Characterization of protein microstructure by various chromatographic techniques

Lakshmi P. Pathange

Dissertation submitted to Faculty of

Virginia Polytechnic and State University

In partial fulfillment of the requirements of the requirements for the degree of the

Doctor of Philosophy

in

Biological Systems Engineering

Dr. Chenming (Mike) Zhang (committee chair),

Dr. David R. Bevan,

Dr. Timothy J. Larson,

Dr. John S. Cundiff,

Dr. David H. Vaughan

Feb., 28, 2007

Blacksburg, Virginia

Keywords: protein microstructure, metal affinity, ion exchange, chromatography

Characterization of protein microstructure by various chromatographic techniques

Lakshmi P. Pathange

Abstract

Due to the rising health care costs and with the advent of biogenerics, there is a growing demand to develop new and reliable techniques to characterize proteins and biopharmaceuticals. In addition, characterization aids in understanding the intricate relationship between a protein's structure and its function. To address this challenge, two protein structural parameters, 1) amino acid surface area and 2) amino acid microstructure, were chosen to be investigated. Two chromatographic techniques, 1) ion exchange chromatography (IEC) and 2) immobilized metal affinity chromatography (IMAC), were used to characterize the above-mentioned protein structure parameters.

The model protein chosen for our work is T4 lysozyme. The protein consists of 164 amino acids with molecular weight ~ 18 kD. SYBYL 7.1 software was used to generate *in silico* point mutants. Two categories of protein variants (point mutants) were generated using site-directed protein mutagenesis. The goal for generating point mutants was to obtain mutants that vary in the two structural parameters. The first category point mutants vary in the surface accessibility of a surface accessible histidine residue. The second category point mutants predominantly vary in protein net charge and the amino acid microstructure. In total, seventeen point mutants were generated: 1) category I consists of seven variants that vary predominantly in their histidine surface accessibility, and were obtained by replacing a charged amino acid residue at different locations on the surface of the protein molecule, and 2) category II consists of ten variants that vary in

both net charge and charge distribution were obtained by replacing charged and neutral amino acid residues at different locations (different microenvironments) on the protein surface.

PCR technique was used to generate the point mutants. Gene and protein sequencing were employed to confirm the veracity of point mutation. CD and Lysozyme activity assays were performed to determine whether or not the 3D structure of all the protein variants was intact. Zonal analysis was used to obtain the binding strength values of all seventeen variants in IMAC with copper as the immobilized metal ions, and gradient elution method was used to obtain the relative retention times (rRT) values of all the variants in IEC.

The seven lysozyme variants generated in category I each contains one surface histidine residue. In IMAC, there is a correlation between the surface accessibility of the lone surface histidine and the protein's binding strength with R^2= 0.76. In IEC, the correlation between the protein's microstructure, which predominantly consists the surface accessibility of the histidine residue, and the protein's retention times was R^2= 0.95. However, there were few outlier variants (e.g. variant K83H) which did not follow the correlations. The variations presented by few outlier variants can be attributed to the presence of intramolecular bonds, which restrict the mobility of the amino acid side chains and subsequently hinder the specific interaction between the amino acid residue and chromatographic media.

For category II variants, short and medium range charge perturbations around the sole histidine residue in T4 lysozyme were engineered within 15 Å distance of histidine. There was a strong correlation (R^2 = 0.96) between the theoretical $\Delta\Delta G_{Elec}$ values,

calculated using simple Coulomb's law, and the experimental $\Delta\Delta G_B$ values, which were obtained by measuring the protein binding strength values using IMAC. Similar correlation (R^2= 0.93) was obtained between the change in net charge (-2 to +2 units) and the relative retention times in IEC. Similarly, there were few variants (e.g. S136K, R76D) that did not follow the trends. The deviations of the few outlier variants can be attributed to the presence of unique microstructure effects around the histidine residue. These microstructure effects were quantified in IMAC as $\Delta\Delta G_{Micro}$, and in IEC they were quantified by the change in rRT values.

In summary, all seventeen variants had different binding strengths and rRT values indicating the variation in the protein structure around the histidine residue. Our work reveals that it is possible to capture the microstructural effects of a protein through the combination of protein molecular modeling and simple chromatographic experiments.

Attribution

I am very grateful to my academic advisor, Dr. Chenming (Mike) Zhang, who has always inspired me to give only the best of my abilities to complete the task at hand. He has been always a source of constant encouragement in completing my Ph.D. work. In addition, I would like to thank him for his considerable patience, time and guidance through out my graduate study. I sincerely thank Dr. David Bevan, who as served on my Ph.D. committee and has been the sole reason for my endeavor to pursue graduate work at Virginia Tech. I would like to thank all the committee members without whose constant encouragement and their belief in me has always allowed me to focus my work without interruptions. Specifically, I would like to thank Dr. Timothy Larson and Ms. Janet Donahue of the Biochemistry Dept., for their patience, time and careful guidance during the initial stages of my Ph.D. work. Special thanks to Dr. John Cundiff and Dr. David Vaughan for their constant encouragement to complete my dissertation. I also want to thank Dr. Robert White of the Biochemsitry Dept., for his invaluable suggestions during my Ph.D. I also want to thank Dr. Mostaghimi, Head of the Biological Systems Engineering Dept., for his constant encouragement and providing me with valuable resources that were vital to the completion of my doctorial dissertation. I gratefully acknowledge computing time on the System X supercomputer in the terascale computing facility at Virginia Tech. I also want to thank Jefferess Memorial Trust for partially funding this work.

Acknowledgements

I humbly attribute the completion of my Ph.D. degree to the grace and the kind blessings of God almighty. All the credit goes to him, who has given me strength, power and courage to pursue and complete my Ph.D. degree. I would like to acknowledge my parents who have always stood beside me and constantly encouraged me to complete my degree. I also want to thank all my colleagues in my lab, Scott Buswell, Ray Lilee, Chris Holler and Brad Matanin for their suggestions and cooperation. I would like to thank Aubrey Murden, Amy Eagan, Tom Glass and Kim Harich for their valuable contribution in completing my dissertation. Last but not the least I want to acknowledge all the folks who have directly or indirectly have inspired me in ways that can be both regrettable and commendable.

Table of contents

List of Tables

Chapter III

Chapter IV

Chapter V

Appendix E

List of Figures

Chapter III

Chapter IV

Chapter V

Appendix E

Chapter I: Introduction

Characterization of protein structure is important for the development of biotherapeuticals.[1, 2] The rapid growth of the biopharmaceutical industry has resulted in a greater demand to develop/evolve new and reliable techniques to characterize recombinant therapeutic proteins. For example, due to the lack of proper characterization techniques and rising health care costs in U.S., there is an intense debate over the need to use 'biogenerics' or 'off-patent biotech products'.

Unlike small compounds, biopharmaceuticals are an ensemble of large, flexible, and complex molecules. As a result, the process of producing biopharmaceuticals is a complex process.[3, 4] The inherent challenges associated with biopharmaceutical production are 1) the type of the extraneous expression host, which expresses proteins that are a heterogeneous mixture, consisting of variants of the natural protein,[5] 2) the type of manufacturing process, which is very complex compared to conventional pharmaceutical products, involving specialized techniques such as genetic engineering, fermentation, extraction and purification,[6, 7] and 3) the need to conduct expensive clinical trials to determine the efficacy and toxicity of the biopharmaceuticals.[1] For example, the biopharmaceutical IFN-β_{1a}, used for treating multiple sclerosis, when produced at two different manufacturing sites (with the same process conditions), still elicited different immunogenic response in humans during clinical trials.[8] Evidently, such results have reinforced the argument that "process is the product"[6] in the biopharmaceutical industry. Consequently, characterizing these biopharmaceuticals is very important. Therefore, there is a urgent need to augment the current array of analytical techniques either by

developing new techniques or extending the currently available techniques such that they can characterize subtle structural changes taken place in proteins.[4, 5, 7]

Currently, several analytical techniques (e.g. ion exchange and gel chromatography, mass spectrometry, gel electrophoresis) are used for protein characterization based on physiochemical parameters (e.g. size, charge, molecular weight).[7] However, none of them are capable of characterizing[4] and predicting protein's biological and clinical properties.[8] For example, due to the difference in the manufacturing process there was a notable difference in the immunogenic properties of the biopharmaceutical IFN-β_{1a}, which is administered in treating multiple sclerosis. In order to address this problem, efforts should be focused on developing new separation techniques that are based on the structural parameters of the protein.

Objectives: The overall goal of my Ph.D. thesis is to determine and understand parameters that aid in elucidating protein structure-function relationship [12, 13]. The two parameters chosen were 1) relative surface area (at same pI) and 2) variation in the microenvironment of different surface accessible amino acid residues (change in net charge resulting in different pI). Based on these two parameters two categories of structural variants were generated via site-directed protein mutagenesis. Here, chromatographic techniques such as IMAC and ion exchange chromatography (IEC) were employed to quantify these parameters. Thus, the main hypothesis of our experiments was to analyze subtle variation in the protein charge including its position on the protein's surface.

References

1. Koths, K., Recombinant Proteins for Medical Use - the Attractions and Challenges. *Current Opinion in Biotechnology* **1995,** 6, (6), 681-687.

2. Kleinberg, M.; Mosdell, K. W., Current and future considerations for the new classes of biologicals. *American Journal of Health-System Pharmacy* **2004,** 61, (7), 695-708.
3. Schellekens, H., When biotech proteins go off-patent. *Trends in Biotechnology* **2004,** 22, (8), 406-410.
4. Schellekens, H., Follow-on biologics: challenges of the 'next generation'. *Nephrology Dialysis Transplantation* **2005,** 20, 31-36.
5. Chicz, R. M.; Regnier, F. E., Immobilized-Metal Affinity and Hydroxyapatite Chromatography of Genetically Engineered Subtilisin. *Analytical Chemistry* **1989,** 61, (15), 1742-1749.
6. Herrera, S., Biogenerics standoff. *Nature Biotechnology* **2004,** 22, (11), 1343-1346.
7. Chirino, A. J.; Mire-Sluis, A., Characterizing biological products and assessing comparability following manufacturing changes. *Nature Biotechnology* **2004,** 22, (11), 1383-1391.
8. Schellekens, H.; Ryff, J. C., 'Biogenerics': the off-patent biotech products. *Trends in Pharmacological Sciences* **2002,** 23, (3), 119-121.
9. Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G., Hydrogen-Bonding and Biological Specificity Analyzed by Protein Engineering. *Nature* **1985,** 314, (6008), 235-238.
10. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry A* **1997,** 101, (16), 3005-3014.
11. Jiang, K. Y.; Pitiot, O.; Anissimova, M.; Adenier, H.; Vijayalakshmi, M. A., Structure-function relationship in glycosylated alpha-chymotrypsin as probed by IMAC and IMACE. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1999,** 1433, (1-2), 198-209.
12. Pathange, L. P.; Bevan, D. R.; Larson, T. J.; Zhang, C. M., Correlation between protein binding strength on immobilized metal affinity chromatography and the histidine-related protein surface structure. *Analytical Chemistry* **2006,** 78, (13), 4443-4449.
13. Pathange, L. P.; Bevan, D. R.; Zhang, C. M., In preparation. **2007**, (13).

Chapter II: Literature Review

Well characterized proteins are of immense importance as they can be used as therapeutic proteins,[1] drug development targets and also aid in understanding the proteome.[2] With the advances in recombinant DNA technology, more and more therapeutic proteins are produced in extraneous expression hosts.[3] These therapeutic proteins, produced by the recombinant DNA technology, are known as biopharmaceuticals or biotherapeutic proteins. Though the growth of the biopharmaceutical industry has been very impressive (the market is estimated to be in tens of billions of dollars),[4] the increscent challenge has been to develop new and reliable techniques to characterize these recombinant proteins,[5, 6] because recombinant proteins when expressed in extraneous hosts may or may not assume the native format of the natural protein. With the growing debate over the need to use 'biogenerics' or 'off-patent biotech products', there is a considerable interest in developing techniques that aid in characterizing biopharmaceuticals. In addition, these new and/or evolving techniques are in great demand due to the rising health care costs in U.S.[7]

The inherent challenges with biopharmaceuticals are: 1) the type of the extraneous expression host, which expresses proteins that are a heterogeneous mixture, consisting of variants of the natural protein (especially in structure),[8] 2) the type of manufacturing process, which is very complex compared to convention pharmaceutical products, involving specialized techniques such as genetic engineering, fermentation, extraction and purification,[7, 9] and 3) the need to conduct expensive clinical trials to determine the efficacy and toxicity of the biopharmaceuticals.[3, 10] Though the current array of analytical techniques,[10] which are based on proteins' physiochemical parameters

(size, charge, molecular weight) are very impressive, they still lack the ability to characterize and predict proteins' biological and clinical properties. In addition, our ability to fully understand the relationship between structure and function is not yet complete. As a result, any new subtle changes that might occur in the structure of the biopharmaceutical could adversely affects its function, especially when used as a drug. Therefore, the crux of the problem is that the available analytical techniques that are based on only few basic physiochemical parameters do not elucidate the whole structural information of a protein. For example there is no analytical technique that is powerful enough to characterize mis-folded proteins.

This problem of characterization is augmented by the fact that all proteins exist in solution as an ensemble of closely related microstates (subtle variations of the native structure).[11-13] They can be imagined as flexible and dynamic (''breathing'') structures. Due to the dynamic nature of the (thermodynamic) equilibrium states, the ensemble-like protein structures are believed to be conformationally heterogeneous.[14-16] Characterization of these macroscopic heterogeneous states is extremely complex and attempts have been made to describe them in biophysical concepts such as elasticity, protein quakes and energy landscapes.[16] Even though it is hard to quantitate the macroscopic heterogeneity/degree of flexibility,[17] there is a need to consider the extent of the role played by flexibility in predicating/discovering the function of a protein/enzyme. Knowledge of such intricate details of a protein would elucidate the mechanisms involved in the areas such as signal transduction, protein-protein interaction and protein-ligand binding.[13, 18]

One of the reasons for this apparent flexibility displayed by proteins is due to the presence of electrostatic forces between the individual amino acid residues. It is believed that electrostatic forces play a major role in maintaining both structural and functional properties of proteins.[19-24] Intramolecular forces such salt bridges and hydrogen bonds (primarily electrostatic forces) not only help in maintaining a protein's structure[25] but also aid in its function by stabilizing the charged transition states during catalytic reactions.[26] For example, it has been proposed that salt bridges play an important role in the biological/physiological function of hemoglobin and pyruvate kinase.[27] Another important parameter that needs to be included in the discussion of electrostatic forces is the effect/function of protons (H^+). Protons, due to their negligible size, act as an important local denaturing agent[24] mainly due to its ability to change the electrostatic properties (charge vs. uncharged) of the particular site (amino acid residue) on a protein surface. Therefore, one way to study the flexibility effects in a protein structure is by measuring the pK_a values of the titratable groups (e.g. aspartate, histidine). However, due to environmental effects in structured proteins, the titratable amino acid (here histidine) residues display different pK_a values from those in small peptides or free amino acids.[8, 25, 28-34] Understanding and quantifying the effects of these factors on the pK_a of a reporter titratable amino acid (here, histidine) would allow possible inference of the microstructure around the titratable groups in proteins, which is otherwise difficult to obtain by other analytical methods.

On the other hand, one of the results due to the heterogeneous nature of the protein structure is that the SA of the titratable groups in different protein ensembles may vary and the extent of variation in SA would be maxima on the surface of a protein. Also

due to the ensemble nature of the protein, researchers have proposed the existence of certain sites on the surface of a protein, in addition to the well defined ligand binding sites (such as active sites), known as 'hot spots".[35, 36] Though these sites are not exactly the allosteric sites on the protein, they still subtly alter the surface accessibility (SA) and the orientation of surrounding amino acids (microenvironment) and sometimes even disrupting the active site. For example, Savvides and Karplus demonstrated that the organic molecule 6-hydroxy-3-oxo-3H-xanthene-9-propionic acid (XAN) alters the conformation of not only the active site but also several surrounding amino acids such as Phe-78, Met-79, His-75 and His-82 in human glutathione reductase.[37] It is believed that these 'hot spots' play an important role in protein's function, especially in protein-protein interaction or protein ligand binding reactions.[35, 38] Understanding and quantifying such minor variations is very important especially when the final product (pure protein) is used for therapeutic purposes. It is in these cases that the SA of the titratable groups can be used as a parameter for analytical purposes. Thus, SA of titratable groups can be used as an analytical parameter in IMAC to characterize heterogeneous protein structures.

Surface accessibility or solvent accessible area of an amino acid residue in a protein is the static area of the amino acid residue exposed to the solvent calculated from the X-ray crystal structure of the protein.[39, 40] SA is an important concept in quantifying the protein's surface structure as it elucidates the associated chemical properties of the amino acid residues (such as hydrophobic or hydrophilic).[41] Suckau et al., in their effort to probe surface topology of a protein by using selective chemical modification and mass spectrometry, suggested that there is a relation between surface structures of the amino

acids with their reactivity.[42] SA in proteins has been used as a parameter to estimate thermodynamic parameters such as entropy and enthalpy.[15]

Thus, structural parameters chosen for this work are surface accessibility and microstructure of individual amino acid residues. A range of these structural parameters were introduced into the protein structure via site-directed protein mutagenesis. Based on these two parameters two categories of structural variants were generated via site-directed protein mutagenesis. Here, chromatographic techniques such as IMAC and ion exchange chromatography (IEC) were employed to quantify these parameters.

Site directed protein mutagenesis: The relationship between the structure and function of a protein is complex. Site-directed protein mutagenesis has proven to be a very valuable tool in elucidating this relationship between the structure and function of proteins. It has also been demonstrated that point mutations (structural variation) aid in protein purification process (function).[43, 44] Point mutation (site-directed mutagenesis) modifies a target protein by replacing selected surface amino acids with amino acids with special properties. By replacing one surface lysine with glutamic acid, the purification of T4 lysozyme from canola extract can be greatly enhanced when cation exchange chromatography is used in separation.[44]. Chicz and Regnier demonstrated that by replacing Gly166 with His on subtilisin, the protein's retention on immobilized metal affinity chromatography (IMAC) was significantly increased.[8] Furthermore, Todd et al. showed that yeast expressed point mutated iso-1-cytochrome *c* maintained its structural integrity and conformation as the native protein when amino acids 4 and 8 were replaced with histidines, and the apparent binding affinity of the proteins containing multiple accessible histidines with immobilized copper ions in IMAC can be increased as much as

a factor of 1000.[45] These examples demonstrate that point-mutation can be an effective way to alter protein structure. Here we have utilized site-directed mutagenesis as a tool to introduce histidine residue, which acts as a reporter amino acid.

Immobilized metal affinity chromatography (IMAC): The ability of few amino acids such as histidine and cysteine to form stable complexes (coordinate covalent bond) with certain transition metals (Ni, Cu, Zn and Co) has been exploited to develop a new separation technique known as immobilized metal affinity chromatography.[46-48] IMAC is a very versatile technique to purify proteins. It has been previously used to characterize natural, recombinant and modified proteins.[49] In IMAC, a transition metal ion such as copper is immobilized in the stationary phase by a chelating agent such as iminodiacetic acid (IDA).[46] Copper ion was selected for our investigation because of the Cu-IDA column's ability to retain a protein with a single histidine. The interaction between a protein and an IMAC column depends on the number and distribution of histidines on a protein surface.[50]

Coordination chemistry of metals: Generally in aqueous solution metal ions are solvated by water molecules. But when a stronger base approaches the metal ion, the water molecule is replaced by the ligand (base) forming a coordination covalent bond (a complex). A metal ion's ability to form complexes can be explained by its versatile energy levels (orbital) which are a function of their local chemical environment. Such a conclusion was derived from various theories such as crystal field, valence bond, ligand field and molecular orbital theory.[51] For example, according to the crystal field theory when ligands are aligned along the axes, as in octahedral complexes, the five degenerate'd' orbitals of a metal ion are split in to two sets, mainly due to simple

electrostatic interactions. The first set contains two higher energy (e_g) orbitals and the second contains three lower energy (t_{2g}) orbitals. The electrons in the e_g orbitals are less stable and therefore hybridize to form d^2sp^3 orbitals leading to the formation of an octahedral structure.

Coordination chemistry between the chelating agent and the metal ion: When two or more atoms of the same ligand bind to a single metal ion, then the resulting complex is called a chelate. A chelate is more stable than a complex (single bond from each ligand) due to the loss in free energy due to the formation of ring like structure. This gain in the free energy is due to the entropic gain by the release of water molecules and the enthalpy loss is due to orbital overlap and reduction in the electrostatic repulsions between ligands in complexes.[52]

Generally metal ions are immobilized to matrices via chelating agent such as iminodiacetic acid (IDA), carboxymethylated aspartic acid (CM-ASP) and tris (carboxymethyl) ethylene diamine (TED). The metal ions immobilized are thus referred as 'tamed' metals.[47, 53] The chemistry of tamed metal ions is much different from the free metals in solutions. Free metal ions can complex with many protein molecules forming large complexes, while tamed metal ions can interact with fewer protein ligands due to a substantial loss in its degrees of freedom and additional steric hindrances created by the supporting matrix.

The stability of the metal-chelate is determined by the number of atoms involved in a ring structure, and the number of the members in the ring structure is partially determined by the contribution if metal ions via the coordinate bond. So if a metal ion devotes most of its available coordinate sites to the chelate, then the metal-chelate might

be very stable but renders the metal-protein interaction weaker. Thus, optimal coordination chemistry should be used to attain a balance in distributing the metal coordinate sites between the chelate and the protein ligand. Generally a maximum of 5 or 6 membered ring is preferable. According to Porath,[47] the order of the stability of the metal-chelating is: Me-TED > Me-CM-ASP > Me-IDA.

However, this order is reversed for the affinity between the metal-chelate and the protein. Apart from the nature of the chelating agent, the density of the chelating ligand influences the adsorption process in IMAC. This could be due to the multipoint interactions between the chelate-metal and the protein or chelate-metal and the chelate, leading to the heterogeneous nature of adsorption in IMAC.

Coordination chemistry between the metal ion and the amino acid in a protein:

Metal ions in general are Lewis acids. According to Pearson's hard and soft acid and bases theory (HSAB), the bond strength associated with acids or bases depends on their intrinsic hardness or softness. If an electron can be easily moved from its (an acid or a base) orbitals then the acid or base is treated as soft, while if the electrons are firmly within their orbitals then the acid or base is treated as hard. Some of the factors that determine whether an acid or a base is hard or soft are size, bond types, oxidation state and electronegativity or electropositivity.

Soft metals such as Ag^{+}, Au^{+}, Hg^{+}, Cu^{+}, Hg^{+}, Pt^{++} and Cd^{++} prefer to bind with soft bases such as sulphur containing groups and cyanide ions. Hard metals such as K^{+}, Na^{+}, Ca^{++} and Mg^{++} prefer to bind with hard bases such as ammonia, amines, sulfates, carbonates, phosphates and perchlorates. While the border line acids represented by Fe^{3+},

Co^{2+}, Zn^{2+}, Ni^{2+} and Cu^{2+} prefer to bind with borderline bases such as aromatic nitrogen.[52] The chelated metal in IMAC can undergo three types of interactions:

1) Chelated-Metal (hard acid) + Protein- COO- (hard base) ---- ionic
2) Chelated-Metal (borderline acid) + Protein- N: (borderline base) ---- coordinate
3) Chelated-Metal (soft acid) + Protein-SH - (soft base) ---- covalent

Off the above three interactions, the most widely desired interaction in IMAC is coordinate bond where the metals used are border line metals. Generally Cu^{2+} and Co^{2+} can be oxidized in solutions, as against Zn^{2+} and Ni^{2+} which are more electrochemically stable, but, when immobilized, all the above mentioned metals are stable (for example Cu^{2+} is the most versatile and strong adsorbent). Cr^{3+} and Co^{3+} metals are not generally employed because of their slow kinetics. Protein affinity for metals increases as the number of the d orbital increases as the following, except for Zn^{2+} [48], $Cu^{2+} > > Ni^{2+} > Co^{2+} \sim Zn^{2+}$. In IMAC the amino acids which are able to bind with border line metals (except Fe^{3+}) are primarily histidine groups while to some extent by cysteines, tryptophans and the amino terminal of the protein, and in contrast, Fe^{3+}binds with phosphoproteins (or phosphoserines).

Some of the important factors affecting coordinate bond: Even after chelating a metal, chelating agents still retain some ion exchange properties at lower salt concentrations. Similarly, aspartate and glutamate can contribute to protein retention by ion exchange mechanism. So having a salt at certain concentration (> 0.5 M) is important for the coordinate characteristics to dominate in an IMAC process. On the other hand, ligand density is another important factor that affects protein retention. The solution pH has a tremendous effect on the affinity of a protein-metal binding. According to

Andersson and Sulkowski,[54] the α–amino terminal in a protein at pH > 8 has a binding affinity to a metal ion similar to that of a histidine residue. And at pH 5-7, the peptides which have an α–amino terminal and surrounded by a number of positive residues could bind to the metal. Lysine containing peptides can also bind to the metal at pH > 9 due to the deprotanation of the ϵ-amino group.

IMAC as a tool to probe the surface accessibility of histidine: IMAC has the potential to be an important tool for protein characterization [55, 56]. There has been a lot of discussion on the exact role played by the surface accessible amino acids during the interaction with chelated metal ion, but there has been no detailed understanding of the mechanism involved [57]. It has been postulated that by maintaining the physiochemical variables such as buffer ions, metal chelates, pH and temperature that affect IMAC,[58] it is possible to probe the effect of the various inherent protein surface properties such as the number, type, surface accessibility, micro chelation (vicinal histidines binding to a single metal ion), macro chelation (multipoint binding) and microenvironment of the electron donor groups.

The elution time for a protein from an IMAC column directly varies with the number of accessible histidines on the protein surface.[59, 60] There have been some studies regarding the generation of multivalent metal binding sites such as His-X3-His motif.[61, 62] By varying the number of accessible histidine sites (single, double and multiple) through site-directed mutagenesis, Todd et al. were able to probe the mechanism involved in protein retention by IMAC.[45] Though the binding strengths of the multivalent interactions depend on both the entropic and enthalpic considerations, there is an extra factor, the

geometric conformation of the protein (as well as the nature of the coordinate interaction involved), which influences these interactions.[62]

It has been mentioned in the literature that the protruding metal ions in IMAC can be used as a probe to assess the surface histidine distribution on a protein.[58, 63] It has also been reported that IMAC can be further exploited to probe changes on the protein surface affected by mis-folding or unfolding, partial digestion or protein-protein interactions.[64] Particularly, IMAC, with IDA (iminodiacetic acid) as the chelating agent, has been successfully used as a facile tool to probe the surface topography of histidine residues.[59, 65] Though there are a number of chelating agents available, IDA was chosen because of its high complex forming ability with metals leading to minimization in metal leaching.[47]

Microstructural effect on IMAC binding processes: Microstructure (conformation and charge) of the amino acids involved in a coordinate bond has an important effect in protein-metal affinity, sometimes rendering the electron donating amino acid chromatographically irrelevant. Berne et al. concluded that His-57 in chymotrypsin is surface accessible but could not bind to the metal ion due to the presence of a hydrogen bond.[66] In general, imidazole ring accessibility and flexibility is an important factor that influences the binding process. The presence of aromatic side chains in the microstructure of accessible histidine enhances the latter's binding ability. This could be either due to a π-π interaction with the metal[65] or might be due to the entropic contribution, wherein the water is released. Meanwhile the presence of negative charges, at neutral pH decrease the histidine binding ability.[67] For example, Wuenschellet al. demonstrated that histidines with minor changes in microenvironment had a considerable effect on the affinity towards the metal ions in IMAC;[68] Chicz and

Regnier used a series of site-directed protein mutagenesis experiments to demonstrated that histidine microenvironment with a radius of 15 Å had a profound effect on the retention time of subtilisin protein.[8]

The pK_a of the imidazole group of histidine changes with its microenvironment, which in turn changes in the electrostatic interactions of the subsystem (microenvironment + histidine). However, to analyze such interaction is very challenging. Nevertheless there has been a considerable endeavor to understand the surface charge effect on the pK_a of the ionizable side chains of the protein.[69] This aspect is important because most of the protein's charged groups are situated on the surface to interact with the solvent and to take part in catalytic reactions. To understand the effect of the environment on the pK_a of the histidine, one has to consider some basic electrostatic theories involved in measuring the pK_a's of a protein. According to Kirkwood,[70] an organic ion is treated as a cavity of low dielectric constant submerged in the continuous solvent with its own dielectric electric constant. Therefore, the free energy of the interaction between two pair of charges q1 and q2 (or a point dipole) separated by a distance 'r' is given in Eq. 1:

$$W = \frac{q_1 \cdot q_2}{D \cdot r} \qquad (1)$$

Where D (units equivalent with dielectric constant) is a function of the position with respect to the interface between the cavity and solvent.[71] When such a model is extended to a protein molecule, the most critical factor influencing the model is the depth of the charge below the surface cavity, thus affecting the D value in Eq. 1. It can be reasoned that the pK_a value of an amino acid depends on the interaction between it and all the surrounding amino acids including water molecules. The values of W, obtained from the

model in Eq. 1, can be correlated to the pK_a difference between amino acid residues, wherein the ΔpK_a can be obtained experimentally.

So whenever the microenvironment of an amino acid is changed, and if there is a change in the value of one of the q's in Eq. 1 (there might be more than two ions (q_1, q_2, q_3, ….q_n)), these changes will affect the value of W and thus affect the pK_a value of the amino acid. With respect to histidine, an over all raise in the negative charge within its microenvironment would raise its pK_a by stabilizing its protonated form. While an overall raise in the positive charge would lower its pK_a by destabilizing the protonated form. For example, at low ionic strengths in subtilism protein, the Asp99 and Glu156 were mutated to serine and lysine, which resulted in lowering the pK_a of histidine by 0.65 and 1.0 units, respectively.[72]

Similarly when a single accessible histidine is the electron donor group on the surface of the protein, the retention time is inversely proportional to the pK_a's of histidines. For example, cytochrome *c* (His-133, pK_a 6.5) elutes earlier than calmodulin (His-107, pK_a 6.1).[52] Bovine-stomach lysozyme c2 (pI 7.5) which is less basic then California quail lysozyme (pI ~11.0) elutes much later than then the latter.[67]

Thus, in addition to measuring the SA, it is important to measure the pK_a values of the titratable groups. These two parameters (pK_a and SA) of a titratable group play an important role in cases where the interaction is solely dependent on the side chain of a particular amino acid residue. Here our studies are centered to investigate whether or not IMAC can be used as a tool to carry out such studies.

Ion exchange chromatography (IEC): In addition to IMAC, another important goal of this dissertation is to elucidate structure-function (a protein's ability to interact

with oppositely charged groups) relationships using ion exchange chromatography (IEC) by correlating the same aforementioned two structural parameters and the resultant functional behavior (binding) in an IEC column. Understanding this correlation is important because the mechanism of protein binding in IEC is different from that in IMAC. The general trend in IEC is that the elution is dependent on the net charge of a protein (or its pI). However it has been demonstrated that net charge is not the only criterion for protein elution in IEC, as indicated by stoichiometric displacement model (SDM).[73, 74] Though the whole protein surface modulates the protein binding process, the specific binding interactions are dominated by certain contact regions.[75] Thus, the contribution of all charged amino acid residues in IEC do not contribute equally to the protein binding process.[73] The main goal of all our IEC experiments was to analyze subtle variation in protein charge including its position on the protein's surface by IEC. Such studies not only will reveal the ability of IEC in capturing subtle changes in protein structure but also may provide insight into the protein binding mechanism in IEC.

References

(1) Copmann, T.; Davis, G.; Garnick, R.; Landis, J.; Lubiniecki, A.; Massa, T.; Murano, G.; Seamon, K.; Zezza, D. *Biopharm-the Applied Technologies of Biopharmaceutical Development* **2001**, *14*, 14-+.

(2) Zachariou, M.; Hearn, M. T. W. *Journal of Protein Chemistry* **1995**, *14*, 419-430.

(3) Koths, K. *Current Opinion in Biotechnology* **1995**, *6*, 681-687.

(4) Walsh, G. *Nature Biotechnology* **2003**, *21*, 865-870.

(5) Green, J. D.; Tsang, L.; Cavagnaro, J. A. *Expert Opinion on Biological Therapy* **2003**, *3*, 1019-1022.

(6) Kleinberg, M.; Mosdell, K. W. *American Journal of Health-System Pharmacy* **2004**, *61*, 695-708.

(7) Herrera, S. *Nature Biotechnology* **2004**, *22*, 1343-1346.

(8) Chicz, R. M.; Regnier, F. E. *Analytical Chemistry* **1989**, *61*, 1742-1749.

(9) Schellekens, H. *Nephrology Dialysis Transplantation* **2005**, *20*, 31-36.

(10) Chirino, A. J.; Mire-Sluis, A. *Nature Biotechnology* **2004**, *22*, 1383-1391.

(11) Whitten, S. T.; Garcia-Moreno, B.; Hilser, V. J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 4282-4287.

(12) Englander, S. W. *Annual Review of Biophysics and Biomolecular Structure* **2000**, *29*, 213-238.

(13) Swain, J. F.; Gierasch, L. M. *Current Opinion in Structural Biology* **2006**, *16*, 102-108.

(14) Vertrees, J.; Barritt, P.; Whitten, S.; Hilser, V. J. *Bioinformatics* **2005**, *21*, 3318-3319.

(15) Thuduppathy, G. R.; Hill, R. B. *Protein Science* **2006**, *15*, 248-257.

(16) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 12570-12575.

(17) Teague, S. J. *Nature Reviews Drug Discovery* **2003**, *2*, 527-541.

(18) Liu, T.; Whitten, S. T.; Hilser, V. J. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 728-738.

(19) Louro, R. O.; Catarino, T.; Paquete, C. M.; Turner, D. L. *Febs Letters* **2004**, *576*, 77-80.

(20) Lee, K. K.; Fitch, C. A.; Lecomte, J. T. J.; Garcia-Moreno, B. *Biochemistry* **2002**, *41*, 5656-5667.

(21) Russell, A. J.; Thomas, P. G.; Fersht, A. R. *Journal of Molecular Biology* **1987**, *193*, 803-813.

(22) Loewenthal, R.; Sancho, J.; Reinikainen, T.; Fersht, A. R. *Journal of Molecular Biology* **1993**, *232*, 574-583.

(23) Allewell, N. M.; Oberoi, H. *Methods in Enzymology* **1991**, *202*, 3-19.

(24) Fitch, C. A.; Whitten, S. T.; Hilser, V. J.; Garcia-Moreno, B. *Proteins-Structure Function and Bioinformatics* **2006**, *63*, 113-126.

(25) Laurents, D. V.; Huyghues-Despointes, B. M. P.; 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.; Pace, C. N. *Journal of Molecular Biology* **2003**, *325*, 1077-1092.

(26) Garcia-Mayoral, M. F.; Perez-Canadillas, J. M.; Santoro, J.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Lacadena, J.; del Pozo, L. M.; Gavilanes, J. G.; Rico, M.; Bruix, M. *Biochemistry* **2003**, *42*, 13122-13133.

(27) Sinha, N.; Smith-Gill, S. J. *Current Protein & Peptide Science* **2002**, *3*, 601-614.

(28) Mehler, E. L.; Fuxreiter, M.; Simon, I.; Garcia-Moreno, E. B. *Proteins-Structure Function and Genetics* **2002**, *48*, 283-292.

(29) Chicz, R. M.; Regnier, F. E. *Journal of Chromatography A* **1990**, *500*, 503-518.

(30) Lee, K. K.; Fitch, C. A.; Garcia-Moreno, B. *Protein Science* **2002**, *11*, 1004-1016.

(31) Huyghues-Despointes, B. M. P.; Thurlkill, R. L.; Daily, M. D.; Schell, D.; Briggs, J. M.; Antosiewicz, J. M.; Pace, C. N.; Scholtz, J. M. *Journal of Molecular Biology* **2003**, *325*, 1093-1105.

(32) Mehler, E. L.; Eichele, G. *Biochemistry* **1984**, *23*, 3887-3891.

(33) Joshi, M. D.; Sidhu, G.; Nielsen, J. E.; Brayer, G. D.; Withers, S. G.; McIntosh, L. P. *Biochemistry* **2001**, *40*, 10115-10139.

(34) Gitlin, I.; Carbeck, J. D.; Whitesides, G. M. *Angewandte Chemie-International Edition* **2006**, *45*, 3022-3060.

(35) Krantz, A. *Trends in Biotechnology* **1998**, *16*, 198-199.

(36) Bonacci *science* **2006**, *312*, 443-446.

(37) Savvides, S. N.; Karplus, P. A. *Journal of Biological Chemistry* **1996**, *271*, 8101-8107.

(38) Niccolai, N.; Spiga, O.; Bernini, A.; Scarselli, M.; Ciutti, A.; Fiaschi, I.; Chiellini, S.; Molinari, H.; Temussi, P. A. *Journal of Molecular Biology* **2003**, *332*, 437-447.

(39) Lee, B. W.; Richards, F. M. *Journal of Molecular Biology* **1971**, *55*, 379-400.

(40) Richmond, T. J. *Journal of Molecular Biology* **1984**, *178*, 63-89.

(41) Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. *Journal of Molecular Biology* **1987**, *196*, 641-656.

(42) Suckau, D.; Mak, M.; Przybylski, M. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 5630-5634.

(43) Mrabet, N. T. *Biochemistry* **1992**, *31*, 2690-2702.

(44) Zhang, C. M.; Glatz, C. E. *Biotechnology Progress* **1999**, *15*, 12-18.

(45) Todd, R. J.; Johnson, R. D.; Arnold, F. H. *Journal of Chromatography A* **1994**, *662*, 13-26.

(46) Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. *Nature* **1975**, *258*, 598-599.

(47) Porath, J. *Trac-Trends in Analytical Chemistry* **1988**, *7*, 254-259.

(48) Porath, J. *Protein Expression and Purification* **1992**, *3*, 263-281.

(49) Ueda, E. K. M.; Gout, P. W.; Morganti, L. *Journal of Chromatography A* **2003**, *988*, 1-23.

(50) Mrabet, N. T. *Methods* **1992**, *4*, 14-24.

(51) Constable, E. C. *Metals and ligand reactivity*; Ellis Horwood Limited: West Sussex, England, 1990.

(52) Wong, J. W.; Albright, R. L.; Wang, N. H. L. *Separation and Purification Methods* **1991**, *20*, 49-106.

(53) Porath, J. a. B., M In *Academic Press Inc.,* ; Chaiken, I. M. e. a., Ed.: New York, 1983, pp 173-190.

(54) Andersson, L.; Sulkowski, E. *Journal of Chromatography* **1992**, *604*, 13-17.

(55) Mrabet, N. T. *Methods: A companion to methods in enzymology* **1992**, *4*.

(56) Hemdan, E. S.; Zhao, Y. J.; Sulkowski, E.; Porath, J. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 1811-1815.

(57) Jiang, W.; Hearn, M. T. W. *Analytical Biochemistry* **1996**, *242*, 45-54.

(58) Hutchens, T. W.; Yip, T. T. *Analytical Biochemistry* **1990**, *191*, 160-168.

(59) Sulkowski, E. *Trends in Biotechnology* **1985**, *3*, 1-7.

(60) Krishnan, S.; Vijayalakshmi, M. A.; Geahel, I. *Journal of Chromatography* **1987**, *397*, 339-346.

(61) Haymore, B. L., Bild, G.S., Salsgiver, W.J., Staten, N.R. and Krivi, G.R. *Methods* **1992**, *4*, 25-40.

(62) Suh, S. S.; Haymore, B. L.; Arnold, F. H. *Protein Engineering* **1991**, *4*, 301-305.

(63) Belew, M.; Porath, J. *Journal of Chromatography* **1990**, *516*, 333-354.

(64) Arnold, F. H. *Bio-Technology* **1991**, *9*, 151-156.

(65) Hemdan, E. S.; Porath, J. *Journal of Chromatography* **1985**, *323*, 255-264.

(66) Berna, P. P.; Mrabet, N. T.; VanBeeumen, J.; Devreese, B.; Porath, J.; Vijayalakshmi, M. A. *Biochemistry* **1997**, *36*, 6896-6905.

(67) Zhao, Y. J.; Sulkowski, E.; Porath, J. *European Journal of Biochemistry* **1991**, *202*, 1115-1119.

(68) Wuenschell, G. E.; Wen, E.; Todd, R.; Shnek, D.; Arnold, F. H. *Journal of Chromatography* **1991**, *543*, 345-354.

(69) Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* **1985**, *314*, 235-238.

(70) Kirkwood, J. *The journal of chemical physics* **1934**, *2*, 351-361.

(71) Tanford, C. *Journal of american chemical soceity* **1957**, *79*, 5348-5352.

(72) Russell, A. J.; Fersht, A. R. *Nature* **1987**, *328*, 496-500.

(73) Kopaciewicz, W.; Rounds, M. A.; Fausnaugh, J.; Regnier, F. E. *Journal of Chromatography* **1983**, *266*, 3-21.

(74) Zhang, C. M.; Glatz, C. E. *Journal of Chromatography A* **2005**, *1069*, 113-118.

(75) Xu, W. S.; Zhou, H.; Regnier, F. E. *Analytical Chemistry* **2003**, *75*, 1931-1940.

(76) Daopin, S.; Soderlind, E.; Baase, W. A.; Wozniak, J. A.; Sauer, U.; Matthews, B. W. *Journal of Molecular Biology* **1991**, *221*, 873-887.

Chapter III: Correlation between protein binding strength on immobilized metal affinity chromatography and the histidine-related protein surface structure

(Reproduced with the permission of Pathange, L. P.; Bevan, D. R.; Larson, T. J.; Zhang, C. M. 2006. Analytical Chemistry, 78, 4443-4449.)

Abstract

Immobilized metal affinity chromatography (IMAC) was investigated for its ability to characterize the histidine-related surface structure of a protein, i.e. histidine surface accessibility and its potential involvement in intramolecular interactions. T4 lysozyme was chosen as the model protein. Seven amino acid sites were selected based on their relative surface accessibility, and they were substituted by histidine via site-directed protein mutagenesis to generate seven T4 lysozyme variants, each containing only one histidine on its surface. IMAC was then used to experimentally quantify the interaction of each lysozyme variant with immobilized copper ions. A direct correlation was shown between the protein binding affinity and the histidine surface accessibility. Of all the lysozyme variants, K83H and K147H showed unusually low binding strength compared with variants having similar histidine surface accessibility. However, with the aid of molecular modeling, their relatively low binding affinities were predicted to be the result of the involvement of the histidine in intramolecular interactions. In contrast to previously reported results, our results showed that lysozyme still binds to the IMAC column, even if its histidine is involved in intramolecular bonding, such as a hydrogen bond, albeit at reduced strength compared with the variant containing a histidine with similar surface accessibility.

Key words: Surface accessibility, binding strength, immobilized metal affinity chromatography (IMAC), histidine, surface structure, intramolecular interaction.

Introduction

With advances in recombinant DNA technology, more and more therapeutic proteins are produced in heterologous expression hosts.[1] Though the growth of the biopharmaceutical industry has been very impressive (the market is estimated to be in tens of billions of dollars),[2] the increasing challenge has been to develop new and reliable techniques to characterize these recombinant therapeutic proteins.[3, 4] Characterization of these recombinant proteins is necessary because biopharmaceutical production is a complex process.[5, 6] For example, depending on the type of the expression host chosen, a recombinant protein may experience variation in expression, folding, post-translational modification, and glycosylation.[5, 7, 8]

Recently, efforts have been focused on developing and extending affinity-based chromatographic techniques to characterize proteins that differ in their structural properties.[7] For example, Qiu and Regnier were able to determine, quantitatively, the extent of glycosylation in glycoproteins using a lectin based affinity chromatographic technique.[9] Furthermore, Jiang et al. employed immobilized metal affinity chromatography (IMAC) and immobilized metal-ion affinity capillary electrophoresis (IMACE) techniques to reveal subtle changes in surface accessibility of the histidine residues in α-chymotrypsin due to glycosylation.[10]

IMAC was developed by Porath and co-workers, primarily as a protein purification technique, about three decades ago.[11] The basic principle involved in IMAC is that certain amino acids (e.g. histidine, cysteine) present on the surface of a protein can be

electron donors to form stable complexes with certain transition metal ions (e.g. Ni^{2+}, Cu^{2+}, Zn^{2+}, Co^{2+}).[11] Among the possible electron donors, histidine is the primary amino acid that interacts with metal ions at neutral pH.[12] The topography of histidine residues on the surface of a protein bears important structural information that can elucidate the functional properties of a protein (e.g. active site histidines in serine protease family[13, 11] and metal binding histidines in metalloproteins[14]). In addition, each protein has a unique distribution of surface histidine residues,[15] which can be exploited as a molecular fingerprint to identify each protein and possibly to differentiate a protein from its isoforms.

Due to the specific interactions between immobilized metal ions and the histidine residues on the surface of a protein, IMAC has been previously investigated as a tool to probe the surface topography of histidine residues,[16] including 1) their surface localization (exposed or buried), 2) their number, and 3) their density.[17] During the histidine-mediated protein-metal ion interactions in IMAC, the following factors can affect the extent of the interaction: 1) the extent of histidine surface accessibility,[11] 2) the number of the surface-accessible histidines,[16] and 3) the microenvironment of histidines, which refers to the amino acids within 15 Å of a particular histidine residue.[13] The importance of histidine microenvironment was demonstrated by Russell and Fresht[13] who calculated the changes in the pK_a values of His-64 in subtilisin variants. Their results indicated that there was a change of 1 pK_a unit of the histidine residue when two acidic amino acids (Asp and Glu) were replaced by two basic amino acids (both Lys). Furthermore, Chicz and Regnier used a series of site-directed protein mutagenesis experiments on subtilisin (His-64) to illustrate that the histidine microenvironment had a profound effect on its retention time in an IMAC column.[8] However, our current understanding of the effects of these factors on protein binding on IMAC is still at a qualitative level. For example, Boden et al. used IMAC as a probe to

characterize the subtle 3D structural differences within the serine protease family by demonstrating a qualitative correlation between the surface accessibility of the histidine residues with their corresponding elution profiles.[18] To develop IMAC as a more powerful tool in obtaining histidine-related protein structural information, it is vital to establish quantitative relationships between the measurable protein-metal binding strength and the effecting factors. Thus, the protein-metal binding strength may be used to derive histidine-related protein structural information, which is otherwise difficult to obtain.

Until now, there has been no systematic study to quantitate the relationship between the surface accessibility of histidine residues and the binding strength of a protein on IMAC. Surface accessibility or solvent accessible surface of an amino acid can be defined as the area accessible to the solvent molecules. In general the surface accessibility values reported in the literature are the areas of amino acid residues, calculated from the X-ray crystal structure of a protein using computational programs. The surface accessibility of amino acids is an important concept in quantifying the protein surface structure as it elucidates the associated chemical properties of the amino acid residues (such as hydrophobic or hydrophilic).[19] However, it does not provide insights into other important aspects of the surface structure of a protein in solution (e.g. strength of a hydrogen bond). For example, His-57 in chymotrypsin was a surface accessible residue and therefore it was expected to bind on an IMAC column, but the experiment results indicated that His-57 did not bind to the IMAC column due to its involvement in intramolecular hydrogen bonds.[11] Thus, IMAC may also be used as a probe to determine the involvement of the histidine residue in intramolecular interactions.

In this report, the main objective is to use an IMAC-Cu^{2+} column to determine the binding strength of T4 lysozyme variants, each of which contains only one surface histidine with

different surface accessibility, and to establish a correlation between the protein-metal ion binding strength and the histidine accessibility. In addition, the effect of microenvironment of the histidine residue on the protein binding strength was also considered. To achieve this objective, a series of single-site mutated variants of T4 lysozyme were generated. The extent of the histidine residue exposed to solvent was measured using the relative surface accessibility, which is often used to compare amino acid residues of various sizes.[20] Zonal analysis[21, 22] experiments were carried out to quantify the interaction between the lysozyme variants and the immobilized copper ion of an IMAC column.

Materials and Methods

Materials

The plasmid pHS1403[23] containing a T4 lysozyme gene was kindly donated by Dr. Brian Matthews (University of Oregon, Eugene, OR). The T4 lysozyme encoded by pHS1403 has two amino acid substitutions, C54T and C97A, yielding a cysteine-free lysozyme.[24, 25] The TSK gel chelate-5PW glass column (5 cm × 5 mm, bead size 10 μm), used to perform the zonal analysis experiments, was purchased from Tosoh Bioscience (Philadelphia, PA). All the other reagents were analytical grade and were purchased from either Fisher Scientific (Hampton, NH) or Sigma (St. Louis, MO).

In silico modeling methods

SYBYL 6.7 (Tripos Inc., St. Louis, MO) was used to calculate the surface accessibility of all amino acids of the T4 lysozyme and to generate the *in silico* lysozyme variants (point mutants), wherein the original amino acid was replaced by a histidine residue. The PDB code for T4 lysozyme used for all calculations was 1L63. The surface accessible values of all amino acids in T4 lysozyme were calculated using a probe with a

radius of 1.93 Å, to mimic the metal ion accessibility in the IMAC column, previously reported by Mrabet et al.[11] Before calculating the surface accessibility of the histidine residue, all the variant structures were energy minimized with the SYBYL energy minimization program. UCSF Chimera was used to visualize and generate figures.[3] The theoretical pK_a value for the histidine residue in all the lysozyme variants was calculated using a program H++ (with both options, the Generalized Born and the Poisson-Boltzmann methods).[26]

Experimental methods: Generation of T4 lysozyme variants

T4 lysozyme gene in the plasmid pHS1403 was used as a template to generate all the variants. Single mutagenic[27] and overlap extension (consisting of two rounds of PCR)[27-29] PCR techniques were used to generated the PCR product containing the point mutation. The mutated gene was ligated into the vector pHS1403, replacing the original T4 lysozyme gene, and then introduced by transformation into the *E. coli* strain RR1.[24] In each case, the CAC histidine codon and the entire gene were confirmed by DNA sequencing.

The production and purification of the lysozyme variants was performed using a modification of a published method.[30] Over-expression of T4 lysozyme variants was induced by isopropyl β-D-thiogalactopyranoside (IPTG). *E. coli* cells containing the mutated gene were grown, induced with IPTG, centrifuged to harvest the cells, and stored at -80 ºC.[30] Frozen cells were resuspended in 20 ml of 20 mM sodium phosphate buffer at pH 7.2 (Buffer A). After adding a few crystals of DNase I, the cells were lysed by sonication (3 × 30 sec with intermittent cooling on ice). After removal of cell debris by centrifugation at 4 ºC, (13000 × g for 30 min), the supernatant fraction was dialyzed

against Buffer A. The dialyzed protein solution was then applied to a XK 16/20 Carboxyl Methyl Sepharose Fast Flow ion exchange column (Amersham Bioscience, Piscataway, NJ) equilibrated with Buffer A. The variants were eluted by a salt gradient (they elute at ~ 0.2 M NaCl in Buffer A containing 0.01% sodium azide). The eluted proteins were then dialyzed against Buffer A for 45 min to lower the salt concentration below 0.2 M NaCl, and the above chromatographic step was repeated with a shallower gradient elution. Before conducting the IMAC experiments, the proteins were concentrated by ultrafiltration using a YM3 membrane (Amicon 8200, Millipore Corp., Bedford, MA) and then followed by a buffer exchange step with 20 mM sodium phosphate, 1 M NaCl at pH 7.0 (Buffer B) in an Econo-pac® desalting column (Bio-Rad laboratories, Hercules, CA). The final protein purity of all variants was more than 95%, determined by SDS-PAGE. The lysozyme concentration (mg/ml) was determined by measuring the absorbance at 280 nm using an extinction coefficient of 1.28.[24] The protein sequence (including the histidine residue) of all the variants, except K43H, was verified by electrospray ionization mass spectrometry (Table 1).

Lysozyme activity assay

The enzyme activity of T4 lysozyme was determined by measuring the rate of clearance of a *Micrococcus lysodeiktikus* cell suspension using a modification of a published method.[31] The *M. lysodeiktikus* cell concentration was 0.5 mg dry weight/ml, obtained by suspending the cells in 20 mM sodium phosphate, 50 mM NaCl (pH 6.0). After adding lysozyme sample (1 μg), the rate of clearance of the *M. lysodeiktikus* cells was measured using microplate reader (Bio-Tek Instruments, Inc., Winooski, VT) as the rate of change in the absorbance at 540 nm. One unit of T4 lysozyme activity is defined

as the amount of lysozyme required to catalyze a decrease in absorbance of 0.001 in one minute.

Zonal analysis

The binding strength (association constant) of the protein towards the immobilized metal ion in the IMAC column was determined using an isocratic elution method known as zonal analysis, where the inhibitor concentration in the protein sample is equal to that in the mobile phase (elution buffer).[21, 22, 32] All experiments were conducted at room temperature on the TSK gel column. The TSK gel column was mounted onto a HPLC system equipped with a quaternary pump, HP 1050 (Agilent Technologies, Palo Alto, CA). All solutions used were filtered and degassed using a sonicator prior to the HPLC experiments. The flow rate of the mobile phase was 0.5 ml/min.

Experimental run: Regeneration of the column

TSK gel column, with iminodiacetic acid as the chelating agent, was regenerated by washing with 3 column volumes (CV) of 50 mM EDTA solution containing Buffer B to elute all the metal ions, followed by washing with (3 CV) distilled water. The column was reloaded with copper ions using (4 CV) 50 mM $CuSO_4$. The column was then washed with (5 CV) 100 mM sodium acetate, 1 M NaCl (pH 4.0) to remove excess copper ions. The unabsorbed copper ions were removed by washing with distilled water (2 CV). The Column was equilibrated with (8 CV) Buffer B containing I mM of imidazole. Then a protein sample of 50 µl of lysozyme variant (1 mg/ml) in Buffer B containing I mM of imidazole, was injected onto the TSK column by an auto sampler equipped on the HP 1050 system.

The column was regenerated before every experimental run. A set of four experimental runs, each of which was run with a particular imidazole concentrations *[I]* (I=3, 6, 9, 12 mM), were carried out for each variant to determining the respective elution volume (V). The protein sample also contained the same concentration of imidazole (as a competitor with protein for binding in IMAC) as that in the elution buffer. This set of experiments was repeated two more times (triplicates). The elution volume in the absence of interaction (V_0) was obtained by performing an isocratic elution (20 mM sodium phosphate, 1 M NaCl, 0.5 M imidazole) of T4 lysozyme (50 µl) in the absence of the copper ion. The presence of NaCl and imidazole will help in preventing any non-specific interaction that might occur between the column matrix and the lysozyme.

Determination of the copper concentration

The concentration of the immobilized Cu^{2+} on the TSK gel column was determined by spectral analysis, using inductively coupled plasma (ICP) spectrometry (Soil Testing Laboratory, Virginia Tech, Blacksburg, VA). The total copper ion concentration was determined on the fresh column and also after a considerable number (~ 55) of experimental runs. To avoid any loss of immobilized Cu^{2+}, the column was recharged before every experimental run.

Quantitative data analysis

The protein binding strength was determined using the following equation (1),[21]

$$\frac{1}{V - V_0} = \frac{1}{K_{aP} \bullet (V_0 - V_m) \bullet B_t} + \frac{K_{aI} \bullet [I]}{K_{aP} \bullet (V_0 - V_m) \bullet B_t} \qquad (1)$$

where V = elution volume to elute half of the protein (peak maximum), V_0 = elution volume in the absence of interaction (0.78 ml), V_m = void volume in the column

determined by blue dextran (0.44 ml), $[I]$ = concentration of the inhibitor (imidazole) in the mobile phase, K_{aI} = the association constant of the imidazole-copper complex, K_{aP} = association constant for the interaction between the protein and the copper ion in IMAC (protein binding strength), and B_t = total immobilized copper ions (23.2 μmol/ml). The elution volume of the protein from the IMAC column varies with the inhibitor concentration $[I]$. Therefore, a plot between $\frac{1}{V - V_0}$ vs. $[I]$ would be linear. K_{aI} is equal to the ratio of the slope to the intercept, and $\frac{1}{K_{aP}}$ can be obtained from the intercept and the independently obtainable parameters V_0, V_m, and B_t.

Results and Discussion

Based on the surface accessibility values, seven sites with varying relative surface accessibility (rSA) were chosen to generate the seven lysozyme variants (Table 2). Six of the seven native amino acids are lysines (K), and the other is arginine (R). The distribution of the seven selected sites on the T4 lysozyme surface, in relation to the amino acids that are catalytically active,[33] is shown in Fig. 1. The rSA values of the histidine residue on each variant range from 15% to 74% (Table 2). Although the rSA of K147H was equal to that of K124H, both were chosen for this study, because molecular modeling predicted the presence of a hydrogen bond between the unprotonated imidazole nitrogen (Nε2) of His-147 and the hydroxyl hydrogen of Tyr-139 in K147H (Fig. 2). Therefore the goal for analyzing the variant K147H, in addition to K124H, was to understand the behavior of the protein molecule in an IMAC column, when its donor amino acid is predictably involved in an intramolecular hydrogen bond. The enzymatic

activity of all the variants is shown in Table 3. All the variants were enzymatically active indicating that their 3D structures were largely intact. These results are consistent with the results reported by Dao-Pin et al., wherein the 3D structures of the T4-lysozyme variants containing K83H were intact.[34]

The microenvironment effects of a histidine residue can be divided into two components, a) the involvement of histidine in intramolecular bonds (hydrogen bonds or salt bridges),[35] and b) the effects of electrostatic environment, which involves short and long range charge-charge interactions.[36] Though the *in silico* pK_a value of the histidine residue, in all the variants, was calculated (data not shown), we were not able to draw any inferences regarding microenvironment effects. Nevertheless, the present set of IMAC experiments were able to reveal the effects of microenvironment due to the first component (the intramolecular interactions), as discussed in the following sections. However, to understand the effects due to the second component, further experimental work is required and is currently underway.

Zonal analysis using IMAC

The total concentration of copper ions immobilized on the TSK gel was 23.2 μmol/ml, which is within the concentration range given by the manufacturer (20 μmol/ml). Also, the total copper ion concentration after a number of experimental runs did not change significantly (~ 2 μmol/ml). The elution profiles of the variants R80H and K147H are shown in Fig. 3. Clearly, these elution profiles are highly reproducible. The time required to half elute the protein (V) for all the variants (the time at which the maximum absorbance occurs) was obtained from the elution profiles. The binding strength, K_{aP}, was calculated from the intercept of the plot of $\frac{1}{V - V_0}$ vs. $[I]$ (Fig. 4).

The K_{aP} indicates the strength of the binding between the immobilized copper ion and the protein (via histidine residue). The K_{aP} value for all the variants is shown in Table 4. It is important to note that the wild-type T4 lysozyme, which contains a histidine residue (His-31) that is buried and engaged in an intramolecular salt bridge,[37] did not bind to the Cu^{2+}-column. The binding strength and the rSA values are plotted in Fig. 5. In general, as the surface accessibility of the (lone) histidine increases, the binding strength of the protein to immobilized copper ions increases. However, both K147H and K83H variants showed significantly lower binding strength compared with the mutants containing histidines with similar surface accessibility.

Variant K147H

In the case of variant K147H, the binding strength is approximately 50% that of K124H. However, through molecular modeling, we predicted that His-147 would engage in a hydrogen bond with Tyr-139 as shown in Fig. 2. Thus, the involvement of a surface histidine in a hydrogen bond appears to significantly reduce the interaction between the protein and the immobilized metal ions. On the other hand, the fact that K147H binds to the IMAC column is a little surprising. Berna et al. showed that in chymotrypsin, His-57 (surface accessible area ~ 64 $Å^2$) had no interaction with the immobilized copper ions due to its participation in intramolecular hydrogen bonds. In contrast, His-41 in chymotrypsin was shown to interact with copper ions, although His-41 has smaller surface accessible area (~ 20 $Å^2$).[11] These conflicting results (between our and that reported by Berna et al.) can be explained by the fact that 1) the number of hydrogen bonds involved may differ, and 2) the strength of all hydrogen bonds is not equal.[38] The strength of a hydrogen bond depends on a number of factors such as 1) the microenvironment of both donor and

acceptor atoms,[38, 39] 2) the type and the stereoelectronic effects of the donating group (e.g. syn- or anti- orientation of the carboxylate ions),[40] 3) the distance and the particular angle of the donor and the accepting atoms,[38, 41] and 4) in the case of hydrogen bonds involving histidines, the proper orientation and the tautomeric state of the imidazole ring.[38]

Variant K83H

For variant K83H, which has a rSA almost equivalent to that of K135H, the binding strength is about 15% that of K135H. As predicted by molecular modeling, there was no evidence of His-83 being involved in a hydrogen bond with any of the neighboring amino acid residues because there was neither a donor nor an acceptor group within 3 Å, which is considered to be the distance threshold for the formation of a hydrogen bond.[42] Nevertheless, results reported by Dao-Pin et al., from the X-ray crystal structure of a T4-lysozyme variant that contained K83H, implicated that the Nδ1 hydrogen of His-83 was involved in a weak hydrogen bond (3.5 Å) with the Oδ1 oxygen of Asn-81.[34] Consequently, the 3D structure (by molecular modeling) of the variant K83H in this study was scrutinized and the presence of a possibly weak intramolecular hydrogen bond (3.63 Å) was discovered between Oδ1 oxygen of Asn-81 and Nδ1 hydrogen of His-83 (Fig. 6). This argument is valid because the imidazole group of histidine residue can exist in the doubly protonated form,[43] which may be stabilized by the electronegative Oδ1 oxygen of Asn-81 (Fig. 6). Thus, this results in the decrease of the binding strength of the variant K83H, when compared to K135H.

Histidine-related protein surface structure derived from protein binding strength

Knowing that the histidine residues in K83H and K147H are involved in intra-molecular interactions, the correlation between the protein binding strength and the rSA is thus generated excluding these two variants. As shown in Fig.5, there is a direct correlation ($R^2 = 0.76$) between the rSA and the protein binding strength. Using this correlation, we can predict not only the surface accessibility of a histidine residue on a protein surface but also the presence of intramolecular hydrogen bonds involving the reporter amino acid residue, histidine, from the measured protein-metal ion binding strength. These intramolecular bonds may or may not be captured by molecular modeling and often can only be confirmed by X-ray crystallography or NMR analysis. In addition, K147H and K83H showed slightly different binding strength with immobilized copper ions. In combination with the reported result that the presence of a hydrogen bond in chymotrypsin (His-57) completely prevented it from interacting with copper ions,[11] it is possible that IMAC, at least with immobilized copper ions, may be used as a tool to detect the strength of a hydrogen bond.

Conclusions

Based on the surface accessibility, seven lysozyme variants were generated, each containing one surface histidine residue. An IDA-Cu^{2+} IMAC column was used to determine the binding strength of these lysozyme variants. The binding strength was shown to have a direct correlation with the rSA of the histidine residues, except for the variants K83H and K147H, whose histidines were shown to be involved in intramolecular hydrogen bonds. Thus, based on the binding strength of a protein, IMAC may be used to determine the histidine-related protein surface structure including the surface accessibility of histidine residues and the presence of hydrogen bonds, which

generally requires an X-ray or NMR structure. It is also possible that this method could be used to evaluate the strength of a hydrogen bond. The results shown in this report strongly suggest that IMAC can be used as a quantitative tool to characterize proteins to derive its histidine-related surface structure. Finally, the experiments reported in this paper were designed to study the influence of surface accessibility on the binding strength of a protein containing only one surface histidine. Our future efforts would be focused on studying the effects of the microenvironment of the histidine residue on the binding strength of a protein, and also consider proteins having multiple surface histidines.

Acknowledgements

We are thankful to Janet Donahue in Dr. Timothy Larson's laboratory (Dept. of Biochemistry, Virginia Tech) for her assistance in DNA cloning. We also thank Dr. Brain Matthews from the Institute of Molecular Biology and the Department of Physics at the University of Oregon (Eugene, OR) for providing us with the DNA of wild-type T4 lysozyme. This work is mainly funded by a grant from Jeffress Memorial Trust, Grant # J-706.

Table 1. The peptide sequences of the six lysozyme variants obtained from electrospray ionization mass spectrometry. The peptide sequences were identified by SEQUEST. The histidine residue is shown in bold letters.

Variant	Peptide position	Peptide sequence
K19H	17-35	IY**H**DTEGYYTIGIGHLLTK
R80H	77-83	GIL**H**NAK
K83H	81-95	NA**H**LKPVYDSLDAVR
K124H	120-135	MLQQ**H**RWDEAAVNLAK
K135H	126-137	WDEAAVNLA**H**SR
K147H	146-154	A**H**RVITTFR

Table 2. Relative surface accessibility (rSA), the percentage of surface accessibility of the residue (side chain) in the folded protein divided by the SA of the residue (side chain) in the unfolded state, of lysozyme variants calculated by using a probe with a radius of 1.93 Å.

Variant	rSA (%) (original amino acid)	rSA (%) (histidine residue)
K19H	62	50
K43H	26	15
R80H	62	74
K83H	60	62
K124H	48	47
K135H	69	63
K147H	49	47

Table 3. Enzymatic activities of purified T4 lysozyme variants. The activities of all the lysozyme variants are compared with the wild type.

Variant	(%) Activity
Wild type	100
K19H	112
K43H	110
R80H	126
K83H	131
K124H	116
K135H	162
K147H	102

Table 4. Protein-metal ion binding strength and the relative surface accessibility (rSA) values for T4 lysozyme variants.

Variant	rSA (%) (histidine residue)	Protein binding strength (K_{aP}) (10^1 M^{-1})
K19H	50	28.4
K43H	15	1.3
R80H	74	59.6
K83H	62	11.4
K124H	47	17.8
K135H	63	74.3
K147H	47	6.4

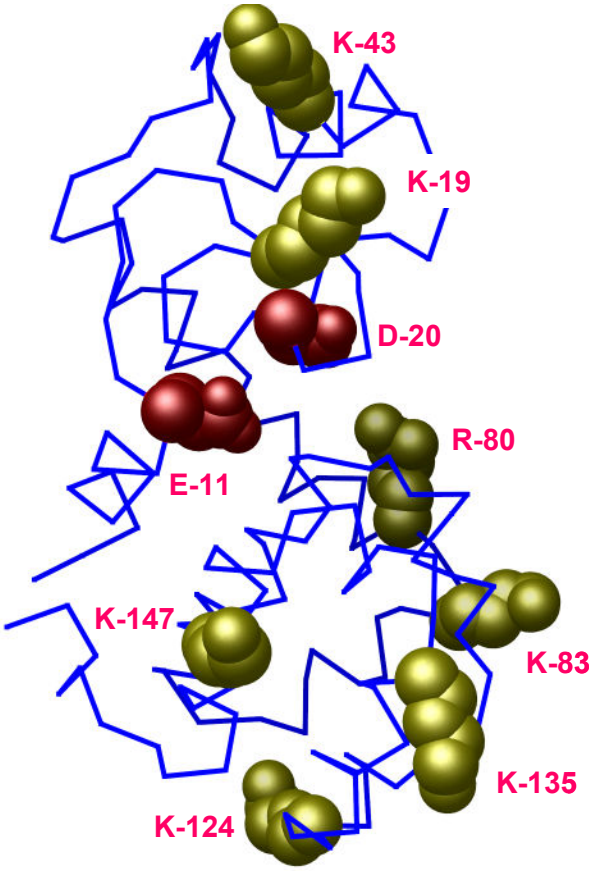


Figure-1: Sites Selected for histidine replacement (amino acid substitutions) on the surface of T4 lysozyme are shown in yellow color. The amino acid residues that are catalytically active in T4 lysozyme (Glu-11 and Asp-20) are shown in red. This figure was generated using UCSF Chimera.

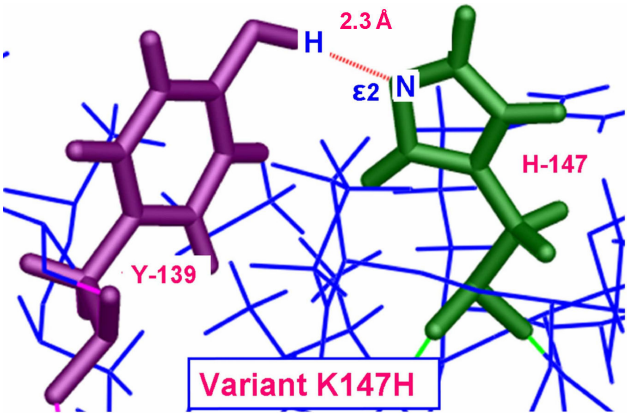
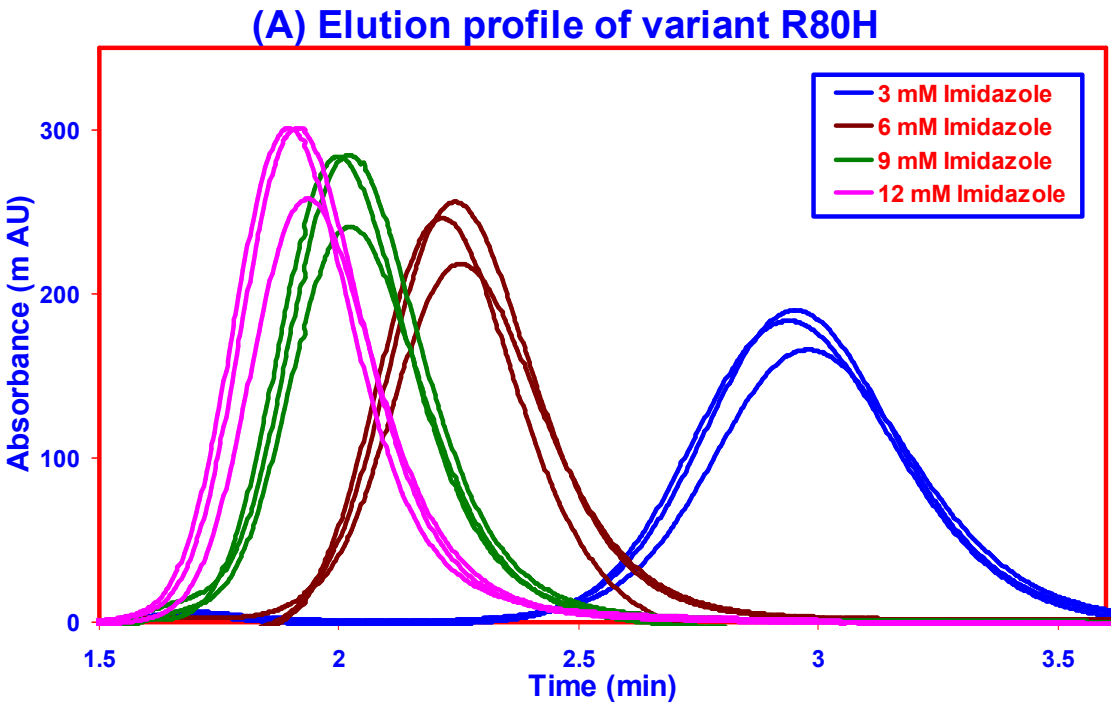


Figure-2: Hydrogen bond (distance 2.32 Å) between the imidazole Nε2 atom of His-147 (donor) and the hydroxyl hydrogen atom of Try-139 (acceptor). The figure was generated using UCSF Chimera. This is a sectional view of the molecule around His-147.


(A) Elution profile of variant R80H
3 mM Imidazole
6 mM Imidazole
9 mM Imidazole
12 mM Imidazole
Absorbance (m AU)
300
200
100
0
1.5
2
2.5
3
3.5
Time (min)

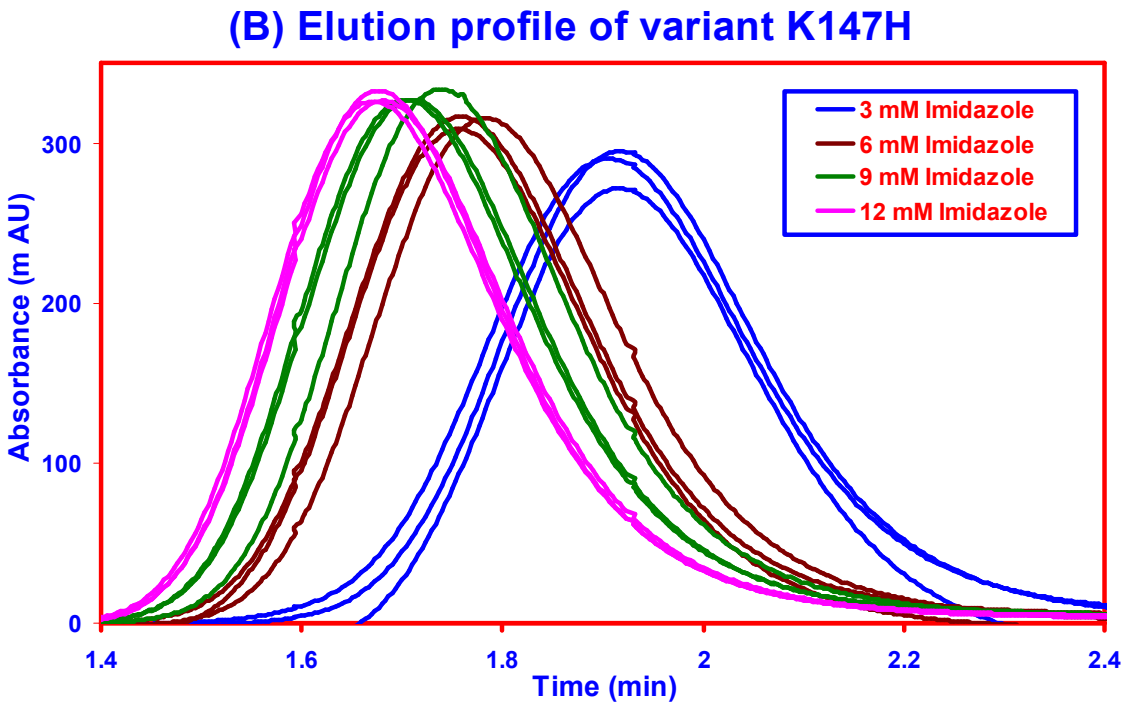


Figure-3: Elution profiles of variants (A) R80H and (B) K147H obtained in triplicate, for the four imidazole concentrations, 3, 6, 9 and 12 mM in the mobile phase. The IMAC column was regenerated for each and every elution profile.

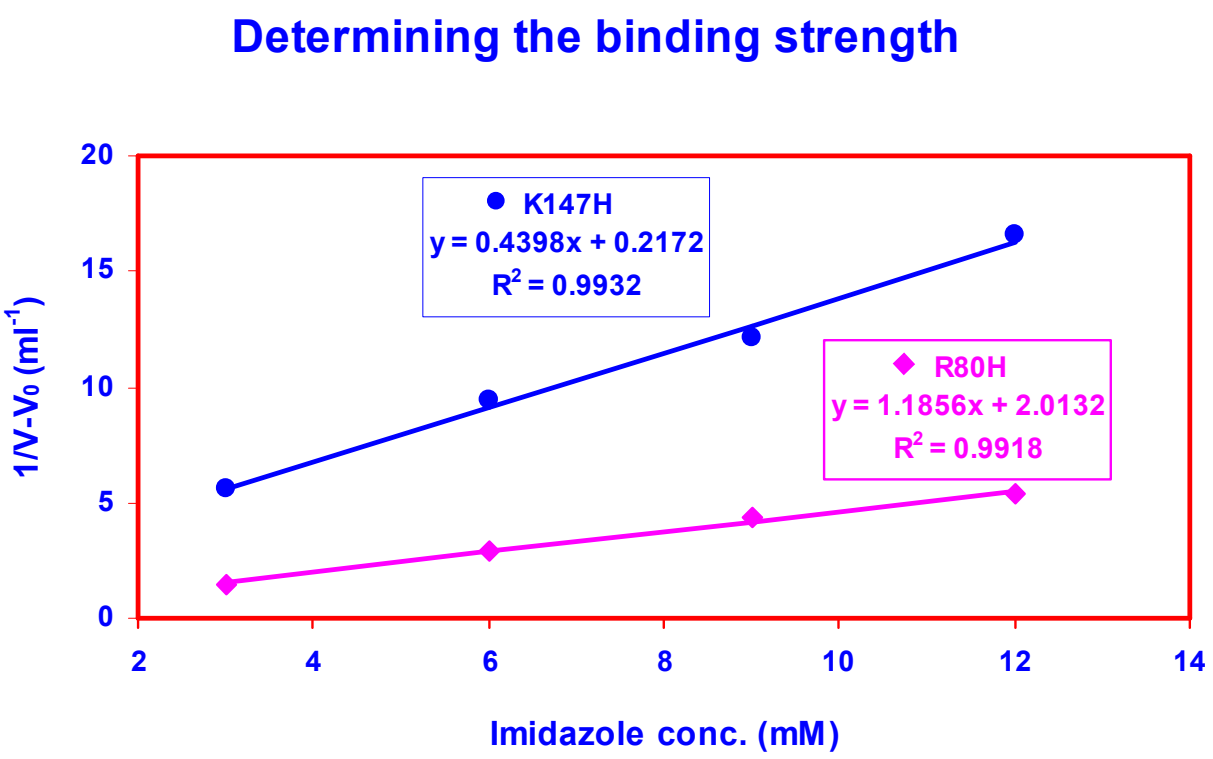


Figure-4: Determination of the $\frac{1}{V - V_0}$ intercept required to calculate the binding strength of variants R80H and K147H.

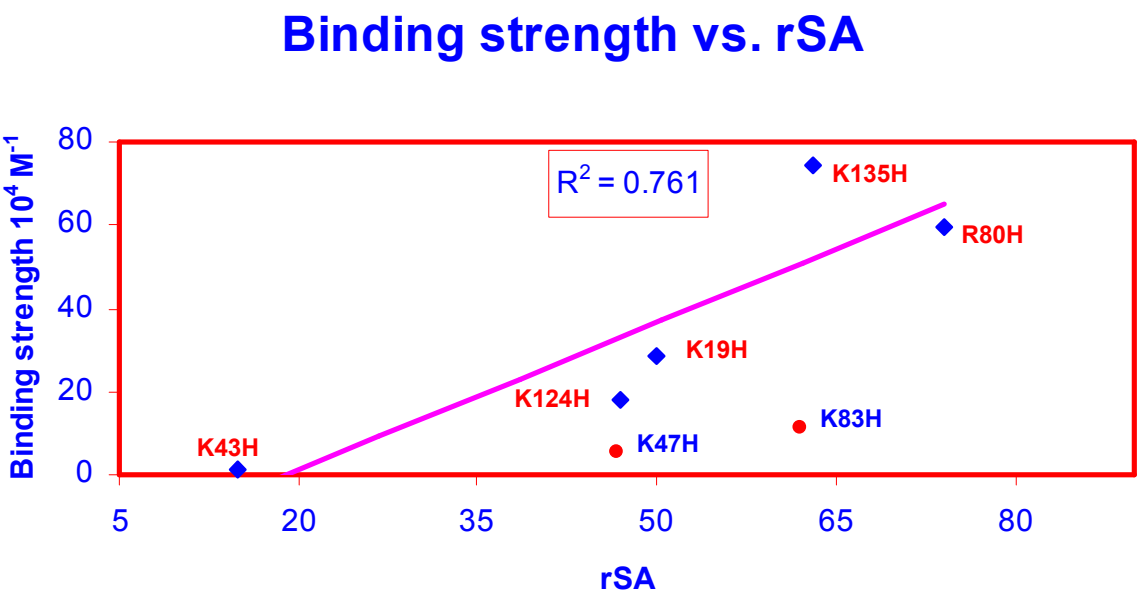


Figure-5: Correlation between the relative surface accessibility (rSA) and the binding strength of T4 lysozyme variants, excluding the two variants K147H and K83H. The histidines of these two variants (K83H and K147H) are involved in intramolecular hydrogen bonds, shown as points that are circled (in red).

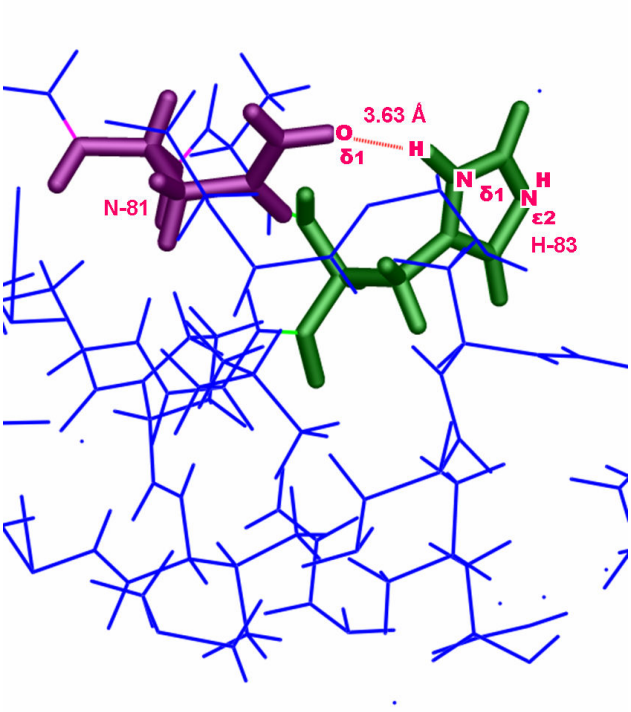


Figure-6: The presence of a weak intramolecular hydrogen bond between Oδ1 oxygen of Asn-81 at distance of 3.63 Å away from the Nδ1 hydrogen of His-83 (doubly protonated form). The figure was generated using UCSF Chimera. This is a sectional view of the molecule around His-83.

References

1. Koths, K., Recombinant Proteins for Medical Use - the Attractions and Challenges. *Current Opinion in Biotechnology* **1995,** 6, (6), 681-687.
2. Walsh, G., Biopharmaceutical benchmarks - 2003. *Nature Biotechnology* **2003,** 21, (8), 865-870.
3. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** 25, (13), 1605-1612.
4. Kleinberg, M.; Mosdell, K. W., Current and future considerations for the new classes of biologicals. *American Journal of Health-System Pharmacy* **2004,** 61, (7), 695-708.
5. Schellekens, H., When biotech proteins go off-patent. *Trends in Biotechnology* **2004,** 22, (8), 406-410.
6. Schellekens, H., Follow-on biologics: challenges of the 'next generation'. *Nephrology Dialysis Transplantation* **2005,** 20, 31-36.
7. Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G., Hydrogen-Bonding and Biological Specificity Analyzed by Protein Engineering. *Nature* **1985,** 314, (6008), 235-238.
8. Chicz, R. M.; Regnier, F. E., Immobilized-Metal Affinity and Hydroxyapatite Chromatography of Genetically Engineered Subtilisin. *Analytical Chemistry* **1989,** 61, (15), 1742-1749.
9. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C., The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry A* **1997,** 101, (16), 3005-3014.
10. Jiang, K. Y.; Pitiot, O.; Anissimova, M.; Adenier, H.; Vijayalakshmi, M. A., Structure-function relationship in glycosylated alpha-chymotrypsin as probed by IMAC and IMACE. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1999,** 1433, (1-2), 198-209.
11. Berna, P. P.; Mrabet, N. T.; VanBeeumen, J.; Devreese, B.; Porath, J.; Vijayalakshmi, M. A., Residue accessibility, hydrogen bonding, and molecular recognition: Metal-chelate probing of active site histidines in chymotrypsins. *Biochemistry* **1997,** 36, (23), 6896-6905.
12. Todd, R. J.; Johnson, R. D.; Arnold, F. H., Multiple-Site Binding Interactions in Metal-Affinity Chromatography .1. Equilibrium Binding of Engineered Histidine-Containing Cytochrome-C. *Journal of Chromatography A* **1994,** 662, (1), 13-26.
13. Russell, A. J.; Fersht, A. R., Rational Modification of Enzyme Catalysis by Engineering Surface-Charge. *Nature* **1987,** 328, (6130), 496-500.
14. Sundberg, R. J.; Martin, R. B., Interactions of histidines and other derivatives with transitional metal ions in chemical and biological systems. *Chemical Reviews* **1974,** 74, 471-516.
15. Meirovitch, H.; Scheraga, H. A., Empirical-Studies of Hydrophobicity .2. Distribution of the Hydrophobic, Hydrophilic, Neutral, and Ambivalent Amino-Acids in the Interior and Exterior Layers of Native Proteins. *Macromolecules* **1980,** 13, (6), 1406-1414.

16. Sulkowski, E., Purification of Proteins by Imac. *Trends in Biotechnology* **1985,** 3, (1), 1-7.
17. Hemdan, E. S.; Zhao, Y. J.; Sulkowski, E.; Porath, J., Surface-Topography of Histidine-Residues - a Facile Probe by Immobilized Metal-Ion Affinity-Chromatography. *Proceedings of the National Academy of Sciences of the United States of America* **1989,** 86, (6), 1811-1815.
18. Boden, V.; Rangeard, M. H.; Mrabet, N.; Vijayalakshmi, M. A., Histidine mapping of serine protease: a synergic study by IMAC and molecular modelling. *Journal of Molecular Recognition* **1998,** 11, (1-6), 32-39.
19. Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C., Interior and Surface of Monomeric Proteins. *Journal of Molecular Biology* **1987,** 196, (3), 641-656.
20. Hamelryck, T., An amino acid has two sides: A new 2D measure provides a different view of solvent exposure. *Proteins-Structure Function and Bioinformatics* **2005,** 59, (1), 38-48.
21. Dunn, B. M.; Chaiken, I. M., Quantitative affinity chromatography, determination of binding constants by elution with competitive inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **1974,** 71, 2382-2385.
22. Dunn, B. M.; Chaiken, I. M., Evaluation of Quantitative Affinity Chromatography by Comparison with Kinetic and Equilibrium Dialysis Methods for Analysis of Nucleotide Binding to Staphylococcal Nuclease. *Biochemistry* **1975,** 14, (11), 2343-2349.
23. Sagermann, M.; Gay, L.; Matthews, B. W., Long-distance conformational changes in a protein engineered by modulated sequence duplication. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** 100, (16), 9191-9195.
24. Poteete, A. R.; Sun, D. P.; Nicholson, H.; Matthews, B. W., 2nd-Site Revertants of an Inactive T4 Lysozyme Mutant Restore Activity by Restructuring the Active-Site Cleft. *Biochemistry* **1991,** 30, (5), 1425-1432.
25. Daopin, S.; Soderlind, E.; Baase, W. A.; Wozniak, J. A.; Sauer, U.; Matthews, B. W., Cumulative Site-Directed Charge-Change Replacements in Bacteriophage-T4 Lysozyme Suggest That Long-Range Electrostatic Interactions Contribute Little to Protein Stability. *Journal of Molecular Biology* **1991,** 221, (3), 873-887.
26. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **2005,** 33, W368-W371.
27. Shen, B., *PCR approaches to DNA mutagenesis and recombinantion*. 2 ed.; Humana Press: New Jersey, 2002; Vol. 192, p 167-174.
28. Higuchi, R., *Using PCR to engineer DNA*. Stockton Press: New York, 1989; p 61-71.
29. Li, S. J.; Chen, P. C.; Li, W.; Hao, X.; Yang, G. F., Establishing mathematical and physical models for the adsorption of biomacromolecules. *Applied Biochemistry and Biotechnology* **2006,** 134, (2), 165-178.
30. Luther, J. R.; Glatz, C. E., Genetically-Engineered Charge Modifications to Enhance Protein Separation in Aqueous 2-Phase Systems - Electrochemical Partitioning. *Biotechnology and Bioengineering* **1994,** 44, (2), 147-153.

31. Balasubramaniam, D.; Wilkinson, C.; Van Cott, K.; Zhang, C. M., Tobacco protein separation by aqueous two-phase extraction. *Journal of Chromatography A* **2003,** 989, (1), 119-129.
32. Johnson, E. M.; Berk, D. A.; Jain, R. K.; Deen, W. M., Diffusion and Partitioning of Proteins in Charged Agarose Gels. *Biophysical Journal* **1995,** 68, (4), 1561-1568.
33. Hardy, L. W.; Poteete, A. R., Reexamination of the Role of Asp20 in Catalysis by Bacteriophage-T4 Lysozyme. *Biochemistry* **1991,** 30, (39), 9457-9463.
34. Daopin, S.; Sauer, U.; Nicholson, H.; Matthews, B. W., Contributions of Engineered Surface Salt Bridges to the Stability of T4 Lysozyme Determined by Directed Mutagenesis. *Biochemistry* **1991,** 30, (29), 7142-7153.
35. Garcia-Mayoral, M. F.; Perez-Canadillas, J. M.; Santoro, J.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Lacadena, J.; del Pozo, L. M.; Gavilanes, J. G.; Rico, M.; Bruix, M., Dissecting structural and electrostatic interactions of charged groups in alpha-sarcin. An NMR study of some mutants involving the catalytic residues. *Biochemistry* **2003,** 42, (45), 13122-13133.
36. Laurents, D. V.; Huyghues-Despointes, B. M. P.; 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.; Pace, C. N., 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). *Journal of Molecular Biology* **2003,** 325, (5), 1077-1092.
37. Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W., Ph-Induced Denaturation of Proteins - a Single Salt Bridge Contributes 3-5 Kcal Mol to the Free-Energy of Folding of T4-Lysozyme. *Biochemistry* **1990,** 29, (9), 2403-2408.
38. Day, P. J.; Shaw, W. V.; Gibbs, M. R.; Leslie, A. G. W., Acetyl Coenzyme-a Binding by Chloramphenicol Acetyltransferase - Long-Range Electrostatic Determinants of Coenzyme-a Recognition. *Biochemistry* **1992,** 31, (17), 4198-4205.
39. Gandour, R. D.; Nabulsi, N. A. R.; Fronczek, F. R., Structural Model of a Short Carboxyl Imidazole Hydrogen-Bond with a Nearly Centrally Located Proton - Implications for the Asp-His Dyad in Serine Proteases. *Journal of the American Chemical Society* **1990,** 112, (21), 7816-7817.
40. Zimmerman, S. C.; Cramer, K. D., Stereoelectronic Effects at Carboxylate - a Syn Oriented Model for the Histidine-Aspartate Couple in Enzymes. *Journal of the American Chemical Society* **1988,** 110, (17), 5906-5908.
41. SYBYL., Online help manual. *Tripos Inc., St. Louis, USA*.
42. Harris, T. K.; Mildvan, A. S., High-precision measurement of hydrogen bond lengths in proteins by nuclear magnetic resonance methods. *Proteins-Structure Function and Genetics* **1999,** 35, (3), 275-282.
43. Bachovchin, W. W., N-15 Nmr-Spectroscopy of Hydrogen-Bonding Interactions in the Active-Site of Serine Protease - Evidence for a Moving Histidine Mechanism. *Biochemistry* **1986,** 25, (23), 7751-7759.

Chapter IV: Quantifying protein microstructure and electrostatic effects on the change in Gibbs free energy of binding in immobilized metal affinity chromatography

Abstract

Electrostatic forces play a major role in maintaining both structural and functional properties of proteins. A major component of protein electrostatics is the interactions between its charged or titratable amino acid residues (e.g. Glu, Asp and His). The titratable groups in a protein, due to the local environmental effects, display different pK_a values from those in peptides or free amino acids. One way to study the effects of electrostatic forces is by measuring the change in the pK_a (ΔpK_a) value of the titratable groups. Because of the heterogeneous nature of protein structure, predicting the ΔpK_a values by different theoretical models is challenging. Here, we report the study of electrostatic forces through experiments using a well-controlled model protein (T4 lysozyme) and its variants. We generated ten T4 lysozyme variants, in which the electrostatic environment of the histidine residue was perturbed by altering charged and neutral amino acid residues at various distances from the histidine residue. The electrostatic perturbations were theoretically quantified by calculating the change in free energy ($\Delta\Delta G_E$) due to the change in electrostatic free energy (ΔE) obtained from simple Coulomb's law. On the other hand, immobilized metal affinity chromatography (IMAC) was used as an experimental tool to quantify these perturbations in terms of protein binding strength or change in free energy of binding due to perturbation ($\Delta\Delta G_B$). Our results demonstrate a good correlation ($R^2 = 0.97$) between the theoretical $\Delta\Delta G_E$ and

experimental $\Delta\Delta G_B$ values. The $\Delta\Delta G_B$ values vary from -0.53 to 0.99 kcal/mol. NMR experiments were also performed, and the data support the results obtained from IMAC. However, among the ten variants studied, the results from three variants do not follow the correlation. The probed histidine residue on these deviant variants was found to be involved in site-specific interactions (e.g. ion pair and steric hindrance), which were further investigated by molecular dynamics simulation. This report demonstrates that the electrostatic ($\Delta\Delta G_{Elec}$) and microstructural effects ($\Delta\Delta G_{Micro}$) in a protein can be quantified by IMAC through surface histidine mediated protein-metal ion interaction, and equally importantly, the unique microstructures around a histidine residue in a protein can also be identified by identifying the abnormal binding behaviors during IMAC.

Key words: Protein microstructure, electrostatic forces, Gibbs free energy, immobilized metal affinity chromatography, protein structure quantification

Introduction

In general, almost all important biochemical processes in living systems require proteins. Proteins are large, heterogeneous, flexible and complex molecules.[1-4] It is believed that there is an intricate relationship between protein structure and function. Consequently any subtle change that might occur in a protein's structure can affect its function, which is generally modulated by the amino acid residues on its surface [5]. One way to characterize the role played by protein surfaces is by calculating its electrostatic energy (or potential).[5] In fact, electrostatics [6] play a major role in maintaining both the structural and functional properties of proteins.[7-11] For example, molecular level interactions such as salt bridges and hydrogen bonds, which are primarily electrostatic in nature [2, 4, 12], are critical not only in maintaining the structure of a protein but also in defining its specific function by stabilizing the charged transition states during catalytic reactions [13-15]. Therefore, it is essential to understand the role played by electrostatic forces to decipher protein structure-function relationships.[9, 16]

Analyzing electrostatic forces not only elucidates structure-function relationships but also aid in engineering proteins rationally.[6] However, theoretical assessment of electrostatic forces has been challenging due to the inability to extend current models to analyze the heterogeneous nature of protein structures.[7, 17-19] Therefore a combined strategy of utilizing both theory (to predict) and experimentation (to verify) is required to understand protein electrostatics.[6, 20-22] One way to study the effects of electrostatic forces is by measuring the pK_a value of titratable amino acid residues.[9, 19, 21] Due to environmental effects in structured proteins, the titratable amino acid residues (here histidine) display different pK_a values from those in peptides or free amino acids.[6, 9, 15, 21,

[23-27] The environmental factors that influence the pK_a values of histidine are: 1) their surface accessibility (hydrophobic / hydrophilic effect) [6, 21, 26, 28], 2) their microenvironment [29] which involves a) the involvement of histidine residues in intramolecular interactions such as salt bridges and hydrogen bonds [21, 26, 30-32], b) histidine's microstructure which consists of short range charge-charge interactions that depends on the local zonal amino (polar / non-polar) acid distribution [33], and c) long range charge-charge interactions and overall protein net charge [7, 14, 15, 26, 32], 3) type, pH and ionic strength of the buffer solution [6, 9, 26], and 4) other factors, which are generally neglected, such as histidine residue's conformational effects [7, 34] including its direction [33] and its side chain flexibility [35, 36]. Understanding and quantifying the effects of these factors on the pK_a of the probe-histidine (a histidine residue directly participates in protein-metal ion interaction) would allow possible inference of the microenvironment around the histidine in proteins, which is otherwise difficult to obtain by other analytical methods.

Currently, the experimental techniques that are used to measure pK_a values of titratable groups are NMR and FT-IR.[21, 31] However, even though these techniques are reliable, they are costly and complicated. Other simpler and indirect techniques such as kinetic analysis using pH dependent catalytic activity have been used to measure the changes in pK_a (ΔpK_a) values.[10, 17, 21] For example, Fersht and co-workers used enzyme activity studies to calculate the change in the (ΔpK_a) values of histidine residue in subtilisin variants, and to quantify the electrostatic effects due to charge variation.[10, 17] However, these indirect techniques cannot be extended to study the pK_a values of titratable groups other than those involved in the catalytic process. Also, it has been

reported that in few cases the changes in the catalytic site are so complex that straightforward site-directed mutagenesis may not yield desired results [31]. Such problems can be addressed by utilizing techniques wherein the histidine group is not involved in the catalytic process. For example, Chicz and Regnier used site-directed protein mutagenesis to induce charge variation in the microenvironment of His-64 in subtilisin to illustrate that immobilized metal affinity chromatography (IMAC) [37-39] could be used to qualitatively distinguish these subtilisin variants via the differences in their retention time.[24] Similar studies were reported in the literature wherein IMAC was sensitive enough to measure variation in the microenvironment of a protein.[40, 41]

Previously we were able to establish a correlation between the protein-metal ion binding strength (K_{aP}) and the histidine surface accessibility.[29] In addition, we were able to demonstrate quantitatively the effect of intramolecular bonds, which is one of the factors comprising the histidine's microenvironment, through the protein-metal ion interaction in IMAC when there is only one histidine on the surface of a protein.[29] Our current goal is to understand the effect of the other factor, namely the electrostatic environment, on the histidine's electrostatic property (such as its pK_a) that is manifested through the protein's binding strength on IMAC.[29] Our hypothesis is that any subtle variation in the histidine's electrostatic environment affects its binding strength values via a change in its pK_a.

More specifically, the overall goal of our present work is to quantify the effects of electrostatic perturbation (charge variation), via change in Gibbs free energy of binding, in the probe-histidine's electrostatic environment. Fersht and co-workers quantified long range electrostatic perturbations (> 12 Å) using enzyme kinetics.[10] Such long range

studies elucidate only charge effects but not any associated structural effects, unless the perturbed amino acid is involved in a network contributing to the catalytic activity. In this report, all the sites chosen for charge modification around the histidine residue were within short or medium range (< 15 Å). A major advantage in perturbing short range residues is that they not only introduce strong electrostatic effects due to the proximity of the charge variation but also may impart site-specific microstructural effects on the to-be probed histidine residue.[42]

Theoretically, the electrostatic perturbation was obtained by calculating the change in net electrostatic energy (ΔE), which represents the cumulative effect of both charge and distance of the perturbation from the histidine. The ΔE values were used to calculate the change in Gibbs free energy due to the change in net electrostatic energy ($\Delta\Delta G_E$). Experimentally, IMAC was used as a quantitative tool to obtain protein binding strength.[29] The protein binding strength values were used to calculate the equivalent ΔpK_a and change in Gibbs free energy of binding due to electrostatic pertubation ($\Delta\Delta G_B$). NMR spectroscopy was also used to measure the ΔpK_a of the histidine.[9, 31, 43, 44] The specific objectives of this article are 1) to obtain a possible correlation between ($\Delta\Delta G_E$) and ($\Delta\Delta G_B$), and to use this correlation as a gauge to discover possible intramolecular interactions involving the binding histidine and the microstructure around it, and 2) to dissect the contribution of microstructural and electrostatic effects to ΔpK_a and $\Delta\Delta G_B$ values.

Materials and Methods

Site selection: Previously, we generated two variants, K135H and R80H, from cysteine-free T4 lysozyme.[29] For this report, these two variants were treated as pseudo wild types

(pWT-135 and pWT-80). Two sets of T4-lysozyme variants were generated by performing site-directed protein mutagenesis on the pWT proteins. The first set of variants has a histidine residue at position 135 (pWT-135), while the second set of variants has a histidine residue at position 80 (pWT-80). The two sites His135 and His80 were chosen because they have similar surface accessibility with high binding strength on IMAC-Cu^{2+} columns.[29] The variants containing His135 and His-80 were chosen as control variants to study whether or not a correlation exists between electrostatic pertubation effects and protein binding strength or $\Delta\Delta G_B$ in an IMAC column. Three sites Asp-127, Ser-136 and Arg-137 around His135 (Fig-1a) and two sites Arg-76 and Asp-108 around His80 (Fig-1b), which vary not only in charge but also in distance from the histidine residue, were chosen (Table 1). The distances reported are the Cα-Cα of original amino acids and the histidine. Instead of the side-chain atoms, Cα atoms were chosen because the orientation of the side-chain in replaced amino acids is hard to predict using molecular modeling tools.[45] All these sites were mutated to two different residues except for site Ser-136 (three mutations) and Glu-108 (one mutation) (Table 1). Therefore, the control set consists of seven variants; while the test set consists of only three variants. As a result, ten variants were generated from five sites, and all variants have a single mutation around the sole histidine residue (Table 1).

In silico molecular modeling methods: *In silico*, all the varaints with point mutation were generated using a program, SYBYL 7.1 (Tripos Inc., St. Louis). In general, the oreintation of the replacing residue is similar to the original residue in the wild-type, unless some specific steric constriants are present in the microstructure of the replacing residues.[46, 47] All variant structures were energy-minimized using the SYBYL program.[29]

The surface accessibility of all the sites was obtained by using a probe with a radius of 1.4 Å (Table 2). To visualize and generate figures, UCSF Chimera and Swiss-Pdb Viewer programs were utilized.[48, 49]

Characterization of protein microstructure: Relative hydrophobicity/hydrophilicity value (rHy), which characterizes the microenvironment around all the sites, was calculated using a modification of a published method [23, 33]. The sum of all the hydrophobic/hydrophilic constants (sHy) of the moieties around these sites (4.25 Å) was calculated manually from the X-ray crystal structure of cystiene free T4 lysozyme (1L63).[33] The rHy values of all the sites were obtained from the following equations [33] and are reported in Table 2.

$$tHy = (1 - SA) \cdot sHy + (SA) \cdot aHy \qquad (1)$$

$$rHy = \frac{tHy}{aHy} \qquad (1a)$$

Where, *SA* is surface accessibility, *tHy* is the total hydrophobicity, and *aHy* is the absolute hydrophilic value in water obtained from a method by Mehler et al.[33]

Molecular dynamics simulation: Molecular Dynamics (AMBER) was used to demonstrate the flexibility of the side chain. MD simulations were done using AMBER8. Structures were prepared for simulation using Leap. Counter ions (Cl^-) were added to neutralize the charge of the protein, and the protein was immersed in a box of TIP3P water, with no protein atom being closer than 9 Å to the edge of the water box. The system was equilibrated by running MD at 300 K of the water molecules and counter ions for 100-ps, after which the energy of the system was minimized, first minimizing the counter ions and water only, and then the entire system. MD of the entire system was begun by heating from 0 K to 300K over 30 ps and then equilibrating the system for an

additional 50 ps, all under constant volume conditions. Constant pressure conditions were then imposed for the remainder of the simulation, which was carried out for a total of 3000 ps. MD and energy minimizations were done using the sander module of AMBER8 running on the System X supercomputer at Virginia Tech.

Experimental: The production and purification of all lysozyme variants with aforementioned mutations (Table 1) was performed following the published method.[29] Mutations in the T4 lysozyme gene were carried out by single mutagenic [50] and overlap extension [50-52] PCR techniques. The point mutation in the lysozyme gene for all ten variants was verified by DNA sequencing. All lysozyme variants were overexpressed in *E. coli* cells, and purified using an XK 16/20 Carboxyl Methyl Sepharose Fast Flow ion exchange column (GE Health Care, Piscataway, NJ).[29] The final protein purity of all variants was determined by SDS-PAGE (> 95%). The protein sequence of all the variants was verified by trypsin digest followed by LC-MS, except for R76S, R137S and E108K (Table 3).

The final protein solution (in 20 mM sodium phosphate, pH 7.0, 0.5 M NaCl and 0.01 % sodium azide) used for conducting IMAC experiments was obtained by a buffer exchange step with Buffer B in an Econo-pac® desalting column (Bio-Rad laboratories, Hercules, CA). The absorbance value at 280 nm ($A^{1mg/ml}$ = 1.28) was used to determine the concentration of all lysozyme variants [53]. The enzyme activity of all ten lysozyme variants was determined by following the published method [29]. Briefly, lysozyme sample (1 μg) was added to a cell suspension (0.5 mg dry weight/mL) of *M. lysodeiktikus* in 20 mM sodium phosphate and 50 mM NaCl (pH 6.0). The enzyme activity is defined as the rate of clearance of *M. lysodeiktikus* cells, which was measured by the rate of change in

absorbance at 540 nm (in a microplate reader, Bio-Tek Instruments Inc., Winooski, VT). The enzyme activity of all ten variants and the wild type lysozyme are reported in Table 4.

CD and NMR spectroscopy: All CD experiments were conducted at room temperature using a J-720 spectropolarimeter (JASCO Incorporated, Easton, MD). The CD spectra were measured from wavelength 190-350 nm in quartz cuvettes of 1-mm path length. The protein concentration of all the samples was around 50 μg/ml. The sample buffer used was 20 mM sodium phosphate (pH 7.0).

For all NMR experiments, the sample buffer solution used was a D_2O solution containing 10 mM D_3PO_4 and 10 mM KCl. Protein samples were prepared using a modification of published methods.[54, 55] The samples containing 6-10 mg of lysozyme were concentrated from 4 ml to 1 ml by centrifugation at 7500×g using disposable Amicon ultra tubes (Millipore, Billerica, MA) having membranes with 5 kDa molecular-weight-cut-off. Then 3 ml of D_2O was added and the above centrifugation step was repeated, followed by addition of 3 ml of sample buffer solution, and then the centrifugation step was repeated. This step was repeated 3 more times. Then the protein samples were heated at 55 ^oC for 30 min to undergo reversible denaturation [54]. The samples were then slowly cooled and stored at 4 ^oC. Two protein stock solutions were prepared by adding NaOD (base stock) and DCl (acid stock). Then by adding varying amounts of protein acid-base stock solution, NMR samples with varying pH values (5.5-8.5) were generated. The pH of all the samples was measured before and after the NMR experiments. The maximum variation in the pH values was ± 0.03. All NMR experiments were conducted at room temperature using a 400-MHz NMR spectrometer.

Determination of protein binding strength (K_{aP}): IMAC experiments were conducted at room temperature on a TSK gel column (Tosoh Bioscience, Philadelphia, PA) mounted onto an HPLC system (HP 1050). The binding strength (protein-metal ion) in the IMAC column was determined using an isocratic elution method known as zonal analysis.[56] By varying the inhibitor (imidazole) concentration in the protein sample and the mobile phase, the association constant was determined using the following equation:[56]

$$\frac{1}{V - V_0} = \frac{1}{K_{aP} \cdot (V_0 - V_m) \cdot B_t} + \frac{K_{aI} \cdot [I]}{K_{aP} \cdot (V_0 - V_m) \cdot B_t} \qquad (2)$$

Where V = elution volume to elute half of the protein (peak maximum), V_0 = elution volume in the absence of interaction (0.79 ml), V_m = void volume in the column determined by blue dextran (0.44 ml), $[I]$ = concentration of the inhibitor (imidazole) in the mobile phase, K_{aI} = the association constant of the imidazole-copper complex, K_{aP} = association constant for the interaction between the protein and the copper ion in IMAC (protein binding strength), and B_t = total immobilized copper ions (20 μmol/ml). The elution volume of the protein from the IMAC column varies with the inhibitor concentration $[I]$. Therefore, a plot between $\frac{1}{V - V_0}$ vs. $[I]$ would be linear. K_{aI} is equal to the ratio of the slope to the intercept, and $\frac{1}{K_{aP}}$ can be obtained from the intercept and the independently obtainable parameters V_0, V_m and B_t.

Theoretical components contributing to Gibbs free energy of binding in IMAC: The interaction or the binding phenomena between lysozyme and the immobilized copper ion in IMAC can be considered as a protein-ligand interaction.[57] The protein-ligand

interaction is a thermodynamic process, which can be characterized by change in Gibbs free energy of binding (ΔG_B). In general ΔG_B can be defined by the following equation,

$$\Delta G_B = \Delta H - T\Delta S \qquad (3)$$

where, ΔH is the change in enthalpy and ΔS is the change in entropy of the interaction. The specific components that contribute to ΔG_B (in Eq. 3) are given in Eq. 4.[58] On the right hand side of Eq. 4, the first three ΔG terms contribute to ΔH, while the next two ΔG terms (inside the brackets) contribute to $T\Delta S$.

$$\Delta G_B = \Delta G_{Dir} + \Delta G_{Micro} + \Delta G_{SA} - (\Delta G_{Des} + \Delta G_{Con}) \qquad (4)$$

ΔG_{Dir} is the ΔG due to direct interactions between protein and ligand (copper ions here), which includes forces such as electrostatic, van der Waals, and intermolecular bonds (hydrogen bonds). ΔG_{Micro} is the ΔG due to specific microenvironment of the amino acid residues (interactions do not directly involve the reporter histidine, but around the histidine) around the histidine group such as intramolecular interactions, hydrophobic interactions and steric effects, which might enhance or hinder the binding process. All these microenvironment effects (hereafter these effects are specifically referred to as ***microstructure or microstructural effects***) may also affect the pK_a of a reporter histidine, and thus the protein-metal interaction. ΔG_{SA} is the ΔG due to the ability of the surface accessible side chain (of histidine residue including its flexibility and orientation) to interact with the immobilized ligand. ΔG_{Des} is the loss of free energy due to desolvation of protein and ligand to release the solvent molecules. ΔG_{Con} is the loss of free energy due to the restrictions of protein and ligand conformations during the binding process. The first three terms in Eq. (4) can be further dissected as:

$$\Delta G_{Dir} = \Delta G_{Elec} + \Delta G_{Vdw} + \Delta G_{Inter} \quad (4a)$$

$$\Delta G_{Micro} = \Delta G_{Int} + \Delta G_{Hydro} + \Delta G_{Steric} \quad (4b)$$

$$\Delta G_{SA} = \Delta G_{Ori} + \Delta G_{Flex} \quad (4c)$$

In Eq. 4a, ΔG_{Elec} describes the charge perturbation around a histidine reporter group, and the perturbation directly affects the pK_a value of the histidine and thus protein-metal ion interaction. It is one of the main parameters studied in this article. As stated above, K_{aP} is used to quantify the adsorption affinities of all variants in the IMAC column. Since the adsorption process is a thermodynamic equilibrium process, the relation between K_{aP} and $\Delta\Delta G_B$ (change in Gibbs free energy of binding due to electrostatic pertubation) can be derived, and their relationship with the change of pK_a (ΔpK_a) can be described by the following equation:[10]

$$1.364\Delta pK_a = -RT \ln \frac{K_{aPv}}{K_{aPc}} = \Delta\Delta G_B \quad (5)$$

Wherein, K_{aPc} is the protein binding strength of the control variants (pWT135 and pWT80) and K_{aPv} is the protein binding strength of all variants in which there is a perturbation in the histidine microenvironment.

Components contributing to ΔG_B due to perturbation ($\Delta\Delta G_B$): In the present set of IMAC experiments the major factors that affect the formation of the coordinate covalent bond between the sole histidine residue on lysozyme and the copper ion are 1) the histidine's pK_a, which is dependent on all the aforementioned factors listed in the introduction section, 2) surface accessibility, flexibility and orientation of the histidine's side chain, 3) histidine's microenvironment, and 4) flexibility of the immobilized ligand. Here, all variants have similar, if not identical, histidine surface accessibility, combining

with the fact that the changes of ΔG_{Des} and ΔG_{Con} are negligible due to the way the variants were generated, $\Delta\Delta G_B$ can be determined by the following equations for a particular variant before and after perturbations take place:

$$\Delta\Delta G_B = \Delta G_{Dir1} + \Delta G_{Micro1} - (\Delta G_{Dir2} + \Delta G_{Micro2}) \qquad (6)$$

Assuming the values of ΔG_{Vdw} and ΔG_{Inter} are negligible for all variants (for IMAC experiments, this is especially true) (as in Eq. 4a), the following equations can be obtained,

$$\Delta\Delta G_B = \Delta\Delta G_{Elec} + \Delta\Delta G_{Micro} \qquad (7)$$

Thus, when there are no significant microstructural effects ($\Delta\Delta G_{Micro} = 0$) around the histidine group, $\Delta\Delta G_B = \Delta\Delta G_{Elec}$. Under this circumstance, Eq. 5 can be extended to the following equation,

$$1.364\Delta pK_a = -RT \ln \frac{K_{aPv}}{K_{aPc}} = \Delta\Delta G_B = \Delta\Delta G_{Elec} \qquad (5a)$$

Moreover, if the $\Delta\Delta G_{Elec}$ value is known, then $\Delta\Delta G_{Micro}$ value can be calculated from Eq. 6 for those interactions in which there is a significant structural contribution to a particular binding event. Then from the obtained $\Delta\Delta G_{Micro}$ value, with the assistance of molecular modeling, it is possible to discover the specific microstructural effects, such as intramolecular interactions, hydrophobic interaction, or steric effects, around the reporter histidine residue.

On the other hand, $\Delta\Delta G_{Elec}$ can be theoretically derived from the basic Coulomb's law.[59] The change in electrostatic force between two charges q_1 *and* Δq_2,

which is the electrostatic perturbation to q_2, can be obtained from Coulomb's law (Eq. 8).[60]

$$\Delta F = -\frac{1}{\varepsilon \cdot 4\Pi\varepsilon_0}\frac{q_1 \cdot \Delta q_2}{r \cdot r} \quad (8)$$

Where, q_1 is the charge of histidine residue at pH 7.0 (0.5 for free histidine), Δq_2 is the net change in charge due to the purterbation in the histidine's microenvironment at pH 7.0, r is the distance between q_1 and Δq_2 (from Cα to Cα), ε_0 is the permittivity of free space, ε is the dielectric constant of the protein surface, which equals 45.[15] The change in electrostatic free energy (ΔE) can be defined as the product of net force (ΔF) and the distance between the two charges q_1 *and* Δq_2 (Eq. 9),[35] and given by the following equation,

$$\Delta E = \Delta F \cdot r = -\frac{1}{\varepsilon \cdot 4\Pi\varepsilon_0}\frac{q_1 \cdot \Delta q_2}{r} \quad (9)$$

Then the change in free energy due to change in electrostatic free energy, $(\Delta\Delta G'_{Elec})$ is given in Eq. 10,[27, 59]

$$\Delta\Delta G'_{Elec} = \Delta E \cdot N \quad (10)$$

Where, N is the Avogadro's number.

Results

All the sites chosen for single point mutations were not critical for protein's structure or activity because none of these sites were involved in any extensive amino acid network including those involved in the protein's catalytic activity.[53, 61, 62] Choosing such isolated residues was necessary because the aim of this study was to understand the structural effects, if any, on the sole probe-histidine were exclusively due to the engineered single point mutation. Circular Dichroism (CD) results (Fig. 2) and enzymatic

assays (Table 4) indicate that the 3D structure of all the lysozyme variants was intact. The $\Delta\Delta G_E$ value of all the variants is reported in Table 5. The elution volume required to half-elute the protein, obtained from the elution profiles (Fig. 3), was used to obtain the intercept by plotting $\frac{1}{V - V_0}$ Vs. [*I*] (Fig. 4).[29] From the intercept the protein binding strength (K_{aP}) value for each variant was calculated (Table 5).

IMAC derived ΔpK_a ***and*** $\Delta\Delta G_B$ ***values:*** The relationship between K_{aPv} and the histidine's pK_a is given in Eq. 5. From Eq. 5, we can infer that as the K_{aPv} value of a variant increases or decreases, the ΔpK_a value of the histidine will proportionally decrease (-) or increases (+).[41] The ΔpK_a values of all the variants vary from -0.4 to +0.7 units (Table 5). With the exception of two variants S136K and D127K (which might include microstructural effects), the approximate ΔpK_a value for an electrostatic perturbation of $\pm$ 1 and $\pm$ 2 units is $\pm$ 0.25 and $\pm$ 0.5 units, respectively. These values are within the range of ΔpK_a values reported by Fersht and co-workers.[10, 17] They reported that when aspartate residue was replaced by serine and lysine, the ΔpK_a value of the histidine residue was +0.4 and +0.6 units, respectively. The ΔpK_a values reported in Table 5 support the general trend observed, wherein as the change in net charge varies from a positive to a negative value (due to electrostatic perturbation), the histidine's pK_a decreases or increases, respectively.[63] As a result, when the net positive charge on a protein increases, there is an increase in the repulsive force experienced by the histidine residue's proton, and consequently, an apparent decrease in the histidine's pK_a is observed.[17, 26] A similar phenomenon is observed in case of a net increase in the negative charge, whereby the proton on the histidine is stabilized resulting in a higher pK_a value

of the histidine residue.[13, 14, 17, 26] In addition, the NMR profiles [8, 9, 31] of the two variants R137D and S136K also support the above-mentioned general trend (Fig. 6). The profile of variant R137D indicates that it has a higher histidine pK_a value when compared to the control histidine (pWT-135), which in turn has a higher histidine pK_a than the variant S136K.

The relationship between $\Delta\Delta G_B$ and K_{aPv} is also given in Eq. 5, which indicates that as the K_{aPv} value increases or decreases, the $\Delta\Delta G_B$ value proportionally decreases (-) or increases (+), respectively. Similar to the general trend reported above for ΔpK_a values, $\Delta\Delta G_B$ values also follow the general trend, wherein as the $\Delta\Delta G_B$ value becomes more negative, the spontaneity of the given reaction (binding between protein and copper) increases corresponding to a greater K_{aPv} value. This general trend is supported by the $\Delta\Delta G'_{Elec}$ values (Table 5), which has an excellent correlation ($R^2 = 0.96$) with the $\Delta\Delta G_B$ values (Fig. 5), when the three variants S136K, S136D, and D127K are excluded, which warrant separate discussion due to their unique structural features.

Microstructure around the three sites 108, 76 and 137: The side chain of the glutamic acid at site 108 (before any site-directed mutagenesis) is not involved in any intramolecular bonds but its orientation is pointing towards the histidine residue (Fig. 7). The distance between the nitrogen in the imidazole ring of the histidine and the carboxyl carbon of glutamic acid (Nε-Cγ is 8.8 Å) is much closer than the Cα-Cα distance (9.6 Å). In cases of the sites 76 and 137 the side chain (amino moieties) of the arginine residue at these two sites is involved in a salt bridge with Asp72 (Nε-Oδ2 is 1.7 Å, Nε- Oδ1 is 2.3 Å and N1H- Oδ1 is 3.2 Å) and Glu22 (N1H- Oε1 is 1.7 Å and N2H- Oε2 is 1.7 Å),

respectively (Fig. 8a and 8b). The residues that replace the original amino acid at these three sites are all surface accessible.

IMAC derived ΔpK_a ***and*** $\Delta\Delta G_B$ ***values at sites 108, 76 and 137:*** In the case of variant E108K, the ΔpK_a and $\Delta\Delta G_B$ values of the histidine residue is -0.39 units and -0.55 kcal/mol, respectively. From Fig. 5, the $\Delta\Delta G_B$ value is primarily due to the ΔpK_a of the histidine and therefore $\Delta\Delta G_B = \Delta\Delta G_{Elec}$ (Eq. 5a). In the cases of the two sites, 76 and 137, although they follow the correlation (Fig. 5), there are some notable differences between their $\Delta\Delta G_B$ values. These two sites can be characterized as having similar microenvironment due to their 1) equivalent relative surface accessible area, 2) involvement in a salt bridge, and 3) equivalent microstructure (*rHy*) (Table 2). However, the ΔpK_a value of the variants at these two sites varies as much as 30% for a serine (-1) substitution and 75% for an aspartate (-2) substitution (Table 5). This difference in the ΔpK_a values for the serine substitution can be attributed to the difference in the orientation of the original arginine residue at these two sites. Due to the arginine's long side chain and its orientation the variation in the distance between the charged moieties of the probe-histidine (Nε2) and the arginine (NH1, NH2) of the two sites is approximately 10.5 Å. In the case of aspartate substitution, assuming that the replaced amino acid side chain mimics the orientation of the side chain of the original amino acid, the distance between His80 and Asp76 (Nε2 – Oδ1 is 5.2 Å and Nε2 – Oδ2 is 6.2 Å) allows for the formation a weak salt bridge (Fig. 9) and thus enhancing the histidine's ΔpK_a value by 75 % when compared with the ΔpK_a of His135, whose distance from Asp137 is 16.7 Å (Fig. 10).

Microstructure around site 127: In general, to stabilize the charge between a pair of oppositely charged residues, the positively charged residue acts as an acid while the negatively charged residue acts as a base.[30] Therefore the pK_a of the positively charged residue is raised while the negatively charged is lowered (e.g. hydrogen bond).[21, 64] On the other hand, to stabilize the charge between a pair of identically charged residues, the amino acids involved have either lower (for positively charged residue, lysine) or higher (for negatively charged residue, aspartate) pK_a values.[21, 65, 66] Thus in both cases, the pK_a of the amino acids involved is altered.

Asp127 is located next (Cα-Cα is 3.86 Å) to another negatively charged residue Glu128, whose distance form the histidine is (Cα-Cα) 11.3 Å, equivalent to that from Asp127 (12.96 Å). In fact, the side chains of the three residues (Asp127 Cδ, His135 Nε and Glu128 Cγ) form a triangle with an angle of 65.8º at His135 Nε (the sole surface histidine) (Fig. 11). Therefore, these two negatively charged residues might be involved in an ion pair of similar charges resulting in an increase in their pK_a value. But from the X-ray crystal structure, these two negatively charged residues are involved in salt bridges with their adjacent arginine residues (Fig. 11). Consequently, although the residues are surface accessible (Table 2), the mobility of their side chains is restricted. On the other hand, it is hard to rationalize their charge effect on the histidine's pK_a value [22], due to their involvement in salt bridges. In addition the presence of Val131, which may act as a screening agent between Asp127 and the histidine [7], complicates the effect of Asp127 on the histidine's pK_a value (Fig. 11).

When a polar serine replaces the aspartic acid, there will be a net increase in the total positive charge on the protein by +1 unit. The charge of Glu128 will be unaffected

due to its immobilized side chain, which is involved in a salt bridge with Arg125. Thus there should be a net decrease in the pK_a value of the histidine. On the other hand, replacing Asp127 with lysine results in the formation of an ion pair of similar charges with Arg154 (Lys 127 Nz - Arg 154 N1H is 4.7 Å) (Fig. 12). This ion pair formation might lower the pK_a value of Lys127 by as many as 4 pH units.[65] Also, assuming the orientation of these residues is similar to that of the original residue, the distance between the charged moieties of Lys127 (Nz) and the histidine (Nε2) is 15.8 Å as against 12.7 Å for serine. Therefore, the effect of lysine replacement on the net protein charge might be much lower than + 2 units [25]. Also the NMR profile of the two variants S127K and S127S is almost overlapping (data not shown), thus indicating an equivalent contribution to the pK_a for the histidine residue. Thus, from the microstructure of Lys-127, it can be inferred that the effect of the lysine and serine substitutions will have similar effects on histidne's pK_a value.

IMAC derived ΔpK_a ***and*** $\Delta\Delta G_B$ ***values at site*** **127:** As expected variant D127S has a $\Delta\Delta G_B$ value greater than pWT-135 (Table 5, indicated by a negative value). On the other hand, variant D127K has the $\Delta\Delta G_B$ value lower than both E108K and D127S variants (Table 5). This is an unexpected result because the net charge change for variant D127K is +2 similar to E108K, while for D127S is +1. Nevertheless, this low $\Delta\Delta G_B$ of D127K can be explained based on the microstructure of Lys127. Due to a lower pK_a of Lys127 and its distance (15.8 Å) from the imidazole ring of histidine, the ΔpK_a of the histidine in variant D127K might be lower than the variant D127S. Thus, it can be concluded that unlike E018K ($\Delta\Delta G_B = \Delta\Delta G_{Elec}$), the D127K $\Delta\Delta G_B$ value depends on both the $\Delta\Delta G_{Elec}$ and $\Delta\Delta G_{Miro}$ (Eq. 7).

Microstructure around site 136: The side chain of serine at site 136 has low surface accessibility (Table 2) as its side chain points into the protein interior (cavity).[7] The cavity around the serine residue, due to the presence of the predominantly non-polar residues such as Ala134, Leu133, Phe114, Trp138 and Tyr139, can be characterized as hydrophobic.[54] However, the *rHy* value at site 136 indicates a hydrophilic microstructure (Table 2). In addition, the side chain of the serine is involved in two hydrogen bonds with its adjacent residues (Fig. 13). Consequently, the mobility or the flexibility of the serine residue is restricted, and thus it can be concluded that the orientation of the side chain (even in solvent) is pointing into the protein interior is the most probable.

The SYBYL generated S136D variant model also orients the side chain of Asp136 into the protein interior. The microstructure (rHy = 0.28) of Asp136 is hydrophilic and its side change is involved in a single hydrogen bond. Thus the orientation of Asp136 side chain can be assumed as similar to the side chain orientation of the serine residue in pWT-135 (Fig. 13 and 14). On the other hand, when lysine replaces serine, there are two possible orientations that need to be considered.[30] In the first case, due to the presence of the cavity and absence of any site-specific steric clashes around the serine residue, the lysine side chain may probably be orientated into the protein interior,[42] as modeled by the SYBYL software (Fig.15). In the second case, due to a) the high desolvation energy costs from a hydrophobic microstructure (cavity) around the lysine residue (rHy = 0.08), b) the high B-values (flexibility) of the lysine residue [45, 67-69], and c) the absence of any hydrogen bonds with its adjacent residues, the side chain of Lys136 is very mobile as demonstrated in Fig. 16. This figure clearly demonstrated that during the course of the simulation, the NZ atom of Lysine is closer (most of the

time) to the ND1 atom of His135 than is the Ser OG atom to the His135 ND1 atom. Therefore the Lys136 side chain possesses the capability to orient itself in a different direction from the serine residue outward towards the probe-histidine at site-135.[70, 71] This hypothesis is supported by the molecular dynamic simulation result shown in Fig. 17. This figure clearly illustrates that unlike the orientation of serine side chain in pWT-135, the orientation of lysine in S136K is pointing away from the protein's interior (in the orientation similar to probe-histidine at site-135).

IMAC derived ΔpK_a and $\Delta\Delta G_B$ values for variants at site 136: In case of variant S136D, the possibility of the aspartate residue orienting itself in the direction of serine is very high because 1) aspartic acid is similar in size as serine and is involved in a hydrogen bond (Fig. 14 comparing to Fig. 13) and 2) the microstructure of aspartate is more hydrophilic (rHy = 0.28 in S136D) than lysine (rHy = 0.08 in S136K). However, on the other hand, the relative hydrophilicity of aspartic acid is much less than serine (rHy = 0.70) and thus might increase it's pK_a value significantly. Consequently, this higher pK_a value of Asp136 (low net negative charge) lowers the total contribution of Asp136 to the ΔpK_a value of His135 in S136D. This lower contribution of Asp136 is captured by Fig. 5, wherein the $\Delta\Delta G_B$ value of S136D is shown as an outlier. Despite the low pK_a value, the $\Delta\Delta G_B$ value of S136D is equivalent to the other -1 charged variants (R137S or R76S), due to its short distance from His135, and most significantly still a positive value.

In the case of variant S136K, there is a net positive increase in $\Delta\Delta G_B$ value opposing to the expected net negative increase (Table 5). This variation in the $\Delta\Delta G_B$ value may not be exclusively due to the ΔpK_a value of variant S136K because none of the aforementioned factors (see the introduction section) affecting the ΔpK_a differ from

other variants. In fact, the NMR profile of S136K is lower than the control variant pWT-135 (Fig. 6), indicating that the pK_a of the histidine residue of variant S136K is lower than pWT-135. Therefore, the variation in the ΔpK_a value might be due to the additional microstructural effects contributing to the $\Delta\Delta G_B$ value (Eq. 7). The $\Delta\Delta G_B$ value of S136K varies by 0.56 kcal/mol when compared with any other +1 overall net charged variant (D127S or R76S or R137S). This variation can be explained by considering the two following scenarios.

Case-1: The basic assumption for this scenario is that the orientation of the lysine side chain is similar to the serine residue that points into the cavity (Fig. 15). Then due to the strong hydrophobic nature of the cavity, the pK_a of Lys136 might be lowered.[54] If this scenario is true, the $\Delta\Delta G_B$ value of S136K (0.22 kcal/mol, a positive value) should be approximately equivalent (or in range of) to that of D127S variant (a negative value). In addition this scenario is not supported by the NMR profile of the variant S136K, which indicates a lower pK_a than pWT-135 (Fig. 5), and thus a positive $\Delta\Delta G_B$ value would be expected. Therefore in order In order to understand this "unreasonably" high deviation in the $\Delta\Delta G_B$ value of S136K, case-2 was considered.

Case-2: Contrary to case-1, the orientation of the lysine residue may be outward towards the histidine at site 135, into the protein exterior (Fig. 17). Due to the bulky nature of lysine and its location (His 135 Cα- Lys 136 Cα is 3.83 Å), the surface accessibility of the histidine residue might be reduced due to 1) steric hindrance or 2) change in its orientation [46]. As the surface accessibility of the histidine decreases, the binding strength of the protein decreases [29] and thereby a decrease in its $\Delta\Delta G_B$ value (Eq. 7). Thus it can be concluded that even though there might be a considerable

electrostatic energy generated due to the charge perturbation (effect on pK_a), the major (or dominant) contribution for $\Delta\Delta G_B$ for variant S136K can be accredited to the steric hindrance effect ($\Delta\Delta G_{Micro}$ = 0.56 kcal/mol) generated by the adjacent lysine residue. On the other hand, lowering of the pK_a value due to inward orientation may be partly responsible (as discussed in case-1) for such a strong deviation in the $\Delta\Delta G_B$ value of S136K.

In order to further investigate the presence of the steric hindrance effect, variant S136M was generated. The methionine residue was selected because it resembles the bulky nature of the lysine residue and is not charged. Also from the hydrophobic parameters, the microstructure of methionine, *sHy* = - 4.7 (*rHy* was not determined), is characterized to be similar to serine (*sHy* = -5.6) and therefore is hydrophilic. Thus, due to the non-polar nature of methionine and the hydrophilic microstructure around it, the orientation of the methionine might be more likely to be outward then inward. Moreover, due to the pair of donor electrons on the sulfur atom and the hydrophobic nature of the methyl moiety in the end of the side chain, the charge of methionine was assigned a value of -0.15. Though the experimental results indicate that the $\Delta\Delta G_B$ value for S136M follows the trend (Table 5), our hypothesis is that both the hydrophobic nature of the methionine (case-1) and the steric hindrance effect (case-2, due the bulky adjacent methionine) were still a contributing factor for the $\Delta\Delta G_B$ value of S136M.

Dissection of $\Delta\Delta G_B$ ***:*** Understanding the importance of each thermodynamic factor that contributes to change in Gibbs free energy (Eq. 4 and 7) is important because they provide significant insight into the protein ligand binding (here, interactions in IMAC). The $\Delta\Delta G_B$ values for variants E108K, D127S, R137S and R137D are primarily due to

$\Delta\Delta G_{Elec}$ since no microstructural effects were present. In the case for site 76, the differences in the $\Delta\Delta G_B$ values (of the variants) when compared with that of the variants at site 137 can be attributed to 1) side chain orientation of the replacing residue ($\Delta\Delta G_{Micro} = \Delta\Delta G_{Steric} =$ 0. 08 kcal/mol for Δq = +1), 2) formation of a weak salt bridge ($\Delta\Delta G_{Micro} = \Delta\Delta G_{Int}$ = 0.15 kcal/mol). At site 127, the two variants D127K and D127S do not follow the additivity in their $\Delta\Delta G_B$ values [17]. This discrepancy in $\Delta\Delta G_B$ values ($\Delta\Delta G_{Micro}$ = 0.21 kcal/mol) can be attributed to 1) the orientation of the long side chain of Lys127, which augments the distance between its charged moiety from His135 by 3 Å when compared with Ser127 and His135, and 2) the presence of the ion pair between two identically charged residues Arg154 and Lys127, which might lower the pK_a of Lys127 (and thus lowering its positive charge). In case of variant S136K, the perturbation is short ranged and bulky. It induces significant $\Delta\Delta G_{Micro}$ = 0.56 kcal/mol, and can be attributed to $\Delta\Delta G_{Hydro}$ (for case-1) and $\Delta\Delta G_{Steric}$ (for case-2). Thus, dissection of $\Delta\Delta G_B$ provides a quantitative contribution or relative importance of each thermodynamic factor to the drives the binding process in IMAC.

Discussion

IMAC can be used as an effective technique to study short range perturbation effects because of the specific nature of the interaction between the histidine's imidazole nitrogen and IMAC-Cu^{2+} ions by forming a coordinate covalent bond. Coordinate covalent bond formation in an IMAC column is a thermodynamic equilibrium process and can be quantified using a parameter called dissociation constant. On the other hand analyzing these short-range perturbations by enzyme activity assays is challenging because the catalytic activity is generally dependent on a network of residues, which can

be both short and long-range. Therefore, the perturbation of a residue in the catalytic network residue might trigger a cascade effect[6] instead of an exclusive single perturbation effect. Also these cascade effects are protein dependent and the values generated thereof are not general.

The dissociation constant, which reflects the protein behavior in an IMAC column, was previously used to measure the structural perturbations in a protein having one surface accessible histidine.[29] In this report, the variation in the dissociation constant value was used to quantify the electrostatic perturbations in the microenvironment of a lone surface histidine. This variation in the dissociation constant was attributed to the ΔpK_a of the histidine and was used to calculate the $\Delta\Delta G_B$ value (Eq. 6). However, since all the electrostatic charge perturbations around the sole histidine residue are in short and medium range, there might be some associated microstructural effects ($\Delta\Delta G_{Micro}$), in addition to electrostatic effects ($\Delta\Delta G_{Elec}$), that might contribute to $\Delta\Delta G_B$ (e.g. site 136). To study the contribution of these individual factors to $\Delta\Delta G_B$ value, quantitative assessment of both the electrostatic forces ($\Delta\Delta G_{Elec}$) and the microstructural affects ($\Delta\Delta G_{Micro}$) on the behavior of the histidine residue were evaluated by employing Eq. 6. In addition, the factor $\Delta\Delta G_{Micro}$ comprises of three factors: 1) the involvement of the histidine in intramolecular bonds with adjacent amino acids ($\Delta\Delta G_{Int}$), 2) the microenvironment of the replaced residue by unfavorable hydrophobic ($\Delta\Delta G_{Hydro}$) environment which may a) alter its pK_a value and b) introduce structural strain [42], and 3) site-specific structural effects that might affect the histidine ($\Delta\Delta G_{Steric}$) by i) the orientation and flexibility of the adjacent residue's side chain that affects the histidine's pK_a,[2, 65, 66, 72] and ii) the size of the adjacent amino acids that might induce

steric hindrance (blocking) effect.[65] Studying and quantifying these individual factors influencing the microstructural effects, due to single point mutations, on the IMAC binding process is very challenging. Here, we have used molecular modeling tools, $\Delta\Delta G'_{Elec}$ and dissected $\Delta\Delta G_B$ values to quantify these factors.

Our study indicates that there are significant microstructural effects that affect IMAC binding process. In addition, we were able to demonstrate that other factors such as flexibility and orientation of not only the histidine residue (side chain) but also neighboring amino acid residues is important in mitigating the protein-metal ion interactions. These results further demonstrate the importance of a flexible or an ensemble protein structure that aids in its function.[36, 73, 74] Quantifying these subtle microstructural effects is important because they play a critical role in important physiological process such as molecular recognition, protein-ligand and protein-protein interactions.[6, 71] Thus, the $\Delta\Delta G_B$ (or the variation of the dissociation constant) values can be used to quantify the change in electrostatic and microstructure of any lone surface histidine. Also it can be concluded that when there were no major microstructural effects influencing the $\Delta\Delta G_B$ value, there might be additivity in the effects due to charge perturbations.[17] On the other hand, this report illustrates the complexity involved in reverse engineering of ion pairs and catalytic sites.[75]

In summary, we have focused our investigation primarily on two directly obtainable parameters. The $\Delta\Delta G_B$ values that were calculated from the measurable protein-metal ion interaction in IMAC, and $\Delta\Delta G'_{Elec}$ from theoretical calculations (obtained from Coulomb's law) from the controlled electrostatic perturbation. A direct correlation between these two parameters should exist if all variants have similar

microstructural effects ($\Delta\Delta G_{Micro} \approx 0$ in Eq. 6). If a significant deviation is observed for a particular variant, $\Delta\Delta G_{Micro}$ can be calculated from the correlation developed. Subsequently, with the assistance of molecular modeling, the specific microstructural effects, such as intramolecular interactions, hydrophobic interaction, or steric effects, around the reporter histidine residue can be determined for the variant. Consequently, the protein microstructures around the reporter histidine (electrostatic interaction involving histidine and the microstructure around it) are quantitatively characterized.

Conclusions

Short and medium range charge perturbations around the sole histidine residue in T4 lysozyme, which is not involved in any catalytic or other network of residues, were engineered within a 15 Å distance. Theoretical $\Delta\Delta G_E$ values were calculated using simple Coulomb's law. Experimental ΔpK_a and $\Delta\Delta G_B$ values were obtained by measuring the protein binding strength values using IMAC. The direct correlation (R^2 = 0.96) between $\Delta\Delta G_E$ and $\Delta\Delta G_B$ indicates that it is possible to predict the electrostatic effect due to mutations $(\Delta\Delta G_{Elec})$, when no other structural effects are involved. The $\Delta\Delta G_B$ value of the three variants (S136K, S136D and D127K) differ from the other similarly charged variants and their own $\Delta\Delta G_E$ value because of the additional site-specific microstructural effects ($\Delta\Delta G_{Micro}$) that contribute to their $\Delta\Delta G_B$ value. The contribution of these individual factors (e.g. electrostatic, microstructural) to $\Delta\Delta G_B$ was dissected by comparing the experimental $\Delta\Delta G_B$ values with the other variants of identical charge perturbations. Electrostatic and microstructural effects due to a single point mutation on $\Delta\Delta G_B$ value of the probe-histidine were quantified. In addition, the contribution of the relevant $\Delta\Delta G_{Micro}$ factor to $\Delta\Delta G_B$ value was identified and

quantified. Quantifying these components of $\Delta\Delta G_B$ will help in not only understanding the structure-function relationship but also engineering or designing tailor-made proteins.

Acknowledgements

We gratefully acknowledge computing time on the System X supercomputer in the terascale computing facility at Virginia Tech. The work is partly funded by a grant from the Jefferess Memorial Trust.

Table 1. The change in net charge and the distance from the histidine residue in both control as well as test site, for all the ten lysozyme variants. The pWT-135 and the pWT-80 represent the pseudo wildtype lysozymes with histidine at site 135 and site 80 respectively.

Protein Variant	**Change in net charge (Units)**	**Cα-Cα distance between the site and the histidine residue (Å)**
Control Variant (pWT-135)	0	N/A
D127S	+1	12.96
D127K	+2	12.96
S136D	-1	3.83
S136K	+1	3.83
S136M	-0.15	3.83
R137S	-1	7.35
R137D	-1	7.35
Test Variant (pWT-80)	0	N/A
R76S	-1	5.97
R76D	-2	5.97
E108K	+2	9.62

Table 2. Relative surface accessibility[1] and relative hydrophilicity values (*rHy*)[2] of all the five sites chosen for single point mutations.

Sites	Relative surface accessibility	Relative hydrophilicity (*rHy*)
Arg-76	0.60	0.84
Glu-108	0.42	0.55
Asp-127	1.00	1.02
Ser-136[3]	0.10	0.70
Arg-137	0.63	0.78

[1] The relative surface area of each amino acid side was directly obtained from SYBYL program

[2] Microenvironments with *rHy* values 1) >0.25 are fairly hydrophilic, 2) < 0.25 are hydrophobic and 3) < 0.0 are extremely hydrophobic

[3] The *aHy* value of tyrosine was used to calculate the *rHy* of Ser-136

Table 3. The peptide sequences of the seven lysozyme variants obtained from electrospray ionization mass spectrometry. The peptide sequences were identified by SEQUEST. The histidine and the replaced residue (point mutation) are shown in bold letters.

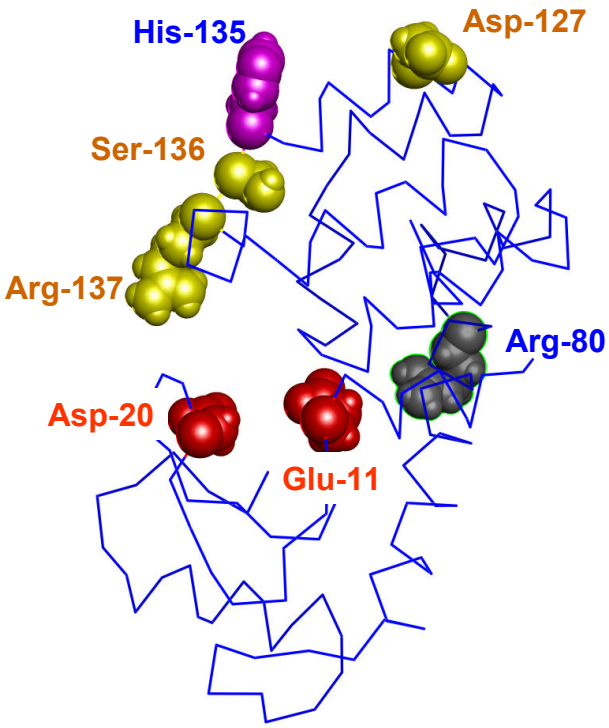
Protein Variant	Peptide position	Peptide sequence
D127S	126-137	W**K**EAAVNLA**H**SR
D127K	126-137	W**K**EAAVNLA**H**SR
S136D	125- 137	RWDEAAVNLA**HD**R
S136K	126- 137	WDEAAVNLA**HK**R
S136M	125- 137	RWDEAAVNLA**HM**R
R137D	125- 145	RWDEAAVNLA**HSD**WYNQTPNR
R76D	61-83	DEAEKLFNQDVDAAV**D**GIL**H**NAK

Table 4. Enzymatic activities of purified T4 lysozyme variants. The activities of all the lysozyme variants are compared with the wild type.

Protein Variant	**(%) Activity**
Wild type	100
pWT-135	162
D127S	132
vD127K	77
S136D	70
S136K	82
S136M	89
R137S	86
R137D	109
pWT-80	126
R76S	120
R76D	124
E108K	102

Table 5. Calculated $\Delta\Delta G_E$ values from Coulomb's law, IMAC protein binding strength (K_{aP}) values, and experimental ΔpK_a and $\Delta\Delta G_B$ values obtained from the K_{aP} values.

Protein Variant	$\Delta\Delta G_E$ **(kcal/mol)**	**Protein binding strength (K_{aPv}) (10^1 M^{-1})**	**Change in pK_a**	$\Delta\Delta G_B$ **(kcal/mol)**
Control Variant (His-135)	0	92.1	0	0
D127S	-0.29	163.4	-0.25	-0.34
D127K	-0.58	138.6	-0.17	-0.24
S136D	0.98	54.8	0.22	0.31
S136K	-0.98	63.4	0.16	0.22
S136M	0.15	70.5	0.12	0.16
R137S	0.51	50.6	0.26	0.35
R137D	1.02	34.8	0.42	0.58
Test Variant (His-80)	0	106.7	0	0
R76S	0.63	48.1	0.34	0.47
R76D	1.25	20.1	0.73	0.99
E108K	-0.78	261.8	-0.39	-0.53



A) The structure of pWT-135

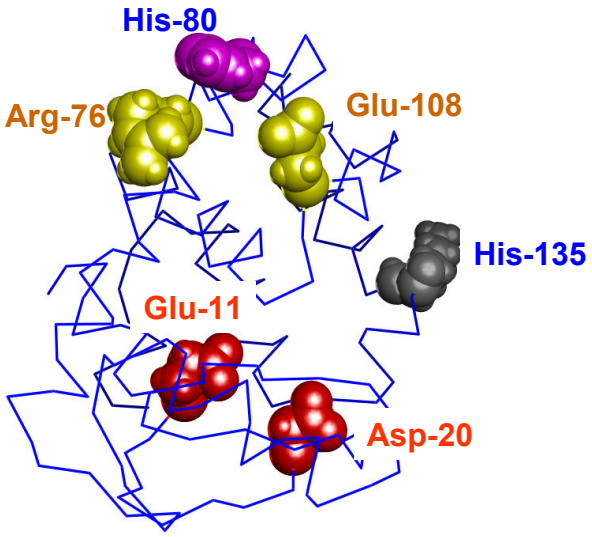


Figure 1 (A and B). The sites shown in red are the residues involved in catalysis. **A)** The three sites (yellow) chosen around the histidine residue, His-135 (magenta), in the control variant. Site 80 is shown in gray. B) The two sites (yellow) chosen around the histidine residue, His-80 (magenta), in the test variant. Site-135 is shown in gray.

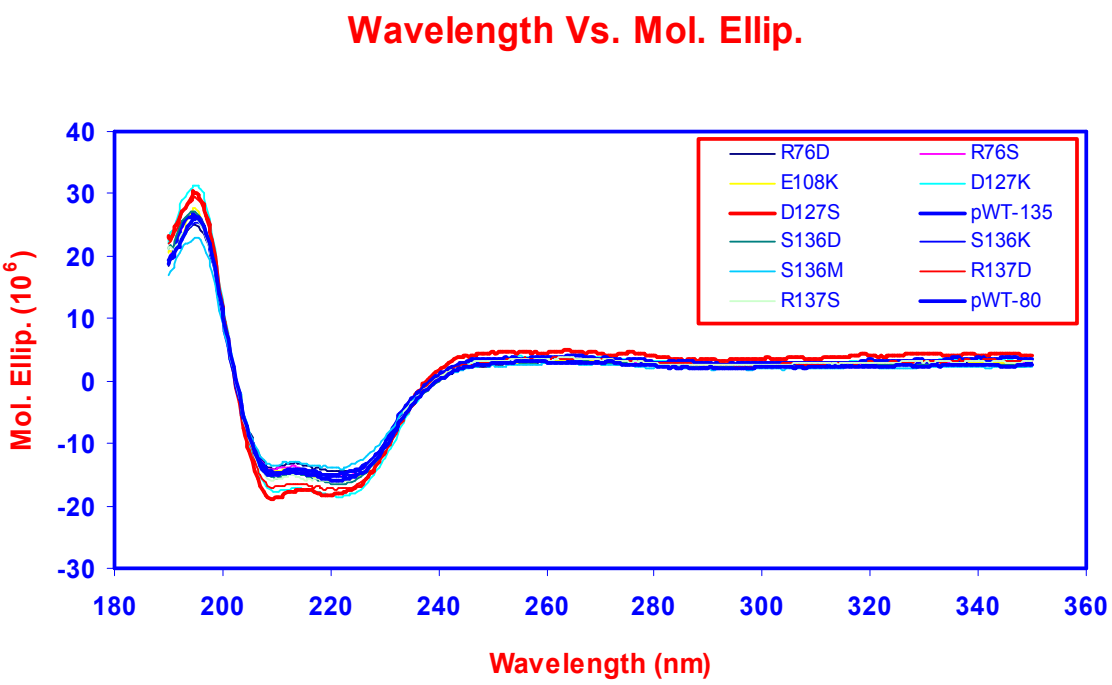


Figure 2. CD spectrum for all the ten variants and both the pseudo wild type lysozyme proteins, pWT-135 and pWT-80.

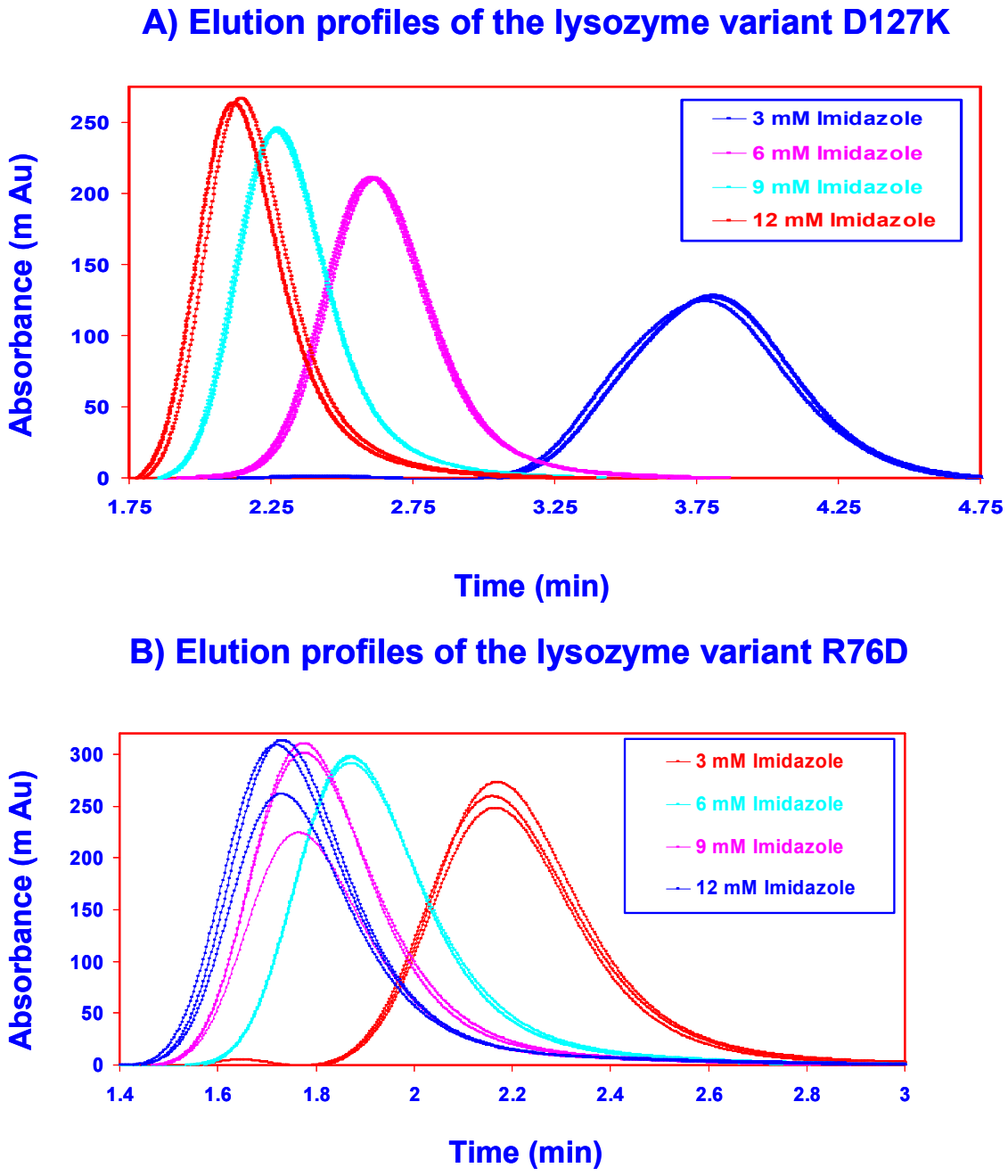


Figure 3 (A and B). Elution profiles of the variants A) D127K and B) R76D for four different concentrations of Imidazole [I].

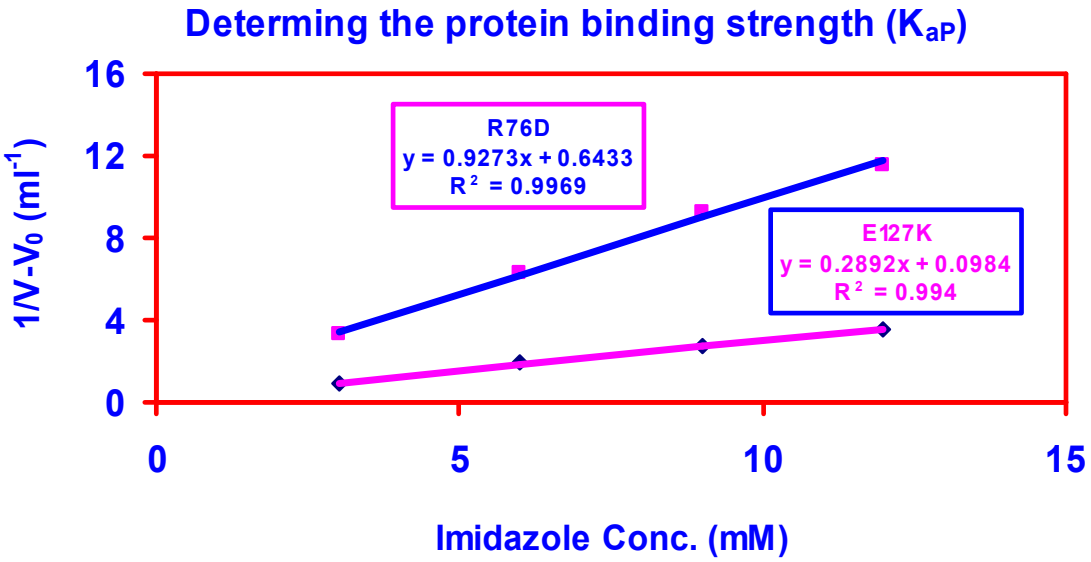


Figure 4. Determination of the $\frac{1}{V - V_0}$ intercept required to calculate the binding strength of variants R76D and E127K.

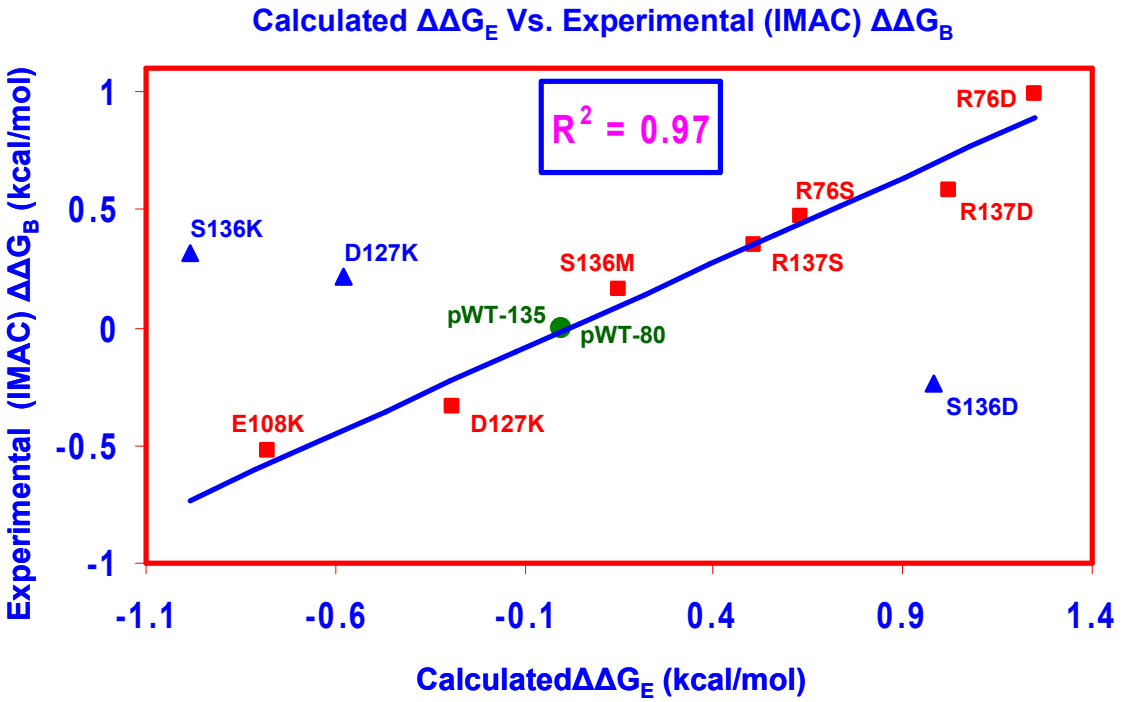


Figure 5. Correlation between the calculated $\Delta\Delta G_E$ values from Coulomb's law and experimental $\Delta\Delta G_B$ values obtained from the K_{aP} values obtained from IMAC, excluding the three variants S136K, S136D and D127K (shown in blue). These three excluded variants (outliers) deviate from the general trend due to the presence of strong site-specific structural (microstructural) effects. The data point indicated in green represents the control variants (pWT-135 and pWT-80).

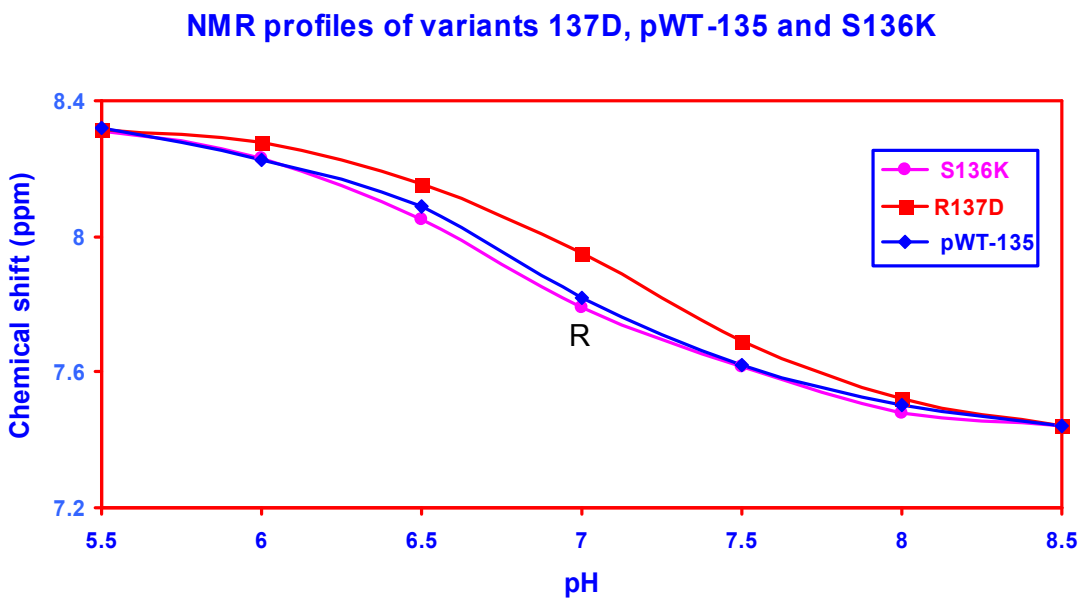


Figure 6. The NMR profiles of three variants R137D (■), pWT-135 (♦) and S136K (●).

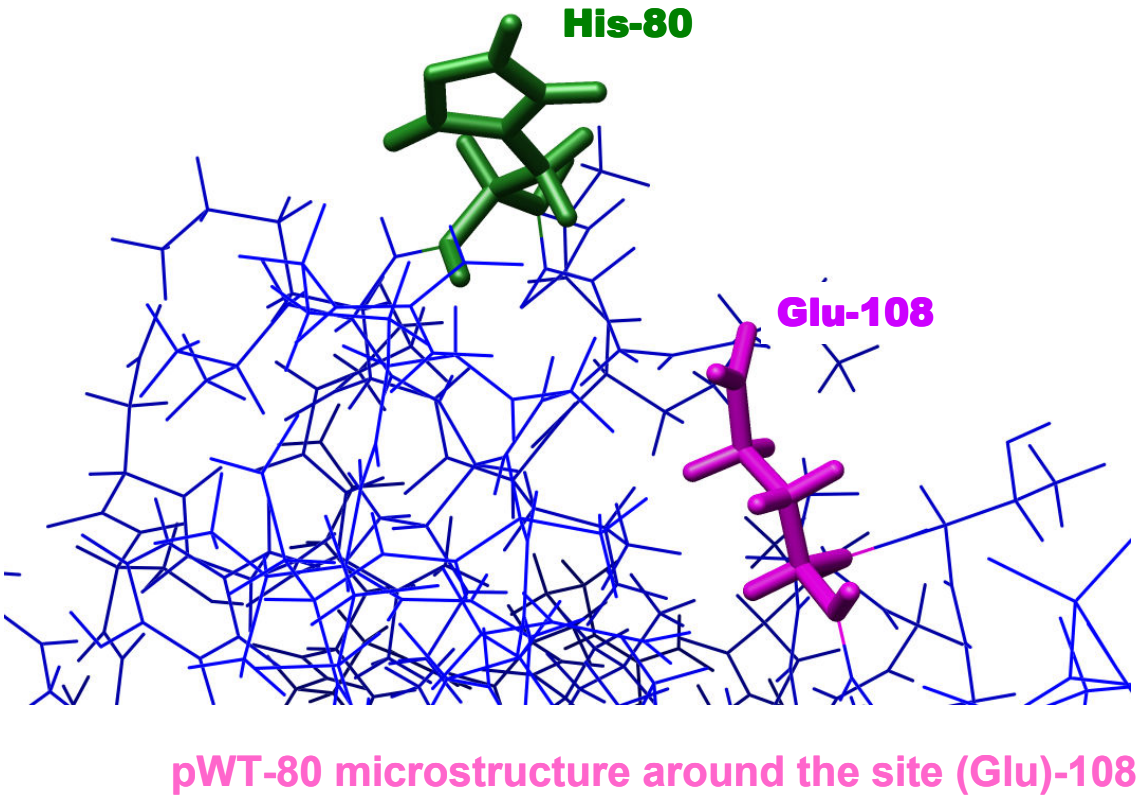
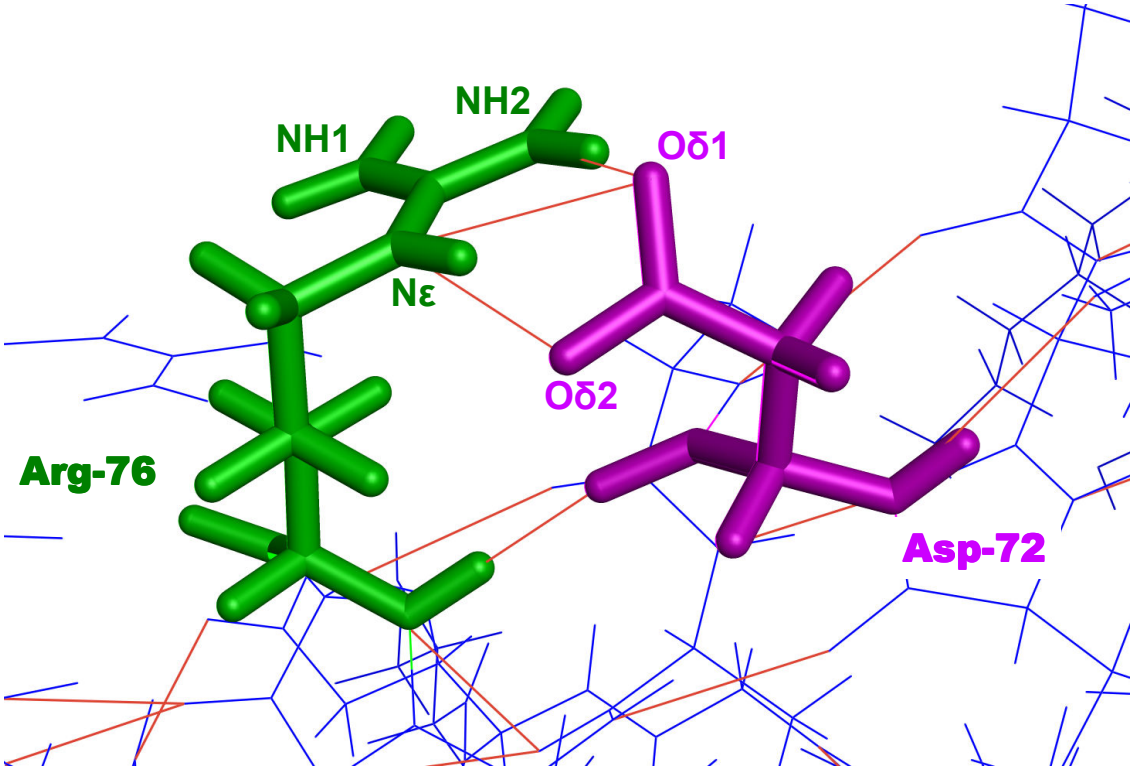


Figure 7. The microstructure of pWT-80, wherein the side chain of Glu108 is oriented towards His80.



A) pWT-80 microstructure around the site (Arg)-76

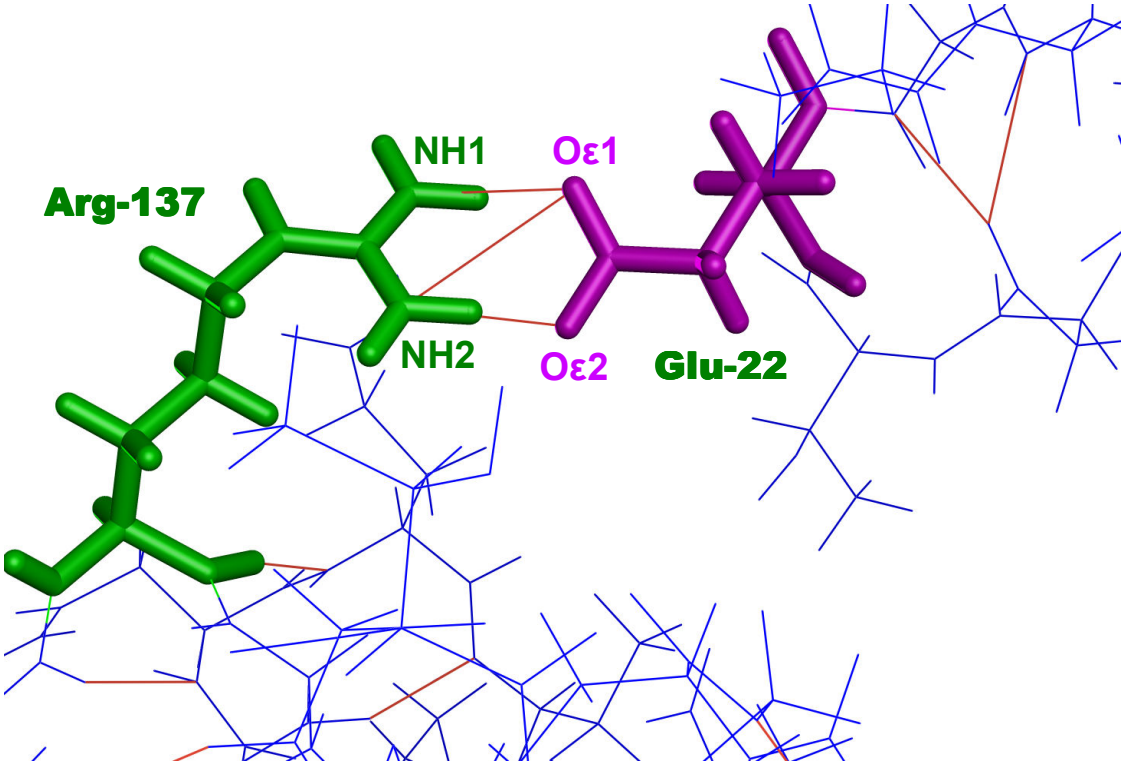


Figure 8 (A and B). The microstructure of A) pWT-80, wherein the side chain of Arg76 is involved in a salt bridge (red lines) with Asp72, and B) pWT-137, wherein the side chain is involved in a salt bridge (red lines) with Glu22.

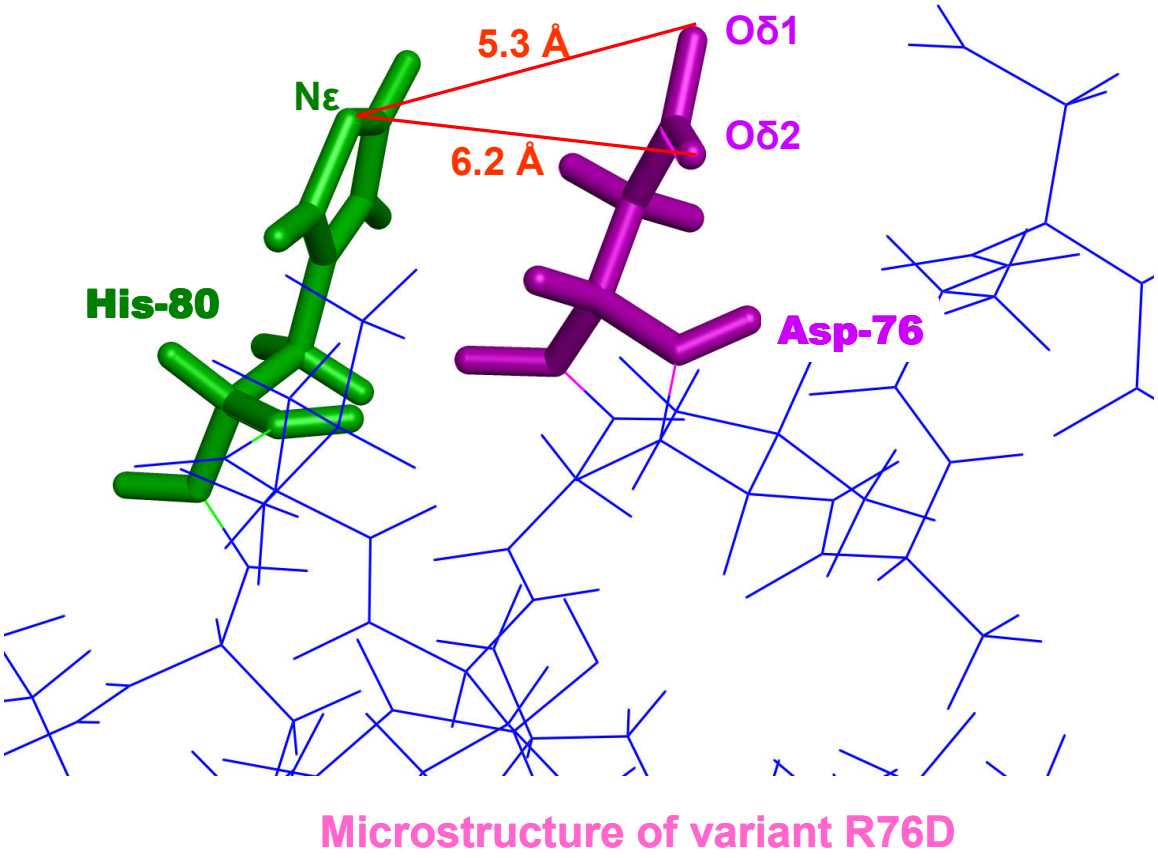


Figure 9. Microstructure of the variant R76D, wherein the side chain of Asp76 is involved in a (weak) salt bridge with the side chain of His80.

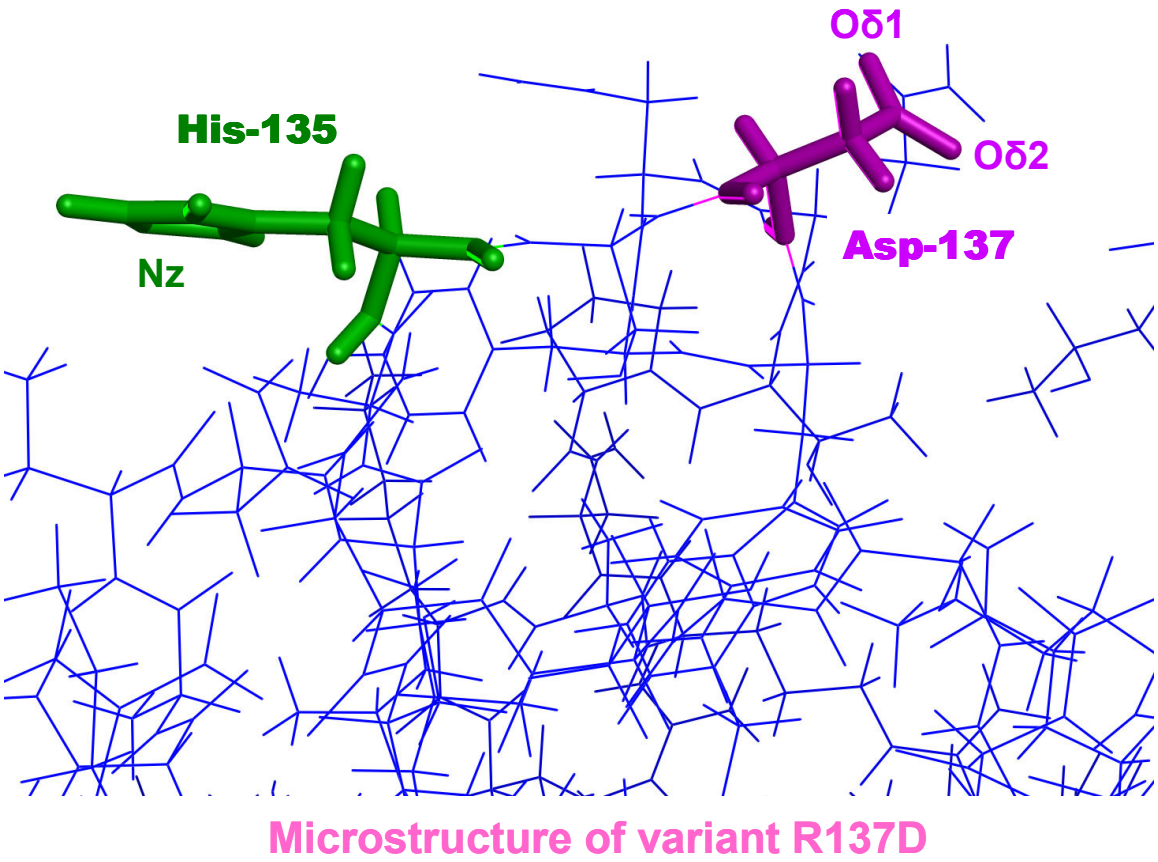


Figure 10. Microstructure of variant R137D, wherein the two charged moieties of His135 and Asp137 are oriented in apposite direction and are 16.7 Å apart.

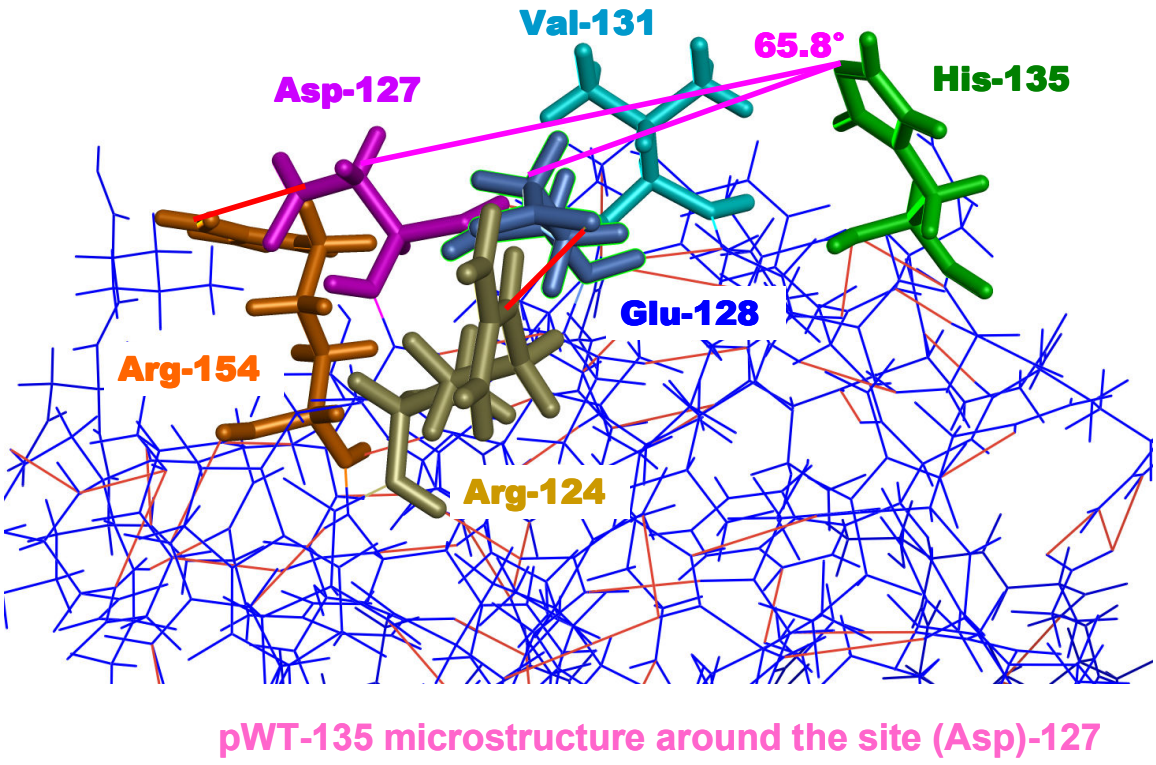


Figure 11. The microstructure of pWT-135 around the site (Asp)-127. The angle between three residues His135, Asp127 and Glu128 is 65.8° (shown in Pink). The ion pair formation between the charged moieties of 1) Asp127 and Arg154 and 2) Glu128 and Arg124 is shown by the line highlighted in red.

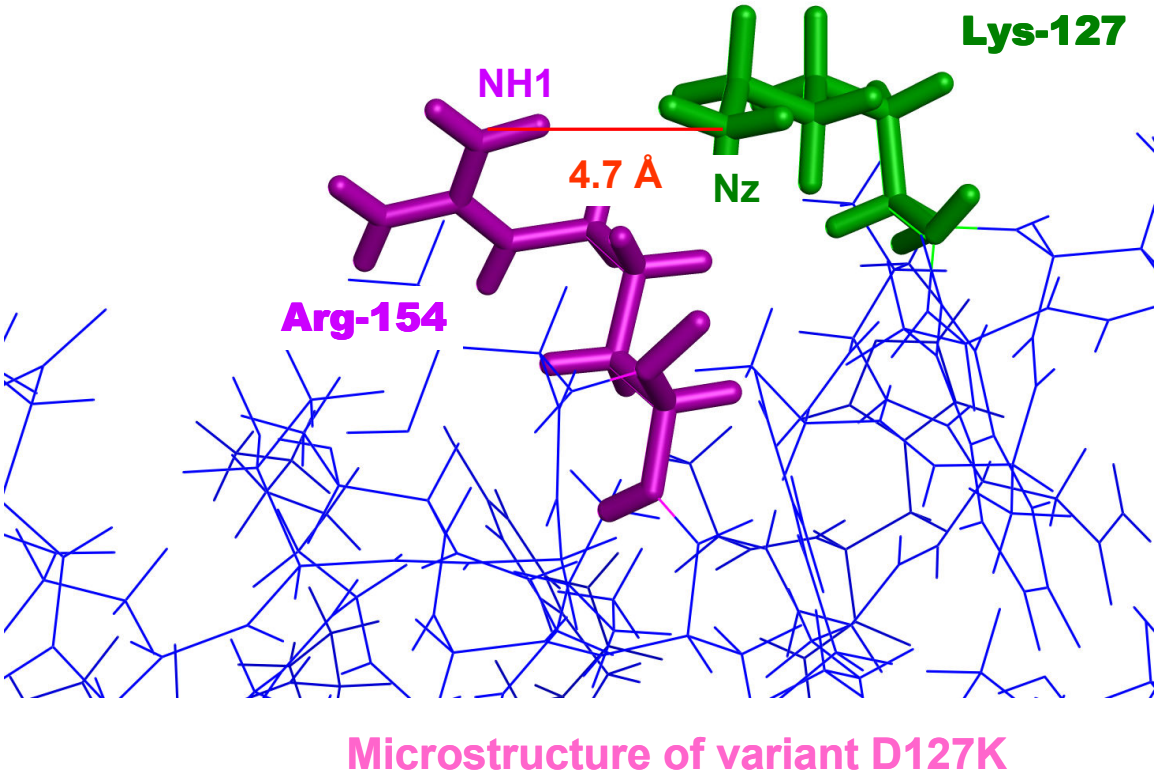


Figure 12. Microstructure of variant D127K, wherein the side chain of Lys127 is involved in an ion pair of similar charges. The distance between the charged moieties is shown by the red line.

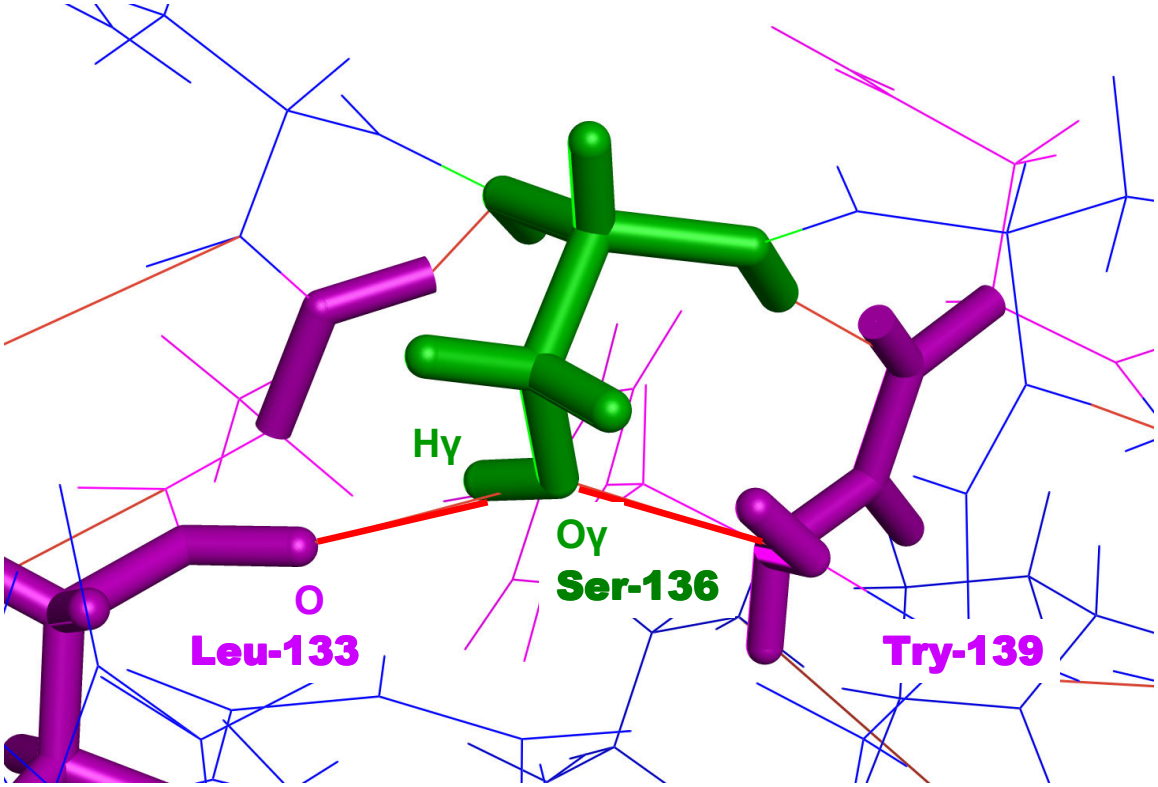


Figure 13. The microstructure of pWT-135, wherein the side chain of Ser136 (OH) is involved in two hydrogen bonds (highlighted in red) with the Try139 (N) and Leu133 (O).

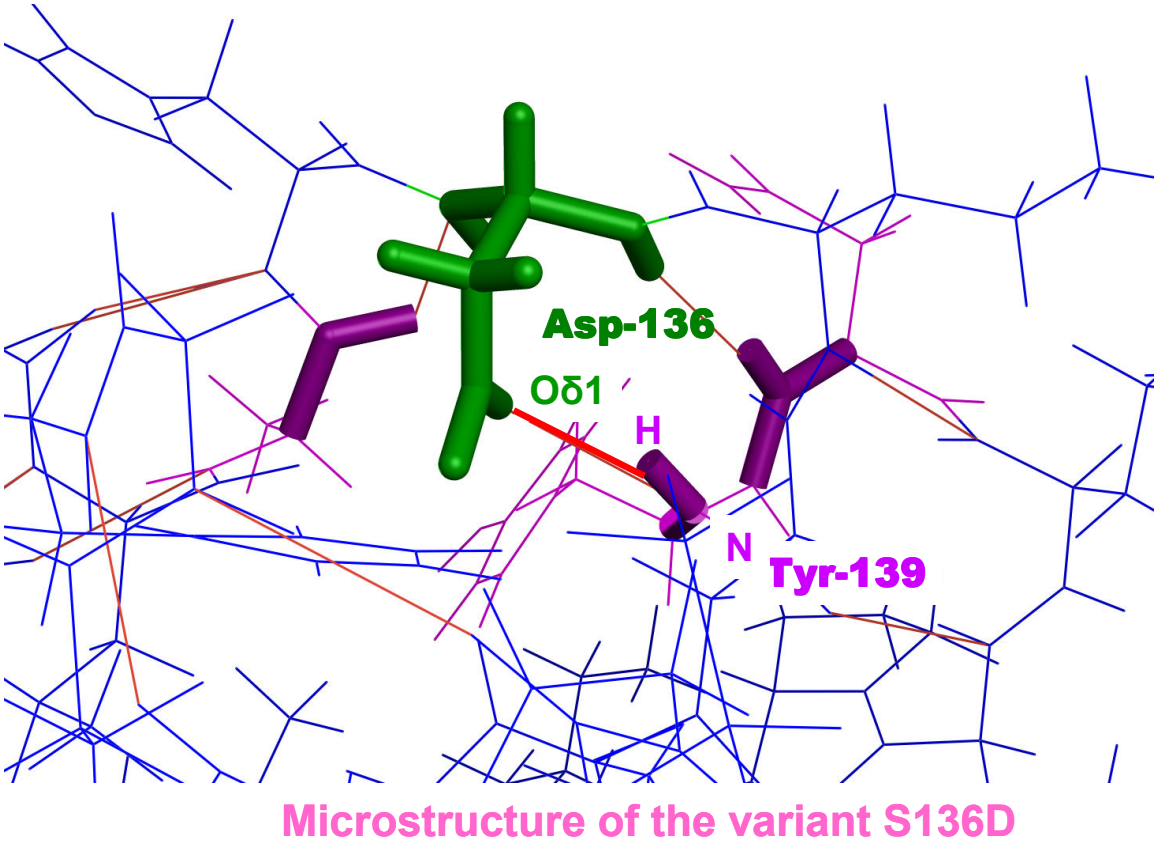


Figure 14. The hydrogen bonds (highlighted in red) formed by the aspartate (O) with Try139 (H-N).

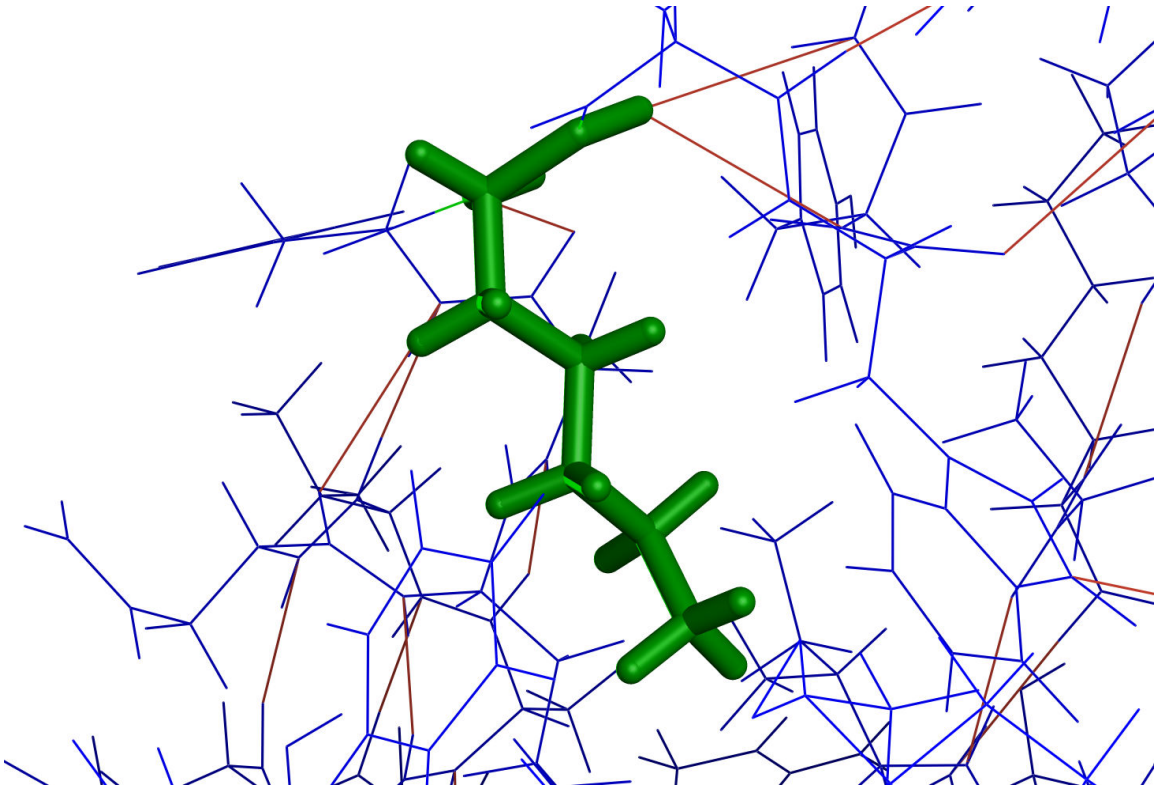


Figure 15. Microstructure of the variant S136K, wherein the side chain of the lysine residue has no hydrogen bonds with its adjacent amino acid residues and is oriented inwards into the protein.

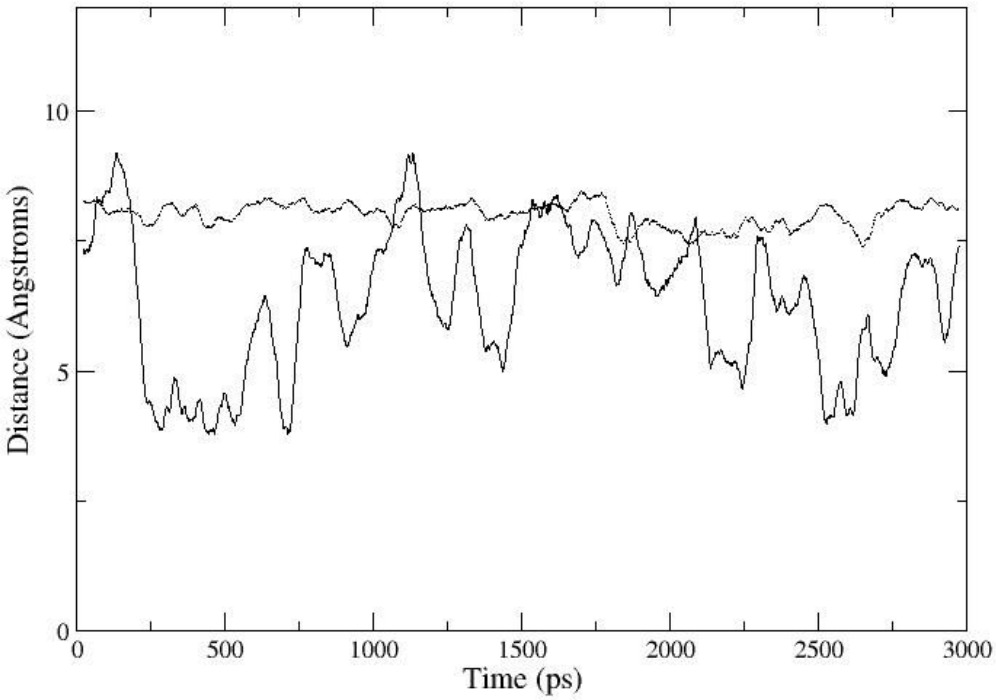


Figure 16. Distance between His135 and residue 136 (in pWT-135 and S136K) during the MD simulations. The lower solid curve depicts the distance between atom ND1 of His135 and atom NZ of Lys-136 in the S136K simulation. The upper dotted line is the distance between atom ND1 of His135 and atom OG of Ser136 in the pWT-135 simulation.

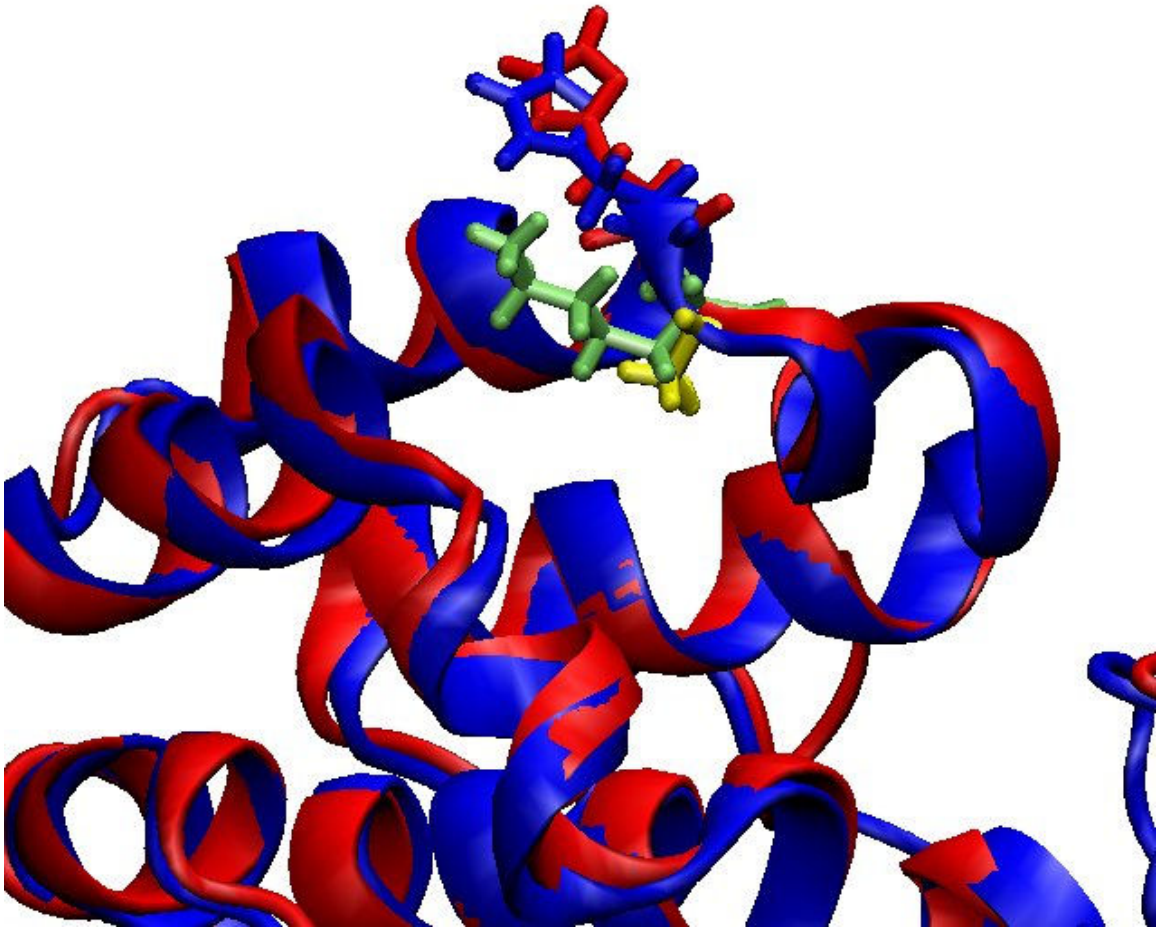


Figure 17. The structures are snapshots at the 3-ns time point in the MD simulations. Structures are illustrated in ribbon format for pWT-135 in blue and S136K in red. TheHis135 residue in each molecule is colored the same as the ribbon backbone. The residue ser136 in pWT-135 is colored in yellow and the lysine that replaces serine in S136K in is colored green.

References

(1) Teague, S. J. *Nature Reviews Drug Discovery* **2003**, *2*, 527-541.
(2) Kumar, S.; Nussinov, R. *Biophysical Journal* **2002**, *83*, 1595-1612.
(3) Schellekens, H. *Nature Biotechnology* **2004**, *22*, 1357-1359.
(4) Kumar, S.; Nussinov, R. *Chembiochem* **2002**, *3*, 604-617.
(5) Sivasankar, S.; Gumbiner, B.; Leckband, D. *Biophysical Journal* **2001**, *80*, 1758-1768.
(6) Gitlin, I.; Carbeck, J. D.; Whitesides, G. M. *Angewandte Chemie-International Edition* **2006**, *45*, 3022-3060.
(7) Russell, A. J.; Thomas, P. G.; Fersht, A. R. *Journal of Molecular Biology* **1987**, *193*, 803-813.
(8) Louro, R. O.; Catarino, T.; Paquete, C. M.; Turner, D. L. *Febs Letters* **2004**, *576*, 77-80.
(9) Lee, K. K.; Fitch, C. A.; Garcia-Moreno, B. *Protein Science* **2002**, *11*, 1004-1016.
(10) Loewenthal, R.; Sancho, J.; Reinikainen, T.; Fersht, A. R. *Journal of Molecular Biology* **1993**, *232*, 574-583.
(11) Allewell, N. M.; Oberoi, H. *Methods in Enzymology* **1991**, *202*, 3-19.
(12) Jeffrey, G. A. *Food Chemistry* **1996**, *56*, 241-246.
(13) Thomas, P. G.; Russell, A. J.; Fersht, A. R. *Nature* **1985**, *318*, 375-376.
(14) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1993**, *32*, 13909-13916.
(15) Laurents, D. V.; Huyghues-Despointes, B. M. P.; 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.; Pace, C. N. *Journal of Molecular Biology* **2003**, *325*, 1077-1092.
(16) Warshel, A. *Biochemistry* **1981**, *20*, 3167-3177.
(17) Russell, A. J.; Fersht, A. R. *Nature* **1987**, *328*, 496-500.
(18) de Kreij, A.; van den Burg, B.; Venema, G.; Vriend, G.; Eijsink, V. G. H.; Nielsen, J. E. *Journal of Biological Chemistry* **2002**, *277*, 15432-15438.
(19) Schutz, C. N.; Warshel, A. *Proteins-Structure Function and Genetics* **2001**, *44*, 400-417.
(20) Alvaro, G.; Russell, A. J. *Methods in Enzymology* **1991**, *202*, 620-643.
(21) Joshi, M. D.; Sidhu, G.; Nielsen, J. E.; Brayer, G. D.; Withers, S. G.; McIntosh, L. P. *Biochemistry* **2001**, *40*, 10115-10139.
(22) Sundd, M.; Iverson, N.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Robertson, A. D. *Biochemistry* **2002**, *41*, 7586-7596.
(23) Mehler, E. L.; Fuxreiter, M.; Simon, I.; Garcia-Moreno, E. B. *Proteins-Structure Function and Genetics* **2002**, *48*, 283-292.
(24) Chicz, R. M.; Regnier, F. E. *Analytical Chemistry* **1989**, *61*, 1742-1749.
(25) Chicz, R. M.; Regnier, F. E. *Journal of Chromatography A* **1990**, *500*, 503-518.
(26) Huyghues-Despointes, B. M. P.; Thurlkill, R. L.; Daily, M. D.; Schell, D.; Briggs, J. M.; Antosiewicz, J. M.; Pace, C. N.; Scholtz, J. M. *Journal of Molecular Biology* **2003**, *325*, 1093-1105.
(27) Mehler, E. L.; Eichele, G. *Biochemistry* **1984**, *23*, 3887-3891.
(28) Edgcomb, S. P.; Murphy, K. P. *Proteins-Structure Function and Genetics* **2002**, *49*, 1-6.

(29) Pathange, L. P.; Bevan, D. R.; Larson, T. J.; Zhang, C. M. *Analytical Chemistry* **2006**, *78*, 4443-4449.
(30) Bastyns, K.; Froeyen, M.; Diaz, J. F.; Volckaert, G.; Engelborghs, Y. *Proteins-Structure Function and Genetics* **1996**, *24*, 370-378.
(31) Garcia-Mayoral, M. F.; Perez-Canadillas, J. M.; Santoro, J.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Lacadena, J.; del Pozo, L. M.; Gavilanes, J. G.; Rico, M.; Bruix, M. *Biochemistry* **2003**, *42*, 13122-13133.
(32) Li, J. W. *Chromatographia* **2005**, *61*, 479-492.
(33) Mehler, E. L.; Guarnieri, F. *Biophysical Journal* **1999**, *77*, 3-22.
(34) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophysical Journal* **2002**, *83*, 1731-1748.
(35) Mehler, E. L. *Journal of Physical Chemistry* **1996**, *100*, 16006-16018.
(36) Whitten, S. T.; Garcia-Moreno, B.; Hilser, V. J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 4282-4287.
(37) Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. *Nature* **1975**, *258*, 598-599.
(38) Porath, J. *Trac-Trends in Analytical Chemistry* **1988**, *7*, 254-259.
(39) Porath, J. *Protein Expression and Purification* **1992**, *3*, 263-281.
(40) Berna, P. P.; Mrabet, N. T.; VanBeeumen, J.; Devreese, B.; Porath, J.; Vijayalakshmi, M. A. *Biochemistry* **1997**, *36*, 6896-6905.
(41) Zhao, Y. J.; Sulkowski, E.; Porath, J. *European Journal of Biochemistry* **1991**, *202*, 1115-1119.
(42) Liu, R. J.; Baase, W. A.; Matthews, B. W. *Journal of Molecular Biology* **2000**, *295*, 127-145.
(43) Daopin, S.; Sauer, U.; Nicholson, H.; Matthews, B. W. *Biochemistry* **1991**, *30*, 7142-7153.
(44) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403-2408.
(45) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. *Proteins-Structure Function and Genetics* **2000**, *39*, 261-268.
(46) Baldwin, E. P.; Hajiseyedjavadi, O.; Baase, W. A.; Matthews, B. W. *Science* **1993**, *262*, 1715-1718.
(47) Tripos Associates Inc., S. L., USA Tripos, Inc.
(48) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *Journal of Computational Chemistry* **2004**, *25*, 1605-1612.
(49) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714-2723.
(50) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *Journal of Physical Chemistry A* **1997**, *101*, 3005-3014.
(51) Higuchi, R., Ed. *Using PCR to engineer DNA.*; Stockton Press: New York, 1989.
(52) Yang, B.; Gerhardt, S. G.; Larson, T. J. *Molecular Microbiology* **1997**, *24*, 511-521.
(53) Rennell, D.; Bouvier, S. E.; Hardy, L. W.; Poteete, A. R. *Journal of Molecular Biology* **1991**, *222*, 67-87.
(54) Barbas, C. F.; Heine, A.; Zhong, G. F.; Hoffmann, T.; Gramatikova, S.; Bjornestedt, R.; List, B.; Anderson, J.; Stura, E. A.; Wilson, I. A.; Lerner, R. A. *Science* **1997**, *278*, 2085-2092.

(55) Tishmack, P. A.; Bashford, D.; Harms, E.; VanEtten, R. L. *Biochemistry* **1997**, *36*, 11984-11994.
(56) Dunn, B. M.; Chaiken, I. M. *Proceedings of the National Academy of Sciences of the United States of America* **1974**, *71*, 2382-2385.
(57) Finette, G. M. S.; Mao, Q. M.; Hearn, M. T. W. *Journal of Chromatography A* **1997**, *763*, 71-90.
(58) Lazaridis, T. *Current Organic Chemistry* **2002**, *6*, 1319-1332.
(59) Warshel, A.; Aqvist, J. *Chemica Scripta* **1989**, *29A*, 75-83.
(60) Pace, C. N.; Huyghues-Despointes, B. M. P.; Briggs, J. M.; Grimsley, G. R.; Scholtz, J. M. *Biophysical Chemistry* **2002**, *101*, 211-219.
(61) Wray, J. W.; Baase, W. A.; Lindstrom, J. D.; Weaver, L. H.; Poteete, A. R.; Matthews, B. W. *Journal of Molecular Biology* **1999**, *292*, 1111-1120.
(62) Matthews, B. W. In *Advances in Protein Chemistry, Vol 46*, 1995; Vol. 46, pp 249-278.
(63) Sternberg, M. J. E.; Hayes, F. R. F.; Russell, A. J.; Thomas, P. G.; Fersht, A. R. *Nature* **1987**, *330*, 86-88.
(64) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 689-704.
(65) Highbarger, L. A.; Gerlt, J. A.; Kenyon, G. L. *Biochemistry* **1996**, *35*, 41-46.
(66) Harris, T. K.; Turner, G. J. *Iubmb Life* **2002**, *53*, 85-98.
(67) Eyal, E.; Najmanovich, R.; Edelman, M.; Sobolev, V. *Proteins-Structure Function and Genetics* **2003**, *50*, 272-282.
(68) Yuan, Z.; Bailey, T. L.; Teasdale, R. D. *Proteins-Structure Function and Bioinformatics* **2005**, *58*, 905-912.
(69) Vieira, E.; Binggeli, A.; Breu, V.; Bur, D.; Fischli, W.; Guller, R.; Hirth, G.; Marki, H. P.; Muller, M.; Oefner, C.; Scalone, M.; Stadler, H.; Wilhelm, M.; Wostl, W. *Bioorganic & Medicinal Chemistry Letters* **1999**, *9*, 1397-1402.
(70) McHaourab, H. S.; Kalai, T.; Hideg, K.; Hubbell, W. L. *Biochemistry* **1999**, *38*, 2947-2955.
(71) McHaourab, H. S.; Lietzow, M. A.; Hideg, K.; Hubbell, W. L. *Biochemistry* **1996**, *35*, 7692-7704.
(72) Tran, N. L.; Colvin, M. E.; Gronert, S.; Wu, W. M. *Bioorganic Chemistry* **2003**, *31*, 271-277.
(73) Laberge, M. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1998**, *1386*, 305-330.
(74) Price, N. C. *Biotechnology and Applied Biochemistry* **2000**, *31*, 29-40.
(75) Hwang, J. K.; Warshel, A. *Nature* **1988**, *334*, 270-272.

Chapter V: Correlation between protein microstructure and retention in cation exchange chromatography - Importance of intramolecular bonds

Abstract

The structure of a protein directly affects its function. Therefore, characterization of recombinant protein structures is important but is a challenging task. One of the important forces that plays a major role in maintaining both structural and functional properties of proteins is electrostatic interactions among different amino acid residues. In this article, cation exchange chromatography was used to study how the microstructure of some charged amino acid residues may affect the protein's retention. Two sets of T4 lysozyme variants were generated. The first set included seven variants that vary in their charge distribution. These variants were obtained by replacing a charged amino acid residue at different sites on lysozyme. The second set included ten variants that vary in both net charge and charge distribution, and these variants were obtained by replacing charged and neutral amino acid residues at different sites on the protein. The microstructure was quantified by calculating the relative hydrophilicity around the replacing amino acid residue. The retention times of all variants were compared with the retention time of the control variants. Among the first set of variants, there was a direct correlation (R^2= 0.95) between the relative hydrophilicity of the replaced amino acid and the protein's retention time, except for two variants (K83H and K124H) whose replacing amino acid residue was involved in intramolecular interactions. For the second set, there was a direct correlation (R^2= 0.93) between the change in net charge (-2 to +2 units) and the retention times. However, the retention times of two variants (R76D and R76S) did

not follow the correlation. We hypothesize that the structure around the replacing charged group is responsible for the deviated protein retention pattern.

Key words: protein microstructure, intramolecular bonds, cation exchange chromatography, lysozyme, protein retention, protein characterization

Introduction

The overall goal of this article is to elucidate structure-function (a protein's ability to interact with oppositely charged groups) relationships using ion exchange chromatography (IEC) by correlating the same aforementioned two structural parameters and the resultant functional behavior (binding) in an IEC column. Understanding this correlation is important because the mechanism of protein binding in IEC is different from that in IMAC. The general trend in IEC is that the elution is dependent on the net charge of a protein (or its pI). However it has been demonstrated that net charge is not the only criterion for protein elution in IEC, as indicated by stoichiometric displacement model (SDM).[1, 2] Though the whole protein surface modulates the protein binding process, the specific binding interactions are dominated by certain contact regions.[3] Thus, the contribution of all charged amino acid residues in IEC do not contribute equally to the protein binding process.[1] The main goal of our experiments was to analyze subtle variation in protein charge including its position on the protein's surface by IEC. Such studies not only will reveal the ability of IEC in capturing subtle changes in protein structure but also may provide insight into the protein binding mechanism in IEC.

Binding mechanism in IEC: To understand the behavior of all the protein variants, it is crucial to understand the binding mechanism of proteins in IEC. IEC is a widely used protein purification technique at both preparative and analytical scale. In fact, it is considered as the work horse of the biopharmaceutical industry.[4] In IEC, the partition of a solute onto a chromatographic column primarily depends on the physicochemical characteristics of both the solute as well as the stationary column. The main components that control the behavior of a protein are[5] 1) protein structure including

charge density, 2) type of stationary phase, 3) pH and type of buffer, and 4) elution conditions (e.g. gradient or step wise increase of salt concentration in the elution buffer). For all our current experiments, the first factor (protein structure) was thoroughly investigated by generating a series of point mutations on a model protein, T4 lysozyme, via site-directed mutagenesis.

The two important aspects, which control a protein's partitioning between the mobile and stationary phase in IEC and thereby determine the important purification factor known as resolution, are 1) thermodynamic equilibrium and 2) mass transport.[6, 7] The former, primarily influenced by the protein structure (electrostatics) for a given stationary phase, determines the protein retention time. The latter affects the column efficiency, which is measured by protein elution peak width or height equivalent to a theoretical plate (HETP).

Generally, thermodynamic equilibrium processes are characterized by change in Gibbs free energy (ΔG), which is comprised of two factors, a) change in enthalpy (bond energy) (ΔH) and b) change in entropy (water structure) ($T\Delta S$). Therefore the presence of an attractive or a repulsive force (that contributes to ΔH) due to a given protein (net protein charge) in a given type of (anion or cation[8]) IEC column affects the ΔG (retention times). The term representing ΔS (desolvation) is generally not a dominating factor in IEC, unlike in the other chromatographic techniques such as reverse phase. Thus, when the structure of a protein is altered then there should be an associated variation in its IEC behavior. Here we have used the converse of this inference to characterize the structural variants of a protein, using retention time in an IEC column.

The other aspect that contributes to the binding mechanism is mass transport. Though this aspect does not directly contribute to the retention times, a brief insight would elucidate some subtle aspects that facilitate or control the precise thermodynamic equilibrium. The adsorption process involves three major steps, they are 1) mass transfer from mobile to stationary phase, (film resistance) controlled by convection (flow rate), 2) pore or surface diffusion controlled by concentration gradient of the protein, and 3) reversible binding (thermodynamic equilibrium) controlled by the nature of the protein structure.[7, 9] Generally, the rate limiting step is believed to be the second step, which is of the order of 10^{-7} cm^2/s for proteins.[6, 7]

According to Johnson et al.,[8] the diffusion of a protein in charged gels is dependent on the following equation:

$$D_e = \Phi * D_b$$

Where D_e is the effective diffusivity in stationary phase, Φ is the partition coefficient and D_b is the diffusivity in the membrane gels. They concluded that effective diffusivity in charged gels is mainly due to partitioning effect. In IEC, Φ is primarily based on the electrostatic properties of both the stationary media and protein molecule[7, 10] for a particular mobile phase. Therefore electrostatics determines Φ, which in turn controls the driving force (protein concentration gradient) for diffusion. In our current experiments, the stationary phase is kept constant, and therefore Φ is primarily dependent on protein structure. Thus, whenever there is a subtle variation (in charge or in microenvironment) in the protein structure, there ought to be a discernable difference in the binding behavior of protein in an IEC column.

Materials and Methods

In silico methods: The model protein chosen for our experiments is a cysteine-free T4 lysozyme (C-WT).[11] Two categories representing two structural parameters were generated by homology modeling using SYBYL 7.1 software.[11] In category-1, seven sites were chosen for site-directed mutagenesis and were based on their relative surface accessibility (rSA) (Table 1). All the seven original amino acid residues were replaced by histidine residue, which varies in the location on the protein surface (Fig.1). The control variant for this category is C-WT. The microenvironment around a surface accessible amino acid residue was characterized by its net charge (Table 1), which was calculated by simple sum of charged residues with in 15 Å distance around the amino acid residue (Table 1). In case of category-II, two point mutants with histidine at sites 135 and 80 (from category I) were chosen as the two control variants pWT-135 and pWT-80, respectively. These control variants have similar surface accessibility with high binding strength on IMAC-Cu^{2+} column, and therefore were chosen as control variants. As a result category II was further subdivided into two subsets with pWT-135 as a control set and pWT-80 as a test set.[12] Five sites were chosen on these two control variants, three in pWT-135 (Fig. 2a) and two in pWt-80 (Fig. 2b) variants. In total ten variants were generated, which vary in net charge (Table 2).

Relative hydrophilicity (rHy): Relative hydrophobicity/hydrophilicity value (rHy), which characterizes the microstructure around all the individual residues or sites, was calculated using a modification of a published method.[13] The sum of all the hydrophobic/hydrophilic constants (sHy) of the moieties around these sites (4.25 Å) was calculated manually from the X-ray crystal structure of cysteine free T4 lysozyme (1L63).[13] In this article the rHy values represent the hydrophilicity nature of the

microstructures, and is obtained from the Equation 1a.[13] The rSA and rHy values for categories I and II are reported in Table 1 and Table 3, respectively. Strongly hydrophilic microstructures are characterized by higher rHy values.

$$\mathrm{tHy} = (1-\mathrm{SA})\cdot \mathrm{sHy} + (\mathrm{SA})\cdot \mathrm{aHy} \quad -----(1a)$$

$$\mathrm{rHy} = \frac{\mathrm{tHy}}{\mathrm{aHy}} \quad -----(1b)$$

Where, SA is surface accessibility, tHy is the total hydrophilicity, and aHy is the absolute hydrophilicity in water obtained from a method by Mehler et al.[14]

Experimental: All variants were overexpressed in *E. coli* and purified by a published method using cation exchange chromatography.[11, 12] From earlier studies,[11, 12] based on circular diochroism (CD) data and enzymatic assays, it was demonstrated that all the seventeen lysozyme variants had intact 3D structures. The nucleotide and amino acid sequences of all the variants were also verified.

Determination of rRT in IEC: A strong cation exchange (100 × 4.6 mm) HPLC column (Analytical Sales and Services Inc, Pompton Plains, NJ), with particle size 6 μm, was used to obtain the retention time of all variants under identical experimental conditions. The sample concentration used was 0.15 mg/ml. After the sample (45 μl) was loaded on to the column, the equilibrating buffer (20 mM sodium phosphate at pH 7.0) was run for 4 column volumes (CV). Then, a two step linear gradient was employed using an elution buffer (1M NaCl, 20 mM sodium phosphate at pH 7.0). In the first step 10% of the elution buffer was reached in 3 CV. In the second step, wherein all seventeen variants and the control variants (C-WT, pWt-135 and pWT-80) eluted, 45% of the elution buffer was reached in 16 CV. The flow rate of the mobile phase was 1 ml/min.

The retention time (RT) values (elution peak maxima), an average of two measurements, were determined at 280 nm. The rRT was determined by the following equation:

$$rRT = \frac{RT_C - RT_V}{RT_C} \times 100 - - - - - (2)$$

Where RT_V stands for the retention time of variants and RT_C stands for retention time of the control variants. For 1) category I it is the retention time of C-WT, and 2) category II it is RT of pWT-135 for subset-1 and pWT-80 for subset-2.

Results

Effect of protein structure on its behavior in IEC: The major components of the protein structure that affects its binding behavior in an IEC column are 1) overall net charge on the protein, 2) the type of individual surface amino acids, 3) the microenvironment, and 4) the microstructure around the surface amino acids (including its own rSA) that are involved in the binding process. In this report, all these four components were theoretically quantified by 1) change in net charge, 2) positively charged, negatively charged, or neutral amino acids, 3) net charge in the microenvironment, and 4) rHy, respectively. Experimentally, all these four components were investigated by replacing selected original amino acid residues with different amino acid substitutions such as histidine, lysine and serine. The category I variants vary specifically in factor 3 and 4, wherein they vary in their molecular microenvironments (same net charge or same pI). The category II variants vary in all four structural components. The effect of these structural components on the binding behavior of the

protein was quantified in terms of relative retention time (rRT). The rRT values for categories I and II are reported in Table 1 and Table 3, respectively.

Category I: In cation exchange chromatography, the elution time of the proteins is directly propomotional to the total net positive charge. All the seven variants eluted earlier than C-WT indicating that the lowered pI (or net positive charge), due to the removal of +1 charge amino acid residue (lysine or arginine) and replacement with +0.5 charge (histidine at pH 7), had an expected impact on the rRT values. However, although the apparent change of the net charge is the same for all variants, the impact was not uniform. The change in rRT values ranges from 0.2% to 11% from the C-WT rRT value (Table 1). This non-uniform retention displayed by category I variants can be attributed to the difference in the microstructure of the original amino acids because the replaced amino acid is same in all the cases. Thus, rHy and the net charge in the microenvironment are the two parameters that differ for all seven sites in the seven variants (Table 1). A correlation between the characterized microstructure (rHy) and its effects on the protein behavior (rRT) is shown in Fig. 3. A direct correlation ($R^2 = 0.93$) between the rHy and the rRT was obtained when the two outlier variants K124H and K83H were excluded. These results illustrates that all these sites excluding the outliers seem to be either a part of a contact region[15] or an individual interacting site. There was no observable trend between the net charge in the microenvironment and the binding behavior for all seven variants (Table 1). Thus, despite having the same net charge or pI, the variants elute at different times during cation exchange chromatography according to the histidine's microstructure (hydrophilicity).

Excluded variants K124H and K83H: These two variants deviate significantly from the trend (Fig. 3). For both proteins, the histidine microstructure did not seem to play an important role in their chromatographic behavior. This indicates that 1) these sites (124 and 83) are not situated in one of the contact regions or 2) there are some other factors that need to be considered such as the location of sites (steric effects) and its involvement in intramolecular interactions. In the case of K124H, the original lysine residue (at 124) in C-WT variant was involved in a salt-bridge (Fig. 4) and therefore its contribution to the binding may be negligible. In addition the orientation of Lys124 is projected into the protein interior. On the other hand, by assuming the orientation of histidine in K124H variant is similar to that of lysine in C-WT variant, we can predict that the histidine's contribution to protein binding in K124H might be also lower than expected. Thus, Lys124 and His124 sites are non-interacting and less-interacting sites, respectively.

In the case of K83H, the Lys83 site in C-WT might be an interacting site due to 1) its high rSA including its favorable orientation, to binding, pointing outward away from the protein's interior and 2) its non-involvement in any intramolecular interactions. However, its replacement amino acid, histidine, is involved in a hydrogen bond (Fig. 5)[11] and therefore its contribution to the binding may be lower (less-interactive site) than other variants such as K135H or R80H. Thus, the presence of intramolecular bond, in addition to rHy, affects the binding behavior of lysozyme variants in IEC.

Category II: As stated earlier, the elution of proteins in cation exchange chromatography is dependant on their pI. Proteins with higher pI elute later than proteins with lower pI. All variants faithfully eluted according to their pI (net charge) when

compared to the pWT (Table 2). These results are similar to those reported in literature (Chicz et al.). The change in rRT values range from -28.5% to 35.1%, from the pWT (control variants) rRT value (Table 2). Assuming negligible microstructure (rHy) effect, the variation in rRT values is approximately 15% per unit change in net charge. But from Table 2, it is evident that the effect of unit net charge is not uniform. When the two variants R76D and R76S were excluded, there is a strong correlation ($R^2 = 0.97$) between change in net charge and rRT (Fig. 6) values, and the strong correlation demonstrates the importance of overall net charge on the proteins. Though the variant S136K is included in this correlation, its rRT value is conspicuously very low when compared to another ±1 change in net charge variant (D127S or R137S). This noticeable difference can be attributed to the replaced lysine residue's low rHy value (0.08), indicating a very hydrophobic microenvironment. The main reason for such a low rHy value in case of lysine is due to 1) the low surface accessibility of its side chain (predicted by SYBYL software (Tripos Associates Inc., St. Louis, USA)) (Fig. 7), which is pointed into the protein interior and 2) its microstructure lowers its pK_a value thus its contribution to total net charge is minimal.

Excluded variants R76S and R76D: The variants at site 76 (R76S and R76D) are similar to the variants at site 137 (R137S and R137D). These two sites can be characterized as having similar microstructural environment due to the following factors: 1) equivalent rSA values, 2) involvement of their arginine residue in a salt bridge (Fig. 8), 3) there is no significant difference between the microenvironmental net charge value of -1 for site 137 and +1 for site 76 (when compared with the category I values in Table 1), and 4) equivalent rHy values (Table 2). As expected, due to the replacement of the

positively charged arginine at these two sites by serine and aspartic acid, the rRT values are negative. But the rRT values of variants at site 76 are inordinately lower than the values at site 137 (Table 2). This variation can be explained by the formation of intramolecular bonds.

In case of variant R76D, Asp76 is between His80 and Asp72 (Fig. 9). Due to the presence of Asp72 and its repulsive force, there is strong possibility that Asp76 forms a salt bridge with His80 (Fig. 9). Thus, the pK_a of aspartate is raised and the total net negative charge on the protein is reduced. Thus, the individual interacting site 76 is now rendered as a less-interacting site. Similarly in case of R76S, as demonstrated by molecular modeling (Fig. 10), there likely is a weak hydrogen bond between Ser76 and Asp72 rendering Asp72 as a less interacting site. Therefore, due to the compromised net negative charge because of the presence of intramolecular interactions, the variants at site 76 eluted much later than the variants at 137.

Discussion

Although it is conceivable that the whole protein surface plays a major role in the chromatographic behavior of a protein, Regnier et al. have demonstrated that there are certain regions or patches on the protein surface known as 'contact regions' that are predominantly involved in the chromatographic binding process.[15, 16] They have also indicated that the regions that are not the contact regions seem to contribute very little to the binding process. Yao et al.[5] were able to use continuous electrostatic modeling program to elucidate the presence of contact regions and its possible effects on the protein behavior in IEC. By measuring the interacting free energy (IFE) values of few regions, they were able to confirm that the front face of cytochrome *c*, which has

abundant positive charges, was the contact region. They also demonstrated a direct correlation between the local surface potentials of a contact region and the retention times in IEC. Therefore it can be inferred that in IEC the presence of anisotropic charge on the protein surface results in contact regions that have high charge densities and thus can contribute to higher enthalpic energies. Therefore any variation in the contact region, such as altering the charge density, affects its thermodynamic favorability in an IEC column.

Another factor, in addition to charge density, which can affect the contact region, is the shape of the protein molecule which induces steric effects during the binding process.[15, 17] These effects can be explained by the heterogeneous physical accessibility, i.e. the steric effects, of all sites on the protein surface to the chromatographic binding (media) sites. The exact nature of this steric effect is still not fully understood due to the poor knowledge of structure-function relationships in protein molecules. The steric effects are experienced due to the inherent heterogeneous nature of the protein molecules, which is also partially responsible to the anisotropic charges distribution on the protein surface. Thus, it is hard to differentiate the effects of charge density and steric contributions to the protein's chromatographic behavior.

Here, we have used the total charge within 15 Å of the surface amino acid as a measure to understand the effect of microenvironment and its location on the 3D structure of lysozyme . Our results indicate that the microenvironment effect on the binding process in the cation exchange chromatography do not follow any particular trend (Table 1). A closer look at these contact regions, which generally consists of one or more charged amino acid residue, might help in elucidating the relationship between

steric and charge effects. We have employed rHy as a parameter to account for the microstructure around the to-be-studied amino acids. For cation exchange chromatography, the contact region mainly consists of charged residues (e.g. lysine and aspartate). These residues that take part in the binding process can be referred as the interacting residues or sites. It is unclear what factors determine whether a site is interacting, less-interacting or non-interacting. However, from our results we list three important parameters that might play a major role in determining whether a site is interacting or non-interacting. They are 1) its charge 2) its side chain surface accessibility, mobility and orientation, and 3) its pK_a , which is influenced by a) its microstructure (rHy) and b) its involvement in intramolecular interactions. Here we have demonstrated that all these three factors determine the binding process in cation exchange chromatography. From category II results, we have demonstrated that there is a general trend that is illustrated by a direct correlation between net charge on the protein retention times (Fig. 5). The surface, mobility and orientation play an important role in determining the retention times of variant K124H. Meanwhile, the influence of microstructure and intermolecular bonds has been illustrated by variants S136K, K83H, R76S, and R76D.

On the other hand, all the above three parameters are dependent on the physical location of a particular site (steric effect) on the 3D structure of a protein. To understand the importance of the location, the IEC binding mechanism needs to be revisited. In IEC, which is a dynamic process due to the forced convective flow (mostly outside the pores), attaining a macroscopic equilibrium (in the whole column) is not possible.[18] However, as discussed in the introduction section, the electrostatic partitioning takes place in the diffusion controlled step inside the pores and therefore at the molecular level there might

be a possibility of a microscopic equilibrium.[18] Since diffusion is inherently a random process and due to the molecular internal energy there is a strong possibility that the all surface amino acids are equally probable to interact with the chromatographic media.[16] In addition, due to the heterogeneous stationary phase ligand distribution and an inherently heterogeneous protein (including anisotropic charge) structure, there is a random interaction between the stationary phase and the protein.

When the binding mechanism in cation exchange chromatography is considered in terms individual interacting sites, their position on the 3D structure of the protein seem to play a minor role and the major role is played by the charges on the protein surface and their (anisotropic) location on the protein molecule (electrostatics). Other factors, whose contribution is hard to characterize also affect chromatographic behavior of a protein in IEC. Some of these factors are 1) hydration effects (entropy), 2) hydrophobic forces that are a part of the microenvironment around the surface amino acids, and 3) intermolecular forces such as Vander Waal forces.[19] Another aspect that is not well understood is the change in protein configuration when it is bound to the stationary matrix.

Conclusions

Two categories of protein variants were generated using site-directed protein mutagenesis. Category I variants primarily vary in their only in their microstructure. The protein microstructure was theoretically characterized in terms of relative hydrophilicity. Category II variants primarily vary in their net charge. All seventeen variants, which consist the two categories, had different rRT values indicating that the underlying thermodynamic process also varies mainly due to variation in the protein structure.

For category I variants, there is a direct correlation between the protein microstructure and the rRT values in IEC. There is a direct correlation (R^2= 0.95) between the microstructure of the charged group and the protein's retention time, except for two variants (K83H and K124H) whose histidine residue was involved in intramolecular bonds. These two excluded variants have a non-interaction (K124H) and less-interacting site (K83H). Hence, the variation in rRT values can be attributed to the characterized protein microstructure.

For category II variants, there is a direct correlation (R^2= 0.93) between the change in net charge (-2 to +2 units) and the retention times, except for two variants (R76D and R76S) whose retention times do not follow the correlation. However, there is no clear correlation between the difference in microstructure (rHy) and the rRT values, except for variant S136K, because the major overriding or dominating parameter in this category is the change in the net charge. On the other hand, due to the compromised net negative charge, which in turn is due to the presence of intramolecular interactions, the variants at site 76 eluted much later than the other similar variants (e.g. site 137). Consequently, this site rendered as a less-interacting site. These excluded variants at site 76 provide insight into the relationship between protein structure and function. Therefore, the presence of intramolecular bonds and unfavorable orientation of amino acid side chain, in addition to rHy, affect the chromatographic binding behavior of lysozyme variants. Thus, IEC may be used as an analytical technique to discover subtle structural variations (charge or microstructure) including intramolecular bonds in proteins by measuring variation in the rRT values.

Table-1: Relative surface accessibility (rSA)[4] of both original and replaced amino acid, net total charge around the surface accessible amino acid with 15 Å, relative hydrophilicity values (rHy)[5] and the retention times (RT, including for C-WT) of all seven variants.

Lysozyme variants	rSA (%) (original amino acid)	rSA (%) (histidine residue)	Net micro-environmental charge	rHy	rRT (%)
C-WT	N/A	N/A	N/A	N/A	0
K19H	62	50	+1	0.54	5.0
K43H	26	15	+2	0.27	11.0
R80H	62	74	+1	0.74	2.1
K83H	60	62	+5	0.62	9.8
K124H	48	47	+3	0.52	0.2
K135H	69	63	+5	0.64	0.7
K147H	49	47	+2	0.51	5.0

[4] the percentage of surface accessibility of the residue (side chain) in the folded protein divided by the SA of the residue (side chain) in the unfolded state, of lysozyme variants calculated by using a probe with a radius of 1.93 Å.

[5] Microenvironments with *rHy* values 1) >0.25 are fairly hydrophilic, 2) < 0.25 are hydrophobic and 3) < 0.0 are extremely hydrophobic

Table 2: The change in net charge and relative retention times (rRT) for all ten variants including the two control variants, pWT-135 and pWT-80.

Protein Variant	Change in net protein charge	Relative retention time (rRT) (%)
Control Variant (His-135)	0	0
D127S	+1	14.3
D127K	+2	35.1
S136D	-1	-21.9
S136K	+1	6.8
S136M	-0.15	5.5
R137S	-1	-19.2
R137D	-1	-28.5
Test Variant (His-80)	0	0
R76S	-1	-3.8
R76D	-2	-6.9
E108K	+2	27.3

Table 3. Relative surface accessibility[6] and relative hydrophilicity values (rHy)[7] of all the five sites chosen for single point mutations.

Sites	Relative surface accessibility	Relative hydrophilicity (rHy)
Arg-76	0.60	0.84
Glu-108	0.42	0.55
Asp-127	1.00	1.02
Ser-136[8]	0.10	0.70
Arg-137	0.63	0.78

[6] The relative surface area of each amino acid side was directly obtained from SYBYL program

[7] Microstructure with rHy values 1) >0.25 are fairly hydrophilic, 2) < 0.25 are hydrophobic and 3) < 0.0 are extremely hydrophobic

[8] The aHy value of tyrosine was used to calculate the rHy of Ser-136

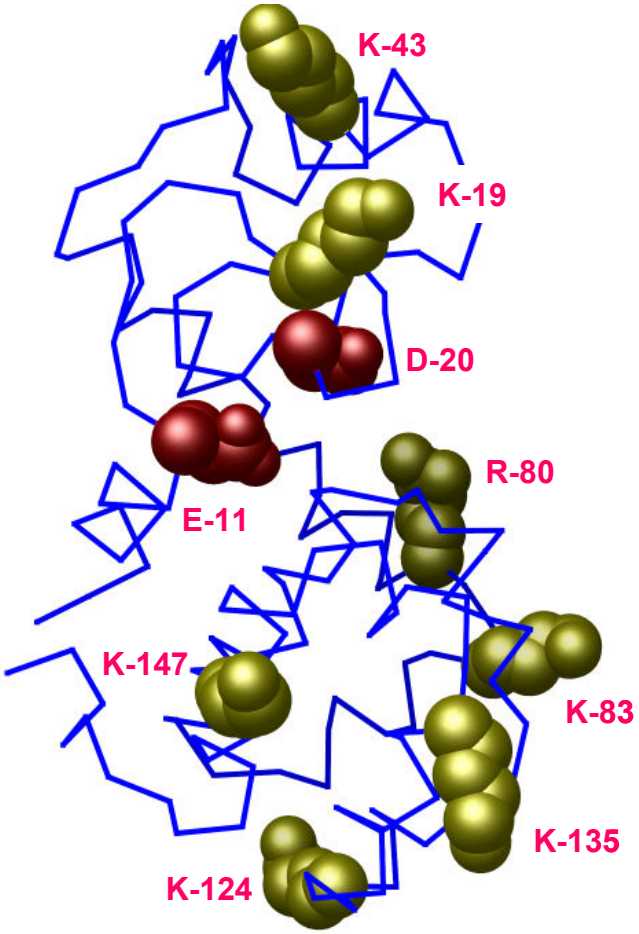
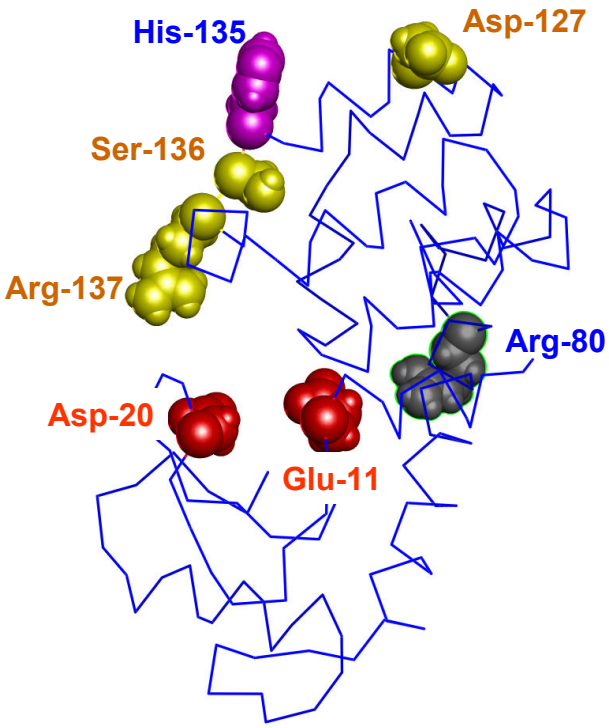
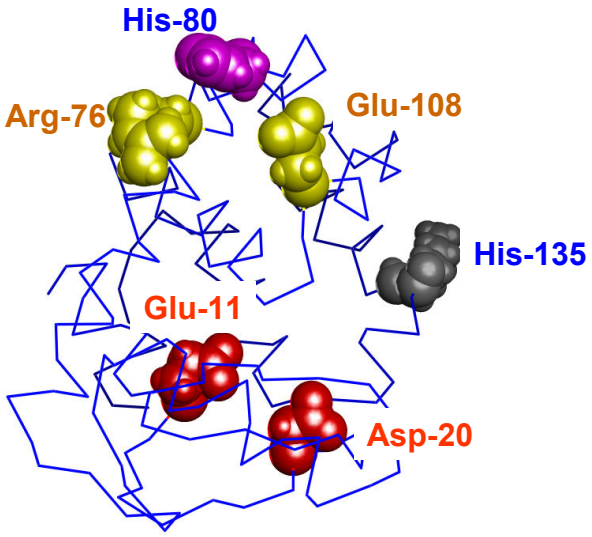


Figure-1: Sites Selected for histidine replacement (amino acid substitutions) on the surface of T4 lysozyme are shown in yellow color. The amino acid residues that are catalytically active in T4 lysozyme (Glu-11 and Asp-20) are shown in red. This figure was generated using UCSF Chimera.



A) The structure of pWT-135



B) The structure of pWT-80

Figure 2 (A and B). The sites shown in red are the residues involved in catalysis. **A)** The three sites (yellow) chosen around the histidine residue, His-135 (magenta), in the control variant. Site -80 is shown in gray. B) The two sites (yellow) chosen around the histidine residue, His-80 (magenta), in the test variant. Site-135 is shown in gray.

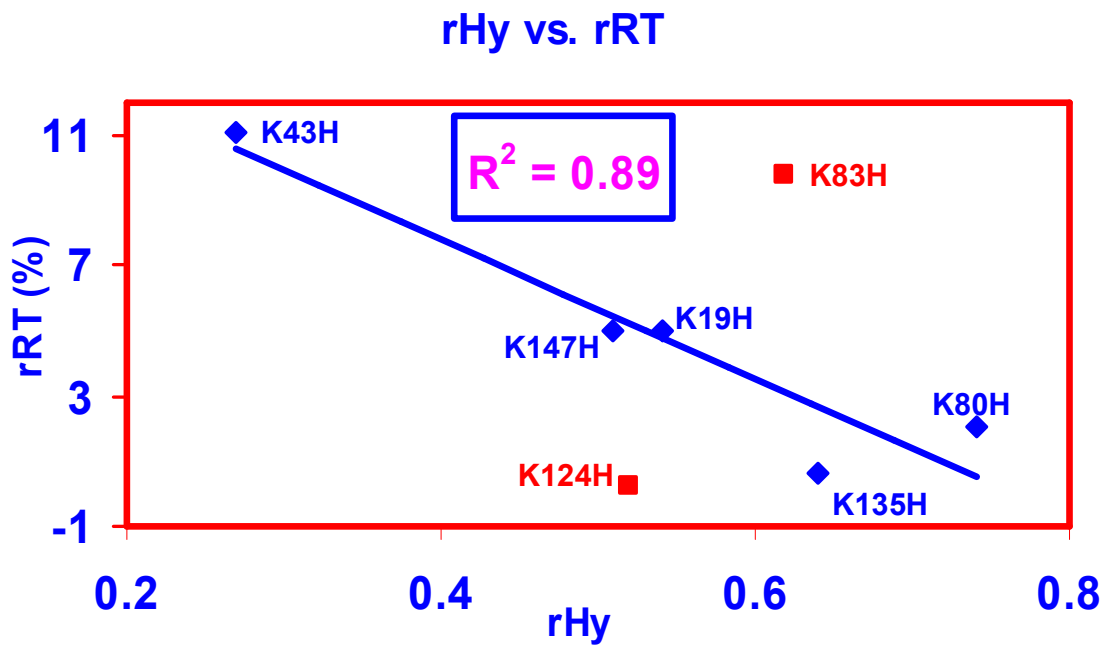


Figure 3. Correlation between the hydrophilic microenvironment (rHy) and the retention time of all seven variants belonging to category-I.

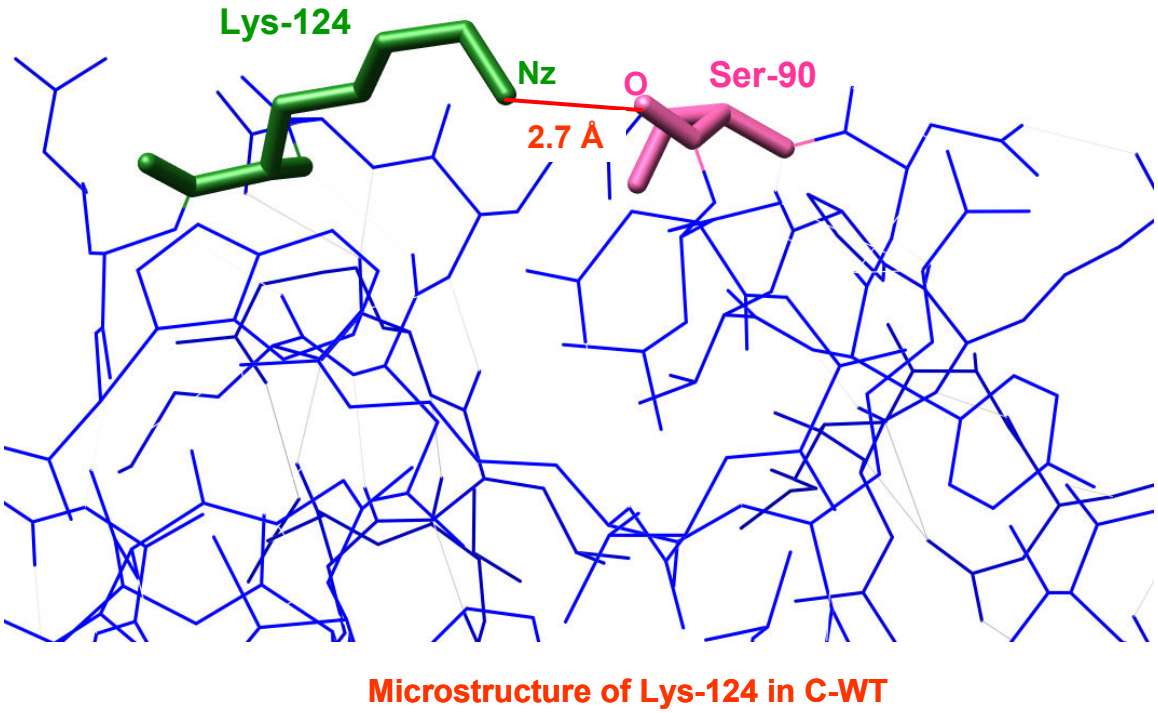


Figure 4. Microstructure of the variant C-WT, wherein the side chain of Lys-124 has involved in a salt bridge with the Ser-90 (oxygen) and is oriented inwards into the protein. Thus there is a high probability that this Site (Lys-124) is a non interacting site in an IEC column.

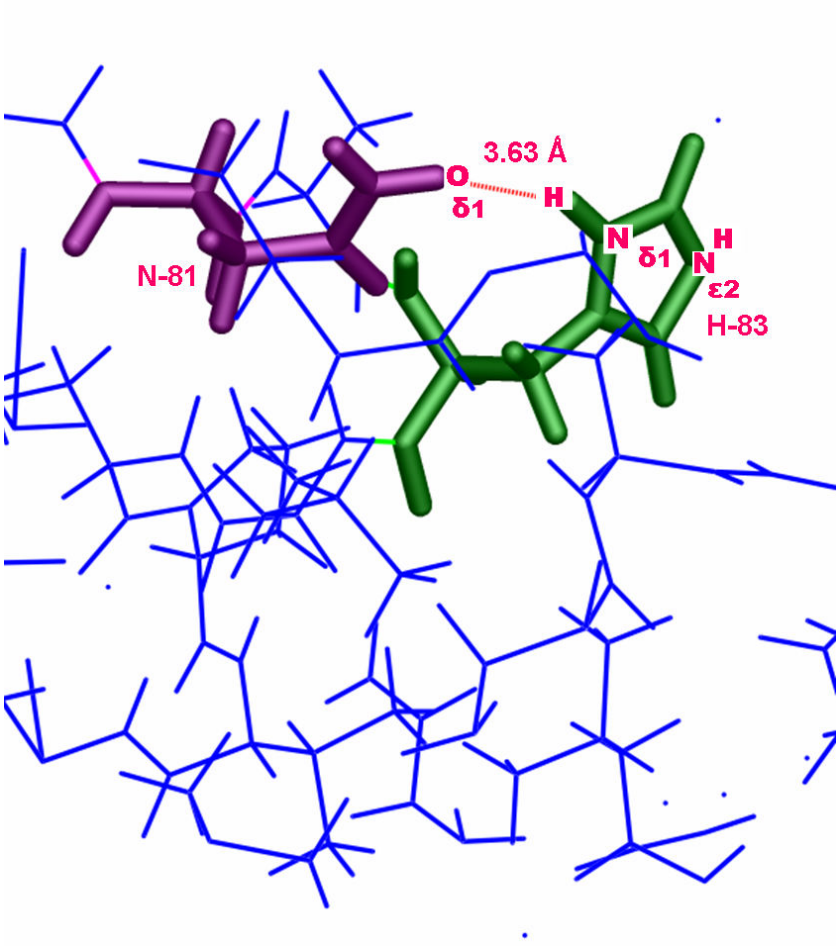


Figure 5: The presence of a weak intramolecular hydrogen bond between Oδ1 oxygen of Asn-81 at distance of 3.63 Å away from the Nδ1 hydrogen of His-83 (doubly protonated form). The figure was generated using UCSF Chimera. This is a sectional view of the molecule around His-83.

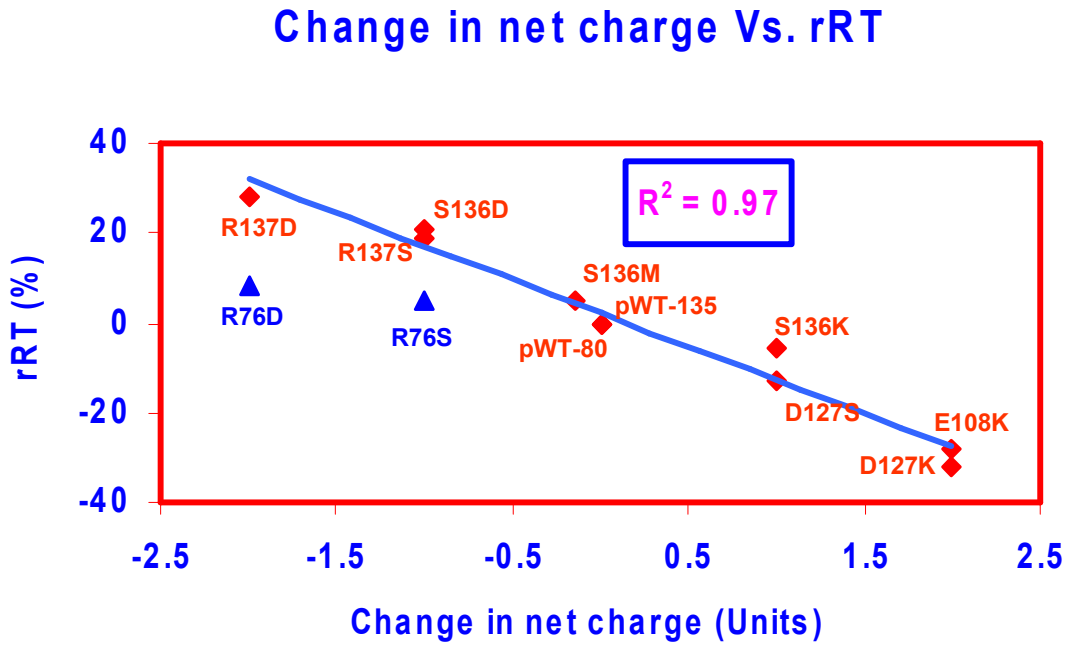


Figure 6. Correlation between the change in net charge and the relative retention time of all seven variants belonging to category-I.

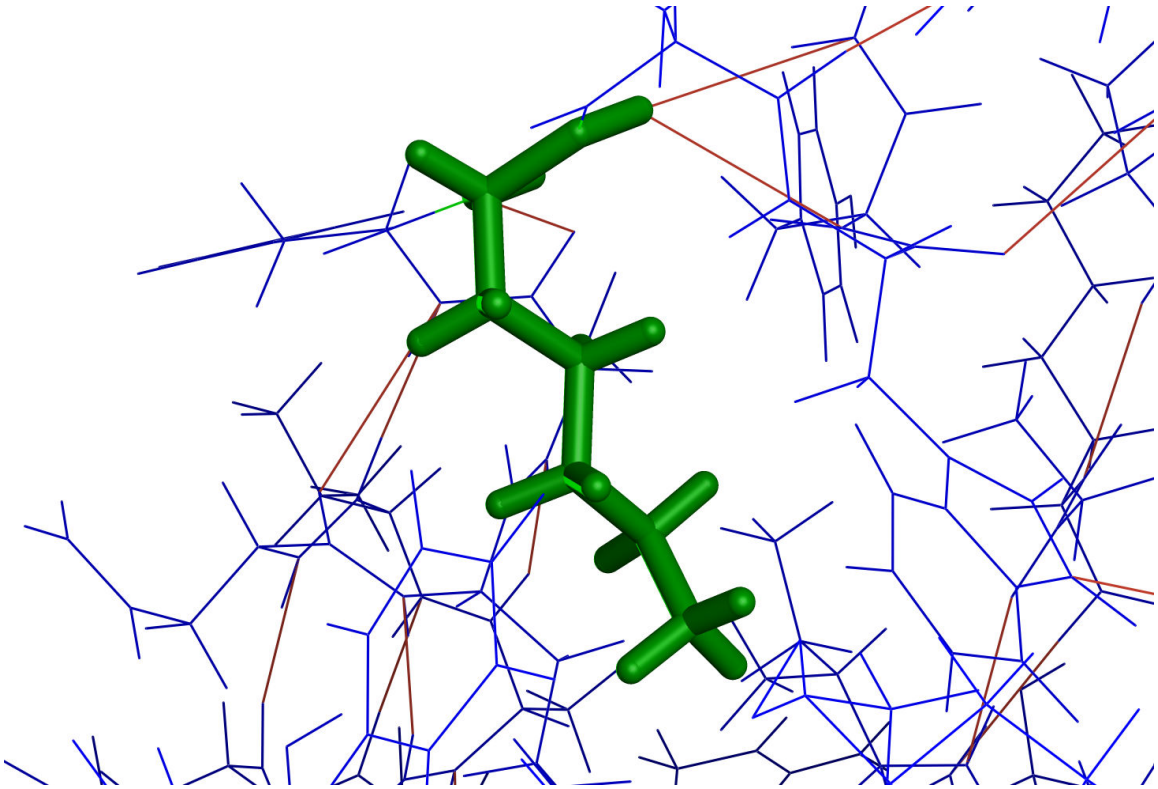
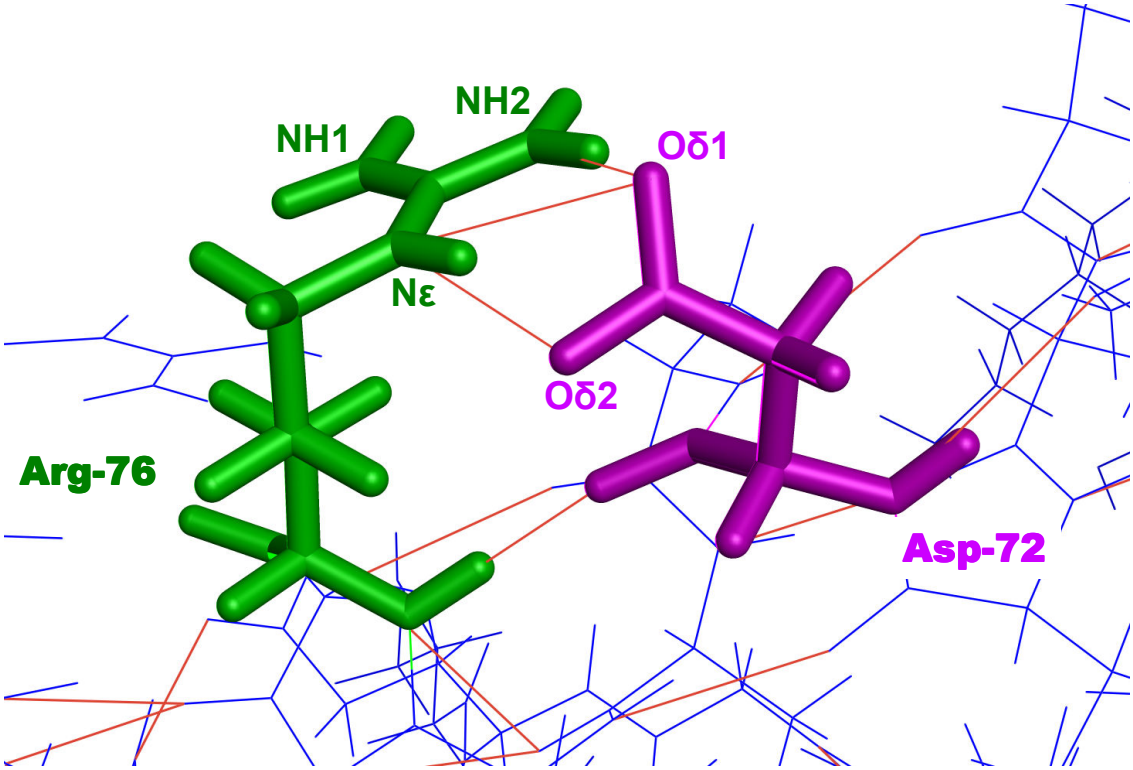
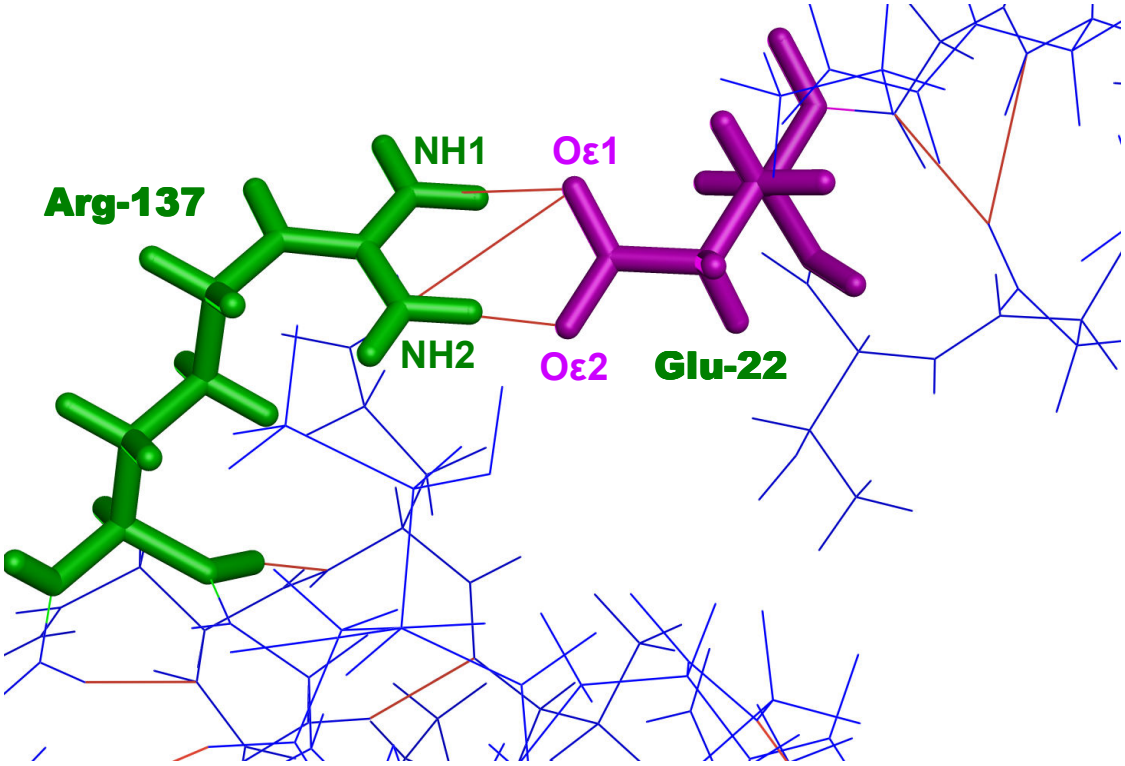


Figure 7. Microstructure of the variant S136K, wherein the side chain of lysine residue has no hydrogen bonds with its adjacent amino acid residues and is oriented inwards into the protein.



A) pWT-80 microstructure around the site (Arg)-76



B) pWT-135 microstructure around the site (Arg)-137

Figure 8 (A and B). The microstructure of A) pWT-80, wherein the side chain of Arg76 is involved in a salt bridge (red lines) with Asp72, and B) pWT-137, wherein the side chain is involved in a salt bridge (red lines) with of Glu22.

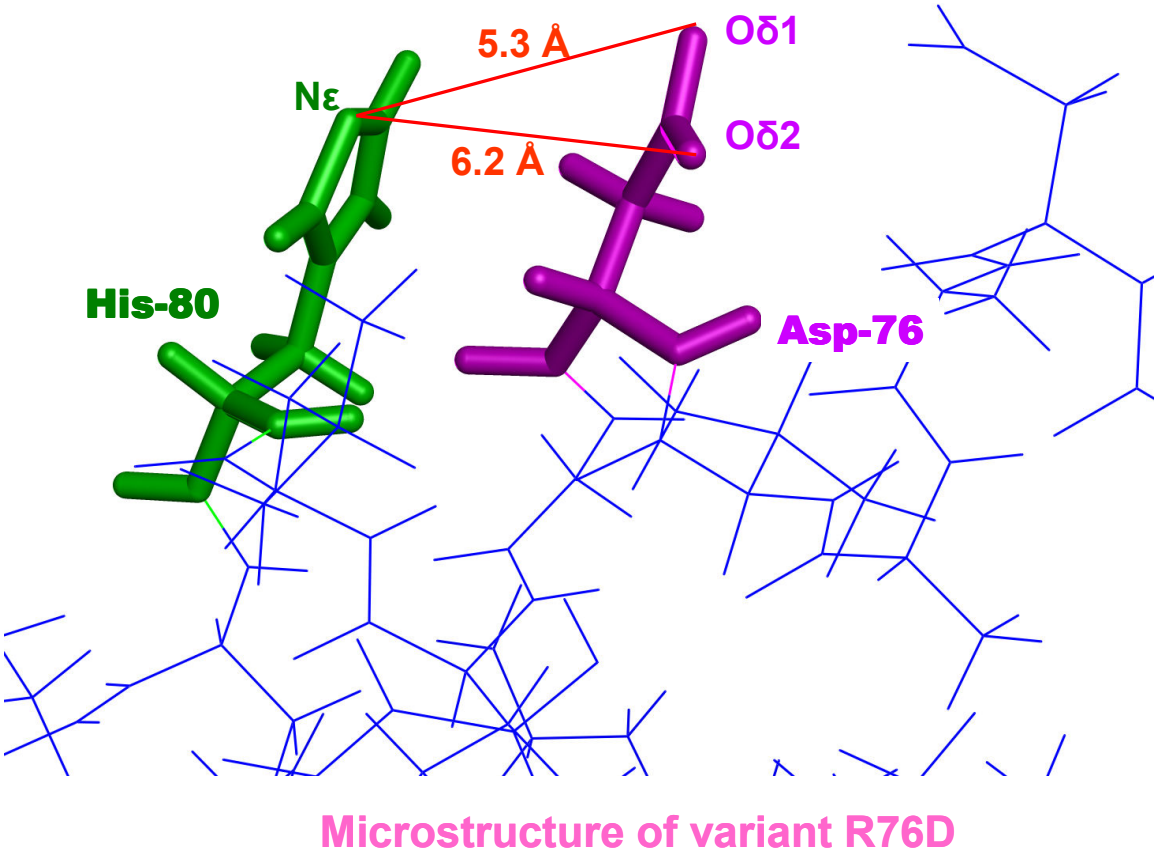


Figure 9. Microstructure of the variant R76D, wherein the side chain of Asp-76 is involved in a (weak) salt bridge with the side chain of His80.

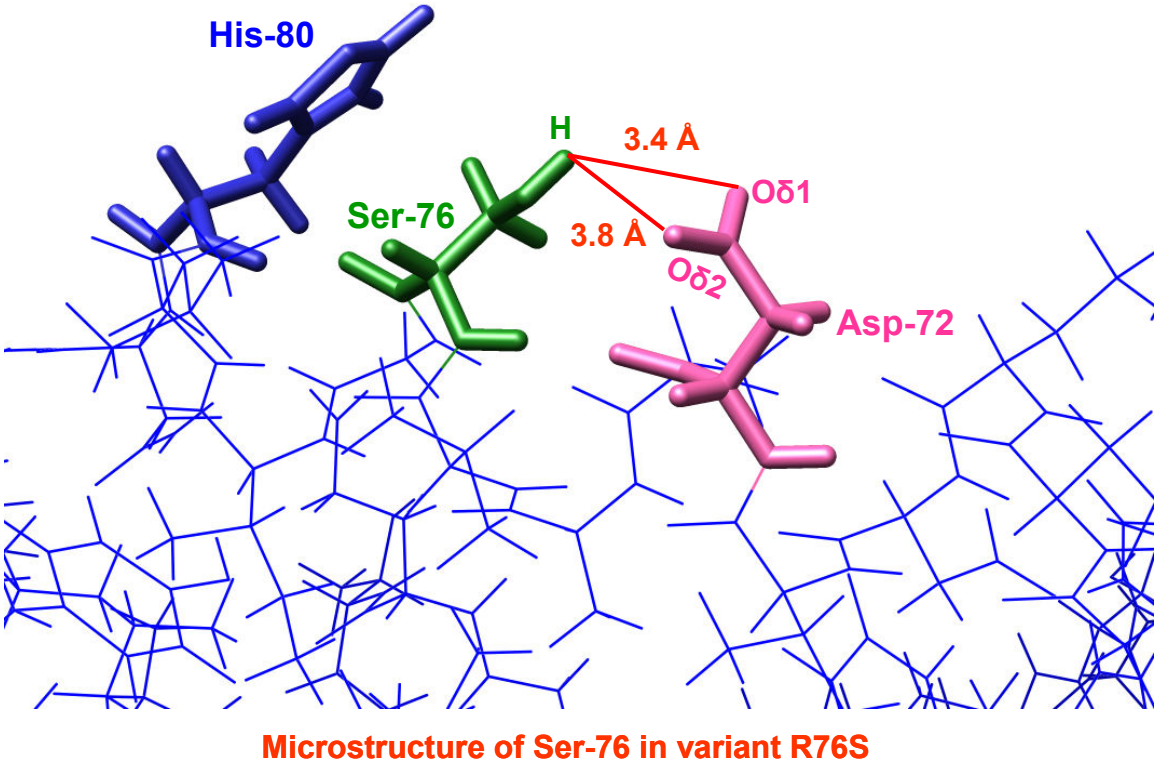


Figure 10. Microstructure of the variant R76S, wherein the side chain of Ser-76 is involved in a (weak) hydrogen bond with the side chain (Oxygen atoms) of Asp-72.

References

(1) Kopaciewicz, W.; Rounds, M. A.; Fausnaugh, J.; Regnier, F. E. *Journal of Chromatography* **1983**, *266*, 3-21.
(2) Zhang, C. M.; Glatz, C. E. *Journal of Chromatography A* **2005**, *1069*, 113-118.
(3) Xu, W. S.; Zhou, H.; Regnier, F. E. *Analytical Chemistry* **2003**, *75*, 1931-1940.
(4) Liapis, A. I.; Grimes, B. A. *Journal of Separation Science* **2005**, *28*, 1909-1926.
(5) Yao, Y.; Lenhoff, A. M. *Analytical Chemistry* **2004**, *76*, 6743-6752.
(6) Wheelwright, S. M. *Protein Purification: Desigh and Scale up of downstream processing*; Wiley-Interscience, 1991.
(7) Carta, G.; Ubiera, A. R.; Pabst, T. M. *Chemical Engineering & Technology* **2005**, *28*, 1252-1264.
(8) Johnson, E. M.; Berk, D. A.; Jain, R. K.; Deen, W. M. *Biophysical Journal* **1995**, *68*, 1561-1568.
(9) Ladisch, M. R. *Bioseparation Engineering: Principles, Practice and Ecionomics*; Wiley-Interscince: New York, 2001.
(10) Hirota, N.; Kumaki, Y.; Narita, T.; Gong, J. P.; Osada, Y. *Journal of Physical Chemistry B* **2000**, *104*, 9898-9903.
(11) Pathange, L. P.; Bevan, D. R.; Larson, T. J.; Zhang, C. M. *Analytical Chemistry* **2006**, *78*, 4443-4449.

(12) Pathange, L. P.; Bevan, D. R.; Zhang, C. M. **2007**.
(13) Mehler, E. L.; Guarnieri, F. *Biophysical Journal* **1999**, *77*, 3-22.
(14) Mehler, E. L. *Journal of Physical Chemistry* **1996**, *100*, 16006-16018.
(15) Regnier, F. E. *Science* **1987**, *238*, 319-323.
(16) Fausnaughpollitt, J.; Thevenon, G.; Janis, L.; Regnier, F. E. *Journal of Chromatography* **1988**, *443*, 221-228.
(17) Chicz, R. M.; Regnier, F. E. *Analytical Chemistry* **1989**, *61*, 2059-2066.
(18) Li, X. P.; Hupp, A. M.; McGuffin, V. L. In *Advances in Chromatography, Vol 45*, 2007; Vol. 45, pp 1-88.
(19) Yao, Y.; Lenhoff, A. M. *Analytical Chemistry* **2005**, *77*, 2157-2165.

Chapter VI: Conclusions and Significance

The two categories of T4 lysozyme variants generated were based on the two structural parameters, 1) predominantly variation in the histidine's surface accessibility (category I), and 2) predominantly variation in the protein net charge (category II). Seven variants (category I) were obtained by replacing a charged amino acid residue at different locations (different surface accessibility) by histidine on the lysozyme. Ten variants (category II) were obtained by replacing charged and neutral amino acid residues at different locations around the histidine residue on the lysozyme. All seventeen variants have different binding strength and rRT values indicating the variation in the protein structure or specifically protein microstructure.

Category I: The binding strength in IMAC was shown to have a direct correlation (R^2= 0.76) with the relative surface accessibility (rSA) of a histidine residue, except for variants K83H and K147H, whose histidines were shown to be involved in intramolecular hydrogen bonds. In ion exchange chromatography, there was a strong correlation (R^2= 0.95) between the protein's microstructure, which predominantly consists the surface accessibility of the charged group, and the protein's retention time, except for two variants K83H and K124H, wherein the presence of intramolecular bonds renders the two sites 83 (in K83H) and 124 (in K124H) as non-interacting and less interacting sites, respectively.

Category II: The direct correlation (R^2 = 0.96) between $\Delta\Delta G_E$ and $\Delta\Delta G_B$ indicates that it is possible to predict the electrostatic effect due to mutations ($\Delta\Delta G_{Elec}$), when no other structural effects are involved. The $\Delta\Delta G_B$ values of the two variants (S136K and D127K) differ from the other similarly charged variants and their own

$\Delta\Delta G_E$ values because of the additional site-specific microstructural effects ($\Delta\Delta G_{Micro}$) that contribute to their $\Delta\Delta G_B$ values. In ion exchange chromatography, there is a direct correlation (R^2= 0.93) between the change in net charge (-2 to +2 units) and the retention times, except for two variants (R76D and R76S). The retention times of these two variants do not follow the correlation due to the presence of intramolecular interactions which compromise net negative charge.

Significance: The results obtained for category I variants on IMAC and IEC indicate that in addition to amino acid surface accessibility, the mobility of amino acid chain is important. From IMAC results it was demonstrated that it is possible to evaluate the strength of a hydrogen bond. The results obtained for category II variants on both IMAC and IEC indicate that there is a broad correlation between the net charge of a protein and the chromatographic parameters (binding strength and rRT value). However, the presence of outlier variants illustrates the significance of microstructure. Thus, a protein's chromatographic binding behavior (function) is dependent on a protein's microstructure, which includes the surface accessibility of a particular amino acid and the orientation and mobility (due to intramolecular bonds) of an amino acid side chain. Thus, based on the binding strength and the rRT values of the variants in both categories, IMAC and IEC may be used to determine the histidine-related protein surface structure including the surface accessibility and microstructure of histidine residues and the presence of hydrogen bonds, which generally requires an X-ray or NMR structure.

In addition, due to the specific nature of the interaction between the imidizole nitrogen and the metal ions in IMAC, it is possible to quantifying the components of $\Delta\Delta G_B$. The contribution of the individual factors (e.g. electrostatic, microstructural) to

$\Delta\Delta G_B$ was dissected by comparing the experimental $\Delta\Delta G_B$ values with the other variants with identical charge perturbations. The contribution of the relevant $\Delta\Delta G_{Micro}$ factor to $\Delta\Delta G_B$ value was identified and quantified. Dissection of $\Delta\Delta G_B$ will help in not only understanding the structure-function relationship of a protein but also engineering or designing tailor-made proteins for various applications.

Appendices

Appendix A: CD spectra of category I

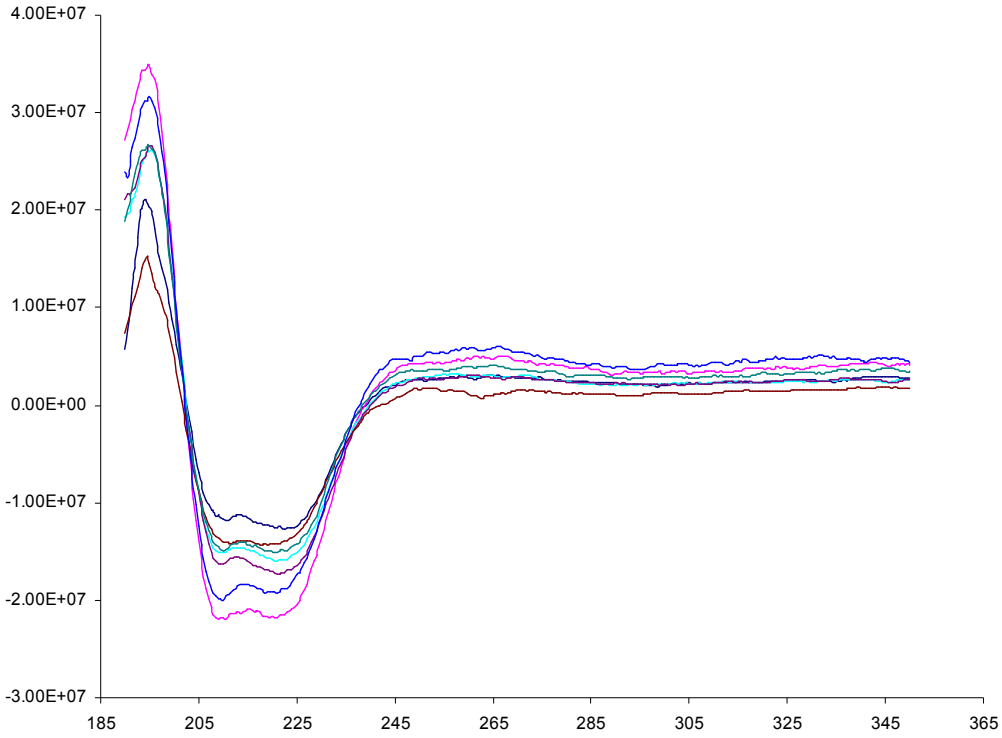


Figure 1. CD spectra of all seven lysozyme variants (of category 1) and wild type.

Appendix B: Raw protein sequencing data obtained electrospray ionization mass spectrometry

Sample	Reference	Protein GI	Peptides Identified by SEQUEST	Notes
TP39B1_tuned	TA- WT	230100	120- 125 MLQQKR 36- 43 SPSLNAAK 53- 65 NTNGVITKDEAEK 77- 83 GILRNAK 138- 145 WYNQTPNR 136- 145 SRWYNQTPNR 148- 154 RVITTFR 44- 52 SELDKAIGR 155- 162 TGTWDAYK 149- 154 VITTFR 9- 16 IDEGLRLK 44- 60 SELDKAIGRNTNGVITK 125- 135 RWDEAAVNLAK 126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR 155- 164 TGTWDAYKNL 148- 162 RVITTFRTGTWDAYK 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 149- 164 VITTFRTGTWDAYKNL 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR 61- 80 DEAEKLFNQDVDAAVRGILR 97- 119 AALINMVFQMGETGVAGFTNSLR	Protein Coverage: 149/164 = 90.9% by amino acid count, 16870/18583 = 90.8% by mass
TP39B2_1	K19H	230101	125-135 RWDEAAVNLAK 126-135 WDEAAVNLAK	Sample was run four times. Unable to locate peptide of interest either with or without the modification. Could also ID as WT, R80H, R83H Protein Coverage of best run: 71/164 = 43.3% by amino acid count, 8123/18592 = 43.7% by mass Peptide fragment resulting from trypsin digest with amino acid substitution of interest is too large to be detected. Could try a GluC digest as an alternative.
TP39B2_2			155-164 TGTWDAYKNL	
TP39B2_3			125-135 RWDEAAVNLAK 126-135 WDEAAVNLAK 155-164 TGTWDAYKNL	
TP39B2_tuned2			36- 43 SPSLNAAK 138- 145 WYNQTPNR 136- 145 SRWYNQTPNR 44- 52 SELDKAIGR 148- 154 RVITTFR 149- 154 VITTFR 125- 135 RWDEAAVNLAK 126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR 155- 164 TGTWDAYKNL 61- 76 DEAEKLFNQDVDAAVR	

			149- 164 VITTFRTGTWDAYKNL	
TP39B3	R80H	230102	36- 43 SPSLNAAK 77- 83 GIL**H**NAK 138- 145 WYNQTPNR 136- 145 SRWYNQTPNR 155- 162 TGTWDAYK 149- 154 VITTFR 148- 154 RVITTFR 9- 16 IDEGLRLK 44- 60 SELDKAIGRNTNGVITK 125- 135 RWDEAAVNLAK 126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR 84- 96 LKPVYDSLDAVRR 84- 95 LKPVYDSLDAVR 155- 164 TGTWDAYKNL 61- 76 DEAEKLFNQDVDAAVR 149- 164 VITTFRTGTWDAYKNL 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 43 DTEGYYTIGIGHLLTKSPSLNAAK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR	Peptide with modification identified by SEQUEST: 77-83 GIL**H**NAK However, I am not 100% confident as it is a fairly small peptide. It is probably there. Many of the fragment ions were present in the MS/MS spectrum but there were some unidentified ions of relative medium abundance as well. Protein Coverage: 134/164 = 81.7% by amino acid count, 15326/18564 = 82.6% by mass
TP39B4	K83H	230103	36-43 SPSLNAAK 138-145 WYNQTPNR 148-154 RVITTFR 44-52 SELDKAIGR 149-154 VITTFR 9-16 IDEGLRLK 125-135 RWDEAAVNLAK 126-135 WDEAAVNLAK 66-76 LFNQDVDAAVR 155-164 TGTWDAYKNL 1-8 MNIFEMLR	Able to positively ID modified peptide manually: NA**H**LKPVYDSLDAVR
TP39B4_tuned3			53- 60 NTNGVITK 36- 43 SPSLNAAK 138- 148 WYNQTPNRAKR 53- 65 NTNGVITKDEAEK 9- 14 IDEGLR 138- 147 WYNQTPNRAK 136- 147 SRWYNQTPNRAK 146- 154 AKRVITTFR 136- 145 SRWYNQTPNR 138- 145 WYNQTPNR 148- 154 RVITTFR 44- 52 SELDKAIGR 149- 154 VITTFR 155- 162 TGTWDAYK 9- 16 IDEGLRLK 44- 60 SELDKAIGRNTNGVITK 125- 135 RWDEAAVNLAK 81- 95 NA**H**LKPVYDSLDAVR	Peptide with modification identified by SEQUEST: 81- 95 NA**H**LKPVYDSLDAVR Protein Coverage: 152/164 = 92.7% by amino acid count, 17108/18580 = 92.1% by mass

			126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR 155- 164 TGTWDAYKNL 149- 164 VITTFRTGTWDAYKNL 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR 96- 119 RAALINMVFQMGETGVAGFTNSLR	
TP39B5	K124H	230104	155- 162 TGTWDAYK 148- 154 RVITTFR 149- 154 VITTFR 126- 135 WDEAAVNLAK 84- 96 LKPVYDSLDAVRR 66- 76 LFNQDVDAAVR 86- 96 PVYDSLDAVRR 84- 95 LKPVYDSLDAVR 155- 164 TGTWDAYKNL 120- 135 MLQQHRWDEAAVNLAK 148- 162 RVITTFRTGTWDAYK 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 149- 164 VITTFRTGTWDAYKNL 20- 35 DTEGYYTIGIGHLLTK	Protein Coverage: 86/164 = 52.4% by amino acid count, 9823/18592 = 52.8% by mass
TP39B5_tuned3			53- 60 NTNGVITK 138- 147 WYNQTPNRAK 138- 145 WYNQTPNR 136- 145 SRWYNQTPNR 146- 154 AKRVITTFR 148- 154 RVITTFR 44- 52 SELDKAIGR 149- 154 VITTFR 66- 76 LFNQDVDAAVR 120- 135 MLQQ**H**RWDEAAVNLAK 155- 164 TGTWDAYKNL 61- 76 DEAEKLFNQDVDAAVR 149- 164 VITTFRTGTWDAYKNL 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 1- 8 MNIFEMLR	Protein Coverage: 86/164 = 52.4% by amino acid count, 10048/18580 = 54.1% by mass Peptide with modification identified by SEQUEST: 120- 135 MLQQ**H**RWDEAAVNLAK
TP39B6	K135 (K13H)	230106	138-145 WYNQTPNR 148-154 RVITTFR 44-52 SELDKAIGR 149-154 VITTFR 126-137 WDEAAVNLA**H**SR 66-76 LFNQDVDAAVR 155-164 TGTWDAYKNL 20-43 DTEGYYTIGIGHLLTKSPSLNAAK 20-35 DTEGYYTIGIGHLLTK 1-8 MNIFEMLR	Peptide with modification identified by SEQUEST: WDEAAVNLA**H**SR
TP39B7	K147H	230107	125- 135 RWDEAAVNLAK 126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR	Protein Coverage: 103/164 = 62.8% by amino acid count, 11677/18592 =

			84- 96 LKPVYDSLDAVRR 155- 164 TGTWDAYKNL 84- 95 LKPVYDSLDAVR 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 149- 164 VITTFRTGTWDAYKNL 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 20- 43 DTEGYYTIGIGHLLTKSPSLNAAK 1- 8 MNIFEMLR 61- 80 DEAEKLFNQDVDAAVRGILR	62.8% by mass Peptide fragment resulting from trypsin digest with amino acid substitution of interest is too small to be detected and identified with any confidence. Could try a GluC digest as an alternative.
			No contaminant proteins from E. coli were found in any of the samples.	

Sample	Reference	Protein GI	Peptides Identified	Notes
TW17C1	K43H	230108	36- 48 SPSLNAA**H**SELDK 149- 154 VITTFR 155- 162 TGTWDAYK 36- 52 SPSLNAA**H**SELDKAIGR 125- 135 RWDEAAVNLAK 84- 95 LKPVYDSLDAVR 66- 76 LFNQDVDAAVR 61- 76 DEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR 120- 124 MLQQK 97- 119 AALINMVFQMGETGVAGFTNSLR	Peptides with modification identified SPSLNAA**H**SELDK SPSLNAA**H**SELDKAIGR
TW17C2	S136D	230109	120- 124 MLQQK 44- 48 SELDK 36- 43 SPSLNAAK 9- 14 IDEGLR 148- 154 RVITTFR 149- 154 VITTFR 44- 52 SELDKAIGR 155- 162 TGTWDAYK 125- 137 RWDEAAVNLAH**D**R 126- 137 WDEAAVNLAH**D**R 84- 95 LKPVYDSLDAVR 66- 76 LFNQDVDAAVR 86- 95 PVYDSLDAVR 61- 76 DEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR 97- 119 AALINMVFQMGETGVAGFTNSLR	Peptides with modification identified WDEAAVNLAH**D**R RWDEAAVNLAH**D**R
TW17C3	S136K	230110	??? ALGR 44- 48 SELDK 120- 124 MLQQK	Peptides with modification identified

			49- 52 AIGR 36- 43 SPSLNAAK 9- 14 IDEGLR 148- 154 RVITTFR 138- 145 WYNQTPNR 149- 154 VITTFR 155- 162 TGTWDAYK 126- 137 WDEAAVNLAH**K**R 126- 136 WDEAAVNLAH**K** 86- 96 PVYDSLDAVRR 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 61- 76 DEAEKLFNQDVDAAVR 1- 8 MNIFEMLR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 97- 119 AALINMVFQMGETGVAGFTNSLR	WDEAAVNLAH**K** WDEAAVNLAH**K**R

Sample	Reference	Peptides Identified	Notes
T63C1	76D	position sequence 36- 43 SPSLNAAK 9- 14 IDEGLR 149- 154 VITTFR 155- 162 TGTWDAYK 125- 135 RWDEAAVNLAK 84- 96 LKPVYDSLDAVRR 126- 135 WDEAAVNLAK 84- 95 LKPVYDSLDAVR 86- 96 PVYDSLDAVRR 66- 83 LFNQDVDAAV**D**GILH**N**AK 17- 35 IYKDTEGYYTIGIGHLLTK 61- 83 DEAEKLFNQDVDAAV**D**GILH**N**AK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR 148- 154 RVITTFR	Peptides with modification identified by Sequest: LFNQDVDAAV**D**GILH**N**AK DEAEKLFNQDVDAAV**D**GILH**N**AK **Protein Coverage:** 103/164 = **62.8%** by amino acid count, 11708/18511 = **63.2%** by mass There were no peptides found without the modifications.
TW63C2	108K	position sequence 155- 162 TGTWDAYK 149- 154 VITTFR 84- 96 LKPVYDSLDAVRR 125- 135 RWDEAAVNLAK 109- 119 **TGVAGFTNSLR** 126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK	I was unable to identify the sequence GIL**H**NAK with the amino acid substitution. However, it is a fairly polar peptide and may not comply with RP-HPLC. I was also unable to find the resulting sequences without the substitution. These would be difficult to see as they would be very short. They were not seen in other sample runs either. The sequence AALINMVFQMG**K** was also not identified directly. However, the sequence TGVAGFTNSLR was identified. This is significant because it indicates that either a Lysine or Arginine directly precedes it because of the Tryptic digest. This indicates that this particular

		20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR	amino acid substitution was most likely successful. **Protein Coverage:** 100/164 = **61.0%** by amino acid count, 11338/18551 = **61.1%** by mass
TW63C3	127K	position sequence 138- 145 WYNQTPNR 148- 154 RVITTFR 44- 52 SELDKAIGR 149- 154 VITTFR 155- 162 TGTWDAYK 126- 137 W**K**EAAVNLA**H**SR 84- 96 LKPVYDSLDAVRR 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 36- 52 SPSLNAAKSELDKAIGR 155- 164 TGTWDAYKNL 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR	Peptide with modifications identified by Sequest: W**K**EAAVNLA**H**SR **Protein Coverage:** 115/164 = **70.1%** by amino acid count, 13120/18593 = **70.6%** by mass There were no peptides found without the modifications.
TW63C4	127S	position sequence 9- 14 IDEGLR 138- 145 WYNQTPNR 149- 154 VITTFR 44- 52 SELDKAIGR 155- 162 TGTWDAYK 36- 48 SPSLNAAKSELDK 125- 137 RW**S**EAAVNLA**H**SR 126- 137 W**S**EAAVNLA**H**SR 84- 96 LKPVYDSLDAVRR 86- 96 PVYDSLDAVRR 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 81- 95 NAKLKPVYDSLDAVR 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 148- 154 RVITTFR 1- 8 MNIFEMLR	Peptides with modifications identified by Sequest: W**S**EAAVNLA**H**SR RW**S**EAAVNLA**H**SR **Protein Coverage:** 126/164 = **76.8%** by amino acid count, 14409/18552 = **77.7%** by mass There were no peptides found without the modifications.
TW63C5	136M	position sequence 36- 43 SPSLNAAK 9- 14 IDEGLR 53- 65 NTNGVITKDEAEK 138- 145 WYNQTPNR 148- 154 RVITTFR	Peptides with modifications identified by Sequest: RWDEAAVNLA**HM**R WDEAAVNLA**HM**R

		149- 154 VITTFR 44- 52 SELDKAIGR 155- 162 TGTWDAYK 84- 96 LKPVYDSLDAVRR 86- 96 PVYDSLDAVRR 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 125- 137 RWDEAAVNLA**HM**R 126- 137 WDEAAVNLA**HM**R 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR	**Protein Coverage:** 123/164 = **75.0%** by amino acid count, 14168/18624 = **76.1%** by mass There were no peptides found without the modifications.
TW63C6	137D	position sequence --- --- 36- 43 SPSLNAAK 53- 65 NTNGVITKDEAEK 9- 14 IDEGLR 149- 154 VITTFR 155- 162 TGTWDAYK 84- 96 LKPVYDSLDAVRR 86- 95 PVYDSLDAVR 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 86- 96 PVYDSLDAVRR 61- 76 DEAEKLFNQDVDAAVR 125- 145 RWDEAAVNLA**H**S**D**WYNQTPNR 126- 145 WDEAAVNLA**H**S**D**WYNQTPNR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR 148- 154 RVITTFR	Peptides with modifications identified by Sequest: RWDEAAVNLA**H**S**D**WYNQTPNR WDEAAVNLA**H**S**D**WYNQTPNR **Protein Coverage:** 114/164 = **69.5%** by amino acid count, 13113/18539 = **70.7%** by mass There were no peptides found without the modifications.
TW63C7	139HA	position sequence --- --- 138- 145 W**A**NQTPNR 36- 43 SPSLNAAK 9- 14 IDEGLR 44- 52 SELDKAIGR 149- 154 VITTFR 155- 162 TGTWDAYK 9- 16 IDEGLRLK	I was unable to identify the sequence AHR with the amino acid substitution. I was also unable to identify the sequence without the substitution. Peptide with modification identified by Sequest: W**A**NQTPNR

		36- 48 SPSLNAAKSELDK 84- 96 LKPVYDSLDAVRR 86- 96 PVYDSLDAVRR 81- 95 NAKLKPVYDSLDAVR 126- 135 WDEAAVNLAK 66- 76 LFNQDVDAAVR 84- 95 LKPVYDSLDAVR 155- 164 TGTWDAYKNL 36- 52 SPSLNAAKSELDKAIGR 61- 76 DEAEKLFNQDVDAAVR 53- 76 NTNGVITKDEAEKLFNQDVDAAVR 17- 35 IYKDTEGYYTIGIGHLLTK 20- 35 DTEGYYTIGIGHLLTK 1- 8 MNIFEMLR ??? GLIR	**Protein Coverage:** 126/164 = **76.8%** by amino acid count, 14249/18488 = **77.1%** by mass There were no peptides found without the modifications.

Note: All the DNA sequences are recorded in the Lab note books

Appendix C: Determination of immobilized copper on IMAC column

Virginia Tech Soil Testing Laboratory
143 Smyth Hall; Blacksburg, Virginia 24061
(540) 231-9806

Results For:	Laskmi Pathange (BSE)	Date Rec'd:	21-Jul-05
Job #:	3414 (water, trace Na Azide)	Date Complete:	10-Aug-05
		ISR #:	N003521

Data reported as mg/L in solution. The "<" indicates concentrations less than the instrument detection limit.

Analyte	Cu	mM	given Liters	m mol	u mol
Wavelength (nm)	**324.754**				
Detection Limit	**0.01**				
Cu original 1	3,467	54.5589	0.00275	0.150037	150.03698
Cu original 2	3,402	53.53602	0.0025	0.1338401	133.84005
New Column 1	492	7.742423	0.0031	0.0240015	24.001511
New Column 2	475	7.4749	0.003	0.0224247	22.4247
Old Column 1	366	5.759607	0.0035	0.0201586	20.158625
Old Column 2	308	4.846883	0.0045	0.021811	21.810972

Virginia Tech Soil Testing Laboratory
143 Smyth Hall; Blacksburg, Virginia 24061
(540) 231-9806

Results For: Laskmi Pathange (BSE) Date Rec'd: 25-Sep-06
Job #: 3593 (water / trace NaAzide) Date Complete: 26-Sep-06
ISR #: N006651

Data reported as mg/L in solution. The "<" indicates concentrations less than the instrument detection limit.

Analyte	**Cu**
Wavelength (nm)	**324.754**
Detection Limit	**0.009**
Sample-1	303.5
Sample-2	285.8
Sample-3	340.3
Sample-4	302.3
Sample-5	3,385

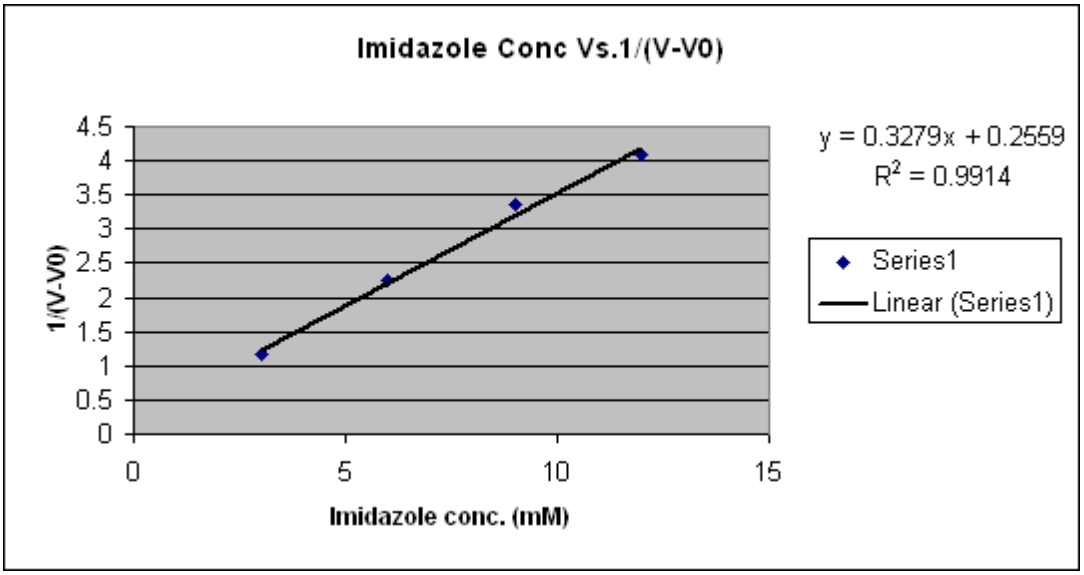
		mM	given ml	L	mM	uM
Sample-1	303.5	4.77607	4.99	0.00499	0.02383	23.8326
Sample-2	285.8	4.49753	4.975	0.004975	0.02238	22.3752
Sample-3	340.3	5.35518	4.22	0.00422	0.0226	22.5988
Sample-4	302.3	4.75718	4.48	0.00448	0.02131	21.3122
standard	**3,385**	53.2685	2.63	0.00263	0.1401	140.096

Appendix-D: Sample calculation using the R137S variant (raw elution) data to obtain the intercept value

Imidazole conc.(I)	run 1(min)	run 2 (min)	run (3)	Average
3 mM	3.264	3.256	3.283	3.267667
6 mM	2.447	2.432	2.442	2.440333
9 mM	2.153	2.149	2.146	2.149333
12 mM	2.04	2.052	2.041	2.044333

V=time*flow rate(0.5 ml/min)	V-V0*		1/(V-V0)
1.633833333	0.85683333	3	1.167088
1.220166667	0.44316667	6	2.256487
1.074666667	0.29766667	9	3.359462
1.022166667	0.24516667	12	4.078858

* Here V0 = 0.777



Note: Calculations for binding strength value is listed in Appendix G

Appendix E: Investigation of the use of molecular modeling to guide site-directed protein mutagenesis

Abstract

Site-directed protein mutagenesis aids in protein purification and the study of protein function. However, the ultimate question one faces before selecting the sites for mutagenesis is whether the side chains of the new amino acids would be able to participate in the desired interaction without altering the protein's tertiary structure. In this report, sites on a model protein surface, T4 lysozyme, were selected for sited-direct mutagenesis using calculated surface accessible area as the selection criterion by molecular modeling and aided by energy minimization determining the possibility of the side chain of the new amino acid in participating in inter-molecular interactions, and the effectiveness of this selection method was evaluated by experimentally quantifying the interaction between the protein variants and immobilized metal (Cu^{2+}) affinity chromatography (IMAC). Our results indicate that the surface accessible area can be used as a base criterion for site-selection during site-directed mutagenesis, but the use of the energy minimization modeling the potential interaction of protein and metal ion should be used with caution.

Keywords: surface accessibility: modeling: IMAC: energy minimization: T4 lysozyme: protein engineering

Introduction

Protein purification is a costly practice and often is the determining factor whether a therapeutic protein can be brought from a research laboratory to the consumer market.[1] To enhance the purification of a target protein from host cell proteins, it is a common practice to modify the protein's binding property to a particular type of separation medium by either fusion technology or site-directed mutagenesis. With fusion technology, DNA encoding an additional polypeptide or another protein is fused to the gene of a target protein.[2,3]. The expressed protein may be purified by techniques utilizing the special properties of the polypeptide and the extra protein.[4,5,6] Currently, it is estimated that more than 50% of the recombinant proteins are purified using histidine fusion tags.[7] Since a fusion tag is simply attached to either the N or C terminal of a protein, no detailed structural information of the target protein is needed. However, the choice of an appropriate fusion tag depends on the factors including host organism, potential adverse effects of the fusion on the target protein's function, and the method of removal. The removal process after protein purification especially complicates a purification process and can potentially offset the advantage gained through the fusion in purification. In addition, some fusions suffer unwanted cleavage by host cell proteases and their stability is unpredictable in certain cases.[8,9]

On the other hand, site-directed mutagenesis (point mutation) modifies a target protein by replacing selected surface amino acids with amino acids with special properties. Point mutations are desirable in protein purification process, as they rarely alter the protein structure or function, dramatically,[10,11] and in many cases these point

mutations induce minor changes restricted to the side-chain conformations of the amino acids surrounding the altered amino acid.[10,12] However, this technique is not used as widely as fusion technology to facilitate protein purification, because detailed structural information of a target protein is needed in order to avoid replacing amino acids that are vital to its bioactivity and there is the uncertainty that the side chain of the replacing amino acid may not behave as expected. Nevertheless, in contrast to fusion technology, the point-mutated protein is usually as stable as the native protein and possesses the same biological function, and no extra purification processes are needed.

Point mutation is a valuable tool for enhanced protein purification, especially when certain fusions are prone to be degraded in certain expression hosts. Zhang and Glatz showed that polyarginine fusions are degraded in canola extract while point mutated T4 lysozymes with the same net charges are stable.[8] By replacing one surface lysine with glutamic acid, the purification of T4 lysozyme from canola extract can be greatly enhanced when cation exchange chromatography is used in separation. Chicz and Regnier demonstrated that by replacing Gly 166 with His on subtilisin, the protein's retention on immobilized metal affinity chromatography (IMAC) was significantly increased.[13] Furthermore, Todd et al. showed that yeast expressed point mutated iso-1-cytochrome *c* maintained its structural integrity and conformation as the native protein when amino acids 4 and 8 were replaced with histidines, and the apparent binding affinity of the proteins containing multiple accessible histidines with immobilized copper ions in IMAC can be increased as much as a factor of 1000.[14] These show that point-mutation can be an effective way to modify a protein to facilitate its purification. However, so far, there still is no a reliable way for selecting the point-mutation sites.

Surface accessibility of amino acids has been used as a criterion in selecting point-mutation sites,[Error! Bookmark not defined.,15] but Berna et al. showed that even with adequate surface accessibility (SA), an amino acid is not guaranteed to be able to interact with a separation medium.[16] However, if the side chain of an amino acid is somewhat buried in the protein structure but is not engaged in any hydrogen bonding and is thus flexible, the flexibility may allow the amino acid to reorient itself to interact with other ligands. Clearly, not only surface accessibility but also factors like flexibility and hydrogen bonding of amino acids need to be considered when selecting sites for successful point mutation to enhance a protein's interaction with a separation medium.

In recent years, there has been a gradual increase in the *in silico* molecular modeling to mimic and understand molecular behavior of large bio-molecules *in vivo* and *in vitro*. Molecular modeling allows simulating the structure of a protein variant based on its native structure, and it has the potential to make the site selection much easier and dependable. It is foreseeable that molecular modeling will gain more popularity in guiding site selection for site-directed mutagenesis since many proteins' structures have been obtained and are available to the public. By means of a spherical probe with predetermined size, the accessibility of all the surface amino acid residues on a protein can be obtained provided the protein's crystal structure is available.[17] Furthermore, whether an amino acid side chain is engaged in hydrogen bonds can also be evaluated by modeling.

Molecular Mechanics use empirical schemes based on equations from classical physics, and thus are fast, consume less computing resources and are applicable to very large systems, such as DNA and proteins. In Molecular Mechanics energy

minimization programs are very widely used.[18,19] Ward and Shepherd used the energy minimization program, MMFF94, to predict and correlate the various changes in the chromatographic behavior with the corresponding changes in the His-tag structure.[20] Liu et al. used molecular simulations (mostly molecular mechanics) to elucidate the underlying mechanism involved in immobilized metal affinity chromatography (IMAC), metal ion binding to His-stag motifs, with varying number of histidines.[21]

We used the above molecular modeling tools to construct a specific ligand with which point-mutated protein may interact, and by carrying out the energy minimization calculation, the interaction of the side chain of a particular amino acid residue with a separation medium can be evaluated. Therefore, molecular modeling can undoubtedly help to deliberately select the sites for site-directed mutagenesis. The inevitable question is how effective it is to use molecular modeling tools to select the sites for successful site-directed mutagenesis, and this question can only be answered by experiments.

If the interaction is specifically dependent on the side chain of a particular amino acid, quantifying the interaction of a point-mutated protein with a separation medium can determine the availability of the side chain of the amino acid for inter-molecular interactions. Immobilized metal affinity chromatography (IMAC) is an ideal tool to carry out this study. In IMAC, a transition metal ion such as copper is immobilized in the stationary phase by a chelating agent such as iminodiacetic acid (IDA).[22] Copper ion was selected for the investigation because of the Cu-IDA column's ability to retain a protein with a single histidine.[23] The interaction between a protein and an IMAC column depends on the number and distribution of histidines on a protein

surface.[24] In fact, IMAC has been used to probe the surface topography of histidine residues in various studies.[16,22,24] Therefore, if a selected amino acid is replaced by histidine, the availability of the imidazole side chain of histidine for binding can be readily evaluated by quantifying the protein's interaction with an IMAC column.

In this report, we based the site selection process on the SA values, and *in silico* energy minimization was used to predict the potential interaction of protein with the separation medium (metal ion in IMAC). The effectiveness of this site selection method was evaluated by determining the binding strengths of the protein in IMAC. Thus, molecular modeling was used to guide the site-selection process for point mutations that were based on high SA values with the goal that the new amino acid residue would participate in the desired interaction favorably.

Materials and Methods

Site selection for protein mutagenesis by molecular modeling: Since the effectiveness of site selection by molecular modeling would be evaluated experimentally by IMAC, we built an *in silico* model consisting of copper ion immobilized to IDA (Fig.1). The model was built, using the molecular builder in the Molecular Operating Environment (MOE) software, similar to the model reported by Mrabet.[17] The model protein chosen for our investigations was T4 lysozyme. From the 3D structure (Protein Data Bank, 1LYD), the surface accessibility of each amino acid was calculated using Turbo-Frodo program (MOE). The amino acids with high surface accessible values were chosen. Homology modeling was used to incorporate point mutations into the lysozyme. Then, energy minimization (EM) was performed using MMFF94 force field in the MOE software, on the complexes consisting of mutated lysozyme with the *in silico* model (Fig.

2). Three sites for protein mutagenesis were selected from the results obtained from the EM calculation (Table 1).

Generating T4 lysozyme mutants with point mutation: Two different strategies were used to generate the three point mutations, K19H, K83H and K135H, where K stands for lysine being replaced and H stands for histidine. Since lysine-83 was located within the vicinity of the restriction site of EcoRI, a polymerase chain reaction (PCR) primer was designed to include both the EcoRI restriction site and the nucleotide substitutions. To obtain the PCR product containing the mutation K83H, a PCR between Primer 1 and Primer 2 was carried out (Fig. 3A).

If there were no viable restriction sites near the selected sites for point mutation, the second strategy illustrated in Fig. 3B was used. Two overlapping DNA fragments were generated. DNA fragment 1 was generated by carrying out a PCR between Primer 1 and Primer 2, while DNA fragment 2 was generated by carrying out a PCR between Primer 3 and Primer 4. To obtain the PCR product containing the point mutation, a PCR between DNA fragment 1 and 2 was carried out.

All the primers required for the PCR were ordered from the Sigma-Genosys (Woodlands, TX). The PCR reaction was done with the touch down method, recommended by Stratagene (Stratagene, La Jolla, CA). The obtained first round PCR products (DNA fragments) were purified and verified using the agarose gel electrophoresis. From the agarose gel picture, the PCR products were quantified. The second round PCR reaction was performed using the same PCR-mix as stated above, except the template used was the DNA fragments in the ratio of 1:1, from the first round with no original wild type (WT) plasmid. Consequently, the PCR products (all three

mutants) and the plasmid (WT) were digested with the restriction endonuclease enzymes. Similar as above, the restricted PCR products and the plasmid (WT) were purified, verified and estimated using agarose gel electrophoresis. Ligation reaction was set-up with 3:1 ratio of insert to the plasmid and was placed at 4 ºC for overnight, to obtain the ligated product (plasmid with the mutated gene).

E. coli. transformation: The ligated products were transformed into competent cells (*E. coli* RRI strain). Approximately 40 ng of the ligation product was suspended in 0.1 ml competent cells and was left on ice for 30 minutes. Then cells were immediately heat shocked at 42 ºC for 2 min, and then about 1 ml of LB media was added. The cells were incubated at 37 ºC for 1 hour. Subsequently, 200 µl aliquot of the LB incubated cells were plated on to the LB_{Amp100} (100 mg ampicillin / 1L of LB), and allowed to grow at 37 ºC for 24 hours.

Individual cells were then picked and plated on the LB_{Amp100} and allowed to grow at 37ºC for 24 hours.

Three individual colonies were picked and grown in 5 ml of liquid LB_{Amp100} medium for overnight. From the over night grown cells, 0.1 ml of the cells were inoculated in another 5 ml liquid LB_{Amp100}. This culture was induced with IPTG at O. D. ~1.0. After induction for one hour, 1ml of the cells was centrifuged and lysed. The over expression of the lysozyme was verified by SDS gel analysis (Fig.3). The remaining cells were used to prepare glycerol stocks (1ml culture + 230 µl of glycerol), stored at -80ºC. The enzyme activity was measured by the clearing of a *Micrococcus lysodeiktikus* cell suspension.[25]

Protein production and purification: From the glycerol stocks, the cells containing the mutated gene were plated and grown in liquid culture as described above. The overnight liquid culture was used to inoculate 1L LB medium. When the O. D. of the 1L LB medium reached ~ 1.2, 40 mg of IPTG was added to induce the cells for protein production. After 1 hr (final O.D. ~ 1.7), the cells were centrifuged at 4000 g for 10 min. Then the cell pellet was collected, and stored at - 80ºC before protein recovery and purification.

The frozen cells (from 1 Liter) were re-suspended in 20 ml of 20 mM sodium phosphate buffer at pH 7. After adding DNaseI, the cells were lysed by sonication. The sonication was done for 30 seconds 3 times with intermittently resting on ice for 30 sec. The lysed cells were then centrifuged for 30 min at 13000 g. The supernatant was decanted into a dialysis tubing with MWCO 12-14000 (spectra / pro®, Spectrum Laboratories, Inc., Rancho Dominguez, CA). Dialysis was performed with 20 mM sodium phosphate buffer pH ~ 7.2. The sample to dialysis buffer ratio was 1: 100. After 1h, the buffer was removed and the above dialysis step is repeated again.

The dialyzed protein solution was then separated by ion exchange chromatography using an Akta Explorer system (Amersham bioscience, Piscataway, NJ). The column and the resin used were XK 16/20 and Carboxyl Methyl Sephrose Fast Flow (Amersham Bioscience, Piscataway, NJ), respectively. The bed volume of the column was 25 ml. The equilibrating buffer used was 20 mM sodium phosphate pH ~ 7.2. The lysozyme was eluted at ~ 0.2 M NaCl in 20 mM sodium phosphate at pH 7.2.

The eluted lysozyme protein was then dialyzed in 20 mM sodium phosphate for 45 min to lower the salt concentration below 0.2 M NaCl. Then the above

chromatographic step was repeated again. The protein solution was concentrated by ultrafiltration (Amicon 8200, Millipore Corp., Bedford, MA). The concentrated protein was dialyzed with the buffer containing 20 mM sodium phosphate and 1 M NaCl, at pH 7. The concentration of the protein was measured at 280 nm by UV spectrophotometer (T4-lysozyme extinction coefficient (ϵ) = 1.28).[26] The protein purity was estimated to be greater than 95%, from SDS gel analysis. The final protein concentration was more than 1mg/mL.

Determining the dissociation constant by Zonal analysis: The dissociation constant of the protein with the metal ion in the IMAC column was determined using zonal analysis.[27] By varying the inhibitor concentration in the mobile phase, the dissociation constant was determined using the following equation.[28]

$$\frac{1}{V - V_0} = \frac{K_d}{V_0 * B_t} + \frac{K_d * [I]}{K_I * V_0 * B_t} \qquad (1)$$

Where V_0 = elution volume in the absence of interaction, [I] = the concentration of the inhibitor (imidazole) in the mobile phase, K_I = the dissociation constant of the imidazole-copper complex, K_d = dissociation constant for the interaction between the protein and the copper ion in IMAC, and B_t = total immobilized copper ions. A plot of $1/ (V-V_0)$ vs [I] would be linear. K_I is equal to the ratio of the intercept to slope, and K_d can be obtained from the intercept and the independently obtainable parameters V_0 and B_t.

The column and the resin used were HR 5/10 and Chelating Sephrose Fast Flow (Amersham Bioscience, Piscataway, NJ), respectively. The column bed height was 3.3 cm and the bed volume was 0.66 ml. The steps reported in Table 1 were followed for automatic runs. The elution buffer used was 1M NACl, 20 mM sodium phosphate, at pH

7. The flow rate of the mobile phase was 0.5 ml/ min. All the solutions used were filtered and degassed. The inhibitor concentration in the sample was varied accordingly with that of the elution buffer (3 mM, 6 mM and 9 mM, in elution buffer). The sample was injected onto the column via the loop mechanism, using a 250µl syringe. The sample volume used was 100 µl. Isocratic elution was carried out to obtain the elution volume (V). V_0 was determined by injecting 100 µl of a protein solution on to the blank column with no copper immobilized in the stationary phase.

Results and Discussion

The sites chosen for the point mutation were based on high SA values (Table 1), since greater SA values generally correspond to greater flexibility of the imidazole group (side chain of histidine residue), which can in turn presumably bind more strongly to the copper ion chelated to the IMAC column [22]. It is important to note here that the SA values chosen are based on protein crystal structure, which is a snap shot of the flexible protein structure present in the solution. And also as described in the earlier section SA cannot be the sole criterion for site selection process, since the microenvironment of the point mutant plays an important role whether the side chain of the mutant participates in the desired interaction favorably.

The amino acid residue chosen to be mutated in all the three mutants was lysine. Lysine was chosen because it is positively charged and mostly located on the surface of the protein. It was hoped that when lysine was replaced by the histidine, it would not alter the globular structure of the protein. All three point mutants had different microenvironment including their SA (Table 1). Molecular modeling (energy minimization) was used to understand this microenvironment affect on the potential

interaction of protein with the metal ion, in IMAC. Thus any difference in the energy minimization values can be attributed to both the SA and the microenvironment of the point mutants.

The yield of the proteins purified by the two-step chromatography was around 30 - 40% (Table 3). The specific activity of all the three mutants were comparable to the wild type protein (Table 4) indicating that the point mutations did not affect the globular structure of the T4 lysozyme. The final concentration of the proteins used to determine the protein binding strength was 1.0 mg/ml. The purity of the concentrated proteins (from SDS-PAGE) was more than 96% (Fig. 4). From the MALDI – TOF MS spectra the protein sequence of all four proteins (including wild type) was confirmed (Fig. 5).

The effectiveness of the molecular modeling was ascertained by comparing the energy minimization values with the protein binding strength values, obtained from the zonal analysis experiments. So, our hypothesis was that the protein binding strength (BS) values will correlate with the energy minimization values.

To determine the BS values, the IMAC column was regenerated before every run, to maintain the identical column conditions for every run. The time required to half elute the protein (V), i.e. the time at which the maximum absorbance occurs, was obtained from the elution profiles. The elution profile of mutant K19H is shown in Fig. 5. Understandably, the concentration of the imidazole [I] is inversely proportional to the time of elution of the protein (Fig. 7). The BS values of all three mutants were obtained from Eq. 1, by plotting $1/ (V-V_0)$ vs [I] (Fig. 8).

The SA, energy minimization and BS values for all the three mutants are tabulated in Table 5. The BS values are comparable to the values cited in the literature.[29]

Though all three mutants have high SA, the energy minimization values of K19H indicated an unfavorable interaction. Thus we expected the protein BS of K19H to be lowest, instead mutant K83H had the lowest protein BS.

In order to understand the discrepancy, the microenvironment of all the point mutants was closely observed. In case of mutant K19H, there were two negatively charged amino acids (ASP-20 and Glu-22) located next to Lys-19 (Fig. 9). This is an important observation because we suspect these negative charged residues might be interacting unfavorably with the IDA (which is negatively charged (Fig. 1)), resulting in unfavorable energy minimization value. However during experimentation, the presence of 1M NaCl in the elution buffer suppresses the non-specific electrostatic interaction, thus eliminating the unfavorable interaction. The low protein BS value for mutant K83H could be due to either the imidazole group of His-83 might not be totally flexible or might be involved in certain intra-molecule interactions, such as hydrogen bonding. Such discrepancy was also reported by Berna et al. [22] for His-57 in bovine chymotrypsin, where they proposed hydrogen bonding could be the main reason for low protein BS.

Though these results are not conclusive in establishing our hypothesis that SA and energy minimization values can be used as a predicting tool, but still they can be used as a guiding tool to generate site-directed protein mutants. Further work is in progress to generate more mutants with varied SA to validate the above hypothesis.

Conclusions

Site-directed protein mutagenesis can be used to facilitate protein purification and to study the protein function, but there is still no general guidance on how the sites should be selected. The main question is whether or not the replacing amino acid side chain

would be able to participate in desired interactions. In this report, site selection for protein mutagenesis was accomplished by identifying surface amino acids with high SA area by an *in silico* IDA-Cu(II) probe. Though there were many probable point mutations for experimentation, the number was drastically reduced by the energy minimization process. The mutated proteins were successfully expressed in *E. coli* cells, and all the mutants maintained comparable activity as the wild-type protein. Wild type protein had no interaction with IMAC; meanwhile all the three mutants interacted with IMAC to varying degree. The protein binding strength was shown to have a direct correlation with the histidine's surface accessible area, and this shows that molecular modeling calculated surface accessible area can be used as a site selection criterion. However, the energy minimization calculation should be used with caution, as K19H still showed considerable binding strength with immobilized metal ions despite the unfavorable prediction from the energy minimization calculation.

Acknowledgement

We thank J. Donahue in Biochemistry Dept. at Virginia Tech for her guidance and help in running the experiments. We thank Dr. Matthew's lab for kindly donating the plasmid containing the wild-type lysozyme. The work is mainly funded by a grant from Jeffress Memorial Trust, Grant # J-706.

Table 1. Sites selected on T4 lysozyme for site-directed mutagenesis after energy minimization modeling the potential protein-metal interaction.

Site Selected for Mutagenesis	Surface Accessible Area, $(Å)^2$	Final Energy
Lys-19	137	+5713
Lys-83	121	-8260
Lys-135	172	-18077

Table 2. The procedure followed for each run before running the protein sample during immobilized metal affinity chromatography experiments.

Solution	Time (min)
50 mM EDTA in Buffer A	12
H_2O	10
50 mM CuSO4	10
0.1 M Sodium Acetate, IM NaCl	17
H_2O	10
Elution buffer	20

Table 3. Stages of purification for Variant K135H

Protein stages (Varaint- K135H)	**Bio-activity (Units/20ul)**	**Bio-activity (Units / ml)**	**Protein Conc. (mg/ml)**	**Specific Activity (Units/mg)**	**Tot. Vol. (ml)**	**Tot. Pro. (mg)**
Cell Lysate	21.7	1085	1.0	1085	23..0	23.0
Dialyzed Sol.	19.3	965	0.87	1109	23.0	20.0
Chromatograph	23.3	1165	0.80	1456	12.0	9.6
Ultra Filter	47.8	2390	1.65	1448	3.2	5.28

Table 4. Specific activity of the purified wild-type T4 lysozyme and the three varaint.

Protein	Wild type	K19H	K83H	K135H
Specific Activity (Units/mg)	**1827**	**1987**	**1653**	**1456**

Table 5. Surface accessibility, energy minimization values (MM) and experimental binding strength values of all the three varaints.

His variants	**Accessible Surface Area (Å)2**	**Final Energy**	**Protein binding strength (*Ka*) (10^{4} M^{-1})**	**Imidazole binding strength (*KaI*) (10^{2} M^{-1})**
His-19	137	+ 5713	50.8	1.3
His-83	121	- 8260	36.7	1.4
His-135	172	- 18077	79.5	2.5

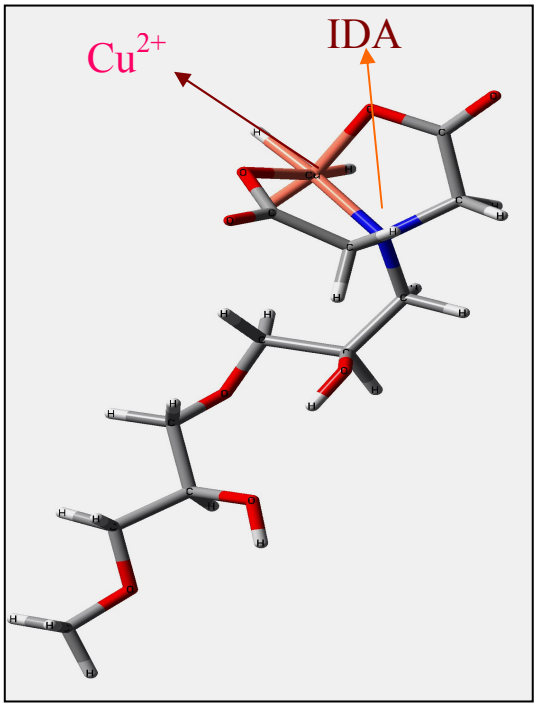


Figure 1. Cu^{2+}-IDA *in silico* model, built by molecular modeling, was used to mimic the Cu^{2+} chelated to IDA. Atoms Cu^{2+}, O and N are indicated in orange, red, and blue, respectively.

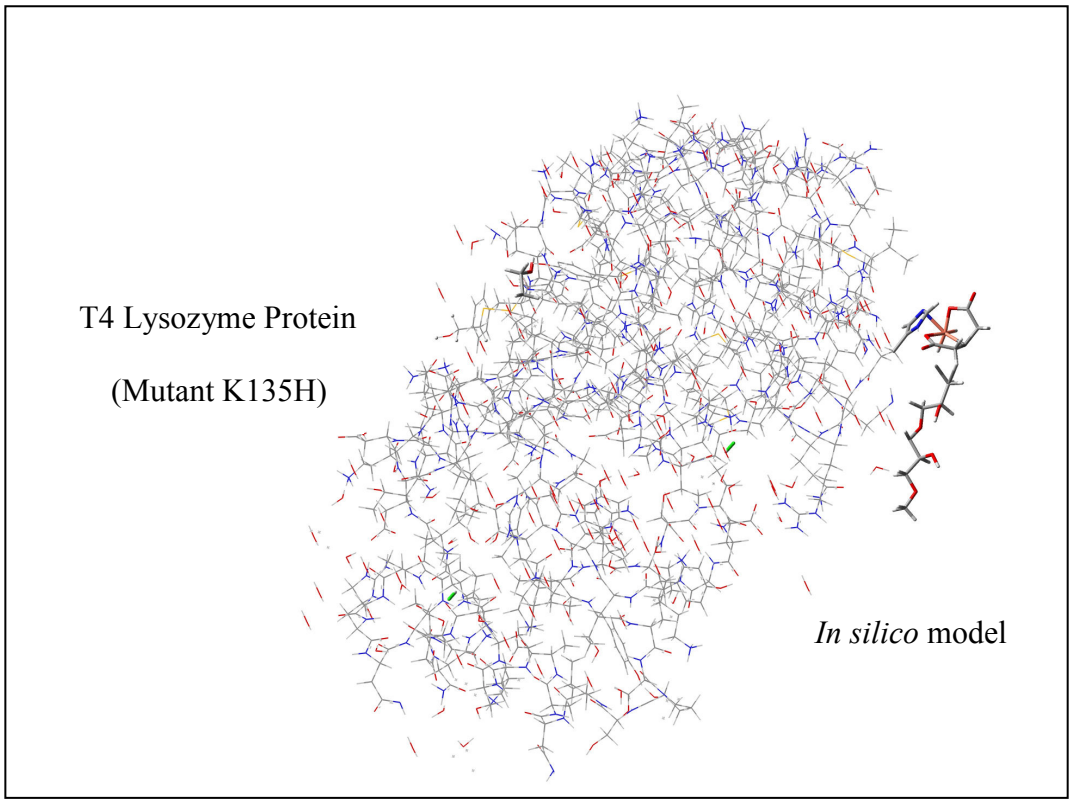


Figure 2. Energy minimization of mutant K135H and *in silico* (IDA) model complex.

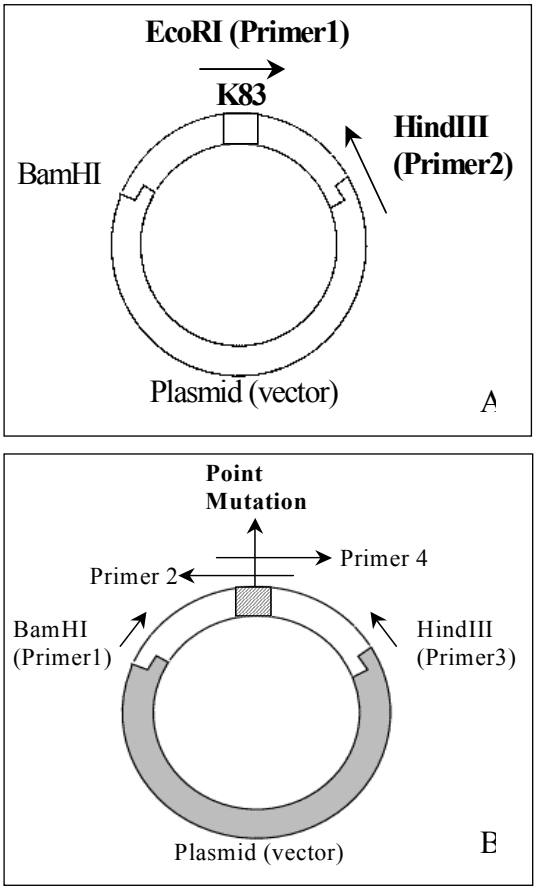


Figure 3. Strategies used to generate point mutants (A) K83H (B) K19H and K135H.

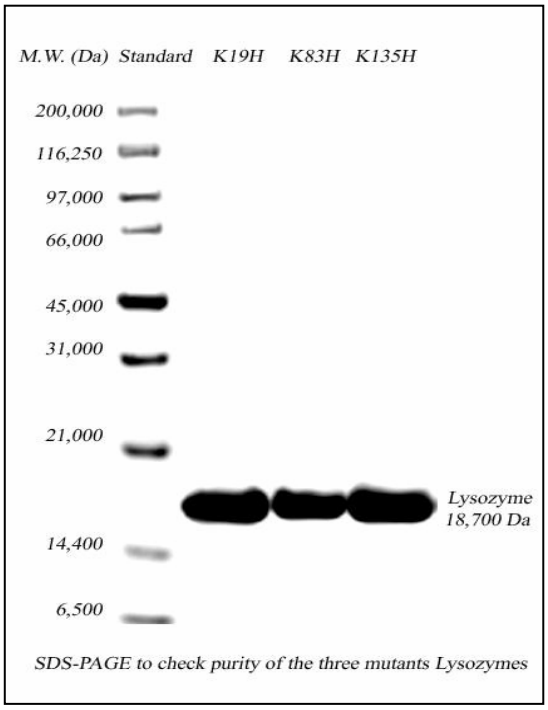


Figure 4. SDS-PAGE to verify the over expression and to check the purity of the three proteins after purification.

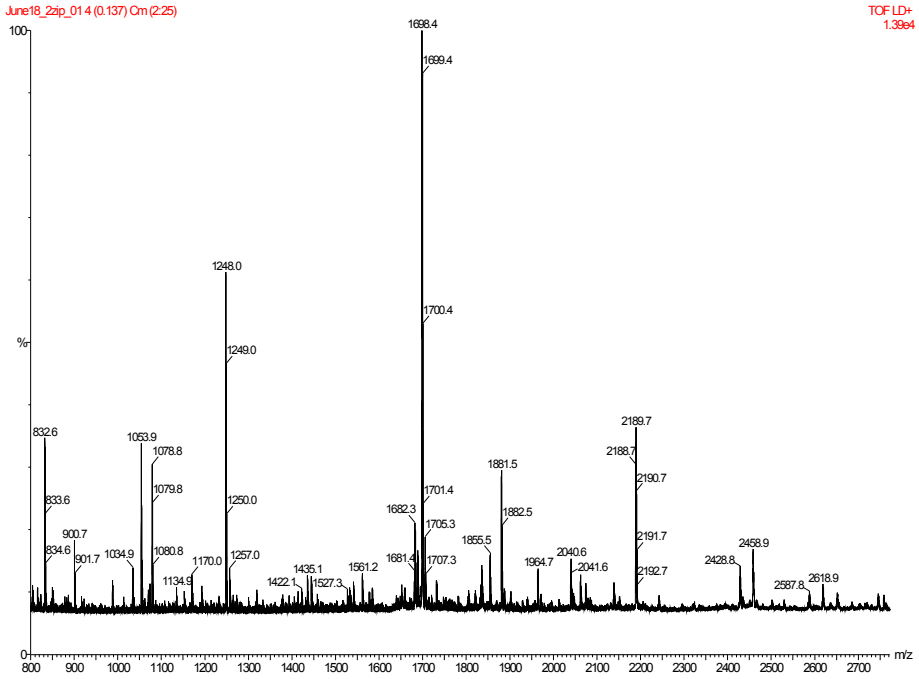


Figure 5. MALDI – TOF MS spectrum of mutant K19H.

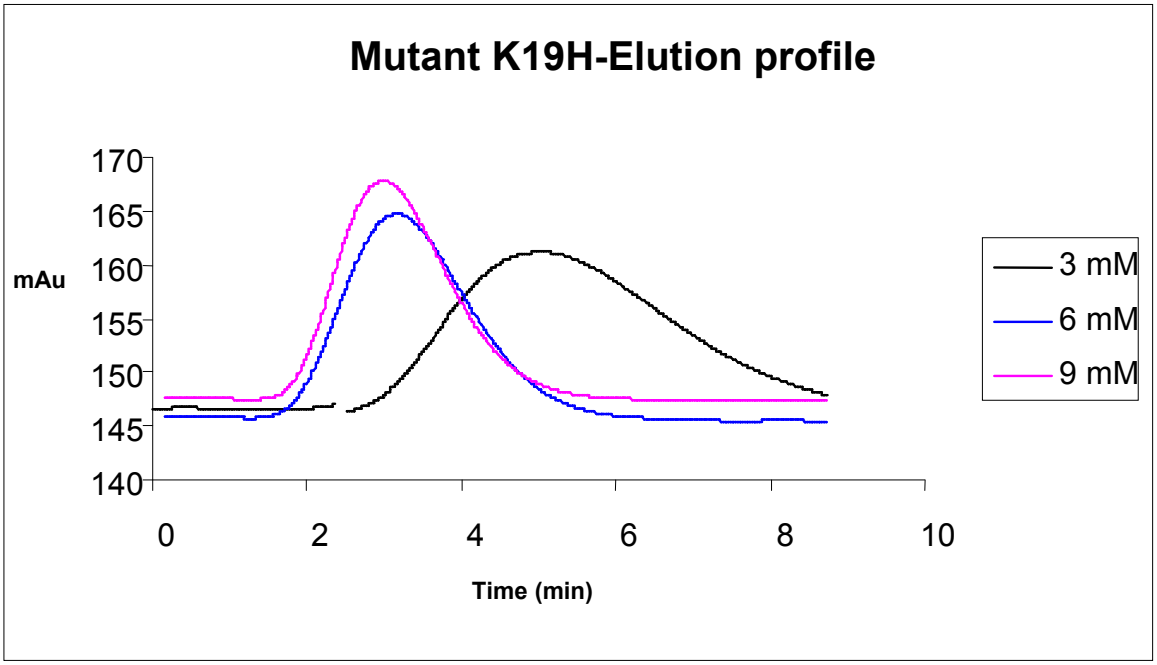


Figure 6. Elution profile of mutant protein (here K19H), obtained by varying the imidazole concentration (3 mm, 6mm, 9mm).

Figure 7.

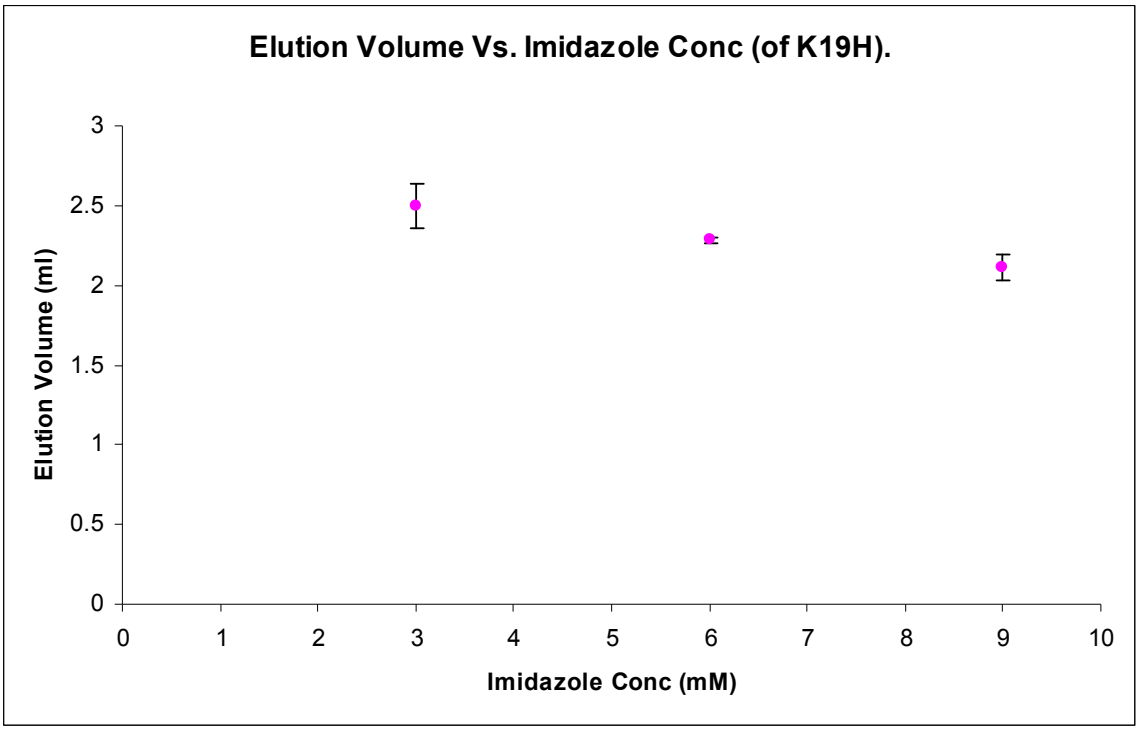


Figure 7. Graph illustrating the relationship between elution volume vs. imidazole concentration of mutant K19H.

Figure 8.

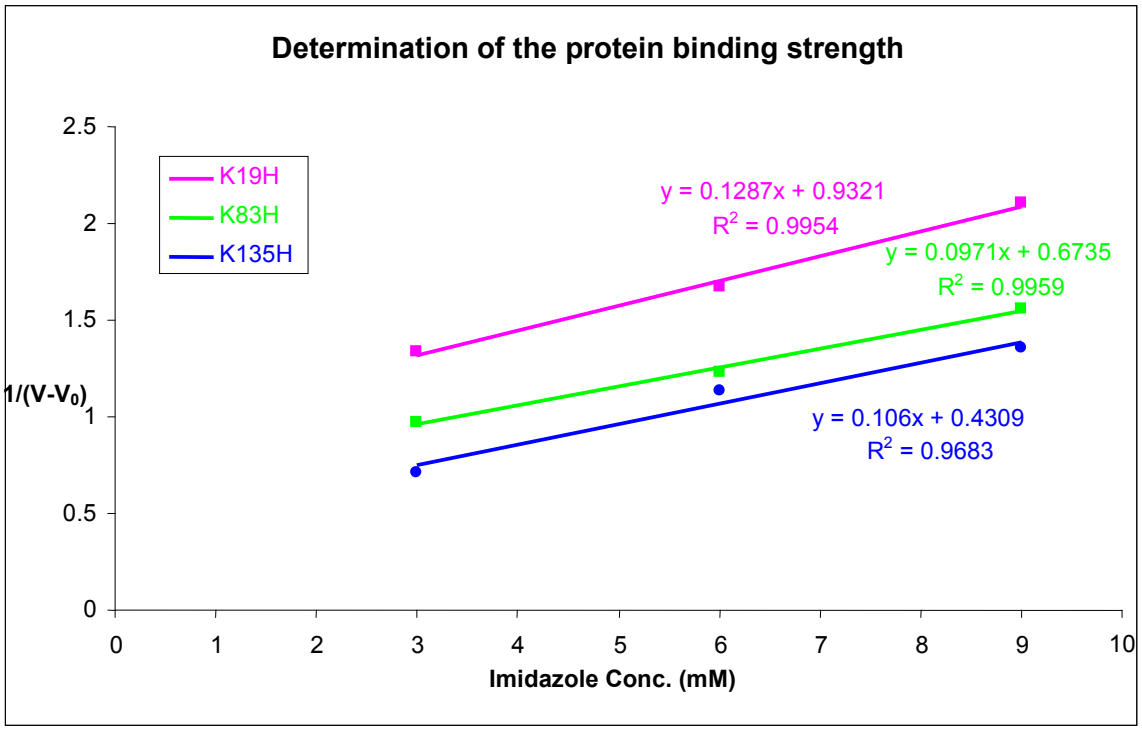


Figure 8. Plot to determine the dissociation constant of the protein metal ion complex (on the IMAC column) via the intercept.

Figure 9.

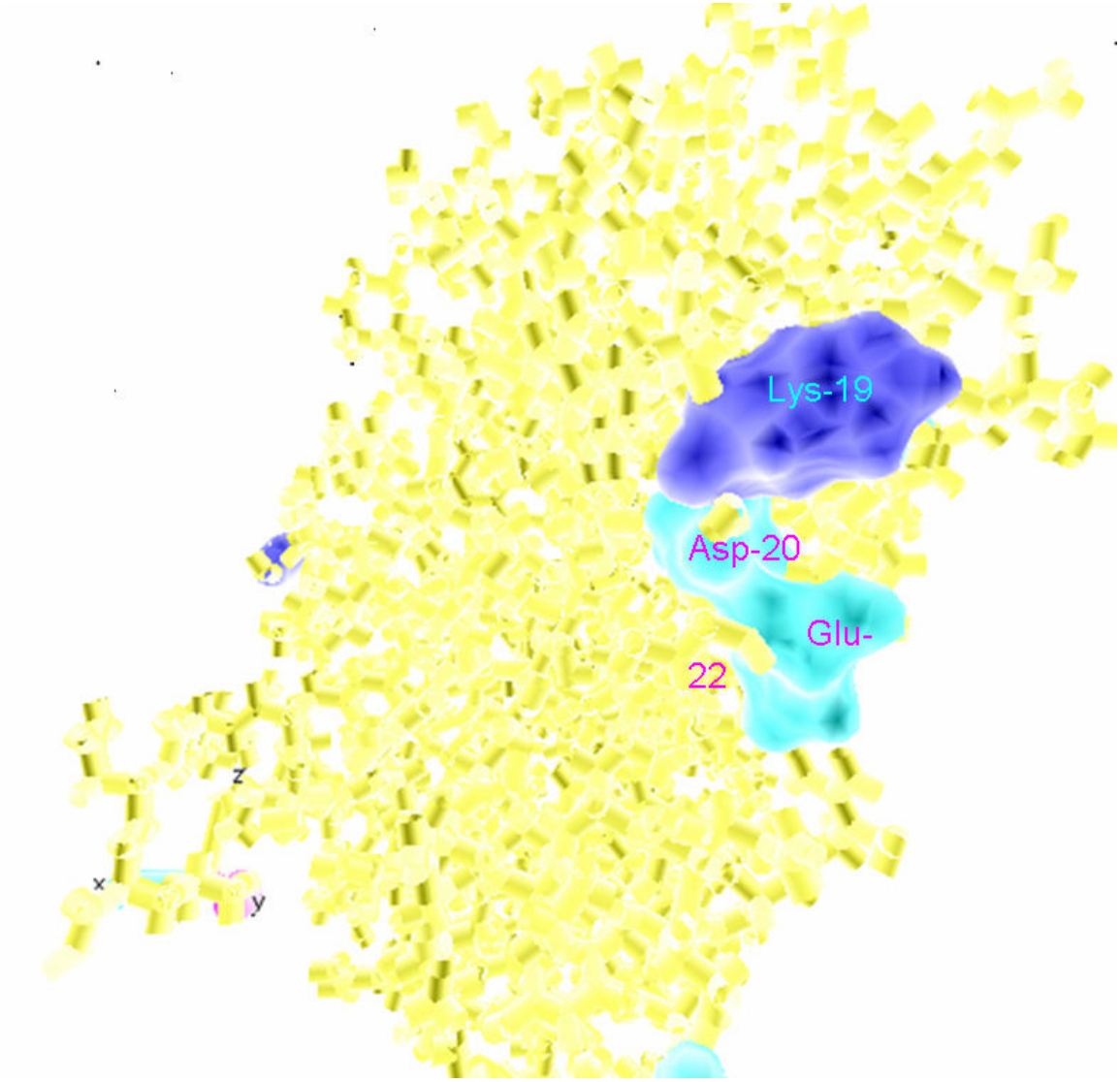


Figure 9. The crystal structure of mutant K19H. Two negatively charged amino acids, Asp-20 and Glu-22, are adjacent to His-19.

References

[1] Wheelwright, S. M. (1989) Designing downstream processes for large-scale protein purification. Bio/Technol. 5, 789-793.

[2] Sassenfeld, H. M. (1990) Engineering proteins for purification. Trends in Biotechnol. 8 (4), 88-93.

[3] Sassenfeld, H. M., Deeley, M., Rubero, J., Shriner, J. and Madani, H. (1991) Genetic alternations which facilitate protein purification: application in the biopharmaceutic industry. In Kelly, J. W. and Baldwin, T. O. (Eds), Application of Enzyme Biotechnology. Plenum Press, New York, pp 251-259.

[4] Zhang, C.-M., Love, R. T., Jilka, J. M. and Glatz, C. E. (2001) Genetic engineering strategies for purification of recombinant proteins from canola by anion exchange chromatography: an example of beta-glucuronidase. Biotechnol. Prog. 17 (1), 161-167.

[5] Zhang, C.-M. and Glatz, C. E. (2000) Suitability of Immobilized Metal Affinity Chromatography for Protein Purification from Canola. Biotechnol. Bioeng. 68, 51-58.

[6] Lien, S., Milner, S. J., Graham, L. D., Wallace, J. C. and Francis, G. (2001) Linkers for improved cleavage of fusion proteins with an engineering α–lytic protease. Biotechnol. Bioeng. 74 (4), 335-343.

[7] Chaga, G. S. (2001) Twenty-five years of immobilized metal ion affinity chromatography: past, present and future. J. Biochem., Biophys., Methods, 49, 313-334.

[8] Zhang, C.-M. and Glatz, C. E. (1999) Process engineering strategy for recombinant protein recovery from canola by cation ion exchange chromatography. Biotechnol. Progr. 15, 12-18.

[9] LaVallie, E. R. and McCoy, J. M. (1995) Gene fusion expression systems in *Escherichia coli*. Current Opinion in Biotechnol. 6, 501-506.

[10] Shortle, D. (1992) Mutational studies of protein structures and their stabilities. Quart. Rev. Biophys. 25, (2), 205-250.

[11] De Filippis, V., Snader, C. and Vriend, G. (1994). Predicting local structural changes that result from point mutations. Protein Eng. 7, (10), 1203-1208

[12] Eyal, E., Najmanovich, R., Edelman, M. and Sobolev, V. (2003) Protein Side-chain rearrangement in regions of point mutations. PROTEINS: structure, Function and Genetics, 50, 272-282.

[13] Chicz, R. M. and Regnier, F. E. (1989) Immobilized-metal affinity and hydroxyapatite chromatography of genetically engineered subtilisin. Anal. Chem. 61, 1742-1749.

[14] Todd, R. J., Johnson, R. D. and Arnold, F. H. (1994) Multiple-site binding interactions in metal-affinity chromatography. I. Equilibrium binding of engineered histidine-containing cytochromes c. J. Chromatogr. A. 662, 13-26.
[15] Sun, D.-P., Soderlind, E., Baase, W. A., Wozniak, J. A., Sauer, U and Matthews, B. W. (1991) Cumulative site-directed charge-change replacements in bacteriophage T4 lyozyme suggest that long-range electrostatic interactions contribute little to protein stability. J. Mol. Biol. 221, 873-887.
[16] Berna, P. P., Mrabet, N. T., Beeumen, J. V., Devreese, B., Porath, J. and Vijayalakshmi, M. A. (1997) Residue accessibility, hydrogen bonding, and molecular recognition: metal-chelate probing of active site histidines in chymotrpsins. Biochem. 36, 6896-6905.
[17] Mrabet, N. T. (1992) One-step purification of *Actinoplanes missouriensis* D-xylose isomerase by high-performance immobilized copper-affinity chromatography: functional analysis of surface histidine by site-directed mutagenesis. Biochem. 31, 2690-2702.
[18] Christensen, I. T. and Jorgenson, F. S. (1997) Molecular mechanics calculations of proteins. Comparison of different minimization strategies. Journal of biomolecular structure and dynamics, 15, 473-488.
[19] Kini, R.M. and Evans, H. J. (1991) Molecular modeling of proteins: A strategy for energy minimization by molecular mechanics in the AMBER force fields. Journal of biomolecular structure and dynamics, 9, 475-488.
[20] Ward, S. M. and Shepherd, R. E. (2000) Comparison of energy-minimized structures of [PdII (mida)] complexes of peptide sequences used in immobilized metal ion affinity chromatography protein separation. Inorganica Chimica, 311, 57-68.
[21] Liu, H. L., Ho, Y. and Hsu, C. M. (2003) Molecular simulations to determine the chelating mechanisms of various metal ions to the his-tag motif: a preliminary study. Journal of biomolecular structure and dynamics, 21, 31-41.
[22] Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. Nature. 258, 598-599.
[23] Hemdan, E. S., Zhao, Y. J., Sulkowski E. and Porath, J. (1989) Surface topography of histidine residues: a facile probe by immobilized metal ion affinity chromatography. Proc. Natl. Acad. Sci. USA, 86, 1811-1815.
[24] Mrabet, N. M. (1992) Physicochemical and structural implications for molecular recognition in immobilized metal affinity chromatography. Methods, 4, 14-24.
[25] Balasubramaniam, D., Wilkinson, C., Van Cott, K. and Zhang, C. M. (2003) Tobacco protein separation by aqueous two-phase extraction. J. Chromatogr. A, 989, (1), 119-129
[26] Poteete, A. R., Dao-Pin, S., Nicholson, H. and Matthews, B. W. (1991) Second-site reverants of an inactive T4 lysozyme mutant restore activity by restructuring the active site cleft. Biochem., 30, 1425-1432.
[27] Dunn, B. M. and Chaiken, I. M. (1974) Quantitative affinity chromatography, determination of binding constants by elution with eompetitive inhibitors. Proc. Nat. Acad. Sci. USA 71, 2382-2385.
[28] Winzor, D. J. (2001) Quantitative affinity chromatography. J. Biochem. Biophys. 49, 99-121.

[29] Belew Belew, M., Yip, T. T., Andersson, L. And Porath, J. (1987) Interaction of proteins with immobilized Cu^{2+}. Journal of Chromatography, 403, 197-206.

Appendix E: Report on up to date work on using SYSBYL to perform energy minimization

The in *silico* IMAC-IDA-Cu model was generated in SPARTAN software with PM (tm) force field (semi empirical). The mutations in the lysozyme model structure (1L63) were engineered by SYBYL. All the attempts using SYBYL and SPARTAN were unsuccessful. Gauss View was used to bind lysozyme to the in silico model (both bonded and non-bonded).

However, by employing constraints on the both bonded and non bonded conditions the minimization was successful in SYBYL. The program used was energy minimization, with force field MMFF94. The charges were introduced on both the *in silico* model and the lysozyme variant by choosing the option of Powell in the energy minimization menu. Energy minimization was accomplished by the steepest descent method, with convergence criterion of 0.05. The results are indicated in the Table 6.

Results and Conclusions

From the Tables 1, 2 and 3, the effect of four parameters (starting bond length, starting angle and bonded or non bonded and dielectric constants) on the energy minimization values was measured. These tables indicate that the effect of angle and starting bond length is significant. But these values do not reproduce the trend that was obtained from IMAC experiment values (Chapter 3, 4 and 6). The main reason for this deviation in the trend obtained from molecular modeling may be due to fact that the SYBYL energy minimization program considers the whole protein structure and not the specific to the microenvironment around the sites of mutation. Therefore the subtle structural variations do not significantly influence the final energy minimization values.

Table 1. All the variants with the *in silico* model were non-bonded and have a starting bond length of 2.2 Å. The various parameters, force constant, starting and final angle, starting and final energy and final bond length, measured during the energy minimization of four lysozyme variants at dielectric constants 20.

Lysozyme variant	Force constant	Starting Angle	Final Angle	Starting energy (kcal/mole)	Final Energy (kcal/mol)	Final Bond length (Å)
K135H	400	0	75.8	-275.2	-514.0	2.273
		90	100.7	52012077.3	-406.2	2.284
	100	0	80.9	-159.1	-443.0	2.378
		90	100.4	52012077.3	-407.0	2.389
	0	0	96.6	-159.1	-445.5	2.943
		90	90.5	52012077.3	-414.6	4.750
K19H	400	0	89.7	-283	31.8	2.295
		90	79.6	3583787.7	-428.4	2.282
	100	0	85.2	-168.3	-429.9	2.423
K83H	400	0	317.9	350350	-428.2	2.28
	100	0	325.6	350350	-431.6	2.405
	0	0	341.4	350350	-435.4	3.043
K43H	100	0	33.5	22926.2	-513.8	2.362

Table 2. . All the variants with the *in silico* model were non-bonded and have a different starting bond length. The various parameters, force constant, starting and final angle, starting and final energy, starting and final bond length, measured during the energy minimization of two lysozyme variants at dielectric constants 20.

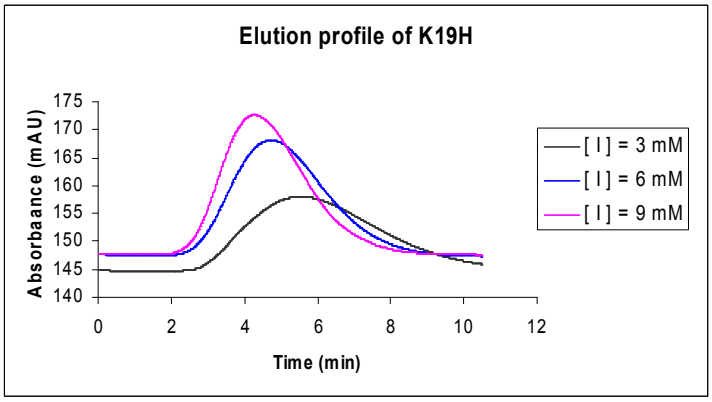
Lysozyme variant	Force constant	Starting Angle	Final Angle	Starting energy (kcal/mole)	Final Energy (kcal/mol)	Initial Bond length (Å)	Final Bond length (Å)
K43H	100	0	33.5	22926.2	-513.8	2.2	2.362
		0	34.3	3135.7	-512.0	2.5	2.317
	0	0	338.5	-438.6	-511.6	3.5	3.51
K135H	100	90	81.5	3143.8	-509.9	2.5	2.32

Table 3. All the variants with the *in silico* model were bonded and have the same starting bond length of 2.2 Å. The various parameters, force constant, starting and final angle, starting and final energy, starting and final bond length, measured during the energy minimization of three lysozyme variants with different dielectric constant.

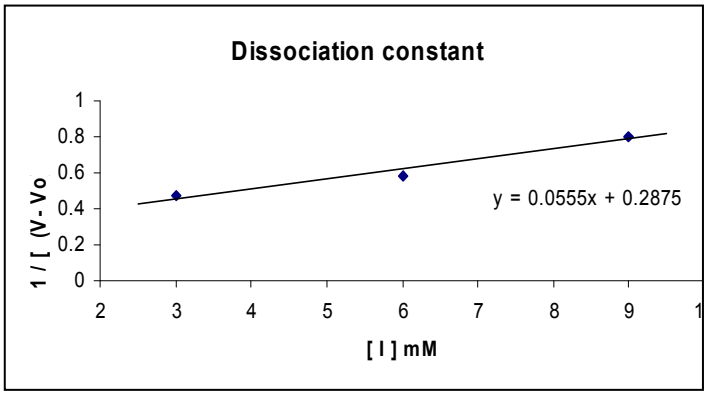
Lysozyme variant	Dielectric constant	Starting Angle	Final Angle	Starting energy (kcal/mole)	Final Energy (kcal/mol)	Initial Bond length (Å)	Final Bond length (Å)
K19H	81	0	248.8	-78.2	-444.6	2.2	1.949
	81	90	336.6	3515478	-445.4	2.2	2.934
K83H	81	0	70.1	350124	-336.8	2.2	1.972
K135H	40	0	119.1	-146.8	-459.8	2.2	1.94
	40	90	134.5	52057653	-444.9	2.2	1.936

Apendix G: Sample calculation to determine the binding strength value

$$K_d = \text{Intercept} \bullet V_0 \bullet [\text{Tcu}]$$
$$= 0.288 \bullet 1.69 \bullet 30$$
$$= 14.5 \ \mu\text{mol/ml}$$
$$\text{Binding strength} = \frac{1}{K_d}$$
$$= 6.9 \bullet 10^{-1} \ \text{M}^{-1}$$



$$K_I = \frac{\text{Intercept}}{\text{Slope}} = \frac{0.288}{0.056}$$
$$= 5.14 \ \text{mM}$$
$$\text{Binding strength} = \frac{1}{K_I}$$
$$= 2.0 \bullet 10^{-2} \ \text{M}^{-1}$$



Appendix H: Possible IMAC mechanism

Porath and Belew [53] suggested that that adsorption in IMAC is a kinetic process involving a two step SN1 mechanism. So when one of the water molecules departs from the metal complex, followed by a protein ligand approaching, the gel matrix with its chelating agent might hinder the total orbital overlap between the metal ion and the protein ligand, thus leading to a weaker bond than the corresponding bond with a free metal. Thus proper coordination chemistry of the metal ions should be considered before choosing for immobilization in an IMAC column. Some of the factors to be considered are 1) availability of the chelating ligands, 2) risk of undesired damage to the protein via oxidation or reduction, and 3) adsorption kinetics, i.e. the speed of the ligand replacement.

VITA

The author, Lakshmi Prasad Pathange, was born on January 27, 1978, in Hyderabad, India. He graduated with a Bachelor of Technology degree in Food Processing and Preservation Technology from college of Technology, Osmania University, in May 2000. In January 2001, he joined Virginia Tech to pursue his Master's degree in the department of Biological Systems Engineering. After obtaining his Master's degree in February 2003, he started pursuing his Ph.D. degree from August 2003 in the same department under the guidance of Dr. Chenming Zhang. After obtaining his doctoral degree he joined Bayer Healthcare Pharmaceuticals in May 2007 as a Process Development Engineer.