The Exploration of Protein Electrostatics Through NMR Chemical Shift Perturbations

by

Efrosini Artikis

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2020

Doctoral Committee:

Professor Charles L. Brooks III, Chair
Associate Professor Tomasz Cierpicki
Assistant Professor Aaron Frank
Professor Ayyalusamy Ramamoorthy
Associate Professor Paul Zimmerman

For my father

ACKNOWLEDGEMENTS

This period has been the most transformative of my life, and for this I have many people to thank.

First, my mentor Dr. Charles L. Brooks III who taught me how to think with a scientific mind. Thank you for helping me build the foundation of my career. I aspire to reach your heights.

Also, I'd like to thank my committee who saw me through the growing pains. Your comments and advice tore me down and built me back up.

To my lab mates, both past and present, with whom I've grown and learned - I look forward to seeing all the science that you will offer the world, and I hope to be your colleague again in the future. I'd like to thank my collaborators, Dr. Duncan MacKenzie and Dr. Elizabeth Meiering from the University of Waterloo for allowing me to share in their work.

To Dr. Hedieh Torabifard and Kate Dyki - we shared many laughs and a few tears, but we made it through all the better. Thank you for your friendship and kindness.

To all the people that have enriched my life during this time especially, Jeff Maltas, Katie Gentry and John Owens - thank you for your support and your encouragement. I won't forget it.

I'd like to thank the Biophysics Program, especially Dr. Sarah Veatch, for giving me the chance to study at Michigan, it has truly been an honor.

Lastly, but perhaps most importantly I'd like to thank my family – Bobby, Mom and Catherina. When I didn't think I could, you told me I would – thanks for being there with me every step of the way.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Understanding the relationship between protein structure and function is paramount to gaining insight into important biological mechanisms. In this context, pH often plays a significant role. The organization of charge within a protein prepares it to form intra-/inter-molecular interactions as the environmental pH changes. Studying the pK_a values of titratable groups in a protein allows us to understand its electrostatic network. Computational pK_a values are influenced by a microscopic environment and are often compared to macroscopic pK_a values derived from NMR experiments. In this work we aim to understand the impact of pH on NMR chemical shift perturbations such that we can bridge the gap between computational and experimental observables.

Peptide model systems have historically been used in NMR spectroscopy to detangle the many contributions which compose the observable NMR chemical shift. Utilizing molecular dynamics simulations, we sampled the conformational preferences of model tripeptides each containing one of the four titratable groups (aspartic acid, glutamic acid, histidine or lysine) in either a protonated or deprotonated state. The conformational ensembles obtained during the simulations were then used to compute pH-dependent NMR chemical shift perturbations for each nucleus in the tripeptides. The perturbations agree well with experimental findings and elucidate the relationship between charge and chemical shift. Furthermore, these results allow for better interpretation of NMR spectra and the possible integration of pH in chemical shift prediction paradigms.

Although random coil chemical shifts serve as the basis for chemical shift prediction and interpretation, the complexity of the protein environment can produce drastically different behaviors. In the second study, we investigate the ability to utilize peptide derived chemical shift perturbations along with through-space electrostatic and conformational effects to compute pH-dependent NMR chemical shifts of the dynamic ensembles produced by constant pH molecular dynamics (CpHMD). Hen egg white lysozyme (HEWL) is an appropriate benchmark protein to probe the efficacy of a new protocol which fortifies the microscopic pK_a values from simulation with macroscopic influences. The inclusion of the pH-dependent chemical shift contribution improved the results from the empirical chemical shift predictor for both ^{15}N and ^{1}H atoms and added dimensionality to the CpHMD simulations informing pH-dependent conformational fluctuations in HEWL. The newly derived macroscopic pK_a values from simulation were directly compared to the experimental pK_a values.

Lastly, hisactophilin, a highly charged protein, is studied in order to identify critical residues that trigger the pH-dependent switching behavior of a post-translational modification. Hisactophilin has an N-terminal myristoyl group which is buried inside the beta-trefoil cavity in pH values greater than 7.5. However, at pH values lower than 6.5, the myristoyl group is accessible and may incorporate itself into an external lipid membrane. The small pH range where this switching behavior occurs likely corresponds with the protonation event of one or a few titratable residues. Implicit and explicit solvent CpHMD simulations allow us to explore the residues involved in the pH-dependent mechanism and formulate conclusions about charge redistribution.

CHAPTER 1

Introduction

1.1 Structural Biology of Proteins

The ability to associate the intricate and intentional architecture of proteins with their evolved function is principal to understanding biology. Recent advances in the field of structural biology have increased our ability to answer foundational questions as more protein structures are deciphered. The main methodologies of this field include X-ray crystallography, cryo-electron microscopy and nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography which was pioneered by giants such as James Sumner, Dorothy Hodgkins, Max Perutz and Rosalind Franklin[1] has seen great improvements such as the use of X-ray free electron lasers (XFEL)[2] and lipidic cubic phase[3]. Not only does XFEL increase structure resolution by allowing for a longer diffraction time of crystals without the detrimental effects of radiation damage[2]; lipidic cubic phase is an elegant solution for the crystallization of difficult proteins such as those that are associated with membranes[3].

Cryo-electron microscopy with the invention of the Titan Krios can now achieve astonishing resolutions up to 3.5Å[4], producing stunning images of virus capsids[5] and large protein machines[6]. NMR spectroscopists have benefited from the superconducting 1.1 GHz magnet whose hyper-sensitivity produces greater separation of resonances resulting in very high-resolution spectra[7]. NMR is unique among the other two techniques, as it allows for the elucidation of both structure and dynamics in real-time. With this capability, NMR has been utilized to investigate problems

such as amyloid aggregation[8], protein folding[9], chaperone activity[10] and many other dynamic protein mechanisms[11].

With the availability of more protein structures, and the push toward atomic resolution, computational methodologies have become necessary not only in the form of noise reduction and image deconvolution, but as a way to provide great insight and guidance on all aspects of the protein structure-function relationship. As computational biophysics serves to propel the structural biology field forward, the need for seamless comparison between computational and experimental observables is imperative.

1.2 Computing pK_a Values

When investigating the protein structure and function relationship, consideration of pH is paramount. The protein's internal network of charge remains poised to respond to changes in the environmental pH. When this balance is perturbed titratable residues experience a protonation event. pK_a values describe residue specific protonation events and reflect aspects of the residue's microenvironment. Thus, the ability to measure or compute pK_a values can illuminate the details of many pH-dependent mechanisms.

Explicit solvent constant pH molecular dynamics simulation in the context of multi-site lambda dynamics ($CpHMD^{MSLD}$) was developed by the Brooks group[12-14] to allow for the calculation of nucleic acid and protein pK_a values. This methodology utilizes a continuous variable, lambda[12], which is allowed to fluctuate between the bounds of 0 and 1[12] representing the protonated and deprotonated states of a titratable residue[14]. For example, an aspartic acid residue is assigned three lambda sites, one monitoring protonation of the first oxygen on the sidechain carboxyl group, the second reports the titration on the second oxygen, and the third represents a fully deprotonated sidechain. The lambda values are coupled to the dynamic steps of the simulation[14], such that both

protonation and conformation can be probed simultaneously[14]. However, because lambda is a continuous variable, unphysical intermediate states may occur and must be disregarded in the final calculations using a lambda cutoff (typically $\lambda > 0.80$ is used)[13]. CpHMD^{MSLD} relies on a hybrid Hamiltonian[12] which describes the number of titrating residues, the protein environment and the coordinates of protonated and deprotonated sites[14] – a complete expression of its potential energy equation can be found in references 12 to 14. The free energy of protonation for model compounds[14] (common titratable residues: aspartic acid, glutamic acid, histidine and lysine), is used to calibrate the environmental pH of the simulation in the form of fixed biases[14]. Parameters for titratable residues with high pK_a values such as cysteine, arginine and tyrosine may also be computed. Furthermore, variable biases can be altered to increase sampling of specific protonation sites[14]. Benchmarking studies of this method indicate good agreement with experiment[14-16], and it has been employed to study proteins in a variety of contexts, such as the identification of catalytic residues in cellobiohydrolases[17] and recently the pH-dependent conformational mechanism of influenza A[18].

1.3 NMR Chemical Shift Prediction

Due to the ability of NMR chemical shifts to comprehensively capture the local and global protein environment, there has been significant effort in recent years to establish methods that robustly predict these observables. In combination with classical descriptions of chemical phenomena[19, 20], empirical observations have been used to inform highly accurate chemical shift prediction programs[21-23]. In addition, many popular predictors also employ machine learning algorithms and are trained with large databases of chemical shifts obtained either from experimental NMR data repositories[21-23], such as the Biological Magnetic Resonance Bank (BMRB)[24] or through numerous quantum mechanical NMR calculations[25]. SHIFTX2, one of the

most popular chemical shift prediction programs can predict backbone and sidechain chemical shifts with an accuracy of R=0.96 and R=0.99 respectively[21]. Utilizing a novel hybrid method, SHIFTX2 builds an extensive feature vector from protein coordinate data and uses an additive regression approach to produce chemical shifts based on structure[21]. It combines these predictions with SHIFTY+ which is a unique algorithm used to predict sequence based chemical shifts by identifying a sequence homolog with available chemical shift data[21]. This dual approach allows for highly accurate chemical shift predictions. Another predictor, SPARTA+[22] emphasizes the relationship between torsion angles and chemical shifts as the input protein sequence is subdivided into tripeptide fragments. Additionally, information about hydrogen bonding interactions, ring current effect, electric field effects and predicted flexibility[22] is included with the peptide fragments to produce a feature vector of 113 nodes[22] from which a trained neural network is able to predict accurate backbone chemical shifts. The work of Robustelli *et. al.*[26] has demonstrated that empirical predictors such as SPARTA+, show improved performance when an ensemble of structures from molecular dynamics is used to compute dynamically averaged chemical shifts, over those predicted from static crystal structures.

The incorporation of machine learning has allowed for unparalleled accuracy in the prediction of NMR chemical shifts[21,22]. However, since the large volume of experimental chemical shifts required for algorithm training are frequently of proteins at physiological pH, non-physiological conditions present a challenge for currently available predictors. As of 2017 there were only 59 entries listed in the BMRB of protein pH titrations, exemplifying the dearth of available pH NMR data. Many of the current predictors utilize the Buckingham equation[20] to express the contribution of electrostatics to the predicted chemical shift.

The Buckingham equation[20], of which will be a focal point of this thesis describes the electrostatic chemical shift contribution (δ_{EF}) with the following expression**:**

$$\delta_{EF} = A_{\parallel} \boldsymbol{E_z} \quad (1)$$

where E_z is the electric field (**E**) component of the nucleus of interest in the z direction, and A_{II} represents the nuclear polarizability constant specific to the bond type[20]. However, the implementation of the Buckingham equation within these programs is limited to a few donor and acceptor pairs[22,23] where the atomic charges remain static[22,23]. Although this maintains the predictor's speed, it does not adequately capture the effects of pH especially for titratable residues. To date there are no available methods to predict pH-dependent NMR chemical shifts of proteins, with the exception of a recently developed predictor POTENCI[27], from the Mulder group which predicts chemical shifts of intrinsically disordered polypeptides in various environmental pH and salt conditions.

1.4 Outline of Thesis

In the following dissertation, the second chapter presents the identification of pH-dependent chemical shift perturbations of model peptides. The subsequent chapter describes the application of the peptide derived chemical shifts to the hen egg white lysozyme protein. Chapter 4 examines the pertinent residues responsible for the pH-dependent switching mechanism of hisactophilin. Finally, the last chapter offers general conclusions and broader applications of this work. Each chapter follows the same structure with an introduction, methods section, results and conclusions followed by the bibliography.

1.5 Bibliography

1. Jaskolski, M., Dauter, Z. and Wlodawer, A., 2014. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *The FEBS journal*, *281*(18), pp.3985-4009.

2. Liu, H. and Lee, W.T., 2019. The XFEL protein crystallography: Developments and perspectives. *International journal of molecular sciences*, *20*(14), p.3421.

3. Caffrey, M., 2015. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallographica Section F: Structural Biology Communications*, *71*(1), pp.3-18.

4. Grigorieff, N. and Harrison, S.C., 2011. Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Current opinion in structural biology*, *21*(2), pp.265-273

5. Zhang, X., Ge, P., Yu, X., Brannan, J.M., Bi, G., Zhang, Q., Schein, S. and Zhou, Z.H., 2013. Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nature structural & molecular biology*, *20*(1), p.105.

6. Kato, H.E., Zhang, Y., Hu, H., Suomivuori, C.M., Kadji, F.M.N., Aoki, J., Kumar, K.K., Fonseca, R., Hilger, D., Huang, W. and Latorraca, N.R., 2019. Conformational transitions of a neurotensin receptor 1–G i1 complex. *Nature*, *572*(7767), pp.80-85.

7. Maeda, H. and Yanagisawa, Y., 2019. Future prospects for NMR magnets: A perspective. *Journal of Magnetic Resonance*, *306*, pp.80-85.

8. Barnes, C.A., Robertson, A.J., Louis, J.M., Anfinrud, P. and Bax, A., 2019. Observation of β-Amyloid Peptide Oligomerization by Pressure-Jump NMR Spectroscopy. *Journal of the American Chemical Society*, *141*(35), pp.13762-13766.

9. Dyson, H.J. and Wright, P.E., 2019. Perspective: the essential role of NMR in the discovery and characterization of intrinsically disordered proteins. *Journal of Biomolecular NMR*, *73*(12), pp.651-659.

10. Yu, X.C., Hu, Y., Ding, J., Li, H. and Jin, C., 2019. Structural basis and mechanism of the unfolding-induced activation of HdeA, a bacterial acid response chaperone. *Journal of Biological Chemistry*, *294*(9), pp.3192-3206.

11. Capraro, D.T., Burban, D.J. and Jennings, P.A., 2020. Unraveling Allostery in a Knotted Minimal Methyltransferase by NMR Spectroscopy. *Journal of Molecular Biology*.

12. Knight, J.L. and Brooks III, C.L., 2011. Applying efficient implicit nongeometric constraints in alchemical free energy simulations. *Journal of computational chemistry*, *32*(16), pp.3423-3432.

13. Knight, J.L. and Brooks III, C.L., 2011. Multisite λ dynamics for simulated structure–activity relationship studies. *Journal of chemical theory and computation*, *7*(9), pp.2728-2739.

14. Goh, G.B., Hulbert, B.S., Zhou, H. and Brooks III, C.L., 2014. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: structure, function, and bioinformatics*, *82*(7), pp.1319-1331.

15. Wallace, J.A. and Shen, J.K., 2011. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *Journal of chemical theory and computation*, *7*(8), pp.2617-2629

16. Goh, G.B., Knight, J.L. and Brooks, C.L., 2012. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *Journal of chemical theory and computation*, *8*(1), pp.36-46.

17. Bu, L., Crowley, M.F., Himmel, M.E. and Beckham, G.T., 2013. Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. *Journal of Biological Chemistry*, *288*(17), pp.12175-12186

18. Torabifard, H., Panahi, A. and Brooks, C.L., 2020. M2 amphipathic helices facilitate pH-dependent conformational transition in influenza A virus. *Proceedings of the National Academy of Sciences*, *117*(7), pp.3583-3591.

19. Haigh, C.W. and Mallion, R.B., 1979. Ring current theories in nuclear magnetic resonance. *Progress in nuclear magnetic resonance spectroscopy*, *13*(4), pp.303-344.

20. Buckingham, A.D., 1960. Chemical shifts in the nuclear magnetic resonance spectra of molecules containing polar groups. *Canadian Journal of Chemistry*, *38*(2), pp.300-307.

21. Han, B., Liu, Y., Ginzinger, S.W. and Wishart, D.S., 2011. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of biomolecular NMR*, *50*(1), p.43.

22. Shen, Y. and Bax, A., 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR*, *48*(1), pp.13-22.

23. Meiler, J., 2003. PROSHIFT: protein chemical shift prediction using artificial neural networks. *Journal of biomolecular NMR*, *26*(1), pp.25-37.

24. Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z. and Nakatani, E., 2007. BioMagResBank. *Nucleic acids research*, *36*(suppl_1), pp.D402-D408.

25. Xu, X.P. and Case, D.A., 2001. Automated prediction of 15N, 13Cα, 13Cβ and 13C′ chemical shifts in proteins using a density functional database. *Journal of biomolecular NMR*, *21*(4), pp.321-333.

26. Robustelli, P., Stafford, K.A. and Palmer III, A.G., 2012. Interpreting protein structural dynamics from NMR chemical shifts. *Journal of the American Chemical Society*, *134*(14), pp.6365-6374.

27. Nielsen, J.T. and Mulder, F.A., 2018. POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *Journal of biomolecular NMR*, *70*(3), pp.141-165.

CHAPTER 2

Peptide NMR Chemical Shifts

This chapter was adapted from:
Artikis, E. and Brooks III, C.L., 2019. Modeling pH-Dependent NMR Chemical Shift Perturbations in Peptides. *Biophysical journal*, *117*(2), pp.258-268.

2.1 Introduction

NMR chemical shifts are composite variables which allow for the holistic description of the atomic and chemical environment of a protein in the context of both dynamics and structure[1,2]. However, it is often difficult to decompose and interpret all of the factors that influence the experimentally measured chemical shifts. Nevertheless, both experimental and computational studies of model peptides have enabled a greater understanding of the relationship between chemical shifts and physical attributes of the peptides and their environments[3-5,54]. In combination with classical descriptions of chemical phenomena[6-9], empirical observations have been used to inform highly accurate chemical shift prediction programs[10-15]. In addition, many popular predictors also employ machine learning algorithms and are trained with large databases of chemical shifts obtained either from experimental NMR data repositories[10-12], such as the Biological Magnetic Resonance Bank (BMRB),[15] or through numerous quantum mechanical calculations[13,14].

Currently, most chemical shift predictors compute absolute shifts at physiological pH (6.5-7.5), since most NMR data repositories do not contain a statistically significant number of pH dependent chemical shift datasets[16] to enable the parameterization of pH conditions outside of this

range. The influences of pH on the shielding of atomic nuclei, which give rise to the observed chemical shift, result from the change in the electronic environment of these nuclei. This may be considered to arise from two mechanisms one local, the chemical environment specific electronic structure change due to the addition or loss of a proton at the site of protonation: effects that occur through charge rearrangement in the bonded environment of the site of protonation state change. The other being a through-space electrostatic effect from the altered electrostatic fields this charge change and any attendant conformational changes produce for nuclei that are not local, in the bonded sense, to the site of change. The electrostatic effects in most chemical shift predictors are primarily modeled by the Buckingham equation (Eq. 1) [9]. This formulism computes the electrostatic chemical shift contribution (δ_{EF})**:**

$$\delta_{EF} = A_{\parallel} E_z \tag{1}$$

where E_z is the electric field (**E**) component of the nucleus of interest in the z direction, and A_{II} represents the nuclear polarizability constant specific to the bond type[9]. This description is primarily useful for through-space effects and is typically applied to amide proton and nitrogens which tend to be most sensitive to solvent-exposure[17]. However, the through-space effects alone do not fully capture the complexity of pH-dependent NMR chemical shift perturbations (CSPs) as noted above, because nuclei closest to the titrating group may be dominated by through-bond shielding effects. It has been demonstrated that nuclei on the titratable residue exhibit prominent CSPs and serve as reporters for pK_a determination in NMR titration experiments[16,22]. In a bioinformatics survey of the existing NMR pH titration datasets deposited in the BMRB, Farrell *et. al* report significant CSPs (0.1 - 0.9 ppm: ^{1}H, 0.7 – 2.9 ppm :^{15}N, and 1.2 - 3.7 ppm:^{13}C) associated with model pK_a values[16].

Pioneering *ab initio* studies on model peptides have greatly contributed to the understanding of NMR chemical shifts and have established guidelines for the interpretation and prediction of these observables[24-27]. In this work, we expand on these efforts by further investigating the relationship between protonation and pH dependent NMR chemical shifts of model tri-peptides (Figure 1). Previous DFT (density functional theory) studies of pH effects have compared absolute NMR chemical shifts to the experimental datasets of random coils[28,29]. Notably, Xu *et. al*[26], using the peptide construct Ace-GGXGG-NH_2 concluded that "gas-phase DFT NMR chemical shifts of neutral titratable groups (X=Asp, Glu, Lys and Arg) produced better agreement with experimental measurements than when compared to the corresponding charged residues"[26]. In this work, we will utilize the recently published NMR dataset from the McIntosh group[30] for the systematic comparison of computed pH-dependent CSPs.

In our approach, molecular dynamics (MD) was carried out to probe the conformational preferences of protonated and deprotonated states of titratable groups Asp, Glu, His and Lys in a tri-peptide construct.

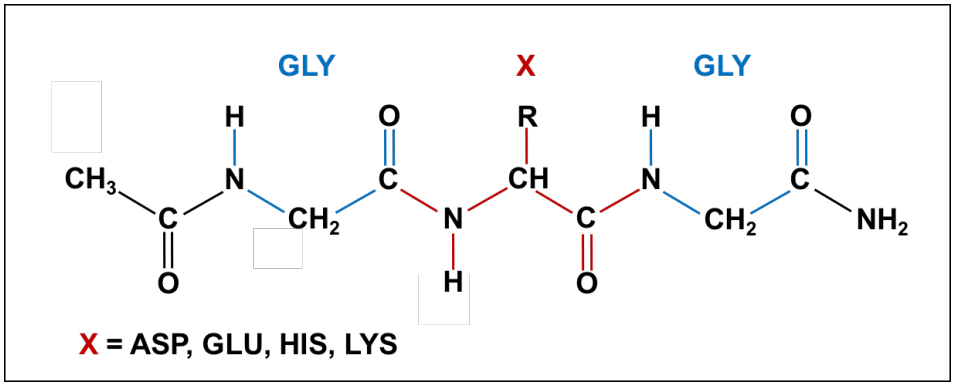


Figure 2.1. Tri-peptide construct used in all calculations adopted from Platzer *et. al*[30]. The peptide with the sequence Ace-GXG-NH_2 where X represents R-groups Asp, Glu, His and Lys is either in a protonated or deprotonated state.

Subsequently, we utilized quantum mechanical (QM) calculations to compute NMR chemical shifts and compare our calculations to the measured chemical shift perturbations as published by

the McIntosh group[30]. Finally, we examine the impact of through-space effects on the amide-protons and carbonyl carbons by both applying the Buckingham equation to the results of the *ab initio* calculations as well as by adding explicit water molecules to the DFT chemical shift calculations. From these results, we hope to better understand the manifestation of pH in NMR CSPs such that we can establish the groundwork for fortifying already available chemical shift predictors with the ability to compute NMR chemical shifts for proteins in a variety of pH conditions.

2.2 Methodology

2.2.1 Building Input Structures

Peptides were built using the CHARMM simulation package[31], with the sequence Ace-GLY-X-GLY-NH_2 where X is the titratable residue of interest (ASP, GLU, HIS, or LYS). Protonation states were altered using the respective CHARMM patches, ASPP, GLUP, LSN, and the terminal ends of the peptides were capped using the ACE and CT2 patches. For each peptide, protonated and deprotonated states were generated, with the exception of histidine for which two deprotonated states were built to represent the variable occupation of the exchangeable proton on either the epsilon or delta nitrogen (HSE, HSD).

2.2.2 Implicit Solvent GBSW Molecular Dynamics Simulations

Structural ensembles of each peptide were produced for a fixed protonation state by running 650 ns of molecular dynamics. The GBSW (a generalized Born model with a simple switching function) implicit solvent model[34] was used and a salt concentration of 50mM was included in order to replicate experimental conditions[30]. All implicit solvent simulations were performed using the CHARMM package and the CHARMM22[32] force field with GBSW-specific CMAP[33] backbone correction. A non-bonded cutoff of 12 Å was applied, as well as SHAKE[35] constraints

to all bonds containing hydrogens. All peptide structures were initially minimized with 100 steps of steepest decent minimization and all MD simulations were performed in triplicate at 298K. Convergence was demonstrated by verifying that each titratable group sampled the same Ramachandran space with similar frequency during all three independent runs (Supplementary Figure 1).

2.2.3 Explicit Solvent Molecular Dynamics Simulations

For further inspection of solvation effects on amide protons and carbonyl carbons, 10 ns of dynamics was performed in explicit solvent utilizing the CHARMM36 force field[36]. The charges in the CHARMM 36 and CHARMM 22 force fields are identical, as are the peptide conformations which were fixed with restraints. The geometry optimized cluster structures (as described below in the cluster analysis section) were solvated in a TIP3P[37] cubic box. The convpdb.pl program from the MMTSB toolset[38] was utilized for this task. Counter ions were added to replicate an experimental salt concentration of 50mM NaCl and constraints were applied to all peptide atoms using the cons fix facility in CHARMM; this ensured that the representative conformations were kept fixed throughout the duration of the runs to correspond to the same conformational states that were used in the optimized quantum mechanical calculations of the chemical shifts. A non-bonded cutoff of 10 Å was applied, as well as SHAKE constraints to all bonds containing hydrogens. These runs were performed in triplicate for the protonated and deprotonated cluster representatives for each peptide.

2.2.4 K-Means Cluster Analysis

Analysis of the implicit solvent trajectories was carried out utilizing the CORREL functionality in CHARMM which computed select torsional angles (ϕ, ψ, χ) of the peptides as a function of time. The ACE and CT2 caps were included in the phi and psi angles of the terminal glycine residues.

The torsional time series was clustered using K-means clustering with a cluster radius cutoff of 100°, which was selected to identify large torsion angle fluctuations[39], the angle flag was employed, to take angle periodicity into account. For all calculations utilizing 'cluster structures', the structures closest in root-mean-square distance to the cluster centroids were selected as representatives of the corresponding subpopulations. The representative structures were given weights based on the number of members in the respective cluster population. As aforementioned, these structures were then used to perform the explicit MD simulations.

2.2.5 Calculation of Chemical Shifts

NMR chemical shielding tensors were computed using the GIAO-DFT methodology in the quantum chemical suite Gaussian 09[40]. Input structures were obtained either directly from the MD simulations or the clustering analysis. Geometry optimization was performed utilizing the B3LYP/6-31+G* level of theory. In order to preserve the desired peptide conformation, the ϕ, ψ, and χ angles used in the clustering analysis were constrained during the optimization procedure. In all of the calculations, solvent effects were incorporated with the use of the integral equation formalism polarizable continuum model (IEF-PCM)[41], specifying a solvent dielectric of 78. NMR chemical shielding tensors of the optimized structures were computed at the B3LYP/6-311++G** level of theory, which has been employed in studies of biomolecules to produce good agreement between computed and measured NMR chemical shifts[26,42]. Diffusivity was added to both of the basis sets to account for the anionic nature of the relevant deprotonated peptides. The CSPs computed for the histidine tautomers were averaged with an 80:20 ratio to favor the Nε tautomer, which has been shown to be more biologically prevalent[30.] The final number of frames for any given peptide is 1,000. NMR chemical shifts (Supplementary Tables 1,2,3) for the peptides were obtained by referencing computed shielding tensors to DSS (sodium 4,4-dimentyl-4-silapentane-1-sulfonate)

shielding tensors, which were calculated at the same level of theory (C^{13}: 186.52 ppm; H^{1}: 30.47 ppm). Nitrogen shielding tensors were referenced to ammonia (^{15}N: 232.77ppm), which was computed utilizing the same procedure as that of the peptides and DSS. From explicit solvent simulations, 100 frames evenly interspersed were selected from each simulation and all TIP3P water molecules within a 3Å radius around any atom in the corresponding peptide was extracted for DFT calculations. The structures containing the water molecules were not geometry optimized (as they were kept fixed throughout the MD simulation) and the chemical shifts were computed at the B3LYP/6-31+G* level of theory. The chemical shifts were then weighted by the representative cluster weights as described in the previous section. For the application of the Buckingham correction, the Buckingham equation[9] was used as shown in equation 1, where the electric field component of each amide proton was computed by CHARMM. An average over 1,000 frames from the explicit solvent simulations was used to compute the average E_z for each (de)protonated peptide. A H^N polarizability constant of 118 ppm*au was applied[17].

2.3 Results and Discussion

2.3.1 Conformational sampling of model peptides

Daggett and coworkers have demonstrated the impact of protonation on the ability for titratable residues to display a conformational preference within glycine and alanine penta-peptides [43,51]. Therefore, in the exploration of the protonation dependence of NMR CSPs, our study began by generating ensembles for each of the tripeptides in fully protonated or deprotonated states in order to account for conformational preferences. Figure 2 depicts the phi (ϕ) and psi (ψ) frequencies sampled for the respective titratable groups over the course of a 650 ns simulation. From the Ramachandran plots, it is apparent that all peptides predominantly sample the right-handed alpha helical conformation (ϕ, ψ: −60°, −45°).

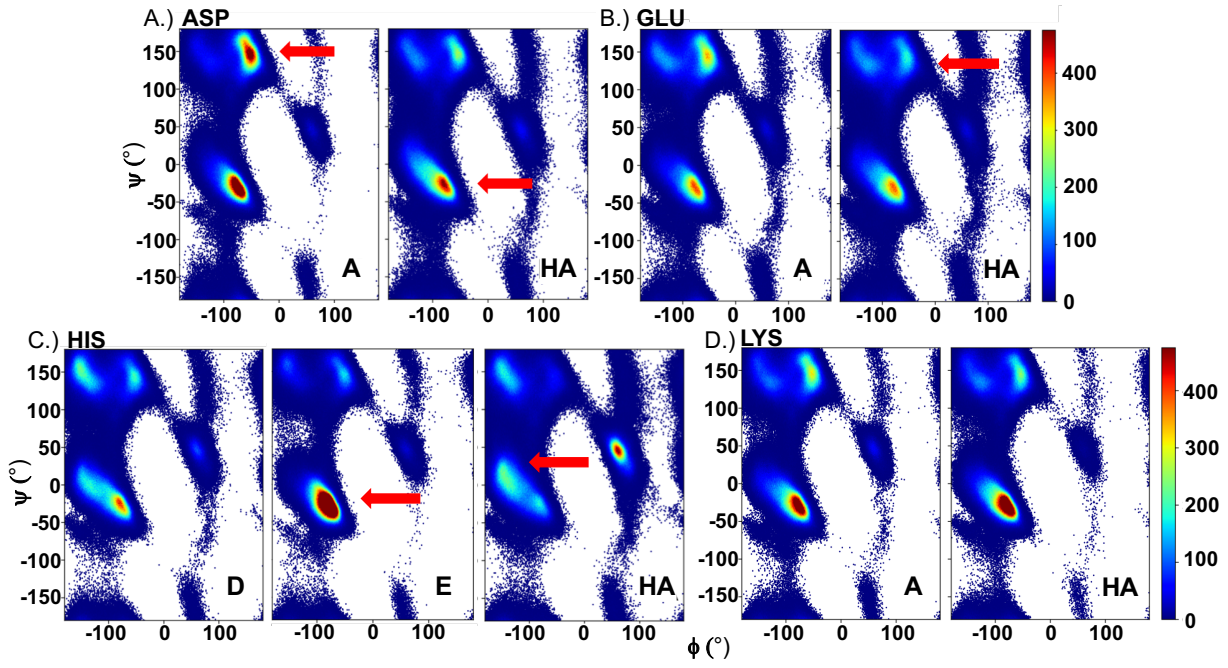


Figure 2.2 Ramachandran plots of titratable residues in tripeptides. The Ramachandran plots for the titratable groups in the peptide Ace-GXG-NH_2 are shown. The dihedral angles for 3 million frames are binned with a grid size of 475. Prominent shifts in sampling are indicated with red arrows. In all panel HA denotes a protonated peptide and A is unprotonated. By visual inspection, it is observed that the right-handed alpha helical region is primarily occupied. The shifts for ASP **(A)** and GLU **(B)** show depopulation of the beta strand region for the PPII space. HIS **(C)** indicates a large variation in sampling between the two tautomers, the Nδ tautomer (D in histidine figure, E shows the Nε tautomer) closely resembling the sampling of the fully protonated histidine. LYS **(E)** does not show any significant sampling variations.

This is to be expected as the tri-peptides are not true random coils[43,51]. Also evident, are the qualitative differences in the sampling between protonated and deprotonated forms of the ASP, GLU and HIS residues. In both states, aspartic acid primarily occupies the right-handed alpha helix followed by the less populated beta sheet and polyproline helical (type I or II) regions. Upon protonation however, there is a significant shift in the occupancy of the dihedral space, and the concentrated turn II population becomes equally dispersed between three states – the beta sheet, turn-II and left-handed alpha helix. An increase in the sampling of the right-handed alpha helix is observed, likely originating from interactions of the negatively charged carboxyl group with neighboring moieties, allowing for the higher prevalence and stabilization of the turn-II structure[43].

A similar trend is observed in the regions sampled by glutamic acid. The turn-I/turn-II regions are highly populated, followed by the beta-sheet space and the right handed alpha helix. The change in protonation shifts the polyproline-like character towards a wider sampling of the right handed helical region. This is anticipated as the side chain is chemically similar to aspartic acid. Presumably, the length of the chain due to the additional carbon adding allows for increased flexibility.

Figure 2 also demonstrates that the dihedral occupation of histidine is vastly different for the two tautomers. Specifically, the epsilon (N^{ε}) tautomer (indicated by E in Figure 2C) samples the right handed alpha helix with greater frequency than any other of the four titratable residues. This is in stark contrast to the delta (N^{δ}) tautomer (D in Figure 2C) which resembles the sampling of the fully protonated histidine, with the majority of the occupied dihedral space near the right-handed alpha helix. Lysine is the only residue which does not exhibit major sampling variations upon protonation. There is a slight increase in the polyproline II character of the deprotonated state, however the length of the side chain appears to diminish any conformational preference. Our results in sampling agree with the findings of Daggett[43] with the exception of the shift in protonated and deprotonated forms of glutamic acid where there is a shift towards the left-handed alpha helix upon protonation. The variation is likely attributable to differences in force fields and methodologies.

2.3.2 Computing pH dependent NMR chemical shift perturbations

In order to introduce dynamical averaging and recapitulate the sampling variations exhibited in the protonated and deprotonated structural ensembles, 1,000 frames were selected (1 frame every 650 picoseconds) from each MD trajectory for the calculation of absolute chemical shifts. The absolute

shifts for each frame were then averaged and the CSP for each nucleus was computed with the following equation

$$\Delta\delta = \delta_A - \delta_{HA} \quad (2)$$

where A and HA denote the deprotonated and protonated residue types, respectively.

Carbon Perturbations.

The results for the CSPs of the ^{13}C nuclei are displayed in Figure 3 and Table 1. At first glance, it is evident that the computed ^{13}C perturbations are significant, ranging from -2.5 ppm to 7.5 ppm. As expected, the largest CSPs are observed in the carbon atoms near the (de)protonation site, and generally all atoms in the titratable group. In the aspartic acid tri-peptide (Figure 3A), the computed deshielding is most evident in the ^{13}Cβ, ^{13}Cγ and ^{13}CO nuclei, which have a downfield shift of 6.9 ppm, 2.4 ppm and 3.2 ppm, respectively. The large perturbations in the backbone ^{13}Cα (3.5 ppm) and side chain ^{13}Cβ (6.9 ppm) atoms suggest notable differences in the secondary structures[4] of the deprotonated and protonated tri-peptide forms. These results mirror the Ramachandran plots in Figure 2, which also illustrate structural differences, suggesting that an average of 1,000 frames sufficiently captures the structural equilibria of the MD ensembles.

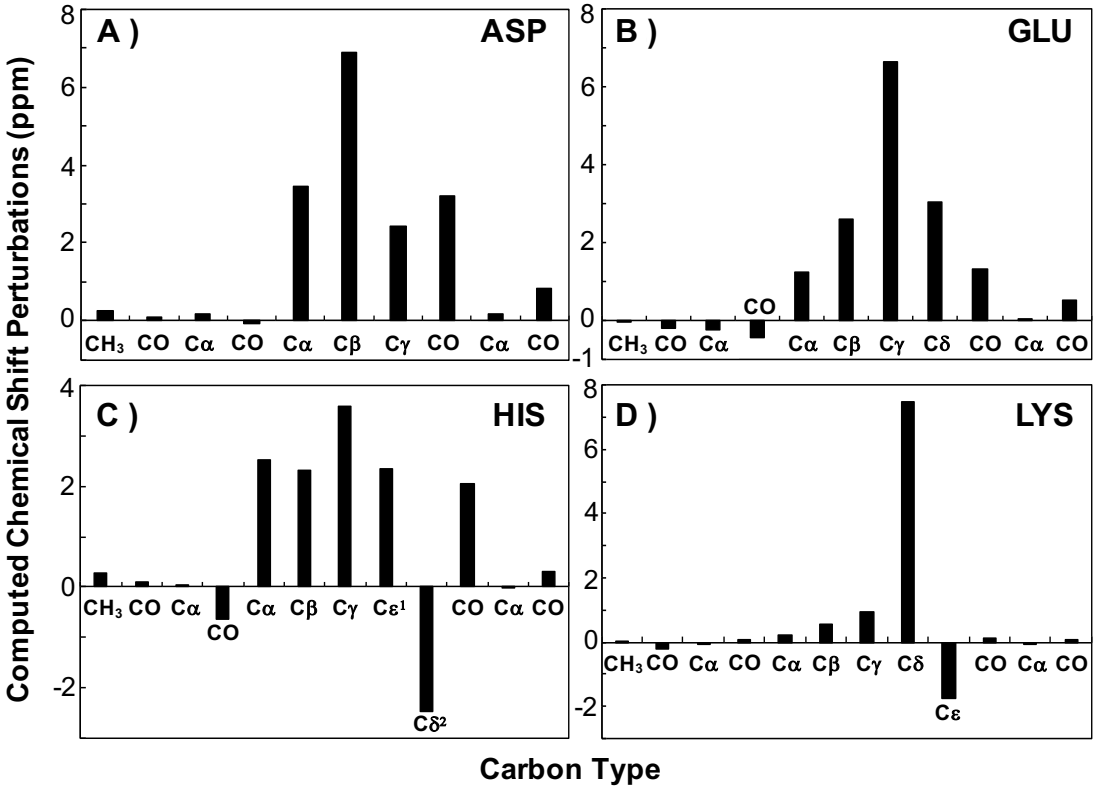


Figure 2.3 Computed ^{13}C pH-dependent CSPs. The computed CSPs for all carbon atoms in the tri-peptides are displayed. Each panel represents an average of 1,000 structural conformations for each protonation state. The residue name is indicated in each plot and the carbon types follow the peptide backbone starting with the ACE cap (CH3). Perturbations for this atom type range from -2.5 ppm to 7.5 ppm, with the largest shifts occurring on the titratable residue.

Of particular importance in pK_a determination is the CSP of the terminal carboxyl carbon, commonly known as a 'reporter nucleus'[30]. Its close proximity to the protonation site enables increased sensitivity for the presence or absence of the proton on the carboxylic oxygen. While the $^{13}C\gamma$ can also be influenced by torsional rotations (seen in solid-state NMR)[45] the majority of the perturbation is generally due to inductive charge effects. In Figure 3, it can be seen that the carbon directly bound to the carboxylic acid (Figure 3A: $^{13}C\beta$ or Figure 3B: $^{13}C\gamma$) experiences the largest protonation-dependent perturbation and not the carboxylic acid. The aspartic acid $^{13}C\gamma$ experiences a downfield CSP of 2.4 ppm upon protonation. In glutamic acid (Figure 3B), a similar trend for the carbon atoms is observed as the most prominent CSPs are seen in the $^{13}C\gamma$

(6.6 ppm) and the $^{13}C\delta$ (3.1 ppm), the two carbons directly adjacent to the titratable moiety. Analogous to the conformational relevance of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts, the shifts of the $^{13}C\gamma$ and $^{13}C\delta$ nuclei in the histidine imidazole ring are often indicative of the residue's tautomeric state[44]. This is exhibited in Figure 3C, as the $^{13}C\delta^2$ perturbation is shifted by -2.5 ppm which suggests a larger population of the N^{ε} tautomer[30]. The CSPs computed for the histidine tautomers were averaged with an 80:20 ratio to favor the N^{ε} tautomer, which has been shown to be more biologically prevalent[30]. This ratio was utilized to replicate the averaging protocol used in the Platzer et. al study[30], against which our computed CSPs are compared. Perhaps the most unambiguous CSP occurs for the $^{13}C\delta$ nucleus of the lysine tri-peptide (Figure 3D). Due to the lack of conformational difference between the charged and uncharged forms, as exhibited by the Ramachandran plots in Figure 2, the $^{13}C\delta$ nucleus, which is shifted by 7.5 ppm, is likely displaying a primarily charge-driven shift.

Generally, the computed CSPs are in good agreement with the measured CSPs (Table 1). The individual Pearson Correlation Coefficients (R) range from 0.83 to 0.90 for the four model peptides and collectively the ^{13}C CPSs have an average mean absolute error (MAE) of 0.73 ppm. This is a much lower MAE than those produced from the *ab intio* calculations of absolute ^{13}C chemical shifts, which can range from 1.5 to 3.4 ppm[42]. Identifying trends in the perturbations of absolute chemical shifts allows for the mitigation of systematic error and basis set dependence. Due to the greater sensitivity of quantum mechanical methods to the geometrical features of the input structure than empirical methods[47], it has been demonstrated that DFT calculations cannot

produce the same accuracy in computed chemical shifts that is observed from empirically derived shifts, when comparing to experimental values. Furthermore, although B3LYP has been successfully employed in previous NMR studies to compute chemical shifts, it is known that B3LYP tends to yield chemical shifts that are systemically biased toward higher fields carboxylic acid nuclei[11]. This phenomenon is apparent in our calculations as the average correlation of the ^{13}C CSPs (R = 0.85) significantly improves (R = 0.91) when not considering carboxylic ^{13}C atom.

Table 2.1 Average computed CSPs compared to experimental CSPs

^{15}N NMR chemical shift perturbations

Residue	R	RMSE (ppm)	MAE (ppm)
ASP	0.95	2.15	1.74
GLU	0.98	0.61	0.52
HIS	0.99	8.63	5.53
LYS	0.99	2.32	1.40

^{13}C NMR chemical shift perturbations

Residue	R	RMSE (ppm)	MAE (ppm)
ASP	0.83	1.57	0.99
GLU	0.86	1.19	0.79
HIS	0.89	0.77	0.52
LYS	0.90	1.10	0.60

$^{1}H^{*}$ NMR chemical shift perturbations

Residue	R	RMSE (ppm)	MAE (ppm)
ASP	0.99	0.10	0.07
GLU	0.98	0.07	0.04
HIS	0.91	0.14	0.10
LYS	0.99	0.67	0.26

*Excluding exchangeable protons

Table 2.1. Computed CSPs over an average from 1,000 frames are correlated to the experimental CSPs provided in Platzer *et al.*[30] Overall, correlations and low MAE values indicate good agreement between computed and experimental CSPs.

Moreover, the calculation of relative shift perturbations instead of absolute shifts allows for the capacity of generalization, as this suggests CSPs are indicative of global trends, which is important when attempting to enhance an existing predictor with pH sensitivity (minimizing the focus on

accuracy). Furthermore, the verification that ^{13}C atom types are highly sensitive to changes in pH enables the incorporation of pH-dependency in predictors that can only predict shifts for carbon atom types, such as CheShift[49] and LarmorD[50].

Nitrogen Perturbations

^{15}N nuclei are ubiquitous in pK_a measurements, as nitrogens are especially sensitive to solvent effects[53]. The computed ^{15}N CSPs are shown in Figure 4. Aspartic (Figure 4A) and glutamic acid (Figure 4B) have modest ^{15}N perturbations at 5.2 ppm and 2.0 ppm, respectively.

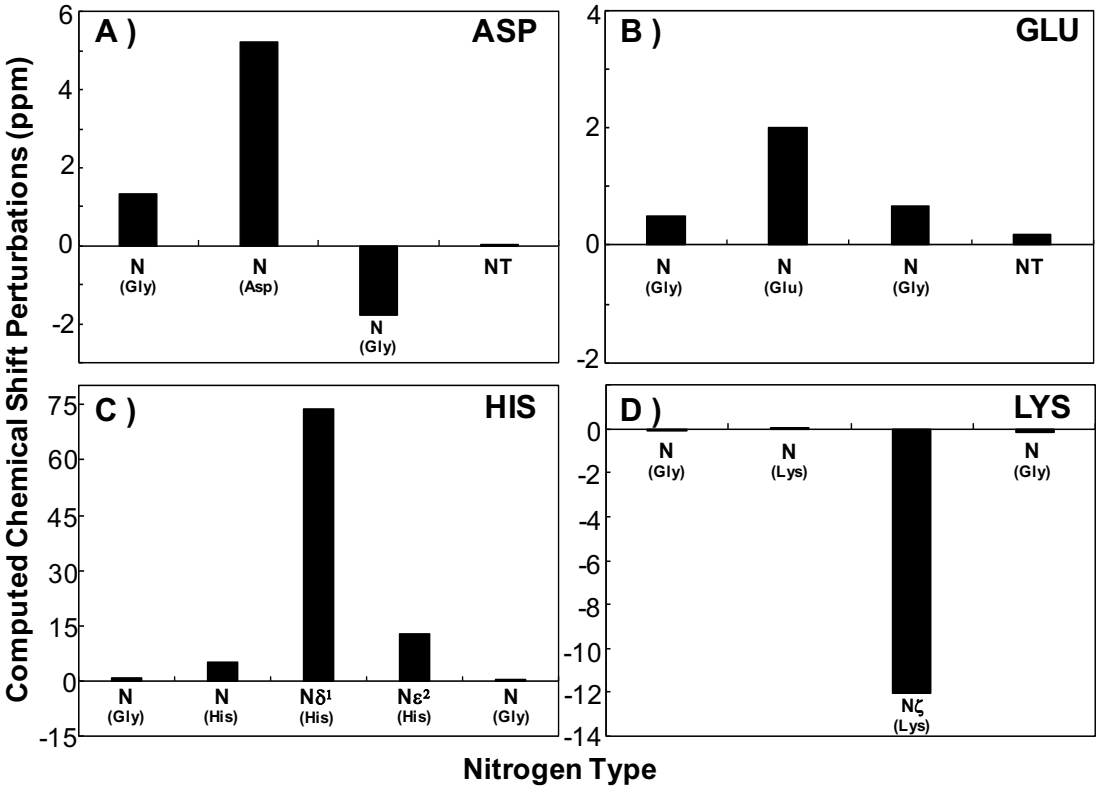


Figure 2.4. Computed ^{15}N pH-dependent CSPs. The computed CSPs for all nitrogen atoms in the tri-peptides are displayed. Each panel represents an average over 1,000 structural conformations for each protonation state. The residue name is indicated in each plot and the nitrogen types follow the peptide backbone starting with the glycine backbone nitrogen to the terminal NT_2 cap (only perturbations for which there is experimental data is displayed). Perturbations for this atom type range from -12.1 ppm to 73.8 ppm, with the largest shifts occurring on the titratable residue.

The increase in aspartic and glutamic acid ^{15}N shifts may be indicative of either an inter or intra-molecular hydrogen bond[30]. Histidine (Figure 4C) displayed a backbone ^{15}N perturbation of the

same magnitude as aspartic and glutamic acid at 5.3 ppm. However, the effect of protonation is most visible in the highly perturbed $^{15}N\delta^1$ resonance, which has a remarkable downfield shift of 73.8 ppm. The alternative protonation site, $^{15}N\varepsilon^2$, is only perturbed by 13.0 ppm. The backbone ^{15}N CSP of lysine is minimal at 0.06 ppm. As aforementioned, there is no significant conformational difference between the protonated and deprotonated peptide form for LYS, and the side chain's length precludes any intramolecular hydrogen bonding. The 'reporter nucleus' for lysine is the $^{15}N\zeta$ atom which exhibits an upfield shift of -12.07 ppm.

Typically, the inclusion of explicit solvent is necessary to compute accurate absolute chemical shifts for this atom type[52]. However, the coarse description of solvent (implicit solvent) that was used in the MD and DFT calculations, proved sufficient to elicit the pH-sensitivity of ^{15}N CSPs suggesting that the primary effect of the perturbation is due to through-bond charge effects. The average correlation of the ^{15}N perturbations is R = 0.99 between the computed and experimental CSPs (Table 1). The average MAE for the ^{15}N perturbations is 2.29 ppm and if the perturbed $^{15}N\delta^1$ resonance of the histidine residue, with a CSP of 73.8 ppm, is neglected the MAE is lower still at 1.22 ppm. This is because the experimental CSP for this atom type is 55.5 ppm[30], a difference of approximately 18 ppm, significantly skewing the computation of the mean average error. Despite this discrepancy in histidine, the overall pH-dependent trends considering all peptide models are very well captured. In comparing the carbon and nitrogen CSP distributions, it is apparent that the through-bond effects sharply diminish with distance. This is especially apparent in the lysine ^{15}N CSPs where the backbone ^{15}N is not meaningfully influenced by the protonation event.

Proton Perturbations

The computed proton perturbations were the least sensitive to pH and are shown in Figure 5. Only non-exchangeable protons (except for amide protons in the imidazole ring of histidine and the lysine side chain) are displayed since hydrogen bonding effects that predominantly influence the $^{1}H^{N}$ nuclei CSPs are not well described with implicit solvent models[50]. For aspartic and glutamic acid, most of the pertinent protons ($^{1}H\alpha$ and $^{1}H\beta$) were shifted upfield. The most pH-dependent CSPs were observed in the protons near the protonation site and were upshifted by a similar magnitude (~ -0.4 ppm) in the aspartic acid, glutamic acid and histidine peptides.

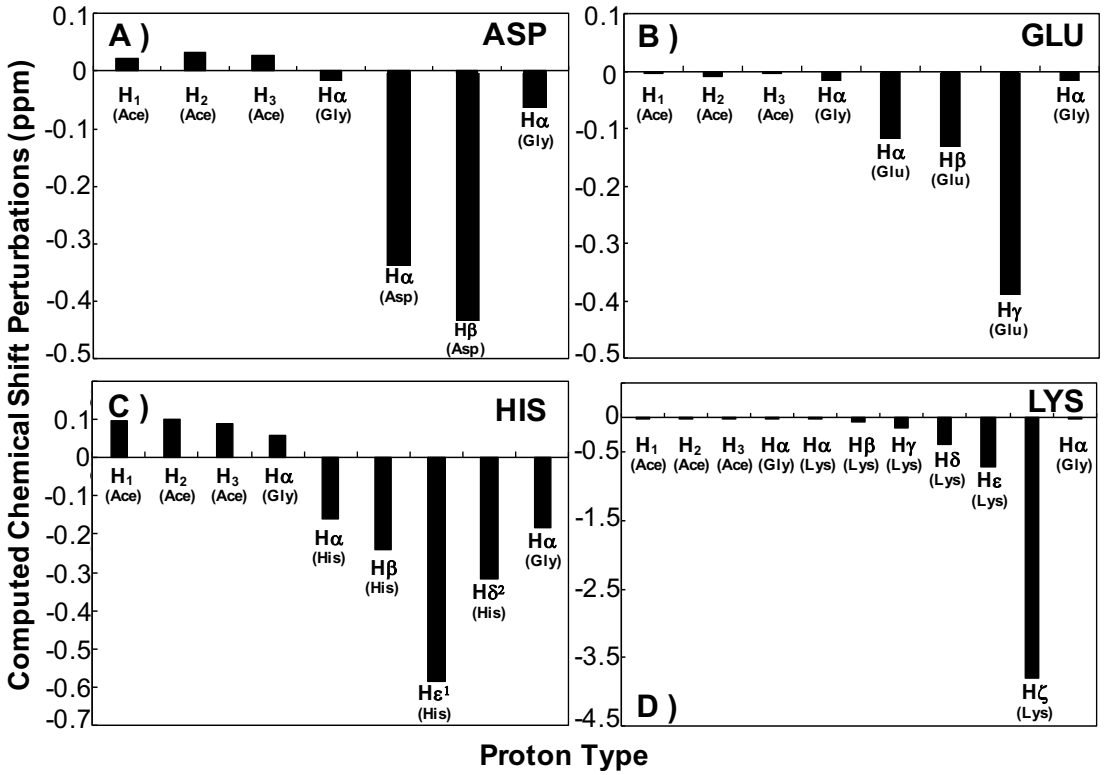


Figure 2.5 Computed ^{1}H pH-dependent CSPs. The computed CSPs for protons in the tri-peptides are displayed. Exchangeable protons are excluded due to their sensitivity to hydrogen bonding. Each panel represents an average of 1,000 structural conformations for each protonation state. The residue name is indicated in each plot and the proton types follow the peptide backbone starting with the Ace cap (only perturbations for which there is experimental data is displayed). Perturbations for this atom type range from -3.8ppm to 0.10 ppm, with the largest shifts occurring on the titratable residue.

The protons on the imidazole ring, particularly $^{1}H\varepsilon^{1}$ and $^{1}H\delta^{2}$ displayed perturbations of -0.58 ppm and -0.31 ppm reproducing the tautomeric trends of the ^{15}N CSPs. Lysine proton perturbations are also indicative of protonation, especially for the $^{1}H\zeta$ which had a shift of -3.8 ppm. It is apparent from Figure 5D that the lysine protons $^{1}H\delta$ and $^{1}H\varepsilon$ also experience pronounced perturbations.

Agreement with experimental values for non-exchangeable protons was generally good with an average correlation of R = 0.98. Most of the protons are highly dependent on secondary structure ($^{1}H\alpha$ and $^{1}H\beta$) and can therefore be used to confirm the sampling in the MD ensembles. Most germane for pH sensitivity are the amide protons, which in addition to describing solvent effects and hydrogen bond formation, tend to rapidly exchange. As previously discussed, although the granularity of the GBSW/PCM solvent approximation does not adversely impact the computation of the ^{15}N CSPs, the exceptional sensitivity of amide protons requires the incorporation of explicit solvent or a long-range electrostatic correction[52] which will be discussed in a subsequent section.

2.3.3 K-means clustering of dihedral space for model peptides

In an attempt to reduce the overall computational cost of the DFT calculations, K-means clustering was used to classify the conformational distributions of the MD ensemble. By utilizing K-means clustering the MD ensemble can be described by a significantly condensed number of structures (12-35 frames). The small number of representative structures can then be weighted according to cluster member population, and therefore can be used to reproduce the conformational distribution of the ensemble. This method was validated by computing the weight-averaged CSPs for the clustered structures. The largest contributing structures represent 5-19% of the general population and, for the majority of the peptides, reproduce the qualitative structures observed in the Ramachandran plots (Figure 2). More importantly, the weighted-average CSPs of the peptide

clusters display reasonable agreement to the original computed CSPs (which CSPs of 1,000 frames were averaged) as well as the experimental results as is seen in Table 2. Because dynamical averaging is crucial in computing accurate NMR chemical shifts[52], there is a slight drop in the Pearson correlation coefficient for the clustered CSPs. Clustering is an important step as it enables a better understanding of the conformational ensemble and the ability to independently tune the weights of each representative to better match experimental CSP observations.

Table 2.2 Computed pH dependent CSPs of clusters compared to experimental CSPs

^{15}N NMR chemical shift perturbations

Residue	R	RMSE (ppm)	MAE (ppm)
ASP	0.80	4.10	3.62
GLU	0.84	1.31	1.10
HIS	0.99	8.21	5.11
LYS	0.99	2.02	1.40

^{13}C NMR chemical shift perturbations

Residue	R	RMSE (ppm)	MAE (ppm)
ASP	0.72	2.18	1.62
GLU	0.87	1.08	0.80
HIS	0.79	1.40	0.99
LYS	0.91	1.46	0.99

^{1}H* NMR chemical shift perturbations

Residue	R	RMSE (ppm)	MAE (ppm)
ASP	0.87	0.23	0.18
GLU	0.90	0.06	0.06
HIS	0.80	0.18	0.16
LYS	0.99	0.67	0.31

*Excluding exchangeable protons

Table 2.2 Weighted averages of the CSPs computed for the clustered structures are correlated to the experimental CSPs provided in Platzer *et al*[30]. Correlations indicate that clustering with a radius of 100° is sufficient to reproduce the MD ensemble for the calculation of CSPs.

2.3.4 Examination of Through-Space Effects

In investigating pH-dependent CSPs thus far, we have focused only on non-exchangeable protons. However, due to the amide proton's exquisite sensitivity to its environment, these nuclei tend to play an important role in the pK_a determination by NMR spectroscopy. Since our calculations utilize implicit solvation models, it was not surprising to see that the inclusion of computed exchangeable proton CSPs significantly decreases the overall ^1H CSP correlation coefficient from R = 0.99 to R=0.54. Incorporation of through-space effects is therefore necessary to improve the computed CSPs of this nucleus type. We first examined the impact of adding explicit water by solving the Buckingham equation (BEQ). Explicit solvent MD simulations were performed for the representative structures determined by the clustering procedure described in the methods section. During the 10 ns simulations, structures were harmonically restrained and only water molecules were allowed to move. The electrostatic interaction energy between the water molecules and the amide proton of interest was computed and the difference in the E_z component between the protonated and deprotonated form of the respective peptide was used in the Buckingham equation. The Buckingham equation accounts predominantly for through-space effects, which are the overriding effects described by the ^1HN perturbations. The CSPs derived from the Buckingham formalism were then added to the CSPs computed by DFT: thereby, accounting for both through-bond and through-space contributions; the results of which are presented in Table 3. It is evident that the correlation coefficient for the amide proton CSPs computed in implicit solvent is quite poor at R = -0.57 compared to experiment. Addition of the BEQ correction slightly improves the correlation to R = -0.44, but does not mitigate the inadequate representation of the amide proton CSPs and suggests that the predominant variable causing the ^1HN perturbations is not a simple electrostatic through-space effect.

When examining the computed absolute chemical shifts of the exchangeable protons, it is observed that the average ^{1}HN absolute chemical shift is approximately 4.5 ppm. This suggests that it is highly unlikely for the majority of the amide protons in the tri-peptide ensembles to be participating in any sort of hydrogen bonding[52] (in this case only intramolecular hydrogen bonding is available as there are no explicit water molecules present). It has been demonstrated by Exner and coworkers[52] that an amide proton engaged in hydrogen bonding may exhibit an absolute chemical shift between 6 and 8 ppm depending of the strength of the bond (i.e. angle, distance)[52]. Hence, to improve the accuracy of the amide proton CSPs, the chemical shifts of the representative structures were recomputed with the addition of a 4Å radius hydration shell around the peptide. The inclusion of explicit water molecules in the DFT calculations allows for a better description of the charge transfer that occurs during a hydrogen bonding. As can be observed in Table 3, the explicit calculations of CSP increase the correlation significantly to R = 0.67, emphasizing the impact of hydrogen bonding even when the degree of solvent averaging utilized in the calculations is small (100 configurations of solvent snapshots from each cluster simulation were used in our quantum mechanical calculations).

Table 2.3 Computed [1]HN CSPs for ASP, GLU and HIS

HN Type	Exp[30]	Implicit	Explicit	Implicit + BEQ
GLY 1	0.02	0.00	0.01	-0.23
ASP 2	-0.17	0.35	-1.03	0.29
GLY 3	-0.03	0.00	0.03	0.17
GLY 1	0.00	0.00	0.01	-0.26
GLU 2	0.12	0.49	-0.18	0.64
GLY 3	0.02	0.36	-0.06	0.68
GLY 1	-0.01	0.19	-0.18	-0.07
HIS 2	-0.20	1.63	-0.39	1.59
R		**-0.57**	**0.67**	**-0.44**

Table 2.3 Weighted CSPs computed for backbone amide protons of ASP, GLU, and HIS compared to available experimental CSPs provided by Platzer *et al*[30]. Implicit denotes DFT calculated CSPs with the utilization of implicit solvent models, and explicit indicates the presence of water molecules in the DFT calculation. BEQ is an abbreviation for the Buckingham equation which was used to compute the through-space contribution and was subsequently added to the implicit CSPs.

Many available chemical shift predictors already utilize a combination of the Buckingham formulism along with an empirically derived hydrogen bonding correction in order to ensure valid HN predictions.

2.4 Conclusion

Using model tri-peptides, we have demonstrated that a combination of molecular dynamics and quantum mechanics allows for the recapitulation of trends observed in experimental pH-dependent NMR chemical shift perturbations. In the context of pH, we simplified chemical shift influences as either through-space or through-bond. Molecular dynamics ensembles of the tri-peptides provide insight into the through-space contributions of charge dependent conformational states. Although the most densely populated dihedral space observed, was that of the right-handed alpha helix - aspartic acid, glutamic acid and histidine sampled alternative subpopulations depending on the side chain's charge. Our findings agree with similar studies illustrating variation in the dihedral distribution of model peptides in similar constructs[43,51].

By considering the conformational variability through dynamical averaging, the observed conformational preference is manifested in the computed CPSs. For the averaged ensemble, ^{13}C, ^{15}N and ^{1}H nuclei had computed CSPs which correlated quite well with experimental values (correlation of R = 0.85, 0.99, 0.98 respectively). Furthermore, the computed MAE and RSME was correspondingly low compared to *ab initio* studies which compare absolute chemical shifts to experimental chemical shifts[42]. Clustering the ensembles allows further insight into the types of secondary structural features that are dominating the through-space contribution of the computed chemical shift – and can serve to reproduce the experimental ensemble by tuning the various cluster weights. More importantly, computing perturbations allows for the identification of systematic trends - for example, in our calculations we observe an average ^{13}C perturbation of approximately 4.5 ppm for aspartic and glutamic acids upon protonation. Interestingly, experimental studies on proteins and peptides report similar ^{13}C chemical shift perturbations for aspartic and glutamic acid at ~3-4 ppm[19,22]. Furthermore, it has been shown that the amide ^{15}N in histidine shifts on the order of ~73ppm[30]. Despite the conformational environment, particular nuclei exhibit consistent perturbations indicating that the through-bond or electronic structure effects due to protonation are the predominant influence when computing pH-dependent CSPs. Not only may these through-bond contributions surpass those of through-space variations in magnitude, but they are predictable (i.e. systematic). Table 4 illustrates this point by indicating generally good agreement with CSPs measure in proteins compared to those seen in peptides[16,22].

Table 2.4. Comparison of Peptide CSPs with Reported Protein CSPs

Residue Type	Atom Type	Protein CSPs[16,22] (ppm)	Peptide CSPs[30] (ppm)	Computed CSPs (ppm)
ASP	CA	1.43	1.4	3.5
	CB	2.3, 3	3	6.9
	CG	3.12, 4	3.2	2.4
	N	1.32	1.5	5.2
	HA	0.12	-0.17	-0.3
	HB	0.25	-0.23	-0.4
GLU	CG	4	3.5	6.6
	CD	3.74, 4	4.1	3.1
	N	0.67	1	2
	HA	0.07	-0.1	-0.1
	HB	0.22	-0.06	-0.1
	HG	0.24	-0.22	-0.4
HIS	CG	2.6	4.2	3.6
	CE1	2.15	2.6	2.4
	N	2.86	1.8	5.3
	ND1	73	56	73
	NE2	2.78, -9	8	13
	HB	0.66	-0.17	-0.2
	HD2	0.33	-0.33	-0.3
	HE1	0.86	-0.92	-0.6
LYS	CE	1.27	1	-1.8
	N	0.65	0.7	0
	NZ	-7	-7.5	-12.1

Table 2.4 Above are the typical heteronuclear pH-dependent chemical shifts compiled from pH titrations of proteins[16,22] compared to those observed experimentally and computationally in peptides.

The complexity of pH cannot be detangled into a single variable as its impact largely permeates the other chemical shift contributors (ie. conformational variation, hydrogen bonding, electrostatics, etc.). Here we demonstrate that in addition to the electrostatic contribution, which accounts for through-space effects, it would be prudent to include a separate variable which describes the inductive effect that is observed when protonation occurs. Hence, when augmenting chemical shift predictors with the ability to consider pH dependent chemical shifts an approach

where δ_{QM} is the chemical shift perturbation for the particular nucleus and residue, scaled by the pK_a of the residue, in addition to the through-space contribution denoted by the Buckingham equation may provide a more complete description of the impact of pH.

Given the complexity of pH-dependent behavior in proteins, implementation of pH effects in chemical shift prediction paradigms is a difficult undertaking. Ultimately, the accuracy of the predictions depends on the ability to reproduce the conformational ensemble of the protein. Furthermore, important considerations include the accurate prediction of pKa values for which CSP corrections may be applied, as well as hydrogen bonding interactions which play a special role in the perturbations of amide protons and nitrogens.

The ability to compute pH-dependent CSPs has the potential to deconvolute NMR spectra, aid in the process of structure refinement and allow for a qualitative metric by which we can use theoretical models to inform experiment.

2.5 Bibliography

1. Cavalli, A., Salvatella, X., Dobson, C.M. and Vendruscolo, M., 2007. Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences*, *104*(23), pp.9615-9620.

2. Kay, L.E., 2011. NMR studies of protein structure and dynamics. *Journal of magnetic resonance*, *213*(2), pp.477-491.

3. Wishart, D.S., 2011. Interpreting protein chemical shift data. *Progress in nuclear magnetic resonance spectroscopy*, *58*(1-2), p.62.

4. Spera, S. and Bax, A., 1991. Empirical correlation between protein backbone conformation and C. alpha. and C. beta. 13C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society*, *113*(14), pp.5490-5492.

5. Wagner, G., Pardi, A. and Wuthrich, K., 1983. Hydrogen bond length and proton NMR chemical shifts in proteins. *Journal of the American Chemical Society*, *105*(18), pp.5948-5949.

6. Wang, Y. and Jardetzky, O., 2002. Investigation of the neighboring residue effects on protein chemical shifts. *Journal of the American Chemical Society*, *124*(47), pp.14075-14084.

7. Haigh, C.W. and Mallion, R.B., 1979. Ring current theories in nuclear magnetic resonance. *Progress in nuclear magnetic resonance spectroscopy*, *13*(4), pp.303-344.

8. Sitkoff, D. and Case, D.A., 1998. Theories of chemical shift anisotropies in proteins and nucleic acids. *Progress in nuclear magnetic resonance spectroscopy*, *32*(2), pp.165-190.

9. Buckingham, A.D., 1960. Chemical shifts in the nuclear magnetic resonance spectra of molecules containing polar groups. *Canadian Journal of Chemistry*, *38*(2), pp.300-307.

10. Han, B., Liu, Y., Ginzinger, S.W. and Wishart, D.S., 2011. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of biomolecular NMR*, *50*(1), p.43.

11. Shen, Y. and Bax, A., 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR*, *48*(1), pp.13-22.

12. Kohlhoff, K.J., Robustelli, P., Cavalli, A., Salvatella, X. and Vendruscolo, M., 2009. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *Journal of the American Chemical Society*, *131*(39), pp.13894-13895.

13. Xu, X.P. and Case, D.A., 2001. Automated prediction of 15N, 13Cα, 13Cβ and 13C′ chemical shifts in proteins using a density functional database. *Journal of biomolecular NMR*, *21*(4), pp.321-333.

14. Larsen, A.S., Bratholm, L.A., Christensen, A.S., Channir, M. and Jensen, J.H., 2015. ProCS15: a DFT-based chemical shift predictor for backbone and Cβ atoms in proteins. *PeerJ*, *3*, p.e1344

15. Markley, J.L., Ulrich, E.L., Berman, H.M., Henrick, K., Nakamura, H. and Akutsu, H., 2008. BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *Journal of biomolecular NMR*, *40*(3), pp.153-155.

16. Farrell, D., Miranda, E.S., Webb, H., Georgi, N., Crowley, P.B., McIntosh, L.P. and Nielsen, J.E., 2010. Titration_DB: Storage and analysis of NMR-monitored protein pH titration curves. *Proteins: Structure, Function, and Bioinformatics*, *78*(4), pp.843-857.

17. Hass, M.A., Ringkjøbing Jensen, M. and Led, J.J., 2008. Probing electric fields in proteins in solution by NMR spectroscopy. *Proteins: Structure, Function, and Bioinformatics*, *72*(1), pp.333-343.

18. Kukić, P., Farrell, D., Søndergaard, C.R., Bjarnadottir, U., Bradley, J., Pollastri, G. and Nielsen, J.E., 2010. Improving the analysis of NMR spectra tracking pH-induced conformational changes: Removing artefacts of the electric field on the NMR chemical shift. *Proteins: Structure, Function, and Bioinformatics*, *78*(4), pp.971-984.

19. Kukic, P., Farrell, D., McIntosh, L.P., García-Moreno E, B., Jensen, K.S., Toleikis, Z., Teilum, K. and Nielsen, J.E., 2013. Protein dielectric constants determined from NMR chemical shift perturbations. *Journal of the American Chemical Society*, *135*(45), pp.16968-16976.

20. Gandhi, N.S., Landrieu, I., Byrne, C., Kukic, P., Amniai, L., Cantrelle, F.X., Wieruszeski, J.M., Mancera, R.L., Jacquot, Y. and Lippens, G., 2015. A Phosphorylation-Induced Turn Defines the Alzheimer's Disease AT8 Antibody Epitope on the Tau Protein. *Angewandte Chemie International Edition*, *54*(23), pp.6819-6823.

21. Boyd, J., Domene, C., Redfield, C., Ferraro, M.B. and Lazzeretti, P., 2003. Calculation of Dipole-Shielding Polarizabilities (σ αβγ I): The Influence of Uniform Electric Field Effects on the Shielding of Backbone Nuclei in Proteins. *Journal of the American Chemical Society*, *125*(32), pp.9556-9557.

22. Hass, M.A. and Mulder, F.A., 2015. Contemporary NMR studies of protein electrostatics. *Annual review of biophysics*, *44*, pp.53-75.

23. Yu, Z., Li, P. and Merz Jr, K.M., 2017. Using ligand-induced protein chemical shift perturbations to determine protein–ligand structures. *Biochemistry*, *56*(18), pp.2349-2362.

24. Ditchfield, R., 1972. On molecular orbital theories of NMR chemical shifts. *Chemical Physics Letters*, *15*(2), pp.203-206.

25. De Dios, A.C., Pearson, J.G. and Oldfield, E., 1993. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science*, *260*(5113), pp.1491-1496.

26. Xu, X.P. and Case, D.A., 2002. Probing multiple effects on 15N, 13Cα, 13Cβ, and 13C′ chemical shifts in peptides using density functional theory. *Biopolymers: Original Research on Biomolecules*, *65*(6), pp.408-423.

27. Vila, J.A., Aramini, J.M., Rossi, P., Kuzin, A., Su, M., Seetharaman, J., Xiao, R., Tong, L., Montelione, G.T. and Scheraga, H.A., 2008. Quantum chemical 13Cα chemical shift calculations for protein NMR structure determination, refinement, and validation. *Proceedings of the National Academy of Sciences*, *105*(38), pp.14389-14394.

28. Prestegard, J.H., Sahu, S.C., Nkari, W.K., Morris, L.C., Live, D. and Gruta, C., 2013. Chemical shift prediction for denatured proteins. *Journal of biomolecular NMR*, *55*(2), pp.201-209.

29. Schwarzinger, S., Kroon, G.J., Foss, T.R., Wright, P.E. and Dyson, H.J., 2000. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *Journal of biomolecular NMR*, *18*(1), pp.43-48.

30. Platzer, G., Okon, M. and McIntosh, L.P., 2014. pH-dependent random coil 1 H, 13 C, and 15 N chemical shifts of the ionizable amino acids: a guide for protein pK a measurements. *Journal of biomolecular NMR*, *60*(2-3), pp.109-129.

31. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.A. and Karplus, M., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), pp.187-217.

32. Best, R.B., Zhu, X., Shim, J., Lopes, P.E., Mittal, J., Feig, M. and MacKerell Jr, A.D., 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *Journal of chemical theory and computation*, *8*(9), pp.3257-3273.

33. Buck, M., Bouguet-Bonnet, S., Pastor, R.W. and MacKerell Jr, A.D., 2006. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophysical journal*, *90*(4), pp.L36-L38.

34. Im, W., Lee, M.S. and Brooks III, C.L., 2003. Generalized born model with a simple smoothing function. *Journal of computational chemistry*, *24*(14), pp.1691-1702.

35. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, *23*(3), pp.327-341.

36. Huang, J. and MacKerell Jr, A.D., 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry*, *34*(25), pp.2135-2145.

37. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, *79*(2), pp.926-935.

38. Feig, M., Karanicolas, J. and Brooks III, C.L., 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling*, *22*(5), pp.377-395.

39. Karpen, M.E., Tobias, D.J. and Brooks III, C.L., 1993. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: analysis of 2.2-ns trajectories of YPGDV. *Biochemistry*, *32*(2), pp.412-420.

40. Frisch, M.J.E.A., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.E. and Nakatsuji, H., 2014. Gaussian~ 09 Revision D. 01.

41. Cossi, M., Barone, V., Mennucci, B. and Tomasi, J., 1998. Ab initio study of ionic solutions by a polarizable continuum dielectric model. *Chemical Physics Letters*, *286*(3-4), pp.253-260.

42. Frank, A., Möller, H.M. and Exner, T.E., 2012. Toward the quantum chemical calculation of NMR chemical shifts of proteins. 2. Level of theory, basis set, and solvents model dependence. *Journal of chemical theory and computation*, *8*(4), pp.1480-1492.

43. Childers, M.C., Towse, C.L. and Daggett, V., 2016. The effect of chirality and steric hindrance on intrinsic backbone conformational propensities: tools for protein design. *Protein Engineering, Design and Selection*, *29*(7), pp.271-280.

44. Abraham, R. and Warne, M., 1997. Proton chemical shifts in NMR. Part 8. 1 Electric field effects and fluorine substituent chemical shifts (SCS). *Journal of the Chemical Society, Perkin Transactions 2*, (2), pp.203-208.

45. Sun, H. and Oldfield, E., 2004. Tryptophan chemical shift in peptides and proteins: a solid state carbon-13 nuclear magnetic resonance spectroscopic and quantum chemical investigation. *Journal of the American Chemical Society*, *126*(14), pp.4726-4734.

46. Pelton, J.G., Torchia, D.A., Meadow, N.D. and Roseman, S., 1993. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Science*, *2*(4), pp.543-558.

47. Sumowski, C.V., Hanni, M., Schweizer, S. and Ochsenfeld, C., 2014. Sensitivity of ab initio vs empirical methods in computing structural effects on NMR chemical shifts for the example of peptides. *Journal of chemical theory and computation*, *10*(1), pp.122-133.

48. Vila, J.A., Arnautova, Y.A., Martin, O.A. and Scheraga, H.A., 2009. Quantum-mechanics-derived 13Cα chemical shift server (CheShift) for protein structure validation. *Proceedings of the National Academy of Sciences*, *106*(40), pp.16972-16977.

49. Frank, A.T., Law, S.M. and Brooks III, C.L., 2014. A simple and fast approach for predicting 1H and 13C chemical shifts: toward chemical shift-guided simulations of RNA. *The Journal of Physical Chemistry B*, *118*(42), pp.12168-12175.

50. Zhu, T., Zhang, J.Z. and He, X., 2013. Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model. *Journal of chemical theory and computation*, *9*(4), pp.2104-2114.

51. Beck, D.A., Alonso, D.O., Inoyama, D. and Daggett, V., 2008. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. *Proceedings of the National Academy of Sciences*, *105*(34), pp.12259-12264.

52. Exner, T.E., Frank, A., Onila, I. and Möller, H.M., 2012. Toward the quantum chemical calculation of NMR chemical shifts of proteins. 3. Conformational sampling and explicit solvents model. *Journal of chemical theory and computation*, *8*(11), pp.4818-4827.

53. Witanowski, M., Sicińska, W., Grabowski, Z. and Webb, G.A., 1990. Study of solvent effects on the nitrogen NMR shieldings of some indolizines. *Magnetic resonance in chemistry*, *28*(11), pp.988-993.

54. Wishart, D.S., Sykes, B.D. and Richards, F.M., 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of molecular biology*, *222*(2), pp.311-333.

CHAPTER 3

Protein NMR Chemical Shifts

3.1 Introduction

Protein electrostatics and the internal organization of charge are crucial in many observed biological phenomena. The ability of titratable residues to respond to environmental pH allows for processes including enzyme catalysis[1], nucleation of aggregation seeds[2], activation of chaperones[3] pH-dependent switching mechanisms[4] and membrane insertion[5]. Understanding the influence of individual titratable residues directly allows for greater insight into how electrostatics facilitate many important biological mechanisms. pK_a values describe the ionization behavior of titratable amino acids and inform aspects of local protein conformation, solvent exposure and neighboring protonation equilibria[5]. Measuring pK_a values experimentally is often accomplished by NMR spectroscopy which is used to elucidate the complex microenvironment of proteins.

In order to circumvent the high cost of carbon nuclei labeling, NMR chemical shifts of amide nitrogen and proton atoms are most commonly monitored during a pH titration and fit to the modified Henderson-Hasselbach equation[6]. This process enables amide nitrogen and proton nuclei to report on the protonation event of the corresponding titratable sidechain and result in a consensus pK_a value for the residue. Although these atoms are commonly referred to as 'reporter nuclei'[5], they do not unambiguously report on the protonation events of their respective sidechains, as chemical shift perturbations due to conformational change, hydrogen bonding, proximity to an aromatic group or a neighboring titration[7] may concurrently arise from the change in

environmental pH[5]. The effects which accompany the protonation event act to perturb the observable chemical shift resulting in a macroscopic pK_a value. The pK_a observed from the NMR pH-titrations therefore is a description of the macroscopic electrostatic environment, as it quite difficult to detangle the various contributors of the measured chemical shift perturbations[7,8].

To compute the microscopic pK_a of a residue, computational methods such as constant pH molecular dynamics[9-11](CpHMD) may be used. CpHMD has been successfully employed in many studies to investigate the impact of pH on protein structure and dynamics[12-14], performing reasonably well when compared to experiment. Briefly, this methodology utilizes a continuous variable, lambda[9], which is allowed to fluctuate between the bounds of 0 and 1[9] representing the protonated and deprotonated states of a titratable residue[10]. The lambda values are coupled to the dynamic steps of the simulation[11], such that both protonation and conformation can be probed simultaneously[10]. However, because lambda is a continuous variable, unphysical intermediate states may occur and must be disregarded in the final calculations using a lambda cutoff (typically $l > 0.80$ is used)[11]. CpHMD^{MSLD} relies on a hybrid Hamiltonian[10] which describes the number of titrating residues, the protein environment and the coordinates of protonated and deprotonated sites – a complete expression of its potential energy equation can be found in references 9 to 11. The free energy of protonation for model compounds[10] (titratable residues: aspartic acid, glutamic acid, histidine and lysine), is used to calibrate the environmental pH of the simulation in the form of fixed biases[10]. Furthermore, variable biases can be altered to increase sampling of specific protonation sites[10]. During these simulations, pH replica exchange[15] may also be incorporated to further explore the conformational landscape.

In order to directly compare the microscopic pK_a computed from the fraction of uprotonated states in the CpHMD^{MSLD} simulation and the macroscopic pK_a from NMR

experiments, chemical shifts must be computed from the conformational ensembles present in the simulation. However, many of the available chemical shift predictors[16-18] do not compute pH-dependent NMR chemical shifts as there is a scarcity of pH titration experiments in the BMRB[19] on which to train machine learning algorithms.

In this work, we implement a protocol to compute the pH-dependent NMR chemical shifts of the conformational ensembles produced by CpHMD^{MSLD}. Utilizing SPARTA+[17] in conjunction with previously computed pH-dependent chemical shift perturbations of model peptides[20] and the Buckingham equation[21], macroscopic pK_a values from simulation can be directly compared to those from NMR pH-titration experiments. This methodology allows for deeper insight of the computational ensemble and better understanding of the protein's electrostatic environment.

3.2 Methodology

3.2.1 Structure Preparation

The starting configuration for hen egg white lysozyme (HEWL) was retrieved from the Ramanadham *et al*. crystal structure[22] in the protein data bank (accession code 2LZT). After crystallographic waters and crystallization solutes were deleted, hydrogen atoms were added with the HBUILD functionality of the CHARMM[23] macromolecular simulation package. Subsequently, the convpdb.pl facility in the MMTSB[24] toolset was utilized to solvate the protein in a 71Å cubic box of TIP3P water[25]. After charge neutralization, the appropriate number of sodium and chloride ions was added to the system, matching the experimental salt concentration of 50mM[6]. The CHARMM36 all-atom protein force field[26] with CMAP correction[27] was applied within CHARMM to cap the protein's N- and C-termini. Additionally, residues ASP, GLU, and HIS were patched to allow for titration and cysteine residues were patched to form disulfide bonds. The system was minimized by 50 steps of steepest descent followed by the adopted basis Newton–Raphson method for 2000 steps.

3.2.2 Explicit Solvent CpHMD^{MSLD} Simulations of HEWL

Explicit solvent CpHMD^{MSLD} with pH replica exchange (pH-REX) was performed in CHARMM, implementing modules BLOCK and REPDSTR. All simulations utilize the CMAP corrected CHARMM36 all-atom protein force field and the CHARMM TIP3P water model. Residues ASP, GLU, and HIS were allowed to sample continuously between protonated and deprotonated states via the continuous variable λ[10]. Lysine was outside of the pH range for which NMR data was available and therefore remained protonated throughout the duration of the simulation. The SHAKE algorithm[28] was used on hydrogen-heavy atom bond lengths and the simulations advanced at a 2-femtosecond time step. A nonbonded cutoff of 15 Å with electrostatic force and van der Waals force switching functions were applied. A Langevin heat bath maintained the temperature at 298K and used a frictional coefficient of 10 ps^{-1}. Ten pH replica windows were used to simulate pH ranges 1 to 10. The exchange in pH was attempted every 500 dynamics steps and each replica was simulated for 20ns after a 1ns equilibration. All biases and parameters were set to reproduce the protocol presented in Goh *et. al*[10]. All simulations were replicated five times.

3.2.3 Calculation of pK_a Values

Lambda values extracted from each pH-replica were used to determine the populations of protonated and deprotonated states. However, because lambda is a continuous variable, only lambda values above a cutoff of 0.80 are considered to be describing physical states[10,11]. The fraction of unprotonated states for a given residue was calculated (S_{un}) and the pK_a was determined using the generalized Henderson-Hasselbach equation: $S_{un} = 1/ (1 + 10^{n\,(pKa - pH)})$ where n is the Hill coefficient[10,11]. Chemical shift perturbations were fit using the modified Henderson-Hasselbach equation presented in the work of Webb *et. al*[6].

3.2.4 Calculation of NMR Chemical Shifts

Frames from each pH-REX CpHMD^{MSLD} simulation were extracted at an interval of 10 picoseconds and titratable groups were either protonated or deprotonated based on the respective lambda values in the corresponding time step. The Buckingham equation was used to compute the impact of long-range electrostatics on the chemical shifts and is expressed by:

$$\delta_{EF} = A_{\parallel} E_z \quad (1)$$

where E_z is the electric field (**E**) component of the nucleus of interest in the z direction, and A_{II} represents the nuclear polarizability constant specific to the bond type[21]. Each residue within a given frame was oriented to the reference frame described in Boyd et. al.[29] such that the same nuclear polarizability constants could be applied to N^H and H^N nuclei (Ax_N: -977 ppm au, Ay_N: 166 ppm au, Ax_{HN}: -89 ppm au, Ay_{HN}: -18 ppm au, respectively). The electric field between the nucleus of interest and the protein system, excluding its corresponding titratable sidechain was computed using CHARMM with a dielectric constant was set to $\varepsilon = 5$.

Subsequently, all water molecules and ions were removed from the snapshots and SPARTA+[17] was employed to compute NMR chemical shifts for backbone N^H and H^N nuclei. To add pH-dependence, chemical shift perturbations that were quantum mechanically computed from tripeptides[20] were scaled by the fraction of unprotonated states for each titratable residue. The chemical shift perturbations describe the through-bond and through-space contribution of a protonation event on the nucleus of interest for a particular titratable group. The expression used to compute the chemical shift perturbations is expressed as:

$$\Delta\delta_{pH}^{N,H} = \Delta\delta_{SPARTA}^{N,H} + S_{un}\left(\Delta\delta_{QM}^{N,H}\right) + \Delta\delta_{EF}^{N,H} \quad (2)$$

where $\Delta\delta_{SPARTA}^{N,H}$ denotes the difference between the chemical shifts computed with SPARTA+ for either N or HN subtracted from the starting pH (pH3), $S_{un}\left(\Delta\delta_{QM}^{N,H}\right)$ describes the scaling of the

quantum computed perturbations (between protonated and deprotonated titratable residues) with the unprotonated fraction of the pH-REX CpHMD^{MSLD} simulations, and $\Delta\delta_{EF}^{N,H}$ is the difference in the Buckingham equation subtracted from pH 3. All computed chemical shifts in this work are a linear average of 20,000 frames.

3.3 Results and Discussion

3.3.1 pK$_a$ Values for HEWL

Hen egg white lysozyme is a small protein with a predominantly helical structure stabilized by four disulfide bonds. An active site facilitated by aspartic acid 52 and glutamic acid 35

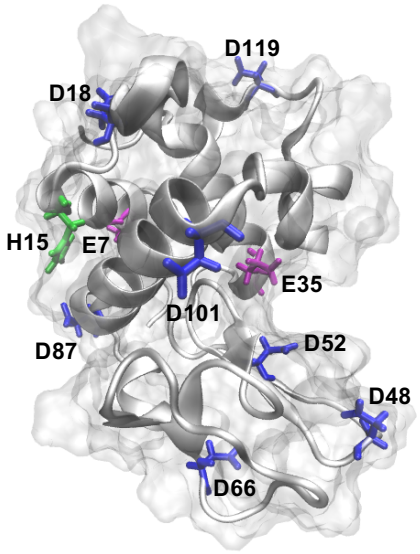


Figure 3.1 Hen Egg White Lyzosyme (PDBID 2LZT) has ten titratable residues in the pH range between 1-10, which include 2 glutamic acid residues (magenta), 1 histidine (green) and 7 aspartic acid residues (blue).

hydrolyzes glycosidic bonds most optimally at pH 5[30]. As seen in Figure 1, apart from the two catalytic residues, HEWL contains six additional aspartic acid residues, one glutamic acid and one histidine residue. Due to the large amount of available experimental data, HEWL is frequently used as a benchmarking protein for case studies, many of which serve to validate the CpHMD methodology. The pK$_a$ values of the titratable groups in HEWL are most commonly compared to those measured by NMR experiments from which consensus pK$_a$ values are extracted[6]. Consensus

pK_a values are important as unambiguous protonation events are experimentally difficult to capture[6]. In this study, pK_a values were obtained using 20ns of explicit pH-REX CpHMD^{MSLD} simulations. To test the convergence of the simulations, computed pK_a values were compared at 5ns intervals as is described in the protocol by Goh *et al*.[10] Convergence of pK_a values was achieved after 5ns and in subsequent intervals showed only small fluctuations of 0.4 pK_a units[10]. Moreover, pK_a variations between five trial runs were on average 0.3 pK_a units. As summarized in Table 1, the computed pK_a values which were fit using the Henderson-Hasselbalch equation, agree well

Table 3.1 pK_a Values of HEWL Computed by 20ns of explicit pH-REX CpHMD^{MSLD}

Residue	Exp pK_a	Comp pK_a	Error
GLU-7	2.6 ± 0.2	3.4 ± 0.1	0.8
HIS-15	5.5 ± 0.2	5.2 ± 0.1	-0.3
ASP-18	2.8 ± 0.3	1.5 ± 0.5	-1.3
GLU-35	6.1 ± 0.4	7.9 ± 0.3	1.8
ASP-48	1.4 ± 0.2	<1	--
ASP-52	3.6 ± 0.3	5.9 ± 0.4	2.3
ASP-66	1.2 ± 0.2	1.3 ± 0.1	-0.1
ASP-87	2.2 ± 0.1	<1	--
ASP-101	4.5 ± 0.1	6.4 ± 0.3	1.9
ASP-119	3.5 ± 0.3	<1	--
RMSE		1.4	
MUE		1.2	

Table 3.1 Computed pK_a values for HEWL represent a linear average of five trials and indicate good agreement with experiment[6] with an RMSE of pK_a 1.4 units and an MUE of 1.2 pK_a units. ASP-48, ASP-87 and ASP-119 do not contribute to the RMSE or MUE as these residues were unable to sample the deprotonated states.

with the consensus pK_as measured by NMR in the work of Webb et al[6]. Generally, the computed pK_a values are also comparable to those obtained by similar pH-REX CpHMD^{MSLD} studies conducted by Goh *et al*.[6], Swails *et al*.[31], and Shen *et al*.[32]

There were however some residues presenting significant sources of error (~2 pK_a units) which included the two catalytic residues GLU-35 and ASP-52, as well as ASP-101. GLU-35 is upshifted by 1.8 pK_a units suggesting that the protonated form of the residue is favored, and that sampling of the deprotonated state does not sufficiently mimic the conformational dynamics of the experiment. GLU-35 is buried in the catalytic cleft of HEWL, but as demonstrated in Figure 2 becomes more solvent exposed as the environmental pH increases. The average hydrogen bond occupancy between the sidechain of GLU-35 and water molecules increases starkly between pH 7 and 9, corresponding with its anomalous pK_a of 7.9. While GLU-35 is buried, it is likely that the

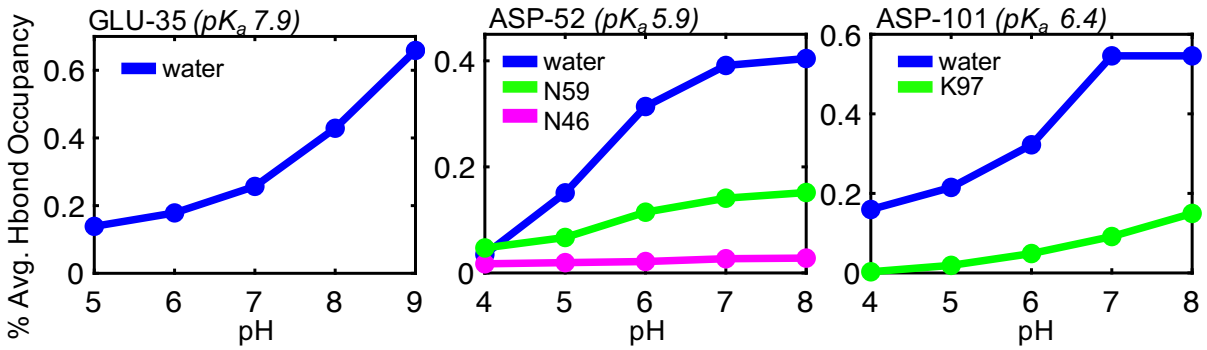


Figure 3.2 Residues with upshifted pK_a Values. The average hydrogen bond occupancy as a function of pH for A) GLU-35, B) ASP-52, and C) ASP-101 is displayed for hydrogen bonds formed with water and intra-molecular residues.

protonated form is maintained but as the sidechain becomes more solvent exposed at pH 7, the deprotonated state can more readily be sampled. These results are summarized in Figure 2. ASP-52 is in a similar electrostatic environment, remaining partially buried while participating in hydrogen bonding interactions with residues ASN-46 and ASN-59. Along with the intra-molecular hydrogen bonds, ASP-52 slightly increases its average hydrogen bonding occupancy with water as the region becomes more solvent exposed. As with GLU-35 this enables sampling of the deprotonated state and the upshifted pK_a of 5.9. Similar analysis and results for ASP-52 and GLU-35 have been reported by the Shen group[32]. Another upshifted residue is ASP-101, located on one of the outward facing helices and largely solvent exposed. The average hydrogen bond occupancy

seems to increase as pH increases, allowing the deprotonated state to be sampled more readily and consequently upshifting the pK_a to 6.4. There is also a hydrogen bond that occurs with LYS-97 which is nominally prevalent.

As seen in Table 1, there are three residues (ASP-48, ASP-87, ASP-119) that have a computed pK_a less than one. This indicates that the deprotonated state is stabilized, such that the protonated state is infrequently sampled. The consistency of the protonation state is of a conformational etiology. The depressed pK_a values for the aspartic acid residues are due to salt-bridge formation with adjacent positively charged residues. The maintenance of the unprotonated state prevents adequate sampling and serves to bias the resulting pK_a.

As seen in Figure 3, ASP-48 forms a salt bridge with the nearby ARG-61 which is sustained throughout the various pH replicas (data not shown). The salt-bridge formation alternates with a strong hydrogen bond formed with the backbone of SER-50, when a different conformation is adopted. Likewise, ASP-87 forms a substantial hydrogen bond with the sidechain of THR-89 which is maintaining the deprotonated conformation. When ASP-87 is not participating in the hydrogen bonding interaction, it forms a salt bridge with HIS-15. This salt-bridge is maintained until the histidine is fully deprotonated at pH values above six.

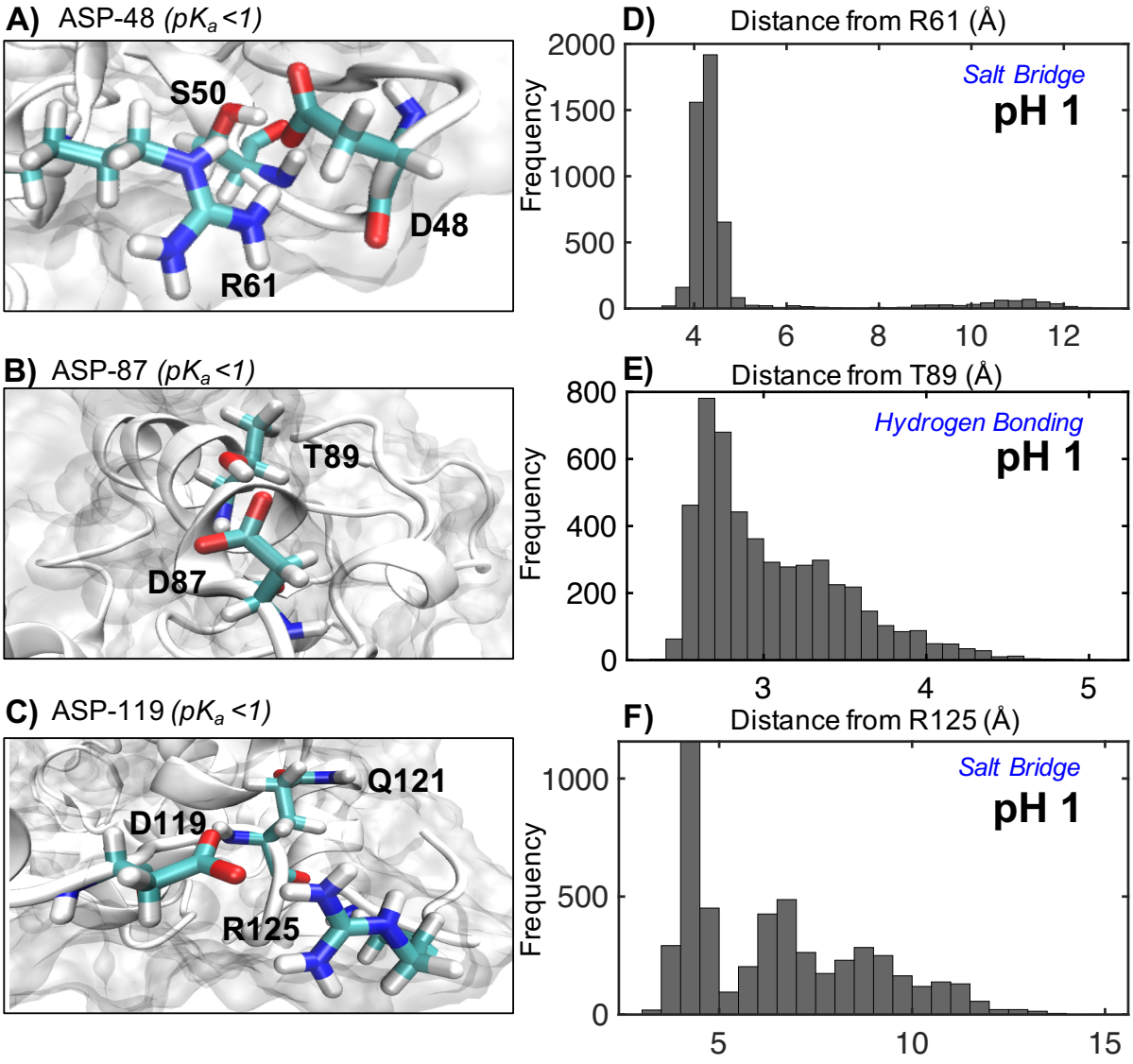


Figure 3.3 Residues with downshifted pK_a Values. Salt-bridge and hydrogen bond formation serves to depress the pK_a values of A) ASP-48, B) ASP-87, and C) ASP-119. The panels on the right-hand side illustrate the relative frequency of the distance between D) ASP-48 and ARG-61, E) ASP-87 and THR-89 and F) ASP-119 and ARG-125. The frames were extracted every 4 picoseconds (5000 frames) from pH replica 1.

This serves to depress the pK_a of ASP-87, but also decreases the pK_a of HIS-15 to 5.4, as compared to the model compound pK_a for histidine which is reported as having a pK_a of 6.45.

The salt-bridge formed by ASP-119 is slightly weaker than the one formed by ASP-48, and its occupancy is significantly lower, however, ASP-119 also forms stabilizing hydrogen bonds with the backbone of GLN-121 and ALA-122. The combination of a salt-bridge and/or hydrogen

bond interaction presents a significant challenge for sampling protonation states and results in computed pK_a values which are shifted. Similar pH-dependent structural observations of HEWL in the context of pK_a calculations have been demonstrated by numerous groups[6,30,31,32], illustrating that microenvironmental factors greatly contribute to the simulated pK_a.

3.3.2 ^{15}N pH-Dependent NMR Chemical Shifts of HEWL

In order to directly compare the microscopic pK_a values obtained from simulation and the apparent pK_a derived from experiments, we utilized SPARTA+ to compute ^{15}N and ^{1}H chemical shifts of the pH-REX CpHMD^{MSLD} ensembles at various pH windows. It has been previously demonstrated that SPARTA+ preforms very well when computing chemical shifts of molecular dynamics ensembles and is sensitive enough to identify changes in backbone dihedrals, variations in hydrogen bonding and differences in the proximity to aromatic-rings[33]. However, because SPARTA+ is not parametrized to predict chemical shifts at non-physiological pH, we incorporated the through-bond and through-space effects of protonation by scaling the pH-dependent chemical shift perturbations of model peptides[20]. Since the pioneering studies of Wüthrich, there have been numerous experimental and computational NMR pH-titration studies performed on model tripeptides and pentapeptides[34-36]. These studies have greatly elucidated the impact of protonation on chemical shift perturbations. In this work, we utilize our previously computed pH-dependent chemical shift perturbations from tripeptide constructs[20]. The peptide perturbation for the amide nitrogen in each corresponding titratable residue is scaled by the unprotonated fraction of states obtained from the pH-REX CpHMD^{MSLD} simulations. This addition supplements the SPARTA+ chemical shift perturbations, which predominantly describe the conformational differences between the ensembles.

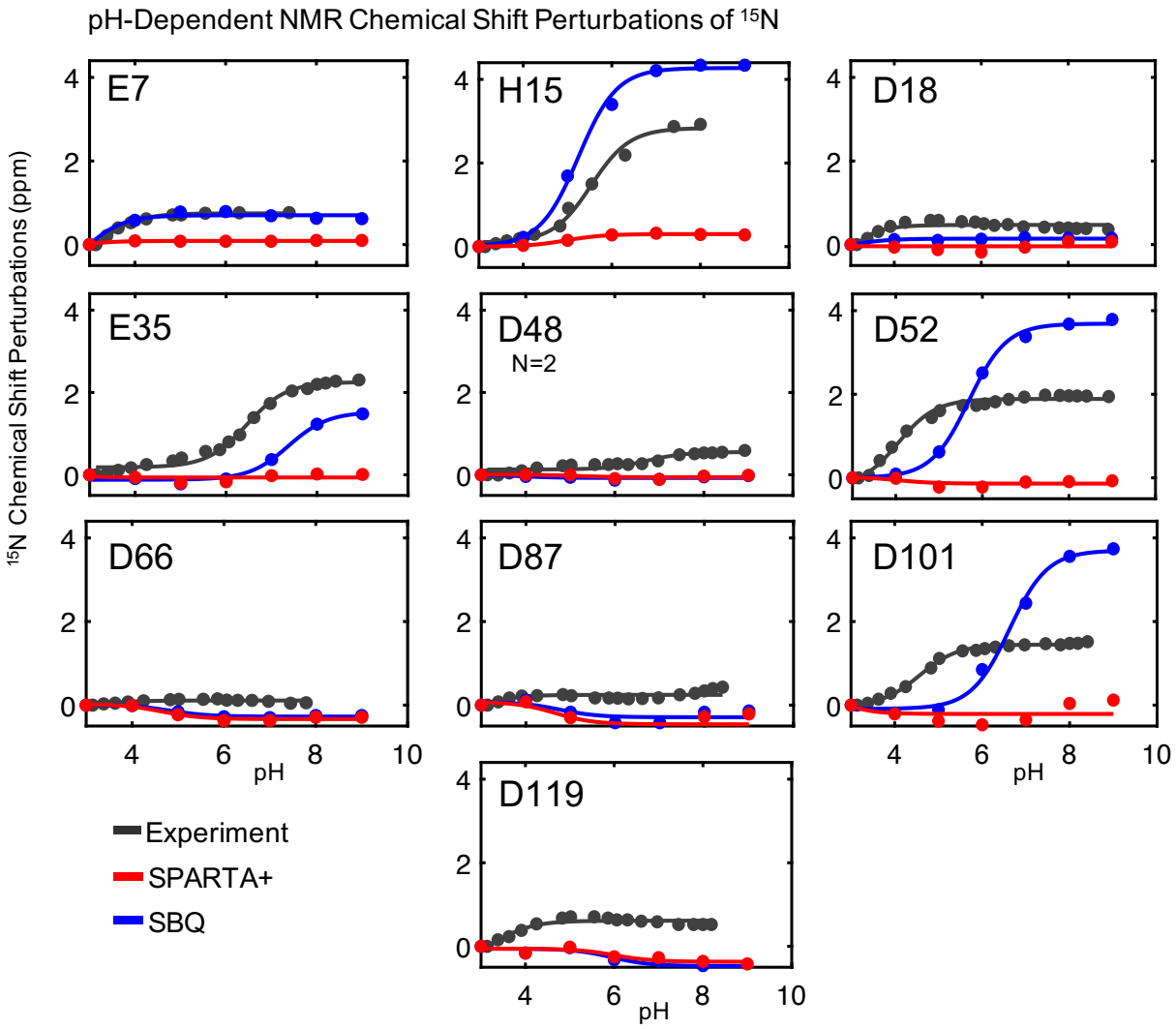


Figure 3.4 ^{15}N pH-Dependent NMR Chemical Shift Perturbations. SPARTA+ (red) was utilized to compute chemical shifts for 20,000 frames extracted from the 20ns explicit solvent pH-REX CpHMD^{MSLD} simulations at each respective pH window. The differences of the chemical shifts at each pH value with respect to the chemical shifts at pH 3, are plotted above. The chemical shift perturbations of SPARTA+ including the pH-dependent perturbations scaled by the unprotonated fraction and the through-bond contribution provided by the Buckingham equation are indicated in blue and abbreviated as SBQ. The experimental ^{15}N chemical shift perturbations are displayed in gray and were calculated by the chemical shifts kindly provided by Dr. Damien Farrell as part of the Titration_DB[37]. All curves are fitted by the modified Henderson-Hasselbach equation for 1-site unless otherwise noted.

To account for electrostatic effects outside of the titratable residue (through-space), the Buckingham equation is applied. All residues, water molecules and ions with the exception of the titratable residue undergoing the protonation event, are included in the Buckingham calculation. Although SPARTA+ includes an electrostatic effect in its predicted chemical shift (computed by

the Buckingham equation), it is not applied to amide nitrogen atoms. Furthermore, the implementation of the Buckingham equation within the program is limited to a few donor and acceptor pairs where the atomic charges remain static. As aforementioned, the Buckingham equation is utilized to compute the electrostatic contribution for each extracted frame, taking into account the fluctuation of atomic charges as protonation equilibria change concomitantly with conformation. Figure 4 shows the average perturbations of 20,000 frames computed by 1) SPARTA+, 2) SPARTA+ including the scaled model peptide pH perturbation and through-bond electrostatic contribution (denoted as SBQ), and 3) the experimental perturbations from the work of Webb *et. al*[6] as reported in the Titration_Database[37].

As expected, the SPARTA+ ^{15}N perturbations do not capture the large changes (~2 ppm) associated with the amide nitrogen atom tracking the titration event of its corresponding side chain. The computed ^{15}N perturbations are approximately 0.1 ppm for residues GLU-7, HIS-15, GLU-35, ASP-52 and ASP-101 all of which titrate within the pH range of 3 to 9 during the pH-REX CpHMD^{MSLD} simulations. Although, HEWL is a relatively stable protein, only pH values between 3 and 9 were considered for the calculation of the pH-dependent NMR chemical shifts. pH values outside of this range would likely result in the partial denaturation of the protein in an experimental setting, and this aspect is not accounted for in the simulations.

SPARTA+ did however, detect pH-dependent variations for residues ASP-18, ASP-66, ASP-87 and ASP-119. All of these aspartic acid residues have a computed pK_a value less than 3 in the pH-REX CpHMD^{MSLD} simulations and hence the observed perturbations cannot stem from a change in the residue's charge. As seen in Figure 5, the chemical shift perturbation of the deprotonated aspartic-acid residues is largely due to a conformational fluctuation. Interestingly, the largest differences in the titration curves of the four aspartic-acid residues occur between pH

values 5 and 7. In this pH region the titration of residues GLU-35, ASP-52, HIS-15 and ASP-101 (all of which are in close proximity to the aspartic acid residues) is altering the charge of the protein and perhaps inducing a localized conformational change which is reflected in the SPARTA+ chemical shift perturbations.

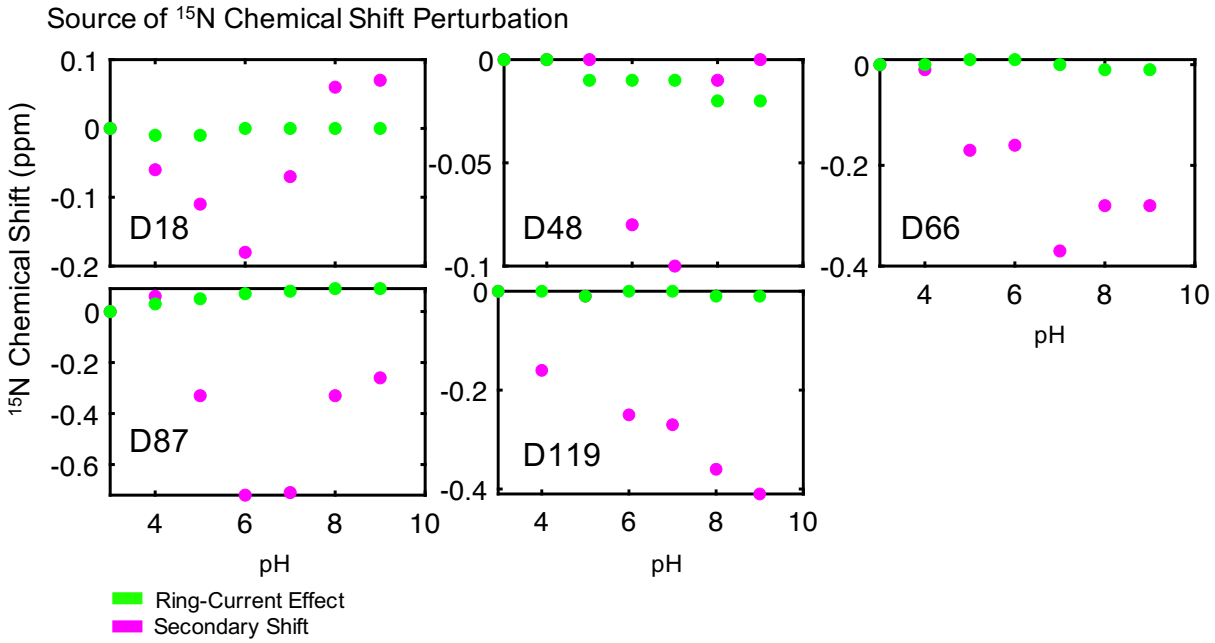


Figure 3.5 Source of SPARTA+ ^{15}N Chemical Shift Perturbations. The SPARTA+ program output specifies the source of each chemical shift contribution (random coil, ring-current effect, electric field effect, secondary-structure). Depicted in the five panels above are the respective contributions of the ring-current effect and secondary-structure for the five non-titrating aspartic acid residues, averaged over 20,000 frames. It is apparent that the majority of the total chemical shift perturbation arises from the secondary-structure component. The random coil value cancels out and the electrostatic field effect is not considered for backbone nitrogen atoms.

The addition of the pH-dependency and through-space contributions do not significantly change the computed ^{15}N perturbations for the residues that are not undergoing titration events. This is surprising in the case of ASP-48, which experimentally possesses a biphasic Henderson-Hasselbach curve. Through mutagenesis, Webb *et al*[6]. determined that the chemical shift perturbation observed by ASP-48 is in reality due to the titration of neighboring residues ASP-52 and GLU-35. Although, their study indicates a robust through-space event, our application of the

Buckingham equation did not capture this phenomenon. In our simulations ASP-52 and GLU-35 are on average 9.60Å and 14.40Å away from ASP-48 respectively, and the long-range effect may be too small to produce a significant perturbation. This may suggest that the predominant conformation in the experiment differs from that sampled in our simulations.

Improvements to the SPARTA+ perturbations can be seen in residues GLU-7, HIS-15, GLU-35, ASP-52, and ASP-101, where the titration events are captured. The predominant contribution for these residues arises from the pH-dependent perturbation that was scaled by the fraction of deprotonated frames from the pH-REX CpHMD^{MSLD} simulations. GLU-7 (0.63 ppm) agrees well with the experimental chemical shift perturbations (0.77 ppm) in the pH range of 3 to 9, however the predicted shift for GLU-35 is underestimated by 0.8 ppm, suggesting that there is a large contribution that is not accounted for in the current dynamical ensemble. In contrast, the magnitudes of the predicted perturbations for HIS-15, ASP-52 and ASP-101 are overestimated by ~2 ppm possibly indicating a systematic error with the peptide perturbations used to approximate pH-dependency. ^{15}N chemical shifts are sensitive to the preceding residue's backbone torsional angles[17] and since nearest neighbor effects were not taken into consideration for the peptide study, the overestimation is perhaps a consequence of the model peptide's two flanking glycines.

For residues ASP-66, ASP-87 and ASP-119, the conformational component is dominant as the addition of Buckingham equation does not alter the unique chemical shift behavior seen in the SPARTA+ analysis. The distance between ASP-66 and ASP-87 in the protein structure is 19Å, and although these residues are not responding to the same influence, they both display a similar line shape with the most notable perturbation of ~ -0.5 ppm at pH 6.

3.3.3 ^{1}H pH-Dependent NMR Chemical Shifts

Due to rapid exchange with solvent, chemical shifts of solvent exposed and titratable protons are difficult to resolve. Although amide protons are not impervious to exchange with solvent, it is typically slower, diminishing the detrimental effect of peak broadening. Along with amide nitrogen chemical shifts, amide protons are frequently used in pH-titrations to track protonation events of titratable residues[7]. However, the impact of hydrogen bonding geometry on the amide proton's chemical shift is significant and may obfuscate pH titration spectra. The work of Exner *et al*[37]. demonstrates a distinct relationship between the amide proton chemical shift and the distance of the proton to the hydrogen bond acceptor, causing a large downfield change up to 6 ppm for a strong hydrogen bond.

Moreover, due to the small resonance signal produced by proton nuclei, pH perturbations are miniscule (0.02 - 0.5 ppm)[6] which further complicates spectral interpretation. When predicting amide proton chemical shifts, SPARTA+ implements the Pardi-Wuthrich formalism[38] which describes an r^{-3} dependence on hydrogen bond length. This expression along with an empirical correction for solvent exposure produces a reasonable agreement with experimental chemical shifts (R= 0.63 for HN atoms)[17].

In this work, pH-dependent NMR chemical shift perturbations for amide proton atoms were calculated for the 20,000 frames extracted from the pH-REX CpHMD^{MSLD} simulations at each pH window between 3 and 9. SPARTA+ was utilized to compute chemical shift perturbations at each pH. For amide protons, SPARTA+ incorporates an electrostatic contribution by utilizing the Buckingham equation. In Figure 6, the plotted SPARTA+ chemical shift perturbations include the electrostatic contribution as computed by the program. However for the SBQ perturbations,

the electrostatic effect was subtracted from the SPARTA+ component, and our own implementation of the Buckingham equation was added.

The through-space contribution of protonation was calculated by scaling pH-dependent chemical shift perturbations of model peptides simulated in either implicit solvent or explicit solvent. Previously in our peptide study, pH-dependent chemical shift perturbations were initially computed from simulations performed in implicit solvent and ultimately yielded poor agreement with measured amide proton perturbations. Not surprisingly, the inclusion of explicit solvent molecules improved the model peptide calculations. For a residue in a random coil, hydrogen bonding with solvent is likely the largest contributor to the amide proton chemical shift. However, in a protein environment an amide proton may be partially or fully buried and as a consequence form primarily intramolecular hydrogen bonds. In the latter case, the pH-dependent NMR chemical shift perturbations may be better represented by the model peptide calculations in implicit solvent. However, if the titratable group is solvent exposed the perturbations computed with explicit solvent would be more adequate.

Figure 6 displays the results for 1) SPARTA+ computed chemical shift perturbations in red, 2) SPARTA+ including the scaled model peptide pH perturbation in implicit solvent and through-bond electrostatic contribution (implicit SBQ - open blue circles with dotted line fit), 3) SPARTA+ including the scaled model peptide pH perturbation in explicit solvent and through-bond electrostatic contribution (explicit SBQ - blue circles with solid line fit), 4) the experimental perturbations from the work of Webb *et. al.* as reported in the Titration_Database. It is initially apparent that the amide proton perturbations are small and average less than 0.5 ppm.

For GLU-7, the explicit SBQ calculation (-0.11 ppm) seems to agree reasonably well with the experimental perturbation (-0.05 ppm) as opposed to the implicit SBQ calculation which over-

predicts the ability of the amide proton to track the sidechain's titration. This behavior suggests that as the pH increases, the GLU-7 amide proton is in a progressively neutral environment and is becoming more shielded. Likewise, for HIS-15 the explicit SBQ (-0.06 ppm) is a better fit with

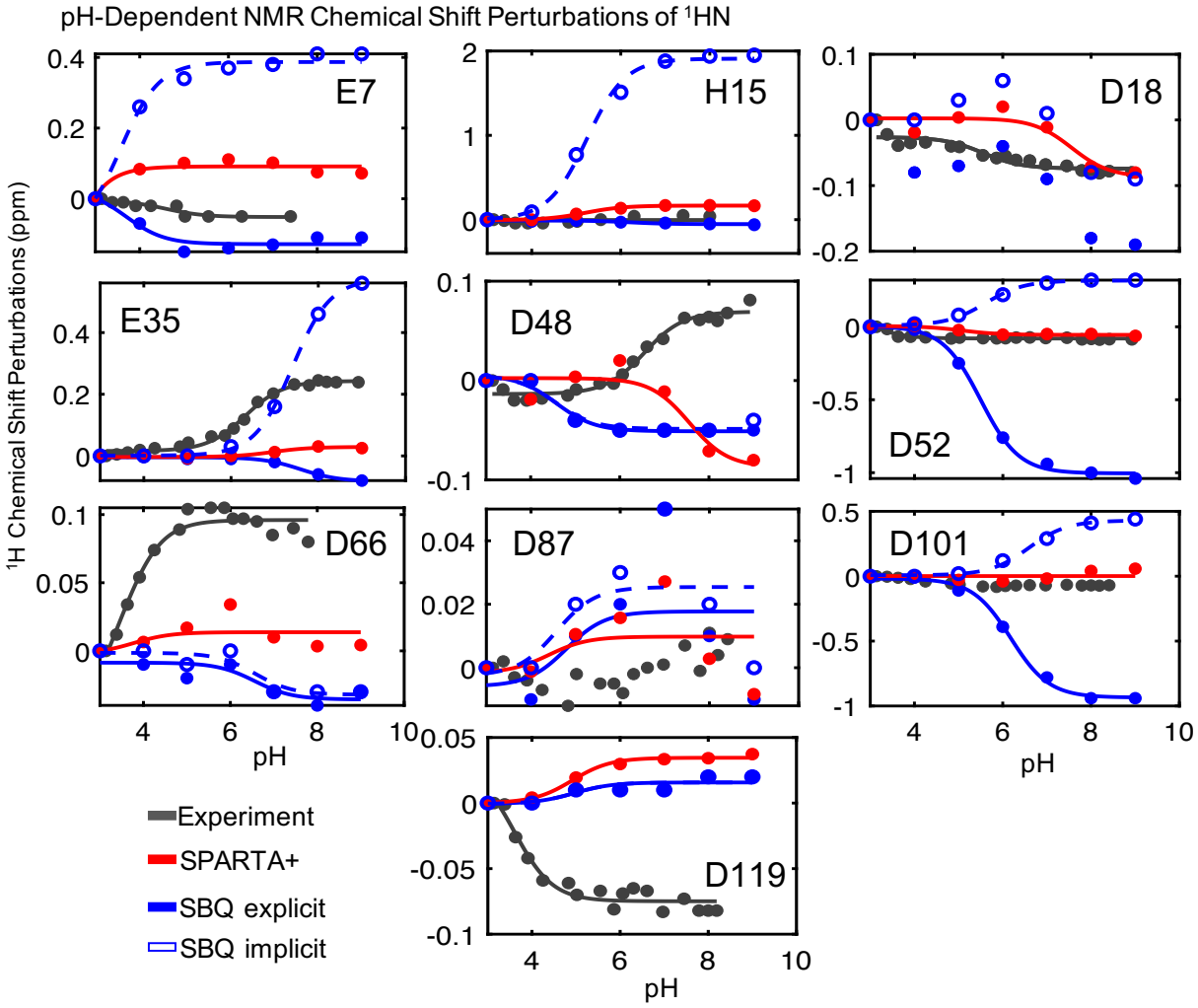


Figure 3.6 ^{1}H pH-Dependent NMR Chemical Shift Perturbations. SPARTA+ (red) was utilized to compute chemical shifts for 20,000 frames extracted from the 20ns explicit solvent pH-REX CpHMD^{MSLD} simulations at each respective pH window. The differences of the chemical shifts at each pH value with respect to the chemical shifts at pH 3, are plotted above. The chemical shift perturbations of SPARTA+ including the pH-dependent perturbations (in explicit or implicit solvent) scaled by the unprotonated fraction and the through-bond contribution provided by the Buckingham equation are indicated in blue and abbreviated as SBQ. The experimental ^{1}H chemical shift perturbations are displayed in gray and were calculated by the chemical shifts kindly provided by Dr. Damien Farrell as part of the Titration_DB. All curves are fitted by the modified Henderson-Hasselbach equation for 1-site unless otherwise noted.

the experimental perturbation (0.04 ppm), although SPARTA+ sufficiently captures the pH trend (0.17 ppm). The ability of SPARTA+ to recapitulate the trend implies that the perturbations are likely a response to a conformational change occurring in the dynamic ensemble around pH 7. Both ASP-18 and ASP-48 are located on flexible loops and display a unique line shape when fitting the SPARTA+ perturbations. The curves indicate a change around pH 7.5 where shielding increases and in turn moves the spectra upfield. For ASP-18 the implicit SBQ (-0.09 ppm) replicates the increase in shielding as seen in the experiment (-0.08 ppm). However, the perturbations for ASP-48 are more complex. The amide nitrogen shifts track the titration of nearby ASP-52 and GLU-35 residues and likewise, for the amide protons Webb et. al. determine that the perturbations arise from the protonation of GLU-35. The application of the Buckingham equation does not account for the titration event of the adjacent GLU-35, and both the implicit and explicit SBQ approaches produce an almost identical curve.

The titration of GLU-35 is also captured by its corresponding amide proton with a perturbation of 0.23 ppm in the experiment, and 0.56 ppm in the implicit SBQ model. This implies that the environment remains partially buried such that the pH-titration is the dominant contributor to the chemical shift. ASP-52 and ASP-101 are overpredicted by both SBQ approaches and SPARTA+ resulted in the most accurate description of the perturbations. ASP-87, ASP-66 and ASP-119 all of which displayed identical titration curves for the amide nitrogen calculations, produced differing perturbations neither of which tracked the titration events observed in the experimental data.

The chemical shifts for the amide protons are notoriously difficult to predict, as they are very sensitive to the microenvironment of the protein. Although the SBQ protocol does not recapitulate all of the experimental trends, it provides useful insight into the dynamic ensembles

produced by the pH-REX CpHMD^{MSLD} simulations as compared to those present in the NMR pH-titration experiment.

3.3.4 pK_a Values from NMR Chemical Shifts

In order to directly compare the micro-pK_a values resulting from the pH-REX CpHMD^{MSLD} simulations to the apparent pK_a values derived from experiment, the Henderson-Hasselbach equation was used to fit pK_a values from the calculated pH-dependent NMR chemical shift perturbations. The resulting pK_a values encompass the protonation events of the titratable residues, as well as the various contributions of the protein's microenvironment. Table 3.2 summarizes these results.

Table 3.2 pK_a Values of HEWL Computed from pH-dependent Chemical Shift Perturbations

Residue	Consensus Exp. pK_a[6]	Comp. pK_a	^{15}N Exp. pK_a	^{15}N Comp. pK_a	^{1}H Exp. pK_a	^{1}H Comp. pK_a
GLU-7	2.6	3.4	3.1	2.8	4.6	3.7, 3.5
HIS-15	5.5	5.2	5.5	5.2	--	6.2, 5.3
ASP-18	2.8	1.5	1.0	3.0	5.5	3.8, 7.3
GLU-35	6.1	7.9	6.4	7.3	6.3	7.6, 7.4
ASP-48	1.4	<1	3.8, 7.3*	3.8	5.9	4.5, 4.5
ASP-52	3.6	5.9	4.0	5.7	2.7	5.5, 5.6
ASP-66	1.2	1.3	3.1	4.8	3.5	6.5, 6.5
ASP-87	2.2	<1	2.1	4.7	--	4.7, 4.5
ASP-101	4.5	6.4	4.5	6.6	4.1	6.2, 6.6
ASP-119	3.5	<1	3.3	5.7	3.6	4.9, 4.9

Table 3.2 Consensus pK_a values from the experimental Webb study are listed in the first column. The second column shows the computed micro-pK_a values from the pH-REX CpHMD^{MSLD} simulations. The $^{15}N/^{1}H$ experimental pK_a columns display the fitted pK_a values for the experimental $^{15}N/^{1}H$ chemical shift perturbations from pH 3 to 9. The ^{15}N computed pK_a values column tabulates the fitted pK_a values from the computed NMR chemical shift perturbations using SBQ and the ^{1}H computed pK_a values column indicates the pK_a values computed for explicit SBQ and implicit SBQ respectively. The asterisk denotes that two sites were fitted.

It is evident that the amide nitrogen and amid proton pK_a values can vary drastically when compared to each other. This occurrence demonstrates ambiguous reporting of the titration event by backbone nuclei. ASP-18 exemplifies the discrepancy as the difference between the amide nitrogen (0.8) and proton (5.5) experimental pK_a apparent is significant. More accurate pK_a apparent values are obtained by monitoring the carbon chemical shift on the sidechain carboxyl moiety. It is important to note that only experimental perturbations between pH values 3 and 9 were used to compute pK_a values. This pH range reflects that of the pH-REX CpHMD^{MSLD} simulations for which NMR chemical shifts were computed. As a result, the experimental pK_a apparent values fit in this study vary slightly from those reported in the work of Webb *et al*.

The ^{15}N computed pK_a values generally agreed with the micro- pK_a values from simulation. Surprisingly, ASP-48, ASP-87 and ASP-119 which had depressed pK_a values due to salt bridge formation, presented upshifted ^{15}N pK_a apparent values. The increase, however, was overestimated and produced an average MUE of 2.3 pK_a units. The upshift in the three aspartic acids was also mirrored in the computed ^{1}H pK_a apparent. GLU-7, HIS-15, GLU-35, ASP-52 and ASP-101 did not express significant variation between the micro- pK_a and the $^{15}N/^{1}H$ pK_a apparent. HIS-15, GLU-35, ASP-52 and ASP-101 are among the only residues that titrate in the pH range between 3 and 9 suggesting that the perturbations of the titration are robust. Interestingly, there was dissimilarity between the computed ^{1}H pK_a apparent using either the implicit or explicit SBQ methods for two residues. HIS-15 exhibits a 1 pK_a unit difference, favoring the implicit pK_a which suggests that the amide proton is in a more buried environment. Similarly, there is a 3 pK_a difference for ASP-18 which agrees best with the explicit SBQ which implies hydrogen bonding with solvent.

3.4 Conclusion and Discussion

In this work, we devise a protocol to compute the pH-dependent NMR chemical shifts of the conformational ensembles produced by CpHMD^{MSLD}. The computed HEWL micro-pK_a values generally agreed well with experiment, however the formation of salt bridges and strong hydrogen bonds prevented adequate sampling for a selection of aspartic acid residues. Due to the variety of conformational effects that perturb the chemical shift in concert with protonation, experimentally derived pK_a apparent values are fundamentally different than those computed from simulation. Computing pK_a values by fitting computed chemical shift perturbations allows for the direct comparison between experiment and simulation. Utilizing the pH-dependent perturbations of model peptides, conformational variation from SPARTA+ and through-space effects from the Buckingham equation, a model to predict chemical shifts was formulated. This method predicted reasonable ^{15}N chemical shift perturbations from which ^{15}N pK_a values were fit. The residues that titrate in the pH range between 3 and 9, exhibited a Henderson-Hasselbach curve with a ^{15}N perturbation of 2-4 ppm. Three aspartic acid residues displayed complex titration behavior stemming from a conformational change in the dynamic ensemble. Amide proton chemical shift prediction presents a significant challenge due to factors such as solvent exposure and hydrogen bonding. We utilized two different SBQ models (implicit and explicit solvent) to account for the variation in solvent exposure for the titratable residues. This approach was successful in computing ^{1}H pK_a values, as the implicit and explicit predictions rarely diverged.

3.5 Bibliography

1. Ellis, C.R. and Shen, J., 2015. pH-dependent population shift regulates BACE1 activity and inhibition. *Journal of the American Chemical Society*, *137*(30), pp.9543-9546.

2. Zahn, R., 2003. The octapeptide repeats in mammalian prion protein constitute a pH-dependent folding and aggregation site. *Journal of molecular biology*, *334*(3), pp.477-488.

3. Ahlstrom, L.S., Dickson, A. and Brooks III, C.L., 2013. Binding and folding of the small bacterial chaperone HdeA. *The Journal of Physical Chemistry B*, *117*(42), pp.13219-13225.

4. Hammond, M.S., Houliston, R.S. and Meiering, E.M., 1998. Two-dimensional 1H and 15N NMR titration studies of hisactophilin. *Biochemistry and cell biology*, *76*(2-3), pp.294-301.

5. Hass, M.A. and Mulder, F.A., 2015. Contemporary NMR studies of protein electrostatics. *Annual review of biophysics*, *44*, pp.53-75.

6. Webb, H., Tynan-Connolly, B.M., Lee, G.M., Farrell, D., O'Meara, F., Søndergaard, C.R., Teilum, K., Hewage, C., McIntosh, L.P. and Nielsen, J.E., 2011. Remeasuring HEWL pKa values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pKa calculations. *Proteins: Structure, Function, and Bioinformatics*, *79*(3), pp.685-702.

7. Nielsen, J.E., 2008. Analyzing protein NMR pH-titration curves. In *Annual reports in computational chemistry* (Vol. 4, pp. 89-106). Elsevier.

8. Hass, M.A., Ringkjøbing Jensen, M. and Led, J.J., 2008. Probing electric fields in proteins in solution by NMR spectroscopy. *Proteins: Structure, Function, and Bioinformatics*, *72*(1), pp.333-343

9. Knight, J.L. and Brooks III, C.L., 2011. Applying efficient implicit nongeometric constraints in alchemical free energy simulations. *Journal of computational chemistry*, *32*(16), pp.3423-3432.

10. Goh, G.B., Hulbert, B.S., Zhou, H. and Brooks III, C.L., 2014. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: structure, function, and bioinformatics*, *82*(7), pp.1319-1331.

11. Khandogin, J. and Brooks, C.L., 2005. Constant pH molecular dynamics with proton tautomerism. *Biophysical journal*, *89*(1), pp.141-157.

12. May, E.R., Arora, K. and Brooks III, C.L., 2014. pH-induced stability switching of the bacteriophage HK97 maturation pathway. *Journal of the American Chemical Society*, *136*(8), pp.3097-3107.

13. Goh, G.B., Laricheva, E.N. and Brooks III, C.L., 2014. Uncovering pH-dependent transient states of proteins with buried ionizable residues. *Journal of the American Chemical Society*, *136*(24), pp.8496-8499.

14. Torabifard, H., Panahi, A. and Brooks, C.L., 2020. M2 amphipathic helices facilitate pH-dependent conformational transition in influenza A virus. *Proceedings of the National Academy of Sciences*, *117*(7), pp.3583-3591.

15. Sugita, Y. and Okamoto, Y., 1999. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, *314*(1-2), pp.141-151.

16. Han, B., Liu, Y., Ginzinger, S.W. and Wishart, D.S., 2011. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of biomolecular NMR*, *50*(1), p.43.

17. Shen, Y. and Bax, A., 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR*, *48*(1), pp.13-22

18. Li, D.W. and Brüschweiler, R., 2012. PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *Journal of biomolecular NMR*, *54*(3), pp.257-265.

19. Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z. and Nakatani, E., 2007. BioMagResBank. *Nucleic acids research*, *36*(suppl_1), pp.D402-D408.

20. Artikis, E. and Brooks III, C.L., 2019. Modeling pH-Dependent NMR Chemical Shift Perturbations in Peptides. *Biophysical journal*, *117*(2), pp.258-268.

21. Buckingham, A.D., 1960. Chemical shifts in the nuclear magnetic resonance spectra of molecules containing polar groups. *Canadian Journal of Chemistry*, *38*(2), pp.300-307.

22. Ramanadham, M., Sieker, L.C. and Jensen, L.H., 1990. Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares. *Acta Crystallographica Section B: Structural Science*, *46*(1), pp.63-69.

23. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.A. and Karplus, M., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), pp.187-217.

24. Feig, M., Karanicolas, J. and Brooks III, C.L., 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling*, *22*(5), pp.377-395.

25. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, *79*(2), pp.926-935.

26. Huang, J. and MacKerell Jr, A.D., 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry*, *34*(25), pp.2135-2145.

27. Best, R.B., Zhu, X., Shim, J., Lopes, P.E., Mittal, J., Feig, M. and MacKerell Jr, A.D., 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of chemical theory and computation*, *8*(9), pp.3257-3273.

28. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, *23*(3), pp.327-341.

29. Boyd, J., Domene, C., Redfield, C., Ferraro, M.B. and Lazzeretti, P., 2003. Calculation of Dipole-Shielding Polarizabilities (σ $\alpha\beta\gamma$ I): The Influence of Uniform Electric Field Effects on the Shielding of Backbone Nuclei in Proteins. *Journal of the American Chemical Society*, *125*(32), pp.9556-9557.

30. Machuqueiro, M. and Baptista, A.M., 2008. Acidic range titration of HEWL using a constant-pH molecular dynamics method. *Proteins: Structure, Function, and Bioinformatics*, *72*(1), pp.289-298.

31. Swails, J.M. and Roitberg, A.E., 2012. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *Journal of chemical theory and computation*, *8*(11), pp.4393-4404.

32. Huang, Y., Yue, Z., Tsai, C.C., Henderson, J.A. and Shen, J., 2018. Predicting catalytic proton donors and nucleophiles in enzymes: How adding dynamics helps elucidate the structure–function relationships. *The journal of physical chemistry letters*, *9*(6), pp.1179-1184.

33. Robustelli, P., Stafford, K.A. and Palmer III, A.G., 2012. Interpreting protein structural dynamics from NMR chemical shifts. *Journal of the American Chemical Society*, *134*(14), pp.6365-6374.

34. Prestegard, J.H., Sahu, S.C., Nkari, W.K., Morris, L.C., Live, D. and Gruta, C., 2013. Chemical shift prediction for denatured proteins. *Journal of biomolecular NMR*, *55*(2), pp.201-209.

35. Schwarzinger, S., Kroon, G.J., Foss, T.R., Wright, P.E. and Dyson, H.J., 2000. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *Journal of biomolecular NMR*, *18*(1), pp.43-48.

36. Platzer, G., Okon, M. and McIntosh, L.P., 2014. pH-dependent random coil 1 H, 13 C, and 15 N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *Journal of biomolecular NMR*, *60*(2-3), pp.109-129.

37. Exner, T.E., Frank, A., Onila, I. and Möller, H.M., 2012. Toward the quantum chemical calculation of NMR chemical shifts of proteins. 3. Conformational sampling and explicit solvents model. *Journal of chemical theory and computation*, *8*(11), pp.4818-4827.

38. Wagner, G., Pardi, A. and Wuethrich, K., 1983. Hydrogen bond length and proton NMR chemical shifts in proteins. *Journal of the American Chemical Society*, *105*(18), pp.5948-5949.

CHAPTER 4

The pH-Dependent Switching Mechanism of Hisactophilin

This work is in collaboration with Dr. Duncan MacKenzie and Dr. Elizabeth Meiering of the University of Waterloo

4.1 Introduction

Protein switches enable the reversible conversion between two functional states based on a response to environmental stimuli[1]. Most commonly, a conformational change coupled with a shift in the equilibrium of protonation, ligand concentration or ion flux, enables a protein to toggle between various functions[2]. This phenomenon is demonstrated in many biological processes, predominantly those involved in regulation of gene expression[3], signaling[4] and luminescence[5]. Due to the exquisite sensitivity required to produce an allosterically driven response, there have been numerous studies in recent years aimed at developing synthetic protein switching models.[6,7] Most recently, the Baker group has developed the first pH-dependent switch, LOCKR[6], which relies on an expansive hydrogen bonding network between histidine residues to trigger a multi-domain conformational change[6].

Understanding the ability of naturally occurring pH-dependent switches to induce a conformational or functional change is important to the advancement of this emerging field. Here, we explore the charge organization of hisactophilin in the context of pH-dependent switching. Hisactophilin is a highly charged protein in the soil amoeba species *Dictyostelium discoideum*[8]. After expression, hisactophilin undergoes a cotranslational modification where a myristoyl moiety

is added to the protein's N-terminal glycine residue[9]. The solvent exposure of the covalently bound myristoyl group is modulated by the cytoplasmic pH[9]. In the 'accessible state' (< pH 6.5) hisactophilin binds to the cell membrane[10] via the solvent exposed myristoyl group, which in turn enables actin recruitment to the membrane[10]. As the environmental pH increases above 6.5, the solvent exposed myristoyl group is buried in the central cavity of the β-trefoil and the protein is in the 'sequestered state'[10].

Because switching occurs between pH 6.5 and 7.5, it is likely that the mechanism is facilitated by the deprotonation of a selection of histidine residues, as free histidine residues have a pK_a of 6.45. Hisactophlin is highly charged, as nearly 45% of the protein is composed of titratable residues including 31 histidine, 6 aspartic acid, 7 glutamic acid and 8 lysine residues. Furthermore, the three-fold axis of symmetry of the protein's β-trefoil[9] architecture enables the titratable sidechains to be grouped together in pairs or triplets, forming stabilizing hydrogen bonds and salt-bridges.

It has been demonstrated by the Meiering group, that the addition of the myristoyl group increases the stability of hisactophilin and that the stabilization is higher in the sequestered state[10] where favorable interactions are formed within the barrel of the protein[11]. Furthermore, NMR studies monitoring amide nitrogen and amide proton chemical shift perturbations suggest that the average pK_a apparent values of the histidines in the myristoylated (pK_a 6) and non-myristoylated (pK_a 7-7.5) hisactophilin differ by approximately 1 pK_a unit[11]. This further suggests that the redistribution of charge within the small pH range of 6.5 to 7.5 plays an important role in the pH-switching mechanism. Importantly, large chemical shift perturbations have been identified in experimental studies, for residues HIS-75 and HIS-91 which are specifically thought to participate in the switching event[11]. Additionally, NOEs between the myristoyl group and the protein have

identified four significant non-titratable residues in the core of the protein cleft[11]. These residues serve to orient the lipid chain in the central cavity of the protein when the system is in the sequestered state.

In this study, we employ constant pH molecular dynamics techniques to clarify the impact of histidine protonation on the pH-switching mechanism observed in hisactophilin. With the use of explicit solvent pH-REX CpHMD^{MSLD} simulations we identify histidine residues that experience large pK_a perturbations upon myristoylation in the pH range between 6.5 and 7.5. Furthermore, we utilize implicit solvent CpHMD to visualize the pH-switching mechanism allowing for a qualitative description of key residues and non-native interactions. Finally, we conduct dynamic network analysis of the myristoylated hisactophilin at pH 7, in order to identify additional residues in the communication interface of the protein.

Understanding how the organization of charge keeps the protein poised on the brink of conformational change will provide insight into the mechanistic details of the pH-dependent switching of hisactophilin. Broadly, this work informs the design of switches responding to electrostatic triggers.

4.2 Methodology

4.2.1 Structure Preparation

The structure of myristoylated hisactophilin was provided by the Meiering group[10,11] and is based on the solution state NMR structure of the non-myristoylated hisactophilin deposited in the protein database by the Holak group (PDB accession code: 1HCD)[8]. The Meiering group[10,11] docked the myristoyl moiety in the central cavity of the 1HCD structure using Chimera and minimized briefly in VMD[24]. The non-myristoylated structure in this work is obtained from the coordinates deposited in the protein database by the Holak group (PDB accession code: 1HCD). For explicit solvent pH-

replica exchange constant pH molecular dynamics (CpHMD[MSLD]), hydrogen atoms were added to both structures with the HBUILD[12] functionality of the CHARMM[12] macromolecular simulation package. Subsequently, the convpdb.pl facility in the MMTSB toolset[13] was utilized to solvate the proteins in a 69Å cubic box of TIP3P water[14]. After charge neutralization, the appropriate number of sodium and chloride ions was added to the systems, to match the experimental salt concentration of 50mM[10]. The CHARMM36 all-atom protein force field[15] with CMAP correction[16] was applied within CHARMM to describe the protein and solvent parameters. Force field parameters for the myristoyl group were built from palmitic acid in the CHARMM27 lipid force field[17] and appended to the force fields used for all simulations of myristoylated hisactophilin. Residues ASP, GLU, and HIS were patched to allow for titration. Lysine residues are outside of the scope of the narrow pH range for which switching occurs, and therefore remained protonated throughout the duration of the simulation. The systems were each minimized by 1000 steps of steepest descent followed by the adopted basis Newton–Raphson method for 1000 steps.

4.2.2 Implicit Solvent GBSW CpHMD Simulation

The implicit solvent CpHMD algorithm was implemented in the CHARMM-OpenMM[12] interface as described in the work by Arthur *et al*[18]. The GBSW (a generalized Born model with a simple switching function) implicit solvent model[19] was used, and a salt concentration of 50mM was included in order to replicate experimental conditions[11]. All implicit solvent simulations were performed using the CHARMM package and the CHARMM22 force field[20] with GBSW-specific CMAP backbone correction. Both initial structures were obtained as described in the previous section and patches on titratable groups ASP, GLU and HIS were applied. Subsequently, structures were minimized for 1000 steps of steepest descent followed by the adopted basis Newton–Raphson method for 1000 steps. A non-bonded cutoff of 12 Å was applied, as well as SHAKE[21] constraints

to all bonds containing hydrogens. Simulations were performed of the myristoylated and non-myristoylated hisactophilin for 20ns at pH values between 6 and 8, in 1 pH unit increments.

4.2.3 Explicit Solvent CpHMD^{MSLD} Simulations of Hisactophilin

For both myristoylated and non-myristoylated hisactophilin, explicit solvent CpHMD^{MSLD} with pH replica exchange (pH-REX) was performed in CHARMM, modules BLOCK and REPDSTR. All simulations utilize the CMAP corrected CHARMM36 all-atom protein force field including the myristoyl parameters as previously described and the CHARMM TIP3P water model. Residues ASP, GLU, and HIS were allowed to sample continuously between protonated and deprotonated states via the continuous variable λ[22]. Lysine is anticipated to titrate outside of the pH range where the switching is observed and hence was not titrated in the simulations. The SHAKE algorithm was used on hydrogen-heavy atom bond lengths and the simulations advanced at a 2-femtosecond time step. A nonbonded cutoff of 15 Å with electrostatic force and van der Waals force switching functions was applied. A Langevin heat bath maintained the temperature at 298K and used a frictional coefficient of 10 ps^{-1}. Fifteen pH replica windows were used to simulate pH ranges 1 to 14 with pH units of 1. The exchange in pH was attempted every 500 dynamics steps and each replica was simulated for 10ns after a 1ns equilibration. All biases and parameters used are those presented in Goh *et. al*[22].

4.2.4 Dynamical Network Analysis

The NetworkView module[23] of VMD[24] was utilized to compute the correlated motions of residues in the pH 7 explicit solvent pH-REX CpHMD^{MSLD} trajectory. All water molecules and ions were removed from the simulation and the protein was aligned based on backbone RMSD. The last 10ns of the simulation were extracted for network analysis. Programs Carma and Catdcd were

downloaded separately and interfaced with VMD for the calculation of nodes and edges. All parameters were kept as default values.

4.3 Results and Discussion

4.3.1 pKa Perturbations between Myristoylated and Non-Myristoylated Hisactophilin

The pH-dependent switching behavior of hisactophilin has been observed to occur at pH 6.5[10] where the accessible, solvent-exposed myristoyl group becomes buried in the central cavity of the β–trefoil and adopts the sequestered conformation[10]. This mechanism is likely facilitated by the deprotonation of a subset of the 31 histidine residues present in hisactophilin[10], as the pK_a of a free histidine residue is 6.45. Explicit solvent pH-REX CpHMD^{MSLD} simulations were performed in order to understand the contribution of titratable residues ASP, GLU and HIS on the overall charge of the protein. Both the myristoylated (MYR) and non-myristoylated (NM) proteins were evaluated such that pK_a values could be computed in the presence and absence of the myristoyl group. The results are summarized in Figure 4.1.

Only the pK_a values for the histidines are plotted as the glutamic and aspartic acid residues have severely depressed pK_a values (<1). Due to the positively charged microenvironment, the glutamic and aspartic acid residues did not sample the protonated state and were negatively charged for the majority of the simulations. It is apparent that the computed pK_a values of the histidine residues span a wide pH range between 2 and 12. NMR pH-titration studies by Meiering

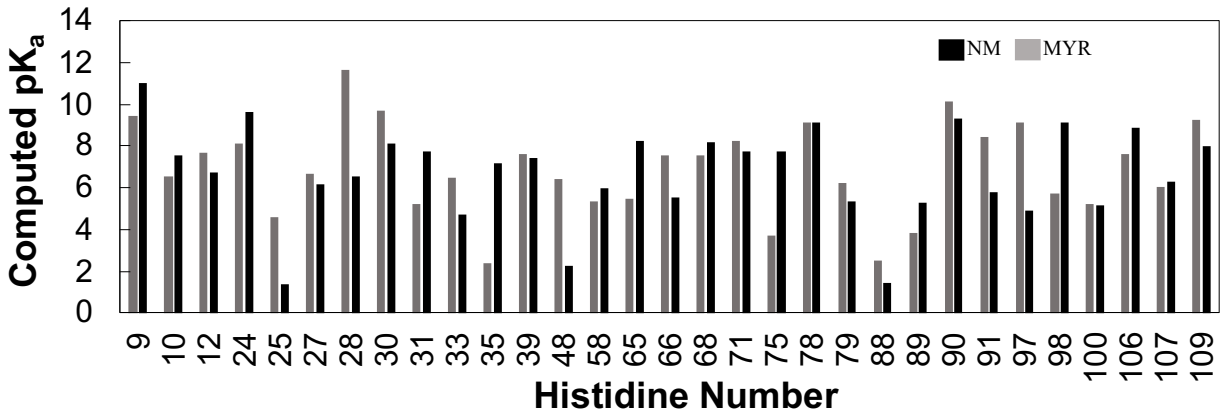


Figure 4.1 pK_a Values Computed for Myristoylated and Non-Myristoylated Hisactophilin. The pK_a values of hisactophilin were computed using 10ns of pH-REX CpHMD^{MSLD} for the nonmyristoylated (NM) hisactophilin protein shown in black and the myristoylated hisactophilin depicted in grey.

and Houliston[25], measured the pK_a apparent of all the residues in the non-myristoylated hisactophilin from ^{15}N HSQC pH titration experiments and observed pK_a values in the range of 5.5 to 8[25]. However, the three-fold axis of symmetry of the β–trefoil allows for the histdines to often be distributed in pairs or triplets throughout the protein. Due to the close proximity of the titratable groups, there is significant crosstalk between the charged residues in the pH titrations. As demonstrated by histidine residues 24, 25 and 27 in Figure 4.1, the close proximity both sequentially and structurally, allows for histidine 24 to be upshifted in both the myristoylated and non-myristoylated proteins; while histidine 25 is depressed to a pK_a of ~5 for the myristoylated protein and below 2 for the non-myristoylated protein. Nearby histidine 27 however, expresses a pK_a similar to that of a model histidine at around 6.45. This phenomenon is due to the simultaneous titration of multiple histidine residues causing large charge fluctuations in the protein's microenvironment and resulting in anomalous pK_a values.

From the pH-REX CpHMD^{MSLD} calculations, the histidine residues that may participate in switching are identified as those demonstrating significant pK_a differences upon myristoylation

between pH 6.5 and pH 7.5. These results are displayed in Figure 4.2. It is apparent that the pK_a differences are not clustered in one particular region of the protein. The large pK_a perturbations can be found mainly on histidines that are solvent exposed and are likely to display conformational flexibility. This is in contrast with the NMR experiments conducted by the Meiering group[11] which identify the largest chemical shift perturbations clustered around histidine residues 75-107[11].

HIS-28, which is close to the opening of the myristoyl cavity undergoes the largest pK_a shift upon myristoylation in the pH-REX CpHMD^{MSLD} results and maintains a positive charge with

Residue	pK_a difference (MYR-NM)
HIS-10	-1.02
HIS-12	0.94
HIS-27	0.51
HIS-28	5.09
HIS-31	-2.49
HIS-33	1.74
HIS-35	-4.82
HIS-39	0.21
HIS-66	2.04
HIS-68	-0.61
HIS-71	0.53
HIS-75	-4.03
HIS-106	-1.28

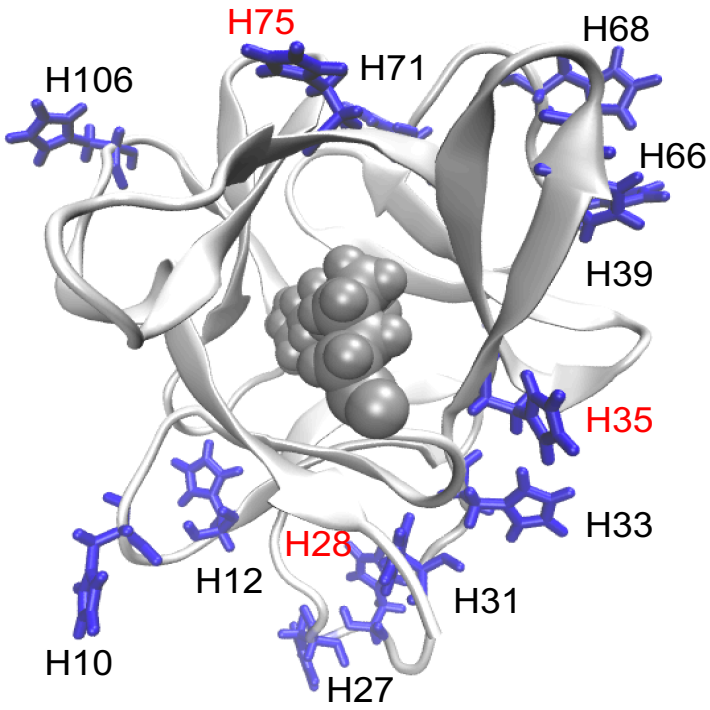


Figure 4.2 Largest pK_a Perturbations between Myristoylated and Non-Myristoylated Hisactophilin between pH range 6.5 and 7.5. The pK_a values of hisactophilin were computed using 10ns of pH-REX CpHMD^{MSLD} and the largest differences in the switching pH range between myristoylated and non-myristoylated hisactophilin is tabulated. The differences are also displayed on the structure of hysactophilin and are colored in blue. Residues colored in red indicate the largest absolute perturbations.

a highly anomalous computed pK_a value of 11.5. However, in the non-myristoylated protein HIS-28 displays a conventional pK_a of ~6.4. Likewise, HIS-35 which is in the vicinity of HIS-28 on the top of the myristoyl cavity, experiences a large (-4.82 pK_a unit) difference upon myristoylation.

HIS-35 is readily deprotonated in the myristoylated form and seems to coordinate with adjacent residue HIS-33 which experiences an upshift in pK_a upon myristoylation. These observations demonstrate the sharing of protons among adjacent titratable groups.

Although the two largest pK_a perturbations were near the N-terminus (residues 28 and 35), the third largest pK_a shift is observed for HIS-75. This residue experiences a 4 pK_a unit difference upon myristoylation favoring the deprotonated form. HIS-75 is one of two residues identified by the experimental NMR studies[11] as having very large amide proton chemical shift perturbation. Fitting the chemical shifts to the Henderson-Hasselbalch equation, Smith et *al*.[11] established that myristoylated hisactophilin exhibits a lower pK_a than the non-myristoylated protein. The pK_a difference between the two was more than 1 pK_a unit and as such HIS-75 and HIS-91 (which also experienced a 1 pK_a unit perturbation) were assumed to play an important role in switching. In our studies, HIS-91 titrated outside of the pH range of 6.5 – 7.5 for both the myristoylated and non-myristoylated hisactophilin.

4.3.2 Implicit Solvent CpHMD Simulations

During the explicit solvent pH-REX CpHMD^{MSLD} simulations, the myristoyl group was rarely observed in the accessible state. Despite sampling the charge distribution which would facilitate the pH-dependent switching mechanism, the 10ns time scale of the explicit solvent simulation was not sufficiently long to observe this transition in our simulations. In order to increase conformational sampling and capture the switching mechanism, 20ns implicit solvent GBSW CpHMD simulations were conducted of myristoylated hisactophilin in the pH range from 6 to 8. Figure 3 displays the results of one set of the CpHMD simulations. Replicate simulations were run 100 times (sets of pH6 to pH8) and only 30% exhibited pH-switching. This may suggest that although implicit solvent simulations allow for fewer degrees of freedom and hence more

conformational sampling, 20ns is still not an adequate amount of time for all the simulations to converge and display pH-switching behavior. Longer runs need to be conducted to verify this hypothesis.

The snapshots in Figure 3, show that at pH 6 the myristoyl group is hovering above the β–trefoil cavity and is frequently interacting with residues HIS-107, GLU-115 and LYS-105. In experimental studies[10], HIS-107 exhibits a large amide proton chemical shifts perturbation, possibly indicating a non-specific interaction with the accessible myristoyl group. The distance of the myristoyl group out of the pocket is quantified by three valine residues that are found in the bottom of the protein's myristoyl cleft. The geometric center of VAL-21, VAL-61 and VAL-101 to the C14 carbon of the myristoyl group is used as a measurement standard for all three pH simulations (similar analysis was performed by Shental-Bechor *et al*[10]). As is demonstrated in Figure 3 panel D, there is a stark bimodal distribution of distance for pH 6 suggesting that the switching occurs in a two-state transition. The myristoyl group is outside of the cavity (as seen in panel A) for a majority of the simulation time and the average distance of the myristoyl group from the valines is 17.5 Å. This distance distribution is not observed for the other pH values.

The myristoyl group buries itself in the hydrophobic protein center at pH 7 as seen in panel B. There are no large secondary structural changes observed between pH 6 and 7, as the four

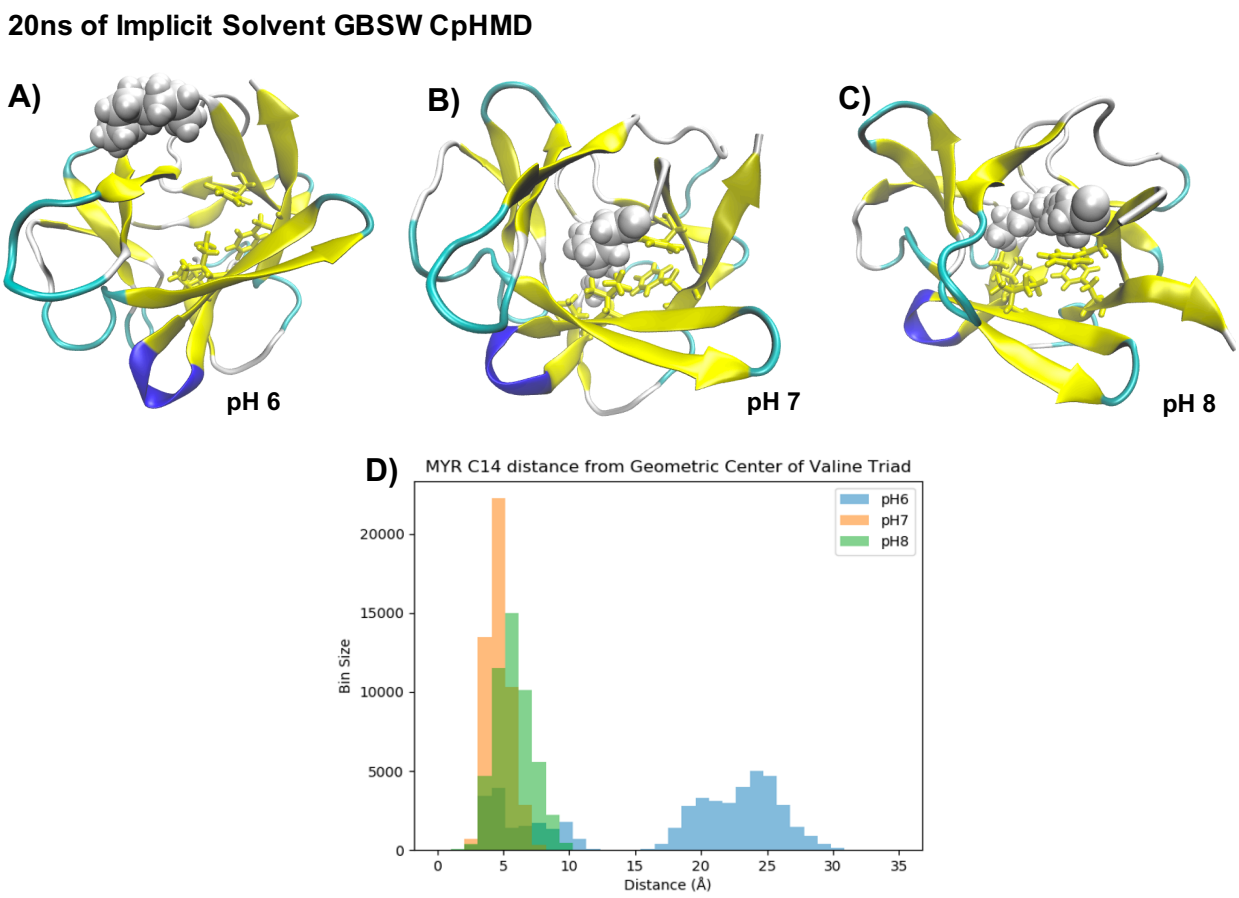


Figure 4.3 Implicit Solvent CpHMD simulations of Myristoylated Hisactophilin. Snapshots of myristoylated hisactophilin from 20ns implicit solvent CpHMD simulations are displayed in panels A, B, and C. The snapshots correspond with pH values 6, 7, and 8 accordingly. The backbone of the protein is color coded to display secondary structure. The residue sidechains presented inside the trefoil barrel are those identified by experiment as important to the switching mechanism. Panel D shows the distribution of distances for the respective pH values, as measured between the geometric center of three valines and the C14 carbon of the myristoyl group.

beta hairpins on each side of the trefoil remain intact. In the pH 7 snapshot, residues PHE-6, PHE-113, ILE-85 and ILE-93 are seen interacting with the myristoyl group. This interaction mirrors the NOEs detected between the myristoyl and the protein by NMR experiments[11]. In subsequent work by the Meiering group, residues PHE-6, ILE-85, ILE-93 which are inside the cavity have been identified as crucial to the pH-switching mechanism as mutating any of the four residues alters or even breaks switching[10]. At pH 7, the myristoyl group is deep inside the protein's cavity with an average distance of 4.68 Å from the valine triad. As expected, the secondary structure is maintained

for the duration of the simulation as denaturation studies have demonstrated that the myristoylated hisactophilin in the sequestered state experiences increased protein stability. Panel C shows the pH 8 conformation of the myristoylated protein. Again, the myristoyl group is buried in the center of the trefoil and is interacting with the four critical residues. The distance distribution although not as uniform as that observed for pH 7, has an average of 5.81Å.

4.3.3 Residues Implicated in Myristoyl Group Switch

Due to the complexity of the charge distribution and conformational plasticity, it is difficult to identify the residues responsible for pH-switching. To better understand the interplay and communication between various residues, we employed dynamic network analysis. Through the VMD plugin NetworkView[23], a dynamic network for the pH 7 explicit solvent pH-rex CpHMD^{MSLD} ensemble was computed. Briefly, the Girvan–Newman algorithm[27] is used to identify subcommunities in which residues have correlated motion[26]. Critical points identify linchpin residues that connect motions of subcommunities[23].

In Figure 4, panel A shows the five distinct subcommunities that display correlated motions. It is readily apparent that the residues are forming numerous connections with each other and only a few display thick edges. Interestingly, the blue subcommunity directly includes the backbone of GLY-2 which is covalently bound to the myristoyl group. Also, within this subcommunity are residues HIS-28 and HIS-35, which exhibit highly perturbed pK_a values upon myristoylation. The thick edges displayed in the gray community can be attributed to stabilizing

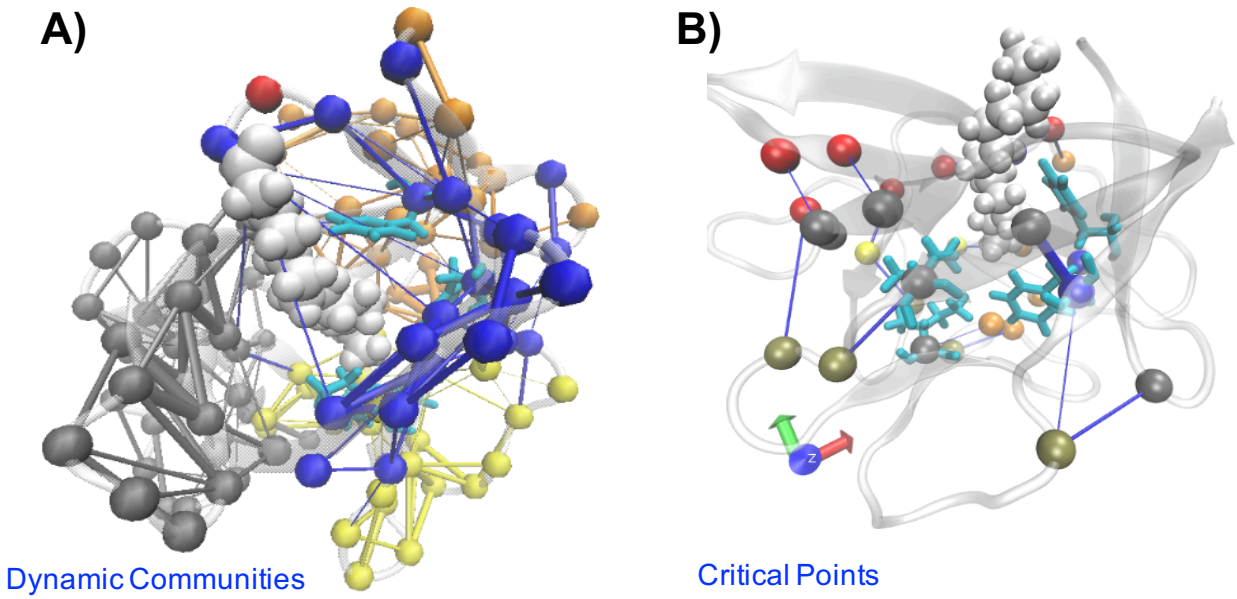


Figure 4.4 Dynamic Network Analysis of Myristoylated Hisactophilin at pH 7. Dynamic network analysis identifies five distinct subcommunities which describe the communication between residues shown in Panel A. Panel B displays the critical nodes which join the various subcommunities.

salt-bridges formed predominantly by GLU-17, GLU-19 and ASP-29 with neighboring histidine and lysine residues. The critical point analysis produces nodes which are important in the global communication of the protein and possibly stability, as they track communication between subcommunities. It is encouraging to see that in panel B, three (PHE-113, ILE-85 and ILE-93) out of the four residues responsible for altering the protein's switching behavior are represented as critical. PHE-113 which is a critical node in the center of the β–trefoil forms an edge with HIS-35 and in turn HIS-66, which are also designated as critical residues. Interestingly, ILE-85 and ILE-93 belong to separate subcommunities, gray and yellow respectively.

4.4 Conclusion and Discussion

Hisactophlin is a highly charged pH-sensor that enables the binding of actin to the intracellular plasma membrane. Understanding the residue specific influence of the titratable groups is invaluable in uncovering the pH-dependent switching mechanism. In this work, we

examine the impact of myristoylation on pK_a values by utilizing pH-REX CpHMD^{MSLD} simulations. We identify 13 histidine residues which exhibit large pK_a fluctuations upon myristoylation. Most of the shifted histidines are solvent exposed and located towards the N-terminus of the protein. In the sequestered state (pH > 6.5), the myristoyl group is buried in the protein cavity which is lined with hydrophobic residues such as phenylalanine (PHE-6, PHE-34, PHE-113), isoleucine (ILE-85, ILE-93) and valine (VAL-21, VAL-61 and VAL-101). Unsurprisingly, NMR studies detect NOEs between the myristoyl and residues PHE-6, PHE-113, ILE-85 and ILE-93[11]. Furthermore, experimental studies demonstrate that the accommodation of the lipid group in the hydrophobic core of the protein allows for global protein stability[10]. When the histidine residues possess a positive charge at pH values lower than 6.5, the myristoyl group is solvent exposed as the protein has large hydrophilic regions. Upon deprotonation of the histidine residues at higher pH, the overall charge of the protein is neutralized especially in the N-terminus, and the myristoyl group can be docked in the hydrophobic cleft of the β–trefoil. Importantly, HIS-75 and HIS-35 which are among the two histidine residues that exhibit largely depressed pK_a values (~3 and ~4 accordingly) upon myristoylation are located on the rim of the myristoyl cavity.

To visualize the switching mechanism, implicit solvent simulations were performed for the switching pH-range (6-8). Only 30% of the implicit solvent simulations that were conducted demonstrated the switching behavior, suggesting that the simulations need to be run longer than 20ns. However, for the simulations where the switching events were captured, distance measurements confirmed that the sequestered and accessible states were sampled. The myristoyl group at pH 6 had the largest displacement from the protein cavity at ~17Å, followed by pH 8 and pH 7 at 5.81 Å and 4.68 Å accordingly. These simulations were integral in qualitative description of the switching behavior as well as identifying non-native interactions formed by the myristoyl

group while in the accessible state. Once ejected from the protein cavity, the myristoyl group adopts a curled conformation and forms transient interactions with the residues of the C-terminus. This behavior agrees well with the large chemical shift perturbations identified by NMR experiments, in the same region of the protein[10].

Finally, dynamic network analysis lends support to the importance of the four residues found in the center of the hisactophilin cavity as PHE-113, ILE-85 and ILE-93 are identified as critical nodes. Although the mechanism of switching is largely controlled by the protonation states of the histidine residues, hydrophobic groups in the barrel of the protein cavity are essential to guide the myristoyl group into the internal hydrophobic protein core. PHE-113 and PHE-6 are staggered on either sides of the protein cleft and demonstrate stabilization of χ_2 angles upon myristoylation (data not shown). Furthermore, the contacts between the phenylalanine residues and the myristoyl group are sustained throughout the pH 7 and pH 8 simulations suggesting the role of these residues in the internal coordination of the myristoyl. Isoleucine residues found at the bottom of the cleft along with the three valine residues serve to further anchor the myristoyl group inside the protein. Mutational studies performed by the Meiering group emphasize the importance of these residues both in providing steric support and an appropriate electrostatic environment, as mutation of these residues causes the pH-switch to break.

Ultimately, the intentional distribution of hydrophobic and hydrophilic regions within a protein allows for a delicate but specific response to environmental pH. Broadly, this principal provides insight into the mechanistic details of hisactophilin and may generally inform the design of synthetic pH-dependent biosensors.

4.5 Bibliography

1. Guntas, G., Mansell, T.J., Kim, J.R. and Ostermeier, M., 2005. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proceedings of the National Academy of Sciences*, *102*(32), pp.11224-11229.

2. Ha, J.H. and Loh, S.N., 2012. Protein conformational switches: from nature to design. *Chemistry–A European Journal*, *18*(26), pp.7984-7999.

3. Bito, H., Deisseroth, K. and Tsien, R.W., 1996. CREB phosphorylation and dephosphorylation: a Ca2+-and stimulus duration–dependent switch for hippocampal gene expression. *Cell*, *87*(7), pp.1203-1214.

4. Dueber, J.E., Yeh, B.J., Chak, K. and Lim, W.A., 2003. Reprogramming control of an allosteric signaling switch through modular recombination. *Science*, *301*(5641), pp.1904-1908.

5. Brakemann, T., Stiel, A.C., Weber, G., Andresen, M., Testa, I., Grotjohann, T., Leutenegger, M., Plessmann, U., Urlaub, H., Eggeling, C. and Wahl, M.C., 2011. A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nature biotechnology*, *29*(10), p.942.

6. Boyken, S.E., Benhaim, M.A., Busch, F., Jia, M., Bick, M.J., Choi, H., Klima, J.C., Chen, Z., Walkey, C., Mileant, A. and Sahasrabuddhe, A., 2019. De novo design of tunable, pH-driven conformational changes. *Science*, *364*(6441), pp.658-664.

7. Ambroggio, X.I. and Kuhlman, B., 2006. Design of protein conformational switches. *Current opinion in structural biology*, *16*(4), pp.525-530.

8. Hanakam, F., Gerisch, G., Lotz, S., Alt, T. and Seelig, A., 1996. Binding of hisactophilin I and II to lipid membranes is controlled by a pH-dependent myristoyl– histidine switch. *Biochemistry*, *35*(34), pp.11036-11044.

9. Habazettl, J., Gondol, D., Wiltscheck, R., Otlewski, J., Schleicher, M. and Holak, T.A., 1992. Structure of hisactophilin is similar to interleukin-1β and fibroblast growth factor. *Nature*, *359*(6398), pp.855-858.

10. Shental-Bechor, D., Smith, M.T., MacKenzie, D., Broom, A., Marcovitz, A., Ghashut, F., Go, C., Bralha, F., Meiering, E.M. and Levy, Y., 2012. Nonnative interactions regulate folding and

switching of myristoylated protein. *Proceedings of the National Academy of Sciences*, *109*(44), pp.17839-17844.

11. Smith, M.T., Meissner, J., Esmonde, S., Wong, H.J. and Meiering, E.M., 2010. Energetics and mechanisms of folding and flipping the myristoyl switch in the β-trefoil protein, hisactophilin. *Proceedings of the National Academy of Sciences*, *107*(49), pp.20952-20957.

12. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.A. and Karplus, M., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), pp.187-217.

13. Feig, M., Karanicolas, J. and Brooks III, C.L., 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling*, *22*(5), pp.377-395.

14. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, *79*(2), pp.926-935.

15. Huang, J. and MacKerell Jr, A.D., 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry*, *34*(25), pp.2135-2145.

16. Best, R.B., Zhu, X., Shim, J., Lopes, P.E., Mittal, J., Feig, M. and MacKerell Jr, A.D., 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *Journal of chemical theory and computation*, *8*(9), pp.3257-3273.

17. Pastor, R.W. and MacKerell Jr, A.D., 2011. Development of the CHARMM force field for lipids. *The journal of physical chemistry letters*, *2*(13), pp.1526-1532

18. Arthur, E.J. and Brooks III, C.L., 2016. Efficient implementation of constant pH molecular dynamics on modern graphics processors. *Journal of computational chemistry*, *37*(24), pp.2171-2180.

19. Im, W., Lee, M.S. and Brooks III, C.L., 2003. Generalized born model with a simple smoothing function. *Journal of computational chemistry*, *24*(14), pp.1691-1702.

20. Chen, J., Im, W. and Brooks, C.L., 2006. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *Journal of the American Chemical Society*, *128*(11), pp.3728-3736.

21. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, *23*(3), pp.327-341.

22. Goh, G.B., Hulbert, B.S., Zhou, H. and Brooks III, C.L., 2014. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: structure, function, and bioinformatics*, *82*(7), pp.1319-1331.

23. Eargle, J. and Luthey-Schulten, Z., 2012. NetworkView: 3D display and analysis of protein· RNA interaction networks. *Bioinformatics*, *28*(22), pp.3000-3001.

24. Humphrey, W., Dalke, A. and Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), pp.33-38.

25. Hammond, M.S., Houliston, R.S. and Meiering, E.M., 1998. Two-dimensional 1H and 15N NMR titration studies of hisactophilin. *Biochemistry and cell biology*, *76*(2-3), pp.294-301.

26. Liao, C., May, V. and Li, J., 2019. Assessment of Conformational State Transitions of Class B GPCRs Using Molecular Dynamics. In *G Protein-Coupled Receptor Signaling* (pp. 3-19). Humana Press, New York, NY.

27. Despalatović, L., Vojković, T. and Vukicevic, D., 2014, May. Community structure in networks: Girvan-Newman algorithm improvement. In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 997-1002). IEEE.

CHAPTER 5

Conclusions and Future Directions

This work outlines the advancement towards incorporating pH into currently available NMR chemical shift prediction schemes. Motivated by the contributions of David Case[1,2] and Harold Scheraga[3-5], quantum mechanical calculations of model systems are used to decompose the complex influence of pH on chemical shifts. Furthermore, the general objective of this work is to develop a set of rules that robustly describe the complex impact of pH on these experimental observables. The ability to predict chemical shifts will ultimately appeal broadly to experimentalists and theorists, as it will allow for better interpretation of both spectra and simulations.

Model peptides are characteristic of early NMR studies, as many fundamental structural relationships were discovered with tri- and penta-peptides. By using a combination of molecular dynamics simulations and quantum mechanical calculations on tri-peptides containing titratable residues, we studied the impact of protonation on chemical shifts. Fully protonated and deprotonated tri-peptide ensembles allowed us to identify conformational characteristics that were distinct to each charge state. As the random coil ensembles converged, chemical shift perturbations were computed and demonstrated good agreement with experiment. The implicit solvent environment allowed for extensive conformational sampling and the elimination of intermolecular hydrogen bonding. The chemical shifts derived from these simulations enabled the elucidation of the pH contribution for ^{15}N and ^{13}C nuclei. However, the dependence of ^{1}HN chemical shifts on

solvent was apparent as the correlations of these results with experimental values was poor. The addition of explicit solvent molecules rectified the initial overestimation of the ^{1}HN shifts and yielding better agreement. The observed pH-dependent chemical shifts of the tri-peptides can be used to fortify available chemical shift prediction paradigms by augmenting the random coil chemical shift values, which are foundational for all chemical shift prediction programs.

Hen egg white lysozyme (HEWL), a favorite for benchmarking due to the plethora of available experimental data[6], served as an appropriate case study to test the ability of peptide chemical shift perturbations to reflect chemical shift variations in the protein environment. Explicit solvent pH-REX CpHMD simulations were used to calculate the micro-pK_a values of the ten titratable residues present in HEWL. The peptide chemical shift perturbations were added to SPARTA+ predicted chemical shifts of the CpHMD ensembles and produced reasonable agreement with experiment for ^{15}N nuclei. As expected, the ^{1}HN atoms proved challenging as both implicit and explicit variations of the pH-dependent peptide correction did not always yield the experimental result. Ultimately the micro-pK_a values differed from the macroscopic pK_a values computed from the chemical shifts suggesting that the latter is more directly comparable to experimentally derived pK_a values. Importantly, this study highlighted some of the short comings with applying the peptide pH-dependent shifts and the steps necessary to refine a pH-prediction model.

Many outstanding challenges in the accurate prediction of pH-dependent chemical shifts remain. Firstly, the pH-contribution is tethered to the many other variables which comprise the chemical shift and may be hard to simultaneously describe with one correction. The ^{15}N and ^{13}C chemical shifts are sensitive to torsional angles of neighboring residues, such that applying the pH-dependent shift from a titratable group in a tri-peptide construct, carries the memory of its

neighboring glycine. The effect of nearest neighbors in this context is not fully deciphered, but for aspartic acid and histidine, the overestimation of ^{15}N perturbations proved to be significant in the HEWL work. Nonetheless, the framework established in the peptide study may be used to further explore these effects.

Moreover, the sensitivity of chemical shifts to protein conformation underscores the requirement for extensive computational sampling. NMR chemical shift measurements represent the average chemical shift of an ensemble of states, and as such prediction paradigms should be applied on a large number of structures. As such, the representation of conformational states in the simulated ensemble must reflect those in the experiment. This is especially true for proteins that exhibit conformational plasticity. In the future, advances in the utilization of GPUs[7] and the development of enhanced sampling techniques[8] will continue to enable theorists to push the bounds of the time limits imposed on atomistic simulations such that more comprehensive protein ensembles can be obtained. Along with extensive sampling, as the accuracy of force fields progresses with the incorporation of more data and machine learning the protein ensembles generated by molecular dynamics simulations will be drastically improved. Ultimately the prediction of pH-dependent NMR chemical shifts encompasses many challenges, predominantly stemming from the ability to capture the experimental conformational ensemble.

In the third study, the pH-dependent switching mechanism of hisactophilin was studied with the use of pH-REX CpHMD and implicit solvent CpHMD. Significant pK_a differences were observed between the myristoylated and non-myristoylated hisactophilin proteins in the switching pH range 6.5 to 7.5. Implicit solvent simulations allowed for the visualization of the pH-switching mechanism and dynamic network analysis hints at the trefoil barrel residues as playing an important role. Further NMR experiments will benefit simulation analysis and help elucidate the

large perturbations which are identified near the N-terminus during the explicit solvent simulations. More specifically, mutational studies of HSP-35 and HSP-78 which are the residues identified in the simulations as large contributors of charge change in the switching pH range will be valuable in better understanding how they facilitate the myristoyl switching mechanism.

As structural biology looks to computation for guidance, the seamless comparison between experimental and computational observables will be paramount. In the context of pH, recent studies comparing pK_a apparent values with the computed micro-pK_a are beginning to emerge[9,10] as theorists look for a way to seamlessly compare experimental and computational observations. However, a method to integrate pH-dependent NMR chemical shifts in currently available predictors must occur in conjunction with the accurate prediction of pK_a values. Similar to the homology calculations conducted by the chemical shift predictor SHIFTX2 with the incorporation of the program SHIFTY[11], pK_a calculation with programs such as PROPKA[12] can be utilized to apply the pH dependent NMR correction as described in Chapters 2 and 3 of this work. In proteins pK_a values of titratable groups may differ from model pK_a values and consequently the applied pH-dependent chemical shift contribution is not constant among all titration events. In the future, the incorporation of a pK_a predictor within the chemical shift prediction algorithm will allow for pH-dependent NMR chemical shift calculations. Furthermore, as computational speed increases with advances in software and hardware, methods such as DFT will not be as time prohibitive and quantum mechanically derived chemical shifts of entire proteins and possibly entire conformational ensembles may be achieved. AF-QM/MM[13] (automated fragmentation quantum-mechanical molecular mechanics) is one current approach which fragments proteins to produce a series of manageably sized DFT jobs which can be submitted in parallel allowing for the computation of NMR chemical shifts for an entire protein. Although this method is still hindered

by time limitations, quantum mechanical studies are very sensitive to input geometry and in the future may provide a useful alternative method for the computation of protein NMR chemical shifts.

5.1 Bibliography

1. Osapay, K. and Case, D.A., 1991. A new analysis of proton chemical shifts in proteins. *Journal of the American Chemical Society*, *113*(25), pp.9436-9444.

2. Moon, S. and Case, D.A., 2007. A new model for chemical shifts of amide hydrogens in proteins. *Journal of biomolecular NMR*, *38*(2), p.139.

3. Vila, J.A., Aramini, J.M., Rossi, P., Kuzin, A., Su, M., Seetharaman, J., Xiao, R., Tong, L., Montelione, G.T. and Scheraga, H.A., 2008. Quantum chemical 13Cα chemical shift calculations for protein NMR structure determination, refinement, and validation. *Proceedings of the National Academy of Sciences*, *105*(38), pp.14389-14394.

4. Villegas, M.E., Vila, J.A. and Scheraga, H.A., 2007. Effects of side-chain orientation on the 13 C chemical shifts of antiparallel β-sheet model peptides. *Journal of biomolecular NMR*, *37*(2), pp.137-146.

5. Martin, O.A., Vila, J.A. and Scheraga, H.A., 2012. Che Shift-2: graphic validation of protein structures. *Bioinformatics*, *28*(11), pp.1538-1539.

6. Webb, H., Tynan-Connolly, B.M., Lee, G.M., Farrell, D., O'Meara, F., Søndergaard, C.R., Teilum, K., Hewage, C., McIntosh, L.P. and Nielsen, J.E., 2011. Remeasuring HEWL pKa values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pKa calculations. *Proteins: Structure, Function, and Bioinformatics*, *79*(3), pp.685-702.

7. Friedrichs, M.S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A.L., Ensign, D.L., Bruns, C.M. and Pande, V.S., 2009. Accelerating molecular dynamic simulation on graphics processing units. *Journal of computational chemistry*, *30*(6), pp.864-872.

8. Friedrichs, M.S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A.L., Ensign, D.L., Bruns, C.M. and Pande, V.S., 2009. Accelerating molecular dynamic simulation on graphics processing units. *Journal of computational chemistry*, *30*(6), pp.864-872.

9. Dobrev, P., Vemulapalli, S.P.B., Nath, N., Griesinger, C. and Grubmüller, H., 2020. Probing the accuracy of explicit solvent constant pH molecular dynamics simulations for peptides. *Journal of Chemical Theory and Computation*, *16*(4), pp.2561-2569.

10. Hofer, F., Dietrich, V., Kamenik, A.S., Tollinger, M. and Liedl, K.R., 2019. pH-Dependent Protonation of the Phl p 6 Pollen Allergen Studied by NMR and cpH-aMD. *Journal of chemical theory and computation*, *15*(10), pp.5716-5726.

11. Wishart, D.S., Watson, M.S., Boyko, R.F. and Sykes, B.D., 1997. Automated 1 H and 13 C chemical shift prediction using the BioMagResBank. *Journal of biomolecular NMR*, *10*(4), pp.329-336.

12. Olsson, M.H., Søndergaard, C.R., Rostkowski, M. and Jensen, J.H., 2011. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of chemical theory and computation*, *7*(2), pp.525-537.

13. He, X., Wang, B. and Merz Jr, K.M., 2009. Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *The Journal of Physical Chemistry B*, *113*(30), pp.10380-10388.

Appendix A.

Supplementary Information for Chapter 2

Table S2.1 Computed absolute chemical shifts of carbon nuclei

Residue	Atom Type	δ Dep. (ppm)	δ Prot. (ppm)	δ Dep. (ppm)	δ Prot. (ppm)
		comp		exp[30]	
ASP	Cα	64.0	60.6	54.3	52.9
	Cβ	50.7	43.9	41.1	38.0
	Cγ	187.4	185.0	180.3	177.1
	CO	182.0	182.0	174.6	174.8
GLU	Cα	63.4	62.2	56.9	56.0
	Cβ	39.1	36.5	30.0	28.5
	Cγ	43.9	37.2	36.1	32.7
	Cδ	191.1	188.1	183.8	179.7
	CO	181.5	182.0	177.0	176.5
HIS	Cα	63.9	61.3	56.7	55.1
	Cβ	37.3	35.0	31.3	28.9
	Cγ	146.6	143.0	135.3	131.0
	Cδ2	127.4	129.8	120.0	120.3
	Cε1	145.4	143.1	139.2	136.6
	CO	181.8	182.4	176.2	174.8
LYS	Cα	63.2	63.0	56.9	56.4
	Cβ	40.3	39.8	33.2	32.8
	Cγ	30.5	29.6	25.0	24.7
	Cδ	42.2	34.7	33.9	28.9
	Cε	51.2	53.0	43.1	42.1
	CO	181.9	181.8	177.5	177.0

Table S2.1 Computed absolute chemical shifts of carbon nuclei. Computed chemical shifts from averaged frames collected from the MD ensembles are tabulated with the corresponding experimental chemical shifts obtained from Platzer *et al.*[30].

Table S2.2 Computed absolute chemical shifts of proton nuclei

Residue	Atom Type	δ Dep. (ppm)	δ Prot. (ppm)	δ Dep. (ppm)	δ Prot. (ppm)
		comp		exp[30]	
ASP	HN	5.66	4.97	8.38	8.55
	Hα	2.91	3.25	4.61	4.78
	Hβ (avg.)	0.92	1.34	2.70	2.93
	Hδ2	--	5.34	--	> 10
GLU	HN	4.97	4.64	8.57	8.45
	Hα	2.81	2.92	4.29	4.39
	Hβ (avg.)	0.53	0.66	2.02	2.08
	Hγ (avg.)	0.65	1.04	2.27	2.49
	Hε2	--	5.22	--	> 10
HIS	HN	6.15	4.75	8.35	8.55
	Hα	3.09	3.25	4.59	4.75
	Hβ (avg.)	1.68	1.92	3.08	3.25
	Hδ2	5.66	5.98	6.97	7.30
	Hε1	6.29	6.87	7.68	8.60
	Hε2	12.13	8.19	--	> 10
LYS	HN	4.71	4.72	--	8.40
	Hα	2.80	2.83	4.30	4.34
	Hβ (avg.)	0.33	0.38	1.78	1.82
	Hγ (avg.)	-0.09	0.06	1.36	1.44
	Hδ (avg.)	0.02	0.41	1.44	1.68
	Hε (avg.)	1.29	2.00	2.60	3.00
	Hζ1	-0.63	3.16	~1 - 2	7.52
	Hζ2	-0.59	3.17	~1 - 2	7.52
	Hζ3	--	3.18	~1 - 2	7.52

Table S2.2 Computed absolute chemical shifts of proton nuclei. Computed chemical shifts from averaged frames collected from the MD ensembles are tabulated with the corresponding experimental chemical shifts obtained from Platzer *et al.*[30].

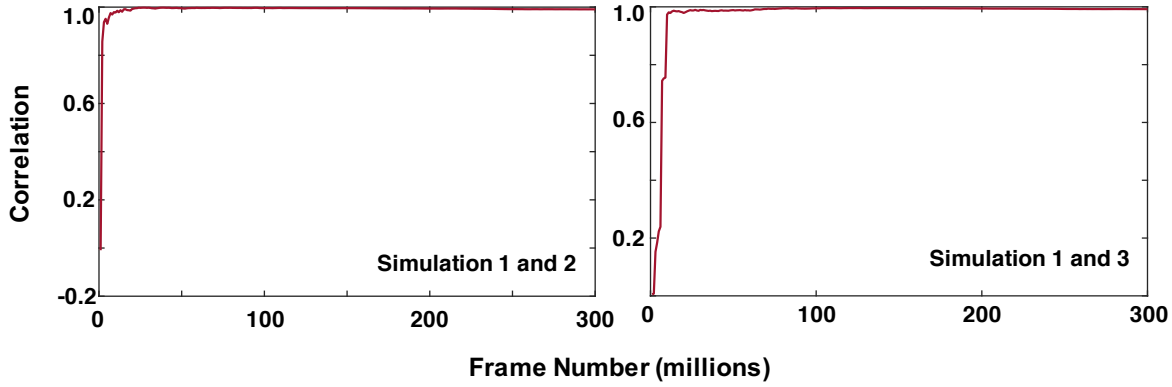
Table S2.3 Computed absolute chemical shifts of nitrogen nuclei

Residue	Atom Type	δ Dep. (ppm)	δ Prot. (ppm)	δ Dep. (ppm)	δ Prot. (ppm)
		comp		exp[30]	
ASP	N	117.3	112.1	120.2	118.7
GLU	N	116.5	114.5	120.9	119.9
HIS	N	114.9	109.6	119.7	117.9
	Nδ1	243.9	170.1	231.3	175.8
	Nε2	179.1	166.1	181.1	173.1
LYS	N	114.3	114.2	121.7	121
	Nζ	10.4	22.4	~25.2	32.7

Table S2.3 Computed chemical shifts from averaged frames collected from the MD ensembles are tabulated with the corresponding experimental chemical shifts obtained from Platzer *et al.*[30].

Figure S2.1 Simulation Convergence

A. Aspartic Acid: Deprotonated



B. Aspartic Acid: Protonated

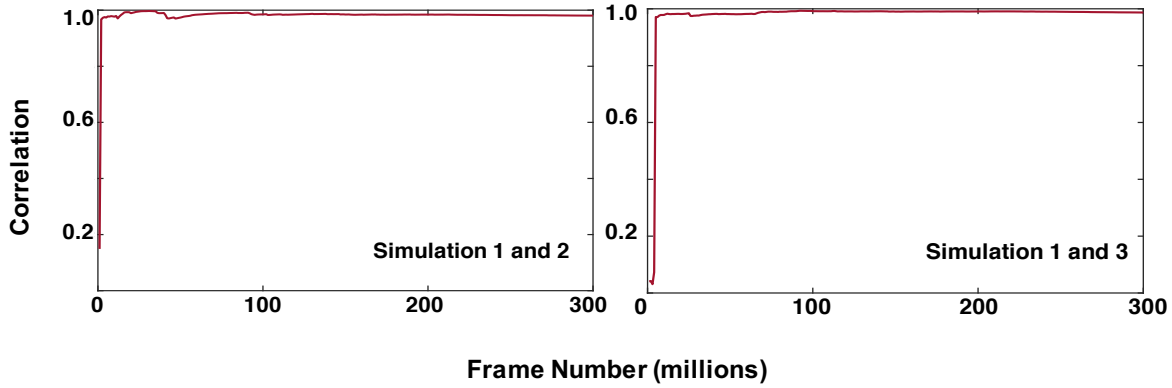


Figure S2.1 Simulation Convergence. The phi and psi angles of titratable groups Asp, Glu, His, and Lys for 300 million frames from respective 650 ns MD simulations were binned (20 x 20 grid size). Binned distribtutions of simulation runs 1 and 2, and runs 1 and 3 are correlated and plotted as a function of frame number. Results for deprotonated (panel A) and protonated (panel B) Asp are displayed. The results from these figures suggest that the convergnece of the phi/psi distributions for all peptides occurs at a correlation at or above 0.94 (data not shown).