Calorimetric studies of the PBX Homeodomain:

Uncovering the Unfolding Mechanism

Hariyanto Darmawan

Department of Chemistry

McGill University, Montreal

August 2009

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of M.Sc.

Abstract

The heat capacity of unfolding of the PBX homeodomain was measured under different solvent conditions using Differential Scanning Calorimetry (DSC). The results were fitted using different models in order to best determine the unfolding mechanism of PBX homeodomain. Solely from the calorimetric data, it was shown that PBX-HD doesn't unfold in a simple 2-state manner. The fitting suggests that it unfolds in a three-state manner. However, DSC calorimetric data alone couldn't provide a conclusive picture of the unfolding mechanism of PBX. Fitting the DSC data together with NMR data, the result suggests that PBX does indeed unfold in a two-state manner, but that the folded state itself undergoes an endothermic transition, leading to a complex thermodynamic signature.

Résumé

La capacité thermique du dépliement du homéodomaine PBX a été mesurée sous différentes conditions de solvant grâce à la Calorimétrie différentielle à balayage. Les résultats ont été ajustés utilisant différents modèles afin de déterminer adéquatement les mécanismes de dépliement du homéodomaine PBX. Les données calorimétriques démontrent que le dépliement du HD-PBX ne se fait pas selon un processus simple de deux étapes. En fait, l'ajustement suggère un processus de dépliement à trois étapes. Cependant, les données calorimétriques provenant du 'DSC' ne peuvent à elles-seules fournir une image concluante du mécanisme de dépliement du HD-PBX. En combinant les données du « DSC » avec celles du RMN, les résultats suggèrent que le HD-PBX se déplie effectivement par un processus en deux-étapes, mais que l'état de repliement elle-même entreprend une transition endothermique amenant à une signature thermodynamique complexe.

Acknowledgement

First and foremost I would like to thank my supervisor, Dr. Anthony Mittermaier, for his guidance and patience in teaching me all the skills and knowledge that make this Master thesis possible. In the span of 2 years, I have learned more than I have had in the previous 6 years, from various laboratory techniques to methods of investigation. Dr. Mittermaier has been more than just my research supervisor; he has also been my mentor.

I would also like to acknowledge my colleagues in the lab: Patrick James Farber, Teresa Miletti, and Jean-Philippe Demers. They have helped make my transition to this lab smooth and lent their hands to teach me the necessary laboratory techniques. Special mention to Patrick James Farber for many helpful discussions and for the NMR data and results to complement my calorimetric study of PBX homeodomain.

I would also like to thank Dr. Masad Damha for his helpful advice on graduate school. Last but not least, I would like to thank all members of McGill Department of Chemistry who have enriched my graduate school experience in their own way.

Table of Content

Chapter I: Introduction

One of the most fundamental questions in biology has always been: *How does protein fold and unfold*? The structure-function relationship of proteins makes it very important to study the mechanism and the forces that hold a protein together in a functional structure. A better understanding of the mechanism of protein folding-unfolding and the forces involved in this process will be important in understanding the function of the protein. The holy grail of protein study is to predict with confidence the structure of a protein strictly from its sequence, and also to understand its dynamics under different physiological environments.

In 1950s, Christian Afinsen postulated that the three dimensional structure of a native (physiologically folded) protein is determined by its amino acid sequence. Simply said, a denatured protein, which has all its secondary and tertiary structures completely disrupted, can refold, under the correct conditions, into its unique functional native structure. Later in the late 60s, Levinthal first proposed what is now known as the Levinthal paradox (*1,2*). Given the astronomical number of possible conformations of unfolded polypeptide chains (far exceeding the number of atoms in the universe), it is impossible that proteins fold via an unbiased search in conformational space, since protein folds in a matter of milliseconds or even microseconds. To answer this paradox, Levinthal argued: "We feel that protein folding is speeded and guided by the rapid formation of local interactions which then determine the further folding of the peptide. This suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process." (*2*)

The study of the folding-unfolding mechanism of proteins and the forces involved is often seen as a study of the protein energy landscape. A protein energy landscape is the free energy of each protein conformation as a function of the degrees of freedom, such as the dihedral bond angles along the peptide backbone. The vertical axis of this energy landscape represents the free energy of a given chain configuration: the sum of hydrogen bonds, ion-pairs, torsion angle energies, hydrophobic and solvation energy, and so on; while the many lateral axes represent the conformational coordinates. Many coordinates, for example the dihedral angles φ_1, ψ_1, φ_2, ψ_2, …, are needed to specify a conformation and each conformation is represented by a point on the multidimensional energy landscape.

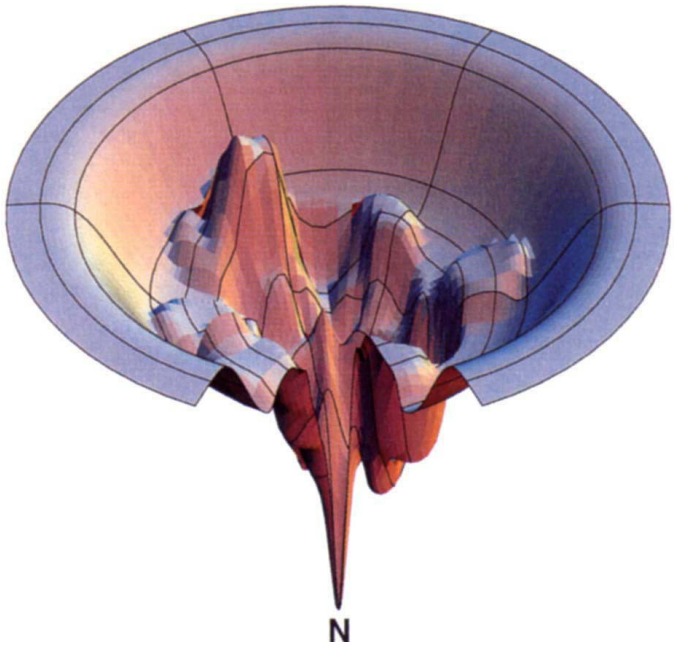

Figure 1.1. Protein energy landscape

An artist's rendition of a protein energy landscape where N is the native structure (*3*). The vertical axis represents the free energy of a given chain configuration. The lateral axes represent the conformational coordinates.

It has been generally agreed that proteins have rugged energy landscapes with varying energy barriers and kinetic traps (*3*); Protein folding can take a multitude of forms: from pathways which have discreet defined states (cooperatively) to folding in the absence of any energy barriers, i.e. down-hill (non-cooperatively). Such a protein energy landscape is depicted schematically in figure 1.1 above (*3*). Of course, this low dimensional figure, where only two conformational coordinates are shown schematically, cannot do justice to the high dimensionality of the configuration space of a protein.

An area of great current interest is the refinement of experimental methods and data analysis or modeling approaches in order to differentiate between proteins that fold cooperatively and non-cooperatively. The recent example of this is the on-going debate on the folding nature of BBL protein, whereby the Munoz and co-workers (*4, 7, 30, 31*) argue that BBL protein folds non-cooperatively while Fersht and co-workers (*5, 32, 33, 34, 35*) argue otherwise.

1.1 Cooperative versus non-cooperative folding

The classical notion of protein folding is based on a chemical reaction scheme where there exists a series of chemical equilibria between distinct thermodynamic states that are individually populated at equilibrium and separated by energy barriers. (*6*) This classical model was aptly elaborated by Privalov in 1986: "… a temperature increase induces the macromolecule to pass through distinct macroscopic states. Arranging all these states according to increasing

enthalpy value, we can formally describe the equilibrium existing between these states by the following sequential scheme:"

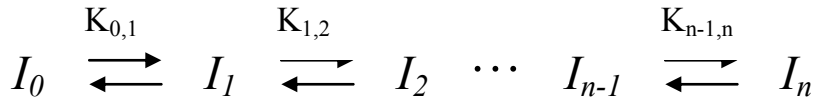

Figure 1.2. Sequential schemes of protein unfolding

Chemical equilibria between ($n+1$) distinct thermodynamic states. $K_{i,j}$ denotes the equilibrium constant between the states i and j; where I_0 is the native state and I_n is the denatured state.

Consequently, the model that has often been used to fit experimental data is based on such chemical phenomenology and thus is biased toward this result. It is therefore desirable to analyze an experimental data with a model that does not make assumption about the existence of a free energy barrier, (*7*) or at least to develop methods to cross-validate the chemical equilibrium model. Although it is still hotly debated in the recent literature whether non-cooperative (downhill) folding of protein occurs for natural proteins (*4,5,8*).

In a cooperative folding, one can picture multiple discreet states (or none in the case of 2-state folding) that the protein accesses in the pathway from a denatured state to a well-defined native state. Whereas in non-cooperative (downhill) folding, the protein undergoes continuous folding without any discreet states being populated; it is a downhill passage across the energy landscape without encountering any significant energy barrier.

Figure 1.3 below depicts the free energy profile for cooperative unfolding that has two wells (folded and unfolded) separated by a barrier. As we change the temperature (or any environmental variable such as denaturant concentration), we bias the system toward the folded or unfolded well. Figure 1.4 below depicts a downhill unfolding where there is no energy barrier. The well is continuously shifted from the folded state to unfolded state (vice versa) by tuning the temperature or other environmental variables.

Population-wise, we can imagine that in cooperative unfolding, only two (or more) distinct states are populated and for downhill unfolding the protein populates an ensemble of states that shifts continuously.

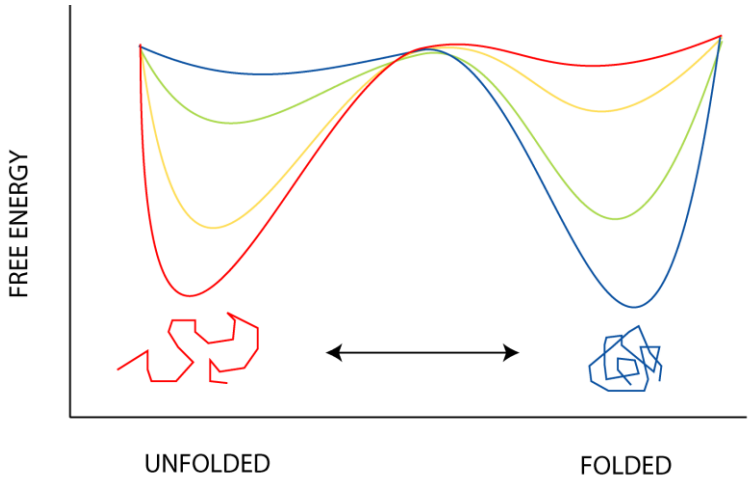

Figure 1.3. Schematic free energy diagram of cooperative folding

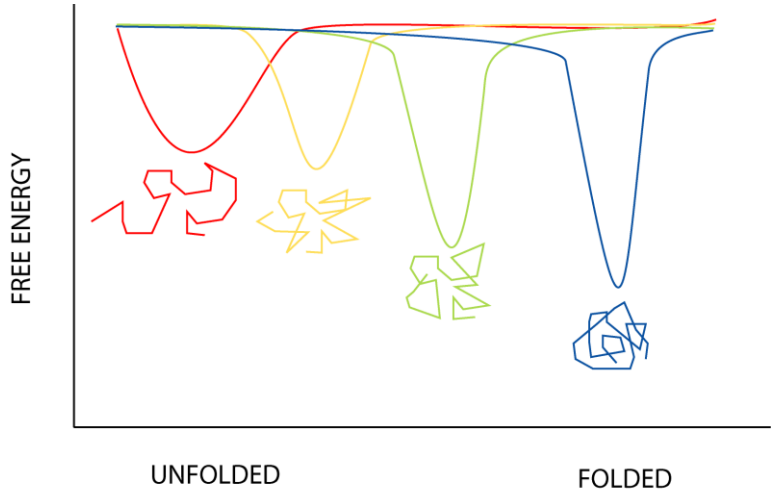

Figure 1.4. Schematic free energy diagram of non-cooperative or downhill folding

1.2 Heat capacity and protein unfolding

Protein thermal denaturation is an endothermic process. The heat of the unfolding represents the enthalpy of the process. This heat absorption results in a heat capacity increase near the melting temperature. One of the methods of direct measurement of the heat of unfolding is Differential Scanning Calorimetry (DSC). DSC measures the heat capacity of protein as a function of temperature. Since the heat capacity determined at constant pressure is a temperature derivative of the enthalpy function:

$$C_P = \frac{dH}{dT} \qquad [1.1]$$

then one can calculate the enthalpy of the protein at any point during the unfolding reaction by integration of the heat capacity:

$$H(T) = H(T_0) + \int_{T_0}^{T} C_P(T)dT \qquad [1.2]$$

where T_0 is reference temperature.

Figure 1.5 below depicts the heat capacity of lysozyme measured with DSC in solutions with different pH values (*6*).

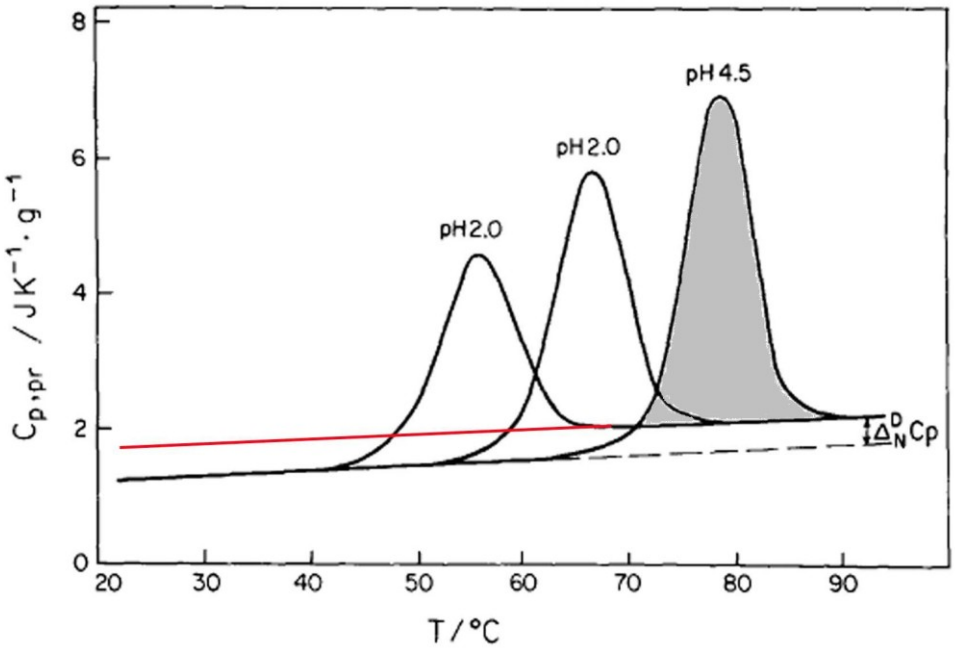

Figure 1.5. Heat capacity of lysozyme in solutions with different pHs (*6*)

The black dashed line is the extrapolation of the pre-transition protein heat capacity, giving the heat capacity of the native (folded) state. The red line is added to highlight the extrapolation of the post-transition protein heat capacity, giving the heat capacity of the denatured (unfolded) state. $\Delta_N^D C_P$ is the heat capacity difference between the native (folded) state and the denatured (unfolded) state. The area under the curve, in gray, is the excess heat due to unfolding.

The excess heat capacity of the protein relative to some state, considered as its "zero" state, is the difference between the heat capacity of the protein and the heat capacity of this "zero" state at the considered temperature *T*.

$$\left\langle \Delta C_P(T) \right\rangle = C_P(T)_{pr} - C_{P,0}(T)_{pr} \qquad [1.3]$$

where $\left\langle \Delta C_P(T) \right\rangle$ is the excess heat capacity, $C_P(T)_{pr}$ is the heat capacity of the protein, and $C_{P,0}(T)_{pr}$ is the heat capacity of the "zero" state at the considered temperature *T*.

The heat capacity of this "zero" state is the population-weighted average of the heat capacities of the pure folded and unfolded states which are extrapolated from the pre-transition and post-transition baselines of the protein heat capacity respectively (black dashed-line and red line in figure 1.5 respectively). Physically, these folded and unfolded baselines represent the heat capacities due to the intensifying vibrations of those states and not the changes in protein structure (*22*).

From the relationship in equation 1.1, the heat absorbed due to unfolding is simply the integration of the area beneath the protein excess heat capacity:

$$\left\langle \Delta H(T) \right\rangle = \int_0^T \left\langle \Delta C_P(T) \right\rangle dT \qquad [1.4]$$

where $\left\langle \Delta H(T) \right\rangle$ is the excess enthalpy of the protein as it unfolds. In the above equation 1.4, the integration is from absolute zero temperature. However, in practice, it is integrated from some temperature where the heat capacity of the protein is equal to that of the folded state and the excess heat capacity is zero.

Using a general relation of statistical thermodynamics below:

$$\left\langle \Delta H(T) \right\rangle = RT^2 \frac{d \ln Q}{dT} \qquad [1.5]$$

where R is the gas constant, Q is the partition function, and T is the temperature, we can obtain all the thermodynamic information on the states realized in the considered temperature range.

The heat capacity of the protein in the unfolded state is generally higher than in the folded state (*9,10*). The main contributor to this increase in heat capacity is the formation of water cage (or hydration) around the hydrophobic regions which are exposed as the native compact structure is disrupted. In other words, increased exposure of hydrophobic region to the solvent results in the increase of heat capacity. The other factor, configurational freedom gained by the polypeptide chain upon disruption of the rigid compact native structure, is found to contribute only a small part of the overall heat capacity change on unfolding (*9,10*). However, physically conformational entropy does play a significant role in the stability of a compact native structure of a protein, along with two other forces: the interactions between groups within a protein molecule, and the interactions between these groups and the surrounding water (hydration) (*11,12*).

1.3 Theory of Differential Scanning Calorimetry

The differential scanning calorimeter measures the heat absorbed by a system as it undergoes endothermic transition, and this absorbed heat is detected as the excess heat capacity. The DSC registers this excess heat capacity as the

temperature is increased or decreased continuously at a constant rate under a constant pressure. In other words, it scans along a certain temperature range by measuring continuously small changes in the heat capacity of the samples, hence the name scanning calorimetry (*13,14*).

Continuous heating and measurement have great advantages over discrete step-wise measurement since it gives more complete information on the heat capacity function. However, the principal problem of continuous heating is that the system measured is not in complete equilibrium. There are two reasons for this non-equilibrium: first, the heat conductivity of the system; second, the rate of temperature-induced processes of the system. The first problem is tackled by constructing a calorimetric cell with a small volume and large surface area. Reducing the volume of the sample usually leads to an increase of the relative measurement error. However, for the past 20 years the rapid development of calorimetry has increased its sensitivity and made possible a measurement of small-volume, dilute samples. To deal with the second problem, one has to make sure that the temperature-induced changes in the samples are not too sharp and the relaxation process of these changes is sufficiently fast, i.e. to choose the appropriate heating rate.

The scanning calorimetry does not measure the absolute heat capacity of the sample. It instead measures the difference heat capacity of the sample against a reference (usually the solvent); hence the name “differential”. The studied sample and the reference sample are heated simultaneously and under the same conditions. Thus, the DSC instruments usually have twin cells, one loaded with

the solution to be studied and the other one with the reference. The heat capacity of the sample is determined relative to the reference.

However, in reality, it is impossible to construct twin cells that are absolutely identical. To overcome this problem, the measurement has to be done twice. First, with both cells filled with the reference solution to determine the "zero" or "base line". Second, with one cell (the sample cell) filled with the sample and the other (the reference cell) with the reference solution. By a comparison of these two recordings, we can eliminate the non-identity of the two cells.

There are two basic methods to measure the heat capacity difference of the two cells. The first one is by measuring the difference in temperature of the cells which are heated with the same power. The second method is by measuring the difference of power necessary to heat both cells to increase at the same rate of temperature, in other words determining the power which compensates for the difference in the heat capacity of the two cells. The second method is better than the first one because we directly measure a parameter (the difference in power) which is proportional to heat capacity difference (*6*). Therefore, the compensation method is mainly used in precision scanning calorimetry.

Another major problem of DSC is the loading of the cells with equal and definite volumes. It is almost impossible to load the cells accurately by weighing the samples. Also, cells that are removable pose another problem since after their removal it is practically impossible to reestablish the original condition due to a number of factors such as contamination on the surface, change in surface contacts between the cells and its respective containers, etc. This problem is

solved by the construction of undismountable cells which are completely filled by samples through the capillary inlets and have fixed definite volumes. The main requirement is that the cells have to be completely filled with sample and contain no microscopic bubbles. One such design comprises capillary tubes wound into a helix, like the one used here (Nano-Differential Scanning Calorimeter III Model CSC 6300) (see figure 1.6).

To prevent bubbles from the cells upon heating, a high pressure is applied by a manostat and the measurements are performed under this constant pressure. A few atmospheres are generally sufficient to compress all the bubbles and prevent their formation on heating (*6*). In addition to this, the excess pressure also increases the temperature range of the calorimetry above 100 Celsius. This is important in studying thermostable macromolecules.

Figure 1.6. Nano-Differential Scanning Calorimeter III Model CSC 6300 and the helical capillary cells (www.tainstruments.com)

The DSC measures the difference in power, ΔP (Watts), which is obtained by subtracting the baseline microwatt data from the sample microwatt data. The relationship between ΔP and the apparent heat capacity difference between the sample and the solvent, ΔC_P^{app} (J/K) is:

$$\Delta C_P^{app}(T) = \Delta P(T)\frac{dt}{dT} \qquad [1.6]$$

where dt/dT is the inverse scan rate (s/K). The apparent heat capacity is usually negative because the heat capacity of the sample is smaller than the heat capacity of the same volume of solvent. From the above equation 1.6, this means that the measured ΔP is usually negative.

The apparent heat capacity is the observed difference in heat capacity between the sample and the solvent of the same volume.

$$\Delta C_P^{app}(T) = C_p(T)_{sample}\, m(T)_{sample} - C_p(T)_{solv}\, \Delta m(T)_{solv} \qquad [1.7]$$

where $C_P(T)_{sample}$ (J/K g) is the partial specific heat capacity of the sample, $m(T)_{sample}$ (gram) is the mass of the sample, $C_P(T)_{solv}$ (J/K g) is the specific heat capacity of the solvent (J/K g), and $\Delta m(T)_{solv}$ (gram) is the mass of the solvent displaced by the sample in solution:

$$\Delta m(T)_{solv} = m(T)_{sample} \times PSV \times \rho_{solv} \qquad [1.8]$$

where ρ_{solv} (g/mL) is the density of the solvent and *PSV* (mL/g) is the Partial Specific Volume of the sample.

Rearranging the equation 1.7, we have:

$$C_p(T)_{sample} = \frac{\Delta P(T) \times \frac{dt}{dT}}{Conc_{sample} \times V_{cell}} + C_p(T)_{solv} \times PSV \times \rho_{solv} \quad [1.9]$$

where $Conc_{sample}$ (g/mL) is the concentration of the sample, V_{cell} (mL) is the volume of the cell (or the total volume of the solution in study).

1.4 About PBX Homeodomain

The transcription factor PBX contains a DNA-binding homeodomain that bears a three amino acid loop extension. It binds DNA cooperatively with the HOX homeodomain, giving rise to heterodimers with greater DNA-binding affinity and specificity than the separate monomer (*29*).

The PBX-HD used in this experiment consists of residues 233-294 of the full length protein, which from now on will be referred to as residues 1-59. The extra 3 amino acids are numbered 23a-c to conform to the consensus numbering scheme.

The molecular weight of the above PBX construct is 7438.4 Dalton, with an extinction coefficient of 11460 M^{-1} cm^{-1} at 280 nm (obtained from ProtParam

tool). It has 14 positively charged residues and 6 negatively charged residues, making it a positively charged protein with a net charge of +8 at pH 7.

1……..10……20……a.b.c……..30……...40……
ARRKRRNFNKQATEILNEYFYSHLSNPYPSEEAKEELAKKCGITVSQVSN
…..50...……..59
WFGNKRIRYKKN

Figure 1.7. Amino acid sequence of PBX-HD construct used in this experiment

The underlined sequences are the three helices that fold around a hydrophobic core. The extra 3 amino acids are numbered 23a-c to conform to the consensus numbering scheme.

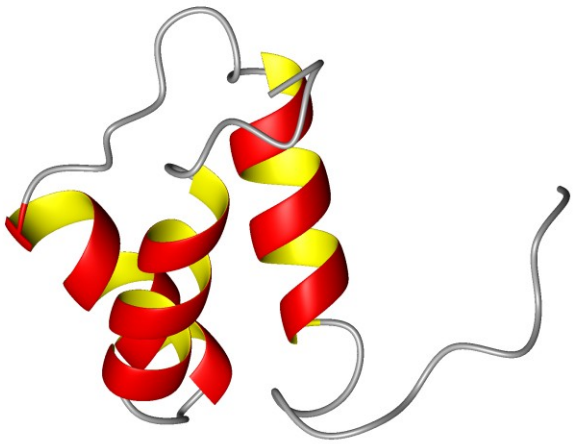

Figure 1.8. Truncated PBX structure (*15*)

The PBX-HD consists of three alpha helices that fold around a hydrophobic core. Based on the structure of truncated PBX-HD (residue 9-59) obtained from solution NMR (*15*), the first alpha helix extends from residue Q11

to Y21, the second from E28 to C38, and the third from V42 to K52. The NMR analysis of the 1-59 construct shows that the N terminal arm is unstructured (*15*).

Chapter 2: Materials and Methods

2.1 Protein preparation

The PBX-HD protein was expressed in freshly transformed Rosetta competent cells. Rosetta competent cells are designed to enhance the expression of proteins that contains codons rarely used in *E. coli* (AUA, AGG, AGA, CUA, CCC, CGG, GGA), and are used to express PBX-HD protein since it contains those rare codons. 100 µL of frozen competent cells were thawed in ice and 2 µL PBX plasmid were added. This was incubated in ice for 5 minutes and then followed directly with a 30 second incubation at 42°C. Immediately after the heat shock, the solution was returned to ice for 2 min and then plated on an LB-agarose medium containing ampicilin (100 µg/mL) and chloramphenicol (34 µg/mL). The plate was then incubated at 37°C for one night. The next day, colonies were observed on the plate. The plate with colonies was stored at 4°C and was used within 1-2 weeks.

For the protein expression, 50 mL of LB solution with ampicilin (100 µg/mL) and chloramphenicol (34 µg/mL) was inoculated with a single colony of the transformed cells. The 50 mL culture was incubated with shaking at 37°C until an OD_{600} of ~0.2 was reached, typically after 6-7 hours. The 50 mL culture was then left at room temperature overnight. The next day, the culture was added to 1 liter of LB with the same concentrations of ampicilin and chloramphenicol. It was incubated with shaking at 37°C until an OD of ~0.6-0.8 was reached. Protein expression was induced through the addition of IPTG to a final concentration of 1mM and proceeded for 4 hours. The cells were harvested by 20 minute of

centrifugation at 8000 rpm in a SLA rotor. The pellet was resuspended in lysis buffer (50 mL 20mM Na-Phosphate buffer pH 6.4, 100mM NaCl, 4M Urea), and stored at -20°C. The next day, the frozen cell pellet was thawed and lysed by two 5 minute period of sonication with a 5 minute intervening period on ice. Cell debris was removed by 30 minute centrifugation at 16000 rpm in a SS-34 rotor.

The protein was purified by ion-exchange chromatography using SP-Sepharose with a flow rate of 2 mL/min. The supernatant was loaded on the column, washed with 20 mL of lysis buffer containing 100mM NaCl and a further 250 mL of the same buffer with 250mM NaCl, during which the OD_{260} of this washing elution was checked to make sure that all of the DNA had come off from the column. This step was crucial to ensure that there is no DNA left that could form a complex with the PBX-HD. The protein was eluted with lysis buffer containing 1M NaCl. It was then dialyzed against 4L refolding buffer (20mM Sodium Phosphate pH 6.4, 5% Glycerol, 200 mM NaCl). At this stage, white precipitates or aggregates formed, mostly composed of contaminants, and they were removed by 15 minute centrifugation at 16000 rpm in a SS-34 rotor. The aggregates were further removed by filtering the protein solution through 0.45 micron filter.

The protein was further purified by size-exclusion FPLC (Fast Protein Liquid Chromatography) using a Superdex 75 10/300 GL column. The protein solution was concentrated first to total volume of 1 mL using Amicon Ultra-15 centrifugal ultrafiltration device prior to injection on the column. The PBX-HD eluted at around 13.0-14.4 mL. (See figure 2.1)

The protein sample was run on SDS-PAGE gel (15% acrylamide) before and after the FPLC procedure (see figure 2.2. and 2.3) and it was clearly shown that there was a significant increase in purity after the FPLC procedure and that many other high molecular weight proteins were removed.

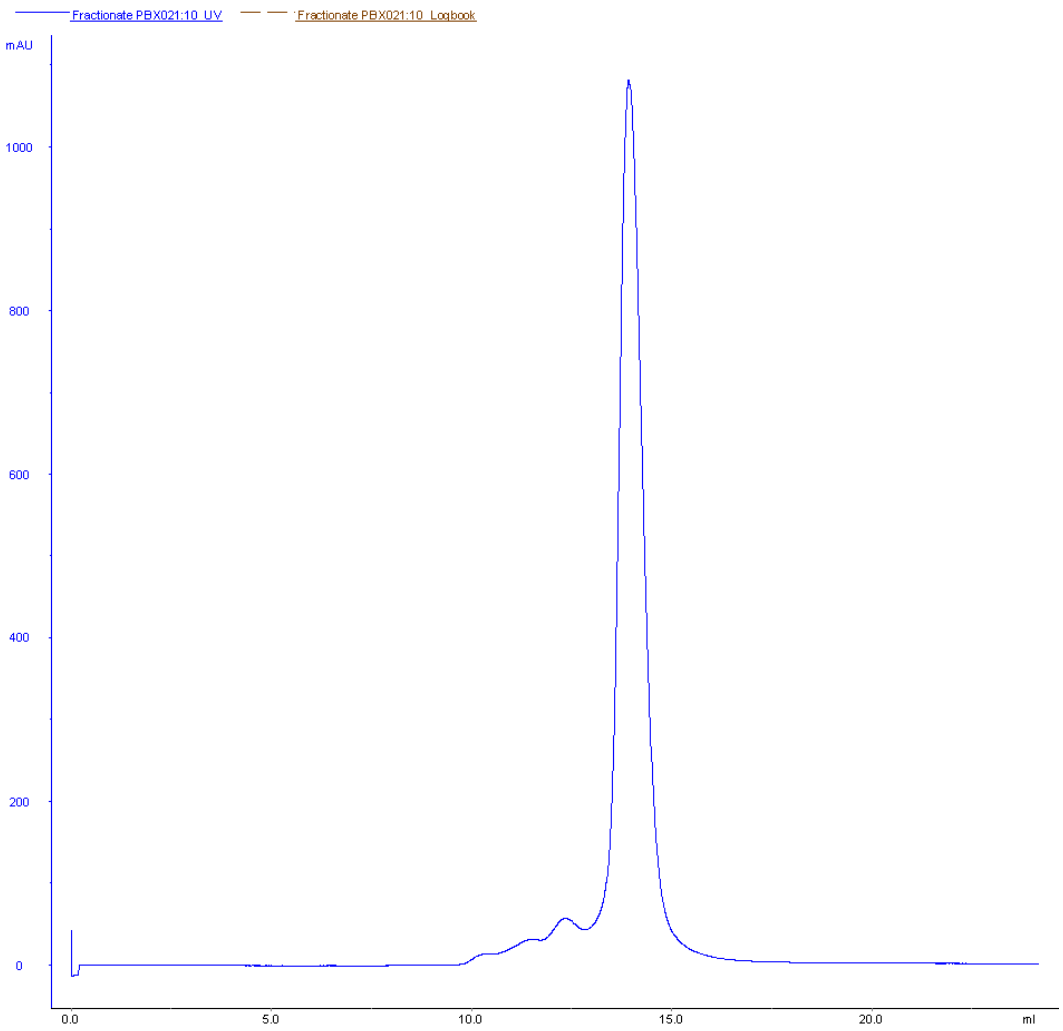

Figure 2.1. FPLC elution diagram of PBX-HD

Y-axis denotes A_{280} reading of the protein, and the X-axis denotes the elution volume. PBX-HD is shown to elute between 13.0 – 14.4 mL.

The fractions that contained PBX-HD were then dialyzed extensively against 4L of the final buffer of interest. After dialysis, the protein was passed through a 0.45 micron filter to remove any aggregates and dust. The dialysis buffer was also passed through a 0.45 micron filter and used as a reference. Approximately 10-15 mg of pure PBX-HD was obtained for every 1 L of culture.

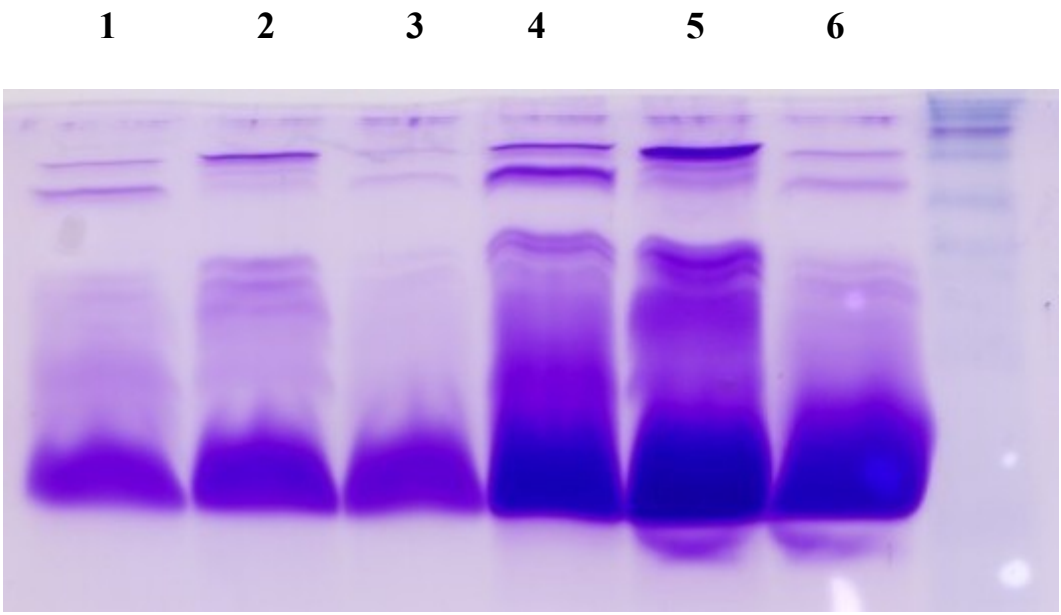

Figure 2.2. SDS Page gel of PBX-HD before size-exclusion FPLC purification
Three protein preparations, right after the ion-exchange chromatography and before the size-exclusion chromatography, were run on 15% acrylamide SDS PAGE. Lane 1, 2, and 3 are the ¼ dilution counterparts of lane 4, 5, and 6. The gel shows a significant amount of high-molecular weight protein contaminants.

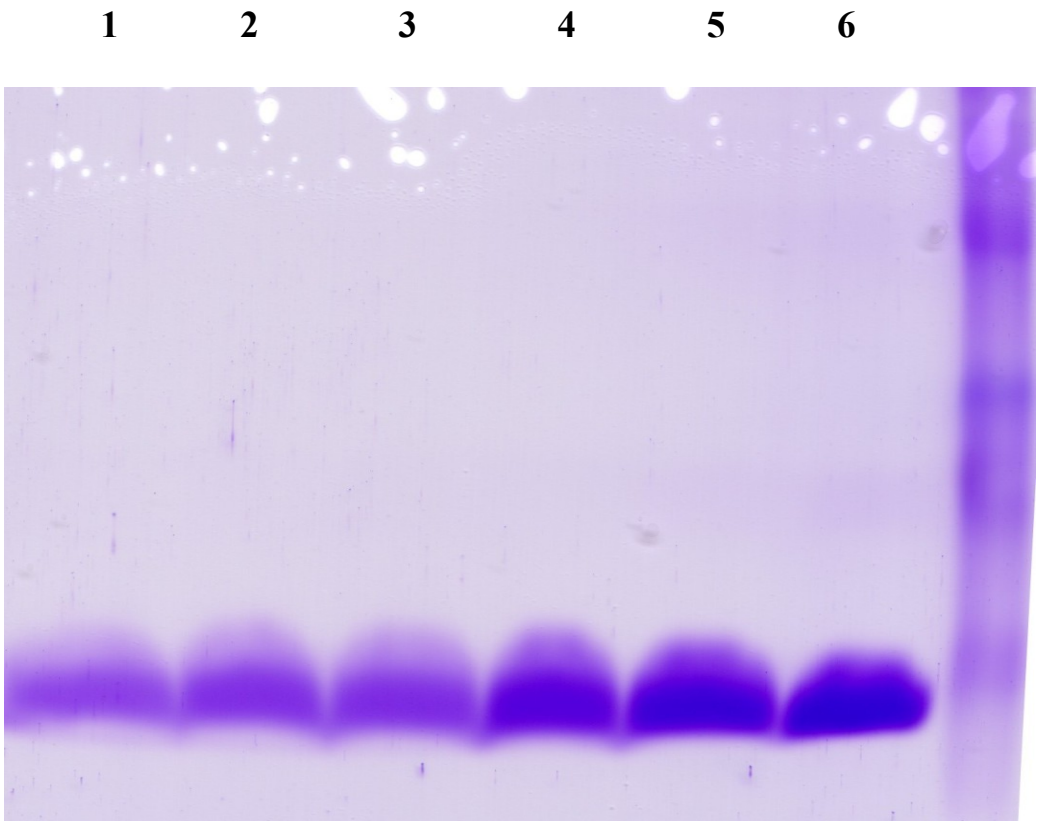

Figure 2.3. SDS Page gel of PBX-HD after size-exclusion FPLC purification
The same three protein preparations as in figure 2.2 were run on 15% acrylamide SDS PAGE after size-exclusion FPLC procedure. Lane 1, 2, and 3 are the ¼ dilution counterparts of lane 4, 5, and 6. The gel shows a significant increase in the purity of the protein of interest.

2.2 DSC Scans of Protein Sample

The DSC instrument used in this experiment is CSC Nano-Differential Calorimeter III (N-DSC III) which was developed by Applied Thermodynamics. It has the operating range of -10°C to 130°C with scan rates up to 2°C/minutes in both heating and cooling directions. For most proteins, intermolecular association can be neglected at concentrations below 10^{-4} M (*14*). The PBX-HD sample was prepared at a concentration of ~1 mg/mL or ~1.34 x 10^{-4} M. The solubility of

dissolved gases in water (O_2 and N_2) decreases as a solution is heated. To prevent the formation of gas bubbles as the sample is heated, both the standard and the sample solutions were degassed prior to DSC measurements using a Microcal Thermovac.

The next step was to condition/equilibrate the cells with the reference buffer before running the baseline. Both reference and sample cells were filled with the reference (buffer solution of the protein preparations). The capillary cells were filled with extreme care to ensure no air bubbles were introduced, because even the smallest bubble could result in erratic heat capacity readings. After loading, the pressure handle was tightened and the sample was pressurized to 3 atm. DSC scans were taken from 0°C to 100°C with a 1°C/minute heating and cooling rate. After this conditioning, the capillary cells were flushed with ~500 mL double-distilled water.

Since the two cells of DSC, reference cell and sample cell, are not identical, it is necessary to run a baseline in order to correct for the small difference between the cells. This baseline measurement is done by filling both cells with the reference buffer and scanning over the entire required range of temperatures. Below figure 2.4 displays three consecutive baselines obtained in heating scans of 20mM Sodium Phosphate pH 6 with a 1°C/minute heating rate from 5°C to 80°C, idling temperature of 5°C (idling temperature is the default temperature when the machine is idle, i.e. it is the initial temperature before the first scan is commenced), equilibration time of 60 seconds (equilibration time is the time allowed for the machine to reach the next initial scan temperature after reaching the end of the heating or cooling scan. The default equilibration time is

600 seconds). The equilibration time was shortened in order to prevent protein aggregation at high temperature. The noise level of this scan is ±0.084 µW standard deviation. If the baseline scans were not reproducible, it was repeated again with fresh buffer.

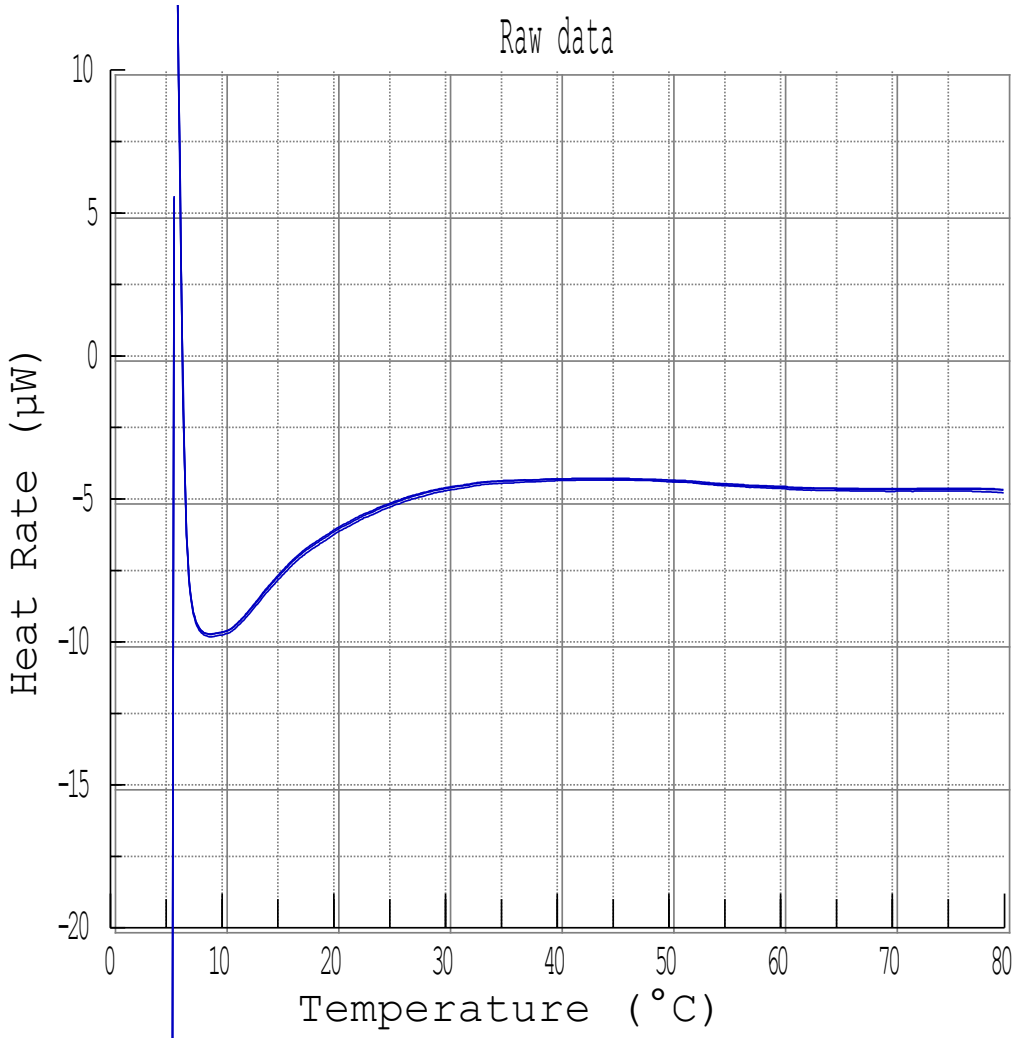

Figure 2.4. DSC baseline scan of reference buffer

Three consecutive heating scans of reference buffer 20 mM Sodium Phosphate pH 6 with a 1°C/minute heating rate from 5°C to 80°C, idling temperature of 5°C, and an equilibration time of 60 seconds.

Following the baseline scans, the cells were flushed with ~500 mL double-distilled water before running the protein sample. Since the protein solution was prone to form foam, an extra care was paid in filling the sample cell with the protein sample. The scan range and parameters for the protein sample runs were identical to those of baseline runs (1°C/minute heating rate from 5°C to 80°C, idling temperature of 5°C, and an equilibration time of 60 seconds). At least 3 consecutive heating and cooling scans were performed. Below, figure 2.5 shows the 3 consecutive heating scans of PBX-HD in 20mM Sodium Phosphate buffer pH 6.

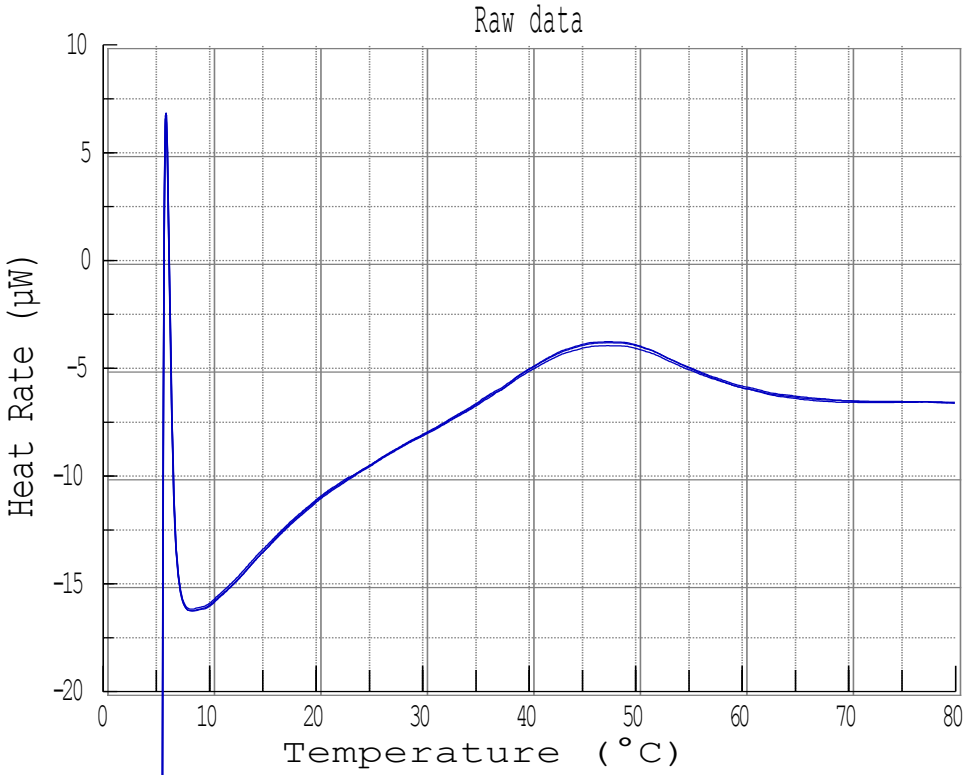

Figure 2.5. DSC scans of PBX-HD protein

Three consecutive heating scans of PBX-HD in 20mM Sodium Phosphate pH 6, with a 1°C/minute heating rate from 5°C to 80°C, idling temperature of 5°C, and an equilibration time of 60 seconds.

Multiple runs were performed for each protein preparation. Between runs, the capillary cells were flushed with ~500 mL double-distilled water.

2.3 Cleaning the cells

In between protein preparations, the cells were cleaned rigorously of any protein aggregates that might form on the surface of the cells. First, the cells were flushed with 1L of 0.1M NaOH solution, followed by 1L water. The cells were then filled with 50% v/v formic acid solution and one heating scan was run from 25°C to 80°C at 1°C/minute with the pressure handle not attached to the calorimeter. The cells were flushed again with 1 L water and then filled with a mixture of 4M NaOH, 15% methanol and 15% residue-free detergent (extran 300). The cells were pressurized to 3 atm and a scan was run from 25°C to 80°C with 1°C/minute heating rate and 0.001°C/minute cooling rate. This allowed the cells to remain heated with the NaOH/MeOH/Detergent solution overnight. The next day, the cells were flushed with 1 L water and filled with 50% formic acid. The same procedure was then repeated but only for few hours instead of overnight. The cells were flushed with 1L water and a water baseline was run to check the reproducibility.

The cells were occasionally washed with an anhydrous tetrahydrofuran to remove any residual organic matter with scan range from 25°C to 50°C with 1°C/minute heating rate and 0.001 °C/minute cooling rate, while leaving the pressure handle open. The THF-cleaning scan was run for few hours.

2.4. Heat capacity calculation

After the baseline data and sample data were obtained, the raw μW data was calculated, normalized, and averaged. Due to differences in each protein preparations and differences in room temperature when the scans were run, and also other uncontrolled factors such as the evaporation of buffers, each scan has different offset, i.e. they are slightly shifted vertically, and they also lead to uncertainties in the overall scaling of the μW data. To correct this, A MATLAB program was written to process the data. Below is an example of the MATLAB program to average and normalize three sets of μW data. Before the raw μW data were processed with this MATLAB program, they were treated to cubic spline interpolation since each of the scans lists power data at slightly different temperatures.

```
dP1=Sample1-Baseline1;
dP2=Sample2-Baseline2;
dP3=Sample3-Baseline3;

dP=mean([dP1',dP2',dP3'])

poly1=polyfit(dP,dP1,1);
poly2=polyfit(dP,dP2,1);
poly3=polyfit(dP,dP3,1);
P1=(dP1-poly1(2))/poly1(1);
P2=(dP2-poly2(2))/poly2(1);
P3=(dP3-poly3(2))/poly3(1);

PSum=[P1;P2;P3];

PNorm=mean(mean(PSum));
for nn=1:length(PSum(:,1))
        PSum(nn,:) = PSum(nn,:)-mean(PSum(nn,:))+PNorm;
End
deltaP=mean(PSum)';
```

First, the baseline μW data (*Baseline1, Baseline2, Baseline3*) were subtracted from the sample μW data (*Sample1, Sample2, Sample3*). Then these data (*dP1, dP2, dP3*) were averaged to give *dP*. The next step is to correct the offset and the overall scaling of the μW data. Scaling must be taken into account, since the protein concentration can change slightly, for instance due to adsorption on the capillary walls or evaporation of the liquid. Each of the subtracted data (*dP1, dP2, dP3*) were fit linearly to the averaged data, *dP*. Subtraction of *poly(2)*, which is the y-intercept of the line, accounts for the offset; while normalization by *poly(1)*, the slope of the fitted line, rescales the data. Finally, they were all averaged to give *deltaP* as the average value of ΔP, from which the heat capacity can be calculated using equation 1.9.

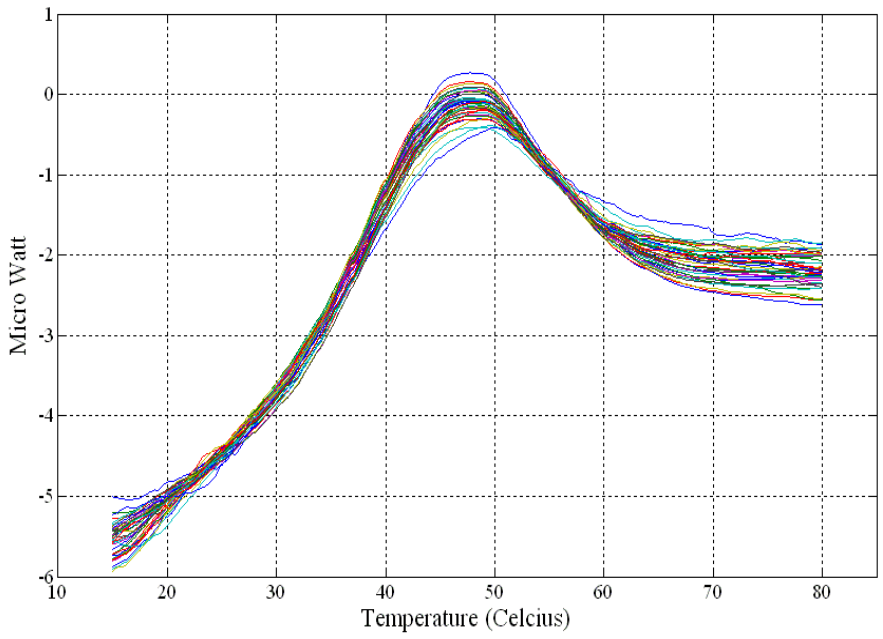

Figure 2.6. Overlay of 46 DSC scans of PBX-HD

μW data of the 46 DSC scans of PBX-HD in 20 mM Sodium Phosphate pH 6. These data have been corrected for the offset and the overall scaling.

For the 20mM Sodium Phosphate pH 6 condition, 4 different protein samples were prepared and 46 DSC scans were performed with concentrations ranging from 0.312 mg/mL to 1.095 mg/mL. Above figure 2.6 is the overlay of a total of 46 scans of the four different protein preparations.

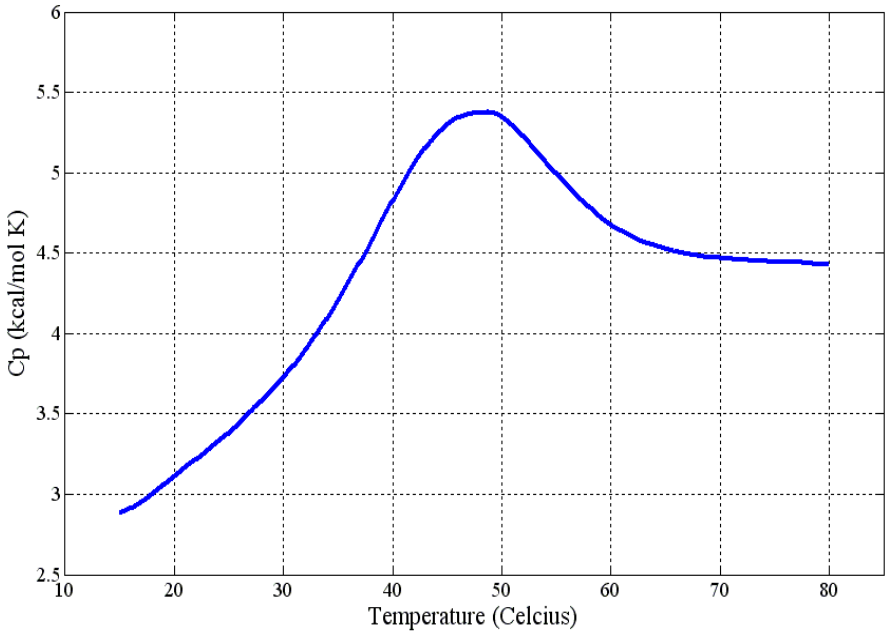

Figure 2.7. Heat capacity of PBX-HD

Heat capacity of PBX-HD in 20 mM Sodium Phosphate pH 6. This heat capacity was obtained from averaging the normalized μW data depicted in figure 2.6 and calculating the heat capacity using equation 1.9

The average value of ΔP was calculated and the heat capacity of the protein was determined using equation 1.9 (see figure 2.7). V_{cell} of the N-DSC III is 0.300 mL. The Partial Specific Volume (*PSV*) of PBX-HD at 25°C was calculated to be 0.726 mg/mL (*16*). The heat capacity and the density of the solvent were assumed to be that of pure water: 4.184 J/K g and 1000 g/L. These three parameters (*PSV*, heat capacity of the solvent, and the density of the solvent)

were assumed to be temperature-independent, a safe assumption since the PSV value is an increasing function and the water density is a decreasing function of temperature, while the heat capacity of water is fairly constant within that operating range of temperature (*36*); thus the total change is very small. Within the range of 5°C to 80°C, the calculated *PSV* of PBX-HD increases by ~5% (*16*), while the water density decreases by ~3% (*36*). Moreover, since these three parameters are largely independent from the ΔP parameter (see equation 1.9), they do not affect the slopes of the heat capacity profile and only the intercept; and during the fitting procedure, this intercept is fitted with an offset.

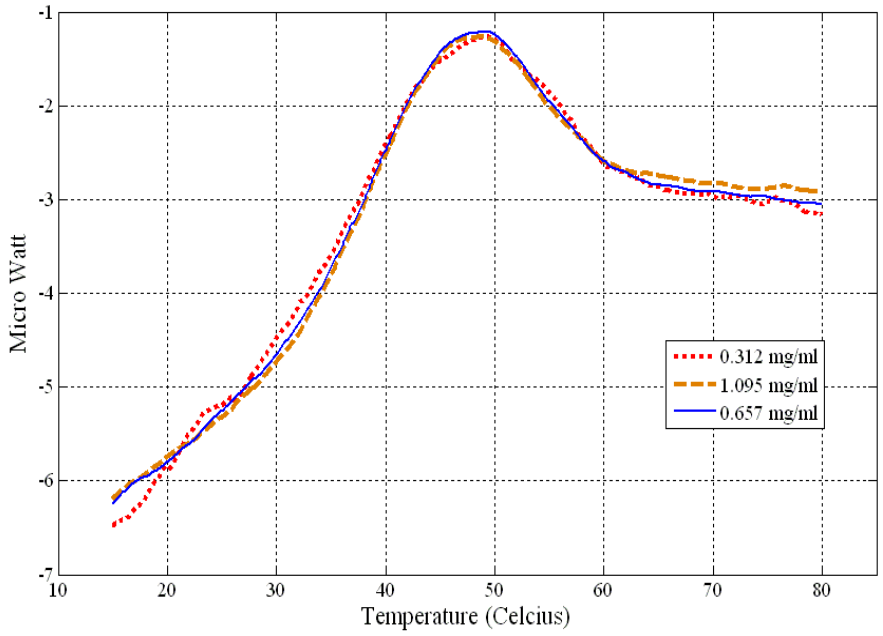

Figure 2.8. DSC runs with different concentrations

PBX-HD was scanned with three different concentrations: 0.312 mg/mL (red dashed line), 0.657 mg/mL (blue line), and 1.095 mg/mL (brown dashed line).

One of the protein samples was used to prepare a series of dilutions at concentration of 1.095 mg/mL, 0.657 mg/mL, and 0.312 mg/mL. Their respective DSC scans showed that the DSC profile is not concentration-dependant at this concentration range (see figure 2.8). Since for most proteins intermolecular association can be neglected at concentrations below 10^{-4} M (*14*), or 0.75 mg/mL of PBX-HD, therefore it can be safely assumed that protein self-association does not contribute to the heat capacity data. The figure also shows that there is more noise at lower concentration of 0.312 mg/mL which was expected due to a lower signal.

2.5 Error analysis

The uncertainties of the heat capacity measurements are obtained from the average deviation of recording of repeated measurements with the following relationship (*13*):

$$C_p(T)_{sample}\, error = \frac{\delta\Delta P(T) \times \frac{dt}{dT}}{Conc_{sample} \times V_{cell}} \Big/ \sqrt{N} \qquad [2.1]$$

where $\delta\Delta P$ is the standard deviation of the μW data and N is the number of measurements. The mean calculated error for the 20mM Sodium Phosphate pH 6 PBX-HD data is ~0.009 kcal/mole K or ~0.2%.

Chapter 3: Analysis

The study of the heat capacity changes is a powerful tool for understanding the unfolding mechanism of a protein, whether it is a two-state process, three-state process, etc, or downhill unfolding. The heat capacity profile of PBX-HD in 20mM Sodium Phosphate pH 6 was fitted with different models and their results were analyzed and compared to each other.

3.1 Discreet state models

In a discreet states model, it is assumed that a temperature increase induces the macromolecule to pass through distinct macroscopic states. It comprises a low-enthalpy folded native state, a high-enthalpy unfolded state and possibly a number of intermediate states. In the case of a simple two-state unfolding process, there is no intermediate state. We can formally describe the equilibrium between all the realized states with the following expression (*14*):

$$I_0 \underset{}{\overset{K_{0,1}}{\rightleftharpoons}} I_1 \overset{K_{1,2}}{\rightleftharpoons} I_2 \cdots I_{n-1} \overset{K_{n-1,n}}{\rightleftharpoons} I_n$$

Figure 3.1. Sequential schemes of protein unfolding

Chemical equilibria between $(n+1)$ distinct thermodynamic states. $K_{i,j}$ denotes the equilibrium constant between the states i and j; where I_0 is the native state and I_n is the denatured state.

The thermodynamic functions for the system above are:

$$K_{0,i} = \prod_{j=1}^{i} K_{j-1,j} \qquad [3.1]$$

$$\Delta_0^i H = H_i(T) - H_0(T) = \sum_{j=1}^{i} \Delta_{j-1}^j H(T) \qquad [3.2]$$

$$\Delta_0^i S = S_i(T) - S_0(T) = \sum_{j=1}^{i} \Delta_{j-1}^j S(T) \qquad [3.3]$$

$$\Delta_0^i C_P = C_{P,i}(T) - C_{P,0}(T) = \sum_{j=1}^{i} \Delta_{j-1}^j C_P(T) \qquad [3.4]$$

where $i = 1, 2, 3, \ldots, n$ and $(n+1)$ is the number of the states; $K_{0,i}$ is the equilibrium constant between state i and the folded native state, $\Delta_0^i H$ is the enthalpy difference between state i and the folded native state, $\Delta_0^i S$ is the entropy difference between state i and the folded native state, and $\Delta_0^i C_P$ is the heat capacity difference between state i and the folded native state.

The relationship between heat capacity and enthalpy and entropy are:

$$C_P = \frac{dH}{dT} \qquad [3.5]$$

$$C_P = T\frac{dS}{dT} \qquad [3.6]$$

Equation 3.5 and 3.6 also apply to changes in enthalpy and entropy by substituting ΔC_p, ΔH, and ΔS:

$$\Delta C_P = \frac{d\Delta H}{dT} \qquad [3.7]$$

$$\Delta C_P = T\frac{d\Delta S}{dT} \qquad [3.8]$$

By integrating equation 3.7 and 3.8, we can calculate the change in enthalpy and entropy of the sequential scheme depicted in figure 3.1 as:

$$\Delta_0^i H(T) = \Delta_0^i H(T_0) + \int_{T_0}^{T} \Delta_0^i C_P(T) dT \qquad [3.9]$$

$$\Delta_0^i S(T) = \Delta_0^i S(T_0) + \int_{T_0}^{T} \Delta_0^i C_P(T) d\ln T \qquad [3.10]$$

where T_0 is the reference temperature

The corresponding partition function is:

$$Q = 1 + \sum_{i=1}^{n} K_{0,i} \qquad [3.11]$$

$$K_{0,i} = e^{\left(-\frac{\Delta_0^i H}{RT} + \frac{\Delta_0^i S}{R}\right)} \qquad [3.12]$$

where $K_{0,i}$ is the equilibrium constant between state 0 (folded state) to a given state i.

And the population of states is:

$$P_i = \frac{K_{0,i}}{Q} \quad ; P_0 + \sum_i^n P_i = 1 \qquad [3.13]$$

where P_0 is the population of the folded state.

The mean excess enthalpy of the system in relation to the initial state can be presented as:

$$\langle \Delta H(T) \rangle = \sum_{i=1}^{n} \Delta_0^i H \times P_i \qquad [3.14]$$

Following the equation 3.7, the excess heat capacity is therefore:

$$\langle \Delta C_P(T) \rangle = \frac{d\langle \Delta H \rangle}{dT} = \sum_{i=1}^{n} \Delta_0^i H \frac{dP_i}{dT} + \sum_{i=1}^{n} P_i \Delta_0^i C_P \qquad [3.15]$$

The calorimetric heat capacity of the protein is the summation of the heat capacity of the folded state, $C_{P,0}$, and the excess heat capacity:

$$C_{p\,cal} = C_{p,0} + \sum_{i=1}^{n} \Delta_0^i H \frac{dP_i}{dT} + \sum_{i=1}^{n} P_i \Delta_0^i Cp \qquad [3.16]$$

where:

$$\frac{dP_i}{dT} = \frac{K_{0,i}}{Q} \frac{\Delta_0^i H}{RT^2} - \frac{K_{0,i}}{Q^2} \sum_{j=1}^{n} K_{0,j} \frac{\Delta_0^j H}{RT^2} \qquad [3.17]$$

The heat capacity difference between the initial folded state 0 and any given state *i* can be calculated by assuming that the changes of the enthalpy and heat capacity are proportional for all the transitions (*6*):

$$\Delta_0^i C_P = \frac{\Delta_0^i H(T_0)}{\Delta_0^n H(T_0)} \Delta_0^n C_P \qquad [3.18]$$

The experimental heat capacity of the PBX-HD in 20mM Sodium Phosphate pH 6 was then fitted with equation 3.16, varying the number of realizable states and determining their thermodynamic parameters. A MATLAB program was written to minimize the following function:

$$\chi^2 = \sum \left(\frac{C_{P\exp} - C_{P cal}}{C_{P\exp} error} \right)^2 \qquad [3.19]$$

where Cp_{exp} is the experimental heat capacity of the protein obtained from equation 1.9, Cp_{cal} is the calculated heat capacity obtained from equation 3.16, and the Cp_{exp} *error* is the experimental error of the protein heat capacity calculated from equation 2.1.

Equation 3.16 and 3.18 show that one of the main problems of determining the thermodynamic parameters of the unfolding process is in specifying the ΔC_P values, the differences in heat capacity between the native, intermediate and unfolded states of the protein over the temperature range of interest. Different ways of specifying these heat capacity baselines gave different pictures of the unfolding process.

3.1.1 Two-state Linear Baseline

The simplest model contains two states with linear baselines for the heat capacities of the native and unfolded states. There are several options for specifying the baselines. In Model 1A, the baseline for the native state is fitted linearly to the initial slope of the protein heat capacity and the unfolded baseline is calculated by summing up the heat capacities of the amino acid residues constituting its polypeptide chain (*17*). The calculated heat capacity baseline for the unfolded state is:

$$C_P\,Unfolded = \begin{bmatrix} 19.741 - 1.833\times10^{-1}\times T + 6.623\times10^{-4}\times T^2 \\ -7.508\times10^{-7}\times T^3 \end{bmatrix} kcal\ K^{-1}\,mol^{-1} \quad [3.20]$$

where T is the temperature in Kelvin

With this model, there were five parameters being fitted: $\Delta^1_0 H(T_0)$, $\Delta^1_0 S(T_0)$, a (the intercept of the native baseline), b (the slope of the native baseline), and Y_o (heat capacity offset). Due to a number of factors such as different protein preparations, vaporization of buffer, fluctuating room temperature and other environmental conditions when the DSC measurements were taken, and the fact that the physical properties of the protein solution (such as the *PSV* of the protein, the heat capacity of the buffer, and the density of the buffer which are needed to calculate the heat capacity in equation 1.9) and the calculated unfolded heat capacity baseline itself (*17*) (equation 3.20) cannot be possibly determined accurately, then an extra parameter, Y_o (heat capacity offset),

has to be added to the experimental protein heat capacity to correct for it. In this fitting, the heat capacity offset is Y_o = 0.145 kcal/mol K (~3% in magnitude of the calculated unfolded heat capacity baseline), which is reasonably small.

The experimental heat capacity fitted quite well with the calculated one (see figure 3.2), with a melting temperature, T_m, defined as the temperature when the folded and unfolded state population is 50%, of 45.3°C, $\Delta_0^1 H(T_m)$ of 32.5 kcal/mol, and $\Delta_0^1 S(T_m)$ of 0.102 kcal/mol K.

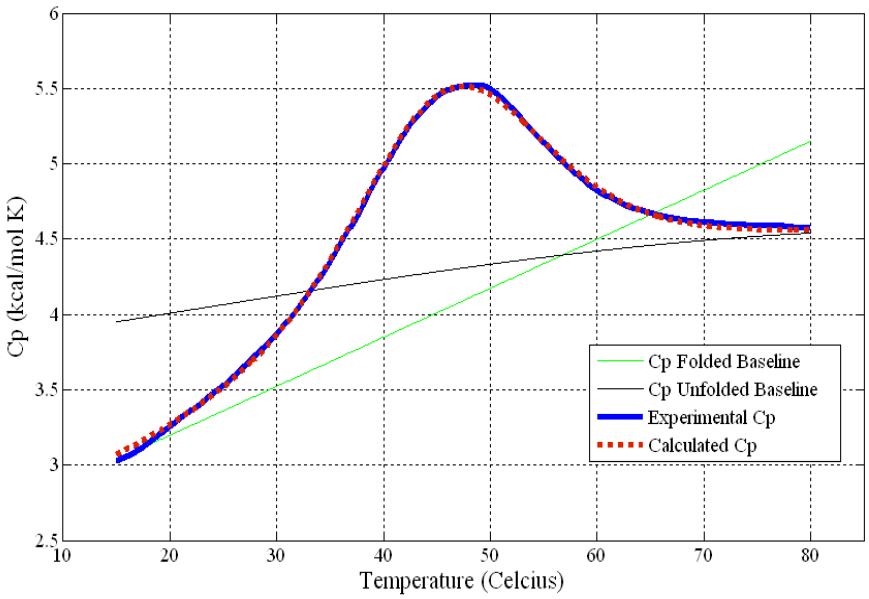

Figure 3.2. Two-state fitting of PBX-HD in 20mM Sodium Phosphate pH 6 buffer, with fitted linear folded baseline and calculated unfolded baseline (Model 1A)

The green line represents the heat capacity of the folded state which is fitted linearly, the black line represents the calculated heat capacity of the unfolded state, the blue line is the experimental heat capacity of PBX-HD, and the dashed red line is the calculated heat capacity. The thermodynamic parameters for this fit are: T_m 45.3°C, $\Delta_0^1 H(T_m)$ 32.5 kcal/mol, and $\Delta_0^1 S(T_m)$ 0.102 kcal/mol K. The residual χ^2 is 3066.

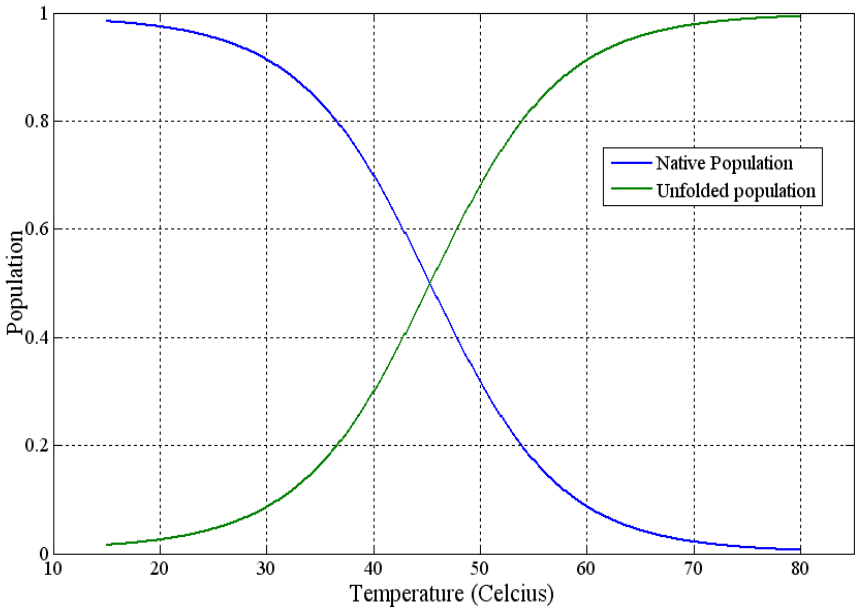

Figure 3.3. Population profile of PBX-HD unfolding in 20mM Sodium Phosphate pH 6, with fitted linear folded baseline and calculated folded baseline (Model 1A)

The green line is the population curve of the folded state and the blue line is the population curve of the unfolded state.

In the second method, model 1B, both native and unfolded baselines are fitted linearly, leading to six adjustable parameters in the minimization function: $\Delta^1_0 H(T_0)$, $\Delta^1_0 S(T_0)$, a (the intercept of the native baseline), b (the slope of the native baseline), g (the intercept of the unfolded baseline), and h (the slope of the unfolded baseline). Since the unfolded baseline is fitted, there is no need to fit the heat capacity offset parameter.

With this model, the experimental heat capacity fitted better (see figure 3.4), with a residual χ^2 of 1500 as opposed to 3066 for the model 1A. According

to model 1B, the melting temperature is 48.0°C, $\Delta_0^1 H(T_m)$ is 32.5 kcal/mol, and $\Delta_0^1 S(T_m)$ is 0.101 kcal/mol K.

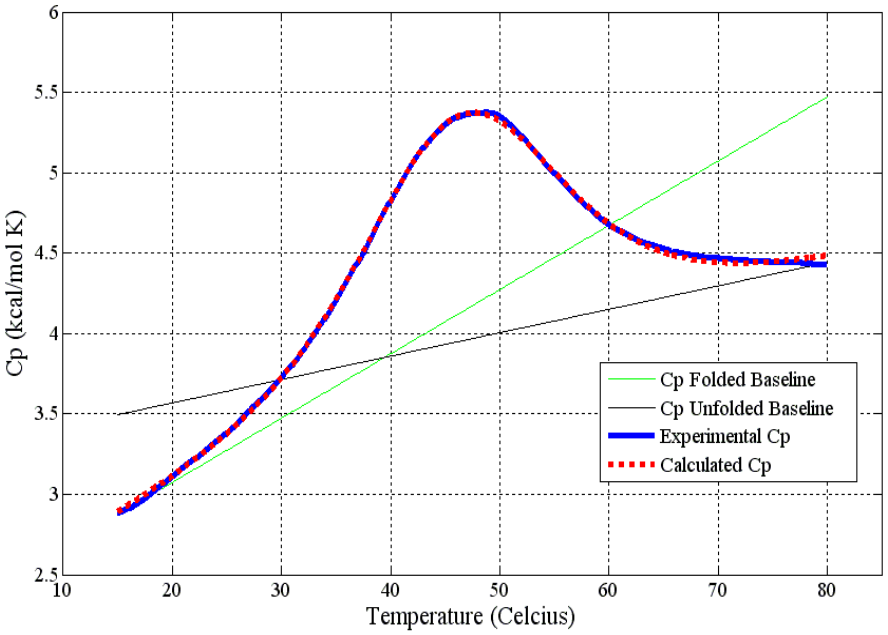

Figure 3.4. Two-state fitting of PBX-HD in 20mM Sodium Phosphate pH 6 buffer, with fitted linear baseline for the native and folded state (Model 1B) The green line represents the heat capacity of the folded state which is fitted linearly, the black line represents the calculated heat capacity of the unfolded state, the blue line is the experimental heat capacity of PBX-HD, and the dashed red line is the calculated heat capacity. The thermodynamic parameters for this fit are: T_m 48.0°C, $\Delta_0^1 H(T_m)$ 32.5 kcal/mol, and $\Delta_0^1 S(T_m)$ 0.101 kcal/mol K. The residual χ^2 is 1500.

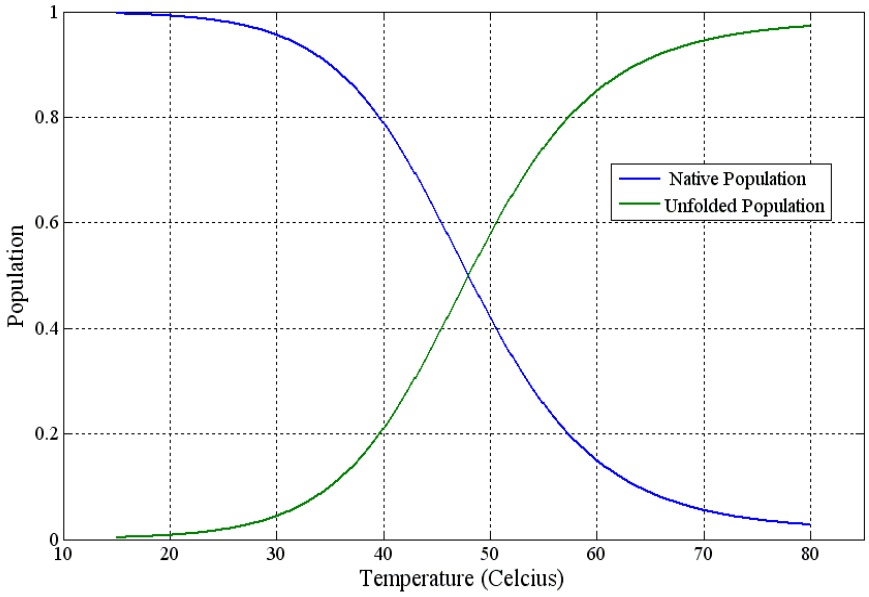

Figure 3.5. Population profile of PBX-HD unfolding in 20mM Sodium Phosphate pH 6, with fitted linear baseline for the native and folded state (Model 1B)

The green line is the population curve of the folded state and the blue line is the population curve of the unfolded state.

Qualitatively and quantitatively, there is no major difference between these two models (see below Table 1). One common characteristic of these two models is that the folded heat capacity baseline crosses the unfolded heat capacity baseline near the unfolding temperature. This means that the difference between the heat capacity of the unfolded state and the native state changes sign to become negative. Physically, this means that the unfolding process predominantly involves the exposure of polar (or hydrophilic) group from the protein core to the solvent instead of the exposure of non-polar (or hydrophobic) group; an unlikely scenario as almost all studies of protein unfolding have shown that the unfolding

of protein is governed predominantly by the exposure of hydrophobic groups from the core of the protein to the solvent.

Table 1. Thermodynamic parameters of two-state fittings of PBX-HD

	T_m (ºC)	$\Delta^1_0 H(T_m)$ (kcal/mol)	$\Delta^1_0 S(T_m)$ (kcal/mol K)
Model 1A	45.3	32.5	0.102
Model 1B	48.0	32.5	0.101

T_m is the melting temperature, defined as the temperature when the population of folded and unfolded state is 50%; $\Delta^1_0 H(T_m)$ is the enthalpy difference between the folded and unfolded state at melting temperature; $\Delta^1_0 S(T_m)$ is the entropy difference between the folded and unfolded state at melting temperature; χ^2 is the residual fit. Model 1A is the two-state model where the folded baseline is fitted linearly and the unfolded baseline is calculated from summing up the individual heat capacities of the amino acid residues of PBX-HD. Model 1B is the two state model where both the folded and unfolded baselines are fitted linearly.

The baseline crossing is due to the steep initial slope which leads to an extrapolated native baseline that crosses the unfolded baseline. Many similar DNA-binding proteins exhibit similar characteristic large initial slopes, for example the DNA binding domain of Sox-5 (*18*) and the leucine zipper of GCN4, bZIP (*19*). This characteristic is markedly different from compact globular proteins that have rigid structures, such as barnase and ubiquitin (*20,21*).

Figure 1.5 that depicts the heat capacity of unfolding lysozyme is an example of a typical heat capacity curve of a well-behaved compact globular protein that undergoes two-state unfolding process. Here, the linear extrapolation

of the heat capacity baselines of the folded state (black dashed line) and the unfolded state (red line) do not cross each other in the whole considered range of unfolding temperature. In other words, the difference in heat capacity of the unfolded state and folded state, ΔC_p, is positive in the unfolding temperature, physically signifying the exposure of hydrophobic region from the core of the protein to the solvent as the protein unfolds. PBX-HD, on the other hand, has a much steeper slope of the heat capacity function in the pre-denaturational temperature range. In this case, extrapolating the heat capacity baseline of the folded state from this steep slope resulted in its crossing the heat capacity baseline of the unfolded state (see figure 3.2 and 3.4, green line for the folded baseline and black line for the unfolded baseline). Therefore, it is clear that for PBX-HD protein, the linear extrapolation of the folded baseline from the initial slope may not be correct. The steep increase of heat capacity at lower temperature cannot be simply due to intensifying vibrations of the protein structure but due to the change in protein structure (*22*). The excess heat capacity at this lower temperature is likely due to a pre-denaturational conformational change in the protein (*22*).

3.1.2 Multi-state model with BPTI-normalized native baseline

Specifying the heat capacity baseline of the folded state is important for determining the number of states in the unfolding mechanism of the protein and the thermodynamic parameters that describes the protein heat capacity. As section 3.1.1 has shown, linear extrapolation of the native baseline leads to the physically

unrealistic conclusion that at some temperature the enthalpy and the entropy of the unfolding process becomes a decreasing function of temperature.

The excess heat capacity needs to be evaluated using a baseline that truly represents the folded heat capacity over the temperature range of interest. Among studied proteins, Bovine Pancreatic Trypsin Inhibitor, BPTI, has a highly stable structure because it is heavily S-S cross-linked. It is therefore sensible to use the heat capacity of BPTI as a reference for the excess heat capacities of other proteins. Using the specific heat capacity of native BPTI over the temperature range of 5-100ºC (Table 2) (*23*), a linear function was fitted (equation 3.21).

Table 2. Specific heat capacity of BPTI in the native state (*23*)

T (^{o}C)	5	25	50	75	100
C_P (J/K g)	1.325	1.447	1.600	1.736	1.888

$$C_P(T)^{BPTI} = \left[1.295 + 5.926 \times 10^{-3} \times (T - 273.15)\right] J\,K^{-1}\,g^{-1} \quad [3.21]$$

where *T* is temperature in Celsius.

Multiplying equation 3.21 with the molecular weight of PBX-HD (7438.4 g/mole), a BPTI-normalized heat capacity of the folded PBX-HD was obtained:

$$C_P(T)^{PBX-HD} = \left[-0.576 + 10.535 \times 10^{-3} \times T\right] kcal K^{-1}\,mol^{-1} \quad [3.22]$$

where *T* is temperature in Kelvin.

Using the BPTI-normalized native baseline of PBX-HD (equation 3.22) and the calculated unfolded baseline of PBX-HD (equation 3.20), the excess heat capacity was evaluated by fitting the PBX-HD heat capacity curve with multi-

state model (Model 2A). Using the discreet-states model as outlined in chapter 3.1 (equation 3.1 to 3.18), the PBX-HD heat capacity curve was fitted with a model where there exists a number of intermediate states, i.e. 3-state, 4-state, 5-state model, and so on.

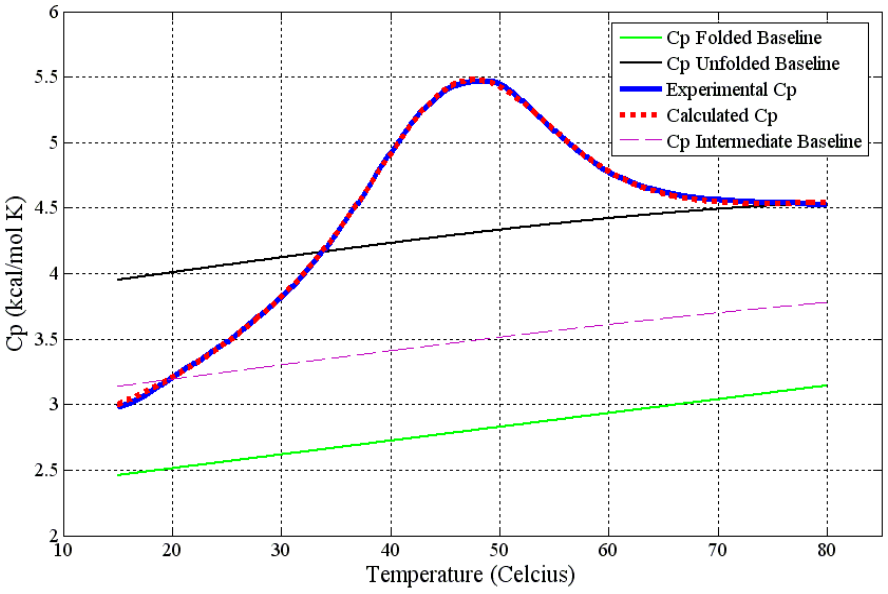

Figure 3.6. Three-state fitting of PBX-HD in 20mM Na-Phosphate pH 6 (Model 2A)

The green line represents the BPTI-normalized heat capacity of the folded state, the black line represents the calculated heat capacity of the unfolded state, the dashed purple line is the heat capacity of the intermediate state, the blue line is the experimental heat capacity of PBX-HD, and the dashed red line is the calculated heat capacity. The thermodynamic parameters for this fit are: T_m 43.4 ºC, $\Delta_0^1 H(T_m)$ 26.8 kcal/mol, $\Delta_0^1 S(T_m)$ 0.0908 kcal/mol K, $\Delta_0^2 H(T_m)$ 58.7 kcal/mol, and $\Delta_0^2 S(T_m)$ 0.192 kcal/mol K. The residual χ^2 is 738.

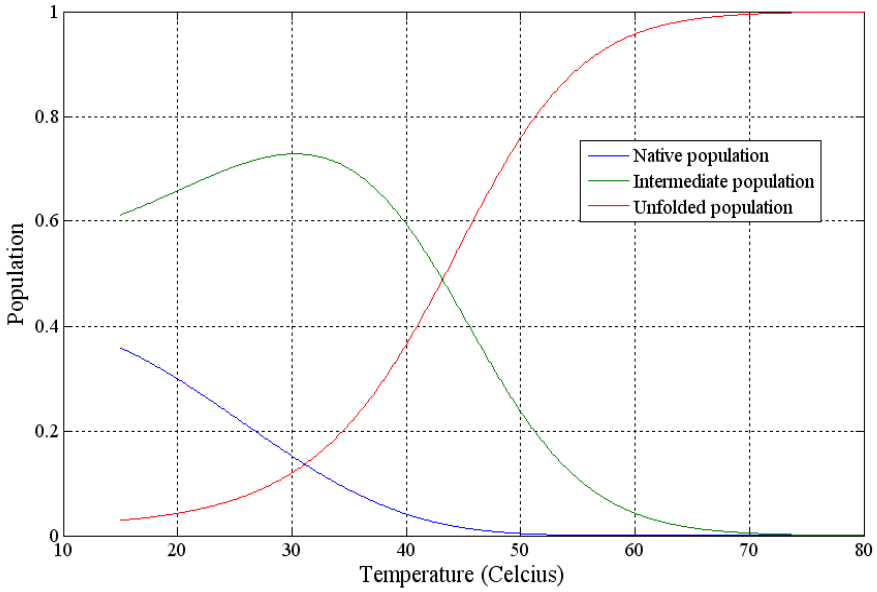

Figure 3.7. Population of three-state unfolding of PBX-HD in 20mM Sodium Phosphate pH 6 (Model 2A)

The blue line is the population of the folded state, the green line is the curve of the intermediate state, and the red line is the population of the unfolded state.

The best fit was found to be three-state model (see figure 3.6), with T_m of 43.4 ºC, $\Delta^1_0 H(T_m)$ of 26.8 kcal/mol, $\Delta^1_0 S(T_m)$ of 0.0908 kcal/mol K, $\Delta^2_0 H(T_m)$ of 58.7 kcal/mol, and $\Delta^2_0 S(T_m)$ of 0.192 kcal/mol K. T_m is taken to the temperature at which the protein is 50% unfolded.

With this model 2A, the excess heat capacity at lower temperature is accounted for. This pre-denaturational change takes place at physiological temperature as the intermediate state is highly populated at that temperature. The transition from native state to intermediate state at this low temperature is an endothermic transition, hence this conformational change in the structure of the protein accounts for the excess heat capacity at low temperature (see figure 3.7).

The model 2A above assumes that the unfolding process takes place in a discreet sequential manner. A different process is possible where the unfolding of PBX-HD proceeds through independent transitions (*14*). Let us consider that the protein has multiple independent domains which can be either in a native state or denatured state; let us call this model independent-transition model or model 2B. Independently, each block can be represented as such:

$$N_i \underset{}{\overset{K_i}{\rightleftharpoons}} U_i$$

Figure 3.8. Independent domains of protein

i = 1, 2, 3, … n where n is the number of independent domains of a protein with a total of 2^n different states. Each of the domain can be either in a native state (N) or denatured state (U) independently of other domains.

The thermodynamic representation of above independent transition model is:

$$\Delta H_{Total} = \sum_{i=1}^{n} \Delta H_i(T) \qquad [3.23]$$

$$\Delta S_{Total} = \sum_{i=1}^{n} \Delta S_i(T) \qquad [3.24]$$

$$\Delta C_{P,Total} = \sum_{i=1}^{n} \Delta C_{P,i}(T) \qquad [3.25]$$

where the total difference in enthalpy, entropy, and heat capacity between the folded state of the protein and the unfolded state of the protein is simply the summation of each of the corresponding thermodynamic parameters of each of the domain *i*.

We can evaluate the difference of enthalpy and entropy of each of the domains according to:

$$\Delta H_i(T) = \Delta H_i(T_0) + \int_{T_0}^{T} \Delta C_{P,i}(T) dT \qquad [3.26]$$

$$\Delta S_i(T) = \Delta S_i(T_0) + \int_{T_0}^{T} \Delta C_{P,i}(T) d \ln T \qquad [3.27]$$

where $\Delta C_{p,i}$ is the heat capacity difference between the folded and unfolded state of domain *i*.

The corresponding partition function for each of the domains is:

$$Q_i = 1 + K_i \qquad [3.28]$$

$$K_i = e^{\left(-\frac{\Delta H_i}{RT} + \frac{\Delta S_i}{R}\right)} \qquad [3.29]$$

where K_i is the equilibrium constant between the folded and unfolded state of each domain *i*. And the population of each of the i^{th} independent domains which are unfolded:

$$P_i = \frac{K_i}{Q_i} \qquad [3.30]$$

Taken from equation 3.15 and 3.16, the heat capacity for each of the independent transitions is:

$$C_{P,i} = C_{P,0} + \Delta H_i \frac{dP_i}{dT} + P_i \Delta C_{P,i} \qquad [3.31]$$

where:

$$\frac{dP_i}{dT} = \frac{K_i}{Q} \frac{\Delta H_i}{RT^2} - \frac{K_i}{Q_i^2} K_i \frac{\Delta H_i}{RT^2} \qquad [3.32]$$

and $C_{P,0}$ is the heat capacity of the folded state of the protein, $\Delta C_{P,i}$ is the heat capacity difference between the folded and unfolded state of domain *i*, *R* is the gas constant, and *T* is the temperature.

Therefore, the total heat capacity is just the summation of the individual heat capacities from each transition:

$$C_P = \sum_{i=1}^{n} C_{P,i} - \left[(n-1)C_{P,0}\right] \qquad [3.33]$$

The heat capacity difference between the folded state and unfolded state of each of domain *i* can be calculated by assuming that the changes of the enthalpy and heat capacity are proportional for all the independent domains (*6*):

$$\frac{\Delta C_{P,i}}{\Delta C_{P,Total}} = \frac{\Delta H_i}{\Delta H_{Total}} \qquad [3.34]$$

where $\Delta C_{p,i}$ is the heat capacity difference between the folded and unfolded state of domain *i*, $\Delta C_{p\ Total}$ is the difference in heat capacity of the folded protein and unfolded protein, ΔH_i is the enthalpy difference between the folded state and

unfolded state of each domain *i*, and ΔH_{Total} is the enthalpy difference between the folded protein and unfolded protein.

A model with two independent transitions (model 2B) fits the data very well with a residual χ^2 comparable to the three-state model, 745 as opposed to 738 for the three-state model (see figure 3.9), with T_m of 43.4 ºC, $\Delta H_1(T_m)$ of 26.6 kcal/mol, $\Delta S_1(T_m)$ of 0.0902 kcal/mol K, $\Delta H_2(T_m)$ of 32.2 kcal/mol, and $\Delta S_2(T_m)$ of 0.102 kcal/mol K. T_m is taken to be the temperature at which both domains are unfolded in 50% of the protein molecules.

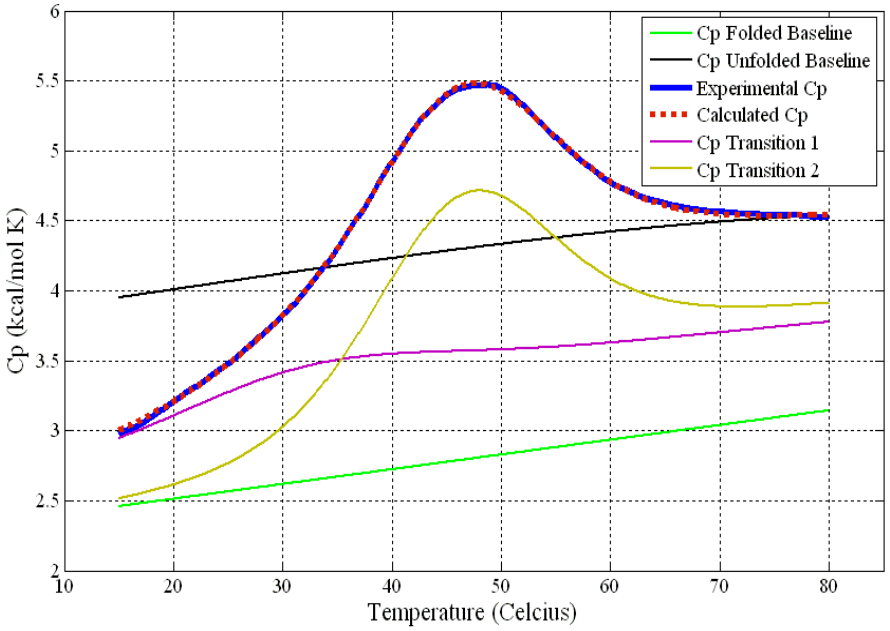

Figure 3.9. Two independent transition fit of PBX-HD in 20mM Sodium Phosphate pH 6 (Model 2B)

The green line represents the BPTI-normalized heat capacity of the folded state, the black line represents the calculated heat capacity of the unfolded state, the purple line is the heat capacity of the transition of the first domain, the brown line is the heat capacity of the transition of the second domain, the blue line is the

experimental heat capacity of PBX-HD, and the dashed red line is the calculated heat capacity which is the summation of the heat capacity first and second transition. The thermodynamic parameters for this fit are: $\Delta H_1(T_m)$ 26.6 kcal/mol, $\Delta S_1(T_m)$ 0.0902 kcal/mol K, $\Delta H_2(T_m)$ 32.2 kcal/mol, and $\Delta S_2(T_m)$ 0.102 kcal/mol K. The residual χ^2 is 745.

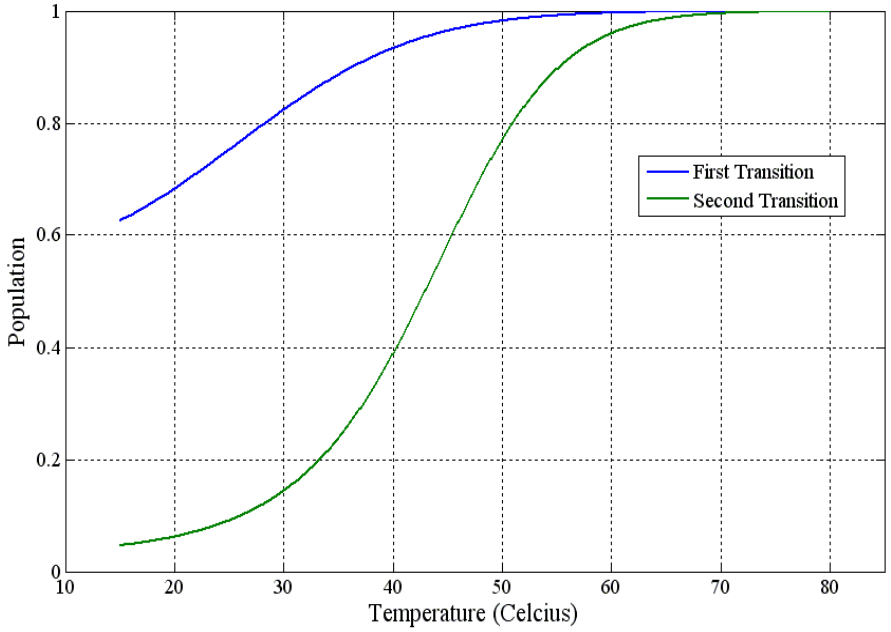

Figure 3.10. Population of the two independent transitions of PBX-HD in 20mM Sodium Phosphate pH 6 (Model 2B)

The blue line is the population curve for the transition of the first domain and the green line is the population curve for the transition of the second domain.

The population profile of the two-transition model (see figure 3.10) shows that the transition of the second domain (green line) only starts to take place after a significant fraction of the first domain unfolded (more than 50%). From this observation, it can be safely assumed that the second transition doesn't happen independently from the first transition; in other words, the second domain unfolds only after the first domain has unfolded. The scheme below (figure 3.11) shows

that of all four possible states in a two-independent transition model, only three states were realized.

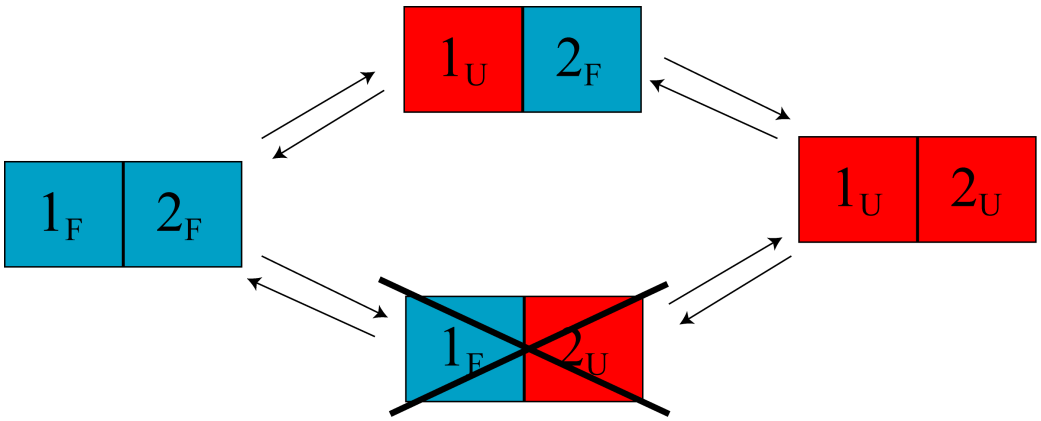

Figure 3.11 Two-independent transition scheme of PBX-HD.
In this model, PBX-HD has two independent domains, 1 and 2, that can unfold independently of each other. When the block is blue, it is folded; when it is red, it is unfolded. The fitting of the PBX-HD heat capacity to the two-independent transition model suggests that of all four possible states, only three states are populated because the second transition does not take place independently from the first transition.

Therefore, the population of unfolded protein (where both domains are unfolded) can be calculated as the product of the first transition and second transition (equation 3.35). The population of the intermediate state (where the first domain is unfolded) will be the population of the first transition minus the population of the second transition (equation 3.36). Finally, the population of the folded protein will simply be the remaining population (equation 3.37).

$$P_{Unfolded} = P_{Transition\ 1} \times P_{Transition\ 2} \qquad [3.35]$$

$$P_{Intermediate} = P_{Transition\ 1} - P_{Transition\ 2} \quad [3.36]$$

$$P_{Folded} = 1 - P_{Unfolded} - P_{Intermediate} \quad [3.37]$$

From the above calculation, the populations of folding intermediates predicted by the two-transition model (2B) are similar to those of the three-state model (2A) (see figure 3.7 and figure 3.12).

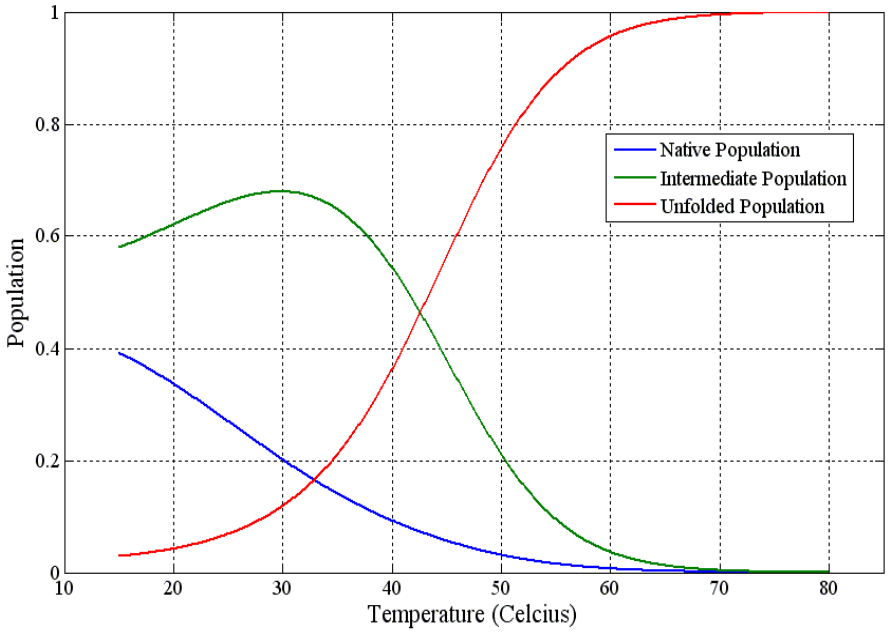

Figure 3.12. Three-state representation of the population profile of two-independent state unfolding of PBX-HD in 20mM Sodium Phosphate pH 6
This population was recalculated from figure 3.10 using equation 3.35, 3.36, and 3.37. The blue line is the population of the folded state, the green line is the curve of the intermediate state, and the red line is the population of the unfolded state.

Furthermore, the thermodynamic parameters of the two-transition model 2B agree very well with the three state model 2A. Based on equation 3.23 and

3.24, the total enthalpy and entropy difference, ΔH_{Total} and ΔS_{Total}, between the folded and unfolded state of the protein is simply the summation of all the ΔH and ΔS of each domain i respectively. This gives $\Delta H_{Total}(T_m)$ of 58.8 kcal/mol and $\Delta S_{Total}(T_m)$ of 0.192 kcal/mol K for the two-transition model 2B. Whereas in the three state model 2A, the enthalpy and entropy difference between the folded and unfolded state, $\Delta_0^2 H(T_m)$ and $\Delta_0^2 S(T_m)$, are 58.7 kcal/mol and 0.192 kcal/mol K respectively. Both models also give the same T_m of 43.4 ºC, the temperature when the population of the unfolded state of the protein is 50%. Below table 3 is the summary table comparing the thermodynamic parameters of both models.

Table 3. Thermodynamic parameters of three state model and two transition model

	Enthalpy (kcal/mol)	Entropy (kcal/mol K)	T_m (ºC)	χ^2
Model 2A: Three states model	$\Delta_0^1 H(T_m) = 26.8$ $\Delta_0^2 H(T_m) = 58.7$	$\Delta_0^1 S(T_m) = 0.0908$ $\Delta_0^2 S(T_m) = 0.192.$	43.4	738
Model 2B: Two transition model	$\Delta H_1(T_m) = 26.6$ $\Delta H_2(T_m) = 32.2$ $* \Delta H_{Total}(T_m) = 58.8$	$\Delta S_1(T_m) = 0.0902$ $\Delta S_2(T_m) = 0.102$ $* \Delta S_{Total}(T_m) = 0.192$	43.4	745

$* \Delta H_{Total}(T_m) = \Delta H_1(T_m) + \Delta H_2(T_m)$ and $\Delta S_{Total}(T_m) = \Delta S_1(T_m) + \Delta S_2(T_m)$

These results from two separate models both suggest that, within the limit of discreet-state model with BPTI-normalized baseline, PBX HD unfolds in a three-state manner

3.2. Variable-barrier model

Classically, the unfolding process of a protein has been described as a series of chemical equilibria between *n* structurally defined macrostates, i.e. cooperative folding. The discrete state model already assumes the existence of an energy barrier between the realizable states. Theoretically, however, a protein can fold non-cooperatively, giving rise to a downhill folding where there exists no energy barrier. Instead of a set of states with well-defined structure, we have a continuum of states.

Muñoz and Sanchez-Ruiz developed a model that is free of assumptions about the existence of a free energy barrier (*7*). In this model, the height of the energy barrier is not presumed but is obtained directly from the experimental data.

The protein folding is described as a continuous distribution of protein states where the partition function is written as:

$$Q = \int \rho(H) \exp\left(-\frac{H}{RT}\right) dH \qquad [3.38]$$

in which the system is described as an ensemble of enthalpy microstates, where H is defined as an enthalpy scale and $\rho(H)$ is the density of the enthalpy microstates. From this equation, the probability density of finding the protein in a microstate of enthalpy H at a given temperature T is:

$$P(H \mid T) = C \cdot P(H \mid T_0) \cdot \exp(-\lambda H) \quad [3.39]$$

where

$$\lambda = \frac{1}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right) \quad [3.40]$$

$$P(H \mid T_0) = C' \cdot \exp\left(-\frac{G_0}{RT_0}\right) \quad [3.41]$$

$$G_0 = -2\beta\left(\frac{H}{\alpha}\right)^2 + |\beta|\left(\frac{H}{\alpha}\right)^4 \quad [3.42]$$

where T_0 is the "characteristic" temperature, R is the gas constant, C is a constant determined by the normalization condition $\left[\int P(H/T)dH = 1\right]$, C' is a constant resulting from the normalization condition $\left[\int P(H/T_0)dH = 1\right]$, and $G_0(H)$ is the free energy functional where α and β are parameters that describe the shape of the free energy function. The β parameter physically determines the height of the energy barrier. The sign of the β parameter determines the observation of two macrostates or a single macrostates. For $\beta > 0$, there are two macrostates, i.e. the folding/unfolding system is cooperative. If $\beta < 0$ or it takes a very small positive value, it is essentially equivalent to non-cooperative or downhill folding.

Finally, the excess heat capacity with reference to the native state can be expressed as:

$$C_P^{Ex} = \frac{d\langle H \rangle}{dT} = \frac{\langle H^2 \rangle - \langle H \rangle^2}{RT^2} \qquad [3.43]$$

where:

$$\langle H^n \rangle = \int H^n P(H \mid T) dH \qquad [3.44]$$

From here, the calculated heat capacity can be obtained (equation 3.45) and a minimization function for the residual χ^2 (equation 3.46) was run on MATLAB program to fit the experimental heat capacity and the calculated heat capacity.

$$C_{P\,cal} = C_{P,0} + C_P^{Ex} \qquad [3.45]$$

$$\chi^2 = \left(\frac{C_{P\exp} - C_{P\,cal}}{C_{P\exp} error} \right)^2 \qquad [3.46]$$

where $C_{P,cal}$ is the calculated heat capacity, $C_{P,0}$ is the heat capacity baseline for the folded state, C_P^{Ex} is the calculated excess heat capacity, $C_{P,exp}$ is the experimental heat capacity from DSC, and $C_{P,exp}$ *Error* is the uncertainties of the heat capacity measurement.

An extra modification to the free energy function (equation 3.43) has to be added. The inherent symmetry of equation 3.43 implies that for large positive β values (i.e. cooperative folding with two macrostates) the two energy wells have the same shape, i.e. they have the same heat capacity. However, this is at odds

with the typical situation of protein folding where the unfolded state (high-enthalpy macrostate) has a higher heat capacity than the folded state (low-enthalpy macrostate). To account for the asymmetry of protein folding, α parameter for negative values of H (α_N) and α parameter for positive values of H (α_P) are introduced, where they can be expressed in terms of the parameters $\Sigma\alpha$ and f by the following relationship.

$$\sum \alpha = \alpha_N + \alpha_P \qquad [3.47]$$

$$\alpha_N = \frac{\sum \alpha \cdot f}{2} \qquad [3.48]$$

where $\Sigma\alpha$ roughly corresponds to the difference in enthalpy between the two wells found at low and high temperature, f is an "asymmetry factor: when f=1, then the heat capacity of the folded and unfolded state is the same, whereas if f<1, then the heat capacity of the folded state is lower than that of the unfolded state. In this fitting, constant f is chosen to be 0.1 as it is well within the appropriate range.

Thus, the model defined above has only three fitted parameters: β that determines the height of the energy barrier, T_0 which is the "characteristic" temperature, and $\Sigma\alpha$ that roughly corresponds to the difference in enthalpy between the two minima found at low and high temperature. In the fitting procedure, T_0 and $\Sigma\alpha$ are fitted to the DSC data of PBX-HD (using the residual χ^2 equation 3.46) using fixed β values and a constant f=0.1. This procedure is repeated for a set of β values to generate a plot of χ^2 vs. β (see figure 3.14). The plot shows that good fit (the low χ^2 value) was only obtained with a negative β

value, which means that PBX-HD unfolds in a non-cooperative or downhill manner. The best fit (χ^2 = 7600) was obtained with such parameters: β = -2.7 kcal/mol, T_0=306.8K, and $\sum\alpha$ = 131 kcal/mol; the calculated and experimental heat capacity profile of PBX-HD is shown in figure 3.13.

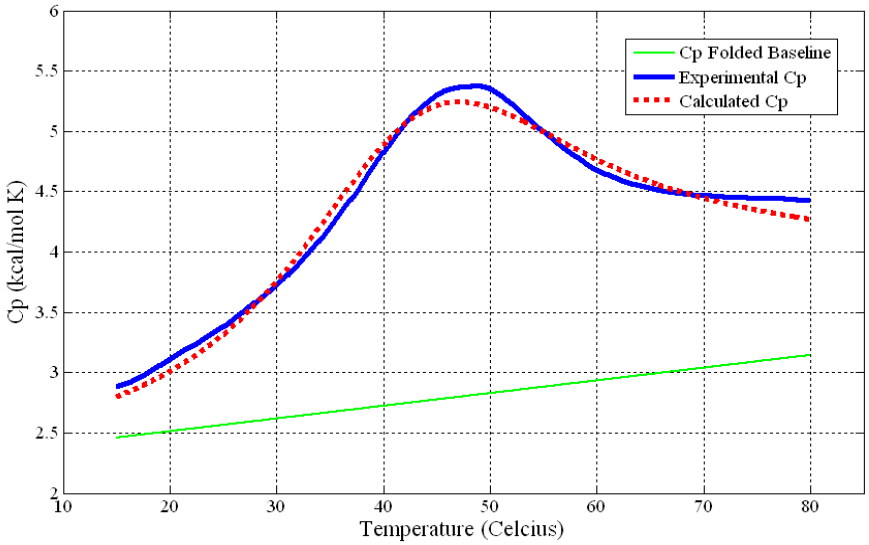

Figure 3.13. Variable-barrier fit of PBX-HD in 20 mM Sodium Phosphate pH 6

The green line is the BPTI-normalized heat capacity of folded PBX, the blue line is the experimental heat capacity, the dashed red line is the calculated heat capacity. The fitting parameters are β = -2.7 kcal/mol, T_0=306.8K, and $\sum\alpha$ = 131 kcal/mol. The residual χ^2 is 7600.

The free energy function of PBX-HD at T_0 is shown in figure 3.15; it is calculated using equation 3.42. The shape of the free energy function is clearly downhill with no energy barrier. Figure 3.16 shows the probability distribution as a function of H for PBX-HD at various temperatures; calculated using equation 3.39. The distribution was shown to be unimodal at all temperatures, with the maximum probability shifting from the low-enthalpy values to high-enthalpy

values as the temperature is increased. The variable-barrier model shows that PBX-HD unfolds non-cooperatively.

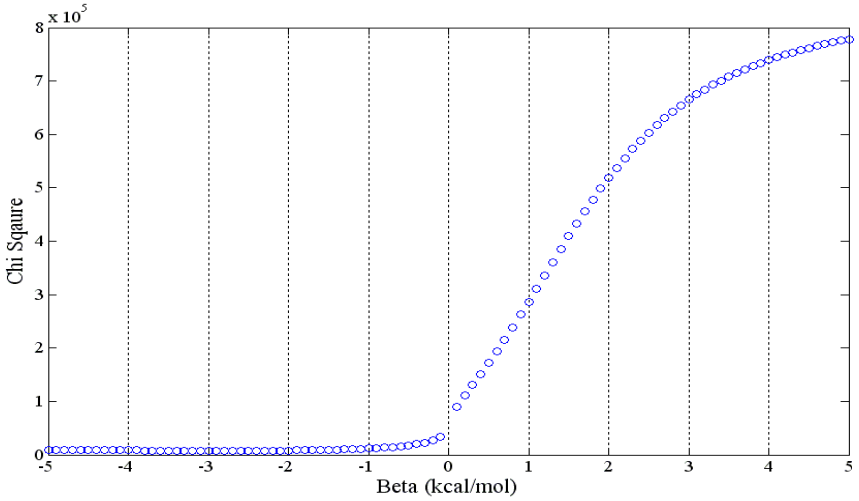

Figure 3.14. Plots of χ^2 vs. β value for the fits of the variable-barrier model of PBX-HD in 20 mM Sodium Phosphate pH 6

PBX-HD heat capacity was fitted with variable-barrier model with a set of β value from 5 to -5 with an increment of 0.1 to generate χ^2 vs. β plot. This plot shows that the best fit, i.e. low χ^2 value, was obtained when β value is negative, signifying a downhill folding.

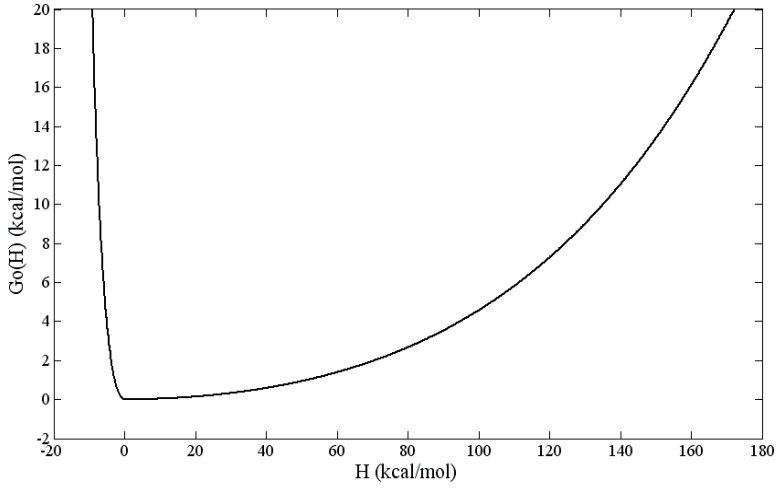

Figure 3.15. Plots of free energy [G_0(H)] versus enthalpy for PBX-HD in 20mM Sodium Phosphate pH 6

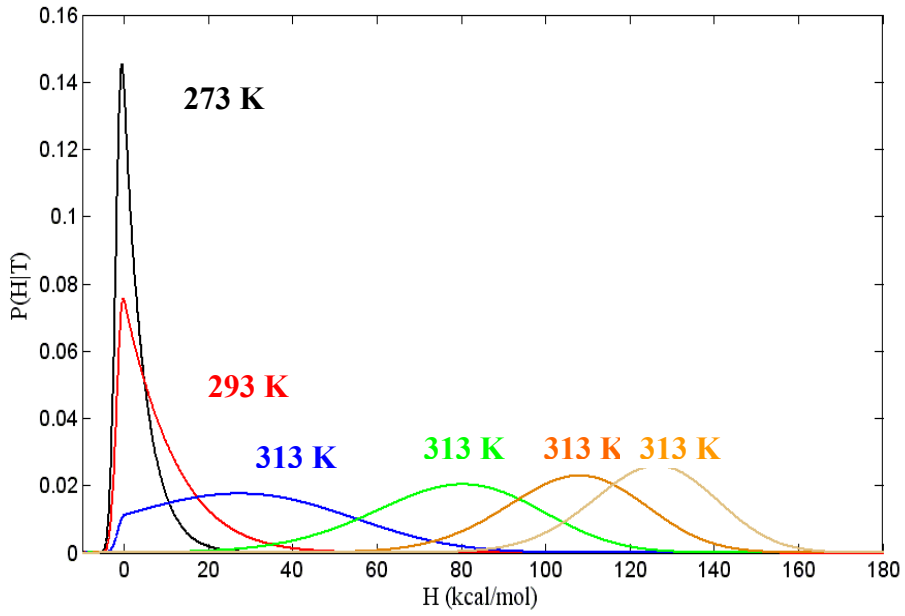

Figure 3.16. Probability distribution of PBX-HD in 20mM Sodium Phosphate pH 6 at different temperatures

3.3 DSC fitting combined with NMR data

As the previous fits (section 3.1, discreet-states models, and 3.2, variable-barrier model) have shown, the interpretation of DSC data is greatly determined by how one defines the heat capacity baseline for the folded and unfolded state. Linear extrapolation of the folded heat capacity suggests two-state folding while BPTI-normalized folded heat capacity suggests three-state folding, while variable barrier model suggests a non-cooperative one-state folding.

In the following fitting, calorimetric data from DSC is combined with the NMR results provided by Patrick Farber. The global fitting of these two different techniques provide greater confidence in the result. This is a fitting method that has never been done before. We compared temperature-dependent signals from

multiple backbone nuclei with DSC data in order to test whether the unfolding of the protein proceeds in a two-state cooperative manner. The calorimetric data were combined with two-dimensional ^{15}N-^{1}H, chemical shift information where the position of the peak is the population-weighted average chemical shift. The NMR data were acquired every 2.5ºC from 10ºC to 65ºC. The DSC heat capacity baselines were fitted to third degree polynomial functions, while the NMR baselines for both proton and nitrogen chemical shifts were fitted to a linear function.

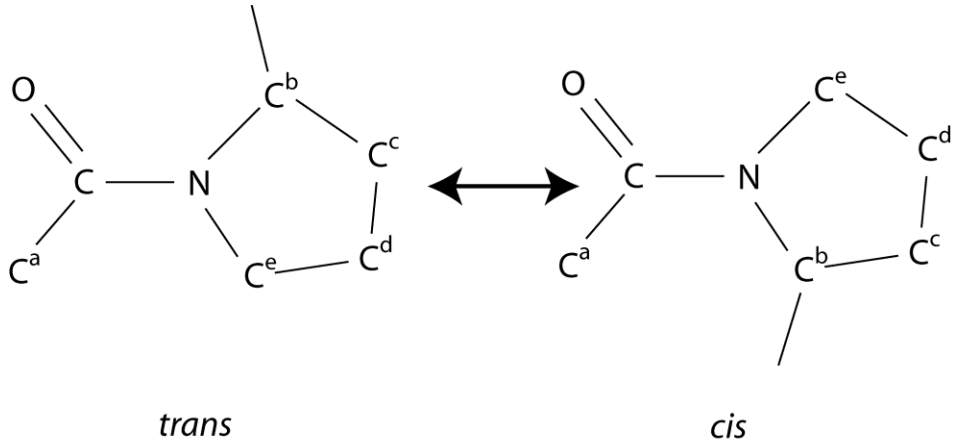

Figure 3.17. *Cis-trans* isomerization of proline

Slow exchange on the NMR timescale was found to occur at higher temperatures (40-60 ºC). This slow exchange most likely corresponds to *cis-trans* isomerization of X-Pro peptide in the denatured state where X denotes any other amino acid group preceding the Proline residue (see figure 3.17). *Cis-trans* isomerization of X-Pro peptide takes place on a slower time scale than protein folding and is a rate-limiting step in the renaturation of protein (*24,25*). PBX-HD

has two proline residues: P24 and P26, and therefore contains two X-Pro bonds: Asp-Pro and Tyr-Pro respectively. When PBX-HD unfolds, these two prolines isomerize to form four *cis-trans* isomeric species (see figure 3.18).

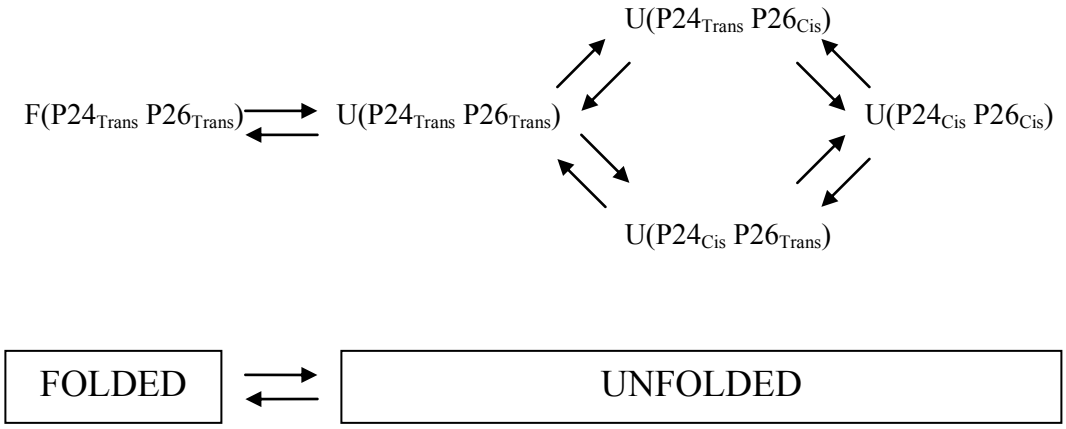

Figure 3.18. *Cis-trans* isomerization of the two proline residues of PBX-HD

All these transitions can be represented by a linear scheme of 5-state transition:

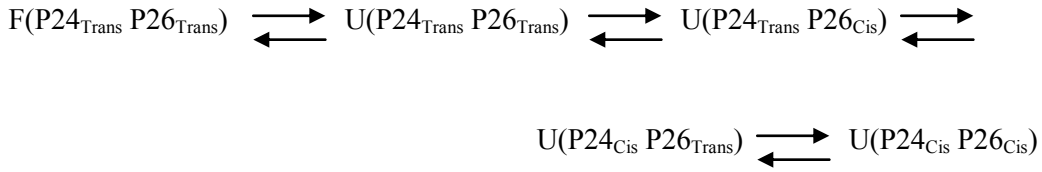

Figure 3.19. 5-state scheme of *cis-trans* isomerization of PBX-HD

From the studies of short peptides (*26*), the enthalpy difference, ΔH, between the *cis* and *trans* form of Tyr-Pro at 298 K were found to be 0.65 kcal mole^{-1}, while the free energy difference, ΔG, is 0.81 kcal mole^{-1}. Similar

thermodynamic values are expected for the unfolded form of PBX-HD (*26*). Using the thermodynamic relationship of $\Delta G=\Delta H-T\Delta S$, the entropy difference of cis-trans isomerization of Tyr-Pro is calculated to be -5.4×10^{-4} kcal mole^{-1} K^{-1}.

Since in the absence of aromatic residue the X-Pro peptide bond was found to be *cis* 7-8% of the time, a value obtained from Ala-Pro measurement (*26*), then it is safe to assume that Asp-Pro, with non-aromatic Asp, will also behave the same. Assuming that the *trans* to *cis* isomerization ΔS of Tyr-Pro and Asp-Pro are similar, -5.4×10^{-4} kcal mole^{-1} K^{-1}, we can calculate the ΔH of the Asp-Pro *cis-trans* isomerization (*26*). With 7.5% *cis* population, the ΔH of Asp-Pro *trans* to *cis* isomerization was calculated to be 1.33 kcal mole^{-1} from the equation $K=\exp(-\Delta H/RT + \Delta S/R)$ where R is the gas constant and T is the temperature at 298K.

Thus, following the description of five-state model (figure 3.19) of the unfolding of PBX-HD and the thermodynamic functions of discreet-state model as outlined in section 3.1, the thermodynamic functions for the whole system are:

$$\Delta_0^1 H(T)_{DSC} = \Delta_0^1 H(T_0)_{DSC} + \int_{T_0}^{T} \Delta C_P(T) dT \qquad [3.49]$$

$$\Delta_0^2 H(T)_{DSC} = \Delta_0^1 H(T)_{DSC} + 0.65\, kcal mol^{-1} \qquad [3.50]$$

$$\Delta_0^3 H(T)_{DSC} = \Delta_0^1 H(T)_{DSC} + 1.33\, kcal mol^{-1} \qquad [3.51]$$

$$\Delta_0^4 H(T)_{DSC} = \Delta_0^1 H(T)_{DSC} + 1.98\, kcal mol^{-1} \qquad [3.52]$$

$$\Delta_0^1 S(T)_{DSC} = \Delta_0^1 S(T_0)_{DSC} + \int_{T_0}^{T} \Delta C_P(T) d \ln T \quad [3.53]$$

$$\Delta_0^2 S(T)_{DSC} = \Delta_0^1 S(T)_{DSC} + (-5.4 \times 10^{-4}) kcalmol^{-1} K^{-1} \quad [3.54]$$

$$\Delta_0^3 S(T)_{DSC} = \Delta_0^1 S(T)_{DSC} + (-5.4 \times 10^{-4}) kcalmol^{-1} K^{-1} \quad [3.55]$$

$$\Delta_0^4 S(T)_{DSC} = \Delta_0^1 S(T)_{DSC} + (-1.08 \times 10^{-3}) kcalmol^{-1} K^{-1} \quad [3.56]$$

where, as depicted in figure 3.18, the state 0 is the folded state of the protein with *trans* form on both of the proline residues, F($P24_{Trans}P26_{Trans}$); state 1 is the unfolded state with *trans* form on both of the proline residues, U($P24_{Trans}P26_{Trans}$); state 2 is the unfolded state with *cis* form on P26 residue, U($P24_{Trans}P26_{Cis}$); state 3 is the unfolded state with *cis* form on the P24 residue, U($P24_{Cis}P26_{Trans}$); and state 4 is the unfolded state with *cis* form on both of the proline residues, U($P24_{Cis}P26_{Cis}$). ΔC_P is the heat capacity difference between the folded and unfolded state of PBX-HD.

A global fitting was then performed where the calorimetric data was fitted with the above thermodynamic functions of the 5-state transition (equations 3.48 to 3.56) as outlined in the discreet-state model in section 3.1; the fast-exchange NMR chemical shift data were fitted with the first-transition thermodynamic parameters that were obtained from DSC fit, $\Delta_0^1 H(T_0)_{DSC}$ and $\Delta_0^1 S(T_0)_{DSC}$, in a two-state cooperative manner since the fast-exchange NMR data only represent the transition from the folded state F($P24_{Trans}P26_{Trans}$) to the unfolded state U($P24_{Trans}P26_{Trans}$).

$$\Delta H(T)_{NMR} = \Delta_0^1 H(T_0)_{DSC} + \int_{To}^{T} \Delta C_p(T) dT \qquad [3.57]$$

$$\Delta S(T)_{NMR} = \Delta_0^1 S(T_0)_{DSC} + \int_{To}^{T} \Delta C_p(T) d\ln T \qquad [3.58]$$

where the population can be then evaluated as such:

$$P_{1,NMR} = \frac{K_{NMR}}{1+K_{NMR}} \quad ; P_{0,NMR} = 1 - P_{1,NMR} \qquad [3.59]$$

$$K_{NMR} = e^{\left(-\frac{\Delta H_{NMR}}{RT} + \frac{\Delta S_{NMR}}{R}\right)} \qquad [3.60]$$

where, $P_{0,NMR}$ is the population of folded state F($P24_{Trans}P26_{Trans}$), $P_{1,NMR}$ is the population of the unfolded state U($P24_{Trans}P26_{Trans}$), K_{NMR} is the equilibrium constant between folded state F($P24_{Trans}P26_{Trans}$) to the unfolded state U($P24_{Trans}P26_{Trans}$).

Since the position of the NMR peaks is the population-weighted average chemical shift, the $P_{0,NMR}$ and $P_{1,NMR}$ was then fitted to both the NMR chemical shifts of nitrogen and proton separately (figure 3.21 and 3.22) in the following manner:

$$\delta^{15}N = \left(P_{0,NMR} \times B_{N,Folded}\right) + \left(P_{1,NMR} \times B_{N,Unfolded}\right) \qquad [3.61]$$

$$\delta^{1}H = \left(P_{0,NMR} \times B_{H,Folded}\right) + \left(P_{1,NMR} \times B_{H,Unfolded}\right) \qquad [3.62]$$

where δ^1N is the nitrogen chemical shift, δ^1H is the proton chemical shift, $B_{N,Folded}$ is the nitrogen baseline of the folded state, $B_{N,Unfolded}$ is the nitrogen baseline of the

unfolded state, $B_{H,Folded}$ is the hydrogen baseline of the folded state, and $B_{H,Unfolded}$ is the hydrogen baseline of the unfolded state; all four baselines were fitted as linear functions of temperature.

Furthermore, the nitrogen and hydrogen baselines for the unfolded state, $B_{N,Unfolded}$ and $B_{H,Unfolded}$, are also fitted to the slow exchange chemical shifts (red triangle in figure 3.21 and 3.22).

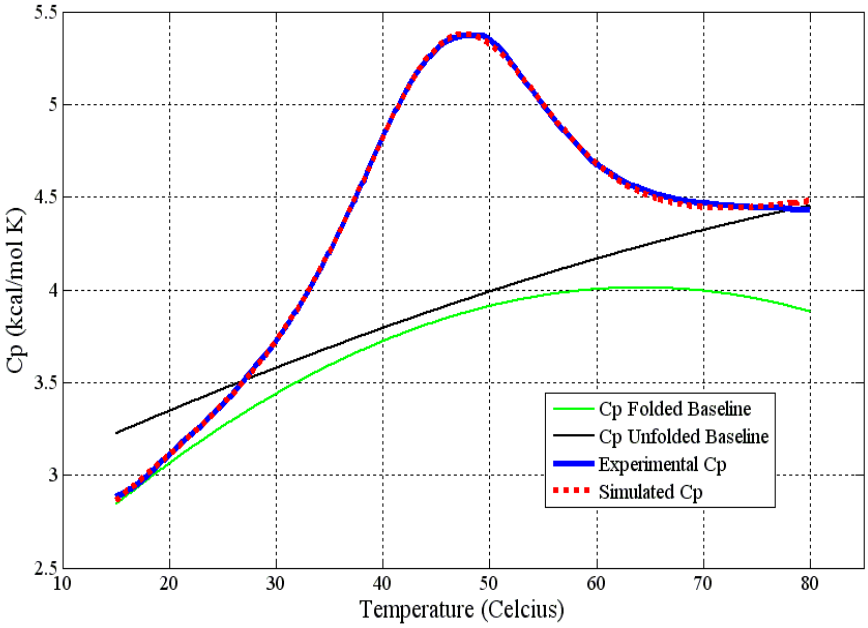

Figure 3.20. Heat capacity fit of PBX-HD in 20 mM Sodium Phosphate pH 6 combined with NMR fit

The green line represents the heat capacity of the folded state which is fitted as a 3rd degree polynomial function, the black line represents the heat capacity of the unfolded state which is fitted as a 3rd degree polynomial function, the blue line is the experimental heat capacity of PBX-HD, and the dashed red line is the calculated heat capacity. The thermodynamic parameters for this calorimetric fit are: T_m 46.5°C, $\Delta_0^4 H(T_m)$ 36.4 kcal/mol, and $\Delta_0^4 S(T_m)$ 0.106 kcal/mol K. The residual χ^2 is 931.

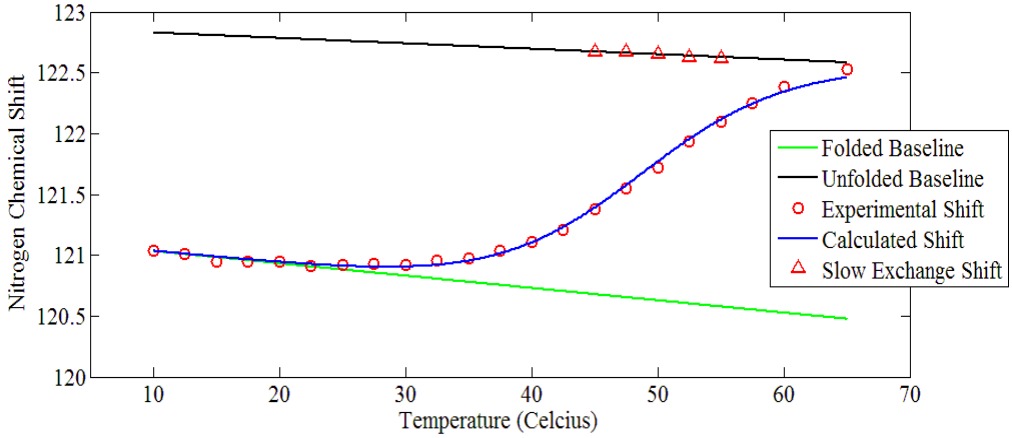

Figure 3.21. Nitrogen chemical shift fit

The red circles are the experimental ^{15}N chemical shifts, the blue line is the calculated fitted ^{15}N chemical shifts (as calculated from equation 3.59), the red triangles are the slow-exchange ^{15}N chemical shifts, the green line is the nitrogen baseline of the folded state ($B_{N,Folded}$), the black line is the nitrogen baseline of the unfolded state ($B_{N,Unfolded}$).

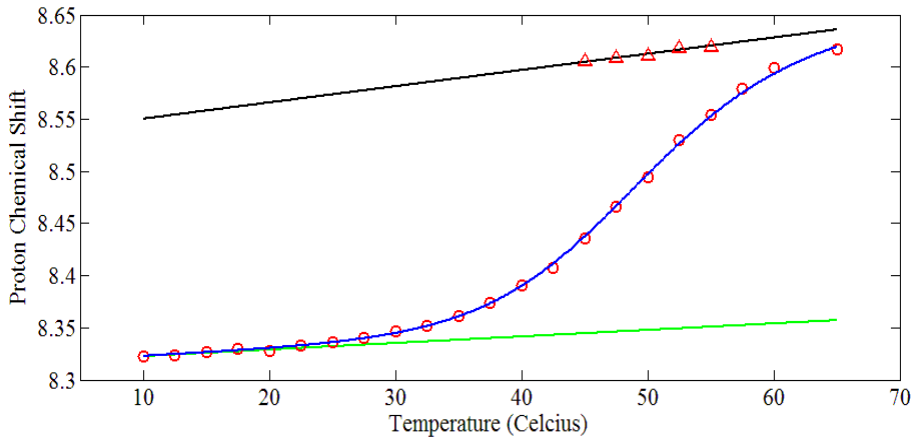

Figure 3.22. Hydrogen chemical shift fit

The red circles are the experimental ^{1}H chemical shifts, the blue line is the calculated fitted ^{1}H chemical shifts (as calculated from equation 3.60), the red triangles are the slow-exchange ^{1}H chemical shifts, the green line is the proton baseline of the folded state ($B_{H,Folded}$), the black line is the proton baseline of the unfolded state ($B_{H,Unfolded}$).

All in all, the calorimetric and NMR data were fitted simultaneously with 18 parameters: $\Delta_0^1 H(T_0)_{DSC}$ and $\Delta_0^1 S(T_0)_{DSC}$, four parameters for the third-degree polynomial function of DSC folded heat capacity baseline, another four for the third-degree polynomial function of DSC unfolded heat capacity baseline, two for each of the linear folded and unfolded baselines of the ^{15}N and ^{1}H chemical shifts. Figure 3.20 above shows that the DSC-NMR global fitting gives a result where the heat capacity baseline of the folded state curves down and does not cross the heat capacity baseline of the unfolded state. The heat capacity difference between the folded and unfolded states remain positive throughout the unfolding process, unlike the results from model 1A and 1B where it becomes negative at low temperature and thus suggests a physically unlikely scenario that the unfolding is driven by the exposure of hydrophilic groups from the core to the solvent. This DSC-NMR global fitting also gives stronger evidence that the unfolding of PBX-HD is two-state and cooperative.

Chapter 4: Further Investigations

PBX-HD in solutions of different pH and ionic strength were also investigated to study how these different conditions affect the stability and the unfolding mechanism of PBX-HD. Here the experimental results are presented. Further investigation is still needed to get a full understanding of these interactions. Calorimetric measurements were also performed on extended PBX-HD, a construct of PBX-HD with a C-terminal extension (*15*).

4.1 pH effect on the thermostability of PBX-HD

The pH of the solvent affects the ionization states of certain amino acid chains, which changes the protein charge distribution and hence its stability. It affects those residues whose p*K*as are different in the folded and unfolded states (if the p*K*as are the same, the residue does not influence the pH dependence of stability) (*27*). Figure 4.1 presents the molar heat capacity of PBX-HD in solutions of different pH: 20 mM Na-Phosphate buffer pH 1.92, 20mM Na-Phosphate buffer pH 3.02, 20mM Acetate buffer pH 5.13, 20mM Na-Phosphate buffer pH 6, and 20mM Na-Phosphate pH 7.

Due to differences in the solvents' heat capacity and also because the samples were run at different times, the different runs have offsets which are accentuated by the very low excess heat capacity of PBX-HD unfolding.

A total of 24 DSC scans of one protein preparation were done for PBX-HD in 20 mM Sodium Phosphate buffer pH 1.92, 20mM Sodium Phosphate buffer pH 3.02, and 20mM Acetate buffer pH 5.13. Each has a mean error of ~0.007 kcal $mole^{-1}$ K^{-1} or ~0.2%, ~0.004 kcal $mole^{-1}$ K^{-1} or ~0.2%, and ~0.006 kcal $mole^{-1}$ K^{-1} or ~0.2% respectively. For 20mM Na-Phosphate pH 7, five different protein preparations were done and 49 DSC scans were performed. The calculated mean error is ~0.01 kcal $mole^{-1}$ K^{-1} or ~0.3%.

From pH range of 2-7, PBX-HD is most stable at pH 6 with its molar heat capacity peak at 48.8 °C. pH 2, 3, 5, and 7 have molar heat capacity peaks at 38.3°C, 35.9°C, 46.9°C, and 43.3°C respectively.

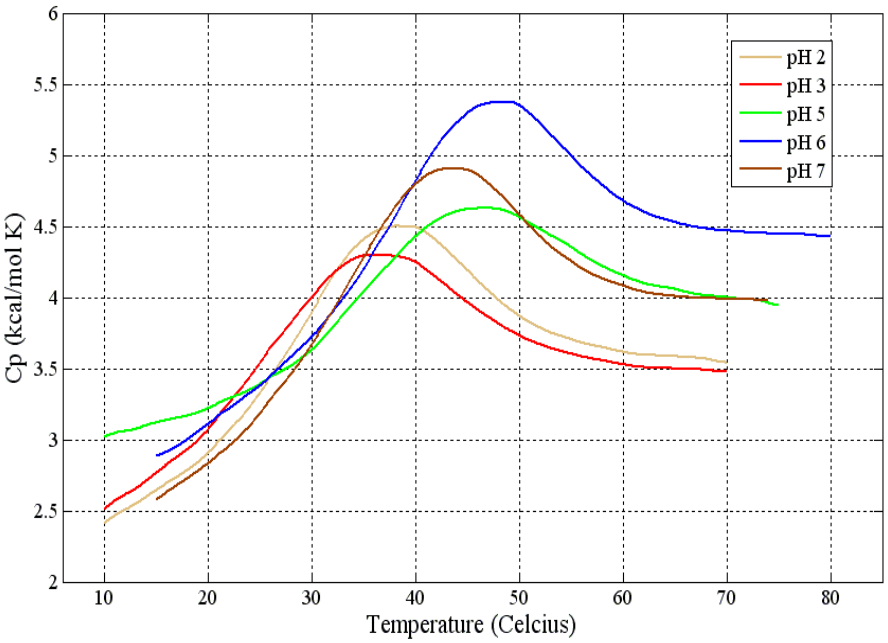

Figure 4.1. Molar heat capacity of PBX-HD in solutions of different pH
20 mM Sodium Phosphate buffer pH 1.92, 20mM Sodium Phosphate buffer pH 3.02, 20mM Acetate buffer pH 5.13, 20mM Sodium Phosphate buffer pH 6, 20mM Sodium Phosphate pH 7.

Normally, the heat capacities of proteins are measured at different pHs in order to determine ΔC_P. At different pHs, proteins have different stability. By calculating the ΔH at each melting temperature T_m, one can determine ΔC_P since it is a derivative of ΔH with respect to T (equation 1.1). In below table 4, the heat capacities of PBX-HD in different pH were fitted to model 1B, two state model with linear baselines of folded and unfolded state. The plot of $\Delta H(T_m)$ vs. T_m (see figure 4.2) does not give a good correlation from which ΔC_P can be determined. This poor correlation suggests that not only is the stability of PBX-HD affected by different pH, the unfolding mechanism of PBX-HD is also affected. Thus, the two-state model 1A was not adequate to explain the real process of denaturation of PBX-HD at all the considered pHs.

Table 4. Thermodynamic parameters of two-state linear baselines fit (model 1A) of PBX-HD at different pH

	T_m (ºC)	$\Delta H(T_m)$ (kcal/mol)	$\Delta S(T_m)$ (kcal/mol K)
20 mM Sodium Phosphate, pH 2	35.2	30.9	0.100
20 mM Sodium Phosphate, pH 3	33.9	28.0	0.0912
20 mM Acetate, pH 5	40.8	27.3	0.0870
20 mM Sodium Phosphate, pH 6	48.0	32.5	0.101
20 mM Sodium Phosphate, pH 7	42.4	32.0	0.101

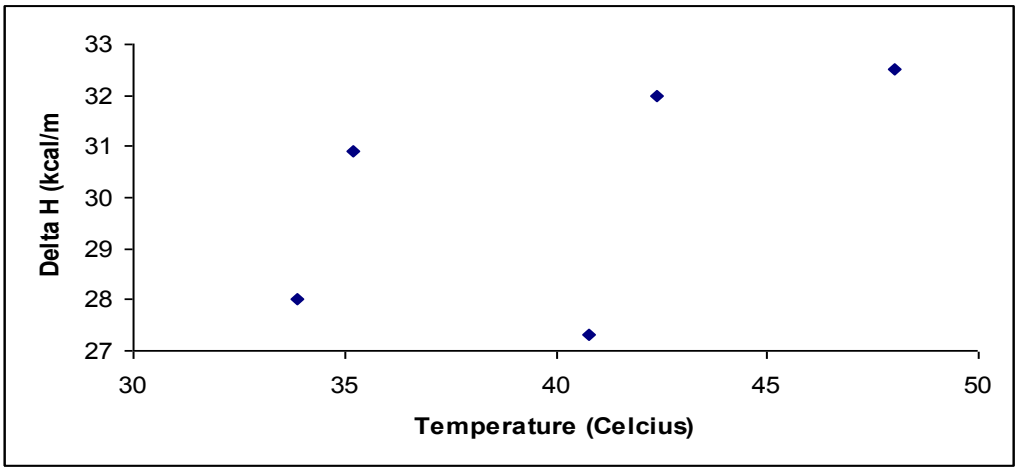

Figure 4.2. $\Delta H(T_m)$ vs. T_m plot of PBX-HD at different pH

$\Delta H(T_m)$ and T_m values from two-state fitting of PBX-HD heat capacities measured at different pH (table 4) were plotted.

4.2 Salt effect on the thermostability of PBX-HD

Ionic interactions in proteins are extremely complex because they can exist on the surface of the proteins, fully buried in the interior of the proteins, or in a partially buried environment. Generally, the effect of the ionic interaction on the stability of proteins can be probed by determining the effects of salt concentration on protein stability.

A preliminary investigation into the effect of ionic interactions on the stability of PBX homeodomain was carried out. Here, the molar heat capacity of PBX-HD was measured in the presence of 1M NaCl, both in 20 mM Sodium Phosphate pH 6 and pH 7 buffer. (see figure 4.3 and 4.4). A total of 12 DSC scans of one protein preparation were done for PBX-HD in 1M NaCl 20 mM Sodium Phosphate buffer pH 6 and 1M NaCl 20mM Sodium Phosphate buffer pH 7. Each

has an error of ~0.01 kcal mole^{-1} K^{-1} or ~0.2% and ~0.006 kcal mole^{-1} K^{-1} or ~0.2% respectively.

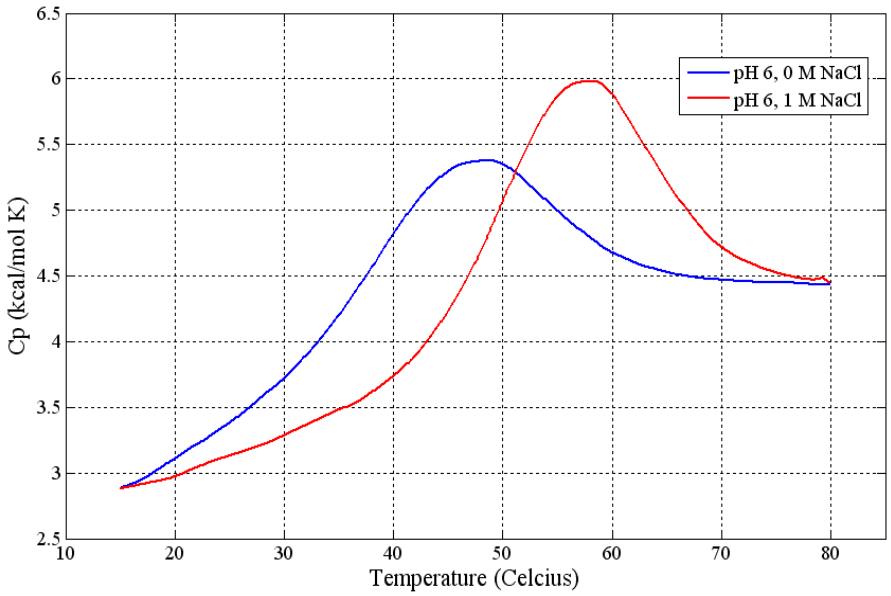

Figure 4.3. Effect of salt on PBX-HD in 20mM Sodium Phosphate pH 6

To align the two curves, the 1M NaCl curve was offset by -0.2 kcal mole^{-1} K^{-1}

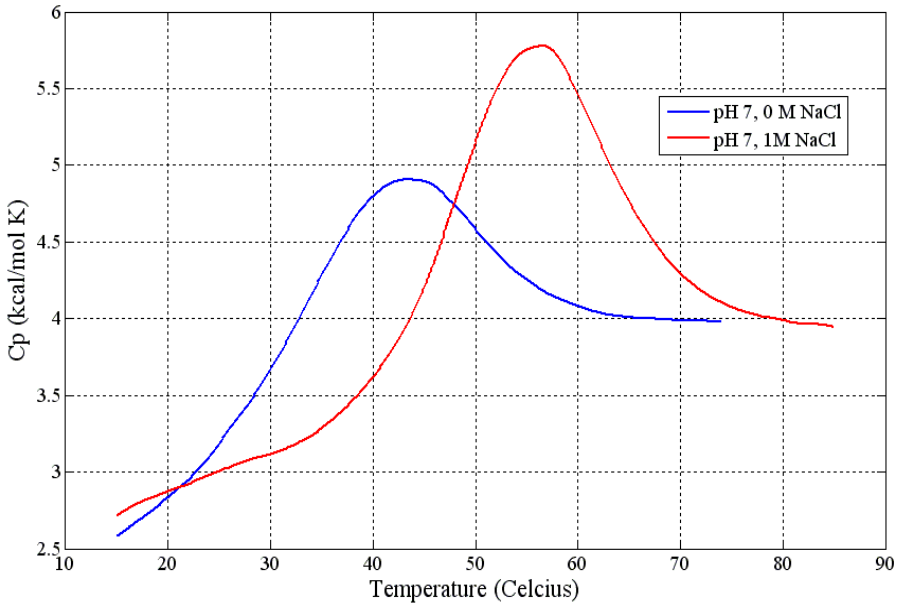

Figure 4.4. Effect of salt on PBX-HD in 20mM Sodium Phosphate pH 7

To line up the two curves, the 1M NaCl curve was offset by 1.1 kcal mole^{-1} K^{-1}

The presence of 1M NaCl clearly stabilized the protein. The molar heat capacity peak for PBX-HD at pH 6 increases from 48.8°C to 57.7°C and for PBX-HD at pH 7 it increases from 43.3°C to 56.7°C. Fitting these heat capacities to model 1A, two-state model with linear fit of folded baseline and calculated unfolded baseline as described in section 3.1.1, the melting temperature and the enthalpy differences in the presence of 1M NaCl increase in general (see table 5).

Table 5. Two-state (Model 1A) thermodynamic parameters of the salt effect on the PBX-HD unfolding

	T_m	$\Delta H(T_m)$	$\Delta S(T_m)$	X^2
pH 6, 0M NaCl	45.3	32.5	0.102	1500
pH 6, 1M NaCl	56.4	41.5	0.126	5143
pH 7, 0M NaCl	41.2	32.1	0.102	970
pH 7, 1M NaCl	55.0	42.6	0.130	5284

There are several factors by which NaCl can stabilize PBX-HD. First, NaCl is a non-denaturing salt that is considered to be Hofmeister salt. It stabilizes proteins by salting out non-polar molecules from aqueous solution (*28, 37*). Secondly, PBX-HD is stabilized through the screening of unfavorable electrostatic interactions. PBX-HD is a strongly charged molecule, with a calculated net charge of +8. Unfavorable electrostatics in the more densely packed folded state would be compensated by an increase in ionic strength, and therefore protein folding is favored (*37, 38*). The highly positively-charged may also be

able to form specific interactions with the Cl^- anion, leading to stabilization. More experimental data, especially DSC data with more varying concentration of NaCl and other salts, are needed to give a clearer picture on the salt effect on the stability of PBX homeodomain.

4.3 Extended PBX-HD

1……..10……20……a.b.c……..30……...40……
ARRKRRNFNKQATEILNEYFYSHLSNPYPSEEAKEELAKKCGITVSQVSN
…..50...……..60……..70…….…78
WFGNKRIRYKKNIGKFQEEANIYAAKTAVTA

Figure 4.5. Amino acid sequence of extended PBX-HD construct
The red amino acid residues are the additional residues in the extended PBX-HD construct.

The extended construct of PBX-HD has 81 residues and a molecular weight of 9429.6 Dalton. The PBX-HD is denoted PBX 1-59 while the extended one is PBX 1-78. It has extra 19 residues at the C-terminus. This C-terminal extension is found to be unstructured in the free form of the protein while it forms a fourth helix in the DNA complex (*15*). Previous study has shown that this C-terminal extension to the PBX-HD results in destabilization of the structure or an increased dynamics in the folded state (*29*). However, the calorimetric data of extended PBX-HD suggests that it is more stable than the regular PBX-HD construct (see figure 4.5). The molar heat capacity peak shifts from 43.3°C to 46.9°C. Fitting model 1A, two-state model with linear baselines as described in

section 3.1.1, to the heat capacities of PBX-HD and extended PBX-HD also shows an increase of melting temperature and enthalpy differences (see table 6). Both suggest that the extended PBX-HD is more thermodynamically stable than the PBX-HD.

This preliminary study suggests that there are residual interactions between the C-terminal extension and the rest of PBX-HD that were not detected in structural studies. It also suggests that structural stability does not necessarily impart thermo-stability.

Table 6. Two-state linear baselines (Model 1A) fitting to PBX-HD and extended PBX-HD in 20 mM Sodium Phosphate pH 7

	T_m (°C)	$\Delta H(T_m)$ (kcal/mol)	$\Delta S(T_m)$ (kcal/mol K)	χ^2
PBX	41.2	32.1	0.102	970
Extended PBX	43.7	38.8	0.122	586

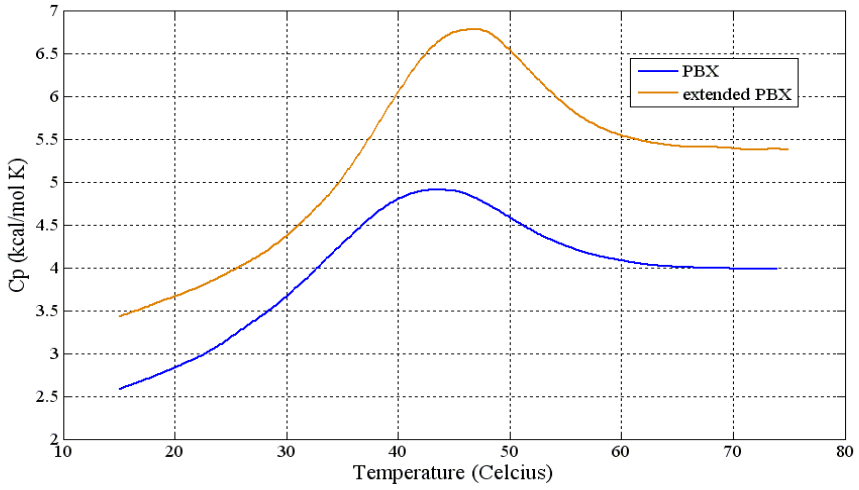

Figure 4.6. Comparison of heat capacity curve of PBX-HD and extended PBX-HD in 20 mM Sodium Phosphate pH 6

Chapter 5: Conclusion

Through the calorimetric study, we have gained insight into the unfolding mechanism of a homeodomain protein, PBX-HD. The heat capacity profile of unfolding PBX-HD is different from those of many well-behaved globular proteins. However, many DNA-binding proteins such as DNA binding domain of Sox-5 (*18*) and the leucine zipper of the GCN4 bZIP (*19*) have similar heat capacity unfolding profiles as PBX-HD: low thermal stability, low excess heat capacity in general, broad curves, and a significantly steeper slope in the pre-denaturational temperature range. The latter characteristic poses a problem when determining the heat capacity baseline because the folded baseline crosses the unfolded baseline near the unfolding temperature (see figure 3.2 and 3.4). This suggests that a linear extrapolation of the folded heat capacity baseline from the DSC curve doesn't truly represent the actual heat capacity of the folded state. When BPTI-normalized baseline was used, both the discreet states model 2A and independent transition model 2B (see figure 3.7 and 3.9) predict that the PBX-HD unfolds in a three-state manner, with the intermediate state populated more than 50% at physiological temperatures, thus accounting for the steep heat capacity curve at that temperature.

However, it becomes clear that one can arrive at many different folding models depending on how one specifies the heat capacity baselines. Combining DSC data with NMR data is one way to deal with this problem. In fitting the DSC and NMR data, no assumptions were made regarding the baselines, beyond that

they may be described by third-order polynomials. This fitting suggests that the unfolding of PBX-HD is two-state with a curved polynomial baseline that does not intersect the unfolded baseline.

The effects of pH and ionic strength on the heat capacity of the unfolding PBX were also measured. From these preliminary studies, it is clear that the pH significantly affects the thermostability of PBX and also its unfolding dynamics (see figure 4.1). The effect of ionic strength, in this case NaCl, seems more straightforward: increase in NaCl concentration results in the increase in PBX-HD stability (see figure 4.3 and 4.4). More experiments and data analysis need to be done on this front to get a full understanding on the role of pH and ionic strength on the unfolding dynamics of PBX.

Last but not least, the heat capacity of extended PBX-HD was also measured and the result shows that it is more stable (see figure 4.6). This is contrary to the previous study (*15,29*) showing that the C-terminal extension of this extended PBX-HD is unstructured and destabilizes the structure. This preliminary results shows that the relationship between this C-terminal extension and the thermostability of the structure is complex. It also suggests that structural stability, as the solution NMR study of PBX shows, does not necessarily impart thermodynamic stability.

To summarize, this calorimetric study of PBX-HD has provided insights not only into the unfolding mechanism of PBX-HD, but also into the study of protein unfolding in general. Though rich in information, calorimetric data itself is not sufficient to provide a strong conclusion as to the unfolding mechanism of a protein, especially one with broad features such as PBX-HD where it is not clear

only from this heat capacity profile whether the unfolding is cooperative or non-cooperative. For now, both DSC and NMR data have strongly suggested that PBX-HD unfolds in a cooperative manner. Another NMR experiment, ^{15}N Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion, conducted by Patrick Farber, has also supported this result.

References

[1] C. Levinthal (1968) Are there pathways for protein folding? *J. Chim. Phys.* 65, 44-45

[2] C. Levinthal (1969) in *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.* Editors: J. T. P. DeBrunner and E. Munck (University of Illinois Press), pp. 22-24

[3] K.A. Dill and Hue Sun Chan (1997) From Levinthal to pathways to funnels. *Nature Structural Biology*. Vol. 4 Number 1, pp. 10-19

[4] M.M. Garcia-Mira, M. Sadqi, N. Fischer, J.M. Sanchez-Ruiz, V. Muñoz (2002) Experimental identification of downhill protein folding. *Science*, 298, 2191-2195

[5] N. Ferguson, P.J. Schartau, T.D. Sharpe, S. Sato, A.R. Fersht (2004) One-state downhill versus conventional protein folding. *J. Mol. Biol.* 344:295-301

[6] P.L Privalov, S.A. Pothekin (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods in Enzymology*, Vol. 131 pp. 4-51

[7] V. Muñoz, J.M. Sanchez-Ruiz (2004). Exploring protein-folding ensembles: A variable-barrier model for the analysis of equilibrium unfolding experiments. *PNAS*, Vol. 101, No. 51, pp. 17646-17651

[8] S.S.Cho, P. Weinkam, P.G. Wolynes (2008) Origins of barriers and barrierless folding in BBL. *PNAS*, Vol. 105, No. 1, pp. 118-123

[9] G.I. Makhatadze, P.L. Privalov (1990) I. Partial molar heat capacity of individual amino acids residues in aqueous solution: hydration effect. *J. Mol. Biol.* 213, 375-384

[10] G.I. Makhatadze, P.L. Privalov (1990) II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects. *J. Mol. Biol.* 213, 385-391

[11] G.I. Makhatadze, P.L. Privalov (1993) Contribution of hydration to protein folding thermodynamics I. The enthalpy of hydration. *J. Mol. Biol.* 232, 639-659

[12] G.I. Makhatadze, P.L. Privalov (1993) Contribution of hydration to protein folding thermodynamics II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* 232, 660-679

[13] P.L. Privalov (1980) Scanning calorimetry for studying macromolecules. *Pure & Applied Chemistry*. Vol. 52, pp. 479-497

[14] P.L. Privalov, S.A. Pothekin (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods in Enzymology*, Vol. 131, pp 4-51

[15] T. Sprules, N. Green, M. Featherstone, K. Gehring (2000) Conformational changes in the PBX homeodomain and C-terminal extension upon binding BDA and HOX-derived YPMQ peptides. *Biochemistry*, 39, 9943-9950

[16] M. Hackel, H.J. Hinz, G.R. Hedwig (1999) Partial molar volumes of proteins: amino acid side-chain contributions derived from the partial molar volumes of some tripeptides over the temperature range 10-90°C. *Biophysical Chemistry*, 82, 35-50

[17] M. Hackel, H.J. Hinz, G.R. Hedwig (1999) A new set of peptide-based group heat capacities for use in protein stability calculations. *Journal of Molecular Biology*, 291, 197-213

[18] P.L. Privalov, I. Jelesarov, C.M. Read, A.I. Dragan, C. Crane-Robinson (1999). The energetics of HMG box interactions with DNA: thermodynamics of the DNA binding of the HMG box from mouse sox-5. *Journal of Molecular Biology*, 294, 997-1013.

[19] A.I. Dragan, P.L. Privalov (2002). Unfolding of a leucine zipper is not a simple two-state transition. *Journal of Molecular Biology*, 321, 891-908

[20] Y.V. Griko, G.I. Makhatadze, P.L. Privalov, R.W. Hartley (1994) Thermodynamics of barnase unfolding. *Protein Science*, 3, 669-676

[21] P.L. Wintrode, G.I. Makhatade, P.L. Privalov (1994) Thermodynamics of ubiquitin unfolding. *Proteins*, 18, 246-253

[22] P.L. Privalov, A.I. Dragan (2006) Microcalorimetry of biological macromolecules. *Biophysical chemistry*, 126, 16-24

[23] G.I. Makhatadze, P.L. Privalov (1995) Energetics of protein structure. *Advances in Protein Chemistry*, 47, 307-425

[24] H.N. Cheng and F.A. Bovey (1977) Cis-Trans equilibrium and kinetic stuides of Acetyl-L-Proline and Glycyl-L-Proline. *Biopolymers*, 16, 1465-1472

[25] C. Grathwohl and K. Wuthrich (1981) NMR studies of the rates of Proline cis-trans isomerization of Oligopeptides. *Biopolymers*, 20, 2623-2633

[26]Wen-Ji Wu and D.P. Raleigh (1997) Local control of peptide conformation: stabilization of *cis* proline peptide bonds by aromatic proline interactions. *Biopolymers*, 45, 381-394

[27] An-Suei Yang and B. Honig (1993) On the pH dependence of protein stability. *Journal of Molecular Biology*, 231, 459-474

[28] R.L. Baldwin (1996) How Hofmeister ion interactions affect protein stability. *Biophysical Journal*, 71, 2056-2063

[29] C. Jabet, R. Gitti, M.F. Summers, C. Wolberger (1999) NMR studies of the Pbx1 TALE homeodomain protein free in solution and bound to DNA: Proposal for a mechanism of HoxB1-Pbx1-DNA complex assembly. *Journal of Molecular Biology*, 291, 521-530

[30] Peng Li, F.Y. Oliva, A.N. Naganathan, V. Muñoz [2009] Dynamics of one-state downhill protein folding. *PNAS*, Vol. 106, No. 1, 103-108

[31] M. Sadqi, D. Fushman, V. Muñoz [2006] Atom-by-atom analysis of global downhill protein folding. *Nature*, Vol. 442, 317-321

[32] N. Ferguson, T.D. Sharpe, C.M. Johnson, P.J. Schartau, A.R. Fersht [2007] Structural Biology: Analysis of 'downhill' protein folding. *Nature*, Vol. 445, E14-E15

[33] E. Arberly, T.J. Rutherford, T.D. Sharpe, N. Ferguson, A.R. Fersht [2009] Downhill versus barrier-limited folding of BBL 1: energetic and structural perturbation effects upon protonation of a histidine of unusually low pKa. *J. Mol. Biol.*, 387, 986-992

[34] H. Neuweller, T.D. Sharpe, C.M. Johnson, D.P. Teufel, N. Ferguson, A.R. Fersht [2009] Downhill versus barrier-limited folding of BBL 2: mechanistic insights from kinetics of folding monitored by independent tryptophan probes. *J. Mol. Biol.*, 387, 975-985

[35] G. Settanni, A.R. Fersht [2009] Downhill versus Barrier-Limited Folding of BBL: 3. Heterogeneity of the Native State of the BBL Peripheral Subunit Binding Domain and Its Implications for Folding Mechanisms. *J. Mol. Biol.*, 387, 993-1001

[36] CRC Handbook of chemistry and physics, 89th Ed., 2008-2009

[37] W.D. Kohn, C.M. Kay, R.S. Hodges [1997] Salt effects on protein stability: two-stranded α-helical coiled-coils containing inter- or intrahelical ion pairs. *J. Mol. Biol.,* 267, 1039-1052

[38] J.H. Carra, P.L. Privalov [1997] Energetics of folding and DNA binding of MATα2 homeodomain. *Biochemistry*, 36, 526