FACTORS DETERMINING THE pKa VALUES OF THE

IONIZABLE GROUPS IN PROTEINS: THEIR INTRINSIC pKas

AND THE EFFECTS OF HYDROGEN BONDING ON BURIED

CARBOXYL GROUPS

A Dissertation

by

RICHARD LEE THURLKILL

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

DOCTOR OF PHILOSOPHY

December 2005

Major Subject: Biochemistry

FACTORS DETERMINING THE pKa VALUES OF THE

IONIZABLE GROUPS IN PROTEINS: THEIR INTRINSIC pKas

AND THE EFFECTS OF HYDROGEN BONDING ON BURIED

CARBOXYL GROUPS

A Dissertation

by

RICHARD LEE THURLKILL

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	C. Nick Pace
Committee Members,	J. Martin Scholtz
	Victoria J. DeRose
	Michael P. Kladde
Head of Department,	Gregory D. Reinhart

December 2005

Major Subject: Biochemistry

ABSTRACT

Factors Determining the pKa Values of the Ionizable Groups in Proteins: Their Intrinsic

pKas and the Effects of Hydrogen Bonding on Buried Carboxyl Groups.

(December 2005)

Richard Lee Thurlkill, B.S., Louisiana Tech University

Chair of Advisory Committee: Dr. C. Nick Pace

A goal of the modern protein chemist is the design of novel proteins with specific activities or functions. One hurdle to overcome is the ability to accurately predict the pKas of ionizable groups upon their burial in the interior of a protein, where they are typically perturbed from their intrinsic pKas. Most discussion of intrinsic pKas is based on model compound data collected prior to the 1960's. We present here a new set of intrinsic pKas based on model peptides, which we think are more applicable than the model compound values. We observe some differences with the model compound values, and discuss these by critically examining the compounds originally used for the dataset. One interaction affecting the pKas of ionizable groups in proteins that is not well understood is the effect of hydrogen bonds. The side chain carboxyl of Asp33 in RNase Sa is buried, forms 3 intramolecular hydrogen bonds, and has a pKa of 2.4 in the folded protein. One of these hydrogen bonds is to the side chain hydroxyl of Thr56. We mutated Thr56 to alanine and valine and observed that the mutations relieves the perturbation on the carboxyl group and elevates its pKa by 1.5 and 2 units, respectively. The side chain carboxyl of Asp76 in RNase T1 is completely buried, forms 3 intramolecular hydrogen bonds to other side chain groups, and has a pKa of 0.5 in the

folded protein. Mutating any of the hydrogen bonding groups to the carboxyl affects its pKa differently, depending on the group mutated. Mutating all of the hydrogen bonding groups, creating a triple mutant of RNase T1, reverses the perturbation on the pKa and elevates it to about 6.4, very near the observed pKa of other carboxyl groups buried in hydrophobic environments. We compared these experimental results with predicted results from theoretical models based on the Solvent Accessibility Corrected Tanford-Kirkwood Equation and the finite difference solution to the linearized Poisson-Boltzmann Equation. The comparisons revealed that these models, most often used by theoreticians, are flawed when typically applied, and some possible improvements are proposed.

DEDICATION

I dedicate this work to my mom and dad, Harry and Inez Thurlkill, who gave to me the ability to dream the dream of completing my PhD. To my wife Cathy, who encouraged me, stood beside me and helped me to complete this work, and to my daughter, Elizabeth, who tolerated the hours and labor that went into this work I am eternally grateful and love dearly.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The ionizable groups in a protein define the acid/base characteristics of that protein. The study of the acid/base properties of proteins began in 1917 when Sörensen showed that egg albumin is an ampholyte.(*1*) More importantly, he showed that it contains several ionizable groups. Most of the common ionizable groups in proteins are located on the side chains of the amino acids.(*2*)

The ionizable groups are important to protein chemists because of their influence on conformational stability, solubility and catalytic activity. The change in the conformational stability of a protein as a function of pH is dependent on the ionizable groups in the folded and the unfolded protein.(*3*) For typical cytosolic proteins, the higher the percentage of ionizable or polar groups the more soluble they are.(*4*) A recent survey showed that amino acids with side chain ionizable groups make up about 25% of all the residues in a typical protein, but residues with side chain ionizable groups make up about 65% of the catalytic residues in the active sites of enzymes.(*5*) This underscores the importance of the ionizable groups of proteins

THE IONIZABLE GROUPS OF PROTEINS

The ionizable groups can be divided into two categories, the neutral acids and the

_____

This dissertation follows the style of *Biochemistry.*

cationic acids. The neutral acids can be modeled by the following equation:

$$HA \xleftrightarrow{Ka} H^+ + A^- \qquad (1)$$

where HA is the ionizable group with a proton bound, A⁻ is the group without a bound

proton, or the conjugate base of HA, H^+ is a proton and Ka is the acid dissociation

constant. The common neutral acids found in proteins and the amino acid where each

group appears is shown in Figure 1. It can be seen that when a neutral acid binds a

proton, the group has a neutral charge.

The cationic acids can be modeled by the following equation:

$$HB^+ \xleftrightarrow{Ka} H^+ + B \qquad (2)$$

where HB^+ is the ionizable group with a proton bound, B is its conjugate base, H^+ is a

proton and Ka is the acid dissociation constant. The common cationic acids found in

proteins and the amino acid where each group appears is shown in Figure 2. It can be

seen that when a cationic acid binds a proton, the group has a positive charge.

THE pKa AND THE HENDERSON-HASSELBALCH EQUATION

The acid dissociation constant, Ka, from eq 1 describes the equilibrium between the

charged form and the neutral form of the respective ionizable group. It is defined as

follows:

$$Ka = \frac{[A^-] \times [H^+]}{[HA]}. \qquad (3)$$

If we take the negative logarithm of both sides of eq 3 and separate terms we get:
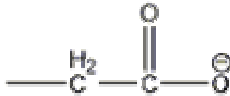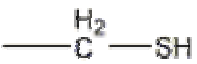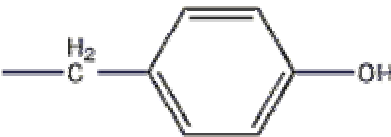
Group	Structure	Intrinsic pKa
Alpha Carboxyl		3.8
Aspartic Acid (Asp)		4
Glutamic Acid (Glu)		4.4
Cysteine (Cys)		9.5
Tyrosine (Tyr)		9.6

Figure 1: The neutral acid ionizable groups of proteins. Each group is shown in its typical protonation state at pH 7.
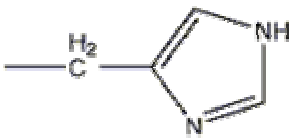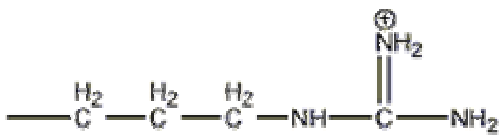
Group	Structure	Intrinsic pKa
Histidine (His)		6.3
Alpha Amino		7.5
Lysine (Lys)		10.4
Arginine (Arg)		12

Figure 2: The cationic acid ionizable groups of proteins. Each group is shown in its typical protonation state at pH 7.

$$-\log(Ka) = -\log[H^+] + (-\log\frac{[A^-]}{[HA]}). \qquad (4)$$

We can apply our definition of pH (pH = -log [H$^+$]) to Ka and to [H$^+$]. If we then

rearrange terms we get:

$$pH = pKa + \log\frac{[A^-]}{[HA]}. \qquad (5)$$

Eq 5 is the commonly written form of the Henderson-Hasselbalch equation, which

defines the relationship between the concentration of an acid and its conjugate base as a

function of pH.

   If [A$^-$] = [HA], then log ([A$^-$]/[HA]) = 0 and eq 5 reduces to:

$$pH = pKa \qquad (6)$$

Therefore, the pKa of an ionizable group is the pH where the concentration of the acid

form of the ionizable group, HA, equals the concentration of its conjugate base, A$^-$, or it

is the pH where half of the ionizable group is protonated and half is deprotonated. If we

know the pKa of an ionizable group and the pH of a solution containing that group, we

can use eq 5 to determine the percentage of the group protonated, HA, and the

percentage deprotonated, A$^-$. We can apply the same discussion of the Henderson-

Hasselbalch equation to cationic acids.

THE INTRINSIC pKa OF IONIZABLE GROUPS

   Each ionizable group has an intrinsic pKa. A group's intrinsic pKa is the pKa of that

group when it is fully solvent exposed and not interacting with any other local group.(*6*)

Estimates of the intrinsic pKas for the ionizable groups found in proteins have been

made by Nozaki and Tanford and are shown in the last column of Figures 1 and 2.(*7*)

The intrinsic pKas determined by Nozaki and Tanford were based on the pKas of

ionizable groups in model compounds.(*6*) The model compounds were chosen based on

the similarity of their structures with the structures of the amino acids containing the

ionizable groups.

Almost any change in the local environment of an ionizable group can perturb the

pKa of that group, resulting in a different, observed pKa. When a protein folds into its

three dimensional conformation, ionizable groups usually remain on, or near, the surface

of the protein where they remain exposed to solvent and typically the perturbations on

their pKa are small, <2 units. Sometimes these groups are sequestered into local

environments removed from solvent. This sequestering of an ionizable group often

results in significant pKa perturbations, >2 units.(*8*)

HOW IONIZABLE GROUP pKas ARE PERTURBED

Perturbations of a group's pKa can result from charge-charge interactions (both

short-range contacts and long-range global effects), burial in a hydrophobic

environment, or hydrogen bonding. If either a neutral acid or a cationic acid is brought

into close contact with a positive charge, the equilibrium described by eq 1 or eq 2 will

shift to the right due to charge-charge interaction. The pKa of the group will be lower as

a result of the interaction. When either type of acid is brought into close contact with a

negative charge the equilibrium of the equation describing the ionization will shift to the

left, resulting in a higher pKa for the group. When a neutral acid is buried in a

hydrophobic region the equilibrium described by eq 1 will shift to the left and the pKa of the group will increase. When a cationic acid is buried in a hydrophobic region we expect that the equilibrium in eq 2 will shift to the right. If the group is exposed to bulk solvent and the solvent conditions are changed to resemble a more hydrophobic environment, the equilibrium will probably not shift because the net charge on both sides of eq 2 are the same. But with the group buried in a hydrophobic environment in a protein, the free H^+ will be allowed to migrate out of the protein to solvent so the charge on the right side of eq 2 becomes zero. The equilibrium of the equation will shift to the right, to the neutral form of the group, and the pKa will decrease as a result. The effect of hydrogen bonding on the apparent pKa of a group is a little more complicated. It depends on whether the ionizable group is a neutral acid or cationic acid, whether the group is the hydrogen bond (h-bond) donor or acceptor or both and whether the hydrogen bond(s) is (are) charge-neutral or charge-charge.

Perturbations on Surface Exposed Groups are Typically Small: The carboxyl group containing residues in Ribonuclease Sa (RNase Sa) that are the most exposed to solvent are Asp17, Asp25, Glu41 and Glu74.(*9*) The carboxyl group pKas for each of these groups have been determined and are 3.72, 4.87, 4.14 and 3.47, respectively.(*10*) Each of these groups is at least 63% solvent exposed, none of the groups form any hydrogen-bonding contacts and the nearest charged group to any is a negative charge more than 6 Å from Asp17. The pKas of these groups are mostly influenced by long-range electrostatic interactions, and the perturbations from intrinsic values are small, in this case less than 1 unit for each.

The carboxyl groups in Ribonuclease T1 (RNase T1) that are the most exposed to solvent are Glu31, Asp49, Asp66 and Glu102. The pKas of these groups are 5.36, 4.22, 3.9 and 5.3, respectively.(*11*) Each group is at least 68% exposed to solvent and even though two of these, Glu31 and Glu102, appear to form electrostatic interactions with other groups in the crystal structure, the perturbations on the pKas of these groups are small, demonstrating that surface charged groups typically are only marginally perturbed, due to high exposure to solvent.

Charge-Charge Interactions Perturb pKas: Long-range electrostatic interactions have been shown to perturb pKas of ionizable groups. A survey of the measured pKas of Asp and Glu carboxyl groups in proteins showed that the average Asp carboxyl in a protein has a pKa of 3.4.(*12*) From Figure 1 the intrinsic pKa of the Asp carboxyl is 4.0. The average pKa for a Glu carboxyl in a protein is 4.1. From Figure 1 the intrinsic pKa for a Glu carboxyl is 4.4. The average pKa for a His imidazole in a protein is 6.5, and from Figure 2 its intrinsic pKa is 6.3.(*13*) The interpretation of these results suggest that at near neutral pH the net charge on the average protein in the survey is near 0 to slightly negative and the pKa of the average His imidazole is 0.2 units higher than its intrinsic value. As we lower the pH the net charge on the average protein becomes positive which perturbs the pKa of the Glu carboxyl 0.3 units lower than its intrinsic value. Further lowering of the pH increases the net positive charge on the protein and the perturbation on the Asp carboxyl, 0.6 units, is larger than the perturbation on the Glu carboxyl. This suggests that long-range electrostatic interactions, and thus the net charge on a protein, influence the pKa of an ionizable group in a protein.

The effects of long-range electrostatic interactions was also demonstrated with an acidic protein, RNase Sa, pI ~ 3.7, and a charge reversal variant, RNase Sa 5K, pI ~10, in which 5 exposed carboxyl groups were replaced with lysines.(*10*) The pKas of all the groups measured in the 5K variant were lower than the pKas in the wild type protein. This was attributed to greater positive charge on the 5K variant of the protein than on the wild type protein.

A salt bridge in T4 lysozyme between His31 and Asp70 results in highly perturbed pKas for both groups. The carboxyl group oxygens of the Asp70 side chain are ~81% buried and the imidazole nitrogens are ~87% buried. The interatomic distance between the carboxyl group and the imidazole group is ~4 Å. The estimated pKa for the side chain carboxyl of Asp70 is 0.5 in the folded protein and 3.7 in the unfolded protein. The estimated pKa for the imidazole group of His 31 is 9.1, folded, and 6.8, unfolded.(*14*) If either residue is mutated to an asparagine the perturbation on the pKa of the other group is relieved and its pKa is near that of the estimated unfolded pKa in the wild type protein. This indicates that the perturbation on each group of the salt bridge is due almost entirely to interaction with the other group of the salt bridge.

Lys115 in acetoacetate decarboxylase from *Clostridium acetobutylicum* supplies the unprotonated amino group that undergoes Schiff-base formation in the mechanism of action for the enzyme. The pKa of the ε-amino group of Lys115 is estimated at 5.9.(*15*) Highbarger and Gerlt showed, by use of a reporter group attached to the side chain of Lys115, that the ε-amino group of Lys116 supplies the interaction that results in the >4 unit perturbation of the Lys115 ε-amino group.(*16*) In the wild type protein the

secondary amine of the reporter group attached to the ε-amino of Lys115 has a pKa of 6.4; in the free compound the pKa is 10.6. In mutant proteins, K116C and K116N, the pKa of the secondary amine is >9.2, while in K116R the pKa is 6.3. This provided strong evidence that the interaction with Lys116 perturbs the pKa of the Lys115 ε-amino group to ~5.9.

Hydrophobic Burial Perturbs pKas: The side chain carboxyl of Asp79 in RNase Sa is a naturally occurring ionizable group that is 85% buried. The nearest charged groups to the carboxyl group are 7.8 Å and 12 Å. The group does not form any intramolecular hydrogen bonds (h-bonds); it does form an h-bond to a crystallographic water molecule. There are two polar groups within 3.5 Å of the carboxyl oxygens but the h-bond predictive program used by our lab, *pfis*, does not predict h-bonds between the carboxyl oxygens of Asp79 and either of these polar groups.(*17*) The pKa of the Asp79 carboxyl is 7.4, among the highest reported pKas for an Asp carboxyl and about 3.5 units higher than the intrinsic value.(*10*) This elevated pKa is primarily a result of its location in a hydrophobic environment, which induces a shift in the equilibrium of proton binding to the neutral acid as described by eq 1. The shift in equilibrium of eq 1 to the left, i.e. the high pKa of the carboxyl group, shows that the protein favors a protonated, or neutral, carboxyl group at this position. Mutational studies at this site have also shown that any other amino acid placed at position 79, except Glu, increases the conformational stability of the protein, in some cases by over 3 kcal/mole.(*18*)

The side chain carboxyl oxygens of Asp26 in *E. coli* thioredoxin are 100% buried. The residue is located at the bottom of a hydrophobic cavity near the active site, and it is

completely conserved within the thioredoxin family. The pKa of the Asp26 carboxyl in oxidized thioredoxin is estimated at 7.5.(*19, 20*) Estimates of the pKa of the group in reduced thioredoxin were originally >9, but later work suggested its value is the same as in the oxidized protein, 7.5.(*21, 22*) Dyson, et al confirmed what was suggested by Langsetmo, et al, that the Asp26 carboxyl group forms an electrostatic interaction with the ε-amino group of Lys57. This interaction apparently lowers the pKa of the Asp26 carboxyl from >9 to ~7.5.(*23*) They also showed that these two groups strongly influence the catalytic activity of the protein as a function of pH through their interaction with the active site cysteine thiols.

Staphylococcal nuclease (SNase) has been used as a model to estimate the effects of placing ionizable groups in the hydrophobic core of a protein. Val66 is a core residue of SNase that is completely removed from bulk solvent. Early mutation studies of Val66 to Lys showed by x-ray crystallography that the ε-amino group of the lysine side chain is completely buried and has a pKa of ~6.4.(*24*) Later studies of the V66K mutant in backgrounds of hyper stable variants of SNase resulted in pKas of 6.35 and 5.76.(*25*) The location of the Lys side chain in the hydrophobic core favors the deprotonated form of the ε-amino group and the pKa is lower than its intrinsic value by about 4.5 units. Burial of a glutamic acid at position 66, V66E, was shown to favor the protonated form of the carboxyl group. The pKa of the Glu side chain in V66E is ~8.9, about 4.5 units higher than its intrinsic value.(*26*) Ile92 is also in the hydrophobic core of the protein and mutations to Glu and Lys result in pKas of the respective ionizable groups of 8.8 and 5.7.(*27*) These are some of the most extremely perturbed pKas reported, and the

agreement between the magnitudes of the perturbations indicate that the shifts are a result of the local environment and are not group dependent.

Hydrogen Bonding Perturbs pKas: The Ser-His-Asp triad of serine proteases provides a good example of the impact of an h-bond on the pKa of a group. An existing h-bond between the His and Asp side chains gets shorter upon binding the substrate peptidyl group. The shortening of the h-bond is thought to polarize the imidazole ring of the His. The increased basicity of the imidazole ring is then able to abstract the hydroxyl proton from the serine side chain, which then performs a nucleophilic attack upon the carbonyl carbon of the substrate's peptide backbone. Under normal conditions the pKa of the His imidazole group is about 7.5, but upon binding substrate the shortened h-bond increases the pKa to an estimated 10-12. This increased pKa is believed high enough to perform the proton abstraction from the Ser hydroxyl, which then performs its nucleophilic attack.(*8*) This demonstrates the variability of impact h-bonds can have on the pKa of an interacting group.

The side chain carboxyl of Asp121 in RNase A has a pKa of 2.4 in the folded protein.(*28*) This perturbation is the result of h-bonds formed with the His119 imidazole and with the amide nitrogen of Lys66. In separate work it was determined that the pKa of the His119 imidazole group in the D121N and D121A mutants is not appreciably different from its pKa in wild type RNase A.(*29*) The results suggested that multiple h-bonds can induce relatively large perturbations on the pKa of an ionizable group, whereas a single h-bond may not be able to.

THEORETICAL PREDICTION METHODS FOR pKas

Biochemists generally use one of two approaches to predict pKas of ionizable groups in proteins, the Solvent-Accessibility Corrected Tanford-Kirkwood Equation (SATK) or the Finite Difference Poisson-Boltzmann Equation (FDPB). Both approaches require some form of a structure, typically a crystal structure and knowledge of the conditions under which the models will be applied, for example the ionic strength of the solution and the dielectric constants of the solution and of the protein interior. A comprehensive discussion of these theoretical models and their development is beyond the scope of this discussion, however I will present a short description of the equations here.

The Solvent-Accessibility Corrected Tanford-Kirkwood Equation: The Solvent-Accessibility Corrected Tanford-Kirkwood Equation (SATK) was developed in an attempt to explain the observations from studies of pH titration curves of proteins.(*30-32*) A simplified form of the equation that describes the interaction energy, E_{ij}, between any two charges in the protein, *i* and *j*, can be written as:(*33*)

$$E_{ij} = \varepsilon^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}) \qquad (7)$$

where A_{ij}, B_{ij}, and C_{ij} are separate functions of the positions of the charges, the dielectric constants of the solvent and protein, and the ionic strength of the solution as defined by Tanford and Kirkwood.(*31*) ε is the unit charge, *b* is the radius of the sphere that represents the protein, *a* is the radius of the sphere that is impenetrable to solvent, which is typically taken as *b* plus the average radius of the ions in the solution (for a solution

with NaCl, $a=b+1.4\ Å$). SA_{ij} is the mean accessibilities to solvent of groups i and j. As written, eq 7 describes the interaction energy between any two charges, but proteins have numerous charged groups. In order to apply the equation to a protein one has to sum the individual interaction energies between each pair of groups. Since interactions between ionizable groups can affect the ionization state of those groups, which may in turn affect the ionization states of other groups, several cycles of computations need to be performed in order to minimize the calculated energies. Computer programs are available to perform these calculations. The program used in our lab is called TKBK, for Tanford-Kirkwood/Bashford-Karplus, and is described by Ibarra-Molero, et al.(*33*) The calculation suite is capable of applying SATK calculations to a protein's crystal structure, along with the Bashford-Karplus reduced set-of-sites approximations for fractional protonations of ionizable groups. The solvent accessibility calculations are determined from the crystal structure of the protein. For each atom of each amino acid of interest, X, the solvent accessible surface area is calculated. The accessible surface area of each atom is used with the accessible surface area of the corresponding atom in the tri-peptide Gly-X-Gly to determine the percent accessibility of the each atom.(*34*) One criticism of applying SATK, which will be shown later, is that it does not take into account the effect of partial charges from polar groups on the ionizable residues. Another criticism is that it assumes that all the ionizable groups are near the surface of the protein. This prevents SATK from modeling deeply buried ionizable groups in a low dielectric environment. That said, however, it has been shown that SATK can predict

reasonably well the pKas of surface exposed residues and the effect of salt on their

pKas.(*35, 36*)

The Finite Difference Poisson-Boltzmann Equation: The linearized Poisson-

Boltzmann Equation was derived from Gauss' Law, which relates the divergence of the

electric field of a point charge to the charge density in solution. The charge density in

solution is described by a Boltzmann distribution. For its application to the electrostatic

interactions in proteins, or macromolecules in general, we can express the equation as

follows:

$$\Delta G_{PB} = \Delta G_{Born} + \Delta G_{Bkgd} + \Delta G_{ij} \qquad (8)$$

where ΔG_{PB} is the electrostatic potential on an ionizable group, i, ΔG_{Born} is the Born self

energy, the energy of moving ionizable group, i, from a medium of one dielectric to a

medium of a different dielectric, in this case from solvent to its position in the three

dimensional structure of the protein. Throughout this report I will use hydrophobic

burial to refer to the Born self energy of burying a group in a hydrophobic environment

in a protein. ΔG_{Bkgd} is the energy of interaction of group i with partial charges from

polar groups within the protein, and ΔG_{ij} is the electrostatic, or Coulombic, interaction

energy between group i and ionizable group j.(*37, 38*) The linearized Poisson-

Boltzmann equation is a partial differential equation. One solution for the equation is by

a finite-difference approach. The University of Houston Brownian Dynamics Suite of

programs (UHBD) uses a finite-difference approach to solve the linearized Poisson-

Boltzmann equation (FDPB).(*39*) In UHBD a three dimensional grid is set up containing

the protein crystal structure which leads to a series of equations relating to the FDPB for

each grid. These series of equations are solved and each of the energy terms in eq 8 is

calculated, leading to the final solution for the FDPB equation for the protein. The

FDPB approach is a more rigorous application than SATK, and FDPB has had

reasonably good success in predicting the pKas of groups near the surface of

proteins.(*10, 36, 40*) There is evidence suggesting that the respective terms in eq 8 do

not properly estimate the energies. The Born Self Energy, or Desolvation Energy, may

overestimate the penalty for burial in a hydrophobic environment.(*10, 36*) An

alternative could be that the values of the dielectric of the interior of a protein that are

commonly used are not correct. The ΔG_{Bkgd} is the main term of interest in the present

work. Previous work has suggested that the ΔG_{Bkgd} is typically underestimated for

ionizable groups that form h-bonds.(*10, 41, 42*)

MODEL SYSTEMS

Intrinsic pKas: As stated earlier, the intrinsic pKas for the ionizable groups were

originally estimated based on model compound pKas.(*6*) Most researchers accept that

the Nozaki and Tanford intrinsic pKas are reasonable estimates, and there is

considerable evidence to suggest they are.(*10-13*) To the best of our knowledge,

however, there is no supporting work based on model peptide studies. We believe a

model peptide system of the form Ac-AAXAA-NH$_2$ is a good model to test the

Nozaki/Tanford intrinsic pKa values.

Asp33 in RNase Sa: RNase Sa is a small monomeric protein, 96 residues, of the α +

β family. It consists of a three-turn α-helix packed against a five-stranded antiparallel β-

sheet, and has one disulfide bond linking residues 7 and 96. A ribbon diagram of its crystal structure is shown in Figure 3A. It was originally isolated from the bacterium *Streptomyces aureofaciens*, and is one of three isozymes made by different strains of the bacteria.(*43*) The thermal and chemical unfolding of the protein has been well characterized.(*9, 44, 45*) It closely follows a two-state unfolding mechanism, and the unfolding is completely reversible, making it a very good system to study protein folding.

The side chain carboxyl of Asp33 forms three intramolecular h-bonds along with an h-bond to a crystallographic water molecule, is completely buried and has the lowest pKa among the Asp or Glu carboxyls in folded RNase Sa, 2.4.(*10*) Having an Asp in that position is critical to the conformational stability of the protein. Either an alanine or an asparagine at position 33 decreases the stability by >4 kcal/mole. The three h-bonds are shown in the ribbon diagram in Figure 3B. They are 3.2 Å to the amide N of Tyr30, 2.8 Å to the amide N of Thr56 and 2.6 Å to γ hydroxyl of Thr56.

Asp76 in RNase T1: RNase T1 was the first microbial ribonuclease discovered. It was isolated from the mold fungus, *Aspergillus oryzae*, which is used in making sake and soy sauce.(*44, 46*) It has 104 residues, 2 disufide bonds linking residues 2 - 6 and 10 – 103. RNase T1 is also a member of the α + β family of proteins, consisting of a four-and-a-half-turn α-helix packed against a four-stranded antiparallel β-sheet. A ribbon diagram of the crystal structure is shown in Figure 4A. The thermal and chemical unfolding of the protein has been extensively studied and well characterized.(*47-51*) It

A

B
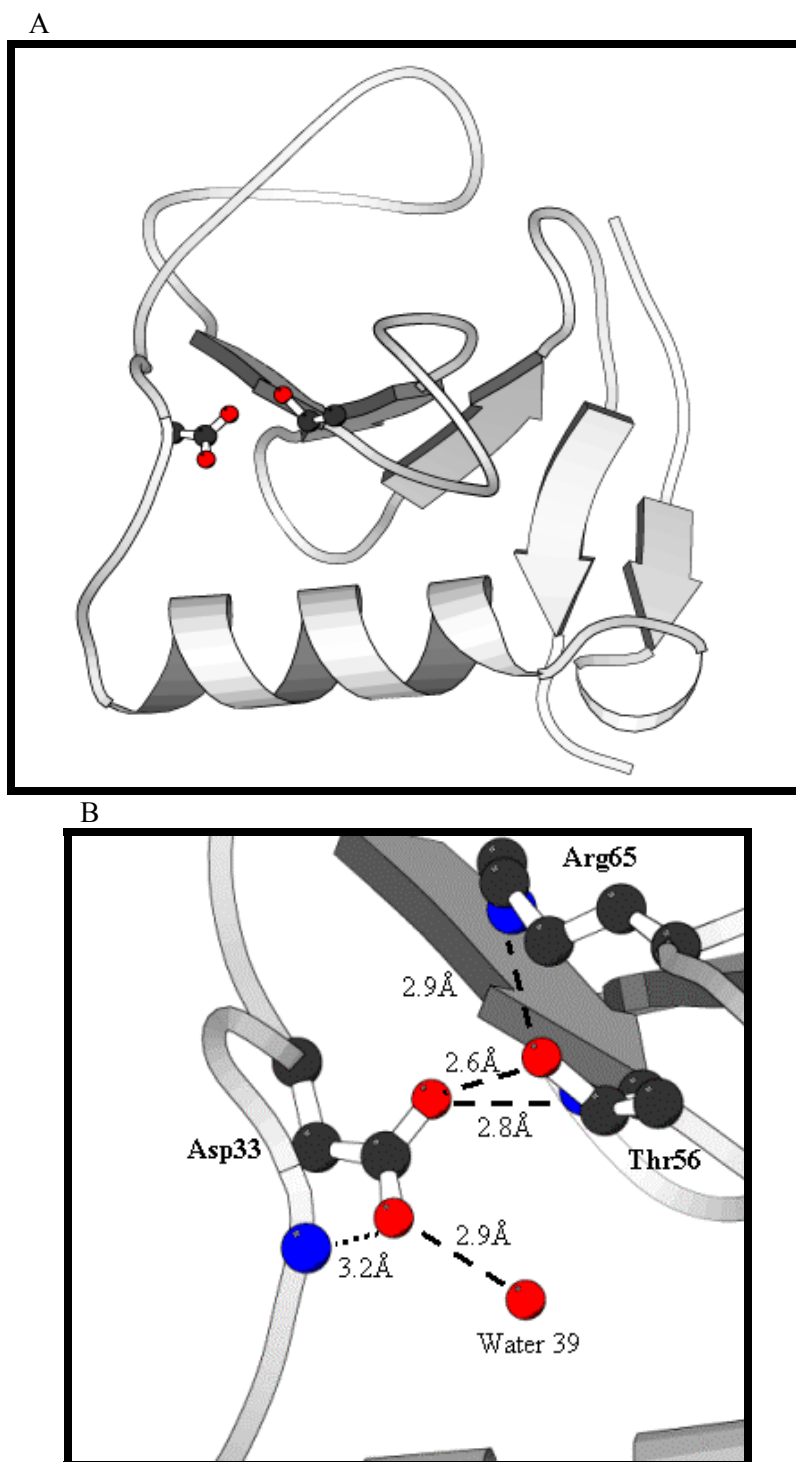
Figure 3: Ribbon diagram of RNase Sa. A: PDB file 1RGG showing the location of Asp 33 and Thr 56.(*52*) B: Close up view of the area around Asp 33 showing the hydrogen bonds formed by the side chain carboxyl of Asp 33 and their distances. These diagrams were made with Molscript.(*53*)

19
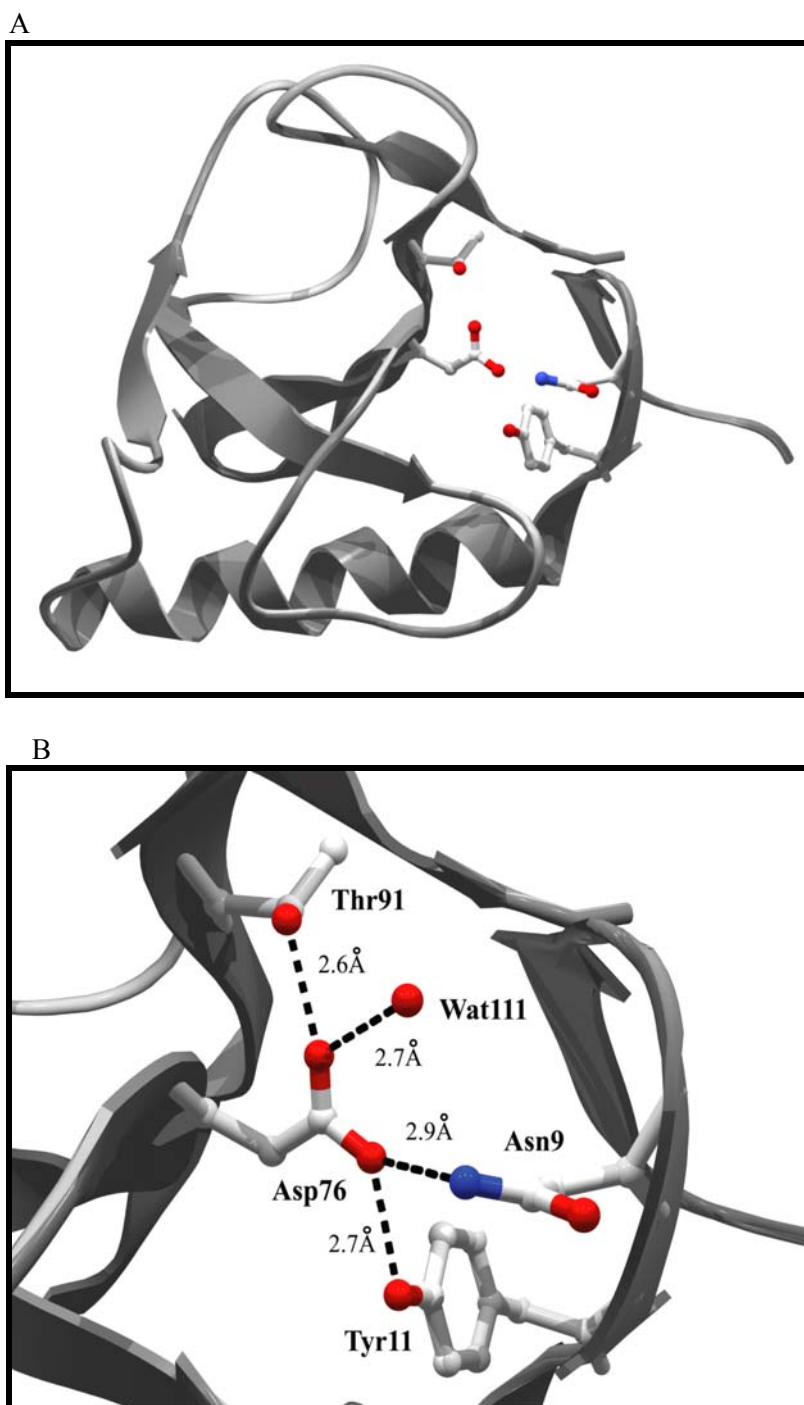
Figure 4: Ribbon diagram of RNase T1. A: PDB file 9rnt showing the location of Asp 76 and the residues around it.(*54*) B: Close up view of the area around Asp 76 showing part of the hydrogen bond network containing Asp 76 and the hydrogen bond distances. These diagrams were made with Swiss PDB Viewer.(*55*)

closely follows a two-state unfolding mechanism and is almost completely reversible, making RNase T1 a very good system to study protein unfolding.

The side chain carboxyl of Asp76 in RNase T1 is completely buried and fully hydrogen bonded. It forms three intramolecular h-bonds along with an h-bond to a highly conserved water molecule. It not only has the lowest pKa in RNase T1, but at 0.5 is one of the most acidic carboxyl groups known in proteins.(56) Asp76 is also critical to the stability of the protein. Mutations to Ser, Ala or Asn reduce the stability of the protein by >3.5 kcal/mole.(56) The h-bonds formed by the side chain carboxyl of Asp76 are shown in Figure 4B. They are 2.9 Å to the δ N of Asn9, 2.7 Å to the phenolic hydroxyl of Tyr11 and 2.6 Å to the γ hydroxyl of Thr91

OBJECTIVES

Our objectives for this work are three fold. In support of the Nozaki and Tanford work on intrinsic pKas, we will determine the intrinsic pKa of each ionizable group in proteins using model peptides. The model peptides are of the form Ac-AAXAA-NH$_2$ where X is the amino acid containing the ionizable group. We will use the RNase Sa and RNase T1 systems to determine the impact of intramolecular h-bonds on the pKa of the buried carboxyls of Asp33 in RNase Sa and Asp76 in RNase T1. We will apply the SATK and FDPB equations to the crystal structures of RNase Sa and RNase T1 to estimate the pKa of the side chain carboxyls of Asp33 and Asp76 in the wild type proteins and mutants where the hydrogen bonds are deleted. We will then compare the predicted results with experimentally determined results.

CHAPTER II

MATERIALS AND METHODS

MATERIALS

Rink Resin, for the C-terminal amidated, and Wang Resin, for the free carboxyl C-terminal peptides, were from Advanced ChemTech (Louisiville, KY). The amino acids were from Advanced ChemTech, and all other reagents were peptide synthesis grade or better and were from either Advanced ChemTech or Sigma-Aldrich (Milwaukee, WI). Purification of the peptides was performed by FPLC on an Äkta FPLC system from Amersham Pharmacia Biotech (Piscataway, NJ) using either reverse phase chromatography with a resource RPC column or by size exclusion chromatography with a Superdex™ Peptide 10/300GL column both from Amersham Pharmacia Biosciences (Piscataway, NJ). All reagents for purification were the best grade available, usually FPLC or HPLC grade and were from either Advanced ChemTech or Sigma-Aldrich.

Protein expression for RNase Sa and mutants was performed with the plasmid pEH100, which has been described elsewhere.(*43*) Protein expression for RNase T1 and mutants was performed with the plasmid pEHT1, which was constructed from pEH100 by removal of the RNase Sa gene with EcoR1 and Xbal and replacing it with the RNase T1 gene from pMc5TPRTQ.(*57*) Oligonucleotide primers for mutagenesis were from Integrated DNA Technologies (IDT, www.idtdna.com). Site directed mutagenesis was performed with a QuickChange™ Site-Directed Mutagenesis Kit from Stratagene (La

Jolla, CA). Minipreps of expressed plasmid were prepared with a QIAprep® Spin

Miniprep Kit from Qiagen, Inc. (Valencia, CA). Luria Broth, Terrific Broth, and Bacto-

Agar were from Difco (Becton Dickinson & Co, Sparks, MD). Expression hosts were

either *E. coli* strain RY1988 (MQ) or *E. coli* strain DS2000 (Described in Appendix

I).(*58*) Isopropyl-β-D-thiogalactoside (IPTG) was from Ambion, Inc (Beverly, CA), and

ampicillin, tetracycline and kanamycin were from Sigma-Aldrich. All reagents used in

the isolation and purification of the proteins and mutants were ACS grade or better and

from Sigma-Aldrich or Fisher Scientific (Pittsburgh, PA).

For the potentiometric analyses, chemicals were reagent grade or better and from

various venders, including Sigma-Aldrich, Fisher Scientific, EM Science (Gibbstown,

NJ), Mallinckrodt (Paris, KY) and VWR, Inc (West Chester, PA). For thermal and urea

unfolding studies of proteins all buffers were ultrapure grade and from Sigma-Aldrich,

J.T. Baker (Phillipsburg, NJ), or ICN Biomedicals (Aurora, OH). Urea was from

Nacalai Tesque (Kyoto, Japan), or ICN Biomedicals. All data was managed and

analyzed using Origin version 6.1 data analysis program from OriginLab Corporation

(Northampton, NJ).

METHODS

Peptides Synthesis and Purification: Peptide synthesis was performed at 0.1mmol

scale and followed typical solid phase peptide synthesis (SPPS) protocols using FMOC-

protected amino acids.(*59, 60*) All reactions were performed manually with a floor

shaker providing agitation. Removal of the protecting group was performed with 20%

piperidine with NMP as the solvent. The coupling reactions were carried out with

HBTU/DIPEA/HOBt activation of the C-terminal carboxyl group to generate the OBt

activated ester form of the amino acid to be coupled. Coupling times were generally 30

minutes to 2 hours. The growing peptide was capped after addition of the third amino

acid with acetic anhydride, and all peptides were capped with acetic anhydride after

deprotection of the final amino acid, except the free N-terminal peptide. Cleavage of the

peptide from the resin support was achieved with a cleavage cocktail: 90% TFA/5%

thioanisole/3% triisopropylsilane/2% anisole. The peptide was separated from the

cleavage cocktail by precipitation with MTBE followed by centrifugation. The crude

peptide was then lyophilized and purified by FPLC as stated above. The identity and

purity of each peptide was confirmed by MALDI-TOF Mass Spectrometry using a

Voyager DE Linear Mass Spectrometer from Applied Biosystems (Farmingham, MA),

courtesy of Larry Dangott of the Protein Chemistry Lab, Texas A&M University.

Protein Expression and Purification: RNase Sa and mutants were expressed and

purified as described previously, with some modifications.(*43, 58*) SP Sephadex was

used as the cation exchange resin with a pH gradient, pH 3.25 to pH 8 in succinate

buffer, to elute the protein.

RNase T1 and mutants were expressed and purified as described previously, with

some modifications.(*57*) DE52 from Whatman (Clifton, NJ) was the anion exchange

resin and the loading buffer was 15mM Tris, 3mM EDTA, pH 7.4. The protein was

eluted from the ion exchange resin with a 0-0.6M NaCl gradient. For some later RNase

T1 preps the RNase Sa purification protocol was used, with excellent results.

For most variants, one of the cell lines above was transformed with gene carrying plasmid and grown in 6L Terrific Broth (TB), with appropriate drugs, distributed in 12 2-L Erlenmeyer Flasks. The flasks were placed in a New Brunswick Incubator Floor Shaker (New Brunswick Scientific, Edison, NJ) set at 30 to 37° C and following induction with IPTG at OD_{600}=0.6, incubated overnight. For those protein variants with suspected thermal unfolding temperatures below about 35° C, an 11.5L capacity New Brunswick Fermentor was placed in a cold room. With the fermentor in the cold room the temperature of the media was maintained at 20-25° C for bacterial growth and protein expression typically using the DS2000 cell line.

Insertion of the desired mutation and amplification onto the gene carrying plasmid were performed following the directions in the QuickChange™ Kit Manual. Upon receipt, mutagenesis primers were dissolved and diluted to a concentration of 5pmol/μl. The thermocycling conditions were those suggested in the QuickChange™ Kit Manual and performed in an Applied Biosystems GeneAmp 2400 PCR System thermocycler. The PCR product was transformed into the supercompetent *E. coli* strain from the kit, and following overnight growth, minipreps of the plasmid were prepared using the Qiagen Miniprep kit. Sequencing of each plasmid was performed by the Gene Technologies Lab, Texas A&M University, and the integrity of each gene was confirmed through the entire sequence. Following expression and purification, the purity of each protein was confirmed to ≥99% by polyacrylamide gel electrophoresis, and mass was confirmed by MALDI-TOF mass spectrometry either as above or by the Laboratory for Biological Mass Spectrometry, Texas A&M University.

Potentiometric pKa Determinations on Peptides: The pKa of the ionizable group in each peptide was determined by potentiometric titration. The experimental design and protocol of this analysis has been described previously, and was applied with minor modifications.(*7, 61*) Our system uses a Hamilton MicroLab 500 syringe pump (Hamilton, Co, Reno, NV) with appropriate syringe for addition of titrant, a Corning Model 450 pH meter (Corning Inc, Corning, NY) and Beckman Futura pH electrode (Beckman Instruments, Fullerton, CA) for pH monitoring, a Fisher Model 9100 refrigerated circulating water bath for constant temperature control and a Thermolyne (Barnstead/Thermolyne, Dubuque, IA) magnetic stirrer for constant stirring during titrations. The Hamilton syringe pump and pH meter are computer controlled so that a preset volume of titrant, usually 2-5 μl, is added at specified intervals and the pH is monitored and recorded at specified times, usually 15-18 sec, within those intervals (The computer program controlling the system is courtesy of Dr. Joe Morgan and Dr. Jay Porter and their ENTC 359 class Fall 2000, Texas A&M University). The titration takes place in a sealed jacketed titration vessel from Metrohm (Brinkman Instruments, Westbury NY) under CO_2 free N_2 atmosphere. The titrants, HCl or NaOH, were standardized with primary standard, trizma base or potassium hydrogen phthalate (KHP), respectively, using Grans Procedures and were typically ~0.2 or ~0.5 M.(*62*) The titrations were performed in aqueous solutions of 0.1 M KCl, to provide constant ionic strength throughout the course of the titration the titrants were prepared in the same concentration KCl, and all solutions were extensively degassed prior to use.

A stock solution of peptide was prepared by dissolving a known mass of peptide in >3.0 ml 0.1 M KCl. One ml of this stock was added to 2 ml 0.1 M KCl in the titration vessel. If the titrant was HCl, the pH was adjusted to pH 4-5 with 5 M HCl and allowed to stir for several minutes to ensure the removal of all CO_2 from the solution. The pH was then adjusted with 5 M NaOH to a pH well above the expected pKa of the group of interest and the titration started. If NaOH was the titrant, the pH was adjusted to pH 4-5 or well below the pKa of the group of interest, whichever was lower, and allowed to stir for a few minutes before the titration was begun. Three independent solvent blanks were performed with 3 ml 0.1 M KCl each day under each set of titration conditions. When plotted, these blanks routinely overlayed and were indistinguishable from each other. The three blanks were averaged and the average was subtracted from each sample titration performed that day under the same conditions. Each peptide was analyzed with three independent titrations. The data output for each titration consisted of the measured pH at each dosing of titrant and the total volume of titrant added after each dosing. From these data, the total moles of titrant added after each dosing were calculated, and by dividing by the moles of peptide in the titration vessel, the moles H^+ taken up or released per mole peptide at each dosing was calculated.

The data, moles H^+ taken up or released per mole peptide were plotted against pH and fit to the following form of the Henderson-Hasselbalch Equation:

$$y = \frac{(10^{(pKa-pH)})}{(1+10^{(pKa-pH)})} . \tag{9}$$

The data were also fit to a form of the Henderson-Hasselbalch Equation, which

includes a term for the cooperativity of proton binding/release, the Hill

Coefficient, according to Markley(*63*):

$$y = \frac{(10^{n(pKa-pH)})}{(1+10^{n(pKa-pH)})} . \tag{10}$$

Where *n* is the Hill Coefficient. If *n*=1, we can assume no cooperativity between the

different ionizable sites in a solution. When *n*>1, positive cooperativity is indicated and

eq 9 is no longer a valid model for fitting the data. When *n*<1, negative cooperativity is

indicated and again eq 9 is not an appropriate representation of the data. Since each of

the peptides analyzed has only one ionizable group present, *n* should be one for every

analysis.

Thermal Unfolding of Proteins by CD: General protocols for thermal unfolding

studies of proteins have been discussed.(*64, 65*) The thermal unfolding of all protein

variants was followed by circular dichroism spectrophotometry (CD) using either an

Aviv 62DS or 202SF spectrometer (Aviv Instruments, Lakewood, NJ). The unfolding

was followed at 234 nm for RNase Sa and mutants and at 244 nm for RNase T1 and

mutants with ~0.1 mg/ml protein solutions in a 1 cm cuvette. The buffers used were the

same as previously described.(*44*) CD measurements for all variants were made at one-

degree intervals over ranges varying from 2° C to 85° C, at heating rates of 5-10 deg.

C/hr with an equilibration time between measurements of 3 min, a 1 nm bandwidth, and

an averaging time of 30 sec per measurement. An example thermal unfolding curve is

shown in Figure 5. Each resulting curve was fit with the following equation.(*44*)
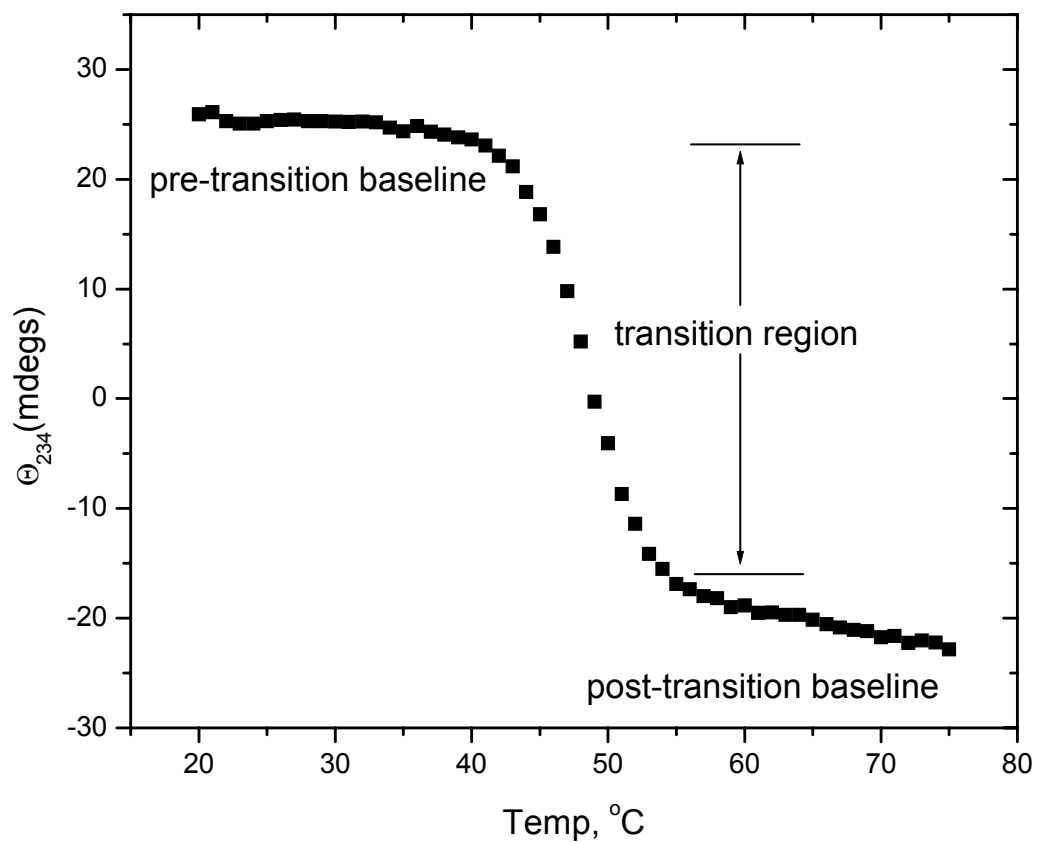
Figure 5: Thermal unfolding curve for RNase Sa monitored by circular dichroism at 234 nm. The transition region of unfolding, along with the pre- and post-transition baselines are shown. The protein concentration was ~0.1mg/ml in 30 mM MOPS, pH 7 in a 1 cm cuvette.

$$y = \frac{(y_f + m_f \cdot T) + (y_u + m_u \cdot T) \cdot \exp(-\Delta H_m \cdot \dfrac{\dfrac{1}{T + 273.15} - \dfrac{1}{T_m + 273.15}}{R})}{1 + \exp(-\Delta H_m \cdot \dfrac{\dfrac{1}{T + 273.15} - \dfrac{1}{T_m + 273.15}}{R})}. \quad (11)$$

Where y_f and y_u are the y-intercept of the pre- and post-transition baselines, and m_f and

m_u are the slopes of the pre- and post-transition baselines of the unfolding curve,

respectively as shown in Figure 5. T is the temperature at each point of measurement,

T_m is the midpoint of the thermally induced unfolding of the protein, also referred to as

the Tmelt and ΔH_m is the van't Hoff enthalpy of unfolding at the T_m. Reversibility of

thermal unfolding was routinely checked by comparing the CD signal of the sample

before heating and again after heating and recooling. Reversibility was judged to be

greater than 95% for all protein variants under the conditions used.

Urea Unfolding of Proteins: General protocols for urea unfolding studies of proteins

have been discussed.(*64, 65*) Urea unfolding studies of RNase Sa have been described

previously, and were performed at 25° C using ~0.1 mg/ml protein solution in 30 mM

MOPS pH 7 in a 1 cm cuvette.(*44*) Unfolding was monitored by CD at 234 nm using

one of the instruments listed above.

Urea unfolding studies of RNase T1 have been described previously.(*56*) They were

performed at 15° C and monitored using ~0.01 mg/ml protein solutions in 30 mM MOPS

pH 7 in a 1 cm cuvette using a SLM 8100 fluorescence spectrometer from SLM

Instruments (Urbana, IL). Excitation was at 278 nm and emission was followed at 320

nm by photon counting with a 60 second averaging time.

Urea unfolding data were plotted as either CD signal or Intrinsic Fluorescence as a function of urea concentration. An example RNase T1 urea unfolding curve is shown in Figure 6 showing the intrinsic fluorescence as a function of urea molarity. Each curve was fit with the following equation:

$$y = \frac{(y_f + m_f \cdot [D]) + (y_u + m_u \cdot [D]) \cdot \exp(m \cdot (\frac{[D] - C_{mid}}{RT}))}{1 + \exp(m \cdot (\frac{[D] - C_{mid}}{RT}))} . \tag{12}$$

Where y_f, m_f, y_u, and m_u are the same variables as in eq 11. R is the Gas Law Constant, T is the temperature of analysis and [D] is the denaturant concentration. m, referred to as the m-value, is the dependence of ΔG on the concentration of denaturant in the transition region of the curve and C_{mid} is the midpoint of the transition region.

Tanford-Wyman Analysis: Tanford applied Wyman's Linked Functions Theory for Substrate-Ligand Binding to the specific case of protein-proton binding.(*3, 66*) From this application he derived the following relationship:

$$K = K_0 (\frac{\prod_{i=1}^{n}(1 + \frac{K_{a,i,D}}{a_{H^+}})}{\prod_{i=1}^{n}(1 + \frac{K_{a,i,N}}{a_{H^+}})}) . \tag{13}$$

Where K is the equilibrium constant for unfolding, K_0 is the equilibrium constant for unfolding at some arbitrary reference point usually taken at $a_H^+ = \infty$, *i.e.*, in the fully protonated state of the protein, a_H^+ is the activity of the proton, and $K_{a,i,D}$, and $K_{a,i,N}$ are the acid dissociation constants for group *i* in the unfolded state and folded state of the protein, respectively.
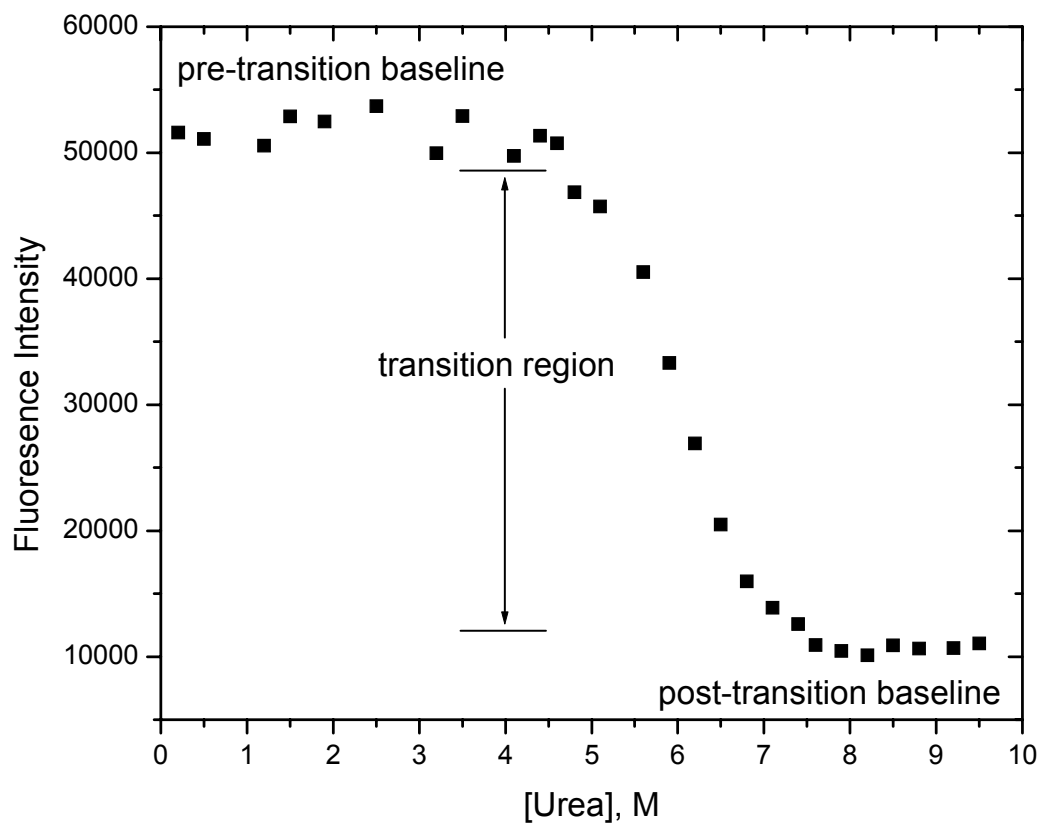
Figure 6: Urea unfolding curve for RNase T1 showing fluorescence intensity as a function of urea concentration. The transition region of unfolding, along with the pre- and post-transition baselines are shown. Each data point represents a separate solution incubated overnight at and measured at 15 °C. Protein concentration was ~0.01 mg/ml in 30 mM MOPS, pH 7 in a 1 cm cuvette. Excitation was at 278 nm and fluorescence emission was measured at 320 nm with 4 nm slitwidths. The fluorescence intensity at each datum point was corrected for the contribution to fluorescence from urea.

Tanford went on to say that if we allow the activity of the proton to equal the proton concentration, which under most mild aqueous conditions is a good approximation, we can allow the proton concentration to be approximated by the pH. We can then convert K to the conformational stability of the protein, ΔG, and eq 13 can be rearranged to:

$$\Delta G([H^+]) = \Delta G_{[H^+] \to \infty} + RT \cdot \ln \sum_{i=1}^{n} \frac{1 + \dfrac{K_{i,N}}{[H^+]}}{1 + \dfrac{K_{i,D}}{[H^+]}}. \tag{14}$$

Where $\Delta G([H^+])$ is the conformational stability of the protein as a function of proton concentration (pH), $\Delta G_{[H^+] \to \infty}$ is the conformational stability of the protein as $[H^+]$ approaches infinity, R is the Gas Law Constant, T is the temperature in Kelvin, and $K_{i,N}$ and $K_{i,D}$ are the acid dissociation constants of group i in the folded and unfolded state of the protein, respectively. From eq 14 the relationship between $[H^+]$ and $\Delta G([H^+])$ shows that if we assume all $K_{i,D}$ for each ionizable group type is the same, the only variable in eq 14 is $K_{i,N}$. In other words, the shape of the curve defined by the dependence of ΔG on pH, is dependent solely on the acid ionization constants, pKas, of the ionizable groups in the folded state of the protein.

If we construct a mutant, via site directed mutagenesis, of your favorite protein, YFP, in which an ionizable group of interest, IGI, is mutated into a group that does not ionize, e.g., an Asp to an Asn, eq 14 for the new protein variant will take the form:

$$\Delta G([H^+]) = \Delta G_{[H^+] \to \infty} + RT \cdot \ln \sum_{i=1}^{n-1} \frac{1 + \dfrac{K_{i,N}}{[H^+]}}{1 + \dfrac{K_{i,D}}{[H^+]}}. \tag{15}$$

Where the summation from i=1 to n-1 subtracts from eq 14 the IGI, which is no longer

present in the new mutant of YFP.

Subtracting eq 15 from eq 14 yields:

$$\Delta\Delta G([H^+]) = \underset{[H^+]\to\infty}{\Delta\Delta G} + RT\cdot\ln(\frac{1+\frac{K_N}{[H^+]}}{1+\frac{K_D}{[H^+]}}).$$ (16)

Where $\Delta\Delta G(H^+)$ is the change in conformational stability of the mutant protein

compared to the wild type protein as a function of pH, $\Delta\Delta G_{[H^+]\to\infty}$ is the change in

conformational stability as the pH approaches zero, and K_N and K_D are the acid

dissociation constants of the ionizable group of interest as it occurs in the folded and

unfolded states of the wild type protein, respectively.

Eq 16 defines a relationship that allows one to estimate the pKa of an ionizable

group in a protein. The protocol can be summarized to the following steps: 1) construct

a mutant of YFP in which the IGI has been mutated into a group which does not bind a

proton, ΔYFP, 2) determine the conformational stability of both YFP and ΔYFP at

several pH values, and 3) determine the difference in conformational stability at each pH

between the two proteins, $\Delta\Delta G$. A plot of $\Delta\Delta G$ as a function of pH can be fit with eq

16. The resulting analysis will yield K_N and K_D, which may then be solved for $pK_{a,N}$ and

$pK_{a,D}$ by taking the negative logarithm. I refer to this analysis as the Tanford-Wyman

analysis in honor of Charles Tanford and Jeffries Wyman. An example Tanford-Wyman

analysis is shown in Figure 7 in which the pKa of Asp76 in RNase T1 is determined.
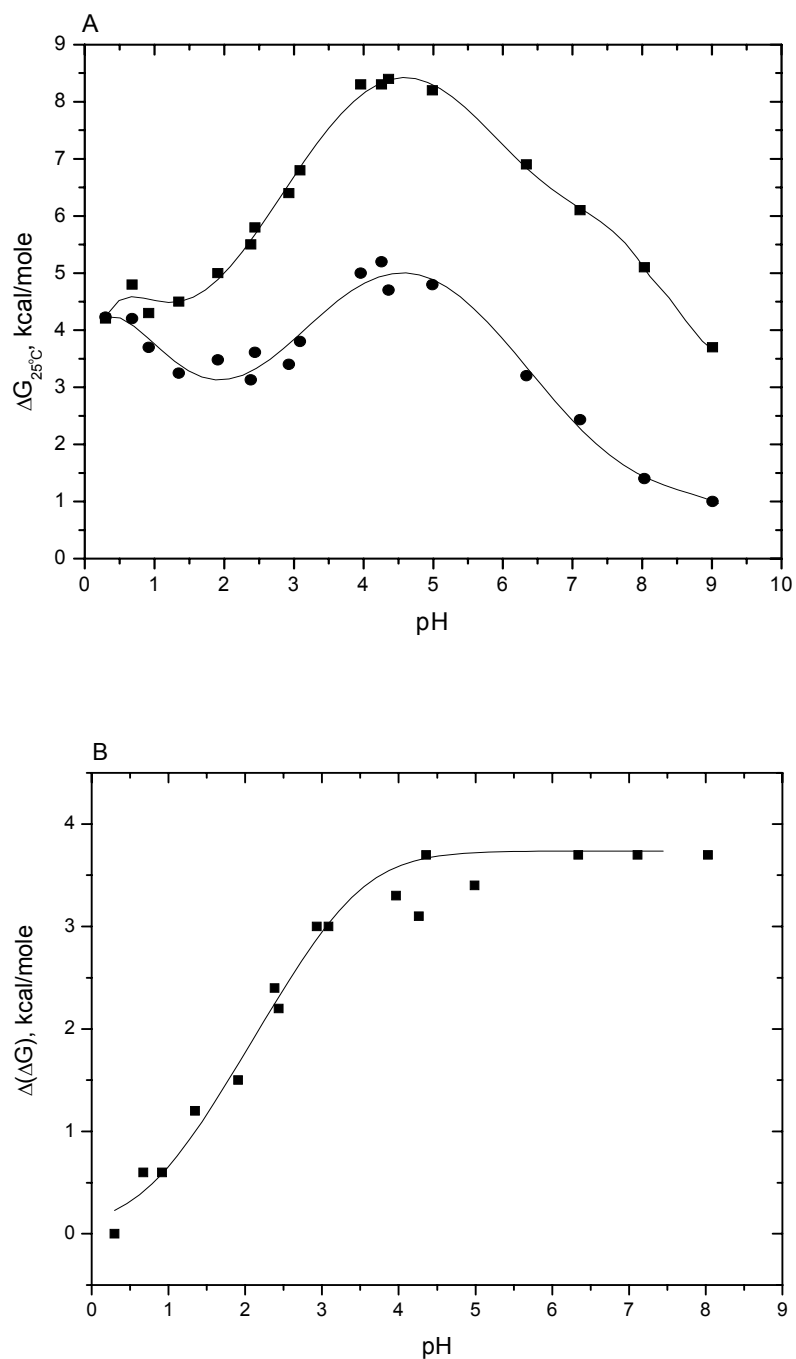
Figure 7: Tanford-Wyman analysis for Asp76 in RNase T1. A: The conformational stability at 25 °C in kcal/mole of RNase T1 wt (■) and T1 D76N (●) as a function of pH. The lines are drawn to guide the eye only. B: $\Delta(\Delta G)$ of the data from A as a function of pH. The solid line is the best fit to the data using eq 16 from pH 0.4 to 7. From this analysis: pKa (folded) = 0.7 ± 0.2 and pKa (unfolded) = 3.4 ± 0.2.

For a more thorough discussion of the assumptions in applying this technique see

Appendix 2.

The conformational stability of each protein at each pH was calculated from thermal

unfolding results using the following form of the Gibbs-Helmholtz Equation:

$$\Delta G(T) = \Delta H_m (1-\frac{T}{T_m}) - \Delta C_p \left[(T_m - T) + T \ln\left(\frac{T}{T_m}\right) \right]. \tag{17}$$

Where $\Delta G(T)$ is the conformation stability of the protein at a given temperature, T. ΔH_m

and T_m were defined by eq 11 and ΔC_p is the change in heat capacity that accompanies

protein unfolding.

Potentiometric Titrations of Proteins: The determination of proton binding curves

for proteins has been described previously.(*7, 25, 61*) The protocol used here is similar

with few exceptions. The protein samples were prepared by dissolving about 20 mg dry

protein in about 3.5 ml degassed 0.1 M KCl solution. The protein solution was dialyzed

overnight against degassed 0.1 M KCl. Following dialysis the protein stock solution was

transferred to a small volume test tube and the KCl dialysate solution was retained for

use as solvent blank. The protein concentration was determined by measuring the

absorbance at 278 nm of three different dilutions of the stock solution using either a

Hewlett-Packard 8452A or an Agilent Technologies 8453 UV/vis spectrophotometer

(Hewlett-Packard, Palo Alto, CA or Agilent Technologies, Waldbronn, Germany). The

molar extinction coefficients were 12,300 M^{-1} cm^{-1} for RNase Sa and mutants and

19,215 M^{-1} cm^{-1} for RNase T1 and mutants and appropriate corrections were made for

those mutants were a chromophore was modified.(*44, 67*) Corrections for light scattering

were applied when necessary according to Pace, et al.(*67*) The concentrations of the

three dilutions were averaged for the final concentration of the protein stock solution.

For each titration 1 ml of protein stock was added to 2 ml of the 0.1 M KCl dialysate.

For each protein variant three separate titrations were performed, and three blank

titrations were performed with the 0.1 M KCl dialysate. The blank titrations were

averaged and the average blank curve was subtracted from each protein titration to yield

three proton binding curves for each protein variant. The three proton binding curves for

each protein variant were analyzed for agreement and were averaged for a final proton

binding curve.

Potentiometric Difference Titrations: Potentiometric difference titrations have been

described elsewhere.(*25*) The technique involves: 1) determine the proton binding curve

for the variant of YFP that contains the ionizable group of interest (IGI), and 2)

determine the proton binding curve for a variant in which the IGI has been mutated to a

group that does not bind a proton, an Asp to Asn mutation for example. This mutation

creates a decrease in proton binding of one proton in the pH region corresponding to the

pKa of the IGI. Subtracting these two binding curves results in a simple one-proton

binding curve that can then be fit with eq 9, yielding the pKa of the IGI. An example of

a potentiometric difference titration is shown in Figure 8, where the pKa of Asp79 in

RNase Sa is determined. The main assumptions and limitations of this technique are

discussed in Appendix 2.

SATK: Calculations based on the Solvent Accessibility Corrected Tanford-

Kirkwood Equation were performed with TKBK.(*33*) The crystal structures used in
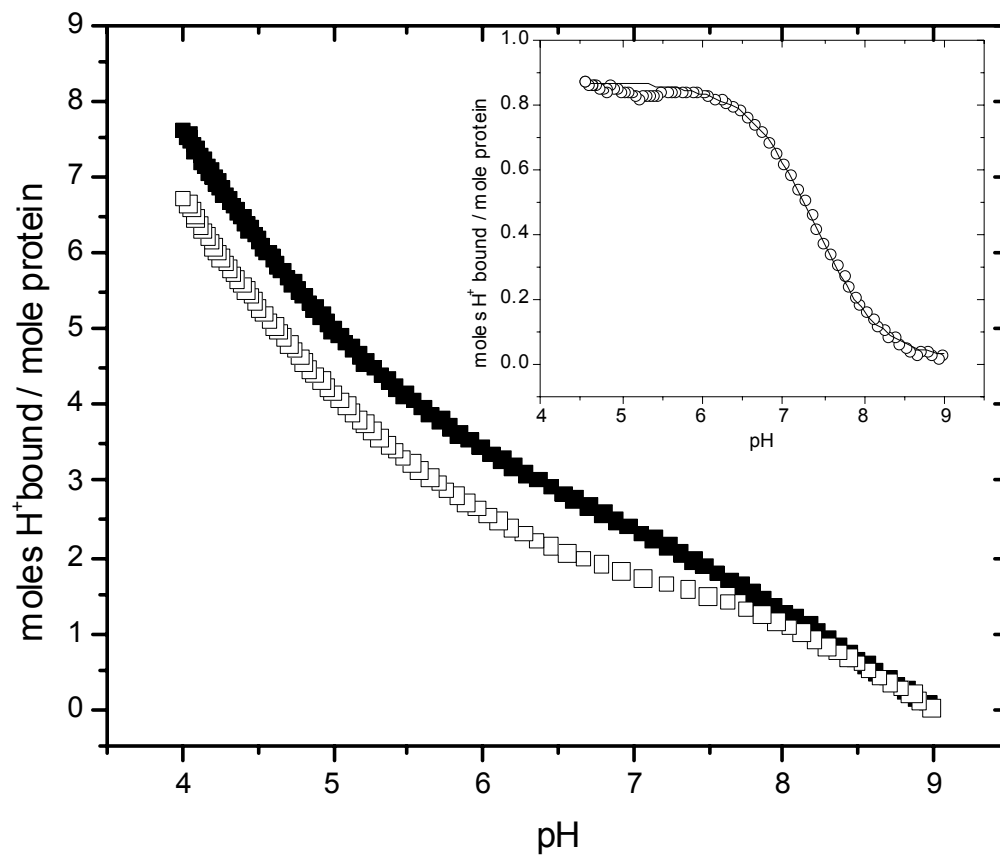
Figure 8: Potentiometric titration curves for RNase Sa (■) and RNase Sa D79N (□). Each curve is the average of three different titrations. (Inset) The difference curve (○) was determined by subtracting the curve for Sa D79N from the curve for RNase Sa. The line (—) is the non-linear least squares fit of eq 1 to the difference curve. The pKa of Asp79 in RNase Sa from this analysis is 7.37, as reported in reference 10. Analytical conditions were 100 mM KCl and 25 °C.

performing these calculations were: RNase Sa – 1RGG and RNase T1 – 9RNT from the Protein Data Bank.(*52, 54, 68*) Mutant forms of the crystal structures were designed by modeling in the mutant side chains with identical side chain conformations as the wild type residue using Swiss PDB Viewer.(*55*) The parameters used for the Accessible Surface Area calculations were as follows: van der Waals radii – those of Chothia; solvent molecule radius – 1.4 Å; number of points - 1000. For these calculations we used a protein dielectric of 4, solvent dielectric of 80.75, temperature - 293.15°C, ionic strength – 0.065 M, pH 7, and with the Gurd correction. Using a protein dielectric of 4 gave the most agreeable pKas to experimental results for all of the residues in the wild type protein, and the changes in the pKa for the mutants were no different regardless of the protein dielectric used.

FDPB: Calculations based on the finite difference solution to the linearized Poisson-Boltzmann Equation were performed with UHBD.(*39*) The crystal structures were the same as with the SATK calculations. The parameters used for the UHBD calculations were protein dielectric – 20, solvent dielectric – 80.75, temperature – 293.15, ionic strength - 65 mM, and 2 Å ion radius.

CHAPTER III

INTRINSIC pKas IN MODEL PEPTIDES

RESULTS

Figure 9 shows the potentiometric titration of Ac-AAEAA-NH$_2$, a representative titration for a group that ionizes at an acidic pH. Non-linear least squares fitting with eq 9 provides the intrinsic pKa of the γ-carboxyl group of glutamic acid, 4.24.

Figure 10 shows the potentiometric titration of Ac-AAHAA-NH$_2$, a representative titration for a group that ionizes near neutral pH. Non-linear least squares fitting with eq 9 yields the intrinsic pKa for the imidazole group of histidine, 6.54.

Figure 11 shows the potentiometric titration of Ac-AAKAA-NH$_2$, a representative titration for a group that ionizes at a basic pH. Non-linear least squares fitting with eq 9 yields the intrinsic pKa for the ε-amino group of lysine, 10.4.

Table 1 shows the intrinsic pKa of each ionizable group in the model peptides from this work. The reported results are the average of at least 3 titrations for each peptide. Each titration was fit for its pKa with both eq 9 and eq 10. The results from fits with eq 9 and eq 10 were within experimental error in every case. The Hill Coefficients, n, from fits with eq 10 were 0.9-1.1 in each case and are not reported in Table 1.
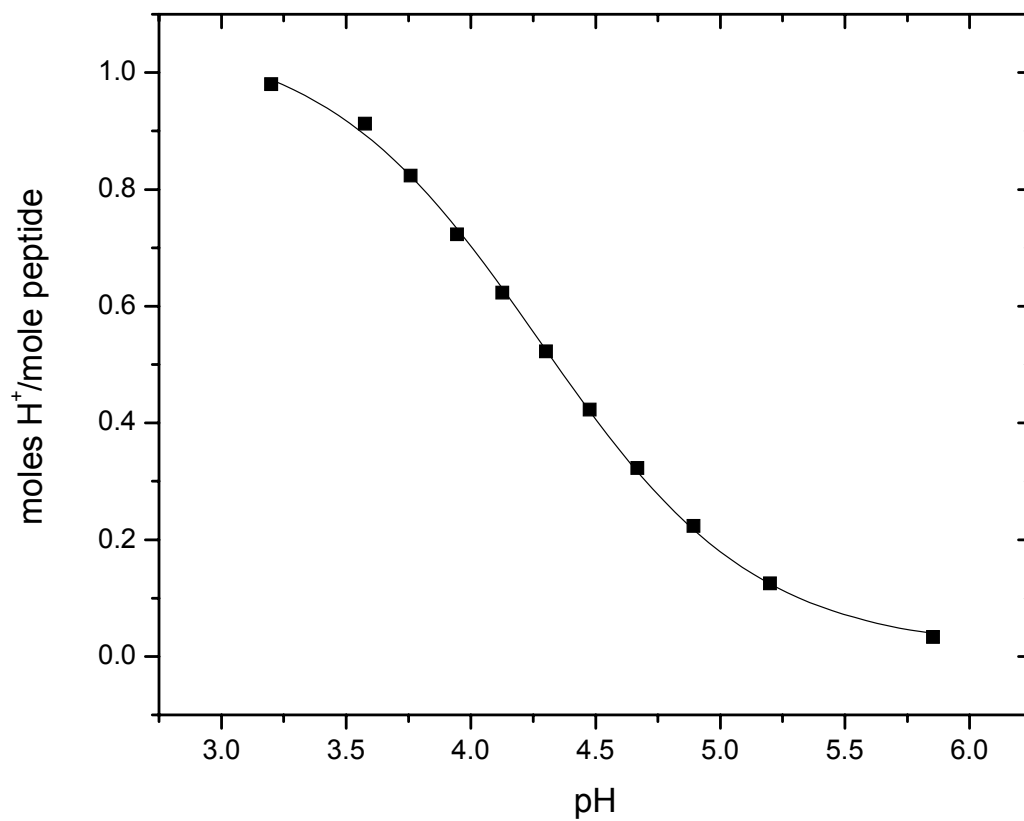
Figure 9: Potentiometric titration curve of the side chain carboxyl in Ac-AAEAA-NH$_2$. The sample was dissolved in 100 mM KCl and titrated with 0.2 M HCl in 100 mM KCl at 25 °C. The line is the best fit of the data with eq 9.
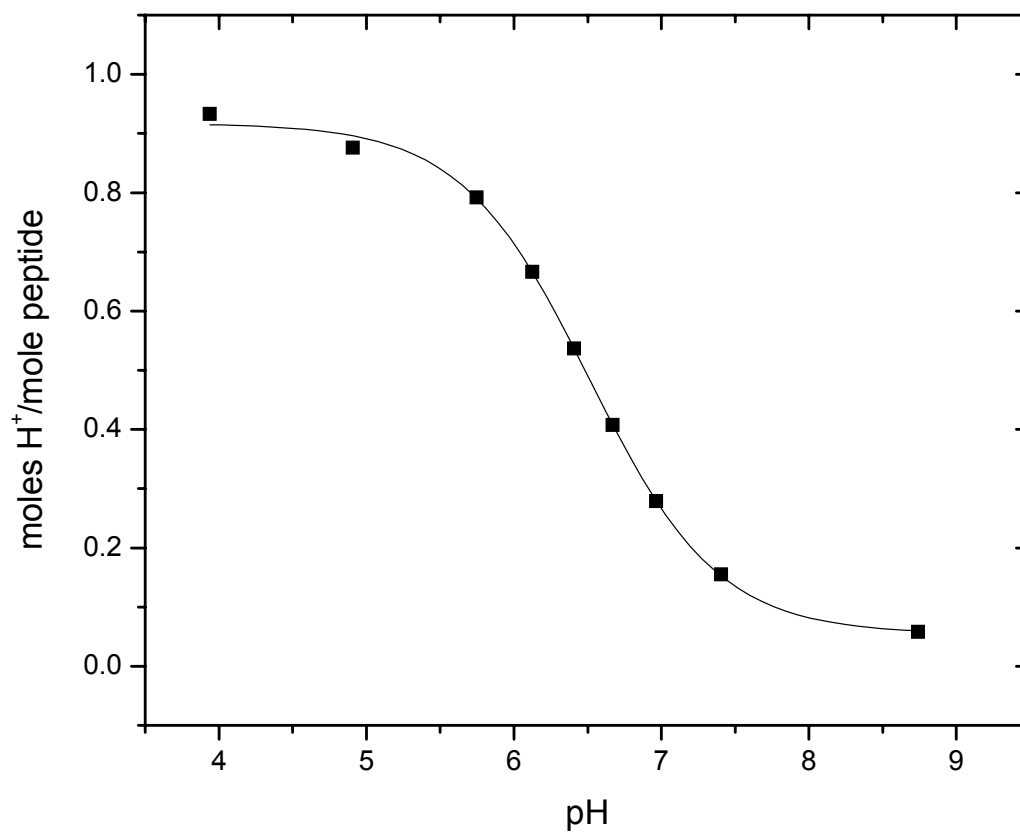
Figure 10: Potentiometric titration curve of the side chain imidazole group in Ac-AAHAA-NH$_2$. The sample was dissolved in 100 mM KCl and titrated with 0.05 M HCl in 100 mM KCl at 25 °C. The line is the best fit of the data with eq 9.
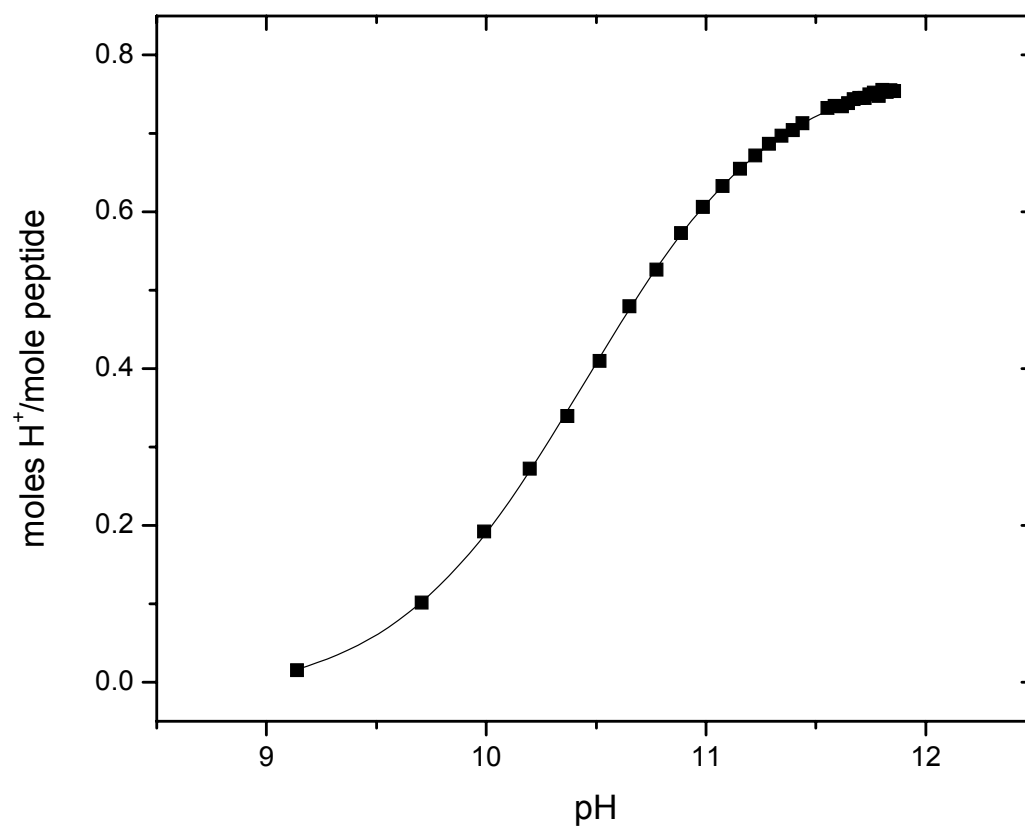
Figure 11: Potentiometric titration curve of the side chain amine in Ac-AAKAA-NH$_2$. The sample was dissolved in 100 mM KCl and titrated with 0.5 M NaOH in 100 mM KCl at 25 °C. The line is the best fit of the data with eq 9.

Table 1: pKas of the Ionizable Groups in Model Peptides

Model Peptide	pKa of Ionizable Group[a]
Ac-AAAAA-CO_2^-	3.67[b]
Ac-AADAA-NH_2	3.67[c]
Ac-AAEAA-NH_2	4.24[d]
Ac-AAHAA-NH_2	6.54[d]
NH_3^+-AAAAA-NH_2	7.97[d]
Ac-AACAA-NH_2	8.51[c]
Ac-AAYAA-NH_2	9.83[c]
Ac-AAKAA-NH_2	10.40[b]

[a] Experimental conditions were 25° C and 100 mM KCl. Each value is the average of three titrations. [b] Error for these values is $\pm$ 0.1. [c] Error for these values is $\pm$ 0.2. [d] Error for these values is $\pm$ 0.05. These errors are not statistical errors, each is the average of the deviations of the measured values.

DISCUSSION

The Intrinsic pKas of the Ionizable Groups of Proteins: The pKas of the ionizable groups of proteins in free amino acids and short peptides have been studied for over eighty years.(*69-73*) They are of major interest to protein chemists because of their role in protein stability, solubility and activity. The dependence of the conformational stability of a protein on pH is dependent on the difference in the pKas of the ionizable groups in the folded protein and the unfolded protein.(*3*) Proteins with higher percentages of polar or charged groups tend to be more soluble in aqueous solution than those higher in non-polar groups.(*4*) Amino acids with an ionizable side chain group are found with much more frequency as catalytic residues than other amino acids.(*5*) Accurate determinations of the pKas of the ionizable groups are very important to understanding the interactions in proteins that affect their pKas, but in order to understand those interactions we must accurately know their intrinsic pKas.

To determine the intrinsic pKas, we constructed as our model system peptides of the form Ac-AAXAA-NH$_2$, where X is an amino acid with an ionizable side chain group. Peptides to represent the N-terminal and C-terminal groups were $^+$NH$_3$-AAAAA-NH$_2$ and Ac-AAAAA-CO$_2^-$, respectively. To eliminate possible long-range electrostatic interactions with the terminal groups, we blocked the N-terminus with an acetyl group, Ac, and the C-terminus with a primary amide, NH$_2$. We have evidence that shows the N-terminus and C-terminus of an unblocked alanine penta-peptide interact and both pKas are perturbed, presumably through long-range electrostatic interactions. (unpublished observations) The choice of penta-peptides was partially one of

convenience. We wanted to place the residue with the ionizable group at the central position of a peptide, where there would be an equal number of amino acids on each side of the residue containing the ionizable group. Tri-peptides have proven somewhat problematic to purify, and we were concerned about possible secondary interactions with longer peptides.(*74*)

It can be seen in Figures 10 and 11 that the moles of protons bound per mole of peptide do not equal 1 as it does in Figure 9. This was observed for several of the model peptides. If there is only one titratable group on each peptide, we should see one mole of protons bound per mole of peptide at the endpoint of the titration. For some of the peptides, as little as 0.7-0.8 moles of protons, and as much as 1.05, mark the endpoint of the titration. These differences from the expected seem a bit high, however titrations performed with different stock solutions and on different days often yielded different proton uptakes but the same results for the pKa and the Hill Coefficient. Therefore we shall attribute these observations to typical lab errors.

The cysteine model peptide deserves discussion because of the possibility of disulfide bond formation prior to or during the titration of the peptide. Titrations of sulfhydryl containing compounds using a very similar procedure to that described here have been performed previously with good rigor.(*75*) The only added precaution to the procedure for this work was to dissolve the peptide in acidified, degassed 0.1M KCl for the stock solution and perform the titration with NaOH. The results obtained here also agree exceptionally well with results from the spectrophotometric analysis of a cysteine containing extended coil peptide in which EDTA was used to minimize oxidation.(*76*)

pKas Determined using Model Systems: Table 2 reports the results of previous work

on the determination of the pKas of ionizable groups in model compounds and peptides.

The ionic strength of solution for some of the results in Table 2 differ markedly from

other results, so in order to make a true comparison between the different works,

corrections for the ionic strength of solution should be made. For a neutral acid group,

whose ionization is described by eq 1, the group's pKa as a function of ionic strength

can be described from the Debye-Huckel Theory as

$$pKa' = pKa - 0.5\sqrt{\mu} \tag{18}$$

where *pKa'* is the measured pKa of the group at the given ionic strength, *pKa* is the pKa

of the group in dilute salt and μ is the ionic strength of the solution.(*77*) For a cationic

acid, whose ionization is described by eq 2, the group's pKa as a function of ionic

strength is

$$pKa' = pKa + 0.5\sqrt{\mu} \ . \tag{19}$$

For purposes of clarity we will define "dilute salt" as the concentration of salt as it

approaches infinite dilution. However, it should be noted that at ionic strengths below

about 10 mM, corrections to infinite dilution do not make significant differences in the

pKa.

The initial work to standardize the intrinsic pKas of the ionizable groups was

performed by Tanford.(*6*) Tanford based his work on model compound data, and where

needed, made corrections to the original results to account for solution ionic strength and

electrostatic interactions with other groups present in the compound. Nozaki and

Table 2: pKas of the Ionizable Groups Previously Reported

Ionizable Group	Nozaki-Tanford[a]	Keim, etal[b]	Richarz and Wüthrich[c]	Bundi and Wüthrich[d]	Corrected for Salt[e]
α-COOH	3.8	3.28	-	-	3.83
Asp-COOH	4	3.91	3.9	4.1	3.83
Glu-COOH	4.4	4.14	4.2	4.5	4.4
His-Im	6.3	6.76	7	6.8	6.38
α-NH$_3$	7.5	8.14	-	-	7.81
Cys-SH	9.5	-	-	-	8.67
Tyr-OH	9.6	10.0	10.2	10.5	9.99
Lys-NH$_3$	10.4	10.47	10.84	10.9	10.24
Arg-Guanidino	12	-	-	-	-

[a] Based on model compound data and corrected for ionic strength to dilute salt, from reference 7 as noted in the text. [b] Based on ^{13}C NMR studies using model peptides of the form Gly-Gly-X-Gly-Gly from references 77-80 as noted in the text. [c] Based on ^{13}C NMR studies using model peptides of the form Gly-Gly-X-Ala, from reference 81 noted in the text. d Based on 1H NMR studies using model peptides of the form Gly-Gly-X-Ala and, where needed, corrected to dilute salt, from references 82 and 83 as noted in the text. [e] This research corrected to dilute salt as stated in the text.

Tanford later modified those early intrinsic pKa values and those reported in Table 2 are

the values usually referred to when we talk about intrinsic pKas.(*7*)

A comparison of our results with the Nozaki/Tanford results in Table 2 shows that

they are in generally good agreement. The pKa value of the β-carboxyl of the aspartic

acid side chain reported here is ~0.2 pH units lower than that of the Nozaki/Tanford

value. This difference is within experimental error and may not be significant. The α-

amino group pKa reported here is ~0.3 pH units higher than the Nozaki/Tanford value.

An examination of the compounds Tanford used for his estimates shows that the pKas

ranged from 6.8 to 7.9, with most of the values 7.6-7.9. If he had limited the compounds

to those in the latter group the average would be 7.76, in good agreement with the value

reported here. The pKa of the side chain sulfhydryl of cysteine reported here is ~0.8 pH

units lower than the Nozaki/Tanford value. The difference here should be attributed to

the lack of suitable model compound data available to Tanford, as he only had two

compounds on which to base his estimate. The pKa of the phenolic OH of tyrosine

reported here is ~0.4 units higher than that of Nozaki/Tanford. The difference here can

also be attributed to a lack of suitable model compound data available to Tanford. One

the three compounds Tanford used for his model compounds is a questionable choice,

polytyrosine. The result used is the observed average pKa extrapolated to zero net

charge of the polymer, and its value is the lowest of the three, 9.5. The average of the

other two compounds, 9.75, is in better agreement with our result reported in Table 2.

The pKa of the ε-amino group of lysine is only slightly lower than the Nozaki/Tanford

value, and we attribute the difference here again to the lack of suitable model

compounds available to Tanford. We did not attempt to study an arginine-centered

peptide primarily because of the high pKa estimated for the guanidino group.

Keim, et al performed a series of NMR studies on glycine based pentapeptides, Gly-

Gly-X-Gly-Gly.(*78-81*) The focus of their research was more to determine and follow

^{13}C chemical shifts as a function of pH for the side chain carbons of the amino acid at

the central position of the peptide, as the residue may appear in an extended random coil,

rather than to actually determine pKas of the ionizable groups. The model peptides they

used were not blocked on the N- and C-termini thus charged groups were still present on

each peptide. The value reported in Table 2 for the C-terminus is the average of all the

C-terminal groups reported in their work. The comparatively low value for the pKa of

the C-terminus is probably mainly due to the presence of the positive charge on the N-

terminus. Likewise, the presence of the negative charge on the C-terminus is probably

perturbing the N-terminus, as its pKa, also the average of all reported values, is elevated.

As stated earlier, long-range electrostatics have been shown to influence measured pKas.

The other pKas reported from the work of Keim, et al are all slightly different from the

Nozaki/Tanford values, except for the imidazole group of histidine, which is ~0.5 units

higher. In their peptide, perhaps the N-terminus is influencing the pKa more strongly

than the C-terminus. The slight differences in the pKas of the other groups from the

Nozaki/Tanford values are probably due to electrostatic influences from the termini.

Richarz and Wüthrich performed a study similar to Keim, et al using a different

model peptide, Gly-Gly-X-Ala, following ^{13}C chemical shifts as function of pH.(*82*)

Likewise, Bundi and Wüthrich used the same model peptide, Gly-Gly-X-Ala, to follow

[1]H chemical shift as a function of pH.(*83, 84*) As with Keim, et al the peptides used by the Wüthrich group were not blocked on the N- and C-termini and the main focus was to determine the chemical shift patterns of the different amino acid side chain atoms as a function of pH in an extended random coil peptide. The results reported from both of these studies are somewhat higher than those of Nozaki/Tanford, especially the imidazole side chain of histidine. The perturbation observed here is probably because the X amino acid is located next to the C-terminal residue. This places the ionizable group in a position where the electrostatic interaction with the C-terminus is stronger than its electrostatic interaction with the N-terminus. The fact that the results are also more perturbed than those of Keim, et al supports this statement.

Intrinsic pKas in Continuum Electrostatic Theory: Calculations based on variations of the Poisson-Boltzman equation and the Solvent Accessible Tanford Kirkwood Equation use intrinsic pKas as first approximations for the pKas of the ionizable groups as they occur in a protein. The results reported here for the intrinsic pKas of the ionizable groups of proteins based on model peptides are similar to the pKas from model compound data proposed by Nozaki and Tanford, however there are some significant differences.

In Table 3, we summarize our results for the intrinsic pKas for the ionizable groups of proteins. We believe that this new set of intrinsic pKas are the best values for researchers to use, especially for first estimates for the pKas of the groups as they appear in proteins.

Ionizable Group	pKa
α-COOH	3.8
Asp-COOH	3.8
Glu-COOH	4.4
His-Im	6.4
α-NH$_3$	7.8
Cys-SH	8.7
Tyr-OH	10.0
Lys-NH$_3$	10.2
Arg-Guanidino	12

Table 3: Intrinsic pKas for the Ionizable Groups of Proteins.

CHAPTER IV

THE EFFECTS OF HYDROGEN BONDING ON THE pKa OF THE SIDE CHAIN CARBOXYL OF ASP33 IN RNASE Sa

RESULTS

Figure 5, Chapter II, shows a typical thermal unfolding curve for RNase Sa as monitored by CD spectrophotometry. This curve shows the typical change in CD signal as RNase Sa unfolds with increasing temperature. Thermal unfolding curves can be converted to the fraction of folded protein as a function of temperature, f_f.(64) Thermal unfolding curves showing f_f as a function of temperature for wild type Sa and the mutants studied here are shown in Figure 12. Curves of this type were fit with eq 11, and the resulting thermodynamic unfolding parameters from these fits are reported in Table 4. Thermal unfolding results from this work are in good agreement with previous results.($44, 45$)

Figure 13 shows a typical urea induced unfolding curve of RNase Sa as monitored by CD. This curve shows how the CD signal of RNase Sa changes with increasing urea concentration. Urea unfolding curves can be converted to the fraction of folded protein as a function of urea concentration, f_f.(64) Urea unfolding curves showing f_f as a function of urea molarity for the RNase Sa variants studied here are shown in Figure 14. Curves of this type were fit with eq 12, and the results are reported in Table 5. Urea unfolding results reported here are in good agreement with results reported previously.(44)
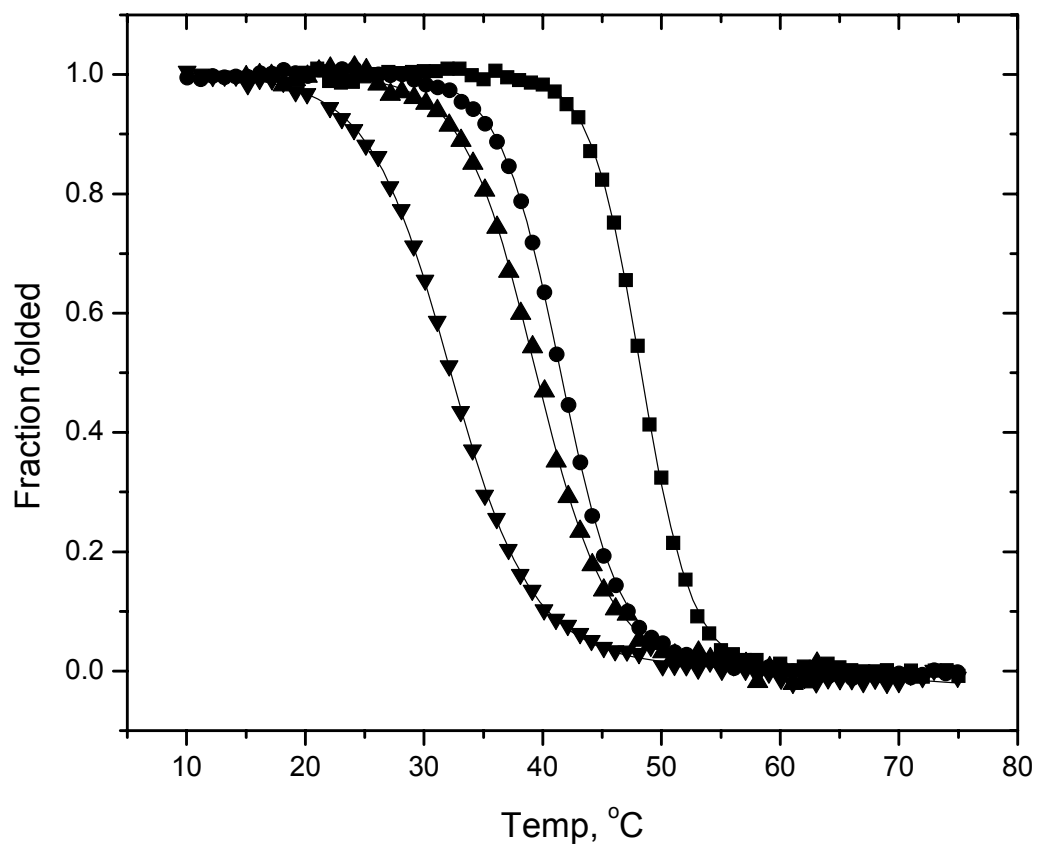
Figure 12: Thermal unfolding curves for RNase Sa wt (■), Sa T56V (●), Sa T56A (▲) and Sa D33N (▼). Performed in 30 mM MOPS, pH 7 showing the fraction of folded protein as a function of temperature. Each line is the best fit of the respective data with eq 11. The results from these analyses are reported in Table 4.
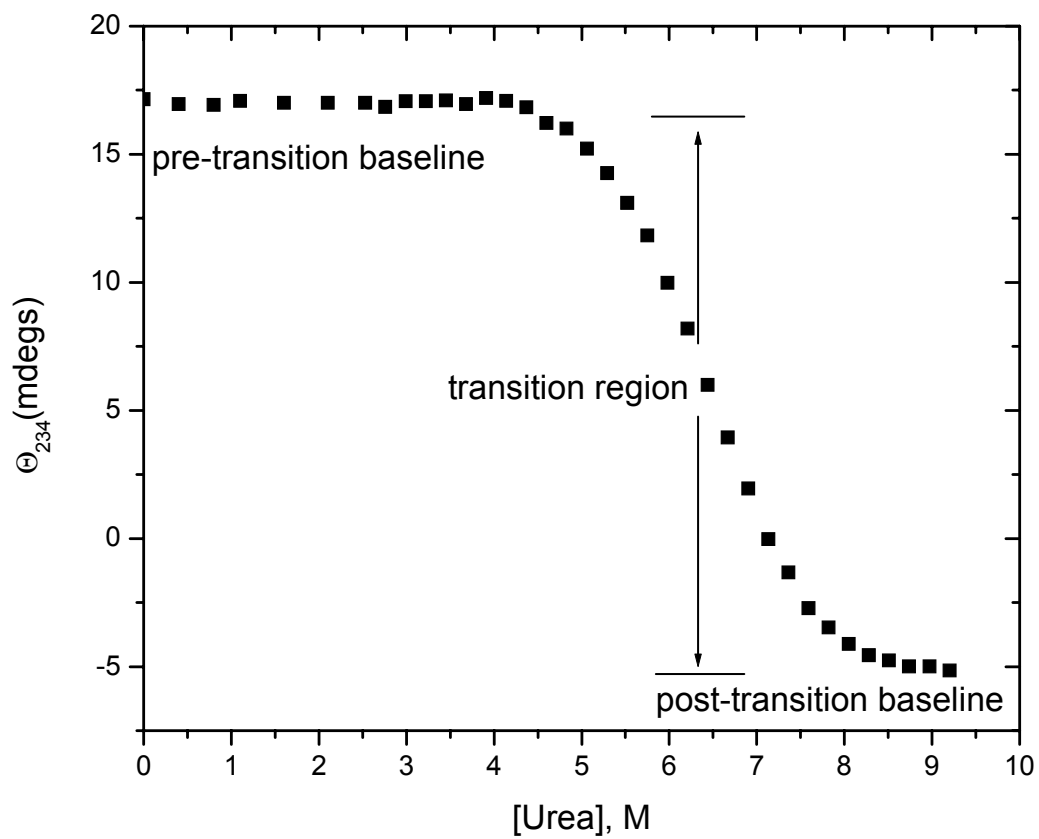
Figure 13: Urea unfolding curve of RNase Sa monitored by circular dichroism. The transition region of unfolding, along with the pre- and post-transition baselines are shown. The protein concentration was ~0.1 mg/ml in 30 mM MOPS, pH 7 and 25 °C in a 1 cm cuvette.
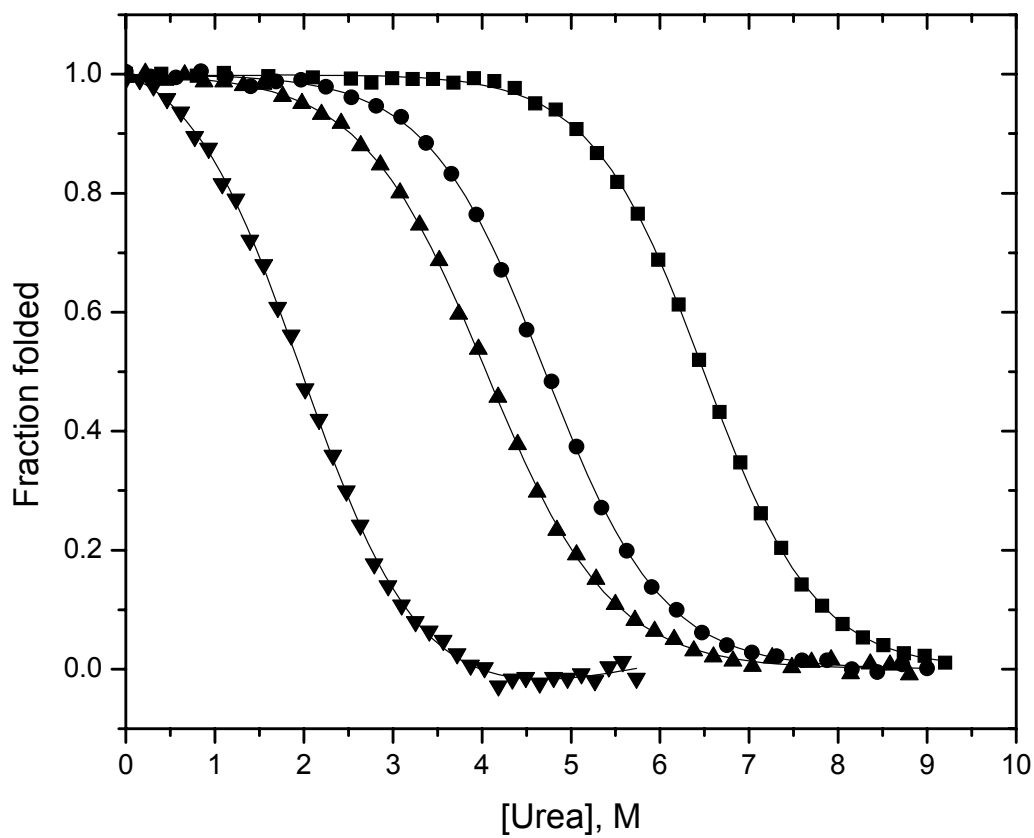
Figure 14: Urea unfolding curves for RNase Sa wt (■), Sa T56V (●), Sa T56A (▲) and Sa D33N (▼). Performed in 30 mM MOPS, pH 7 and 25 °C showing the fraction of folded protein as a function of urea concentration. Each line is the best fit of the respective data with eq 12. The results of these analyses are reported in Table 5.

Table 4: Parameters Characterizing the Thermal Unfolding of RNase Sa and Mutants.[a]

RNase Sa	ΔH_m, kcal/mole	ΔS_m,[b] cal/mole·K	T_m, °C	$\Delta G_{25°C}$,[c] kcal/mole	$\Delta(\Delta G)$,[d] kcal/mole
wild-type	96.7	301	48.2	5.7	-
T56V	83.1	264	41.7	3.7	-2.0
T56A	68.3	218	39.7	2.7	-3.0
D33N	61	199	33.3	1.5	-4.2
D33A	56.5	185	31.8	1.1	-4.6

[a] ΔH_m and T_m were obtained by nonlinear least squares analysis of thermal unfolding curves with eq 11. Protein concentrations were 0.05-0.2 mg/ml in 30 mM MOPS pH 7. Errors are ± 10% for ΔH_m and ± 0.2 for T_m. [b] $\Delta S_m = \Delta H_m/T_m$. [c] $\Delta G_{25°C}$ was obtained from eq 17 with $\Delta C_p = 1.52$ kcal/mole·K. Errors are ± 0.5 kcal/mole. [d] $\Delta(\Delta G) = \Delta G_{25°C (mut)} - \Delta G_{25°C (wt)}$

Table 5: Parameters Characterizing the Urea Unfolding of RNase Sa and Mutants Monitored by Circular Dichroism.[a]

RNase Sa	C_{mid}, M	m, kcal/mol·M	ΔG_{H_2O},[b] kcal/mole	$\Delta(\Delta G)$,[c] kcal/mole
wild-type	6.50	0.95	6.2	-
T56V	4.72	0.90	4.3	-1.7
T56A	4.04	0.86	3.5	-2.3
D33N	2.02	0.88	1.8	-4.3
D33A	1.67	0.77	1.3	-4.6

[a] C_{mid} and m were obtained by nonlinear least squares analysis of urea unfolding curves with eq 12. Protein concentrations were 0.05-0.2 mg/ml in 30 mM MOPS pH 7 at 25°C. Errors for Cmid are ± 0.1, and for m ± 0.05. [b] $\Delta G_{H_2O} = C_{mid} \cdot m$. [c] $\Delta(\Delta G) = (C_{mid (mut)} - C_{mid (wt)}) \cdot m_{(wt)}$.

The pKa of the side chain carboxyl of Asp33 in wild type RNase Sa was determined by the Tanford-Wyman approach for pKa determination. Figure 15A shows the conformational stability as a function of pH for RNase Sa wild type and RNase Sa D33N. The data from A were used to determine the change in conformational stability, $\Delta(\Delta G)$, between the two forms of the protein as a function of pH, shown in Figure 15B. The data shown in Figure 15B were fit with eq 16 and the pKa of the Asp33 carboxyl in the folded, 2.6, and in the unfolded protein, 3.9, was obtained. The results from this analysis are in very good agreement with previous results determined by ^{13}C NMR for the pKa of the Asp33 carboxyl in folded wild type RNase Sa, 2.39.(*10*)

Potentiometric difference titrations were used to determine the pKa of the side chain carboxyl of Asp33 in RNase Sa T56A and Sa T56V. The proton binding curves for Sa T56A and Sa D33N are shown in Figure 16. Also shown in Figure 16 is the difference curve determined by subtracting the Sa D33N curve from the curve for Sa T56A. The difference curve was fit with eq 9 to determine the pKa of the Asp33 carboxyl in Sa T56A, 3.9. The proton binding curves for Sa T56V and Sa D33N along with the difference curve are shown in Figure 17. This difference curve was fit with eq 9 to determine the pKa of the Asp33 carboxyl in Sa T56V, 4.4.

DISCUSSION

In the folded structure of RNase Sa the side chain carboxyl of Asp33 is buried, has no short-range electrostatic interactions but forms three good intramolecular hydrogen bonds. Figure 3A, Chapter I, shows the ribbon diagram of the crystal structure of
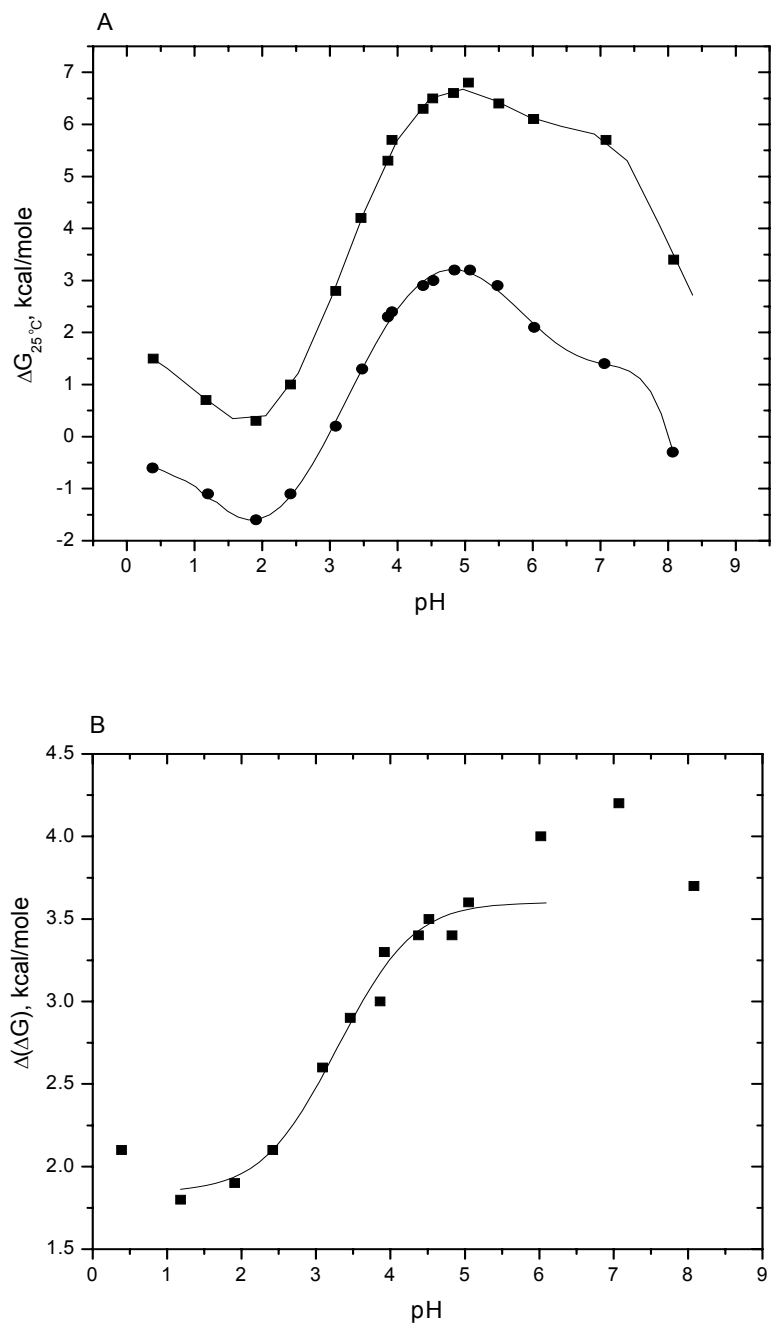
Figure 15: Tanford-Wyman analysis for Asp 33 in RNase Sa. A: The conformational stability at 25 °C in kcal/mole of RNase Sa wt (■) and Sa D33N (●) as a function of pH. The lines are drawn to guide the eye only. B: $\Delta(\Delta G)$ of the data from A as a function of pH (■). The solid line is the best fit to the data using eq 16 from pH 1to 6. From this analysis: pKa (folded) = 2.6 ± 0.2 and pKa (unfolded) = 3.9 ± 0.2.
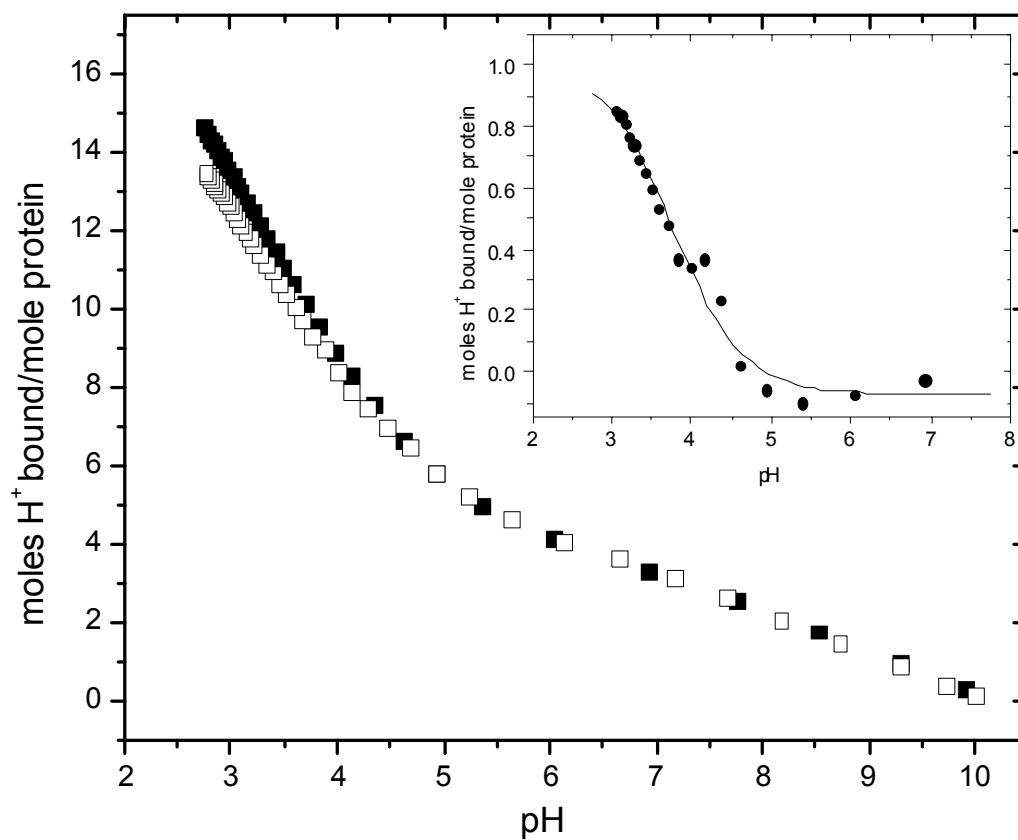
Figure 16: Potentiometric titration curves for RNase Sa T56A (■) and Sa D33N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of about one proton with a corresponding pKa of 3.9 ± 0.2. Experimental conditions were 200 mM KCl and 15 °C, and the difference curve was obtained by a simple subtraction of the curve for D33N from the curve for T56A.
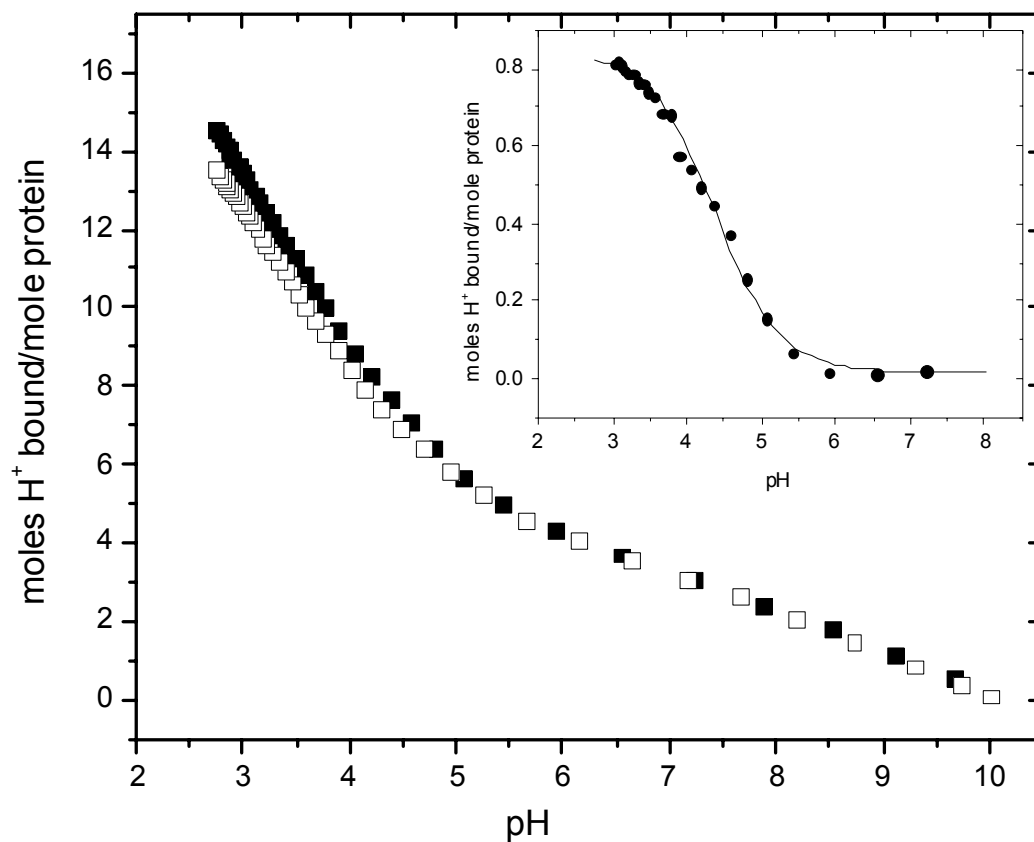
Figure 17: Potentiometric titration curves for RNase Sa T56V (■) and Sa D33N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of about one proton with a corresponding pKa of 4.4 ± 0.2. Experimental conditions were 200 mM KCl and 15 °C, and the difference curve was obtained as described in Figure 16.

RNase Sa with the location of Asp33 relative to the rest of the protein. Figure 3B shows an expanded view of the region around the side chain of Asp33. We see the h-bond contacts that the carboxyl group makes; 2.6 Å to Thr56 Oγ1, 2.8 Å to Thr56 backbone NH, and 3.3 Å to Tyr30 backbone NH. The carboxyl group of Asp33 is about 95% buried and each of the groups it forms an h-bond to is also completely buried. The only h-bond that can be readily eliminated by mutation is the one to the Thr56 OH group. The objective was to mutate Thr56 to a residue whose side chain will not form an h-bond with the side chain carboxyl of Asp33.

Choices of Mutations and Changes in Conformational Stability: The best choice for replacing a threonine residue whose side chain hydroxyl is h-bonded is probably valine. The conformation of the valine side chain, shown in Figure 18, is very similar to the side chain of threonine. The volume of the valine side chain (~87 Å^3) that replaces the threonine side chain (~68 Å^3) should fill the space nicely, whereas an alanine side chain might introduce a cavity of about 32 Å^3.(*85*) From Figure 18 we see that Val is more favorable than Ala for buried sites due to its higher hydrophobicity, but the difference in side chain conformational entropy (TΔS) favors Ala.(*86-88*) The net effect is that Ala should be a better mutation for Thr than Val by about 0.5 kcal/mole. Previous work however, on Thr to Val and Thr to Ala mutants in which the Thr side chain forms at least one h-bond shows that mutations to Val are typically less destabilizing than Ala mutants.(*89-92*) These observed differences between Val and Ala mutants could be due to more buried hydrophobic surface area and better van der Waals interactions of the Val side chain. The increased volume of the valine side chain could introduce steric strain to

	Volume (Å³)	Hydrophobicity (kcal/mole)	Side Chain Entropy (TΔS) (kcal/mole)
Asp → Asn	+8	-0.23	-0.25
Asp → Ala	-26	-1.47	0.78
Thr → Val	+18.5	-1.31	0.65
Thr → Ala	-32	-0.07	1.08

Figure 18: Differences in structure, volume, hydrophobicity, and side chain conformational entropy (TΔS) for the following mutations: Asp to Asn, Asp to Ala, Thr to Val, and Thr to Ala. The differences in volume are from Tsai et al.(*85*) The differences in hydrophobicity are from Pace and are based on the *n*-octanol hydrophobicity scale of Fauchere and Pliska .(*86, 87*) The differences in conformational entropy are based on the mean TΔS values at 300 K as found in Doig and Sternberg.(*88*)

the protein but this seems to be overcome as evidenced by the higher stability for the Val

mutants over the Ala mutants. Thanks to previous work in the lab, we already had

available the plasmids for Sa T56V and Sa T56A. With both of these variants available,

we decided to determine the impact of each mutation on the stability of the protein and

the pKa of the Asp33 carboxyl in the respective mutant.

The side chain OH of Thr56 is 100% buried and forms three intramolecular h-bonds,

to the Oδ1 of Asp 33, and 2.84 Å and 3.01 Å to the Nϵ and Nη1 of Arg 65, respectively.

From Table 4 and Table 5, the change in conformational stability, $\Delta(\Delta G)$, for Sa T56V is

about -1.9 kcal/mole, and the $\Delta(\Delta G)$ for Sa T56A is about -2.7 kcal/mole. The

difference in stability between the Val and Ala mutants is consistent with previously

observed results.(*89-92*) The loss of stability is also relatively consistent with the

estimate that each h-bond is worth between 1-1.5 kcal/mole to the conformational

stability of the protein.(*50, 93*)

In order to determine the pKa of the Asp33 carboxyl using the methods of choice

here it was necessary to construct a mutant in which the aspartic acid side chain is

mutated to a side chain that does not bind a proton. From Figure 18 we see that

asparagine and aspartic acid have very similar side chain conformations and size, and the

hydrophobicity and side chain conformational entropies are also very similar. An Asn

side chain may also be able to reform some of the h-bonds found in the wild type

protein. From Table 4 and Table 5, it can be seen that $\Delta(\Delta G)$ for Sa D33N is about -4.3

kcal/mole and $\Delta(\Delta G)$ for Sa D33A is about -4.5 kcal/mole. This seems to indicate that

none of the h-bonds found in the wild type protein are formed in the folded D33N

protein. The differences in hydrophobicity and TΔS, from Figure 18, suggest that in the absence of hydrogen bonding in the Asn mutant, the Ala mutant should be more stable than the Asn mutant. For the purposes of further discussion, we will assume the h-bonds are not present in D33N since the stabilities of D33N and D33A are within experimental error.

One other point on stability we will address is why the Asp33 mutants are so much more destabilizing than the Thr56 mutants. Both side chain groups form three intramolecular h-bonds in the wild type protein, yet the Thr56 mutants are ~2 kcal/mole more stable than the Asp33 mutants. One possibility is that Asp33 is located in a site more highly specific for an aspartic acid residue, than Thr56 is for a threonine. A sequence alignment of several members of the guanyl specific microbial ribonuclease family shows that both sites are conserved among most of the family members with the exception being RNase Sa. In the alignment the position corresponding to Asp33 in Sa has a tyrosine in almost all other RNases, and the residue corresponding to Thr56 in Sa has a proline in the other RNases. This indicates that both residues are specific for RNase Sa, but does not address why Asp33 has a greater effect on the stability.(*94*) Another possibility is that one or more of the h-bonds observed in the crystal structure for the Thr56 side chain may not be present in solution. Previous work on the solution structure and dynamics of RNase Sa shows that the h-bond between Thr56 Oγ1 and the Nε of Arg65 is not observed in solution.(*95*) This missing h-bond could help to explain the large difference in stability between the Thr56 mutants and the Asp33 mutants.

Main Forces Affecting Perturbed pKas: The main forces perturbing the pKas of ionizable groups are electrostatic interactions with other ionizable groups, burial in a hydrophobic environment and hydrogen bonding. There has been considerable work investigating the effects of charge-charge interactions and of burial in a hydrophobic environment on the pKas of ionizable groups.(*10, 14, 15, 24-27, 35*) These two forces are the most studied of the three and are surely major forces affecting the pKas of perturbed ionizable groups. There has been less work on the effects of hydrogen-bonding, but it also seems to be a primary force perturbing the pKas of some ionizable groups.(*28, 29, 42, 56, 96, 97*)

Previous evidence suggests that burial of a carboxyl group in a hydrophobic environment can elevate the pKa of that group by 3 units or more.(*10, 26, 27*) The results for the pKa of the buried, hydrogen bonded carboxyl of Asp33 in wild type RNase Sa shows that the three h-bonds formed by the side chain carboxyl perturb the pKa to an acidic value. In the absence of other interacting groups in the vicinity, this shows that the h-bonds introduce a perturbation counter to and larger than the hydrophobic burial. The question we consider now is how much one h-bond contributes to the total perturbation.

pKa of the Side Chain Carboxyl of Asp33 in Sa T56A: From Figure 16, the pKa of the side chain carboxyl of Asp33 in Sa T56A is 3.9. This indicates that removal of one hydrogen bond is sufficient to return the pKa of the group to near its intrinsic value. This implies that the two remaining h-bonds effectively perturb the pKa just enough to cancel the perturbation due to burial in a region of low dielectric constant such as a

hydrophobic environment. The proton binding curves shown in Figure 16 are approaching the acid pH limit of the ability to titrate the proteins with good confidence. This introduces greater uncertainty in the analysis. As a result, we attempted to apply the Tanford-Wyman approach to determine the pKa of the Asp33 carboxyl in the T56A mutant. Figure 19A shows the conformational stability, ΔG, as a function of pH for Sa T56A and Sa D33N. Figure 19B shows the $\Delta(\Delta G)$ as a function of pH, which shows that the difference in stability between the two mutant forms of the protein is consistent throughout the pH range studied. One of the suppositions in applying the Tanford-Wyman analysis is that there should be a significant difference in the pKa of the Asp carboxyl group in the folded and in the unfolded protein. In the event that the pKa in the folded and the unfolded T56A protein are similar, the $\Delta(\Delta G)$ between T56A and D33N will show no dependence on pH, as is shown in Figure 19B. If we assume that the pKa of the Asp33 carboxyl in unfolded T56A is nearly the same as in unfolded wild type Sa, see Figure 15, then we can estimate from the current discussion and Figure 19 that the pKa of the Asp33 carboxyl in Sa T56A is about 3.9. This supports the result by potentiometric difference titration that the pKa of the Asp33 carboxyl is 3.9.

pKa of the Side Chain Carboxyl of Asp33 in Sa T56V: From Figure 17, the pKa of the side chain carboxyl of Asp33 in Sa T56V is 4.4. As in the previous discussion, this result shows that the two remaining h-bonds to the carboxyl group perturb the pKa just enough to return it to near its intrinsic value. In an effort to confirm this value, we attempted the Tanford-Wyman analysis on Sa T56V for the Asp33 carboxyl. The results for the conformational stability as a function of pH for Sa T56V and Sa D33N are shown
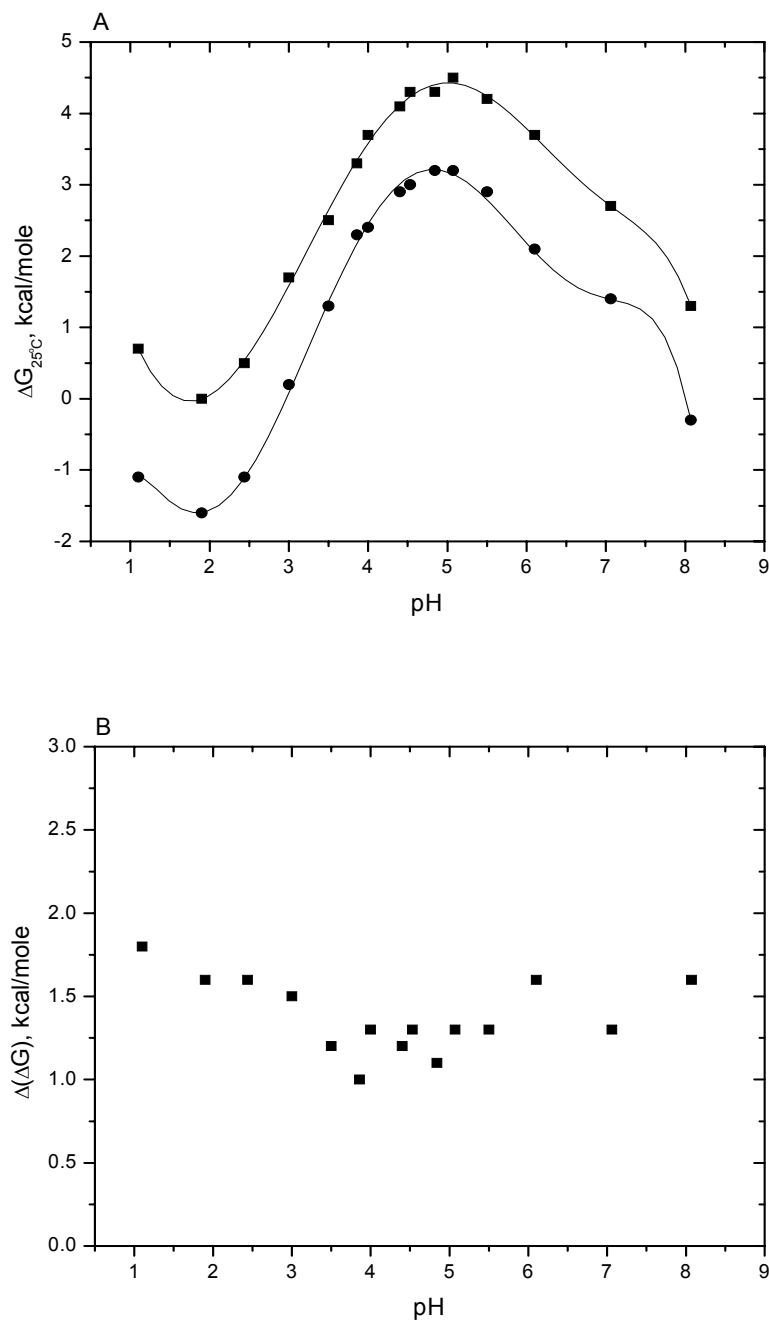
Figure 19: Tanford-Wyman analysis for Asp33 in RNase Sa T56A. A: The conformational stability at 25° C in kcal/mole of RNase Sa T56A (■) and Sa D33N(●) as a function of pH. The lines are intended to guide the eye only. B: Δ(ΔG) of the data from A as a function of pH (■). Δ(ΔG) between the two mutants is relatively constant throughout the pH range studied. This indicates that the pKa of the side chain carboxyl of Asp33 in the folded T56A protein is very similar to the pKa in the unfolded protein.
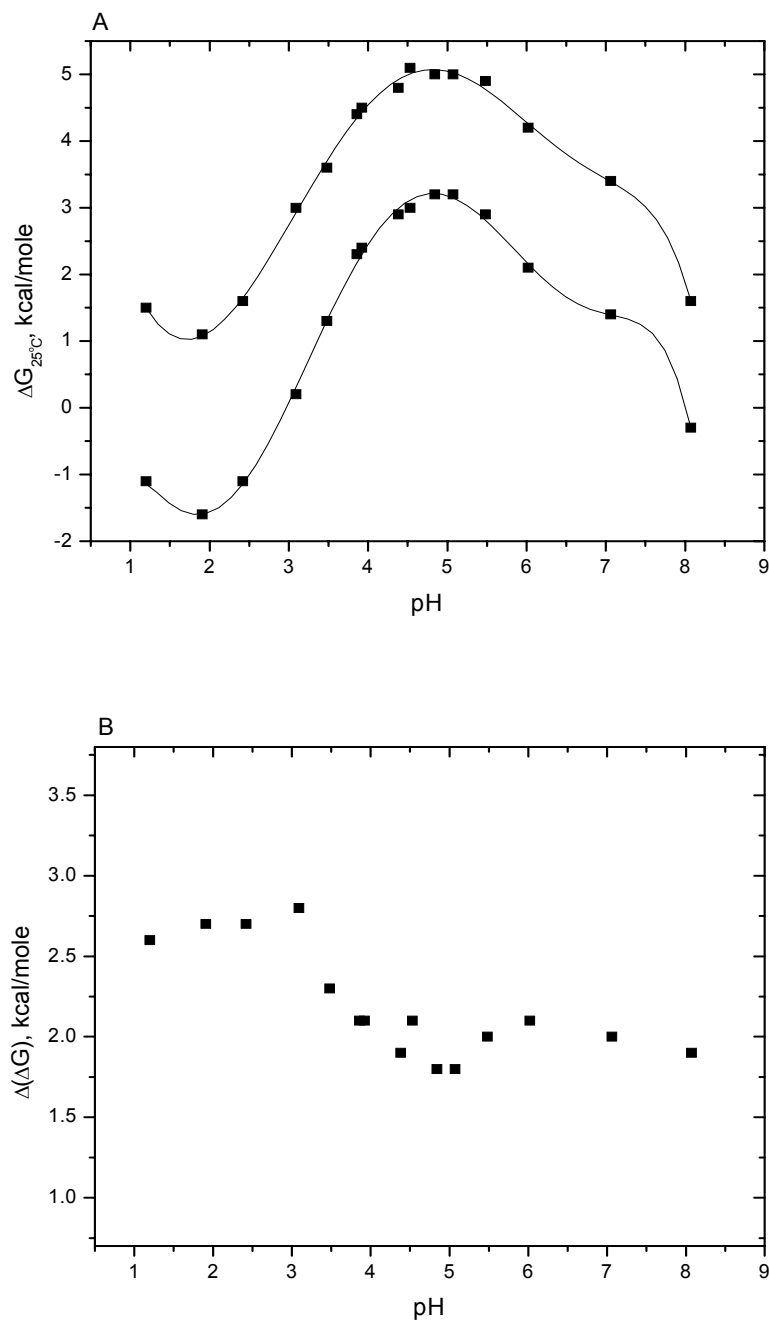
Figure 20: Tanford-Wyman analysis for Asp33 in RNase Sa T56V. A: The conformational stability at 25° C in kcal/mole of Sa T56V (■) and Sa D33N (●) as a function of pH. The lines are to guide the eye only. B: $\Delta(\Delta G)$ of the data in A as a function of pH (■). $\Delta(\Delta G)$ between the two mutants is slightly higher at acidic pHs than at neutral pHs. This suggests that the pKa of the side chain carboxyl of Asp33 in the folded T56V protein is higher than in the unfolded protein, however errors from the stability measurements prevent fitting with eq 16.

in Figure 20A. Figure 20B shows the $\Delta(\Delta G)$ as a function of pH. This plot shows that as the pH increases $\Delta(\Delta G)$ decreases, which is opposite of that observed for the residue in wild type Sa, shown in Figure 15B. Attempted fits of the data in Figure 20B suggest that the pKa of the Asp33 carboxyl in Sa T56V in the folded protein is higher than in the unfolded protein by about 0.4 units. However, assigning definite pKas from the fits is difficult because the decrease in $\Delta(\Delta G)$ over the pH range studied is within experimental error for the stability measurements. If we assume the pKa of the Asp33 carboxyl group is 3.9 in the unfolded protein, as above with the T56A mutant, then the pKa in the folded protein should be about 4.3, which is in excellent agreement with the potentiometric difference titration.

Estimates of the pKa using SATK: Estimates for the pKa of the Asp33 carboxyl in RNase Sa and the mutants studied here are reported in Table 6 along with the experimental results. As can be seen, SATK does not predict any changes in the pKa as a result of mutating Thr56. This was expected since the SATK equations were developed to explain charge-charge electrostatic interactions in proteins and not hydrogen bonding or hydrophobic burial. Some studies have demonstrated reasonably good success using SATK predictions compared with experimental results.(*35, 36, 98, 99*) Each of these studies however, was dealing with charge-charge interactions, and in most cases with groups and interactions near the surface of the protein.

Estimates of the pKa using UHBD: The Finite-Difference Poisson Boltzmann equation, FDPB, as employed in UHBD has been used previously to compare predicted pKas of ionizable groups in proteins to experimental results with reasonably good

Table 6: pKa of the Side Chain Carboxyl of Asp33 in RNase Sa and Mutants.

RNase Sa	pKa	SATK[c]	UHBD[c]
wild-type	2.6[a]	3.9	2.9
T56A	3.9[b]	3.9	2.8
T56V	4.4[b]	3.9	2.8

[a] From Tanford-Wyman analysis only. [b] From Potentiometric Difference Titrations only. [c] Performed as described in Materials and Methods.

Table 7: The Potentials of Interaction and Their Effects on the pKa of the Side Chain Carboxyl of Asp33 in RNase Sa Calculated with FDPB as Employed by UHBD.[a]

Protein Variant	ΔG (Born), kcal/mole	ΔpKa (Born)	ΔG (Bkgd), kcal/mole	ΔpKa (Bkgd)	ΔG (ij), kcal/mole	ΔpKa (ij)	ΔG (Total), kcal/mole	ΔpKa (Total)
Sa wild type	1.54	1.15	-1.46	-1.09	-1.61	-1.2	-1.53	-1.1
Sa T56A	1.53	1.14	-1.53	-1.14	-1.58	-1.18	-1.58	-1.2
Sa T56V	1.54	1.15	-1.54	-1.15	-1.62	-1.21	-1.62	-1.2

[a] The energy values reported here are from the application of FDPB to the crystal structure if RNase Sa, 1RGG. The values are reported as they are described in eq 8. The ΔpKas are determined from

$\Delta(\Delta G) = 1.34 \cdot \Delta pKa.$

success.(*10, 35, 100, 101*) The UHBD predictions from Table 6 show that the FDPB equation seems to detect the presence of hydrogen bonding to the carboxyl group in wild type Sa. The predicted pKa is only 0.3 units different from experimental. However, the predictions for the two h-bond mutants, T56A and T56V, suggest that the hydrogen bonding group is still present. A breakdown of the potentials calculated from the program is given in Table 7. From these potentials we see that the ΔG_{Born} and the ΔG_{Bkgd} terms practically cancel each other for each of the variants, and the ΔG_{IJ} term determines the total potential energy. Initially this was a bit surprising, however previous work with Coulomb's Law calculations showed that the electrostatic interactions on the Asp33 carboxyl group are large and negative, suggesting that FDPB is correctly predicting the effects of the electrostatic interactions in the protein.(*10*)

The ΔG_{Born} term seems to be underestimated for the mutant proteins. I would expect that by changing the local environment of the Asp, the ΔG_{Born} would increase due to the increased hydrophobic environment, which would decrease the local dielectric around the carboxyl group. It probably does not increase because of the way in which UHBD performs the calculations. The program establishes a position in the core of the protein where the dielectric, ε_p, equals a pre-set value, in this case $\varepsilon_p = 20$. The program then generates a grid with gradually increasing dielectrics to the surface of the protein, where the value equals that of the solvent. Asp33 is completely removed from bulk solvent but, it is not in the core of the protein, thus the desolvation penalty on the carboxyl group is not as large as if it were buried in the core. The approach used by UHBD may not be the best approach, and some have argued for and used a quantum mechanical/molecular

mechanical hybrid approach to estimate the micro-dielectric environment around groups in proteins.(*102*) Others have used a semimacroscopic approach which assigns the interior of the protein a set ε_p but treats the polar groups explicitly thus creating a "floating dielectric" for the different regions of the protein.(*103*) Either of these approaches is probably more appropriate than a dielectric grid.

The ΔG_{Bkgd} term seems to be underestimated for the wild type protein. The fact that the ΔG_{Bkgd} term is more favorable in the mutants, where it should be less favorable due to the loss of a hydrogen-bonding partner, makes the calculations questionable. The program we use to predict h-bonds, *pfis*, predicts that the Thr56 hydroxyl serves as an h-bond donor to the Asp33 carboxyl group.(*17*) The presence of the h-bond was also predicted by *Chimera*, a molecular modeling system from UCSF, the NCRR and NIH.(*104*) In order to perform its calculations, UHBD uses CHARMM to add hydrogens to the protein's crystal structure.(*105*) However, in the crystal structure for Sa, 1RGG, from the PDB, hydrogens are already present. Figure 21 shows an expanded view of the area around Asp33 in wild type Sa using the PDB file, 1RGG molecule B, overlaid with the crystal structure after CHARMM adds the hydrogen to the γ-oxygen of Thr56. As can be seen, in its original location the hydrogen is in an optimal position to form an h-bond with a Thr56Oγ1-H-Asp33Oδ1 angle of ~155°. After manipulation with CHARMM the hydrogen has been shifted and at ~73° the angle is no longer optimal. Therefore the ΔG_{Bkgd} term is probably underestimated for wild type Sa because there is no h-bond present to calculate. In fact, there is probably a penalty due to an unfavorable interaction with the partial negative charge from the γ-oxygen of the Thr56 side chain
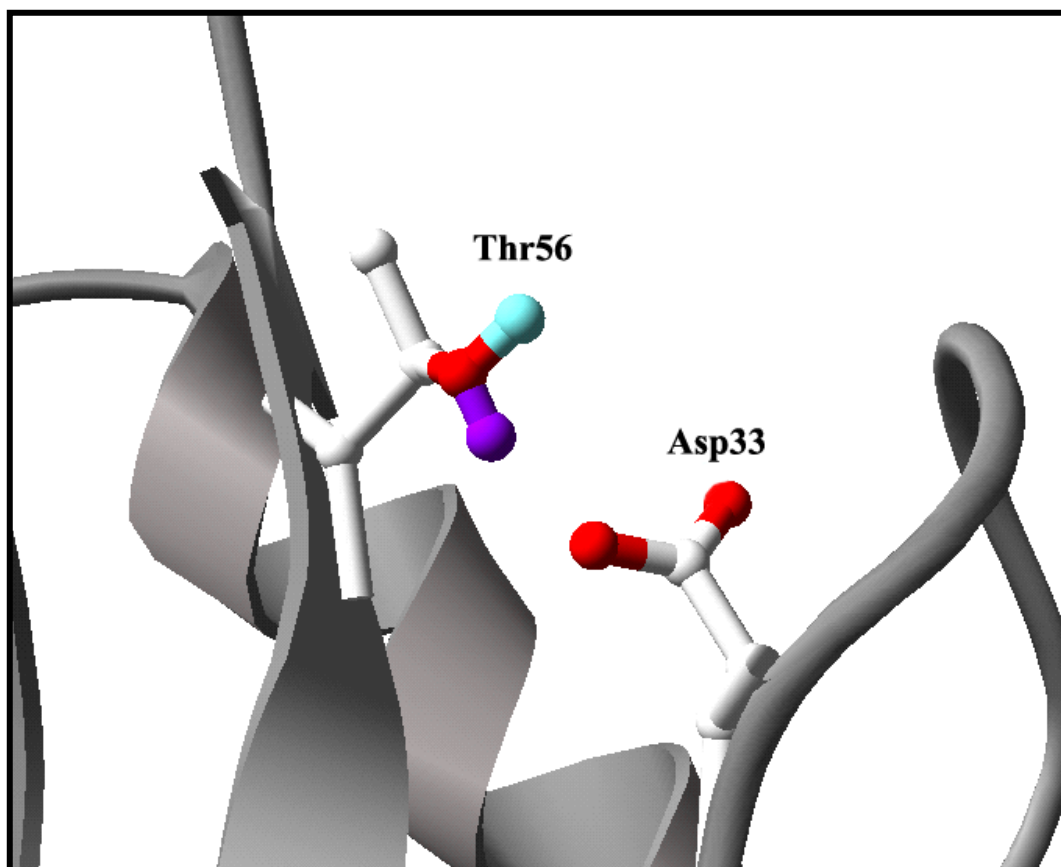
Figure 21: The crystal structure of 1RGG overlaid with 1RGG after the protons have been removed and added back by CHARMM. The purple atom denotes the γ-hydroxyl hydrogen from the original PDB file. The aqua atom is the hydrogen that has been added by CHARMM. As can be seen, when CHARMM adds the hydrogen to the Thr 56 γ-oxygen it does not place it in the original position, consequently UHBD does not detect a favorable hydrogen bonding interaction.

that is relieved with the mutation to Ala or Val, thus the ΔG_{Bkgd} term is more favorable in the mutants. On method to address this problem is to perform a global hydrogen bond optimization analysis on the protein prior to analysis with FDPB, as suggested by Nielson and McCammon.(*106*) This would optimize the placement of the hydrogens and perform the necessary hydrogen bond calculations needed for FDPB.

Concluding Comments: We have shown that the h-bond between the Asp33 carboxyl and the side chain hydroxyl of Thr56 in RNase Sa perturbs the pKa of the carboxyl group down about 1.5 units. The results also show that the choice of mutants is important for studies involving pKas of hydrogen bonded carboxyls, since the pKa of the carboxyl in the Val mutant is about 0.5 units higher than in the Ala mutant. The Solvent Accessibility Corrected Tanford Kirkwood equation is not an appropriate equation to estimate pKas for studies of this nature. Its main usefulness is to describe the effects of charge-charge interactions on ionizable groups near or on the surface of the protein. The Finite-Difference Solution to the Linearized Poisson Boltzmann Equation does a better job of predicting the pKa of the Asp33 carboxyl in the wild type protein. However, its agreement with experiment is suspect because of structural disagreement involving placement of hydrogens by CHARMM that were encountered while interpreting the results from the analysis. One way that could improve our ability to predict the pKas of the ionizable groups is to use micro-dielectric constants instead of gradient dielectrics. Another possibility would be to perform a global proton placement optimization prior to performing calculations with UHBD. Ideally, applying both approaches would probably yield the most agreeable results with experiment.

CHAPTER V

THE EFFECTS OF HYDROGEN BONDING ON THE pKa OF THE SIDE CHAIN CARBOXYL OF ASP76 IN RNASE T1

RESULTS

Figure 22 shows a typical thermal unfolding curve for RNase T1 as monitored by CD spectrophotometry. This curve shows the change in CD signal as RNase T1 unfolds with increasing temperature. Thermal unfolding curves can be converted to the fraction of folded protein as a function of temperature, f_f.(*64*) Thermal unfolding curves showing f_f as a function of temperature for wild type T1 and the mutants studied here are shown in Figures 23 and 24. Curves of this type were fit with eq 11, and the resulting thermodynamic unfolding parameters from these fits are reported in Table 8. Thermal unfolding results from this work are in good agreement with previous results.(*50, 56*)

Figure 6, Chapter II, shows a typical urea induced unfolding curve for RNase T1 as monitored by intrinsic fluorescence intensity. This curve shows how the fluorescence signal of RNase T1 changes with increasing urea concentration. Urea unfolding curves can be converted to the fraction of folded protein as a function of urea concentration, f_f.(*64*) Urea unfolding curves showing f_f as a function of urea molarity for the RNase T1 mutants studied here are shown in Figures 25 and 26. Curves of this type were fit with eq 12, and the results are reported in Table 9. Urea unfolding results reported here are in good agreement with results reported previously.(*50, 56*)
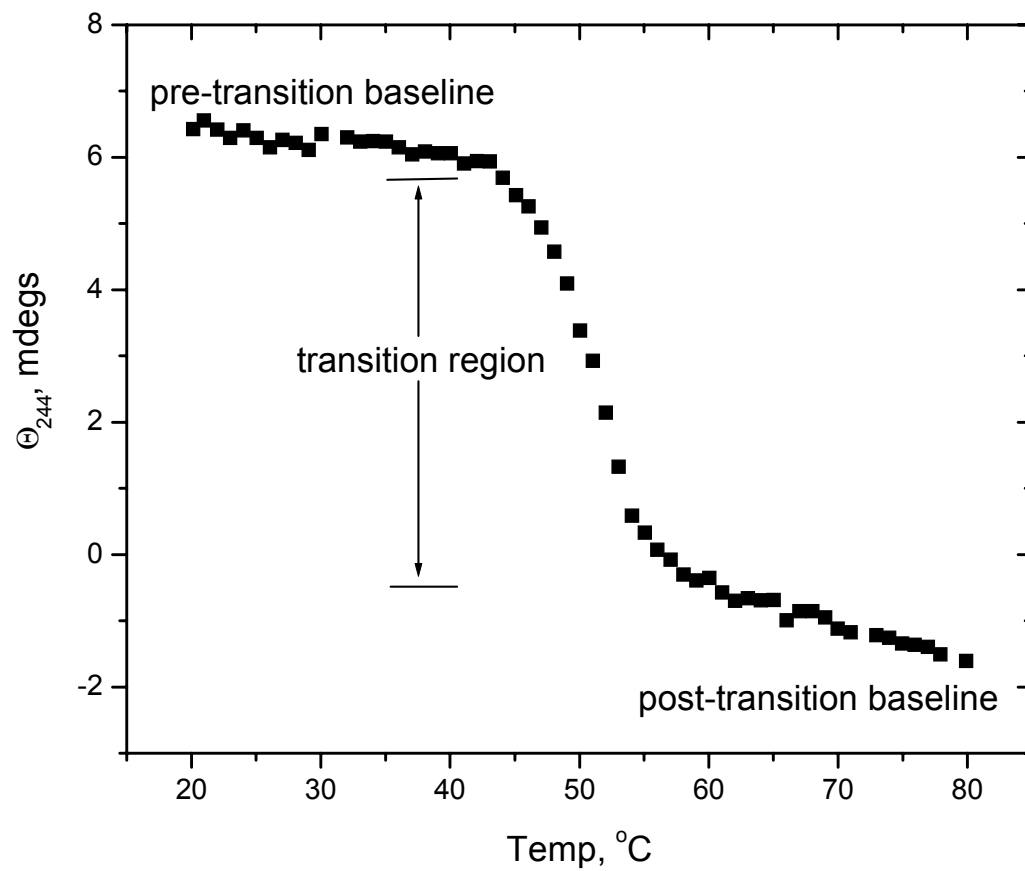
Figure 22: Thermal unfolding curve for RNase T1 monitored by circular dichroism. The transition region of unfolding is shown, along with the pre- and post-transition baselines. The protein concentration was ~0.1 mg/ml in 30 mM MOPS, pH 7 in a 1 cm cuvette.

Figure 23: Thermal unfolding curves for RNase T1 wt (■), T1 N9A (●), T1 Y11F (▲), T1 T91V (▼) and T1 D76N (♦). Performed in 30 mM MOPS, pH 7 showing the fraction of folded protein as a function of temperature. Each line is the best fit of the respective data with eq 11. The results from these analyses are reported in Table 8.

Figure 24: Thermal unfolding curves for RNase T1 N9A Y11F (■), T1 N9A T91V (●), T1 Y11F T91V (▲) and T1 Triple Mutant (▼). Performed in 30 mM MOPS, pH 7 showing the fraction of folded protein as a function of temperature. Each line is the best fit of the respective data with eq 11. The results from these analyses are reported in Table 8.

Table 8: Parameters Characterizing the Thermal Unfolding of RNase T1 and Mutants.[a]

RNase T1	ΔH_m, kcal/mole	ΔS_m,[b] cal/mole · K	T_m, °C	$\Delta G_{25°C}$,[c] kcal/mole	$\Delta(\Delta G)$,[d] kcal/mole	Expected $\Delta(\Delta G)$[e]
wild-type	98.1	0.303	50.8	6.1	-	-
N9A	99.2	0.308	48.7	5.8	-0.6	-
Y11F	99.8	0.314	44.8	5.2	-1.8	-
T91V	81.0	0.259	39.8	3.2	-3.3	-
D76N	70.9	0.228	37.5	2.4	-3.7	-
N9A Y11F	90.4	0.287	41.6	4.0	-2.8	-2.4
N9A T91V	66.9	0.218	34.4	1.8	-4.3	-3.9
Y11F T91V	56.5	0.186	30.6	1.0	-5.1	-5.1
TM	67.0	0.222	28.4	0.7	-5.4	-5.7

[a] ΔH_m and T_m were obtained by nonlinear least squares analysis of thermal unfolding curves with eq 3. Protein concentrations were 0.05-0.2 mg/ml in 30 mM MOPS pH 7. Errors are $\pm$ 10% for ΔH_m and $\pm$ 0.2 for T_m. [b] $\Delta S_m = \Delta H_m/T_m$. [c] $\Delta G_{25°C}$ was obtained from eq 9 with $\Delta C_p = 1.65$ kcal/mole · K. Errors are $\pm$ 0.5 kcal/mole. [d] For the four mutants in which ΔH_m is within 20% of wt, $\Delta(\Delta G) = (T_{m(mut)} - T_{m(wt)}) \cdot \Delta S_{m(wt)}$. For the four mutants in which ΔH_m is not within 20% of wt, $\Delta(\Delta G) = \Delta G_{25°C(mut)} - \Delta G_{25°C(wt)}$. [e] The expected $\Delta(\Delta G)$ based on the $\Delta(\Delta G)$s of the single mutants.
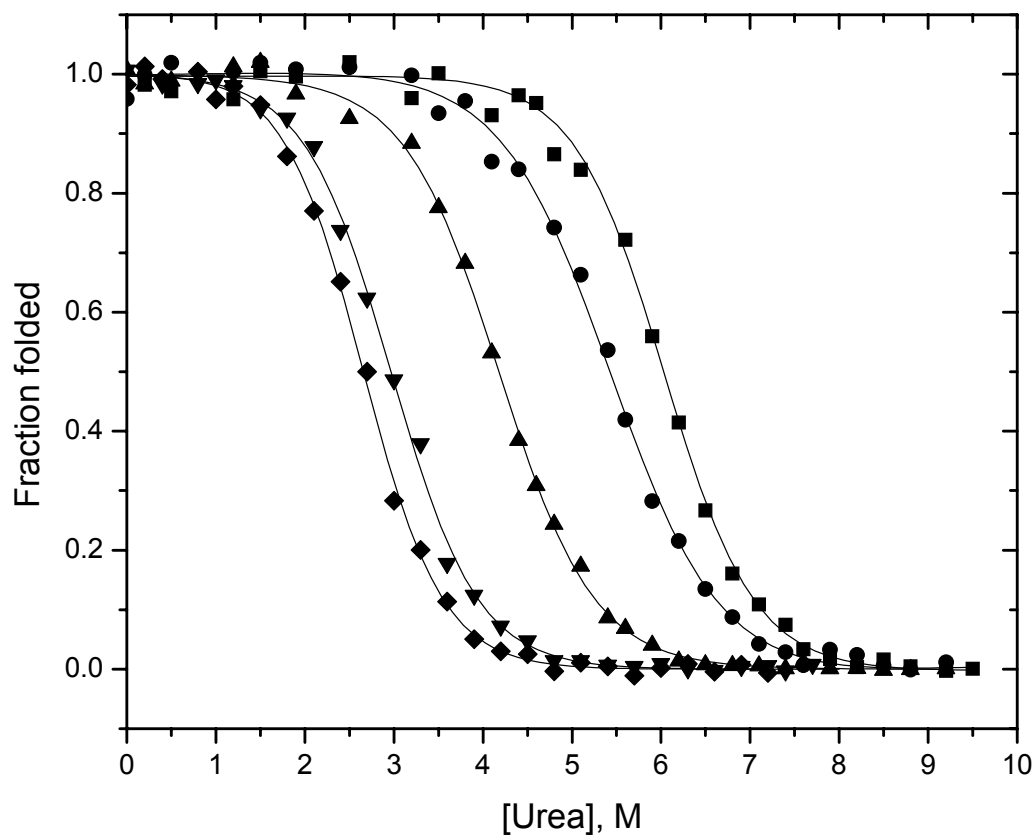
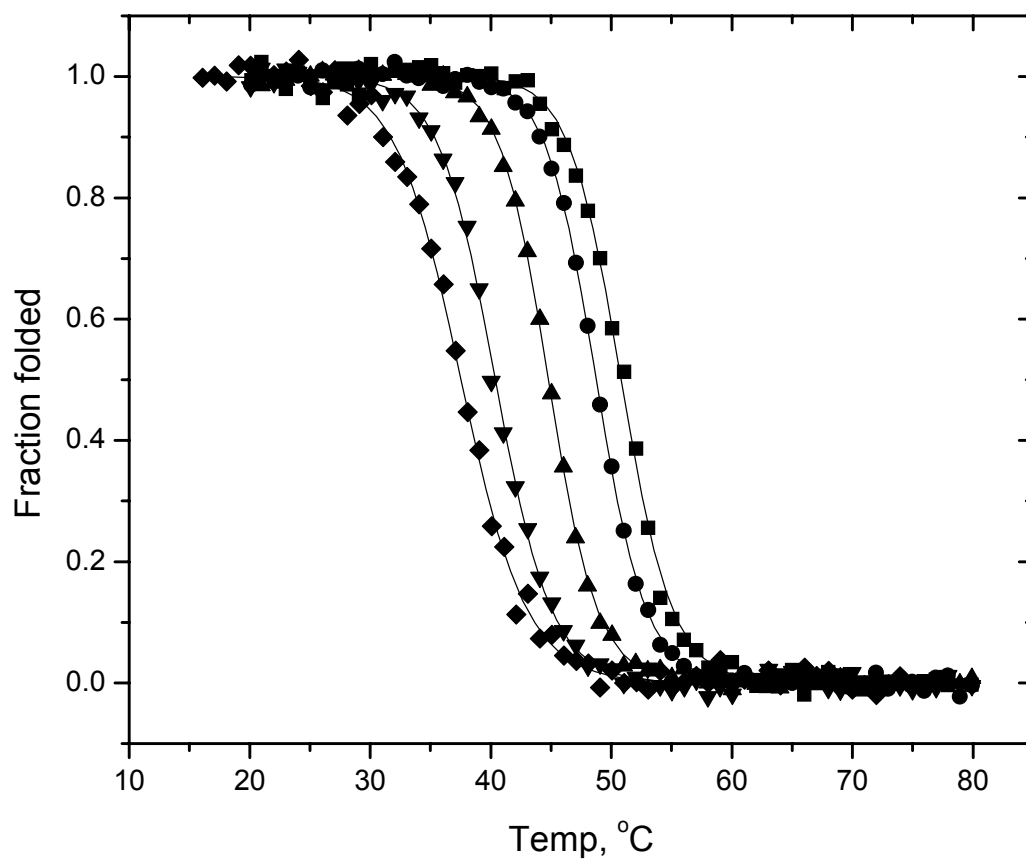Figure 25: Urea unfolding curves for RNase T1 wt (■), T1 N9A (●), T1 Y11F (▲), T1 T91V (▼) and T1 D76N (♦). Performed in 30 mM MOPS, pH 7 and 15 °C showing the fraction of folded protein as a function of urea concentration. Each line is the best fit of the respective data with eq 12. The results from these analyses are reported in Table 9.
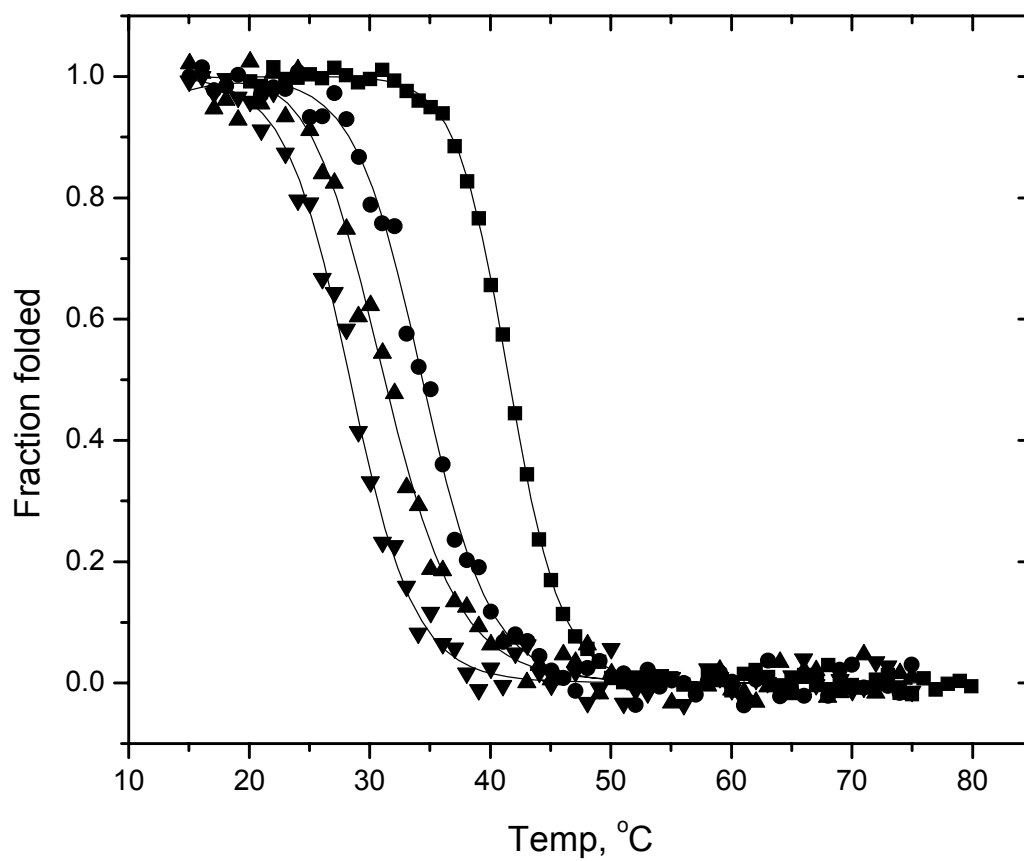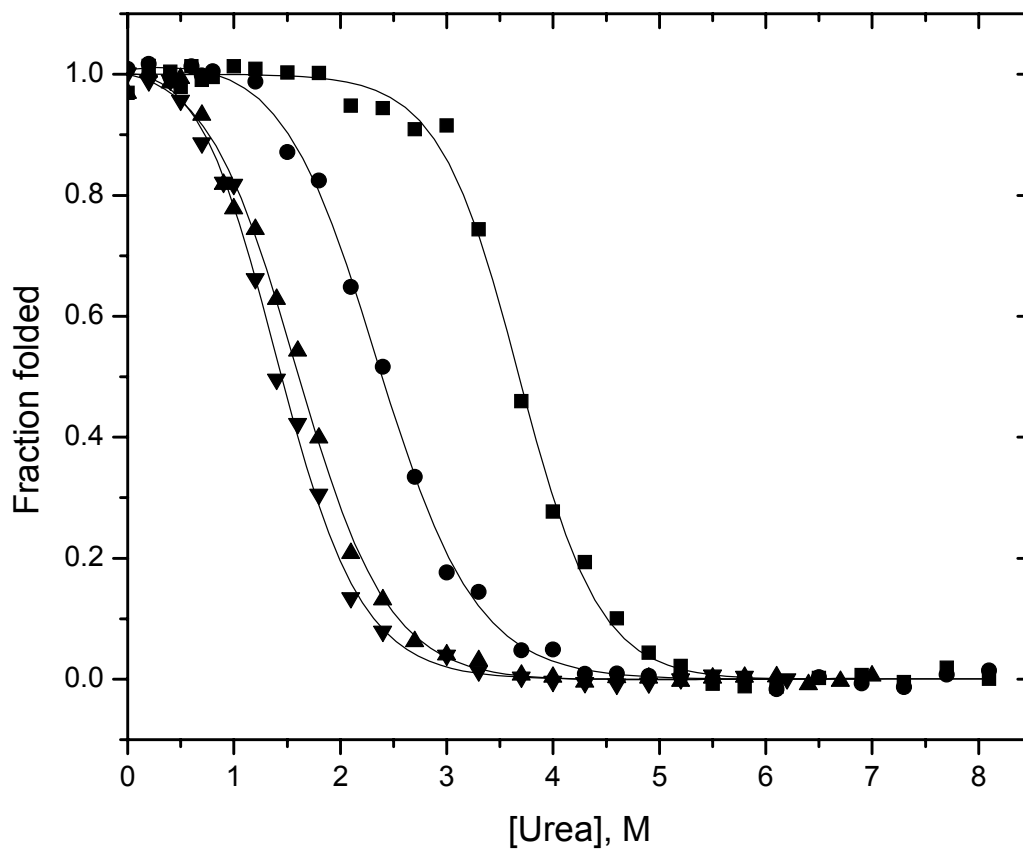
Figure 26: Urea unfolding curves for RNase T1 N9A Y11F (■), T1 N9A T91V (●), T1 Y11F T91V (▲) and T1 Triple Mutant (▼). Performed in 30 mM MOPS, pH 7 and 15 °C showing the fraction of folded protein as a function of urea concentration. Each line is the best fit of the respective data with eq 12. The results from these analyses are reported in Table 9.

Table 9: Parameters Characterizing the Urea Unfolding of RNase T1 and Mutants Monitored by Intrinsic Flourescence.[a]

RNase T1	C_{mid}, M	m, kcal/mol · M	ΔG_{H_2O},[b] kcal/mole	$\Delta(\Delta G)$,[c] kcal/mole	Expected $\Delta(\Delta G)$[d]
wild-type	6.02	1.06	6.4	-	-
N9A	5.44	1.05	5.7	-0.6	-
Y11F	4.18	1.11	4.6	-2.0	-
T91V	2.97	1.22	3.6	-3.2	-
D76N	2.65	1.33	3.5	-3.6	-
N9A Y11F	3.68	1.58	5.8	-2.5	-2.6
N9A T91V	2.35	1.28	3.0	-3.9	-3.8
Y11F T91V	1.60	1.41	2.3	-4.7	-5.2
TM	1.39	1.51	2.1	-4.9	-5.8

[a] C_{mid} and m were obtained by nonlinear least squares analysis of urea unfolding curves with eq 4. Protein concentrations were ~0.01 mg/ml in 30 mM MOPS pH 7 at 15°C. Errors for Cmid are ± 0.1, and for m ± 0.05. [b] $\Delta G_{H_2O} = C_{mid} \cdot m$. [c] $\Delta(\Delta G) = (C_{mid\ mut} - C_{mid\ wt}) \cdot m_{wt}$.
[d] The expected $\Delta(\Delta G)$ based on the $\Delta(\Delta G)$ of the single mutants.

The pKa of the side chain carboxyl of Asp76 in wild type RNase T1 was determined by the Tanford-Wyman approach for pKa determination. Figure 7A, Chapter II, shows the conformational stability, ΔG, as a function of pH for RNase T1 wild type and T1 D76N. The data from A were used to determine the difference in conformational stability, $\Delta(\Delta G)$, between the two variants of the protein as a function of pH, shown in Figure 7B. The data shown in Figure 7B were fit with eq 16 and the pKa of the Asp76 carboxyl in the folded, 0.7, and in the unfolded protein, 3.4, were obtained. The results from this analysis are in good agreement with previous results determined by Tanford-Wyman analysis using urea unfolding studies at 15 °C, 0.5 and 3.7, respectively.(*56*)

The pKa of the side chain carboxyl of Asp76 in T1 N9A was determined by the Tanford-Wyman approach. Figure 27A shows the ΔG as a function of pH for N9A and D76N. These data were used to determine $\Delta(\Delta G)$ as a function of pH as shown in Figure 27B. The data shown in Figure 27B were fit with eq 16 and the pKa of the Asp76 carboxyl in folded N9A, 1.7, and in unfolded N9A, 3.3, were obtained.

Potentiometric difference titrations were used to determine the pKa of the side chain carboxyl of Asp76 in T1 Y11F, T1 T91V and the double mutant T1 N9A Y11F. Proton binding curves for T1 Y11F and D76N are shown in Figure 28. Also shown in Figure 28 is the difference curve determined by subtracting the curve for D76N from the curve for Y11F. The difference curve was fit with eq 9 to determine the pKa of the Asp76 carboxyl in Y11F, 4.0. Proton binding curves for T1 T91V and D76N along with their difference curve are shown in Figure 29. The difference curve was fit with eq 9 to determine the pKa of the Asp76 carboxyl in T91V, 4.2. Proton binding curves for T1

Figure 27: Tanford-Wyman analysis for Asp76 in RNase T1 N9A. A: The conformational stability at 25 °C in kcal/mole of RNase T1 N9A (■) and T1 D76N (●) as a function of pH. The lines are drawn to guide the eye only. B: $\Delta(\Delta G)$ of the data from A as a function of pH. The solid line represents the best fit to the data using eq 16 from pH 0.6 to 7. From this analysis: pKa (folded) = 1.7 ± 0.2 and pKa (unfolded) = 3.3 ± 0.2.
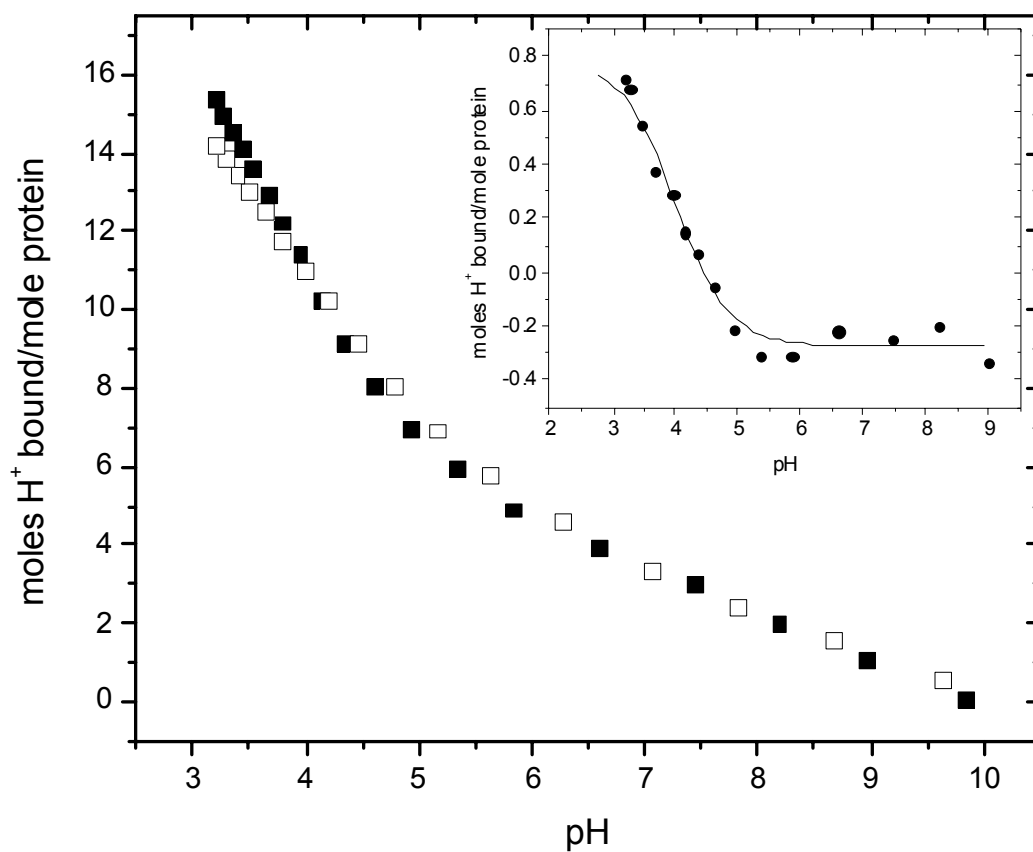
Figure 28: Potentiometric titration curves for RNase T1 Y11F (■) and T1 D76N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of one proton with a corresponding pKa of 4.0 ± 0.2. Experimental conditions were 100 mM KCl and 15 °C.
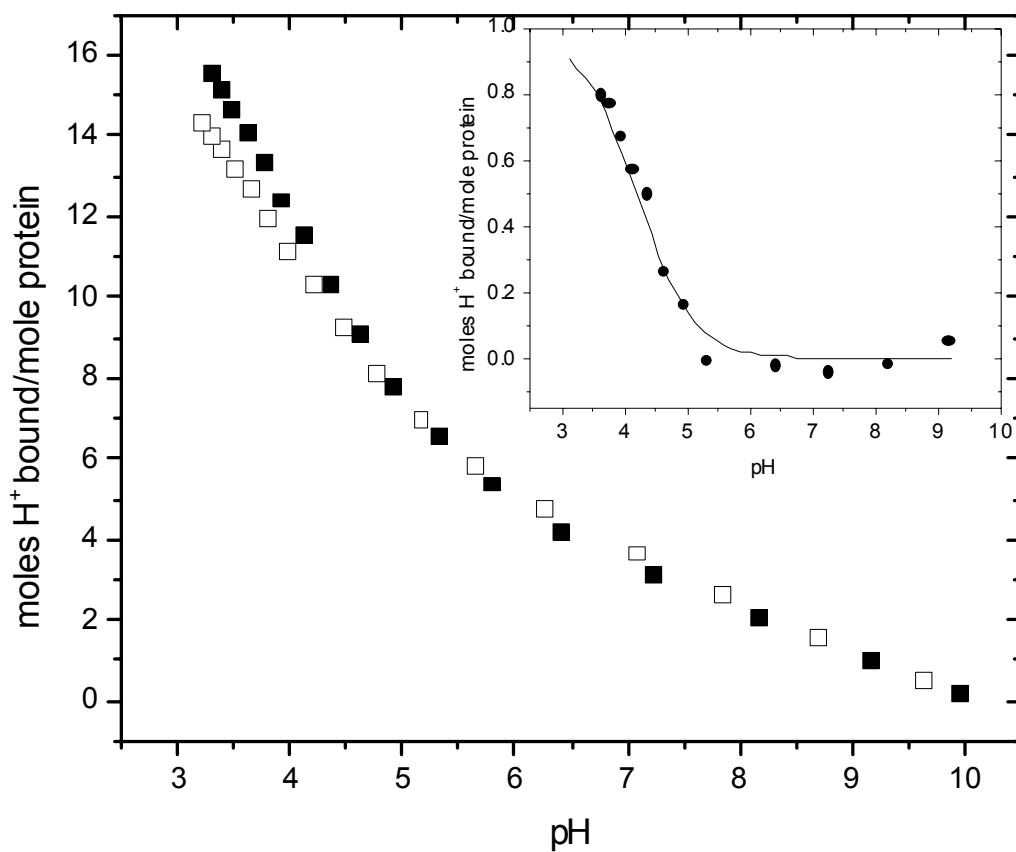
Figure 29: Potentiometric titration curves for RNase T1 T91V (■) and T1 D76N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of one proton with a corresponding pKa of 4.2 ± 0.2. Experimental conditions were 100 mM KCl and 15 °C.
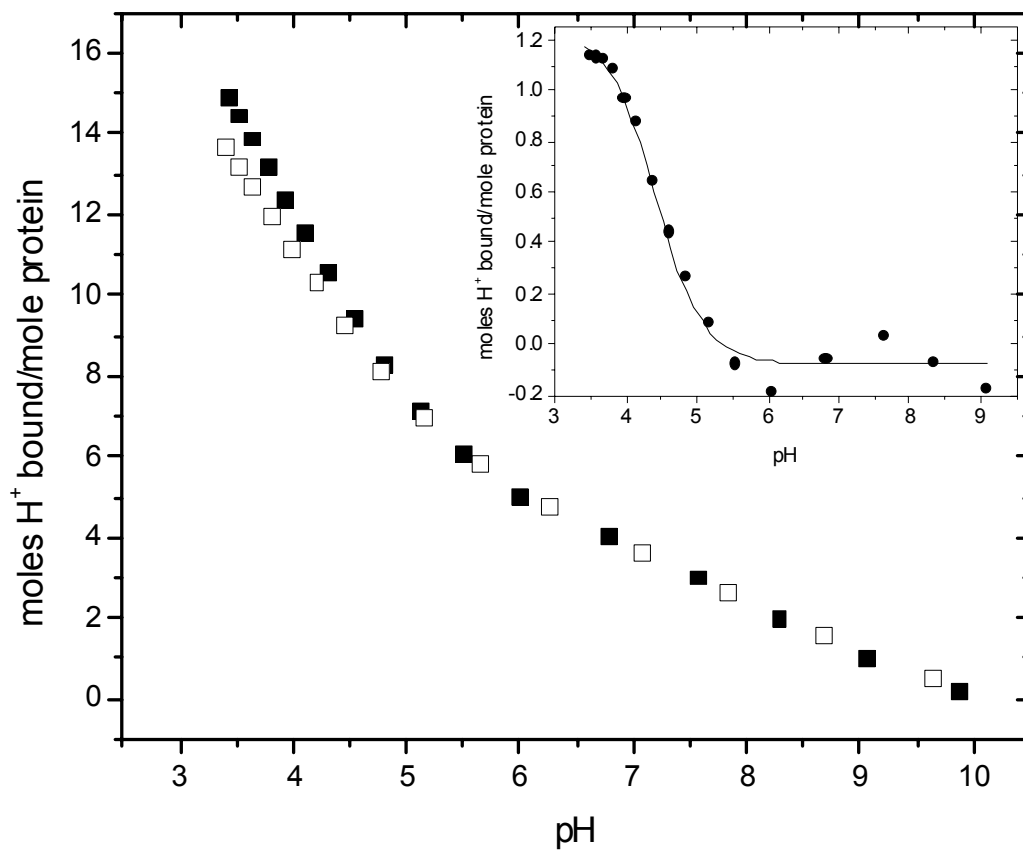
Figure 30: Potentiometric titration curves for RNase T1 N9A Y11F (■) and T1 D76N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of one proton with a corresponding pKa of 4.4 ± 0.2. Experimental conditions were 100 mM KCl and 15 °C.

N9A Y11F and D76N along with their difference curve are shown in Figure 30. The difference curve was fit with eq 9 to determine the pKa of the Asp76 carboxyl in N9A Y11F, 4.4.

The pKa of the side chain carboxyl of Asp76 in T1 N9A T91V was determined by both the Tanford-Wyman analysis and by potentiometric difference titrations. The Tanford-Wyman analysis is shown in Figure 31, where in part A, ΔG as a function of pH is shown for N9A T91V and D76N. The data from A were used to determine $\Delta(\Delta G)$ as a function of pH which is shown in part B of Figure 31. The data in part B were fit with eq 16 and the pKa of the Asp76 carboxyl obtained, 4.5 in the folded protein and 3.4 in the unfolded protein. Figure 32 shows the proton binding curves for N9A T91V and D76N along with their difference curve. The difference curve was fit with eq 9 to determine the pKa of the Asp76 carboxyl in N9A T91V, 5.2.

The pKa of the side chain carboxyl of Asp76 in T1 Y11F T91V was determined by both the Tanford-Wyman analysis and by potentiometric difference titrations. The Tanford-Wyman analysis is shown in Figure 33, where in part A, ΔG as a function of pH is shown for Y11F T91V and D76N. The data from A were used to determine $\Delta(\Delta G)$ as a function of pH which is shown in part B of Figure 33. The data in part B were fit with eq 16 and the pKa of the Asp76 carboxyl obtained, 5.6 in the folded protein and 3.7 in the unfolded protein. Figure 34 shows the proton binding curves for Y11F T91V and D76N along with their difference curve. The difference curve was fit with eq 9 to determine the pKa of the Asp76 carboxyl in Y11F T91V, 6.2.
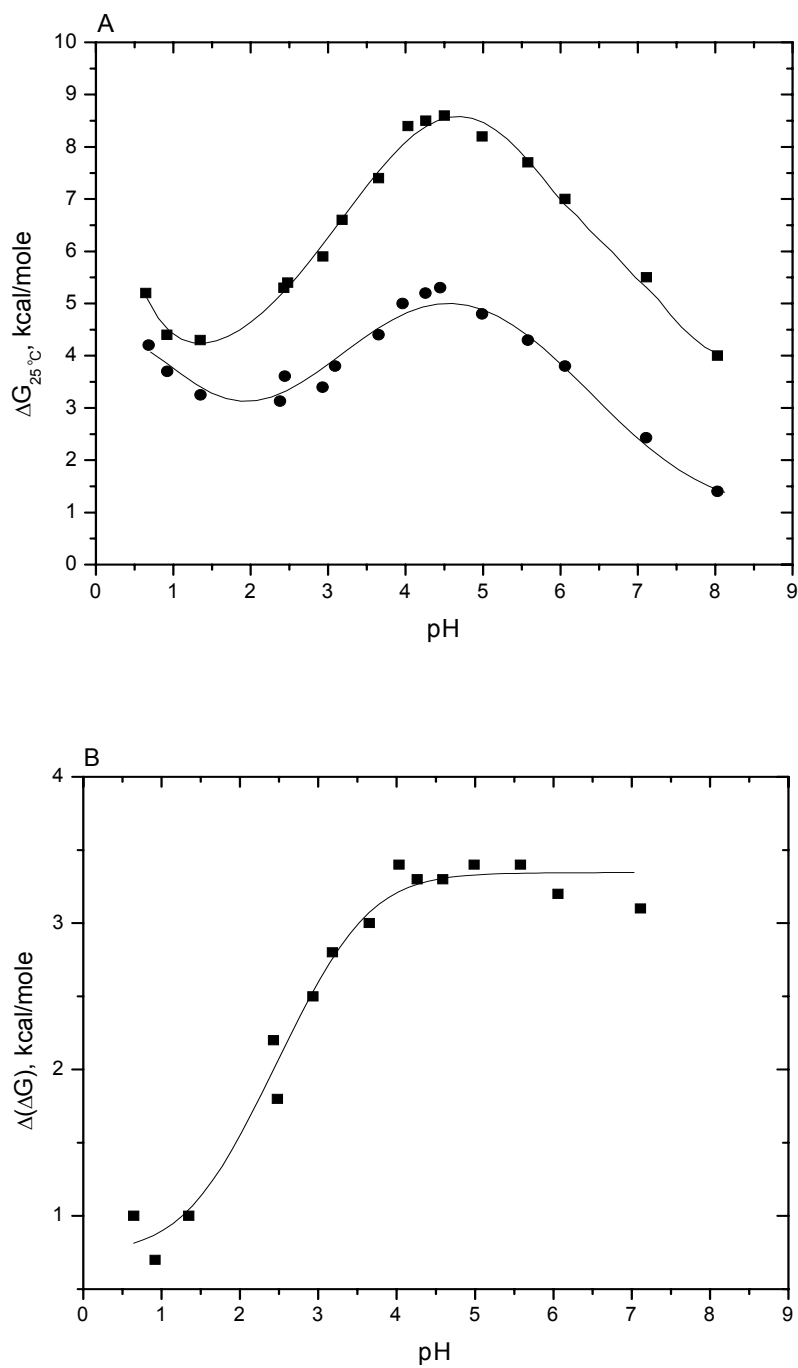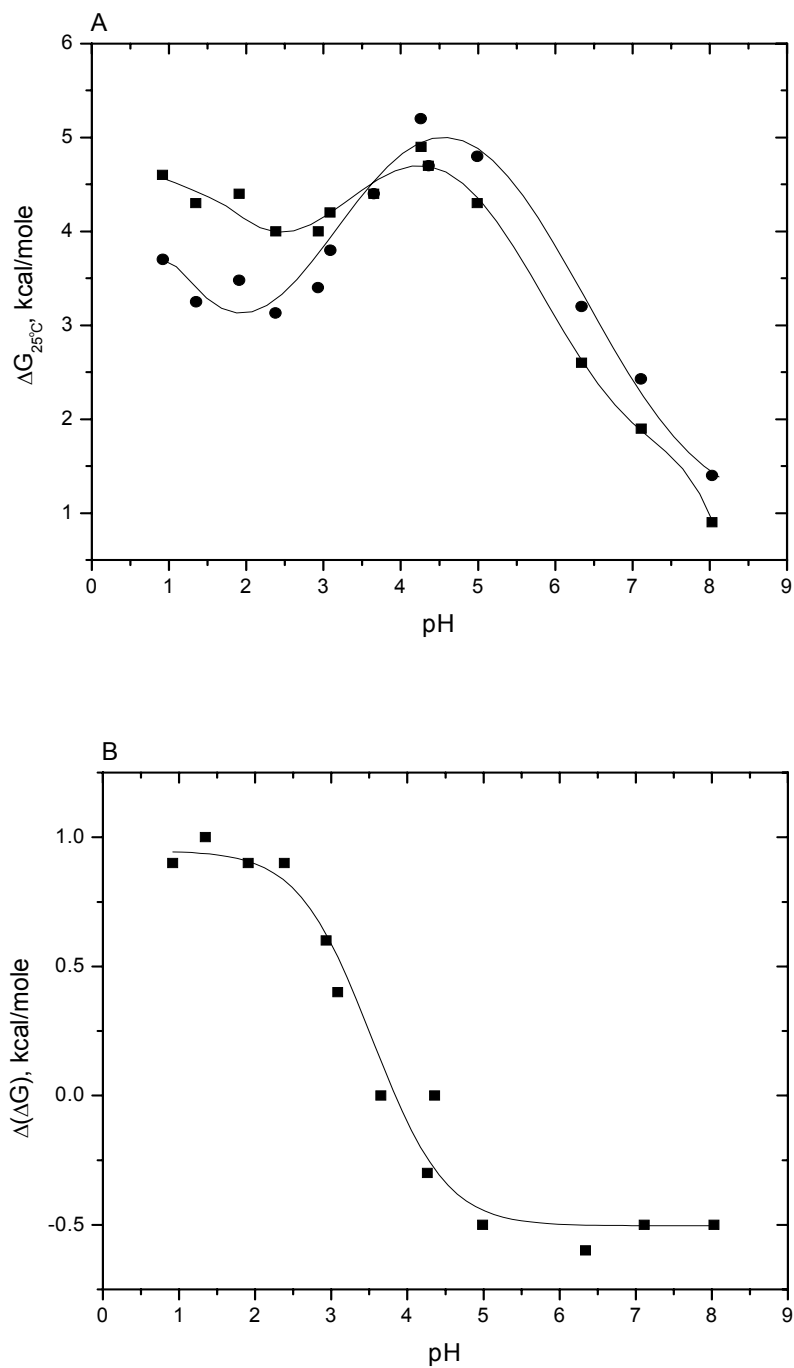
Figure 31: Tanford-Wyman analysis for Asp 76 in RNase T1 N9A T91V. A: The conformational stability at 25 °C in kcal/mole of RNase T1 N9A T91V (■) and T1 D76N (●) as a function of pH. The lines are drawn to guide the eye only. B: $\Delta(\Delta G)$ of the data from A as a function of pH. The solid line represents the best fit to the data using eq 16 from pH 0.9 to 8. From this analysis pKa (folded) = 4.5 ± 0.2 and pKa (unfolded) = 3.4 ± 0.2.
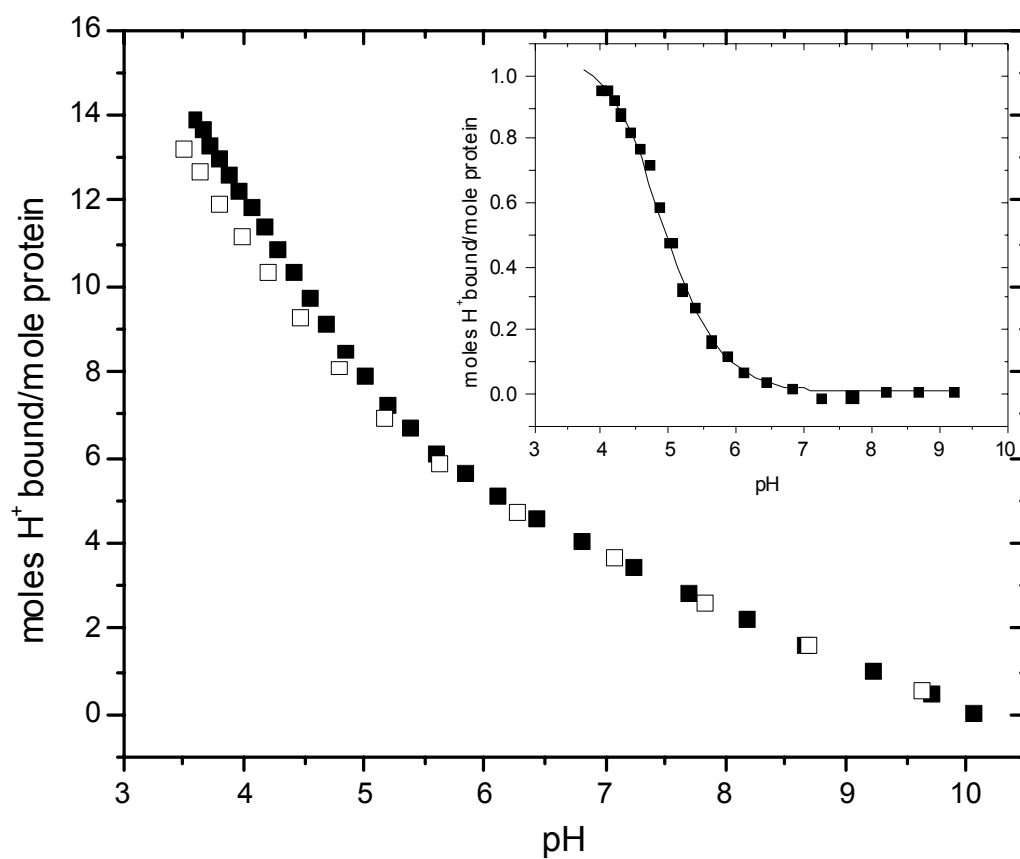
Figure 32: Potentiometric titration curves for RNase T1 N9A T91V (■) and T1 D76N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of one proton with a corresponding pKa of 5.2 ± 0.2. Experimental conditions were 100 mM KCl and 15 °C.
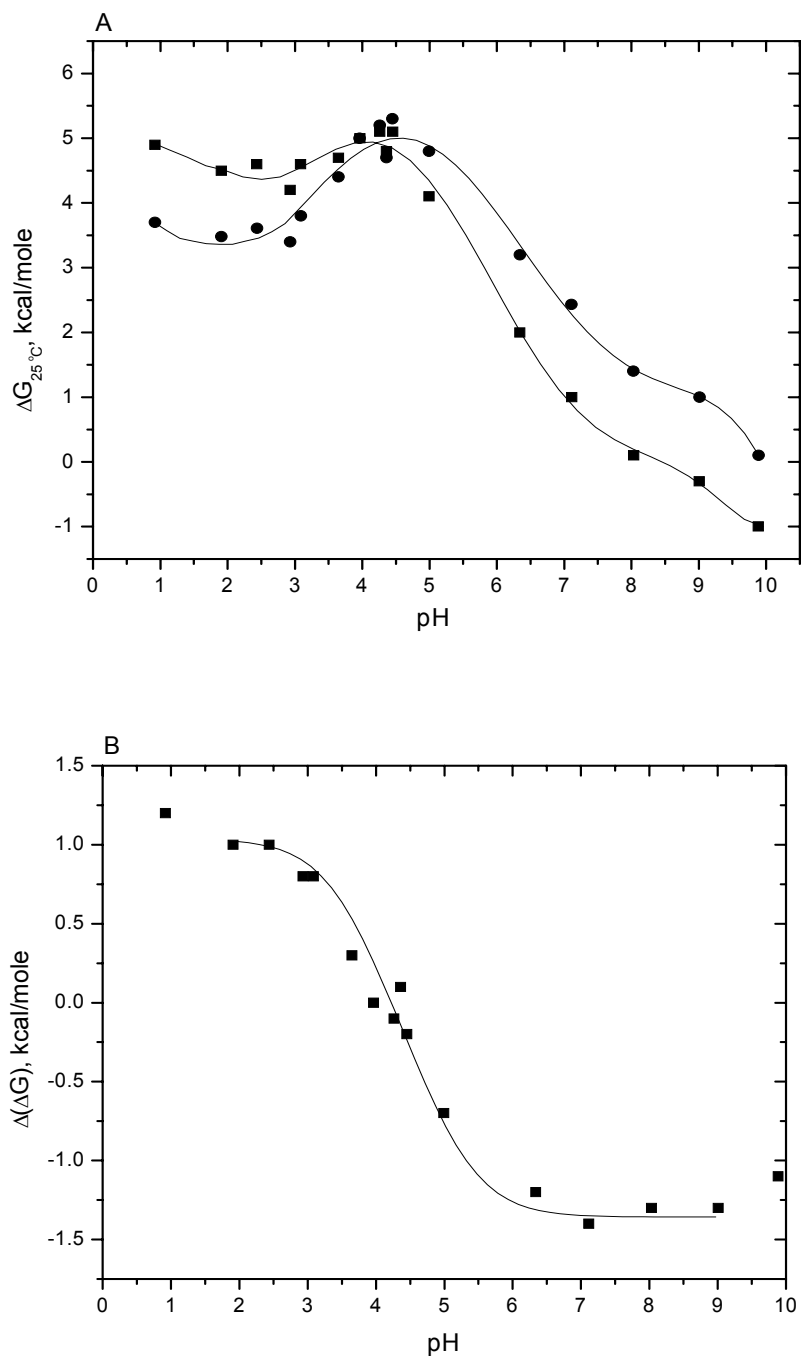
Figure 33: Tanford-Wyman analysis for Asp 76 in RNase T1 Y11F T91V. A: The conformational stability at 25 °C in kcal/mole of RNase T1 Y11F T91V (■) and T1 D76N (●) as a function of pH. The lines are drawn to guide the eye only. B: $\Delta(\Delta G)$ of the data from A as a function of pH. The solid line represents the best fit to the data using eq 16 from pH 2 to 9. From this analysis: pKa (folded) = 5.6 ± 0.2 and pKa (unfolded) = 3.7 ± 0.2.
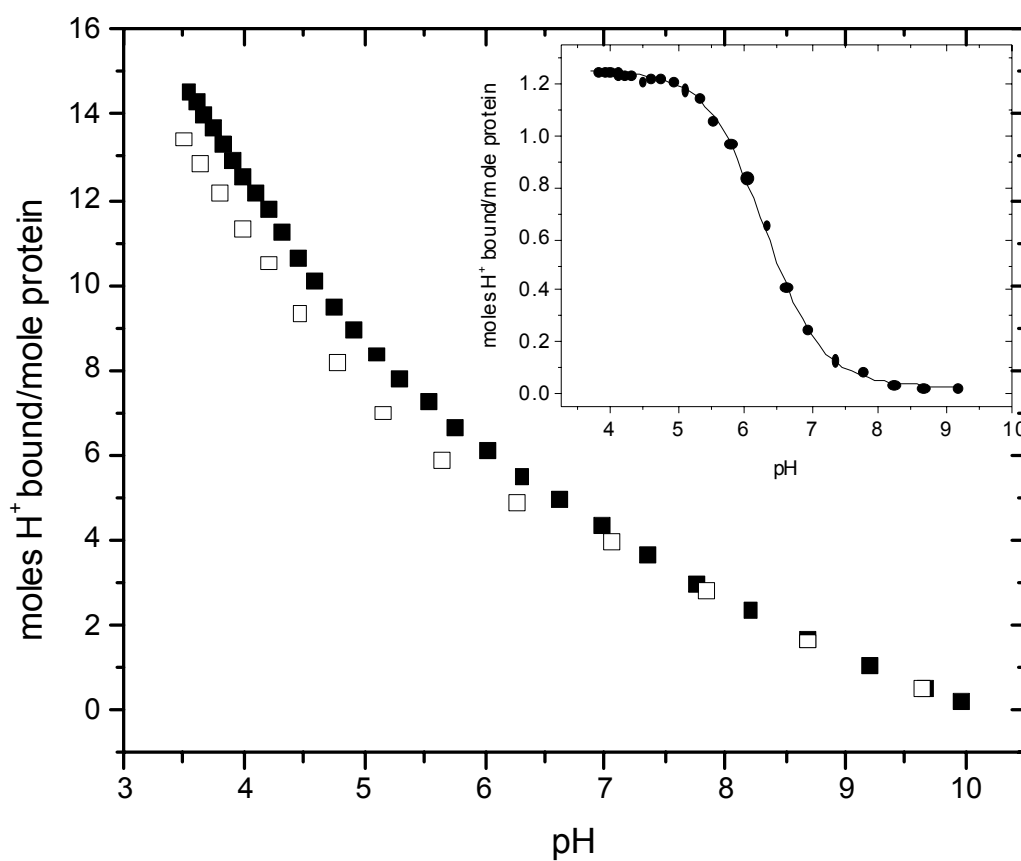
Figure 34: Potentiometric titration curves for RNase T1 Y11F T91V (■) and T1 D76N (□). Each curve represents the average of three independent titrations.   (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of one proton with a corresponding pKa of 6.2 ± 0.2. Experimental conditions were 100 mM KCl and 15 °C.
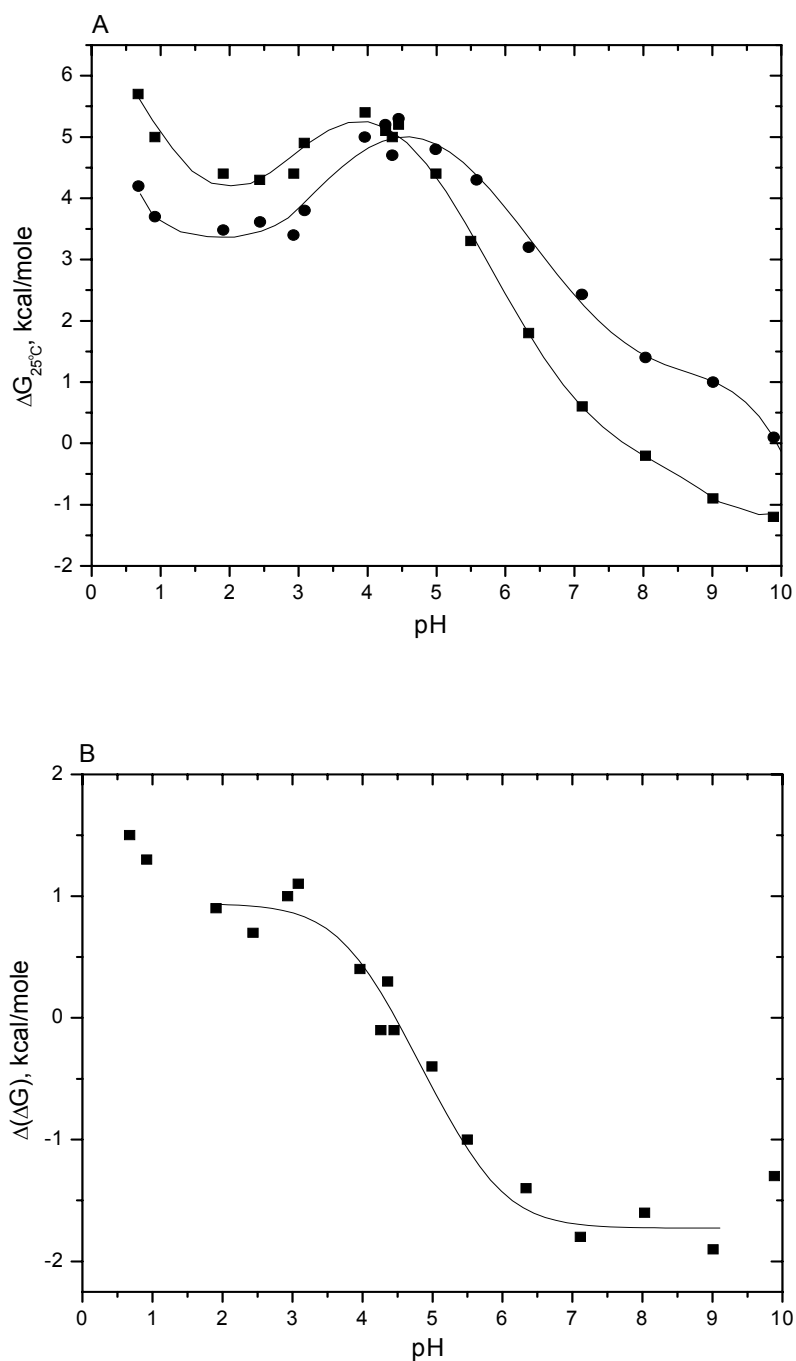
Figure 35: Tanford-Wyman analysis for Asp 76 in RNase T1 TM. A: The conformational stability at 25 °C in kcal/mole of RNase T1 TM (■) and T1 D76N (●) as a function of pH. The lines are drawn to guide the eye only. B: $\Delta(\Delta G)$ of the data from A as a function of pH. The solid line represents the best fit to the data using eq 16 from pH 2 to 9. From this analysis: pKa (folded) = 6.2 ± 0.2 and pKa (unfolded) = 3.8 ± 0.2.
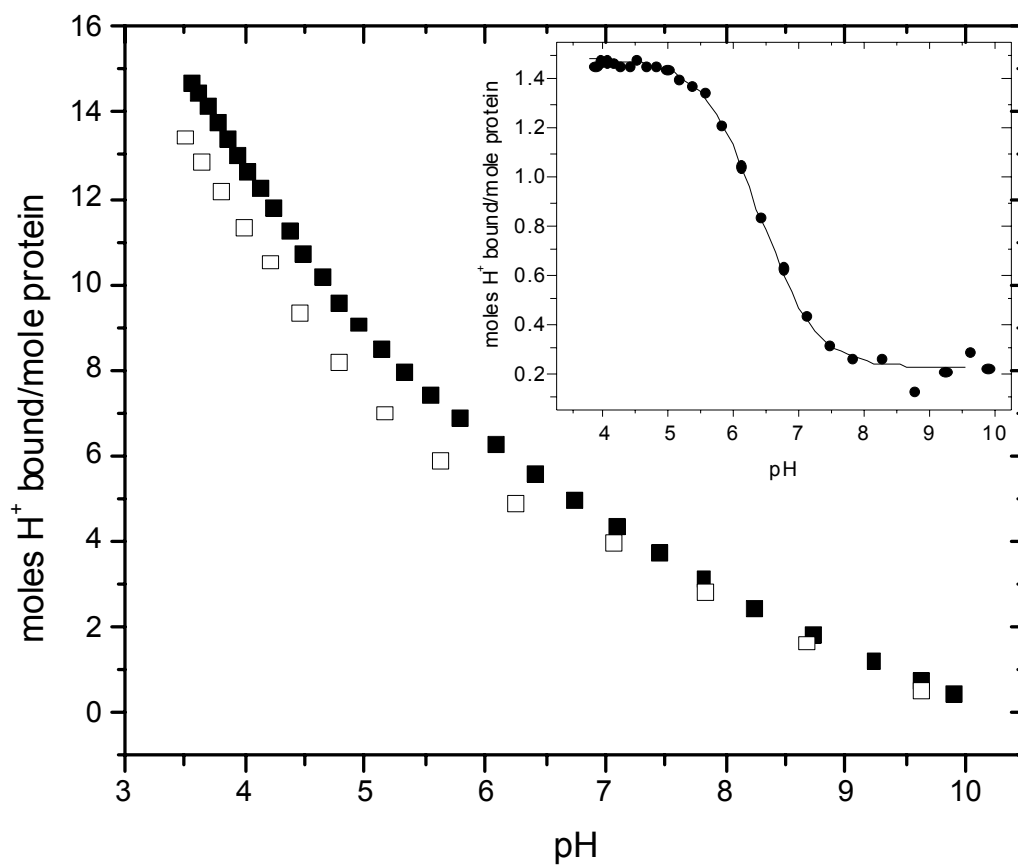
Figure 36: Potentiometric titration curves for RNase T1 TM (■) and T1 D76N (□). Each curve represents the average of three independent titrations. (Inset) The solid line (—) is the best fit of the difference curve (●) with eq 9, which shows an uptake difference of one proton with a corresponding pKa of 6.5 ± 0.2. Experimental conditions were 100 mM KCl and 15 °C.

The pKa of the side chain carboxyl of Asp76 in T1 TM, N9A Y11F T91V, was determined by both the Tanford-Wyman analysis and by potentiometric difference titrations. The Tanford-Wyman analysis is shown in Figure 35, where in part A, ΔG as a function of pH is shown for T1 TM and D76N. The data from A were used to determine $\Delta(\Delta G)$ as a function of pH which is shown in part B of Figure 35. The data in part B were fit with eq 16 and the pKa of the Asp76 carboxyl obtained, 6.2 in the folded protein and 3.8 in the unfolded protein. Figure 36 shows the proton binding curves for T1 TM and D76N along with their difference curve. The difference curve was fit with eq 9 to determine the pKa of the Asp76 carboxyl in T1 TM, 6.5.

DISCUSSION

In the folded structure of RNase T1, the side chain carboxyl of Asp76 is buried, makes no short-range electrostatic interactions but forms three good intramolecular hydrogen bonds. Figure 4A, Chapter I, shows the ribbon diagram based on the crystal structure of RNase T1 with the location of Asp76 shown relative to the rest of the protein. Figure 4B shows an expanded view of the region around the side chain of Asp76. The h-bond contacts that the carboxyl group makes are; 2.9 Å to Asn9 Nδ2, 2.7 Å to Tyr11 OH, and 2.6 Å to Thr91 Oγ1. The carboxyl group of Asp76 is about 98% buried and each group it forms an h-bond to is also mostly buried. The advantage of Asp76 for this research is that each of the groups that donate an h-bond to the side chain carboxyl can be readily modified to a non hydrogen bonding group by mutagenesis.
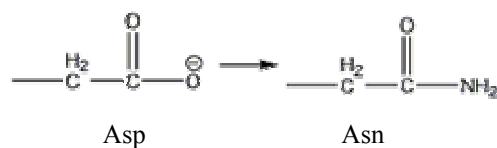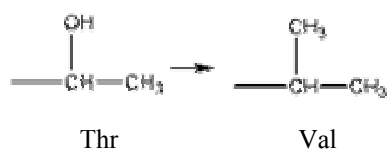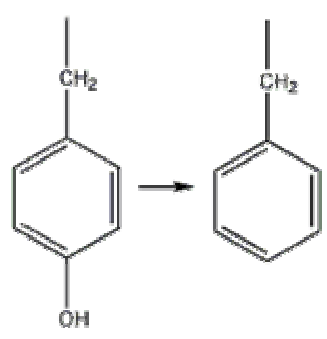
	Volume (Å^3)	Hydrophobicity (kcal/mole)	Side Chain Entropy (TΔS) (kcal/mole)
Asp → Asn	+8	-0.23	-0.25
Asn → Ala	-34	-1.24	1.03
Thr → Val	+18.5	-1.31	0.65
Tyr → Phe	-4	-1.13	0.51

Figure 37: Differences in structure, volume, hydrophobicity, and side chain conformational entropy (TΔS) for the following mutations: Asp to Asn, Asn to Ala, Thr to Val, and Tyr to Phe. The differences in volume are from Tsai et al.(*85*) The differences in hydrophobicity are from Pace and are based on the *n*-octanol hydrophobicity scale of Fauchere and Pliska.(*86, 87*) The differences in conformational entropy are based on the mean TΔS values at 300 K as found in Doig and Sternberg.(*88*)

Figure 37 lists the mutations used for this project. The amino acid chosen to replace Asn9 was Ala.(*50*) Asn to Ala mutations may introduce a cavity of about 34 Å^3. This cavity is large enough to accommodate a water molecule, which in turn could h-bond to the carboxyl of Asp76. In the absence of a water molecule, the cost of cavity formation could be compensated by more favorable hydrophobicity and side chain conformational entropy, TΔS, for the Ala. The side chain conformation and volume of Val suggests that it might be a good choice to replace Asn as well. The obvious replacement for a Tyr is Phe.(*107*) The conformations and volumes of the two side chains are nearly identical; removal of the Tyr OH may introduce only about a 5 Å^3 cavity. The side chain of Phe has a more favorable hydrophobicity and TΔS, but Phe is still the best replacement. The best choice for replacing a Thr residue whose side chain hydroxyl is h-bonded is probably Val. The conformation of the Val side chain is very similar to the side chain of Thr. The volume of the Val side chain is a little larger, and the hydrophobicity and TΔS are more favorable. For a more in depth discussion on Thr to Val mutations see Chapter IV.

General Observations on the Conformational Stability: The m-values of the mutants, especially the higher order mutants, are substantially higher than the m-value of the wild type protein. This could indicate an intermediate in the unfolding pathway of wild type T1, however there is considerable evidence to suggest that the unfolding of T1 and its mutants closely approach a two-state unfolding mechanism.(*50, 51*) The higher m-values more likely result because the mutants unfold to a greater extent than the wild type protein as suggested by Shortle.(*108*) Higher m-values for mutants were also

observed by Giletto and Pace with Asp76 mutants.(*56*) They suggested that the charged Asp76 side chain interacts with other ionizable groups in the unfolded protein. When they lowered the pH below the pKa of the Asp76 carboxyl in the unfolded state, they observed that the m-values of the Asp76 mutants were no longer elevated relative to wild type. The results here suggest that it is the hydrogen bonding network around Asp76 that is responsible for any residual structure in unfolded T1, and disruption of the network allows the mutant protein to unfold to a greater extent than wild type. In the previous work, lowering the pH for wild type could have disrupted the residual hydrogen bond network by protonating the carboxyl group in the unfolded protein.

An analysis with double mutant cycles of the changes in conformational stability of the double and triple mutants shows that all of the mutations are relatively additive. The results are summarized in the last column of Tables 8 and 9. This is a bit surprising when one considers that each of the three residues, Asn9, Tyr11 and Thr91, are all interrelated through the hydrogen bond network with Asp76. The only significant deviation is with the Triple Mutant.

Conformational Stability Measurements: Previous studies of hydrogen bonding mutants suggest that an h-bond contributes about 1-1.5 kcal/mole to the conformational stability of a protein.(*45, 50*) The side chain amide of Asn9 is 58% buried and forms only the intramolecular hydrogen bond with the Oδ2 of Asp76. The amide group also forms an intermolecular h-bond to a highly conserved water molecule, WAT111 in the PDB crystal structure, 9RNT.(*54*) From the thermal unfolding data, Table 8, and the urea unfolding data, Table 9, the change in conformational stability, Δ(ΔG), for T1 N9A is

-0.6 kcal/mole. This is in good agreement with previous results for this mutant.(*50, 109*)

Langhorst, et al solved the crystal structure of T1 N9A, and showed that a new water

molecule replaces the amide group of Asn9.(*109*) This water molecule h-bonds to the

Oδ2 of Asp76 and to another new water molecule, which in turn forms an h-bond to the

backbone NH of Ala75. These new water molecules probably make a favorable

contribution to the conformational stability to the mutant protein, and could explain why

the Δ(ΔG) of N9A is less than the expected 1-1.5 kcal/mole. The side chain OH of

Tyr11 is ~99% buried and forms only the one h-bond to the Oδ2 of Asp76. The Δ(ΔG)

for Y11F is about -1.9 kcal/mol, and is in excellent agreement with previous results.(*50*)

The side chain OH of Thr91 is ~99% buried and forms two h-bonds in addition to the h-

bond to the side chain carboxyl of Asp76. They are 3.3 Å to the Oγ1 of Thr93 and 2.9 Å

to the amide N of Thr93. The Δ(ΔG) for T91V is -3.3 kcal/mole. The local environment

and interactions for the side chain carboxyl of Asp76 have been discussed earlier. The

Δ(ΔG) for D76N is -3.7 kcal/mole and is in good agreement with previous results.(*56*)

Each of these results for single mutants is in good agreement with the estimate that each

h-bond contributes about 1-1.5 kcal/mole to the conformational stability of a protein.

From Tables 8 and 9, the Δ(ΔG) for each double mutant and the triple mutant agree

reasonably well with estimates based on the Δ(ΔG) from the single mutants. The Δ(ΔG)

for the double mutant T1 N9A Y11F is -2.7 kcal/mole, which agrees very well with the

estimate of -2.5 based on the Δ(ΔG) values of the single mutants. Since the two side

chain groups h-bond to the same oxygen, Oδ2, we might expect that the Δ(ΔG) of the

double mutant would show some cooperativity toward stability, but no cooperativity is

observed. If we assume that the new water molecules found in T1 N9A are present in N9A Y11F, we may rationalize that these waters are exerting a stabilizing effect in the double mutant, eliminating any cooperativity toward destabilization between the hydrogen bond forming groups. The $\Delta(\Delta G)$ for T1 N9A T91V is -4.1 kcal/mole. This value agrees well with the estimate of -3.9 based on the $\Delta(\Delta G)$ values of the single mutants. The h-bonds formed by the Asn9 and Thr91 side chains are to the different oxygens of the carboxyl group, thus we might expect that the contribution of each to the stability would be additive, and it is. The $\Delta(\Delta G)$ for T1 Y11F T91V is -4.9 kcal/mole. This value agrees well with the estimate of -5.2 based on the $\Delta(\Delta G)$ values of the respective single mutants. As above, the h-bonds formed by the Tyr11 and Thr91 side chains are to the different oxygens of the Asp76 carboxyl group, which suggests that the contribution of each group to the stability of the double mutant might be additive, and they are. The $\Delta(\Delta G)$ of the triple mutant (TM) is -5.2 kcal/mole. This value is a little low based on the sum of the individual single mutants, -5.8. The experimental error for these measurements is about $\pm$ 0.5 kcal/mole, so this estimate can probably be considered within experimental error of the measured value and additivity is maintained. However, it is possible that the extra water molecules found in T1 N9A are not present in the triple mutant, and the positive cooperativity with stability observed for the triple mutant is a more accurate representation of the $\Delta(\Delta G)$ for removal of the hydrogen bond forming groups to the Oδ2 of Asp76.

Main Forces affecting Perturbed pKas: The main forces acting on ionizable groups responsible for perturbing their pKas are electrostatic interactions with other ionizable

groups, burial in a hydrophobic environment and hydrogen bonding. There has been considerable work investigating the effects of charge-charge interactions and the effects of burial in a hydrophobic environment on the pKas of ionizable groups.(*10, 14, 15, 24-27, 35*) These two forces are the most studied of the three and are surely major forces affecting the pKas of perturbed ionizable groups. There has been less work on the effects of hydrogen bonding, but it also seems to be a primary force perturbing the pKas of some ionizable groups.(*28, 29, 42, 56, 96, 97*)

Previous evidence suggests that burial of a carboxyl group in a hydrophobic environment can elevate the pKa of that group by 3 units or more.(*18, 27*) The results for the pKa of the buried, fully hydrogen bonded carboxyl of Asp76 in wild type RNase T1 shows that the three intramolecular h-bonds formed by the side chain carboxyl perturb the pKa to a very acidic value. In the absence of other interacting groups in the vicinity, this shows that the h-bonds introduce a perturbation counter to and larger than the hydrophobic burial. The question to consider here is how much each h-bond contributes to the total perturbation on the carboxyl group and what effect the removal of all the h-bonds has on the pKa.

pKa of the Side Chain Carboxyl of Asp76 in the Single Mutants: The pKa of the side chain carboxyl of Asp76 in wild type RNase T1 has been previously determined and discussed.(*56*) The analysis was repeated here for technique confirmation, and is shown in Figure 7, Chapter I. The pKa of the Asp76 carboxyl group in T1 N9A from Figure 27 is 1.7. Based on the results for the side chain carboxyl of Asp33 in RNase Sa, Chapter IV, we might expect that the pKa of the Asp76 carboxyl group in the single mutants

would be about 4. The pKa in Y11F and T91V approximate this, but N9A seems to be a special case. The results suggest that the interaction between the carboxyl group and the amide nitrogen is weaker than the interactions with the hydroxyls from either of the other side chains. The Nδ2 of Asn9 has a lower electronegativity than the oxygen atoms of Tyr11 or Thr91. This lower electronegativity means less polarity of the N-H bond. The h-bond is also the longest of the three, 2.9 Å. These two factors translate into a weaker h-bond between the Asp76 carboxyl group and the side chain amide of Asn9. This might explain the lower pKa for the carboxyl group in N9A. The side chain amide of Asn9 is more solvent exposed than the side chain groups of either Tyr11 or Thr91, 58% compared to ~99% for both Tyr11 and Thr91. This added exposure might allow the pKa in the mutant N9A to more closely approach the intrinsic value for an Asp side chain, especially since the Ala side chain is considerably smaller than the Asn side chain. On the other hand the added burial of the other two single mutants, Y11F and T91V, combined with the larger mutant side chains would add more hydrophobic interaction. This would be expected to elevate the pKa of the carboxyl, which may be occurring in each of the cases here, and is supported by the results for the Thr56 mutants of RNase Sa, Chapter IV. In the crystal structure of T1 N9A extra water molecules were observed which may be affecting the pKa of the carboxyl group.(*110*) WATA, as designated by Langhorst, et al, replaces the side chain amide of Asn9 in N9A and forms an intermolecular h-bond to the carboxyl group of Asp76. The h-bond formed with this water may be exerting a perturbation on the pKa of the side chain carboxyl similar to an h-bond with a side chain group. All these interactions are probably contributing to the

relatively low pKa of the side chain carboxyl of Asp76 in T1 N9A. As a result, more work should be done in an attempt to isolate the main causes for the somewhat unusual pKa of the carboxyl of Asp76 in T1 N9A.

The pKa of the Asp76 carboxyl group in T1 Y11F and in T1 T91V was determined by potentiometric difference titrations, as shown in Figures 28 and 29, respectively. The potentiometric difference titration approach was not attempted on T1 N9A because of the low pKa observed from the Tanford-Wyman analysis. We attempted the Tanford-Wyman approach to estimate the pKa in T1 Y11F, and the conformational stability, ΔG, as a function of pH for Y11F and T1 D76N are shown in Figure 38A. As can be seen in part B of Figure 38, the $\Delta(\Delta G)$ between the two protein variants does not change noticeably over the pH range studied. As discussed in Chapter IV, one of the assumptions in using the Tanford-Wyman approach for pKa determination is that the pKa of the group of interest in the folded protein should be considerably different from its pKa in the unfolded protein. This difference in pKa will result in a dependence in pH of $\Delta(\Delta G)$ between the two variants as can be seen in Figure 27 for T1 N9A. The fact that $\Delta(\Delta G)$ between Y11F and D76N in Figure 38B is not dependent on pH indicates that the pKa of the Asp76 carboxyl in folded Y11F is close to the pKa of the group in unfolded Y11F. If we assume that the pKa of the carboxyl group in unfolded Y11F can be approximated by the pKa of the group in unfolded wild type T1 then we can estimate that the pKa in folded Y11F is about 3.5.

From Figure 39 we see that $\Delta(\Delta G)$ as a function of pH for T91V and D76N is also independent of pH. The same argument applied to Figure 38 for Y11F can be applied to
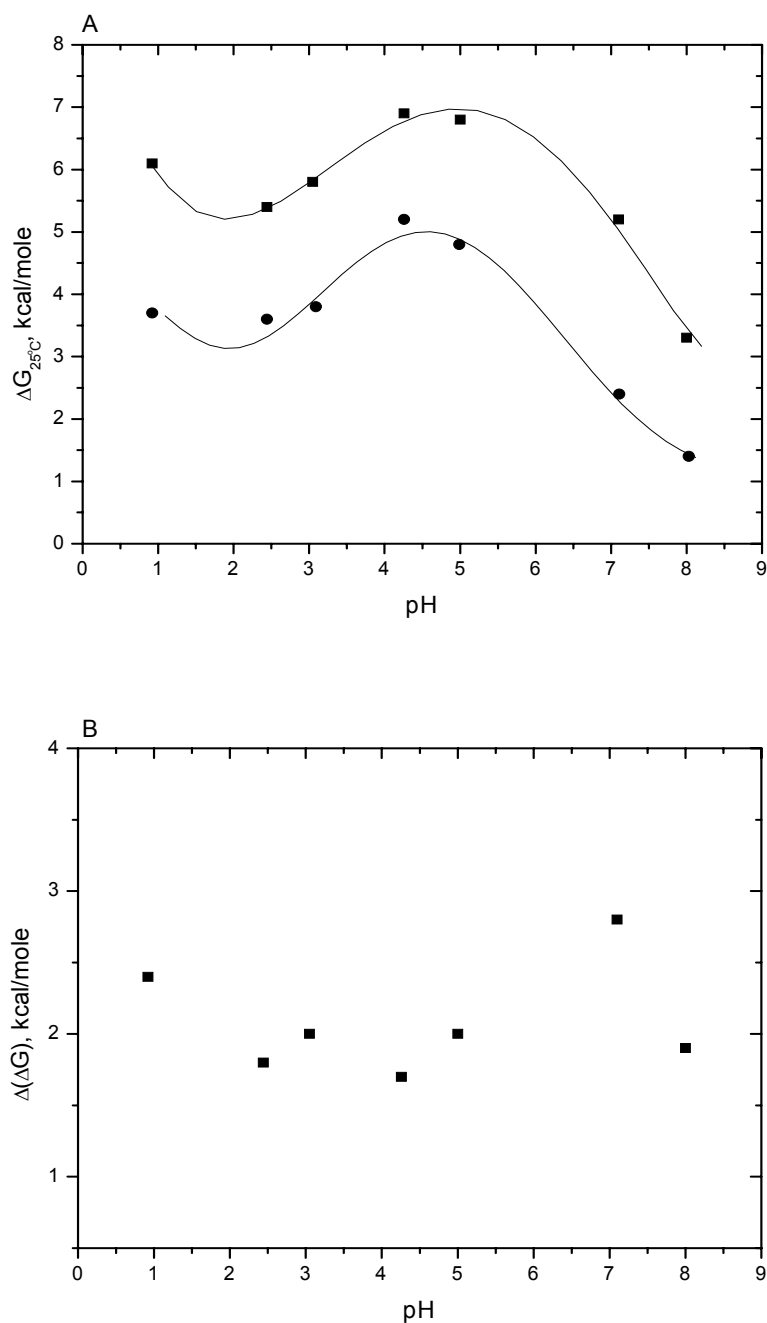
Figure 38: Tanford-Wyman analysis for Asp76 in RNase T1 Y11F. A: The conformational stability at 25 °C in kcal/mole of RNase T1 Y11F (■) and RNase T1 D76N (●) as a function of pH. The lines are intended to guide the eye only. B: $\Delta(\Delta G)$ of the data in A as a function of pH (■). $\Delta(\Delta G)$ between the two mutants is relatively constant throughout the pH range studied. This indicates that the pKa of theAsp76 carboxyl in the folded Y11F protein is relatively similar to its pKa in the unfolded protein.
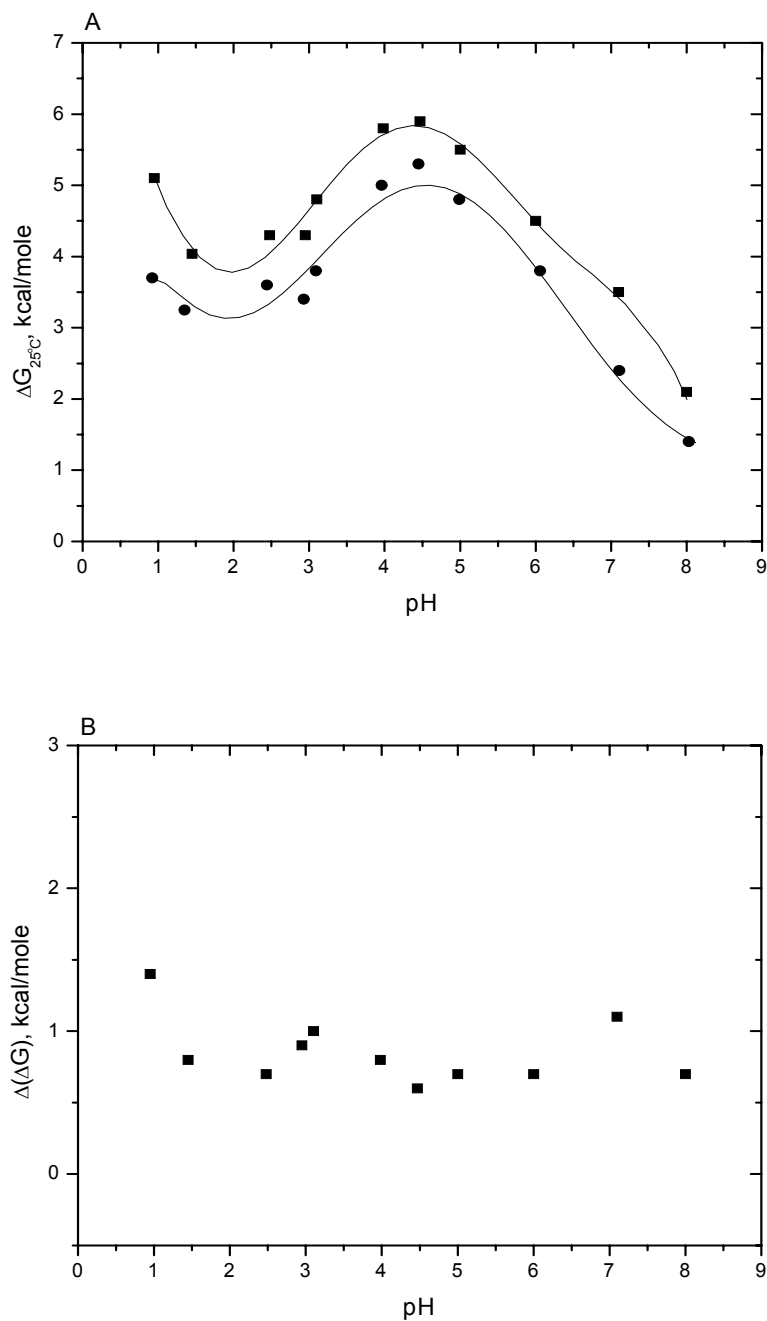
Figure 39: Tanford-Wyman analysis for Asp76 in RNase T1 T91V. A: The conformational stability at 25 °C in kcal/mole of RNase T1 T91V (■) and RNase T1 D76N (●) as a function of pH. The lines are intended to guide the eye only. B: $\Delta(\Delta G)$ of the data in A as a function of pH(■). $\Delta(\Delta G)$ between the two mutants is relatively constant throughout the pH range studied. This indicates that the pKa of the Asp76 carboxyl in the folded T91V protein is relatively similar to its pKa in the unfolded protein.

Figure 39 for T91V, and we can estimate that the pKa of the Asp76 carboxyl in T91V is about 3.5.

The estimates for the pKa of the Asp76 carboxyl in T1 Y11F and T1 T91V based on the attempted Tanford-Wyman analysis are a little low when compared with the results from the potentiometric difference titrations. The potentiometric difference titrations may be estimating the respective pKa a little high because of the pH where the ionization of the carboxyl group is taking place. The proton binding curves for the proteins shown in Figure 28 and in Figure 29 are approaching the acid pH limit for accurate determination. As a result, greater uncertainty is introduced in the analysis. Despite the differences, the results from the potentiometric difference titrations and Tanford-Wyman analysis for both Y11F and T91V, respectively, can be considered in reasonably good agreement.

pKa of the Side Chain Carboxyl of Asp76 in the Double Mutants and T1 TM: The pKa of the side chain carboxyl of Asp76 in T1 N9A Y11F is estimated at 4.4 by potentiometric difference titrations. Tanford-Wyman analysis yielded results very similar to those for Y11F and T91V. Figure 40A shows ΔG as a function of pH for N9A Y11F and D76N. From Figure 40B, the $\Delta(\Delta G)$ for the two protein variants is independent of pH, indicating that the pKa of the Asp76 carboxyl in folded N9A Y11F is similar to the pKa in the unfolded protein. Applying the same rationale as above, we can estimate the pKa of the Asp76 carboxyl to be about 3.5, which differs a bit from the result of the potentiometric difference titration. If the effect on the pKa of the carboxyl group of the two hydrogen bonding groups is additive we should expect the pKa to be
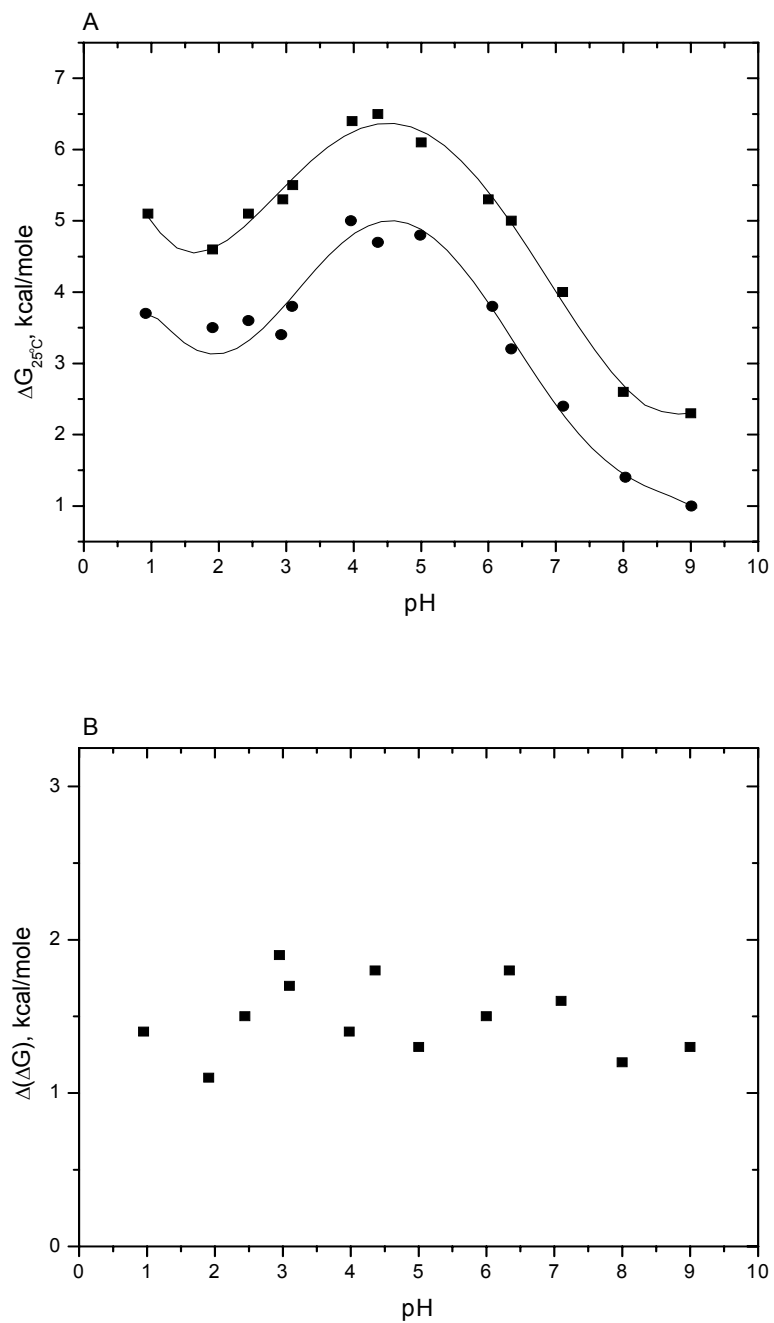
Figure 40: Tanford-Wyman analysis for Asp76 in RNase T1 N9A Y11F. A: The conformational stability at 25 °C in kcal/mole of RNase T1 N9A Y11F (■) and RNase T1 D76N (●) as a function of pH. The lines are intended to guide the eye only. B: Δ(ΔG) of the data in A as a function of pH. Δ(ΔG) between the two mutants is relatively constant throughout the pH range studied. This indicates that the pKa of Asp76 in the folded N9A Y11F protein is relatively similar to its pKa in the unfolded protein

about 4.5-5. This estimate is within error of the experimental result of 4.4, which again is probably a little high. We conclude that the two single mutations show some negative cooperativity toward the pKa of the carboxyl group. This cooperativity could be a result of greater accessibility of the carboxyl group to solvent. Asn9 and Tyr11 h-bond to the carboxyl of Asp76 via the $O\delta2$ oxygen, so in the double mutant that oxygen has no hydrogen-bonding partner left. If the new waters observed in the N9A single mutant are present in N9A Y11F, they may allow easier sampling of bulk solvent, especially by the $O\delta2$ oxygen. This would make the observed pKa of the carboxyl group lower than expected.

The pKa of the side chain carboxyl of Asp76 in T1 N9A T91V was determined by both Tanford-Wyman analysis and potentiometric difference titration. The average pKa from both techniques is 4.9. If the effect of the two hydrogen bonding groups in this double mutant are additive, we would expect the pKa to be about 5.2, which is in reasonable agreement with the observed result.

The pKa of the side chain carboxyl of Asp76 in T1 Y11F T91V was determined by both Tanford-Wyman analysis and potentiometric difference titration. The average pKa from both techniques is 5.9. Taking into account the effects of the two single mutations we should observe a pKa of about 7.5. The difference here underscores the possible errors in determining the pKas of the carboxyl group in Y11F and T91V with only the potentiometric difference titrations. Tyr11 is an h-bond donor to the $O\delta2$, and Thr91 is an h-bond donor to the $O\delta1$ of the side chain carboxyl of Asp76. Both of these interactions are fully buried and neither mutation should introduce a cavity at the

mutation site, therefore there should be no buried water molecules involved in the double mutant. The $\Delta(\Delta G)$ measurements of the single mutants are additive in the double mutant. There is no readily available explanation as to why the sum of the effects of the single mutants on the pKa of the carboxyl is more than one unit higher than the measured pKa of the carboxyl in the double mutant. As stated earlier, the results for the pKa of the Asp76 carboxyl in Y11F and T91V based on the potentiometric difference titrations are probably high. Based on the Tanford-Wyman analysis, the pKa of the carboxyl in each of these mutants could be 3.0-4.0. This implies that an estimate of the pKa of Y11F T91V based on the pKa results from the Y11F and T91V single mutants could be from 5.0 to over 7.5. Estimates on the lower end of this range would agree with experiment well and suggest no cooperativity between the side chains on the pKa of the carboxyl group, which is probably the case.

The pKa of side chain carboxyl of Asp76 in T1 TM was determined by both Tanford-Wyman analysis and potentiometric difference titration. The average pKa from the two techniques is 6.4. If we take into account the effects of the single site mutations on the pKa of the side chain carboxyl we may expect the pKa to be about 7.3 to over 8. This suggests that in the triple mutant there is some negative cooperativity between the groups and their affects on the pKa of the carboxyl group. This negative cooperativity could be from the apparent negative cooperativity from Asn9 and Tyr11. As a result we could be seeing the effects of new water molecules inserted into the structure of the protein. It is also possible that we have created an unstable pocket in the protein consisting of an ionizable group surrounded by hydrophobic groups. This pocket could

be experiencing some local repulsion and expansion. We do observe a marked decrease in relative activity, however this was expected. His92 is the active site acid for the protein. It is located in a loop segment that positions the active site right next to the groups under study here. The water mediated hydrogen bond network that includes each of the groups involved in this study has been studied previously, and it was observed that any disruptions in this network adversely affects the activity of the protein.(*110*) Therefore it was not surprising that the observed relative activity of several of these mutants was less than the activity of wild type. Most likely, the somewhat lower than expected pKa of the carboxyl group in T1 TM is a result of slight expansion of the pocket containing Asp76 and the possible invasion of new water molecules into the structure of the protein.

Estimates of the pKa using SATK: Estimates for the pKa of the side chain carboxyl of Asp76 in RNase T1 and the mutants studied here are reported in Table 10 along with the experimental results. As can be seen, SATK does not predict any changes in the pKa as a result of mutating any of the single residues or combinations of them. This was expected since the SATK equations were developed to explain charge-charge electrostatic interactions in proteins and not hydrogen bonding or even hydrophobic burial. Some studies have demonstrated reasonably good success using SATK predictions compared with experimental results.(*35, 36, 98, 99*) Each of these studies however, was dealing with charge-charge interactions, and in most cases with groups and interactions near the surface of the protein.

Table 10: pKa of the Side Chain Carboxyl of Asp76 in RNase T1 and Mutants.

RNase T1	pKa	SATK[d]	UHBD[d]
wild-type	0.7[a]	4.2	2.8
N9A	1.7[a]	4.2	3.4
Y11F	4.0[b]	4.2	3.1
T91V	4.2[b]	4.2	2.8
N9AY11F	4.4[b]	4.2	3.7
N9AT91V	4.9[c]	4.2	3.5
Y11FT91V	5.9[c]	4.2	3.2
TM	6.4[c]	4.2	3.8

[a] From Tanford-Wyman analysis only. [b] From Potentiometric Difference Titrations only. [c] From the average of the Tanford-Wyman analysis and the Potentiometric Difference Titrations. [d] Performed as described in Materials and Methods.

Estimates of the pKa using UHBD: The Finite-Difference Poisson Boltzmann

equation, FDPB, as employed in UHBD has been used previously to compare predicted

pKas of ionizable groups in proteins to experimental results with reasonably good

success.(*10, 35, 100, 101*) The UHBD predictions from Table 10 show that the FDPB

equation seems to detect the presence of some of the hydrogen bonds to the carboxyl

group of Asp76 in wild type T1. The predicted pKa for the carboxyl group in wild type

is about 1 unit lower than its intrinsic value, but about 2 units higher than experimental.

This suggests that some interactions are perturbing the group, but FDPB is not predicting

the full magnitude of the perturbations. An examination of the UHBD predictions on the

mutants studied here should help shed some light on where the program fails.

The predicted pKa for the carboxyl group in N9A is 3.4, which is higher than the

experimental value by about 1.7 units. When compared to the predicted value for wild

type, however, the prediction for the pKa in N9A is encouraging. The difference

between the measured carboxyl pKa in N9A and in wild type is about 1 unit, while the

difference in the predicted pKas between N9A and wild type is about 0.6 units. This

suggests that FDPB predicted the h-bond between the Asp76 carboxyl and Asn9 side

chain amide in the wild type protein. It also suggests that FDPB predicted the loss of

that h-bond in N9A almost equal to the experimental results. The loss of the h-bond to

the carboxyl of Asp76 is supported by the less favorable ΔG_{Bkgd} for N9A as seen in

Table 11. The ΔG_{IJ} is essentially unchanged from wild type, which is expected. The

ΔG_{Born} term is a little more favorable for the mutant, but is probably within error for the

calculations. This discussion did not take into account the new water molecules

Table 11: The Potentials of Interaction and Their Effects on the pKa of the Side Chain Carboxyl of Asp76 in RNase T1 Calculated with FDPB as Employed by UHBD.[a]

Protein Variant	ΔG (Born), kcal/mole	ΔpKa (Born)	ΔG (Bkgd), kcal/mole	ΔpKa (Bkgd)	ΔG (*ij*), kcal/mole	ΔpKa (*ij*)	ΔG (Total), kcal/mole	ΔpKa (Total)
T1 wild type	1.39	1.04	-2.08	-1.55	-0.97	-0.72	-1.66	-1.2
T1 N9A	1.21	0.9	-1.25	-0.93	-0.87	-0.65	-0.83	-0.6
T1 Y11F	1.39	1.04	-1.72	-1.28	-0.89	-0.66	-1.22	-0.91
T1 T91V	1.39	1.04	-2	-1.49	-0.95	-0.71	-1.56	-1.16
T1 N9AY11F	1.21	0.9	-0.91	-0.68	-0.68	-0.51	-0.38	-0.28
T1 N9AT91V	1.21	0.9	-1.17	-0.87	-0.76	-0.57	-0.72	-0.54
T1Y11FT91V	1.39	1.04	-1.62	-1.21	-0.86	-0.64	-1.09	-0.81
T1 TM	1.21	0.9	-0.82	-0.61	-0.65	-0.49	-0.26	-0.19

[a] The energy values reported here are from the application of FDPB to the crystal structure if RNase T1, 9RNT. The values are reported as they are described in eq 8. The $\Delta pKas$ are determined from

$\Delta G = 1.34 \cdot \Delta pKa$, where $1.34 = RT \cdot 2.303$ at T=293 K, and $\Delta pKa = pKa - 4.0$.

observed in the crystal structure for N9A. UHBD does not typically use crystallographic water molecules in its calculations. Had we analyzed the crystal structure of N9A with UHBD and utilized the water molecules, the ΔG_{Bkgd} might have been as favorable as wild type, while ΔG_{IJ} and ΔG_{Born} should have remained unchanged. The predicted ΔpKa would have probably been closer to the experimental value of 1 as well.

The predicted pKa for the carboxyl group in Y11F is 3.1, which agrees reasonably well with experiment. The disagreement is when we compare this prediction to the predicted pKa of the carboxyl group in wild type T1. The predicted ΔpKa for the carboxyl in Y11F from wild type is 0.3 units. This is considerably less than the experimental difference of about 3.0 units. The calculations do not predict that the removal of the hydrogen bonding group with the mutation of Tyr11 to Phe will significantly increase the pKa. An examination of the local structure shown in Figure 41 shows that the hydrogen added by CHARMM to the Tyr11 phenyl oxygen was added in a good position to form an h-bond (the donor-H-acceptor angle is ~160°). Perhaps the $\pi-$electron cloud of the phenyl ring contributes to the background component of the calculations so that the contribution of the h-bond is underestimated in the calculations. From Table 11, the ΔG_{Bkgd} for the mutant protein is less favorable than the ΔG_{Bkgd} for wild type T1, but not as much less favorable as the ΔG_{Bkgd} for N9A. This suggests that even though the h-bond formed between the Tyr11 side chain and Asp76 side chain is shorter than the h-bond between Asn9 and Asp76, it is a weaker h-bond. The ΔG_{IJ} and ΔG_{Born} terms are essentially unchanged from wild type T1, which is expected.
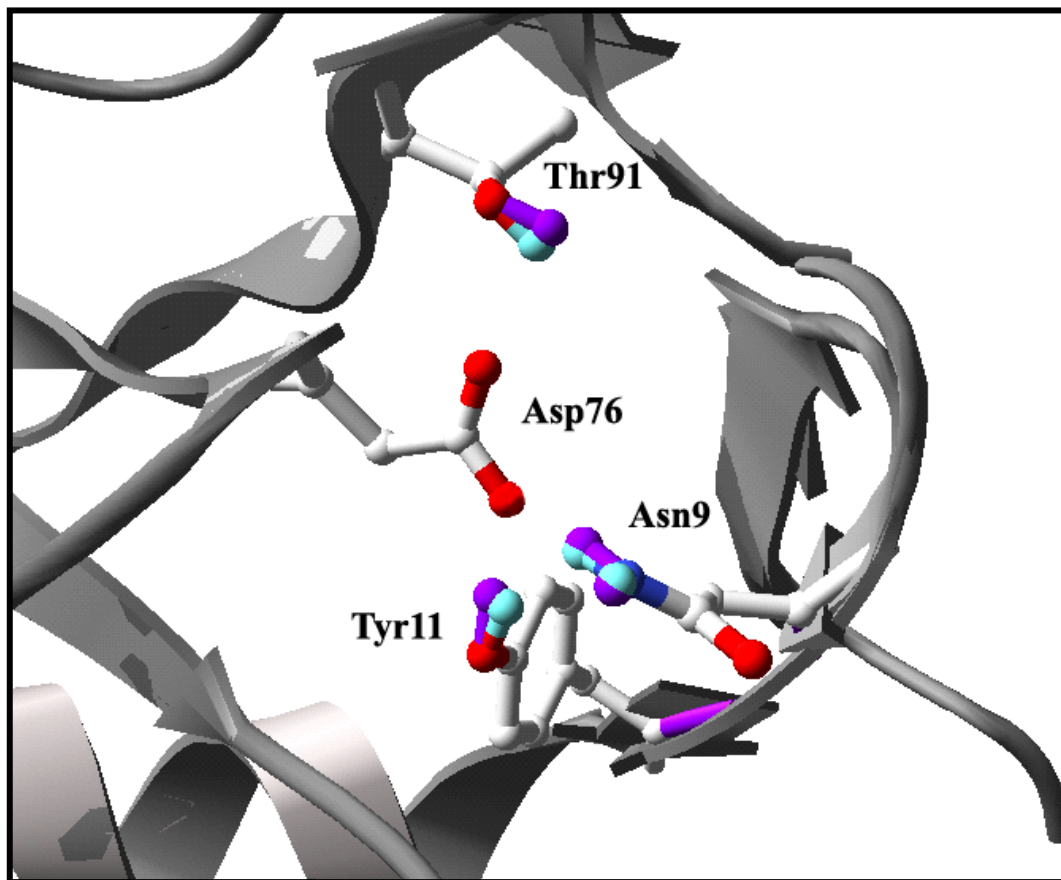
Figure 41: The crystal structure of 9RNT overlaid with 9RNT after the protons have been added by CHARMM. The purple atoms are hydrogens we added with Insight 2. Prior to running *pfis* on a protein we add the hydrogens with Insight. The aqua atoms are hydrogens added by CHARMM in preparation for the UHBD calculations. We see here that the locations of the atoms added by the two programs are close, but they do differ a bit.

The predicted pKa for the carboxyl group of Asp76 in T91V is 2.8, which clearly shows that no h-bond was predicted to exist between the carboxyl group and the side chain OH of Thr91in wild type T1. Figure 41 shows the positioning of the hydrogen on Thr91 by CHARMM. The donor-H-acceptor angle for the h-bond between Thr91 and Asp76 is about 109.5°. This may not be an optimal angle for h-bond formation, but is acceptable for forming a weak h-bond. Nevertheless, UHBD did not predict the presence of an h-bond between the carboxyl of Asp76 and the hydroxyl of Thr91. This is also shown by the results shown in Table 11, where the different ΔG components for T91V are unchanged from those of wild type T1.

The UHBD predictions for the double and triple mutants follow in line with the predictions for the single mutants. The differences from wild type for the ΔG_{Bkgd} and ΔG_{Born} components determined by UHBD for the single mutants are additive for each of the respective double and triple mutants. The ΔG_{IJ} component is not additive. As the calculations are performed, the ΔG_{Born} and ΔG_{Bkgd} terms are determined. Then the ΔG_{IJ} term is calculated based on the newly estimated pKa. As ΔG_{Born} and ΔG_{Bkgd} change from one mutant to another, there will be a change in ΔG_{IJ} to reflect the different initial pKa estimates. So, it is not surprising that as the predicted pKa changes from one mutant to another ΔG_{IJ} also changes.

The results of the mutational studies here show that the ΔG_{Bkgd} component underestimated the contributions of the polar groups, *i.e.* hydrogen bonding, to the total electrostatic potential on the carboxyl group of Asp76 in each case but one. The comparison of N9A, experimental to predicted, should be considered critically due to the

presence of new water molecules in the mutant structure. Placement of the hydrogens on the polar groups by CHARMM may be partially responsible for the underestimation. The electrostatic component, ΔG_{IJ}, of the potential changes a little for the different mutants, but this is easily explained by the changes observed for ΔG_{Born} and ΔG_{Bkgd}.

The ΔG_{Born} term appears underestimated as well. This is evidenced by the fact that there is no change in the ΔG_{Born} values between wild type T1 and Y11F or T91V or Y11F T91V or T1 TM. In each of these mutants we would expect the desolvation penalty for burial of the carboxyl group of Asp76 to increase due the increased hydrophobic environment around the group in the mutants. This is one weakness of UHBD. The program uses a simple grid for estimating the dielectric constant of the interior of the protein. Mutating a hydrophilic group to a hydrophobic group should alter the local dielectric environment of the carboxyl group and ΔG_{Born} should reflect this. By using a simple grid pattern, UHBD does not make the distinction between the mutants, thus the dielectric is not adjusted. Alternative approaches for estimating the micro-dielectric environment within the regions of a protein would greatly benefit applications of UHBD.(*102*)

Concluding Comments: We have shown that the three intramolecular hydrogen bonds to the side chain carboxyl of Asp76 in RNase T1 perturb its pKa by almost 6 pH units and that the contribution to the pKa of each h-bond is dependent on the group contributing the h-bond. The Solvent Accessibility Corrected Tanford Kirkwood equation is not an appropriate equation to estimate pKas for studies involving hydrogen-bonded groups. The Finite-Difference Solution to the Linearized Poisson Boltzmann

Equation does a better job of predicting the pKa of the Asp76 carboxyl in the wild type protein, and it predicted reasonably well the change in electrostatic potential around the carboxyl group in T1 N9A, although this apparent agreement should be critically examined. Its agreement with the experimental results of the other single mutants was not as good, as the program failed to detect the loss of one h-bond, T1 T91V, and failed to adjust the Born desolvation penalty of either mutant. Discussion of applying UHBD to the double and triple mutants follows very closely the discussion on the single mutants.

CHAPTER VI

CONCLUSIONS

The ionizable groups in a protein define the acid/base characteristics of that protein. The ionizable groups are important to protein chemists because of their influence on the conformational stability, solubility and catalytic activity of proteins. The pKa of an ionizable group is determined by the electrostatic potential on that group and is a reflection of its local environment, which can have dramatic affects on the group's measured pKa. In the event that a group has no other groups interacting with it and the group is fully solvent exposed, the measured pKa of that group is its intrinsic pKa.

The intrinsic pKas of the ionizable groups of proteins based on model peptides reported here are similar to the pKas from model compound data proposed by Nozaki and Tanford. There are a few differences, and some of the compounds Tanford used as models are questionable, which helps to explain most of the observed differences. We present a new set of intrinsic pKas based on the results from our model peptide system, which we feel are better representative of the intrinsic pKas of the ionizable groups of proteins.

The results from the fully hydrogen bonded carboxyl of Asp33 in RNase Sa show that removal of one h-bond can increase the pKa of a buried carboxyl group by about 1.5 units. These results show also that one hydrophobic group, in this case a Val side chain as compared to an Ala side chain, can further increase the pKa of a carboxyl by about 0.5 units.

The results from the fully hydrogen bonded carboxyl of Asp76 in RNase T1 show that the contribution of each h-bond to the perturbed pKa of the carboxyl is between 1 and about 3.3 units. The pKa of the carboxyl group in the double mutants is about the expected value based on the results from the single mutants. The pKa of the carboxyl group in the triple mutant, in which all three hydrogen-bonding partners to the Asp76 carboxyl are mutated to hydrophobic groups, is 6.4, almost 6 units higher than the value in wild type T1.

The Solvent Accessibility Corrected Tanford Kirkwood equation is not an appropriate equation to estimate pKas for studies involving hydrogen bonded ionizable groups. Its main usefulness is to describe the effects of charge-charge interactions on ionizable groups. The Finite-Difference Solution to the Linearized Poisson Boltzmann Equation, as implemented in UHBD, does a better job of predicting the pKas of the buried ionizable groups in a protein. The results generated with UHBD should be examined carefully however. We have shown that the necessity to add hydrogens to a protein's crystal structure can have a large impact on the predicted pKas of ionizable groups that form h-bonds. We have also shown that the estimate of the Born desolvation energies on ionizable groups may not be accurately determined since replacing local polar groups with hydrophobic groups have little or no effect on the calculated ΔG_{Born}.

This work is a start in investigating a subject that has been neglected by researchers over the years, *i.e.* the contribution of hydrogen bonding to the perturbation of pKas. Considerable work is yet to be done, however, before any conclusions may be drawn.

REFERENCES

1.   Sorensen, S. P. L., Hoyrup, M., Hempel, J., and Palitzsch, S. (1917) Studies on proteins. II. The capacity of egg-albumin to combine with acids or bases, *Compt. rend. trav. lab. Carlsberg 12*, 68-163.

2.   Rawn, J. D. (1983) *Biochemistry*, 1st ed., Harper and Row, New York, NY.

3.   Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation, *Adv. Protein Chem. 24*, 1-95.

4.   Cohn, E. J., and Edsall, J. T. (1965) in *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions. Am. Chem. Soc. Monograph No. 90* pp 569-585, Hafner Publishing Co., New York, NY.

5.   Bartlett, G. J., Porter, C. T., Borkakoti, N., and Thornton, J. M. (2002) Analysis of catalytic residues in enzyme active sites, *J. Mol. Biol. 324*, 105-21.

6.   Tanford, C. (1962) The interpretation of hydrogen ion titration curves of proteins, *Advan. Protein Chem. 17*, 69-165.

7.   Nozaki, Y., and Tanford, C. (1967) Examination of titration behavior, *Methods Enzymol. 11*, 715-734.

8.   Harris, T. K., and Turner, G. J. (2002) Structural basis of perturbed pKa values of catalytic groups in enzyme active sites, *IUBMB Life 53*, 85-98.

9.   Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A., and Pace, C. N. (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa, *Protein Sci. 10*, 1206-15.

10.   Laurents, D. V., Huyghues-Despointes, B. M., Bruix, M., Thurlkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S., Rico, M., Briggs, J. M., Antosiewicz, J. M., Scholtz, J. M., and Pace, C. N. (2003) Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2), *J. Mol. Biol. 325*, 1077-92.

11. Spitzner, N., Lohr, F., Pfeiffer, S., Koumanov, A., Karshikoff, A., and Ruterjans, H. (2001) Ionization properties of titratable groups in ribonuclease T1 I. pKa values in the native state determined by two-dimensional heteronuclear NMR spectroscopy, *Eur. Biophys. J. 30*, 186-197.

12. Forsyth, W. R., Antosiewicz, J. M., and Robertson, A. D. (2002) Empirical relationships between protein structure and carboxyl pKa values in proteins, *Proteins 48*, 388-403.

13. Edgcomb, S. P., and Murphy, K. P. (2002) Variability in the pKa of histidine side-chains correlates with burial within proteins, *Proteins 49*, 1-6.

14. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme, *Biochemistry 29*, 2403-8.

15. Schmidt, D. E., Jr., and Westheimer, F. H. (1971) PK of the lysine amino group at the active site of acetoacetate decarboxylase, *Biochemistry 10*, 1249-53.

16. Highbarger, L. A., Gerlt, J. A., and Kenyon, G. L. (1996) Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115, *Biochemistry 35*, 41-6.

17. Hebert, E. J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K. S., Dauter, Z., and Pace, C. N. (1998) Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1, *Biochemistry 37*, 16192-200.

18. Trevino, S. R., Gokulan, K., Newsom, S., Thurlkill, R. L., Shaw, K. L., Mitkevich, V. A., Makarov, A. A., Sachettini, J. C., Scholtz, J. M., Pace, C. N. (2005) Asp79 makes a large, unfavorable contribution to the stability of RNase Sa, *J. Mol. Biol. (In Press)*.

19. Langsetmo, K., Fuchs, J. A., and Woodward, C. (1991) The conserved, buried aspartic acid in oxidized Escherichia coli thioredoxin has a pKa of 7.5. Its titration produces a related shift in global stability, *Biochemistry 30*, 7603-9.

20. Langsetmo, K., Fuchs, J. A., Woodward, C., and Sharp, K. A. (1991) Linkage of thioredoxin stability to titration of ionizable groups with perturbed pKa, *Biochemistry 30*, 7609-14.

21.   Wilson, N. A., Barbar, E., Fuchs, J. A., and Woodward, C. (1995) Aspartic acid 26 in reduced Escherichia coli thioredoxin has a pKa > 9, *Biochemistry 34*, 8931-9.

22.   Jeng, M. F., and Dyson, H. J. (1996) Direct measurement of the aspartic acid 26 pKa for reduced Escherichia coli thioredoxin by 13C NMR, *Biochemistry 35*, 1-6.

23.   Dyson, H. J., Jeng, M. F., Tennant, L. L., Slaby, I., Lindell, M., Cui, D. S., Kuprin, S., and Holmgren, A. (1997) Effects of buried charged groups on cysteine thiol ionization and reactivity in Escherichia coli thioredoxin: structural and functional characterization of mutants of Asp 26 and Lys 57, *Biochemistry 36*, 2622-36.

24.   Stites, W. E., Gittis, A. G., Lattman, E. E., and Shortle, D. (1991) In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core, *J. Mol. Biol. 221*, 7-14.

25.   Garcia-Moreno, B., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S., and Stites, W. E. (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein, *Biophys. Chem. 64*, 211-24.

26.   Dwyer, J. J., Gittis, A. G., Karp, D. A., Lattman, E. E., Spencer, D. S., Stites, W. E., and Garcia-Moreno, E. B. (2000) High apparent dielectric constants in the interior of a protein reflect water penetration, *Biophys. J. 79*, 1610-20.

27.   Fitch, C. A., Karp, D. A., Lee, K. K., Stites, W. E., Lattman, E. E., and Garcia-Moreno, E. B. (2002) Experimental pK(a) values of buried residues: analysis with continuum methods and role of water penetration, *Biophys. J. 82*, 3289-304.

28.   Quirk, D. J., Park, C., Thompson, J. E., and Raines, R. T. (1998) His...Asp catalytic dyad of ribonuclease A: conformational stability of the wild-type, D121N, D121A, and H119A enzymes, *Biochemistry 37*, 17958-64.

29.   Quirk, D. J., and Raines, R. T. (1999) His ... Asp catalytic dyad of ribonuclease A: histidine pKa values in the wild-type, D121N, and D121A enzymes, *Biophys. J. 76*, 1571-9.

30.   Tanford, C. (1955) The electrostatic free energy of globular protein ions in aqueous salt solution, *J. Phys. Chem. 59*, 788-93.

31. Tanford, C., and Kirkwood, J. G. (1957) Theory of protein titration curves. I. General equations for impenetrable spheres, *J. Am. Chem. Soc. 79*, 5333-9.

32. Shire, S. J., Hanania, G. I., and Gurd, F. R. (1974) Electrostatic effects in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin, *Biochemistry 13*, 2967-74.

33. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability, *Biochemistry 38*, 8138-49.

34. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol. 55*, 379-400.

35. Huyghues-Despointes, B. M., Thurlkill, R. L., Daily, M. D., Schell, D., Briggs, J. M., Antosiewicz, J. M., Pace, C. N., and Scholtz, J. M. (2003) pK values of histidine residues in ribonuclease Sa: effect of salt and net charge, *J. Mol. Biol. 325*, 1093-105.

36. Kao, Y. H., Fitch, C. A., Bhattacharya, S., Sarkisian, C. J., Lecomte, J. T., and Garcia-Moreno, E. B. (2000) Salt effects on ionization equilibria of histidines in myoglobin, *Biophys. J. 79*, 1637-54.

37. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) Prediction of pH-dependent properties of proteins, *J. Mol. Biol. 238*, 415-36.

38. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1996) The determinants of pKas in proteins, *Biochemistry 35*, 7819-33.

39. Davis, M. E., Madura, J. D., Luty, B. A., and McCammon, J. A. (1991) Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program, *Computer Physics Communications 62*, 187-97.

40. Lee, K. K., Fitch, C. A., and Garcia-Moreno, E. B. (2002) Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein, *Protein Sci. 11*, 1004-16.

41.   Forsyth, W. R., Gilson, M. K., Antosiewicz, J., Jaren, O. R., and Robertson, A. D. (1998) Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain, *Biochemistry 37*, 8643-52.

42.   Forsyth, W. R., and Robertson, A. D. (2000) Insensitivity of perturbed carboxyl pK(a) values in the ovomucoid third domain to charge replacement at a neighboring residue, *Biochemistry 39*, 8067-72.

43.   Hebert, E. J., Grimsley, G. R., Hartley, R. W., Horn, G., Schell, D., Garcia, S., Both, V., Sevcik, J., and Pace, C. N. (1997) Purification of ribonucleases Sa, Sa2, and Sa3 after expression in Escherichia coli, *Protein Expr. Purif. 11*, 162-8.

44.   Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W., and Grimsley, G. R. (1998) Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3, *J. Mol. Biol. 279*, 271-86.

45.   Takano, K., Scholtz, J. M., Sacchettini, J. C., and Pace, C. N. (2003) The contribution of polar group burial to protein stability is strongly context-dependent, *J. Biol. Chem. 278*, 31790-5.

46.   Sato, K., and Egami, F. (1957) [The specificity of T1 ribonuclease.], *C R Seances Soc. Biol. Fil. 151*, 1792-6.

47.   Pace, C. N., Laurents, D. V., and Thomson, J. A. (1990) pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1, *Biochemistry 29*, 2564-72.

48.   Plaza del Pino, I. M., Pace, C. N., and Freire, E. (1992) Temperature and guanidine hydrochloride dependence of the structural stability of ribonuclease T1, *Biochemistry 31*, 11196-202.

49.   Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci. 8*, 1843-9.

50.   Shirley, B. A., Stanssens, P., Hahn, U., and Pace, C. N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1, *Biochemistry 31*, 725-32.

51. Yu, Y., Makhatadze, G. I., Pace, C. N., and Privalov, P. L. (1994) Energetics of ribonuclease T1 structure, *Biochemistry 33*, 3312-9.

52. Sevcik, J., Dauter, Z., Lamzin, V. S., and Wilson, K. S. (1996) Ribonuclease from Streptomyces aureofaciens at atomic resolution, *Acta Crystallogr. D. Biol. Crystallogr. 52*, 327-44.

53. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *Journal of Applied Crystallography 24*, 945-9.

54. Martinez-Oyanedel, J., Choe, H. W., Heinemann, U., and Saenger, W. (1991) Ribonuclease T1 with free recognition and catalytic site: crystal structure analysis at 1.5 A resolution, *J. Mol. Biol. 222*, 335-52.

55. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis 18*, 2714-23.

56. Giletto, A., and Pace, C. N. (1999) Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1, *Biochemistry 38*, 13379-84.

57. Shirley, B. A., and Laurents, D. V. (1990) Purification of recombinant ribonuclease T1 expressed in Escherichia coli, *J. Biochem. Biophys. Methods 20*, 181-8.

58. Shaw, K. L. (2000) Reversing the net chage of RNase Sa, in *Biochemistry and Biophysics* pp 68, Texas A&M University, College Station, TX.

59. (2000) in *Novabiochem Catalog* (Dorner, B., and White, P., Ed.) pp S1-S120.

60. Chan, W. C., White, P. D. (2000) Fmoc solid phase peptide synthesis: A Practical Approach, in *The Practical Approach Series* (Hames, B. D., Ed.) pp 346, Oxford University Press, Oxford.

61. Huang, Y., and Bolen, D. W. (1995) Probes of energy transduction in enzyme catalysis, *Methods Enzymol. 259*, 19-43.

62. Rossotti, F. J. C., and Rossotti, H. (1965) Potentiometric titrations using Gran plots-a textbook omission, *J. Chem. Ed. 42*, 375-8.

63.   Markley, J. L. (1975) Observation of Histidine residues in proteins by means of nuclear magnetic resonance spectroscopy, *Accounts of Chemical Research 8*, 70-80.

64.   Grimsley, G. R., Huyghues-Despointes, B. M. P., Pace, C. N., and Scholtz, J. M. (2004) in *Purifying Proteins for Proteomics* (Simpson, R. J., Ed.) pp 535-566, Cold Srping Harbor Laboratory Press, Cold Spring Harbor, NY.

65.   Pace, C. N., and Scholtz, J. M. (1997) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 299-321, Oxford University Press, London.

66.   Wyman, J., Jr. (1964) Linked Functions and Reciprocal Effects in Hemoglobin: A Second Look, *Adv. Protein Chem. 19*, 223-86.

67.   Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci. 4*, 2411-23.

68.   Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-42.

69.   Harris, L. J. (1923) The titration of amino-and carboxyl-groups in amino-acids, polypeptides, etc. PartsI-II.-Investigations with aqueous solutions., *Proceedings of the Royal Society of London, Series B 95*, 440-484.

70.   Greenstein, J. P. (1931) Studies on the peptides of tervalent amino acids. I. Titration constants of histidyl-histidine and of aspartyl-aspartic acid, *J. Biol. Chem. 93*, 479-94.

71.   Ellenbogen, E. (1952) Dissociation constants of peptides. I. A survey of the effect of optical configuration, *J. Am. Chem. Soc. 74*, 5198-201.

72.   Tanokura, M., Tasumi, M., and Miyazawa, T. (1976) Proton nuclear magnetic resonance studies of histidine-containing di- and tripeptides. Estimation of the effects of charged groups on the pKa value of the imidazole ring, *Biopolymers 15*, 393-401.

73.   Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896-903.

74.   Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., and Baldwin, R. L. (1990) Relative helix-forming tendencies of nonpolar amino acids, *Nature 344*, 268-70.

75.   Ryklan, L. R., and Schmidt, C. L. A. (1944) The apparent dissociation constant of homocysteine and homocystine, *Archives of Biochem. 5*, 89-98.

76.   Kortemme, T., and Creighton, T. E. (1995) Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family, *J. Mol. Biol. 253*, 799-812.

77.   Zief, M., and Edsall, J. T. (1937) Studies in the physical chemistry of amino acids, peptides and related substances. IX. The dissociation constants of some amino acid derivatives, *J. Am. Chem. Soc. 59*, 2245-8.

78.   Gurd, F. R. N., Keim, P., Glushko, V.G., Lawson, P.J., Marshall, R.C., Nigen, A.M., Vigna, R.A. (1972) in *Chemistry and Biology of Peptides* (Meienhofer, J., Ed.) pp 45-49, Ann Arbor Science Publishers, Ann Arbor, MI.

79.   Keim, P., Vigna, A., Marshall, R. C., and Gurd, F. R. N. (1973) Coordination complexes and catalytic properties of proteins and related substances. .52. Carbon-13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of aliphatic amino acids, *J. Biol. Chem. 248*, 6104-13.

80.   Keim, P., Vigna, R. A., Morrow, J. S., Marshall, R. C., and Gurd, F. R. N. (1973) Carbon 13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of serine, threonine, aspartic and glutamic acids, asparagine, and glutamine, *J. Biol. Chem. 248*, 7811-18.

81.   Keim, P., Vigna, R. A., Nigen, A. M., Morrow, J. S., and Gurd, F. R. N. (1974) Carbon-13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of methionine, proline, arginine, and lysine, *J. Biol. Chem. 249*, 4149-56.

82. Richarz, R., and Wuethrich, K. (1978) Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH, *Biopolymers 17*, 2133-41.

83. Bundi, A., and Wuethrich, K. (1979) Proton NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH, *Biopolymers 18*, 285-97.

84. Bundi, A., and Wuethrich, K. (1979) Use of amide proton NMR titration shifts for studies of polypeptide conformation, *Biopolymers 18*, 299-311.

85. Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. (1999) The packing density in proteins: standard radii and volumes, *J. Mol. Biol. 290*, 253-66.

86. Pace, C. N. (1995) Evaluating contribution of hydrogen bonding and hydrophobic bonding to protein folding, *Methods Enzymol. 259*, 538-54.

87. Fauchere, J. L., and Pliska, V. (1983) Hydrophobic parameters p of amino acid side chains from the partitioning of N-acetyl-amino acid amides, *Eur. J. Med. Chem. 18*, 369-75.

88. Doig, A. J., and Sternberg, M. J. (1995) Side-chain conformational entropy in protein folding, *Protein Sci. 4*, 2247-51.

89. Green, S. M., Meeker, A. K., and Shortle, D. (1992) Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state, *Biochemistry 31*, 5717-28.

90. Byrne, M. P., Manuel, R. L., Lowe, L. G., and Stites, W. E. (1995) Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease, *Biochemistry 34*, 13949-60.

91. Main, E. R., Fulton, K. F., and Jackson, S. E. (1998) Context-dependent nature of destabilizing mutations on the stability of FKBP12, *Biochemistry 37*, 6145-53.

92. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., and Sarai, A. (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants, *Nucleic Acids Res. 32*, D120-1.

93.   Myers, J. K., and Pace, C. N. (1996) Hydrogen bonding stabilizes globular proteins, *Biophys. J. 71*, 2033-9.

94.   (1997) in *Ribonucleases: Structures and Functions* (D'Alessio, G., Riordan, James, Ed.) pp 670-672, Academic Press, NY.

95.   Laurents, D., Perez-Canadillas, J. M., Santoro, J., Rico, M., Schell, D., Pace, C. N., and Bruix, M. (2001) Solution structure and dynamics of ribonuclease Sa, *Proteins 44*, 200-11.

96.   Swint-Kruse, L., and Robertson, A. D. (1995) Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry 34*, 4724-32.

97.   Sundd, M., Iverson, N., Ibarra-Molero, B., Sanchez-Ruiz, J. M., and Robertson, A. D. (2002) Electrostatic interactions in ubiquitin: stabilization of carboxylates by lysine amino groups, *Biochemistry 41*, 7586-96.

98.   Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M., and Makhatadze, G. I. (1999) Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface, *Biochemistry 38*, 16419-23.

99.   Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering, *J. Mol. Biol. 327*, 1135-48.

100.   Pace, C. N., Huyghues-Despointes, B. M., Briggs, J. M., Grimsley, G. R., and Scholtz, J. M. (2002) Charge-charge interactions are the primary determinants of the pK values of the ionizable groups in Ribonuclease T1, *Biophys. Chem. 101-102*, 211-9.

101.   Lee, K. K., Fitch, C. A., Lecomte, J. T., and Garcia-Moreno, E. B. (2002) Electrostatic effects in highly charged proteins: salt sensitivity of pKa values of histidines in staphylococcal nuclease, *Biochemistry 41*, 5656-67.

102.   Li, H., Robertson, A. D., and Jensen, J. H. (2004) The determinants of carboxyl pKa values in turkey ovomucoid third domain, *Proteins 55*, 689-704.

103.   Schutz, C. N., and Warshel, A. (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models?, *Proteins 44*, 400-17.

104.   Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera-a visualization system for exploratory research and analysis, *J. Comp. Chem. 25*, 1605-12.

105.   Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comp. Chem. 4*, 187-217.

106.   Nielsen, J. E., and McCammon, J. A. (2003) On the evaluation and optimization of protein X-ray structures for pKa calculations, *Protein Sci. 12*, 313-26.

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

108.   Shortle, D. (1995) Staphylococcal nuclease: a showcase of m-value effects, *Adv. Protein Chem. 46*, 217-47.

109.   Langhorst, U., Backmann, J., Loris, R., and Steyaert, J. (2000) Analysis of a water mediated protein-protein interactions within RNase T1, *Biochemistry 39*, 6586-93.

110.   Langhorst, U., Loris, R., Denisov, V. P., Doumen, J., Roose, P., Maes, D., Halle, B., and Steyaert, J. (1999) Dissection of the structural and functional role of a conserved hydration site in RNase T1, *Protein Sci. 8*, 722-30.

APPENDIX I

DS2000

DS2000 is a cell line construct of David Schell. The cell line was constructed by mating *E. coli* ATCC® 55244 (*tonA ptr3 ΔphoA ΔE15 Δ(argF-lac)169 degP41 ΔompT*), with the F-positive *E. coli* RY2700, courtesy of Ry Young, Department of Biochemistry and Biophysics, Texas A&M University, from which the F' episome (*lacI^q Tet^R*) was transferred. This mating produced a cell line resistant to kanamycin and tetracycline, and deficient of periplasmic proteases, with the final genotype: *tonA ptr3 ΔphoA ΔE15 Δ(argF-lac)169 degP41 ΔompT lacI^q Tet^R* . With DS2000 we are able to express proteins that are destabilized and export them to the periplasm of the cell with confidence that the proteins will not be degraded in the periplasm by proteases. We have mass spectral evidence that shows that proteins expressed in DS2000 are of the correct mass, indicating that cleavage of the *phoA* leader sequence is properly occurring sometime during the prep.

APPENDIX II

ASSUMPTIONS TO THE TECHNIQUES

There are three assumptions to applying potentiometric difference titrations that need to be addressed. One assumption is that the ionizable group of interest, IGI, is **not** so critical to the conformational stability of the protein that the mutation causes the mutant to be unfolded under nominal conditions of temperature and salt concentration in the pH range of interest. Since our interest is the pKa of the IGI in the native state of your favorite protein, YFP, we must ensure that the mutant protein, as well as the wild type protein, is in its native conformation throughout the pH range of the potentiometric titration. One way to test this is to determine the thermal melting temperature of both variants as a function of pH.

A second assumption is that a mutation does not dramatically alter the structure of the protein. Any changes in the structure of the protein caused by a mutation might change the local environment of one or more ionizable groups, leading to shifts in the pKas of those groups. A shift in the pKa of a group other than the IGI probably would not result in a full proton difference between the binding curves, because the group is still present in both variants of the protein. Shifts in the pKas of several groups, however, might affect the difference titration resulting in protons, or partial protons, bound or released at various pH values. This can result in a distorted difference plot or even in a plot that is not interpretable.

The final assumption is that the IGI has limited interactions with other ionizable groups in the native state of the protein. Any such interactions would probably perturb the pKa of the IGI, but would also perturb the pKa of the other group or groups as well. Mutating the IGI to a non-ionizable group could remove this perturbation resulting in a shift in the pKa of the other groups so that the analysis would detect both the absence of the IGI and the shift in pKa of the other interacting groups. As stated in the previous paragraph, any shifts in the pKa of other groups could alter the shape of the resulting difference curve to yield more or less than one proton difference, and could do so at various regions of pH thus making the analysis difficult.

There are some exceptions to the above noted assumptions that can be taken into account. For example, in the event that as Asp side chain is interacting with a Lys side chain resulting in a lower pKa for the carboxyl group and a higher pKa for the ε-amino group the difference titration may still be performed for either of these groups. The rationale is that the pKa of the remaining group in the ΔIGI mutant protein would still be far enough removed from the pKa of the IGI in the YFP wild type that the difference curve in the region near the expected IGI proton binding would not be affected. The results however should be critically analyzed and care should be exercised in the interpretation of the results. Arguments for the second and third assumptions are also applicable to Tanford-Wyman Analysis.

VITA

Name: Richard Lee Thurlkill

Address: Medical Biochemistry and Genetics
Texas A&M University System Health Science Center
440 Reynolds Medical Building
1114 TAMU
College Station, TX 77843-1114
tel 979-845-6834
fax 979-847-9481
rich@pauling.tamu.edu

Education:
B.S., Louisiana Tech University, Ruston, LA, 1986 (Chemistry)
PhD, Texas A&M University, College Station, TX, expected Fall, 2005 (Biochemistry)

Publications:
1. Grimsley, GR, Shaw, KL, Fee, LR, Alston, RW, Huyghues-Despointes, BM, Thurlkill, RL, Scholtz, JM, and Pace, CN. *Increasing protein stability by altering long-range coulombic interactions.* Protein Sci, 1999. **8**(9): p. 1843-9.
2. Holtman, CK, Thurlkill, R, and Pettigrew, DW. *Unexpected presence of defective gIpR alleles in various strains of Escherichia coli.* J Bacteriol, 2001. **183**(4): p. 1459-61.
3. Laurents, DV, Huyghues-Despointes, BM, Bruix, M, Thurlkill, RL, Schell, D, Newsom, S, Grimsley, GR, Shaw, KL, Trevino, S, Rico, M, Briggs, JM, Antosiewicz, JM, Scholtz, JM, and Pace, CN. *Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pl=3.5) and a basic variant (pl=10.2).* J Mol Biol, 2003. **325**(5): p. 1077-92.
4. Huyghues-Despointes, BM, Thurlkill, RL, Daily, MD, Schell, D, Briggs, JM, Antosiewicz, JM, Pace, CN, and Scholtz, JM. *pK values of histidine residues in ribonuclease Sa: effect of salt and net charge.* J Mol Biol, 2003. **325**(5): p. 1093-105.
5. Thurlkill, RL, Scholtz, JM, Pace, CN, Cross, DA. *The pKa of fentanyl varies with temperature: Implications for acid-base management during extremes of body temperature.* J Cardiothor Vasc An, 2005. (In press).
6. Trevino, SR, Gokulan, K, Newsom, S, Thurlkill, RL, Shaw, KL, Mitkevich, VA, Makarov, AA, Sachettini, JC, Scholtz, JM, Pace, CN. *Asp79 makes a large, unfavorable contribution to the stability of RNase Sa.* J Mol Biol, 2005. (In press).