdot 10^{-5}$

FIGURE 4-2: Eyring plot for the two RdPf unfolding kinetic phases (k_1 , faster kinetic phase; k_2 , slower kinetic phase) as detected by fluorescence experiments, at pH 2.0. Linear least squares curve fitting was performed on both traces.

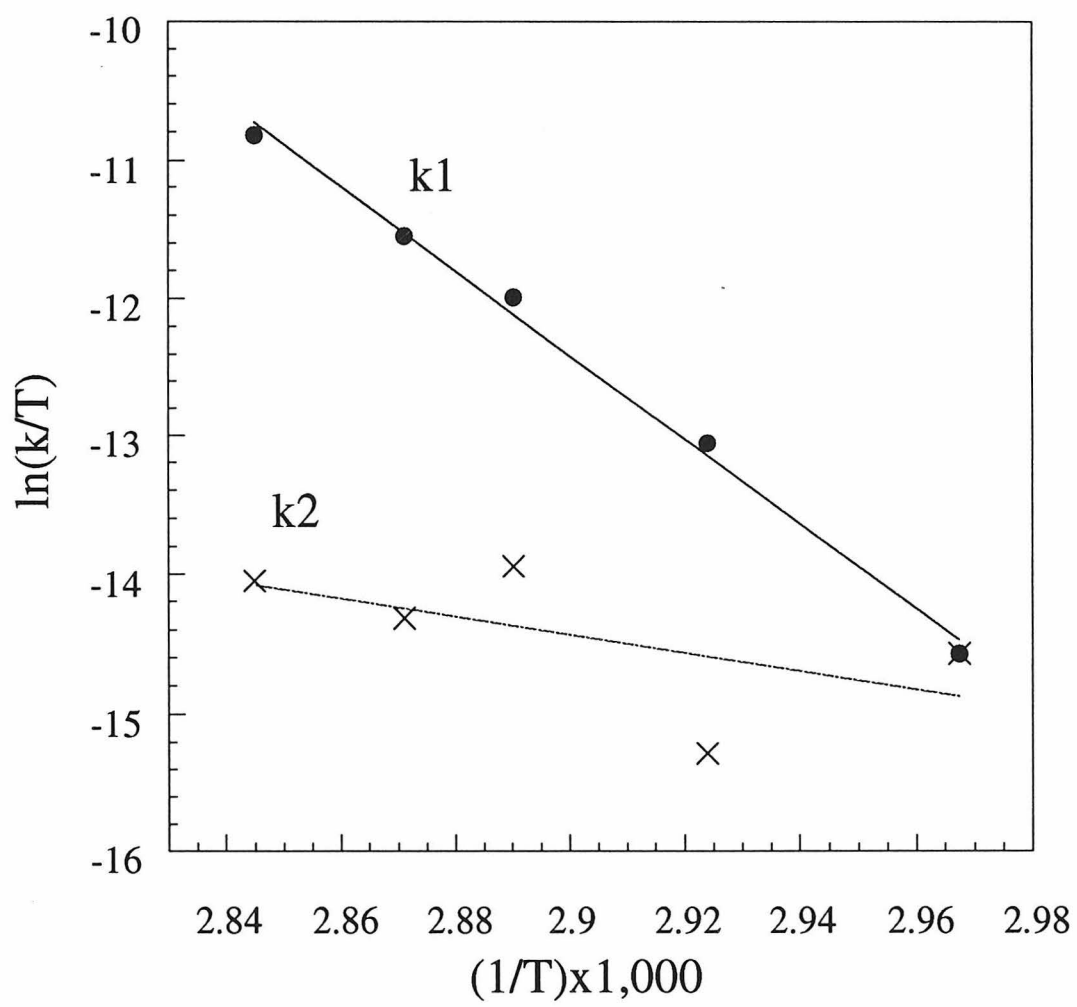
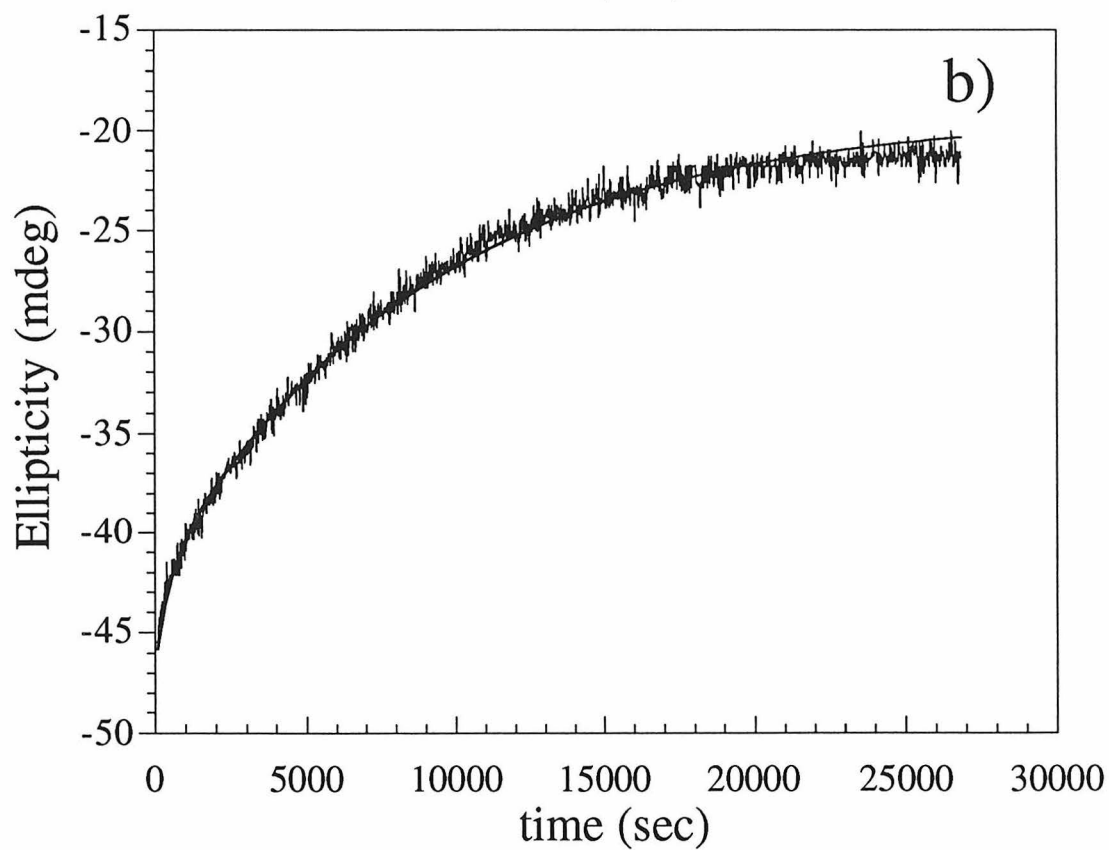
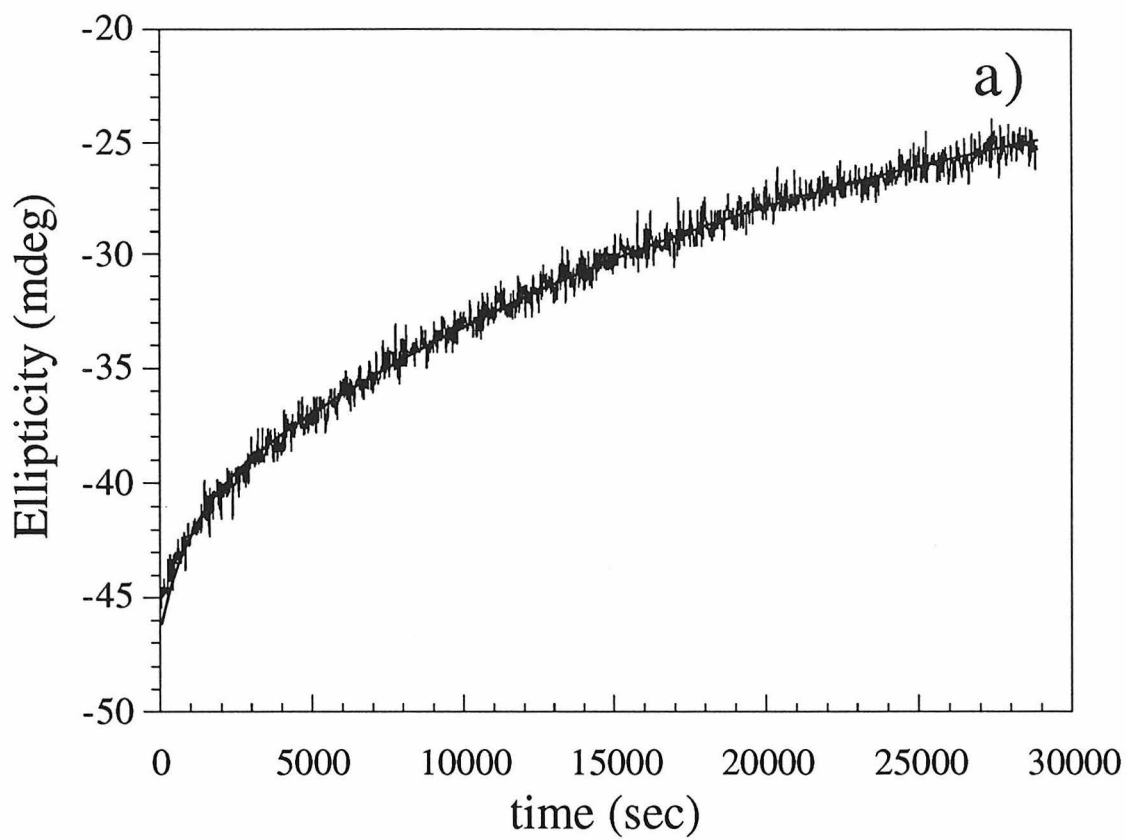
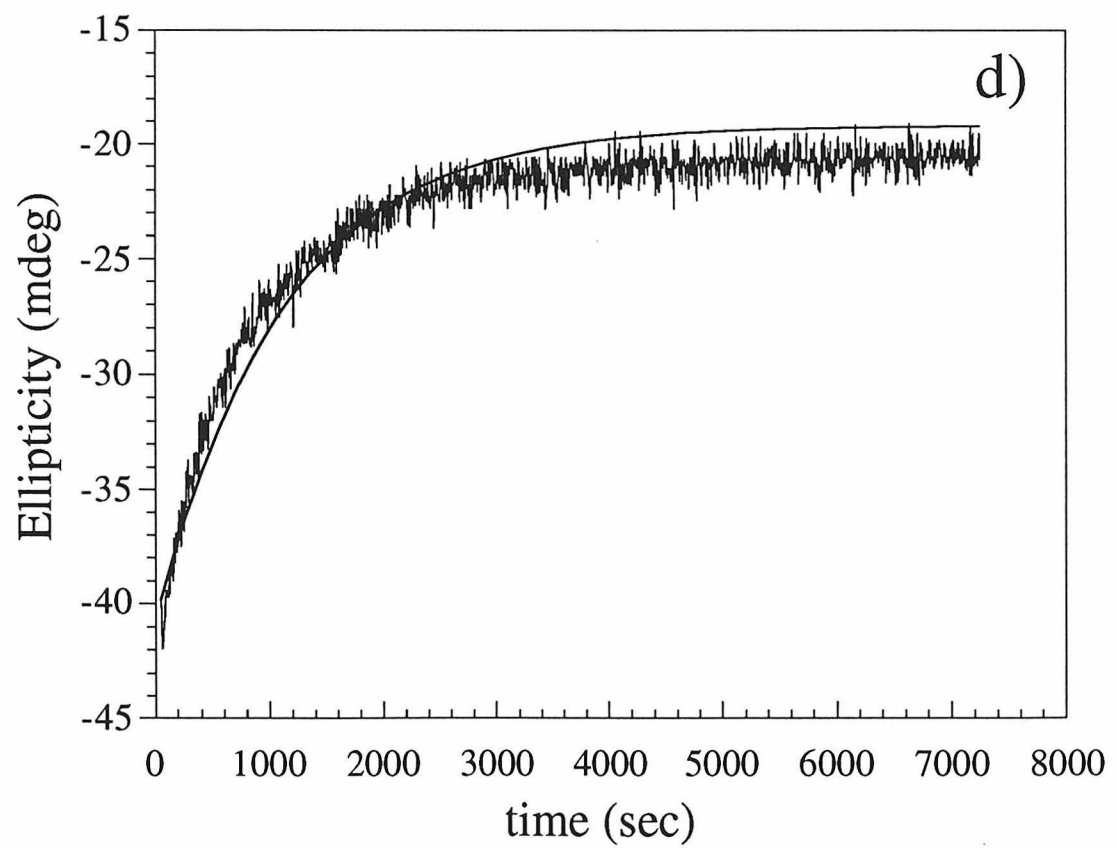
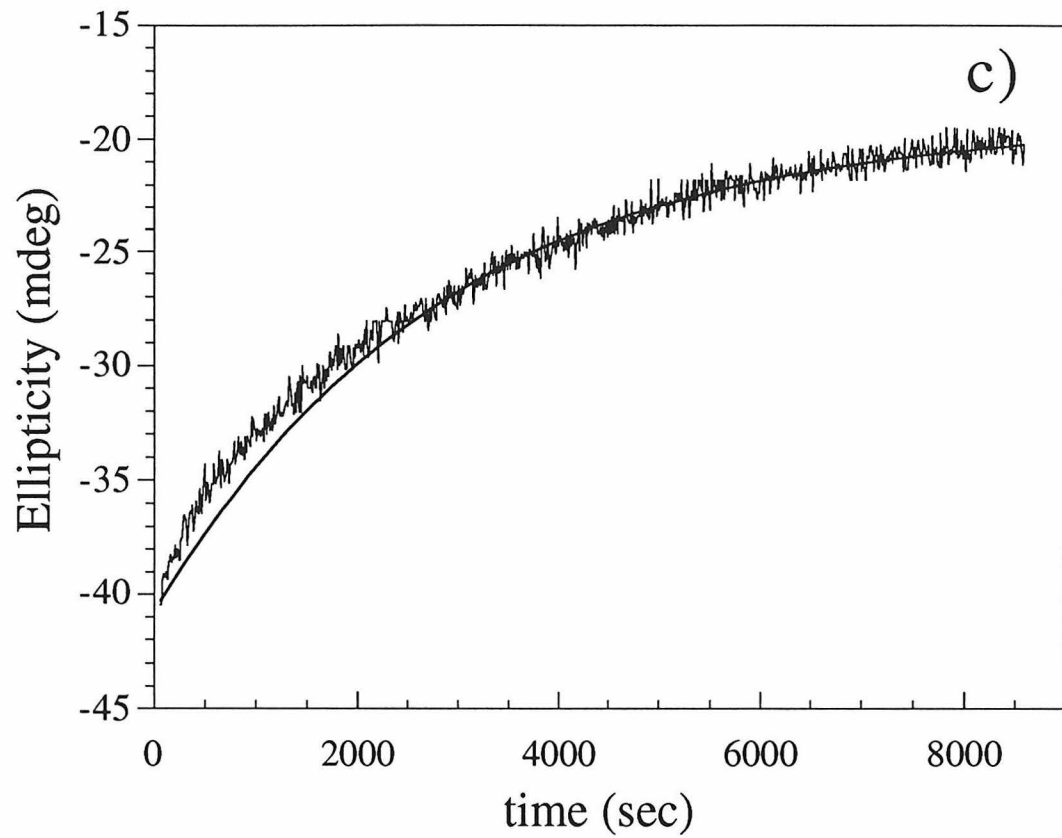
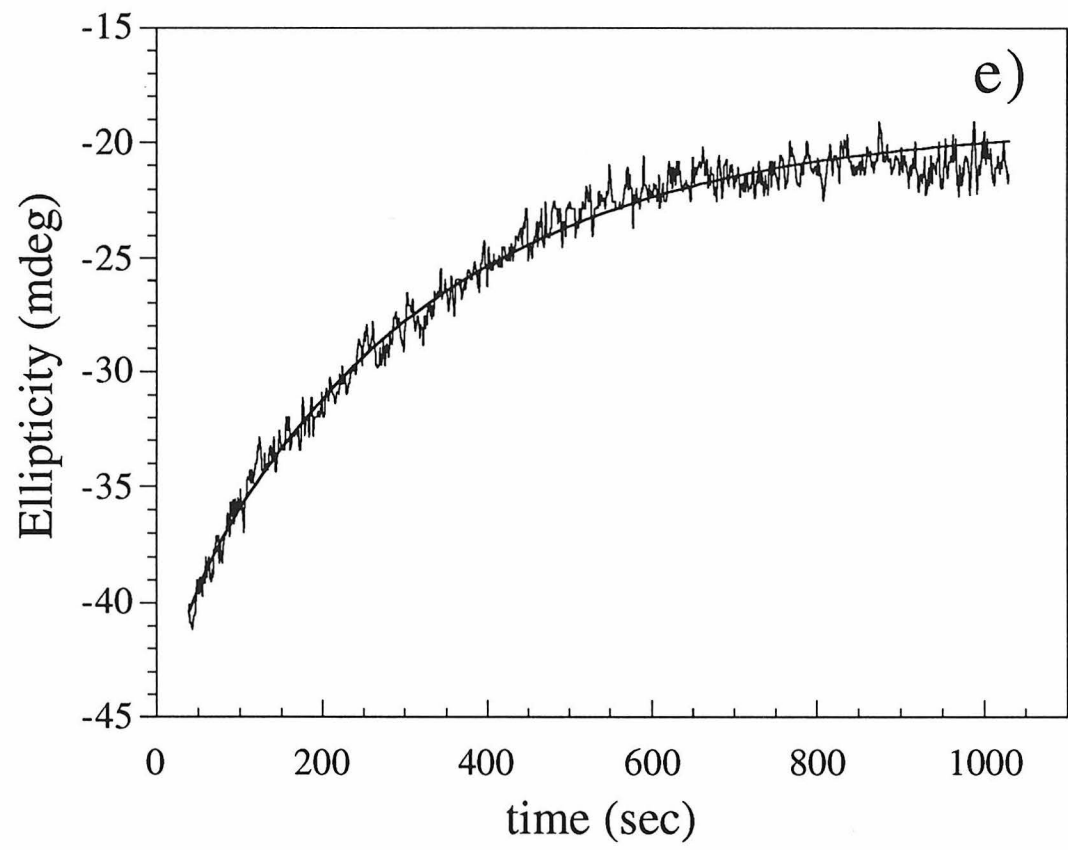


FIGURE 4-3: RdPf unfolding kinetics as followed by circular dichroism spectroscopy. The experimental pH is 2.0. The experimental temperatures are: a) 63.6°C, b) 66.3°C, c) 70.1°C, d) 75.7°C, and e) 78.5°C. Non linear least squares curve fitting was performed. The kinetic model and the equation used for curve fitting are described in the "Materials and Methods" section.



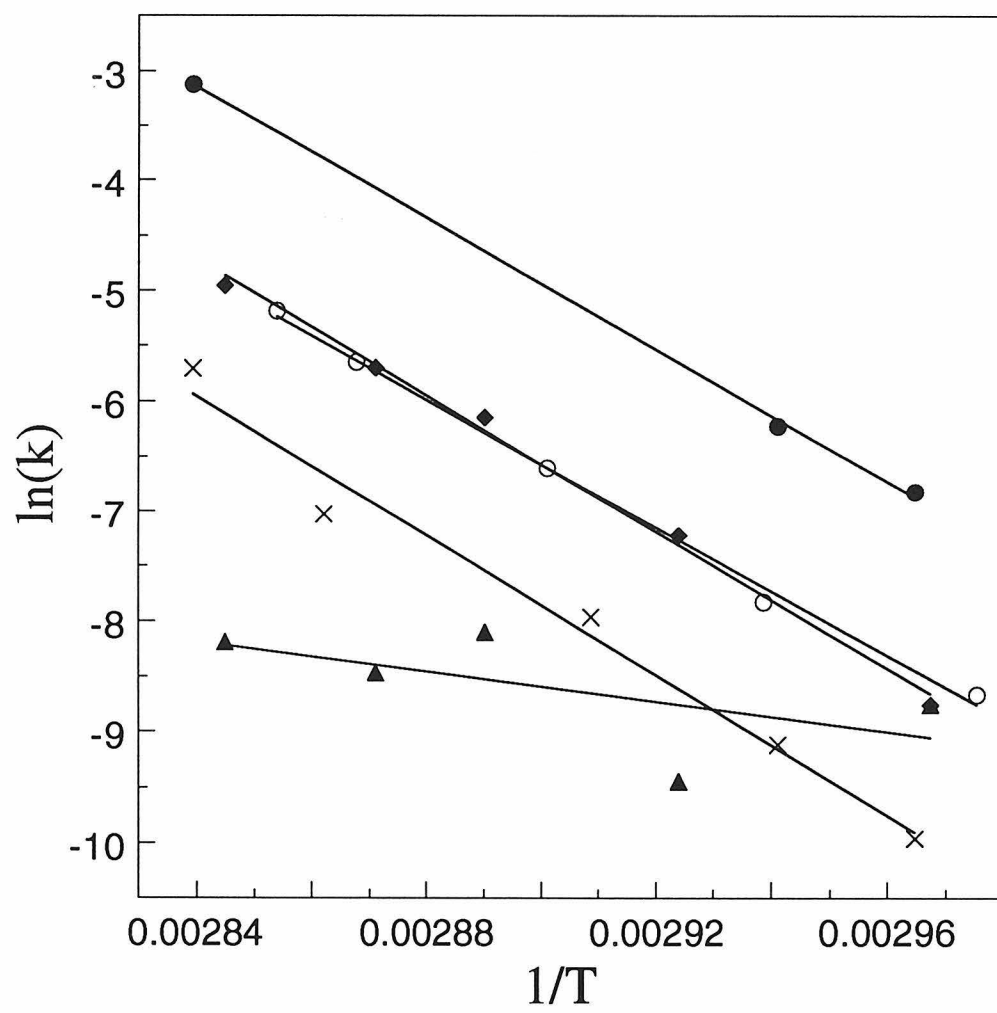




biexponential expression 4-7. Excellent fits are obtained at all experimental temperatures. The first-order unfolding rate constants derived from curve fitting for the CD data are presented in Table 4-1. We observed that the values obtained for the first kinetic constant (k_1) tend to become unreasonably high as the temperatures has approached. This can be explained since, according to the kinetic trend established by the k_1 values at lower temperature, the expected lifetimes of the native state at higher temperatures range between 90 sec (at 78.5°C) and 170 sec (70.1°C). These times are definitely comparable with the dead time of the CD measurements (about 60 sec). Therefore, the corresponding kinetic values cannot be accurately defined at temperatures higher than 70°C. Having taken this into consideration, we have not included the k_1 values obtained at 70.1°C and 75.7°C in the Arrhenius curve fit for this kinetic phase. However, we have retained the k_1 value at 78.5°C since, due to good manual technique, the dead time for the corresponding measurement turned out to be only 32 sec. It is, nonetheless, important to remember that the Arrhenius curve for this CD phase is affected by considerably large error.

A comparative Arrhenius plot for RdPf unfolding as followed by absorption, Trp fluorescence emission and CD spectroscopies is presented in Fig. 4-4. The analysis of this plot is fundamental to proposing a mechanism for the unfolding process. More will be said on this point in the next section.

FIGURE 4-4: Comparative Arrhenius plots for RdPf unfolding, monitored by different spectroscopic techniques: ● circular dichroism data, first phase; ○ absorption data; ◆ fluorescence data, first phase; x circular dichroism data, second phase; ▲ fluorescence data, second phase.



DISCUSSION

We have extended the kinetic studies on RdPf presented in the previous chapter by recording the unfolding kinetics by Trp fluorescence emission and by far-UV CD spectroscopies, at pH 2. Kinetic studies of chapter 3 are based on absorption measurements at wavelengths corresponding to the ligand to metal RdPf charge transfer transitions, i.e., in the visible region of the spectrum. These measurements are a good probe for the Fe^{3+} environment. By taking advantage of the fact that a) Trp fluorescence emission is capable of probing the chemical environment surrounding the Trp amino acid(s) in a protein and, b) far-UV CD is sensitive to the amount of secondary structure of a polypeptide, we have been able to gain a more complete picture of the unfolding event. It has been possible to find that RdPf unfolds, i.e., it loses its secondary and tertiary structures, on timescales comparable to those associated with Fe^{3+} release. More importantly, distinct kinetic events have been identified along the unfolding pathway. While the absorption data (see chapter 3), which display monophasic kinetics, suggest that Fe^{3+} loss takes place in a single event, the fluorescence denaturation profile is more complex, since it is biphasic and it clearly points to the formation of an intermediate (Figs 4-1). This constitutes the first example of an unfolding intermediate observed by fluorescence spectroscopy. This is particularly interesting, since protein unfolding almost universally takes place as an all-or-none event (Creighton, 1990) with no detectable kinetic intermediates (Chen et al., 1989; Chen & Matthews, 1994; Kuwajima et al., 1989; Mücke & Schmid, 1994;

Segawa & Sugihara, 1984; Serrano et al., 1992). Previous kinetic studies on protein denaturation have generally shown a single kinetic phase, which obviously coincides with the rate-determining step, with no lag period. Furthermore, probes of either secondary structure or tertiary structure produce the same rate constant (Schmid, 1992). A rare exception is given by the NMR experiments recently performed by Baldwin and coworkers (Kiefhaber et al., 1995) on ribonuclease A, where the existence of a lag phase preceding the unfolding NMR kinetic trace was observed. However, unfolding of ribonuclease A appears to be a monophasic process when followed by other spectroscopies (Kiefhaber et al., 1995), unlike what we have seen for RdPf. This general trend should be contrasted with the behavior of typical spectroscopic probes in the refolding reactions of most small proteins: several kinetic intermediates are often detected in this case (Dobson et al., 1994). The fact that we were able to detect kinetic intermediates before the unfolding transition state is not so important *per se*, but, rather, because it enables us to gain important information on the RdPf folding (or refolding) pathways after the transition state has been “hit.” This has generally not been possible before with mesophilic proteins where any “unfolding” intermediates, if present, are not sufficiently populated to be detected in real time. The reason why we could detect these intermediates is essentially due to the fact that all RdPf unfolding steps are slow. This allows buildup of sufficient populations of intermediates.

Going back to our experiments, by simple comparison of reaction rates at a similar temperature, it is apparent that the first fluorescence phase for RdPf unfolding

TABLE 4-2: Activation parameters for the unfolding of RdPf, as followed by UV-vis absorption, fluorescence and circular dichroism spectroscopies.

Technique	E_a	A	$\Delta H^\ddagger$	$\Delta S^\ddagger$	$\Delta G^\ddagger$
	(kcal mol ⁻¹)	(sec ⁻¹)	(kcal mol ⁻¹)	(cal mol ⁻¹ deg ⁻¹)	(kcal mol ⁻¹)

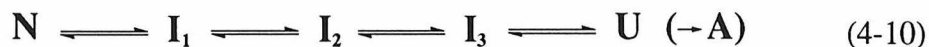
UV-vis	61 ^{a)}	4x10 ³⁶ a)	60 ^{a)}	101 ^{a)}	25
Fluorescence ^{b)}	61	1x10 ³⁶	61	104	25
CD ^{c)}	47	3x10 ²⁶	47	60	26

a) Average value of three individual experiments; b) First phase; c) second phase.

corresponds to the same event as the one detected by UV-vis. Accordingly, the relative Arrhenius plots are, within experimental error, virtually superimposable (Fig. 4-4). The activation parameters from UV-vis are also very similar to those from fluorescence (Table 4-2). Kinetic traces of typical fluorescence protein unfolding processes are characterized by a decrease in the observed signal with time. This is usually interpreted in terms of solvent quenching due to increased exposure of the Trp-containing hydrophobic probe to the solvent, in the unfolded form. In our case, the observed fluorescence quantum yield increases, as the intermediate is formed. This is due to Fe^{3+} , a strong fluorescence quencher, being released during this step. The second fluorescence phase corresponds to an increased exposure of the two RdPf Trp to the solvent, which is reflected into a gradual decrease in fluorescence quantum yield. This hypothesis is confirmed by the CD experiments, which also show biphasic kinetics with different kinetic constants (Fig 4-3). The activation parameters derived from the absorption, CD and fluorescence data are presented in Table 4-2. Activation data for the less accurate Arrhenius plots (compare data in Fig. 4-4) have been omitted from Table 4-2. Both preexponential factors (A) and activation energies (E_a) are very high for all the kinetic phases reported in Table 4-2. Furthermore, as already noted above, the activation parameters for UV-vis and fluorescence first phase suggest that the two probes describe the same process. Additionally, both activation entropy and enthalpy due to partial release of secondary structure (second CD phase) are smaller than for the step involving iron loss during unfolding, suggesting that fewer structural rearrangements are required to lose helical content than to release Fe^{3+} . A rigorous

comparison of activation parameters is not possible because of the associated experimental error (an estimate of the size of this error is obtained upon analysis of Table 3-2). However, this is not so relevant, since what differentiates these data from those for mesophilic proteins are not so much the individual values of A or E_a but, rather, the fact that the overall rates are very slow. These slow rates arise as a result of a combination of smaller A and/or larger E_a , compared to proteins from mesophiles.

The comparative Arrhenius plots for RdPf unfolding presented in Fig. 4-4 reveal that the unfolding of RdPf is a complex process. Analysis of Fig. 4-4 suggests that the simplest scheme consistent with the experimental data comprises the following sequential model:

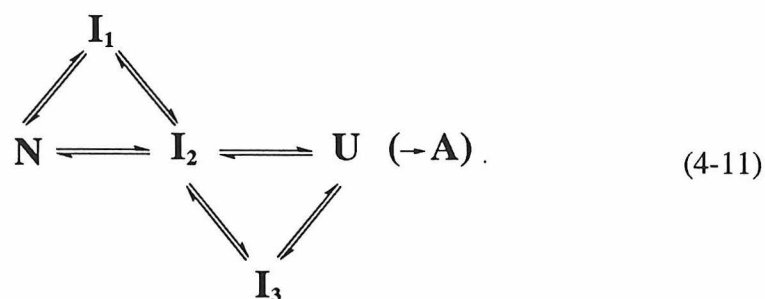


The native state partially loses some secondary structure (I_1 , first CD phase), Fe^{3+} is then released (I_2 , first fluorescence phase, absorption data), more secondary structure is then lost (I_3 , second CD phase), followed by formation of the unfolded state (U). Irreversible aggregation either coincides with the last kinetic step, or it immediately follows the formation of the denatured species. According to Fig. 4-4, the transition state for unfolding is crossed as Trp gets more exposed to the solvent, i.e., during the last step of scheme 4-10. This is what is commonly called an “early transition state” along the folding pathway, since most structural elements (i.e., secondary structure and Fe^{3+} assembly) build up later during folding. This is true, of course, if one assumes

that folding and unfolding pathways are essentially identical, except for the direction in which they are traveled. The principle of microscopic reversibility suggests that this is the case, unless any distinct irreversible bottlenecks in both directions are present in RdPF's complex free energy hypersurface.

In thermodynamic terms, the folding/unfolding heat capacity change is commonly believed to be proportional to the difference in hydrophobicity between native and denatured states (Privalov & Makhatadze, 1990). Similarly, the hydrophobic character of a folding/unfolding transition state can be qualitatively evaluated, with respect to the starting species, by the magnitude of the activation heat capacity ($\Delta C_p^\ddagger$) for the rate-determining step (Oliveberg et al., 1995). This, in turn, is proportional to the curvature of the Arrhenius plot for the rate-determining step. It is clear from Fig. 4-4 that, in general, the unfolding transition states for the various kinetic steps do not involve significant changes in hydrophobicity (i.e., the Arrhenius plots are linear) with respect to the ground state for that step. However, since the experimental data points for the Arrhenius plot of the rate-determining step are more scattered than usual, we cannot draw conclusions on this point based on our data. Since the rate-determining step involves exposure of the hydrophobic core to the solvent, it may well be that $\Delta C_p^\ddagger$ is different from zero for this step.

Folding pathways are potentially extremely complex. A more elaborate model takes into account the fact that some of the pathways may take place in parallel with others. We constrained this alternative model to account for the requirement that I_2 has to be formed along the path, as shown below:



Both models 4-10 and 4-11 are consistent with the observed kinetic data. It is not possible to definitely say which of the two models is correct. However, in light of the available experimental evidence, we favor the first one on the grounds that the CD data satisfactorily fit a biexponential expression. In case the second model were operative, the CD kinetic trace would, in principle, fit best to a quadriexponential expression. Nonetheless, in the case where the kinetic rates for the different steps were relatively similar, it would be virtually impossible to discriminate between the two mechanisms.

CONCLUDING REMARKS

The present studies suggest that RdPf unfolding proceeds via a complex multistep pathway, that starts with partial loss of secondary structure, followed by Fe^{3+} release. In a further step, more secondary structure elements are released. The process ends with exposure of the Trp environment (i.e. the hydrophobic core) to the solvent, followed by, or concomitant with, irreversible aggregation. Our data indicate

that, unlike most mesophilic proteins, the rate-determining step occurs late along the unfolding pathway (i.e. early along the folding pathway).

REFERENCES

- Adams, M. W. W. (1993) *Annu. Rev. Microbiol.* 47, 627.
- Adams, M. W. W. & Kelly, R. M. (1995) *Chem. & Eng. News Dec.* 18, 32.
- Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. & Adams, M. W. W. (1991) *Biochemistry* 30, 10885.
- Cavagnero, S., Zhou, Z.-H., Adams, M. W. W. & Chan, S. I. (1995) *Biochemistry* 34, 9865.
- Chen, B., Baase, W. A. & Schellman, J. A. (1989) *Biochemistry* 28, 691.
- Chen, X. & Matthews, C. R. (1994) *Biochemistry* 33, 6356.
- Creighton, T. E. (1990) *Biochem. J.* 270, 1.
- Creighton, T. E. (1993) in *Proteins - structures and molecular properties* Eds.) 310, W. H. Freeman and Company,
- Dobson, C. M., Evans, P. A. & Radford, S. E. (1994) *Trends Biochem. Sci.* 19, 31.
- Flam, F. (1994) *Science* 265, 471.
- Jaenicke, R. (1996) *FASEB J.* 10, 84.
- Kiefhaber, T., Labhardt, A. M. & Baldwin, R. L. (1995) *Nature* 375, 513.
- Kuwajima, K., Mitani, M. & Sugai, S. (1989) *J. Mol. Biol.* 206, 547.
- Lecomte, J. T. J. & Matthews, C. R. (1993) *Protein Eng.* 6, 1.
- Matthews, C. R. (1993) *Ann. Rev. Bioch.* 62, 653.
- Mücke, M. & Schmid, F. X. (1994) *Biochemistry* 33, 12930.

- Oliveberg, M., Tan, Y.-J. & Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* 92, 8926.
- Privalov, P. L. & Makhatadze, G. I. (1990) *J. Mol. Biol.* 213, 385.
- Rehaber, V. & Jaenicke, R. (1993) *FEBS Lett.* 317, 163.
- Schmid, F. X. (1992) in *Protein Folding* (Creighton, T. E., Eds.) 197, Freeman, New York.
- Schultes, V. & Jaenicke, R. (1991) *FEBS Lett.* 290, 235.
- Segawa, S.-I. & Sugihara, M. (1984) *Biopolymers* 23, 2473.
- Serrano, L., Matouschek, A. & Fersht, A. R. (1992) *J. Mol. Biol.* 224, 805.
- Stetter, K. O., Fiala, G., Huber, G., Huber, R. & Segerer, A. (1990) *FEMS Microbiol. Rev.* 75, 117.
- Stetter, K. O., König, H. & Stackebrandt, E. (1983) *System. Appl. Microbiol.* 4, 535.
- Woese, C. R., J., M. L. & Fox, G. E. (1978) *J. Mol. Evol.* 11, 245.

CHAPTER 5.

Unfolding Kinetics and the Protein Folding Problem

ABSTRACT

This chapter introduces some theoretical considerations on the nature of the protein folding problem, as viewed from the reverse path, i.e., unfolding. As long as one can determine the shape of the folding energy landscape, or more realistically, its main features, unfolding data should yield valuable information about folding pathways. The concept of folding funnel, recently introduced for small single-domain proteins, dictates that folding is driven by a potential energy slide coupled to an entropy slide. We will briefly explain the origins and main features of these funnels, and relate them to the global events (i.e., formation of secondary structure, hydrophobic core, etc.) that take place upon folding. The experimental data presented in chapter 4 on rubredoxin from *Pyrococcus furiosus* will be considered in this framework. Of particular interest is the positioning of the folding/unfolding transition state along the path. In conclusion, protein unfolding is going to be presented as an elementary unimolecular process, and viewed according to RRKM semiclassical unimolecular reaction rate theory.

HYPERCOMPLEX POTENTIAL ENERGY SURFACES AND FOLDING FUNNELS

Protein folding is a complex process, which involves breakage and formation of many noncovalent bonds to reach a status, called a denatured or unfolded form, in

which the amount of residual secondary and tertiary structure is minimal. A protein, from the physical chemist's point of view, is a large polyatomic molecule, held together by covalent and noncovalent forces among the various atoms. The potential energy surface which describes these interactions is a complex hypersurface in multidimensional space. Its dimensionality is of the order of N , the number of atoms that compose the protein. If one were able to construct this hypersurface from first principles, one would know the conformational space of the molecule and its vibrational behavior. If the internal energy of the protein were known, its dynamic behavior and the population of the various energy levels belonging to the hypersurface would be determined as well. Most importantly, with the potential energy hypersurface in hand, one would be able to define the minimum energy pathway for going from unfolded to native state, i.e., the folding mechanism. However, going from native to unfolded state (and vice versa) in a protein does not necessarily occur by a single route. There are, at least in principle, multiple concurrent pathways. Understanding the folding problem, then, reduces to characterizing the dynamics of an ensemble over an ensemble of possible routes.

The total internal energy of a protein is a function of temperature. The temperature of the protein is assumed to be that of its solvent bath, at thermal equilibrium and in liquid phase. The potential energy hypersurface, in the case of a protein, contains two sub-regions which identify its native and denatured state. The two sub-regions are populated to a different extent depending on the total internal energy of the protein. A protein undergoes a phase transition from native to unfolded

state upon redistribution of its populations in the various wells of the hypersurface. This is what happens in a thermally induced transition. The spreading of populations among the energy wells is governed by the Boltzmann distribution law. However, this is only true in subregions where the barriers among the wells are thermally accessible, i.e., where there are no kinetic biases. This is a very important concept, and it is essential to understand folding pathways. It also suggests how nature overcomes the so called "Levinthal paradox" (Levinthal, 1969), which states that a protein would take several hundred years to fold if it were to explore all of its conformational hyperspace. This is, of course, in striking contrast with what occurs in nature. It is also worth noticing that denaturant-induced unfolding processes are less predictable, because denaturing agents (e.g., urea, guanidine hydrochloride) perturb the shape of the potential energy hypersurface. This shape perturbation is, for denaturant-induced transitions, what induces a population redistribution in favor of the unfolded species.

Due to the exceedingly high number of variables involved, it is impossible, in practice, to precisely characterize the potential energy hypersurface of a protein. Chemists are then faced with the problem of trying to find working approximations that yield a simplified, yet correct, description of a complicated phenomenon. There have been a number of attempts, in the past, based on simplified mappings of protein energy landscapes. These were mainly due to Levitt and Warshel (Levitt & Warshel, 1975), Honeycutt and Thirumalai (Honeycutt & Thirumalai, 1990), Skolnick (Skolnick & Kolinski, 1991), and Shakhnovich (Shakhnovich et al., 1991). All of these models rely on a highly simplified description of the protein molecule, which

neglects such essential features as hydrogen bonds, secondary structure and side chain conformational degrees of freedom.

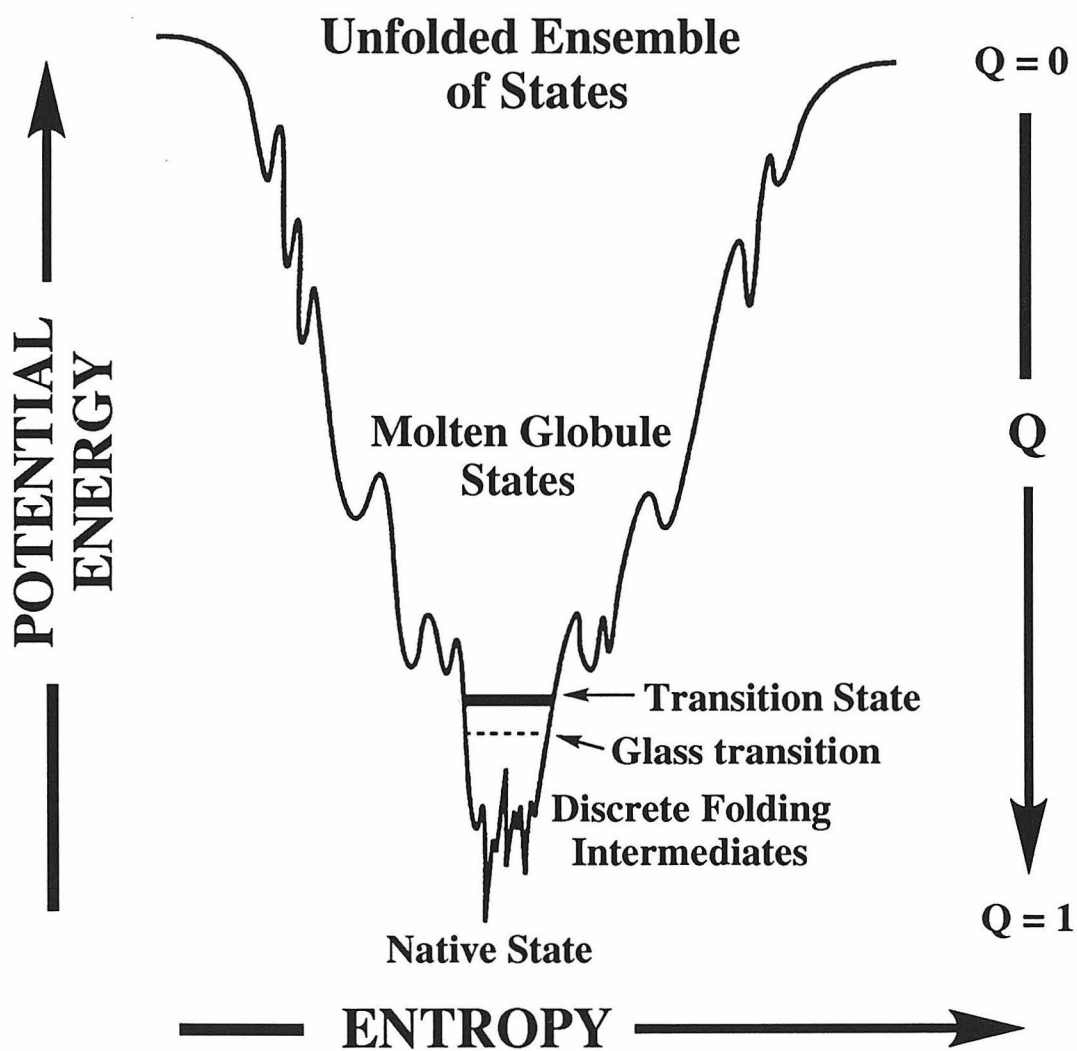
More recently, Onuchic, Wolynes and coworkers proposed a method that provides a simplified statistical description of the topography of the free energy landscape (Bryngelson et al., 1995; Leopold et al., 1992; Onuchic et al., 1995; Wolynes et al., 1995). This description takes advantage of the fact that phase transitions depend only on gross features of the energy functions. The authors refer to the case of phase diagrams for the evaporation of liquids as disparate as water and xenon. These diagrams can be superimposed, after a simple scaling factor is introduced. They claim that this concept, known as the law of corresponding states (Lifshitz & Pitaevskii, 1980), is widespread in nature and extensible to proteins. This allows a simplified description of the protein potential energy landscape. The method has recently become very popular, since it has yielded excellent agreement with observed rates of folding. We have decided to introduce it here because it provides a convincing qualitative description of protein energy landscapes, in accordance with a physically sound picture of the folding phenomenon. The model is based on a simplified description of a protein, which is represented as a 27-unit chain on a cubic lattice. Each amino acid is represented as a single entity, located on one of the lattice edges. The residue types do not correspond directly to the naturally occurring amino acids but are chosen to provide chemical differentiation, based on a hydrophilicity/hydrophobicity scale. More detailed amino acid features can also be added (Onuchic et al., 1995) so that this lattice model can be used to simulate real

proteins more effectively than previous treatments. The landscape is characterized by three parameters:

- 1) The mean square interaction energy fluctuation, ΔE^2 . This parameter measures the ruggedness of the energy landscape.
- 2) A gradient towards the folded state, δE_s .
- 3) An effective configurational entropy, S_L , describing the search problem size.

The entity of these parameters is estimated from experimental data, theory, and folding simulations. The model also relies on the principle of minimal frustration, according to which the possible intramolecular interactions of a kinetically foldable protein have fewer conflicts than generally expected. The folding process is simulated at room temperature by examining the lattice kinetics by the Metropolis algorithm (Leopold et al., 1992). During the simulation, single residues are randomly moved to nearby vacant lattice sites, with the constraint that the covalent chain bonds remain intact. Potential energy and number of nonbonded contacts are measured at each step of the simulation. Small single-domain proteins with foldable sequences follow converging pathways which lead to the lowest energy and maximally compact conformers. Energy landscapes resulting from these simulations resemble a funnel in a potential energy-entropy space (Fig. 5-1). The main feature of folding, as described by this model, is the convergence of pathways which takes place as folding occurs. This

FIGURE 5-1: Folding energy landscape, better known as "folding funnel", according to the model of Wolynes *et al.* (Wolynes et al., 1995). Modified after (Onuchic et al., 1995).



implies a progressive reduction of the number of variables involved in defining the folding energy landscape, which therefore becomes “funneled” as folding occurs. The folding process is characterized by a progressive free energy loss, which arises from the fact that the potential energy (a quantity which is proportional to enthalpy) gradient completely overcomes the decrease in entropy. Entropy loss can be visualized as the decrease in funnel width as folding proceeds. A significant portion of the enthalpy/entropy collapse takes place upon formation of the molten globular forms, i.e., as a consequence of initial short range secondary structure formation during folding.

The reaction coordinate of the process follows the number of native contacts that are formed (defined as Q). Q ranges from 0 ($Q = 0$ for an unfolded protein) to 1 ($Q = 1$ for a native protein). This choice of reaction coordinate is questionable, according to the criteria established by Wigner in 1938 (Wigner, 1938), in line with transition state theory. Reaction coordinate has been defined by Wigner as the $(2N-1)$ -dimensional surface, separating the reactants region from the products regions of a $2N$ -dimensional phase space, with fewest recrossings, i.e., with the minimum flux across it. N represents the number of atoms composing the system.

The dips (i.e., the ruggedness), which are visible along the sides of the funnel, represent local minima in which polypeptides can become temporarily trapped. Experimental data entered in the model are what ultimately determines the extent, ruggedness and slope of folding funnels.

No single pathway takes place: multiple pathways in energy space are predicted by this model. However, a nice and realistic feature is that these pathways converge towards similar molecular conformations as the process progresses. This goes hand in hand with the observed entropy decrease as Q gets closer to 1. It is good to keep in mind that each folding funnel comprises many pathways.

One surprising feature of this model is that the predicted kinetic barriers for folding are very small. They amount to about 3 kT (kT is of the order of 0.6 kcal/mol. Its precise value depends on the temperature of the system). However, the authors claim that the process is slowed by the thermodynamic bottleneck due to the fact that multiple pathways take place at the same time. This point is not clearly stated by the authors, and it needs to be reexamined, in our opinion.

The funnel model predicts folding as taking place by an initial protein collapse, which is concomitant with secondary structure formation. This takes place in the upper region of the funnel, where entropy (i.e., protein mobility and lack of ordered structure) is still very elevated. It is predicted that this event takes place in less than a millisecond. As secondary structure gets established, one enters the molten globule region, which is characterized by about 30% native contacts ($Q = 0.3$). As entropy decreases and pathways converge, the transition state is approached. The late stages of folding consist of mere conformational rearrangements among relatively similar conformers, to find the lowest energy structure. These late stages, predicted by Onuchic and coworkers to be formed in times of the order of seconds, are the intermediates that most experimental studies have detected. It is interesting that this

model predicts that late intermediates are mainly due to final conformational rearrangements, while the most significant part of the process takes place much earlier. It is hoped that future experimental efforts, including studies currently underway in the Chan group, will shed light on the intriguing early folding events. It is likely that these are going to be harder to characterize in real proteins than in simple peptides, due to the inherently higher configurational entropy as the primary sequence gets longer. Local folding wells contribute to the ruggedness of the landscape, according to the funnel model. Experimental studies focusing on characterizing the landscape ruggedness and the distribution of nucleation sites during folding initiation seem particularly promising.

WHAT UNFOLDING EXPERIMENTS REVEAL (OR DO NOT REVEAL) ABOUT FOLDING

A typical folding model, which has been invoked in many experimental studies, is the so called two-state mechanism (Schellman, 1987). According to this model, a protein can only assume two classes of states: a) a folded state in which a polypeptide undergoes fluctuations among an ensemble of highly ordered structures, and b) an unfolded state which is composed of a large number of random conformations. According to the two-state model, the conformational transitions within each state are much faster than those among different states. This directly implies that the kinetic barriers among conformers within each state are much smaller

than the kinetic barriers among the two different states. Furthermore, this also suggests that the two barriers between folded and unfolded states, i.e., those relative to the folding and unfolding process, are of comparable magnitude. This can certainly be tested experimentally by verifying that the transition among the states is reversible. The expected folding funnel for this situation closely resembles a triangle.

We have brought this up here because we want to ask whether unfolding experiments can teach us anything about the folding mechanism. The question can be rephrased as to **whether unfolding pathways coincide with folding pathways**. Rapid conformational interconversions among folded and unfolded substates ensure that the transition will give the same ensemble of products no matter from which substate the process gets started. Now, within the simple two-state model, one could have either sequential or parallel pathways. In the case where pathways are sequential, and the process is reversible, both folding and unfolding experiments yield the same information. However, it is interesting to consider that, when parallel pathways are present, an overall-reversible process may have one or more virtually irreversible steps associated with it. For a protein, these steps, if present, may coincide with strongly-downhill folding sub-events. Although one could say that pathways in both directions are the same given that microscopic reversibility applies, the irreversible step(s) will only be appreciably traveled by the protein in the folding direction. In this case, the overall process may still be reversible, and microscopic reversibility apply, as long as alternative paths with lower free energy barriers are available in the unfolding direction. The funnel model shows, as expected, that folding occurs downhill. This is

consistent with the possibility that some irreversible or pseudo-irreversible pathways are present in the folding direction. In the case of a multistep mechanism, the situation is qualitatively the same. Therefore, in the presence of parallel processes, refolding may preferentially occur via a less steep pathway than folding.

It follows from the previous discussion that, even if the information gained by folding and unfolding experiments may not be identical, the study of both processes yields very useful complementary data to explore and experimentally characterize some features of protein energy landscapes. Furthermore, unfolding processes, being uphill, have better chances to capture early folding intermediates, which occur too fast in folding experiments to be detected. This is particularly fortunate in the rare cases where unfolding experiments give rise to multiexponential kinetic traces, indicating the presence of intermediates. This is precisely what happens for rubredoxin from *Pyrococcus furiosus* (RdPf), as we have seen in chapter 4. This is a lucky circumstance, probably to be ascribed to the slow overall kinetics displayed by this hyperthermophilic protein. One important point to be noticed is that, in general, only intermediates that occur before the rate-determining step of a reaction sequence can be kinetically detected. This may impose some restrictions to the detection of all intermediates involved in the folding/unfolding process.

Next, we would like to reexamine the results of chapter 4 in light of what we have learned on folding funnels. Our experiments detected two unfolding intermediates, which we called I_1 and I_2 , which are structurally not so different from the native state. I_1 corresponds to rubredoxin with some secondary structure

disruption, and I_2 is similar to I_1 except for the fact that iron has also been released. These intermediates may well be placed on the bottom of the folding funnel, where all of the discrete folding intermediates and, to some extent, the native state, are likely to be relatively similar to each other.

The third RdPf unfolding intermediate (I_3) has lost more secondary structure, and it resembles a collapsed early folding intermediate, which would have to be placed very high on the vertical coordinate of the folding funnel, somewhere above the molten globule region. This is very interesting because such an early folding intermediate is hard to detect along the folding pathway, with current experimental tools. Another interesting aspect of our findings is that the unfolding transition state, which lies between I_3 and the unfolded state, is late along the unfolding pathway. This is clearly in contrast with the predictions of the funnel model and, in general, with previous experimental findings on mesophilic proteins. This apparent contradiction may be resolved by considering that:

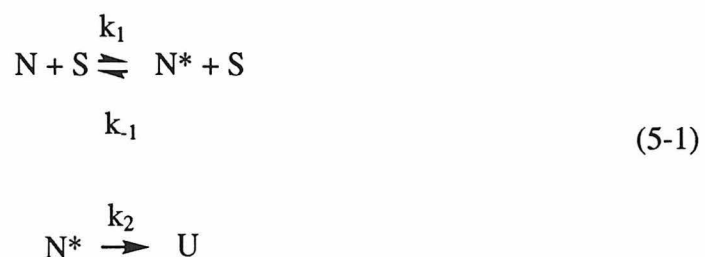
- a) The unfolding transition state may well be different from the folding transition state. This possibility has been mentioned before. If this is the case the folding funnel should be redrawn to indicate two transition states: one close to the unfolded form (for unfolding) and another close to the native state (for folding).
- b) Unfolding of hyperthermophilic proteins may simply have overall different features from that of mesophilic species.

Another interesting point is that our unfolding data are consistent with hydrophobic collapse preceding secondary structure formation during folding, at least at temperature greater than about 68°C (see Fig. 4-4). However, this may be another point in support of two different classes of pathways for the forward and the reverse processes.

PROTEIN UNFOLDING VIEWED ACCORDING TO UNIMOLECULAR REACTION RATE THEORY

Protein unfolding is essentially a unimolecular process of a polyatomic substrate. The following discussion considers the simple case of a two-step unfolding transition, starting from a single native protein conformer (N), and leading to a generic unfolded state (U). This allows neglecting the complications arising from multiple species and pathways.

As any chemical reaction, unfolding requires a given amount of internal energy, which is transferred to the protein by collision with one molecule of a chemically inert species (S). We can then describe unfolding as follows:

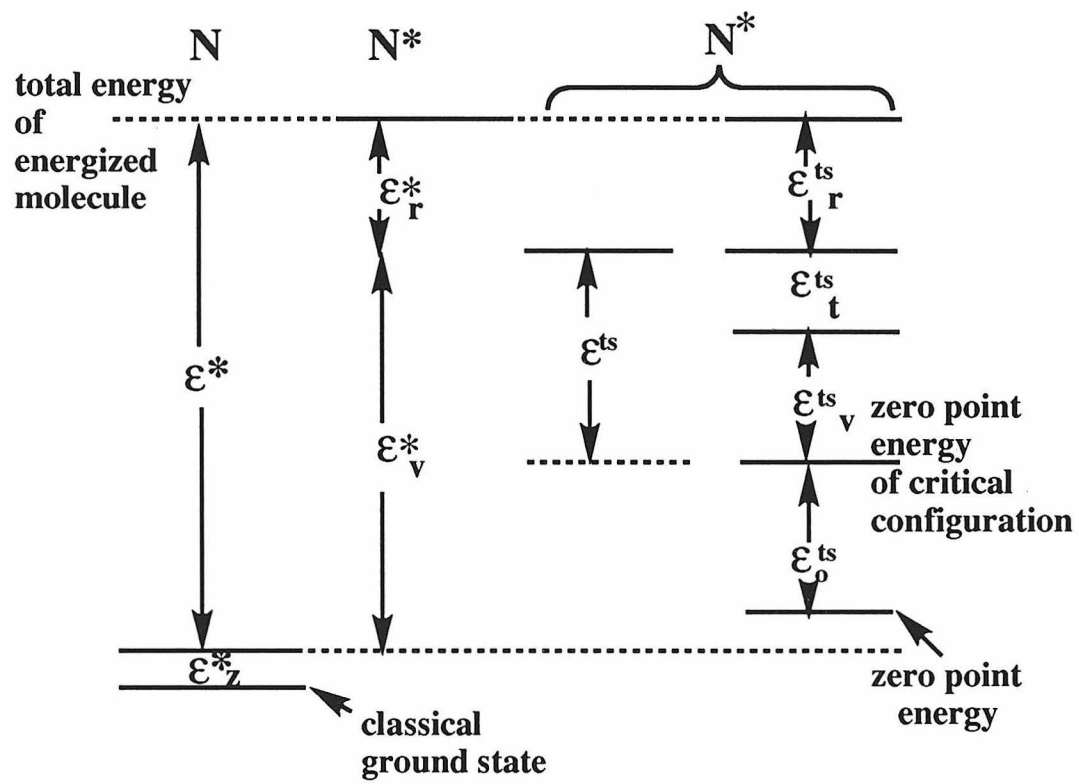


N^* represents the energized protein. The native state receives an excess energy by collision with a single chemically inert species, thus becoming energized. This excess energy is then rapidly and randomly distributed among the different vibrational and rotational normal modes of the protein. One of these modes corresponds to motion along the unfolding coordinate. Upon receiving the necessary energy in this special normal mode, unfolding is started. The energy levels involved in this process are outlined in Fig. 5-2. A detailed theory of gas phase reaction rates has been formulated by Marcus (Marcus, 1952), who extended an older theory by Rice Ramsperger and Kassel. Marcus' formulation is known as RRKM theory. It is based on statistical considerations and it has been used successfully to make predictions for gas phase processes. Interestingly, the rate constant k_2 , according to RRKM theory, depends on the number of vibrational degrees of freedom of the molecule that are involved in the reaction. Usually, all vibrational degrees of freedom give a contribution, for small molecules. This theory has been used successfully even for very large molecules at low pressures (Bernshtein & Oref, 1994). One could envision applying the RRKM formula

$$k_2 = \frac{1}{h} \frac{G(\epsilon^{ts})}{N(\epsilon^*_v)} , \quad (5-2)$$

where h is the Plank's constant, $G(\epsilon^{ts})$ is the sum of states for the active degrees of freedom in the transition state, and $N(\epsilon^*_v)$ is the density of states for the active degrees

FIGURE 5-2: Diagram of the energies for the RRKM theory. ϵ^* is the internal energy gained upon collision; ϵ_v^* and ϵ_r^* are the vibrational and rotational components of ϵ^* , respectively. $\epsilon^{ts} + \epsilon_0^{ts}$ is the total energy of the critical configuration of the transition state. ϵ_r^{ts} is the internal rotational energy of the transition state; ϵ_t^{ts} is the energy of translational motion along the reaction coordinate. Finally, ϵ_v^{ts} is the vibrational energy of the transition state.



of freedom in the reactant. $G(\epsilon^{ts})$ is equal to

$$G(\epsilon^{ts}) = \frac{(\epsilon^{ts} + a\epsilon_z^{ts})^{(s-1)}}{(s-1)! \prod_{i=1}^{s-1} h\nu_i^{ts}}, \quad (5-3)$$

where s is the number of vibrational normal modes, ϵ^{ts} is the energy of the critical configuration of the transition state, ϵ_z^{ts} is the zero-point energy of the critical configuration of the transition state, and ν_i^{ts} is a single adiabatic (i.e., active) vibrational frequency of the transition state. The parameter a is a function of s and ν_i^{ts} , and its value ranges between 0 and 1. $N(\epsilon^*_v)$ is equal to

$$N(\epsilon^*_v) = \frac{(\epsilon^* + a\epsilon_z^*)^{(s-1)}}{(s-1)! \prod_{i=1}^s h\nu_i^*} \left[1 - \frac{dW(\epsilon')}{d\epsilon'} \right], \quad (5-4)$$

where ϵ^* is the internal energy gained upon collision, ν_i^* is an adiabatic vibrational frequency of the reactant, ϵ' is a parameter whose value ranges between 0.1 and 1. Finally, $W(\epsilon')$ is a simple empirical function of ϵ' . Expression 5-4 is the widely used semiempirical expression for $N(\epsilon^*_v)$ by Rabinovitch and coworkers.

This treatment could lead to prediction of unfolding rates as a function of molecular size, and to a better understanding of the number and frequency of vibrational degrees of freedom that need to be excited to disrupt the protein molecule upon unfolding. It is expected that only low frequency vibrations (ranging from 40 to

a few hundreds of cm^{-1}) contribute to the process, since only weak noncovalent interactions need to be broken.

One problem that we encountered, when considering doing some of these calculations, is the application of the above concepts to a process that takes place in solution. In this case, scheme 5-1 needs to be modified to take into account the fact that a strong excess of S molecules (i.e., the solvent) is present. These S molecules are involved in strong energizing and deenergizing collisions with the reactant. To a first approximation, solution state can be approximated by high pressure gas phase. This corresponds, in practice, to a case where **all** steps in eq. 5-1 are fully reversible. However, RRKM theory merges into transition state theory at high pressures (i.e., $> 10^{-4}$ atm). Therefore, as opposed to the low pressure case, the high pressure situation does not regard reaction of the energized species N^* , which has a selected number of excited normal vibrational modes, as peculiar. Upon reequilibration of U to form N^* , the energizing preequilibrium shown in eq. 5-1 is no longer a necessary requirement to generate N^* in solution. It is interesting to notice that the unimolecular rate expression according to transition-state theory is a simple function of the ratio of the vibrational partition function of the reactants and the vibrational partition function of the transition-state. The vibrational partition function (Q_v) is, in turn, a function of the number of vibrational normal modes, according to the expression:

$$Q_v = \prod_{i=1}^s \frac{1}{1 - \exp(-h\nu_i / k_B T)} , \quad (5-5)$$

where ν_i represents a generic vibrational frequency and k_B is the Boltzmann constant. It is therefore conceivable that one could make unfolding rate predictions if it were possible to estimate Q_v , to a reasonable degree of approximation, for both the native state and the unfolding transition state. Only low frequency vibrations ($\nu_i \ll k_B T$) would have to be included in the calculation.

The application of reaction rate theories to learn about specific events that promote protein unfolding is an exciting new area of research, and we are certain that, in the near future, a lot of progress is going to be made in the field.

CONCLUDING REMARKS

This chapter has presented some of the implications of unfolding experiments for the protein folding problem. Our findings on RdPf can be interpreted in terms of the funnel model of protein folding. We have examined the possibility that the transition state for RdPf unfolding may differ from the transition state for folding. Unfolding kinetic data yield very valuable mechanistic information and important insights into the pathways that are available within the complicate protein energy landscape. Finally, we introduced the potential of RRKM theory to describe and predict protein unfolding kinetics in solution.

REFERENCES

- Bernshtein, V. & Oref, I. (1994) *J. Phys. Chem.* 98, 136.
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) 21, 167.
- Honeycutt, J. & Thirumalai, D. (1990) *Proc. Natl. Acad. Sci. USA* 87, 3526.
- Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* 89, 8721.
- Levinthal, C. (1969) in *Mossbauer Spectroscopy in Biological Systems* (De Brunner, P., Tsibris, J., Munck, E., Eds.) 22, University of Illinois Press, Urbana, IL.
- Levitt, M. & Warshel, A. (1975) *Nature* 253, 694.
- Marcus, R. (1952) *J. Chem. Phys.* 20, 359.
- Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3626.
- Schellman, J. A. (1987) *Ann. Rev. Biophys. Chem.* 16, 115.
- Shakhnovich, E., Farztdinov, G., Gutin, A. M. & Karplus, M. (1991) *Phys. Rev. Lett.* 67, 1665.
- Skolnick, J. & Kolinski, A. (1991) *J. Mol. Biol.* 221, 449.
- Wigner, E. (1938) *Trans. Faraday Soc.* 34, 29.
- Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995) *Science* 267, 1619.