Electrostatic Contributions to the Thermodynamics of Ribonuclease P Protein Folding

by

Pamela Lynnette Mosley

Department of Chemistry
Duke University

Date:_______________________

Approved:

___________________________
Terrence G. Oas, Supervisor

___________________________
Katherine J. Franz

___________________________
Michael J. Therien

___________________________
Steve W. Baldwin

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2016
ABSTRACT

Electrostatic Contributions to the Thermodynamics of Ribonuclease P Protein Folding

by

Pamela Lynnette Mosley

Department of Chemistry
Duke University

Date:_______________________
Approved:

___________________________
Terrence G. Oas, Supervisor

___________________________
Katherine J. Franz

___________________________
Michael J. Therien

___________________________
Steve W. Baldwin

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2016
Abstract

Electrostatic interactions are of fundamental importance in determining the structure and stability of macromolecules. For example, charge-charge interactions modulate the folding and binding of proteins and influence protein solubility. Electrostatic interactions are highly variable and can be both favorable and unfavorable. The ability to quantify these interactions is challenging but vital to understanding the detailed balance and major roles that they have in different proteins and biological processes. Measuring pK_a values of ionizable groups provides a sensitive method for experimentally probing the electrostatic properties of a protein.

pK_a values report the free energy of site-specific proton binding and provide a direct means of studying protein folding and pH-dependent stability. Using a combination of NMR, circular dichroism, and fluorescence spectroscopy along with singular value decomposition, we investigated the contributions of electrostatic interactions to the thermodynamic stability and folding of the protein subunit of Bacillus subtilis ribonuclease P, P protein. Taken together, the results suggest that unfavorable electrostatics alone do not account for the fact that P protein is intrinsically unfolded in the absence of ligand because the pK_a differences observed between the folded and unfolded states are small. Presumably, multiple factors encoded in the P protein sequence account for its IUP property, which may play an important role in its function.
Dedication

To my mother and father, Nell and James Mosley, who have always loved and supported me. To my aunt and uncle, Margaret and Jerome Rogers, and cousin, Wesley Rogers, thank you for everything.
Table of Contents

Abstract ......................................................................................................................................... iv

List of Tables .................................................................................................................................. x

List of Figures ............................................................................................................................... xi

List of Schemes ...........................................................................................................................xiv

List of Abbreviations .................................................................................................................. xv

Acknowledgements ...................................................................................................................xvi

1. Introduction ............................................................................................................................... 1

1.1 Protein electrostatics ........................................................................................................ 1

1.2 The ionizable groups in proteins .................................................................................... 2

1.3 pKₐ and the Henderson-Hasselbalch equation ............................................................ 4

1.3.1 Intrinsic pKₐ values ..................................................................................................... 7

1.3.2 Calculating free energies from pKₐ shifts ................................................................. 8

1.4 Molecular determinants of perturbed pKₐ values in proteins ...................................... 11

1.4.1 Charge-charge interactions ...................................................................................... 11

1.4.1.1 Electrostatic modeling of charge-charge interactions ....................................... 13

1.4.2 Charge-dipole interactions ....................................................................................... 21

1.4.3 Born effect ................................................................................................................... 23

1.4.4 Protein reorganization .............................................................................................. 25

1.4.5 Ionic strength and the Debye-Hückel effect .......................................................... 26

1.4.6 Metal ions ................................................................................................................... 27
2.4.2 Determination of pKₐ values using NMR spectroscopy .............................................64
2.4.3 Calculation of pH-dependent stability .......................................................................68
2.4.4 Determination of tautomeric state populations .......................................................69

3. Salt sensitivity of histidine pKₐ values in P protein ......................................................71
   3.1 Introduction ..................................................................................................................71
   3.2 Results and discussion ...............................................................................................73
      3.2.1 The pKₐ values of sucrose-folded P protein His are depressed ......................73
      3.2.2 Comparison of the pKₐ values of unliganded unfolded and unliganded sucrose-folded P protein His ............................................................................................79
      3.2.3 Electrostatic repulsion is greater in sucrose-folded state than sulfate-folded state ........................................................................................................................................80
      3.2.4 The pKₐ values of sucrose-folded P protein His have a strong ionic strength dependence .........................................................................................................................80
   3.3 Conclusions ..................................................................................................................81
   3.4 Materials and methods ..............................................................................................82
      3.4.1 Expression and purification of P protein ...............................................................82
      3.4.2 Determination of pKₐ values using NMR spectroscopy .......................................84

4. The effect of dielectric constant on the protonation of the histidine side chain.............87
   4.1 Introduction ..................................................................................................................87
   4.2 Results and discussion ...............................................................................................89
   4.3 Conclusions and future directions ...............................................................................93
   4.4 Materials and methods ..............................................................................................96
      4.4.1 Chemicals and reagents .......................................................................................96
4.4.2 Measurement of pKₐ values using NMR spectroscopy ........................................... 96

5. Probing the folding intermediate of P protein via singular value decomposition ...... 98
   5.1 Introduction ..................................................................................................................... 98
     5.1.1 Singular value decomposition ............................................................................... 100
   5.2 Results and discussion ................................................................................................. 104
     5.2.1 Urea-induced unfolding of P protein ................................................................... 104
     5.2.2 TMAO-induced folding of P protein ................................................................. 107
     5.2.3 Sucrose-induced folding of P protein ............................................................... 111
   5.3 Conclusions and future directions ............................................................................. 113
   5.4 Material and methods .................................................................................................. 117
     5.4.1 Chemicals and reagents .......................................................................................... 117
     5.4.2 Expression and purification of P protein ............................................................. 118
     5.4.3 Circular dicroism and fluorescence spectroscopy .............................................. 119
     5.4.4 Singular value decomposition ............................................................................... 120

Appendix A ........................................................................................................................ 121

References ........................................................................................................................ 132

Biography ........................................................................................................................ 144
List of Tables

Table 1.1: pKₐ values for ionizable groups in proteins ............................................................ 7

Table 2.1: Histidine pKₐ values and Hill coefficients of N-acetyl-L-histidine methylamide and three histidines in unfolded, intermediate, and sulfate-bound folded P protein .............................. 48

Table 3.1: pKₐ values and Hill coefficients of L-histidine, N-acetyl-L-histidine methylamide and three histidines in unfolded, sulfate-folded and sucrose-folded P protein .......................................................... 78

Table S.1: Histidine pKₐ values and Hill coefficients in unfolded, intermediate, and sulfate-bound folded P protein based on fitting $^{13}\text{C} \epsilon_1$ data ................................................................. 130

Table S.2: Solvent exposure of the histidine sidechains in P protein, relative to Gly-His-Gly, calculated as described in Materials and Methods ................................................................. 130

Table S.3: $^1\text{H}$ and $^{13}\text{C}$ chemical shifts of internal standard L-Histidine, Pyridine, and P protein histidine residues ........................................................................................................ 131
List of Figures

Figure 1.1: Ionizable groups in proteins ................................................................. 3

Figure 1.2: An example of a titration curve .......................................................... 6

Figure 1.3: Electric field produced by positive charge Q ..................................... 15

Figure 1.4: Equipotential lines around a point charge ......................................... 17

Figure 1.5: Timeline of different methods developed for the determination of pKₐ values ................................................................. 29

Figure 2.1: Electrostatic surface potential of the sulfate-bound folded state of B. subtilis RNase P protein ........................................................ 38

Figure 2.2: Titration curves following the proton resonance of pyridine in the absence of 20 mM sodium sulfate ......................................................... 42

Figure 2.3: H₁ P protein histidine chemical shift vs H₁ L-histidine chemical shift titration curves ................................................................. 46

Figure 2.4: Calculated free energy of the U→I reaction based on ΔpKₐ = pKₐ - pKᵢₐ.57

Figure 2.5: H₁ P protein histidine chemical shift vs H₁ L-histidine chemical shift titration curves ................................................................. 58

Figure 3.1: H₁ P protein histidine vs H₁ L-histidine chemical shift titration curves..... 74

Figure 3.2: L-Histidine (internal standard) titration curves ..................................... 76

Figure 3.3: H₁ N-acetyl-L-histidine methylamide vs H₁ L-histidine chemical shift titration curve ................................................................. 77

Figure 4.1: H₁ His peptide vs H₁ L-histidine chemical shift titration curve .......... 91

Figure 4.2: Coupled deprotonation and transfer of a histidine side chain scheme .... 95

Figure 5.1: Linear fit line reduces two-dimensional data into one ......................... 102
Figure 5.2: Perpendicular regression line is a poor approximation of the original data points ........................................................................................................................................................................... 103

Figure 5.3: P protein tryptophan fluorescence emission spectra at pH 7 and various urea concentrations ........................................................................................................................................................................... 105

Figure 5.4: The SVD basis vectors obtained from a FL urea titration of P protein in 10 mM L-histidine, 10 mM pyridine, 20 mM sodium sulfate at pH 7.0 and 25°C ............................................................ 106

Figure 5.5: The fits of the urea-dependent basis vectors to a three-state model .......... 108

Figure 5.6: (A) Relative populations and (B) FL spectra of each of the three thermodynamic species of P protein ........................................................................................................................................................................... 109

Figure 5.7: P protein tryptophan fluorescence emission spectra at pH 7 and various TMAO concentrations ........................................................................................................................................................................... 110

Figure 5.8: The SVD basis vectors obtained from a FL TMAO titration of P protein in 10 mM L-histidine, 10 mM pyridine at pH 7.0 and 25°C ................................................................................................................. 112

Figure 5.9: Far-UV CD spectra at pH 7 and various sucrose concentrations .......... 113

Figure 5.10: The SVD basis vectors obtained from a CD sucrose titration of P protein in 20mM sodium cacodylate at pH 7.0 and 25°C ........................................................................................................................................................................... 114

Figure 5.11: The fits of the sucrose-dependent basis vectors to a three-state model ....... 115

Figure 5.12: (A) Relative populations and (B) CD spectra of each of the three thermodynamic species of P protein ........................................................................................................................................................................... 116

Figure S.13: Titration of NMR sample .................................................................................... 121

Figure S.14: L-Histidine (internal standard) titration curves under different buffer conditions ........................................................................................................................................................................... 122

Figure S.15: $^{13}$C$^\epsilon$ 1 P protein histidine versus $^{13}$C$^\epsilon$ 1 L-Histidine chemical shifts titration curves ........................................................................................................................................................................... 123

Figure S.16: NMR resonances from all three histidine residues in the unfolded state of P protein ........................................................................................................................................................................... 124
Figure S.17: $^1$H $^1$N-acetyl-L-Histidine methylamide versus $^1$H $^1$L-Histidine chemical shifts titration curves ................................................................................................................ 125

Figure S.18: NMR resonance from intermediate peak in unfolded state of P protein .... 126

Figure S.19: $^1$H-$^13$C HSQC unfolded P protein spectra in various pH buffers.............. 127

Figure S.20: Pyridine (internal standard) and P protein histidine residue titration curves ..................................................................................................................................................... 128

Figure S.21: Determination of the tautomeric state of the neutral form of the protein and model compound histidines............................................................................................................................................................................ 129
List of Schemes

Scheme 1.1: Thermodynamic cycle for analyzing pKₐ shifts ................................................. 9

Scheme 1.2: Thermodynamic cycle for calculating the transfer of a single charge from water to the interior of a protein ........................................................................................................ 24

Scheme 2.1: Thermodynamic cycle for coupled proton binding and conformational change ................................................................................................................................. 36

Scheme 4.1: Thermodynamic cycle for the deprotonation and transfer of histidine....... 92
List of Abbreviations

RNase - ribonuclease

P protein – protein subunit of *Bacillus subtilis* RNase P

IUP – intrinsically unfolded protein

NMR – nuclear magnetic resonance

HSQC – heteronuclear single quantum coherence

TMAO - trimethylamine N-oxide

CD – circular dichroism

FL – fluorescence

SVD – singular value decomposition
Acknowledgements

There are many people who I would like to thank that have helped me throughout my time here at Duke. First, I would like to thank my advisor, Terry Oas, for all of his guidance and utmost support. Terry has been instrumental in my growth and development as a critical scientific thinker and I will always be grateful for the opportunity to work with him. I would like to thank all of my dissertation committee members for their support and helpful advice and feedback: Katherine Franz, Michael Therien, and Steven Baldwin. Thank you to all of my Oas labmates. Also, I would like to thank the Structural Biology and Biophysics and GAANN/Pathway to PhD programs at Duke for fellowship support and the Duke Chemistry department for financial support.

I am also thankful to my undergraduate advisor, Eric Toone, for giving me the opportunity to work in his lab and my mentor, Allison Schmitt, for taking me under her wing, allowing me to be a part of the camelid antibody project, and being such a great mentor and friend. Thank you to my other Toone labmates for an amazing undergraduate research experience.

Finally, I would like to thank my family and friends for all of their unconditional love and support throughout these years. I would have never made it this far without them. Thank you for everything!
1. Introduction

1.1 Protein electrostatics

In protein chemistry, interactions that arise from polar and charged groups are most often referred to simply as “electrostatics”. It has become clear that there are several processes that are influenced by electrostatic interactions such as the folding and stability of proteins\textsuperscript{1-4}, formation of protein-protein\textsuperscript{5} and protein-nucleic acid\textsuperscript{6} complexes, binding of substrates by enzymes and enzyme catalysis\textsuperscript{3, 7, 8}, precipitation and solubilization of proteins\textsuperscript{9, 10}, redox processes\textsuperscript{11}, denaturation of proteins at extreme pH values\textsuperscript{2}, and ion channel transport\textsuperscript{12, 13}. The ability to quantify these interactions is vital to understanding the detailed balance and major roles that electrostatics play in biological macromolecules and processes. However, accomplishing this task is not trivial.

It is difficult to quantitatively describe protein electrostatic interactions in aqueous solution for several reasons. First, the dielectric properties of a protein-solvent system are not homogeneous\textsuperscript{14}. Spatial variations in dielectric constant are important because the free energy of solvation of a charge and free energy of interaction between charges depend on this important property. Second, the pK\textsubscript{a} value of a charged group on a protein may not only depend heavily on its interaction with one other ionizable or polar group but many groups, both in solution and on the protein. As a consequence, acid-base equilibria can be cooperative and therefore quite complex\textsuperscript{15}. Third, electrostatic
interactions are long-range or the charged groups are separated by distances that are
similar to the dimensions of the protein. Therefore, large numbers of interactions must
be taken into consideration when calculating the total electrostatic free energy\textsuperscript{16}. Finally,
specific ion effects such as the Hofmeister effect\textsuperscript{17} can affect the properties of proteins
and are not well understood but need to be included in the complete portrayal of
electrostatic interactions.

1.2 The ionizable groups in proteins

On average, 29% of amino acids in proteins have charged side chains when
reversibly protonated or deprotonated by water or other bases and acids in solution\textsuperscript{18}.
There are eleven groups (glutamate (Glu), aspartate (Asp), histidine (His), lysine (Lys),
arginine (Arg), cysteine (Cys), tyrosine (Tyr), serine (Ser), threonine (Thr), N-terminus
amino, and C-terminus carboxyl) that ionize in the pH range of 1 to 14 (Figure 1.1). The
general definition of acids and bases suggested by Johannes Bronsted and Thomas
Lowry in 1923 states that an acid is a proton donor and a base is a proton acceptor\textsuperscript{19}.
Here, we will think of the eleven groups as all acids or proton donors and divide them
into two sub-groups: the cationic acids and the neutral acids\textsuperscript{20}. For example, according to
the Bronsted-Lowry definition, cationic acids could also be thought of as neutral bases
but to keep the nomenclature consistent, they will be referred to as cationic acids
throughout this chapter.
Figure 1.1: Ionizable groups in proteins

Ionizable groups can be divided into two groups: (A) cationic acids and (B) neutral acids.
The cationic acids (His, Lys, Arg, and N-terminus amino) can be represented by the following equation:

\[ A^+H \leftrightarrow H^+ + A \]  

(1.1)

where \( A^+H \) is the ionizable group with a proton bound, \( H^+ \) is the proton, and \( A \) is the conjugate base of \( A^+H \). It should be noted in Equation 1.1 that when a cationic acid is protonated, it is positively charged. For His, Lys, Arg, and N-terminus amino, their side chains are positively charged below their pK\(_a\) and neutral above their pK\(_a\). The neutral acids (Glu, Asp, Cys, Tyr, Thr, Ser, and C-terminus carboxyl) can be represented by the following equation:

\[ HA \leftrightarrow H^+ + A^- \]  

(1.2)

where HA is the ionizable group with a proton bound, \( H^+ \) is the proton, and \( A^- \) is the conjugate base of HA. It should be noted in Equation 1.2 that when a neutral acid is protonated, it is uncharged. For Glu, Asp, Cys, Tyr, Thr, Ser, and C-terminus carboxyl, their side chains are neutral below their pK\(_a\) and negatively charged above their pK\(_a\).

1.3 pK\(_a\) and the Henderson-Hasselbalch equation

The equilibrium expression for the reaction given in Equation 1.1, described by the acid dissociation constant, K\(_a\), is
\[ K_a = \frac{[H^+][A]}{[A^+H]} \]  \hspace{1cm} (1.3)

Equation 1.3 can then be rearranged to give

\[ [H^+] = K_a \frac{[A^+H]}{[A]} \]

Next, taking the negative log of both sides of the expression yields

\[ -log[H^+] = -logK_a - log \frac{[A^+H]}{[A]} \]

Using the definitions of pH \((pH = -log[H^+])\) and \(pK_a\) \((pK_a = -logK_a)\), this equation becomes

\[ pH = pK_a - log \frac{[A^+H]}{[A]} \]

Finally, inverting the term \(-log \frac{[A^+H]}{[A]}\) yields the Henderson-Hasselbalch equation:

\[ pH = pK_a + log \frac{[A]}{[A^+H]} \]  \hspace{1cm} (1.4)

It relates pH and \(pK_a\) to the equilibrium concentrations of the ionizable group \([A^+H]\) and its conjugate acid \([A]\). This equation shows why the \(pK_a\) of an acid is equal to the pH of the solution at the midpoint of its titration curve (Figure 1.2). At that point, \([A^+H] = [A]\), and Equation 1.4 reduces to

\[ pH = pK_a + log (1) = pK_a + 0 = pK_a \]
This titration curve is created by plotting a measured quantity versus pH. The equivalence point corresponds to the pKₐ value.

Therefore, the pKₐ of an ionizable group is equal to the pH of the solution where the concentration of the group equals the concentration of its conjugate base. Equation 1.4 is the most commonly used equation to calculate pKₐ values and it can also be used to calculate pH if the pKₐ and the molar ratio of the group and its conjugate base are known and calculate the molar ratio of the group and its conjugate base if the pH and pKₐ are known. The same discussion can be applied to Equation 1.2 for neutral acids as well.
1.3.1 Intrinsic pK\textsubscript{a} values

All ionizable groups have an intrinsic pK\textsubscript{a}. In 1962, the term intrinsic pK\textsubscript{a} was introduced by Tanford in early studies that aimed at interpreting titration curves of proteins. He defined the term as the pK\textsubscript{a} value an ionizable group would have when the net charge on the molecule is zero. In other words, if the group were completely exposed to the solvent and not affected by the presence of any polar or charged group, this would be its intrinsic pK\textsubscript{a}. Table 1.1 lists the intrinsic pK\textsubscript{a} values measured by Nozaki and Tanford\textsuperscript{21} determined using model compounds at 25°C. “Corrections for

<table>
<thead>
<tr>
<th>Group</th>
<th>Nozaki and Tanford\textsuperscript{a}</th>
<th>Average pK\textsubscript{a}\textsuperscript{b}</th>
<th>Low pK\textsubscript{a}\textsuperscript{b}</th>
<th>High pK\textsubscript{a}\textsuperscript{b}</th>
<th>No. of measurements\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-terminus carboxyl</td>
<td>3.8</td>
<td>3.3 ± 0.8</td>
<td>2.4</td>
<td>5.9</td>
<td>22</td>
</tr>
<tr>
<td>Asp</td>
<td>4.0</td>
<td>3.5 ± 1.2</td>
<td>0.5</td>
<td>9.2</td>
<td>139</td>
</tr>
<tr>
<td>Glu</td>
<td>4.4</td>
<td>4.2 ± 0.9</td>
<td>2.1</td>
<td>8.8</td>
<td>153</td>
</tr>
<tr>
<td>His</td>
<td>6.3</td>
<td>6.6 ± 1.0</td>
<td>2.4</td>
<td>9.2</td>
<td>131</td>
</tr>
<tr>
<td>N-terminus amino</td>
<td>7.5</td>
<td>7.7 ± 0.5</td>
<td>6.8</td>
<td>9.1</td>
<td>16</td>
</tr>
<tr>
<td>Cys</td>
<td>9.5</td>
<td>6.8 ± 2.7</td>
<td>2.5</td>
<td>11.1</td>
<td>25</td>
</tr>
<tr>
<td>Tyr</td>
<td>9.6</td>
<td>10.3 ± 1.2</td>
<td>6.1</td>
<td>12.1</td>
<td>20</td>
</tr>
<tr>
<td>Lys</td>
<td>10.4</td>
<td>10.5 ± 1.1</td>
<td>5.7</td>
<td>12.1</td>
<td>35</td>
</tr>
<tr>
<td>Arg</td>
<td>12.0</td>
<td>not calculated</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} pK\textsubscript{a} values are listed in Table 3 of Ref 21.
\textsuperscript{b} This is a summary of 541 pK\textsubscript{a} values tabulated in Ref 23.
electrostatic and other effects were made to yield pKₐ values representative of the unperturbed group”²¹. pKₐ values for Ser and Thr were not measured. Ser and Thr each have an intrinsic pKₐ of ~13²². For comparison, Table 1.1 lists the average pKₐ values calculated from 541 reported pKₐ values in the literature for 78 different folded proteins under various conditions²³. The differences between these two columns reflect how a pKₐ values can be shifted away from its intrinsic value when in a protein. This shift can be due to several factors that will be discussed in Section 1.4 of this chapter. Intrinsic pKₐ values are beneficial to know and serve as decent models for the unperturbed pKₐ values of ionizable side chains in proteins.

1.3.2 Calculating free energies from pKₐ shifts

The acid dissociation constant for either reaction described in Equations 1.1 and 1.2 is related to the standard reaction Gibbs free energy, ΔG°, by

\[
\Delta G^\circ = -RT \ln K_a = -RT \times 2.303 \log K_a = RT \times 2.303 \text{ pK}_a \tag{1.5}
\]

where R is the gas constant and T is temperature. Hence, the pKₐ value of a titratable site can be used to calculate the free energy of proton release or the free energy to deprotonate that particular group. If this pKₐ is compared to the intrinsic pKₐ of the
group, the double free energy difference between the two can be determined. This process of analyzing pKₘ shifts is depicted in Scheme 1.1 as a thermodynamic cycle

\[
\Delta \Delta G = \Delta G_{\text{protein}} - \Delta G_{\text{intrinsic}}
\]

Scheme 1.1: Thermodynamic cycle for analyzing pKₘ shifts

where P-His⁺ and His⁺ are the protonated histidine side chain in a protein and model compound in solution, respectively, P-His and His are the deprotonated histidine side chain in a protein and model compound in solution, respectively, and \(\Delta \Delta G = \Delta G_{\text{protein}} - \Delta G_{\text{intrinsic}}\) is the double free energy difference. Equation 1.5 can be substituted into \(\Delta G_{\text{protein}}\) and \(\Delta G_{\text{intrinsic}}\) of the \(\Delta \Delta G\) equation to get

\[
\Delta \Delta G = RT \frac{2.303}{pK_a} - RT \frac{2.303}{pK_{a_{\text{intrinsic}}}}
\]
which can be further simplified to

\[
\Delta \Delta G = 2.303 \, RT \, (pK_{a_{\text{protein}}} - pK_{a_{\text{intrinsic}}})
\]  

(1.6)

Anderson, et al. found that in T4 lysozyme, His 31 has a pK$_{a}$ value of 9.1 in the native state and 6.8 in the unfolded state. Similarly, the pK$_{a}$ of Asp 70 is shifted to 0.5 in the native state and to a value of 3.5-4.0 in the unfolded state. The pK$_{a}$ values of His and Asp are greatly perturbed from their intrinsic values. These shifts show that the salt bridge formed between these two residues stabilizes the protein by approximately 3-5 kcal/mol\textsuperscript{24}. It is important to note that this free energy is not Equation 1.6. The delta pK$_{a}$ is not the difference between the pK$_{a}$ value of the group in a particular state of the protein and its intrinsic value but rather the difference between the pK$_{a}$ value of the group in the native state and unfolded state. This idea of coupled (un)folding and proton release can be illustrated as a thermodynamic cycle too and will be discussed further in Chapter 2.

How small or large the shift of a pK$_{a}$ value away from its intrinsic value will depend on where the ionizable group is located on the protein. Many experimental studies of these perturbations reported have shown that the shifts are usually small if the groups are on the surface of the protein but significantly large if these groups are
partially or fully buried in the protein interior, and it is important that we understand why.

### 1.4 Molecular determinants of perturbed pK$_a$ values in proteins

In proteins, the three main factors that will perturb the pK$_a$ values of ionizable groups from their intrinsic values are charge-charge interactions, charge-dipole interactions, and burial in a hydrophobic environment (Born effect). Protein reorganization, ionic strength, and metal ions are known to influence pK$_a$ values as well.

#### 1.4.1 Charge-charge interactions

If the cationic or neutral acid described in Equations 1.1 and 1.2 were positioned close to a positive charge, the reaction would shift to the right favoring the deprotonated state of the group and decrease its pK$_a$. The opposite would occur if these groups were placed in a negatively charged environment. The reaction would shift to the left favoring the protonated state of the group and increase its pK$_a$. There are many examples of the effect of Coulombic interactions on pK$_a$ values in the literature such as the active site Cys in Ubc13, a ubiquitin-containing enzyme, that has an elevated pK$_a$ of 11.1 because it is surrounded by acidic residues$^{25}$. Another example is Glu 73 that participates as a general base at the catalytic site in barnase and has a pK$_a$ value of 2.1$^{26}$. This pK$_a$ is depressed
because it is positioned near three positively charged groups: Lys 27, Arg 83, and Arg 87. Laurents, et al. showed that charge-charge interactions are the chief determinants of the pKₐ values of exposed ionizable residues on the surface of ribonuclease Sa.

Charge-charge interactions can be short-, medium-, and long-ranged. Short attractive interactions (<5 Å) between nearby oppositely charged groups are called ion pairs or salt bridges. Attractive ion pairs have been shown to contribute favorably to protein stability such as the Lys 11 – Glu 34 salt bridge in ubiquitin. The CutA1 protein from hyperthermophile Pyrococcus horikoshii is very stable with a Tₘ of approximately 150°C at pH 7. Thirty ion pairs have been found in the monomer of P. horikoshii and are thought to be critical to its stability.

The force between a charged cationic or neutral acid and another charge can be calculated as a function of the distance between them using Coulomb’s law:

\[
F(r) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{\varepsilon \varepsilon_0^2 r^2}
\]

where the \(q_i\) and \(q_j\) is the charge on the cationic or neutral acid and other charged group respectively, \(\varepsilon_0\) is the electric permittivity of free space, \(\varepsilon\) is the dielectric constant of the medium, and \(r\) is the distance between the two charges. Since energy is the integral of work over distance, the work extracted or expended to bring two charges within a
certain distance from one another or separate them is the integral of the Coulombic force from infinity to a given distance \( r \). Consequently, the potential energy between the two charges can be calculated by

\[
\Delta U(r) = -\int_r^\infty F(r)dr = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{er_{ij}}
\]

(1.8)

One can use Equation 1.8 to calculate crude estimates of electrostatic energies in proteins but this is challenging because the intercharge distance, \( r \), is actually a distribution of values that depends on the ensemble of structures of the protein in solution. Also, the dielectric constant of the medium surrounding charges is not uniform but can vary dramatically from that of free solution\(^{14,32}\). (For example, when two charges are 4.2Å apart on the surface of a protein, solvent exposed (\( \epsilon=78 \)), \( \Delta U = 1.0 \text{ kcal/mol} \) versus when they are in the interior of protein (\( \epsilon=4 \)), \( \Delta U = \sim20 \text{ kcal/mol} \).) This latter challenge is one of the main reasons why electrostatic modeling of proteins and creating accurate potential maps is so difficult.

1.4.1.1 Electrostatic modeling of charge-charge interactions

Suppose we have a positive charge \( Q \) that produces an electric field, \( \vec{E} \), over all space (where we assume \( \epsilon=1 \)), which is depicted in Figure 1.3 by the arrows that can be connected to form lines. This field is greater closer to \( Q \) and weaker farther away from
Q. We know this because of the density of electric field lines, which is the number of lines crossing a surface perpendicular to the lines divided by the area of that surface. For example, let’s draw two circles around Q which represent a small sphere of radius \( r_1 \) and an even bigger sphere of radius \( r_2 \) centered on Q in three-dimensional space (Figure 1.3). A certain number of lines cross the smaller sphere and the same number of lines also cross the bigger sphere but they are less dense. So, the field lines spread as we go further out radially from Q. Let’s say that 1 coulomb of charge will get \( \frac{1}{\varepsilon_0} \) lines. Since we have Q coulombs of charge, we will have \( \frac{Q}{\varepsilon_0} \) lines and if we look at either one of the spheres, the line density will be

\[
\text{line density} = \frac{Q/\varepsilon_0}{4\pi r^2} = \frac{Q}{4\pi \varepsilon_0 r^2} = \vec{E} \tag{1.9}
\]

and will be equal to the strength of the electric field. So, if we wanted to know what the \( \vec{E} \) is at some point some distance away from Q, we could use Equation 1.9 to calculate it. Now, suppose we put a positive charge, q, at that particular point. The force exerted on q by Q is given by

\[
\vec{F} = q \times \vec{E} \tag{1.10}
\]

Thus, the force q feels is the electric field at that point times q. Then, if we substitute Equation 1.9 into Equation 1.10 for \( \vec{E} \), we get Equation 1.7 or Coulomb’s law (Woah!).
Figure 1.3: Electric field produced by positive charge $Q$

The electric field lines spread as you go further out radially from the positive charge $Q$ which is represented by the decreasing length of the arrows.
It is important to note that $\vec{F}$ like $\vec{E}$ is a vector, it has both magnitude and direction, so when two charges have the same sign the force will point outwards and will be positive. When the charges have opposite sign, the force will point inwards and will be negative.

What if we wanted to know the electrostatic potential $Q$ creates at the position of $q$? Because the potential is the integral of the electric field over distance or the work done moving $q$ from infinity to that position, the potential at $q$ will be

$$
\Delta V(\vec{r}) = - \int \vec{E} \cdot d\vec{r} = \frac{q}{4\pi \varepsilon_0 r} 
$$

(1.11)

Note that the charge of $q$ is not involved in Equation 1.11 and the potential is a property related to the electric field itself and not $q$. So, we can think of $Q$ as a source charge and $q$ as a test charge. Figures 1.4a and 1.4b show the equipotential lines that can be drawn around $Q$, which are contour lines that trace lines of equal potential. In comparison to a topographical map, one can think of them as the lines on the map which trace lines of equal altitude of the landscape. If we think of this in three dimensions as shown in Figure 1.4c, Figures 1.4a and 1.4b are two-dimensional slices of the peak which correspond to different heights on the electric field landscape. Then, if we wanted to calculate the electric potential energy or energy of interaction between $q$ and $Q$, we would just simply multiply the potential by the charge $q$, and when we do that we get
Equation 1.8 (Double woah!). In other words, the potential energy is just the force acting on q times the distance q is from Q.

![Figure 1.4: Equipotential lines around a point charge](image)

Equipotential lines can be plotted to get a contour map of the electric field landscape produced by a point charge.

Let us take this example one final step further: What if Q is not just a single positive charge but a system of charges (like a protein!)? All of the charged components of Q will affect q and because of the principle of superposition, the potential at the
position of \( q \) due to a system of charges will be equal to the sum of the charged components individual potentials. Now imagine using a computer program model to calculate the potential at different positions along the surface of \( Q \) where each position has a corresponding electrostatic potential value which could be mapped out and visually represented using a color spectrum where for instance red indicates a low or negative potential energy and blue represents a high or positive potential energy.

One model that has been widely used is based on solving the linearized Poisson-Boltzmann (PB) equation\(^{33-35}\). Imagine, instead of having discrete charges like the previous given example with the positive charge \( Q \), there is a continuous distribution of charge represented by a charge density, \( \rho(\vec{r}) \), which is just the sum over all point sources of charge divided by the volume. An approach to calculating electric potentials is to relate the potential to the charge density which gives rise to it by using two different relationships\(^{31}\): (1) the divergence relationship (which is one of Maxwell’s equations) that relates the electric field to the charge density:

\[
\nabla \cdot \vec{E} = \frac{\rho(\vec{r})}{\varepsilon_0}
\]  

(1.12)
and (2) the gradient relationship (which is true because of another one of Maxwell’s
equations that states $\nabla \times \vec{E} = 0$ or the curl of the electric field is equal to zero) that relates
the electric field to the potential:

$$\vec{E} = -\nabla V(\vec{r})$$  \hspace{1cm} (1.13)

If we substitute Equation 1.13 into Equation 1.12, we get Poisson’s equation$^{36, 37}$

$$\nabla \cdot \vec{E} = \nabla \cdot -\nabla V(\vec{r}) = \nabla^2 V(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon_0}$$  \hspace{1cm} (1.14)

This equation employs two types of fields: $V(\vec{r})$, which is called a scalar field because it is
position-dependent but has no direction; and $\vec{E}$, which is called a vector field because it is
position-dependent but also has a direction at every position. Eq. 1.14 also employs the
operator $\nabla$, or “del”, which is the vector differential operator or three dimensional first
derivative. This operator is used in two different ways in Eq. 1.14. When del is
multiplied by a scalar field, as in $\nabla V(\vec{r})$, the operation is called the gradient; the gradient
operation turns a scalar field into a vector field. When del is dotted on a vector field, as
in $\nabla \cdot \vec{E}$, the operation is called the divergence; the divergence operation turns a vector
field into a scalar field $\left( -\frac{\rho(\vec{r})}{\varepsilon_0} \right)$. The result of crossing del with a vector field is called the
curl; the curl operation turns a vector field into another vector field. $\nabla^2$, called the
Laplacian operator, is the three dimensional second derivative and is the divergence of
the gradient of a function. Expressing the Laplacian in different coordinate systems to take advantage of the symmetry of a charge distribution helps in the solution of the potential.

The solution to Poisson’s equation gives the potential for a given continuous charge distribution in a homogenous medium. The linearized PB equation arises because it allows one to calculate the potential for inhomogeneous mediums

\[ \nabla^2 V(\vec{r}) = \kappa^2 V(\vec{r}) \]  

(1.15)

where \( \kappa^2 \) is the Debye-Hückel screening parameter which describes the exponential decay of the potential in the solvent. According to Debye-Hückel theory\textsuperscript{38, 39}, since ions are free to move around in solution, they will arrange themselves into ion clouds where a high density of ions will surround a certain ion of opposite charge. These “layers” of ions will partially reduce or screen the Coulombic force that the ion feels from other charges located near it. The thickness of this layer is called the Debye-Hückel screening length and is described by

\[ \frac{1}{\kappa} = \sqrt{\frac{\varepsilon \varepsilon_0 k_B T}{2 N_A e^2 I}} \]  

(1.16)

where \( k_B \) is Boltzmann’s constant, \( N_A \) is Avogadro’s number, \( e \) is the elementary charge, and \( I \) is ionic strength. Beyond this length or distance, the ions are considered non-
interacting. In the PB model, solute atoms are treated explicitly as particles with a low dielectric constant (~2-4) with point partial charges at atomic positions. The solvent surrounding the solute is treated as a high dielectric constant (~80) continuum.

1.4.2 Charge-dipole interactions

_Hydrogen Bonds_: Ionizable groups can hydrogen bond or interact with the partial charges or dipoles on neighboring polar groups. Hydrogen bonds generally contribute 1-2 kcal/mol to the stability of a protein\(^{40}\). The effect of hydrogen bonding on a pK\(_a\) value will depend on whether the interaction is more favorable with the deprotonated state of the group which again will lower the pK\(_a\) or if it is more favorable with the protonated state, the pK\(_a\) will be elevated. For His 18 coordinated to the iron in horse cytochrome c, its pK\(_a\) value is 2.4. This extremely and unusually low pK\(_a\) is due to the group being buried and that it forms two hydrogen bonds to Ala 15 and Pro 30\(^{41}\). The side chain carboxyl of Asp 76 in ribonuclease T1 also has a very low pK\(_a\) value, 0.6\(^{42}\). This was hypothesized to be due to the three intramolecular hydrogen bonds that it forms to the side chains of Asn 9, Tyr 11, and Thr 91. To determine whether these hydrogen bonds were responsible for lowering the pK\(_a\), the hydrogen bonds were removed one at a time and the pK\(_a\) of Asp 76 was measured. When only one of the hydrogen bonds is removed,
the pKₐ increases to approximately 3.3. When two hydrogen bonds are removed, the pKₐ increases to about 5.1. When all three of them are removed, the pKₐ increases to 6.4. Therefore, when there are no hydrogen bonds present, the pKₐ is elevated 2.6 units above the intrinsic pKₐ of Asp. Moreover, this increase in pKₐ is because of the Born effect because Asp 76 is buried and shielded from the negative net charge of the protein near neutral pH.

**Helix Dipoles:** The alpha-helices in proteins have a dipole moment due to the alignment of the dipoles of the peptide units and can shift pKₐ values. The dipole of a helix runs from the negatively charged C-terminal to the positively charged N-terminal. An alpha helix can interact with surrounding groups in two different ways. First, the dipoles of the peptide units of the helix are available to hydrogen bond with nearby molecules. Second, “the helix dipole effect” can be viewed as an electrostatic effect where a partial net charge exists at the helix termini and nearby groups are perturbed by this field. Sali, et al. showed that the proton attached to the Ne atom on the imidazole ring of His 18 makes a hydrogen bond with the backbone carbonyl oxygen of Gln 15 residue in barnase. A study by Perutz, et al. provided evidence of the second perspective where the pKₐ of His 97 in hemoglobin is 7.8 compared to 6.6 which is
characteristic of free histidines on the surface of proteins. This high pKₐ is accounted for by the protonated histidine’s interaction with the negative pole at the C-termini of helices F and FG.

1.4.3 Born effect

In the interior of a protein, an ionizable group is not exposed to solvent (i.e., water) like it would be if it were located on the surface of the protein. The dielectric constant of this hydrophobic environment is much lower in comparison with the dielectric constant of water (~78 at 25°C) and it is energetically unfavorable for an ionizable residue to be charged. The neutral state of the group will be favored and the pKₐ values of cationic acids will be depressed and those of neutral acids will be elevated when they are partially or completed buried. The Born formula can be used to calculate a simple estimate of Gibbs free energy of solvation described as the work obtained by transferring a charged sphere of radius a from vacuum to a medium with a dielectric constant ε:

\[
E_{\text{solv}} = \frac{N_A Z^2 e^2}{8\pi \varepsilon_0 a} \left(1 - \frac{1}{\varepsilon}\right) \quad (1.17)
\]

where \(N_A\) is Avogadro’s constant, \(Z\) is the charge of the ion, \(e\) is elementary charge, and \(\varepsilon_0\) is the electric permittivity of free space. Thus, the free energy required to transfer a
cationic acid (single charged ion) with a radius of 2Å from water to the interior of a protein would be $\Delta G_{\text{transfer1}} = E_{\text{solv}}(\text{Pinterior}) - E_{\text{solv}}(\text{water}) = 20 \text{ kcal/mol}$ and can be described by the thermodynamic cycle depicted in Scheme 1.2. This calculation explains why a charged group is rarely observed buried in a nonpolar environment.

Studies of staphylococcal nuclease (SNase) illustrate the Born effect in proteins. When Val 66, which is located in the hydrophobic core of SNase, is mutated to Asp, the carboxyl group has a pK$_a$ value of 8.9. This is approximately 5.0 units higher than its
intrinsic pK$_a$. When Val 66 is replaced with a Lys, the amino group has a pK$_a$ of 5.5 which is approximately 5.0 units lower than its intrinsic pK$_a$ value. Garcia-Moreno and coworkers concluded that “the shift in the pK$_a$ value of Asp 66 is governed by the loss of hydration of the carboxylic group in the buried state that is not offset by interactions with charges or with polar atoms of the protein”$^{46}$. The highest pK$_a$ reported for an Asp residue is 9.2 for Asp 26 in *E. coli* thioredoxin where the side chain is completely buried near the active site$^{47}$. The highest pK$_a$ measured for a His residue is 9.2 for His 72 in bovine protein tyrosine phosphatase. This elevated pK$_a$ is due to the imidazole group being buried and participating in charge-charge interactions with several nearby negatively charged groups$^{48}$.

### 1.4.4 Protein reorganization

A more subtle effect that can shift a pK$_a$ is protein reorganization when the protonation state changes. This effect for ionizable side chains that are solvent exposed is small in comparison to reorganization of the solvent. However, for side chains that are partially or fully buried, this effect can be greater. This reorganization can be small and consist of just slight reorientations of a polar group or can lead to a large conformational change like the one observed in hemoglobin due to the Bohr effect$^{44}$. 

---

25
1.4.5 Ionic strength and the Debye-Hückel effect

The ionic strength of the solution does influence measured $pK_a$ values of ionizable groups. If we take Equation 1.3 and rewrite it in terms of activities, we get

$$K_a = \frac{a_A a_{H^+}}{a_{A^+H}}$$

For a neutral species, $a_A$ can be replaced with $[A]$. If we rearrange the terms and use the definition of activity, $a_{H^+A} = \gamma[A^+H]$, the above equation becomes

$$a_{H^+} = K_a \frac{\gamma[A^+H]}{[A]}$$

Then, if we take the negative logarithm of this expression, we get

$$pH = pK_a + log \frac{[A]}{[A^+H]} + log \gamma_{A^+H}$$

(1.18)

The Debye-Hückel limiting law\textsuperscript{38,39} allows us to determine the activity coefficient of an ion in a dilute solution at low ionic strength:

$$log\gamma_i = -\frac{1.824 \times 10^6}{\varepsilon \tau^{3/2}} z_i^2 \sqrt{I}$$

(1.19)

where $z$ is the charge number of ion species $i$, $I$ is the ionic strength of the solution, and $\varepsilon$ is the dielectric constant, and $T$ is temperature. Because ions in solution act together, the activity coefficient obtained using Equation 1.19 is a mean activity coefficient. If the
solution being studied is aqueous (ε=78) and at room temperature (T=25°C=298K),

Equation 1.19 reduces to

$$\log \gamma_i = -0.509z_i^2 \sqrt{I}$$

Ionic strength can be calculated using the following equation:

$$I = \frac{1}{2} \sum z_i^2 c_i$$  \hspace{1cm} (1.20)

where c is molar concentration of ith ionic species present and z is its charge. According to Debye-Hückel theory and by looking at Equations 1.18 - 1.20, we can see that activity coefficients, γ, of different species in solution depend on ionic strength and pK_a depends on activity coefficients, which means that ionic strength will affect pK_a. Specific examples of how ionic strength perturbs pK_a will be discussed in Chapter 2.

1.4.6 Metal ions

Proteins coordinate metal ions with sulfur, nitrogen, and oxygen biological ligands. The bond between a metal and ligand depends on many things including the nature of the valence orbitals of the ligands, the effective nuclear charge, coordination number, and geometry of the metal ion. The ability of a metal ion to effectively compete with a proton in ligand binding strongly depends on the strength of the metal-ligand bond. Metals influence pK_a values and can even bind ligands at pH values significantly
below their pKₐ values. For example, coordination at the deprotonated nitrogen atom of the histidine imidazole side chain lowers the pKₐ value of the protonated nitrogen by about 2 log units due to an inductive effect.

1.5 Methods for the determination of pKₐ values

Over the past century, there have been numerous methods developed to determine pKₐ values (Figure 1.5). Only a few of them will be discussed here.

*Potentiometry:* Potentiometry is one of the most commonly used techniques for pKₐ determination. In a potentiometric titration, a known quantity of reagent like NaOH or HCl is added stepwise to a solution of analyte. The change in hydrogen ion activity upon reaction is measured with a pH electrode. Then, by plotting the hydrogen ion activity versus the volume added, a sigmoidal curve such as the one in Figure 1.2 is created where the midpoint of the curve is equal to the potential at equilibrium. This hydrogen ion activity can be converted by using standards with known pH. Two major drawbacks of potentiometric titrations are that they require a large amount of sample and the pH electrode readings can be affected by high temperatures, extreme pH values, and nonpolar-polar solvent mixtures.
Figure 1.5: Timeline of different methods developed for the determination of pKₐ values
**Voltammetry:** In contrast to potentiometry, voltammetry is not widely used to determine pKₐ values. Actually, its use for this particular reason did not begin until the 1960s. In this method, a changing voltage is applied over the sample solution and the resulting current is measured. When the voltage reaches the reduction potential of the analyte, the current will increase, and then decrease because of the depletion of the analyte. A reference with known electrochemical behavior is usually added to the sample and the shifts of the peaks of the reference upon addition of the base or acid is then used to determine the pKₐ value. One of the challenges of this technique is the need for an electro-active molecule which is soluble in a conductive solvent. Nevertheless, it is more useful when measuring pKₐ values in less polar solvents, which is more difficult to do by potentiometric titration.

**Calorimetry:** There are two calorimetric techniques that have been used for the measurement of pKₐ values: Isothermal Titration Calorimetry (ITC) and Isothermal Titration Microcalorimetry (ITM). For the former, an acid-base titration is conducted inside the calorimeter while the energy needed to keep the temperature constant is measured. Alternatively, ITC has been used to monitor the time derivative heat change resulting from an event such as a conformational change or ligand binding. The data
obtained can then be integrated with respect to time to generate an isotherm for the reaction. Thermodynamic parameters such as enthalpy change, ΔH, can be extracted and used to indirectly calculate the pKₐ. For ITM, the heat released upon reaction is measured in buffer solutions with different pH values. Then, a sigmoidal curve is created by plotting the minima or maxima versus pH. The amplitude of the minimum or maximum at different pH values is proportional to the degree of dissociation.

**UV/Vis spectrometry:** In 1966, Wigler, et al. were the first to use ultraviolet (UV) absorption spectroscopy to measure the pKₐ values of di-protic compounds like 3-hydroxypyridine. For a UV pH measurement, the presence of a chromophore near the ionization site is necessary. The spectra of the protonated and deprotonated group will differ and any wavelength with the exception of the isosbestic point can be used for the determination of the pKₐ. Also, if the absorption of two different wavelengths is measured at a variable pH and the ratio in absorption at those two wavelengths is plotted against pH, the pKₐ can be determined from the sigmoidal curve obtained. One of the wavelengths has to belong to the chromophore and the second is a reference that does not change as a function of pH.
**Nuclear magnetic resonance:** NMR is a very powerful technique because it offers a direct and site-specific method to observe different nuclei (such as $^1$H, $^{13}$C, $^{31}$P) and determine acid dissociation constants of ionizable groups. Not only simple small molecules such as free amino acids but also even more complex molecules such as enzymes and large proteins can be characterized with NMR\textsuperscript{57,58}. Initially, NMR was used to identify the site of deprotonation of an acid by decreasing the temperature of the sample to slow down the proton equilibrium on the NMR timescale in order to observe two distinct peaks for the protonated acid and its conjugate base. In later years, the technique was used to calculate pK\textsubscript{a} values. A mean chemical shift, $\delta_{obs}$, or populated weighted average of the protonated acid and its conjugate base is observed when the pH dependent equilibrium is fast on the NMR timescale. For these conditions, Equation 1.4 can be rewritten as:

$$pH = pK_a + \log\left(\frac{\delta_{obs} - \delta_{HA}}{\delta_A - \delta_{obs}}\right)$$

Equation 1.21 can then be transformed into the Hill form of the Henderson-Hasselbalch equation:

$$\delta_{obs} = \frac{\delta_{HA} + \delta_A(10^{pH-pK_a})}{1 + 10^{pH-pK_a}}$$

Equation 1.22
where \( n \) is the Hill coefficient\(^{59}\). If Equation 1.22 is used to fit the titration curve generated by plotting chemical shift versus pH, the pK\(_{a}\) can be determined.

Unfortunately, one of the disadvantages of NMR is that it does require expensive and sophisticated instrumentation.

**1.6 Overview of work**

This dissertation consists of four additional chapters, each describing experimental investigations into the contributions of electrostatic interactions to the thermodynamic stability and folding of proteins using the protein subunit of *Bacillus subtilis* ribonuclease P, P protein, as a model system. Electrostatic interactions, as mentioned previously, are known to affect protein stability and can be both stabilizing and destabilizing. They depend on the geometry and separation of the residues, the degree of exposure to the solvent, and on neighboring residues. One way to probe electrostatic interactions is to measure the pK\(_{a}\) values of ionizable groups of proteins and investigate their molecular determinants. The types of interactions these ionizable groups will have will depend on their protonation state. Hence, the pK\(_{a}\) values of charged groups are valuable reporters of their electrostatic microenvironment.
The experiments described in Chapter 2 examine the role electrostatic forces play in the energetics of P protein. NMR spectroscopy was used to measure the pKₐ values of the three histidine residues in all three states of P protein and assess the molecular perturbants of these pKₐ values. The experiments in Chapters 3 and 4 address the effects of salt and dielectric constant on pKₐ values, respectively. The results from these studies can be used to help guide the development of computational models that can accurately calculate salt and dielectric constant effects in proteins. NMR spectroscopy was used to measure the pKₐ values of the histidine residues in P protein and small model compounds under different solvent and salt conditions. In Chapter 5, circular dichroism and fluorescence spectroscopy were used to monitor the equilibrium folding of P protein. The results of these experiments along with the findings of Chapters 2 and 3 are used to understand the pH-dependent stability of P protein and gain a more complete description of its folding intermediate.
2. Electrostatic energetics of *Bacillus subtilis* ribonuclease P protein determined by nuclear magnetic resonance-based histidine pKₐ measurements

2.1 Introduction

Electrostatic interactions are an important determinant of biomolecular structure and function and play a central role in modulating stability. Ionizable group pKₐ values provide a sensitive and specific probe of the electrostatic component of the energetics of salt bridges; buried charged groups; catalytic groups in enzyme active sites; and protein conformational equilibria. The pH dependence of protein stability is reflected in differences in pKₐ values of charged groups in the folded and unfolded states. Histidine residues in particular play a critical role in protein folding and function. With pKₐ values near the physiological pH range, they can act as an acid or base and exist in either neutral or positively charged states. Thus, the protonation state of histidine actively controls charge balance and in many cases is strongly coupled to the stability of a protein.

Addition of a proton to an ionizable group in a protein is equivalent to a ligand binding reaction, where the ligand is a proton. The effect of this binding on protein
stability results from thermodynamic coupling of folding and binding reactions, as depicted in Scheme 2.1. In this scheme, C₁ and C₂ represent alternative protein

conformations and C₁H⁺ and C₂H⁺ represent the proton-bound forms of each. Numerous experimental studies have examined the effect of proton binding on the stability of proteins by measuring pKₐ values⁴,⁵⁸,⁶⁵,⁶⁶. To date, most of these NMR investigations have focused on the native state and less on the unfolded state because its low population under physiological conditions makes observation in standard experiments difficult. Denaturing conditions and peptide fragments have been used as a model for the unfolded state and demonstrate that pKₐ values measured with simple model
compounds do not accurately represent the values in the unfolded state\textsuperscript{47, 68}. Tollinger, et al. measured both unfolded and folded pK\textsubscript{a} values for the N-terminal SH3 domain from the \textit{drk} protein under non-denaturing conditions taking advantage of the fact that it exists in an approximate 1:1 equilibrium between native and unfolded conformations at neutral pH\textsuperscript{69}. Here, we take advantage of the intrinsically-unfolded (IUP) property of the protein subunit of \textit{Bacillus subtilis} RNase P holoenzyme (P protein) in order to measure the histidine pK\textsubscript{a} values (H3, H22, and H105) of its unfolded (U), partially folded (I), and sulfate-bound folded (F) states. The locations of these groups are shown in Figure 2.1. The histidine residues are solvent accessible and well distributed on the protein in all three conformational states.

Like many other RNA-binding proteins, P protein is highly charged. It is comprised of 11 Arg, 19 Lys, 3 His, 4 Asp and 9 Glu residues. P protein is predominantly unfolded when not bound to its cognate RNA or small molecule anions. Under these low ionic strength conditions the population of U and I is \textasciitilde 85\% and 15\%, respectively, with no detectable folded protein.\textsuperscript{70} The tendency to unfold to U or I under these conditions was previously attributed to the large net charge, which would produce more unfavorable electrostatic interactions in the more compact folded state than in the
Figure 2.1: Electrostatic surface potential of the sulfate-bound folded state of *B. subtilis* RNase P protein

(Protein Data Bank entry 1A6F)\(^7\) (A) The front and side views of the electrostatic surface potential of P protein show that this protein is highly basic. Positively and negatively charged surfaces are colored blue and red, respectively. (B) The front and side views of the ribbon model of P protein show the three histidine residues in green—His 3, His 22, and H105—and a close-up of the sulfate binding site located between the second α-helix and N-terminus. The electrostatic surface potential was calculated using the program APBS\(^7\) and visualized in Chimera\(^7\).
unfolded state. Addition of various small anions (e.g., sulfate) or the osmolyte trimethylamine N-oxide (TMAO) induce the protein to fold to its native α-β structure. Circular dichroism, NMR paramagnetic relaxation enhancement, and stopped-flow fluorescence experiments along with isothermal titration calorimetry established that P protein has two high-affinity anion binding sites. This property provides experimental control of the conformation of P protein, thereby allowing determination of the pKₐ values of individual conformational states.

Our previous kinetic and equilibrium studies of TMAO-induced folding of P protein as well as NMR experiments demonstrated the presence of a partially folded intermediate in its folding mechanism. ¹H-¹⁵N HSQC NMR spectra collected as a function of pH showed increasing pH, from 5.0 to 7.0 in steps of 1 pH unit, shifts the U-I conformational equilibrium toward I in the absence of sulfate and the I-F equilibrium toward F in the presence of sulfate. These effects were attributed to the unfavorable electrostatic interactions in the more compact forms of the protein. Stopped-flow studies showed that pyrophosphate ligand binding to P protein shifts the conformational equilibrium toward I and F by both increasing the folding rate constants
and decreasing the unfolding rate constants. The order in which binding and conformational change occur depends on ligand concentration.

In the work described here, we applied an accurate and precise NMR titration approach using internal reference compounds to determine the pKₐ values of the three histidines in RNase P protein in the U, I and F conformational states. Because proton binding affinities are coupled to conformational change, it is possible to estimate the electrostatic contribution to the conformational energetics from pKₐ values. We conclude that the effect of electrostatics on conformational equilibrium, while detectable, is insufficient to fully explain the IUP property of P protein.

2.2 Results and discussion

2.2.1 Measuring pKₐ values without a pH electrode

Although they are relatively straightforward to conduct, conventional NMR-monitored pH titration experiments require multiple samples and a pH measurement before and after acquisition of each spectrum. The titration experiments presented here do not require pH measurement because an internal standard or reference acts as an NMR-detected pH meter. Several compounds have been used in the past to monitor the pH of a biomolecular NMR sample. Using this approach, only one sample is needed.
to generate an entire titration curve and it is not removed from the NMR tube between points. This procedure conserves protein and allows the use of a higher concentration NMR sample. The pK$_a$ of the reference must be measured separately in the same buffer used for the protein titration and can be obtained by fitting chemical shifts versus pH (as measured by a carefully calibrated pH electrode in a ~30 mL solution) to Equation 2.2 (see Materials and Methods). Once the chemical shifts of the desired residues and reference from the protein titration are determined, the resulting parametric plots can be fit to determine protein pK$_a$ values using the following equation:

$$\delta_p(\delta_r) = \frac{n_A^{pK_A-pK_R}(\delta_r-\delta_{RH})^{n_A^{pK_R}}} {1+10^{(pK_A-pK_R)(\delta_r-\delta_{RH})^{n_A^{pK_R}}}} + \delta_{AH}$$

(2.1)

where $\delta_A$ and $\delta_{AH}$ are the plateau values of the protein histidine chemical shifts in the basic and acidic pH limits, respectively; $\delta_R$ and $\delta_{RH}$ are the plateau values of the reference chemical shifts in the basic and acidic pH limits, respectively; $\delta_p$ and $\delta_r$ are the observed chemical shifts of the protein histidine and reference, respectively; $n_R$ and $n_A$ are the Hill coefficients for the reference and protein histidine, respectively; and $pK_R$ and $pK_A$ are the dissociation constants for the reference and protein histidines, respectively. ($pK_R$ and $n_R$ were fixed to the values obtained from the fits of the internal standard.
Figure 2.2: Titration curves following the proton resonance of pyridine in the absence of 20 mM sodium sulfate

The black lines are the best fitting results using (A) Equation 2.2 and (B) Equation 2.1 along with the previously calculated pKₐ of internal standard L-histidine of 6.08 ± 0.01 (see Figure S14). The estimated pKₐ of pyridine was (A) 5.34 ± 0.01 and (B) 5.35 ± 0.04.
shown in Figure 2S.) These fits reveal the difference between the pKₐ of the protein residue and pKₐ of the reference (ΔpKₐ = pKₐ^P - pKₐ^R). If the two pKₐ values are the same, a plot of the protein versus reference chemical shifts is a diagonal line with a slope = 1. If the two pKₐ values are different, there is a deviation from linearity whose curvature (concave or convex) depends on the sign of ΔpKₐ (see inset of Figure 2.2B). It is important to note that the internal standard chosen must have a pKₐ value that is close (within ± 1 pH unit interval) to the pKₐ value of the protein ionizable group or the parametric curve will be too steep to allow accurate estimation of the pKₐ. This method differs from the closest previous method in that the chemical shifts of the titrating standard are directly used in a parametric equation (Equation 2.1) to determine the unknown pKₐ values. Also, all acid base equilibria must be in the fast exchange regime in order for the chemical shift equations to hold as written. The equilibria observed here were in fast exchange because neither multiple resonances nor extensive line broadening due to intermediate exchange was observed.

One of the advantages of using an internal standard to measure sample pH rather than a pH electrode is that the apparent Hill coefficient of the histidine of interest is not influenced by errors in pH measurement, which can lead to non-unitary Hill
coefficients in cases where there is no neighboring titratable charge. As long as the $pK_a$ and $n$ values of the internal standard reference compound are well known, the Hill coefficient estimated from fits of parametric plots of protein His chemical shift versus internal standard chemical shift to Equation 2.1 should accurately reflect the electrostatic environment of the protein His.

### 2.2.2 $pK_a$ values of pyridine determined from two methods are in good agreement

To establish the feasibility of the approach described above, we first measured the $pK_a$ values of L-histidine and pyridine in the same buffer used for the P protein titration using 1D $^1$H NMR. First, $pK_a$ values were obtained in the conventional fashion by making separate samples whose pH was determined by pH electrode. The pyridine $^1$H-C$_{3,5}$ chemical shift for each sample was plotted versus pH and fit to Equation 2.2. In addition, the $^1$H-C$_{3,5}$ chemical shift for pyridine was plotted versus the $^1$H-$^1$3C$_{11}$ chemical shift for L-histidine and fit to Equation 2.1. The $pK_a$ values determined for pyridine from the two methods are in very good agreement. The pH electrode measurement gave a pyridine $pK_a$ value of 5.34 ± 0.01 and our method gave 5.35 ± 0.04 (Figure 2.2). Therefore, having established that our approach provides accurate $pK_a$ values, we next focused on the three histidines in P protein.
2.2.3 P protein His pK\textsubscript{a} values determined without a pH electrode

Chemical shifts of the \textsuperscript{1}H-\textsuperscript{13}C resonances of P protein histidines were determined by \textsuperscript{1}H-\textsuperscript{13}C HSQC NMR spectroscopy, over a pH range of 4.0–8.0 using a series of 0.05-0.4 μl increments of NaOH, added via a 5.0 μl positive displacement syringe whose needle was extended with polyethylene tubing long enough to reach below the meniscus of the NMR sample (see Figure S13). Care was taken to assure that all of injected NaOH solution reached the sample, which was subsequently well mixed by centrifugation. The frequencies of both nuclei (\textsuperscript{1}H and \textsuperscript{13}C\textsubscript{ε1}) were recorded for each protein histidine. The reference L-histidine and the resulting parametric correlations were fit to Equation 2.1 to determine the pK\textsubscript{a} values of the protein histidines. The absolute pK\textsubscript{a} values of the protein histidines were determined using a pK\textsubscript{a} value of the internal standard, L-histidine, measured using 1D \textsuperscript{1}H NMR (Figure S14). Figure 2.3 shows the protein histidine \textsuperscript{1}H chemical shifts versus L-histidine chemical shifts titration curves for sulfate-bound folded (A) and unfolded (B) P protein. For \textsuperscript{1}H\textsubscript{ε1} chemical shifts, if the pK\textsubscript{a} of the protein histidine is lower than that of L-histidine, the plot will curve below the diagonal line and if it is greater, it will curve above. The opposite trends are observed with the \textsuperscript{13}C\textsubscript{ε1} plots (Figure S15). A total of 30 and 31 spectra were collected for
Figure 2.3: $^1$H$^1$ P protein histidine chemical shift vs $^1$H$^1$ L-histidine chemical shift titration curves

$^1$H$^1$ P protein histidine chemical shift vs $^1$H$^1$ L-histidine (internal standard) chemical shift titration curves for (A) sulfate-bound state and (B) unfolded P protein.
the pH titrations of sulfate-bound folded and unfolded P protein, respectively. NMR experiments with histidine to alanine substitution mutants were used to assign residue numbers to each of the three histidine resonances in the F and U state spectra of P protein (see Figure S16).

2.2.4 The pKₐ values of H3, H22, and H105 are depressed in U and F

Measured pKₐ values are sensitive probes of the electrostatic environment of a protonatable group. If a protonated histidine side chain is involved in a favorable electrostatic interaction such as a salt bridge, it will be harder to deprotonate and will have a higher pKₐ value than histidines in an uncharged environment. On the other hand, if the histidine side chain is near other positive charges, it will be easier to deprotonate and will have a lower pKₐ value. Nonspecific Coulombic interactions experienced by the protonated histidines in a polypeptide with a net charge of +17 should be repulsive and result in depressed pKₐ values. As a model of a histidine side chain in an electrostatically neutral environment, we used N-acetyl-L-histidine methylamide. The pKₐ values measured for this compound were 6.44 ± 0.02 and 6.52 ± 0.03, in the two buffers used in this study (Figure S17). When compared to the model compound N-acetyl-L-histidine methylamide, all three histidines have pKₐ values that
Table 2.1: Histidine pKₐ values and Hill coefficients of N-acetyl-L-histidine methylamide and three histidines in unfolded, intermediate, and sulfate-bound folded P protein

<table>
<thead>
<tr>
<th>Buffer Conditions</th>
<th>State pKᵢ</th>
<th>State Hill Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Na₂SO₄]</td>
<td>20 ± 1 mM</td>
<td>0</td>
</tr>
<tr>
<td>[NaCl]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ionic Strength †</td>
<td>80 ± 1.4 mM</td>
<td>20 ± 1 mM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Histidine</th>
<th>pKᵢᴺ</th>
<th>pKᵢᴵ</th>
<th>pKᵢᵁ</th>
<th>pKᵢᵀ</th>
<th>pKᵢᴺ</th>
<th>pKᵢᴵ</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Ac-His-NH-Me</td>
<td>6.52 ± 0.03</td>
<td>6.44 ± 0.02</td>
<td>6.44 ± 0.02</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>His 3</td>
<td>6.31 ± 0.01</td>
<td>5.71 ± 0.02d</td>
<td>5.71 ± 0.02d</td>
<td>6.16 ± 0.02</td>
<td>6.25 ± 0.01</td>
<td>6.23 ± 0.03</td>
</tr>
<tr>
<td>His 22</td>
<td>6.03 ± 0.01</td>
<td>5.41 ± 0.25</td>
<td>5.99 ± 0.02</td>
<td>5.95 ± 0.02</td>
<td>6.11 ± 0.03</td>
<td>5.99 ± 0.01</td>
</tr>
<tr>
<td>His 105</td>
<td>5.52 ± 0.01</td>
<td>5.76 ± 0.02d</td>
<td>5.76 ± 0.02d</td>
<td>5.46 ± 0.01</td>
<td>5.61 ± 0.01</td>
<td>5.73 ± 0.02</td>
</tr>
<tr>
<td>Histidine</td>
<td>nᴺ</td>
<td>nᴵ</td>
<td>nᵁ</td>
<td>nᵀ</td>
<td>nᴺ</td>
<td>nᴵ</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>N-Ac-His-NH-Me</td>
<td>1.09 ± 0.05</td>
<td>1.09 ± 0.05</td>
<td>1.09 ± 0.05</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>His 3</td>
<td>1.02 ± 0.04</td>
<td>1.04 ± 0.05d</td>
<td>1.04 ± 0.05d</td>
<td>0.99 ± 0.03</td>
<td>0.93 ± 0.02</td>
<td>1.03 ± 0.07</td>
</tr>
<tr>
<td>His 22</td>
<td>1.07 ± 0.03</td>
<td>1.09 ± 0.21</td>
<td>1.02 ± 0.05</td>
<td>1.00 ± 0.03</td>
<td>0.95 ± 0.05</td>
<td>1.03 ± 0.03</td>
</tr>
<tr>
<td>His 105</td>
<td>1.01 ± 0.03</td>
<td>1.04 ± 0.05d</td>
<td>1.04 ± 0.05d</td>
<td>0.98 ± 0.02</td>
<td>0.95 ± 0.03</td>
<td>0.98 ± 0.05</td>
</tr>
</tbody>
</table>

†Solutions contained 10 mM L-histidine, 10 mM pyridine in 90% H₂O, 10% D₂O at 25 °C with either 0 mM or 20 mM sodium sulfate. The pKᵢ values were obtained by fitting the data to Equation 2.1. The ¹H ¹H shifts were used to determine the pKᵢ values. Errors represent 95% confidence limits and were obtained as described in Materials and Methods. †Ionic strengths were calculated including all buffer components. †N-acetyl-L-histidine methylamide. †Values based on the inability to distinguish U and I resonances for these residues.
are lower in both the U and F states of P protein. The $^{1}H$ and $^{13}C$ NMR chemical shifts were used to determine the P protein histidine pK$_a$ values shown in Tables 2.1 and S1, respectively. The Hill coefficients, which reflect deviations from ideal single-site titration and errors in pH are presented in Tables 2.1 and S1. These pK$_a$ values are similar to those measured in the folded state of another positively charged protein, staphylococcal nuclease (SNase) (pI > 10) at low ionic strength$^{83}$. The observed pK$_a$ depressions are also in contrast to the negatively charged protein RNase T1 (pI = 3.5) where all three of its histidine residues have elevated pK$_a$ values ranging from 7.2 to 7.8$^{84}$.

P protein was examined over a relatively small pH range, 4.0-8.0, corresponding to the expected titration range for histidine, glutamate, and the N-terminus. The average pK$_a$ values for glutamate and the N-terminus are 4.2 and 7.7, respectively$^{85}$. These values are more than 1 pH unit away from any of the pK$_a$ values measured for the P protein histidines. For this reason, we conclude that the titration of glutamate residues and the N-terminus does not significantly affect the pK$_a$ measurements of the histidines.

2.2.5 Electrostatic repulsion is greater in the more compact folded state

H105 titrates with a pK$_a$ value that is 0.3 units lower in F than in U, as expected because the positive charge density should be greater in the more compact and less
dynamic F state. As seen in Figure 2.1, H105 is located in the folded structure on the third α-helix and its surrounding microenvironment is electropositive. H105 is the least solvent exposed of the three histidine residues in F, which places it in a lower dielectric environment (Table S2). This amplifies the charge-charge interactions between itself and other charged groups like K109 (6.3 Å away in F) and contributes to its lower pK$_a$ value.

2.2.6 P protein histidine side chains are well-solvated in both U and F

Although the pK$_a$ values of the three histidines in P protein are depressed in all states, the depression is not as great as might be expected for such a basic protein. This moderate suppression could be explained by effective electrostatic screening by solvent water, which would minimize electrostatic interactions with other charges on the protein. Two lines of evidence support this conclusion. First, the Hill coefficients we measured for all three histidines in all states are very close to 1, indicating that the co-titration of the other histidines and any carboxylates with elevated pK$_a$’s has no effect on the protonation of each His site. Second, the tautomeric states of all three histidines in both U and F are nearly identical to that of N-acetyl-L-histidine methylamide, which presumably only interacts with water to determine its tautomer equilibrium (Figure S21). These results strongly suggest that P protein histidines are surrounded by water
molecules whose dipole moments effectively screen the charged form of the side chain from strong electrostatic interactions with the rest of the protein. This effect minimizes the energetic perturbation caused by protonation of the histidines. If such strong screening occurs at other charged residues, electrostatics may play a relatively small role in the low stability of unliganded P protein. This conclusion calls into question the usual assumption that a large net charge predisposes a protein to be intrinsically unfolded.

2.2.7 The H3-SO$_4$ salt bridge contributes moderately to the stability of the sulfate-folded state

In contrast to H105, the pK$_a$ of H3 is 0.6 units higher in F than in U, which can be rationalized by examining the crystal structure of P protein shown in Figure 2.1B$^{71}$. H3, along with R9 and R68, coordinate the sulfate ion bound at the anion binding site between the second $\alpha$-helix and the N-terminus in the folded protein. It is apparent that the pK$_a$ of H3 should be increased by the salt bridge it forms with the sulfate ion (3.4 Å away). This region of P protein offers a favorable environment for sulfate association and the crystal structure of Reiter, et al. suggests that its cognate binding partner, P RNA, forms a similar salt bridge with the phosphate backbone of the RNA and the region of the protein where H3 is located$^{86}$. H3 likely plays a key role in both of these binding reactions. However, the difference in pK$_a$ values between U and F only
corresponds to a -0.82 ± 0.02 kcal/mol contribution to the folding free energy due to proton binding at H3 (see Scheme 2.1). This is a relatively small contribution to the stability of the protein and is insufficient to elevate the pKₐ of H3 above that of N-acetyl-L-histidine methylamide and is a relatively small fraction of the stability of the sulfate-bound folded protein. In general, solvent-exposed salt bridges seem to play a relatively small role in stabilizing proteins while buried ion pairs make much larger contributions. The surface salt bridge in ubiquitin between K11 and E34 was estimated to contribute favorably by 0.86 kcal/mol²⁹. In both P protein and ubiquitin the stabilization is modest in comparison to the buried ion pairs found in chymotrypsin and T4 lysozyme that stabilize those proteins by 3 to 5 kcal/mol²⁴, ⁶².

The moderate stabilization by the H3-SO₄ salt bridge is apparently not enhanced by a hydrogen bond between the H3 side chain and a sulfate oxygen, which would affect the tautomeric state populations if it were present. Based on model compounds, deprotonation of histidine to its N⁺H tautomeric form leads to a chemical shift change of ~ -2 ppm (i.e., upfield) for the ¹³C₀², whereas the opposite change of ~ +7 ppm results for formation of the N⁺¹H tautomer⁸⁷. Upfield changes in ¹³C₀² shift observed for H3, H22, and H105 with increasing pH (Figure S21) indicates that the N⁺²H tautomer is
preferentially adopted in both the unfolded and sulfate-bound folded states. Although the crystal structure of the folded protein shows H3 in close proximity to a bound sulfate\textsuperscript{35}, the distance between the N\textsuperscript{\delta1} and the closest sulfate oxygen is 2.9 Å, which is much farther than His N-H—O H-bonds observed in proteins\textsuperscript{46}. Apparently the P protein histidines are well-solvated by water and therefore adopt the same tautomer preferred by the model compound.

**2.2.8 The P protein H22 residue is influenced by an \(\alpha\)-helix dipole**

Unlike H3, the pK\textsubscript{a} of H22 remains unchanged and its pK\textsubscript{a} in F is within uncertainty of its pK\textsubscript{a} in U. H22 is located on the negatively charged C-terminal end of the first \(\alpha\)-helix in P protein that extends from residues 13 to 22. The orientation of H22 with respect to the \(\alpha\)-helix (it is flipped out away from the helix, see Figure 2.1B) suggests that the histidine residue does not likely form a hydrogen bond with a carbonyl oxygen of the backbone. Thus, the lack of folding-induced depression of the H22 pK\textsubscript{a} value may be due to a favorable electrostatic interaction with the helix dipole that compensates for the presumed increase in positive charge density. Such helix favorable His\textsuperscript{+}-helix dipole interactions have been previously reported\textsuperscript{43,88}.
2.2.9 Helices 1 and 3 are unfolded in I

An extra histidine peak, which we assigned to H22 (see Figures S6 and S7), was observed in the P protein NMR spectra when no ligand was present. The intensity of this peak increases from below the detection limit at pH values below 5.5 to approximately 20 ± 5% of the H22 U peak at pH 6.5. This observation suggests that the extra peak corresponds to a H22 in the I state because previous studies showed that the population of I increases with pH. The two sharp H22 crosspeaks observed in the HSQC spectrum indicate that U and I are in slow exchange relative to the difference in frequency ($k_{ex} << 250 \text{ s}^{-1}$), which is consistent with a previous measurement of the rate constants of this reaction by stopped-flow fluorescence ($k_{ex} = k_{UI} + k_{IU} = 2.1 \pm 0.3 \text{ s}^{-1}$). The relatively large frequency difference between the U and I H22 peaks at most pH values indicates that the residue is found in quite different magnetic environments in the two species. H22 is located at the end of the first α-helix while H3 and H105 are located on the first and third α-helices, respectively, and the absence of extra peaks for residues H3 and H105 suggests that they are in very similar magnetic environments in U and I. This observation suggests that helices 1 and 3 are likely unfolded in I, which is consistent with our previous observations. The pKₐ measurements reveal that the pKₐ of H22 in
the intermediate state is depressed relative to those in both the unfolded and folded states. In contrast to the compensating effects of the helix dipole and charge density of F described above, the absence of helix 1 in I precludes this compensation and the increased intramolecular charge repulsion due to the folded central β-sheet and helix 2 in I is the likely cause of the depressed pKₐ of H22.

### 2.2.10 H22 is the source of the pH dependence of the unliganded U to I equilibrium

Given the histidine pKₐ values for both U and I in the absence of sulfate, one can predict the pH dependence of the U⇌I equilibrium. Scheme 2.1 can be used to depict the coupled folding and deprotonation reactions between the U and I states at a particular histidine. U with a protonated histidine is represented as C₁H⁺, protonated I as C₂H⁺, deprotonated U as C₁ and deprotonated I as C₂. The effect of histidine deprotonation on the free energy of the U⇌I equilibrium can be calculated using the equation given in Scheme 2.1⁶⁸ (see Materials and Methods). Because H3 and H105 have single NMR resonances, we assume that their magnetic environments and therefore electrostatic environments are nearly identical in U and I, which is consistent with our previous studies indicating that these residues are found in the unfolded regions of I. On this basis, we assume that only the deprotonation of H22 perturbs the relative populations of
unliganded U and I between pH 4 and 8 and shifts the U⇌I equilibrium toward I. The pKₐ of H22 for the I state is lower than for the U state (corresponding to ΔG⁻→⁺ - ΔG⁺→⁻⁺ = -0.86 ± 0.09 kcal/mol, see Scheme 2.1) and contributes to the stability profile of the protein by stabilizing I relative to U as the pH increases (Figure 2.4). This lower pKₐ value indicates more favorable interactions when H22 is deprotonated in I than in U. The free energy of forming I from U becomes even more positive (unfavorable) at low pH when H22 is protonated, which suggests that H22 is the primary source of the pH dependence of the unliganded U to I equilibrium in the physiological pH range. We can also conclude that lowering the net charge by three units does not significantly populate the unliganded F state.

2.2.11 The pKₐ values of sulfate-folded P protein His have a weak ionic strength dependence

The effect of ionic strength was determined by measuring the pKₐ values of the sulfate-bound folded state in the presence of 0.15 M, 0.5 M and 1.0 M NaCl (panels A through C of Figure 2.5). The results show that the histidine residues in P protein have a low ionic strength dependence and shift their pKₐ values by only 0.1-0.2 units between
Figure 2.4: Calculated free energy of the U→I reaction based on \( \Delta pK_a = pK'_{U_a} - pK''_{U_a} \).

The point at pH 7 is derived from equilibrium cotitration experiments using both TMAO and urea\(^7\). The curve is calculated using this estimate and the pK\(_a\) values listed in Table 2.1 and Equation 2.4 (see Materials and Methods).
Figure 2.5: $^1$H$^\varepsilon_1$ P protein histidine chemical shift vs $^1$H$^\varepsilon_1$ L-histidine chemical shift titration curves

$^1$H$^\varepsilon_1$ P protein histidine chemical shift vs $^1$H$^\varepsilon_1$ L-histidine (internal standard) chemical shift titration curves for the sulfate-folded state in the presence of (A) 0.15M, (B) 0.5M, and (C) 1.0M NaCl. The gaps are from peaks that could not be resolved because they overlapped with another peak.
0.08 M and 1.08 M ionic strength (Table 2.1). P protein has a large net positive charge and screening by solution anions might be expected to reduce the pKₐ depression, as in the case of wild-type SNase where three out of four histidine residues exhibited increases of approximately 1 full pKₐ unit when the ionic strength was increased from 0 M to 1.5 M. The authors of this study concluded that in the case of positively charged SNase, high salt sensitivity arises from screening of long-range Coulombic interactions, which in a highly charged protein can add up to yield substantial effects. The lack of a strong ionic strength dependence of histidine pKₐ values in P protein, whose net charge is similar to that of SNase, is therefore surprising. A likely explanation for the low ionic strength dependence of P protein is the association of sulfate ions with many different sites on the protein surface in addition to the high-affinity specific binding sites observed in the crystal structure. The apparent affinities of sulfate and chloride at the high affinity sites differ by a factor of 100, meaning that 20 mM sulfate can only be displaced from these sites by 2 M chloride. If this ratio of sulfate and chloride affinities applies to the non-specific sites also, then the effect of adding chloride would be minimal at concentrations below 2 M. By this scenario, divalent sulfate interacts weakly but favorably with most or all of the positively charged groups in P protein, including
protonated histidines. In support of this interpretation, previous studies have determined the pKₐ values of histidine side chain of model compounds, myoglobin, and lysozyme and observed increases by 0.3-0.4 units between 0.02 M and 1.5 M NaCl. Kao, et al. attributed these changes to favorable interactions between the charged form of the imidazole moiety and solution counteranions at high salt concentrations rather than screening of unfavorable electrostatic interactions. We interpret the low ionic strength dependence of P protein histidine pKₐ values in a similar way: favorable interactions between the divalent sulfate ion and the positive surface of P protein effectively eliminate inter-residue electrostatic interactions. Similar interactions in the cell may play the same role, thus suggesting that unfavorable electrostatics may have little or nothing to do with the low stability of P protein and its intrinsically unfolded property.

2.3 Conclusions and future directions

In conclusion, our data show that the conformational equilibrium between the unfolded and partially folded intermediate states of P protein is favored by the deprotonation of H22. The magnitude of this stabilization is energetically equivalent to the effect of deprotonation of all three histidines on the U to F equilibrium, although the origins of these two stabilizations are probably different. In contrast to another highly
basic protein, SNase, ionic strength does not have a significant effect on the histidine pKₐ values of folded P protein in the presence of 20 mM Na₂SO₄, perhaps due to non-specific association of sulfate with positively charged side chains. This result implies that pairwise electrostatic interactions in folded P protein are weak and furthermore, suggests that in general, electrostatic forces do not play a large role in the energetics of this protein.

In addition, it is interesting to note that the intramolecular charge repulsion in U appears to only be somewhat relieved by polypeptide expansion and flexibility because all of the histidine pKₐ values are still depressed in comparison to model compounds. In contrast to previous studies of electrostatic interactions in the unfolded state, the pKₐ values of negatively charged groups were 0.3 units lower than those of model compounds, suggesting that the net Coulombic interactions in the U state were stabilizing⁶⁸, ⁹⁰, ⁹¹. It is unlikely that this depression in P protein can be explained in terms of local interactions with other positively charged residues nearby in the amino acid sequence. Of the three histidines, only H22 has neighboring basic groups and it has the least depressed pKₐ in U. It is possible that these low pKₐ values are governed by solvation effects in addition to high net charge. Coulomb’s law, \( U = \sum \frac{kq_1q_2}{\varepsilon r_{12}} \), reveals that

\[ 61 \]
the energy of interaction between an ionizable group and another charge on a protein can be influenced by the charge of either group, the distance between them, and the local dielectric constant of the medium that they reside in. The presence of the unfolded polypeptide chain could reduce the solvent accessibility of the histidine side chains, thereby reducing the polarizability of the environment around the histidines. This partial desolvation would decrease the dielectric constant, which would decrease the histidine pK$_a$ values. In future studies we plan to obtain more quantitative estimates of the effect of the dielectric constant on histidine pK$_a$ values by employing our method to measure the pK$_a$ values of histidine in model compounds and polypeptides as a function of solvent dielectric constant.

2.4 Materials and methods

2.4.1 Expression and purification of P protein

The F107W version of P protein used in our previous studies was the default background used here$^70$. The histidine variants of P protein (H3A, H22A, and H105A) were generated in this background. These site-directed mutational plasmids were constructed using the QuikChange procedure (Stratagene). The F107W and histidine variants were overexpressed in *Escherichia coli* [BL21 (DE3) pLysS] cells and purified via
Method 1 as previously described with the following modifications (Bacillus subtilis strain 168; UniProtKB ID: P25814). A single colony was inoculated into 50 mL of LB medium containing 50 μg/mL kanamycin and 17 μg/mL chloramphenicol and grown at 37°C until the OD600 reached 0.8-0.9. The culture was diluted into 2 L M9 minimal media cultures containing the same antibiotics and all 20 amino acids. All amino acids were the L-isomer and included: 0.8 mM Ala, 0.4 mM Arg, 0.4 mM Asn, 0.4 mM Asp, 0.1 mM Cys, 0.6 mM Glu, 0.6 mM Gln, 0.8 mM Gly, 0.4 mM Ile, 0.8 mM Leu, 0.4 mM Lys, 0.2 mM Met, 0.4 mM Phe, 0.4 mM Pro, 10.0 mM Ser, 0.4 mM Thr, 0.1 mM Trp, 0.2 mM Tyr, and 0.6 mM Val. For the F107W variant, histidine labeling was effected by substitution of the unlabeled compound in the medium with [ring-2-13C]-L-histidine (40 mg/L). For the histidine mutants, the minimal media cultures were enriched with 2 g/L 13C glucose and 1 g/L 15N ammonium chloride. The culture was incubated at 37°C until the OD600 reached 0.8-1.0. Expression of P protein was induced by the addition of 0.4 M isopropylthio-β-D-galactopyranoside (IPTG) and the culture was incubated for an additional 4-5 hours at 37°C. The cells were harvested by centrifugation at 4°C (7300 x g, 20 min) and resuspended in 20mL of lysis buffer (50 mM Tris pH 8.0, 5 mM EDTA, 10% glycerol) per liter of culture broth. The cells were then lysed by passing the cell
suspension through a French pressure cell at 17,000 p.s.i. The pooled fractions of P protein that eluted from the second CM Sepharose column were concentrated using Thermo Scientific Pierce Protein concentrators, 9K Molecular Weight Cut Off (MWCO), down to a volume of about 1 mL. Then, the remaining protein was buffer exchanged into 6 M Guanidine HCl, 50 mM Tris pH 7.5 and concentrated down to a volume of about 750 μL. The concentrated sample was loaded on to an S-100 column equilibrated with 6 M Guanidine HCl, 50 mM Tris pH 7.5 to remove ethylenediaminetetraacetic acid (EDTA) bound to the protein. Fractions containing P protein were pooled and stored at -80°C. The mass spectrum of the [ring-2-13C]-L-histidine labelled P protein and each histidine variant was determined to ensure that it contained the desired substitution. The protein NMR samples were dialyzed extensively against water and then the desired buffer. Protein purity was assessed by SDS-PAGE with Coomassie blue detection. Protein concentrations were determined by method of Edelhoch with an extinction coefficient of 11460 M⁻¹ cm⁻¹ at 280 nm.

2.4.2 Determination of pKₐ values using NMR spectroscopy

All one and two-dimensional NMR spectra were recorded on an 800 MHz Varian INOVA spectrometer at 25°C with a triple resonance probe. The NMR sample used for
the pH titration of unliganded P protein contained 500 μM $^{13}$C-$^{1}$-histidine ring labeled F107W P protein, 10 mM L-histidine, 10 mM pyridine, 10% D$_2$O, 1 mM trimethylsilyl propanoic acid (TMSP), and 0.05% sodium azide. The NMR samples used for the assignment of peaks in the unfolded spectra consisted of 300 μM $^{15}$N and $^{13}$C doubly uniformly labelled F107W H3A, H22A, or H105A P protein, 10 mM L-histidine, 10 mM pyridine, 10% D$_2$O, 1 mM trimethylsilyl propanoic acid (TMSP), and 0.05% sodium azide. The sample used for the pH titration of liganded P protein was the same as the one for unliganded P protein but also contained 20 mM sodium sulfate. pH titrations of liganded P protein were performed with no added NaCl and in 0.15, 0.5, and 1.0 M NaCl. All spectra were referenced against the position of the TMSP peak, whose pH-dependence was accounted for$^{95, 96}$. $^{1}$H-$^{13}$C HSQC spectra were collected from pH 4.0-8.0. Pyridine was present in the sample to ensure that the solution was well buffered throughout the entire pH range and was chosen because it is not an anion and will not induce folding of P protein. 1 M NaOH was titrated into the NMR tube using a 5.0 μL Hamilton syringe with polyethylene tubing attached to the tip of the needle. Two-dimensional $^{1}$H-$^{13}$C HSQC experiments were conducted with a spectral width in the $^{1}$H dimension of 8012 Hz and 2051 complex points and a spectral width of 10056 Hz and 64
complex points in the $^{13}$C dimension. Both dimensions were apodized with a shifted squared sine-bell window function and Fourier transformed using NMRPipe and NMRDraw suite of programs. Zero filling was also used in the $^{13}$C dimension (10056/256=39 Hz per point). The spectra were exported as text files and imported into a Mathematica™ notebook used to fit the multiple histidine resonances to two dimensional Gaussian distributions to determine their $^1$H and $^{13}$C chemical shifts (see the PeakPicking.nb Mathematica™ notebook included in the Supporting Information in Appendix A). The pK$_a$ values of histidines were obtained by fitting plots of protein histidine chemical shifts versus reference L-histidine chemical shifts to Equation 2.1 (see Results and Discussion) using the program Mathematica™. The pK$_a$ values for the reference L-histidine under conditions where there was no added NaCl and in 0.15, 0.5, and 1.0 M NaCl were obtained by fitting plots of the chemical shifts of the $^{13}$C proton resonance versus pH with the following equation using Mathematica™

$$
\delta_{obs}(pH) = \frac{\delta_A(10^{n(pH-pK_a)})+\delta_{HA}}{1+10^{n(pH-pK_a)}}
$$

(2.2)

where $\delta_A$ and $\delta_{HA}$ are the plateau values of the chemical shifts in the basic and acidic pH limits, respectively; $\delta_{obs}$ is the observed chemical shifts, and n is the Hill coefficient$^{59,97}$. The uncertainties in the pK$_a$ values and Hill coefficients were obtained from error
propagation of the standard errors given by the Mathematica™ NonlinearModelFit function from the fits of the reference compound titrations and the protein versus reference titrations and are reported as 95% confidence limits, ± two standard errors.

The change in ionic strength over the course of a titration is very small. For the protein titrations in the absence and presence of sulfate, the change in ionic strength was calculated to be 0.012 ± 0.01 M and 0.011 ± 0.01 M, respectively.

The chemical shifts of L-histidine in the presence and absence of protein when compared at low and high pH are within experimental error and suggest that the internal standard does not interact with the protein. In addition, the sulfate-bound folded state data was fitted using pyridine as the internal standard and the results are identical to those obtained using L-histidine as the internal standard (shown in Figure S20), which suggests that neither pyridine nor L-histidine interacts with the protein.

The solvent accessibilities of the histidine side chains were calculated using GETAREA98.
2.4.3 Calculation of pH-dependent stability

If the pKₐ values are known for two different conformational states of a protein, then it is possible to estimate the pH dependent stability of the protein and the following Tanford-Wyman linkage relationship applies:

\[
\frac{\partial \Delta G^i}{\partial pH} = 2.303 (RT) \Delta Q
\]

where \( \Delta Q \) is the difference in number of protons bound to the two conformational states which in this case are the intermediate and unfolded states. The integrated form of the equation can be expressed as a summation over the unfolded and intermediate state pKₐ values, provided that the titration behavior can be described as a set of independently titrating sites:

\[
\Delta \Delta G^o (pH - pH_{ref}) = RT \sum_{i=1}^{N} ln \left( \frac{(1 + 10^{pH - pK_{a}^{iI}})(1 + 10^{pH_{ref} - pK_{a}^{iI}})}{(1 + 10^{pH_{ref} - pK_{a}^{iU}})(1 + 10^{pH - pK_{a}^{iU}})} \right)
\]

Here, the superscripts I and U refer to the intermediate and unfolded state, respectively, i identifies the residue, N is the number of histidines in the protein, R is the gas constant, T is the temperature in Kelvin, \( pK_{a}^{iI} \) and \( pK_{a}^{iU} \) are the pKₐ values for the \( i^{th} \) group in the intermediate and unfolded state, respectively. \( pH_{ref} \) is the pH at an arbitrarily chosen reference pH, chosen here to be pH 7.0. To generate the calculated curve, the \( \Delta \Delta G_{U \rightarrow I} \) values were subtracted from the value, 1.0 ± 0.2 kcal/mol, obtained
from the experimentally measured stability from urea denaturation and TMAO folding experiments at pH 7.0\(^{70}\). This is why the experimental value and calculated curve agree at pH 7. Over the pH range of 4.0–8.0 covered in this study, histidine is the only group likely to titrate and thus one can predict the pH dependence of the U⇌I equilibrium if the histidine pK\(_a\) values for both U and I in the absence of sulfate are known.

### 2.4.4 Determination of tautomeric state populations

Both NMR samples used for the tautomer assignments consisted of 300 μM \(^{13}\)C uniformly labelled F107W P protein, 10 mM L-histidine, 10 mM pyridine, 10% D\(_2\)O, 1 mM trimethylsilyl propanoic acid (TMSP), 0.05% sodium azide. The tautomeric state populations of the neutral form of the protein histidines and N-Ac-His-NH-CH\(_3\) were determined in the presence and absence of 20 mM NaSO\(_4\) by measuring the chemical shifts of the \(^{13}\)C\(_{\delta1}\) nuclei in the His sidechains as a function of pH. The spectra were exported as text files and imported into the PeakPicking.nb Mathematica™ notebook included in the Supporting Information in Appendix A. This notebook was used to fit the multiple histidine resonances to two-dimensional Gaussian distributions to determine the chemical shifts of the observed \(^{13}\)C\(_{\delta2}\) doublets, which were then averaged.
These data were fitted to the following equation using the NonlinearModelFit function of Mathematica™:

$$\delta_{\text{His}}(\text{pH}) = \frac{\delta_{\text{AH}} + (\delta_{\text{AH}} + 7.1 - 9.2F_\epsilon)10^{\text{pH} - pK_a}}{1 + 10^{\text{pH} - pK_a}}$$ (2.5)

where $\delta_{\text{His}}$ is the chemical shift of the His $^{13}$C$\delta_1$; $\delta_{\text{AH}}$ is the chemical shift of the fully protonated His side chain; $pK_a$ is the $pK_a$ value of the His, and $F_\epsilon$ is the fraction of the His side chain in the $N^{\epsilon2}H$ tautomeric state, as opposed to the alternative $N^{\delta1}H$ state. Equation 2.5 is based on a change in chemical shift upon deprotonation of -2.1 ppm for the $N^{\epsilon2}H$ tautomer and +7.1 ppm for the $N^{\delta1}H$ tautomer given by Platzer, et al$^{38}$. Because the results given in Table 2 indicate Hill coefficients very close to 1, they have been omitted from Equation 2.5.
3. Salt sensitivity of histidine pK\textsubscript{a} values in P protein

3.1 Introduction

The effects of salt on pK\textsubscript{a} values of ionizable groups or proteins in general are complex. Salt can perturb a pK\textsubscript{a} by binding to a group specifically\textsuperscript{100} or screening a charge-charge interaction between groups. The concept of screening of electrostatic interactions by ionic strength in proteins was first introduced by Linderstrom-Lang in 1924\textsuperscript{101}. This screening is referred to as the Debye-Hückel screening\textsuperscript{36} and does not depend on the chemical nature of the salt. It is rare to observe salt screen buried electrostatic interactions unless the water and salt can infiltrate the interior of the protein significantly. There are instances reported in the literature where salt does not seem to have much of an effect such as the studies that indicate some surface salt bridges in proteins are insensitive to NaCl\textsuperscript{44,102}. The work of Dominy, et al. suggests that NaCl can screen long-range interactions more efficiently than short-range ones\textsuperscript{103}. On the contrary, there are examples of processes in proteins where salt seems to have a profound effect such as the interactions between human hemoglobin and anions that are important for the regulation of its functional properties\textsuperscript{104}. Salt can also indirectly affect the stability of a protein by changing the properties of water which is known as the Hofmeister effect\textsuperscript{17}. 

71
This effect explains why some ions stabilize proteins (eg. NaCl) while others destabilize them (eg. guanidinium chloride). Hofmeister ions salt-in polar groups and salt-out nonpolar groups.

Our interest in understanding the effect of ionic strength on pK\textsubscript{a} values stems from a previous study where we looked at this particular effect on the histidine pK\textsubscript{a} values of P protein. The effect of ionic strength was determined by measuring the pK\textsubscript{a} values of the sulfate-bound folded state in the presence of 0.15 M, 0.5 M and 1.0 M NaCl. The results show that the histidine residues in P protein have a small ionic strength dependence and their pK\textsubscript{a} values are shifted by only 0.1-0.2 units between 0.08 M and 1.08 M ionic strength. This shift was a surprising result because the pK\textsubscript{a} values of most histidines in small peptides and in myoglobin increase on average by 0.3 units between 0.02 M and 1.5 M NaCl\textsuperscript{89}. Also, the pK\textsubscript{a} values of three out of four histidines in SNase, a highly basic protein like P protein, were found to increase by a full unit between 0.01 M and 1.5 M KCl\textsuperscript{83}. We believe that the low ionic strength dependence of P protein histidine pK\textsubscript{a} values is due to favorable interactions between the divalent sulfate ions and the positive surface of P protein that effectively eliminate the inter-residue electrostatic interactions.
In order to elucidate further the source of this small ionic strength dependence, the histidine pKₐ values of the sucrose-folded state of P protein were measured in the presence of 1.0 M NaCl. Sucrose is an osmolyte that increases the thermodynamic conformational stability of a protein by shifting its folding equilibrium to favor the folded state. Osmolytes such as this one stabilize the folded state because of the unfavorable interactions that form between itself and the peptide backbone. Using sucrose allowed us to fold P protein in the absence of sulfate. Understanding the screening by salts, such as NaCl, of the charge-charge interactions in P protein is important because screening by intracellular salt concentrations may modulate its stability and interactions in vivo.

3.2 Results and discussion

3.2.1 The pKₐ values of sucrose-folded P protein His are depressed

pH titrations of the sucrose-folded state of P protein over a pH range of 4.0-8.0 were performed by 2D ^1H-^13C HSQC NMR spectroscopy using the NMR-monitored titration method previously described that utilizes internal reference compounds and a parametric fitting method. L-histidine was used as the internal standard in the experiments presented here. A total of 20 and 19 spectra were collected for the pH
Figure 3.1: $^1H^1P$ protein histidine vs $^1H^1$ L-histidine chemical shift titration curves

$^1H^1P$ protein histidine chemical shift vs $^1H^1$ L-histidine (internal standard) chemical shift titration curves for sucrose-folded P protein in (A) 0 M and (B) 1 M NaCl.
titrations of sucrose-folded P protein in 0 M and 1 M NaCl, respectively. Figure 3.1 contains the $^1$Hε1 P protein histidine chemical shift versus L-histidine chemical shift titration curves for each of the three histidine residues in (A) 0 M and (B) 1 M NaCl. These curves were fit to the modified Henderson-Hasselbalch equation previously described to determine the pKₐ values of the protein histidines. The absolute pKₐ values of the protein histidines were determined using a pKₐ value of the internal standard, L-histidine, measured using 1D $^1$H NMR (Figure 3.2). The pKₐ values for the protein histidines and Hill coefficients are listed in Table 3.1.

As a model of a histidine side chain in an electrostatically neutral environment, we used N-acetyl-L-histidine methylamide. The pKₐ value measured for this compound in 2 M sucrose and in the absence of NaCl was 6.22 ± 0.02 (Figure 3.3). When compared to the model compound, N-acetyl-L-histidine methylamide, all three histidines have pKₐ values that are lower in the sucrose-folded state of P protein in the absence of NaCl. This observed depression is what we expected to see in such a highly basic protein like P protein where the Coulombic interactions experienced by the protonated histidines would be repulsive.
L-Histidine titration curves under sucrose-folded conditions in the (A) absence and (B) presence of NaCl. The estimated pKₐ values of L-Histidine were (A) 6.16 ± 0.02 and (B) 5.88 ± 0.02.
Figure 3.3: $^1\text{H}$$^\varepsilon_1$ N-acetyl-L-histidine methylamide vs $^1\text{H}$$^\varepsilon_1$ L-histidine chemical shift titration curve

$^1\text{H}$$^\varepsilon_1$ N-acetyl-L-histidine methylamide vs $^1\text{H}$$^\varepsilon_1$ L-histidine chemical shift titration curve in 2 M sucrose and 0 M NaCl. The estimated pK$_a$ of N-acetyl-L-histidine methylamide was 6.22 ± 0.02.
Table 3.1: pKₐ values and Hill coefficients of L-histidine, N-acetyl-L-histidine methylamide and three histidines in unfolded, sulfate-folded and sucrose-folded P protein

<table>
<thead>
<tr>
<th></th>
<th>Sulfate Buffer Conditions</th>
<th>Sucrose Buffer Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Na₂SO₄]</td>
<td>20 ± 1 mM</td>
<td>20 ± 1 mM</td>
</tr>
<tr>
<td>[Sucrose]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>[NaCl]</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ionic Strengthb</td>
<td>80 ± 1.4 mM</td>
<td>20 ± 1 mM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Histidine</th>
<th>State pKₐ</th>
<th>State Hill Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Histidine</td>
<td>6.16 ± 0.01</td>
<td>0.98 ± 0.01</td>
</tr>
<tr>
<td>N-Ac-His-NH-Me</td>
<td>6.52 ± 0.03</td>
<td>1.09 ± 0.05</td>
</tr>
<tr>
<td>His 3</td>
<td>6.31 ± 0.01</td>
<td>1.07 ± 0.04</td>
</tr>
<tr>
<td>His 22</td>
<td>6.03 ± 0.01</td>
<td>1.07 ± 0.03</td>
</tr>
<tr>
<td>His 105</td>
<td>5.52 ± 0.01</td>
<td>1.01 ± 0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Histidine</th>
<th>nF</th>
<th>nF</th>
<th>nF</th>
<th>nF</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Histidine</td>
<td>1.03 ± 0.01</td>
<td>1.05 ± 0.02</td>
<td>1.08 ± 0.04</td>
<td>0.95 ± 0.05</td>
</tr>
<tr>
<td>N-Ac-His-NH-Me</td>
<td>1.09 ± 0.05</td>
<td>1.09 ± 0.05</td>
<td>1.13 ± 0.06</td>
<td>0.93 ± 0.02</td>
</tr>
<tr>
<td>His 3</td>
<td>1.02 ± 0.04</td>
<td>1.03 ± 0.07</td>
<td>0.99 ± 0.03</td>
<td>0.93 ± 0.02</td>
</tr>
<tr>
<td>His 22</td>
<td>1.07 ± 0.03</td>
<td>1.02 ± 0.05</td>
<td>1.00 ± 0.03</td>
<td>0.95 ± 0.05</td>
</tr>
<tr>
<td>His 105</td>
<td>1.01 ± 0.03</td>
<td>0.98 ± 0.05</td>
<td>0.98 ± 0.02</td>
<td>0.95 ± 0.03</td>
</tr>
</tbody>
</table>

bSolutions contained 10 mM L-histidine, 10 mM pyridine in 90% H₂O, 10% D₂O at 25 °C with either 0 mM or 20 mM sodium sulfate or either 0 M or 2 M sucrose. The pKₐ values were obtained by fitting the data to Equation 2.1 in Ref 107. The ¹H-¹ shifts were used to determine the pKₐ values. Errors represent 95% confidence limits and were obtained as described in Materials and Methods. bIonic strengths were calculated including all buffer components. cN-acetyl-L-histidine methylamide.
3.2.2 Comparison of the $pK_a$ values of unliganded unfolded and unliganded sucrose-folded P protein His

In our previous study, we measured the $pK_a$ values of the histidine residues for unliganded unfolded P protein which are listed in Table 3.1. These values can be compared to the unliganded sucrose-folded histidine $pK_a$ values in 0 M NaCl presented here. H105 has a depressed $pK_a$ upon folding of the protein which is consistent with its surrounding microenvironment being electropositive. In addition, the electrostatic repulsion is greater in the more compact and less dynamic folded state. For H3, we only observed a 0.2 unit elevation of the $pK_a$ in the sucrose-folded state unlike in the sulfate-folded state (which was 0.6 units higher). This is primarily due to the absence of the salt bridge that is formed between the sulfate ion and the H3 side chain. According to the crystal structure of folded P protein$^{71}$, H3, along with R9 and R68, coordinate the sulfate ion bound at the anion binding site between the second $\alpha$-helix and the N-terminus. The lack of significant folding-induced depression of the $pK_a$ of H22 in the sucrose-folded state is consistent with our previous explanation of why the $pK_a$ of H22 in the sulfate-folded state remains unchanged and is within uncertainty of its $pK_a$ in unfolded state. The favorable interaction that H22 has with an $\alpha$-helix dipole diminishes the $pK_a$ depression. Overall, the observed $pK_a$ differences between the folded and unfolded state
are small for all three histidines which suggests that unfavorable electrostatics alone do not account for the fact that RNase P protein is intrinsically unfolded in the absence of an anionic ligand.

**3.2.3 Electrostatic repulsion is greater in sucrose-folded state than sulfate-folded state**

All three histidine pKₐ values are lower in the sucrose-folded state than in the sulfate-folded state of P protein in the absence of NaCl. The previously measured pKₐ values for the protein histidines of sulfate-folded P protein are listed in Table 3.1. This is what we expected to see since in the sucrose-folded state there are no sulfate ions to screen the electrostatic interactions between the histidine residues and other positively charged groups in P protein. Thus, the positive charge density should be greater in the sucrose-folded state than the sulfate-folded state.

**3.2.4 The pKₐ values of sucrose-folded P protein His have a strong ionic strength dependence**

The effect of ionic strength was determined by measuring the pKₐ values of the sucrose-folded state in the presence of 1 M NaCl (Figure 3.1). The results show that 1 M NaCl shifts the pKₐ values of the histidine residues of P protein by 0.3-0.8 units (Table 3.1). This shift is comparable to the shift seen in the case of wild-type SNase where three
out of the four histidines residues exhibited increases of approximately 1 full pKₐ unit when the ionic strength was increased from 0 M to 1.5M. Overall, these results suggest that P protein does indeed have a strong ionic strength dependence. In our previous study of the sulfate-folded state of P protein, most likely screening by sulfate anions reduced the histidine pKₐ depression and resulted in the small pKₐ shift of only 0.1-0.2 units between 0.08 M and 1.08 M ionic strength (Table 3.1).

3.3 Conclusions

In conclusion, the three histidines (H3, H22, and H105) in the sucrose-bound folded state of P protein in the absence of salt have pKₐ values depressed by 0.34, 0.35, and 1.23 units, respectively, relative to that of the model compound, N-acetyl-L-histidine methylamide. These measured pKₐ values are consistent with repulsive interactions between protonated histidine side chains and the net positive charge of the protein. In comparison to another highly charged basic protein, SNase, ionic strength does have a significant effect on the histidine pKₐ values of sucrose-folded P protein (in the absence of 20 mM sodium sulfate). This effect is most likely due to the fact that there are no sulfate ions to interact with the many different sites on the protein including the high-affinity specific binding sites observed in the crystal structure and effectively screen the
NaCl. However, anionic ligands in the cell may play this role. Taken together, these observations further denote that unfavorable electrostatic interactions may only play a small part in the intrinsically unfolded property of P protein.

3.4 Materials and methods

3.4.1 Expression and purification of P protein

The F107W variant was overexpressed in *Escherichia coli* [BL21 (DE3) pLysS] cells and purified via Method 1 as previously described with the following modifications (Bacillus subtilis strain 168; UniProtKB ID: P25814). A single colony was inoculated into 50 mL of LB medium containing 50 μg/mL kanamycin and 17 μg/mL chloramphenicol and grown at 37°C until the OD600 reached 0.8-0.9. The culture was diluted into 2 L M9 minimal media cultures containing the same antibiotics and all 20 amino acids. All amino acids were the L-isomer and included: 0.8 mM Ala, 0.4 mM Arg, 0.4 mM Asn, 0.4 mM Asp, 0.1 mM Cys, 0.6 mM Glu, 0.6 mM Gln, 0.8 mM Gly, 0.4 mM Ile, 0.8 mM Leu, 0.4 mM Lys, 0.2 mM Met, 0.4 mM Phe, 0.4 mM Pro, 10.0 mM Ser, 0.4 mM Thr, 0.1 mM Trp, 0.2 mM Tyr, and 0.6 mM Val. Histidine labeling was effected by substitution of the unlabeled compound in the medium with [ring-2-13C]-L-histidine (40 mg/L). The culture was incubated at 37°C until the OD600 reached 0.8-1.0. Expression of P protein
was induced by the addition of 0.4 M isopropylthio-β-D-galactopyranoside (IPTG) and the culture was incubated for an additional 4-5 hours at 37°C. The cells were harvested by centrifugation at 4°C (7300 x g, 20 min) and resuspended in 20mL of lysis buffer (50 mM Tris pH 8.0, 5 mM EDTA, 10% glycerol) per liter of culture broth. The cells were then lysed by passing the cell suspension through a French pressure cell at 17,000 p.s.i. The pooled fractions of P protein that eluted from the second CM Sepharose column were concentrated using Thermo Scientific Pierce Protein concentrators, 9K Molecular Weight Cut Off (MWCO), down to a volume of about 1 mL. Then, the remaining protein was buffer exchanged into 6 M Guanidine HCl, 50 mM Tris pH 7.5 and concentrated down to a volume of about 750 μL. The concentrated sample was loaded on to an S-100 column equilibrated with 6 M Guanidine HCl, 50 mM Tris pH 7.5 to remove ethylenediaminetetraacetic acid (EDTA) bound to the protein. Fractions containing P protein were pooled and stored at -80°C. The mass spectrum of the [ring-2-13C]-L-histidine labelled P protein was determined to ensure that it contained the desired substitution. The protein NMR samples were dialyzed extensively against water and then the desired buffer. Protein purity was assessed by SDS-PAGE with Coomassie blue
detection. Protein concentrations were determined by method of Edelhoch with an extinction coefficient of 11460 M$^{-1}$ cm$^{-1}$ at 280 nm.

### 3.4.2 Determination of pH values using NMR spectroscopy

All one and two-dimensional NMR spectra were recorded on an 800 MHz Varian INOVA spectrometer at 25°C with a triple resonance probe. The NMR samples used for the pH titrations of sucrose-folded P protein contained 300 μM $^{13}$C$^{1}$-histidine ring labeled F107W P protein, 10 mM L-histidine, 10 mM pyridine, 10% D$_2$O, 1 mM trimethylsilyl propanoic acid (TMSP), 0.05% sodium azide, and either 0 M or 1 M NaCl. All spectra were referenced against the position of the TMSP peak, whose pH-dependence was accounted for. $^1$H-$^{13}$C HSQC spectra were collected from pH 4.0-8.0. Pyridine was present in the sample to ensure that the solution was well buffered throughout the entire pH range and was chosen because it is not an anion and will not induce folding of P protein. 1 M NaOH was titrated into the NMR tube using a 5.0 μL Hamilton syringe with polyethylene tubing attached to the tip of the needle.

Two-dimensional $^1$H-$^{13}$C HSQC experiments were conducted with a spectral width in the $^1$H dimension of 8012 Hz and 2051 complex points and a spectral width of 10056 Hz and 64 complex points in the $^{13}$C dimension. Both dimensions were apodized
with a shifted squared sine-bell window function and Fourier transformed using NMRPipe and NMRDraw suite of programs. Zero filling was also used in the $^{13}$C dimension ($10056/256 = 39$ Hz per point). The spectra were exported as text files and imported into a Mathematica™ notebook used to fit the multiple histidine resonances to two dimensional Gaussian distributions to determine their $^1$H and $^{13}$C chemical shifts (see the PeakPicking.nb Mathematica™ notebook included in the Supporting Information in Appendix A). The pK$_a$ values of histidines were obtained by fitting plots of protein histidine chemical shifts versus reference L-histidine chemical shifts as previously described$^{60}$ using the program Mathematica™. The pK$_a$ values for the reference L-histidine under conditions where there was no added NaCl and in 1.0 M NaCl were obtained by fitting plots as previously described$^{60}$. The uncertainties in the pK$_a$ values and Hill coefficients were obtained from error propagation of the standard errors given by the Mathematica™ NonlinearModelFit function from the fits of the reference compound titrations and the protein versus reference titrations and are reported as 95% confidence limits, ± two standard errors.
The chemical shifts of L-histidine in the presence and absence of protein when compared at low and high pH are within experimental error and suggest that the internal standard does not interact with the protein.
4. The effect of dielectric constant on the protonation of the histidine side chain

4.1 Introduction

The pK_a value of an ionizable group is not only determined by the charge distribution in a protein or geometry and flexibility of other polar and charged groups, but also by the properties of the solvent. The dielectric constant of a solvent is a measure of its ability to reduce the electric field across it and to oppose the electrostatic force between the charges separated by it\(^\text{109}\). Solvents with high dielectric constants very efficiently separate or insulate ions from each other. In contrast, low dielectric constant media less effectively decrease the electrostatic force between charged species. Polar substances such as water usually possess high dielectric constants while nonpolar substances have relatively low dielectric constants. For example, the dielectric constants for water and ethanol at room temperature are 78.5 and 24.3, respectively, while benzene has a dielectric constant of 2.28\(^\text{109, 110}\).

In a polar solvent, a solute ion is solvated or has concentric shells of many solvent dipoles with their oppositely charged poles oriented towards the ion due to electrostatic attraction. This structure distributes the charge of the ion and weakens the electric field of the ion thereby polar solvents dissolve electrolytes easily. In other words,
two ions “feel” each other much less in a high dielectric constant solvent. In a nonpolar medium, the favorable electrostatic interaction between two unlike charges would be much larger because the dielectric constant is smaller. Nevertheless, this favorable contribution is outweighed by the larger energetic cost of placing charges in a nonpolar environment. The charged state of a group will be favored more in a high dielectric constant medium over a lower one. This is primarily why charged groups are excluded from hydrophobic interior of a protein and instead are found at the surface where they can interact with bulk water. The very few exceptions are those that have functional significance such as charges in the active site of enzymes.

To obtain more quantitative estimates of the effect of dielectric constant on pKₐ values, we plan to employ our previously described method in Chapter 2 to measure the pKₐ value of histidine in different model compounds as a function of solvent dielectric constant. From the unfolded P protein NMR titration experiment results discussed in Chapter 2, it was noted that when the protein unfolds the intramolecular charge repulsion appears to only be somewhat relieved by polypeptide expansion and flexibility because all of the histidine pKₐ values are still depressed in comparison to model compounds. It is unlikely that this depression in P protein can be explained in
terms of local interactions with other basic residues nearby in the primary sequence. Of the three histidines, only H22 has neighboring basic groups and it has the least depressed pK$_a$ in U. It is possible that these low pK$_a$ values are governed by solvation effects in addition to high net charge. Coulomb's law reveals that the energy of interaction between an ionizable group and another charge on a protein can be influenced by the charge of either group, the distance between them, and the local dielectric constant of the medium that they reside in. The presence of the unfolded polypeptide chain could reduce the solvent accessibility of the histidine side chains, thereby reducing the polarizability of the environment around the histidines. This desolvation would lessen the dielectric constant, which would lower the histidine pK$_a$ values.

**4.2 Results and discussion**

In order to determine the effect a polypeptide chain would have on the dielectric constant of the medium surrounding an ionizable group, we measured the pK$_a$ value of histidine in a Gln-Gln-Gln-His-Gln-Gln-Gln peptide (His peptide) with both termini blocked (N-terminal acetylation and C-terminal amidation) in order to eliminate the possibility of pK$_a$ perturbations through electrostatic interactions with a charged group.
There are no charges on the peptide except for the protonated histidine side chain. The peptide serves as a simplified mimic to a protein polypeptide chain and should give a local environment for the histidine similar to that expected in an unfolded, uncharged protein. Glutamine was chosen to ensure that the compound was polar and soluble in aqueous solution. The measured pKₐ value for the His peptide was 6.34 ± 0.02 (Figure 4.1). In comparison to the pKₐ of N-acetyl-L-histidine methylamide reported in Chapter 2 (6.44 ± 0.02), this value is lower, which agrees with our hypothesis that the extra residues of the peptide chain diminish the solvation of the histidine side chain and decrease the effective dielectric constant in the vicinity of the histidine side chain. As a result, this reduction of the dielectric constant depresses the pKₐ of histidine.

Shown in Scheme 4.1 is the thermodynamic cycle that describes the deprotonation of histidine in both the model compound and His peptide, where $^+HA_{mod}$ and $^+HA_{pep}$ are protonated His in the model compound and His peptide, respectively, $A_{mod}$ and $A_{pep}$ are deprotonated His in the model compound and His peptide, respectively. If we calculate the double free energy difference between the two deprotonation reactions or $\Delta \Delta G = \Delta G_{depro}^{pep} - \Delta G_{depro}^{mod}$, which can be further simplified to $\Delta \Delta G = 2.303 \times RT (pK_{a_{pep}} - pK_{a_{mod}})$, we get $\Delta \Delta G = -0.14 \text{ kcal/mol}$. Due to the
Figure 4.1: $^1H^1$ His peptide vs $^1H^1$ L-histidine chemical shift titration curve

The estimated pKₐ of the His peptide was $6.34 \pm 0.02$. 
relationship between the free energies of the cycles, this $\Delta \Delta G$ of deprotonation is also equal to the double free energy difference between the two vertical transfer reactions. Hence, $\Delta \Delta G \text{mod} = \Delta G^\text{mod}_{\text{depro}}$ will also be negative. This agrees with our expectations because transferring the histidine from the model compound to the peptide will be more favorable when it is deprotonated if the environment around the histidine has a lower dielectric constant in the peptide than it does in the model compound and $\Delta G^\text{trans}_{\text{mod} \rightarrow \text{pep}}$ will be less negative than $\Delta G^\text{trans}_{\text{mod} \rightarrow \text{pep}}$ and the resulting double free energy difference will be negative.
4.3 Conclusions and future directions

In a more hydrophobic environment, the proton attached to the histidine would leave more readily to make the neutral species, and thus decrease its pK\textsubscript{a}. Our results suggest that the polypeptide chain of a protein affects the solvation of the side chain of an ionizable group which will therefore shift the pK\textsubscript{a} value of that group. In the case of histidine, the chain will shift its pK\textsubscript{a} value down. In order to gain more insight about this effect, we plan to also measure the histidine pK\textsubscript{a} values of the model compound, N-acetyl L-histidine methylamide, and His peptide in the presence of 50% ethanol or in a lower dielectric solvent\textsuperscript{111}. The method previously described in Chapter 2 will be used to measure the \(\Delta pK\textsubscript{a}\) values and the pH of the solution will not have to be measured. We will only be able to deduce the sign of the double free energy differences of these reactions and not the magnitude of the free energy of each reaction. We hypothesize that the \(\Delta pK\textsubscript{a}\) or the \(\Delta\Delta G\) of deprotonation between the model compound and His peptide will be smaller because the effect of the polypeptide chain should be more subtle in an apolar solvent of low dielectric constant. Once the \(\Delta pK\textsubscript{a}\) is known, the thermodynamic cycle in Scheme 4.1 that depicts the reactions in water can be combined with another cycle that describes the reactions in a water-ethanol solvent to form a cube, as shown in
Figure 4.2, and help us to predict whether or not it would be favorable to transfer a protonated or deprotonated histidine from a high to low dielectric constant environment (or visa versa). We expect that the ethanol-water solvent will be less effective at solvating the charged histidine side chain than a 100% water solvent and therefore, the free energies that describe the transfer of protonated histidine from a high to low dielectric will be unfavorable (these are represented by the dotted lines on the left side of the cube in Figure 4.2). The free energies that describe the transfer of deprotonated histidine will be slightly less unfavorable in comparison to the former case because the histidine is not charged (these are represented by the dotted lines on the right side of the cube in Figure 4.2).
Figure 4.2: Coupled deprotonation and transfer of a histidine side chain scheme

The front or green and back or black thermodynamic cycles represent the coupled deprotonation and transfer of a histidine side chain from a model compound to His peptide in a low and high dielectric constant solvent, respectively. The dotted lines represent equilibrium arrows for the transfer of a histidine side chain in the model compound or His peptide from a high to a low dielectric constant solvent (or visa versa).
4.4 Materials and methods

4.4.1 Chemicals and reagents

N-acetyl-methyl amide was purchased from Toronto Research Chemicals Inc. The (Gln)$_3$-His-(Gln)$_3$ peptide was purchased from GenScript. The liquid chromatography-mass spectrum of the (Gln)$_3$-His-(Gln)$_3$ peptide was obtained to ensure purity and that it contained the desired substitutions.

4.4.2 Measurement of pK$_a$ values using NMR spectroscopy

All two-dimensional NMR spectra were recorded on an 800 MHz Varian INOVA spectrometer at 25°C with a triple resonance probe. The NMR samples used for the pH titrations of N-acetyl L-histidine methyl amide and the (Gln)$_3$-His-(Gln)$_3$ peptide contained 1 mM N-acetyl L-histidine methyl amide or .5mM (Gln)$_3$-His-(Gln)$_3$ peptide, 10 mM L-histidine, 10 mM pyridine, 10% D$_2$O, 1 mM trimethylsilyl propanoic acid (TMSP), and 0.05% sodium azide. All spectra were referenced against the position of the TMSP peak, whose pH-dependence was accounted for$^{95,96}$. $^1$H-$^{13}$C HSQC spectra were collected from pH 4.0-8.0. Pyridine was present in the sample to ensure that the solution was well buffered throughout the entire pH range. 1 M NaOH was titrated into the
NMR tube using a 5.0 μL Hamilton syringe with polyethylene tubing attached to the tip of the needle.

Two-dimensional $^1$H-$^{13}$C HSQC experiments were conducted with a spectral width in the $^1$H dimension of 8012 Hz and 2051 complex points and a spectral width of 10056 Hz and 64 complex points in the $^{13}$C dimension. Both dimensions were apodized with a shifted squared sine-bell window function and Fourier transformed using NMRPipe and NMRDraw suite of programs. Zero filling was also used in the $^{13}$C dimension ($10056/256=39$ Hz per point). The spectra were exported as text files and imported into a Mathematica™ notebook used to fit the multiple histidine resonances to two dimensional Gaussian distributions to determine their $^1$H and $^{13}$C chemical shifts. The pK$_a$ values of histidines were obtained by fitting plots of protein histidine chemical shifts versus reference L-histidine chemical shifts as previously described using the program Mathematica™. The uncertainties in the pK$_a$ values and Hill coefficients were obtained from error propagation of the standard errors given by the Mathematica™ NonlinearModelFit function from the fits of the reference compound titrations and the protein versus reference titrations and are reported as 95% confidence limits, ± two standard errors.
5. Probing the folding intermediate of P protein via singular value decomposition

5.1 Introduction

To decipher the folding pathway and understand the mechanism by which the amino acid sequence directs folding of a protein to its native conformation, it is imperative to characterize protein folding intermediates. Many kinetics studies demonstrate that partially folded forms do exist for many proteins\(^{112}\). However, their low population and transient nature preclude detailed structural, dynamic, and thermodynamic studies. A few of the approaches taken to investigate partially folded proteins include studies of proteolytic fragments\(^ {113}\) and native-state hydrogen exchange NMR\(^ {114}\). Another potentially useful method is to examine the folding mechanism of proteins that have intermediates that are significantly populated under equilibrium conditions\(^ {115}\).

One protein whose equilibrium (un)folding mechanism involves an intermediate is the protein subunit of *Bacillus subtilis* RNase P, P protein. Previous kinetic and equilibrium studies of TMAO-induced folding of P protein as well as NMR experiments demonstrated the presence of a partially folded intermediate in its folding mechanism\(^ {70,75}\). \(^ {1}H\)-\(^ {15}N\) HSQC NMR spectra collected as a function of pH showed increasing pH, from
5.0 to 7.0 in steps of 1 pH unit, shifts the U-I conformational equilibrium toward I in the absence of sulfate and the I-F equilibrium toward F in the presence of sulfate. These effects were attributed to the unfavorable electrostatic interactions in the more compact forms of the protein. Stopped-flow studies showed that pyrophosphate ligand binding to P protein shifts the conformational equilibrium toward I and F by both increasing the folding rate constants and decreasing the unfolding rate constants. The NMR pH titration experiment of the unfolded state of P protein from Chapter 2 also provided further evidence that this protein does undergo a three-state equilibrium folding reaction and suggests that the intermediate state has its first and third α-helices unfolded.

In this chapter, we seek to obtain further information on the nature of both the unliganded and liganded forms of the intermediate of P protein. Singular value decomposition (SVD) multiwavelength analysis of tryptophan fluorescence (FL-) and circular dichroism (CD)-monitored equilibrium unfolding or folding of P protein data are performed to pursue this goal.
5.1.1 Singular value decomposition

SVD takes a high dimensional, highly variable set of data points and reduces it to a lower dimensional space that exposes the substructure of the original data more clearly and orders it from most variation to the least. This original data set, called matrix A, is made up of n rows and p columns. For example, it can be a data matrix of fluorescence spectra from a urea titration of a protein where a spectrum of the protein at a particular urea concentration can be represented as a column of numbers in which each element is an intensity at a particular wavelength. If the spectrum had 300 wavelengths and the titration included 40 different urea concentrations, matrix A would have 300 rows and 40 columns. SVD breaks down this matrix A into the product of three matrices (see Equation 5.1) – an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V:

\[ A = USV^T \] (5.1)

U and V are orthogonal matrices which means that their columns are orthonormal sets (or a set of orthonormal vectors also called basis vectors). The columns of U and V are called the left-singular vectors and right-singular vectors of A, respectively. The columns of U or left-singular vectors of A are the eigenvectors of \( AA^T \)
(matrix A multiplied by the transpose of matrix A) and form the orthonormal basis for
the row space of information of matrix A. The columns of V or the right-singular vectors
are the eigenvectors of $A^TA$ (the transpose of matrix A multiplied by matrix A) and
form the orthonormal basis for the column space of information of matrix A. W is a
diagonal matrix with nonnegative elements on the diagonal. These nonnegative
elements are called the singular values of A or weights and are the square roots of the
eigenvalues for the vectors of U and V, which are eigenvectors of $AA^T$ and $A^TA$,
respectively. Each weight refers to the importance and relative amount of information in
the first columns of U and V. Each successive column of U and V accounts for a fraction
of the remaining information in the original data matrix A. Only a small number of the
leading columns form the basis of U or V and the additional columns contain mostly
noise.

In other words, SVD provides a convenient way for breaking down a matrix,
which perhaps contains some data we are interested in, into simpler, meaningful pieces.
For example, consider the two-dimensional data points in Figure 5.1. If we take this
data, put it into a matrix, and perform an SVD we find the singular values: $s_1=5.02$ and
$s_2=0.15$. The second singular value is so much smaller than the first. Presumably, the
second value is due to noise in the data and would be ideally zero. In that case, the matrix would only have one significant singular value. This makes sense because the best fit line running through the points in Figure 5.1 is the best approximation of the original data with a one-dimensional object, a line. Suppose we drew another best fit line that was perpendicular to the first (see Figure 5.2). This line captures as much of the variation as possible along the second dimension of the original data set. It does not do as great of a job as the first line of approximating the original data because it corresponds to a dimension exhibiting less variation. It is possible to use regression lines
Figure 5.2: Perpendicular regression line is a poor approximation of the original data points

like these to generate a set of uncorrelated data points that will show subgroupings in the original data that were not noticeable initially. Thus, SVD can be viewed as a method for transforming correlated variables into a set of uncorrelated ones that better show the relationships among the original data. Also, SVD is a way to identify and order the dimensions along which data points exhibit the most variation. And ultimately, once we have pinpointed where the most variation is, the best approximation of the original data using fewer dimensions can be found.
5.2 Results and discussion

CD and FL spectroscopy were used to monitor a series of urea-induced unfolding or osmolyte-induced (TMAO and sucrose) folding of P protein at different pH values. Then singular value decomposition (SVD) multiwavelength analysis was used to decompose the data in basis vectors that can be fit to a three state equilibrium folding model to calculate the thermodynamic parameters for the folding of P protein. Next, these parameters were used to reconstruct the CD and FL spectra and calculate the relative populations of all three species.

5.2.1 Urea-induced unfolding of P protein

Figure 5.3 shows the FL-monitored urea-induced unfolding of sulfate-folded P protein at pH 7.0. As the urea concentration is increased, the folding equilibrium is pushed towards the unfolded state. The tryptophan FL emission spectrum of folded P protein displays a maximum at 308nm which decreases in intensity and red shifts to 345nm and then increases in intensity again as the urea concentration is increased or as the protein unfolds. This data was collected from 300-360 nm at 0.2 nm intervals (300 wavelengths total) at 40 different urea concentrations from 0-8.9 M urea. Due to noise caused by the fluorimeter, the last 33 wavelengths were removed as well as 5 of the urea
concentrations. As a result, the “corrected” data consists of 268 wavelengths and 35 urea concentrations.

The spectra were subjected to SVD analysis which decomposed the FL data into wavelength- and urea-dependent basis vectors (Figure 5.4). Shown in blue in Figure 5.4 are the singular values each of the vectors are given. The analysis revealed three pairs of basis vectors that appeared nonrandom and had statistically significant weights. It is important to note that the basis vectors do not necessarily correspond to the spectrum or denaturation profile of a particular species of the protein.
Figure 5.4: The SVD basis vectors obtained from a FL urea titration of P protein in 10 mM L-histidine, 10 mM pyridine, 20 mM sodium sulfate at pH 7.0 and 25°C

The left column shows the five-wavelength dependent basis vectors (U) while the right column shows the five urea-dependent basis vectors (V). The numbers in blue represent the weights for each pair of U and V basis spectra. The protein concentration was 5 μM.
A three state equilibrium folding model was used to fit the urea-dependent vectors of these three pairs in order to obtain the thermodynamic parameters (ΔG_{IU}, ΔG_{IU}, m_{IU}, and m_{FI}) for the unfolding of P protein (Figure 5.5). These parameters along with the wavelength-dependent vectors were then used to reconstruct the FL spectra of all three states of P protein and calculate the relative populations of U, I and F (Figure 5.6). Unfortunately, the fit produced thermodynamic parameters with large standard errors which suggest that the parameters are poorly determined. We believe this is most likely due to the folded state and the intermediate having such a similar FL spectrum.

Therefore, in an effort to help the fit, this experiment can be repeated using CD to monitor the urea-induced unfolding of P protein because the CD spectrum of the folded and intermediate state should not be similar. Both fitting this CD data individually and global fitting this data with the FL data could possibly give better determined thermodynamic parameters.

### 5.2.2 TMAO-induced folding of P protein

In the previous section, our aim was to look at the liganded intermediate state (in the presence of an anionic ligand) of P protein. Additionally, we wanted to observe the unliganded intermediate state and so we used TMAO as an osmolyte to fold the protein
Figure 5.5: The fits of the urea-dependent basis vectors to a three-state model

The best fit lines of the data are shown in blue.
Figure 5.6: (A) Relative populations and (B) FL spectra of each of the three thermodynamic species of P protein

(A) These curves were calculated from the three-state fit parameters. (B) The FL spectra at 0 M urea were reconstructed from the SVD analysis and fits from Figure 5.5.
in the absence of an anionic ligand. Shown in Figure 5.7 is the FL-monitored TMAO-induced folding titration of P protein at pH 7.0. As the TMAO concentration is increased, the folding equilibrium is pushed towards the folded state. The tryptophan FL emission spectrum of unfolded P protein displays a maximum at 345 nm which decreases in intensity and blue shifts to 307 nm and then increases in intensity again as the TMAO concentration is increased or as the protein folds. This data was collected from 300-360 nm at 0.2 nm intervals (300 wavelengths total) at 30 different TMAO concentrations from 0-1.5 M TMAO.

Figure 5.7: P protein tryptophan fluorescence emission spectra at pH 7 and various TMAO concentrations
The spectra were subjected to SVD analysis which decomposed the FL data into wavelength- and TMAO-dependent basis vectors (Figure 5.8). Shown in blue in Figure 5.8 are the singular values each of the vectors are given. The analysis revealed two pairs of basis vectors that appeared nonrandom and had statistically significant weights. The third basis vector appeared to be slightly noisy and more random that the others. However, it was decided that further analysis of this data would not be helpful because TMAO is not an ideal osmolyte for this study where we want to look at the folding equilibrium of P protein at different pH values. According to a study conducted by Singh, et al., the effect of TMAO on protein stability is pH-dependent. The authors concluded that TMAO destabilizes proteins at pH values below its pKₐ of 4.66 ± 0.10 and stabilizes proteins at pH values above its pKₐ. Therefore, sucrose was selected as the new osmolyte for this study because its effect on protein stability is independent of pH.

5.2.3 Sucrose-induced folding of P protein

Figure 5.9 depicts a sucrose-induced folding titration of P protein monitored by CD. The far-UV CD spectrum of folded P protein displays minima at 220 and 208 nm which appear as the protein folds. This data was collected at pH 7 from 195-260 nm at 1 nm intervals (65 wavelengths total) at 34 sucrose concentrations. The spectra were
Figure 5.8: The SVD basis vectors obtained from a FL TMAO titration of P protein in 10 mM L-histidine, 10 mM pyridine at pH 7.0 and 25°C

The left column shows the four-wavelength dependent basis vectors (U) while the right column shows the five TMAO-dependent basis vectors (V). The numbers in blue represent the weights for each pair of U and V basis spectra. The protein concentration was 5 μM.
subjected to SVD analysis and the resulting basis vectors were fit to a three state model (Figure 5.10). As with the urea FL data, the thermodynamic parameters obtained from the sucrose CD data were poorly determined by the model. The same experiment could be repeated and monitored by FL. The FL data could be fit individually and both the CD and FL data could be globally fit in an effort to provide more information for the model to more accurately determine the parameters.

5.3 Conclusions and future directions

SVD analysis of FL and CD data was performed to gain insight about the relative stability of the intermediate to the unfolded and folded state of P protein in both the absence and presence of an anionic ligand. Unfortunately, the resulting basis vectors
Figure 5.10: The SVD basis vectors obtained from a CD sucrose titration of P protein in 20mM sodium cacodylate at pH 7.0 and 25°C

The left column shows the three-wavelength dependent basis vectors (U) while the right column shows the three sucrose-dependent basis vectors (V). The numbers in blue represent the weights for each pair of U and V basis spectra. The protein concentration was 5 μM.
Figure 5.11: The fits of the sucrose-dependent basis vectors to a three-state model

The best fit lines of the data are shown in blue.
Figure 5.12: (A) Relative populations and (B) CD spectra of each of the three thermodynamic species of P protein

(A) These curves were calculated from the three-state fit parameters. (B) The CD spectra at 0 M sucrose were reconstructed from the SVD analysis and fits from Figure 5.11.
from SVD were not able to be accurately fit to a three state model. The thermodynamic parameters obtained were poorly determined by the model. Presumably, this could be due to the model’s inability to tease apart the intermediate and folded state FL spectra or the intermediate and unfolded state CD spectra because they are similar at pH 7. It could also be because there are not enough parameters that are known to allow the model to accurately fit the data. In addition, the population of the intermediate state may be too low for CD and FL to detect under these conditions. It would be interesting to try a different approach, relaxation CPMG NMR experiments, to look at the exchange processes between the folded or unfolded and intermediate states and use this as a way to estimate the populations (and rate constants) of the exchanged states.

5.4 Material and methods

5.4.1 Chemicals and reagents

Ultrapure urea was purchased from Nacalai Tesque Inc. TMAO was from Fluka. Cacodylic acid and sucrose were from Sigma. To remove the background fluorescence from TMAO, a 3 M stock TMAO solution was mixed with activated charcoal for 12 hours. The charcoal was then removed by filtration using a 0.22 μM syringe filter. Urea
and TMAO concentrations were measured by refractive index using the equations from
the following references, respectively\textsuperscript{118,119}.

\section*{5.4.2 Expression and purification of P protein}

The F107W variant was overexpressed in \textit{Escherichia coli} [BL21 (DE3) pLysS] cells
and purified via Method 1 as previously described\textsuperscript{92} with the following modifications
(\textit{Bacillus subtilis} strain 168; \textit{UniProtKB ID: P25814}). A single colony was inoculated into
50 mL of LB medium containing 50 \(\mu\text{g/mL}\) kanamycin and 17 \(\mu\text{g/mL}\) chloramphenicol
and grown at 37°C until the OD600 reached 0.8-0.9. The culture was diluted into 2 1 L LB
media cultures containing the same antibiotics. The culture was incubated at 37°C until
the OD600 reached 0.8-1.0. Expression of P protein was induced by the addition of 0.4 M
isopropylthio-\(\beta\)-D-galactopyranoside (IPTG) and the culture was incubated for an
additional 3-4 hours at 37°C. The cells were harvested by centrifugation at 4°C (7300 x g,
20 min) and resuspended in 20mL of lysis buffer (50 mM Tris pH 8.0, 5 mM EDTA, 10%
glycerol) per liter of culture broth. The cells were then lysed by passing the cell
suspension through a French pressure cell at 17,000 p.s.i. The pooled fractions of P
protein that eluted from the second CM Sepharose column were concentrated using
Thermo Scientific Pierce Protein concentrators, 9K Molecular Weight Cut Off (MWCO),
down to a volume of about 1 mL. Then, the remaining protein was buffer exchanged into 6 M Guanidine HCl, 50 mM Tris pH 7.5 and concentrated down to a volume of about 750 μL. The concentrated sample was loaded on to an S-100 column equilibrated with 6 M Guanidine HCl, 50 mM Tris pH 7.5 to remove ethylenediaminetetraacetic acid (EDTA) bound to the protein. Fractions containing P protein were pooled and stored at -80°C. Protein purity was assessed by SDS-PAGE with Coomassie blue detection. Protein concentrations were determined by method of Edelhoch with an extinction coefficient of 11460 M⁻¹ cm⁻¹ at 280 nm.

5.4.3 Circular dicroism and fluorescence spectroscopy

An Aviv model 202 CD spectrometer was employed to collect CD wavelength spectra. Samples were scanned from 260 to 195 nm in 1 nm increments and averaged for 4 seconds at each wavelength. The protein concentration was 5 μM in 20 mM sodium cacodylate at pH 7 in the presence or absence of TMAO using a cuvette with a path length of 1mM at 25°C.

Fluorescence spectra were recorded on a Shimazdu spectrofluorometer using a cuvette with a path length of 1 cm. The temperature was controlled at 25°C by a circulating water bath. The excitation wavelength was 280 nm and emission spectra
were collected from 300 to 360 nm at 0.2 nm increments. An automated Hamilton titrator was used to perform all titrations.

### 5.4.4 Singular value decomposition

The CD and FL data were decomposed into wavelength- and denaturant- or osmolyte-dependent basis vectors by the SVD algorithm. The m x n data matrix A, consisting of m wavelengths and n denaturant or osmolyte concentrations, is represented by Equation 5.1 where U is a m x n matrix whose columns are orthonormal basis vectors representing the wavelength spectrum, S is an n x n diagonal matrix of singular values or weights of the basis spectra, and V is an n x n matrix whose orthonormal columns represent the denaturant or osmolyte dependence of the basis vectors. In selecting the number of significant basis sets to represent the data matrix, consideration was given to the singular values and the autocorrelation of the corresponding basis vectors. The basis vectors representing the denaturant or osmolyte dependence were fit to a three-state equilibrium folding model. Spectra of F, I, and U extrapolated to 0 M denaturant or osmolyte were obtained by reconstruction of the data matrix from these basis vectors and their thermodynamic parameters. The relative populations of each species were calculated from the three-state fit parameters.
Appendix A

Chapter 2 Supporting Information

Figure S.13: Titration of NMR sample

With a 0.5–0.4 μl aliquot of NaOH using a 5.0 μl positive displacement syringe extended with polyethylene tubing.
Figure S.14: L-Histidine (internal standard) titration curves under different buffer conditions

In addition to the components indicated, each sample had 10 mM pyridine, 90% H₂O, and 10% D₂O. The $^{13}$C ε₁ protons were used to determine the pKₐ values and Hill coefficients for L-Histidine, listed in panel F. Uncertainties are based on the standard errors of the fitted parameters.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>L-Histidine pKₐ</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6.08 ± 0.01</td>
<td>1.05 ± 0.02</td>
</tr>
<tr>
<td>B</td>
<td>6.16 ± 0.01</td>
<td>1.03 ± 0.01</td>
</tr>
<tr>
<td>C</td>
<td>6.07 ± 0.01</td>
<td>0.99 ± 0.01</td>
</tr>
<tr>
<td>D</td>
<td>6.11 ± 0.01</td>
<td>0.99 ± 0.01</td>
</tr>
<tr>
<td>E</td>
<td>6.11 ± 0.01</td>
<td>0.98 ± 0.01</td>
</tr>
</tbody>
</table>
Figure S.15: $^{13}\text{C}^{1}$ P protein histidine versus $^{13}\text{C}^{1}$ L-Histidine chemical shifts titration curves

$^{13}\text{C}^{1}$ P protein histidine chemical shifts versus $^{13}\text{C}^{1}$ L-Histidine chemical shifts titration curves for (A) sulfate-bound folded and (B) unfolded P protein. H3, H22, and H105 are represented as the blue, red, and green curves, respectively. The purple curve in (B) represents the H22 intermediate peak.
Figure S.16: NMR resonances from all three histidine residues in the unfolded state of P protein

NMR resonances from all three histidine residues in the unfolded state of P protein were assigned by comparing spectra of (A) F107W P protein to spectra of two variants, (B) H105A and (C) H22A, that each lack a different histidine residues. All spectra were collected at a sample pH of 4. Assignment spectra for the sulfate-bound folded state are not shown.
Figure S.17: $^1H$ N-acetyl-L-Histidine methylamide versus $^1H$ L-Histidine chemical shifts titration curves

$^1H$ N-acetyl-L-Histidine methylamide chemical shifts versus $^1H$ L-Histidine chemical shifts titration curves in the (A) presence and (B) absence of 20 mM sodium sulfate. The estimated pK$_a$ of N-acetyl-L-Histidine methylamide was (A) 6.52 ± 0.03 and (B) 6.44 ± 0.02.
NMR resonance from intermediate peak in the unfolded state of P protein was assigned by comparing spectra of F107W P protein to spectra of two variants, (A) H3A and (B) H22A, that each lack a different histidine residue. The left sides of panels (A) and (B) correspond to spectra that were collected at a sample pH of 5.85 and 6.10, respectively.
Figure S.19: $^1$H-$^{13}$C HSQC unfolded P protein spectra in various pH buffers

The intensity of the H22 intermediate peak (H22 I) increases as the pH is increased.
Figure S.20: Pyridine (internal standard) and P protein histidine residue titration curves

(A) Pyridine (internal standard) and (B) P protein histidine residue titration curves in the presence of 20 mM sulfate using pyridine as the internal standard. Inset table lists the best-fit pKₐ values, using pyridine as an internal standard instead of L-histidine.
Figure S.21: Determination of the tautomeric state of the neutral form of the protein and model compound histidines

In (A) 20 mM NaSO\textsubscript{4} (folded P protein) and (B) 0 mM NaSO\textsubscript{4} (unfolded P protein). All P protein histidines in both folding and unfolding conditions show very similar tautomeric state populations to the model compound N-acetyl-L-Histidine-methylamide, indicating that the protein histidines are highly solvated and form water hydrogen bonds under all conditions. The error bars represent 95% confidence levels obtained from peak fitting. The beige bands represent 95% confidence levels in the best-fit chemical shift. The fitting function is described in Materials and Methods.
Table S.1: Histidine pKₐ values and Hill coefficients in unfolded, intermediate, and sulfate-bound folded P protein based on fitting $^{13}$C⁰ data.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$pK_a^F$</th>
<th>$n^F$</th>
<th>$pK_a^I$</th>
<th>$n^I$</th>
<th>$pK_a^U$</th>
<th>$n^U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>His 3</td>
<td>6.32 ± 0.02</td>
<td>1.10 ± 0.06</td>
<td>5.73 ± 0.04</td>
<td>0.94 ± 0.09</td>
<td>5.73 ± 0.04</td>
<td>0.94 ± 0.09</td>
</tr>
<tr>
<td>His 22</td>
<td>6.10 ± 0.05</td>
<td>0.99 ± 0.12</td>
<td>5.40 ± 0.55</td>
<td>0.91 ± 0.41</td>
<td>6.11 ± 0.04</td>
<td>0.95 ± 0.09</td>
</tr>
<tr>
<td>His 105</td>
<td>5.55 ± 0.05</td>
<td>0.90 ± 0.03</td>
<td>5.80 ± 0.03</td>
<td>0.96 ± 0.08</td>
<td>5.80 ± 0.03</td>
<td>0.96 ± 0.08</td>
</tr>
</tbody>
</table>

*Solutions contained 10 mM L-Histidine, 10 mM pyridine in 90% H₂O, 10% D₂O at 25°C with either 0 mM or 20 mM sodium sulfate. The pKₐ values were obtained by fitting the data to Eq 5. The $^{13}$C⁰ shifts were used to determine the pKₐ values. Uncertainties represent the 95% confidence intervals and were obtained as described in Materials and Methods.

Table S.2: Solvent exposure of the histidine sidechains in P protein, relative to Gly-His-Gly, calculated as described in Materials and Methods

<table>
<thead>
<tr>
<th>Residue</th>
<th>Relative accessible surface area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>His 3</td>
<td>70</td>
</tr>
<tr>
<td>His 22</td>
<td>61</td>
</tr>
<tr>
<td>His 105</td>
<td>57</td>
</tr>
</tbody>
</table>
Table S.3: $^1$H and $^{13}$C chemical shifts of internal standard L-Histidine, Pyridine, and P protein histidine residues

<table>
<thead>
<tr>
<th>Residue</th>
<th>$^1$H (ppm)</th>
<th>$^{13}$C (ppm)</th>
<th>$^1$H (ppm)</th>
<th>$^{13}$C (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Histidine (pH~4)</td>
<td>8.66 ± 0.01</td>
<td>136.25 ± 0.03</td>
<td>8.65 ± 0.01</td>
<td>136.26 ± 0.02</td>
</tr>
<tr>
<td>L-Histidine (pH~8)</td>
<td>7.72 ± 0.01</td>
<td>138.59 ± 0.03</td>
<td>7.72 ± 0.01</td>
<td>138.59 ± 0.02</td>
</tr>
<tr>
<td>Pyridine (pH~4)</td>
<td>8.08 ± 0.01</td>
<td>129.47 ± 0.05</td>
<td>8.09 ± 0.01</td>
<td>129.49 ± 0.03</td>
</tr>
<tr>
<td>Pyridine (pH~8)</td>
<td>7.45 ± 0.01</td>
<td>126.58 ± 0.02</td>
<td>7.46 ± 0.01</td>
<td>126.57 ± 0.01</td>
</tr>
<tr>
<td>His 3 (pH~4)</td>
<td>8.61 ± 0.01</td>
<td>135.81 ± 0.06</td>
<td>8.62 ± 0.01</td>
<td>135.99 ± 0.02</td>
</tr>
<tr>
<td>His 3 (pH~8)</td>
<td>7.69 ± 0.01</td>
<td>138.33 ± 0.07</td>
<td>7.67 ± 0.01</td>
<td>137.98 ± 0.04</td>
</tr>
<tr>
<td>His 22 (pH~4)</td>
<td>8.60 ± 0.01</td>
<td>135.82 ± 0.05</td>
<td>8.70 ± 0.01</td>
<td>136.28 ± 0.05</td>
</tr>
<tr>
<td>His 22 (pH~8)</td>
<td>7.67 ± 0.01</td>
<td>138.48 ± 0.08</td>
<td>7.74 ± 0.01</td>
<td>138.46 ± 0.09</td>
</tr>
<tr>
<td>His 105 (pH~4)</td>
<td>8.54 ± 0.01</td>
<td>135.65 ± 0.06</td>
<td>8.65 ± 0.01</td>
<td>136.14 ± 0.07</td>
</tr>
<tr>
<td>His 105 (pH~8)</td>
<td>7.69 ± 0.01</td>
<td>138.52 ± 0.06</td>
<td>7.81 ± 0.01</td>
<td>137.99 ± 0.05</td>
</tr>
</tbody>
</table>

The imidazole ring is fully protonated at (pH~4) or deprotonated at (pH~8). Uncertainties represent the 95% confidence intervals and were obtained as described in Materials and Methods.
References


115. Gualfetti, P. J.; Bilsel, O.; Matthews, C. R. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of
tryptophan synthase from Escherichia coli. Protein Science: A Publication of the Protein Society 1999, 8, 1623-1635.


Biography

Pamela Lynnette Mosley was born on August 26, 1988 in Hackensack, New Jersey. She graduated from Duke University in May 2010 with a Bachelor of Science in Chemistry. In August 2010, Pamela began graduate school at Duke University and joined the lab of Terrence Oas. While at Duke, she was awarded the Structural Biology and Biophysics Fellowship and the Pathway to PhD Fellowship. Pamela has authored two publications:
