A COMPARATIVE STUDY OF HPR PROTEINS FROM EXTREMOPHILIC ORGANISMS

A Dissertation

by

ABBAS RAZVI SYED ALI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2005

Major Subject: Biochemistry

A COMPARATIVE STUDY OF HPR PROTEINS FROM EXTREMOPHILIC ORGANISMS

A Dissertation

by

ABBAS RAZVI SYED ALI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	J. Martin Scholtz
Committee members	C. Nick Pace
	James C. Hu
	Gary R. Kunkel
Head of Department	Gregory D. Rienhart

December 2005

Major Subject: Biochemistry

ABSTRACT

A Comparative Study of HPr Proteins from Extremophilic Organisms.

(December 2005)

Abbas Razvi Syed Ali, B.Sc., Osmania University;

M.S., University of Pune

Chair of Advisory Committee: Dr. J. Martin Scholtz

A thermodynamic study of five homologous HPr proteins derived from organisms inhabiting diverse environments has been undertaken. The aim of this study was to further our understanding of protein stabilization in extremes of environment. Two of the proteins were derived from moderate thermophiles (*Streptococcus thermophilus* and *Bacillus staerothermophilus*) and two from haloalkaliphilic organisms (*Bacillus halodurans* and *Oceanobacillus iheyensis*); these proteins were compared with HPr from the mesophile *Bacillus subtilus*. Genes for three of these homologous HPr proteins were for the first time cloned from their respective organisms into expression vectors and they were over-expressed and purified in *Escherichia coli*. Stability measurements were performed on these proteins under a variety of solution conditions (varying pH, salinity and temperature) by thermal and solvent induced denaturation experiments. Stability curves were determined for every homologue and these reveal very similar conformational stability (ΔG) for these homologues at their habitat temperatures. The *Bst*HPr homologue is the most thermostable and also has the highest ΔG_{25}; the stability of other homologues was ranked as *Bst*>*Bh*>*St*>*Bs*>*Oi*HPr. Other key thermodynamic parameters, like ΔC_p, have been estimated for all the homologues and it was found that these values are identical within errors of estimation. Also, it was

found that the values of T_S are very similar for these homologues. Together these observations allow us to propose a thermodynamic mechanism toward achieving higher T_m.

The crystal structures of the *Bst*HPr and a single tryptophan-containing variant (*Bst*F29W) of this homologue are also reported here. Also reported is a domain-swapped dimeric structure for the *Bst*F29W variant, together with a detailed investigation into the solution oligomeric nature of this protein. The crystal structure of *Bst*HPr is analyzed to enumerate various stabilizating interactions like hydrogen bonds and salt-bridges and these were compared with those for the mesophilic homologue *Bs*HPr. Finally, an analysis of sequence alignments together with structural information for these homologues has allowed design of numerous variants of both *Bs* and *Bst*HPr. A detailed thermodynamic study of these variants is presented in an attempt to understand the origins of the differences in stability of the HPr homologues.

DEDICATION

to my parents,

for being such good examples and to my M.S. advisor, Dr. V. Sitaramam, for getting me interested in research

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Scholtz for all his faith in me, for allowing me to pursue projects of my choosing and his guidance, which helped immensely. I would like to thank Dr. Pace for his encouragement and invaluable advice. I greatly appreciate all the advice from my committee members Dr. Hu and Dr. Kunkel. I thank all members of my lab for their support, including past members Drs. Jason Schmittschmitt and Richard Thurlkill from whom I learned a great deal. I thank Jennifer Dulin for all her help at various stages of this work. I would also like to thank my friends, soon to be Drs. Peter and Suraj, for being great friends and my family for their support and patience.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES xi

CHAPTER

I INTRODUCTION 1

An overview of thermodynamic parameters 6
Construction of database 8
Classification of data and example cases 17
General observations on enhanced thermostability 24

II MATERIALS AND METHODS 27

Cloning of HPr homologues 28
Mutagenesis 36
Protein overexpression and purification 37
Thermodynamic stability measurements 40
Studies on protein oligomeric state 49

III COMPARATIVE STUDY OF HPr HOMOLOGUES 53

Introduction 53
Results 55
Discussion 72

IV STRUCTURAL ANALYSIS OF HPr HOMOLOGUES AND THEIR SOLUTION CHARACTERIZATION 79

Introduction 79
Methods for protein crystallography 80

CHAPTER Page

Results 82
Discussion 100

V DETERMINANTS OF THERMOSTABILITY IN *Bst*HPr 106

Introduction 106
Results 107
Discussion 120

VI SUMMARY 126

REFERENCES 129

VITA 154

LIST OF FIGURES

FIGURE Page

1. A stability curve for a hypothetical protein. 7
2. Stability curves showing different methods to achieve a higher T_m 9
3. A diagrammatic representation of the 'four-primer cloning method'. 30
4. A representative PCR amplification schedule, shown here for amplification of *St*HPr gene 31
5. Ribbon diagram of *Bst*HPr protein. 42
6. Circular dichroism spectra for native, denatured and renatured *St*HPr. 43
7. Representative thermal denaturation transition for *St*HPr 45
8. Representative urea denaturation transition for *St*HPr 47
9. A ribbon diagram showing the superposition of the wild-type *Bs* and *Bst*HPr structures 56
10. Representative thermal, urea and GuHCl denaturation unfolding curves for *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr 58
11. Stability curves (ΔG versus T) for the *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr proteins. 62
12. The changes in conformational stability (ΔΔG) as a function of [NaCl] for the *Oi*, *Bs*, *St*, *Bh*, and *Bst*HPr proteins. 65
13. Conformational stability, ΔG (25°C) of *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr proteins as a function of pH. 67
14. Sequence alignments for *Bs*, *St*, *Bst*, *Oi* and *Bh*HPr proteins 70
15. Ribbon diagram showing superposition of *Bs*, *Bst*, *Sf* and F29W *Bst*HPr structures. . 84
16. Ribbon diagram showing a superposition of structures for monomeric and dimeric forms of F29W *Bst*HPr. 86

FIGURE Page

17. Electron density maps for sections of the Gly54-Lys62 region in the monomeric (panel a) and dimeric (panel b) forms of the F29W *Bst*HPr. 88

18. NMR spectra from the diffusion measurements by pulsed filed gradient NMR on the F29W variant of *Bst*HPr..... 91

19. Radius of hydration (R_h) values measured by diffusion NMR experiments as a function of protein size. 93

20. The elution profile for a mixture of F29W and wild type *Bst*HPr from a typical size exclusion chromatography experiment..... 94

21. A typical equilibrium radial absorbance profile from analytical ultracentrifugation experiment on the F29W variant of *Bst*HPr..... 96

22. Ribbon diagram comparing the domain-swapped dimer structures of the Crh protein (panel a) and F29W variant of *Bst*HPr (panel b)..... 103

23. Estimation of ΔC_p for *Bst*HPr by a Kirchoff analysis of ΔH vs. T_m..... 109

24. Ribbon diagram of the *Bst*HPr showing the differences in sequence with respect to *Bs*HPr..... 112

25. Illustrative results from thermal, urea and GuHCl denaturation experiments on the wild-type *Bs* and *Bst*HPr proteins and their most stabilized and destabilized variants. 115

LIST OF TABLES

TABLE Page

1. A compilation of thermodynamic parameters for homologous proteins derived from mesophiles and thermophiles. 11

2. Primers used for cloning of *St*HPr gene into vector pLOI1803. 29

3. Primers used for cloning of *Oi*HPr gene into vector pUC (HPr). 33

4. Primers used for cloning of *Bh*HPr gene into vector pLOI1803. 34

5. Parameters characterizing urea, GuHCl and thermal denaturations for *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr at pH 7 59

6. Parameters characterizing the stability of the *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr ptoteins. 63

7. Summary of data collection and refinement statistics for *Bst*HPr, F29W *Bst*HPr and domain-swapped structure of F29W *Bst*HPr. 83

8. Parameters characterizing urea and thermal denaturation experiments on *Bs*, *Bst* and F29W *Bst*HPr proteins. 97

9. A comparison of structural features between *Bs*, *Bst* and F29W variant of *Bst*HPr. 99

10. Parameters characterizing urea, GuHCl and thermal denaturations and ΔC_p estimates for *Bs* and *Bst*HPr at pH 7 110

11. Compilation of stability data from urea, GuHCl and thermal denaturation experiments on variants of *Bs* and *Bst*HPr 116

CHAPTER I

INTRODUCTION

Life on earth extends to the entire surface of the planet except for centers of volcanic activity, from deep-sea hydrothermal vents to the heights of the Himalayas, from the boiling waters of hot springs to the cold expanses of Antarctica. Extremophiles are organisms inhabiting these extremes of environment. The nature of their adaptations is diverse and is the basis of their classification into various classes such as temperature adaptations - psychrophiles (cold adapted organisms) to hyperthermophiles, high salinity adaptations - halophiles, pH adaptations - acidophiles and alkaliphiles, pressure adaptation - barophiles to name a few. Proteins from extremophilic organisms have been the focus of several studies, starting with early studies by Perutz and colleagues (*1, 2*). Biochemists have tried over the years to understand how these organisms are able to survive these harsh environmental conditions and proteins have been recognized to play a vital role in their survival, as they do in all other organisms. In case of adaptations to extremes of pH, salinity and pressure, the role of membrane components and protective small molecules has been acknowledged (*3*) and studied in detail (*4*). For temperature adaptations, however, it has been recognized (*5*) that the environmental stress cannot be avoided by compensatory mechanisms involving membranes or small molecule and thus the cellular machinery (proteins and nucleic acids) has to adapt. For this reason, proteins from temperature-adapted organisms have been the focus of several studies and reviews (*6-11*).

This dissertation follows the style of *Biochemistry*.

Studies on protein stability are essentially explorations of the sequence-structure-stability relationship with stability being the measured thermodynamic quantity. Stability is, in essence, the consequence of structure, which defines interactions that afford stability. Sequence, in turn, defines structure as studied in the field of protein folding. Sequence is also the variable which organisms change as they evolve to adapt their proteins to the environments they inhabit. The sequence-structure-stability triad can also be viewed as a cause and effect phenomenon, where the sequence defines structure and structure allows interactions in three-dimensional space that produce stability. It is worth emphasizing however, that it is only the sequence that is altered by organisms to modulate their structures and hence stability characteristics.

Proteins perform important tasks in all biological systems, and they do so by maintaining a specific globular conformation. This functional state of proteins is marginally stabilized in a balancing act of opposing forces. The players in this balancing act have long been identified (*12*), although their relative contributions have been debated (*13-17*). The major stabilizing forces include the hydrophobic effect and hydrogen bonding while conformational entropy favors the unfolded state. The forces stabilizing the functional state outweigh the disruptive forces marginally in a folded protein, usually in the range of 5 to 10 kcal mol^{-1}. The conformational stability of a protein is thermodynamically defined as the free energy change, ΔG, accompanying the native $\leftrightarrow$ denatured state transition, and since the native state is usually taken as the reference state, the value of ΔG is positive for a stable folded protein. Measurements of and studies on protein stability have remained important over several decades owing to central role these macromolecules play in maintaining life, and their involvement in several diseases affecting humans. Traditionally the stability of

proteins has been estimated experimentally by perturbing the native state using temperature or solvent additives like chaotropes as denaturants and following this 'reaction' by direct (calorimetric) and indirect (spectroscopic) probes. For further detail, the reader is referred to a recent articles dedicated to this subject (*18, 19*).

The availability of three-dimensional structures of proteins has been instrumental to our knowledge of protein stability and the forces involved. Structures not only allow enumeration of stabilizing interactions like hydrogen bonds and electrostatic interactions, but they also aid in theoretical studies which attempt to correlate features like buried surface areas with magnitude of stabilizing forces like the hydrophobic effect (*20*). Also, structures have been used to measure ΔASA or change in solvent accessible surface area upon unfolding of a protein. The ΔASA has for long been correlated with thermodynamic quantities like ΔC_p (*21-23*) or change in heat capacity upon protein unfolding and m value or the [denaturant] dependence of the unfolding transition in solvent denaturation experiments (*24*). The availability of protein structures has also reinforced the fact that the folded conformation is held together by large numbers of weak non-covalent interactions between the amino acids that compose them. The fact that proteins are only marginally stable in general makes the ability of thermophilic proteins to function particularly intriguing because unlike membranes, which show some heterogeneity in their building block lipids (*25*), proteins are composed of the same 20 amino acids irrespective of the organism's habitat. Although a few amino acid modifications, like lysine methylation (*26, 27*) have been reported, specifically in the genus *Sulfolobus,* their contributions to protein stability are debated (*28, 29*).

The interest in proteins from thermophiles is multifaceted, from applications in industry, where proteins have been applied as catalysts in various processes as diverse as starch processing (*30*) to petroleum refining (*31*), to more fundamental studies of academicians trying to understand how these proteins are able to function at extremes of temperature. Enzymes from hyperthermophilic organisms have proven vital to the success of PCR based techniques (*32*); other enzymes from these organisms have been used widely in industry (*33*).

Over the years, researchers have studied pairs of homologous proteins from mesophilic and thermophilic organisms in an attempt to understand the difference(s) in their thermostability. These studies have compared the three-dimensional structures whenever possible, counted the number of stabilizing interactions, made mutations and have tried to come up with an unified theory of thermostabilization which would explain the ability of proteins from thermophiles to function in their environments. Although these attempts have been largely unsuccessful in providing such a unifying theory, they have provided us with some strategies proteins employ to maintain functionality at extremes of temperature. These strategies include presence of increased number of hydrogen bonds, salt bridges, improved core packing, shorter and or tighter surface loops, enhanced secondary structure propensities, oligomerization among others. There have been numerous reviews and studies that have compiled successes and failures of these studies (*34-38*). Information gleaned from these studies has been used to rationally design variants of proteins with desired properties (*39, 40*). Computational algorithms have also been used with some success to rationally design proteins with enhanced thermostability (*41-43*). Another relatively novel technique used to design proteins is directed evolution; this technique is especially

advantageous because it requires no *a priori* knowledge of stabilizing/destabilizing interactions like the rational design methodology. Directed evolution can sometimes be used to evolve enzyme activity under a desired condition, for example at high temperature, extremes of pH, salinity etc. This technique has been applied to a large number of proteins with a good measure of success (*44-50*).

Several studies have found that proteins from thermophiles use different thermodynamic strategies for stabilization in terms of the way they modulate key parameters like ΔC_p, the change in heat capacity that accompanies the protein unfolding reaction, ΔH, the change in enthalpy associated with protein folding and ΔG, the free energy of protein stabilization. Proteins from thermophiles alter their sequence in a way such that it optimizes the interactions holding their conformations together; these optimizations in turn alter the thermodynamic parameters mentioned above in a way that 'tunes' the stability characteristics to the organism's habitat. As protein chemists, we can measure these cardinal parameters and learn about the strategies employed in thermostabilization.

Here we have compiled results from studies reporting thermodynamic characterization of proteins (or domains) from thermophilic species. More specifically, thermodynamic parameters have been compiled so comparisons can be made with values from mesophilic homologues where possible and conclusions drawn on the mode of thermostabilization employed in each case. Moreover, we hope to draw conclusions based on these results with respect to the most common methods of stabilization employed by thermophilic proteins and possible reasons for the same.

An overview of thermodynamic parameters

In this review, we have compiled data from studies where stability curves are reported for thermophilic proteins and, preferably, from reports that compare pair of homologous proteins from mesophilic and thermophilic organisms. Since this review focuses on thermodynamics of protein stability and protein stability curves in particular, a brief introduction to these concepts is in order. Becktel and Schellman (*51*) introduced protein stability curves, showing plots of free energy of stabilization (ΔG) as a function of temperature (see Figure 1 for a representative stability curve). Such data are described by a modified version of the Gibbs-Helmholtz equation and important thermodynamic parameters can be determined (equation 1).

$$\Delta G(T) = \Delta H_m\left(1 - T/T_m\right) - \Delta C_p\left[\left(T_m - T\right) + T\ln\left(T/T_m\right)\right] \qquad (1)$$

where, ΔG(T) is the free energy at a temperature T, ΔH_m is the enthalpy at T_m, ΔC_p is the change in heat capacity associated with the unfolding of the protein, T_m is the melting temperature or the temperature at mid-point of transition from native to denatured state in a thermal denaturation. Other parameters of interest that can be calculated using modifications to equation (1) include T_S and ΔG_S, where T_S is the temperature of maximum stability or temperature where the change in entropy between native and denatured states is zero and ΔG_S is the conformational stability at this temperature.

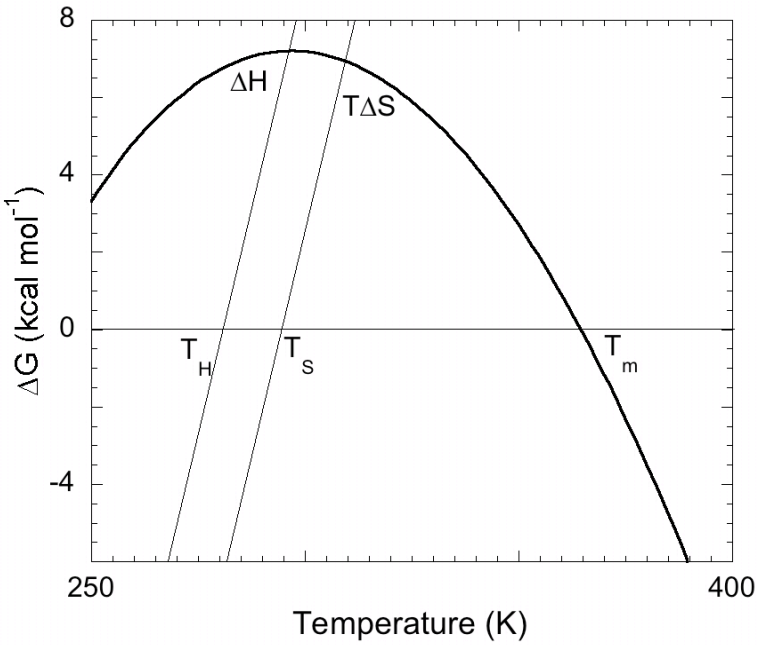


Figure 1: A stability curve for a hypothetical protein. Some key thermodynamic parameters are also marked on the plot. For an explanation of the terms used, refer to the text.

Nojima et al. (*52*) first proposed three different methods of modulating the stability curve of a protein to achieve higher thermostability. For a description of these three methods see Figure 2. Briefly, a hypothetical mesophilic protein can I) raise the stability curve to higher ΔG so it now has a higher T_m II) broaden its stability curve so it now intersects the abscissa at a higher temperature (a higher melting temperature, T_m) or III) shift the stability curve to the right (to higher temperatures). All three methods of achieving higher thermostability have been observed in nature, some independently and others in combination. Each of these methods has an underlying thermodynamic explanation. For example, increasing the value of ΔH_S (the change in enthalpy measured at T_S) without compensating changes in ΔS will result in a similar stability curve, but with higher ΔG values at all temperatures (method I). A broadened stability curve (method II) is caused by a reduced ΔC_p. Lowering the ΔS or the change in entropy for the folding transition shifts the T_S (temperature of maximum stability or the temperature at which change in entropy is zero) to higher temperatures and has the effect of shifting the stability curve to the right (method III).

Construction of database

We performed a literature search to find experimental thermodynamic characterization of proteins from thermophilic organisms. Our search focused on studies reporting a comparison of thermodynamic data on homologous proteins from thermophiles and mesophiles. These results were supplemented with results from a search for publications that cited the pioneering work of Nojima et al. (*52*). Finally, the results of literature search were augmented with data from the Protherm database (*53*) where possible.

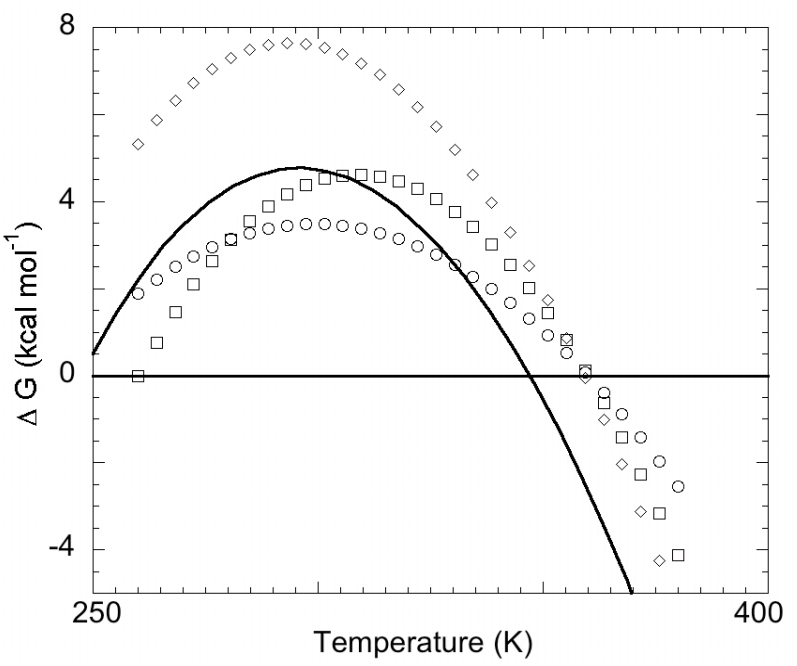


Figure 2: Stability curves showing different methods to achieve a higher T_m. Starting with a stability curve for a hypothetical mesophilic protein (solid line), the protein may increase T_m by shifting the curve up (method I, diamonds), by making the curve flatter (method II, circles), or by shifting the curve to the right (method III, squares).

In all, we were able to find 23 proteins, for which conclusions concerning the thermodynamic mode of stabilization have been made or can be, based on the information provided. To this set of proteins we have added the three cases of the HPr homologues, results for which are reported in Chapter III. The data for these 26 cases of protein homologues have been compiled in Table 1.

Of the 26 sets of proteins for which data has been compiled, 19 cases make comparisons with a homologous protein from a different organism, four make comparisons with a collection of similarly sized proteins and the remaining three do not make comparisons with other proteins. Most of the data is from CD and DSC experiments, either in the presence or absence of denaturants like urea or GuHCl; only in the case of ferredoxin proteins HD exchange NMR was used.

The transitions from native to denatured state followed a two state behavior for most proteins in our compilation. Although 9 proteins were dimers in solution, they were shown to follow the two-state model of folding, and the remaining 17 were monomeric and also followed the two-state model. There is however one pair of proteins namely the isopropyl malate dehydrogenases from *Thermus thermophilus* and *Escherichia coli*, where the former follows a two state and the latter follows a three-state unfolding model (entry 13 in Table 1).

Table 1: A compilation of thermodynamic parameters for homologous proteins derived from mesophiles and thermophiles. The data was compiled from a literature survey as described in the text.

	Protein	Source[a]	PDB id (#Residues)	T_m (ºC)	ΔC_p [b] (kcal mol^{-1} K^{-1})	ΔG [c] (kcal mol^{-1})	T_S [d] (ºC)	Mode of Stabilization[e]	Reference
1	CspB	*Thermotoga maritima*	1G6P (67)	83.1	0.22			I and II*	(*54*), (*55*), (*56*)
		Bacillus subtilus	1CSP (67)	54	0.88	ΔG_S: *5.5*	*-6*		
2	Rnase H	*Thermus thermophilus*	1RIL (166)	86	1.8	ΔG_S: 12.7	20	I and II	(*57*),(*58*)
		Escherichia coli	1JXB (155)	66	2.7	ΔG_S: 7.5	24		
3	SSo7D	*Sulfolobus solfataricus*	1SSO (64)	98	0.63	ΔG_S: 7	9	I and II*	(*29*),(*59*)
		SH3 domain proteins		69-80	0.69-0.81	ΔG_{25}: 3-4	16-23		
4	CheY	*Thermotoga maritima*	1TMY (120)	101	1.17	ΔG_S: 9.54	29	I and II	(*60*)
		Bacillus subtilus	– (120)	55	2.34	ΔG_S: 3.15	27		
5	Ribosomal protein L30E	*Thermococcus celer*	1H7M (100)	93.8	1.27	ΔG_S: *12*	*17*	I and II	(*61*)
		Saccharomyces cervisiae	1CN7 (104)	45.7	2.51	ΔG_S: *3.5*	*17*		

Table 1: Continued

	Protein	Source[a]	PDB id (#Residues)	T_m (ºC)	ΔC_p [b] (kcal mol^{-1} K^{-1})	ΔG [c] (kcal mol^{-1})	T_S [d] (ºC)	Mode of Stabilization[e]	Reference
6	Aspartate aminotransferase	*Sulfolobus solfataricus*	(362)		2.87	ΔG_{25}: 16.8		I and II	(*62*)
		Cytosolic pig heart	(413)		4.78	ΔG_{25}: 13.8			
7	Phosphoglycerate kinase	*Thermotoga maritima*	1VPE (398)	85		ΔG_{20}: 28.9		I and II*	(*63*) (*64*), (*65*)
		Saccharomyces cervisiae	3PGK (416)	60		ΔG_{20}: 6			
8	Phosphoglycerate kinase	*Thermus thermophilus*	1V6S (390)			ΔG: 12		I and II*	(*52*)
		Saccharomyces cerevisiae	3PGK (416)	60		ΔG_{20}: 6			
9	Sac7d	*Sulfolobus acidocaldarius*	1SAP (66)	91	0.5–0.86 [f]	ΔG_S: *6.5*	*22*	II	(*66*)
		Other small proteins		54-90	0.62-1.6	ΔG_{25}: 2.3-8.9			
10	Ssh10B	*Sulfolobus shibattae*	– (97)	125	0.9	ΔG_{25}: 30.1	*15*	II	(*67*), (*68*)
		Arc repressor-bacteriophage P22	1ARQ (106)	54	1.53	ΔG_{25}: 11	*19*		

Table 1: Continued

	Protein	Source[a]	PDB id (#Residues)	T_m (ºC)	ΔC_p [b] (kcal mol^{-1} K^{-1})	ΔG [c] (kcal mol^{-1})	T_S [d] (ºC)	Mode of Stabilization[e]	Reference
11	HU DNA binding protein	*Thermotoga maritima*	1B8Z (90)	101.9	0.76	ΔG_S: 6.8	20	II*	(*69*)
12	Glutamate dehydrogenase domain II	*Thermotoga maritima*	2TMG (149)	69.5	1.4	ΔG_S: 3.7	33	II	(*70*)
13	Isoprpopyl malate dehydrogenase	*Thermus thermophilus*	1OSI (345)	57	1.73	ΔG_S: 15.8	31	II	(*71*)
		Escherichia coli	1CM7 (363)	107	20.7	ΔG_S: 32.7	25		
14	Histidine containing protein (HPr)	*Bacillus staerothermophilus*	1Y4Y (88)	88.9	1.37	ΔG_S: 8.2	27	I	(*72*)
		Bacillus subtilus	2HPR (88)	74.4	1.33	ΔG_S: 5.2	24.1		
15	Histidine containing protein (HPr)	*Streptococcus thermophilus*	– (87)	77	1.28	ΔG_S: 6.3	22	I	see Chapter III
		Bacillus subtilus	2HPR (88)	74.4	1.33	ΔG_S: 5.2	24.1		
16	Histidine containing protein (HPr)	*Bacillus halodurans*	– (87)	82.2	1.3	ΔG_S: 6.7	25	I	see Chapter III
		Bacillus subtilus	2HPR (88)	74.4	1.33	ΔG_S: 5.2	24.1		
17	Histone proteins (rHMfa)	*Methanothermus fervidus*	1B67 (68)	104	2.16	ΔG_S: 15.5	35	I	(*73*)
		Methanobacterium formicicum (rHFoB)	– (67)	74.8	2.55	ΔG_S: 7.2	32		
18	Cellulase	*Themomonospora fusca*$E2_{cd}$	1TF4 (~270)	72.2		ΔG_{30}: 11.2		I*	(*74*)
		Cellulomonas fimi $CenA_{P30}$	1GU3 (~270)	56.4		ΔG_{30}: 4.3			

Table 1: Continued

	Protein	Source[a]	PDB id (#Residues)	T_m (ºC)	ΔC_p [b] (kcal mol^{-1} K^{-1})	ΔG [c] (kcal mol^{-1})	T_S [d] (ºC)	Mode of Stabilization[e]	Reference
19	Histone (rHMfB)	*Methanothermus fervidus*	1A7W (69)	113	1.87	ΔG_S: 14.6	40	I, II and III	(*73*)
		Methanobacterium formicicum (rHFoB)	– (67)	74.8	2.55	ΔG_S: 7.2	32		
20	Rubredoxin	*Pyrococcus furiosus*	1BRF (53)	176-195		ΔG_L: 15	*50*	I, II* and III	(*75*), (*76*)
		GB1, BPTI and *C. pasteurianum* rubredoxin							
21	O^6Methyl guanine- DNA methyl transferase (MGMT)	*Thermococcus kodakiensis*	1MGT (174)	98.6	1.24	ΔG_S: 10.2	29.5	I, II* and III	(*77*)
		E*scherichia* coli C terminal domain of ADA	1SFE (180)	43.8	1.77	ΔG_S: 4	7.4		
22	Ferridoxin	*Thermotoga maritima*	1VJW (60)	125.4	0.86	ΔG_S: 9.32	45	I, II* and III	(*78*)
		Bovine (fdx), Rabbit ($Cytb_5$), Yeast (isoCyt C)		50-70		ΔG_S: 4.8–6	0–25		

Table 1: Continued

	Protein	Source[a]	PDB id (#Residues)	T_m (ºC)	ΔC_p [b] (kcal mol^{-1} K^{-1})	ΔG [c] (kcal mol^{-1})	T_S [d] (ºC)	Mode of Stabilization[e]	Reference
23	Histone protein (rHPyA1)	*Pyrococcus* GB3a	(67)	114	2.39	ΔG_S: 17.2	44	I and III	(*73*)
		Methanobacterium formicicum (rHFoB)	(67)	74.8	2.55	ΔG_S: 7.2	32		
24	DHFR	*Thermotoga maritima*	1CZ3 (172)		5.35	ΔG_{25}: 30.11	41	I and III	(*79*), (*80*), (*81*)
		Escherichia coli	1RX1 (159)			ΔG_{25}: 6.1	15		
25	Cytochrome c-552	*Thermotoga maritima*	1C52 (131)		1.2–1.7	ΔG_{25}: 28.5		I* and III	(*82*)
		Bovine, Horse and *C kruseii*			1.5–2.5	ΔG_{25}: 13–15			
26	Farnesyl diphosphate/geranyl-geranyl diphosphate synthase	*Thermococcus kodakiensis*	– (343)	91	2.03	ΔG_S: 3.82	60	II and III*	(*83*)

[a] Source of the homologous proteins. The name of the organism is provided in cases where homologous proteins were compared; the upper row contains information on the thermophilic homologue. In other cases, where the thermophilic protein was compared with a collection of similarly sized proteins, the names of the proteins and their sources have been provided.

[b] Heat capacity change associate with protein unfolding. In cases where comparisons were made with multiple proteins, the range of ΔC_p values is given with lower and upper bound values. This convention is followed in all the data columns.

Table 1: Continued

[c] Free energy of stabilization (ΔG) which is usually ΔG_S or the ΔG at T_S. In other cases, the subscript indicates the temperature at which ΔG was measured. ΔG_L is the ΔG at T_L, the habitat temperature of the organism in question. Values in italics indicate values that have been calculated using variations of the Gibbs-Helmholtz equation.

[d] The temperature of maximum stability or the temperature where change in entropy between native and denatured states is zero. Values in italics indicate values that have been calculated by us using variations of the Gibbs-helmholtz equation.

[e] Method of stabilization as postulated by Nojima et al. (*52*). The three methods are:

Method I: Higher overall ΔG, shifting the curve up; Method II: Reduced ΔC_p, flattening the curve or Method III: Higher T_S, shifting the curve to the right. Some entries are marked with asterisks; these distinguish cases where the conclusions are based on insufficient data or entries where magnitude of difference between homologues is too small to support the conclusion drawn.

[f] see text for an explanation of ΔC_p values reported for this protein.

In looking at all the data, we found the average increase in T_m between the mesophilic and thermophilic protein was 32ºC, for the 15 cases where data are available. In the case of ΔG, data were available for 19 of the 26 cases and the average increase in ΔG of stabilization for the thermophilic homologue was 8.8 kcal mol^{-1}; (for the ΔG values, the average reported includes ΔG values listed in column 6 of Table 1 irrespective of the temperature at which they were measured; for both ΔT_m and ΔΔG values entry 13 was not included). We also compared the sequence identity for the protein homologues where sequences were available. For this purpose, we used the CLUSTALW (*84*) program as implemented on the EBI server (http://www.ebi.ac.uk/clustalw/) with default settings. For the 16 cases where alignments were possible, the average sequence identity was 51%. The alignment scores varied from 11% identity (for Aspartate aminotransferases from *Sulfolobus solfataricus* and pig heart cytosol), to 82% (for Histone proteins rHMfa and rHFoB). The average value of 51% identity appears quite high considering the diversity of the proteins and the sources they are derived from.

Classification of data and example cases

We have classified the proteins compiled here based on the methods for thermostabilization proposed by Nojima et al. (*52*). However, classification into the three groups as proposed originally by these authors was not entirely possible because proteins use different combinations of these methods making three classes inadequate. Also, for one of the methods proposed (method III), representative examples were not found in literature. For these reasons we have grouped proteins based on the methods of stabilization or combinations thereof used and arranged these groups in descending order

of the number of occurrences, with the group with highest number of reported occurrences first and so on. Based on this method of classification, we have six groups of proteins. What follows is a brief description of each of these groups with details of an example study for each case.

Stabilization by increased ΔG and reduced ΔC_p (methods I and II)

For the set of proteins included in this study, it was found that the combination of increased ΔG and reduced ΔC_p was the most commonly used way to achieve a higher T_m. Of the 26 sets of proteins, there are eight cases where this combination of stabilizing effects has been identified as the cause for enhanced thermostability in the thermophilic homologue. The proteins in this group are quite diverse with a range of sizes from the small (67 residue) cold shock proteins to the large (398 residues) glycolytic enzyme phosphoglycerate kinase. This spread of size and function in proteins belonging to this group shows that this combination of methods for enhancing thermostability is widely applicable.

A representative example from this group is the RNase H enzyme from *Thermus thermophilus (Tt)* and *Escherichia coli (Ec)* (*57*). RNase H is a small enzyme, which cleaves RNA from RNA-DNA hybrids; the protein from the thermophile (*Tt* RNase H) is 166 residues in length and shares a high sequence identity (52%) with its 155 residue mesophilic homologue (*Ec*RNase H). High-resolution structures are also available for the two proteins and the structures are very similar. To understand the thermodynamic basis of the difference in stability between the two proteins, GuHCl denaturation experiments as a function of temperature were performed to obtain stability curves for these proteins. The data reveal

that *Tt*RNase H is indeed more thermostable and it is so because of a lowered ΔC_p and a higher ΔG over a broad range of temperatures. The ΔC_p for the thermophilic protein *Tt* RNase H is 0.9 kcal mol K^{-1} lower than that for *Ec*RNase H. The ΔG_S, the maximum free energy of stabilization for the thermophilic protein is over 5 kcal mol^{-1} higher than that of the mesophilic homologue, the T_S however, is very similar for the two proteins.

Stabilization by reduced ΔC_p (method II)

For the set of proteins included in this study, stabilization by reduced ΔC_p was the second most common method to attain a higher T_m. Five proteins showed a lowered ΔC_p compared to their mesophilic homologues and thus have broader stability curves and thus remain folded over a wider range of temperatures. Proteins in this group show some diversity in terms of their size and function, from the small DNA binding proteins like the 66 residue Sac7d to the large enzyme isopropyl malate dehdrogenase (IPMD) which contains 345 residues. Three of the five proteins in this group are small DNA binding proteins, one is a sub-domain of the enzyme phosphoglycerate kinase and the last protein in this group is IPMD.

As a representative example of this group, consider the Sac7d protein from *Sulfolobus acidocaldarius* (*66*). Sac7d is a small DNA binding protein that is highly basic and whose structure has been solved by NMR spectroscopy. In the absence of a mesophilic homologue for Sac7d, comparisons have been made with a variety of small proteins (*66*). Sac7d is stable of over a broad range of pH (0 to 10) and DSC experiments have been performed over this pH range to estimate a ΔC_p value from Kirchoff analysis. Solvent denaturation experiments with GuHCl were also performed and a global fit to this data

yielded a ΔC_p. This estimate for ΔC_p was found to be higher than that obtained from the DSC data and the authors provide an excellent discussion on possible causes for this disparate observation (*66*). In any case, either value of ΔC_p produces stability curves that look very similar to those for other mesophilic proteins and the use of either value of ΔC_p does not cause significant differences in stability at 80ºC, the habitat temperature of the organism. Estimates of free energy of stabilization at 80ºC, the T_L (habitat temperature of the organism) reveal that the protein is only marginally stable at this temperature (1.6 kcal mol^{-1}). The lower estimate of ΔC_p however, helps explain the enhanced thermal stability of this protein and is also the reason for its inclusion in this group.

Stabilization by a higher overall ΔG (method I)

For the set of proteins in this study, stabilization by higher ΔG was found to be just as common as stabilization by reduced ΔC_p. In five of the 26 cases in this study, proteins from thermophilic organisms showed a higher ΔG over a broad range of temperatures compared to their mesophilic homologues, thus shifting the stability curve up and achieving a higher T_m in the process. This group contains the three cases of HPr homologues, two pairs of small archaeal histones and the enzyme cellulase. Most proteins in this group are small in size, like the archaeal histones (67 residues) or HPr homologues (88 residues) except for the cellulase catalytic domains which are both ~270 amino acids long. The composition of this group, with three of the six proteins being HPr homologues, precludes much insight into the diversity of this class in terms of both size and function.

As a representative example from this group, consider the HPr proteins from *Bacillus staerothermophilus* (*Bst*) and *Bacillus subtilus* (*Bs*) (see Chapter III for their full

characterization). HPr or histidine containing protein is involved in the bacterial phosphotransferase system (PTS) sugar transport pathway. The thermophilic variant is from *Bacillus staerothermophilus* (T_L: 58ºC) and the mesophilic variant is from *Bacillus subtilus* (T_L: 20ºC). The *Bs*HPr protein is the same size (88 residues), shows high sequence identity (72%) and has a very similar structure as the mesophilic homologue. The *Bst*HPr protein, however, has a higher T_m (~15ºC) and a larger ΔG_S (~3.2 kcal mol^{-1}). The complete analysis of the stability curves from denaturation experiments reveal that the ΔC_p values for the two HPr proteins that are very similar at 1.3 kcal mol^{-1} K^{-1}. The T_S value for the two proteins also are similar at 24 and 27ºC (*Bs* and *Bst*HPr respectively). Therefore, this pair of protein is a nearly perfect example for stabilization by method I or by a higher overall ΔG.

The archaeal histone proteins use different methods to gain thermostabilty

We now consider the case of four archaeal histone homologues studied by Li et al. (*73*) three of these homologues were derived from hyperthermophilic archaea and the fourth from a mesophilic archaeon. The thermophilic histones, rHMfa, rHMfB (both derived from *Methanothermus fervidus*) and rHPyA1 (from *Pyrococcus* strain GB3a) were compared to the mesophilic homologue rHFoB (from *Methanobacterium formicicum*). The curious feature of these proteins is that each of the thermophilic homologues uses a different thermodynamic approach to achieve a higher T_m compared to the mesophilic homologue. The archaeal histone rHMfa for example, belongs to the group of proteins discussed above, where thermostabilization is achieved by a higher ΔG (method I).

The next group of thermophilic proteins achieves higher T_m by a combination of all three methods, and the archaeal histone rHMfB is representative of this class. This histone

attains a higher T_m by combining a reduced ΔC_p with higher T_S and ΔG (methods I, II and III). Three other thermophilic proteins were found to use this approach to increase their T_m; two of them are metal cluster containing proteins ferridoxin and rubredoxin. The third is O^6Methyl guanine DNA methyl transferase enzyme, which is also the largest protein in this group (174 residues). Given the composition of this group, it appears that smaller proteins commonly use this strategy for thermostabilization.

As mentioned earlier, all thermophilic archaeal histones were compared with their mesophilic homologue rHFoB. rHFoB shares 80% sequence identity with rHMfB, both proteins like the other histones studied here, are dimers in solution and unfold in a two state manner to two unfolded monomers. Both DSC experiments and circular dichroism experiments have been used to construct stability curves for these proteins and the two ΔC_p estimates from these experiments are in good agreement with theoretical estimates (*23*) for the thermophilic homologue for which a structure is available. The ΔC_p for rHMfB, the thermophilic homologue is 1.9 kcal $mol^{-1}K^{-1}$ and is lower than that for rHFoB (2.6 kcal $mol^{-1}K^{-1}$). The ΔG_S for the thermophilic rHMfB is 14.6 kcal mol^{-1}, which is more than twice that for rHfoB the mesophilic homologue. This difference in ΔG_S causes the stability curve for rHMfB to be shifted to higher ΔG values. The T_S for rHMfB (40ºC) is significantly greater than that for rHFoB (32ºC), this difference in T_S causes the stability curve for rHMfB to be shifted to higher temperatures. Together these thermodynamic features of the rHMfB protein provide for a combination of the three methods of stabilization, causing the T_m to be 113ºC, significantly higher than that for rHFoB (74.8ºC).

Finally, the third thermophilic histone homologue rHPyA1 represents a small group of proteins that achieve high T_m by combining a higher overall ΔG with a higher T_S

(methods I and III). Two other thermophilic proteins use this approach to achieve higher T_m: DHFR and Cytochrome c-552 from *Thermotoga maritima*. rHPyA1 is the same length as the mesophilic homologue, rHFoB (67 residues) and shares 57% sequence identity. DSC and CD thermal melts were used to obtain ΔC_p estimates for the two proteins, these values are very similar for the two proteins and are identical within error. The ΔC_p for rHPyA1 is 2.4 kcal mol^{-1} K^{-1} and for rHFoB is 2.6 kcal $mol^{-1}K^{-1}$. The ΔG_S for rHPyA1 is however much larger at 17.2 kcal mol^{-1} than the 7.2 kcal mol^{-1} for rHFoB. The T_S for rHPyA1 is 44ºC, which is significantly higher than that for rHFoB (32ºC). The high ΔG_S and T_S for rHPyA1, in comparison with the mesophilic homologue rHFoB, make this protein a good example for thermostabilization by a combination of methods I and III.

Stabilization by reduced ΔC_p and higher T_S (methods II and III)

The only case of a thermophilic protein using a combination of a reduced ΔC_p and a higher T_S to achieve a higher T_m is farnesyl diphosphate/geranylgeranyl diphosphate synthase. This is a dimeric enzyme from *Thermococcus kodakiensis* that has 343 residues (*83*). GuHCl denaturation experiments performed at different temperatures have been used to measure ΔG and the data have been combined to construct a stability curve for this enzyme. Other thermodynamic parameters like, T_m (91ºC), ΔC_p (2 kcal mol^{-1} K^{-1}), ΔG_S (3.8 kcal mol^{-1}) and T_S (60ºC), were estimated from the stability curve. From comparisons with other thermostable proteins, the authors conclude that their enzyme has achieved a higher T_m by a combination of lower ΔC_p with higher T_S. No details of the comparisons made have been provided.

General observations on enhanced thermostability

A comprehensive collection of data comparing protein homologues from thermophiles and mesophiles was presented here. This broad compilation lends itself to some conclusions regarding the methods of thermostabilization adopted by proteins from thermophiles. It has allowed a ranking of different modes of thermostabilization originally proposed by Nojima et al. (*52*) in terms of their occurrence. We present here the results of this ranking and offer some interpretations of the findings.

The most common way to attain a higher T_m in proteins from thermophiles is to raise the stability curve to higher values of ΔG or to higher intrinsic stability at all temperatures; 77% of the thermophilic proteins in this study use higher ΔG either independently or in conjunction with other stabilizing effects as a way to attain a higher T_m. The next most popular method to attain a higher T_m is to lower ΔC_p, or to make the stability curve flatter (70% of thermophilic proteins in this study). Lastly, the lowest number of occurrences (31% of the thermophilic proteins in this study) was reported for cases where the thermophilic protein exhibits a higher T_S compared to the mesophilic homologue. However, as mentioned earlier in the introduction, these modes of stabilization namely the reduction of ΔC_p, increase in T_S or increased ΔG, are only effects caused by changes in the sequence which evolution has iterated and perfected over time. Thus, it is pertinent to rationalize the observations regarding the popularity of the different modes based on this evolutionary perspective, and how it would be more difficult to adopt one strategy rather than the other in terms of changes to the sequence.

It appears that stabilization by shifting the stability curve to higher temperatures by means of a higher T_S might require specialized changes to the sequence because a higher T_S is achieved by reducing the ΔS, or change in entropy between the folded and unfolded states. Since this requires that either the entropy of the denatured state be reduced relative to the folded state or the entropy of the folded state be enhanced to more closely match that of the denatured state, it will require rather precise changes in the sequence that affect one of the two states only. This could be a reason for the low number of occurrences of stabilization reported for method III or increased T_S. For example, constraining a certain loop in the denatured state by introduction of proline residues or introduction of glycine residues in structured regions of a protein can reduce ΔS, by reducing configurational entropy of the denatured state or increasing it for the native state respectively. Specialized mutations like these will reduce the ΔS of folding resulting in higher T_S. The other curious feature of proteins stabilized by high T_S was that a majority of them were non-enzymatic proteins. This suggests these proteins might be more tolerant of rigid conformations that might be afforded by an enhanced T_S, unlike a lowered ΔC_p which makes for a shallow stability curve (or lower ΔG at high and low temperatures) a high T_S (method III) or high ΔG (method I) both provide for substantially higher ΔG at high temperatures, structures with high ΔG are more stabilized and thus more rigid.

To increase the ΔG of a protein on the other hand, more options are available such that any number of interactions like salt bridges, H bonds or hydrophobic interactions may be added by single amino acid changes. Similarly, to attain a lower ΔC_p the sequence can be altered in many ways to provide for tighter core packing or promote structured clusters that persist in the denatured state. It has been shown that the ΔC_p is strongly correlated with

amount of solvent accessible surface area exposed when the protein unfolds (ΔASA) (*22-24*). This means that if changes in sequence promote structured regions in the denatured state, the ΔASA for this protein will be lower and will result in a lower ΔC_p. For example, in the case of the Ribonuclease H protein from *Thermus thermophilus*, it was suggested that a structured cluster in the denatured state caused the reduced ΔC_p for this protein (*85, 86*).

In conclusion, the data compiled here suggest that there is no single thermodynamic method which thermophilic proteins use to enhance their thermostability and it appears that proteins use combination of different methods to increase their T_m. Both increasing ΔG and lowering ΔC_p appear to be the most common methods to enhance thermostability.

CHAPTER II

MATERIALS AND METHODS

Reagents for routine protein purification were of reagent grade and were procured from various vendors, which included Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA) and EM Science (Gibbstown, NJ). Reagents and buffers used for spectroscopic experiments were of analytical grade or better and were from the same vendors. For solvent denaturation experiments, urea was obtained from Nacalai Tesque (Kyoto, Japan) or ICN Biomedicals (Aurora, OH) and guanidinium hydrochloride (GuHCl) was from ICN Biomedical (Aurora, OH).

The oligonucleotides used for mutagenesis and cloning were from IDT Integrated DNA technologies (Corallville, IA). Site directed mutagenesis was performed using the QuikChange™ kit from Stratagene (La Jolla, CA). Enzymes used in recombinant DNA methodologies like restriction digests and PCR amplification of DNA were from New England Biolabs (Ipswich, MA), Promega (Madison, WI) or Stratagene (La Jolla, CA). Plasmids for over-expression of HPr homologs were derived either from pUC (HPr) (*87*) or pLOI 1803 (*88*). Cell cultures for cloning of different HPr homologues were obtained either from ATCC® (American type culture collection, Manassas, VA) or DSMZ (German collection of Microorganisms and Cell Cultures, Braunschweig, Germany). Plasmid and other DNA purification kits were procured from Qiagen (Valencia, CA). Plasmid DNA was sequenced by the Texas A&M Gene Technologies Lab.

For protein over-expression the ES7R strain (*87*), (*89*), of *Escherichia coli* was used. For mutagenesis and other routine molecular biology procedures, like cloning, plasmid

purification etc., XL1-blue (Stratagene) strain of *Escherichia coli* cells were used. Culture media for bacterial cultures were obtained from DIFCO (Becton Dickinson, Sparks, MD) and resins for column chromatography, like DEAE Sephacel for ion exchange and Sephadex G50 for gel filtration chromatography were procured from Amersham Biosciences (Picatsaway, NJ). The purity of protein preparation was confirmed by MALDI-TOF mass spectrometry by the Texas A&M University Laboratory for Biological Mass spectrometry.

The acquisition of most data was facilitated by software provided by the instrument manufacturers. Data analysis was done using the Kaleidagraph® software (Synergy software, Reading, PA), ProFit® software (QuantSoft software, Uetikon am See, Switzerland) and Origin® software (OriginLab, Northhampton, NJ). For protein visualization, structure manipulation, plot making and other structural calculations the SPDB viewer (*90*) was primarily used. Other software packages used for these purposes were Molscript (*91*), ICM Molsoft™, Chimera (*92*), and Insight II® (Accelrys, San Diego, CA). For accessible surface area calculations p*fis* (*93*), WHATIF (*94*) or Chimera (*92*) were used.

Cloning of HPr homologues

Plasmids for the over expression of two HPr homologues from *Bacillus subtilus* (*Bs*) (*87*) and *Bacillus staerothermophilus* (*Bst*) (*88*) were available. For other HPr homologues from *Streptococcus thermophilus* (*St*), *Bacillus halodurans* (*Bh*) and *Oceanobacillus iheyensis* (*Oi*), the HPr gene was cloned into appropriate vectors to create plasmids for protein over expression.

Cloning of *Streptococcus thermophilus* (*St*) HPr

The *St*HPr gene was cloned into the pLOI 1803 vector using the four-primer cloning method using a modification of the method published by Howorka and Bayley (*95*). A diagrammatic representation of the procedure is shown in Figure 3.

To obtain the gene coding for *St*HPr, we purified genomic DNA from *S thermophilus* cells, which were obtained from the ATCC® (BAA-491™). The freeze-dried culture was rehydrated with trypticase soy broth (BD 236950), supplemented with 5% (v/v) defibrinated sheep blood. A 30 ml overnight culture (grown for 12 to 16 hours) was used to prepare genomic DNA using standard methods (*96*). The genomic DNA was used as a template in a PCR reaction to amplify the HPr gene. The primers used for *St*HPr gene amplification had overhangs that are complementary to the vector sequence; similarly the primers used to amplify the pLOI 1803 vector DNA had overhangs complementary to the *St*HPr gene. The primers used are shown in Table 2.

Table 2: Primers used for cloning of *St*HPr gene into vector pLOI1803.

*St*HPr insert amplification
*St*HPr insert + (sense primer)
5′-G GGA ATG ATG AAC ATG GCT TCT AAA GAT TTC-3′
Vector Overhang *St*HPr gene
*St*HPr insert – (antisense primer)
5′-G CGT TTC CCA TTA TGC CAA TCC TTC-3′
Vector overhang *St*HPr gene
Vector amplification
Vector +
5′-GGA TTG GCA TAA TGG GAA ACG CAA TCC-3′
*St*HPr gene overhang Vector
Vector –
5′-C TTT AGA AGC CAT GTT CAT CAT TCC CC-3′
*St*HPr gene overhang Vector

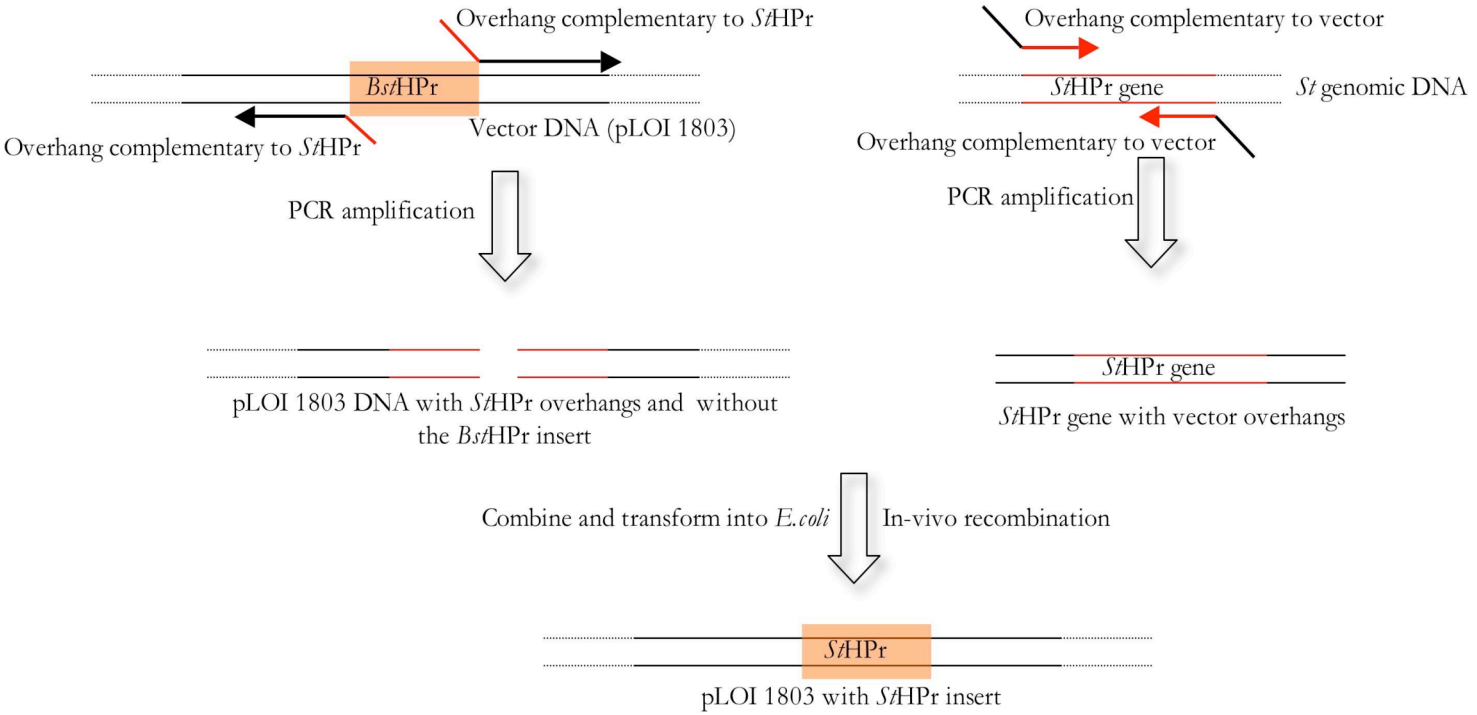


Figure 3: A diagrammatic representation of the 'four-primer cloning method'. The figure illustrates the cloning of the *St*HPr gene into the pLOI 1803 vector. The pLOI 1803 vector contains the *Bst*HPr insert, which was replaced by the *St*HPr gene.

For amplification of the *St*HPr insert, Taq polymerase was used with 30 cycles of amplification, a one minute extension time, 30-second annealing time, 30 second melting time at 95ºC, with the amplification cycles preceded by a 2 minute hold at 95ºC. A typical PCR amplification schedule is shown in Figure 4.

95ºC	95ºC	T_m-5ºC	68ºC	**68ºC**	**4ºC**
2 min	30 sec	30 sec	1 min	**1 min**	**∞**
Melt	Melt	Annealing	Extension	**Extension**	**Hold**
		30 cycles			

Figure 4: A representative PCR amplification schedule, shown here for amplification of *St*HPr gene. The temperatures used and cycle times are typical for amplifications using the Taq polymerase and for short products.

The annealing temperature was usually set 5ºC lower than the melting temperature of the primer dimers. The extension temperature was that suggested by the enzyme manufacturer; similarly the PCR reaction mixture (primer, template and dNTP concentration) was prepared according to suggestions by the enzyme manufacturer. For amplification of the vector DNA the Herculase® polymerase mix (Stratagene) was used, with 30 cycles of amplification. The reaction mixture and thermocycler setup was according to recommendations of the enzyme manufacturer, specifically those for targets ≤ 10 kb in length (the length of pLOI 1803 vector is ~5 kb). The success of the PCR amplifications

was confirmed by analyzing 2 μl of the reaction product on a 0.7% agarose gel for the vector amplification and a 2% agarose gel for the insert amplification reaction. The appearance of high intensity DNA bands of appropriate size on agarose gels with ethidium bromide staining confirmed the success of amplification reaction.

The PCR products from both reactions were then mixed and transformed into chemically competent XL1-blue *Escherichia coli* cells. Briefly, 5 μl each of the vector and insert amplification reactions were mixed with 190 μl of competent cells and incubated on ice for 30 minutes, following which they were transferred to a heat block at 42ºC and held there for one minute. The entire transformation mix was then plated onto a LB agar plate supplemented with 50 μg/ml ampicillin. The plate was incubated for at least 16 hours after which the transformants were tested for presence of desired plasmid by colony PCR. For the colony PCR screening of transformants, 12 μl reactions were set up and the primers were those used for amplification of *St*HPr gene. The enzyme used for amplification was Taq polymerase and the reaction conditions were similar to those used earlier for the amplification of *St*HPr gene. The template DNA for these reactions was provided by the bacterial colony that was picked from the selective agar plate. A control reaction was also performed where the *Bst*HPr gene was amplified from pLOI 1803 plasmid under conditions identical to those used for the colony PCR screening. The entire PCR reactions were analyzed on 2% agarose gels and positive clones were chosen based on the presence of DNA bands of appropriate size. The colonies with positive clones were cultured overnight in Luria broth (LB) and plasmid DNA was purified using the QIAprep Spin Miniprep Kit (Stratagene) and sequenced to confirm the presence of and orientation of the *St*HPr insert.

Cloning of *Oceanobacillus iheyensis* (*Oi*) HPr

The *Oi*HPr gene was cloned into the pUC (HPr) vector using the same four-primer cloning method used for cloning the *St*HPr gene. The gene coding for *Oi*HPr was obtained from *O iheyensis* genomic DNA, which was purified from cultures obtained from the DSMZ (Braunschweig, Germany). This culture was rehydrated in marine broth (BD 2216) and a 30 ml overnight culture of these cells was used for a genomic DNA preparation, using the standard methods used earlier (*96*), with the only difference in the protocol was being the cell lysis step, where the cells had to be lysed by passing them through a french press cell at 1200 psi. This was necessary because the standard alkaline lysis procedure was not successful in lysing these bacteria. The primers used for cloning of the *Oi*HPr gene are shown in Table 3:

Table 3: Primers used for cloning of *Oi*HPr gene into vector pUC (HPr).

*Oi*HPr gene amplification
*Oi*HPr + (sense primer)
5′-GT TGG GGA AAT ACA ATG AAA TCA CAA ACA TTT AC-3′
Vector overhang *Oi*HPr
*Oi*HPr – (antisense primer)
5′-GG CTA CCC GGG CTA TTC ACC AAG ATG C-3′
Vector overhang *Oi*HPr
Vector amplification
Vector + (sense primer)
5′-GGT GAA TAG CCC GGG TAG CCA AAG-3′
*Oi*HPr overhang Vector
Vector – (antisense primer)
5′-GT TTG TGA TTT CAT TGT ATT TCC CCA ACT TAT AGG-3′
*Oi*HPr overhang Vector

The PCR amplification reaction, subsequent transformation and screening protocols used were identical to those used for cloning of the *St*HPr gene.

Cloning of *Bacillus halodurans* (*Bh*) HPr

The gene for *Bh*HPr was cloned into the pLOI 1803 vector using restriction enzyme digests and subsequent ligation into the vector DNA. To obtain the gene coding for *Bh*HPr, it was PCR amplified from genomic DNA that was prepared from a culture of *B halodurans*. The culture was obtained from ATCC (BAA-125™) in a vacuum dried form and was rehydrated in trypticase soy broth (BD 236950). A 30 ml overnight culture was used to prepare genomic DNA according to standard methods (*96*). The genomic DNA was used as template in PCR reaction to amplify the HPr gene. The primers used for *Bh*HPr gene amplification had overhangs that were the recognition sequence for restriction enzyme Bgl II. The primers used are shown in Table 4.

Table 4: Primers used for cloning of *Bh*HPr gene into vector pLOI1803.

*Bh*HPr insert amplification
*Bh*HPr insert + (sense primer)
5′- GAG TAGA TCT ATG GTT GAA AAA CAA G-3′
Bgl II *Bh*HPr
*Bh*HPr – (antisense primer)
5′- T GAG AGA TCT TTA TTC CTT TTC GAT GAA TG-3′
Bgl II *Bh*HPr
Vector amplification
Vector + (sense primer)
5′- G AGT AGA TCT TGG GAA ACG CAA TC-3′
Bgl II Vector
Vector – (antisense primer)
5′- T GAG AGA TCT GTT CAT CAT TCC CC-3′
Bgl II Vector

The enzymes used and reaction conditions for both *Bh*HPr insert and vector amplification were similar to those described above in the case of *St* and *Ot*HPr. The success of the PCR reactions was confirmed with appropriate agarose gels. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), and eluted in sterile deionized water. The eluate was concentrated to 20 µl in a speedvac assembly. The restriction digest with BglII was performed in a 20µl reaction with 5µl of purified PCR product using conditions recommended by the enzyme manufacturer (NEB). The digested DNA for both insert and vector reactions was treated with Shrimp Alkaline Phosphatase (SAP) for 15 minutes at 37ºC followed by inactivation at 72ºC for 5 minutes. These reaction products were purified with a PCR purification kit and the products eluted in deionized water followed by concentration to 10 µl. The entire purified reaction products were used to setup a ligation reaction with T4 DNA ligase (Promega). The ligation reaction was incubated at room temperature for 2 hours and then the entire reaction contents were used to transform 80µl of chemically competent XL1 blue *Escherichia coli* cells according to standard methods. The transformation mixture was plated on the LB plates containing ampicillin and the plates incubated overnight. The transformants obtained were screened for the presence of plasmid with *Bh*HPr insert according to methods described earlier. The positive clones were processed further to obtain plasmid DNA, which was sequenced to confirm the presence of the correct plasmid with the *Bh*HPr insert in the proper orientation. The orientation of *Bh*HPr insert was of special concern in this cloning experiment because of an inherent handicap in the protocol; namely the use of a single restriction enzyme to digest both ends in the construct. This trade-off in efficiency was however useful to simplify the experimental protocol.

Mutagenesis

Site-directed variants for all proteins were produced using the QuickChange™ kit from Stratagene. The kit uses the overlap extension by PCR method to make entire copies of the template plasmid now incorporating the altered nucleic acid bases. For single-site mutants of a protein, a pair of complementary primers incorporating the desired mutated bases was employed with the length of these mutagenic primers adjusted such that the melting temperature for the primer dimers is close to 60ºC. The primers were designed using the software tools provided by oligonucleotide manufacturer, Integrated DNA technologies (http://scitools.idtdna.com/scitools/Applications/OligoAnalyzer/). The concentrations of primers, dNTPs and template DNA in the PCR reactions were as recommended by Quikchange™ kit protocol (Stratagene). Following the amplification reaction, the presence of product DNA of desired length was confirmed on a 0.7% (w/v) agarose gel. As a control, for both size of the product and yield, 3 μl of diluted template (volume of template used in the PCR was adjusted to 50 μl with de-ionized water) was also analyzed on the agarose gel. The size of the product should be similar to the template plasmid DNA except the PCR product is mostly linear DNA, and hence, runs slower on the agarose gel compared with the super-coiled template DNA; the band for the PCR product should be more intense than that for diluted template which indicates a successful amplification reaction.

The PCR product was digested with DPN1, a restriction endonuclease that specifically cleaves methylated DNA. Since the PCR product does not contain methylated DNA, this step removes most of the template DNA used in the amplification. This is useful

because it improves the chances of obtaining mutant plasmid instead of wild-type or template plasmid. The product of the digestion reaction is then transformed into chemically competent XL1 blue *Escherichia coli* cells. Following the heat shock treatment and subsequent incubation, 20 μl of the transformation mixture was plated onto selective LB agar plates containing ampicillin (50 μg/ml). The plates were incubated overnight, after which transformant colonies were picked and cultured in 5 ml of LB supplemented with ampicillin (50 μl/ml) for plasmid purification. The plasmid DNA was sequenced to confirm the presence of desired mutation-containing DNA.

Protein overexpression and purification

The ES7R strain of *Escherichia coli* was used for over-expression of the HPr proteins. The genotype of ES7R cells is designed for HPr over-expression because these cells do not produce any endogenous HPr since they contain a defect in the PTS operon (*89*). They also contain the TETr gene, which confers tetracycline resistance on these cells. The HPr homologues were over-expressed from either the pUC (HPr) or pLOI 1803 derived vectors. The PTS operon encoded in both these over-expression vectors is under the control of its native promoter and the protein is constitutively expressed and does not require induction. All HPr homologues were over-expressed and purified by slight modifications to the protocol for *Bst*HPr, published earlier (*72*). The yield of pure protein in each case was approximately 60-70 mg/L of culture. The yield for *Oi*HPr was lower partly due to accumulation of some protein in the inclusion bodies. However the yield was sufficient (30 mg/L of culture) for our purposes.

Purification of *Bst*HPr

Electrocompetent ES7R cells were transformed with the pLOI 1803 plasmid by electroporation in BioRad cuvettes. A single colony was picked from a selective agar plate and used to start a 5 ml LB culture containing 50 μg/ml of ampicillin and this culture was later used to inoculate a 6-liter batch of LB containing the required antibiotics (50 μg/ml of ampicillin and 5 μg/ml of tetracycline). After growing the cultures for 14-16 hours with vigorous agitation at 37ºC, the cells were harvested by centrifugation. Cell pellets were resuspended in minimal volume of 2x TE (20 mM Tris, 2 mM EDTA, pH 8.4). The cells were lysed by passing them through a french press twice at 1200 psi and the supernatant was collected after centrifugation in JA-20 (Beckmann) rotor for 30 minutes at 15000 rpm, 4°C. The volume of supernatant was adjusted to 150 ml with 2x TE, to the suspension 22.4% (w/v) ammonium sulfate was added slowly over an hour on ice. The suspension was then centrifuged at 10,000 rpm in a KA-10 rotor (Kompspin) for 30 minutes, 4ºC, and the supernatant obtained was dialyzed extensively against 2x TE buffer. The pH of resulting solution was adjusted to 8.4 and it was applied to DEAE resin (100 ml bed volume for a 6 L preparation) pre-equilibrated in the same buffer and the flow-through solution was collected. The ion exchange resin together with lysate was transferred to a 250 ml separatory funnel and all the remaining steps in the ion exchange procedure were performed in the funnel. The resin was washed with one additional bed volume of buffer and the two solutions were pooled and lyophilized. The lyophilized sample was resuspended in minimal amount of 20 mM ammonium bicarbonate and applied to a G-50 Sepharose column (~800 ml bed volume in a Kontes Chromaflex® column measuring 4 ft in length and 3" in outer diameter). Fractions were analyzed by SDS PAGE and those containing the HPr protein

were pooled and lyophilized. The purity and identity of the protein was confirmed by SDS PAGE and MALDI TOF mass spectrometry.

Purification of *Bs*HPr

The *Bs*HPr protein was over-expressed using the pUC (HPr) in ES7R *Escherichia coli* (*87, 97*). The cells were cultured and processed like in the *Bst*HPr purification with only slight modifications. Following the ammonium sulfate precipitation and dialysis step, the pH of the lysate was adjusted to 7.2 instead of 8.4 as in the case of *Bst*HPr purification. The ion exchange step was also altered so that the pH of the DEAE resin was adjusted to 7.2, the lysate was applied to the resin and the resin washed with 2X TE (pH 7.2) and the flow through was collected in both cases. The resin was then washed with two bed volumes of 0.2 M NaCl in 2X TE; the eluate was pooled with the flow through fraction collected earlier. The remaining steps of the protocol were identical to those used for *Bst*HPr.

Purification of *St*HPr

The *St*HPr protein was over-expressed from a pLOI 1803 derived vector in ES7R *Escherichia coli*. The growth, harvest and cell lysis steps were as described earlier for *Bst*HPr purification. However, for the purification of *St*HPR the pH of 2x TE buffer was adjusted to 7.2 and the DEAE resin was washed with two bed volumes of 0.1M NaCl in 2x TE following the application of lysate (at pH 7.2) and subsequent wash with 1 bed volume of 2x TE. The remaining steps of the purification protocol were identical to those for *Bst*HPr purification.

Purification of *Bh*HPr

The *Bh*HPr protein was over-expressed from a pLOI 1803 derived vector in ES7R *Escherichia coli*. The growth, harvest and cell lysis steps were as described earlier for *Bst*HPr purification. For the purification of *Bh*HPr the pH of the 2x TE buffer was adjusted to 8.4 like that for *Bst*HPr purification. However, following application of lysate (adjusted to pH 8.4), the resin was washed with 2 bed volumes of 0.15 M NaCl after the resin was washed with 1 bed volume of 2x TE. The remaining steps of the purification protocol were identical to those for *Bst*HPr purification.

Purification of *Oi*HPr

The *Oi*HPr protein was over-expressed from a pUC (HPr) derived vector in ES7R *Escherichia coli*. For purification of *Oi*HPr the protocol used was identical to that used for *Bs*HPr protein. Since the yields for this homologue are lower, we suggest growing the 6L culture at 30ºC instead of the 37ºC used for other proteins and use of column chromatography for the ion exchange step with a 0 to 1M NaCl gradient instead of the 0.2M NaCl wash.

Thermodynamic stability measurements

The model protein used for this study HPr, has a significant α helix content as seen in the X-ray crystal structure for *Bs* and *Bst* homologues (see Figure 5). This high α helix content provides for an excellent probe to follow the global conformation or native state of the protein by circular dichroism (CD) spectroscopy. Consequently, the CD spectra shown for the *St*HPr protein in Figure 6 show strong characteristic negative bands at 208 and 222

nm, typical for α helices. Figure 6 also shows CD spectra for denatured/unfolded protein, unfolded in this case by raising the temperature to 95ºC; it can be seen that the spectrum for denatured protein loses significant signal strength especially at 222 nm. These observations provide the basis for use of CD signal at 222 nm as a probe for protein conformation. We have used this probe in the majority of our experiments in this study, although other spectroscopic probes, like fluorescence emission intensity at 350nm have also been used in some experiments.

Protein stability measurements by equilibrium denaturation experiments

To measure the stability of HPr homologues and their variants, we have performed equilibrium denaturation experiments using either chaotropes like urea or guanidine hydrochloride (GuHCl) or temperature as the denaturants. The circular dichroism signal at 222 nm was used to monitor the unfolding transitions of the proteins on either an Aviv 62DS or Aviv 202SF (Aviv Biomedical, Lakewood, NJ) spectropolarimeter equipped with temperature control and stirring units.

The thermal unfolding transitions were followed by heating the protein samples from 25 to 99°C in 1ºC steps, with 3 or 5-minute equilibration intervals at each temperature; the data were averaged for one minute at each temperature step. Thermal transitions were observed for protein samples at concentrations of 4.5 to 6.5 μM in 10 mM NaPi buffer at pH 7; 10 mm path length quartz cells were used. The reversibility of the thermal transitions was confirmed by measuring the CD signal at 25°C after the completion of the unfolding transition and the magnitude of CD signal compared with that at the beginning of the experiment was used to calculate percentage of reversibility. Data from a

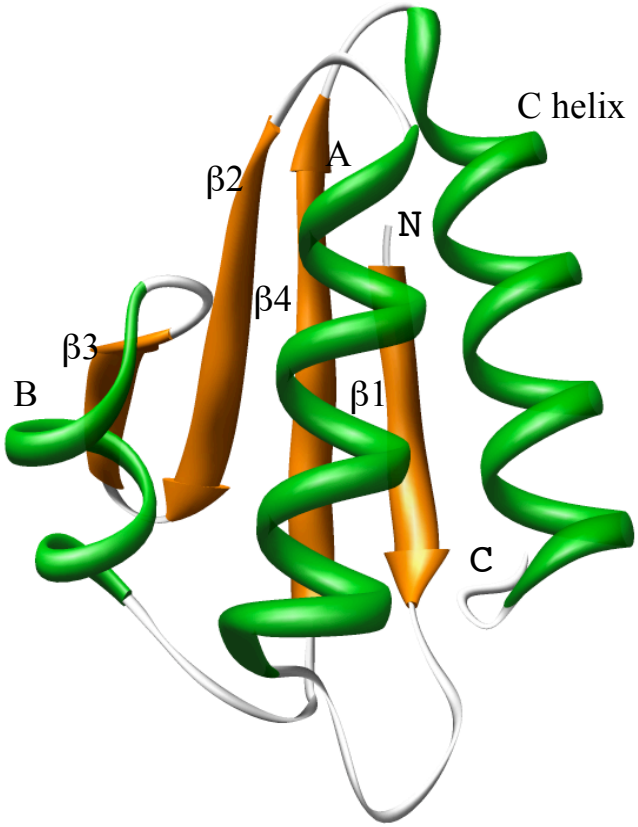


Figure 5: Ribbon diagram of *Bst*HPr protein. The PDB accession code for the structure used here is 1Y4Y; indicated on the plot are the secondary structure elements (helices A through C and β strands 1 through 4) and N and C termini for the protein. The image was created using the UCSF Chimera package (*92*).

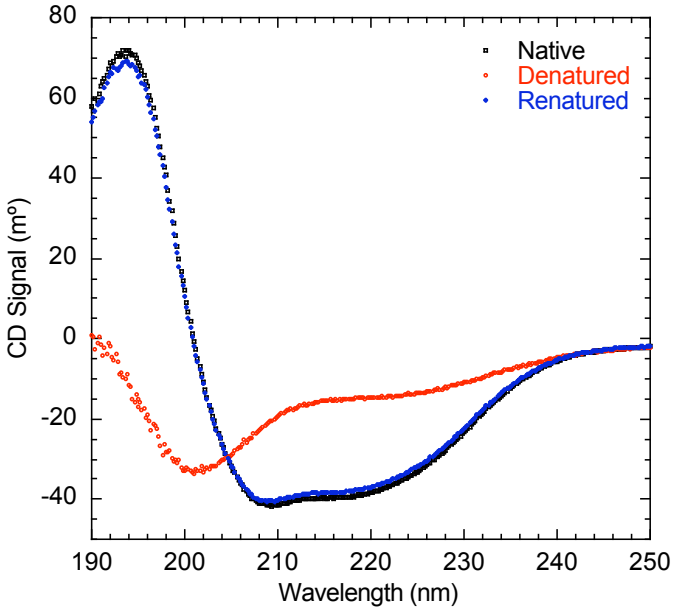


Figure 6: Circular dichroism spectra for native, denatured and renatured *St*HPr. All spectra were recorded in 10mM Sodium Phosphate buffer at pH 7. The 'native' spectrum was recorded at 25ºC; the 'denatured' spectrum was recorded by heating the sample to 95ºC. Finally, the 'renatured' spectrum was recorded after the sample was cooled back down to 25ºC, the sample was allowed to equilibrate for 3 minutes following a change in temperature. The raw CD signal expressed in millidegrees (mº) is shown for each spectrum.

typical thermal denaturation experiment are shown in Figure 7; the entire thermal denaturation profile can be described by equation 2, which is a variation of the van't Hoff equation:

$$\theta_{obs} = \frac{\left(\theta_N + m_N * T\right) + \left(\theta_D + m_D * T\right) * \exp\left(\frac{-\Delta H_m}{R} * \left(\frac{1}{T + 273.15} - \frac{1}{T_m + 273.15}\right)\right)}{1 + \exp\left(\frac{-\Delta H_m}{R} * \left(\frac{1}{T + 273.15} - \frac{1}{T_m + 273.15}\right)\right)} \quad (2)$$

where, θ_{obs} is the observed elipticity at 222 nm, θ_N is the native state intercept, θ_D is the denatured state intercept, m_N and m_D are the slopes of the native and denatured state baselines, T is the temperature, T_m is the temperature at the mid point of the unfolding transition, ΔH_m is the van't Hoff enthalpy at T_m and R is the gas constant in kcal mol^{-1} K^{-1}. A least squares fit of the above equation, therefore, resolves best-fit estimates for T_m and ΔH_m values for the protein.

The solvent denaturation experiments, with urea or GuHCl as the denaturant, were performed using similar protein concentrations as the thermal unfolding experiments. Solvent denaturation experiments were performed using a computer controlled Hamilton Microlab® 500 two-syringe titrator assembly since equilibration times were fairly small. Typically, the experiment started with folded protein dissolved in an appropriate buffer solution in the cuvette and denatured protein in concentrated chaotrope solution in a volumetric flask and both the folded and denatured protein are at the same concentration. These two solutions are connected through the titrator assembly, which adds small increments of concentrated denaturant solution to the cuvette with concomitant removal of the same volume from the cuvette, keeping the protein concentration in the cuvette constant. This process was repeated several times with increasing volumes of the denaturant

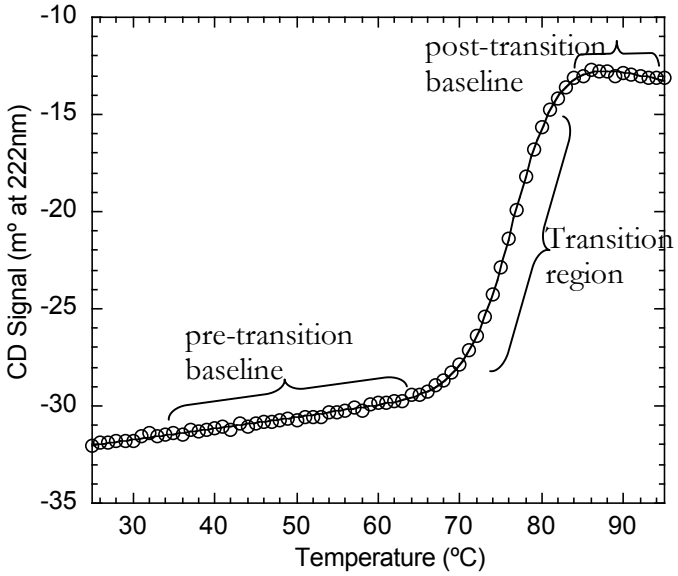


Figure 7: Representative thermal denaturation transition for *St*HPr. The three regions of typical curve are indicated. The line through the data is the best fit curve to equation 2. The buffer used was 10 mM sodium phosphate at pH 7 and the protein concentration was ~4.5 μM. The raw CD signal expressed in millidegrees (mº) is shown.

solution being added in each step until the set target concentration of denaturant was achieved. Each step of this 'titration' is separated by an equilibration period of five minutes following which the CD signal at 222 nm is averaged over a minute and recorded. The number of steps in the titration was adjusted to achieve a good dispersion of data over the entire denaturant concentration range. Data from a typical solvent denaturation experiment are shown in Figure 8 and the curve is a least squares fit of equation 3. This equation allows us to calculate a free energy of stabilization (ΔG) and m-value or the slope of the unfolding transition through:

$$\theta_{obs} = \frac{\left(\theta_N + a_N * [D]\right) + \left(\theta_D + a_D * [D]\right) * \exp\left(m * \frac{[D] - C_{mid}}{RT}\right)}{1 + \exp\left(m * \frac{[D] - C_{mid}}{RT}\right)} \quad (3)$$

where, θ_{obs} is the observed elipticity at 222 nm, θ_N is the native state intercept, θ_D is the denatured state intercept, a_N and a_D are the slopes of the native and denatured state baselines, [D] is the concentration of the denaturant, m is the m-value, C_{mid} is the [D] at the mid-point of the unfolding transition, R is the gas constant and T is the temperature.

Applying the linear extrapolation method (LEM) (*98, 99*) and assuming a two-state transition between the native and denatured states, the free energy of stabilization in water can be calculated according to equation (4).

$$\Delta G_{H_2O} = m * C_{mid} \quad (4)$$

The experiments studying ionic strength dependence of protein stability were performed using urea as the denaturant and both the buffer and the urea solutions contained identical concentrations of NaCl. Similarly, the experiments studying pH dependence of protein stability were done using urea as the denaturant. In this case the urea

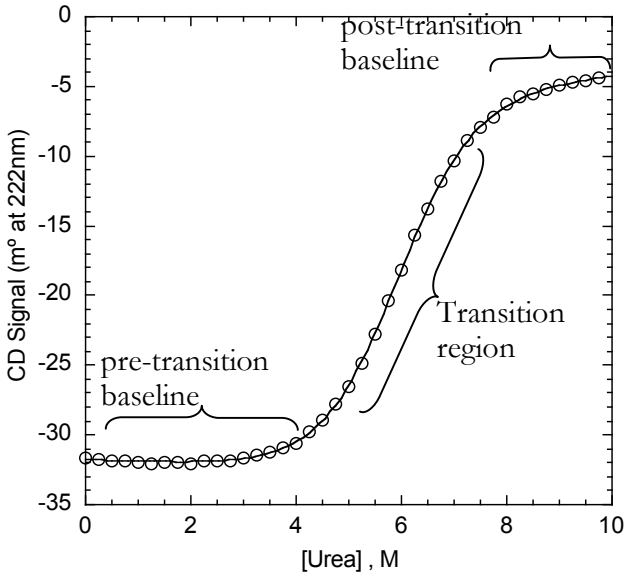


Figure 8: Representative urea denaturation transition for *St*HPr. The three regions of typical curve are indicated. Line through the data is the best fit curve to equation 3. The buffer used was 10 mM sodium phosphate at pH 7, at 40ºC, and the protein concentration used was ~4.5 μM. The raw CD signal expressed in millidegrees (mº) is shown.

stock and protein solution in the cuvette were prepared in the same buffer mix of sodium citrate, sodium phosphate and sodium borate (final concentration of 10 mM) adjusted to the same pH.

Stability curves

To obtain 'stability curves' (*51*) for the HPr homologues or their variants, the ΔG from urea denaturation experiments at temperatures between 5 and 50ºC were combined with data from transition region of a thermal denaturation experiment. The data were fit to a form of the Gibbs-Helmholtz equation (equation 1) according to methods of Pace and Laurents (*100*).

Differential scanning calorimetry (DSC)

DSC experiments were performed on *Bs* and *Bst*HPr proteins to estimate ΔC_p and native and denatured state C_p values. All measurements were performed using the Microcal VP-DSC calorimeter (Microcal, Northampton, MA). Protein concentration used was in the range of 20 to 100 μM. The buffers used for DSC experiments were carboxylic acid based buffers with minimal heats of ionization (*101*). All protein samples were dialyzed against two liters of buffer with one buffer change and the buffer at the end of the dialysis was used as the reference solution. The temperatures over which the samples were scanned depended on the pH of the sample and ranged from 25 to 120ºC. A Scan rate of 60ºC/hour was used for all experiments and no scan rate dependence was detected for scan rates between 30 and 90ºC/hour. At least two consecutive scans were performed for every sample to test for reversibility. Buffer scans were subtracted from the thermograms and data were normalized

for protein concentration, following which data were analyzed using Origin software and the two-state folding model. For the measurement of native and denatured state C_p the methods of Kholodenko and Freire (*102*) were employed.

Studies on protein oligomeric state

Pulsed-field gradient NMR

Pulsed-field gradient (PFG) NMR experiments were performed for the wild type *Bs* and *Bst*HPr protein together with the F29W variant of *Bst*HPr, to study the oligomeric nature of these proteins in solution. Protein samples for the wild-type and F29W *Bst*HPr proteins were prepared at concentrations of 1 mM and 2 mM, in 10 mM sodium phosphate buffer in ^2H_2O; pH* was adjusted to 7.0 in all cases. All samples also included 20 μl of 1% dioxane in ^2H_2O. NMR data were acquired on Varian Inova 500 spectrometer (Biomolecular NMR Laboratory, Texas A&M University) equipped with triple-resonance pulsed-field gradient probe heads with linear amplifiers. We used the PG-SLED sequence (*103*) incorporating composite sine gradient pulses. All the data were acquired at 25°C. The lengths of the diffusion gradient and the stimulated echo were optimized for each sample to give a total decay of 75-80% from the initial signal. Each measurement was repeated 3-5 times to improve quality of the data and to allow for an estimation of the errors. NMR spectra were acquired with 9K complex points, with a spectral width of 6092 Hz. Data processing was performed using the VNMR software from Varian. Peak heights were measured and compiled as a function of gradient strength using macros written in the lab. The decay in peak heights for the protein signals was fit by a single exponential in most

cases, however peaks for dioxane required fitting by a double exponential in some cases, where the rate constants obtained on the slower phase matched the rate constants obtained for protein peaks measured independently.

Size exclusion chromatography

Analytical size exclusion chromatography experiments were performed on *Bst*HPr and its F29W to study the oligomeric nature of these proteins. A sample consisting of a mixture of wild-type *Bst*HPr and the F29W variant of the protein was analyzed with size exclusion chromatography. An equimolar mixture at a final protein concentration of 330 μM was applied to a Superdex-200 16/60 column on an AKTAexplorer HPLC system from Amersham Biosciences. The mobile phase (TBS: 50 mM Tris, pH 7.5, 150 mM NaCl) flow rate was set at 0.5 ml/min and the elution peaks were detected by UV absorbance at 280 nm. The column was calibrated using the following proteins as molecular weight standards: bovine serum albumin (66 KDa), ovalbumin (45 KDa), carbonic anhydrase (29 KDa), myoglobin (17.5 KDa), aprotinin (6.5 KDa). Blue dextran was used to estimate the void volume (V_o) and acetone was used to estimate the included volume (V_i). Partition coefficients were calculated for each of the standards and the samples studied in accordance with the equation

$$K_d = \frac{(V_e - V_o)}{V_i} \qquad (5)$$

where, V_e is the elution volume of the protein sample, V_o is the void volume represented by elution volume of blue dextran and V_i is the volume of solvent inside the gel matrix

available to low molecular weight compounds, evaluated as the difference between the elution volume of acetone and the elution volume of blue dextran.

Analytical ultracentrifugation

The solution oligomeric states of *Bst*HPr and the F29W variant were probed by sedimentation equilibrium analysis using analytical ultracentrifugation. The proteins were studied at concentrations of 250, 500 and 750 μM for the F29W variant and 500 μM and 1 mM for the wild type protein. The buffer systems used for the F29W variant were 10 mM NaPi pH 7 and 50 mM Na Citrate pH 6.5; for the wild type protein 10 mM NaPi pH 7 was used as the buffer. For the F29W variant, additional experiments with 0.5 and 1 M NaCl were performed in the Na Citrate buffer system. Experiments were performed at 25°C and 10°C. Protein samples were analyzed by SDS PAGE after the centrifugation experiments to determine if protein denaturation or fragmentation had occurred.

Sedimentation equilibrium experiments were performed using a Beckmann XLA centrifuge, a Ti60 rotor and two sectored epoxy centerpieces with quartz windows. The wavelength for absorbance measurements was based on the protein concentrations in question, making sure that all measurements were in the linear range of the detector (0.4 – 1.4). Data were collected in radial step size of 0.001 cm with 10 replicates to improve data quality. In all the experiments, scans were performed until equilibrium was reached; this was ascertained by overlaying subsequent scans. At least three overlaying scans were obtained in all cases, and rotor speeds varied between 10,000 and 25,000 RPM. The data were analyzed using Origin software with macros supplied by Beckman. Single and multiple species models were used; the goodness of fit in each case was evaluated by randomness in

residuals and their magnitude. The partial specific volume was calculated using the SEDENTERP software program (version 1.08).

CHAPTER III

COMPARATIVE STUDY OF HPr HOMOLOGUES

Introduction

The folded conformations of proteins are stabilized by various forces (including hydrophobic, electrostatic and hydrogen bonding interactions) acting in concert to offset alternative forms, which are favored chiefly by conformational entropy. Though our understanding of these forces has improved significantly since the early days of protein folding (*104*), studies of proteins derived from extremophiles provide novel insights into the role that these forces play in stabilizing their native conformation. Organisms inhabiting extremes of environment are often called extremophiles. This is a broad term that describes organisms that have adapted to extremes of temperature, salinity, pH, pressure etc. Numerous studies have focused on the ability of proteins from these organisms to survive these extreme environmental conditions (*6-11*). As a result of this continued interest, many hypotheses have been proposed to explain the enhanced stability of these proteins. These include stabilization by increased number of hydrogen bonds, improved core packing, stiffer surface loops, etc. Although these and other modes of stabilization are independently and/or collectively applicable in many cases, a unifying 'recipe' for conferring thermostability on a protein remains elusive. Of all the adaptations to extreme environments, thermophilic adaptations are of particular interest to protein chemists because survival in extremes of temperature requires adaptation (*3*), whereas in other kinds of stress (like pH, osmotic), the organisms can survive by 'avoiding' these stress factors by

compensatory mechanisms like pH regulation (*105*) or synthesis of low molecular weight 'osmoprotectants' (*106, 107*).

Proteins from extremophiles are also of potential interest to biotechnology as they provide a myriad of advantages over their mesophilic homologues; these proteins are more tolerant to harsher conditions of temperature, solvent composition (presence of additive or organic solvents), pH and salinity (*33, 108*). These harsh conditions are advantageous for industrial applications because they provide for higher reaction rates, higher substrate concentrations and lower viscosity while also reducing chances of microbial contamination. To confer these advantages on proteins of mesophilic origin, directed evolution (*45, 46*) and other protein engineering (*109*) experiments have been used.

On a more fundamental level, proteins from extremophilic organisms provide good model systems to study the forces responsible for protein stability. Indeed, many studies have focused on teasing out determinants of enhanced stability of these proteins by comparing them with homologues derived from species inhabiting more moderate habitats. The most fruitful of these studies have relied on detailed thermodynamic comparisons of the homologous pair of proteins together with comparison of their three-dimensional structures (*110*). These thermodynamic comparisons require measurements of free energies of stabilization (ΔG) under a wide range of solution conditions involving variations of temperature, solution pH and ionic strength.

In this report, we present such a detailed analysis of five HPr homologues all derived from gram-positive bacteria, inhabiting very different habitats. These HPr homologues are derived from *Bacillus subtilus (Bs), Streptococcus thermophilus (St), Bacillus staerothermophilus (Bst), Bacillus halodurans (Bh*) and *Oceanobacillus iheyensis (Oi)*. *Bacillus subtilus* is

a mesophile, *Streptococcus thermophilus* and *Bacillus staerothermophilus* are moderate thermophiles, *Bacillus halodurans* is a haloalkaliphilic thermophile (*111, 112*) and *Oceanobacillus iheyensis* is an extremely halotolerant alkaliphile (*113*). As mentioned above, these homologous proteins are adapted to varied environments and we hope to learn about the adaptations involved by detailed thermodynamic analyses of each protein and comparisons between them. HPr or histidine containing protein is a small (~88 amino acid) monomeric protein that is involved in the PTS sugar transport pathway in bacteria. The availability of high-resolution crystal structures for two homologues (*72, 114*) (see Figure 9 for a Cα overlay of *Bs* and *Bst*HPr structures) together with their relatively small size, lack of disulfide bonds and prosthetic groups makes these proteins good model systems for folding studies. The HPr homologues compared here show high sequence identity and yet show remarkable diversity in their thermodynamic behavior.

Results

Conformational stability of HPr homologues

The conformational stability of all the proteins in this study was determined through the analysis of solvent and thermal denaturation curves, where CD spectroscopy at 222 nm was used as a probe to follow the global conformation of the protein. Solvent denaturation experiments were performed using urea or guanidine hydrochloride (GuHCl). The transitions from native to denatured states in all cases were cooperative with flat native and denatured state baselines. In case of thermal denaturations, at least 90% reversibility was

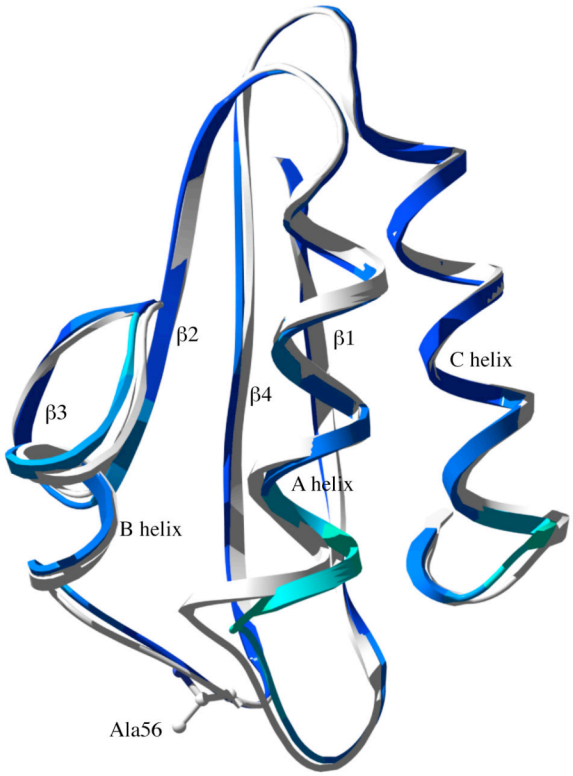


Figure 9: A ribbon diagram showing the superposition of the wild-type *Bs* and *Bst*HPr structures. Overlays of the structures were generated by a best fit superposition of the Cα atoms in the crystal structures for *Bs* (2HPR) and *Bst*HPr (1Y4Y). The *Bst*HPr chain is colored according to magnitude of RMSD between the Ca atoms for the two proteins, blue being low RMSD and cyan high. Shown in ball and stick is Ala 56 of *Bs*HPr. The overlay was done with Swiss PDB viewer (*90*) and rendered by POVRAY™.

observed (see Chapter II) and the transitions were independent of protein concentration over the studied concentration range of 5 to 50μM.

Typical results from thermal, urea, and GuHCl denaturation experiments are presented in Figure 10, panels a through c, and the thermodynamic parameters obtained for each HPr homologue are presented in Table 5. Urea and GuHCl denaturation experiments were performed at 25ºC in 10 mM sodium phosphate buffer at pH 7. Thermal denaturation experiments were performed in the same buffer also at pH 7.

The thermal denaturation experiments do not provide a direct measure of protein stability in terms of a free energy of stabilization (ΔG) in the absence of a measured ΔC_p (change in heat capacity upon protein folding). However, they do provide useful information in terms of melting temperatures (T_m) and change in enthalpy accompanying protein unfolding (ΔH). Variations of these parameters among proteins studied can provide useful insights into forces stabilizing the structure. Results from these experiments (Figure 10 panel a) show that HPr homologues have a wide range of thermostability, the melting temperature (T_m) shows a spread of ~30ºC, with *Oi*HPr (closed circles in Figure 10) having the lowest T_m at 58ºC and *Bst*HPr (open circles in Figure 10) having the highest T_m at 88.3ºC. The order of thermostability from thermal denaturation experiments is: *Bst*>*Bh*>*St*>*Bs*>*Oi*HPr. The ΔH values for all the proteins are listed in Table 5; again *Oi*HPr has the lowest ΔH and *Bst*HPr has the highest.

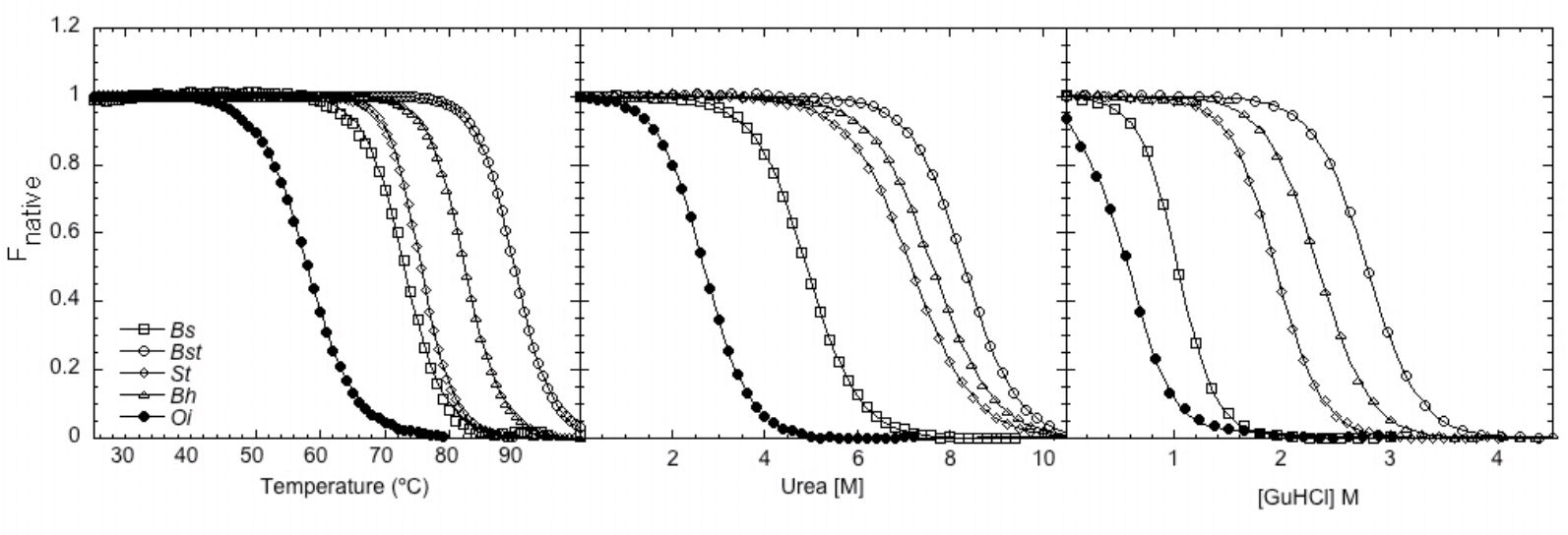


Figure 10: Representative thermal, urea and GuHCl denaturation unfolding curves for *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr. The symbols used are: *Oi* (●), *Bs* (□), *St* (◇), *Bh* (△) and *Bst*HPr (○), for thermal (panel a), urea (panel b) and GuHCl (panel c) denaturation experiments. All experiments were performed in 10 mM sodium phosphate at pH 7 and the data in panels b and c were collected at 25° C. Curves through the data are fits of equation 2 (panel a) or equation 3 (panels b and c) to the data. Data from CD at 222nm was normalized to fraction native using equations mentioned above.

Table 5: Parameters characterizing urea, GuHCl and thermal denaturations for *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr at pH 7.

	Urea denaturations		GuHCl denaturations		$\Delta\Delta G_{urea-GuHCl}$	Thermal denaturations	
Protein	**ΔG** (kcal mol^{-1})	**m-value**[a] (kcal mol^{-1} M^{-1})	**ΔG** (kcal mol^{-1})	**m-value**[b] (kcal mol^{-1} M^{-1})	(kcal mol^{-1})	**T_m** (ºC)	**ΔH**[c] (kcal mol^{-1})
***Bs*HPr**	5.00 (±0.16)	1.1	3.41 (±0.04)	3.3	1.6	73.0 (±0.6)	76.7
***St*HPr**	6.30 (±0.05)	0.9	5.37 (±0.10)	2.7	0.9	75.6 (±0.1)	95.4
***Bst*HPr**	8.20 (±0.20)	1.0	6.75 (±0.05)	2.4	1.4	88.3 (±0.8)	98.6
***Oi*HPr**	3.37(±0.20)	1.3	1.40 (±0.10)	2.6	2.0	58.0 (±0.8)	59
***Bh*HPr**	6.80 (±0.10)	0.9	5.35 (±0.20)	2.3	1.5	82.3 (±0.6)	87.2

All experiments were performed in 10 mM NaPi and the solvent denaturation experiments were performed at 25ºC. All values reported are averages and the values in parenthesis are standard deviations.

[a] Standard deviations for m value from urea denaturations were ≤ 5%

[b] Standard deviations for m value from GuHCl denaturations ≤ 5%

[c] Standard deviations in ΔH were usually ≤ 5% of the average value reported.

The results from urea denaturation shown in panel b of Figure 10, give a free energy of stabilization (ΔG) for the HPr proteins. The trends in ΔG values are similar to the ones reported in the T_m and ΔH values from thermal denaturation experiments, *Bst*HPr is found to be the most stable homologue and *Oi*HPr the least, with a difference of ~5 kcal mol^{-1} at 25ºC. The m values for urea denaturation were found to be very similar for most proteins and were found to be close to 1.0 kcal $mol^{-1}M^{-1}$, however *Oi*HPr has a significantly higher m value at 1.3 kcal mol^{-1} M^{-1}. Results from GuHCl denaturation experiments returned a ΔG of 6.8 and 1.4 kcal mol^{-1} for *Bst* and *Oi*HPr respectively. As in the case of urea denaturation experiments, the m values for the proteins are very similar, around 2.7 kcal mol^{-1} M^{-1}. *Bs*HPr appears to be the outlier with an m value of 3.3 kcal mol^{-1} M^{-1}. The $\Delta\Delta G_{urea\text{-}GuHCl}$ or the difference in stability as measured by urea and GuHCl denaturation experiments is the lowest for *St*HPr and highest for *Oi*HPr proteins at 0.9 and 2.0 kcal mol^{-1}, the other HPr homologues have $\Delta\Delta G_{urea\text{-}GuHCl}$ value close to 1.5 kcal mol^{-1}. This parameter has been used in the past as a preliminary estimate of electrostatic contributions to conformational stability of the proteins (*115*) and we propose to use it similarly.

Based on the results presented this far, we can conclude that the *Bst*HPr protein is the most stable homologue at 25ºC and is also most thermostable. For a more thorough characterization over a range of temperatures, we proceeded to construct a 'stability curve' for each HPr protein.

Stability curves for HPr homologues

Becktel and Schellman first described stability curves for proteins (*51*); these are plots of ΔG vs. temperature. Various methods of measuring stability can be used to construct stability curves. Here, we use the method described by Pace and Laurents where data from thermal and solvent denaturation experiments are combined to obtain a stability curve (*100*). The stability curves for the proteins are shown in Figure 11. The HPr homologues have very similarly shaped stability curves and appear to be shifted up or down on the y-axis according to the habitat temperature of the organism in question. The stability curve for *Bst*HPr is shifted up the most and the lowest curve is for *Oi*HPr. Stability curves also provide measures of other important thermodynamic properties, including the ΔC_p, which dictates the curvature of a stability curve. The values of ΔC_p are very similar among the different homologues (Table 6), and so are the values of T_S, the temperature of maximum stability or the temperature where change in entropy between native and denatured states is zero (Table 6). The T_S values for HPr homologues are close to 25ºC, with *Oi* and *St*HPr being the lowest at 22ºC. These values are, however the same within error.

The other parameter of interest that can be obtained from a stability curve is the free energy of stabilization at habitat temperature (ΔG_L). As the name suggests, it is simply the conformational stability of the protein at the temperature where the organism lives. The ΔG_L for the HPr homologues is very similar and is close to 5.0 kcal mol^{-1}, except for *Oi*HPr, which has a lower ΔG_L (3.2 kcal mol^{-1}). The similar ΔG_L values for these strongly support the corresponding state hypothesis, which was first proposed by Somero (*116-119*).

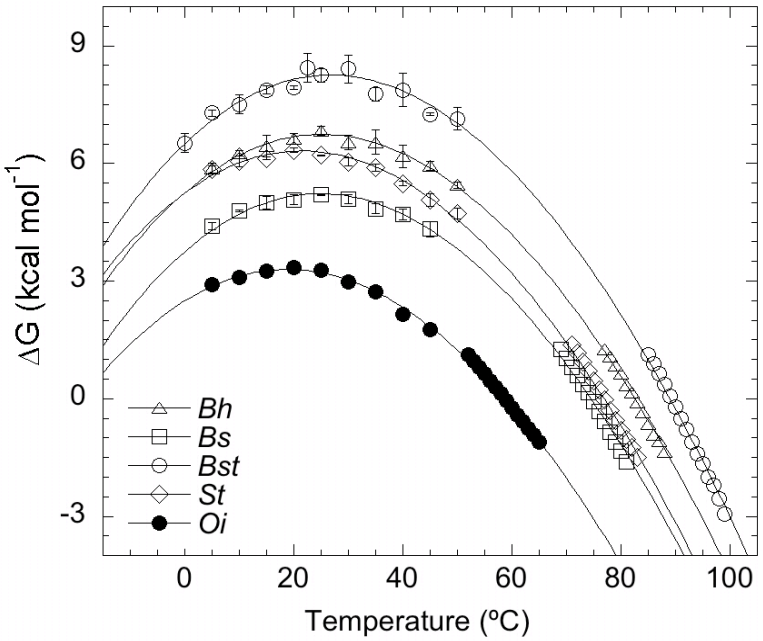


Figure 11: Stability curves (ΔG versus T) for the *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr proteins. The data at lower temperatures (0-50ºC) are from results of urea denaturation experiments at those temperatures, and the data at higher temperatures are from thermal denaturation experiments in the absence of urea. Curves through the data are fits of equation 1 to the combined data for a given protein. Symbols used follow the same convention as in Figure 10. The error bars for data in the low temperature range depict standard deviation from repeat measurements; for some data, error bars are smaller than the symbol. The best fit parameters describing the data are presented in Table 6.

Table 6: Parameters characterizing the stability of the *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr ptoteins.

Protein	ΔG_{25}	ΔG_S [a]	ΔG_L	T_L	T_S [b]	T_m	ΔC_p [c]
	(kcal mol^{-1})	(kcal mol^{-1})	(kcal mol^{-1})	(ºC)	(ºC)	(ºC)	(kcal $mol^{-1}K^{-1}$)
***Bs*HPr**	5.00 (±0.16)	5.2	4.8	20	24.1	74.4	1.33
***St*HPr**	6.30 (±0.05)	6.3	4.58	50	22	77	1.28
***Bst*HPr**	8.20 (±0.20)	8.2	5	58	27	88.9	1.34
***Oi*HPr**	3.40 (±0.20)	3.3	3.2	20	22	58.7	1.25
***Bh*HPr**	6.80 (±0.10)	6.7	4.87	55	25	82.2	1.3

Conformational stability at 25ºC, ΔG_{25} is the same as that reported in Table 5 from urea denaturation experiments, ΔG_S is the stability at T_S, which is the temperature of maximal stability, ΔG_L is the stability at habitat temperature of the organism (T_L), T_m is the melting temperature or the temperature where ΔG equals 0. Values for T_m and ΔC_p are best-fit estimates from fits of equation 1 to the data defining the stability curve. The values of ΔG_S, ΔG_L are from equation 1, values of T_L were those reported in literature.

[a] standard deviation in ΔG values obtained from fits of equation 1 ranged from 0.1 to 0.2 kcal mol^{-1}

[b] standard deviation in temperature estimates from fits of equation 1 were usually 0.1ºC

[c] errors in ΔC_p estimates are usually $\leq$10%

This hypothesis postulates that proteins balance their conformational stability with a need to be flexible, since flexibility is crucial to catalysis and other structural fluctuations required for protein function. Conformational stability, on the other hand, promotes rigidity in protein structures owing to the enhanced number of stabilizing interactions, like hydrogen bonds, hydrophobic interactions and ion-pair interactions as is usually observed. The similarity of ΔG_L values for the HPr homologues is testimony to this concept, where these proteins appear to have tuned their thermodynamic characteristics to provide for identical stability in each of their habitats.

Electrostatic contributions to protein stability

Salt effects on protein stability: To explore how electrostatic interactions contribute to the stability of the HPr proteins, we determined the conformational stability under various conditions of solution ionic strength and pH. For the study of ionic strength dependence, we have used urea denaturation experiments for the HPr proteins at various NaCl concentrations ranging between 0 and 1.2 M, in the presence of 10 mM sodium phosphate at pH 7 at 25ºC. We have presented the data from these experiments as a plot of normalized free energy of stabilization versus [NaCl], combined into one plot for the different homologues in Figure 12. We have normalized the free energies of stabilization in terms of the stability in the absence of salt; the stability for an HPr protein at a particular salt concentration is reported as the difference between stability in the absence of any added salt and stability at that salt concentration.

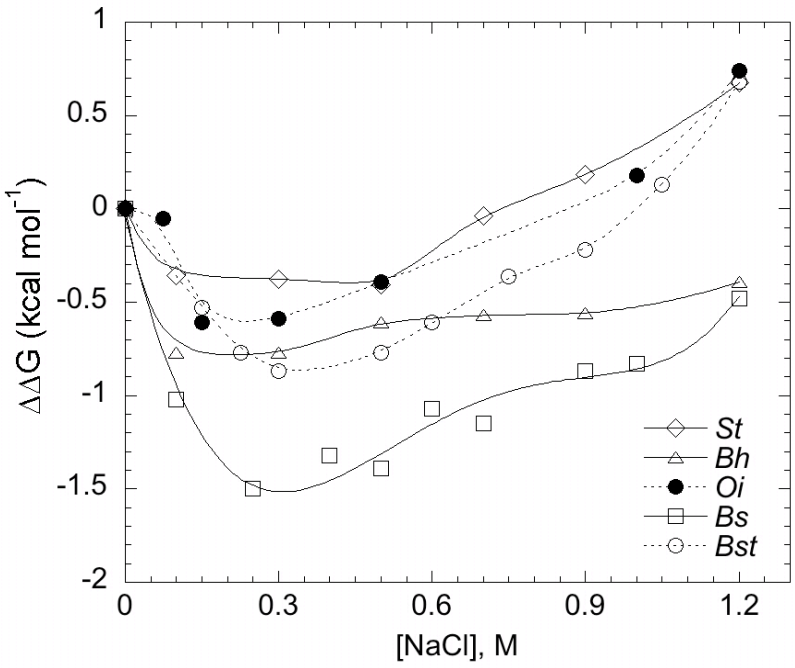


Figure 12: The changes in conformational stability (ΔΔG) as a function of [NaCl] for the *Oi*, *Bs*, *St*, *Bh*, and *Bst*HPr proteins. Stability measurements were made using urea denaturation experiments performed in increasing [NaCl] at 25ºC in 10 mM NaPi, pH 7. The stability data obtained were normalized with respect to stability of the protein in the absence of added NaCl ($\Delta\Delta G[NaCl] = \Delta G[NaCl] - \Delta G_{water}$). The curves through the data are only meant to guide the eye and symbols used follow the convention used in Figure 10. The solid curves through some of the data are meant to distinguish HPr homologues with different stability responses to added NaCl from the other homologues (see text for details).

The HPr homologues show varied stability behavior in the presence of added NaCl. The *Bst*, *Bs*, *Bh* and *Oi*HPr proteins all lose stability in the low [NaCl] (0 to 0.3 M NaCl) range, however *St*HPr appears to be relatively unaffected. In higher [NaCl] concentrations, above 0.3 M, all HPr homologues except *Bh*HPr show increased stability. The relative magnitudes of change in stability for the different HPr homologues in both concentration domains appear to be similar, except for *Bs*HPr, which in low [NaCl] shows a more pronounced decrease in stability (open squares in Figure 12). Whereas other HPr homologues show a ~0.9 kcal mol^{-1} drop in stability in low [NaCl], *Bs*HPr loses 1.5 kcal mol^{-1} in stability. In high [NaCl], all HPr homologues except *Bh*HPr (open triangles in Figure 12) show ~1 kcal mol^{-1} increase in relative stability when the [NaCl] is increased from 0.3 M to 1.2 M. The *Bh*HPr protein appears to be indifferent to increasing [NaCl] beyond 0.3 M, where the stability of the protein remains essentially the same within error.

pH effects on protein stability: We have investigated the effect of solution pH on the stability of the HPr homologues by performing urea denaturation experiments at various pH values from 3 through 10. For these experiments, we used a mixture consisting of sodium citrate, sodium phosphate and sodium borate at a final concentration of 10 mM as the buffer, which allowed us to perform experiments through the pH range in the same buffer. Urea denaturation experiments were performed at 25ºC using this buffer system and the results are presented in Figure 13 as a plot of ΔG as a function of solution pH. Again, we have combined the results for the HPr protein into one panel.

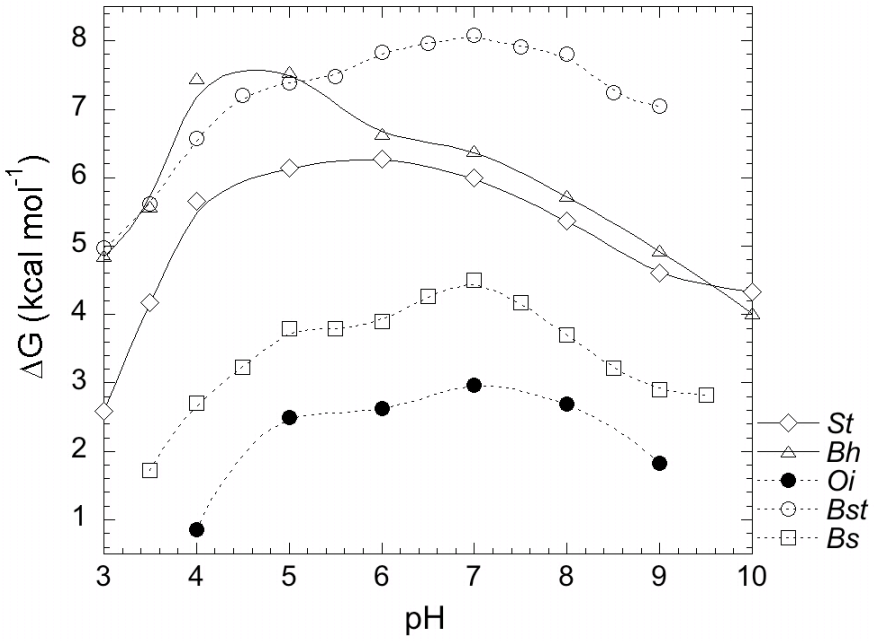


Figure 13: Conformational stability, ΔG (25ºC) of *Oi*, *Bs*, *St*, *Bh* and *Bst*HPr proteins as a function of pH. Urea denaturation experiments were used to measure stability of the proteins at different pH at 25ºC. The buffer used was a mixture of sodium citrate, sodium phosphate, and sodium borate at final concentration of 10 mM. The curves through the data are meant only to guide the eye and symbols used follow the convention used in Figure 10. The solid curves through some of the data are meant to distinguish HPr homologues with different stability responses to solution pH from the other homologues (see text for details).

We find that three of the five HPr homologues namely, *Bs*, *Bst* and *Oi*HPr (squares, open circles and closed circles respectively in Figure 13) show maximal stabilities at pH 7. The *Oi*HPr protein was completely unfolded at pH 3 and 10 at 25ºC and we could not collect any useful data at these pH values. The drop in stability between pH 7 and 3 is ~3.2 kcal mol^{-1} for *Bs*HPr and ~3 kcal mol^{-1} for *Bst*HPr; *Oi*HPr loses 2.1 kcal mol^{-1} in stability when the pH is reduced from 7 to 4. The stability also decreases at pH higher than 7 for these proteins. This decrease in stability is however most pronounced for *Bs*HPr when pH is increased from 7 to 9 (1.60 kcal mol^{-1}), than it is for *Oi* or *Bst*HPr (1.10 and 1.03 kcal mol^{-1}, respectively). The other two HPr homologues namely *St* and *Bh*HPr do not show stability maxima at pH 7. Instead *St*HPr (diamonds in Figure 13) is most stable at pH 4 and 5 and *Bh*HPr is most stable between pH 5 and 6 (triangles in Figure 13). The *St*HPr protein loses more stability when pH is reduced from 6 to 3 (3.7 kcal mol^{-1}) than *Bh*HPr, which loses 2.7 kcal mol^{-1} when the pH is reduced from 5 to 3. In the alkaline pH domain, however, *Bh*HPr is more destabilized when the pH is raised from 5 to 10 (3.5 kcal mol^{-1}) than *St*HPr, which loses 2 kcal mol^{-1} in stability when the pH is raised from 6 to 10. In general, it can be said that all the HPr homologues behave similarly in the acidic pH domain when the solution pH is reduced from where they are most stable. In other words, all HPr homologues are destabilized to the same extent in acidic conditions. However, in alkaline solutions, *St* and *Bh*HPr show larger relative decreases in stability from their maxima when compared with the other three HPr homologues.

To summarize results presented this far: we have established that the HPr protein from *Bacillus staerothermophilus* is the most stable of all the HPr homologues studied here as it is the most thermostable homologue and also has the highest conformational stability. We

have also established a trend in stabilities of the HPr homologues *Bst*>*Bh*>*St*>*Bs*>*Oi*HPr, and this trend holds true for both conformational stability as measured by solvent denaturation experiments and thermal stability. Stability curves for these HPr homologues have also been determined and it is clear that the magnitude of stability for each homologue is a reflection on the thermal environment inhabited by the organism. In terms of electrostatic contributions to protein stability, it was shown that most of the homologues show common behavior in terms of their response to solution pH and salinity; however, *St* and *Bh*HPr appear to be outliers with respect to these trends in stability. We now proceed to a brief description of sequence alignments for these HPr homologues and then a discussion of the results.

Sequence alignments

For sequence alignments we have grouped the HPr homologues as temperature-adapted proteins (*St* and *Bst*HPr) and haloalkaliphiles (*Oi* and *Bh*HPr). For both groups we have aligned the protein sequences with the sequence of *Bs*HPr, the mesophilic homologue. We have also included an alignment for all five HPr homologues. The results generated by CLUSTALW (*84*) are shown in Figure 14. In panel a of Figure 14 we show an alignment of the *Bs*, *St* and *Bst*HPr proteins. Whereas *St*HPr has 87 residues the other two homologues both have 88 residues, the CLUSTALW alignment score for *St* with *Bs*HPr was 64 and for *Bst* with *Bs*HPr was 70. For the most part the sequence of these homologues show conservative and semi-conservative replacements up to residue 69, after which we see more

a

```
Bs  MAQKTFKVTADSGIHARPATVLVQTASKYDADVNLEYNGKTVNLKSIMGVMSLGIAKGAEITISASGADENDALNALEETMKSEGLGE 88
St  MASKDFHIVAETGIHARPATLLVQTASKFASDITLEYKGKAVNLKSIMGVMSLGVGQGADVTISAEGPDADDAIAAIAETMTKEGLA- 87
Bst MAEKTFKVVSDSGIHARPATILVQTASKFNSEIQLEYNGKTVNLKSIMGVMSLGIPKGATIKITAEGADAAEAMAALTDTLAKEGLAE 88
    **.* *::.:::********:*******: ::: ***:**:**************: :** :.*:*.*.*  :*: *: :*: .***.
```

b

```
Bs MAQKTFKVTADSGIHARPATVLVQTASKYDADVNLEYNGKTVNLKSIMGVMSLGIAKGAEITISASGADENDALNALEETMKSEGLGE 88
Oi MKSQTFTITAETGVHARPATLLVNKAGQFDSEIEVSYKGKQVNLKSIMGVMSLGIPKGAEIEVSADGKDEEDALNGVAEVIK-EHLGE 87
Bh MAEKTFTITAETGIHARPATQLVNKAGQYSSEITLEYKGKAVNLKSIMGVMSLGVGKGAQVTIKAEGSDEAEALKGIEEVIK-EGLGE 87
   * .:**.:**::*:****** **:.*.::.::: :.*:** ************: ***:: :.*.* ** :**:.: *.:* * ***
```

c

```
Bs  MAQKTFKVTADSGIHARPATVLVQTASKYDADVNLEYNGKTVNLKSIMGVMSLGIAKGAEITISASGADENDALNALEETMKSEGLGE 88
Bst MAEKTFKVVSDSGIHARPATILVQTASKFNSEIQLEYNGKTVNLKSIMGVMSLGIPKGATIKITAEGADAAEAMAALTDTLAKEGLAE 88
St  MASKDFHIVAETGIHARPATLLVQTASKFASDITLEYKGKAVNLKSIMGVMSLGVGQGADVTISAEGPDADDAIAAIAETMTKEGLA- 87
Oi  MKSQTFTITAETGVHARPATLLVNKAGQFDSEIEVSYKGKQVNLKSIMGVMSLGIPKGAEIEVSADGKDEEDALNGVAEVIK-EHLGE 87
Bh  MAEKTFTITAETGIHARPATQLVNKAGQYSSEITLEYKGKAVNLKSIMGVMSLGVGKGAQVTIKAEGSDEAEALKGIEEVIK-EGLGE 87
    *  . * . ...*******  **. * .. ... . *.** *************. .** . . * * *  .*.  . . .  * *
```

Figure 14: Sequence alignments for *Bs*, *St*, *Bst*, *Oi* and *Bh*HPr proteins. Panel a shows an alignment for the *St* and *Bst*HPr proteins with *Bs*HPr. Panel b shows an alignment for the *Oi* and *Bh*HPr proteins with *Bs*HPr and panel c shows an alignment for all the HPr homologues studied here. Underlined segments correspond to the C helix. All alignments were generated by CLUSTALW (*84*) and the symbols used are standard notations for identity (*), conserved (:), semi-conserved (.) or non-conserved ( ) changes between aligned sequences.

non-conserved replacements. This latter part of the sequence corresponds to the last element of secondary structure called the C helix. Also, of interest is residue 56, which in *Bst*HPr is a proline. This is curious because residue 56 is part of a surface loop connecting two elements of secondary structure (see Figure 9 for position of residue 56 in the protein structure) and presence of prolines in such loops has been shown to contribute positively to protein stability (*120*). Of similar interest is the presence of a proline residue at position 68 in *St*HPr. This is the only HPr homologue of the five studied here that has a proline at this position.

The C helix sequences (residue 69 through 85) were analyzed for helix content using the AGADIR algorithm (*121*). The results from AGADIR ranked the *St*HPr sequence as having the highest helix content followed by the *Bst* and *Bs*HPr sequences, among temperature-adapted homologues. Similarly, the AGADIR algorithm was used to analyze the C helix sequences from *Oi* and *Bh*HPr, and this analysis ranked *Bh*HPr higher than *Oi*HPr in terms of helix content. In fact *Oi*HPr ranked the lowest amongst all five homologues and *Bh*HPr ranked the highest in helix content amongst the HPr homologues.

The sequence alignment of haloalkaliphilic HPr homologues, *Oi* and *Bh*HPr with *Bs*HPr is shown in panel b of Figure 14. Both *Oi* and *Bh*HPr have 87 residues whereas *Bs*HPr has 88; also, *Oi*HPr is the only HPr homologue described in this study that retains the N-terminal methionine in the over-expressed protein. The CLUSTALW alignment score for *Oi*HPr with *Bs*HPr was 56 and that for *Bh*HPr with *Bs*HPr was 64. *Oi*HPr is also peculiar in having a lysine at position 2. No other HPr homologue has a lysine at this position, and all the other homologues described here have an alanine at this position (see panel c in Figure 14 for a sequence alignment of all HPr homologues). The C helix

sequences of *Oi* and *Bh*HPr are more similar to the *Bs*HPr C helix than the C helices of *St* and *Bst*HPr. *Oi*HPr is also peculiar in having a proline at residue 56 like *Bst*HPr. These are the only homologues that have a proline at this position. The others have either a glycine (*St* and *Bh*HPr) or an alanine (*Bs*HPr) at this position. The number of acidic residues is also the highest in *Oi*HPr (14 Asp and Glu residues), compared with other homologues, which have 10 (*Bst*HPr), 11 (*St*HPr) or 12 (*Bh* and *Bs*HPr). This is not surprising given the established role of acidic residues especially Glu, that are capable of binding more water than other residues. This helps create a hydration shell that keeps the protein solvated in high salt conditions, where other proteins lacking excess negative charges would aggregate (*122-125*).

Discussion

Proteins from extremophiles provide unique insights into the forces involved in protein stability. The environmental stresses exerted on these proteins cause them to adapt in ways that are not seen in proteins from organisms inhabiting moderate environments. Thermophiles are of special interest because proteins from these organisms have to adapt to the elevated habitat temperatures, unlike other environmental stresses that can be circumvented by compensatory mechanisms. We have presented here a thermodynamic characterization of five HPr homologues derived from organisms inhabiting diverse environments. We find that, *Bst*HPr is the most stable of all the homologues. It has a ΔG of 8.2 kcal mol^{-1}, which is ~5 kcal mol^{-1} higher than that for the least stable homologue *Oi*HPr. Also, *Bst*HPr is the most thermostable homologue, it has a T_m of 88.3ºC which is ~30ºC higher than that of *Oi*HPr. The *Bst*HPr is more stable than the other homologues at all temperatures as shown by the stability curve. Solvent denaturation experiments also provide

an m value and this parameter has been correlated with the size of a protein and the amount of surface area it exposes upon unfolding (*24*). Given the similar size of these homologues it not very surprising that the m values are the same for most proteins. However, outliers like *Oi*HPr and *Bs*HPr, make an interesting observation and warrant a closer look, if possible with structural insights.

Protein stability curves also provide a thermodynamic description for the higher thermostability exhibited by proteins from thermophiles (see Chapter II for a detailed discussion on stability curves). In the case of HPr homologues however, it is seen that the stability curves are shifted to either higher or lower ΔG values depending on the habitat temperature of the organism, without any significant changes in ΔC_p or T_S. The stability curves for HPr homologues also reveal a very similar ΔG_L or free energy of stabilization at habitat temperature T_L. Although not unique, this observation is notable for the proximity of ΔG_L values of HPr homologues; for example ΔG_L have been reported for the RNase H homologues (*57*) from *Escherichia coli* and *Thermus thermophilus*, but the agreement in ΔG_L is not as remarkable in that case. In the case of *Oi*HPr, we feel that the extreme haloalkaliphilic nature of this organism requires high salinity together with high pH for optimal stability. Since we have not reproduced these conditions *Oi*HPr shows significantly lower stability at all temperatures.

Structural comparisons between pairs of proteins derived from mesophiles and thermophiles have proven to be indispensable in comparative studies of the kind we have undertaken here. Structural comparisons point toward stabilizing interactions, which might be present specifically in the thermophilic homologue. A structural comparison between two of the five HPr homologues for which structures are available was reported earlier

(Table 9 in Chapter IV). It was found that the two proteins have very similar structures with very similar number of hydrogen bonds and salt bridge interactions, within errors of estimation and taking into account the resolution of the structures.

In recent statistical surveys comparing proteins from mesophiles and thermophiles, it was found that of all the parameters compared, the only feature of proteins from thermophiles that shows a positive correlation with growth temperature was the number of interactions of electrostatic nature (*38, 126*). Proteins from thermophiles have been shown to exhibit higher numbers of surface charges, which are purported to be involved in stabilizing interactions like salt bridges and side-chain hydrogen bonds (*2, 127-130*). Given this fairly established role of electrostatic interactions in stability of proteins from thermophiles, we felt it was pertinent to investigate the role of these forces in our set of homologues. To this end, we have measured protein stability for all the HPr homologues in different solution conditions, varying ionic strength or pH.

Salt, (NaCl in the case of this study) can have varied effects on the stability of a protein. These effects are mainly due to a) Debye-Hückel screening of electrostatic interactions contributing to the protein's stability b) specific ion binding either to the folded or unfolded states c) salting-in and salting-out effects (*131-134*). In this study, we investigated the first of the three effects mentioned above: the Debye-Huckel screening of electrostatic interactions. This was possible in our case for two reasons: 1) the HPr homologues studied here are almost identical in size and all homologues have similar non-polar amino acid content except *Oi*HPr, which has a lower non-polar content. 2) data obtained from earlier studies in the lab have ruled out high affinity ion binding to specific sites in the wild-type *Bs*HPr protein (Schmittschmitt and Scholtz unpublished results). We

also eliminated specific ion binding for *Bst*HPr as well, and we assumed this to be the case for the other homologues. With this perspective, we proceeded with a comparison of stability behavior of HPr homologues using the relatively neutral Hoffmeister salt, NaCl. The differential effects of [NaCl] in the 0-1 M range on protein stability are a good measure of the different electrostatic effects operating in proteins being compared. Beyond 1 M concentrations, the Hoffmeister effects begin to dominate and these effects are independent of the particular protein, acting instead by increasing the surface tension of water, adding to the hydrophobic force.

Our results show that each HPr homologue behaves slightly differently in response to added NaCl. Broadly speaking, however, *Bs*, *Bst* and *Oi*HPr show a similar trend in stability, whereas *St*HPr is indifferent to NaCl at low concentrations, and *Bh*HPr is indifferent to NaCl at high concentrations. It might also be useful to compare these results by grouping proteins according to their habitat conditions, as in temperature-adapted proteins (*St* and *Bst*HPr) and haloalkaliphilic proteins (*Bh* and *Oi*HPr), and comparing them with the model mesophilic protein *Bs*HPr. It can be seen that *Bst*HPr shows stability features similar to *Bs*HPr, although the magnitude of decrease in stability is higher for *Bs*HPr suggesting rather counter intuitively that the mesophilic homologue has more optimized electrostatic interactions than the thermophilic *Bst*HPr. In the case of *St*HPr, the other moderately thermophilic HPr homologue, it appears that electrostatic interactions play a relatively small role in the stability since screening by NaCl has little effect (the $\Delta\Delta G_{0\text{-}0.3}$ or the difference in stability at 0 and 0.3 M NaCl, for the for *St*HPr is 0.37 kcal mol^{-1} and for *Bst*HPr is 0.77 kcal mol^{-1}). This result is also counterintuitive like the case of *Bst*HPr protein, where it appears that a thermophilic protein has fewer contributions from

electrostatic forces when compared with a mesophilic protein. In the case of haloalkaliphilic proteins (*Bh* and *Oi*HPr), simplistic interpretations of results would be naïve since the role of membranes and ion-specific exchange mechanisms is fairly well established in these organisms and a realistic estimate of cytoplasmic ion strength and pH conditions is far from trivial (*108*). Nevertheless, it is tempting to point that in 1.5 M NaCl, the *Oi*HPr protein is almost as stable as its *Bs*HPr homologue, especially since *Oceanobacillus halodurans* has demonstrated viability in twice as concentrated NaCl solutions. *Bacillus halodurans* has also demonstrated viability in high [NaCl] (~2 M) and curiously the *Bh*HPr protein appears indifferent to added NaCl in this high concentration range. It is worth reiterating however, that these observations do not take into consideration the actual cytoplasmic ionic strength conditions operating in these organisms.

We have also investigated the electrostatic contributions to protein stability by comparing the pH dependence of protein stability. pH perturbations can favor either the folded or unfolded states of a protein, and is a consequence of side-chain ionizable groups having different pKa values in either the native or denatured states of a protein. For example, if a side-chain carboxyl group has a reduced pKa in the native state, the denatured state binds protons more strongly than the native state. Therefore as the pH decreases the denatured state becomes more favored. In RNase A for example, the Asp 121 side chain has a pKa of 2.4 in the native state, which is depressed by over one pKa unit from the pKa in the denatured state (*135*). pH perturbations also affect protein stability by causing changes in ionization states of side-chain groups involved in favorable or unfavorable interactions, thereby causing changes in the stability. We hoped to investigate any differences in such interaction among the HPr proteins; and thus we have studied the

stability of the different homologues as function of pH. A comparison of such data for the HPr homologues puts the proteins in two groups. The first group is comprised of *Bs*, *Bst* and *Oi*HPr. These proteins all have maximal stabilities at pH 7. The second group comprises *St* and *Bh*HPr and both of these proteins have maximal stability in the acidic pH range; while *St*HPr shows maximum stability at pH 5 and 6, *Bh*HPr shows maximum stability at pH 4 and 5. Among the temperature-adapted proteins, *St*HPr is an outlier as seen in the case of salt dependent stability studies. The other two homologues, *Bs* and *Bst*HPr both have maximal stabilities at pH 7. Interestingly, the *St* and *Bh*HPr proteins also showed atypical behavior in the salt dependence studies. The concurrence of these results from pH and salt dependence studies warrants a more detailed analysis of the electrostatic forces operating in the *St* and *Bh*HPr proteins, involving mutagenesis experiments on these homologues, we hope to pursue such a study in the future. These studies will be greatly aided by availability of high-resolution structures for the *St* and *Bh*HPr proteins; work is in progress toward this end in our labs.

Finally, we have used the differences in stability measured by urea and GuHCl denaturation experiments ($\Delta\Delta G_{urea\text{-}GuHCl}$) for the wild-type proteins as a first approximation of electrostatic forces acting in that protein. As mentioned earlier, this parameter has been used by Kohn et al in an elegant study to estimate the electrostatic component to the stability of leucine zipper coiled coil variants (*115*). A difference in $\Delta\Delta G_{urea\text{-}GuHCl}$ for a homologue from the chosen model protein will point to differences in electrostatic forces operating in these cases. When comparing the $\Delta\Delta G_{urea\text{-}GuHCl}$ values, we find that the HPr homologues fall into mainly one group with two outliers. *Bs*, *Bst* and *Bh*HPr all have a $\Delta\Delta G_{urea\text{-}GuHCl}$ of ~1.5 kcal mol^{-1}, *St*HPr has a low value of 0.9 kcal mol^{-1} and *Oi*HPr has a

high value of 2.0 kcal mol^{-1}. In the case of *Oi*HPr however, we are skeptical about the ΔG estimate from GuHCl denaturation data because of the paucity of data in the pre-transition region at 25ºC. It is not surprising that *St*HPr is an outlier with respect $\Delta\Delta G_{urea\text{-}GuHCl}$ given our findings in the case of pH and salt dependence studies, where this protein was an outlier. That being said, it is surprising that the $\Delta\Delta G_{urea\text{-}GuHCl}$ value for *Bh*HPr is in good agreement with those for other HPr.

We have presented here a detailed characterization of five HPr homologues derived from organisms inhabiting diverse environments as a first step toward understanding the origins of their varied stability behaviors. To our knowledge this is one of the few studies that provide thermodynamic characterization of protein homologues derived from temperature, high salinity and pH-adapted organisms. This work will be followed up with structural information and mutagenesis experiments, which will provide more insights into the molecular basis of adaptation of these proteins to their diverse habitats. The insights from an analysis of sequence alignment will be useful in further studies on these protein homologues.

CHAPTER IV

STRUCTURAL ANALYSIS OF HPr HOMOLOGUES AND THEIR SOLUTION CHARACTERIZATION*

Introduction

Proteins from extremophilic organisms pose a significant challenge to our understanding of protein folding. These proteins exhibit unusually high stabilities and are extremely well adapted to their environments. Several studies have focused on the issue of enhanced stability of proteins from thermophilic organisms (*11, 37, 57, 126, 136, 137*). These studies have proposed several mechanisms for achieving thermostability in these proteins; however, none of these provide a universal paradigm for the increased stability of most proteins from thermophiles.

To elucidate the mechanisms of thermal adaptation in proteins, one of the principal approaches has been to provide a direct comparison of thermodynamics and folding between a thermophilic protein and its mesophilic counterpart. To completely understand these comparisons, it is desirable to have structural information to help understand the molecular determinants of the differences in folding and stability. Here, we used the histidine-containing phosphocarrier protein (HPr) as our model system. It is an attractive system because we have characterized the folding and stability of the mesophilic homologues of this protein from *Bacillus subtilis* (*Bs*HPr) and *Escherichia coli* (*138, 139*) and

* Reprinted with permission from 'The HPr proteins from the thermophile *Bacillus staerothermophilus* can form domain-swapped dimers', by Sridharan S, Razvi A, Scholtz J. M. and Sacchettini J. C. *Journal of Molecular Biology*, 346, 919-931. Copyright 2005 Elsevier.

the *pts*H gene encoding the HPr protein has been cloned and over-expressed from the thermophilic organism *Bacillus stearothermophilus* (*88*). Here, we report the purification, solution properties, conformational stability and the three-dimensional structure of the HPr protein from *Bacillus stearothermophilus* (*Bst*HPr) and a variant that adds a single Trp to the protein, F29W.

Crystal and NMR solution structures for the mesophilic variants of HPr are available (*114, 140-143*). The availability of high resolution crystal structures for the thermophilic HPr might allow us to enumerate and compare the factors that might impart increased thermostability, such as differences in the number of ion-pair interactions, hydrogen bonds and buried surface area. Here, we have solved the crystal structure of *Bacillus stearothermophilus* HPr and a site-directed mutant with decreased thermostability.

Methods for protein crystallography[1]

Crystallization

The wild-type and F29W variant of the *Bst*HPr protein were crystallized using ammonium sulfate as the precipitant. All crystals were grown using the hanging drop vapor diffusion method at 18°C by mixing equal volumes (3-5 μl) of protein (20-30 mg/ml) and precipitant. The wild- type protein was dissolved in 100 mM citrate buffer pH 6.3 and crystallized in 2 M ammonium sulfate (Hampton Research Screen I, condition 32). Crystals of the monomeric F29W variant were obtained using 2 M ammonium sulfate, 100 mM phosphate-citrate pH 4.2 (Wizard Screen II, condition 9) with the protein dissolved in 20

[1] The crystallography experiments were conducted by Sudharsan Sridharan in the labs of James C. Sacchettini in the Department of Biochemistry and Biophysics at Texas A&M University. These authors also wrote the sections in this manuscript on crystallography.

mM Tris pH 7.0. The domain-swapped dimeric variant formed crystals in 68-74 % saturated ammonium sulfate with the protein dissolved in citrate buffer pH 6.3.

Structure determination

Paratone-N was used as the cryoprotectant for all crystals. X-ray intensity data were collected at Beamline 19 at the Advanced Photon Source (APS), Argonne National Lab, Illinois and the Center for Advanced Microstructures and Devices (CAMD), Baton Rogue, Louisiana. Data processing, integration and scaling were done using the HKL2000 software package (*144*). Both the monomeric wild-type and the F29W variant of *Bst*HPr crystals belong to the space group $I4_122$, with 3 molecules per asymmetric unit and 46% solvent content with cell dimensions of a = b = 74.83 Å and c = 179.44 Å for the wild-type and a = b = 75.71 Å and c = 177.78 Å for the variant. The domain-swapped dimeric variant crystallized in the I222 space group with cell dimensions
a = 34.86 Å, b = 63.45 Å and c = 82.88 Å with one molecule per asymmetric unit.

An initial model of the *Bst*HPr structure was obtained by submitting the sequence to SwissModel server (*145*) and this model was used for finding molecular replacement solutions with the MolRep (*146*) program in the CCP4 suite (*147*). Structure refinements and map calculations were performed using the CCP4 program suite (*147*) and the TB Consortium Bias Removal Server (*148*)and model building was done using Xtalview (*149*). At each phase of refinement, the R-free and R-factor improvement was monitored to make sure that the model was not over-refined.

Results

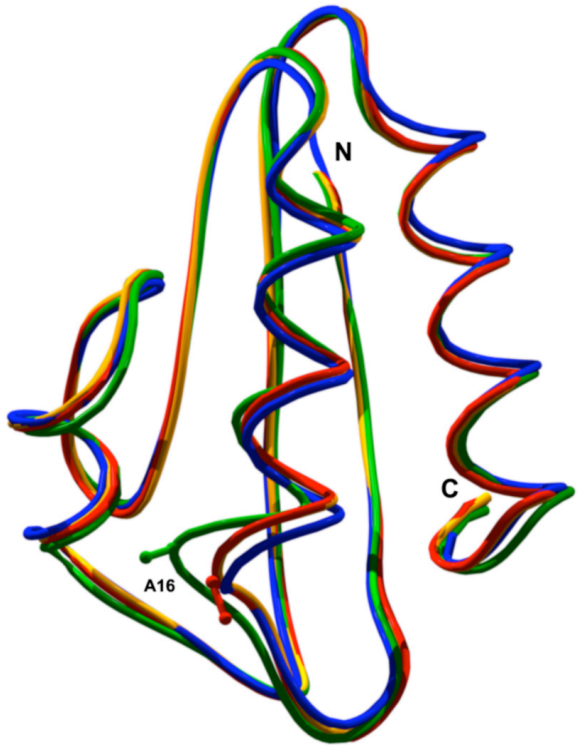
Crystal structures of wild-type and the F29W variant of *Bst*HPr

The X-ray crystal structures of the wild-type *Bst*HPr and the F29W variant have been solved using molecular replacement. Data collection and refinement statistics are given in Table 7. The wild-type protein crystallized in the space group $I4_122$ with unit cell dimensions a = b = 74.83 Å and c = 179.44 Å using 2 M ammonium sulfate as the precipitant. The F29W mutant crystallized in two different space groups, despite the fact that similar conditions were employed in the two independent crystal trials and these conditions were also similar to the wild-type protein crystallization condition. One form is the same as that of the wild-type protein with the same space group ($I4_122$) and very similar cell dimensions (a = b = 75.71 Å; c = 177.78 Å) and crystallized using 2 M ammonium sulfate as the precipitant. The overall fold of the wild-type and this form of the F29W variant is an open-faced β-sandwich just as in other HPr proteins of known structure in which a β sheet is packed against three α helices (*114, 141-143*). Figure 15 shows an overlay of Cα traces of the F29W variant and the wild-type *Bst*HPr proteins. The structures are virtually identical and the overall RMS deviation between the Cα atoms is only 0.2 Å. Figure 15 also shows overlays of the Cα traces of the wild-type *Bst*HPr and those of two mesophilic HPr proteins from *Bacillus subtilis* HPr (2HPR (*114*), referred to as *Bs*HPr throughout the rest of the chapter) and *Streptococcus faecalis* HPr (1PTF (*141*), referred to as *Sf*HPr in the rest of the chapter).

Table 7: Summary of data collection and refinement statistics for *Bst*HPr, F29W *Bst*HPr and domain-swapped structure of F29W *Bst*HPr.

	Wild-type *Bst*HPr	**F29W *Bst*HPr monomer**	**F29W *Bst*HPr dimer**
Data collection :			
Lattice type	Tetragonal	Tetragonal	Orthorhombic
Space group	$I4_122$	$I4_122$	I222
Cell dimensions (a, b, c Å)	74.83, 74.83, 179.44	75.71, 75.71, 177.78	34.86, 63.45, 82.88
Resolution (Å)	70.0-2.0	50-1.64	50.0-2.0
Completeness % (last shell)	99.7 (99.9)	98.4 (90.5)	94.4 (70.9)
I/σ(I) (last shell)	17.1 (3.2)	34.7 (2.5)	31.6 (2.0)
No. of reflections	124959	301503	32271
No. of unique reflections	17682	31548	6174
R_{sym} % (last shell)	11.2 (55.3)	6.6 (79.4)	5.2 (24.4)
Redundancy	7.1	9.6	5.2
Refinement :			
Resolution (Å)	20.0-2.0	20.0-1.65	25.0-2.0
No. of reflections in test set	899	1589	290
No. of reflections in working set	16731	29874	5866
No. of protein molecules per asu	3	3	1
No. of protein atoms in asu	1896	1942	645
No. of water molecules in asu	286	372	73
R-factor (%)	21	20.5	21.9
R_{free} (%)	28.2	24.6	29.7
RMS deviations			
Bond lengths (Å)	0.016	0.017	0.011
Bond angles (degrees)	1.406	1.559	1.388
Average B-factors (Å^2)			
Main chain	19.5	24.5	44.7
Side chain and waters	26.21	32.02	52.7
RMS B main chain	1.44	1.42	1.7
RMS B side chain	3.29	3.29	3.75

The PDB accession codes for these structures are 1Y4Y, 1Y51 and 1Y50 respectively.




```
Bst   -AEKTFKVVSDSGIHARPATILVQTASKFNSEIQLEYNGKTVNLKSIMGVMSLGIPKGAT 60
Bs    -AQKTFKVTADSGIHARPATVLVQTASKYDADVNLEYNGKTVNLKSIMGVMSLGIAKGAE 60
Sf    MEKKEFHIVAETGIHARPATLLVQTASKFNSDINLEYKGKSVNLKSIMGVMSLGVGQGSD 60
Crh   -VQQKVEVRLKTGLQARPAALFVQEANRFTSDVFLEKDGKKVNAKSIMGLMSLAVSTGTE 60
        :: ..:  .:*::*****::** *.:: ::: ** .**.** *****:***.:  *:
Bst   IKITAEGADAAEAMAALTDTLAKEGLAE 87
Bs    ITISASGADENDALNALEETMKSEGLGE 87
Sf    VTITVDGADEAEGMAAIVETLQKEGLAE 88
Crh   VTLIAQGEDEQEALEKLAAYVQEEV--- 84
      :.: ..* *  :.:  :   : .*
```

Figure 15: Ribbon diagram showing superposition of *Bs*, *Bst*, *Sf* and F29W *Bst*HPr structures. The best-fit superpositions were generated using the Swiss-PDB viewer (*90*) program and images were rendered using POVRAY™. Cα trace overlays for *Bst* (red), *Bs* (blue), *Sf* (green) and the F29W variant of *Bst*HPr (yellow) are shown. Ala16 is shown in a ball and stick representation for the *Sf* and *Bst*HPr proteins. The sequence alignments for *Bst*HPr, *Bs*HPr, *Sf*HPr and *Bs*Crh generated using the CLUSTALW program (*84*) are also shown.

The sequence alignments of these HPr proteins, and a related protein Crh (see below), are also shown. The average RMS differences between the Cα atoms of the wild-type *Bst*HPr and *Bs*HPr and *Sf*HPr are 0.61 Å and 0.62 Å, respectively. *Bs*HPr shows the largest Cα RMS deviations from the wild-type *Bst*HPr structure between residues 13 and 18 with an average deviation of 1.41 Å and the highest deviation (2.12 Å) at the active site His15. As in *Sf*HPr, Ala16 is in a sterically strained conformation in both the wild type and the F29W variant of *Bst*HPr. The backbone carbonyl group of His15 points directly towards the ring of Pro18 in *Sf*HPr and *Bst*HPr, whereas in *Bs*HPr it forms a hydrogen bond with the backbone nitrogen of Ala19.

Domain-swapped dimeric form of the F29W variant of *Bst*HPr

The F29W variant of *Bst*HPr also crystallized in a second crystal form using 2.38 to 2.50 M ammonium sulfate as the precipitant (68-74 % saturation). The crystals belong to the I222 space group with cell dimensions a = 34.86 Å, b = 63.45 Å and c = 82.88 Å. The structure of the F29W variant that formed in this space group shows a domain-swapped dimer in the asymmetric unit. The overall fold of each half of the dimer is very similar to the monomer of the wild-type protein, but formed by portions from two different polypeptide chains, as observed in other domain-swapped dimers (*150*). Figure 16 shows a ribbon representation of the dimer superimposed on the monomeric form of the F29W variant of *Bst*HPr. When residues 2-52 of the monomer are superimposed over the corresponding residues in one molecule of the dimer and residues 56-88 are superimposed on the other molecule, the RMS deviation between the Cα atoms was found to be 0.72 Å.

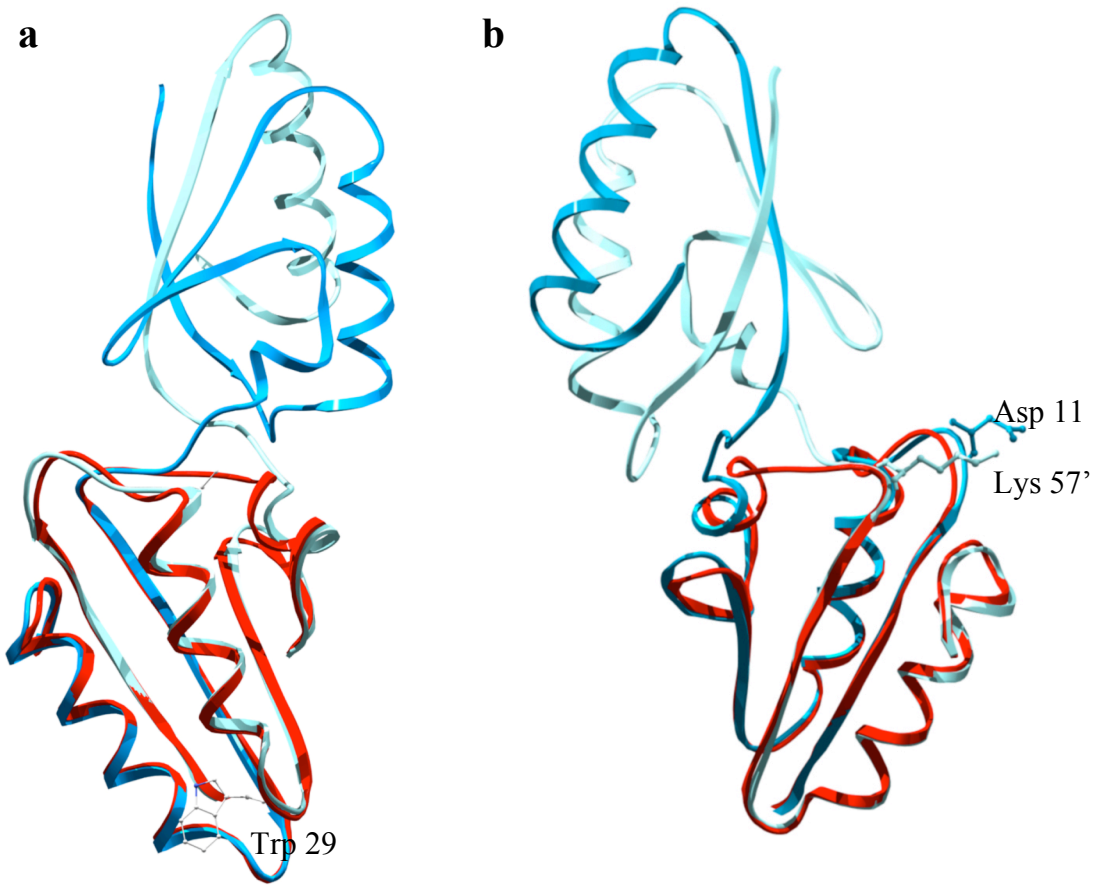


Figure 16: Ribbon diagram showing a superposition of structures for monomeric and dimeric forms of F29W *Bst*HPr. Two views of the overlay are shown; the monomer is shown in red and the dimer chains are shown in light and dark blue. Panel a shows the side chain of Trp29 of the monomer in a ball and stick representation. Panel b shows the side chains of Asp11 and Lys57' of the partner monomers involved in an intermolecular salt bridge interaction. The Cα superpositions of the monomer and dimer structures were performed using the LSQKAB (*151*) script of the CCP4 suite of programs (*147*); the ribbon diagrams were made using the Swiss-PDB viewer program and rendered using POVRAY™.

To avoid model-bias, residues 52-56 were omitted from the model during map calculation using the TB consortium Bias Removal Server (*148*). Clear electron density was seen for this region of the model showing that this is not an artifact (Figure 17). Gly54 is the pivotal residue in the formation of the domain-swapped dimer. The phi-psi torsion angles for Gly54 in the dimer are 81.8° and -163.1°, whereas Gly54 has phi/psi values of 52.5°/47.8°, 53.2°/50.2°, 60°/46.4° and 64.1°/56.3° in wild-type *Bst*HPr, the monomeric F29W variant, *Bs*HPr and *Sf*HPr, respectively. This change in phi-psi angle for Gly54 in the dimeric F29W mutant results in about a 180 degree rotation and swapping of residues 56 to 88, the C-terminal β strand and α helix of one monomer, for the corresponding elements in the other monomer. The dimer interface thus has extensive intermolecular interactions.

Significant structural changes occur at the dimer interface of the domain-swapped form when compared to the monomeric protein. Residues at or near the dimer interface (11-17, 38-40 and 48-52) show RMS deviations from the monomeric structures of 1.34 Å, 1.93 Å and 1.13 Å respectively. One region not close to the dimer interface that shows significant deviations is at the extreme C-terminus of the protein (residues 83-85), which shows an RMS deviation of 1.07 Å from the monomeric form.

The Cα of the active site residue His15 in the dimer is displaced 1.2 Å from its position in the monomeric structure and its imidazole ring is oriented away from the dimer interface to avoid a clash with Gly54. To accommodate the movement of His15 in the dimer, structural rearrangements have to be made to residues on either side of His15. The most significant of these is the relieving of the torsional strain of Ala16 seen in the *Sf*HPr (*141*), the wild-type and F29W *Bst*HPr monomeric structures. In these monomeric structures, Ala16 is found to be in a disallowed region of the Ramachandran plot, however

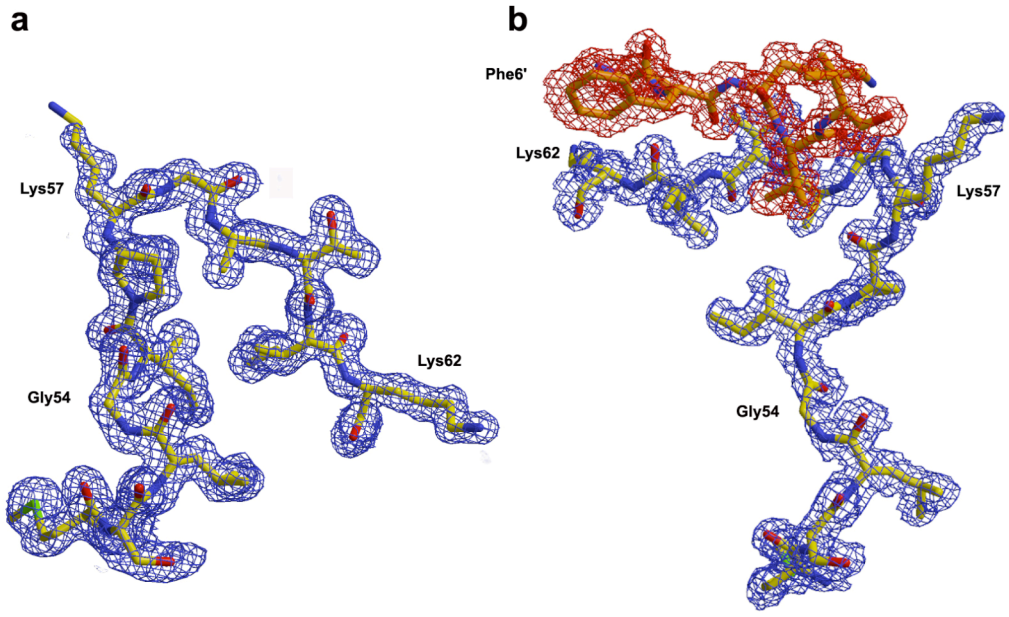


Figure 17: Electron density maps for sections of the Gly54-Lys62 region in the monomeric (panel a) and dimeric (panel b) forms of the F29W *Bst*HPr. For the dimer residues 52-56 were omitted during map calculation to avoid model bias. In panel b, the densities for the two chains of the dimer are shown in blue and red. The prime symbol on Phe6' is used to indicate that this residue is from the partner monomer chain. Density map representations were made using Xtalview (*149*) and rendered using Imagemagick™ software.

in the F29W *Bst*HPr dimer, the backbone of Ala16 adopts an α-helical conformation and forms part of helix A just as in *Bs*HPr. In the dimer, Ala16 moves towards the side chain of Met51 forcing the latter to move away from Ala16. This causes the B helix to be shorter in the dimer by one residue as the backbone of Leu53 is distorted from its helical conformation and now forms part of the loop connecting the C-helix and β strand 4.

A similar propagation of structural rearrangements can be seen in residues 36-40. If Tyr37 were to be in the same position in the dimer as it is in the monomer, its side-chain hydroxyl oxygen would be only 2.3 Å from the Cβ of Pro56' of the other monomer. Thus, in the dimer, Tyr37 moves so that this distance increases to 3.3 Å. There is also a movement of the β-turn comprising Asn38-Gly39-Lys40 towards the dimer interface. The side-chain of Asn38 is held in position by an H-bond between its N∂2 and backbone CO of Leu53. This rearrangement also moves the side chain of Lys40 to prevent a steric clash with the imidazole ring of His15' of the other monomer.

Characterization of solution oligomeric state of wild-type *Bst*HPr and the F29W variant

A number of biochemical techniques including analytical ultracentrifugation, diffusion measurements using NMR spectroscopy, gel filtration chromatography, native PAGE and CD spectroscopy were used to identify the solution oligomeric state of native and mutant *Bst*HPr. These experiments were performed in a number of different solution conditions including variations of temperature, pH, and ionic strength and protein concentration.

Diffusion measurements by pulsed-field gradient NMR

Pulsed-field gradient (PFG) NMR experiments on samples of wild-type *Bst*HPr and the F29W variant at protein concentrations of 1 and 2 mM and two salt concentrations (0.75 and 1.5 M NaCl) were conducted. As described in the methods section, the PG-SLED pulse sequence produces a series of 1D spectra, each recorded at a different gradient strength (Figure 18). The intensities of the signals(s) from these spectra are plotted against gradient strength (g) and fit by the equation:

$$s(g) = Ae^{-dg^2} \tag{6}$$

to obtain a decay rate (d) which is proportional to the diffusion coefficient D. Hydrodynamic radii (R_h) can be calculated from the diffusion coefficients using the Stokes-Einstein relationship; however we have used the method described by Wilkins et al. (*152*) where a small reference molecule with a known R_h (dioxane) is included in the samples. The decay rates are obtained for both protein ($d_{protein}$) and dioxane (d_{ref}) simultaneously and these are used to calculate R_h for protein using:

$$R_h^{protein} = \frac{d_{ref}}{d_{protein}}\left(R_h^{ref}\right) \tag{7}$$

where, $R_h^{protein}$ is the protein hydrodynamic radius and R_h^{ref} is the radius of the reference molecule. We have used R_h^{ref} for dioxane reported by Wilkins et al. (*152*) which is based on calibration experiments with lysozyme using radius of gyration values obtained from SAXS studies.

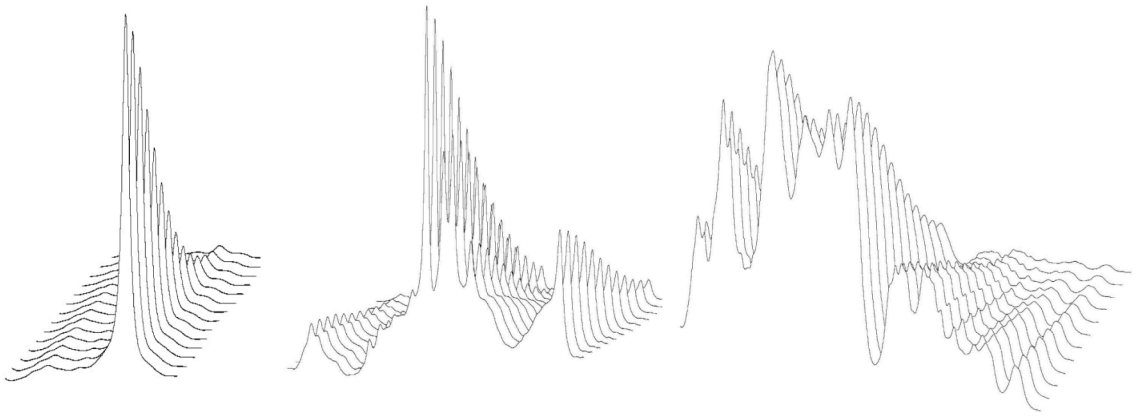


Figure 18: NMR spectra from the diffusion measurements by pulsed feild gradient NMR on the F29W variant of *Bst*HPr. The first set of peaks on the left shows the region of spectrum for the size reference dioxane (3.75 ppm) and only the first fourteen spectra of the twenty-four were plotted in this case. The other two sets of peaks are from the main aliphatic region of the spectra (3-0.4 ppm). While all spectra in the series were used in the analysis, only alternate spectra are plotted here for clarity. The slower decay rate of protein peaks compared to the reference molecule is evident.

It has been shown (*152*) that the measured hydrodynamic radii for folded globular proteins show a strong correlation with length of the polypeptide chain and the experimental data can be described by the empirical relationship:

$$R_h = \left(4.75 \pm 1.11\right) N^{0.29 \pm 0.02} \qquad (8)$$

where N is the number of residues in the protein. Figure 19 shows a plot of measured Rh values for our HPr proteins and some other proteins from the literature (*152*). The curve represents the expression shown in equation 8. It is clear that the dimensions obtained for the wild-type *Bst*HPr protein and the F29W variant with or without added salt are in close agreement with the values expected for proteins of this size and are significantly different from the values expected for a dimer with twice the number of amino acid residues.

Size exclusion chromatography

Size exclusion chromatography (SEC) was employed to determine the oligomeric state of the HPr proteins in solution. Figure 20 shows the elution profile of an equimolar mixture of wild-type *Bst*HPr and the F29W variant at a total concentration of 330 μM using the methods described below. Both proteins elute in a single symmetric peak indicating that they have a similar oligomeric state in solution. The inset to the figure shows the calibration curve for the column determined as described in the methods section (Chapter II). Under these experimental conditions, the apparent molecular weight estimated from the calibration curve for our samples is 10.9 KDa, very close to the expected molecular weight of the *Bst*HPr monomer (wild type = 9035 Da; F29W = 9068 Da) and not that of a dimer. There

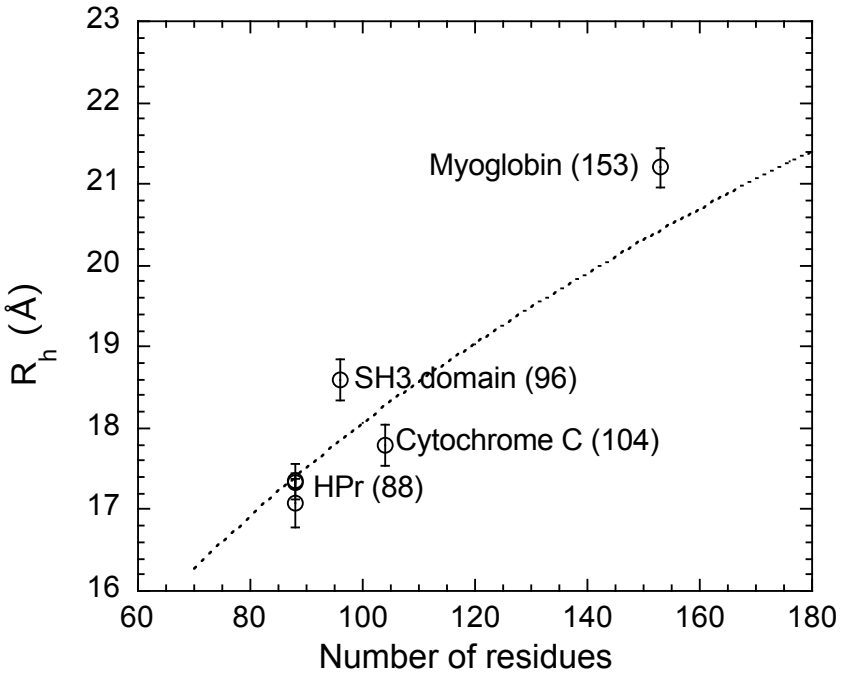


Figure 19: Radius of hydration (R_h) values measured by diffusion NMR experiments as a function of protein size. Experimentally obtained values are shown as symbols and numbers in parenthesis indicate number of residues in the said protein. R_h values for *Bst*HPr are from this study, while the other proteins shown for comparison are data collected by Wilkins et al. (*152*). The dotted line shows the empirical estimate of R_h based on equation (8).

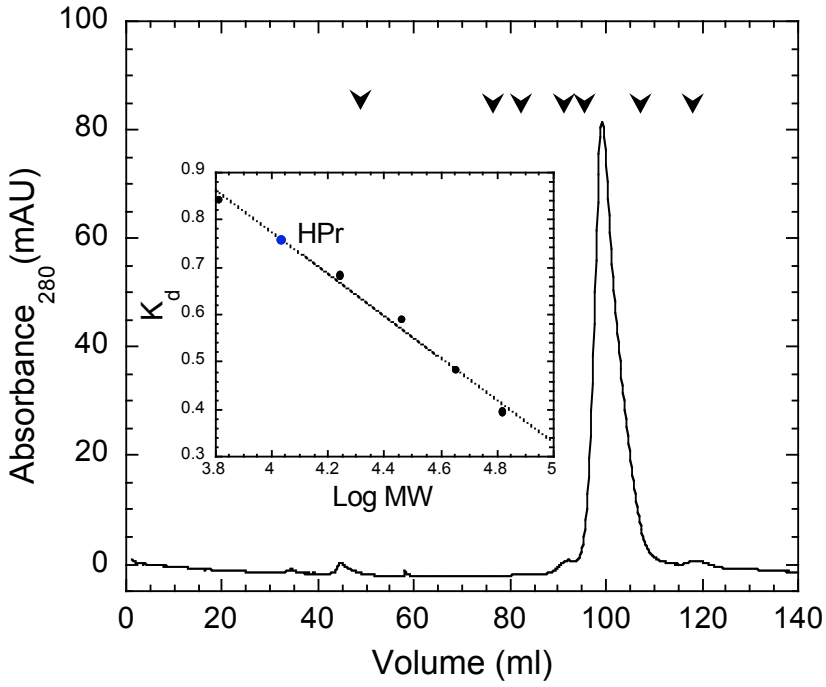


Figure 20: The elution profile for a mixture of F29W and wild type *Bst*HPr from a typical size exclusion chromatography experiment. V_e for the calibrants used are indicated by arrows (see Chapter II). The inset shows the calibration curve for our Superdex 200 column; the K_d for HPr elution is indicated. The calibrants used were blue dextran, BSA, ovalbulmin, carbonic anhydrase, myoglobin, aprotinin and acetone.

was no indication of protein eluting at the position for a dimer, based on the standard curve.

Analytical ultracentrifugation

Sedimentation equilibrium experiments using the analytical ultracentrifuge on the wild-type and F29W variant of *Bst*HPr were performed under a variety of solution conditions. In all cases, the data were fit well by a single ideal species model (Figure 21). There is no evidence of oligomerization in these experiments, the scans are fairly linear at all rotor speeds tested and show very little curvature if any at high rotor speeds. Attempts at fitting the data with multiple species models or single ideal species model with molecular weight fixed at dimer values yielded poor fits to the data with large non-random residuals (data not shown).

Thermodynamic stability measurements

The free energy of stabilization (ΔG) for the wild-type and F29W *Bst*HPr proteins were determined from the analysis of urea-induced unfolding reactions as monitored by circular dichroism spectroscopy. In all cases, the unfolding/refolding transitions show a single cooperative and reversible step, as observed for the mesophilic HPr proteins from *Bacillus subtilis* (*138*) and *Escherichia coli* (*139*). The values for ΔG obtained from the analyses are collected in Table 8 along with the corresponding m values.

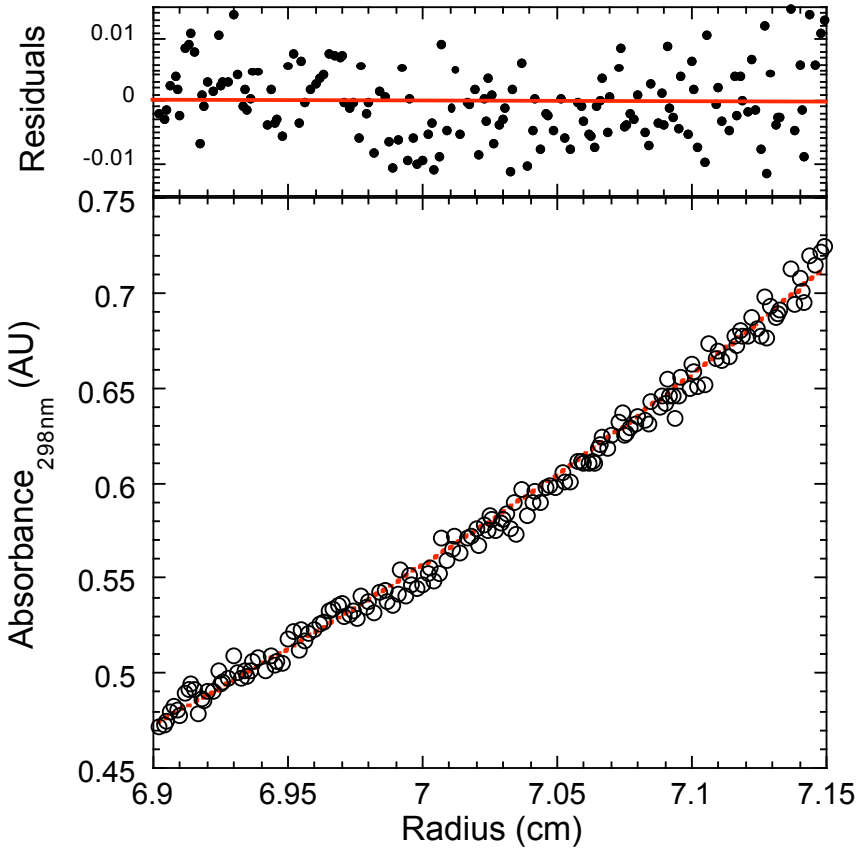


Figure 21: A typical equilibrium radial absorbance profile from analytical ultracentrifugation experiment on the F29W variant of *Bst*HPr. The oligomeric nature of the F29W variant of *Bst*HPr was probed in 10 mM NaPi, pH 7 at a final concentration of 500 μM at 25°C. The equilibrium radial absorbance profile measured at 15,000 rpm is shown as filled circles. The broken red line is a fit curve using an ideal single species model, which gave an estimated molecular weight of 9380±60 Da; residuals to the fit are shown above.

Table 8: Parameters characterizing urea and thermal denaturation experiments on *Bs*, *Bst* and F29W *Bst*HPr proteins.

Protein	Urea Denaturation[a]		Thermal Denaturation[b]
	ΔG (kcal mol^{-1})	m-value[c] (cal mol^{-1} M^{-1})	T_m (ºC)
Wild-type *Bst*HPr	8.23 (± 0.19)	1.0	88.3 (± 0.8)
***Bst*F29W**	7.37 (± 0.18)	1.1	83.7 (± 0.9)
Wild-type *Bs*HPr	5.00 (± 0.16)	1.1	73.0 (± 0.8)

[a] The urea denaturation experiments were performed at 25°C in 10 mM sodium phosphate buffer at pH 7 using techniques described earlier (*18*).

[b] Thermal denaturation experiments were performed in 10 mM sodium phosphate buffer at pH 7. Thermal denaturation experiments were also performed on the F29W variant as a function of protein concentration between 6 and 50 μM and in the presence of 750 mM NaCl; the T_m values obtained in the presence of NaCl were identical and in the range of 85.3 ± 0.8°C.

[c] Standard deviations for m value from urea denaturations were ≤ 5%

The tryptophan containing variant F29W is destabilized by roughly 1 kcal mol^{-1}, but the m values for the variant and the wild-type proteins are nearly identical. Furthermore, the m values are consistent with proteins the size of the HPr monomer and not a dimer *(24)*.

The proteins were also subjected to thermal denaturation experiments at different concentrations between 6 and 50 μM. As expected, the thermophilic protein has a higher melting temperature (T_m) than the mesophilic protein (88 and 73°C, respectively). Consistent with the results of the urea denaturation experiments, the F29W variant of the *Bst*HPr protein is less stable than the wild-type protein. In all cases, the proteins showed apparent two-state transitions with linear pre- and post-transition baselines. The T_m values were independent of protein concentration (Table 8) as expected for a monomeric protein. In addition, the pre-transition baselines at the different protein concentrations were very flat and did not show any sign of events depicting possible transitions from a dimeric form to monomer or unfolded state (data not shown).

With the availability of high resolution crystal structures for *Bs*HPr and *Bst*HPr, we performed a detailed comparison of the structural features in an attempt to explain the observed differences in conformational stability between the proteins. We analyzed the structures to determine the differences in the number of hydrogen bonds, salt bridges and change in accessible surface areas between these proteins. As seen in Table 9, there are no significant differences in any of the parameters. Therefore, unlike other comparisons of thermophilic and mesophilic proteins *(11, 37, 57, 126, 136, 137)*, we do not have any direct correlation with any of these structural features and the differences in stability between the HPr proteins studied here.

Table 9: A comparison of structural features between *Bs*, *Bst* and F29W variant of *Bst*HPr.

	*Bs*HPr	*Bst*HPr	F29W *Bst*HPr
Hydrogen Bonds[a]			
main chain to main chain	60	61	61
main chain to side chain	16	12	14
side chain to side chain	7	7	7
Change in accessible area upon folding[b] (ΔASA Å^2)			
Polar	2960 – 4415	3060 – 4425	3010 – 4375
Apolar	1315 – 2650	1325 – 2775	1360 – 2795
Salt Bridges[c] (ion pairs)	11	10	11

PDB accession codes for structures used are, 2HPR (*Bs*HPr), 1Y4Y (*Bst*HPr) and 1Y51 (F29W *Bst*HPr).

[a] The hydrogen bond calculations were performed using program *pfis* (*93*).

[b] The lower and upper estimates of ΔASA were estimated by subtracting the lower (or upper) estimate for unfolded surface area according to methods of Rose et al. (*153, 154*) from the surface area in the folded state calculated using program *pfis* (*93*).

[c] Salt bridges were determined using the WHAT IF suite of programs (*94*) available on the www at http://swift.cmbi.kun.nl/WIWWWI.

Discussion

We have successfully expressed and purified the HPr protein from the thermophilic organism *Bacillus stearothermophilus* and have solved its structure using x-ray crystallographic methods. We have also produced a single tryptophan variant and solved its structure as well. The wild-type *Bst*HPr protein has 70% sequence identity and 85% sequence similarity with the mesophilic *Bs*HPr protein (Figure 15), and therefore it is not too surprising that the three dimensional structures are nearly identical. Even in the region that shows the largest sequence divergence, the C-terminal α-helix, the structures are very similar (Figure 15).

The F29W variant crystallized in two different forms, one of which is a monomer like all other known HPr proteins and the second is a domain-swapped dimer. It is interesting to note that both the monomer and dimer forms of the F29W variant crystallized from similar conditions using ammonium sulfate as the precipitant. Although we could not detect any dimers in solution, the formation of crystals of the dimer suggests that the population of this form may have increased in the crystallization drop over time. The high concentration (20-30 mg/ml) of the protein used seems to be necessary as crystallization trials using a concentration of 15 mg/ml or less were unsuccessful in producing crystals.

The structure of the monomeric form of the F29W variant is very similar to all the other HPr proteins (Figure 15), while the dimeric form shows the classic features of a domain-swapped dimer. Domain swapping has been reported in proteins of diverse structure and function (for recent reviews, see Rousseau et al. (*155*) and Liu et al. (*150*)). In a recent review on domain swapping, Eisenberg's group (*150*) has suggested that domain

swapping can occur in any protein under appropriate conditions, the only prerequisite being that the termini of the protein are free. Also, as has been noted before (*155*), the monomeric and domain-swapped dimeric forms of proteins often have very similar structures outside of the hinge loop region that connects the two monomers.

Domain swapping has been proposed as a mechanism for amyloid fibril formation (*156, 157*) and two human proteins known to be amyloidogenic, the prion protein (*158*)and cystatin C (*159, 160*), crystallize as domain-swapped dimers. These two proteins also appear to be dimers in solution. Eisenberg's group has extensively discussed the role of domain swapping in amyloid formation and suggest that amyloid fibril formation might very well share common features with domain swapping (*150, 156, 161*). Our findings that an HPr protein can form a domain-swapped dimer is especially noteworthy as we recently reported that the HPr proteins from *Escherichia coli* and *Bacillus subtilis* can form amyloid fibrils (*162*). Furthermore, we can detect dimers, trimers and tetramers in the solutions where amyloid fibrils are forming (Schmittschmitt and Scholtz, unpublished data); we are in the process of characterizing the structures of these intermediates. Together these results suggest that domain swapping might play a critical role in amyloid formation for the HPr proteins.

A domain-swapped dimer has recently been observed in the crystal structure of an HPr homolog, Crh (*163*). Crh is found only in gram-positive bacteria and, like HPr, is known to interact with catabolite control protein A (CcpA). The authors of this study have modeled two HPr molecules from *Bacillus subtilis* on to the dimer and demonstrated an extended contacting surface between HPr monomers. Together with sequence comparisons of other bacterial homologues and the aforementioned modeling studies, they raise the interesting possibility that HPr itself maybe better suited to interact with CcpA as a dimer.

The structure of the Crh dimer involves a swap of the N-terminal β-strand between the two interacting monomers. Figure 22 compares the domain-swapped dimeric structures for Crh and the F29W variant of *Bst*HPr in panels a and b, respectively. The structure we report here for the *Bst*HPr dimer is more intertwined and involves swapping of 32 residues of the C-terminus β-strand 4 and the C-helix. The F29W domain-swapped dimer of *Bst*HPr involves many more residues than the domain-swapped Crh dimer, even though the overall structures of the monomers and the dimers are very similar.

In our structure of the domain-swapped dimer of the F29W variant of *Bst*HPr, there is an intermolecular salt-bridge between the side chains of Lys57 and Asp11' (Figure 16b). This interaction is an intramolecular one in the case of *Bs*HPr, and the monomeric *Bst*HPr structures, and was predicted to be important in the domain-swapped dimeric structure of the HPr homolog, Crh from *Bacillus subtilis* (*163*). In the case of Crh, the dimeric form is populated in solution and its formation is temperature sensitive, with low temperature favoring the dimeric form (*163*). Our structure of the domain-swapped F29W *Bst*HPr variant and the domain-swapped Crh structure reported earlier (*163*), represent another example of proteins with similar folds that show completely different types of domain swapping. In these cases, the swapped domains involve different elements of secondary structures at opposite ends of the polypeptide chain. To our knowledge, different type of domain swaps for the same protein were first demonstrated for RNase A, a protein long-known to form dimers, trimers and even tetramers (*164-166*). Eisenberg's group has extensively studied the RNase A system and has structurally characterized two different domain-swapped dimers, a domain-swapped trimer and have developed models for another trimer and the tetrameric form of domain-swapped oligomers (*150, 161, 167*). In the case

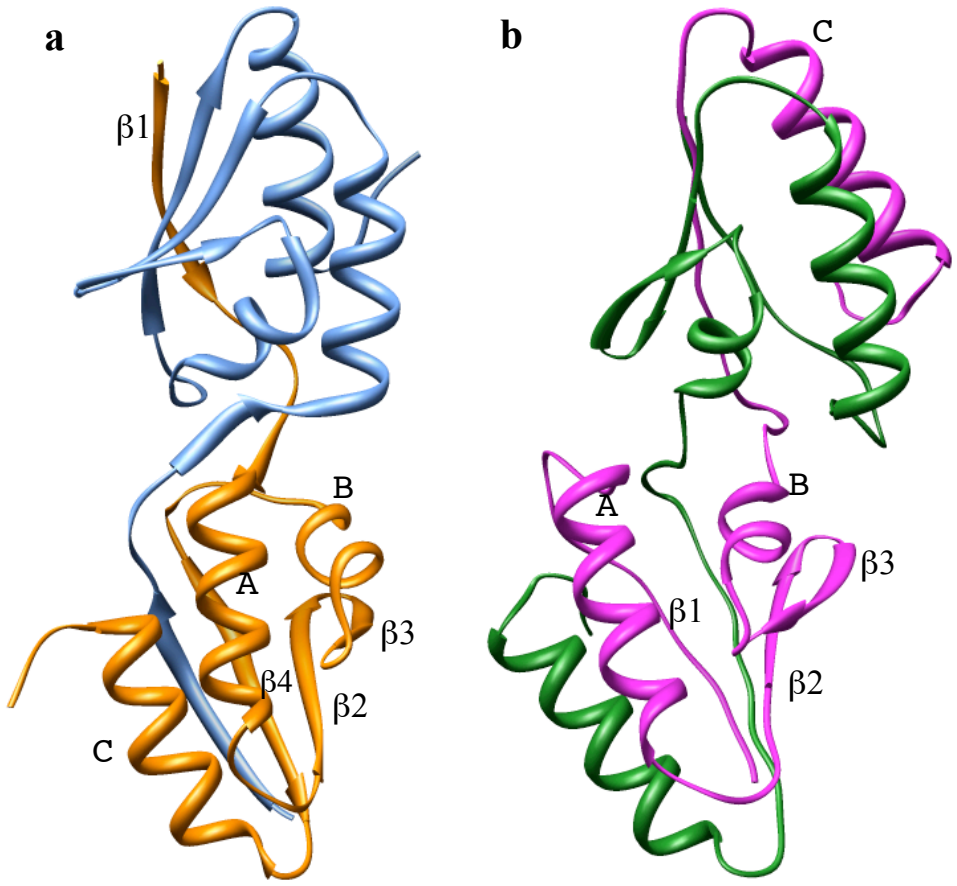


Figure 22: Ribbon diagram comparing the domain-swapped dimer structures of the Crh protein (panel a) and F29W variant of *Bst*HPr (panel b). PDB accession codes for the structures used: 1MU4 for Crh dimer and 1Y50 for the F29W *Bst*HPr protein. Alpha helices (A, B and C) and β-strands (β1, β2, β3 and β4) are labeled for one of the monomers. Ribbon representation and rendering was done using the Chimera software (*92*).

of the two dimers of RNase A, one involves the swap of the N-terminal helix and the other swaps the C-terminal β-strand. The HPr and Crh dimers are thus similar to the RNase A protein in that different regions of the polypeptide can be involved in domain-swapping for structurally related proteins.

We employed a number of techniques including analytical ultracentrifugation, diffusion measurements using NMR spectroscopy, gel filtration chromatography, native PAGE and CD spectroscopy to determine the hydrodynamic properties of the proteins and to determine if we could detect any dimeric structure in solution. Our studies indicate that all the HPr proteins, including the F29W variant of the *Bst*HPr protein, are monomeric in solution under all the conditions we explored. We cannot rule out the possibility that a small fraction of any of the proteins is dimerizing, but certainly the majority of the species in solution behaves as expected for a monomeric protein the size of HPr. Therefore, the domain-swapped does not appear to be stable in solution, in contrast to the dimer observed for the related Crh protein (*163*), even though the HPr dimer has a more extensive swap in terms of the number of residues compared to the Crh dimer. Further studies will be needed to determine which features of the two proteins are responsible for the different behaviors of these two related proteins.

Finally, we looked in detail at the high resolution structures of the monomeric HPr proteins to compare molecular interactions and properties in an attempt to understand the differences in stability between the different variants. First, mutations involving substitution of wild-type residues with a tryptophan are usually destabilizing, probably due to the large size of the aromatic side chain, which causes some changes in the local packing at the host site. A survey of tryptophan variants compiled in the Protherm database (*53*), and recent

data from our laboratory (*168*), reveals that of the 24 variants in which either tyrosine or phenylalanine residues are replaced by a tryptophan, only 3 are found to be stabilizing. Our data are consistent with this general trend. However, there are no apparent changes in the number of hydrogen bonds or ion pairs or the expected change in surface area upon folding that could explain the enhanced stability of the *Bst*HPr protein as has been suggested to be important in other systems (*11, 37, 57, 126, 136, 137*). Further work will be required to provide a molecular description of the changes in stability for the different HPr proteins.

CHAPTER V

DETERMINANTS OF THERMOSTABILITY IN *Bst*HPr

Introduction

The processes and factors involved in stabilizing the native conformations of proteins represent one focus of the field of protein folding. This question is complicated because it is difficult to discern contributions from individual interactions in the highly cooperative system that is the folded state of a protein. In spite of this inherent difficulty, great progress has been made toward understanding the role of forces involved keeping proteins folded in their native state. To a large extent, the success can be attributed to our ability to create and study protein variants using site-directed mutagenesis (*169-173*). In the case of studies on thermophilic proteins as well, numerous attempts at uncovering key interactions responsible for their ability to function at high temperatures have relied on directed mutagenesis and characterization of variants created (*49, 174-176*). These studies are aided by the availability of high-resolution structures of the proteins, which help researchers identify interactions and test their contributions by mutating the residues involved (*110, 177, 178*). In this report, a study of this nature is presented. We have used the high-resolution crystal structures to design variants of a pair of mesophilic and thermophilic homologous proteins whose thermodynamic stability features have been previously established. Our goal is to identify key interactions in the *Bst*HPr protein, which might be responsible for the enhanced thermostability of this protein.

Our model system is the homologous HPr proteins from mesophile *Bacillus subtilus* (*Bs*HPr) and thermophile *Bacillus staerothermophilus* (*Bst*HPr). These proteins are very amenable to folding studies due to their small size, lack of disulfide bonds or cofactor requirements and also because structures are available for both homologues (*72, 114*). The HPr homologues compared here show very high sequence identity, yet the mesophilic protein is significantly less stable than the thermophilic HPr. In previous publications, we have established the higher stability of *Bst*HPr and have also ruled out electrostatic interactions as a major contributing factor to this difference (see Chapter III for a characterization of the wild-type proteins). We have reported a detailed thermodynamic characterization of the wild-type proteins, including their 'stability curves' (*51*) and parameters like ΔC_p (change in heat capacity upon protein unfolding) derived from them. In this report, we attempt to find key interactions in *Bst*HPr, which are responsible for its enhanced thermostability by designing and characterizing variants of both the *Bs* and *Bst*HPr proteins. We also sought to confirm the similar ΔC_p values reported earlier for these homologues by attempting measurement of this cardinal thermodynamic parameter using calorimetry.

Results

Calorimetric estimation of ΔC_p

Differential scanning calorimetry (DSC) has been used traditionally to measure ΔC_p with the Kirchoff analysis providing a method to estimate ΔC_p from DSC thermograms. In this method, a series of DSC thermograms are obtained in the presence of a stability

perturbant like a chaotrope, salt or variation in pH and the T_m and ΔH obtained from these different thermograms are determined and a linear fit to these data thus provides a slope, which is the ΔC_p (*179*).

Modulating the ΔC_p value of a protein has been established as a very effective way to achieve a higher T_m in many proteins from thermophilic organisms (see Chapter I for a discussion on this topic); hence a thorough study of this parameter is warranted for our set of homologous HPr proteins. As mentioned earlier, we have reported ΔC_p values derived from stability curves for these homologues and found them to be very similar. However, we wanted another independent estimate of ΔC_p for these proteins so we could rule out with confidence a difference in this value for the two proteins.

We have obtained DSC thermograms for *Bst*HPr at various solution pH ranging from 1.5 through 7 and the parameters obtained are plotted in Figure 23. The data obtained span a wide temperature range and give a reasonably good linear fit and the slope, ΔC_p, is 1.5 kcal mol^{-1} K^{-1}, which is same as the value estimated from stability curves (see Table 10 for a compilation of these values). The ΔC_p for *Bs*HPr could not be measured calorimetrically because of aggregation problems at high temperatures under most of the solution conditions tested.

Sequence comparisons and design of mutants

The HPr proteins from the mesophile *Bacillus subtilus* (*Bs*) and thermophile *Bacillus staerothermophilus* (*Bst*) each have 88 amino acids and show a very high level of sequence identity (72%). A sequence alignment for these proteins generated using CLUSTALW (*84*)

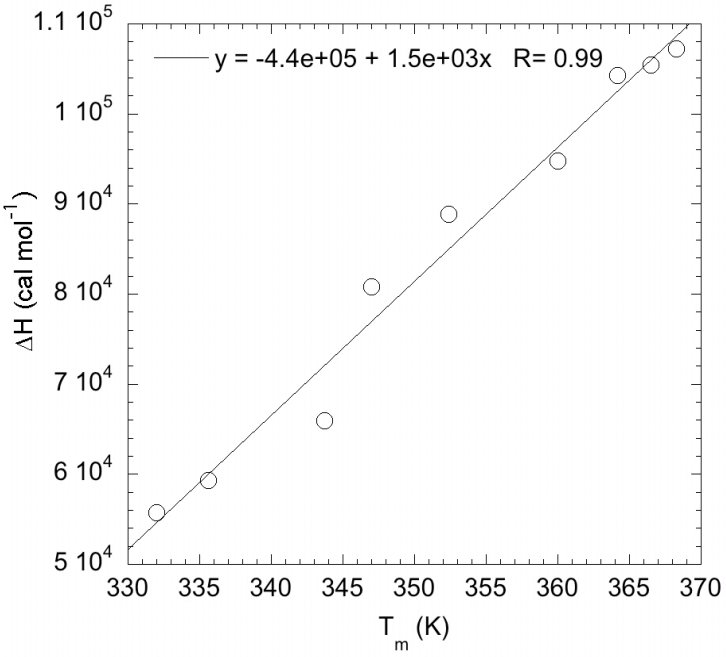


Figure 23: Estimation of ΔC_p for *Bst*HPr by a Kirchoff analysis of ΔH vs. T_m. The data were compiled from DSC thermograms performed at pH ranging from 1.5 through 7. A linear fit to the data shows a high correlation coefficient and slope of the fit (ΔC_p) is in close agreement with value estimated by the Pace-Laurents (*100*) method (see Figure 11 and Table 10).

Table 10: Parameters characterizing urea, GuHCl and thermal denaturations and ΔC_p estimates for *Bs* and *Bst*HPr at pH 7.

	Urea denaturations		GuHCl denaturations		$\Delta\Delta G_{urea-GuHCl}$	Thermal denaturations		
Protein	**ΔG** (kcal mol^{-1})	**m-value**[a] (kcal mol^{-1} M^{-1})	**ΔG** (kcal mol^{-1})	**m-value**[b] (kcal mol^{-1} M^{-1})	(kcal mol^{-1})	**T_m** (ºC)	**ΔH**[c] (kcal mol^{-1})	**ΔC_p**[d] (kcal mol^{-1}K^{-1})
***Bst*HPr**	8.20 (±0.20)	1.0	6.75 (±0.05)	2.4	1.4	88.3 (±0.8)	98.6	1.34 (1.5)
***Bs*HPr**	5.00 (±0.16)	1.1	3.41 (±0.04)	3.3	1.6	73.0 (±0.6)	76.7	1.33

All experiments were performed in 10 mM NaPi and the solvent denaturation experiments were performed at 25ºC. All values reported are averages of results from multiple experiments. The values in parenthesis are standard deviations of results of multiple experiments.

[a] Standard deviations for m value from urea denaturations were ≤ 5%

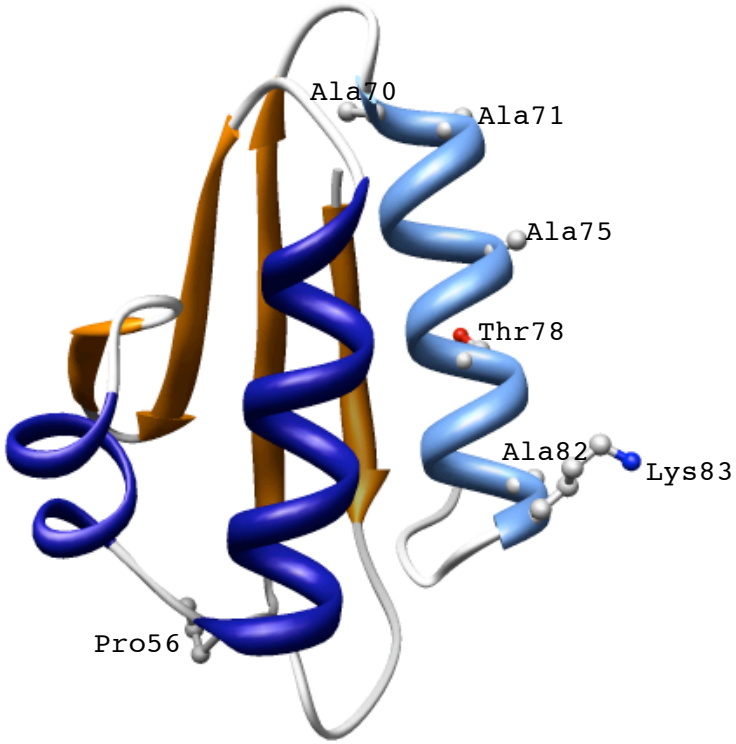
[b] Standard deviations for m value from GuHCl denaturations were ≤5%

[c] Standard deviations in ΔH were usually ≤ 5% of the average value reported.

[d] ΔC_p value estimated from best-fit parameters to stability curve data are reported here, for *Bst*HPr the value in parenthesis is ΔC_p estimated by Kirchoff analysis of DSC data. The error associated with ΔC_p is usually ≤ 10%

is shown in Figure 24; this alignment shows that the sequences are identical for the most parts but for two regions. The first region is a cluster of conservative replacements in residues 29-34 and a cluster of mostly non-conserved replacements in the region of residues 70-83. The latter region of the sequence corresponds to the last element of secondary structure in the protein referred to as the C helix. The replacements that occur in the *Bst*HPr C helix are mostly helix favoring and occur on the solvent exposed side. Thus, in terms of helix propensity scales, it can be posited that the C helix from *Bst*HPr fares better than that from *Bs*HPr. This conjecture is also supported by predictions from the AGADIR algorithm (*121*) (data not shown). The other residue that stands out in sequence alignments between *Bs*, *Bst* and other HPr proteins (from *Streptococcus thermophilus St, Bacillus halodurans Bh,* not shown) is residue 56. Most HPr homologues have either a glycine or alanine at position 56 but *Bst*HPr is peculiar in having a proline. Residue 56 lies on a surface exposed loop (see Figure 24) connecting the B helix to β-strand four. It has been shown that prolines are preferred in loop regions because they reduce conformational strain, probably adding to conformational stability of the protein (*120, 180*).

We have used these insights from the sequence analysis to create variants of *Bs* and *Bst*HPr in an attempt to tease out the determinants of higher stability shown by protein from the thermophile. The aim of the mutagenesis experiments was to find stabilizing or destabilizing mutations in either wild-type protein, which would point toward specific stabilizing interactions that occur in *Bst*HPr. Since all the mutations we made for this study involve the C helix of both homologues either independently or in conjunction with mutations to residue 56, we propose a nomenclature scheme that will be brief and will best describe these variants:




```
Bs  MAQKTFKVTADSGIHARPATVLVQTASKYDADVNLEYNGKTVNLKSIMGVMSLGIAKGAE 60
Bst MAEKTFKVVSDSGIHARPATILVQTASKFNSEIQLEYNGKTVNLKSIMGVMSLGIPKGAT 60
    **:*****.:**********:********:::::**********************.***
Bs  ITISASGADENDALNALEETMKSEGLGE 88
Bst IKITAEGADAAEAMAALTDTLAKEGLAE 88
    *.*:*.***  :*: ** :*: .***.*
```

Figure 24: Ribbon diagram of the *Bst*HPr showing the differences in sequence with respect to *Bs*HPr. The C helix is colored light blue and residues shown in ball and stick are sites of non-conservative replacements; also shown in ball and stick is proline 56. The PDB file 1Y4Y was used to render this image using the UCSF Chimera package (*92*). Also shown is the sequence alignment for *Bs* and *Bst*HPr and the C helix sequences have been underlined.

$$C_{\alpha}^{Bst}\,Bs^{Ala56} \qquad (9)$$

where, C_α indicates the secondary structure element in question (for example, C_α for the C α-helix or 1_β for β-strand 1), with the superscript indicating the protein from which it is derived (*Bs*, *Bst* or hyb for a hybrid of the two proteins), followed by the protein homologue where these mutations are made and finally the superscript to this term indicating the identity of the residue at position 56.

The C helix was of the greatest interest in our mutational design since this region showed highest sequence diversity between the two proteins. We started by mutating all the sites of non-conservative replacements in the C helix of *Bs*HPr to those found in *Bst*HPr, the resulting variant being $C_\alpha^{hyb}Bs^{Ala56}$. In the next step, we replaced the remaining residues in the C helix of *Bs*HPr with residues from *Bst*HPr; the resulting variant was $C_\alpha^{Bst}Bs^{Ala56}$. A complementary set of mutants was also made in the *Bst*HPr protein; all divergent residues in the C helix (70-83) of this protein were mutated to their *Bs* counterparts to give a chimera we call $C_\alpha^{Bs}Bst^{Pro56}$. These mutants should help ascertain if the C helix of *Bst*HPr is necessary and sufficient to confer the protein's enhanced stability.

Another set of mutants investigates the role of proline in the surface exposed loop connecting the B helix to β-strand four. The wild-type alanine in *Bs*HPr was replaced with a proline (*Bs*A56P); this mutation was also made in the $C_\alpha^{hyb}Bs^{Ala56}$ and $C_\alpha^{Bst}Bs^{Ala56}$ backgrounds. In addition, the wild-type proline in *Bst*HPr was mutated to a glycine and an alanine in the wild-type protein and to an alanine in the $C_\alpha^{Bs}Bst^{Ala56}$ background.

We present below results for each mutant protein obtained from the three denaturation (urea, GuHCl and thermal) methods used; the results are also compiled in

Table 11. As representative examples we have shown the results from thermal and solvent denaturation experiments on the wild-type proteins in Figure 25. Also shown in this figure are results for the most stabilized *Bs*HPr variant ($C_\alpha^{hyb}Bs^{Pro56}$) and the most destabilized variant of *Bst*HPr ($C_\alpha^{Bs}Bst^{Ala56}$). We will also compare the $\Delta\Delta G_{urea\text{-}GuHCl}$ value (difference in ΔG estimated by urea and GuHCl denaturation experiments) for a variant with that of the respective wild-type protein to estimate the role of electrostatic contributions to stability.

C helix variants of *Bs* and *Bst*HPr

$C_\alpha^{hyb}Bs^{Ala56}$: This *Bs*HPr variant has mutations in its C helix to *Bst* residues at sites of non-conservative replacements: residues 70, 71, 75, 78, 82 and 83. These residues have also been shown in ball and stick in the ribbon diagram for *Bst*HPr in Figure 24. Results of urea denaturation experiments show that the protein is as stable as the wild-type *Bs*HPr protein. The thermal denaturation experiments returned a T_m of 73.6ºC and ΔH of 92 kcal mol^{-1}. The ΔH is significantly higher than the wild type protein. The $\Delta\Delta G_{urea\text{-}GuHCl}$ was found to be 1.1 kcal mol^{-1}; this is lower compared to the wild-type value, which is 1.6 kcal mol^{-1}.

$C_\alpha^{Bst}Bs^{Ala56}$: This variant of *Bs*HPr is essentially a chimera in which the *Bst* C helix is present in the *Bs*HPr background. This protein was found to be less stable than the wild-type protein with a ΔG of 4.3 kcal mol^{-1} from urea denaturation experiments and a significantly higher m value of 1.2 kcal mol^{-1} M^{-1}.

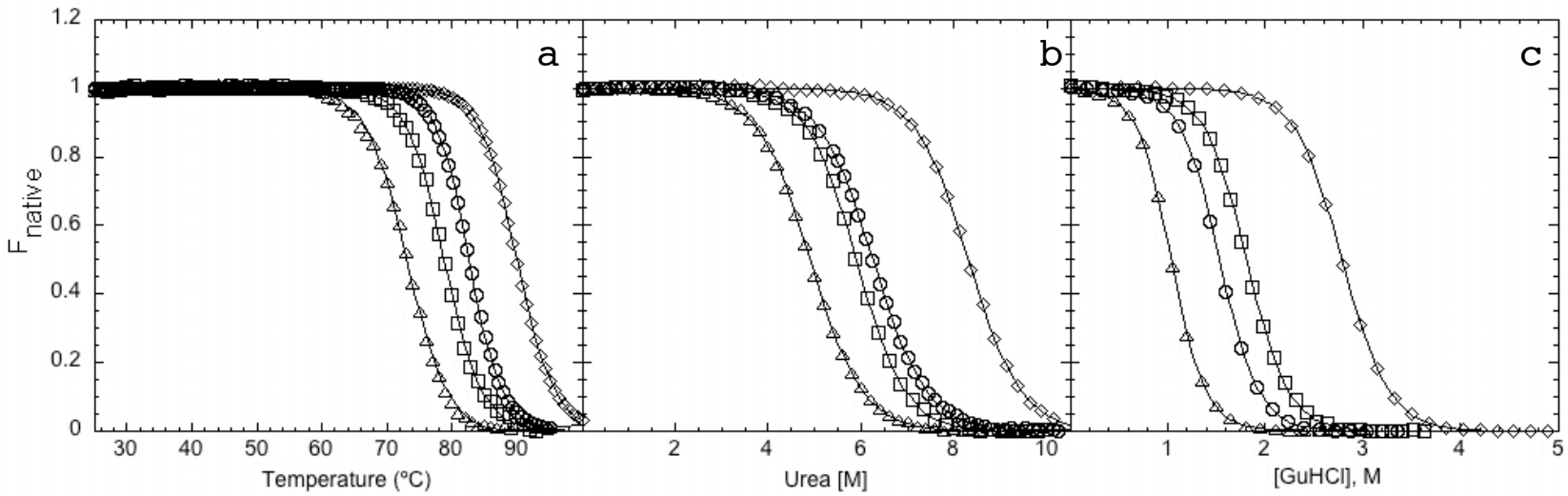


Figure 25: Illustrative results from thermal, urea and GuHCl denaturation experiments on the wild-type *Bs* and *Bst*HPr proteins and their most stabilized and destabilized variants. Panel a shows results from thermal denaturation experiments on *Bs* (△), $C_\alpha^{hyb}Bs^{Pro56}$ (□), $C_\alpha^{Bs}Bst^{Ala56}$ (○) and *Bst*HPr (◇) proteins (for a description of nomenclature methodology see equation 9). Similarly panels b and c show data from urea and GuHCl denaturation experiments respectively; symbols used follow the convention used in panel a. All experiments were carried out in 10 mM sodium phosphate at pH 7, solvent denaturation experiments were performed at 25ºC. Curves through the data are fits of equation 2 in panel a and equation 3 in panels b and c to the data.

Table 11: Compilation of stability data from urea, GuHCl and thermal denaturation experiments on variants of *Bs* and *Bst*HPr.

	Urea denaturations		GuHCl denaturations				Thermal denaturations	
Protein	**ΔG** (kcal mol^{-1})	**m-value**[a] (kcal mol^{-1}M^{-1})	**ΔG** (kcal mol^{-1})	**m-value**[b] (kcal mol^{-1}M^{-1})	**$\Delta\Delta G_{urea-GuHCl}$** (kcal mol^{-1})	**ΔG_{WT}-ΔG_{mut}**[c] (kcal mol^{-1})	**T_m** (ºC)	**ΔH** (kcal mol^{-1})
Wild Type *Bst*	8.2 (±0.2)	1.0	6.8	2.4	1.4	0.0	88.3 (±0.8)	97.4 (±4.2)
***Bst* P56G**	7.5 (±0.2)	1.0	5.60 (±0.16)	2.60 (±0.03)	1.9	-0.8	84.3 (±0.3)	98.5 (±6.7)
***Bst* P56A**	7.50 (±0.12)	1.0	5.60 (±0.24)	2.40 (±0.06)	1.9	-0.8	83.6 (±1.0)	94.8 (±2.6)
C_α^{Bs}*Bst*^{Pro56}	7.0 (±0.11)	0.9	5.5 (±0.2)	2.50 (±0.09)	1.5	-1.2	92.1 (±1.1)	83.1 (±2.4)
C_α^{Bs}*Bst*^{Ala56}	6.40 (±0.04)	1.0	4.5 (±0.1)	2.9 (±0.1)	1.9	-1.8	82.9 (±0.4)	102.1 (±4.2)
Wild Type *Bs*	5.00 (±0.16)	1.1	3.50 (±0.07)	3.30 (±0.04)	1.5	0.0	73.0	77.6 (±0.9)
***Bs* A56P**	5.60 (±0.29)	1.0	4.20 (±0.06)	3.10 (±0.02)	1.4	0.6	78.2 (±1.6)	87.8 (±2.3)
C_α^{hyb}*Bs*^{Ala56}	5.10 (±0.16)	1.1	3.90 (±0.14)	2.90 (±0.04)	1.2	0.1	73.3 (±0.4)	88.6 (±4.8)
C_α^{hyb}*Bs*^{Pro56}	6.20 (±0.32)	1.1	4.90 (±0.04)	2.70 (±0.04)	1.3	1.2	78.7 (±0.0)	89.0 (±5.5)
C_α^{Bst}*Bs*^{Ala56}	4.30 (±0.29)	1.2	2.90 (±0.03)	3.00 (±0.08)	1.4	-0.7	67.3 (±0.3)*	
C_α^{Bst}*Bs*^{Pro56}	5.10 (±0.05)	1.1	4.5 (±0.3)	2.90 (±0.06)	0.6	0.2	74.0 (±1.1)	74.0 (±5.5)

Table 11: Continued

All experiments were done in 10 mM sodium phosphate buffer at pH 7; solvent denaturation experiments were performed at 25ºC. All values reported are averages of results from multiple experiments and the values in parenthesies are the measured standard deviations.

[a] Standard deviations for m value from urea denaturations ranged between 0.005 and 0.04 kcal mol^{-1} M^{-1}

[b] Standard deviations for m value from GuHCl denaturations ranged between 0.03 and 0.1 kcal mol^{-1} M^{-1}

[c] The difference in ΔG obtained from urea denaturations for mutant from the respective wild type protein (*Bs* or *Bst*HPr)

*Thermal denaturation for this variant was not reversible due to aggregation problems and we were unable to measure the ΔH for this variant, thus the T_m reported is an apparent value.

The apparent T_m for this protein was 67.5ºC and the ΔH could not be determined due to aggregation problems. The difference in stabilities from urea and GuHCl denaturation experiments $\Delta\Delta G_{urea\text{-}GuHCl}$ was found to be very similar to the wild-type protein at 1.5 kcal mol^{-1}.

$C_\alpha{}^{Bs}BstP^{ro56}$: This variant of *Bst*HPr is complementary to the protein discussed above; it is a chimera of *Bs*HPr C helix in *Bst* background. It was found that this protein is significantly less stable than the wild-type *Bst*HPr protein with a ΔG of 7.0 kcal mol^{-1} from urea denaturation experiments. The m value was slightly lower than the wild-type protein at 0.9 kcal $mol^{-1}M^{-1}$. Thermal denaturation experiments on this variant revealed that surprisingly this protein had a higher T_m than the wild-type protein at 91.5ºC although the ΔH was significantly lower at 82 kcal mol^{-1}. The $\Delta\Delta G_{urea\text{-}GuHCl}$ for this protein was found to be slightly higher than the wild-type protein at 1.7 kcal mol^{-1}.

Residue 56 variants of *Bs* and *Bst*HPr

*Bs*P56A: This variant of *Bs*HPr replaces the wild-type alanine at residue 56 with a proline. This mutation stabilized the protein by 0.6 kcal mol^{-1}. The m value of this mutant was slightly lower than the wild-type protein at 1.0 kcal $mol^{-1}M^{-1}$, which is the m value of *Bst*HPr. Thermal denaturation of this protein revealed a slightly higher T_m at 77ºC, and a significantly elevated ΔH at 89 kcal mol^{-1} compared to the wild-type protein. The $\Delta\Delta G_{urea\text{-}GuHCl}$ for this protein was 1.3 kcal mol^{-1}, which is slightly lower than that for the wild-type protein.

$C_\alpha{}^{hyb}Bs^{Pro56}$: This variant of *Bs*HPr replaces the alanine in the $C_\alpha{}^{hyb}Bs^{Ala56}$ variant with a proline at position 56. The addition of proline to the $C_\alpha{}^{hyb}Bs^{Ala56}$ variant appears to

significantly stabilize the protein as revealed by all three denaturation methods; in fact this is the most stabilized variant of *Bs*HPr we created in this study. The results from denaturation experiments are shown in Figure 25 for this protein. The urea denaturation experiments returned a ΔG of 6.2 kcal mol^{-1} with an m value close to that of the wild-type protein. Thermal denaturation experiments returned a significantly higher T_m of 78.7ºC and a ΔH of 89 kcal mol^{-1}. However, when compared to the $C_\alpha^{hyb}Bs^{Ala56}$ variant alone the T_m is higher with a very similar ΔH. The $\Delta\Delta G_{urea\text{-}GuHCl}$ for this variant is 1.3 kcal mol^{-1}, which is close to the wild-type protein value.

$C_\alpha^{Bst}Bs^{Pro56}$: This variant of *Bs*HPr adds a proline to the $C_\alpha^{Bst}Bs^{Ala56}$ chimera. The results of urea denaturation experiments on this variant reveal a ΔG of 5.1 kcal mol^{-1} and an m value of 1.1 kcal mol^{-1} M^{-1}, both of which are close to those for the *Bs*HPr protein. The T_m for this protein is 74.8ºC, slightly higher than the wild-type protein, but the ΔH remains lower than the wild-type protein at 70 kcal mol^{-1}. The $\Delta\Delta G_{urea\text{-}GuHCl}$ value for this variant is only 0.4 kcal mol^{-1}, this represents the most drastic difference from the value for the wild-type protein (1.6 kcal mol^{-1}) for any variant in this study.

*Bst*P56G and *Bst*P56A: these variants of *Bst*HPr introduce either a glycine or alanine at residue 56. The ΔG determined by urea denaturation experiments is very similar for these two variants at 7.5 kcal mol^{-1}. The m values are also very similar at 1.02 and 1.05 kcal mol^{-1} M^{-1}, for *Bst*P56G and *Bst*P56A respectively, which in turn is very close to the value for the wild-type protein. Thermal denaturation experiments returned a T_m of 84.3 and 83.6ºC for the P56G and P56A variant respectively. The ΔH for *Bst*P56G was 98.5 kcal mol^{-1} and for the P56A variant the ΔH was 94.8 kcal mol^{-1}. The $\Delta\Delta G_{urea\text{-}GuHCl}$ values for these variants are 1.9 kcal mol^{-1}.

$C_\alpha^{Bs}Bst^{Ala56}$: This variant of *Bst*HPr introduces an alanine at position 56 of the C helix swap chimera $C_\alpha^{Bs}Bst^{Pro56}$. It appears that the replacement of the proline at position 56 with an alanine significantly destabilizes the protein as revealed by all three denaturation methods; in fact this is the most destabilized variant of *Bst*HPr. The data from thermal and solvent denaturation experiments on this protein are included in Figure 25. The urea denaturation experiments return a ΔG of 6.4 kcal mol^{-1} and an m value of 1.03 kcal mol^{-1} M^{-1}. The T_m determined by thermal denaturation experiments is 83ºC with a ΔH of 102 kcal mol^{-1}. The $\Delta\Delta G_{urea\text{-}GuHCl}$ value for this variant is slightly higher than *Bst*HPr at 1.9 kcal mol^{-1}.

Discussion

We are interested in determining the molecular basis for the enhanced thermostability of the HPr protein from the thermophile *Bacillus staerothermophilus*. Several hypotheses have been proposed to explain the enhanced thermostability of proteins, these include improved hydrogen bonding, improved electrostatic interactions, improved core packing, tighter surface loops etc. We have tried to systematically investigate as many of these hypotheses by using the HPr proteins from *Bs* and *Bst*, in this study and previous work on these homologues.

Structural comparisons between pairs of proteins derived from mesophiles and thermophiles have proven to be indispensable in comparative studies of the kind we have undertaken here (*177, 181, 182*). Structural comparisons point toward stabilizing interactions, which might be present specifically in the thermophilic homologue. A structural comparison for the two HPr homologues in this study was reported earlier (Table 9 in Chapter IV). It was found that the two proteins have very similar structures with very

similar number of hydrogen bonds and salt-bridge interactions, within errors of estimation and taking into account the resolution of the structures.

High-resolution crystal structures also allow the estimation of theoretical values of ΔC_p. We have used the available structures to calculate ΔC_p using the methods of Myers et al. (*24*), Spolar et al. (*22*)and Murphy et al. (*23*). The ΔC_p for *Bs* and *Bst*HPr was estimated by the three methods listed above, using the lower and upper estimates of ΔASA (change in solvent accessible surface area listed in Table 9 of Chapter IV) according to methods of Creamer et al. (*153, 154*). The theoretical estimates of ΔC_p bracket the experimental value within errors of the estimate. However, the measured ΔC_p for *Bst*HPr is consistently close to the upper bound estimates of all three methods, unlike the *Bs*HPr value which is close to the average of upper and lower estimates. As mentioned earlier, the upper and lower bounds of the estimated ΔC_p are derived from low and high estimates of ΔASA. These estimates are based on different models for the unfolded state. The lower estimate assumes a more compact denatured state composed of peptide fragments derived from native protein states, while the upper estimate assumes a more expanded native state. A more detailed description of the methods is available in publications from George Rose's group (*153, 154*). From this observation, it can be proposed that the unfolded state for *Bst*HPr is more expanded than it is for *Bs*HPr. However, it can also be argued that this difference in ΔC_p originates from differences in ΔASA caused by differences in the native state. More insights into this problem can be obtained by experimental methods involving more detailed calorimetric measurements of native and denatured state C_p and also by theoretical methods that measure the extent of packing in the native states of proteins. The role of core packing in enhancing thermostability of proteins has been debated in literature (*183-185*);

based on computational methods involving a measure of voids and cavity volumes, authors have found evidence for and against this conjecture. Experiments to measure the native and denatured state C_p were successful for *Bst*HPr only (data not shown); the *Bs*HPr protein was not amenable to calorimetric analyses under the conditions tested. These experimental difficulties with *Bs*HPr thus preclude any meaningful conclusions on the extent of packing in the native and denatured states of the two HPr homologues.

To broadly compare contributions from electrostatic interactions to protein stability we have used the $\Delta\Delta G_{urea\text{-}GuHCl}$ value; Kohn et al. (*115*) have used this parameter in an earlier study for similar purposes. A difference in $\Delta\Delta G_{urea\text{-}GuHCl}$ for a variant from the respective wild-type protein would point to possible differences in electrostatic forces acting in that variant. Most variants studied have $\Delta\Delta G_{urea\text{-}GuHCl}$ values similar to their respective wild-type proteins. The few variants with significantly different $\Delta\Delta G_{urea\text{-}GuHCl}$ ($C_\alpha{}^{Bst}Bs^{Pro56}$, for example), fail to provide useful insights into the underlying aim of this study and hence have not been pursued with more rigorous analyses.

Since the HPr proteins show a high level of sequence identity (72%), we have exploited the limited diversity to create a number of variants of both HPr proteins. We hoped to create a variant of either *Bs* or *Bst*HPr, which would add/lose critical interactions that result in a gain/loss of 3.2 kcal mol^{-1} (the difference in free energy of stabilization between *Bst* and *Bs*HPr) of free energy of stabilization. We started with mutations in the C helix of the protein, which seemed promising, since most of the non-conserved replacements occurred here. Accordingly, swap chimeras of C helix in *Bs* and *Bst*HPr background were made and characterized thermodynamically.

The results of this characterization show that the C helix alone cannot confer the enhanced thermostability of *Bst*HPr upon *Bs*HPr. Moreover, the C helix swap in *Bs* background actually destabilized the protein. The C helix swap chimera in the *Bst*HPr background was destabilized by 1.2 kcal mol^{-1}. However, this variant had a higher T_m and lower ΔH than *Bst*HPr itself. The ΔC_p for this variant was measured calorimetrically (data not shown) and it was found to be lower (1.03 kcal $mol^{-1}K^{-1}$) than *Bst*HPr in magnitude. Given the established correlations between m value or ΔC_p and protein size (*24*), the difference in ΔC_p/m values among HPr homologues and their variants seems significant given the similar size and structure of these proteins. Hence, we propose that the C helix does play a role in the compactness of the native or denatured states in *Bst*HPr. However, a complementary effect in the *Bs*HPr swap chimera was not observed in terms of a ΔC_p due to experimental difficulties, and the m value change was marginal. Therefore we cannot draw any general conclusions regarding contributions of the C helix to the packing of HPr homologues. In the *Bs*HPr background two variants were created with mutations in the C helix: $C_\alpha^{hyb}Bs^{Ala56}$ and $C_\alpha^{Bst}Bs^{Ala56}$. The former was a more conservative variant where only the non-conserved changes in the C helix were mutated to residues in *Bst*HPr (see Figure 24 for the residues mutated). The $C_\alpha^{hyb}Bs^{Ala56}$ variant of *Bs*HPr was found to be as stable as *Bs*HPr, albeit with a higher ΔH. The $C_\alpha^{Bst}Bs^{Ala56}$ variant was found to be less stable than the wild-type protein as mentioned earlier. The m value was slightly higher than the wild-type value and the ΔH could not be measured due to aggregation problems.

Another set of variants involves investigation into the role of residue 56. In general, the contribution of a proline at residue 56 appears to be 0.7 (±0.1) kcal mol^{-1}. The replacement of the native proline in *Bst*HPr by a glycine or alanine destabilizes the protein

by that amount and replacement of native alanine by a proline in *Bs*HPr stabilizes the protein to the same degree. This trend was also expected in the case of residue 56 variants in the helix swap chimera background, and was found to be true in most cases except for the $C_{\alpha}^{hyb}Bs^{Ala56}$ background where the alanine to proline mutation adds ~1.1 kcal mol^{-1} to the stability of $C_{\alpha}^{hyb}Bs^{Ala56}$ making it the most stabilized variant of *Bs*HPr created in this study. In the $C_{\alpha}^{Bst}Bs^{Ala56}$ background, however, the A56P mutation adds only 0.8 kcal mol^{-1}, which is the expected contribution. Similarly, in the $C_{\alpha}^{Bs}Bst^{Ala56}$ variant the replacement of the wild-type proline reduces the stability by 0.6 kcal mol^{-1}, making this the most destabilized variant of *Bst*HPr created in this study.

The results presented in this report support a role for the C helix in the stability of the *Bst*HPr protein; however, these contributions appear to be context dependent since placing this C helix in *Bs*HPr does not produce a stabilized version of the mesophilic protein. This conjecture is supported by the fact that the C helix swap chimera in the *Bs*HPr background is significantly destabilized, whereas a more conservative swap of C helix residues is tolerated without causing changes in stability. The other focus of our study was the role of residue 56 in the stability of these proteins. We found that for the most part a proline at position 56 contributes independently to the stability except for *Bs*HPr variant with the partial C helix swap. In this variant, it appears that the presence of proline at position 56 acts synergistically with the C helix mutations in stabilizing the *Bs*HPr protein. We have also investigated the role of proline at position 56 with fluorescence experiments studying the changes in fluorescence emission properties caused by this mutation in a variety of protein backgrounds (Razvi and Scholtz unpublished results). We found that this residue has a pronounced effect on fluorescence properties of the proteins, alluding to a

role for this residue in packing of HPr proteins in general or the neighborhood of the chromophore (a tryptophan introduced at position 37) at the least.

We have established that the C helix plays a role in the stability of the *Bst*HPr protein, although this role does not appear to be independent of other interactions. We have also shown that a proline at position 56 is beneficial to the stability of both HPr homologues. However, neither of these sequence peculiarities entirely account for the difference in stability between the HPr homologues. This observation prompts a closer look at the interactions between the C helix and its neighbors. Investigations into this effect by mutagenesis experiments will benefit from structural information on the variants created here (crystallization trials are in progress for these variants) and we will follow the leads provided. We also attempted to study the role of core packing in the HPr homologues by calorimetric studies on these homologues. Unfortunately, the *Bs*HPr homologue was not amenable to calorimetric experiments and this has precluded support for our findings from the theoretical analysis of ΔC_p.

Based on the experiments presented here and from previous work, it appears that no single paradigm, like presence of increased number of salt bridges or enhanced secondary structure, content can explain the enhanced stability of *Bst*HPr. Although, *Bst*HPr shows higher stability than *Bs*HPr at all temperatures as shown by the stability curves, it is probably a sum of several small stabilizing effects that cause the enhanced stability of *Bst*HPr.

CHAPTER VI

SUMMARY

A thermodynamic characterization of HPr homologues was undertaken here. These homologues were derived from organisms inhabiting diverse environmental conditions; two of these homologues are from moderate thermophiles (*Streptococcus thermophilus* and *Bacillus staerothermophilus*), one from a thermophilic haloalkaliphile (*Bacillus halodurans*) and the fourth from an extreme haloalkaliphile (*Oceanobacillus iheyensis*). The conformational stability characteristics of these homologues were compared with those of HPr from *Bacillus subtilus*, a mesophile. The comparison of stability characteristics allowed a ranking of these proteins, which established *Bst*HPr as the most stable and *Oi*HPr as the least stable of the proteins compared, under the conditions used. *Bst*HPr proved to be the most thermostable and also exhibited highest conformational stability at 25ºC. Stability comparisons were also made under various conditions of solution pH and [NaCl], which allowed us to estimate electrostatic contributions to conformational stability in each homologue. The results of such an analysis revealed unique characteristics for each protein, and the data led us to a few counterintuitive conclusions. For example, the *Bs*HPr protein, the mesophilic homologue appears to have highest contributions to stability from electrostatic interactions. This observation is counterintuitive given the established role of electrostatic interactions as major contributors to stability of proteins from thermophiles and the anticipated role that they might play in the proteins from halophilic organisms (*2, 127-130, 186*). A few other observations, like the low stability of *Oi*HPr in low [NaCl], or the indifference of *Bh*HPr to

added NaCl at high concentrations, can however be interpreted as reflective of their adaptations to habitat environments.

Stability curves were also determined for each homologue, and the data revealed interesting features of these homologues. One of the striking observations was the close agreement among values of ΔC_p and T_S values among these proteins. These observations from stability curve data allowed us to discern the thermodynamic mechanism used to achieve high T_m. The other curious observation from stability curve data was the close agreement among ΔG_L values for these proteins; this observation is testimony to the concept of 'corresponding states' first proposed by Somero (*116*). We have discussed in detail the thermodynamic mechanisms used by proteins from thermophiles to attain higher T_m. Our discussion was based on data compiled from literature studies on pair of homologous proteins from mesophilic and thermophilic organisms. This compilation has led us to draw conclusions on the use of the different mechanisms first proposed by Nojima et al for obtaining higher thermostability (*52*). The data show that to achieve higher T_m, proteins from thermophiles most commonly increase ΔG at all temperatures, therefore moving the stability curve up so that it now intersects the temperature abscissa at higher temperatures, thus providing higher T_m. Also, the data reveal that proteins from thermophiles use the third mechanism, namely shifting the stability curve to the right by virtue of higher T_S (temperature of maximum stability), the least to achieve a higher thermostability.

The crystal structures of *Bst*HPr and a variant were also reported here. The variant (F29W *Bst*HPr), proved to crystallize in two distinct conformations; one was almost identical to the wild-type protein, the second was domain-swapped dimeric conformation.

The significance of domain swapping with respect to amyloidogenesis was discussed in detail. Also, results from a comprehensive solution characterization of this variant failed to detect presence of this domain-swapped dimeric form under solution conditions. The availability of high-resolution crystal structures for *Bst*HPr allowed structural comparisons with *Bs*HPr for which a structure was available and these comparisons reveal very similar structures for these homologues. Structures for these homologues have also allowed calculation of theoretical ΔC_p values for these homologues, which provide interesting insights into the packing of their structures. However, these leads could not be followed up with experiments in the case of *Bs*HPr because this protein is not amenable to calorimetric analysis, thus precluding any support to the theoretical analysis of ΔC_p.

An analysis of sequence alignments for *Bs* and *Bst*HPr provided some insights into determinants of thermostability in *Bst*HPr and these were followed with mutagenesis experiments to prepare variants whose stability characteristics were studied. Although the study of these variants revealed interesting features regarding the stability of *Bst*HPr protein, it failed to account entirely for the difference in stability between the *Bs* and *Bst*HPr proteins. Work is in progress toward gathering structural information for these variants and we hope to gain more insights from these studies that will elucidate interactions responsible for enhanced thermostability of *Bst*HPr.

The results presented here should be useful in improving our understanding of protein stability in extreme environments in general. These results establish the stability features of homologous proteins adapted to diverse environments comprehensively, and their interpretation will be greatly aided by structural information and mutagenesis experiments.

REFERENCES

(1) Perutz, M. F., and Raidt, H. (1975) Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature 255*, 256-259.

(2) Perutz, M. F. (1978) Electrostatic effects in proteins. *Science 201*, 1187-1191.

(3) Jaenicke, R. (1991) Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem 202*, 715-728.

(4) van de Vossenberg, J. L., Driessen, A. J., and Konings, W. N. (1998) The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles 2*, 163-170.

(5) Jaenicke, R., and Zavodszky, P. (1990) Proteins under extreme physical conditions. *FEBS Lett 268*, 344-349.

(6) Sterner, R., and Liebl, W. (2001) Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol 36*, 39-106.

(7) Jaenicke, R. (1998) What ultrastable globular proteins teach us about protein stabilization. *Biochemistry (Mosc) 63*, 312-321.

(8) Vieille, C., and Zeikus, J. G. (1996) Thermozymes: identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol 14*, 183-190.

(9) Rees, D. C., and Adams, M. W. (1995) Hyperthermophiles: taking the heat and loving it. *Structure 3*, 251-254.

(10) Somero, G. N. (1995) Proteins and temperature. *Annu Rev Physiol 57*, 43-68.

(11) Argos, P., Rossman, M. G., Grau, U. M., Zuber, H., Frank, G., and Tratschin, J. D. (1979) Thermal stability and protein structure. *Biochemistry 18*, 5698-5703.

(12) Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem 14*, 1-63.

(13) Rose, G. D., and Wolfenden, R. (1993) Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu Rev Biophys Biomol Struct 22*, 381-415.

(14) Creighton, T. E. (1992) Folding and binding. *Curr Opin Struct Biol 2*, 1-5.

(15) Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry 29*, 7133-7155.

(16) Honig, B. (1999) Protein folding: from the Levinthal paradox to structure prediction. *J Mol Biol 293*, 283-293.

(17) Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J 10*, 75-83.

(18) Grimsley, G. R., Huyghues-Despointes, B. M. P., Pace, C. N., and Scholtz, J. M. (2003) Measuring the conformational stability of a protein, in *Purifying Proteins for Proteomics: A Laboratory Manual* (Simpson, R. J., Ed.) pp 535-566, Cold Spring Harbor Press, Cold Spring Harbor, New York.

(19) Lopez, M. M, and Makhatadze, G. I. (2002) Differntial scanning calorimetry. *Methods in Molecular Biology 173*, 113-119.

(20) Pace, C. N. (1992) Contribution of the hydrophobic effect to globular protein stability. *J Mol Biol 226*, 29-35.

(21) Livingstone, J. R., Spolar, R. S., and Record, M. T., Jr. (1991) Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry 30*, 4237-4244.

(22) Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry 31*, 3947-3955.

(23) Murphy, K. P., and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv Protein Chem 43*, 313-361.

(24) Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci 4*, 2138-2148.

(25) Russell, N. J., and Fukunaga, N. (1990) A comparison of thermal adaptation of membrane lipids in psychrophilic and thermophilic bacteria. *FEMS Microbiology Reviews 75*, 171-182.

(26) Choli, T., Henning, P., Wittmann-Liebold, B., and Reinhardt, R. (1988) Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the Archaebacterium, *Sulfolobus solfataricus*. *Biochim Biophys Acta 950*, 193-203.

(27) Kimura, M., Kimura, J., Davie, P., Reinhardt, R., and Dijk, J. (1984) The amino acid sequence of a small DNA binding protein from the archaebacterium *Sulfolobus solfataricus*. *FEBS Lett 176*, 176-178.

(28) McAfee, J. G., Edmondson, S. P., Datta, P. K., Shriver, J. W., and Gupta, R. (1995) Gene cloning, expression, and characterization of the Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry 34*, 10063-10077.

(29) Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B., and Ladenstein, R. (1996) Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile *Sulfolobus solfataricus*. *J Mol Biol 264*, 1132-1144.

(30) Dong, G., Vieille, C., Savchenko, A., and Zeikus, J. G. (1997) Cloning, sequencing, and expression of the gene encoding extracellular alpha-amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Appl Environ Microbiol 63*, 3569-3576.

(31) Borgne, S.L., and Quintero, R. (2003) Biotechnological processes for the refining of petroleum. *Fuel Processing Technology 81*, 155-169.

(32) Hogrefe, H. H., Cline, J., Lovejoy, A. E., and Nielson, K. B. (2001) DNA polymerases from hyperthermophiles. *Methods Enzymol 334*, 91-116.

(33) Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev 65*, 1-43.

(34) Matthews, B. W. (1993) Structural and genetic analysis of protein stability. *Annu Rev Biochem 62*, 139-160.

(35) Petsko, G. A. (2001) Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol 334*, 469-478.

(36) Vogt, G., and Argos, P. (1997) Protein thermal stability: hydrogen bonds or internal packing? *Fold Des 2*, S40-46.

(37) Vogt, G., Woell, S., and Argos, P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol 269*, 631-643.

(38) Szilagyi, A., and Zavodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure Fold Des 8*, 493-504.

(39) Bryan, P. N. (2000) Protein engineering of subtilisin. *Biochim Biophys Acta 1543*, 203-222.

(40) Eijsink, V. G., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., van den Burg, B., and Vriend, G. (2004) Rational engineering of enzyme stability. *J Biotechnol 113*, 105-120.

(41) Malakauskas, S. M., and Mayo, S. L. (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol 5*, 470-475.

(42) Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) De novo protein design: towards fully automated sequence selection. *J Mol Biol 273*, 789-796.

(43) Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005) Computational thermostabilization of an enzyme. *Science 308*, 857-860.

(44) Kuchner, O., and Arnold, F. H. (1997) Directed evolution of enzyme catalysts. *Trends Biotechnol 15*, 523-530.

(45) Arnold, F. H., Wintrode, P. L., Miyazaki, K., and Gershenson, A. (2001) How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci 26*, 100-106.

(46) Wintrode, P. L., and Arnold, F. H. (2000) Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv Protein Chem 55*, 161-225.

(47) Eijsink, V. G., Gaseidnes, S., Borchert, T. V., and van den Burg, B. (2005) Directed evolution of enzyme stability. *Biomol Eng 22*, 21-30.

(48) Sieber, V., Pluckthun, A., and Schmid, F. X. (1998) Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol 16*, 955-960.

(49) Martin, A., Sieber, V., and Schmid, F. X. (2001) In-vitro selection of highly stabilized protein variants with optimized surface. *J Mol Biol 309*, 717-726.

(50) Wunderlich, M., Martin, A., Staab, C. A., and Schmid, F. X. (2005) Evolutionary protein stabilization in comparison with computational design. *J Mol Biol 351*, 1160-1168.

(51) Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves. *Biopolymers 26*, 1859-1877.

(52) Nojima, H., Ikai, A., Oshima, T., and Noda, H. (1977) Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J Mol Biol 116*, 429-442.

(53) Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., and Sarai, A. (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res 32*, D120-121.

(54) Schindler, T., and Schmid, F. X. (1996) Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry 35*, 16833-16842.

(55) Jacob, M., Holtermann, G., Perl, D., Reinstein, J., Schindler, T., Geeves, M. A., and Schmid, F. X. (1999) Microsecond folding of the cold shock protein measured by a pressure-jump technique. *Biochemistry 38*, 2882-2891.

(56) Wassenberg, D., Welker, C., and Jaenicke, R. (1999) Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J Mol Biol 289*, 187-193.

(57) Hollien, J., and Marqusee, S. (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry 38*, 3831-3836.

(58) Kanaya, S., and Itaya, M. (1992) Expression, purification, and characterization of a recombinant ribonuclease H from *Thermus thermophilus* HB8. *J Biol Chem 267*, 10184-10192.

(59) Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I., and Ladenstein, R. (1998) Thermal unfolding of small proteins with SH3 domain folding pattern. *Proteins 31*, 309-319.

(60) Deutschman, W. A., and Dahlquist, F. W. (2001) Thermodynamic basis for the increased thermostability of CheY from the hyperthermophile *Thermotoga maritima*. *Biochemistry 40*, 13107-13113.

(61) Lee, C. F., Allen, M. D., Bycroft, M., and Wong, K. B. (2005) Electrostatic interactions contribute to reduced heat capacity change of unfolding in a thermophilic ribosomal protein l30e. *J Mol Biol 348*, 419-431.

(62) Arnone, M. I., Birolo, L., Pascarella, S., Cubellis, M. V., Bossa, F., Sannia, G., and Marino, G. (1997) Stability of aspartate aminotransferase from *Sulfolobus solfataricus*. *Protein Eng 10*, 237-248.

(63) Jaenicke, R., and Bohm, G. (2001) Thermostability of proteins from *Thermotoga maritima*. *Methods Enzymol 334*, 438-469.

(64) Grattinger, M., Dankesreiter, A., Schurig, H., and Jaenicke, R. (1998) Recombinant phosphoglycerate kinase from the hyperthermophilic bacterium *Thermotoga maritima*: catalytic, spectral and thermodynamic properties. *J Mol Biol 280*, 525-533.

(65) Hu, C. Q., and Sturtevant, J. M. (1989) A differential scanning calorimetric study of the binding of sulfate ion and of Cibacron blue F3GA to yeast phosphoglycerate kinase. *Biochemistry 28*, 813-818.

(66) McCrary, B. S., Edmondson, S. P., and Shriver, J. W. (1996) Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. *J Mol Biol 264*, 784-805.

(67) Xu, S., Qin, S., and Pan, X. M. (2004) Thermal and conformational stability of Ssh10b protein from archaeon *Sulfolobus shibattae*. *Biochem J 382*, 433-440.

(68) Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nat Struct Biol 1*, 518-523.

(69) Ruiz-Sanz, J., Filimonov, V. V., Christodoulou, E., Vorgias, C. E., and Mateo, P. L. (2004) Thermodynamic analysis of the unfolding and stability of the dimeric DNA-binding protein HU from the hyperthermophilic eubacterium *Thermotoga maritima* and its E34D mutant. *Eur J Biochem 271*, 1497-1507.

(70) Consalvi, V., Chiaraluce, R., Giangiacomo, L., Scandurra, R., Christova, P., Karshikoff, A., Knapp, S., and Ladenstein, R. (2000) Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*. *Protein Eng 13*, 501-507.

(71) Motono, C., Oshima, T., and Yamagishi, A. (2001) High thermal stability of 3-isopropylmalate dehydrogenase from *Thermus thermophilus* resulting from low ΔC_p of unfolding. *Protein Eng 14*, 961-966.

(72) Sridharan, S., Razvi, A., Scholtz, J. M., and Sacchettini, J. C. (2005) The HPr proteins from the thermophile *Bacillus stearothermophilus* can form domain-swapped dimers. *J Mol Biol 346*, 919-931.

(73) Li, W. T., Grayling, R. A., Sandman, K., Edmondson, S., Shriver, J. W., and Reeve, J. N. (1998) Thermodynamic stability of archaeal histones. *Biochemistry 37*, 10563-10572.

(74) Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R., and Shoichet, B. K. (1999) Comparing the thermodynamic stabilities of a related thermophilic and mesophilic enzyme. *Biochemistry 38*, 2570-2576.

(75) Hiller, R., Zhou, Z. H., Adams, M. W., and Englander, S. W. (1997) Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 degrees C. *Proc Natl Acad Sci U S A 94*, 11329-11332.

(76) Hernandez, G., and LeMaster, D. M. (2001) Reduced temperature dependence of collective conformational opening in a hyperthermophile rubredoxin. *Biochemistry 40*, 14384-14391.

(77) Shiraki, K., Nishikori, S., Fujiwara, S., Hashimoto, H., Kai, Y., Takagi, M., and Imanaka, T. (2001) Comparative analyses of the conformational stability of a hyperthermophilic protein and its mesophilic counterpart. *Eur J Biochem 268*, 4144-4150.

(78) Pfeil, W., Gesierich, U., Kleemann, G. R., and Sterner, R. (1997) Ferredoxin from the hyperthermophile *Thermotoga maritima* is stable beyond the boiling point of water. *J Mol Biol 272*, 591-596.

(79) Dams, T., and Jaenicke, R. (1999) Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemistry 38*, 9169-9178.

(80) Ionescu, R. M., Smith, V. F., O'Neill, J. C., Jr., and Matthews, C. R. (2000) Multistate equilibrium unfolding of *Escherichia coli* dihydrofolate reductase:

thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. *Biochemistry 39*, 9540-9550.

(81) Ohmae, E., Kurumiya, T., Makino, S., and Gekko, K. (1996) Acid and thermal unfolding of *Escherichia coli* dihydrofolate reductase. *J Biochem (Tokyo) 120*, 946-953.

(82) Nojima, H., Hon-Nami, K., Oshima, T., and Noda, H. (1978) Reversible thermal unfolding of thermostable cytochrome c-552. *J Mol Biol 122*, 33-42.

(83) Fujiwara, S., Yamanaka, A., Hirooka, K., Kobayashi, A., Imanaka, T., and Fukusaki, E. (2004) Temperature-dependent modulation of farnesyl diphosphate/geranylgeranyl diphosphate synthase from hyperthermophilic archaea. *Biochem Biophys Res Commun 325*, 1066-1074.

(84) Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*, 4673-4680.

(85) Guzman-Casado, M., Parody-Morreale, A., Robic, S., Marqusee, S., and Sanchez-Ruiz, J. M. (2003) Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H. *J Mol Biol 329*, 731-743.

(86) Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M., and Marqusee, S. (2003) Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci U S A 100*, 11345-11349.

(87) Anderson, J. W., Bhanot, P., Georges, F., Klevit, R. E., and Waygood, E. B. (1991) Involvement of the carboxy-terminal residue in the active site of the histidine-containing protein, HPr, of the phosphoenolpyruvate: sugar phosphotransferase system of *Escherichia coli*. *Biochemistry 30*, 9601-9607.

(88) Lai, X., and Ingram, L. O. (1995) Discovery of a ptsHI operon, which includes a third gene (ptsT), in the thermophile *Bacillus stearothermophilus*. *Microbiology 141 (Pt 6)*, 1443-1449.

(89) Waygood, E. B., Reiche, B., Hengstenberg, W., and Lee, J. S. (1987) Characterization of mutant histidine-containing proteins of the phosphoenolpyruvate: sugar phosphotransferase system of *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol 169*, 2810-2818.

(90) Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis 18*, 2714-2723.

(91) Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst 24*, 946-950.

(92) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem 25*, 1605-1612.

(93) Hebert, E. J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K. S., Dauter, Z., and Pace, C. N. (1998) Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. *Biochemistry 37*, 16192-16200.

(94) Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph 8*, 52-56, 29.

(95) Howorka, S., and Bayley, H. (1998) Improved protocol for high-throughput cysteine scanning mutagenesis. *Biotechniques 25*, 764-772.

(96) Kalia, A., Rattan, A., and Chopra, P. (1999) A method for extraction of high-quality and high-quantity genomic DNA generally applicable to pathogenic bacteria. *Anal Biochem 275*, 1-5.

(97) Nicholson, E. M., Peterson, R. W., and Scholtz, J. M. (2002) A partially buried site in homologous HPr proteins is not optimized for stability. *J Mol Biol 321*, 355-362.

(98) Greene, R. F., Jr., and Pace, C. N. (1974) Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem 249*, 5388-5393.

(99) Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry 27*, 8063-8068.

(100) Pace, C. N., and Laurents, D. V. (1989) A new method for determining the heat capacity change for protein folding. *Biochemistry 28*, 2520-2525.

(101) Goldberg, R. N., Kishore, N., and Lennen, R. M. (2002) Thermodynamic quantities for the ionization reactions of buffers. *J Phys Chem Ref Data 31*, 231-370.

(102) Kholodenko, V., and Freire, E. (1999) A simple method to measure the absolute heat capacity of proteins. *Anal Biochem 270*, 336-338.

(103) Gibbs, S. J., Johnson, J., and Charles S. (1991) A PFG NMR experiment for accurate diffusion and flow studies in the presence of eddy currents. *J Mag Reson (1969) 93*, 395-402.

(104) Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science 181*, 223-230.

(105) Booth, I. R. (1985) Regulation of cytoplasmic pH in bacteria. *Microbiol Rev 49*, 359-378.

(106) Timasheff, S. N. (2002) Protein hydration, thermodynamic binding, and preferential hydration. *Biochemistry 41*, 13473-13482.

(107) Timasheff, S. N. (1993) The control of protein stability and association by weak interactions with water: how do solvents affect these processes? *Annu Rev Biophys Biomol Struct 22*, 67-97.

(108) Horikoshi, K. (1999) Alkaliphiles: some applications of their products for biotechnology. *Microbiol Mol Biol Rev 63*, 735-750.

(109) Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G., and Eijsink, V. G. (1998) Engineering an enzyme to resist boiling. *Proc Natl Acad Sci U S A 95*, 2056-2060.

(110) Cheung, Y. Y., Lam, S. Y., Chu, W. K., Allen, M. D., Bycroft, M., and Wong, K. B. (2005) Crystal structure of a hyperthermophilic archaeal acylphosphatase from *Pyrococcus horikoshii*-structural insights into enzymatic catalysis, thermostability, and dimerization. *Biochemistry 44*, 4601-4611.

(111) Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S., and Horikoshi, K. (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res 28*, 4317-4331.

(112) Takami, H., and Horikoshi, K. (1999) Reidentification of facultatively alkaliphilic *Bacillus* sp. C-125 to *Bacillus halodurans*. *Biosci Biotechnol Biochem 63*, 943-945.

(113) Lu, J., Nogi, Y., and Takami, H. (2001) *Oceanobacillus iheyensis* gen. nov., sp. nov., a deep-sea extremely halotolerant and alkaliphilic species isolated from a depth of 1050 m on the Iheya Ridge. *FEMS Microbiol Lett 205*, 291-297.

(114) Herzberg, O., Reddy, P., Sutrina, S., Saier, M. H., Jr., Reizer, J., and Kapadia, G. (1992) Structure of the histidine-containing phosphocarrier protein HPr from *Bacillus subtilis* at 2.0-A resolution. *Proc Natl Acad Sci U S A 89*, 2499-2503.

(115) Kohn, W. D., Kay, C. M., and Hodges, R. S. (1995) Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper. *Protein Sci 4*, 237-250.

(116) Somero, G. (1978) Temperature adaptation of enzymes - biological optimization through structure-function compromises. *Ann Rev Eco Sys 9*, 1-29.

(117) Akke, M. (2004) Out of hot water. *Nat Struct Mol Biol 11*, 912-913.

(118) Zavodszky, P., Kardos, J., Svingor, and Petsko, G. A. (1998) Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc Natl Acad Sci U S A 95*, 7406-7411.

(119) Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., and Kern, D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol 11*, 945-949.

(120) Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci U S A 84*, 6663-6667.

(121) Munoz, V., and Serrano, L. (1997) Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers 41*, 495-509.

(122) Kuntz, I. D., Jr., and Kauzmann, W. (1974) Hydration of proteins and polypeptides. *Adv Protein Chem 28*, 239-345.

(123) Dym, O., Mevarech, M., and Sussman, J. L. (1995) Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science 267*, 1344-1346.

(124) Rao, J. K., and Argos, P. (1981) Structural stability of halophilic proteins. *Biochemistry 20*, 6536-6543.

(125) Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003) Unique amino acid composition of proteins in halophilic bacteria. *J Mol Biol 327*, 347-357.

(126) Chakravarty, S., and Varadarajan, R. (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry 41*, 8152-8161.

(127) Xiao, L., and Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol 289*, 1435-1444.

(128) Karshikoff, A., and Ladenstein, R. (2001) Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem Sci 26*, 550-556.

(129) Spassov, V. Z., Karshikoff, A. D., and Ladenstein, R. (1995) The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci 4*, 1516-1527.

(130) Elcock, A. H. (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J Mol Biol 284*, 489-502.

(131) Vonhippel, P. H., and Wong, K. Y. (1964) Neutral salts: the generality of their effects on the stability of macromolecular conformations. *Science 145*, 577-580.

(132) Baldwin, R. L. (1996) How Hofmeister ion interactions affect protein stability. *Biophys J 71*, 2056-2063.

(133) Nandi, P. K., and Robinson, D. R. (1972) The effects of salts on the free energy of the peptide group. *J Am Chem Soc 94*, 1299-1308.

(134) Nandi, P. K., and Robinson, D. R. (1972) The effects of salts on the free energies of nonpolar groups in model peptides. *J Am Chem Soc 94*, 1308-1315.

(135) Quirk, D. J., Park, C., Thompson, J. E., and Raines, R. T. (1998) His...Asp catalytic dyad of ribonuclease A: conformational stability of the wild-type, D121N, D121A, and H119A enzymes. *Biochemistry 37*, 17958-17964.

(136) Ladenstein, R., and Antranikian, G. (1998) Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv Biochem Eng Biotechnol 61*, 37-85.

(137) Clark, A. T., McCrary, B. S., Edmondson, S. P., and Shriver, J. W. (2004) Thermodynamics of core hydrophobicity and packing in the hyperthermophile proteins Sac7d and Sso7d. *Biochemistry 43*, 2840-2853.

(138) Scholtz, J. M. (1995) Conformational stability of HPr: the histidine-containing phosphocarrier protein from *Bacillus subtilis*. *Protein Sci 4*, 35-43.

(139) Nicholson, E. M., and Scholtz, J. M. (1996) Conformational stability of the *Escherichia coli* HPr protein: test of the linear extrapolation method and a thermodynamic characterization of cold denaturation. *Biochemistry 35*, 11369-11378.

(140) Napper, S., Delbaere, L. T., and Waygood, E. B. (1999) The aspartyl replacement of the active site histidine in histidine-containing protein, HPr, of the *Escherichia coli* Phosphoenolpyruvate: sugar phosphotransferase system can accept and donate a phosphoryl group. Spontaneous dephosphorylation of acyl-phosphate autocatalyzes an internal cyclization. *J Biol Chem 274*, 21776-21782.

(141) Jia, Z., Vandonselaar, M., Hengstenberg, W., Quail, J. W., and Delbaere, L. T. (1994) The 1.6 Å structure of histidine-containing phosphotransfer protein HPr from *Streptococcus faecalis*. *J Mol Biol 236*, 1341-1355.

(142) Kalbitzer, H. R., Gorler, A., Li, H., Dubovskii, P. V., Hengstenberg, W., Kowolik, C., Yamada, H., and Akasaka, K. (2000) ^{15}N and ^{1}H NMR study of histidine

containing protein (HPr) from *Staphylococcus carnosus* at high pressure. *Protein Sci 9*, 693-703.

(143) Kalbitzer, H. R., and Hengstenberg, W. (1993) The solution structure of the histidine-containing protein (HPr) from *Staphylococcus aureus* as determined by two-dimensional ^{1}H-NMR spectroscopy. *Eur J Biochem 216*, 205-214.

(144) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276, 307-326.

(145) Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res 31*, 3381-3385.

(146) Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr 56 Pt 12*, 1622-1624.

(147) Collaborative Computational Project, Number 4. (1994) The CCP4 Suite: programs for protein crystallography. *Acta Cryst D 50*, 760-763.

(148) Reddy, V., Swanson, S. M., Segelke, B., Kantardjieff, K. A., Sacchettini, J. C., and Rupp, B. (2003) Effective electron-density map improvement and structure validation on a Linux multi-CPU web cluster: the TB structural genomics consortium bias removal web service. *Acta Crystallogr D Biol Crystallogr 59*, 2200-2210.

(149) McRee, D. E. (1999) XtalView/Xfit--a versatile program for manipulating atomic coordinates and electron density. *J Struct Biol 125*, 156-165.

(150) Liu, Y., and Eisenberg, D. (2002) 3D domain swapping: as domains continue to swap. *Protein Sci 11*, 1285-1299.

(151) Kabsch, W. (1976) A solution for the best notation to relate two sets of vecotrs. *Acta. Cryst. A 32*, 922-923.

(152) Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry 38*, 16424-16431.

(153) Creamer, T. P., Srinivasan, R., and Rose, G. D. (1997) Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry 36*, 2832-2835.

(154) Creamer, T. P., Srinivasan, R., and Rose, G. D. (1995) Modeling unfolded states of peptides and proteins. *Biochemistry 34*, 16245-16250.

(155) Rousseau, F., Schymkowitz, J. W., and Itzhaki, L. S. (2003) The unfolding story of three-dimensional domain swapping. *Structure (Camb) 11*, 243-251.

(156) Schlunegger, M. P., Bennett, M. J., and Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem 50*, 61-122.

(157) Klafki, H. W., Pick, A. I., Pardowitz, I., Cole, T., Awni, L. A., Barnikol, H. U., Mayer, F., Kratzin, H. D., and Hilschmann, N. (1993) Reduction of disulfide bonds in an amyloidogenic Bence Jones protein leads to formation of "amyloid-like" fibrils in vitro. *Biol Chem Hoppe Seyler 374*, 1117-1122.

(158) Knaus, K. J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W. K., and Yee, V. C. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol 8*, 770-774.

(159) Staniforth, R. A., Giannini, S., Higgins, L. D., Conroy, M. J., Hounslow, A. M., Jerala, R., Craven, C. J., and Waltho, J. P. (2001) Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *Embo J 20*, 4774-4781.

(160) Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001) Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat Struct Biol 8*, 316-320.

(161) Liu, Y., Gotte, G., Libonati, M., and Eisenberg, D. (2001) A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol 8*, 211-214.

(162) Schmittschmitt, J. P., and Scholtz, J. M. (2003) The role of protein stability, solubility, and net charge in amyloid fibril formation. *Protein Sci 12*, 2374-2378.

(163) Juy, M., Penin, F., Favier, A., Galinier, A., Montserret, R., Haser, R., Deutscher, J., and Bockmann, A. (2003) Dimerization of Crh by reversible 3D domain swapping induces structural adjustments to its monomeric homologue HPr. *J Mol Biol 332*, 767-776.

(164) Gotte, G., Bertoldi, M., and Libonati, M. (1999) Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme. *Eur J Biochem 265*, 680-687.

(165) Libonati, M., Bertoldi, M., and Sorrentino, S. (1996) The activity on double-stranded RNA of aggregates of ribonuclease A higher than dimers increases as a function of the size of the aggregates. *Biochem J 318 (Pt 1)*, 287-290.

(166) Crestfield, A. M., Stein, W. H., and Moore, S. (1962) On the aggregation of bovine pancreatic ribonuclease. *Arch Biochem Biophys Suppl 1*, 217-222.

(167) Liu, Y., Hart, P. J., Schlunegger, M. P., and Eisenberg, D. (1998) The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-A resolution. *Proc Natl Acad Sci U S A 95*, 3437-3442.

(168) Alston, R. W., Urbanikova, L., Sevcik, J., Lasagna, M., Reinhart, G. D., Scholtz, J. M., and Pace, C. N. (2004) Contribution of single tryptophan residues to the fluorescence and stability of ribonuclease Sa. *Biophys J 87*, 4036-4047.

(169) Takano, K., Scholtz, J. M., Sacchettini, J. C., and Pace, C. N. (2003) The contribution of polar group burial to protein stability is strongly context-dependent. *J Biol Chem 278*, 31790-31795.

(170) Pace, C. N., Horn, G., Hebert, E. J., Bechert, J., Shaw, K., Urbanikova, L., Scholtz, J. M., and Sevcik, J. (2001) Tyrosine hydrogen bonds make a large contribution to protein stability. *J Mol Biol 312*, 393-404.

(171) Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science 255*, 178-183.

(172) Xu, J., Baase, W. A., Baldwin, E., and Matthews, B. W. (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci 7*, 158-177.

(173) Schmittschmitt, J. P., and Scholtz, J. M. (2004) The side chain of aspartic acid 69 dictates the folding mechanism of *Bacillus subtilis* HPr. *Biochemistry 43*, 1360-1368.

(174) Lebbink, J. H., Consalvi, V., Chiaraluce, R., Berndt, K. D., and Ladenstein, R. (2002) Structural and thermodynamic studies on a salt-bridge triad in the NADP-binding domain of glutamate dehydrogenase from *Thermotoga maritima*: cooperativity and electrostatic contribution to stability. *Biochemistry 41*, 15524-15535.

(175) Goedken, E. R., and Marqusee, S. (2001) Native-state energetics of a thermostabilized variant of ribonuclease HI. *J Mol Biol 314*, 863-871.

(176) Bogin, O., Levin, I., Hacham, Y., Tel-Or, S., Peretz, M., Frolow, F., and Burstein, Y. (2002) Structural basis for the enhanced thermal stability of alcohol dehydrogenase mutants from the mesophilic bacterium *Clostridium beijerinckii*: contribution of salt bridging. *Protein Sci 11*, 2561-2574.

(177) Oikawa, K., Nakamura, S., Sonoyama, T., Ohshima, A., Kobayashi, Y., Takayama, S. J., Yamamoto, Y., Uchiyama, S., Hasegawa, J., and Sambongi, Y. (2005) Five amino acid residues responsible for the high stability of *Hydrogenobacter thermophilus* cytochrome c552: reciprocal mutation analysis. *J Biol Chem 280*, 5527-5532.

(178) Tanner, J. J., Hecht, R. M., and Krause, K. L. (1996) Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-

glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å resolution. *Biochemistry 35*, 2597-2609.

(179) Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv Protein Chem 33*, 167-241.

(180) Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1993) The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. *Science 262*, 1715-1718.

(181) Hasegawa, J., Yoshida, T., Yamazaki, T., Sambongi, Y., Yu, Y., Igarashi, Y., Kodama, T., Yamazaki, K., Kyogoku, Y., and Kobayashi, Y. (1998) Solution structure of thermostable cytochrome c-552 from *Hydrogenobacter thermophilus* determined by ^{1}H-NMR spectroscopy. *Biochemistry 37*, 9641-9649.

(182) Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R., and Jaenicke, R. (2000) The crystal structure of dihydrofolate reductase from *Thermotoga maritima*: molecular features of thermostability. *J Mol Biol 297*, 659-672.

(183) Russell, R. J., Hough, D. W., Danson, M. J., and Taylor, G. L. (1994) The crystal structure of citrate synthase from the thermophilic archaeon, *Thermoplasma acidophilum*. *Structure 2*, 1157-1167.

(184) Britton, K. L., Baker, P. J., Borges, K. M., Engel, P. C., Pasquo, A., Rice, D. W., Robb, F. T., Scandurra, R., Stillman, T. J., and Yip, K. S. (1995) Insights into thermal stability from a comparison of the glutamate dehydrogenases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Eur J Biochem 229*, 688-695.

(185) Karshikoff, A., and Ladenstein, R. (1998) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng 11*, 867-872.

(186) Elcock, A. H., and McCammon, J. A. (1998) Electrostatic contributions to the stability of halophilic proteins. *J Mol Biol 280*, 731-748.

VITA

Name: Abbas Razvi Syed Ali

Address: c/o Dr. J. Martin Scholtz

440 Reynolds Medical Building

1114 TAMU, College Station, TX 77843-1114

Email address: ali@pauling.tamu.edu

Education: B.Sc., Microbiology, Botany and Chemistry, Osmania University, 1997

M.S., Biotechnology, University of Pune, 1999

Ph.D., Biochemistry, Texas A&M University, 2005