UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Thermodynamics of Charge Regulation: From Small Molecules to Polymers and Surfaces

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Alejandro A. Gallegos

September 2022

Dissertation Committee:
Dr. Jianzhong Wu, Chairperson
Dr. Younjin Min
Dr. Chia-en Chang

The Dissertation of Alejandro A. Gallegos is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

This dissertation is the result of my five difficult, yet rewarding years as a Ph. D. student under the supervision of Dr. Jianzhong Wu. I owe a lot to Dr. Wu for his continued supervision and support during my Ph. D. and his ability to guide us during our challenging moments. We have had many fruitful discussions on a wide range of topics and new ideas which has helped me to better understand myself and my role as a researcher. I have learned a lot from him on how to run my own research group in the future and specifically how to comprehend the big picture, yet focus on the important problems that need attention. I look forward to our continued collaboration in the years to come.

I would like to express my gratitude to Dr. Chia-en Chang and Dr. Younjin Min for their time serving on my dissertation committee. I would also like to say thank you to Dr. Kandis Abdul-Aziz and Dr. Bryan Wong for serving on my qualifying examination committee. Although none of my collaborative work with the Fluid Interface Reactions Structures and Transport (FIRST) Center on the physics and design of supercapacitors is included in this dissertation, they have played a key role in my academic and personal development and I thank them for this. Within Dr. Jianzhong Wu's group, I have had the pleasure to interact with many wonderful and smart people. I would first like to thank my colleague and friend, Dr. Musen Zhou, who has been invaluable to my academic and personal growth. I am truly grateful for his friendship and I wish him the best in his new position. I would also like to thank Dr. Kun Liu for the many years of guidance and support she gave during my time as an undergraduate and Ph. D. student in Dr. Wu's group. Lastly, I am very fortunate to have started my research career under the supervision of Dr. Xian

Kong and Dr. Cheng Lian and would like to acknowledge them for their guidance.

Most importantly, I would like to express my deepest and most sincere thanks to my partner, Sabrina, for her unconditional love and support during my time as a Ph. D. student. She has been my rock through it all and I would not have succeeded without her. To my two kids, Elora and Nathaniel, I am truly grateful to have you both in my life. Your love and laughter motivate me to be my best and help me through even the most trying of times. To my parents, I thank you for the love and guidance you have given me in life. You have taught me how to face and overcome the difficulties I face and always pushed me to succeed in what I do.

The content of this dissertation, in part or in full, is a reprint of the materials as it appears in *Journal of Chemical & Engineering Data* (Volume 65, Issue 12, Pages 5630-5642, November 2020), *Soft Matter* (Volume 17, Pages 9221-9234, August 2021), *AIChE Journal* (Volume 68, Issue 2, September 2021), and *The Journal of Chemical Physics* (Volume 155, Issue 24, Pages 241102, December 2021).

The work completed in this dissertation is financially supported by the U.S. National Science Foundation Graduate Research fellowship under Grant No. NSF-DGE-1326120 and the U.S. National Science Foundation Harnessing the Data Revolution (HDR) Big Ideas Program under Grant No. NSF 1940118. Parts of computation in this dissertation used resources of the National Energy Research Science Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy, under Contract DE-AC02-05CH11231.

ABSTRACT OF THE DISSERTATION

Molecular Thermodynamics of Charge Regulation:
From Small Molecules to Polymers and Surfaces

by

Alejandro A. Gallegos

Doctor of Philosophy, Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, September 2022
Dr. Jianzhong Wu, Chairperson

Charge regulation of monomers, polymers, and surfaces have attracted great interest in recent years because of its fundamental and practical importance to many technological applications such as bioadhesion and drug delivery. The ability to tune the properties of these ionizable species opens avenues to "smart" systems which can respond to changes in pH, salt concentration, and other environmental factors. A better understanding of when and why the electrostatic charge of these species emerges and fluctuates will allow for the more rational design to meet targeted needs. Unfortunately, the chemistry at the microscopic scale is difficult to discern and an adequate theoretical tool is needed to provide insights into the mechanisms governing the charge regulation of these species.

The purpose of this dissertation is to advance the knowledge of the charge regulation through the innovation of new theoretical methods to characterize the properties of ionizable systems. In particular, we emphasize the applications of our new theoretical methods to describe the interfacial phenomena. These new tools pave the way towards realistic models to understand bioadhesion, thermal energy storage, drug delivery, waste-

water treatment and many more applications. We first developed a molecular thermodynamic model that is able to describe the charge regulation and activity coefficients of amino acids in excellent agreement with experimental data. Next, we extended the thermodynamic model using classical density functional theory to predict the interfacial behavior of amino acids at inorganic surfaces. Different from conventional methods, our model accounts for key correlation effects that dictate the charge behavior of amino acids near an interface. We also developed a coarse-grained model that can be used to predict the ionization of weak polyelectrolytes in a solution with physically realistic parameters. In order to describe weak polyelectrolytes in non-uniform fluids such as near a surface, we developed a theoretical tool known as the Ising density functional theory (iDFT), which bridges the gap between the site-binding model and polymer density functional theory. We demonstrated that iDFT is able to capture the adsorption isotherms of polypeptides near ionizable inorganic surfaces in good agreement with experimental data. The last two chapters of this dissertation are focused on theoretical developments for incorporating long-range intrachain correlations that are neglected in conventional treatments of polymers in a bulk solution or near an interface. The inclusion of long-range interactions provides an accurate description of the coupling of polymer charge and conformation in weak polyelectrolyte systems. Lastly, we combined iDFT with the single-chain-in-mean-field algorithm to capture the long-range two-body correlation effects in non-uniform fluids. The computational framework developed in this dissertation opens up new opportunities for engineering design of ionizable molecular systems.

Table of Contents

List of Figures

List of Tables

List of Supporting Figures

List of Supporting Tables

Chapter 1. Introduction

1.1 Scope of research

Understanding the charge behavior of weak polyelectrolytes is critical to technological applications such as bioadhesion[1, 2], drug delivery[3-5], wastewater treatment[6], and thermal energy storage[7]. However, due to the long-range nature of electrostatic interactions and the resulting intrachain correlations, it remains a theoretical challenge to accurately describe weak polyelectrolytes in solutions owing to the intrinsic coupling between ionization and polymer conformation. A poor description of these effects may lead to erroneous predictions of "smart" polymeric systems to achieve their specific functionality such as targeted drug delivery and controlled release by taking advantage of the changes in the ionization and structure of the polymer backbone in response to the pH of the intracellular fluids[8, 9]. In addition, the behavior of the polymer also usually feels external stimuli through an interface, particularly one that is charged, which influences the properties of the polymer. Careful considerations of these effects are necessary for technological design.

The ionization and conformation of a weak polyelectrolyte chain are dictated by the segment-level behavior through its inter- and intrachain interactions. A successful description of these properties requires an understanding on not only how the polymer interacts within itself, but also how the polymer, and more specifically its individual sites, interacts with the local environment. Thus, it is important to accurately describe the individual segments of the polymer including these residues when they are free in the solution (i.e., not in a polymer chain). Besides the interactions within the fluid, the surface

itself may be ionizable through the protonation and deprotonation of its functional groups. The chemistry occurring at the interface is relevant for a broad set of applications since many surfaces form a hydroxyl layer that can either be protonated (positively charged) or deprotonated (negatively charged) depending on its local environment (e.g., pH or salt concentration).

This dissertation is focused on the development of new physics-based models and its integration into new theoretical methods to achieve a comprehensive picture of charge regulation from small molecules to polymers and surfaces. We are interested in the charge behavior of monomers to understand their interactions with other species in solution so that a better description of polymers is possible. In addition, the interaction of monomers with surfaces is extremely relevant to the food industry that has to deal with surface fouling due to the adsorption of the monomers[10]. Conventional methods fail to capture the correlation effects due to excluded volume and electrostatics that are important for the interfacial phenomena[11, 12]. Our models provide a means to capture these effects and go beyond these methods to accurately predict the adsorption of amino acids to inorganic surfaces across a myriad of solution conditions. With an accurate description of charge regulation in monomers and surfaces, we next focused on the charge regulation of polymers which distinguish themselves from their monomeric counterparts through the intrachain interactions. Previous methods relied on fitting parameters within the site-binding model or considering a mean-field approach to charge regulation (i.e., monomers within a chain are independent of one another)[13]. A major focus of this dissertation was to demonstrate that the key correlation effects due to inter- and intrachain interactions could be

successfully incorporated into a molecular thermodynamic model allowing for semi-quantitative prediction of the properties of polymers in bulk and inhomogeneous conditions. Specifically, we developed new theoretical tools like Ising density functional theory (iDFT) to accurately capture the charge behavior of weak polyelectrolytes[14]. The combination of new coarse-grained models and new theoretical tools demonstrated in this dissertation opens up the opportunity to accurately and efficiently describe the charge regulation of monomers, polymers, and surfaces.

1.2 Recent progress in the field of charge regulation

Charge regulation has received significant attention in recent years due to its presence in a myriad of practical and theoretical applications. While it has always been known that the behavior of ionizable molecules is likely different near interfaces than in a bulk solution, a faithful description of this phenomena was difficult because of the diverse number of interactions and correlation effects at play. In this section, we outline the progress made in the description of charge regulation for small molecules, polymers, and surfaces.

1.2.1 Charge regulation of small molecules

The classical treatment of small molecules is via the Henderson-Hasselbach equation[15]:

$$\mathrm{pH} = \mathrm{pK} + \log\left(\frac{c_{A^-}}{c_{AH}}\right) \tag{1.1}$$

which states that the dissociation of the acid (written in terms of concentration c_i) from its neutral form to ionized form is related to the pH of the solution and the thermodynamic equilibrium constant K governing the reaction. Unfortunately, this equation only holds in the dilute (i.e., ideal) limit where each species has no interactions with all other species in the solution. In general, we must account for the interactions that occur between the different species and the rest of the solution via the activity coefficient γ_i. A generalized form of the previous equation can be rewritten in terms of the activity of a species: $a_i = c_i \gamma_i$:

$$\mathrm{pH} = \mathrm{pK} + \log\left(\frac{a_{A^-}}{a_{AH}}\right) = \mathrm{pK} + \log\left(\frac{c_{A^-}}{c_{AH}}\right) + \log\left(\frac{\gamma_{A^-}}{\gamma_{AH}}\right) \tag{1.2}$$

where we emphasize that the pH is related to the activity of the proton (viz. $\mathrm{pH} = -\log a_{H^+}$). Thus, there is one additional term that accounts for the thermodynamic non-ideality resulting from the intermolecular interactions by the protonated and deprotonated form of the acid. Since the concentrations and pH can generally be measured experimentally, the thermodynamic equilibrium constant and the activity coefficients term are typically lumped together into an *apparent* equilibrium constant[16]:

$$\mathrm{pK_{app}} = \mathrm{pK} + \log\left(\frac{\gamma_{A^-}}{\gamma_{AH}}\right). \tag{1.3}$$

The apparent equilibrium constant describes the ionization behavior of the acid at the solution conditions (e.g., salt concentration). It is often used as benchmark for theoretical

methods to test their capability in predicting the influence of solution conditions on the charge regulation of the monomers.

Conventional theoretical approaches to describe the activity coefficients for the titration of monomers have relied on the specific ion interaction (SIT) model[17, 18] or Pitzer's equation[19]. Both rely on the Debye-Hückel (DH) theory to describe the electrostatic interactions between species, particularly at dilute salt conditions. It predicts that, at low salt concentrations, the activity coefficients of charged species (modeled as point charges) sharply falls as the ionic strength increases[20]. Because of the strong deviation from thermodynamic ideality even at low salt concentrations, the equilibrium constants cannot be reliably extrapolated from theoretical methods that do not reproduce the DH limiting law. As the DH theory becomes problematic when the salt concentration increases, the activity coefficients can be described by using various forms of the extended DH theory[20, 21]. At moderate to high salt concentrations, electrostatics alone is not enough to describe the apparent equilibrium constant measured experimentally. The reason for this failure can be attributed to the excluded volume effect which must be accounted for in some way whether through a linear term like in SIT or a more complicated term like in Pitzer's equation[22-26]. Over the years, a number of thermodynamic models have been developed for electrolyte solutions beyond semi-empirical modifications of the DH theory[27, 28]. Among them, the mean-spherical approximation (MSA) represents a popular choice owing to its computational simplicity to describe the ion-size effects and electrostatic correlations[29]. Importantly, it obeys the limiting law at low electrolyte concentrations similar to various empirical modifications of the DH theory[30]. While

existing methods do reasonably well at predicting either the activity coefficient or the charge regulation, none do well in predicting both together. Chapter 3 focuses on this problem and the coarse-grained model we developed to remedy the inconsistency.

As of yet, we have only focused on the bulk solution where the behavior of the different species is not influenced by an external stimulus. Many practical applications do involve interfaces such as adsorption, adhesion, etc.; thus, we must also understand charge regulation when the system is no longer uniform (i.e., inhomogeneous fluid) due to the presence of the interface. In this case, the typical approach is to employ a localized version of the Henderson-Hasselbalch equation in which the pH is defined as

$$\mathrm{pH}(\mathbf{r}) = -\log c_{H^+}(\mathbf{r}), \tag{1.4}$$

thus, the proton concentration at position **r** in the inhomogeneous fluid describes the local ionization behavior of the acid[13]:

$$\mathrm{pH}(\mathbf{r}) = \mathrm{pK} + \log\left[\frac{c_{A^-}(\mathbf{r})}{c_{AH}(\mathbf{r})}\right]. \tag{1.5}$$

Therefore, the degree of ionization of the acid at position **r** depends only on the proton concentration which can be determined by some mean-field approach (e.g., Poisson-Boltzmann[12], self-consistent field theory[31], etc.). For example, the Poisson-Boltzmann method gives the following expression for the distribution of protons in the inhomogeneous fluid:

$$c_{H^+}(\mathbf{r}) = 10^{-\mathrm{pH}} \exp\left[-\beta e \psi(\mathbf{r})\right], \tag{1.6}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant and T is the absolute temperature. Here, $\psi(\mathbf{r})$ is the mean electrostatic potential at position **r** which describes the direct coulomb contribution between the ionic species and the surface charge. Specifically, we can solve for the mean electrostatic potential via the Poisson equation:

$$\nabla^2 \psi(\mathbf{r}) = -4\pi l_B \sum_i \rho_i(\mathbf{r}), \tag{1.7}$$

where $l_B = \beta e^2 / 4\pi\varepsilon_0\varepsilon_r$ is the Bjerrum length (0.714 nm for water at room temperature), e is the elementary charge, ε_0 and ε_r is the vacuum permittivity and relative permittivity, respectively. The density profile of species i, $\rho_i(\mathbf{r})$, can be self-consistently determined with the mean electrostatic potential in the case of Poisson-Boltzmann method or through another means if additional contributions are accounted for like in density functional theory[32-34].

The use of $\mathrm{pH}(\mathbf{r})$, however, does not agree with the thermodynamics since pH should be defined in terms of the proton activity, not the local concentration, and it would therefore be constant throughout the entire fluid whether near the interface or the bulk. Instead, we can use the activities to arrive at the appropriate expression. The activity coefficient is related to the excess chemical potential by $\gamma_i = \exp(\mu_i^{ex})$, where the excess chemical potential measures the thermodynamic non-ideality due to the interactions present in the system. At the Poisson-Boltzmann level, $\mu_i^{ex}(\mathbf{r}) = \beta e \psi(\mathbf{r}) Z_i$ where Z_i is the valence. Therefore, the activity ratio for the charged to uncharged form of the acid is given by

$$\frac{a_{A^-}}{a_{AH}} = \frac{c_{A^-}}{c_{AH}} \frac{\gamma_{A^-}}{\gamma_{AH}} = \frac{c_{A^-}}{c_{AH}} \exp\left[-\beta e \psi(\mathbf{r})\right]. \tag{1.8}$$

Thus, we find that the ionization of the acid in the inhomogeneous fluid is determined from

$$\mathrm{pH} = \mathrm{pK} + \log\left[\frac{c_{A^-}(\mathbf{r})}{c_{AH}(\mathbf{r})}\right] - \beta e \psi(\mathbf{r}) \ln 10. \tag{1.9}$$

Interestingly, this expression is actually equivalent to *incorrect* expression given earlier (at the Poisson-Boltzmann level), except that this new expression was derived in a thermodynamically consistent manner. It is easy to see that a general expression can be given instead for acids as

$$\mathrm{pH} = \mathrm{pK} + \log\left[\frac{c_{A^-}(\mathbf{r})}{c_{AH}(\mathbf{r})}\right] + \beta\left[\mu^{ex}_{A^-}(\mathbf{r}) - \mu^{ex}_{AH}(\mathbf{r})\right] \ln 10, \tag{1.10}$$

where the specific terms that contribute to the excess chemical potential depend on the model. In the work considered in Chapter 4, we consider the contributions of hard-sphere effects, electrostatic correlations, and short-range attractions to the excess chemical potential through classical density functional theory[35].

1.2.2 Charge regulation of polymers

In conventional methods, the charge behavior of polymers has typically been treated in the same manner as that of monomers (i.e., by treating the monomers within a polymer chain as independent and only considering the intermolecular interactions)[13, 36]. However, a realistic description of polymers requires the inclusion of the intrachain correlations. While the conventional methods do provide insight into the properties of weak polymers in the bulk and near an interface[37-41], they are missing key physics that

results from the explicit consideration of interactions between segments within a single polymer chain. To start, we first consider weak polyelectrolytes immersed in a bulk solution. The conventional approach is through the site-binding model, specifically the nearest-neighbor site-binding model, which truncates the interactions between all segments within the polymer chain to only those that are adjacent to one another[42]. Despite the obvious simplification of the intrachain correlations, it provides a quick and efficient calculation of the thermodynamic properties of weak polymers and it has shown to be in excellent agreement with experimental titration data.

The basic idea of the site-binding model is to write the Hamiltonian for the polymer as

$$H(\mathbf{S}) = \sum_{i=1}^{M} \left[(\mathrm{pH} - \mathrm{pK}) \ln 10 + \Delta\mu_i^{ex} \right] s_i + \sum_{i=1}^{M-1} \Delta W_{i,i+1} s_i s_{i+1} \tag{1.11}$$

where the first term contains the chemical contribution (i.e., pH-pK term) due to the deprotonation of the sites as well as an additional contribution due to the difference in excess chemical potential between the charged form and uncharged form. If the Hamiltonian contains only the first summation, then it becomes equivalent to the expression given in the monomeric case in the previous section. However, the introduction of the second summation which sums over adjacent pairs is the intrachain correlation term. Successful description of weak polyelectrolytes including the step-like feature in its titration curve like in poly(maleic acid) can be attributed to this term[43]. In the conventional methods, the effects of solution conditions cannot be predicted and therefore the method relies on the fitting of $\Delta\mu_i^{ex}$ and $\Delta W_{i,i+1}$ to match the experimental work. The

purpose of Chapter 5 is to account for the solution effects using a molecular thermodynamic model within the framework of the nearest-neighbor site-binding model.

More recently, research efforts have focused on the inclusion of intrachain correlations beyond the nearest-neighbor since the conformation and ionization of polymers depend on the long-range intrachain correlation effects[44-46]. In this case, the Hamiltonian can be written as

$$H(\mathbf{S}) = \sum_{i=1}^{M} \left[(\mathrm{pH} - \mathrm{pK}) \ln 10 + \Delta\mu_i^{ex} \right] s_i + \sum_{i=1}^{M-1} \Delta W_{i,i+1} s_i s_{i+1} + \sum_{i=1}^{M-2} \sum_{j>i+1}^{M} \Delta\Psi_{i,j} s_i s_j \quad (1.12)$$

where the third term accounts for the long-range intrachain correlation effects. Unfortunately, direct evaluation of the thermodynamic properties is quite difficult when intrachain correlations beyond the nearest-neighbor are included. Others have proposed using a mean-field approach to include an effective potential that accounts for the average behavior of the long-range interactions[44] or to evaluate the Hamiltonian using Monte Carlo simulations[42]. In Chapter 6, we demonstrate a hierarchical model to determine the thermodynamic properties of weak polyelectrolyte solutions. As discussed earlier, the charge regulation of weak polymers in the non-uniform fluid relies on mean-field approach which does not account for the intrachain correlations. To improve upon this, we have developed a new theoretical tool shown in Chapter 7. Ising density functional theory (iDFT) accounts for the ionization state and position of all monomers in the weak polyelectrolyte chain. Therefore, we are able to include intrachain correlation effects between monomers in their given states and not treat them by their average behavior like in conventional methods. While iDFT is formally exact, we must approximate the

intrachain correlations to the nearest-neighbor level in order to perform the calculations. To remedy this, we utilize the single-chain-in-mean-field algorithm with the iDFT framework to capture the long-range interactions within the polymer chain in a fast and robust manner as shown in Chapter 9.

1.2.3 Charge regulation of surfaces

In conventional methods, the charge behavior of polymers has typically been treated in the same manner as that of monomers (i.e., by treating the monomers within a polymer chain as independent and only considering the intermolecular interactions)[13, 36]. However, a realistic description of polymers requires the inclusion of the intrachain correlations. While the conventional methods do provide insight into the properties of weak polymers in the bulk and near an interface[37-41], they are missing key physics that results from the explicit consideration of interactions between segments within a single polymer chain. To start, we first consider weak polyelectrolytes immersed in a bulk solution. The conventional approach is through the site-binding model, specifically the nearest-neighbor site-binding model, which truncates the interactions between all segments within the polymer chain to only those that are adjacent to one another[42]. Despite the obvious simplification of the intrachain correlations, it provides a quick and efficient calculation of the thermodynamic properties of weak polymers and it has shown to be in excellent agreement with experimental titration data.

The basic idea of the site-binding model is to write the Hamiltonian for the polymer as

$$H(\mathbf{S}) = \sum_{i=1}^{M} \left[(\mathrm{pH} - \mathrm{pK}) \ln 10 + \Delta\mu_i^{ex} \right] s_i + \sum_{i=1}^{M-1} \Delta W_{i,i+1} s_i s_{i+1} \tag{1.13}$$

where the first term contains the chemical contribution (i.e., pH-pK term) due to the deprotonation of the sites as well as an additional contribution due to the difference in excess chemical potential between the charged form and uncharged form. If the Hamiltonian contains only the first summation, then it becomes equivalent to the expression given in the monomeric case in the previous section. However, the introduction of the second summation which sums over adjacent pairs is the intrachain correlation term. Successful description of weak polyelectrolytes including the step-like feature in its titration curve like in poly(maleic acid) can be attributed to this term[43]. In the conventional methods, the effects of solution conditions cannot be predicted and therefore the method relies on the fitting of $\Delta\mu_i^{ex}$ and $\Delta W_{i,i+1}$ to match the experimental work. The purpose of Chapter 5 is to account for the solution effects using a molecular thermodynamic model within the framework of the nearest-neighbor site-binding model.

More recently, research efforts have focused on the inclusion of intrachain correlations beyond the nearest-neighbor since the conformation and ionization of polymers depend on the long-range intrachain correlation effects[44-46]. In this case, the Hamiltonian can be written as

$$H(\mathbf{S}) = \sum_{i=1}^{M} \left[(\mathrm{pH} - \mathrm{pK}) \ln 10 + \Delta\mu_i^{ex} \right] s_i + \sum_{i=1}^{M-1} \Delta W_{i,i+1} s_i s_{i+1} + \sum_{i=1}^{M-2} \sum_{j>i+1}^{M} \Delta\Psi_{i,j} s_i s_j \tag{1.14}$$

where the third term accounts for the long-range intrachain correlation effects. Unfortunately, direct evaluation of the thermodynamic properties is quite difficult when intrachain correlations beyond the nearest-neighbor are included. Others have proposed using a mean-field approach to include an effective potential that accounts for the average behavior of the long-range interactions[44] or to evaluate the Hamiltonian using Monte Carlo simulations[42]. In Chapter 6, we demonstrate a hierarchical model to determine the thermodynamic properties of weak polyelectrolyte solutions. As discussed earlier, the charge regulation of weak polymers in the non-uniform fluid relies on mean-field approach which does not account for the intrachain correlations. To improve upon this, we have developed a new theoretical tool shown in Chapter 7. Ising density functional theory (iDFT) accounts for the ionization state and position of all monomers in the weak polyelectrolyte chain. Therefore, we are able to include intrachain correlation effects between monomers in their given states and not treat them by their average behavior like in conventional methods. While iDFT is formally exact, we must approximate the intrachain correlations to the nearest-neighbor level in order to perform the calculations. To remedy this, we utilize the single-chain-in-mean-field algorithm with the iDFT framework to capture the long-range interactions within the polymer chain in a fast and robust manner as shown in Chapter 9.

1.3 Dissertation organization

The dissertation is focused on the description of charge regulation using density functional theory for the ionization behavior of monomers, polymers, and surfaces. There are in total 10 chapters included in this dissertation.

Following the introduction in Chapter 1, Chapter 2 describes the basic concepts and formulations of charge regulation and density functional theory (DFT), in particular, DFT applied to weak polyelectrolytes in which the multibody configuration (viz. state and spatial position of each segment) is accounted for.

Chapter 3 presents a predictive model for the electrostatic behavior of natural amino acids at different solution conditions by incorporating the thermodynamic non-ideality governing its interactions and charge regulation in solution. The thermodynamic model allows for predictions of the activity coefficients of amino acids and salt ions as well as the apparent equilibrium constants for each amino acid as a function of salt concentration. Importantly, we can extrapolate to determine the intrinsic equilibrium constant (i.e., the equilibrium constant at infinite dilution).

Chapter 4 investigates the behavior of amino acids at inorganic surfaces by using the coarse-grained model developed in Chapter 3. The model predicts the adsorption of different types of amino acids (viz. acidic, basic, and neutral) at various amino acid concentration, salt concentration, and pH with great success. The molecular theory accounts for the physical binding and surface associations as well as the deprotonation/protonation of the surface.

Chapter 5 proposes a molecular thermodynamic model to describe the thermodynamic non-ideality governing the ionization behavior of weak polyelectrolytes. By combination of the site-binding model and classical density functional theory, the effects of local ionic environment on both the inter- and intrachain correlations can be accounted for. The thermodynamic model quantifies the ionization behavior of weak polyelectrolytes over a broad range of molecular architectures and solution conditions at moderate to high salt concentration.

Chapter 6 presents a new formulation of polymer density functional theory by combining it with the Ising model for charge regulation. The so-called Ising DFT (iDFT) accounts for the multibody position and state of the entire polymer (i.e., its configuration) in the inhomogeneous fluid. It includes the effect of inter- and intrachain correlations due to excluded-volume, chain connectivity, and electrostatic interactions.

Chapter 7 utilizes iDFT to determine the polypeptide adsorption to an inorganic surface using a thermodynamic model based off the coarse-grained amino acid model. We predict the polypeptide adsorption to the rutile surface at a semi-quantitative level indicating that the key physics (i.e., intra- and interchain interactions) are correctly captured.

Chapter 8 improves upon the conventional descriptions of weak polyelectrolytes by accounting for the influence of long-range interactions on the conformation and ionization of the polymer. By building off the model proposed in Chapter 5, we can accurately predict the salt concentration effect on the titration of poly(acrylic acid), even at dilute salt

conditions. In addition, we can describe how the conformation of the polymer is influenced by salt concentration, pH (or degree of ionization), salt type, etc.

Chapter 9 combines Ising density functional theory with the single-chain-in-mean-field (SCMF) algorithm so that the long-range correlations, particularly those in weak polyelectrolytes, in inhomogeneous fluids can be accounted for. Different from the conventional approaches using the SCMF algorithm, the conformation and ionization is directly sampled from the mean-field determined by iDFT that incorporates the correlation effects due to excluded volume and electrostatic interactions. We demonstrate that our model correctly captures the conformation and ionization behavior of polymers near interfaces.

Bibliography

1. Waite, J.H., *Mussel adhesion – essential footwork.* Journal of Experimental Biology, 2017. **220**(4): p. 517-530.

2. Stewart, R.J., et al., *The role of coacervation and phase transitions in the sandcastle worm adhesive system.* Advances in colloid and interface science, 2017. **239**: p. 88-96.

3. MacEwan, S.R. and A. Chilkoti, *Applications of elastin-like polypeptides in drug delivery.* Journal of Controlled Release, 2014. **190**: p. 314-330.

4. González-Aramundiz, J.V., et al., *Polypeptides and polyaminoacids in drug delivery.* Expert Opinion on Drug Delivery, 2012. **9**(2): p. 183-201.

5. Talmadge, J.E., *The pharmaceutics and delivery of therapeutic polypeptides and proteins.* Advanced drug delivery reviews, 1993. **10**(2-3): p. 247-299.

6. Bolto, B. and J. Gregory, *Organic polyelectrolytes in water treatment.* Water Research, 2007. **41**(11): p. 2301-2324.

7. Seitz, S. and H. Ajiro, *Self-assembling weak polyelectrolytes for the layer-by-layer encapsulation of paraffin-type phase change material icosane.* Solar Energy Materials and Solar Cells, 2019. **190**: p. 57-64.

8. Priya James, H., et al., *Smart polymers for the controlled delivery of drugs – a concise overview.* Acta Pharmaceutica Sinica B, 2014. **4**(2): p. 120-127.

9. Jaganathan, S., *Bioresorbable polyelectrolytes for smuggling drugs into cells.* Artificial Cells, Nanomedicine, and Biotechnology, 2016. **44**(4): p. 1080-1097.

10. Schnöing, L., W. Augustin, and S. Scholl, *Fouling mitigation in food processes by modification of heat transfer surfaces: A review.* Food and Bioproducts Processing, 2020. **121**: p. 1-19.

11. Gama, M.d.S., A.G. Barreto, and F.W. Tavares, *The binding interaction of protein on a charged surface using Poisson–Boltzmann equation: lysozyme adsorption onto SBA-15.* Adsorption, 2021. **27**(7): p. 1137-1148.

12. Zhulina, E.B. and O.V. Borisov, *Poisson–Boltzmann Theory of pH-Sensitive (Annealing) Polyelectrolyte Brush.* Langmuir, 2011. **27**(17): p. 10615-10633.

13. Gonzalez Solveyra, E., et al., *Theoretical modeling of chemical equilibrium in weak polyelectrolyte layers on curved nanosystems.* Polymers, 2020. **12**(10): p. 2282.

14. Gallegos, A., G.M.C. Ong, and J. Wu, *Ising density functional theory for weak polyelectrolytes with strong coupling of ionization and intrachain correlations.* The Journal of Chemical Physics, 2021. **155**(24): p. 241102.

15. Atkins, P.W., J. De Paula, and J. Keeler, *Atkins' physical chemistry molecular thermodynamics and kinetics*. Eleventh edition. ed. 2019, Oxford, United Kingdom ; New York, NY: Oxford University Press. xix, 445 pages.

16. De Stefano, C., et al., *Chemical speciation of amino acids in electrolyte solutions containing major components of natural fluids.* Chemical Speciation & Bioavailability, 1995. **7**(1): p. 1-8.

17. Guggenheim, E.A. and J.C. Turgeon, *Specific interaction of ions.* Transactions of the Faraday Society, 1955. **51**(0): p. 747-761.

18. Bretti, C., C. Foti, and S. Sammartano, *A new approach in the use of SIT in determining the dependence on ionic strength of activity coefficients. Application to some chloride salts of interest in the speciation of natural fluids.* Chemical Speciation & Bioavailability, 2004. **16**(3): p. 105-110.

19. Pitzer, K.S., *Activity Coefficients in Electrolyte Solutions*. 2018.

20. Bickmore, B.R. and M.C.F. Wander, *Activity and Activity Coefficients*, in *Encyclopedia of Geochemistry: A Comprehensive Reference Source on the Chemistry of the Earth*, W.M. White, Editor. 2018, Springer International Publishing: Cham. p. 21-23.

21. Vilariño, T., et al., *Effect of ionic strength on the protonation of various amino acids analysed by the mean spherical approximation.* Journal of the Chemical Society, Faraday Transactions, 1997. **93**(3): p. 413-417.

22. Bretti, C., et al., *Modeling solubility and acid-base properties of some polar side chain amino acids in NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures.* Fluid Phase Equilibria, 2018. **459**: p. 51-64.

23. Bretti, C., et al., *Modeling solubility and acid–base properties of some amino acids in aqueous NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures.* SpringerPlus, 2016. **5**(1): p. 928.

24. Bretti, C., et al., *Some thermodynamic properties of dl-Tyrosine and dl-Tryptophan. Effect of the ionic medium, ionic strength and temperature on the solubility and acid–base properties.* Fluid Phase Equilibria, 2012. **314**: p. 185-197.

25. Kim, H.T. and W.J. Frederick, *Evaluation of Pitzer ion interaction parameters of aqueous electrolytes at 25.degree.C. 1. Single salt parameters.* Journal of Chemical & Engineering Data, 1988. **33**(2): p. 177-184.

26. De Stefano, C., et al., *The interaction of amino acids with the major constituents of natural waters at different ionic strengths.* Marine Chemistry, 2000. **72**(1): p. 61-76.

27. Lee, L.L., *Molecular thermodynamics of electrolyte solutions*. 2008, Singapore ; Hackensack, NJ: World Scientific. xii, 251 p.

28. Hossain, N., et al., *Revisiting electrolyte thermodynamic models: Insights from molecular simulations.* Aiche Journal, 2018. **64**(10): p. 3728-3734.

29. Blum, L., *Mean spherical model for asymmetric electrolytes.* Molecular Physics, 1975. **30**(5): p. 1529-1535.

30. Maribo-Mogensen, B., G.M. Kontogeorgis, and K. Thomsen, *Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions.* Industrial & Engineering Chemistry Research, 2012. **51**(14): p. 5353-5363.

31. Lyatskaya, Y.V., et al., *Analytical self-consistent-field model of weak polyacid brushes.* Macromolecules, 1995. **28**(10): p. 3562-3569.

32. Wu, J., et al., *A classical density functional theory for interfacial layering of ionic liquids.* Soft Matter, 2011. **7**(23): p. 11222-11231.

33. Jiang, J., V.V. Ginzburg, and Z.-G. Wang, *Density functional theory for charged fluids.* Soft Matter, 2018. **14**(28): p. 5878-5887.

34. Yu, Y.-X., J. Wu, and G.-H. Gao, *Density-functional theory of spherical electric double layers and ζ potentials of colloidal particles in restricted-primitive-model electrolyte solutions.* The Journal of Chemical Physics, 2004. **120**(15): p. 7223-7233.

35. Gallegos, A. and J. Wu, *Molecular thermodynamics for amino‐acid adsorption at inorganic surfaces.* AIChE Journal, 2021: p. e17432.

36. Landsgesell, J., et al., *Simulations of ionization equilibria in weak polyelectrolyte solutions and gels.* Soft Matter, 2019. **15**(6): p. 1155-1185.

37. Willott, J.D., et al., *Behavior of Weak Polyelectrolyte Brushes in Mixed Salt Solutions.* Macromolecules, 2018. **51**(3): p. 1198-1206.

38. Nap, R.J., S.H. Park, and I. Szleifer, *Competitive calcium ion binding to end-tethered weak polyelectrolytes.* Soft Matter, 2018. **14**(12): p. 2365-2378.

39. Uhlík, F., et al., *Modeling of Ionization and Conformations of Starlike Weak Polyelectrolytes.* Macromolecules, 2014. **47**(12): p. 4004-4016.

40. Gong, P., J. Genzer, and I. Szleifer, *Phase Behavior and Charge Regulation of Weak Polyelectrolyte Grafted Layers.* Physical Review Letters, 2007. **98**(1): p. 018302.

41. Léonforte, F., U. Welling, and M. Müller, *Single-chain-in-mean-field simulations of weak polyelectrolyte brushes.* The Journal of Chemical Physics, 2016. **145**(22): p. 224902.

42. Koper, G.J.M. and M. Borkovec, *Proton binding by linear, branched, and hyperbranched polyelectrolytes.* Polymer, 2010. **51**(24): p. 5649-5662.

43. Kawaguchi, S., et al., *Dissociation behavior of poly(fumaric acid) and poly(maleic acid). II. Model calculation.* Macromolecules, 1990. **23**(3): p. 731-738.

44. Blanco, P.M., et al., *Coupling of charge regulation and conformational equilibria in linear weak polyelectrolytes: Treatment of long-range interactions via effective short-ranged and pH-dependent interaction parameters.* Polymers, 2018. **10**(8): p. 811.

45. Blanco, P.M., et al., *Effect of charge regulation and conformational equilibria in the stretching properties of weak polyelectrolytes.* Macromolecules, 2019. **52**(21): p. 8017-8031.

46. Blanco, P.M., et al., *Role of charge regulation and fluctuations in the conformational and mechanical properties of weak flexible polyelectrolytes.* Polymers, 2019. **11**(12): p. 1962.

Chapter 2. Basic formulations of physics-based models

In this dissertation, physics-based models are developed to provide fast and accurate evaluation of molecular properties in bulk and inhomogeneous fluids for monomers, polymers, and surfaces. In this chapter, the basic concepts and formulations of physics-based models used in this dissertation are introduced.

2.1 Density Functional Theory

Density functional theory (DFT) is an effective computational tool to study the microscopic structure and thermodynamic properties of complex fluids, such as polymers, colloids, and ionic liquids at uniform as well as inhomogeneous conditions. Different from molecular simulations, DFT provides analytical relations between the microscopic structure and thermodynamic quantities from a molecular perspective. Also different from simulation methods, practical application of DFT requires the formulation of an approximate free energy functional F^{ex} in terms of the one-body density profile $\rho(\mathbf{r})$. Its numerical implementation relies on variational calculus to obtain the equilibrium density profile and subsequently thermodynamic variables. While the equilibrium density profile can be obtained from molecular simulations, simulation of thermodynamic properties such as the free energy calculation is extremely time consuming. The density functional methods can be naturally applied to systems with multiple length scales and various inter-molecular interactions that are difficult to handle by using alternative computational methods. This chapter presents an overview of the basic concepts of DFT and its applications to various complex fluids.

2.1.1 One body density profile

In DFT, thermodynamic properties are often represented in terms of the functionals of the molecular density profiles. A density profile describes the average distribution of molecules in a many-body system. For a system containing *N* identical particles, the microscopic instantaneous particle density $\hat{\rho}(\mathbf{r})$, which counts the number of particles at a position **r** is defined as

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \tag{2.1}$$

where δ is the Dirac-Delta function

$$\delta(\mathbf{r} - \mathbf{r}') = \begin{cases} \infty, & \text{if } \mathbf{r} = \mathbf{r}' \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

The Dirac-Delta function is subject to the normalization condition

$$\int \delta(\mathbf{r} - \mathbf{r}') d\mathbf{r}' = 1 \tag{2.3}$$

The equilibrium density profile is defined as an ensemble average of the instantaneous density

$$\rho(\mathbf{r}) = \hat{\rho}(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle \tag{2.4}$$

For a one-component open system at fixed temperature *T* and volume *V*, the equilibrium one-body density profile is related to the Hamiltonian of the system and the grand partition function Ξ. The latter is given as

$$\Xi = \sum_N \frac{1}{N!\Lambda^{3N}} \int \exp\left\{-\beta\left[\Gamma(\mathbf{r}^N) + N\mu + \sum_{i=1}^{N}\phi(\mathbf{r}_i)\right]\right\} d\mathbf{r}^N \tag{2.5}$$

where Λ is the thermal wavelength, $\Gamma(\mathbf{r}^N)$ stands for the total interaction potential of N particles at configuration $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$, μ is the bulk chemical potential of particle, $\beta = 1/(k_B T)$ where k_B is the Boltzmann constant, and $\phi(\mathbf{r})$ is the one-body external potential.

By substituting Eq. (1.4) into Eq. (1.5), the one-body density profile $\rho(\mathbf{r})$ is related to the grand partition function by

$$\rho(\mathbf{r}) = \frac{1}{\Xi}\sum_N \frac{1}{N!\Lambda^{3N}} \int \sum_{i=1}^{N}\delta(\mathbf{r}-\mathbf{r}_i)\ \exp\left\{-\beta\left[\Gamma(\mathbf{r}^N) + N\mu + \sum_{i=1}^{N}\phi(\mathbf{r}_i)\right]\right\} d\mathbf{r}^N \tag{2.6}$$

The right-hand side of Eq. (1.6) is related to the functional derivative of the grand partition function with respect to the external potential. After rearrangement, the one-body density profile then given by

$$\rho(\mathbf{r}) = -\frac{1}{\beta\Xi}\frac{\delta\Xi}{\delta\phi(\mathbf{r})} = -\frac{\delta \ln \Xi}{\delta\beta\phi(\mathbf{r})} \tag{2.7}$$

The grand potential Ω is defined by the grand partition function

$$\beta\Omega = -\ln \Xi \tag{2.8}$$

Substituting Eq. (1.8) into Eq. (1.7) yields

$$\rho(\mathbf{r}) = \frac{\delta\Omega}{\delta\phi(\mathbf{r})} \tag{2.9}$$

2.1.2 Ising density functional theory: excess free energy for weak polyelectrolyte systems

We can describe a weak polyelectrolyte configuration (i.e., the position $\mathbf{r}_i$ and state s_i of all M segments in the polymer chain) through the multibody vector $\mathbf{X} = (\mathbf{R}, \mathbf{S})$ where $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_M)$ is the position vector and $\mathbf{S} = (s_1, s_2, ..., s_M)$ is the state vector. The valence $Z(s_i)$ for each segment in the polymer chain is dependent upon the state of that segment. The polymer density profile is given by $\rho(\mathbf{X})$ and the segment density profile by $\rho_i(\mathbf{x}_i)$ where $\mathbf{x}_i = (\mathbf{r}_i, s_i)$. All other non-polymeric species density profile is given by $\rho_\alpha(\mathbf{r})$. For weak polyelectrolyte systems represented by tangent hard-sphere chains of ionizable segments and charged hard spheres for monomeric ions, the excess Helmholtz energy is conventionally approximated by the thermodynamic perturbation theory (TPT)

$$F^{ex} \approx F_m^{ex}[\rho_{k,i}(\mathbf{r}, s_i), \rho_\alpha(\mathbf{r})] - \sum_k \int d\mathbf{X}_k \rho_k(\mathbf{X}) k_B T \ln y_k(\mathbf{X}_k) \tag{2.10}$$

where $F_m^{ex}[\rho_{k,i}(\mathbf{r}, s_i), \rho_\alpha(\mathbf{r})]$ denotes the excess Helmholtz energy of the corresponding monomeric system and $y_k(\mathbf{X})$ is the multibody cavity correlation function[1]. The former takes into account contributions due to hard-sphere repulsion (viz., excluded volume effects) as described by the modified fundamental measure theory[2], the mean electrostatic potential, and electrostatic correlations[3].

The excess Helmholtz energy due to hard-sphere repulsion is described through modified fundamental measure theory (MFMT)[2] by

$$F^{hs} = k_B T \int \Phi^{hs} \left[n_\alpha \left(\mathbf{r}' \right) \right] d\mathbf{r}' \tag{2.11}$$

where the reduced excess Helmholtz energy density Φ^{hs} is a function of the weighted averages of the monomeric density distributions $\rho_i(\mathbf{r})$,

$$n_\alpha \left(\mathbf{r} \right) = \sum_i n_{\alpha_i} \left(\mathbf{r} \right) = \sum_i \int \rho_i \left(\mathbf{r}' \right) \omega_i^{(\alpha)} \left(\mathbf{r} - \mathbf{r}' \right) d\mathbf{r}' . \tag{2.12}$$

The weight functions $\omega_i^{(\alpha)}$ are geometric descriptions of a hard sphere (i.e., the volume, the surface area, and the mean radius of curvature)[4]. The expression for each weight function are given as

$$\omega_i^{(3)} \left(\mathbf{r} \right) = \theta \left(\frac{\sigma_i}{2} - r \right) \tag{2.13}$$

$$\omega_i^{(2)} \left(\mathbf{r} \right) = \delta \left(\frac{\sigma_i}{2} - r \right) \tag{2.14}$$

$$\omega_i^{(V2)} \left(\mathbf{r} \right) = \frac{\mathbf{r}}{r} \delta \left(\frac{\sigma_i}{2} - r \right) \tag{2.15}$$

where $\omega_i^{(3)}$ characterizes the volume, $\omega_i^{(2)}$ characterizes the surface area, and $\omega_i^{(V2)}$ characterizes the variance across the particle surface. The other weight functions can be related to these three as follows:

$$\omega_i^{(1)} \left(\mathbf{r} \right) = \frac{\omega_i^{(2)} \left(\mathbf{r} \right)}{2\pi\sigma_i} \tag{2.16}$$

$$\omega_i^{(0)} \left(\mathbf{r} \right) = \frac{\omega_i^{(2)} \left(\mathbf{r} \right)}{\pi\sigma_i^2} \tag{2.17}$$

$$\omega_i^{(V1)}(\mathbf{r}) = \frac{\omega_i^{(V2)}(\mathbf{r})}{2\pi\sigma_i} \tag{2.18}$$

The final expression for the excess Helmholtz energy density is given by

$$\begin{aligned}\Phi_{hs}^{ex} = &-n_0 \ln(1-n_3) + \frac{n_1 n_2 - \vec{\mathbf{n}}_{V1}\cdot\vec{\mathbf{n}}_{V2}}{1-n_3} + \\ &\left(n_2^{\,3} - 3n_2 \vec{\mathbf{n}}_{V2}\cdot\vec{\mathbf{n}}_{V2}\right)\frac{n_3 + (1-n_3)^2\ln(1-n_3)}{36\pi n_3^{\,2}(1-n_3)^2}\end{aligned} \tag{2.19}$$

The two vector weighted densities, $\mathbf{n}_{V1}$ and $\mathbf{n}_{V2}$, will vanish in the bulk limit and thus, Φ^{hs} is identical to the Boublik-Mansoori-Carnahan-Starling-Leland (BMCSL) equation of state[5]. The functional derivative of excess Helmholtz free energy due to the hard-sphere repulsion yields the excess chemical potential due to the hard sphere repulsion, μ_i^{hs}, given by

$$\mu_i^{hs}(\mathbf{r}) = \frac{\delta F^{hs}}{\delta\rho_i(\mathbf{r})} = k_B T \sum_\alpha \int \frac{\partial\Phi^{hs}}{\partial n_\alpha(\mathbf{r}')}\omega_i^{(\alpha)}(\mathbf{r}'-\mathbf{r})\,d\mathbf{r}'. \tag{2.20}$$

The mean electrostatic potential describes the electrostatic interactions between all charges in the system (i.e., charged monomers, ions, and those fixed on an interface). We can write the excess Helmholtz free energy due to the long-range Coulomb interaction as

$$F^c = \int \psi(\mathbf{r}')q(\mathbf{r}')\,d\mathbf{r}' - \frac{\varepsilon_r\varepsilon_0}{2}\int\left[\nabla\psi(\mathbf{r}')\right]^2 d\mathbf{r}', \tag{2.21}$$

The total charge distribution of the system is given by $q(\mathbf{r}) = \sum_k \sum_{i=1}^{M_k} \sum_{\{s_{k,i}\}} Z_k(s_{k,i}) e\rho_{k,i}(\mathbf{r}, s_{k,i}) + \sum_\alpha Z_\alpha e\rho_\alpha(\mathbf{r}) + f(\mathbf{r})$ where the first summation is over all segments in their different states of the polymers, the second

summation is over the free ions (viz. salt ions, protons, and hydroxyl ions), and $f(\mathbf{r})$ defines the position(s) of the fixed charges within the system. For a single charged wall, $f(\mathbf{r}) = Q\,\delta(z)$, where Q is the surface charge density of the planar surface fixed at $z = 0$. It should be noted that by expressing the free energy in this manner, we account for the charge component of the external potential induced by the interface. The mean electrostatic potential can be determined from solution of the Poisson equation[6]:

$$\nabla^2 \psi(\mathbf{r}) = -4\pi l_B\, q(\mathbf{r}) \tag{2.22}$$

The excess Helmholtz free energy due to electrostatic correlations is given by a quadratic functional Taylor expansion with respect to a monomeric reference fluid of densities $\{\rho_i^{ref}\}$[7]:

$$\begin{aligned} F^{el}\left[\{\rho_t(\mathbf{r})\}\right] \approx \int f^{el}\left[\{\rho_t^{ref}(\mathbf{r}')\}\right] d\mathbf{r}' + \sum_j \int \Delta\rho_j(\mathbf{r}')\,\mu_i^{el}\left[\{\rho_t^{ref}(\mathbf{r})\};\mathbf{r}'\right] d\mathbf{r}' \\ -\frac{k_B T}{2}\sum_j\sum_k \int\int \Delta\rho_j(\mathbf{r}')\,\Delta\rho_k(\mathbf{r}'')\,c_{ij}^{(2),el}\left[\{\rho_t^{ref}(\mathbf{r})\};\mathbf{r}',\mathbf{r}''\right] d\mathbf{r}' d\mathbf{r}'' \end{aligned} \tag{2.23}$$

where $\Delta\rho_i(\mathbf{r}) = \rho_i(\mathbf{r}) - \rho_i^{ref}(\mathbf{r})$ and f^{el}, μ_i^{el}, and $c_{ij}^{(2),el}$ are the free energy density, excess chemical potential, and direct correlation function (DCF), respectively, due to electrostatic correlations evaluated using the reference fluid. These expressions can be evaluated using the mean spherical approximation which is a novel approach to incorporate the finite size effect on electrostatic interactions[8]. The analytical expressions were derived for a homogeneous charged hard sphere solution. The free energy density due to electrostatic correlations can be written as

$$\beta f^{el}\left[\left\{\rho_t^{ref}\left(\mathbf{r}\right)\right\}\right]=\frac{\Gamma^3}{3\pi}-l_B\sum_i\rho_i^{ref}Z_i\left(\frac{Z_i\Gamma+\eta\sigma_i}{1+\Gamma\sigma_i}\right) \tag{2.24}$$

$$\eta=\frac{1}{H\left(\Gamma\right)}\sum_i\frac{\rho_i^{ref}Z_i\sigma_i}{1+\Gamma\sigma_i} \tag{2.25}$$

$$H\left(\Gamma\right)=\sum_i\frac{\rho_i^{ref}\sigma_i^3}{1+\Gamma\sigma_i}+\frac{2}{\pi}\left(1-\frac{\pi}{6}\sum_i\rho_i^{ref}\sigma_i^3\right) \tag{2.26}$$

The MSA screening parameter Γ is determined on the basis that it minimizes the free energy

$$\frac{\partial f^{el}}{\partial\Gamma}=0 \tag{2.27}$$

Satisfaction of this condition results in an implicit equation for Γ

$$\Gamma^2=\pi l_B\sum_i\rho_i^{ref}\left(\frac{Z_i-\eta\sigma_i^2}{1+\Gamma\sigma_i}\right)^2 \tag{2.28}$$

The excess chemical potential due to electrostatic correlations is

$$\beta\mu_i^{el}=-l_B\left[\frac{Z_i^2\Gamma}{1+\Gamma\sigma_i}+\eta\sigma_i\left(\frac{2Z_i-\eta\sigma_i^2}{1+\Gamma\sigma_i}+\frac{\eta\sigma_i^2}{3}\right)\right] \tag{2.29}$$

Lastly, the DCF can be approximated by the interaction between two charged shells[9]

$$c_{ij}^{(2),el}\left(\mathbf{r},\mathbf{r}'\right)=Z_iZ_j\frac{l_B}{4}\frac{1}{b_ib_j}\frac{\left[|\mathbf{r}-\mathbf{r}'|^2-2|\mathbf{r}-\mathbf{r}'|\left(b_i+b_j\right)+\left(b_i-b_j\right)^2\right]}{|\mathbf{r}-\mathbf{r}'|} \tag{2.30}$$

where $b_i=\sigma_i/2+1/2\Gamma$. The success of this approach can be attributed to the choice in the reference fluid used. Gillespie et al. proposed a reference fluid that accounted for the

inhomogeneous distribution of ions in the pore but satisfied the local charge neutrality requirement that MSA requires[3]. The reference fluid is given by

$$\rho_i^{ref}(\mathbf{r}) = \int \bar{\rho}_i(\mathbf{r}')\, w(\mathbf{r},\mathbf{r}')\, d\mathbf{r}' \tag{2.31}$$

where $\bar{\rho}_i(\mathbf{r})$ is spatially averaged density distribution for species *i* that is charge neutral at every point **r** and has the same ionic strength as the local ion distribution $\rho_i^*(\mathbf{r})$, i.e.,

$$0 = \sum_i \bar{\rho}_i(\mathbf{r}) Z_i \tag{2.32}$$

and

$$\sum_i \bar{\rho}_i(\mathbf{r}) Z_i^2 = \sum_i \rho_i^*(\mathbf{r}) Z_i^2 \,. \tag{2.33}$$

The weight function *w* is the normalized form of the volume weight $\omega_i^{(3)}$ given early:

$$w(\mathbf{r},\mathbf{r}') = \frac{\theta\left[\,|\mathbf{r}'-\mathbf{r}| - R_f(\mathbf{r})\right]}{4\pi R_f^3(\mathbf{r})/3}. \tag{2.34}$$

The capacitance radius R_f is an approximation of the electrostatic length scale over which we average and is given by[3]

$$R_f(\mathbf{r}) = \frac{\sum_i \bar{\rho}_i(\mathbf{r})\sigma_i / 2}{\sum_i \bar{\rho}_i(\mathbf{r})} + \frac{1}{2\Gamma(\mathbf{r})}. \tag{2.35}$$

where $\Gamma(\mathbf{r})$ is the MSA parameter given earlier and is evaluated at every point **r** using the reference fluid density. Since the reference fluid density is a function of itself through the capacitance radius, we must iterate until self-consistency between the reference fluid

densities and capacitance radius is achieved. The excess chemical potential due to electrostatic correlations at position **r** in the inhomogeneous fluid can be expressed as

$$\beta\mu_i^{el}\left[\left\{\rho_k(\mathbf{r})\right\}\right] \approx \beta\mu_i^{el}\left[\left\{\rho_k^{ref}(\mathbf{r})\right\}\right] - \sum_j \int \Delta\rho_j(\mathbf{r}')\, c_{ij}^{(2),el}\left[\left\{\rho_k^{ref}(\mathbf{r})\right\};\mathbf{r},\mathbf{r}'\right] d\mathbf{r}' . \quad (2.36)$$

With the excess Helmholtz energy for the corresponding monomeric system outlined, we now turn our attention to the contribution resulting from polymerization, i.e., the change in free energy due to the chain connectivity of monomers. The excess Helmholtz energy due to chain connectivity is given by

$$\beta F^{ch} = -\sum_k \int d\mathbf{X}_k \rho_k(\mathbf{X}_k) k_B T \ln y_k(\mathbf{X}_k) \quad (2.37)$$

where $\rho_k(\mathbf{X})$ and $y_k(\mathbf{X})$ are the multi-body polymer and cavity correlation function, respectively, for polymer *k*. Wertheim's thermodynamic perturbation theory (TPT) provides a systematic procedure to approximate the intersegment correlations owing to chain connectivity[1]. In its lowest-order approximation (TPT1), the multibody-cavity correlation function (CCF) is represented by a superposition of the two-body correlation functions in the monomeric fluid:

$$\ln y_k(\mathbf{X}_k) \approx \sum_{i=1}^{M_k-1} \ln y_k(\mathbf{x}_{k,i}, \mathbf{x}_{k,i+1}) . \quad (2.38)$$

It is important to note that we still consider the state of each monomer in an adjacent pair. It should be noted that by writing the CCF in this manner, we are implicitly assuming that any intramolecular interactions beyond the neighbors are sufficiently screened such that

$$\beta\mu_k^{\infty}\left(\mathbf{X}_k\right)-\ln y_k\left(\mathbf{X}_k\right)\approx\sum_{i=1}^{M_k-1}\beta u_k^{\infty}\left(s_{k,i},s_{k,i+1}\right)-\sum_{i=1}^{M_k-1}\ln y_k\left(\mathbf{x}_{k,i},\mathbf{x}_{k,i+1}\right), \tag{2.39}$$

where the first term on the righthand side is given by

$$\beta u_k^{\infty}\left(s_{k,i},s_{k,i+1}\right)=\frac{l_B}{\sigma_k\left(s_{k,i},s_{k,i+1}\right)}Z_k\left(s_{k,i}\right)Z_k\left(s_{k,i+1}\right). \tag{2.40}$$

Unfortunately, an exact formulation of the CCF in an inhomogeneous fluid is unknown at present, thus we must use further approximations to estimate this value. Consistent with versions of PDFT[10, 11], we can estimate the inhomogeneous CCF by a geometric mean of the bulk CCF of two segments at contact evaluated at position $\mathbf{r}_{k,i}$ and $\mathbf{r}_{k,i+1}$:

$$y_k\left(\mathbf{x}_{k,i},\mathbf{x}_{k,i+1}\right)\approx\sqrt{y_k^c\left[\left\{n_{\alpha,j}\left(\mathbf{r}_{k,i}\right)\right\},s_{k,i},s_{k,i+1}\right]y_k^c\left[\left\{n_{\alpha,j}\left(\mathbf{r}_{k,i+1}\right)\right\},s_{k,i},s_{k,i+1}\right]} \tag{2.41}$$

where the CCF at contact y_k^c accounts for the state of the segments i and i+1. It can be evaluated from MSA using a reference system of charged hard spheres[12]:

$$\begin{aligned}y_k^c\left(n_\alpha,s_i,s_{i+1}\right)=&\left[\frac{1}{1-n_3}+\frac{n_2\sigma_k\left(s_{k,i}\right)\sigma_k\left(s_{k,i+1}\right)\left(1-\mathbf{n}_{V2}\cdot\mathbf{n}_{V2}/n_2^2\right)}{4\sigma_k\left(s_{k,i},s_{k,i+1}\right)}\right]\\&\times\exp\left\{\frac{l_B}{\sigma_k\left(s_{k,i},s_{k,i+1}\right)}\left[Z_k\left(s_{k,i}\right)Z_k\left(s_{k,i+1}\right)-Z_k'\left(s_{k,i}\right)Z_k'\left(s_{k,i+1}\right)\right]\right\}\end{aligned} \tag{2.42}$$

where $Z_k'(s_{k,i})=[Z_k(s_{k,i})-\eta\sigma_k(s_{k,i})^2]/[1+\Gamma\sigma_k(s_{k,i})]$ and Γ is the MSA screening parameter defined in terms of weighted densities n_α from FMT[13]. Plugging the previous set of equations into the excess Helmholtz energy due to chain connectivity results in the following expression:

$$\beta F^{ch} = -\frac{1}{2}\sum_{k}\sum_{i=1}^{M_k-1}\sum_{\{s_{k,i}\}}\sum_{\{s_{k,i+1}\}}\int\left[n_{0,k,i}\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right) + n_{0,k,i+1}\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)\right] \times \ln y_k^c\left[\left\{n_{\alpha,j}\left(\mathbf{r}'\right)\right\}, s_{k,i}, s_{k,i+1}\right] d\mathbf{r}' \tag{2.43}$$

where the weighted densities $n_{0,k,i}(\mathbf{r}', s_{k,i}, s_{k,i+1})$ and $n_{0,k,i+1}(\mathbf{r}', s_{k,i}, s_{k,i+1})$ are the FMT weighted densities given by

$$n_{0,k,i}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right) = \int \frac{\rho_{k,i}\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)}{4\pi\sigma_k^2\left(s_{k,i}\right)}\delta\left[\frac{\sigma_k\left(s_{k,i}\right)}{2} - |\,\mathbf{r}-\mathbf{r}'\,|\right] d\mathbf{r}' \tag{2.44}$$

and

$$n_{0,k,i+1}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right) = \int \frac{\rho_{k,i+1}\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)}{4\pi\sigma_k^2\left(s_{k,i+1}\right)}\delta\left[\frac{\sigma_k\left(s_{i+1}\right)}{2} - |\,\mathbf{r}-\mathbf{r}'\,|\right] d\mathbf{r}' . \tag{2.45}$$

The first density term $\rho_{k,i}(\mathbf{r}, s_{k,i}, s_{k,i+1})$ is for monomer i in state $s_{k,i}$ at position $\mathbf{r}$ given that the adjacent segment i+1 is in state $s_{k,i+1}$ for the polymer k. The second density term $\rho_{k,i+1}(\mathbf{r}, s_{k,i}, s_{k,i+1})$ describes the monomer i+1 in state $s_{k,i+1}$ given that the adjacent segment i is in state $s_{k,i}$. These terms are needed to account for the different possible state combinations of two adjacent monomers. The functional derivative of the excess free energy due to chain connectivity with respect to the density profile of the polymer is given by

$$\frac{\delta F^{ch}}{\delta\rho_k\left(\mathbf{X}_k\right)} = -\frac{1}{2}k_B T\sum_{i=1}^{M_k-1}\int \ln y_k^c\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)\omega_{i,s_{k,i}}^{(0)}\left(\mathbf{r}'-\mathbf{r}_{k,i}\right)d\mathbf{r}' - \frac{1}{2}k_B T\sum_{i=2}^{M_k}\int \ln y_k^c\left(\mathbf{r}', s_{k,i-1}, s_{k,i}\right)\omega_{i,s_{k,i}}^{(0)}\left(\mathbf{r}'-\mathbf{r}_{k,i}\right)d\mathbf{r}' + \sum_{i=1}^{M}\mu_{k,i}^{ch}\left(\mathbf{r}_{k,i}, s_{k,i}\right) \tag{2.46}$$

where the first two terms are the summation over the intrachain contribution of segment *i* and segment *i*+1 in their respective states for polymer *k*. The last term is denoted as the excess chemical potential due to chain connectivity as it is a one-body term that depends only on the state and position of segment *i*. It is given by

$$\mu_{u,t}^{ch}\left(\mathbf{r},s_{u,t}\right)=-\frac{1}{2}\sum_{k}\sum_{i=1}^{M_k-1}\sum_{\{s_{k,i}\}}\sum_{\{s_{k,i+1}\}}\int\left[n_{0,k,i}\left(\mathbf{r}',s_{k,i},s_{k,i+1}\right)+n_{0,k,i+1}\left(\mathbf{r}',s_{k,i},s_{k,i+1}\right)\right] \times\sum_{\alpha}\frac{\partial\left(\ln y_k^c\left[\left\{n_{\alpha,j}\left(\mathbf{r}'\right)\right\},s_{k,i},s_{k,i+1}\right]\right)}{\partial n_{\alpha,u,t}\left(s_{u,t}\right)}\omega_{t,s_{u,t}}^{(\alpha)}\left(\mathbf{r}'-\mathbf{r}\right)d\mathbf{r}' \tag{2.47}$$

The density terms that depend upon the paired state can be determined by

$$\rho_{k,i}(\mathbf{r},s_{k,i},s_{k,i+1})=\int d\mathbf{X}_k'\delta(\mathbf{r}-\mathbf{r}_{k,i}')\delta_{s_{k,i}s_{k,i}'}\delta_{s_{k,i+1}s_{k,i+1}'}\rho_k(\mathbf{X}_k') \tag{2.48}$$

and

$$\rho_{k,i+1}(\mathbf{r},s_{k,i},s_{k,i+1})=\int d\mathbf{X}_k'\delta(\mathbf{r}-\mathbf{r}_{k,i+1}{}')\delta_{s_{k,i}s_{k,i}'}\delta_{s_{k,i+1}s_{k,i+1}'}\rho_k(\mathbf{X}_k') \tag{2.49}$$

for i=1, M_k-1.

The multibody integration for these two can be simplified using a set of propagator functions:

$$\rho_{k,i}(\mathbf{r},s_{k,i},s_{k,i+1})=\exp\left\{-\beta\left[\lambda_{k,i}\left(\mathbf{r},s_{k,i}\right)-\mu_k\right]\right\} \times\left[\sum_{\{s_{k,i-1}'\}}G_{k,i}^{L}\left(\mathbf{r},s_{k,i-1}',s_{k,i}\right)\right]G_{k,i}^{R}\left(\mathbf{r},s_{k,i},s_{k,i+1}\right) \tag{2.50}$$

and

$$\rho_{k,i+1}(\mathbf{r}, s_{k,i}, s_{k,i+1}) = \exp\left\{-\beta\left[\lambda_{k,i+1}\left(\mathbf{r}, s_{k,i+1}\right) - \mu_k\right]\right\} \times G^L_{k,i+1}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right)\left[\sum_{\{s'_{k,i+2}\}} G^R_{k,i+1}\left(\mathbf{r}, s_{k,i+1}, s'_{k,i+2}\right)\right], \quad (2.51)$$

where $\lambda_{k,j}(\mathbf{r}, s_{k,j})$ is the one-body potential for segment j in state $s_{k,j}$ of polymer k at position $\mathbf{r}$ and describes the influence of intermolecular interactions:

$$\lambda_{k,i}\left(\mathbf{r}, s_{k,i}\right) = \frac{\delta F^{hs} + \delta F^{el} + \delta F^{cp}}{\delta \rho_{k,i}\left(\mathbf{r}, s_{k,i}\right)} + \mu^{ch}_{k,i}\left(\mathbf{r}, s_{k,i}\right) + \Delta\mu^{H}_{k,i}\left(s_{k,i}\right) + V^{ext}_{k,i}\left(\mathbf{r}, s_{k,i}\right) + e\psi\left(\mathbf{r}\right) Z_k\left(s_{k,i}\right) \quad (2.52)$$

The propagator functions denoted by G account for the influence of all other segments in the chain and incorporate the cavity correlation term. They are expressed as

$$G^L_{k,i}\left(\mathbf{r}, s_{k,i-1}, s_{k,i}\right) = \exp\left[-\beta\tau_{k,i}\left(\mathbf{r}, s_{k,i-1}, s_{k,i}\right)/2 + \beta u^{\infty}_k\left(s_{k,i-1}, s_{k,i}\right)\right] \times \int \exp\left[-\beta v^B_k\left(\mathbf{x}'_{k,i-1}, \mathbf{x}_{k,i}\right) - \beta\lambda_{k,i-1}\left(\mathbf{r}', s_{k,i-1}\right)\right] \times \exp\left[-\beta\tau_{k,i-1}\left(\mathbf{r}', s_{k,i-1}, s_{k,i}\right)/2\right] \sum_{\{s_{k,i-2}\}} G^L_{k,i-1}\left(\mathbf{r}', s_{k,i-2}, s_{k,i-1}\right) d\mathbf{r}' \quad (2.53)$$

$$G^R_{k,i}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right) = \exp\left[-\beta\tau_{k,i}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right)/2 + \beta u^{\infty}_k\left(s_{k,i}, s_{k,i+1}\right)\right] \times \int \exp\left[-\beta v^B_k\left(\mathbf{x}_{k,i}, \mathbf{x}'_{k,i+1}\right) - \beta\lambda_{k,i+1}\left(\mathbf{r}', s_{k,i+1}\right)\right] \times \exp\left[-\beta\tau_{k,i+1}\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)/2\right] \sum_{\{s_{k,i+2}\}} G^R_{k,i+1}\left(\mathbf{r}', s_{k,i+1}, s_{k,i+2}\right) d\mathbf{r}' \quad (2.54)$$

where $\sum_{\{s'_{k,0}\}} G^L_{k,1}(\mathbf{r}, s'_{k,0}, s_{k,1})$ and $\sum_{\{s'_{k,M_{k+1}}\}} G^R_{k,M_k}(\mathbf{r}, s_{M_k}, s'_{M_k+1})$ are equal to 1. The terms $\tau_{k,i}(\mathbf{r}, s_{k,i}, s_{k,i+1})$ and $\tau_{k,i+1}(\mathbf{r}, s_{k,i}, s_{k,i+1})$ are the weighted intrachain correlation between segment i and i+1 in their respective states given by

$$\tau_{k,i}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right) = k_B T \int \frac{\ln y_k^c\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)}{\pi \sigma_k\left(s_{k,i}\right)^2} \delta\left[\frac{\sigma_k\left(s_{k,i}\right)}{2} - |\mathbf{r}' - \mathbf{r}|\right] d\mathbf{r}' \qquad (2.55)$$

and

$$\tau_{k,i+1}\left(\mathbf{r}, s_{k,i}, s_{k,i+1}\right) = k_B T \int \frac{\ln y_k^c\left(\mathbf{r}', s_{k,i}, s_{k,i+1}\right)}{\pi \sigma_k\left(s_{k,i+1}\right)^2} \delta\left[\frac{\sigma_k\left(s_{k,i+1}\right)}{2} - |\mathbf{r}' - \mathbf{r}|\right] d\mathbf{r}' . \qquad (2.56)$$

The division by two in Eq. (S52) and (S53) is because each segment in a connection makes a half contribution within the approximation of the two-body correlation function [see Eq. (S40)]. The density profile for a polymer segment in a given state $\rho_{k,i}(\mathbf{r}, s_{k,i})$ can then be determined by the summation over the states of the adjacent monomer via Eq. (S47), i.e., $\rho_{k,i}(\mathbf{r}, s_{k,i}) = \sum_{\{s'_{k,i+1}\}} \rho_{k,i}(\mathbf{r}, s_{k,i}, s'_{k,i+1})$, besides the end-segment which is determined via Eq. (S48), i.e., $\rho_{k,i+1}(\mathbf{r}, s_{k,i+1}) = \sum_{\{s'_{k,i}\}} \rho_{k,i+1}(\mathbf{r}, s'_{k,i}, s_{k,i+1})$. The summation over the different states gives the density of a segment at position **r** [viz. $\rho_{k,i}\left(\mathbf{r}\right) = \sum_{\{s'_{k,i}\}} \rho_{k,i}(\mathbf{r}, s'_{k,i})$].

The bulk chemical potential is obtained from the application of Wertheim's TPT1 equation of state to weak polyelectrolytes[1]:

$$\mu_k\left(\mathbf{S}_k\right) = k_B T \ln \rho_k^b\left(\mathbf{S}_k\right) - k_B T \ln y_k\left(\mathbf{S}_k\right) + \sum_{i=1}^{M_k} \mu_{k,i}^{ex,b}\left(s_{k,i}\right) \qquad (2.57)$$

where the bulk density of a given state $\rho_k^b(\mathbf{S}_k)$ must be determined by evaluating the semi-grand partition function Ξ_k. It can be expressed as

$$\Xi_k = \sum_{\{\mathbf{S}'_k\}} \exp\left[\ln y_k\left(\mathbf{S}'_k\right) - \beta\mu_k^{\infty}\left(\mathbf{S}'_k\right) - \beta\mu_k^{H}\left(\mathbf{S}'_k\right) - \beta\sum_{i}^{M_k} \mu_{k,i}^{ex,b}\left(s'_{k,i}\right)\right] \qquad (2.58)$$

where we must evaluate over all possible states of the polymer accounting for the excess chemical potential of that state as well as the contribution due to the chemical equilibrium for that state $\mu_k^H(\mathbf{S}_k)$. A procedure for calculating the partition function of any polymer architecture that does not contain a loop can be found in our previous work[14]. It should be noted that chemical equilibrium is satisfied, i.e.,

$$\beta\mu_k\left(\mathbf{S}_k=0\right)=\beta\mu_k\left(\mathbf{S}'_k\right)-\mu_k^{\infty}\left(\mathbf{S}'_k\right)-\mu_k^{H}\left(\mathbf{S}'_k\right) \tag{2.59}$$

where the left-hand side refers to the fully uncharged polymer state. The bulk density of a polymer in a given state $\mathbf{S}_k$ can then be determined by

$$\rho_k^b\left(\mathbf{S}_k\right)=\rho_k^b\frac{\exp\left[\ln y_k\left(\mathbf{S}_k\right)-\beta\mu_k^{\infty}\left(\mathbf{S}_k\right)-\beta\mu_k^{H}\left(\mathbf{S}_k\right)-\beta\sum_{i=1}^{M_k}\mu_{k,i}^{ex,b}\left(s_{k,i}\right)\right]}{\Xi_k} \tag{2.60}$$

where $\rho_k^b=\sum_{\{\mathbf{S}'_k\}}\rho_k^b(\mathbf{S}'_k)$. Thus, the chemical potential of a polymer in state $\mathbf{S}_k$ can alternatively be expressed as

$$\mu_k\left(\mathbf{S}_k\right)=k_BT\ln\rho_k^b-\ln\Xi_k-\mu_k^{\infty}\left(\mathbf{S}_k\right)-\mu_k^{H}\left(\mathbf{S}_k\right)\ . \tag{2.61}$$

Bibliography

1. Wertheim, M.S., *Thermodynamic perturbation theory of polymerization.* The Journal of Chemical Physics, 1987. **87**(12): p. 7323-7331.

2. Yu, Y.-X. and J. Wu, *A modified fundamental measure theory for spherical particles in microchannels.* The Journal of Chemical Physics, 2003. **119**(4): p. 2288-2295.

3. Gillespie, D., M. Valiskó, and D. Boda, *Density functional theory of the electrical double layer: the RFD functional.* Journal of Physics: Condensed Matter, 2005. **17**(42): p. 6609-6626.

4. Rosenfeld, Y., *Free-energy model for the inhomogeneous hard-sphere fluid mixture and density-functional theory of freezing.* Physical Review Letters, 1989. **63**(9): p. 980-983.

5. Mansoori, G.A., et al., *Equilibrium Thermodynamic Properties of the Mixture of Hard Spheres.* The Journal of Chemical Physics, 1971. **54**(4): p. 1523-1525.

6. Boda, D. and D. Gillespie, *Calculating the electrostatic potential profiles of double layers from simulation ion density profiles.* Hungarian Journal of Industry and Chemistry, 2013. **41**(2): p. 123-130.

7. Kierlik, E. and M.L. Rosinberg, *Density-functional theory for inhomogeneous fluids: Adsorption of binary mixtures.* Physical Review A, 1991. **44**(8): p. 5025-5037.

8. Blum, L., *Mean spherical model for asymmetric electrolytes.* Molecular Physics, 1975. **30**(5): p. 1529-1535.

9. Blum, L. and Y. Rosenfeld, *Relation between the free energy and the direct correlation function in the mean spherical approximation.* Journal of statistical physics, 1991. **63**(5): p. 1177-1190.

10. Li, Z. and J. Wu, *Density Functional Theory for Planar Electric Double Layers: Closing the Gap between Simple and Polyelectrolytes.* The Journal of Physical Chemistry B, 2006. **110**(14): p. 7473-7484.

11. Yu, Y.-X. and J. Wu, *Density functional theory for inhomogeneous mixtures of polymeric fluids.* The Journal of Chemical Physics, 2002. **117**(5): p. 2368-2376.

12. Jiang, J., et al., *A molecular-thermodynamic model for polyelectrolyte solutions.* The Journal of chemical physics, 1998. **108**(2): p. 780-784.

13. Liu, K., P. Zhang, and J. Wu, *Does capillary evaporation limit the accessibility of nonaqueous electrolytes to the ultrasmall pores of carbon electrodes?* The Journal of Chemical Physics, 2018. **149**(23): p. 234708.

14. Alejandro Gallegos, G.M.C.O., Jianzhong Wu, *Thermodynamic non-ideality in charge regulation of weak polyelectrolytes.* 2021.

Chapter 3. Charge regulation of natural amino acids in aqueous solution

The aim of this work is to develop a predictive model for describing the electrostatic behavior of twenty natural amino acids under diverse solution conditions. A coarse-grained model is proposed to account for the key ingredients of thermodynamic nonideality arising from the interaction of amino acids with various solvated ions in an aqueous solution including the molecular volume exclusion effects, solvent-mediated electrostatic interactions, van der Waals attraction, and short-ranged hydrophobic and hydration forces. With a small set of parameters characterizing the intermolecular interactions and dissociation equilibrium, the thermodynamic model is able to correlate extensive experimental data for the activity coefficients and solubility of amino acids in pure water and in aqueous sodium chloride solutions. Moreover, it predicts apparent equilibrium constants for the charge regulation of all amino acids in excellent agreement with experimental data. Importantly, the thermodynamic model allows for the extrapolation of the intrinsic equilibrium constants for amino-acids conversion among different charge states (*viz*., negative, neutral and positive charges) thereby enabling the prediction of the charge behavior for all natural amino acids under arbitrary solution conditions.

3.1 Introduction

Natural amino acids may exist as cations, zwitterions or anions in an aqueous solution depending on pH and other thermodynamic conditions including the ionic strength and solution composition. The charge regulation, *viz*., a precise control of the electrostatic states of amino acids, is important not only for understanding their solution behavior but

it has also direct implications to a broad range of biochemical applications of polypeptides and proteins such as bioadhesives, smart materials, and anti-fouling coatings[1].

The speciation of amino acids is intrinsically affiliated with the protonation/deprotonation of the terminal amino and carboxyl groups as well as the ionizable side chains such as those from Lys, Arg and Glu. Whereas the degree of proton association/dissociation of a simple acid or base can be readily determined from the Henderson–Hasselbalch equation[2], predicting the electrostatic states of amino acids is much more complicated due to the coupling of acid-base reactions. As a matter of fact, conventional experimental methods are not able to determine the acid-base equilibrium constants for the individual functional groups of amino acids precisely. Significant variations can often be found in the literature for the experimental values of these fundamental constants[3],[4]. Mostly, the uncertainties can be attributed to neglecting the thermodynamic ideality due to the interactions of amino acids with themselves and with other ionic species in the solution. For practical convenience, the thermodynamic nonideality may be lumped together leading to the definition of so-called *apparent* equilibrium constants[5].

For a weak acid or base with a single ionizable group, the apparent equilibrium constant is equivalent to the true thermodynamic equilibrium constant in the limit of infinite dilution (*viz*. in an ideal solution). However, measuring the equilibrium constant for amino acids at infinite dilution is not practical; the reported values from the literature are usually obtained at salt concentrations as low as 10 mM. The type of experimental procedures (e.g., titration or spectroscopy) can also factor in to the difference in the

equilibrium constants[6]. In principle, the equilibrium constants can be predicted from quantum-mechanical (QM) calculations. However, existing computational methods provide only order-of-magnitude estimates for the formation free energies and the numerical results are highly sensitive to the selection of approximations including the system size and specific basis sets[7]. Reliable theoretical tools are yet to be developed such that they can predict apparent equilibrium constants well correlated with experimental data at finite salt concentrations and allow for the extrapolation of the experimental results to amino acids at infinite dilutions.

For a particular acid-base equilibrium, the *apparent* equilibrium constant amalgamates the true thermodynamic equilibrium constant with the activity coefficients of all pertinent chemical species. The experimental values are closely related to the thermodynamic nonideality due to solvent-mediated interactions among ionic as well as non-ionic species. Because of the long-range nature of electrostatic correlations, the activity coefficients can vary quite significantly even at low salt concentrations. Besides, various types of solvent-mediated short-range interactions could make the *apparent* equilibrium constant highly sensitive to the identity of salt ions.

In a dilute solution, amino-acid dissociation deviates from the ideal behavior mostly dominated by their electrostatic interactions with ionic species. In this case, the Debye-Hückel (DH) theory provides a reasonable description of the activity coefficients. It predicts that, at low salt concentrations, the activity coefficients of charged species (modeled as point charges) sharply falls as the ionic strength increases[8]. Because of the strong deviation from thermodynamic ideality even at low salt concentrations, the

equilibrium constants cannot be reliably extrapolated from theoretical methods that do not reproduce the DH limiting law. As the DH theory becomes problematic when the salt concentration increases, the activity coefficients can be described by using various forms of the extended DH theory. The revised forms retain the strong dependence of activity coefficients on the ionic strength when the salt concentration is sufficiently small, but predict a slower decay in the activity coefficient at moderate and high salt concentrations. In comparison with experimental data, the extended DH theories are typically valid up to about 0.1 M for a monovalent electrolyte solution. Beyond which the mean activity coefficient of the electrolyte shows a minimum and increases at higher salt concentrations. In the latter case, the short-range interactions between ionic species become important, thus additional terms must be included to describe the activity coefficients quantitatively[9].

For applications to acid-base equilibrium, a conventional method to represent the activity coefficients of various ionic species at high salt concentrations is given by the specific ion interaction (SIT) model[10, 11]. In this method, the activity coefficients of individual ions are predicted by the DH theory augmented with an additional term for non-electrostatic interactions that is linearly proportional to the ionic strength. The proportionality coefficient is typically a function of the ionic strength and varies with the identity of ionic species. Despite its simplicity, SIT performs reasonably well in capturing the activity coefficients associated with acid-base equilibrium in simple electrolytes. It has been applied to predicting the charge regulation of acids and bases whereby the activity coefficient of a neutral species is assumed to be only a linear function of the ionic strength[6, 12, 13]. A slightly more rigorous approach can be taken to describe non-

electrostatic interactions through the virial expansion of the non-electrostatic component of the excess Gibbs energy similar to the Pitzer's equation[14]. The modified SIT model provides a near quantitative description of the activity coefficients for aqueous solutions of both monovalent salts and more complex electrolytes. Regrettably, the quantitative performance of this model is compromised by a large number of empirical parameters that are often unavailable for different amino acids[15]. Not surprisingly, Pitzer's equation does well to capture the charge regulation of many acids and bases if sufficient experimental data are available to fix the adjustable parameters; albeit it involves a more complex procedure for numerical fitting than that for the original SIT model[16].

Over the years, a number of thermodynamic models have been developed for electrolyte solutions beyond semi-empirical modifications of the DH theory[17, 18]. Among them, the mean-spherical approximation (MSA) represents a popular choice owing to its computational simplicity to describe the ion-size effects and electrostatic correlations[19]. MSA provides an accurate representation of both thermodynamic and transport properties of electrolyte solutions using only the ionic diameters as the adjustable parameters[20, 21]. Importantly, it obeys the limiting law at low electrolyte concentrations similar to various empirical modifications of the DH theory[22]. Besides, further improvements can be accomplished by including the ion pairing effects[23]. MSA has been previously applied to modeling the charge regulation of acids and bases[24-26]. For example, Vilariño et al. used MSA to correlate the apparent equilibrium constants of certain amino acids with satisfactory results[25, 27]. From a practical perspective, one additional benefit that MSA has over alternative theories is that the ion diameters have an intuitive

physical meaning and typically fall into the expected ranges for the size of hydrated ions. Typically, an anion is less hydrated than a cation of similar diameter and the same absolute charge, thus the Pauling radius provides a reasonable estimate of the anion diameter. In contrast, cation hydration is usually nonnegligible and its diameter must be adjusted to match the experimental data[28]. Remarkably, MSA is able to correlate both thermodynamic and transport coefficients quantitatively for many electrolyte systems up to high concentrations. Further improvement can be achieved by making the cation diameter and/or the dielectric constant concentration-dependent[20]. The concentration dependence of the ion diameter or the dielectric constant is understandable and may be determined from independent experiments[29, 30]. On the one hand, the change in the diameter of a hydration ion can be attributed to enhanced electrostatic screening as the salt concentration increases[31]. On the other hand, the dielectric constant of an electrolyte solution usually falls as more salt is added because the solvated ions reduce the orientational polarization of water molecules[30, 32].

It should be noted that the original MSA model provides a poor correlation of the activity coefficients of amino acids in aqueous solutions. For example, MSA fails to predict the reduction of activity coefficients of both glycine and sodium chloride when more glycine is added to a sodium chloride solution[33]. Because glycine exists predominately as zwitterions near a neutral pH, the primitive model would predict an increase of the activity coefficients for both glycine and sodium chloride. Understandably, an aqueous solution of amino acids is not a simple electrolyte. The thermodynamic properties depend not only on the solute size and long-range electrostatic interactions but also on the short-

ranged forces such as van der Waals (vdW) attraction, hydrogen bonding, and various forms of water-mediated interactions. Alternative models have been proposed to incorporate these interactions such that a quantitative correlation can be achieved for the activity coefficients of amino acids in pure water and in aqueous solutions[6, 34-37]. For example, a perturbation theory was proposed to correlate the activity coefficients of both amino acids and peptides in pure water[35]. With the primitive model of electrolyte solutions as the reference, two perturbation terms were introduced to account for thermodynamic properties due to short-ranged interactions: one for the dispersion forces as represented by the Lennard-Jones (LJ) potential, and the other for the angle-averaged dipole-dipole interactions between amino-acid molecules. While the dipole moments were determined from quantum-mechanical calculations, two adjustable parameters were introduced in the perturbation theory, *viz*., the size and energy parameters of the LJ potential for each amino acid or peptide. Quantitative correlations were established for the activity coefficients of a number of amino acids in liquid water. The perturbation approach was extended to study the activity coefficients of amino acids in sodium chloride solutions[34]. In the latter case, the mean activity coefficient of the electrolyte was described by the MSA model with concentration-dependent cation diameters[31], and the perturbation terms include additional thermodynamic contributions due to the LJ attraction, the dipole-induced dipole and the charge-dipole interactions between amino acids and the electrolyte. With the model parameters previously determined for amino acids in pure water and two additional adjustable parameters, the perturbation theory was able to correlate the activity coefficients for both amino acids and sodium chloride quantitatively.

Whereas most electrolyte theories assume water as a dielectric continuum, a number of so-called non-primitive models have been proposed over the years[38, 39]. Held et al. considered, in addition to the explicit representation of water molecules, the anisotropic nature of amino-acid molecules by using the perturbated-chain version of the statistical associating fluid theory (PC-SAFT)[40]. With a generic model accounting for hard-sphere repulsion, dispersion attraction, and association of polar groups, PC-SAFT was able to correlate various thermodynamic properties (e.g., densities, vapor-pressure depression, activity coefficients, and solubility) for many amino acids and oligopeptides in pure water. The thermodynamic model was extended to amino acids-electrolyte aqueous systems by combining DH and the PC-SAFT equations[36]. They captured the thermodynamic properties of electrolyte/amino acid/water solution by using only parameters determined from the pure component case with mixing rules that do not include any correction terms. The ability to predict the mixture behavior using pure component parameters is beneficial because data on mixtures is often limited or even not available. Later, Bang et al. studied a wide variety of these mixture systems using a combination of MSA and PC-SAFT equations[41]. It was found that the formation of ion-pair complexes is indispensable to describe the activity coefficients of both amino acids and the electrolytes. While PC-SAFT provides an excellent correlation of diverse experimental results for electrolyte systems, it involves a large set of adjustable parameters.

The purpose of this work is to develop a coarse-grained model that can be used to predict the charge behavior of all twenty natural amino acids under various solution conditions. While there has been extensive work on acid-base equilibrium and activity

coefficients of electrolyte solutions, a predictive model has yet to be developed for charge regulation in aqueous solutions of natural amino acids. In order to derive reliable equilibrium constants for protonation/deprotonation reactions, such a model should well capture the key forces underlying interactions between amino acids and all ionic species in an aqueous medium. While a precise description of hydration and water-mediated forces in microscopic details is beyond the scope of the present work, their contributions to thermodynamic properties can be described with semi-empirical models such that the parameters can be fixed by comparison with experimental data. The coarse-grained approach allows us to best estimate the equilibrium constants for transition among different electrostatic states of natural amino acids.

3.2 Thermodynamic models and methods

3.2.1 Electrostatic states of amino acids

Each amino acid consists of a carboxyl group and an amine group that can be dissociated in an aqueous solution through acid-base equilibrium. These functional groups, along with possible ionizable groups in the side chain, are responsible for the titration behavior of amino acids. Therefore, amino acids can exist in multiple states of electrostatic charge depending on pH and other solution conditions.

Conventionally, natural amino acids are divided into three types according to the charge propensity of the side chain, *viz*., neutral, acidic, and basic. Neutral amino acids contain no ionizable group in their side chains and exist predominately in charge neutrality over a broad range of solution conditions. Meanwhile, an acidic amino acid contains an acidic functional group in its side chain (e.g. glutamic acid has a carboxyl group in its side

chain) which is deprotonated near a neutral pH. Similarly, a basic amino acid contains a basic functional group in the side chain and typically bears a positive charge at pH = 7.0.

Because each amino-acid molecule contains at least two ionizable groups, we can express the acid-base equilibrium in terms of multi-step protonation reactions:

$$A_{i-1} + H^+ \leftrightarrow A_i \tag{3.1}$$

where A_i refers to the amino acid in a given charge state. The amino-acid valence satisfies $Z_{A_i} = Z_{A_0} + i$ with Z_{A_0} being the charge for the amino acid in its fully deprotonated state. For a neutral or basic amino acid, Z_{A_0} = -1; while an acidic amino acid molecule has a valence of -2 in its fully deprotonated state. Accordingly, in the fully protonated state, each neutral or acidic amino acid has a valence of +1 and each basic amino acid has a valence of +2.

The equilibrium constant for the protonation reaction is related to the solution composition and activity coefficients

$$K_i^T = \frac{c_{A_i}}{c_{H^+} c_{A_{i-1}}} \frac{\gamma_{A_i}}{\gamma_{H^+} \gamma_{A_{i-1}}} \tag{3.2}$$

where K_i^T represents the equilibrium constant for the amino acid in charge state *i*, superscript *T* denotes true thermodynamic constant, c_i is the molar concentration of species *i*, and γ_i is the corresponding activity coefficient. The equilibrium constant is a thermodynamic quantity defined by the change in the chemical potentials of reactants and products at their reference states, i.e., each species in an ideal solution at unit molar

concentration. The activity coefficients account for the effect of solvent-mediated interactions among all species in the solution.

Whereas the solution composition can be monitored with various analytical tools, it is not feasible to directly measure the true thermodynamic equilibrium constant because the activity coefficients are often unknown. As a result, the equilibrium constant is conventionally expressed in terms of its apparent value in most practical applications

$$K_i' = \frac{c_{A_i}}{c_{H^+} c_{A_{i-1}}} = K_i^T \frac{\gamma_{A_{i-1}} \gamma_{H^+}}{\gamma_{A_i}}. \tag{3.3}$$

At infinite dilution, the activity coefficients are negligible and there is no difference between apparent and true equilibrium constants. In general, the *apparent* equilibrium constant is influenced not only by the change of the chemical potential for individual species but also by the solvent-mediated interactions.

The apparent equilibrium constant provides direct information on the speciation of the amino acid at a given solution condition. From the equilibrium constant and mathematical relations between the activity coefficients and the solution composition, Eq. (3) allows us to calculate the distributions of amino acids in different charge states. For fitting the thermodynamic model with experimental results, it is convenient to express the equilibrium constants in the logarithmic form:

$$\log K_i' = \log K_i^T + \log \gamma_{A_{i-1}} + \log \gamma_{H^+} - \log \gamma_{A_i}. \tag{3.4}$$

The experimental results for acid-base equilibrium are typically reported in terms of the solution concentrations and apparent equilibrium constants. Therefore, a thermodynamic

model is needed to calculate the activity coefficients and extrapolate the true equilibrium constants.

3.2.2 A coarse-grained model for the aqueous solutions of amino acids

As mentioned above, a number of models can be used to represent the thermodynamic properties of amino acids in aqueous solutions, ranging from those with atomic details to coarse-grained approaches with the solvent treated implicitly as a dielectric continuum[5, 40, 42, 43]. Nonetheless, understanding the charge regulation of amino acids remains a challenge because it requires quantitative predictions of both equilibrium constants and activity coefficients over a broad range of conditions. In this work, we propose an augmented primitive model (APM) to represent amino acids in aqueous systems. The coarse-grained model describes amino acids in different charge states as hard spheres with different charge valences. Given a solution condition, the system composition can be determined from the protonation/deprotonation equilibrium self-consistently.

For simplicity, we lump all short-range interactions between amino acids and ionic species in an aqueous environment in terms of a square-well potential. While the simple model lacks atomic details and does not reflect the anisotropic nature of amino acid molecules, it provides a flexible framework to represent thermodynamic nonideality due to the transition between different electrostatic states of amino acids. As discussed above, the primitive model has been routinely used to correlate the activity coefficients of aqueous electrolyte solutions. We expect that similar results can be accomplished by adding a square-well potential to account for different types of solvent-mediated interactions.

Despite extensive research, a first-principles approach to predicting water-mediated interactions remains theoretically challenging. The adoption of an effective potential allows us to capture the essential physics without evoking the microscopic details[19, 44].

With amino acids in aqueous systems represented by APM, the activity coefficient for each solute species can be decomposed into contributions from the molecular excluded-volume effects, electrostatic correlation, and water-mediated interactions

$$\ln \gamma_i = \beta\mu_i^{ex} = \beta\mu_i^{hs} + \beta\mu_i^{el} + \beta\mu_i^{sw}. \quad (3.5)$$

In Eq. (5), μ_i^{ex} stands for the excess chemical potential, *i.e.*, deviation from the chemical potential of species *i* in an ideal solution; $\beta = 1/(k_B T)$, k_B is the Boltzmann constant and *T* the absolute temperature, μ_i^{hs}, μ_i^{el}, and μ_i^{sw} are contributions to the excess chemical potential due to hard-sphere repulsion, electrostatic correlation, and solvent-mediated interactions, respectively. The expression for each contribution has been discussed previously and can be found in Supporting Information[22, 44, 45].

The parameters in our activity-coefficient model include the hard-sphere diameters for each amino acid in different charge states, and the energetic parameters for short-range attraction between different pairs of solute species. The determination of these parameters will be discussed in the following section. As discussed previously, the hard-sphere diameters are ascribed to the excluded volumes of hydrated molecules thereby depending on the charge states. We find that, approximately, both the hydration shell thickness and water-mediated interactions are invariant with the change in salt concentration. At a fixed

temperature, all the model parameters are assumed to be independent of solution conditions.

3.2.3 Evaluation of model parameters

We calibrate the activity-coefficient model for twenty natural amino acids in aqueous solutions of sodium chloride. A similar procedure can be applied to aqueous systems containing other ionic species. The sodium chloride system is considered in this work because extensive data are available for all amino acids.

To determine the activity coefficients and the apparent equilibrium constant of acid-base equilibrium, we need the size parameters for all solute species and the interaction energies. For amino acids dissolved in a NaCl solution, the solute species include amino acids at different charge states, sodium and chloride ions, and protons or hydroxide ions. The parameters affiliated with amino acids can be fixed by fitting the APM model with their activity coefficients and solubility data in pure water and in aqueous NaCl solutions. As discussed previously, the equilibrium constants for amino-acid transition between different charge states can be determined by extrapolation of the apparent equilibrium constants to those corresponding to infinite dilution.

In addition to amino acids, we need to fix the model parameters for salt ions, protons and hydroxide ions. Previous studies indicate that the activity coefficients of electrolyte solutions can be well correlated with the hard-sphere diameters of cations and anions within the mean spherical approximation (MSA)[20, 28]. At high salt concentrations, the performance can be further improved by making the cation diameter

and/or the dielectric constant decrease with the ionic strength. In this work, we are mainly concerned with protonation/deprotonation of amino acids in aqueous electrolytes of moderate concentrations. As shown in Figure S2, the MSA model is able to reproduce the mean activity coefficient for sodium chloride from 0.0 to 1.0 M with the chloride ion diameter fixed at 3.62 Å (its Pauling diameter) and that of the sodium ion equal to 3.22 Å[46]. Throughout this work, the hard-sphere diameter for hydronium ions is fixed at 5.00 Å because it provides a reasonable estimation of the apparent equilibrium constant for natural amino acids when we ignore the activity coefficients for amino acids. This value falls within range of hydronium ion diameter previously estimated by others[20, 27]. The diameter of hydroxide ions was determined to be 3.22 Å based off comparison to the activity coefficients of sodium hydroxide solutions (see Figure S2).

Throughout this work, the parameters for amino acids were estimated by fitting experimental data using MATLAB R2018b. These parameters, along with comparison of the numerical results and experimental data, are presented in Supporting Information (Table S1). In all cases, the following objective function (OF) was minimized:

$$\mathrm{OF} = \sum [X_i^{exp} - X_i^{cal}]^2 \tag{3.6}$$

where X refers to the experimental data for the activity coefficients of amino acids, solubility data, or the apparent equilibrium constants. In certain cases, such as systems containing amino acids in charged states, it was necessary to combine the objective function for different data sets in order to best fit all experimental results.

3.3 Results and discussion

3.3.1 Activity coefficients for amino acids

Given the added complexity in determining the activity coefficients of amino acids with multiple charge states, we first consider neutral amino acids in aqueous sodium chloride solutions around their isoelectric points. In this case, we can correlate the experimental data in terms of three adjustable parameters: the hard-sphere diameter for each amino acid and the energy parameters for amino acid interaction with itself and with the salt ions. These parameters are listed in Table S1.

Figure 1 presents a comparison between theoretical correlations and experimental data for the activity coefficients of four representative neutral amino acids at different salt concentrations with the solution pH equal to the respective isoelectric points. Similar comparisons for other neutral amino acids and for the activity coefficients of NaCl are presented in the SI. In all cases, our model provides a quantitative description of how the activity coefficient of amino acid is influenced by its own concentration and by the changes in the solution conditions. For each neutral amino acid, the hard-sphere diameter is typically 10 to 20% smaller than the molecular diameter determined from the van der Waals (vdW) volume[47]. The smaller effective size may be attributed to the non-spherical shape and the zwitterionic nature of neutral amino acids[48]. Interestingly, the tendency of the hard-sphere diameter to approach that corresponding to the vdW volume increases with the hydrophobicity of the amino acids (e.g. Ala → Val → Leu). The trend may be attributed to the compression of the water layer in the solvation shell as the amino acid becomes more hydrophilic[49]. The choice of the vdW volume to describe the molecular

size of the neutral amino acids provides a semi- to near quantitative fitting of the activity coefficient data, thus to a certain degree demonstrating the physical significance of our model parameters.

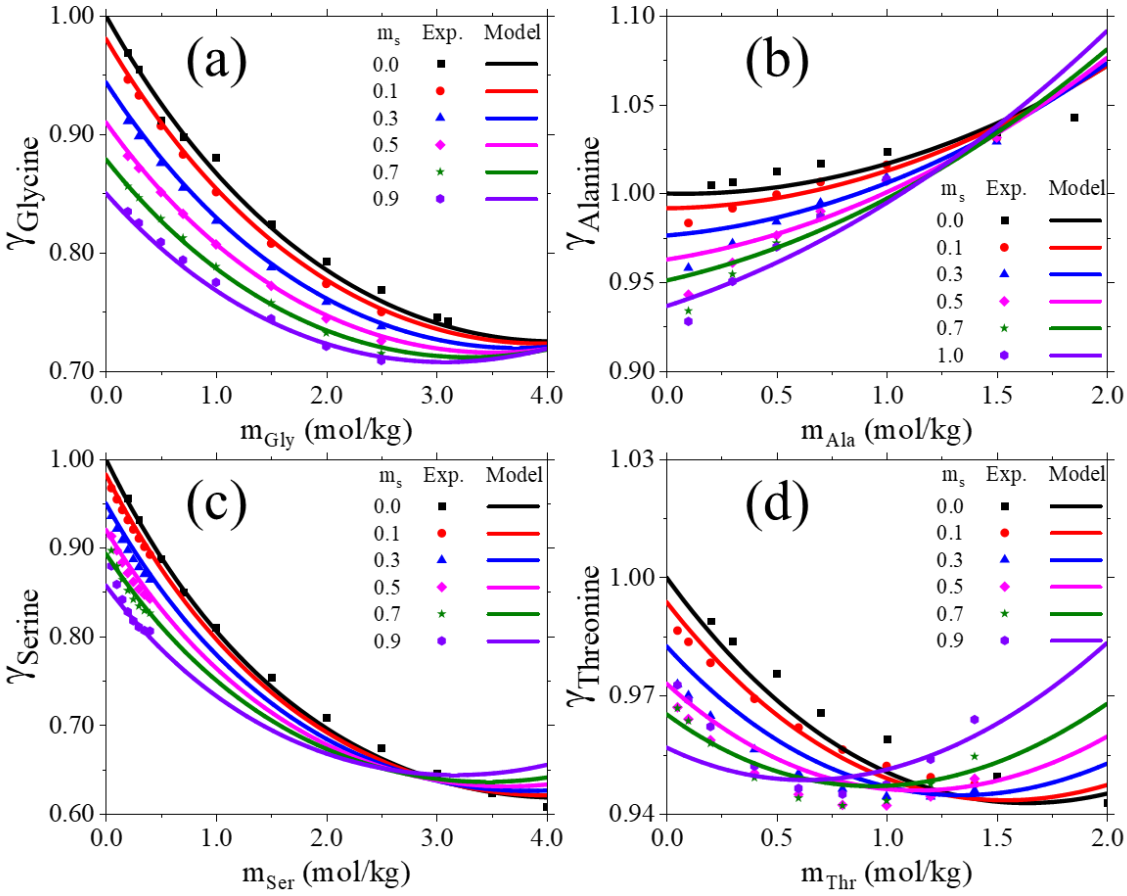

Figure 3.1 Activity coefficients of (a) glycine (pI = 5.98), (b) alanine (pI = 6.18), (c) serine (pI = 5.72) and (d) threonine (pI = 5.72) in NaCl aqueous solutions from experiment[34, 50-52] (symbols) and from theoretical correlations of the coarse-grained model (solid lines).

Figure 1a shows that the activity coefficient of glycine falls significantly as the concentration of amino acid or NaCl increases. The decline in the activity coefficient may be attributed to the high polarity of glycine molecules and to the salting out effect of sodium

chloride. Because the amino acid is modeled as a neutral particle in this work, the negative deviation from the ideal solution behavior implies that the attractive contribution has a greater effect than the repulsion or the excluded-volume effects. At high salt molality, the different curves in Figure 1a collapse almost into a single point suggesting that the hard-sphere repulsion outcompetes the attraction contributions. The non-linear dependence is consistent with the prediction of the hard-sphere model for the excess chemical potential[45].

Figure 1b shows that the activity coefficient of alanine increases with its concentration but decreases with more added salt. In comparison to glycine, however, the dependence on the salt concentration is significantly smaller. These trends are captured in our model by using a non-electrostatic attraction energy between the amino-acid molecules lower than that for their interaction with the salt ions. The decrease in the attraction energy from Gly to Ala can be attributed to hydrophobic effect due to the addition of the methyl group in Ala's side chain, which causes the water layer surrounding the amino acid to become more structured. The higher ordering of the solvation shell weakens the interaction of the amino acid with itself and salt ions in the solution. A similar trend was observed for the change in the interaction energy from Ala to Val and from Ser to Thr as shown in Table S1. In both cases, the energy parameters falls as the carbon number of the side chain increases and the trend agrees with the findings by Khoshkhabari et al.[35].

If instead of a methyl group, a hydroxyl group is added to alanine's side chain so that it becomes serine, the activity coefficient will behave similarly to that of glycine (Figure 1c). In fact, the concentration effect on the activity coefficient of serine is much

more substantial than that for glycine in terms of both the addition of amino acid and salt ions. This stronger dependence on the solute concentration is somewhat expected due to the increased capability of hydrogen bonding and dipole-dipole interactions that the additional hydroxyl group can contribute. Interestingly, if the addition to alanine were instead a weak carboxyl group like in the case of aspartic acid, the interaction strength would decrease (Table S1). While this may seem to oppose the previous statement regarding the increase of attraction due to the polar functional group, it may rather be explained by the interaction between the ionic carboxyl group and the ionic amine group, which weakens the intermolecular interactions[53]. If the carboxyl group is replaced by an amide group like that in asparagine, the attraction energy greatly increases due to the amide group not participating in ionic interactions with the amine group.

Figure 1d shows the activity coefficient of threonine, which differs from serine by one methyl group. It exhibits an initial decrease in the activity coefficient with added threonine or salt, but the trend is reversed when the threonine molality is greater than about 1.0 mol/kg. The non-monotonic trend is expected because of the reduced attractions of the amino acid with itself and with the salt in comparison to serine. While the discrepancy between the correlated model and experimental data is noticeable, we consider the overall performance as satisfactory given the simplicity of this model.

For amino acids in aqueous electrolyte solutions, experimental data are often available for the activity coefficients of both amino acids and ionic species. Figure S4 compares our correlated model of the mean activity coefficients of NaCl with the experimentally determined values for those systems considered in Figure 1. While these

data were not used in the fitting, our coarse-grained model is able to describe the mean activity coefficients of NaCl in good agreement with experiment. For all four amino acids considered, the mean activity coefficient of sodium chloride decreases with an increase in the molality of the amino acid or the salt ions. The decline in the mean activity coefficients can be attributed both to the water-mediated attraction between amino acids and salt ions and to the electrostatic correlation effects. It is known experimentally that, when amino acid is added to a NaCl solution, the dielectric constant increases due to the high polarity[54]. Because the dielectric constant is inversely proportional to the strength of electrostatic correlations (see Eq. S10), the addition of amino acids reduces electrostatic attraction between salt ions thereby reducing its contribution to the mean activity coefficient of sodium chloride. Indeed, one may improve the theoretical performance by considering the dielectric behavior of the solution due to added amino acids.

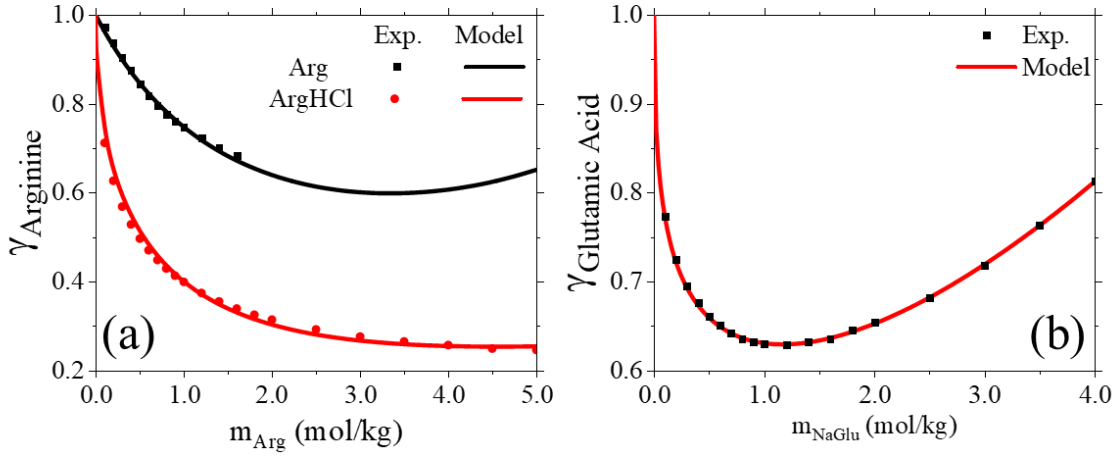

Figure 3.2 Activity coefficients of amino acids versus the molality for (a) arginine and its salted form (ArgHCl) in pure water and (b) for glutamic acid in salted form (NaGlu). Symbols are from experiments[53] and the lines are theoreticalcorrelations.

The computational procedure to obtain the parameters for acidic and basic amino acids is more complicated than that for neutral amino acids. This is because, even at the isoelectric point, there are more than one charge states in coexistence for an acidic or basic amino acid in an aqueous solution. In order to determine the activity coefficients of both amino acids and salt ions, we must consider the speciation of the amino acid at each solution condition. The composition of the amino acid at different charge states are solved from the condition of chemical equilibrium in couple with the mass and charge balance conditions. The activity coefficient for the amino acid is then given by a partial contribution from each charge state (see Eq. S16). For example, at its isoelectric point (pI = 3.24), approximately 88% of glutamic acid exists in the neutral state, and 6% in the positively and negatively charged states, respectively. While the neutral state has a dominant composition, the contribution from the charged states to the activity coefficient of glutamic acid cannot be neglected due to the importance of electrostatic interactions at low salt concentrations. For acidic and basic amino acids, the thermodynamic properties are often reported in the salted form (e.g. sodium glutamate, NaGlu). In that case, the amino acid exists predominately in the charged state (e.g. Glu^-) and the activity coefficient measured experimentally has an equal contribution from the amino acid and the salt ion (see Eq. S25).

In Figure 2, we present a comparison of our correlated model with the experimental data for the activity coefficients of arginine and glutamic acid. One noticeable difference between the amino acid in its salted and pure forms is that the salted form exhibits a much smaller activity coefficient due to the electrostatic contribution. The excellent agreement

exemplifies the power of the MSA model to account for electrostatic correlations over the entire concentration range considered in this work. Interestingly, for the activity coefficient of arginine hydrochloride (ArgHCl) shown in Figure 2a, it does not display the characteristic increase due to the hard-sphere repulsion, even up to 5 mol/kg of amino acid. The stark difference may be attributed to short-range attractions between large organic compounds absent in simple electrolyte solutions. As shown in Table S1, glutamic acid has a self-attraction energy significantly lower than that of arginine. As a result, it shows a minimum in the activity coefficient curve similar to that for simple electrolytes. While the effects of hard-sphere repulsion and electrostatic correlations are well accounted for within the MSA model, inclusion of short-range attractions is essential to properly capture the thermodynamic behavior of amino acids in aqueous electrolyte solutions.

3.3.2 Apparent equilibrium constants

Unlike the true equilibrium constant, the apparent constants depend on pH, and the concentrations of amino acids and salt. These two quantities are equivalent only for amino acids at infinite dilutions. Nevertheless, the apparent equilibrium constants provide insight into the charge regulation of amino acids in response to the changes of the solution conditions. Qualitatively, they can be used to replace the true thermodynamic equilibrium constants of protonation/deprotonation. Because of the thermodynamic nonideality as represented by the activity coefficients of all pertinent species, we expect that apparent equilibrium constants would lead to good quantitative performance only over a narrow range of solution conditions.

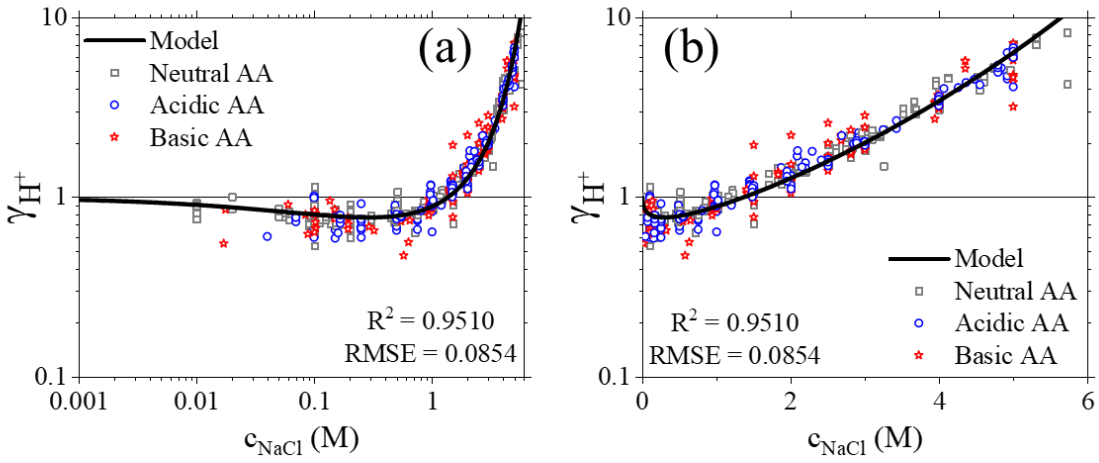

Figure 3.3 The activity coefficient of protons (hydronium ions) extracted from the apparent equilibrium constants of natural amino acids in NaCl solutions (Eq. 7). Here the solid line corresponds to the MSA prediction with the hard-sphere diameter for hydronium ions equal to 5.00 Å, and the symbols are based on the experimental data for the apparent equilibrium constants with the activity coefficients of amino acids calculated self-consistently.

To facilitate a connection between apparent and true equilibrium constants, we may express their ratio in terms of the activity coefficient of protons (*viz*., hydronium ions):

$$\gamma_{H^+} = \frac{K_i'}{K_i^T}\frac{\gamma_{A_i}}{\gamma_{A_{i-1}}}. \tag{3.7}$$

For an amino acid at low concentration, the ratio of the activity coefficients in the protonated and deprotonated state is approximately a constant, mostly determined by the change of the valence due to protonation/deprotonation. Meanwhile, the proton activity is mainly determined by its interaction with salt ions, insensitive to the types of amino acids at low concentration. Figure 3 shows a comparison of the theoretical prediction with the experimental results for apparent equilibrium constants with the activity coefficients of the amino acids predicted self-consistently. Approximately, the experimental data points

collapse into a single curve for all natural amino acids. The scattering of data around the line for the proton activity coefficient shows experimental errors and non-specific salt effects not considered in the model. As the salt concentration approaches zero, the curve approaches the ideal limit (i.e. $\gamma_i = 1$). In this case, the apparent equilibrium constant would be equivalent to the true thermodynamic equilibrium constant. Most experimental results are obtained at sodium chloride concentrations at or greater than 0.01 M. Because experimental measurements of the apparent equilibrium constants become more difficult at lower salt concentrations, a reliable theoretical model is needed to extrapolate an accurate value of the true equilibrium constant for the protonation/deprotonation equilibrium of each amino acid.

We have compared how well the model describes the apparent equilibrium constants as a function of the salt concentration for different types of amino acids. Figure 4 presents the results for alanine, a neutral amino acid. For each protonation step, the apparent equilibrium constant depends on the sodium chloride concentration. Since the size of alanine in its neutral state was determined previously by comparison with experimental activity coefficient data, there is only one adjustable parameter for each equilibrium reaction, *viz*., the size of the amino acid in its respective charge state. The parameters are shown in Table S1. As expected, alanine in its neutral state has the smallest volume, which agrees with the partial molar volume data for neutral amino acids in different charge states[55].

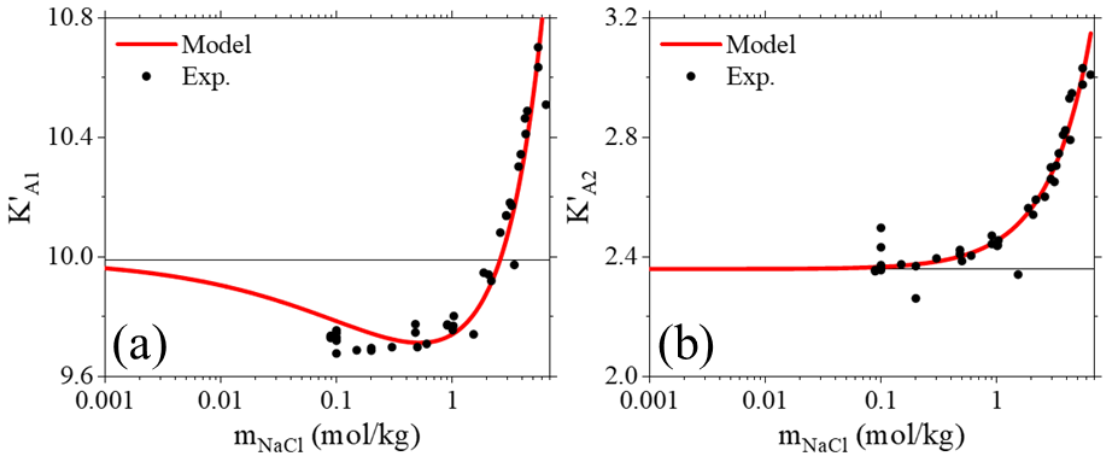

Figure 3.4 Apparent equilibrium constants for alanine versus NaCl concentration according to experiments[3, 16] (symbols) and our model (lines). (a) $Ala^- + H^+ \rightarrow Ala^0$ and (b) $Ala^0 + H^+ \rightarrow Ala^+$.

As shown in Figure 4, the apparent equilibrium constant for the first protonation step ($Ala^- + H^+ \rightarrow Ala^0$) exhibits a minimum as the salt concentration increases. However, such a minimum disappears for the second protonation step ($Ala^0 + H^+ \rightarrow Ala^+$). Instead, the apparent equilibrium constant increases monotonically with more added salt. The different trends arise from the imbalance between the absolute charge of reactants and products (*i.e.*, A^- and H^+ versus A^0) for the first protonation step while the second step has the same total absolute charge on both sides of the reaction. In the former case, there is an electrostatic contribution to the apparent equilibrium constant which would favor the direction toward more charged species. As a result, the apparent equilibrium constant decreases with an increase in the salt concentration in the dilute regime. For the second protonation step, there is equal charge for the reactants and the product, leading to the cancelation of the electrostatic contributions. At higher salt concentrations, both steps exhibit a characteristic increase in the apparent equilibrium constants with the ionic

strength. This can be explained by the larger contribution of the hard-sphere repulsion to the activity coefficients of the chemical species. The equilibrium shifts towards the side that has the least volume occupancy, i.e., the product side in both protonation steps. In other words, the excluded volume effects lead to the increase in the apparent equilibrium constant. A similar trend can be observed for other neutral amino acids as shown in Figures S11-14. Clearly, our model is able to well account for how the apparent equilibrium constant changes with the sodium chloride concentration for all neutral amino acids. It thus enables extrapolation of the true equilibrium constants from the experimental data.

Next, we consider how the apparent equilibrium constants for acidic amino acids change with the sodium chloride concentration by taking glutamic acid as an example. For an acidic amino acid, the protonation reaction takes place in three steps with the valence increasing from -2 to 1. Figure 5 shows a comparison between our model and the experimental data. For glutamic acid, the second and third protonation steps are similar to the two protonation steps for a neutral amino acid (e.g., alanine). The appearance of a minimum in the apparent equilibrium constant also highlights the importance of the charge distribution on both sides of the chemical reaction. In a dilute solution, the electrostatic interaction is the governing factor to determine how the apparent equilibrium constant changes with the salt concentration. By contrast, the trend is little variant with the specific functional groups being protonated. In all cases, we see a sharp increase in the apparent equilibrium constant at high salt concentrations. Meanwhile, the behavior at low salt concentrations is mainly determined by electrostatic effects on protonation reactions. The electrostatic effects also explain why the apparent equilibrium constant for the first

protonation step is significantly reduced in compared to that for the second step for glutamic acid and other acidic amino acids. Because the electrostatic component of the activity coefficient of an ionic species scales with the squared valence, glutamic acid in the divalent anion state is significantly more difficult to protonate at low salt concentrations (i.e. the equilibrium reaction favors the A^{2-} species) compared to that for the monovalent anion state (Figure 5b). Overall, the agreement of our correlated model with the experimental data is quite satisfactory.

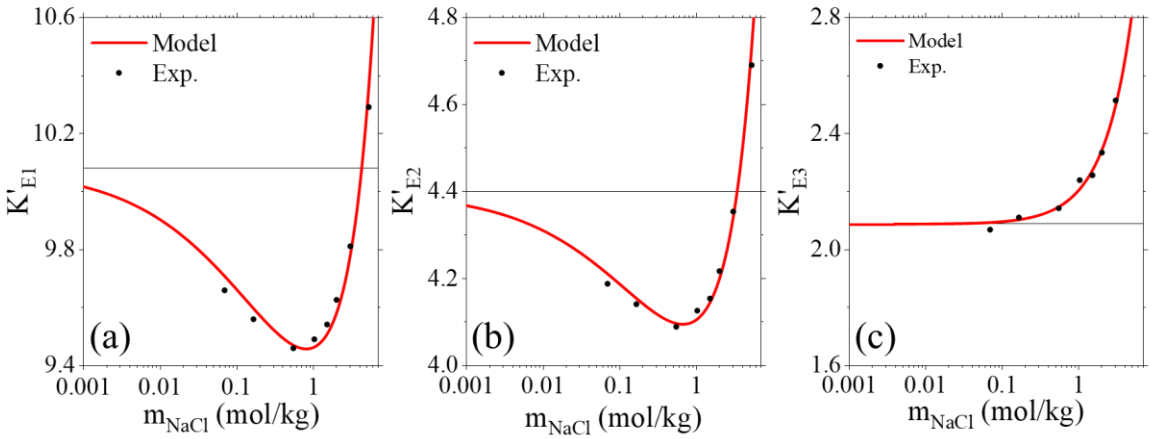

Figure 3.5 Apparent equilibrium constants of protonation for glutamic acid from experiments[12] (symbols) and model (lines). (a) $Glu^{-2} + H^{+} \rightarrow Glu^{-1}$, (b) $Glu^{-1} + H^{+} \rightarrow Glu^{0}$, and (c) $Glu^{0} + H^{+} \rightarrow Glu^{+}$.

Lastly, we consider the protonation of the basic amino acids in aqueous sodium chloride solution. Figure 6 shows the apparent equilibrium constants for the protonation of histidine from negative charge to divalent positive charges. The dependence of the apparent equilibrium constants on the salt concentration looks quite similar to that of alanine for the first two steps of protonation. However, histidine has one additional protonation step leading to a divalent cation state. By a comparison of Figure 6b and 6c, we see that both

apparent equilibrium constants exhibit a monotonic increase with the salt concentration. The third protonation step shows a more rapid increase in the apparent equilibrium constant than the second protonation step at low salt concentrations. As discussed earlier, the different effect results from the electrostatic contribution which favors the divalent cation form similar to the case with glutamic acid. Unlike a divalent anion in the glutamic acid case, however, the divalent cation form appears on the product side of the chemical equilibrium leading to a more rapid increase in the apparent equilibrium constant. The electrostatic behavior is well captured by our model demonstrating its capability in describing all steps of protonation for different amino acids.

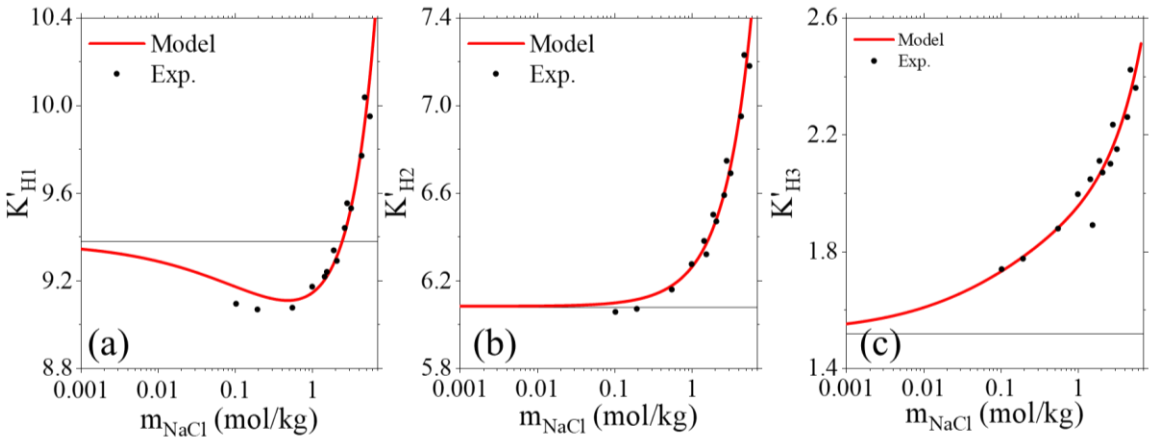

Figure 3.6 Apparent equilibrium constants of protonation for histidine from experiments[12] (symbols) and model (lines). ((a) $His^{-1} + H^{+} \rightarrow His^{0}$, (b) $His^{0} + H^{+} \rightarrow His^{+1}$ and (c) $His^{+1} + H^{+} \rightarrow His^{+2}$.

As mentioned earlier, the hard-sphere diameter of an amino acid depends on its charge state. The exact value is obtained by fitting the experimental data for apparent equilibrium constants for different types of protonation. We use only one adjustable parameter for each protonation step because the size of the amino acid in its neutral state

can be determined by comparison with the activity coefficient data. For example, the hard-sphere diameter for alanine is 4.91 Å in its neutral state. By correlation with the experimental data for the apparent equilibrium constants, we found that the diameters of alanine in the monovalent cation and anion states to be 5.76 Å and 5.44 Å, respectively. Qualitatively, the neutral state is the smallest in size in good agreement with the experimental data for the partial molar volumes of amino acids in different charge states [55, 56].

For a neutral amino acid, we can intuitively understand the increase of the hydration diameter as it becomes charged because of the appearance of ion-dipole interaction between the solute and the solvent. However, this trend does not hold for acidic or basic amino acids. We found that glutamic acid has the smallest hard-sphere diameter in the monovalent anion form. Similarly, histidine has the smallest hard-sphere diameter in the monovalent cation form. The change in the molecular size with the valence also agrees with the experimental data for the partial molar volumes of amino acids referenced earlier. For these cases, amino acid hydration in different charge states cannot be simply explained in terms of the electrostatic interactions. The hard-sphere diameter reflects an interplay between the side chain charge (and amino acid as a whole) and the hydrogen bonding capability of the functional group in the protonated and deprotonated forms. It has been found[57] that the carboxyl group decreases its partial molar volume upon deprotonating. The trend is opposite for an amine group which leads to a higher partial molar volume when it is deprotonated. The first protonation step of an amino acid involves the protonation of the amine group thus the hard-sphere diameter will decrease according to

the change in the partial molar volume. A basic amino acid has an additional amine group in comparison to alanine and its protonation from neutral to a positive charge reduces the size. As the protonation of the carboxyl group increases the size, an acidic amino acid should have a minimum size in the monovalent negatively charged state.

3.3.3 Thermodynamic equilibrium constants

In order to properly characterize the charge regulation of amino acids at arbitrary solution conditions, we need the true thermodynamic equilibrium constants for all protonation reactions, which are intrinsic properties for each amino acid depending only on temperature. These fundamental constants provide a basis to predict when amino acids will switch from one state to another due to the protonation of their functional groups. As mentioned earlier, these thermodynamic equilibrium constants are quite difficult to accurately measure and the reported values rely on extrapolation techniques with the activity coefficients correlated with various theoretical models. While conventional methods like Pitzer's equation and SIT are useful to correlate the activity coefficients, these models involve a large number of empirical parameters that are unavailable for many amino acids. By contrast, our model uses only a small number of parameters to account for the size effects and electrostatic correlations. All these parameters have clear physical significance and can be estimated from alternative experiments or obtained by correlation with a relatively small amount of experimental data. Therefore, our model allows us to extrapolate the true equilibrium constants of protonation for natural amino acids at room temperature

We demonstrated in the previous sections that our coarse-grained model well captures the electrostatic behavior of amino acids in sodium chloride solutions. Based on the calibration of our model with apparent equilibrium constants from the literature, we are able to obtain a set of thermodynamic equilibrium constants for protonation of all twenty amino acids (Table 1). These constants provided the best fit to the experimental data when they are extrapolated to the infinite dilution limit. While the results were extracted from experiments in NaCl solutions, the true thermodynamic equilibrium constants are independent of the properties of NaCl thus useful to determine the charge states of amino acids under other electrolyte conditions.

Table 3.1 Recommended equilibrium constants for the protonation of natural amino acids.

Amino Acid	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$
Ala	9.99	2.36	-
Arg	12.65[a]	9.01	1.75
Asn	9.02	2.13	-
Asp	10.17	3.97[a]	1.93
Cys	10.65	8.34[a]	1.85
Gln	9.24	2.15	-
Glu	10.08	4.40[a]	2.09
Gly	9.73	2.24	-
His	9.38	6.08[a]	1.52
Ile	9.82	2.34	-
Leu	9.82	2.34	-
Lys	10.95[a]	9.09	1.82
Met	9.41	2.15	-
Phe	9.27	2.14	-
Pro	10.73	1.95	-
Ser	9.29	2.16	-
Thr	9.20	2.23	-
Trp	9.60	2.30	-
Tyr	10.63[a]	9.30	2.10
Val	9.76	2.30	-

[a]The pK_a value for the side chain.

Tables S2-S4 compare the equilibrium constants obtained from this work with those extrapolated by using other methods like Pitzer's equation and SIT. In Table S2, we compare the predictions for certain neutral amino acids using different theoretical methods. We see that there is not much variation between the different methods for the two protonation steps. This is expected because all methods provide a good fitting of the activity coefficients. One key advantage of our model is that all parameters have clear physical meanings and can be fitted with a small amount of data. A similar finding is found for those amino acids with acidic side chains (Table S3) and those with basic side chains (Table S4). The general agreement across different thermodynamic models further highlights the benefit of our model in terms of its simplicity.

3.3.4 Speciation of amino acids in pure water and in sodium chloride solution

As discussed before, amino acids in water can exist in terms of multiple charge states due to the protonation of the functional groups. The change between different charge states is governed by the thermodynamic equilibrium constant. In section 3.3, we proposed a set of thermodynamic equilibrium constants for all twenty amino acids based off the correlations of our model to the apparent equilibrium constants measured by experiments. These constants can be employed under the assumption that the solution is ideal to study how the solution pH influences the speciation of amino acids. Alternatively, the apparent equilibrium constants measured experimentally can be used instead to demonstrate the effect that the non-ideality on the speciation. This is appropriate if the amino acid concentration is dilute compared to the electrolyte such that the protonation to its different states does not impact the solution conditions, particularly the ionic strength. This is also

true when the solution pH is adjusted since there will be noticeable changes in the ionic strength at very low or very high pH. With the true equilibrium constants and the thermodynamic model to account for interaction between solute species, we can now predict the speciation of amino acids in pure water and in NaCl solution at any concentrations.

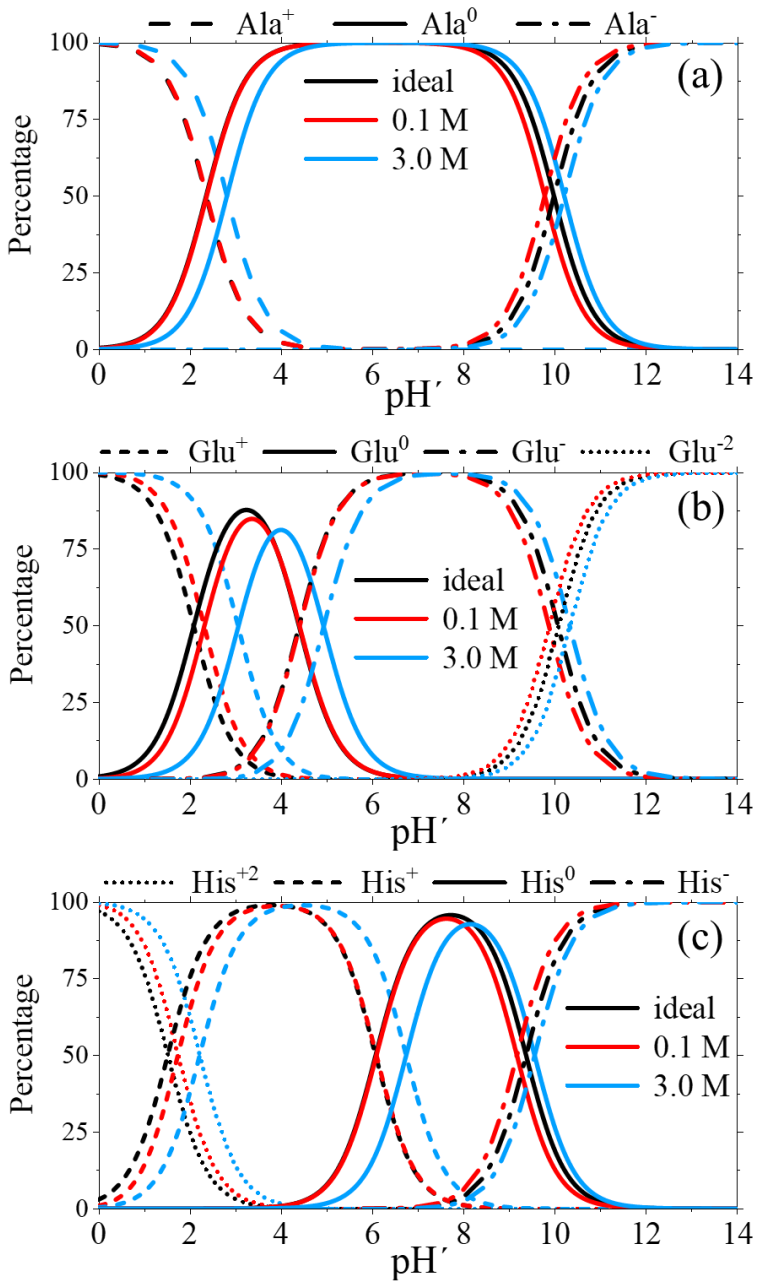

Figure 3.7 Speciation diagrams for three types of amino acids: (a) neutral, (b) acidic, and (c) basic as represented by alanine, glutamic acid, and histidine, respectively. Shown here are the percentages of amino acids in different charge states at ideal condition and in 0.1 and 3.0 M sodium chloride concentrations.

We may elucidate the speciation of amino acids by considering three representative amino acids in different NaCl concentrations. Figure 7 shows the speciation diagrams of alanine, glutamic acid, and histidine as the representatives of the neutral, acidic, and basic amino acids. The speciation is plotted in terms of $pH' = -\log[H^+]$, i.e. in terms of the logarithm of the concentration of free protons in the solution. A smooth transition between different states of amino acids is observed as pH´ increases. Figure 7a shows that alanine can speciate into three different charge states (i.e. Z = -1, 0, +1). In the pH range between 4 and 8, alanine exists entirely in the neutral state. At either lower or higher pH´, alanine exists in only two of the three possible charge states. Near the isoelectric point (pI = 6.18), we only need to consider alanine as a neutral species. During the transition from one charge state to another, there is an intersection between the two lines which reflects the apparent equilibrium constant at these conditions. Clearly, the apparent equilibrium constant changes with the salt concentration. In the first protonation step (i.e. $Ala^- \rightarrow Ala^0$), the pH´ required to protonate 50% of alanine to its neutral state shifts to a lower value as the salt concentration increases from the ideal-solution condition to 0.1 M, which reflects the decrease in the apparent equilibrium constant as shown in Figure 7a. However, a further increase in the salt concentration results in a higher pH´ (i.e., a lower hydrogen concentration) to reach the same level of protonation. As discussed previously, this reflects the shift towards the product side of the first protonation step in order to reduce the volume of the chemical species. For the second protonation step, there is no significant difference between the ideal and 0.1 M case due to the hard-sphere contribution being relatively

insignificant at low salt concentrations. A further increase in the salt concentration leads to the second protonation of alanine occurring at higher pH′ values.

As shown in Figure 7b and c, glutamic acid and histidine exist in four different charge states when they are dissolved in to an aqueous solution. In the region where the pH′ is between 6 to 8, glutamic acid exists in its monovalent anionic state which has a narrower pH′ window in comparison to that for alanine. On the other hand, the neutral state only reaches about 88% of the speciation of glutamic acid around its isoelectric point (i.e. pH′ = 3.24). This is the reason why it was necessary to consider the speciation of glutamic acid in calculating the activity coefficients. For acidic or basic amino acids, we could not consider only the predominate state like that used for the neutral amino acids. It is also interesting to note the large shift in the position of the isoelectric point and the decrease in the percentage that exists in the neutral state as the salt concentration increases. Figure 6c shows the speciation of histidine in pure water and two NaCl solutions. It exhibits a higher presence of the neutral state than that for glutamic acid because the protonation steps are not as close together with each other. As was seen for glutamic acid, there can be significant shifts in the speciation as a result of the increase in the salt concentration.

Another way to look at the speciation of amino acids is the pH effect on the average charge of the amino acid. Figure 8 shows how the average charge of alanine, glutamic acid, and histidine changes with pH′ for the ideal case and two NaCl concentrations. As expected, we see clear influence of the salt concentration on the net charge of the amino acids. For alanine, the second protonation step shows the onset of charge at higher pH′ values than that for the salt-free case. Glutamic acid, on the other hand, shows a noticeable

change in the net charge curve when the salt concentration increases (Figure 8b). At higher salt concentrations, we can observe how the narrowing and shift of the pH´ window for the glutamic acid in its neutral state results in a significant change in the net charge in the range of pH´ between 2 to 6. While histidine does not exhibit as close of protonation steps as glutamic acid, the increase in salt concentration still makes noticeable changes in the net charge in the transition regions (Figure 8c).

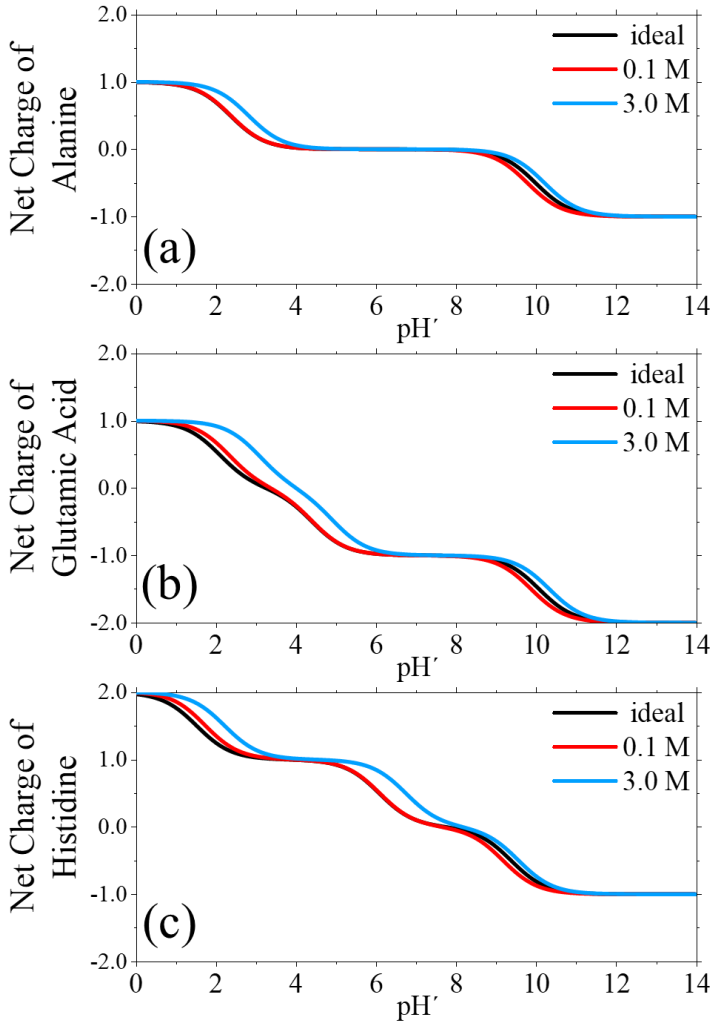

Figure 3.8 The net charge of (a) neutral, (b) acidic, and (c) basic amino acids versus pH´ (i.e. the hydrogen concentration) at three solution conditions: pure water, 0.1 M and 3.0 M sodium chloride solutions.

3.4 Conclusion

In this work, we have developed an augmented primitive model (APM) that captures the thermodynamic properties of all twenty natural amino acids in aqueous sodium chloride solutions. In addition to the volume exclusion effects due to the finite size of the solute species and electrostatic correlations, APM accounts for the water-mediated attraction of amino acids with themselves and with salt ions. The activity coefficients are described in terms of the mean-spherical approximation (MSA) for the primitive model and a mean-field approximation for the short-range attraction represented by the square-well potential.

We first demonstrated that APM is able to describe the activity coefficients of amino acids in pure water and in aqueous sodium chloride solutions. This was done by correlating the size parameters of each amino acid in the neutral state and the interaction energy with itself and with the salt ions. The model is able to quantitatively capture the experimental data for all natural amino acids. More importantly, the parameters determined through the fitting procedure maintain physical significance (e.g. the size of the amino acid in its neutral state is consistent with that determined from the van der Waals volume). We find that the short-range attraction tends to outcompete the repulsive excluded volume contribution to the activity coefficient leading to a reduction in the activity coefficient upon addition of either amino acid or salt. In addition, the addition of polar groups results in an increase in the interaction energy while hydrophobic groups diminish the interaction energy. Through the use of MSA to account for the finite size effects and electrostatic correlations, our model can well capture the change in the activity coefficient between the

neutral state and charged state of the amino acid by comparison to the experimental data for amino acids in their natural and salted (e.g. ArgHCl) forms.

Lastly, we considered the speciation of amino acids by comparison of the theoretical predictions with experimental data for the apparent equilibrium constants in aqueous sodium chloride solutions. Our model is able to well describe the apparent equilibrium constants for all protonation steps of neutral, acidic, and basic amino acids by correlating the size of the amino acid in its different charge states and the thermodynamic equilibrium constants (i.e., the apparent equilibrium constants in the limit of infinite dilution). The change in size between the different charge states of amino acids is in good agreement with the partial molar volume data determined from experiments. The simplicity of our coarse-grained model and its ability to accurately capture the charge regulation of all amino acids provide confidence in our proposed set of the equilibrium constants of different protonation reactions for all natural amino acids. These fundamental constants will be valuable in determining the charge states of amino acids and understanding their biochemical properties under complicated solution conditions.

Supporting Information

S3.1. Model parameters

Table S 3.1 The hard-sphere diameters in units of angstrom and the interaction energy in units of k_BT for amino acids in different charge states based off the correlation to experimental data for the activity coefficient, solubility, and apparent equilibrium constant.

Amino Acid	σ_{vdW}	$\sigma_{Z=-2}$	$\sigma_{Z=-1}$	$\sigma_{Z=0}$	$\sigma_{Z=+1}$	$\sigma_{Z=+2}$	ε_{AA}	ε_S
Ala	5.48	-	5.44	4.91	5.76	-	0.47	0.69
Arg	6.85	-	6.67	6.28	5.44	6.02	0.63	0.84
Asn	6.13	-	5.88	5.46	5.97	-	0.81	0.99
Asp	6.10	6.07	5.04	5.43	6.28	-	0.15	0.80
Cys	5.85	6.33	5.87	5.08	4.81	-	1.64	0.89
Gln	6.41	-	6.45	5.72	6.14	-	0.65	0.85
Glu	6.38	6.36	5.56	5.68	6.07	-	0.12	0.65
Gly	5.09	-	5.77	4.74	5.61	-	0.84	0.89
His	6.09	-	6.58	6.09	5.71	6.81	0.81	0.87
Ile	6.42	-	7.21	6.42	6.61	-	0.15	0.54
Leu	6.42	-	7.21	6.42	6.61	-	0.15	0.54
Lys	6.58	-	6.93	5.68	5.08	6.42	0.67	0.44
Met	6.43	-	5.82	5.59	6.18	-	2.38	0.48
Phe	6.72	-	7.28	6.72	6.33	-	0.15	0.55
Pro	5.92	-	5.50	5.00	5.50	-	0.32	0.72
Ser	5.66	-	5.70	5.06	6.00	-	0.92	0.82
Thr	5.99	-	6.01	4.97	5.29	-	0.57	0.65
Trp	7.25	-	7.56	7.25	7.71	-	0.15	0.60
Tyr	6.84	7.24	7.07	6.84	6.28	-	0.15	0.53
Val	6.14	-	7.04	5.95	6.59	-	0.33	0.50

S3.2. Activity coefficient equations

1) Hard-sphere repulsion

The contribution to the activity coefficient due to the hard-sphere repulsion is represented by the Mansoori-Carnahan-Starling-Leland equation[45]

$$\ln \gamma_i^{hs} = A + \frac{\sigma_i}{2} B + \pi \sigma_i^2 C + \pi \frac{\sigma_i^3}{6} D \tag{S3.1}$$

where the coefficients are:

$$A = -\ln\left(1 - n_3\right) \tag{S3.2}$$

$$B = \frac{n_2}{1 - n_3} \tag{S3.3}$$

$$C = \frac{n_1}{1 - n_3} + \frac{n_2^2}{12\pi n_3}\left[\frac{\ln\left(1 - n_3\right)}{n_3} + \frac{1}{\left(1 - n_3\right)^2}\right] \tag{S3.4}$$

$$D = \frac{n_0}{1 - n_3} + \frac{n_1 n_2}{\left(1 - n_3\right)^2} - \frac{n_2^3}{36\pi n_3^3}\left[2\ln(1 - n_3) + \frac{n_3}{\left(1 - n_3\right)} + \frac{n_3\left(1 - 3n_3\right)}{\left(1 - n_3\right)^3}\right] \tag{S3.5}$$

and the weighted densities are given as

$$n_0 = \sum_j \rho_j \tag{S3.6}$$

$$n_1 = \frac{1}{2}\sum_j \rho_j \sigma_j \tag{S3.7}$$

$$n_2 = \pi \sum_j \rho_j \sigma_j^2 \tag{S3.8}$$

$$n_3 = \frac{\pi}{6}\sum_j \rho_j \sigma_j^3 \tag{S3.9}$$

where ρ_j and σ_j are the number density and size of the species, respectively.

2) Electrostatic correlation

The electrostatic correlations are accounted by using the mean spherical approximation (MSA)[19].

$$\ln \gamma_i^{el} = -l_B \left[\frac{Z_i^2 \Gamma}{1+\Gamma\sigma_i} + \eta\sigma_i \left(\frac{2Z_i - \eta\sigma_i^2}{1+\Gamma\sigma_i} + \frac{\eta\sigma_i^2}{3} \right) \right] \tag{S3.10}$$

where $l_B = \beta e^2 / 4\pi\varepsilon_0\varepsilon_r$ is the Bjerrum length (7.14 Å for water at room temperature), $\beta = 1/k_B T$ is the inverse of the Boltzmann factor, k_B, and the absolute temperature T, e is the elementary charge, and ε_0 and ε_r are the vacuum permittivity and the relative dielectric constant of the solution, respectively, Z_i is the valence of the species, and Γ is the MSA screening parameter which is given by

$$\Gamma^2 = \pi l_B \sum_j \rho_j \left(Z_j^{eff} \right)^2 \tag{S3.11}$$

$$Z_j^{eff} = \frac{Z_j - \eta\sigma_j^2}{1+\Gamma\sigma_j} \tag{S3.12}$$

$$\eta = \frac{1}{H(\Gamma)} \sum_j \rho_j \frac{Z_j \sigma_j}{1+\Gamma\sigma_j} \tag{S3.13}$$

$$H(\Gamma) = \sum_j \rho_j \frac{\sigma_j^3}{1+\Gamma\sigma_j} + \frac{2}{\pi} \left(1 - \frac{\pi}{6} \sum_j \rho_j \sigma_j^3 \right) \tag{S3.14}$$

3) Solvent-mediated interactions

The mean-field approximation for the square-well potential results in the following contribution to the activity coefficient:

$$\ln \gamma_i^{att} = -\frac{4\pi}{3} \sum_j \rho_j \varepsilon_{ij} \lambda_{ij}^3 \sigma_{ij}^3 \tag{S3.15}$$

where ε_{ij} is the interaction energy between the different species in solution, λ_{ij} is the square-well width multiplier which is fixed to a typical value of 1.2, and σ_{ij} is the distance between two species at contact. To reduce the complexity of the model, we have assumed that $\varepsilon_{ij}\lambda_{ij}\sigma_{ij}^3$ is a constant for amino acid in different states. As a result, this term makes no contribution to the chemical equilibrium between different charge states.

The activity coefficient of an amino acid is dependent on the dissociation of the amino acid into the different charge states and should be expressed as

$$\ln \gamma_{AA} = \sum_{i=\{states\}} \alpha_i \ln \gamma_{AA,i} \tag{S3.16}$$

To calculate the dissociation for each species, we must define the solution conditions. In general, we utilize a combination of a mass balance and charge balance to determine the concentration of each species. The mass balance for the amino acid is given as

$$c_A = \sum_{j=\{states\}} c_{A,j} \tag{S3.17}$$

and the charge balance is given as

$$[H^+] - [OH^-] + \sum_j c_{A,j} Z_j + \sum_k c_k Z_k = 0 \tag{S3.18}$$

where $[H^+]$ is the hydrogen concentration and $[OH^-]$ is the hydroxide concentration. The concentrations of protons and hydroxide ions are coupled with the water dissociation reaction

$$H_2O \rightleftharpoons H^+ + OH^- \tag{S3.19}$$

with the equilibrium constant being given as $pK_w = 13.997$ at 25 °C[24]. The last term on the left-hand size of Eq. S18 accounts for the contribution of other species such as added salt or counterions to the amino acid. We can define the dissociation into state *i* as

$$\alpha_i = \frac{c_{A,i}}{c_A} \tag{S3.20}$$

By combining the mass balance equation and the expression for the equilibrium constant, we can determine the dissociation for state *i* from the following expression:

$$\alpha_i = \frac{c_{H^+}^i \prod_{j=1}^{i} K_j'}{1 + \sum_{j=\{states\}} c_{H^+}^j \prod_{k=1}^{j} K_k'} \tag{S3.21}$$

If activity coefficients are neglected, then the dissociation into each state is known automatically given that the equilibrium constants are available. In general, this equation must be solved numerically by coupling with an expression for the activity coefficient.

S3.3. Mass balance and charge balance

There are three cases that need to be considered when solving the mass balance and charge balance. All are valid at any background salt concentration as they are immaterial to the charge balance equation since they satisfy charge neutrality.

1) Amino acid in pure water

$$[H^+]-[OH^-]+c_A\alpha_0\left[Z_{A,0}+\sum_{j>0}Z_j c_{H^+}^j\prod_{k=1}^{j}K_k'\right]=0 \qquad \text{(S3.22)}$$

For this case, the concentration of the amino acid determines the hydrogen concentration in the solution, and therefore the pH of the solution. It typically takes up to 100 mM of the amino acid before the isoelectric point is reach (see Figure S1).

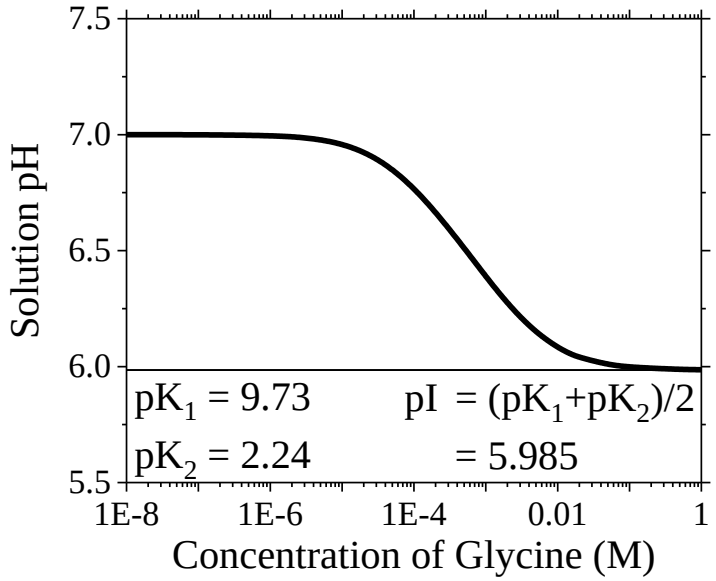

Figure S 3.1 The change in solution pH due to added glycine for an ideal solution. The isoelectric point, pI, is given as pI = (pK₁ + pK₂)/2.

2) Amino acid in pH adjusted water

$$[H^+]-[OH^-]+c_A\alpha_0\left[Z_{A,0}+\sum_{j>0}Z_j c_{H^+}^j\prod_{k=1}^{j}K_k'\right]=-c_cZ_c \qquad \text{(S3.23)}$$

In this case, the pH is fixed to a specific value so it is the counterion (e.g. chloride ion to the proton) concentration that must be determined.

3) Amino acid added as a salt (e.g. sodium glutamate)

$$[H^+]-[OH^-]+c_A\alpha_0\left[Z_{A,0}+\sum_{j>0}Z_j c_{H^+}^j\prod_{k=1}^{j}K_k'\right]=-c_AZ_c \qquad \text{(S3.24)}$$

The concentration of the counterion (right hand side of Eq. S24) to the amino acid is the same as the amino acid concentration and therefore, the unknown is the hydrogen ion concentration. Lastly, if the amino acid is added as a salt, then the activity coefficient measured experimentally shows an equal contribution from both the amino acid and the salt ion:

$$\gamma_{Salt-AA}=\sqrt{\gamma_{AA}\gamma_{Salt\,ion}} \qquad \text{(S3.25)}$$

With the outlined theoretical method, we are able to solve for any system containing dissolved amino acid in aqueous system. This provides the theoretical foundation to compare directly to experimental data for activity coefficients, solubility, and apparent equilibrium constants in pure water and aqueous systems.

S3.4. Comparison with experimental data

Experimental data is often presented in the molality scale instead of the concentration used in our model. The molality scale describes the solution by the weight of the solvent instead of the volume of solution. While the differences are negligible at dilute conditions, the concentration can be significantly smaller than the molality when the molality of the solute is large. The conversion from molality to concentration units is done by the standard expression:

$$m_i = \frac{c_i}{d_L - 0.001\sum_i^n M_i c_i} \tag{S3.26}$$

where M_i is the molecular weight of the species i, m_i is the molality in *mol/kg* and c_i is the concentration in *mol/L*. This requires knowledge of the density of solution which can typically be described empirically. The density of solution is expressed as

$$d_L = d_w + f\left(c_s\right) + g\left(m_{AA}\right) \tag{S3.27}$$

where d_w is the density of pure water and the functions f and g describe the effect of salt concentration and amino acid molality on the density of the solution. The activity coefficient can be converted from the molality scale to the molarity scale by

$$\gamma_i^{(c)} = \frac{m_i d_w}{c_i}\gamma_i^{(m)} \tag{S3.28}$$

Lastly, the apparent equilibrium constants can be converted by

$$K'^{(c)} = K'^{(m)} / \left(d_L - c_s M_s\right) \tag{S3.29}$$

With these expressions, we can convert between the commonly used molality scale in experiments and the concentration scale to fit the parameters within our model.

The experimental data for the activity coefficient [33-36, 40, 46, 50-53, 58-60], solubility [6, 12, 61-66], and apparent equilibrium constant [3, 6, 12, 16, 24-27, 67-75] can be found in the references.

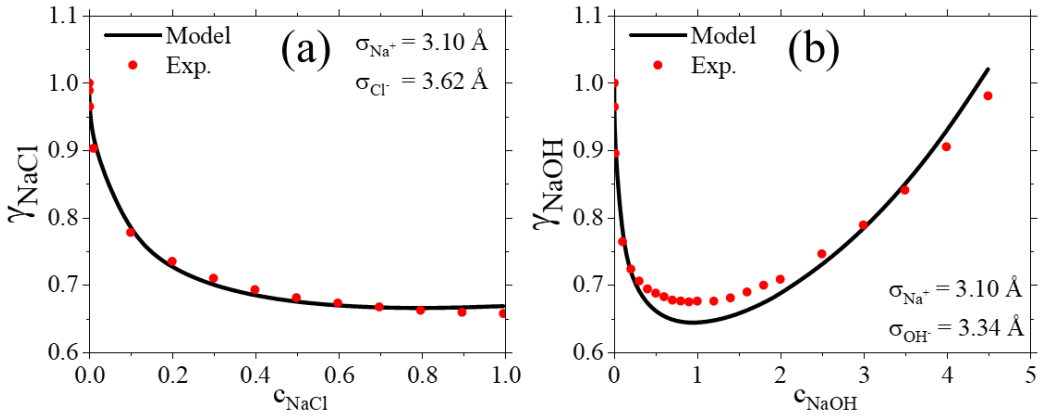

Figure S 3.2 Comparison of predicted activity coefficient for (a) sodium chloride and (b) sodium hydroxide compared to the respective experimentally determined activity coefficient.

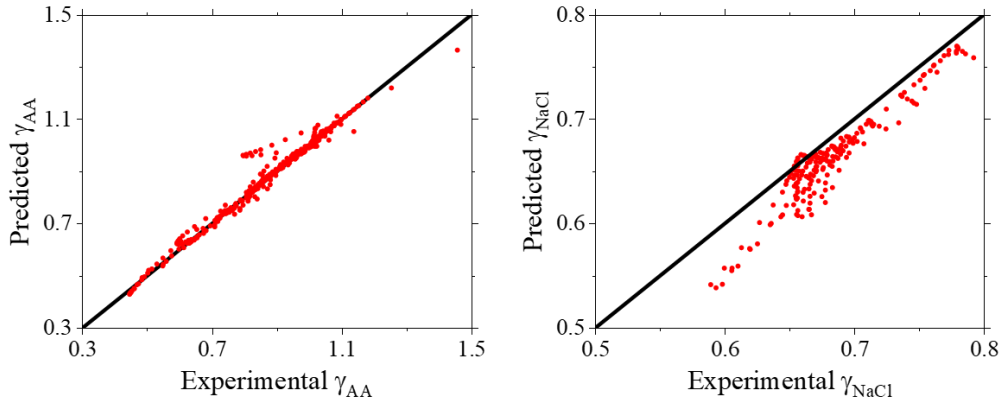

Figure S 3.3 Comparison of predicted activity coefficient for (a) amino acids and (b) sodium chloride compared to the respective experimentally determined activity coefficient.

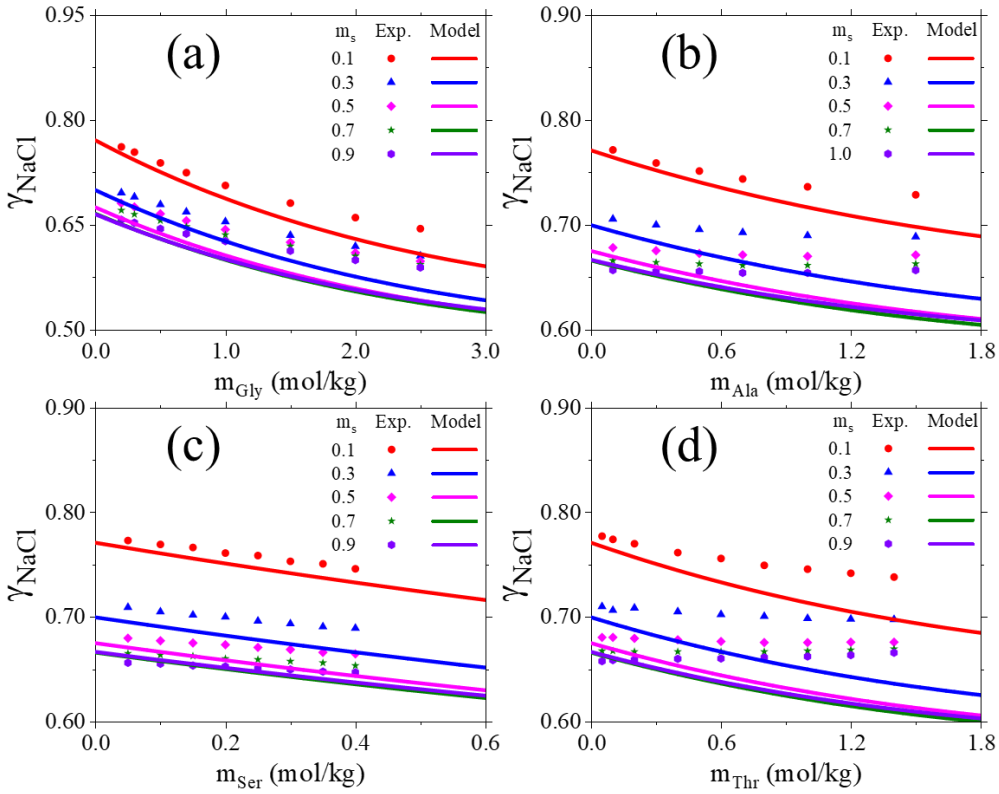

Figure S 3.4 Activity coefficient of sodium chloride in amino acid solutions of (a) glycine, (b) alanine, (c) serine and (d) threonine at different amino acid and sodium chloride molalities by experiment (symbols) and model (lines).

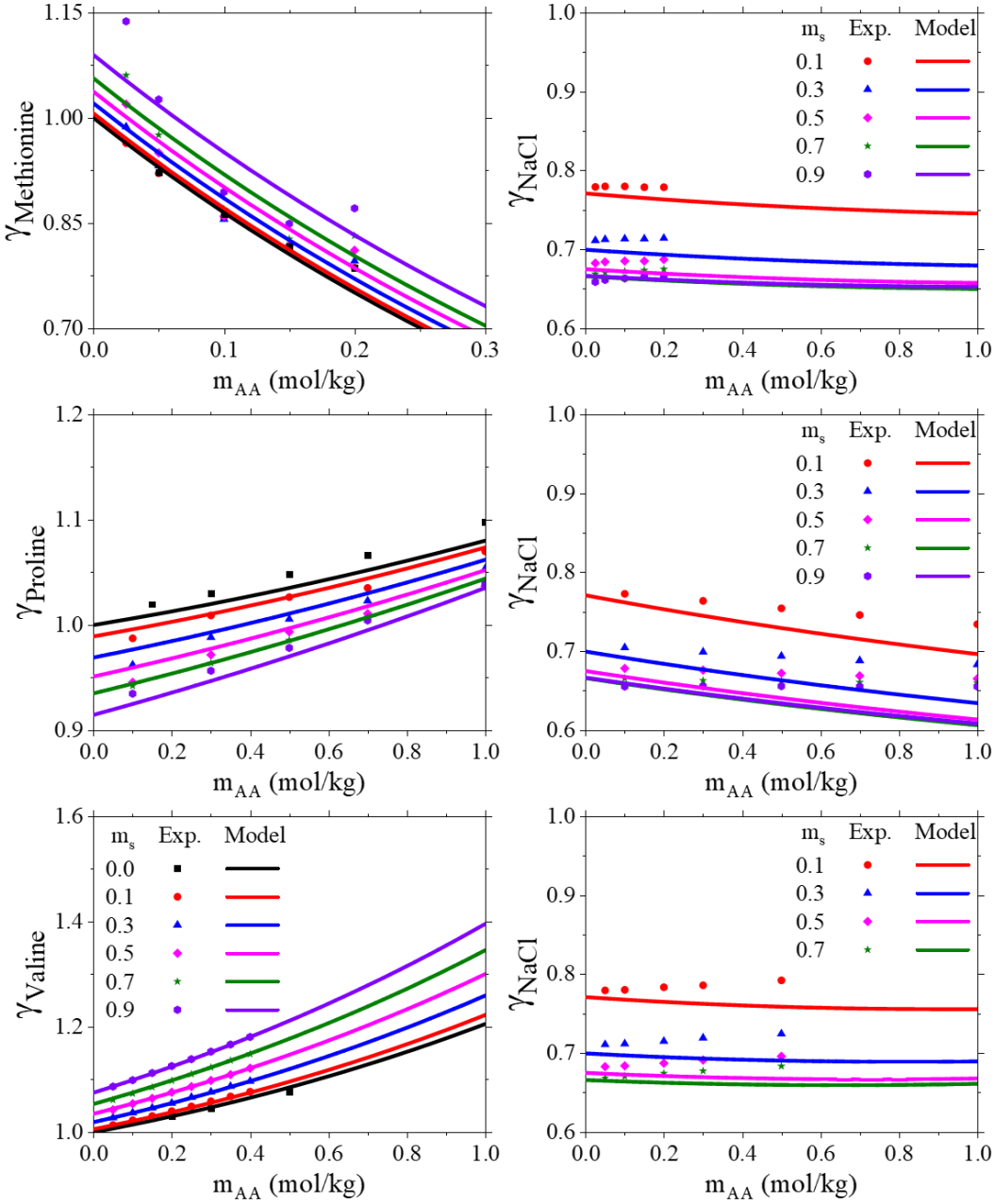

Figure S 3.5 Activity coefficient of the amino acid and sodium chloride in amino acid solutions of methionine, proline, and valine at different amino acid and sodium chloride molalities by experiment (symbols) and model (lines).

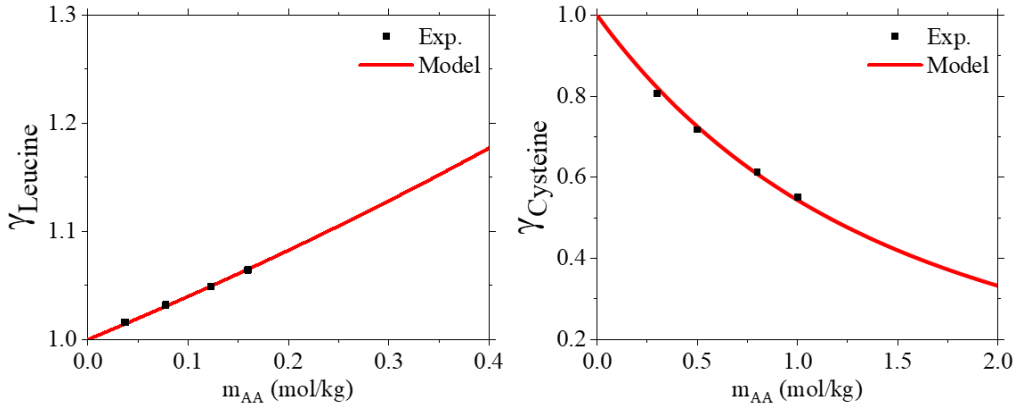

Figure S 3.6 Activity coefficient of leucine and cysteine in pure water determined by experiment (symbols) and model (lines).

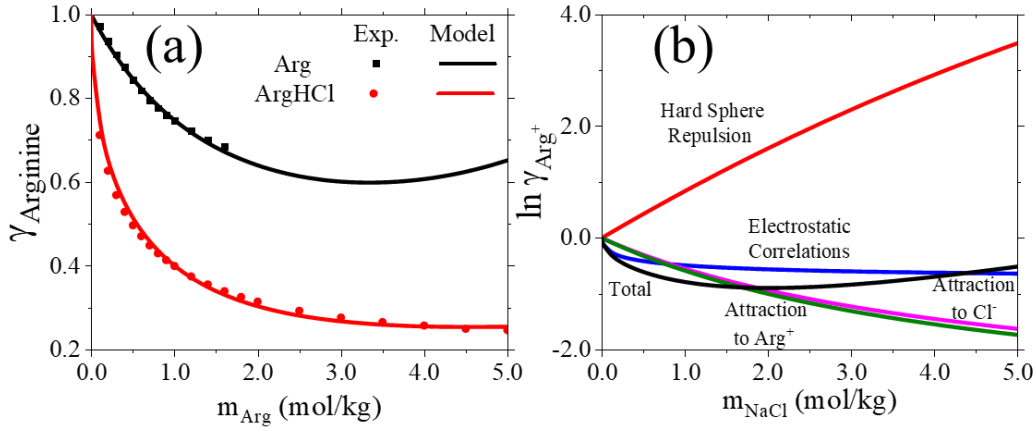

Figure S 3.7 (a) Experimental activity coefficient of arginine in pure water (squares) and arginine hydrochloride (circles) and the model (lines). (b) Contribution of each component to the activity coefficient of arginine in its monovalent charged state.

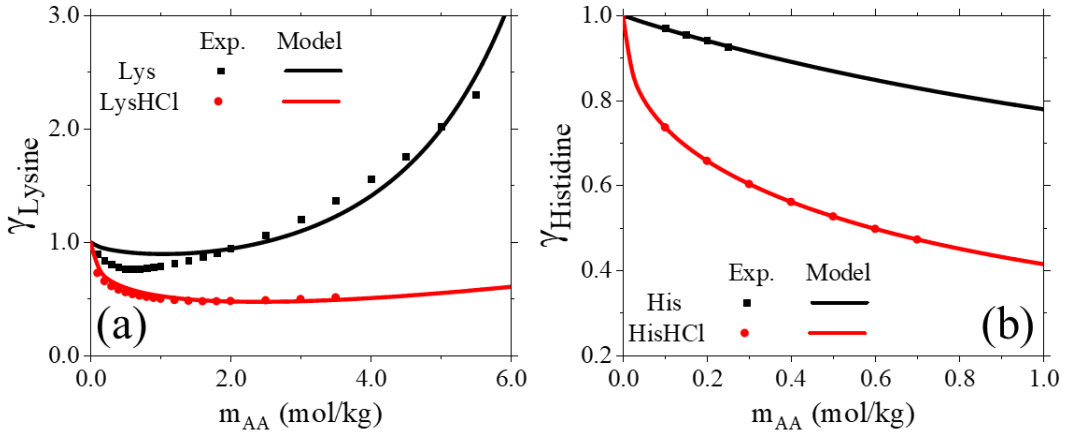

Figure S 3.8 (a) Experimental activity coefficient of lysine in pure water (squares) and lysine hydrochloride (circles) and the model (lines). (b) Experimental activity coefficient of histdine in pure water (squares) and histidine hydrochloride (circles) and the model (lines).

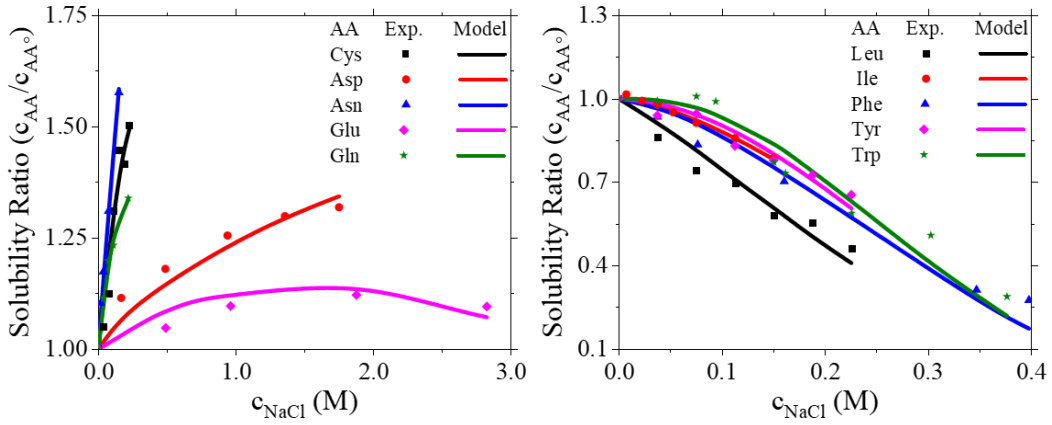

Figure S 3.9 Solubility ratio of amino acids in sodium chloride. (a) Salting in of amino acids and (b) salting out of amino acids.

S3.5. Apparent Equilibrium Constants

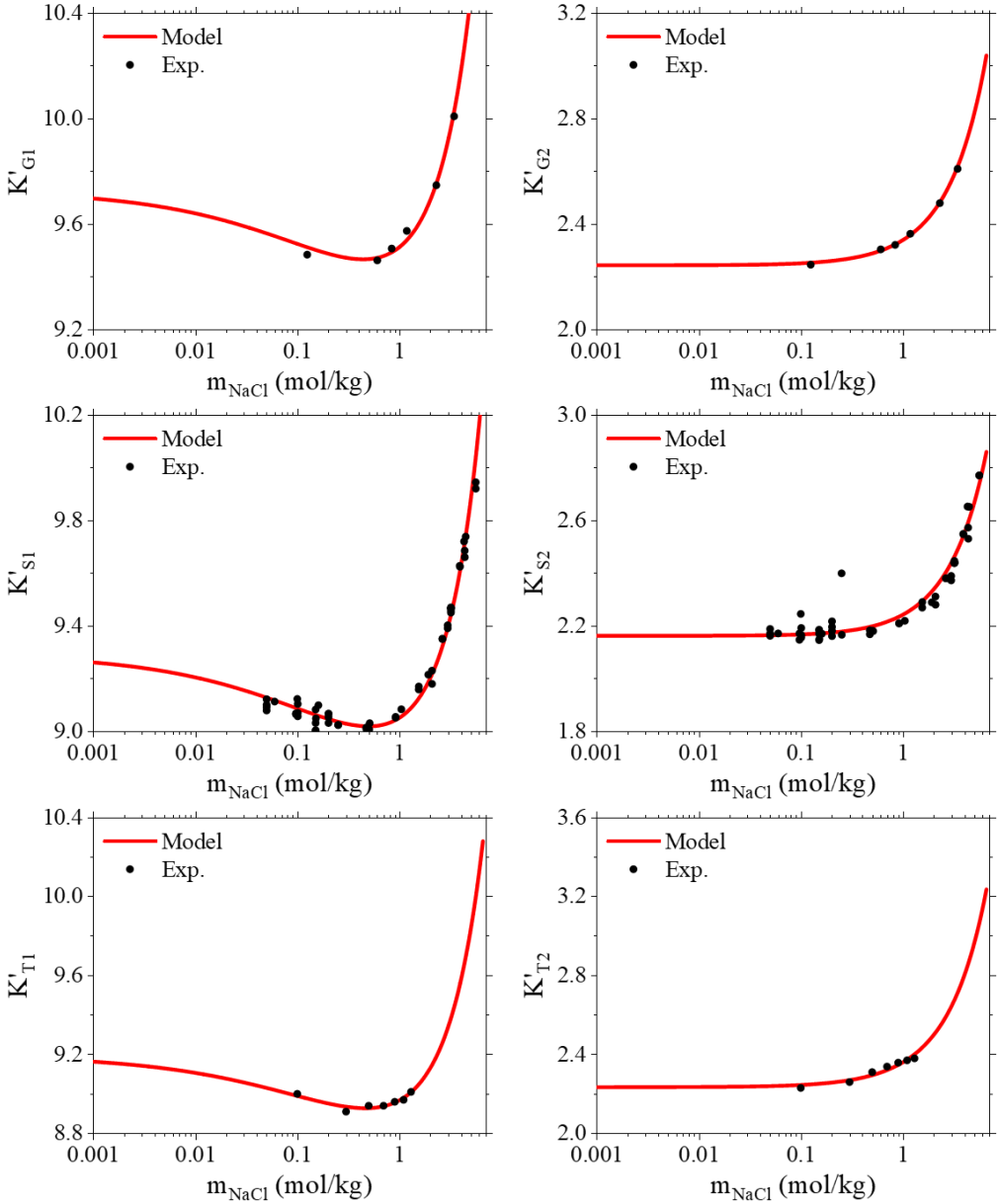

Figure S 3.10 Model comparison to apparent equilibrium constant data for glycine (G), serine (S), and threonine (T).

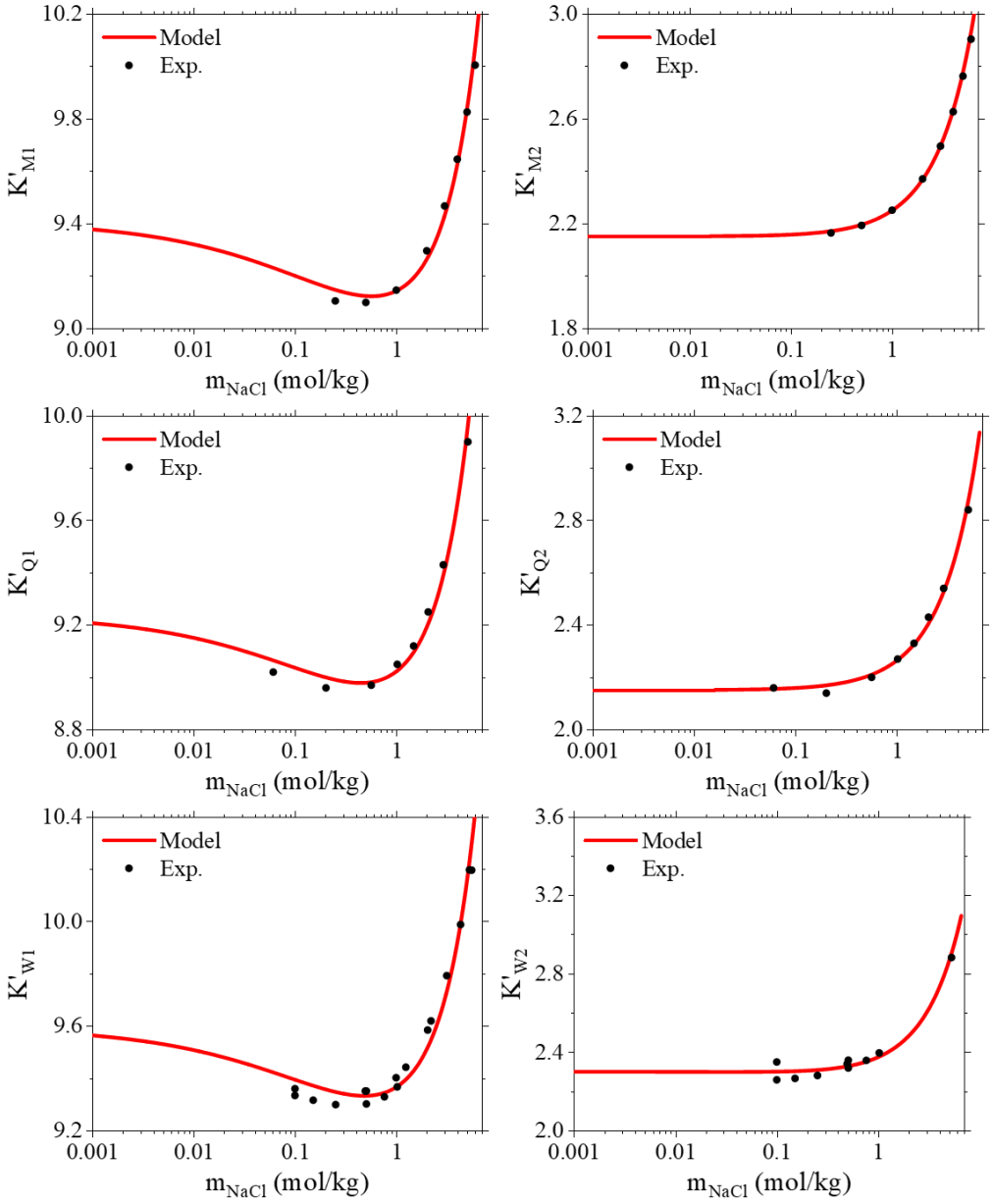

Figure S 3.11 Model comparison to apparent equilibrium constant data for methionine (M), glutamine (Q), and tryptophan (W).

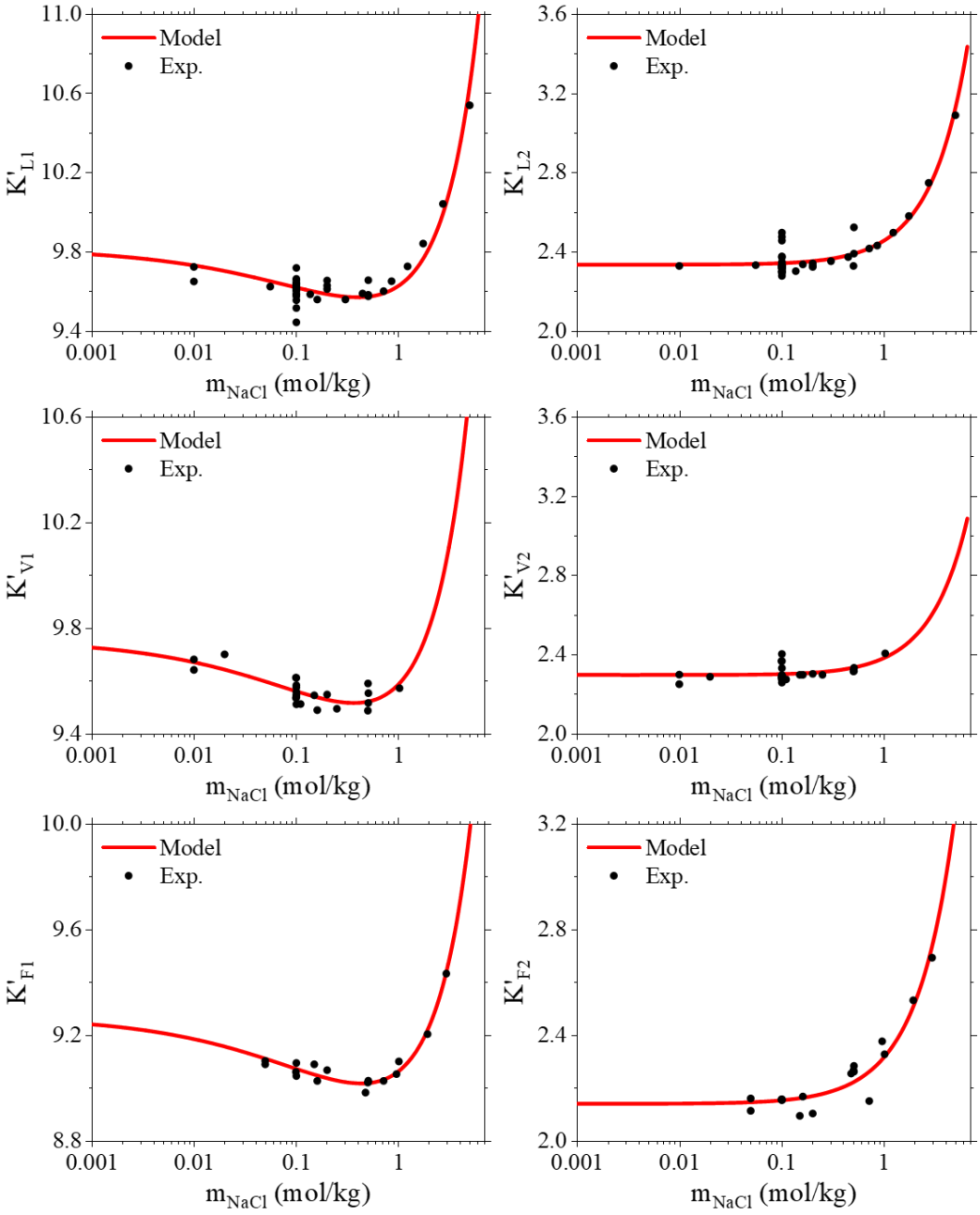

Figure S 3.12 Model comparison to apparent equilibrium constant data for leucine (L), valine (V), and phenylalanine (F).

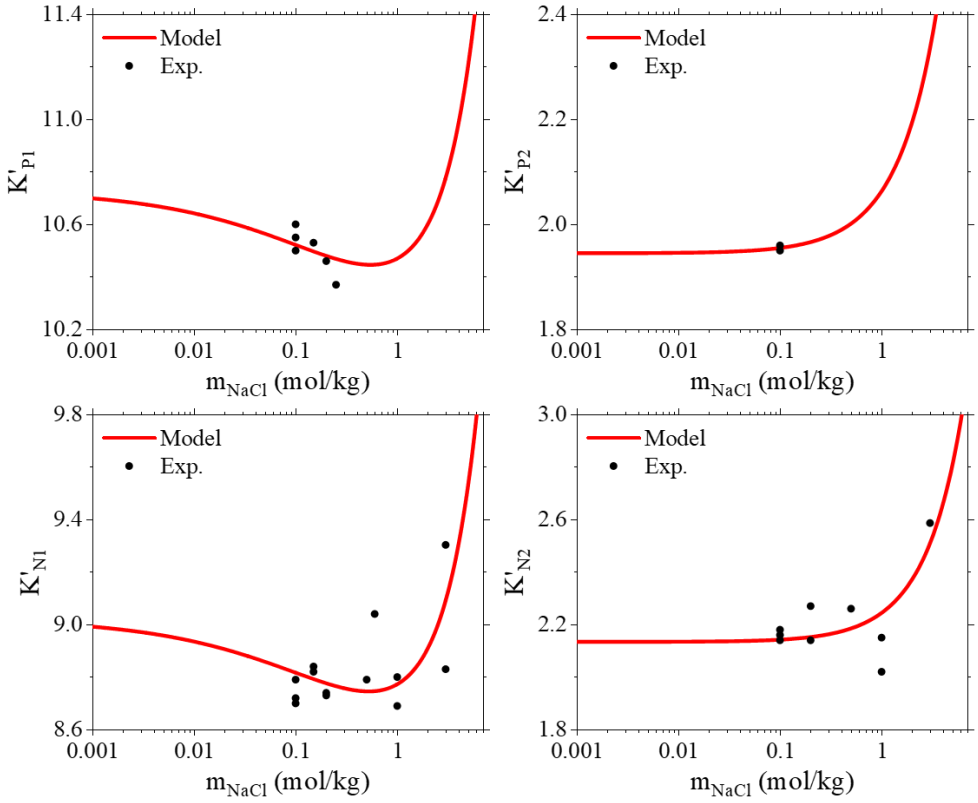

Figure S 3.13 Model comparison to apparent equilibrium constant data for proline (P) and asparagine (N).

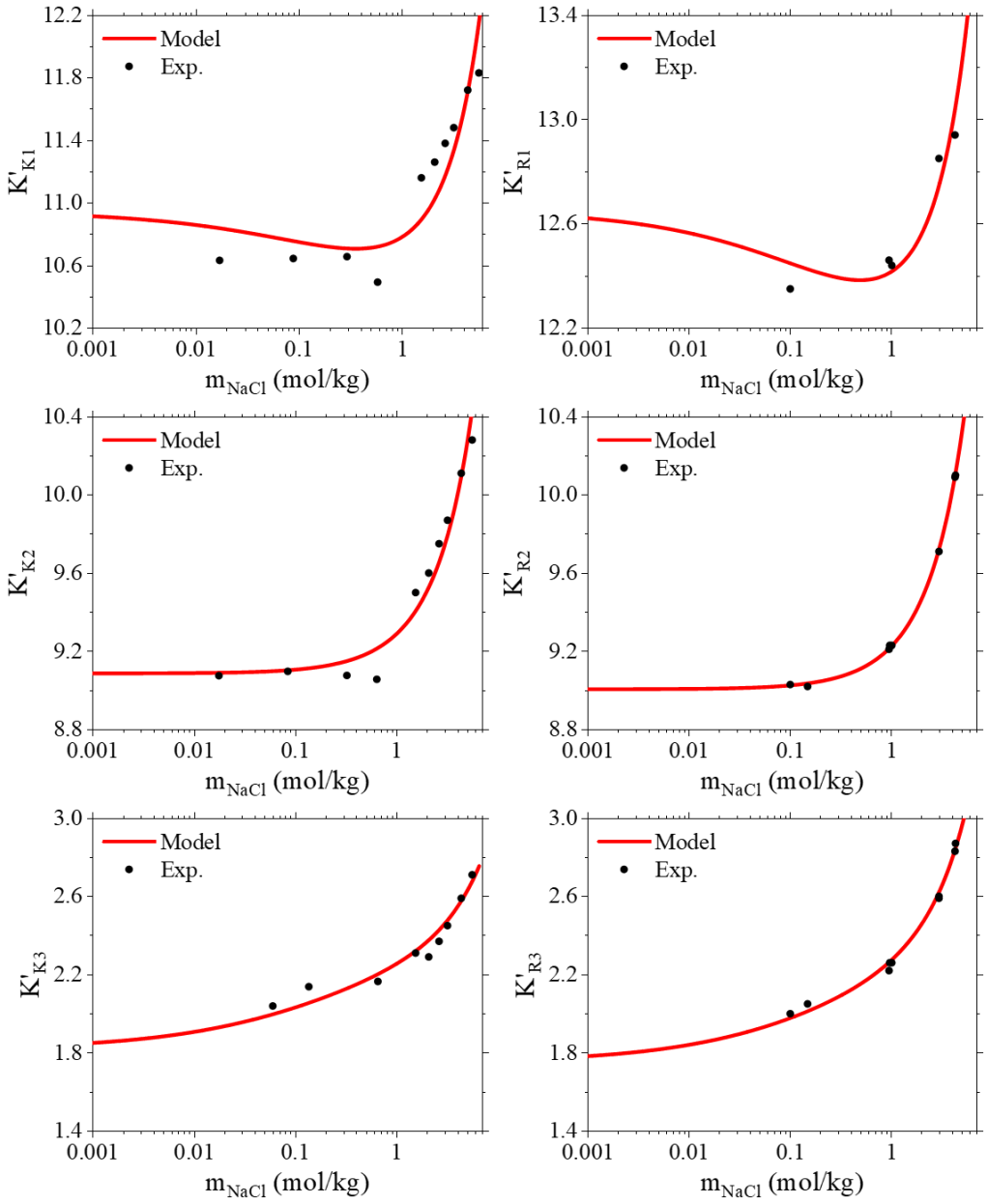

Figure S 3.14 Model comparison to apparent equilibrium constant data for lysine and arginine.

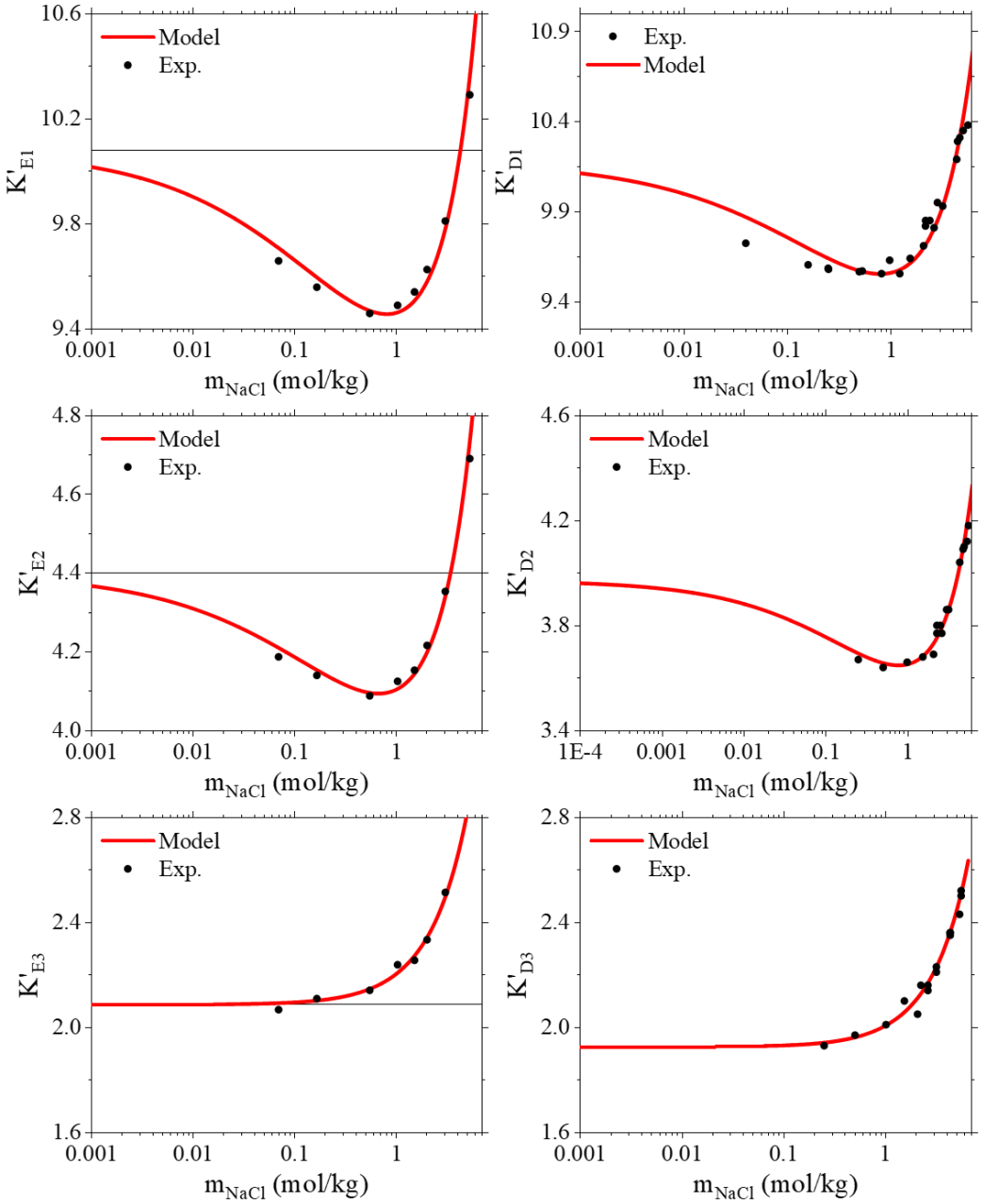

Figure S 3.15 Model comparison to apparent equilibrium constant data for glutamic acid and aspartic acid.

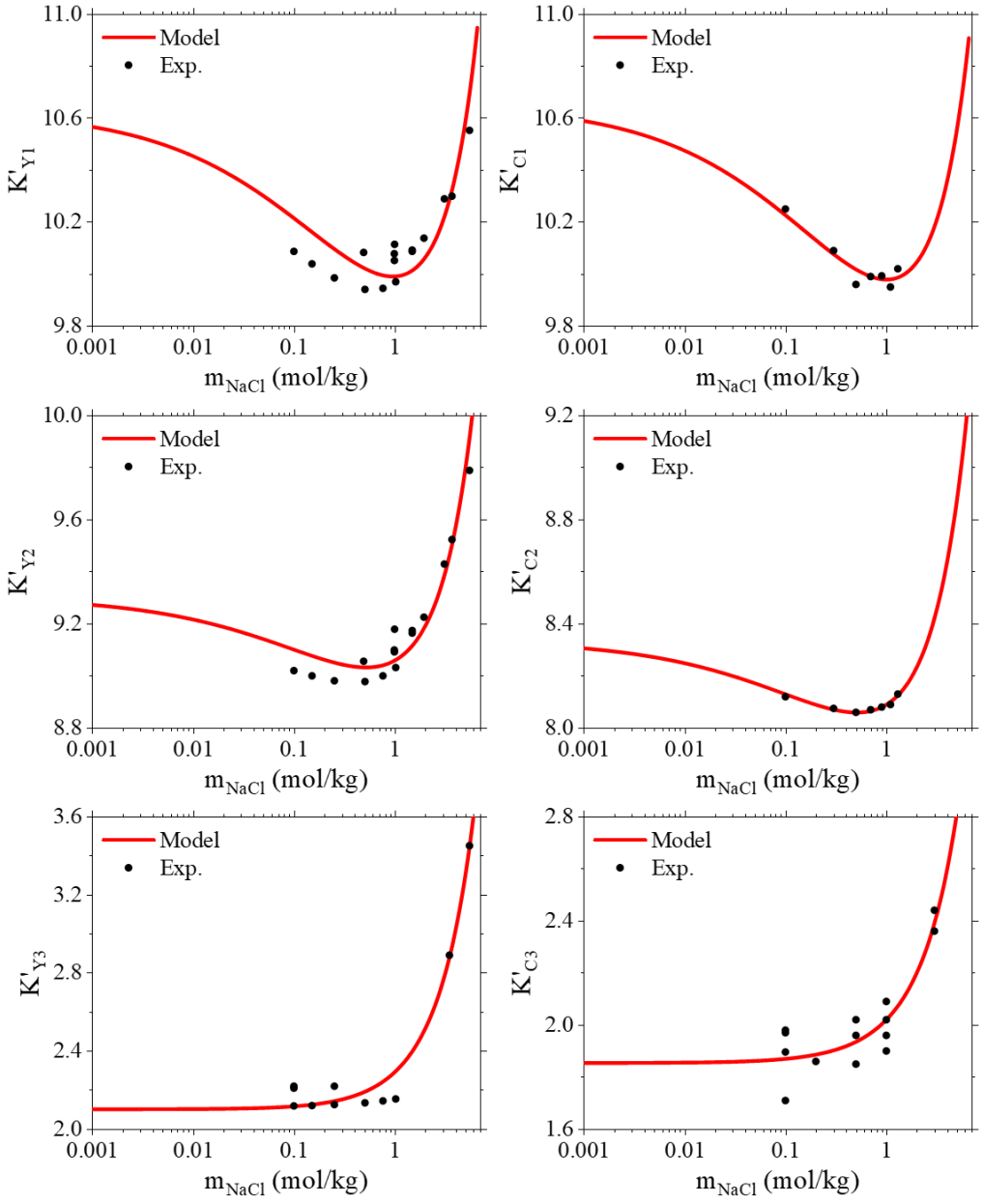

Figure S 3.16 Model comparison to apparent equilibrium constant data for tyrosine and cysteine.

Table S 3.2 Comparison of different theoretical methods for the predicted thermodynamic equilibrium constants of some neutral amino acids

Model	Glycine		Alanine		Serine	
	$\log K_1^T$	$\log K_2^T$	$\log K_1^T$	$\log K_2^T$	$\log K_1^T$	$\log K_2^T$
MSA	9.73	2.24	9.99	2.36	9.29	2.16
Pitzer	9.78	2.35	9.86	2.35	9.21	2.19
SIT	9.78	2.33	9.91	2.36	9.26	2.16
Other	9.85	2.54	-	-	9.28	2.18

Table S 3.3 Comparison of different theoretical methods for the predicted thermodynamic equilibrium constants of acidic amino acids

Model	Aspartic Acid			Glutamic Acid			Tyrosine		
	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$
MSA	10.17	3.97	1.93	10.08	4.40	2.09	10.63	9.30	2.10
Pitzer	10.01	3.88	1.92	9.98	4.39	2.19	-	-	-
SIT	10.09	3.93	1.92	10.06	4.39	2.17	10.47	9.21	2.16
Other	-	-	-	9.69	4.18	2.32	-	-	-

Table S 3.4 Comparison of different theoretical methods for the predicted thermodynamic equilibrium constants of basic amino acids

Model	Arginine			Histidine			Lysine		
	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$	$\log K_1^T$	$\log K_2^T$	$\log K_3^T$
MSA	12.65	9.01	1.75	9.38	6.08	1.52	10.95	9.09	1.82
Pitzer	-	-	-	9.28	5.97	1.59	10.80	9.09	1.77
SIT	12.63	9.01	1.77	9.30	6.03	1.53	-	-	-
Other	11.94	9.05	1.97	-	-	-	11.10	9.47	2.27

Bibliography

1. Gitlin, I., J.D. Carbeck, and G.M. Whitesides, *Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis.* Angewandte Chemie-International Edition, 2006. **45**(19): p. 3022-3060.

2. Atkins, P.W., J. De Paula, and J. Keeler, *Atkins' physical chemistry molecular thermodynamics and kinetics*. Eleventh edition. ed. 2019, Oxford, United Kingdom ; New York, NY: Oxford University Press. xix, 445 pages.

3. Bretti, C., et al., *Modeling solubility and acid-base properties of some amino acids in aqueous NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures.* SpringerPlus, 2016. **5**(1): p. 928-928.

4. Berthon, G., *Critical evaluation of the stability constants of metal complexes of amino acids with polar side chains (Technical Report).* Pure and Applied Chemistry, 1995. **67**(7): p. 1117-1240.

5. De Stefano, C., et al., *Chemical speciation of amino acids in electrolyte solutions containing major components of natural fluids.* Chemical Speciation & Bioavailability, 1995. **7**(1): p. 1-8.

6. Bretti, C., et al., *Some thermodynamic properties of dl-Tyrosine and dl-Tryptophan. Effect of the ionic medium, ionic strength and temperature on the solubility and acid–base properties.* Fluid Phase Equilibria, 2012. **314**: p. 185-197.

7. Gómez-Bombarelli, R., et al., *Computational Calculation of Equilibrium Constants: Addition to Carbonyl Compounds.* The Journal of Physical Chemistry A, 2009. **113**(42): p. 11423-11428.

8. Bickmore, B.R. and M.C.F. Wander, *Activity and Activity Coefficients*, in *Encyclopedia of Geochemistry: A Comprehensive Reference Source on the Chemistry of the Earth*, W.M. White, Editor. 2018, Springer International Publishing: Cham. p. 21-23.

9. Elizalde, M.P. and J.L. Aparicio, *Current theories in the calculation of activity coefficients—II. Specific interaction theories applied to some equilibria studies in solution chemistry.* Talanta, 1995. **42**(3): p. 395-400.

10. Guggenheim, E.A. and J.C. Turgeon, *Specific interaction of ions.* Transactions of the Faraday Society, 1955. **51**(0): p. 747-761.

11. Bretti, C., C. Foti, and S. Sammartano, *A new approach in the use of SIT in determining the dependence on ionic strength of activity coefficients. Application to some chloride salts of interest in the speciation of natural fluids.* Chemical Speciation & Bioavailability, 2004. **16**(3): p. 105-110.

12. Bretti, C., et al., *Modeling solubility and acid-base properties of some polar side chain amino acids in NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures.* Fluid Phase Equilibria, 2018. **459**: p. 51-64.

13. Bretti, C., et al., *Modeling solubility and acid–base properties of some amino acids in aqueous NaCl and (CH3)4NCl aqueous solutions at different ionic strengths and temperatures.* SpringerPlus, 2016. **5**(1): p. 928.

14. Pitzer, K.S., *Activity Coefficients in Electrolyte Solutions*. 2018.

15. Kim, H.T. and W.J. Frederick, *Evaluation of Pitzer ion interaction parameters of aqueous electrolytes at 25.degree.C. 1. Single salt parameters.* Journal of Chemical & Engineering Data, 1988. **33**(2): p. 177-184.

16. De Stefano, C., et al., *The interaction of amino acids with the major constituents of natural waters at different ionic strengths.* Marine Chemistry, 2000. **72**(1): p. 61-76.

17. Lee, L.L., *Molecular thermodynamics of electrolyte solutions*. 2008, Singapore ; Hackensack, NJ: World Scientific. xii, 251 p.

18. Hossain, N., et al., *Revisiting electrolyte thermodynamic models: Insights from molecular simulations.* Aiche Journal, 2018. **64**(10): p. 3728-3734.

19. Blum, L., *Mean spherical model for asymmetric electrolytes.* Molecular Physics, 1975. **30**(5): p. 1529-1535.

20. Simonin, J.-P., L. Blum, and P. Turq, *Real Ionic Solutions in the Mean Spherical Approximation. 1. Simple Salts in the Primitive Model.* The Journal of Physical Chemistry, 1996. **100**(18): p. 7704-7709.

21. Simonin, J.-P. and O. Bernard, *Organic electrolyte solutions: Modeling of deviations from ideality within the binding mean spherical approximation.* Fluid Phase Equilibria, 2018. **468**: p. 58-69.

22. Maribo-Mogensen, B., G.M. Kontogeorgis, and K. Thomsen, *Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions.* Industrial & Engineering Chemistry Research, 2012. **51**(14): p. 5353-5363.

23. Bernard, O. and L. Blum, *Binding mean spherical approximation for pairing ions: An exponential approximation and thermodynamics.* The Journal of Chemical Physics, 1996. **104**(12): p. 4746-4754.

24. Vilarino, T. and M.E.S. de Vicente, *Theoretical calculations of the ionic strength dependence of the ionic product of water based on a mean spherical approximation.* Journal of Solution Chemistry, 1997. **26**(9): p. 833-846.

25. Vilariño, T. and M.E. Sastre de Vicente, *Protonation of Glycine in Saline Media: Evaluation of the Effect of Ionic Strength by Use of the Mean Spherical Approximation.* The Journal of Physical Chemistry, 1996. **100**(40): p. 16378-16384.

26. Vilariño, T., J.L. Barriada, and M.E. Sastre de Vicente, *The mean spherical approximation and the prediction of the size of the species involved in an ionization equilibrium in saline media.* Physical Chemistry Chemical Physics, 2001. **3**(6): p. 1053-1056.

27. Vilariño, T., et al., *Effect of ionic strength on the protonation of various amino acids analysed by the mean spherical approximation.* Journal of the Chemical Society, Faraday Transactions, 1997. **93**(3): p. 413-417.

28. Triolo, R., J.R. Grigera, and L. Blum, *Simple electrolytes in the mean spherical approximation.* The Journal of Physical Chemistry, 1976. **80**(17): p. 1858-1861.

29. Waluyo, I., et al., *The structure of water in the hydration shell of cations from x-ray Raman and small angle x-ray scattering measurements.* The Journal of chemical physics, 2011. **134**(6): p. 064513-064513.

30. Nörtemann, K., J. Hilland, and U. Kaatze, *Dielectric Properties of Aqueous NaCl Solutions at Microwave Frequencies.* The Journal of Physical Chemistry A, 1997. **101**(37): p. 6864-6869.

31. Lu, J.-F., Y.-X. Yu, and Y.-G. Li, *Modification and application of the mean spherical approximation method.* Fluid Phase Equilibria, 1993. **85**: p. 81-100.

32. Gavish, N. and K. Promislow, *Dependence of the dielectric constant of electrolyte solutions on ionic concentration: A microfield approach.* Physical Review E, 2016. **94**(1): p. 012611.

33. Soto-Campos, A.M., M.K. Khoshkbarchi, and J.H. Vera, *Activity coefficients of the electrolyte and the amino acid in water + NaNO3 + glycine and water + NaCl + dl-methionine systems at 298.15 K.* Biophysical Chemistry, 1997. **67**(1): p. 97-105.

34. Khoshkbarchi, M.K. and J.H. Vera, *A Perturbed Hard-Sphere Model with Mean Spherical Approximation for the Activity Coefficients of Amino Acids in Aqueous Electrolyte Solutions.* Industrial & Engineering Chemistry Research, 1996. **35**(12): p. 4755-4766.

35. Khoshkbarchi, M.K. and J.H. Vera, *A Simplified Perturbed Hard-Sphere Model for the Activity Coefficients of Amino Acids and Peptides in Aqueous Solutions.* Industrial & Engineering Chemistry Research, 1996. **35**(11): p. 4319-4327.

36. Held, C., et al., *Measuring and modeling aqueous electrolyte/amino-acid solutions with ePC-SAFT.* The Journal of Chemical Thermodynamics, 2014. **68**: p. 1-12.

37. Breil, M.P. and J. Mollerup. *Thermodynamics, Experimental, and Modelling of Aqueous Electrolyte and Amino Acid Solutions*. 2001.

38. Theiss, M. and J. Gross, *Nonprimitive Model Electrolyte Solutions: Comprehensive Data from Monte Carlo Simulations.* Journal of Chemical and Engineering Data, 2020. **65**(2): p. 634-639.

39. Smith, W.R., et al., *Recent progress in the molecular simulation of thermodynamic properties of aqueous electrolyte solutions.* Fluid Phase Equilibria, 2018. **466**: p. 19-30.

40. Held, C., L.F. Cameretti, and G. Sadowski, *Measuring and Modeling Activity Coefficients in Aqueous Amino-Acid Solutions.* Industrial & Engineering Chemistry Research, 2011. **50**(1): p. 131-141.

41. Bang, C.-H., H.-K. Choi, and B.-S. Lee, *Modeling of activity coefficients of amino acid and electrolyte in aqueous solutions.* Journal of Molecular Liquids, 2016. **223**: p. 1-9.

42. Conway, O., et al., *Development of transferable coarse-grained models of amino acids.* Molecular Systems Design & Engineering, 2020. **5**(3): p. 675-685.

43. Dasetty, S., J.K. Barrows, and S. Sarupria, *Adsorption of amino acids on graphene: assessment of current force fields.* Soft Matter, 2019. **15**(11): p. 2359-2372.

44. Jin, Z., Y. Tang, and J. Wu, *A perturbative density functional theory for square-well fluids.* The Journal of Chemical Physics, 2011. **134**(17): p. 174702.

45. Mansoori, G.A., et al., *Equilibrium Thermodynamic Properties of the Mixture of Hard Spheres.* The Journal of Chemical Physics, 1971. **54**(4): p. 1523-1525.

46. Robinson, R.A. and R.H. Stokes, *Tables of osmotic and activity coefficients of electrolytes in aqueous solution at 25° C.* Transactions of the Faraday Society, 1949. **45**(0): p. 612-624.

47. Zhao, Y.H., M.H. Abraham, and A.M. Zissimos, *Fast Calculation of van der Waals Volume as a Sum of Atomic and Bond Contributions and Its Application to Drug Compounds.* The Journal of Organic Chemistry, 2003. **68**(19): p. 7368-7373.

48. Edward, J.T. and P.G. Farrell, *Relation between van der Waals and Partial Molal Volumes of Organic Molecules in Water.* Canadian Journal of Chemistry, 1975. **53**(19): p. 2965-2970.

49. Eisenberg, D., et al., *Hydrophobic moments and protein structure.* Faraday Symposia of the Chemical Society, 1982. **17**(0): p. 109-120.

50. Lee, B.S. and K.C. Kim, *Measurements and Modeling of the Activity Coefficients and Solubilities of L-alanine in Aqueous Electrolyte Solutions.* Korean Chemical Engineering Research, 2010. **48**(4): p. 519-533.

51. Khoshkbarchi, M.K., A.M. Soto-Campos, and J.H. Vera, *Interactions of DL-serine and L-serine with NaCl and KCl in aqueous solutions.* Journal of Solution Chemistry, 1997. **26**(10): p. 941-955.

52. Soto-Campos, A.M., M.K. Khoshkbarchi, and J.H. Vera, *Interaction of DL-threonine with NaCl and NaNO3 in aqueous solutions: e.m.f. measurements with ion-selective electrodes.* The Journal of Chemical Thermodynamics, 1997. **29**(5): p. 609-622.

53. Bonner, O.D., *Osmotic and activity coefficients of sodium and potassium glutamate at 298.15 K.* Journal of Chemical & Engineering Data, 1981. **26**(2): p. 147-148.

54. Wyman, J. and T.L. McMeekin, *The Dielectric Constant of Solutions of Amino Acids and Peptides.* Journal of the American Chemical Society, 1933. **55**(3): p. 908-914.

55. Rao, M.V.R., M. Atreyi, and M.R. Rajeswari, *Estimation of partial molar volumes of α-amino acids in water.* Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 1984. **80**(8): p. 2027-2032.

56. Rao, M.V.R., M. Atreyi, and M.R. Rajeswari, *Partial molar volumes of .alpha.-amino acids with ionogenic side chains in water.* The Journal of Physical Chemistry, 1984. **88**(14): p. 3129-3131.

57. Mishra, A.K. and J.C. Ahluwalia, *Apparent molal volumes of amino acids, N-acetylamino acids, and peptides in aqueous solutions.* The Journal of Physical Chemistry, 1984. **88**(1): p. 86-92.

58. Bong-Seob, L. and K. Ki-Chang, *Modeling and Measurements of the activity coefficients and solubilities of amino acids in the L-valine/electrolyte and L-proline/electrolyte aqueous solutions.* 2012. **50**.

59. Tsurko, E.N., R. Neueder, and W. Kunz, *Water Activity and Osmotic Coefficients in Solutions of Glycine, Glutamic Acid, Histidine and their Salts at 298.15 K and 310.15 K.* Journal of Solution Chemistry, 2007. **36**(5): p. 651-672.

60. Bonner, O.D., *Osmotic and activity coefficients of some amino acids and their hydrochloride salts at 298.15 K.* Journal of Chemical & Engineering Data, 1982. **27**(4): p. 422-423.

61. Carta, R. and G. Tola, *Solubilities of l-Cystine, l-Tyrosine, l-Leucine, and Glycine in Aqueous Solutions at Various pHs and NaCl Concentrations.* Journal of Chemical & Engineering Data, 1996. **41**(3): p. 414-417.

62. Ferreira, L.A., E.A. Macedo, and S.P. Pinho, *KCl effect on the solubility of five different amino acids in water.* Fluid Phase Equilibria, 2007. **255**(2): p. 131-137.

63. Bretti, C., et al., *Solubility and Activity Coefficients of Acidic and Basic Nonelectrolytes in Aqueous Salt Solutions. 1. Solubility and Activity Coefficients of o-Phthalic Acid and l-Cystine in NaCl(aq), (CH3)4NCl(aq), and (C2H5)4NI(aq) at Different Ionic Strengths and at t = 25 °C.* Journal of Chemical & Engineering Data, 2005. **50**(5): p. 1761-1767.

64. Lee, C.-Y., et al., *Effect of pH on the solubilities of divalent and trivalent amino acids in water at 298.15K.* Fluid Phase Equilibria, 2013. **343**: p. 30-35.

65. Mondal, S., et al., *Study of the Solubility and Transfer Thermodynamics of d,l-Phenylalanine in Aqueous Sodium Chloride and d,l-Serine in Aqueous Sodium Nitrate Solutions.* Journal of Solution Chemistry, 2016. **45**(12): p. 1755-1772.

66. Wang, J., et al., *Solubility of d-Aspartic Acid and l-Aspartic Acid in Aqueous Salt Solutions from (293 to 343) K.* Journal of Chemical & Engineering Data, 2010. **55**(4): p. 1735-1738.

67. Rey, F., et al., *Influence of the ionic strength on the ionization of amino acids.* Journal of Chemical & Engineering Data, 1989. **34**(1): p. 35-37.

68. Bhushanavathi, P., et al., *Effect of Solvent on Protonation Equilibria of L-Proline and L-Valine in 1, 2-Propanediol-Water Mixtures.* E-Journal of Chemistry, 2012. **9**(2): p. 517-524.

69. Vilariño, T. and M. E. Sastre de Vicente, *The mean spherical approximation methodology applied to the acid–base equilibria of glycine in artificial seawater.* Physical Chemistry Chemical Physics, 1999. **1**(10): p. 2453-2456.

70. Gharib, F. and L.A. Dogaheh, *Ionic Strength Dependence of Formation Constants: Complexation of Cysteine with Molybdenum(VI).* Journal of Chemical & Engineering Data, 2003. **48**(4): p. 999-1003.

71. Hamborg, E.S., J.P.M. Niederer, and G.F. Versteeg, *Dissociation Constants and Thermodynamic Properties of Amino Acids Used in CO2 Absorption from (293 to 353) K.* Journal of Chemical & Engineering Data, 2007. **52**(6): p. 2491-2502.

72. Nagai, H., K. Kuwabara, and G. Carta, *Temperature Dependence of the Dissociation Constants of Several Amino Acids.* Journal of Chemical & Engineering Data, 2008. **53**(3): p. 619-627.

73. Sharma, V.K., et al., *Dissociation Constants of Protonated Cysteine Species in NaCl Media.* Journal of Solution Chemistry, 2002. **31**(10): p. 783-792.

74. Sharma, V.K., et al., *Dissociation constants of protonated methionine species in NaCl media.* Biophysical Chemistry, 2003. **105**(1): p. 79-87.

75. Kochergina, L.A., S.G. Grosheva, and O.N. Krutova, *Thermochemical study of acid-base interactions in L-asparagine aqueous solutions.* Russian Journal of Inorganic Chemistry, 2011. **56**(9): p. 1481.

Chapter 4. Molecular thermodynamics for amino-acid adsorption at inorganic surfaces

The interaction of polypeptides and proteins with an inorganic surface is intrinsically dependent on the interfacial behavior of amino acids and sensitive to solution conditions such as pH, ion type, and salt concentration. A faithful description of amino-acid adsorption remains a theoretical challenge from a molecular perspective due to the strong coupling of local thermodynamic nonideality and inhomogeneous ionization of both the adsorbate and substrate. Building upon a recently developed coarse-grained model for natural amino acids in bulk electrolyte solutions, here we report a molecular theory to predict amino-acid adsorption on ionizable inorganic surfaces over a broad range of solution conditions. In addition to describing the coupled ionization of amino acids and the underlying surface, the thermodynamic model is able to account for both physical binding and surface associations such as hydrogen bonding or bidentate coordination. It is applicable to all types of natural amino acids regardless of the solution pH, salt type and concentration. The theoretical predictions have been validated by extensive comparison with experimental data for the adsorption of acidic, basic, and neutral amino acids at rutile (α-TiO_2) surfaces.

4.1 Introduction

Understanding the interaction of amino acids with inorganic surfaces is a prerequisite for studying the interfacial behavior of polypeptides and proteins under various solution conditions. By reducing the complexity of the polymeric systems to their building

blocks, we may identify key components of surface forces dictating both the adsorption and functionality of biomolecules at the interface[1-3]. However, despite its simplicity in comparison with proteins, the adsorption of amino acids in aqueous solutions is not yet fully understood from a molecular perspective[4]. The difficulty arises not only from multifaceted interactions between amino acids and an inorganic surface but also from the variation of both the surface charge and the degree of ionization of the functional groups in response to the changes in solution conditions such as pH, ion type, and salt concentration.

In addition to amine and carboxyl groups, natural amino acids consist of different side chains that vary in size and hydrophobicity. The diverse characteristics of these functional groups leads to complex interactions with inorganic surfaces in an aqueous medium. Conventionally, the surface forces are described in terms of electrostatic interactions, van der Waals and hydrophobic attractions, hydration forces, and various forms of surface association or chemical bonding. The intricate interplay of these physical and chemical interactions is coupled with protonation/deportation and multi-body correlations due to the inhomogeneous distributions of ionic species (and solvent molecules) near the surface. For most systems of practical interest, a first principles approach to predicting such interactions is beyond the reach of current computation capabilities. As a result, coarse-grained models may serve as a reasonable starting point to describe the adsorption and surface behavior.

A number of experimental techniques can be used to study the adsorption of amino acids at inorganic surfaces[5-8]. One such technique is infrared (IR) spectroscopy, which

provides valuable insight into the mechanisms of amino-acid adsorption through the comparison of the spectra of adsorbed species versus that in the bulk solution. For example, IR measurements revealed glutamic acid and aspartic acid binding with γ-Al_2O_3 through oxygen atoms from one or both carboxyl groups[9]. An indirect way to measure amino-acid adsorption is by monitoring its concentration in the bulk phase[7, 10-13]. The adsorption isotherm, i.e., the amount of amino acid adsorbed on the surface as a function of the *equilibrium* concentration in the bulk solution, can be determined from the reduction of the amino-acid concentration in the bulk solution. Radiotracer experiments are applicable to adsorption from dilute solutions whereby the depletion method is not possible[14]. With the help of a semi-empirical model such as the Langmuir equation, the adsorption isotherms provide valuable insight into the strength of surface binding and the maximum occupancy of amino acids. Whereas conventional adsorption models may be employed as a benchmark to compare the adsorption of amino acids at different surfaces, the model parameters are typically dependent on solution conditions thus have little predictive capability.

Complementary to experimental investigations, molecular modeling offers valuable details to bridge the knowledge gap between microscopic and macroscopic observations. Whereas atomistic models can be used to describe amino-acid adsorption through molecular dynamics (MD) simulation[15-17], coarse-grained models are often more convenient for the development of analytical methods. By capturing the essential features of intermolecular interactions and surface forces, coarse-grained models are able to provide quantitative predictions in good agreement with experimental observations. For

example, Vlasova and Golovkova studied the adsorption of basic amino acids to highly dispersed silica using a surface-complexation model that describes the amino-acid molecules at the silica surface through a Stern model of the electric double layer[18]. Two parameters were used for each amino acid, *viz*., the equilibrium constants governing the reaction of amino acids with a neutral or a negatively charged surface. With the assumption that these parameters were dependent upon the salt concentration but insensitive to the pH, the theoretical predictions of adsorption isotherms were found in quantitative agreement with experimental data. Similarly, Jonsson et al. proposed a surface-complexation model to describe the adsorption isotherms of L-aspartate and L-glutamate to a rutile surface at various solution conditions[19, 20]. They proposed a reaction scheme that involved association of the acidic amino acids with the rutile surface through the binding of either one or both carboxyl groups to one or four surface titanium sites, respectively. Like the Langmuir adsorption isotherm, the surface-complexation models neglect thermodynamic non-ideality due to excluded volume effects and electrostatic correlations that are important in describing the charge regulation of amino acids both in the bulk solution and near the surface.

In order to accurately describe the adsorption of amino acids to inorganic surfaces, we must account for thermodynamic non-ideality due to interactions between amino acids and all ionic species in the solution. In addition, it is essential to capture the charge behavior of amino acids as well as the ionization of the inorganic surfaces that play a significant role in the adsorption process. In our previous work, we developed a coarse-grained model for aqueous solutions of amino acids that appropriately takes into account

bulk interactions[21]. The thermodynamic model provides a faithful description of the charge behavior of amino acids as a function of pH and solution conditions. In addition, we proposed a coarse-grained model to describe the charge regulation of various ionizable surfaces by an explicit consideration of the inhomogeneous ion distributions and surface reactions[22-24]. Herein, we combine these models to examine the adsorption behavior of amino acids at the rutile (α-TiO_2) surface. The molecular-thermodynamic model allows us to predict the adsorption of various amino acids in response to the changes in solution conditions. This predictive model fills the gap between phenomenological models to describe the adsorption of amino acids and all-atom or first-principles simulations. Importantly, the theoretical procedure may provide a foundation for future studies of the interaction of polypeptides and proteins with various inorganic surfaces under diverse solution conditions.

4.2 Thermodynamic models and methods

4.2.1 A coarse-grained model for amino-acid adsorption

In this work, we are interested in developing a coarse-grained model applicable to amino- acid adsorption at an ionizable surface over a broad range of solution conditions. Toward that end, we employ an augmented primitive model (APM) that explicitly accounts for the different charged states of natural amino acids[21]. Due to the variation of hydration structure in response to the deprotonation or protonation of ionizable sites, each amino-acid molecule has a unique hard-sphere diameter σ_i and valence Z_i in different charged states.

Figure 1 shows a schematic representation of amino acids near an organic surface as described by the APM model. As in the primitive model (PM) that is conventionally used to describe the thermodynamic properties of aqueous electrolyte solutions, all amino acids and ionic species are represented by charged hard spheres and the solvent by a dielectric continuum of relative primitivity $\varepsilon_r = 78$, which corresponds to that for liquid water at the ambient condition. Because of ionization, each amino acid exists as a mixture of molecules in discrete integer-value charge states. In addition to the variation of the valence in response to pH changes, the augmented aspect of the primitive model is the inclusion of various well-recognized but poorly-understood water-mediated interactions among amino acids and other ionic species in the solution as well as their short-range interactions with the surface. In this work, such interactions are described through a square-well model. While the phenomenological description neglects atomic details and the anisotropic nature of amino-acid molecules, it provides a flexible framework to quantify the interfacial behavior of various natural amino acids in aqueous solutions by treating the square-well width and attraction energy as adjustable parameters. We demonstrated in our previous work that the augmented model is able to reproduce both the titration behavior and the thermodynamic properties of amino-acid solutions in good agreement with experimental observations[21].

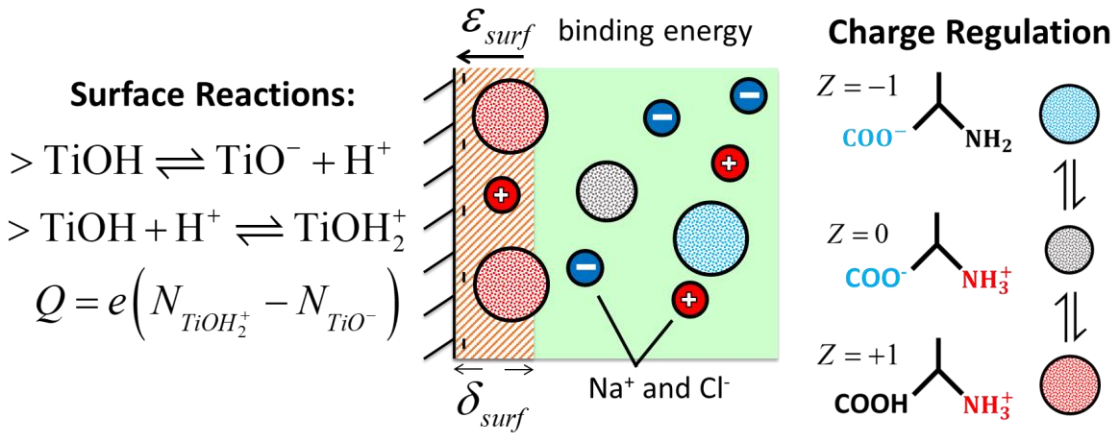

Figure 4.1 A coarse-grained model for the adsorption of natural amino acids on a rutile surface. In an aqueous solution, an amino-acid molecule may exist in different charged states dependent on the solution condition. The rutile surface may develop a net charge through deprotonation or protonation of the hydroxyl sites.

The coarse-grained model can be similarly applied to the adsorption of amino acids at an inorganic surface. In this case, the surface is represented by a hard wall with the charge density determined from the dissociation equilibrium of ionizable sites and ionic distributions self-consistently[22-24]. In addition to the electrostatic interactions and excluded volume effects, the surface interacts with amino acids through semi-empirical short-range interactions (e.g., hydrophobic attraction and hydrogen bonding). As in the PM description of electrolyte solutions, the van der Waals attraction is relatively insignificant in comparison with other forms of surface forces. Because our model accounts for ionization and the electrostatic charges of amino-acid molecules explicitly, we neglect the polarity effects on intermolecular and surface interactions. Our results suggest that the remanent electrostatic effects are effectively accounted for by short-range bindings. For simplicity, we fix the range of the short-range interaction between the surface and all amino acids at δ_{surf} = 0.4 nm from the edge of the hard wall (see Figure 1). The choice of four

angstroms is somewhat arbitrary in that a shift of the range of attraction can always be compensated by changing the energy parameter. Nevertheless, the chosen value appears reasonable from physical considerations because both hydrophobic interactions and hydrogen bonding are most significant when the amino acid is in intimate contact with the surface. It should be noted that δ_{surf} is immaterial to the identity of amino acids.

According to APM, the external potential due to the inorganic surface is expressed as

$$V_i^{ext}(z) = \begin{cases} \infty & z < \sigma_i / 2 \\ -\varepsilon_{surf,i} & \sigma_i / 2 \le z \le \delta_{surf} \\ 0 & else \end{cases} . \qquad (4.1)$$

In Eq. (1), the first line on the right accounts for the hard-wall repulsion, which limits the access of a chemical species (viz., ions and amino acids) to the solid phase beyond the point of contact; and the second line accounts for the short-range attraction. The latter is represented by a square-well potential with a constant energy in the immediate vicinity of the surface. Because of the excluded-volume effects, the closest distance for an amino acid from the surface depends on its hard-sphere diameter. As a result, the range of non-electrostatic surface attraction is different for different amino acids. On the other hand, the square-well energy, ε_{surf}, accounts for specific binding of an amino acid molecule with the surface. We expect that this parameter should be relatively insensitive to the changes in solution conditions (e.g., salt concentration), but may vary with temperature, because it reflects the local properties of the underlying chemical species. The results here are reported for T=298.15 K. The surface energy may also be used to account for adsorption

of amino acids due to specific chemical bonding with the surface. To a certain degree, our description of amino-acid binding with an inorganic surface is equivalent to surface complexation modeling mentioned above[25, 26]. One benefit of our thermodynamic approach is that it takes into account the solution effects on amino-acid adsorption.

4.2.2 Charge regulation for inorganic surfaces

In an aqueous solution, inorganic surfaces are typically terminated with hydroxyl groups due to the chemical adsorption of water molecules[27]. The deprotonation or protonation of these functional groups leads to a surface charge that can be regulated by adjusting the solution pH or salt concentration. Even for the same surface, the charge density may differ in both magnitude and sign depending on the solution conditions.

Due to its broad use as bioimplants, titanium dioxide is commonly used as a model inorganic material for studying amino-acid adsorption. In particular, experimental studies are mostly focused on rutile (100) and (110) surfaces as single crystals with these orientations are readily available. For a rutile surface, the deprotonation and protonation reactions are given by[28]

$$\text{-TiOH} \rightarrow \text{-TiO}^- + \text{H}^+, \tag{4.2}$$

$$\text{-TiOH} + \text{H}^+ \rightarrow \text{-TiOH}_2^+. \tag{4.3}$$

The equilibrium constants for these two reactions, K_D and K_P, respectively, are connected with the solution $\text{pH} = -\log a_{H^+}$ and the activities of the surface sites:

$$K_D = \frac{N_{TiO^-}}{N_{TiOH}} \frac{\gamma_{TiO^-}}{\gamma_{TiOH}} a_{H^+}, \tag{4.4}$$

$$K_P = \frac{N_{TiOH_2^+}}{N_{TiOH}} \frac{\gamma_{TiOH_2^+}}{\gamma_{TiOH}} \frac{1}{a_{H^+}} \tag{4.5}$$

where N_i refers to the number of surface site *i* per unit area, and γ_i is the activity coefficient of the corresponding surface site.

The activity coefficient of each surface group accounts for its physical interactions with the environment[29]. In this work, we assume that the ratio of the activity coefficients for each surface group in its different charge states to be governed by the local electrostatic potential (i.e., the surface potential, ψ_s). Thus, the surface reaction equilibrium can be re-expressed in terms of apparent equilibrium constants, K_D' and K_P',

$$K_D' = K_D \frac{\gamma_{SOH}}{\gamma_{SO^-}} = K_D \exp\left(\beta e \psi_s\right), \tag{4.6}$$

$$K_P' = K_P \frac{\gamma_{SOH}}{\gamma_{SOH_2^+}} = K_P \exp\left(-\beta e \psi_s\right). \tag{4.7}$$

The surface charge density is related to the total number density of the site and can be determined from the mass and charge balance:

$$N_{sites} = N_{SOH} + N_{SO^-} + N_{SOH_2^+}, \tag{4.8}$$

$$Q = -e\left(N_{SO^-} - N_{SOH_2^+}\right) \tag{4.9}$$

where N_{sites} refers to the total number of surface groups per unit area, *e* is the elementary charge, and Q is the surface charge density.

The combination of Eqs. (4-9) leads to an explicit expression for the surface charge density in terms of the total number of available sites per unit area, the apparent equilibrium constants, and the proton activity:

$$Q = -e\,N_{sites}\,\frac{K_D'/a_{H^+} - K_P'\,a_{H^+}}{1 + K_D'/a_{H^+} + K_P'\,a_{H^+}}. \tag{4.10}$$

Since the apparent equilibrium constants change with the surface potential, the surface charge density depends on the inhomogeneous ion distributions near the surface. As a result, the charge regulation of amino acids is more complex than that in the bulk solution due to the interfacial behavior of all charged species playing a key role in determining the local solution condition.

4.2.3 Ionization of amino acids in an inhomogeneous environment

Each amino acid molecule consists of a carboxyl group and an amine group which can be deprotonated and protonated, respectively, at suitable solution conditions. Neutral amino acids are referred to as those that do not contain any ionizable functional groups in the side chain. On the other hand, acidic and basic amino acids contain an additional functional group in their side chains that can be deprotonated and protonated, leading to an additional negative and positive charge, respectively. As a result of the acid-base equilibrium, amino acids exist in multiple charged states dependent on the pH and other solution conditions.

For each amino acid, the acid-base equilibrium can be described in terms of multi-step protonation reactions

$$A_{i-1} + H^+ \rightarrow A_i \tag{4.11}$$

where A_i refers to the amino acid in a given charged state. The amino-acid valence satisfies $Z_{A_i} = Z_{A_0} + i$ with Z_{A_0} being the charge for the amino acid in its fully deprotonated state. For a neutral or basic amino acid, $Z_{A_0} = -1$; while an acidic amino acid has a valence of -2 in its fully deprotonated state. Accordingly, in the fully protonated state, each neutral or acidic amino acid has a valence of +1, and each basic amino acid has a valence of +2.

Similar to that in the bulk solution, the equilibrium constant for the protonation of amino acids in an inhomogeneous environment (viz., near the surface) is related to the local solution composition and activity coefficients

$$K_i^T = \frac{\rho_{A_i}(z)}{\rho_{A_{i-1}}(z)} \frac{\gamma_{A_i}(z)}{\gamma_{A_{i-1}}(z)} \frac{1}{a_{H^+}} \tag{4.12}$$

where K_i^T represents the equilibrium constant for the amino acid in charge state *i*, $\rho_i(z)$ is the number density of species *i* at perpendicular distance *z* from the surface, and γ_i is its local activity coefficient. The equilibrium constant is a thermodynamic quantity defined by the change in the chemical potentials of reactants and products at their corresponding reference states, i.e., each species in an ideal solution at unit molar concentration[30]. These values were obtained in our previous work based off correlations with experimental data for all natural amino acids in bulk NaCl solutions[21]. The activity coefficients account for the effect of solvent-mediated interactions among all chemical species in the solution.

Within the augmented primitive model (APM), the local activity coefficient can be expressed as

$$k_B T \ln \gamma_i = \mu_i^{ex} = \mu_i^{hs} + \mu_i^{sw} + \mu_i^{el} + \psi Z_i e + V_i^{ext} \tag{4.13}$$

where μ_i^{ex} stands for the local excess chemical potential, *i.e.*, deviation from the chemical potential of species *i* in an ideal solution; k_B is the Boltzmann constant, and *T* is the absolute temperature. μ_i^{hs}, μ_i^{el}, and μ_i^{sw} are contributions to the local excess chemical potential due to hard-sphere repulsion, electrostatic correlation, and solvent-mediated interactions, respectively. The last two terms, V_i^{ext} and ψ, are the external potential due to the interface (e.g., hard wall and specific binding to the surface) and the local electrostatic potential, respectively. The expression for each contribution has been discussed previously and can be found in Supporting Information[31-33].

4.2.4 Classical density functional theory

Classical density functional theory (cDFT) provides a generic mathematical procedure to describe the inhomogeneous distributions of ionic species including amino acids near inorganic surfaces. At a given temperature *T* and the bulk densities of all ionic species { ρ_i^b }, cDFT predicts the local density, $\rho_i(z)$, of each species by minimizing the grand potential

$$\Omega = F + \sum_i \int \left[V_i^{ext}(z') - \mu_i \right] \rho_i(z') dz' \tag{4.14}$$

where V_i^{ext} and μ_i are the external potential and the chemical potential of species *i*, respectively, and F denotes the total intrinsic Helmholtz energy. The latter consists of an ideal-gas contribution and the excess arising from intermolecular interactions. Similarly, the chemical potential is composed of an ideal part, which is directly related to the number density of the species, and an excess part as discussed in the prior section.

For systems considered in this work, the intermolecular interactions include the hard-sphere repulsion, electrostatic correlations, solvent-mediated interactions, and direct coulombic interactions. As the chemical potential is uniform throughout the system, any change in the local environment directly influences the local densities of all species as predicted by the Euler-Lagrange equation:

$$k_B T \ln \rho_i^b + \mu_i^{ex,b} = k_B T \ln \rho_i(z) + \mu_i^{ex,\mathrm{int}}(z) + \psi(z) Z_i e + V_i^{ext}(z). \qquad (4.15)$$

The left side of Eq. (15) corresponds to the chemical potential of species *i* in the bulk, whereas the right side is that at position *z* in the inhomogeneous fluid (i.e., near the interface). Similar to that in the bulk solution, the excess chemical potential $\mu_i^{ex,int}$ accounts for intermolecular interactions other than the direct Coulomb contribution. The latter is related to the mean electrostatic potential, ψ, which includes contributions from Coulomb interactions between all ionic species and the surface. The last term on the right side of Eq. (15), V_i^{ext}, denotes the non-electrostatic interactions with the surface, i.e., the hard-wall and short-range attraction as given by Eq. (1). It is worth noting that the chemical equilibrium between the different electrostatic states of the amino acid near the surface is naturally satisfied within our cDFT calculations.

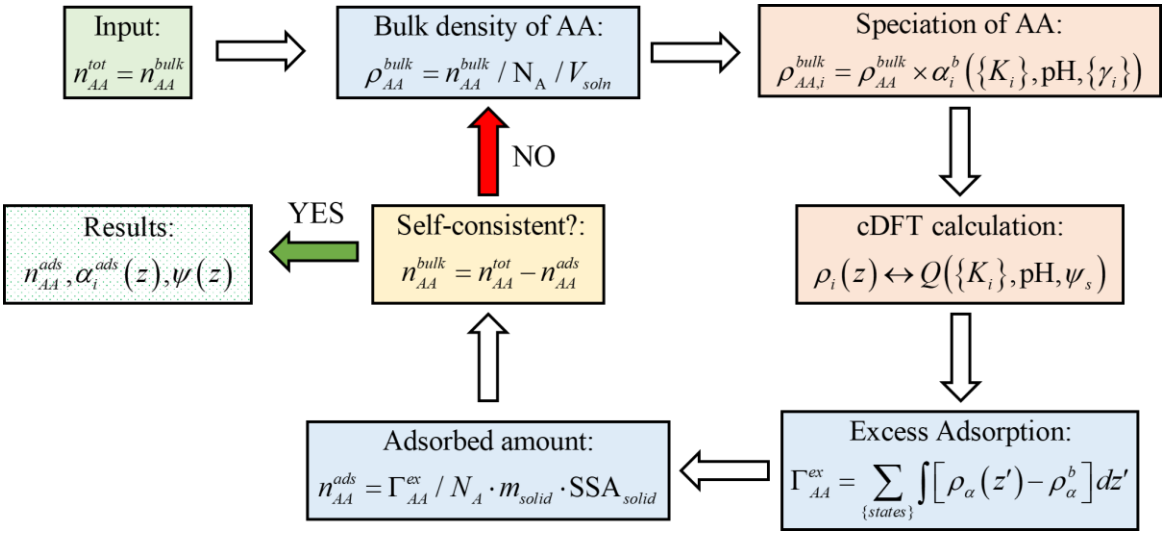

Figure 4.2 A flowchart of the computational procedure to determine the adsorption of amino acids on an inorganic surface. In consistent with experimental measurements of adsorption isotherms, each cDFT calculation is carried out at a fixed temperature, volume, and the total moles of amino acid n_{AA}^{tot} in the system.

To facilitate direct comparison of the theoretical results with experimental data, we must perform the cDFT calculations for the adsorption of amino acids at an inorganic surface at a fixed total number of amino-acid molecules in the solution. Figure 2 presents a flowchart of our calculation procedure. The total amount of amino acid in the solution includes contributions from both the adsorbed amount on the surface as well as that in the bulk solution, i.e., $n_{AA}^{tot} = n_{AA}^{bulk} + n_{AA}^{ads}$. However, the cDFT calculation must be performed at a fixed bulk chemical potential, i.e., the bulk density of each species must be known *a priori*. To convert the moles of amino acid in the solution to a bulk density, we must know the solution volume (V_{soln}). We can then relate the total number of moles of the amino acid to its corresponding bulk density by $\rho_{AA}^{bulk} = n_{AA}^{bulk} / N_A / V_{soln}$, where N_{A} is Avogadro's number. Because we explicitly consider the chemical equilibrium between the different

ionization states of each amino acid, we must also determine the bulk density for the amino acid at each ionization state. The degree of dissociation at state *i* is defined by $\alpha_i = \rho_{A_i} / \sum_j \rho_{A_j}$, where the summation extend over all possible charged states of the amino acid. By coupling the degree of ionization with the expressions for protonation/deprotonation equilibrium, we can determine the relative presence of each charged state for the amino acid from

$$\alpha_i = \frac{a_{H^+}^i \prod_{j=1}^{i} K'_j}{1 + \sum_{j=\{states\}} a_{H^+}^j \prod_{k=1}^{j} K'_k} \tag{4.16}$$

where $K'_j = K_j^T \gamma_{A_{j-1}} / \gamma_j$ is the apparent equilibrium constant defined by the thermodynamic equilibrium constant and the ratio of the bulk activity coefficients.

Eq. (16) can be solved numerically by using known equilibrium constants for different functional groups of the amino acid and the augumented primitive model (APM) to account for thermodynamic non-ideality. Once the bulk density of the amino acid at each charged state has been determined from $\rho_{AA_i}^{bulk} = \rho_{AA}^{bulk} \alpha_i$, cDFT calculation is then performed to determine the adsorption of the amino acid at the inorganic surface. The chemical equilibrium for the protonation/deprotonation of amino acids in the inhomogeneous environment will be automatically satisfied as shown by Eq. (15). During the cDFT calculations, we determine the density profiles of the amino-acid molecules in different charged states along with the distributions of salt ions. The charge density of the inorganic surface can be calculated from the condition of charge neutrality as shown in Eq.

(9). The surface excess for each amino acid, i.e., the amount adsorbed at the surface, is given by

$$\Gamma_{AA}^{ex} = \sum_{\{states\}} \int \left[\rho_\alpha (z') - \rho_\alpha^b \right] dz' . \tag{4.17}$$

Given the total mass (m_{solid}) and the specific surface area (SSA) of the solid, the excess adsorption can be related to the total adsorbed amount by

$$n_{AA}^{ads} = \Gamma_{AA}^{ex} / N_A \cdot m_{solid} \cdot \text{SSA}_{solid} . \tag{4.18}$$

As mentioned above, experimental measurements of adsorption isotherms are often reported in terms of the total moles of amino acid in the solution. The sum of amino acids in the bulk and that at the surface must be equal to the total moles in the solution. In our cDFT calculations, we iterate the cDFT calculations with the mole amount in the bulk given by $n_{AA}^{bulk} = n_{AA}^{tot} - n_{AA}^{ads}$ and solve for n_{AA}^{ads} from Eq. (18) until the quantity is converged. After convergence, we can make direct comparison of the theoretical results with experimental data for the adsorbed amount of each amino acid given that the total amino acid in the solution is fixed.

4.3 Results and discussion

In the following, we provide 3 case studies to illustrate the application of our coarse-grained model to describe the effects of solution conditions on the adsorption of amino acids on different inorganic surfaces. We first consider the charge regulation of rutile in different salt solutions. Rutile can be either positively or negatively charged depending upon the solution pH. Next, we investigate the adsorption isotherms on rutile

(110) surface for the three charge classes of natural amino acids: acidic, basic, and neutral. We find that the adsorption of these amino acids is highly dependent upon pH, amino acid concentration, and salt concentration in the bulk solution. Due to the protonation/deprotonation of both the amino acid molecules and the rutile surface, the adsorption isotherm typically shows a maximum at an intermediate pH value that is dependent upon the salt type and concentration. In all cases, we find that our coarse-grained model performs well in capturing the adsorption behavior of natural amino acids in different solution conditions.

4.3.1 Surface charge of rutile in aqueous NaCl and $CaCl_2$

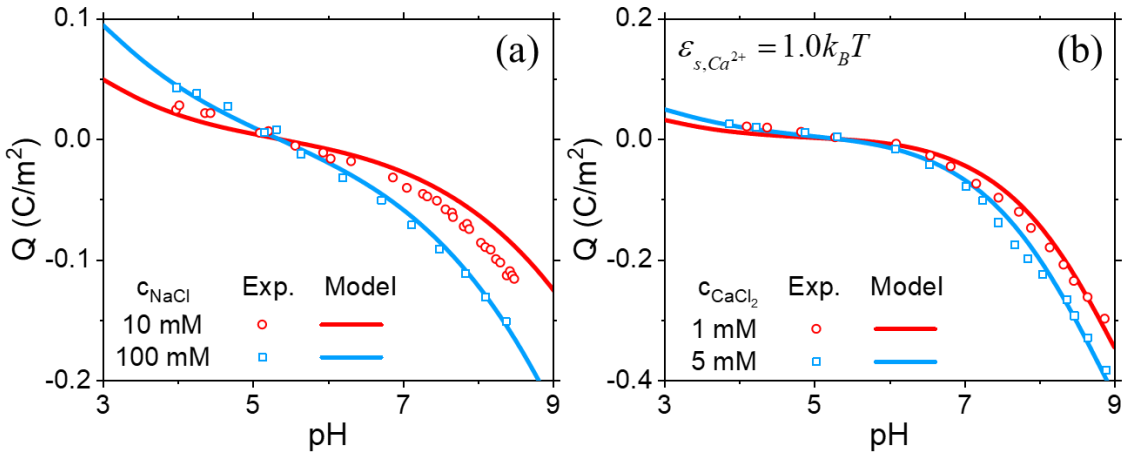

Figure 4.3 The surface charge density of rutile as a function of pH in (a) NaCl solution and (b) $CaCl_2$ solution from experiment[19, 34] (symbols) and from the prediction of the coarse-grained model (lines). The surface site density is fixed at 3 nm^{-2} in agreement with Bahri et al.[19]. The deprotonation and protonation constants are pK_D=6.5 and pK_P=-4.1, respectively.

Charge regulation is important for understanding the adsorption of amino acids at rutile (and other inorganic) surfaces as the adsorption is highly sensitive to electrostatic interactions. Figure 3 shows the surface charge density of rutile in sodium chloride and

calcium chloride aqueous solutions at two representative salt concentrations. By employing the augmented primitive model (APM) for aqueous electrolytes, we can describe the surface charge at different pH and salt concentrations by using only the thermodynamic equilibrium constants for protonation and deprotonation. These two constants are inherently coupled with the isoelectric point of the surface which can be related by $pI = (pK_D - pK_P)/2$. We set the isoelectric point to pI=5.3 consistent with previous experimental findings[35]. Based on this, the deprotonation and protonation constants were determined to be pK_D = 6.5 and pK_P = -4.1, respectively. The surface charge is highly dependent on pH and, to a lesser extent, on sodium chloride concentration. The addition of salt ions to the solution reduces the electrostatic repulsion between surface sites leading to an increase in the surface charge at a fixed pH[17]. APM slightly overestimates electrostatic repulsion between the surface sites as shown by the underestimation of the surface charge at 10 mM. A more sophisticated approach to describe the surface reactions (e.g., multi-site or ion-complexation) could lead to more accurate prediction of the surface charge regulation. As can be seen from Figure 3, the rutile surface is positively charged at low pH (i.e., below its isoelectric point) and negatively charged at high pH. The regulation of the surface charge provides a convenient way to promote or inhibit the adsorption of amino acids or charged organic molecules. We see that APM provides a quantitative description of the effects of the monovalent electrolyte on the surface charge of rutile. A good agreement of the theoretical prediction and experiment was also found for other inorganic surfaces[22, 24].

We next consider the surface charge of rutile in calcium chloride solutions. As shown in Figure 3b, the deprotonation of the rutile surface (i.e., increases in negative charge) is less dependent on pH in the presence of calcium ions. The trend is noticeable despite the fact that the concentration of calcium ions is one order of magnitude smaller than that of sodium ions (e.g., 1 mM $CaCl_2$ vs. 10 mM NaCl). Such behavior can be attributed to the stronger electrostatic binding between the negatively charged surface sites and the divalent cation. On the other hand, when rutile is positively charged, the influence of the cation valence is negligible as it is the monovalent chloride ions dictating the surface charge density. To match the experimental data for rutile in the presence of calcium chloride, we used the same equilibrium constants determined for rutile in sodium chloride solutions. However, this led to a consistent underestimation of the surface charge above the isoelectric point. Previous studies have demonstrated the importance of calcium binding to surface hydroxyl sites beyond electrostatic attraction[36]. It has also been suggested that the hydration of the calcium ion near the surface is different from that in the bulk[37]. To account for such effects, we introduced a short-range attraction of magnitude ε_{surf} = 1.0 k_BT between the calcium ions and the rutile surface to mimic this non-electrostatic binding behavior. The inclusion of this short-range attraction leads to a perfect description of the ionization behavior of rutile in the presence of calcium chloride.

4.3.2 Adsorption of acidic amino acids to rutile

The adsorption of amino acids at inorganic surfaces is complicated by the fact that both the amino acid and surface are ionizable and that the surface forces are highly sensitive to the solution pH and local ion concentrations. Due to competing electrostatic

interactions, the adsorption isotherm is often nonmonotonic with respect to pH and shows a strong dependence on solution conditions (e.g., salt type and concentration). To elucidate, Figure 4 presents the adsorption of glutamic acid and aspartic acid at a rutile surface versus pH at several sodium chloride and amino acid concentrations. The adsorption of both acidic amino acids shows a maximum near the pK_a value of the amino-acid side chain. At this condition, approximately 50% of the acidic amino acids carry a negative charge. A reduction of the pH below the pK_a would shift the equilibrium to the neutral state while an increase in pH would result in the amino acids existing predominately in their negatively charged state. Meanwhile, the rutile surface exhibits a positive charge below its isoelectric point of pI=5.3, which leads to a pH window where the adsorption is favored by the electrostatic interaction between the surface and the amino acid. Surprisingly, significant adsorption is observed even at pH 9 where there is a strong electrostatic repulsion between the acidic amino acids and the negatively charged surface. In this case, the adsorption can be attributed to hydrogen bonding between the carboxyl groups from the amino acids and the surface hydroxyl sites[38, 39]. Such interaction can be modeled through a square-well potential between the amino acid and the surface, irrespective of the charge status. We determined the fitting parameter to be ε_{surf} =9.6 k_BT based off a comparison with the experimental adsorption data for 0.5 mM amino acid and 100 mM sodium chloride solution. Due to similarities in the chemical structure of both acidic amino acids, it is expected that the two will interact similarly with the rutile surface. Therefore, we did not distinguish the surface binding energy between the two amino acids which, as demonstrated by Figure 4, was found to be satisfactory.

As expected, the solution condition has an important effect on the adsorption of acidic amino acids at the rutile surface. In general, the adsorbed amount increases with the total concentration of amino acid in the solution. For example, the mean lateral separation between glutamic acid on the rutile surface at pH 3 decreases from approximately 3 nm (~0.2 μmol/m^2) to 1 nm (~1.52 μmol/m^2) when the concentration of glutamic acid is increased from 0.1 mM to 2.0 mM. Such effect is most noticeable at low pH because the opposite surface charge leads to a stronger driving force for the adsorption. While the amino-acid concentration has little effect on the solution pH at which the maximum in adsorption occurs, the salt concentration results in a notable shift in the pH at which the adsorption is maximized. The variation of the salt concentration will also influence the adsorption of charged amino acids because of the electrostatic screening effect.

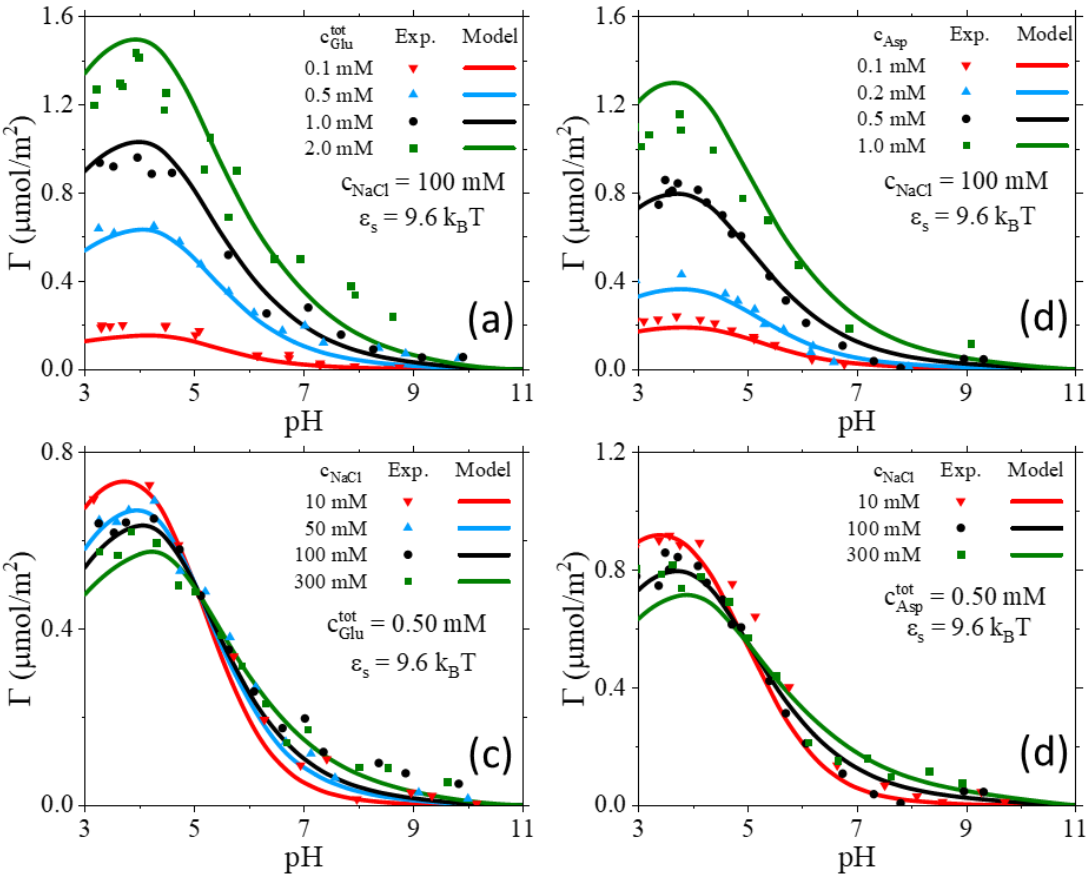

Figure 4.4 The pH-dependent adsorption of glutamic acid and aspartic acid on the rutile surface at (a and b) various total concentrations of amino acids in a 100 mM aqueous NaCl solution and (c and d) several NaCl concentrations in a 0.50 mM aqueous amino acid solution. The symbols are from experiment[19], and the lines are predictions of the coarse-grained model.

It has been well documented that the addition of salt ions to the solution will promote ionization of the inorganic surface by decreasing the electrostatic repulsion between the surface sites (as shown previously in Figure 3). However, it is not clear how the distribution of charged states for the amino acid is affected by the increase in salt concentration. While the ionized forms of the amino acid would be typically favored by increasing the salt concentration in the bulk solution, the reduced Coulomb interaction with the surface sites may be more impactful on amino-acid speciation. Since an increase in the

salt concentration shifts the maximum adsorption to a higher pH (i.e., the chemical equilibrium shifts to the negatively charged state of the amino acid), we expect that, at a fixed pH, the addition of salt inhibits amino acids in the ionized state. This is because a weaker interaction with the positively charged surface by the negatively charged amino acid reduces the preference of the amino acid in its anionic state.

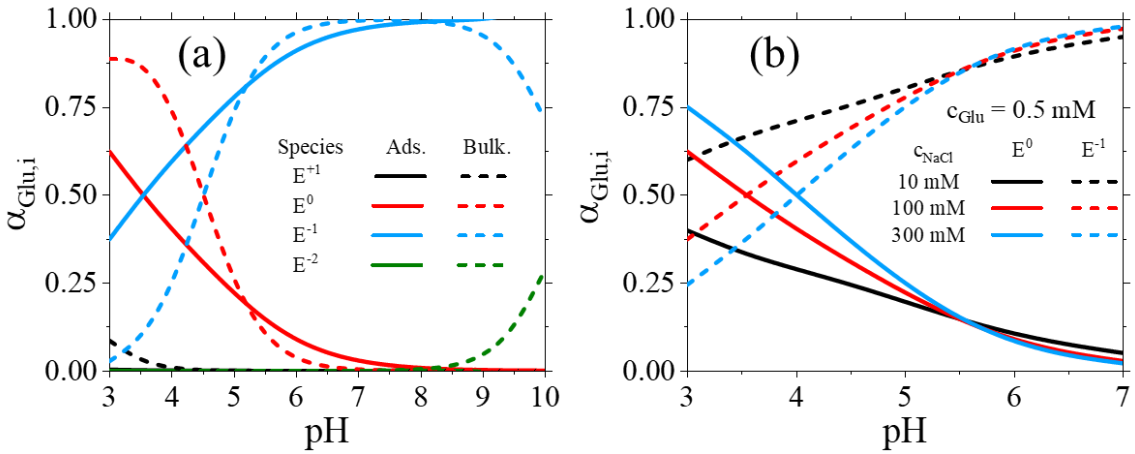

Figure 4.5 (a) The pH-dependent speciation of glutamic acid at the rutile surface and in the bulk solution. The total concentration of glutamic acid is 0.5 mM, and the NaCl concentration is 100 mM. (b) Speciation of adsorbed glutamic acid into its two predominant states, neutral and negatively charged, in the pH range of 3 to 7 at three NaCl concentrations (10, 100, and 300 mM).

The interfacial behavior of amino acids at inorganic surfaces are dependent on their ionization states near the surface. In most cases, the dominate charge state at the surface is significantly different from that in the bulk solution due to the selective interactions of the surface with amino acid molecules. To demonstrate this, we show in Figure 5a the pH-dependent speciation of glutamic acid in the bulk solution and on the rutile surface in 100 mM sodium chloride solution. At low pH, the amino acid is dominated by its negatively charged state due to strong Coulomb interactions with the positively charged surface sites

on rutile while the amino acid mostly exists in its neutral state in the bulk solution (i.e., far from the surface). There is a small reversal in this trend as the pH is increased because the rutile surface transitions to negatively charged and repulses the anionic state of the amino acid. At higher pH, the monovalent anionic state is maintained somewhat more than that in the bulk solution to avoid the repulsion between the negatively charged surface and amino-acid molecules in the divalent anionic state. Based off these results, it also becomes clear why there is a maximum in the adsorption at low pH seen in Figure 4. While the surface has more positive charge as pH falls, it adsorbs less amino acid even though it is negatively charged. This interplay between a stronger electrostatic attraction but a lower concentration of amino acids in the anionic state leads to the maximum adsorption in responding to the pH changes. Although the positively charged surface promotes the ionization of the acidic amino acid relative to the bulk solution, the increased electrostatic force is insufficient to compensate the reduction in the concentration of a negatively charged amino acid.

Figure 5b shows the influence of salt concentration on the speciation of the adsorbed amino acid molecules in neutral and negatively charged states. It is clear that the salt concentration plays a key role on speciation at low pH where the adsorption is mostly driven by electrostatic forces. At low pH, an increase in the salt concentration disfavors the presence of the amino acid in the negatively charged state. Interestingly, an opposite trend is observed as the salt concentration in the bulk solution rises[40]. In the former case, the addition of salt weakens the electrostatic attraction between the rutile surface and amino acid molecules (*viz*. through the reduction of the mean electrostatic potential). The

electrostatic interaction is responsible for the increased population of the anionic state near the surface while the bulk solution is dominated by amino-acid molecules in the neutral state. The increase in the salt concentration reduces the mean electrostatic potential at the surface leading to a reduction of the local concentration of the amino acid in the negatively charged state. At these conditions (e.g., low pH and small salt concentration), the ionic excluded-volume effects and electrostatic correlations are relatively insignificant in the ionization behavior of the acidic amino acid. Therefore, an accurate description of charge regulation for both the amino acid and the surface is essential for understanding the adsorption of amino acids at the rutile surface.

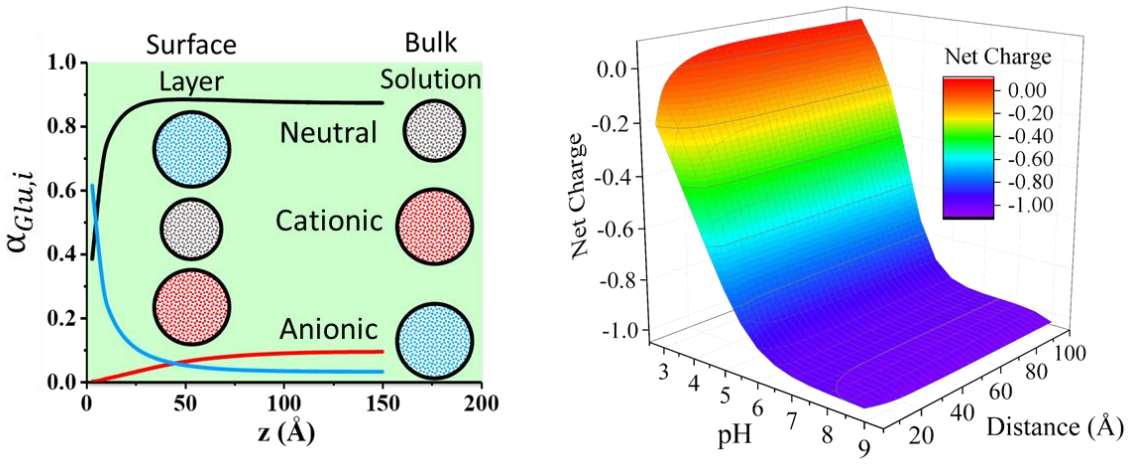

Figure 4.6 (a) The speciation of glutamic acid as a function of the distance from the surface at pH=3. (b) The net charge of the glutamic acid as a function of the distance from the surface and pH. In both cases, the total concentration of glutamic acid is 0.5 mM, and the NaCl concentration is 100 mM.

The presence of rutile, particularly when it is highly charged, leads to a significantly different speciation of the amino acid near the surface from that in the bulk solution. The variation in local ionization may span up to a few nanometers depending on the salt

concentration. In Figure 6a, we present theoretical predictions for the speciation of glutamic acid as a function of the distance from the surface at pH 3. In this case, the rutile surface is positively charged as previously shown in Figure 3. The strong electrostatic attraction between the amino acid in the negatively charged state and the positively charged surface shifts the protonation equilibrium of the amino acid to the anionic state. Meanwhile, the amino acid in neutral and cationic states, which dominate in the bulk solution, are depleted near the surface. While amino acids typically show monolayer adsorption and thus the long-range ionization behavior is of less significance, the strong variation in speciation near the surface results in drastic differences between the composition of the adsorbed layer and that in the bulk solution. On the other hand, the long-range ionization behavior may find more importance in the adsorption of polypeptides and proteins due to strong intrachain correlations.

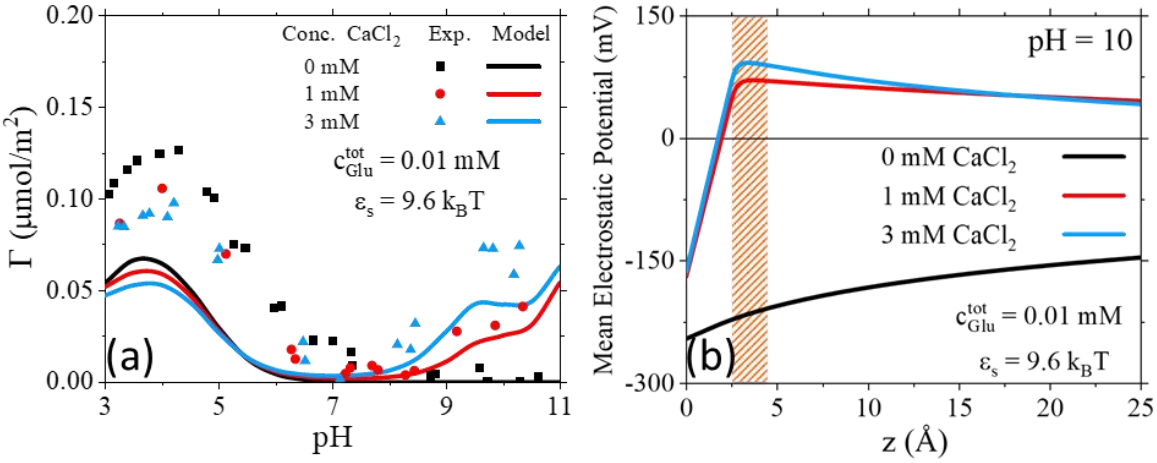

Figure 4.7 (a) The pH-dependent adsorption of glutamic acid on the rutile surface at different concentrations of $CaCl_2$ in 0.01 mM aqueous glutamic acid solution from experiment[36] (symbols) and the theoretical prediction (lines). (b) The mean electrostatic potential as a function of the distance from the rutile surface for the adsorption of glutamic acid at pH = 10 from salt-free and 1 or 3 mM $CaCl_2$ solutions. The shaded region corresponds to a monolayer of glutamic acid adsorbed at the surface.

We show in Figure 6b how the net charge of glutamic acid varies with the distance from the rutile surface and pH. In general, the net charge is shifted towards the charged state that favors interaction with the surface. At low pH, the glutamic acid has a positive net charge in the bulk solution while the rutile surface is also positively charged. In this case, the net charge is reduced as the amino acid is positioned close to surface. The opposite occurs at high pH with the net charge increasing near the surface since rutile is negatively charged at these conditions. Again, an accurate description of the charge regulation for both the surface and amino acid is important to predict how changes in the environment affect the amino-acid adsorption.

Next, we consider the role of multivalent ions in facilitating the adsorption of amino acids. Because of electrostatic correlations, significant adsorption may take place in the presence of multivalent ions despite that the amino acid and the surface have the same sign of electric charge. Figure 7 shows the adsorption of glutamic acid at the rutile surface in a salt-free solution and two calcium chloride solutions of different concentrations. For both glutamic acid and calcium ions, the surface energy parameters are the same as those determined previously. For the salt-free case, glutamic acid exhibits a maximum adsorption at low pH with little to no adsorption when pH>7. Qualitatively, the pH effect is consistent with the adsorption at higher amino-acid concentrations as previously shown in Figure 4. The addition of calcium ions to the solution leads to significant adsorption of the glutamic acid at high pH when both the amino acid and the rutile surface are negatively charged.

To understand the origin of attraction, we show in Figure 7b the mean electrostatic potential as a function of the distance at pH 10 for the three aqueous solutions considered in Figure 7a. For the salt-free case (i.e., 0 mM $CaCl_2$), the mean electrostatic potential is strongly negative at the surface and decays slowly with the distance as predicted by a conventional electric double layer (EDL) model. However, in the presence of calcium ions, the surface potential is significantly reduced and the local electrostatic potential switches in sign (i.e., from negative to positive) signaling charge inversion. In other words, the surface charge is overcompensated by the oppositely charged ions through specific association and electrostatic correlations[41]. The charge inversion leads to the adsorption of the negatively charged glutamic acid which experiences an effective attraction to the overcompensated "positive" surface despite itself being negatively charged. Although the agreement between experiment and theoretical predictions is only semi-quantitative, Figure 7 clearly demonstrates that our model captures the important physics regarding the effect of multivalent ions on the adsorption of amino acids at inorganic surfaces.

4.3.3 Adsorption of basic amino acids

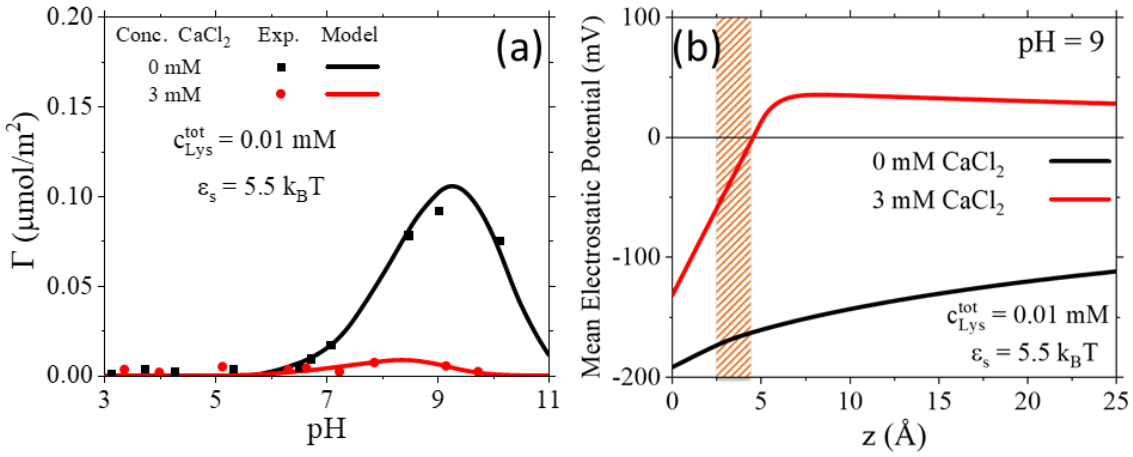

Figure 4.8 (a) The pH-dependent lysine adsorption at the rutile surface from experiment[36] (symbols) and the theoretical prediction (lines). (b) The mean electrostatic potential as a function of the distance from the rutile surface in two aqueous solutions at pH=9. Approximately, the shaded area corresponds to a monolayer of lysine at the rutile surface.

We now turn our attention to the adsorption of lysine, a basic amino acid, at the rutile surface. In the bulk solution, lysine molecules mostly exist in its positively charged state, except for high pH (i.e., pH>9) where its backbone amine group will deprotonate shifting the equilibrium to its neutral state. Further increase in the pH will result in the side-chain amine group also deprotonating and the anionic state of lysine becoming dominant. Figure 8 shows the adsorption of lysine at the rutile surface in a salt-free solution and that in a 3 mM calcium chloride solution. While aspartic acid and glutamic acid display significant adsorption when they have the same electric charge as the surface, lysine shows no adsorption at similar conditions. Interestingly, no significant adsorption is observed even when the surface is neutral or weakly charged with a negative sign. The non-adsorption behavior can be attributed to the limited binding capability of its side chain with

the rutile surface. In contrast to the carboxyl group in the side chain of aspartic or glutamic acid, there is less significant binding between the amine group in the lysine side chain with the hydroxyl groups from the rutile surface. In our coarse-grained model, the non-electrostatic effect is captured by a lower binding energy of each lysine molecule with the surface, ε_{surf} =5.5 k_BT, compared with 9.6 k_BT for the acidic amino acids. As a result, the electrostatic interactions play a more dominant role in lysine adsorption at the rutile surface.

When the solution pH is above the isoelectric point of the rutile surface (pI = 5.3), lysine molecules and the surface have opposite charges and the electrostatic attraction leads to strong adsorption of the amino acid. However, the adsorption shows a maximum at pH 9 beyond which it falls as a result of the increased prevalence of lysine molecules in the neutral state. The situation is similar to that found for acidic amino acids at low pH. In the presence of calcium ions, the adsorption of lysine drops dramatically compared to the salt-free case. This is not surprising because the divalent cations outcompete the monovalent lysine. As the non-electrostatic binding energy is relatively insignificant in comparison with that for acidic amino acids, lysine adsorption is more sensitive to the screening effects due to salt ions in the solution.

Figure 8b shows how the local electrostatic potential varies with the distance from the rutile surface. As discussed above, the presence of calcium ions leads to a significant reduction of the surface electric potential. The charge inversion hinders the adsorption of lysine molecules at the rutile surface even though the amino acid and the surface have

opposite charges. The competition in adsorption between the amino acid and the multivalent ions to the rutile surface is well captured by our coarse-grained model.

4.3.4 Adsorption of neutral amino acids

Lastly, we consider the adsorption of a neutral amino acid, dihydroxyphenylalanine (DOPA), to the rutile surface. From the practical perspective, understanding DOPA adsorption is highly important due to its extensive use in bioadhesive materials[42-44]. DOPA exhibits excellent adhesive and cohesive properties through its strong binding to inorganic surfaces by a bidentate bonding of its two hydroxyl groups in the side chain and complexation with multivalent ions by its aromatic group, respectively[45]. While most neutral amino acids show little or no adsorption at inorganic surfaces[18], DOPA displays strong adsorption that is highly dependent on the solution pH and salt concentration.

As demonstrated by spectroscopy measurements[46], DOPA binds with an inorganic surface in one of two configurations: lying flat or standing up. When a DOPA molecule is in its flat configuration, it interacts with the surface through its carboxyl group in the backbone and the two hydroxyl groups in its side chain. Meanwhile, the standing up orientation allows only for the interaction of the two hydroxyl groups with the surface. Since these configurations are dependent upon the availability of specific surface sites, the solution pH has noticeable effects on both the adsorption and configuration of DOPA due to the protonation or deprotonation of the surface sites.

In order to accurately describe the adsorption of DOPA to an inorganic surface, we must acknowledge the existence of these two orientations at the surface. Barhi et al.

employed an extended triple layer model that describes DOPA attachment to the surface through either four or two surface sites[47]. The former corresponds to the lying flat orientation where the two phenolic oxygens and one of the carboxylate oxygens bind to the titanium atom on the surface while the other carboxylate oxygen is hydrogen bonded to a hydroxyl group at the surface. On the other hand, the complexation with two surface sites involves one of the phenolic oxygens bonding to the titanium atom while the other phenolic oxygen is involved with hydrogen bonding to the surface. The surface complexation scheme can be described through the following chemical reactions[47]:

$$4-\mathrm{TiOH}+\mathrm{H_3DP}\leftrightarrow\left(-\mathrm{TiOH}\right)-\mathrm{Ti_3DP}+3\mathrm{H_2O} \tag{4.19}$$

$$2-\mathrm{TiOH}+\mathrm{H_3DP}\leftrightarrow\left(-\mathrm{TiOH_2^+}\right)-\mathrm{TiHDP^-}+\mathrm{H_2O} \tag{4.20}$$

where H_3DP represents DOPA in its neutral state. The thermodynamic equilibrium constants for these reactions can be related to the activities of the pertinent species by

$$K_{(-\mathrm{TiOH})-\mathrm{Ti_3DP}}=\frac{a_{(-\mathrm{TiOH})-\mathrm{Ti_3DP}}a_{\mathrm{H_2O}}^3}{a_{-\mathrm{TiOH}}^4 a_{\mathrm{H_3DP}}}\exp\left[\beta e\Delta\psi_{r,1}\right], \tag{4.21}$$

$$K_{\left(-\mathrm{TiOH_2^+}\right)-\mathrm{TiHDP^-}}=\frac{a_{\left(-\mathrm{TiOH_2^+}\right)-\mathrm{TiHDP^-}}a_{\mathrm{H_2O}}}{a_{-\mathrm{TiOH}}^2 a_{\mathrm{H_3DP}}}\exp\left[\beta e\Delta\psi_{r,2}\right]. \tag{4.22}$$

The inclusion of the exponential term results from the electrical work involved with moving ions and/or water dipoles to and from the surface[26]. The electrostatic work to displace water dipoles from the surface is given by $\Delta\psi_{\mathrm{H_2O}}=-n_{\mathrm{H_2O}}(\psi_s-\psi_\beta)$, where $n_{\mathrm{H_2O}}$ represents the stoichiometric coefficient of water molecules on the right-hand side of the reactions in Eqs. (19) and (20), ψ_s and ψ_β refer to the electric potentials at the surface

and at the plane at which the DOPA molecule is adsorbed (i.e., when in contact with the surface, $z = \sigma_{DP} / 2$). The electrical work involved in the first reaction is

$$\Delta\psi_{r,1} = 3\psi_s - 3\psi_\beta - 3\left(\psi_s - \psi_\beta\right) = 0 \tag{4.23}$$

where the three terms in the middle correspond to changes in the potentials experienced by the three H^+ ions adsorbing to the surface, the DP^{3-} ion adsorbing to the β-plane, and three H_2O molecules desorbing from the surface, respectively. Thus, there is no contribution by the electrical work involved with moving ions and water molecules to and from the surface. On the other hand, the electrical work of the second reaction is

$$\Delta\psi_{r,2} = 2\psi_s - 2\psi_\beta - \left(\psi_s - \psi_\beta\right) = \psi_s - \psi_\beta \tag{4.24}$$

where the three terms in the middle correspond to changes in the electric potentials experienced by the two H^+ ions adsorbed at the surface, the HDP^{2-} ion adsorbed at the β-plane, and H_2O molecule desorbed from the surface, respectively. Based off the fitting with the experimental data, we determined the negative logarithmic equilibrium constants for the two reactions to be 11.6 and 5.8, respectively. The pK_a values are slightly different from those determined by Bahri et al. (11.8 and 6.4, respectively). The difference is expected because we consider thermodynamic non-ideality in addition to the direct Coulomb interactions.

We show in Figure 9a and b the adsorption of DOPA to the rutile surface versus the solution pH at different salt and total DOPA concentrations. The adsorption of most neutral amino acids shows little sensitivity to pH since they are not directly influenced by changes in the surface charge[48]. This is valid except at very low or high pH values when

the ionized state of the amino acid becomes dominate. For example, a neutral amino acid typically carries a positive charge at pH<2 while the rutile surface is also positively charged. Similar to other neutral amino acids, the DOPA adsorption is reduced when the ionized states are favored at extreme pH values. However, DOPA adsorption varies with pH even though the molecule is entirely in its neutral state. Because of its association with the surface sites (i.e., TiOH), the sites in protonated ($TiOH_2^+$) or deprotonated (TiO^-) states will impact DOPA adsorption.

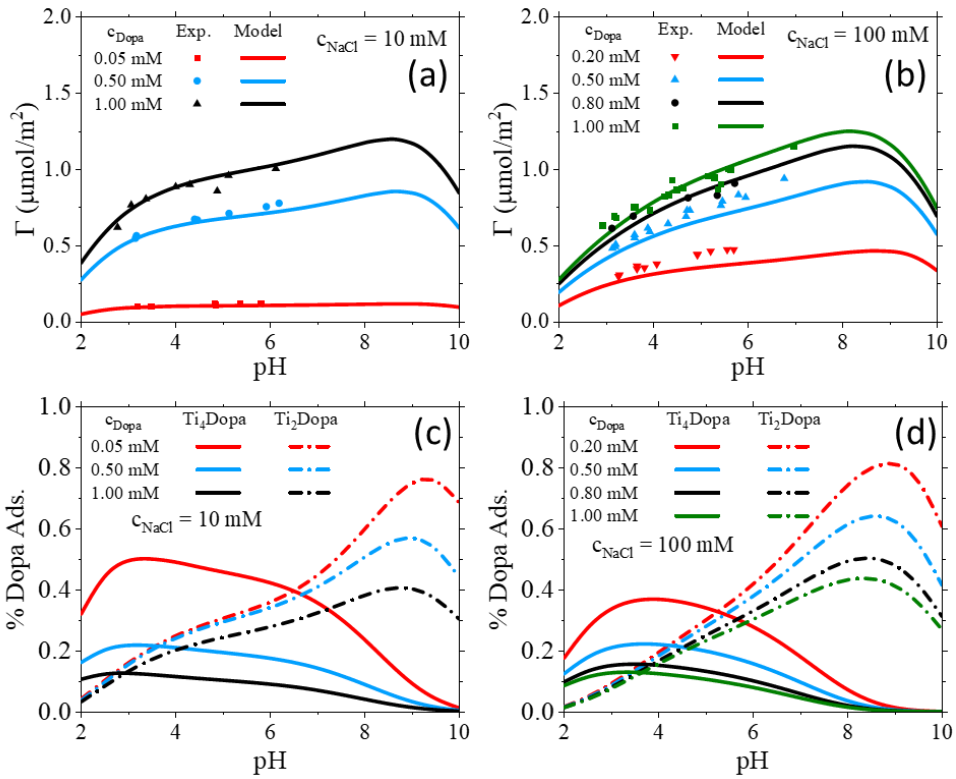

Figure 4.9 The adsorption of dihydroxyphenylalanine (DOPA) at the rutile surface versus pH in (a) 10 mM and (b) 100 mM NaCl solutions. Different lines are theoretical predictions corresponding to various total DOPA concentrations, the symbols are from experiment[47]. The speciation of DOPA at the rutile surface as a function of pH predicted according to the thermodynamic model. Ti_nDOPA refers to the binding of DOPA to n titanium surface sites as given by Eqs. (21) and (22).

To understand further why DOPA adsorption increases as the pH rises, we may consider the distribution of adsorbed DOPA according to the surface configurations:

$$n_{DP}^{ads} = n_{(>\mathrm{TiOH})>\mathrm{Ti_3DP}} + n_{(>\mathrm{TiOH_2^+})>\mathrm{TiHDP^-}}. \tag{4.25}$$

Based off the conditions of chemical equilibrium as shown in Eqs. (21) and (22), we may relate the surface concentrations of DOPA in different configurations to the number density of surface hydroxyl groups:

$$n_{DP}^{ads} = \mathrm{K}_1' n_{\mathrm{TiOH}}^4 c_{\mathrm{DP}} + \mathrm{K}_2' \exp\left[-\beta e \Delta\psi_{r,2}\right] n_{\mathrm{TiOH}}^2 c_{DP} \tag{4.26}$$

where indices 1 and 2 refer to the chemical reactions involving four and two surface sites, respectively, and the primes denote the apparent equilibrium constants, i.e., after the conversions from activity to the number density or concentration. Eq. (26) indicates that DOPA adsorption is intrinsically related to the availability of surface hydroxyl sites and the DOPA concentration at the surface. As the number density of surface hydroxyl sites decreases as the pH deviates from the isoelectric point, we expect that the DOPA adsorption will be maximized near the isoelectric point. Meanwhile, the exponential factor in the second term on the right side of Eq. (26) explains why DOPA adsorption increases with pH beyond the isoelectric point. When the surface is positively charged, $\Delta\psi_{r,2}$ is greater than zero; and it is less than zero when the surface is negatively charged, i.e., the surface potential is always greater in magnitude than that at the adsorption plane of DOPA. Because the magnitude of $\Delta\psi_{r,2}$ increases as the pH deviates more from the isoelectric point, the exponential term approaches unity from below as the pH increases towards the isoelectric point. Further increase in the pH will increase the absolute value of the

exponential factor and thus improve the adsorption of DOPA. Therefore, we conclude that the increase in DOPA adsorption with pH results from the favorable electrostatic work due to the complexation of the DOPA molecule with the surface. The importance of electrostatic interactions in the binding of a neutral DOPA molecule with the rutile surface explains why the salt concentration plays a significant role in the adsorption.

Figure 9c and d present the fraction of DOPA in two surface configurations. The electrostatic work is a contributing factor for DOPA transition from lying flat to standing up as the pH rises. The increase in salt concentration tends to shift the transition between the two surface configurations to a lower pH due to electrostatic screening. Another major factor in determining the surface configuration of DOPA is its local number density. When the surface concentration of DOPA increases, it lowers the pH at which the transition takes place due to the competition of DOPA molecules with each other at the surface. At small surface concentrations, the competition for the surface sites is less important, which explains why the DOPA molecules are mostly lying flat at the surface. Clearly, the surface complexation model provides a satisfactory description of DOPA adsorption to the titanium surface and valuable insight into the mechanisms of DOPA attachment to inorganic surfaces in an aqueous solution.

4.4 Conclusion

There have been substantial interests in understanding the interaction of polypeptides and proteins with inorganic surfaces due to their applications to the design of bioadhesives and the remediation of biofouling. In the present study, we have initiated a theoretical work for such possibilities by employing a coarse-grained model for amino

acids that captures their adsorption at an inorganic surface under different solution conditions. By leveraging our previous work to account for the key physics governing charge regulation and the interaction of amino acids with ionic species in aqueous solutions, we developed a molecular-thermodynamic framework that is able to predict the adsorption of amino acids on inorganic surfaces driven by either electrostatic binding or surface complexation. The thermodynamic model integrates ionization equilibrium for both amino acids and inorganic surfaces with the classical density functional theory (cDFT) that facilitates an accurate description of the inhomogeneous distributions of amino acids and ionic species near inorganic surfaces. Importantly, this model accounts for both chemical and physical interactions between amino acids and an inorganic surface, which is key to capturing the changes in adsorption due to variation in solution conditions.

To demonstrate the effectiveness of this molecular thermodynamic model, we compare theoretical predictions directly with experimental adsorption data for different charge types (i.e., acidic, basic, and neutral) of amino acids on a rutile surface. We first showed that the deprotonation and protonation of hydroxyl sites at the inorganic surface can be well described by coupling the chemical equilibrium of the surface sites with cDFT calculation for ion distributions. We then investigated the adsorption behavior of different amino acids at the rutile surface in aqueous solutions that are varied in pH, amino acid concentration, salt type and concentration. We found that amino-acid adsorption typically shows a maximum an intermediate pH value that is dependent upon the salt type and concentration. While the increase in salt concentration promotes the ionization of the surface sites, it also weakens the electrostatic interaction between amino acid and the

surface. Therefore, the addition of salt reduces the adsorption of amino acid and shifts the pH to a higher value where the maximum adsorption takes place. For acidic amino acids, the non-electrostatic surface binding makes a significant contribution to the adsorption even when the electrostatic charges of the amino acid and the surface are of the same sign. On the other hand, a basic amino acid like lysine shows much weaker dependence on non-electrostatic surface binding and the adsorption is mostly driven by electrostatic attraction from the surface. A maximum adsorption takes place at high pH when the amino acid and surface are oppositely charged. Lastly, we considered the adsorption of DOPA, a neutral amino acid, to the rutile surface through a combination with the surface complexation model that accounts for the two configurations of DOPA molecules at the rutile surface.

An accurate description of the charge regulation for both amino acids and the underlying surface in a highly inhomogeneous environment plays an important role in understanding the adsorption behavior of amino acids, particularly the acidic and basic amino acids, to inorganic surfaces. The interaction between amino acids and the surface also affects equilibrium between different charged states leading to the speciation amino acid molecules significantly different from that in the bulk solution. Because of the shift in speciation, the amino acids adsorbed at a highly charged surface may exist in an ionized state of opposite charge of the surface while those in the bulk solution are entirely neutral. Although the molecular-thermodynamic model employs a number of semi-empirical parameters, it provides a predictive description of thermodynamic non-idealities that are relevant to describe the environmental effects on both amino-acid adsorption and chemical equilibrium. In the future, we plan to extend this molecular-thermodynamic framework to

describing the adsorption of polypeptides and flexible proteins which are of keen interest for practical applications.

Bibliography

1. Balkenende, D.W.R., S.M. Winkler, and P.B. Messersmith, *Marine-inspired polymers in medical adhesion.* European Polymer Journal, 2019. **116**: p. 134-143.

2. Adamczyk, Z., *Protein adsorption: A quest for a universal mechanism.* Current Opinion in Colloid & Interface Science, 2019. **41**: p. 50-65.

3. Mahmoudi, M., *Debugging Nano–Bio Interfaces: Systematic Strategies to Accelerate Clinical Translation of Nanotechnologies.* Trends in Biotechnology, 2018. **36**(8): p. 755-769.

4. Schwaminger, S., et al., *Experimental characterization and simulation of amino acid and peptide interactions with inorganic materials.* Engineering in Life Sciences, 2018. **18**(2): p. 84-100.

5. Costa, D., L. Savio, and C.M. Pradier, *Adsorption of Amino Acids and Peptides on Metal and Oxide Surfaces in Water Environment: A Synthetic and Prospective Review.* The Journal of Physical Chemistry B, 2016. **120**(29): p. 7039-7052.

6. Kitadai, N., T. Yokoyama, and S. Nakashima, *ATR-IR spectroscopic study of L-lysine adsorption on amorphous silica.* Journal of Colloid and Interface Science, 2009. **329**(1): p. 31-37.

7. Lambert, J.F., *Adsorption and polymerization of amino acids on mineral surfaces: a review.* Orig Life Evol Biosph, 2008. **38**(3): p. 211-42.

8. Shchelokov, A., et al., *Adsorption of Native Amino Acids on Nanocrystalline TiO2: Physical Chemistry, QSPR, and Theoretical Modeling.* Langmuir, 2019. **35**(2): p. 538-550.

9. Greiner, E., et al., *Adsorption of l-glutamic acid and l-aspartic acid to γ-Al2O3.* Geochimica et Cosmochimica Acta, 2014. **133**: p. 142-155.

10. Begonja, S., et al., *Adsorption of cysteine on TiO2 at different pH values: Surface complexes characterization by FTIR-ATR and Langmuir isotherms analysis.* Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2012. **403**: p. 114-120.

11. El Shafei, G.M.S. and N.A. Moussa, *Adsorption of Some Essential Amino Acids on Hydroxyapatite.* Journal of Colloid and Interface Science, 2001. **238**(1): p. 160-166.

12. Imamura, K., et al., *Adsorption Behavior of Amino Acids on a Stainless Steel Surface.* Journal of Colloid and Interface Science, 2000. **229**(1): p. 237-246.

13. O'Connor, A.J., et al., *Amino acid adsorption onto mesoporous silica molecular sieves.* Separation and Purification Technology, 2006. **48**(2): p. 197-201.

14. Horányi, G., *Adsorption of primary amino compounds at platinum electrodes: A survey of radiotracer studies.* Electrochimica Acta, 1990. **35**(6): p. 919-928.

15. Hughes, Z.E. and T.R. Walsh, *What makes a good graphene-binding peptide? Adsorption of amino acids and peptides at aqueous graphene interfaces.* Journal of Materials Chemistry B, 2015. **3**(16): p. 3211-3221.

16. Dasetty, S., J.K. Barrows, and S. Sarupria, *Adsorption of amino acids on graphene: assessment of current force fields.* Soft Matter, 2019. **15**(11): p. 2359-2372.

17. Hoefling, M., et al., *Interaction of amino acids with the Au(111) surface: adsorption free energies from molecular dynamics simulations.* Langmuir, 2010. **26**(11): p. 8347-51.

18. Vlasova, N. and L. Golovkova, *The adsorption of amino acids on the surface of highly dispersed silica.* Colloid Journal, 2004. **66**(6): p. 657-662.

19. Jonsson, C.M., et al., *Attachment of l-Glutamate to Rutile (α-TiO2): A Potentiometric, Adsorption, and Surface Complexation Study*. Langmuir, 2009. **25**(20): p. 12127-12135.

20. Jonsson, C.M., et al., *Adsorption of l-aspartate to rutile (α-TiO2): Experimental and theoretical surface complexation studies.* Geochimica et Cosmochimica Acta, 2010. **74**(8): p. 2356-2367.

21. Gallegos, A. and J. Wu, *Charge Regulation of Natural Amino Acids in Aqueous Solutions.* Journal of Chemical & Engineering Data, 2020. **65**(12): p. 5630-5642.

22. Kong, X., et al., *Molecular Theory for Electrokinetic Transport in pH-Regulated Nanochannels.* The Journal of Physical Chemistry Letters, 2014. **5**(17): p. 3015-3020.

23. Ong, G.M.C., A. Gallegos, and J. Wu, *Modeling Surface Charge Regulation of Colloidal Particles in Aqueous Solutions.* Langmuir, 2020. **36**(40): p. 11918-11928.

24. Yang, J., et al., *Understanding surface charge regulation in silica nanopores.* Physical Chemistry Chemical Physics, 2020. **22**(27): p. 15373-15380.

25. Sverjensky, D.A. and K. Fukushi, *A predictive model (ETLM) for As(III) adsorption and surface speciation on oxides consistent with spectroscopic data.* Geochimica et Cosmochimica Acta, 2006. **70**(15): p. 3778-3802.

26. Sverjensky, D.A. and K. Fukushi, *Anion Adsorption on Oxide Surfaces: Inclusion of the Water Dipole in Modeling the Electrostatics of Ligand Exchange.* Environmental Science & Technology, 2006. **40**(1): p. 263-271.

27. Tamura, H., et al., *Mechanism of Hydroxylation of Metal Oxide Surfaces.* Journal of Colloid and Interface Science, 2001. **243**(1): p. 202-207.

28. Akratopulu, K.C., C. Kordulis, and A. Lycourghiotis, *Effect of temperature on the point of zero charge and surface charge of TiO2.* Journal of the Chemical Society, Faraday Transactions, 1990. **86**(20): p. 3437-3440.

29. Loux, N.T., *Extending the diffuse layer model of surface acidity behaviour: III. Estimating bound site activity coefficients.* Chemical Speciation & Bioavailability, 2009. **21**(4): p. 233-244.

30. Gómez-Bombarelli, R., et al., *Computational Calculation of Equilibrium Constants: Addition to Carbonyl Compounds.* The Journal of Physical Chemistry A, 2009. **113**(42): p. 11423-11428.

31. Mansoori, G.A., et al., *Equilibrium Thermodynamic Properties of the Mixture of Hard Spheres.* The Journal of Chemical Physics, 1971. **54**(4): p. 1523-1525.

32. Maribo-Mogensen, B., G.M. Kontogeorgis, and K. Thomsen, *Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions.* Industrial & Engineering Chemistry Research, 2012. **51**(14): p. 5353-5363.

33. Jin, Z., Y. Tang, and J. Wu, *A perturbative density functional theory for square-well fluids.* The Journal of Chemical Physics, 2011. **134**(17): p. 174702.

34. Jang, H.M. and D.W. Fuerstenau, *The specific adsorption of alkaline-earth cations at the rutile/water interface.* Colloids and Surfaces, 1986. **21**: p. 235-257.

35. Kosmulski, M., *Isoelectric points and points of zero charge of metal (hydr)oxides: 50years after Parks' review.* Advances in Colloid and Interface Science, 2016. **238**: p. 1-61.

36. Lee, N., D.A. Sverjensky, and R.M. Hazen, *Cooperative and Competitive Adsorption of Amino Acids with Ca2+ on Rutile (α-TiO2).* Environmental Science & Technology, 2014. **48**(16): p. 9358-9365.

37. Misra, R.P., et al., *Theory of Surface Forces in Multivalent Electrolytes.* Langmuir, 2019. **35**(35): p. 11550-11565.

38. Imamura, K., et al., *Contribution of acidic amino residues to the adsorption of peptides onto a stainless steel surface.* Journal of Colloid and Interface Science, 2003. **267**(2): p. 294-301.

39. Nagayasu, T., et al., *Effects of carboxyl groups on the adsorption behavior of low-molecular-weight substances on a stainless steel surface.* Journal of Colloid and Interface Science, 2004. **279**(2): p. 296-306.

40. De Stefano, C., et al., *The interaction of amino acids with the major constituents of natural waters at different ionic strengths.* Marine Chemistry, 2000. **72**(1): p. 61-76.

41. de Vos, W.M. and S. Lindhoud, *Overcharging and charge inversion: Finding the correct explanation(s).* Advances in Colloid and Interface Science, 2019. **274**: p. 102040.

42. Waite, J.H., *Mussel adhesion – essential footwork.* The Journal of Experimental Biology, 2017. **220**(4): p. 517.

43. Stewart, R.J., T.C. Ransom, and V. Hlady, *Natural underwater adhesives.* Journal of Polymer Science Part B: Polymer Physics, 2011. **49**(11): p. 757-771.

44. Kord Forooshani, P. and B.P. Lee, *Recent approaches in designing bioadhesive materials inspired by mussel adhesive protein.* Journal of polymer science. Part A, Polymer chemistry, 2017. **55**(1): p. 9-33.

45. Moulay, S., *Recent Trends in Mussel-Inspired Catechol-Containing Polymers. Part 1 (A Review.* 2018.

46. Gulley-Stahl, H., et al., *Surface complexation of catechol to metal oxides: an ATR-FTIR, adsorption, and dissolution study.* Environmental science & technology, 2010. **44**(11): p. 4116-4121.

47. Bahri, S., et al., *Adsorption and Surface Complexation Study of L-DOPA on Rutile (α-TiO2) in NaCl Solutions.* Environmental Science & Technology, 2011. **45**(9): p. 3959-3966.

48. Krohn, J.E. and M. Tsapatsis, *Phenylalanine and Arginine Adsorption in Zeolites X, Y, and β.* Langmuir, 2006. **22**(22): p. 9350-9356.

Chapter 5. Thermodynamic non-ideality in charge regulation of weak polyelectrolytes

Polymer ionization differs from that for monomeric counterparts due to intramolecular correlations. Such effects are conventionally described in terms of the site-binding model that accounts for short-range interactions between neighboring sites. With an apparent equilibrium constant for each ionizable group and the nearest-neighbor energy as adjustable parameters, the site-binding method is useful to correlate experimental titration curves when the long-ranged site–site interactions are insignificant. This work aims to describe the electrostatic behavior of weak polyelectrolytes in aqueous solutions on the basis of the intrinsic equilibrium constants of the individual ionizable groups and the solution conditions underlying thermodynamic non-ideality. A molecular thermodynamic model is proposed for the protonation of weak polyelectrolytes by incorporating the classical density functional theory into the site-binding model to account for the effects of the local ionic environment on both inter- and intra-chain correlations. By an extensive comparison of theoretical predictions with experimental titration curves, we demonstrate that the thermodynamic model is able to quantify the ionization behavior of weak polyelectrolytes over a broad range of molecular architecture and solution conditions.

5.1 Introduction

The functionality of polyelectrolytes is closely affiliated with electrostatic interactions. In contrast to strong polyelectrolytes, weak polyelectrolytes present ionizable

groups with the degree of ionization in accordance with the solution pH and the local chemical environment. The pH-regulated charge behavior is advantageous for applications in "smart" polymeric systems to achieve specific functionality such as targeted drug delivery and controlled release in response to the pH variation of intracellular fluids[1, 2]. Such behavior can also be utilized in wastewater treatment[3], thermal energy storage[4], solid-liquid separation[5], and underwater adhesion[6, 7], to name only a few. Clearly, electrostatic interactions play a central role in diverse applications of weak polyelectrolyte systems.

The dependence of site ionization on solution conditions makes electrostatic correlation in weak polyelectrolyte systems more complex than that in strong polyelectrolytes. For weak polyelectrolytes, charge regulation is significantly different from the ionization of monomeric compounds because it is coupled not only with interactions with other chemical species in the solution but also with the ionization of neighboring sites[8]. For weak homo-polyelectrolytes such as polyacids and polybases, the ionization is hindered by the Coulombic repulsion among neighboring charged monomers[9]. On the other hand, polyampholytes are more easily ionized than their monomeric counterparts due to favorable attraction between neighboring acidic and basic groups because of the opposite charge[10]. For many weak polyelectrolytes, the titration curve exhibits a step-like shape caused by the sequential ionization of monomers along the polymer chain. In general, the degree of ionization is lower than that for their monomeric counterparts due to intra-chain correlations. For example, poly(maleic acid), PMA, exhibits a two-step titration behavior resulting from strong interactions among its high-

density carboxyl groups along the polymer backbone[11]. At moderate to high salt concentrations, the ionization takes place at alternating sites and must overcome strong electrostatic repulsion with the neighboring charged groups[8]. In weak polyelectrolyte solutions, the charge regulation depends not only on intra-chain correlations but also on intermolecular interactions due to the presence of other chemical species. In that case, we need a thermodynamic model to account for the solution condition such as salt concentration and the chemical details of individual ionic species.

Theoretical modeling of weak polyelectrolytes has a long history and varies greatly in term of complexity[12, 13]. Recent studies of weak polyelectrolyte titration are mostly based on Monte Carlo (MC) simulations, typically within the framework of a coarse-grained representation of the polymer chains[14-17]. For example, Laguecir et al. investigated the ionization of linear poly(acrylic acid) using a freely-jointed hard-sphere-chain model along with a screened Debye-Hückel potential for electrostatic interactions between charged segments[18]. Excellent agreement was achieved in comparison with experimental titration curves for the polyacid in dilute electrolyte solutions. These authors also demonstrated that, at the same pH, the degree of ionization is reduced as the molecular weight (MW) increases due to the long-range correlation effects. A similar model was used by Uhlík et al. to study the ionization behavior of star-like weak polyelectrolytes in a salt-free solution[19]. It was found that, when the number of arms varies from 2 to 10 while fixing the length of each arm at 100 segments, the overall ionization of the polymer is reduced as the number of arms increases due to the increased monomer density near the molecular center. In comparison with analytical methods, one major advantage of MC

simulation is that, given a polymer model, it is able to account for both the intra- and inter-molecular correlations exactly. However, molecular simulation of weak polyelectrolyte systems typically avoids an explicit consideration of electrostatic interactions with protons or hydroxyl ions because the inclusion of such effects would significantly increase the computational cost. By contrast, mean-field methods are able to explicitly account for all ionic species in the solution and computationally more efficient than simulation[20, 21]. Recently, Uhlík et al. compared the predictions of the self-consistent-field theory (SCFT) with simulation data[19]. While SCFT provides a reasonable description of the configuration of weak polyelectrolyte chains, it fails to capture the titration curves for star-like weak polyelectrolytes. Some success was found by the application of SCFT to the experimental data for linear poly(acrylic acid); however, it required the consideration of a two-site model with one site containing one carboxyl group and the other site containing two carboxyl groups (divalent when fully deprotonated)[22].

In this work, we develop a liquid-state method to describe the titration of weak polyelectrolytes by incorporating the classical density functional theory (cDFT) with the conventional site-binding model[23]. By accounting for the free energy of ionization for each segment in a local solution environment, cDFT enables a faithful description of the salt effects on weak polyelectrolyte charging that are typically ignored in conventional methods. With the intrinsic protonation/deprotonation constant as the input, which can be fixed from fitting with experimental data for monomeric systems or predicted by first-principles methods, the thermodynamic model provides a quantitative description of the degree of ionization for individual segments in weak polyelectrolyte systems with arbitrary

chain topologies. We will present the theoretical details and numerical results in the reminder of this article as follows. We first outline the conventional site-binding model for describing the titration of weak polyelectrolyte chains. Next, cDFT is introduced to account for the effect of local solution environment on the ionization of individual segments by using a coarse-grained model for polymers and electrolyte solutions. The numerical results are then discussed in comparison with experimental titration curves for linear and branched polyelectrolytes. After demonstrating the theoretical capability of the molecular thermodynamic model for diverse weak polyelectrolyte systems, we summarize the main results and offer perspectives on possible further developments and new applications.

5.2 Thermodynamic models and methods

5.2.1 Site-binding model

The site-binding model is built upon the Ising chain in statistical mechanics. It describes the ionization of weak polyelectrolytes in terms of the one-body potential and nearest-neighbor interactions[8, 24]. Within this model, the ionizable groups are treated as proton-binding sites with their spatial arrangements along the polymer backbone dependent upon the polymer architecture (e.g., linear or branched). Each site can exist in charged or neutral state with the difference in energy depending on the solution pH and the apparent protonation constant (*viz.*, pK_i'). The interaction between neighboring sites can be represented by a screened Coulomb potential that partially accounts for the solution conditions[8]. The site-binding model performs well to correlate experimental titration curves with the proton binding constant and nearest-neighbor interaction energy treated as

adjustable parameters. One major limitation of the site-binding model is that it is not able to account for polymer conformations. Theoretical improvements are mostly focused on site-site interactions beyond the immediate neighbors[25]. However, less attention has been directed at the thermodynamic effects on both the one-body potential and intra-chain correlations. In Section 2.2, we will outline a method to account for this thermodynamic non-ideality due to changes in the inter- and intramolecular correlations with salt concentration. We first begin with an outline of the original site-binding model and its extension to arbitrary polymer architecture.

Formally, the Hamiltonian for the site-binding model is a quadratic function of state variable for the ionizable sites

$$\Gamma\left(s_1,...,s_M\right)=\sum_{i}^{M}\Delta\mu_i s_i+\sum_{i}^{M-1}W_{i,i+1}^{P}s_i s_{i+1} \tag{5.1}$$

where s_i stands for the microstate of ionizable site i, and M is the total number of ionizable sites from each weak polyelectrolyte chain. $s_i = 0$ if site i is not charged and $s_i = 1$ otherwise. The first term on the right side of Eq. (1), $\Delta\mu_i$, accounts for the reversible work to convert a monomer from the neutral state to the charged state. The ionization can be achieved by either adding a proton to a basic site or removing a proton from an acidic site. Approximately, the change in the intrinsic free energy, i.e., the free energy of a single site at the single-chain reference state, is the same as that for the monomer at infinite dilution

$$\beta\Delta\mu_i=\beta\mu_i\left(s_i=1\right)-\beta\mu_i\left(s_i=0\right)=\begin{cases}\left[\mathrm{pK}_i'-\mathrm{pH}\right]\ln 10, & \text{for an acid}\\ \left[\mathrm{pH}-\mathrm{pK}_i'\right]\ln 10, & \text{for a base}\end{cases} \tag{5.2}$$

where pK_i' is the negative logarithm of the apparent protonation constant for segment *i*, K_i', $pH = -\log a_{H^+}$ with a_{H^+} being the proton activity, and $\beta = 1/k_B T$ where k_B is the Boltzmann constant, and *T* is the absolute temperature. In the conventional site-binding model, the apparent equilibrium constant depends not only on the ionization reactions of individual groups but also on the local solution condition. It should be clarified that the apparent protonation constant is for that of an individual segment and not for the pH at which the polymer is half-charged as it is typically referred to in experiments. In the latter case, the polymer architecture and inter- and intramolecular correlations dictate the pH at which the polymer is half-charged. Such monomer-level information is not accessible experimentally. The second term on the right side of Eq. (1), $W_{i,i+1}^{p}$, takes into account an additional work due to the electrostatic (and non-electrostatic) correlations between neighboring sites. This term differentiates the titration of weak polyelectrolytes from that for their monomeric counterparts. Owing to the electrostatic screening effects, the strength of the electrostatic interaction will vary with the salt concentration. On the other hand, non-electrostatic interactions, which can result from hydrogen bonding, electron sharing, etc., are less sensitive to the salt concentration. The relative strength of the short- and long-range interactions is dependent on the distance between neighboring sites. The formulation of the site-binding model in this manner provides a clear incorporation of the intermolecular correlations through $\Delta\mu_i$ and the intramolecular correlations through $W_{i,i+1}^{p}$

.

The degree of ionization is obtained from the thermal average of all possible charge states of the ionizable groups based on the semi-grand canonical partition function

$$\Xi = \sum_{s_1,\ldots,s_M} e^{-\beta\Gamma(s_1,\ldots,s_M)} \quad . \tag{5.3}$$

The summation extends to all combinations of uncharged and charged sites on the weak polyelectrolyte chain. For each ionizable site, the degree of ionization can be determined by the ensemble average:

$$\alpha_k = \langle s_k \rangle = -\frac{\partial \ln \Xi}{\partial \beta \mu_k} = \frac{\sum_{\{s_{i\neq k}\}} e^{-\beta\Gamma(\{s_i\},s_k=1)}}{\sum_{\{s_i\}} e^{-\beta\Gamma(\{s_i\})}} \quad . \tag{5.4}$$

For a weak homo-polyelectrolyte with M identical ionizable sites, the average degree of ionization is calculated from $\alpha = \sum_k \alpha_k / M$.

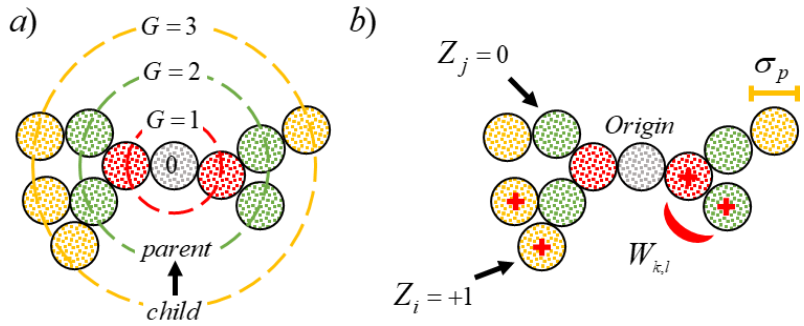

Figure 5.1 (a) Schematic of polymer architecture where the polymer 'originates' from a segment denoted origin. Each generation refers to the number of parent-child relations between the segment and the origin. The origin is designated as generation 0. Any segment without any children is referred to as an endling. (b) Schematic of a weak polymer represented by freely jointed hard spheres that can be in either neutral of charged states of valence Z=0 and Z=+1 for a basic monomer (or Z=-1 for an acidic monomer), respectively. Here, W stands for the potential of mean force between neighboring segments and σ_p is the size of the monomer.

In principle, Eq. (4) can be evaluated by enumerating all possible combinations of the charge states or by applying MC simulation methods. Whereas such procedures are straightforward for short polymers, direct enumeration of all possible charge states of ionizable groups can quickly become computationally prohibitive as the chain length increase because the number of charge states scales with 2^M. From an analytical perspective, the transfer matrix technique provides an elegant approach to evaluating the partition function. For linear polyacids or polybases, analytical expressions can be derived to describe the ionization behavior[8]. The method can be generalized for all architectures of weak polyelectrolytes by organizing the polymer as 'originating' from a specific segment as shown in Figure 1. In this model, we first assign a polymer segment (no special requirement on this segment) as the *origin (*i.e., the segment in which all other segments branch from). Then, any segment immediately connected to the origin segment are specified as generation 1 (i.e., the child of generation 0). Thus, the origin segment is also the parent to these segments in generation 1. If the segments in generation 1 have children, then they will be the parents to the segments in generation 3. However, if the parent-child relation ends (i.e., there is no child for a particular generation path), then this last segment is denoted as an *endling* and is the child of the previous generation. A segment is at a *generation G* if the parent-child relation transfers *G* times between this segment and origin. For consistency, we define the generation of the origin as zero. The key point of this distinction is that we can find one and only one path from the origin to the segment of interest, or from the segment of interest to the origin.

If the ionizable sites in a weak polyelectrolyte are arranged in a parent-child structure, a *restricted* partition function can be defined for parent *p* with *k* children each with their own restricted partition functions[26]

$$\Xi'_p\left(s_p\right) = z_p^{s_p} \prod_{k=\text{child}(p)} \left[\sum_{\{s_k\}} u_{pk}^{s_p s_k} \Xi'_k\left(s_k\right) \right]. \tag{5.5}$$

Eq. (5) corresponds to the partition function of the parent with a fixed (restricted) value of the state variable s_p for the parent segment, where z_p is the reduced activity defined as $z_p = \mathrm{K}'_p a_{H^+}$ and $u_{pk} = \exp[-\beta W_{pk}]$. The product considers the contribution of different children, and the summation is over the different possible states of the child *k*. Recursive application of the *restricted* partition function to the entire polymer leads to an exact partition function for weak polyelectrolytes of arbitrary tree-type architecture. To carry out such calculations, we introduce variable $\tau_p = \Xi'_p(1) / \Xi'_p(0)$ and the recursion relation[27]

$$\tau_p = \frac{\Xi'_p(1)}{\Xi'_p(0)} = z_p \prod_{k=child(p)} \frac{\left[\Xi'_k(0) + u_{pk}\Xi'_k(1)\right]}{\left[\Xi'_k(0) + \Xi'_k(1)\right]} = z_p \prod_{k=child(p)} \frac{\left[1 + u_{pk}\tau_k\right]}{\left[1+\tau_k\right]} \ . \tag{5.6}$$

The total partition function for the entire polymer can then be written in a recursive manner by designating a specific segment as the origin, *o*, (and is equivalent to the partition function of the origin):

$$\Xi = (1+\tau_o) \prod_{k=\text{child}(o)} \Xi_k \tag{5.7}$$

where Ξ_k is the partition function for the child k who is parent to its children *j*:

$$\Xi_k = (1+\tau_k) \prod_{j=\text{child}(k)} \Xi_j \tag{5.8}$$

The average number of segments that are ionized, $\nu_r = \sum_i^M \alpha_i$, can be found from the application of Eq. (4) to Eq. (7) which leads to the following recursive relation[27]

$$\nu_i = \frac{b_i \tau_i}{1+\tau_i} + \sum_{k=\text{child}(i)} \nu_k \tag{5.9}$$

where

$$b_i = \frac{z_i}{\tau_i}\frac{\partial \tau_i}{\partial z_i} = 1 + \sum_{k=\text{child}(i)} \frac{(u_{ik}-1)\tau_k b_k}{(1+u_{ik}\tau_k)(1+\tau_k)} \ . \tag{5.10}$$

The recursion relation is carried out from every endling to the origin generation by generation, which means that we must first calculate the quantities at maximum *generation G*, and then calculate these at levels *G*-1, *G*-2, …, down to generation 1. The variable τ_i and b_i are initialized to z_i and 1, respectively, for all endling segments. This results in a straightforward and fast procedure to determine the degree of ionization for any weak polyelectrolyte.

Parent-child resummation is most useful for polymers with highly branched structures. While this technique is applicable to polymers of any architecture that can be arranged such that there is only one unique parent to a child (i.e., no loops), it is not necessary for linear chains such as poly(acrylic acid) because an analytical expression for the partition function is available. When stereochemistry must be considered and the connection types (i.e., mesomeric and racemic) are chosen according to a prescribed

probability distribution, we can also use the parent-child resummation method with one of the end segments of the linear chain assigned as the origin.

5.2.2 Molecular thermodynamic model

As discussed above, the conventional site-binding model accounts for the solution effects on weak polyelectrolyte ionization in terms of two adjustable parameters, i.e., the ionization free energy of an isolated monomer and the nearest-neighbor interaction energy. Whereas the theoretical procedure is convenient for fitting the experimental titration curves, a thermodynamic model is needed to relate these parameters to the solution conditions. Towards that end, we employ the primitive model for polyelectrolyte solutions to account for electrostatic interactions, short-range attractions and molecular excluded volume effects. As shown schematically in Figure 1, each weak polyelectrolyte chain is represented by a freely jointed tangent hard-sphere chain such that each segment is ionizable dependent upon the local solution condition. We assume that the hard-sphere diameter for each segment is not influenced by the ionization status. Unlike typical coarse-grained models used in MC simulation, the salt ions are explicitly considered as charged hard spheres while the solvent is represented by a dielectric background. At room temperature, the dielectric constant of liquid water is set equal to 78.4. The hydrated diameter for the salt ions were taken from Simonin et al.[28]. Because the segment size is assumed to be independent of the ionization status, the calculated results were relatively insensitive to the ion diameters for all systems considered in this work.

Thermodynamic non-ideality plays a significant role in both one- and two-body potentials of the site-binding model. In general, we may decompose the one-body energy

into chemical and physical contributions, i.e., the apparent equilibrium constant can be separated into contributions due to protonation reaction and non-bonding interactions:

$$\mathrm{pK}_\mathrm{i}' = \mathrm{pK}_\mathrm{i}^\mathrm{T} + \frac{1}{k_B T \ln 10}\left[\mu_i^{ex}\left(s_i = 1\right) - \mu_i^{ex}\left(s_i = 0\right)\right] . \tag{5.11}$$

The first term on the right side of Eq. (11) is affiliated with the thermodynamic equilibrium constant for the ionization of an isolated monomer (i.e., independent of charge status of the polymer) at a reference state of infinite dilution. In principle, the thermodynamic equilibrium constant can be predicted from the Gibb's energy of formation for individual monomers by using first-principles methods. However, quantum-mechanical calculations are computationally demanding and the results are sensitive to the theoretical details[29]. Alternatively, the thermodynamic constant can be determined by fitting the thermodynamic model with experimental titration curves for specific polymer systems under consideration.

The second term on the right side of Eq. (11) represents the change in the excess chemical potential of the monomer upon ionization. This term arises from the physical interactions of the ionizable site with all other chemical species in the environment thus changes with the solution condition. Within the thermodynamic framework discussed above, the change in the excess chemical potential can be expressed in terms of the electrostatic correlations (el) and local electric potential (charge)

$$\Delta\mu_i^{ex} = \mu_i^{ex}\left(s_i = 1\right) - \mu_i^{ex}\left(s_i = 0\right) = \Delta\mu_i^{el} + \Delta\mu_i^{charge} . \tag{5.12}$$

Eq. (12) does not include contributions due to volume exclusion because the segment size is assumed to be the same regardless of whether the ionizable site is charged or neutral. The neglect of the size effect is obviously a simplification since the hydration structure may change upon protonation or deprotonation[30]. However, the change in the particle size would have substantial influence on the excluded-volume effects only in concentrated electrolyte solutions[31]. Besides, the hydration of a segment within a polymer chain is most likely not the same as that for an isolated monomer. With the assumption that ionization has negligible excluded-volume effects, the change in excess chemical potential is therefore dominated by electrostatic interactions and correlation effects.

By treating the polymer as ionizable hard-sphere chains, we can derive analytical expressions for both the local and bulk thermodynamic properties of weak polyelectrolytes from the classical density functional theory (cDFT)[32]. The excess chemical potential due to electrostatic correlations results from interactions with other ionizable groups in the same polymer chain and with all charged species in the solution. For each polymer segment, the ionization leads to a non-uniform distribution of free ions and other polymer segments in its surrounding environment. Such effects can be described through the mean-spherical approximation (MSA)[33]. Importantly, MSA reduces to the Debye-Hückel (DH) limiting law at dilute electrolyte concentrations[34].

According to MSA, the change in excess chemical potential due to electrostatic correlations between the ionized monomer and the uncharged monomer is expressed as

$$\beta\Delta\mu_i^{el} = -l_B\left[\frac{Z_i^2\Gamma + 2Z_i\eta\sigma_i}{1+\Gamma\sigma_i}\right] \tag{5.13}$$

Where $l_B = \beta e^2 / 4\pi\varepsilon_0\varepsilon_r$ is the Bjerrum length (7.14 Å for liquid water at room temperature), *e* is the elementary charge, and ε_0 and ε_r are the vacuum permittivity and the dielectric constant of the solvent, respectively; σ_i and Z_i are the hard-sphere diameter and valence of the ionic species, η is a parameter related to the asymmetry of the system (i.e., differences in size and valence of the ionic species), and Γ is the MSA screening parameter. Approximately, Γ is proportional to the square root of the ionic strength similar to that for the Debye screening parameter. As a result of this square root dependency, the importance of electrostatics becomes relevant even at dilute conditions. A full set of MSA equations is given in Supporting Information.

The polymer ionization also leads to a change in the segment excess chemical potential affiliated with the local electric potential, ϕ^{local}. For an uncharged segment, the local electric potential is approximately zero unless it is directly connected with a charged segment. Upon ionization, the local electric potential will be dependent on the segment size and the salt concentration. The change in the excess chemical potential due to the change in the local electric potential is given by

$$\Delta\mu_i^{charge} = eZ_i\Delta\phi_i^{local} = eZ_i\left[\phi_i^{local}(I) - \phi_i^{local}(I \to 0)\right], \tag{5.14}$$

where Z_i is the valence of the charged segment and *I* is the ionic strength $I = \sum_k c_k Z_k^2 / 2$ and c_k is the molar concentration of species *k*. The local electric potential must be taken

relative to that corresponding to the infinite dilution (i.e., $I \to 0$) because it is used as the reference state for the thermodynamic equilibrium constant. The contribution to the change in excess chemical potential is always negative since the electrostatic energy is the largest at infinite dilution. In principle, the local electric potential is also dependent upon the position of the ionizable site within the polymer backbone. Because such considerations would be cumbersome from a computational perspective, we treat all segments identical and assume that the local electric potential is the same as that corresponding to an isolated monomer in the polyelectrolyte system. The latter can be estimated from cDFT calculations by fixing a particle of an absolute charge of *e* and hard-sphere diameter σ at the origin. As demonstrated in our previous work[32], cDFT provides an accurate description of the inhomogeneous distribution of ions in an aqueous environment.

As indicated before, the ionization of weak polyelectrolytes differs significantly from its monomeric counterparts primarily due to the presence of electrostatic interactions between neighboring sites. For homopolymers, the intrachain correlation impedes the charging of ionizable groups in the same polymer chain and gives rise to a step-like titration curve as commonly seen in polyacids[11]. To describe the interaction between adjacent sites, we must account for the extra work to form a pair of charges. This extra work can be separated into non-electrostatic and electrostatic components, $W_{i,i+1}^{P} = u_{i,i+1} + W_{i,i+1}^{el}$. The non-electrostatic component can be estimated by fitting the experimental data across different solution conditions. The non-electrostatic energy is difficult to predict; however, it can be expected to be relatively independent of the salt concentration but may be dependent on the salt type. Thus, it is still unique to the chemical system. This non-

electrostatic energy originates from chemical specific interactions such as the breakdown of hydrogen bonding between neighboring pairs[35] or the stabilization of charge like that seen in PMA[11]. Such effects go beyond the primitive model employed here and allow for coarse-grained inclusion of non-electrostatic interactions.

On the other hand, the electrostatic component can be estimated from the first-order thermodynamic perturbation theory (TPT)[36].

$$\beta W_{i,i+1}^{el}\left(\sigma_{i,i+1}\right) = -\ln g_{i,i+1}^{el}\left(\sigma_{i,i+1}\right) \tag{5.15}$$

where $\sigma_{i,j} = (\sigma_i + \sigma_j)/2$, and $g_{i,j}^{el}$ stands for the electrostatic component of the contact value of the radial distribution function between segments *i* and *j*. The latter is often approximated by the radial distribution function for a reference system of charged hard spheres[37]

$$g_{i,j}^{el}\left(\sigma_{i,j}\right) = \exp\left[-\frac{l_B}{\sigma_{i,j}}\frac{Z_i - \eta\sigma_i^2}{1+\Gamma\sigma_i}\frac{Z_j - \eta\sigma_j^2}{1+\Gamma\sigma_j}\right]. \tag{5.16}$$

At infinite dilution, the radial distribution function reduces to the exponential of the contact value of the electrostatic pair potential between neighboring segments. Intuitively, Eq. (15) can be understood as the electrostatic work to form an ionic bond as represented by the potential of mean force. As the salt concentration increases, two segments of the same charge become less repulsive due to electrostatic screening, leading to an increased contact value for the radial distribution function. It is important to note that TPT takes into account the chain-connectivity effect on ionization only at the nearest-neighbor level. Nevertheless, unlike the original site-binding model (viz., Ising chain) that ignores the chain

conformation completely, the thermodynamic model takes into account the intra-chain correlations through the excess chemical potential of monomers and chain connectivity. Because such correlations are considered at the nearest-neighbor level, our model is not able to describe the conformation effects adequately. The influence of polymer conformation on ionization is an important problem for future development but beyond the scope of this study.

The thermodynamic procedure discussed above allows us to predict the effects of solution conditions on the one-body energy and nearest-neighbor interaction parameters used in the site-binding model. The one-body energy includes a contribution affiliated with the change in the excess chemical potential of each *monomer* due to ionization. This term accounts for the reversible work to place the monomer at the position of ionization from its ideal reference state. In other words, the excess chemical potential arises from the interactions of each ionizable site with all other species in the solution. Because the excess chemical potential neglects the intra-chain correlation among neighboring sites, TPT is used at the nearest-neighbor level to account for the additional work to bring charged monomers together in the same polymer chain. There is no double-counting in the one-body and nearest-neighbor interactions. The modified site-binding model can be readily extended to weak polyelectrolytes with different types of ionizable groups. In comparison with that for a monomer, the ionization of a polymer segment can be either prohibited or promoted in a heterogeneous weak polyelectrolyte because the extra work to ionize adjacent sites is positive for like-charges and negative for unlike-charges. In addition to electrostatic interactions and excluded volume effects, it can be expected that other non-

electrostatic interactions may also be relevant. For example, polyacids such as poly(fumaric acid) exhibit nearest-neighbor interactions that are dependent upon the salt cations in the solution[38]. It was suggested that the variation in the nearest-neighbor energy originates from differences in cation hydration. In addition, the loss of hydrogen bonding mediated by the surrounding water molecules can also be a contributing factor[39]. We expect that, by accounting for the non-ideality in both the electrostatic and non-electrostatic components of the reversible work of ionization, the thermodynamic model will be able to describe the titration behavior of weak polyelectrolytes with molecular parameters independent of the solution conditions.

5.3 Results and discussion

In the following, we provide three case studies to illustrate applications of our thermodynamic model to describe the effects of solution conditions on ionization of weak polyelectrolytes with different molecnvular architectures. We first consider the effects of salt types and ionic strength on ionization of poly(acrylic acid), a linear polyacid with a relatively low density of ionizable groups. In this case, the titration behavior has been well studied and can be described with various mean-field methods owing to the weak interactions between neighboring sites. Next, we consider two polyacids with a higher linear charge density, poly(maleic acid) and poly(fumaric acid), which differ in their stereochemical arrangement of the carboxyl groups. Unlike poly(acrylic acid), these polyacids exhibit step-like titration behavior that cannot be described by mean-field methods[11]. Lastly, we investigate the titration properties of dendrimers at different generations that are composed of titratable amine groups. Normally, the site-binding

description of dendrimers requires many adjustable parameters due to large changes in the local chemical environment from the outer to inner regions of the dendrimer. We find that the application of our thermodynamic model to predicting the titration of weak polyelectrolytes is most successful at moderate to high salt conditions. As low salt concentration, a discrepancy emerges between the theoretical results and experimental data due to interactions beyond immediate neighbors that are not explicitly accounted for in our thermodynamic model.

5.3.1 Poly(acrylic acid)

Poly(acrylic acid) (PAA) is a linear weak polyelectrolyte that shows no apparent step-like ionization in the titration curve. Conventional models are able to describe the titration curves provided that both the binding constant and nearest-neighbor energy are allowed to change with solution conditions. Figure 2 shows the titration curves of PAA (MW=88 kg/mol[20]) in four types of alkali aqueous solutions (lithium, sodium, potassium, and cesium) at different salt concentrations. The ionization behavior is not significantly influenced by the salt type except at the lowest salt concentration (e.g., 5 mM). Except for the case of 5 mM salt concentration, our thermodynamic model successfully reproduces the experimental data for all alkali metals. As the salt concentration increases, the polyacid begins to ionize at a lower pH. Due to the electrostatic screening effect, a narrowed pH window is observed between the neutral and fully charged polymer chain. The changes in the position and shape of the titration curves are reflected by the decrease in both the apparent equilibrium constant and the work to ionize two adjacent sites together as determined from the conventional site-binding model (see Table S1-S4), respectively.

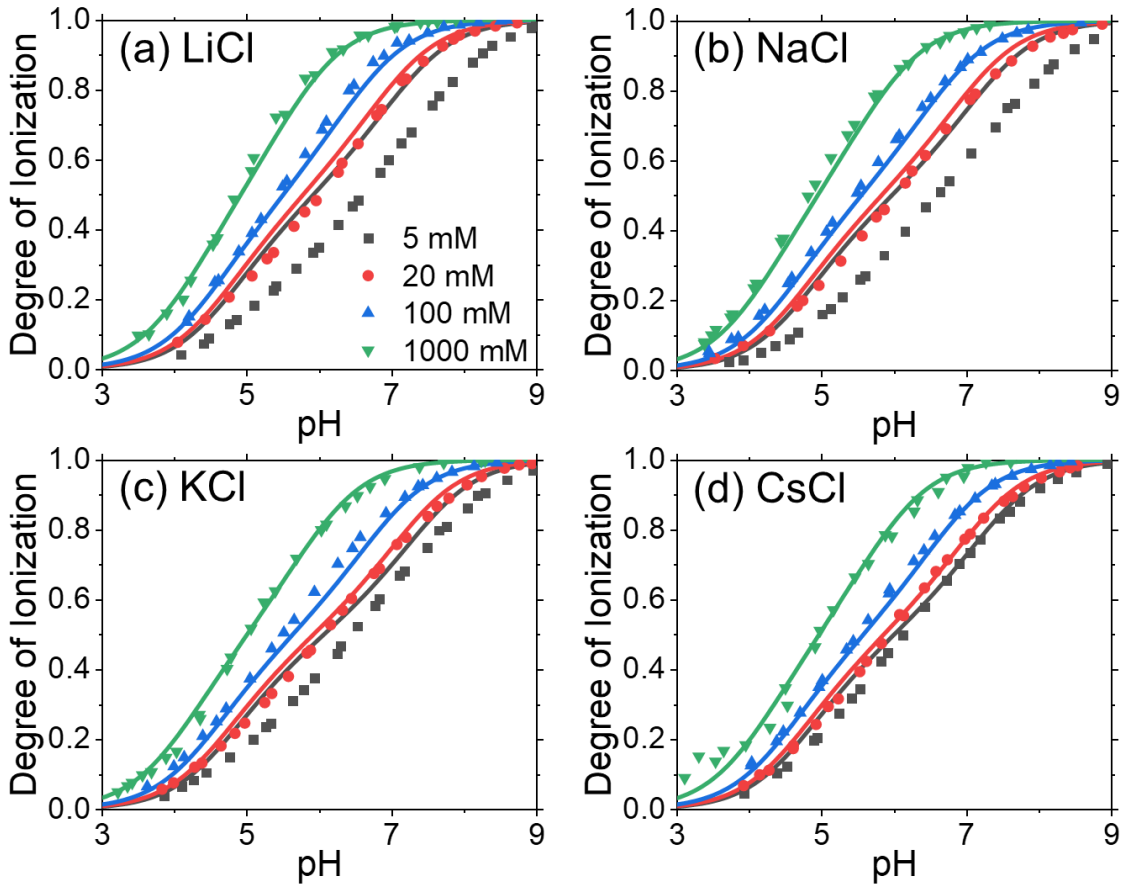

Figure 5.2 Titration curves for poly(acrylic acid) at 25 °C in aqueous solutions of (a) lithium chloride, (b) sodium chloride, (c) potassium chloride, and (d) cesium chloride solutions from experiment[20] (symbols) and theoretical correlations (lines). As marked in panel (a), different symbols stand for different salt concentrations. The theoretical correlations determined a thermodynamic equilibrium constant of 5.22 and the monomer size was 3.56 Å, 3.37 Å, 3.05 Å, and 3.25 Å, respectively, for the alkali salts.

At the lowest salt concentration (5 mM), the large discrepancy between theory and experiment is unsurprising because the site-binding model neglects intra-chain correlations beyond the nearest-neighbor level. When interactions beyond adjacent sites become more significant as the salt concentration falls, the monomer ionization is subject to an additional potential due to the presence of non-adjacent sites. For homopolymers, the degree of ionization is reduced due to the additional repulsion. Since the nearest-neighbor model

neglects long-range intra-chain correlations, it overpredicts the ionization weak polyelectrolytes in comparison with experiment and the deviation is most significant at low salt concentration. With the apparent binding constant and nearest-neighbor energy treated as adjustable parameters, the site-binding model can capture the titration curve at 5 mM salt concentration (see Figures S1-S4). However, the good agreement should be ascribed to the weak correlation between these sites that allows for the overall titration curve to be approximated at the two-body level when using two adjustable parameters. In particular, the initial ionization of the polymer is missed which can be credited to an overestimation of the pK' to compensate for the neglect of interactions between non-neighboring segments. The mean-field approximation neglects the explicit charge of individual sites. Instead, it treats the energy of each site as an average of the overall interactions[20]. By contrast, our coarse-grained model demonstrates the importance of long-range interactions that must be considered for a fully predictive model of the titration curves even at low, but not infinitely dilute, salt concentrations.

In correlation of the experimental titration curves for poly(acrylic acid), we use only the thermodynamic equilibrium constant and the monomer size (i.e., the distance between sites) as the adjustable parameters. Unlike the conventional site-binding model, these parameters are intrinsic properties of the thermodynamic system independent of the solution conditions. It was found that the addition of a non-electrostatic pair interaction did not improve the quality of the fitting results (see Figure S3). For the results shown in Figure 2, the experimental data for the salt concentration of 5 mM were not included in the fitting procedure because our model does not adequately account for intra-chain

correlations between next-nearest neighbors and beyond. The thermodynamic equilibrium constant was constrained to be the same for all alkali metals because it is an intrinsic property of the polymer (i.e., independent on the specific salt present in the solution). The pK_a value, 5.22 ± 0.02, is greater than that corresponding to an isolated monomer, pK_a = 4.756 because of the influence of neighboring segments[40]. The elevated pK_a value agrees with previous quantum-mechanical calculations[41, 42]. The polymerization leads to an increased equilibrium constant for deprotonation, i.e., the ionizable sites are less acidic in the polymer than the corresponding isolated monomeric counterparts. The apparent equilibrium constant shows a stronger dependence on salt concentration than found for simple carboxylic acids (e.g., $\Delta pK' = pK' - pK$ is approximately -0.6 for PAA versus -0.2 for acetic acid[43]). This can likely be attributed to the high local charge of a polymer compared to that of a simple acid such that the screening effect is more significant.

According to our thermodynamic model, the distance between adjacent sites was dependent upon the cation and was found to be 3.56 Å, 3.37 Å, 3.05 Å, and 3.25 Å with a 95% confidence interval of 0.12 Å for lithium, sodium, potassium, and cesium cation, respectively. The salt-type dependence of the distance between the neighboring ionizable groups could be attributed to the differences in hydration of the alkali ions which influences the hydration of the polymer sites when the cations approach the negatively charged monomers in PAA. The specific chemical details that lead to this difference are not captured by the primitive model. For example, there is a noticeable difference between the titration curves for lithium chloride versus cesium chloride at the same salt concentration of 5 mM. One consideration is that even though both cations are monovalent, they differ

significantly in hydrated diameters (4.76 Å versus 1.89 Å)[28]. The larger hydration shell for lithium ions is partially responsible for the larger distance between adjacent sites due to their condensation at the polymer backbone. At high salt concentrations, the specific ion effect becomes less important because the electrostatic interactions are dominated by the overall screening of all ionic species in the system. Further improvement of our model would be possible by adopting an effective potential to distinguish different alkali ions besides size and valence[31]. Nonetheless, our results demonstrate that the incorporation of thermodynamic non-ideality into the nearest-neighbor site-binding framework can correctly account for the titration behavior of poly(acrylic acid) in different aqueous solutions.

5.3.2 Poly(maleic acid) and poly(fumaric acid)

Both poly(maleic acid), PMA, and poly(fumaric acid), PFA, have ionizable groups with a linear density significantly larger than that of poly(acrylic acid). As a result, their titration curves exhibit a step-like behavior which is characteristic for many weak polyelectrolytes. The steps in each titration curve result from alternating ionization of adjacent sites and can be attributed to the Coulombic repulsion and non-electrostatic effects such as hydrogen bonding or electron sharing between functional groups[11, 44]. However, PFA shows significant asymmetry in its titration curve below and above 50% of ionization and a less prominent step-like feature. Interestingly, PMA and PFA differ only in the stereochemical arrangement of their carboxyl groups. For PMA, every second bond is racemic (i.e., neighboring carboxyl groups are on the same side along the C-C bond) and about one-third from the remaining ones (i.e., the odd bonds) are mesomeric (i.e.,

neighboring carboxyl groups are on the opposite side across the C-C bond). By contrast, in PFA every second bond is mesomeric and only about half of the remaining ones are racemic[44, 45]. Due to the closer distance between adjacent sites for a racemic bond compared to that for a mesomeric bond, we expect that, on average, the interaction between neighboring sites in PMA will be stronger than that in PFA. It is important to clarify that whether the odd bond is racemic or mesomeric can only be given in terms of probability. For example, the probability of the odd bond being racemic is 0.655 for PMA and that for PFA is 0.54[38]. Thus, there is no regular pattern in the PMA backbone structure. Previous application of the site-binding model with regularity in the structure found the resulting titration curves to be unsatisfactory[44]. We expect that the poor agreement is due to the absence of stretches in racemic bonds and mesomeric bonds in PMA and PFA, respectively.

The site-binding model can be applied to describe the titration behavior of PMA and PFA by accounting for the different possible connections (i.e., racemic or mesomeric bonds). The assumption of different connections between adjacent sites on the chain means different strengths for the pair interactions, dependent upon if the connection is by a racemic or mesomeric bond. Following the method proposed by Groot et al.[38], we may select the odd bonds to be either racemic or mesomeric by drawing pseudorandom numbers using the propagation rate P_r as a decision threshold level. The degree of dissociation converges after evaluating for approximately 8000 consecutive sites on the chain. Since the two weak polyelectrolytes are different only in the sequence and the numbers of racemic and mesomeric bonds, we assume that the distance between neighboring sites

depends only on the type of bonds but not the polymer. In that case, the same set of model parameters can be used for both PMA and PFA. We obtained these parameters by fitting the thermodynamic model with the titration data at different solution conditions[11, 38, 44].

Unfortunately, due to the use of a hard-sphere chain model, the different distances between racemic and mesomeric bonds introduces an inconsistency. Because the hard-sphere diameter of the individual segments cannot account for this difference, we make the monomer size equal to the distance between two sites connected by a racemic bond. The work to charge two adjacent sites connected by a mesomeric bond is then considered through an analytical approximation by Henderson et al.[46]

$$\begin{aligned}\beta W_{i,i+1}^{el}\left(d_{i,i+1}\right) &= -\ln g_{i,i+1}^{el}\left(d_{i,i+1}\right) \\ &= \frac{l_B}{d_{i,i+1}}\frac{Z_i-\eta\sigma_i^2}{1+\Gamma\sigma_i}\frac{Z_{i+1}-\eta\sigma_{i+1}^2}{1+\Gamma\sigma_{i+1}}\exp\left[-\kappa\left(d_{i,i+1}-\sigma_{i,i+1}\right)\right]\end{aligned} \tag{5.17}$$

where $d_{i,i+1}$ is the distance between two sites connected by a mesomeric bond and $\kappa=\sqrt{8\pi l_B I}$. The pK_a values for PMA and PFA were 4.47 and 4.16 with a 95% confidence interval of ± 0.04, respectively, and the distances between adjacent sites characterizing the racemic and mesomeric bonds were determined to be 0.69 ± 0.01 Å and 1.95 ± 0.06 Å, respectively. Because PMA and PFA have a higher linear density of ionizable sites compared with PAA, the distances between neighboring sites are much smaller than that for PAA (e.g., 3.57 Å for LiCl). At such short distances, the assumption that the Coulomb interaction is governed by the bulk dielectric constant of 78.5 is likely not valid[47]. If a lower dielectric constant were assumed to govern this interaction, the distance parameter

would need to be larger in order to reproduce the same interaction energy. Considering that the description of inhomogeneous dielectric behavior cannot be well described through conventional methods, we use the small distances as a zeroth-order approximation. The lower dielectric constant of the interchain interaction may play a role in PAA as well, but to a lesser degree owing to the larger distance between adjacent sites.

In addition to the electrostatic component of the pair interaction, our thermodynamic model includes a non-electrostatic pair potential of -4.53 ± 0.18 k_BT between adjacent sites connected by each racemic bond in PMA. It is important to note that this energy will not vary with salt concentration and counteracts some of the electrostatic repulsion ($W^{el} \sim +10\ k_BT$). It has been suggested that the nearest-neighbor attraction originates from the local stabilization of charge where a proton can be shared between an uncharged carboxyl group and a neighboring ionized group connected by a racemic bond[35, 48]. Such interaction does not exist between adjacent sites connected by a mesomeric bond since it is sterically impossible. Due to the stereochemistry of PFA, two racemic bonds never occur in a row. On the other hand, the racemic bonds in PMA often appear in series leading to significant charge stabilization by proton sharing. Thus, we see that stereochemistry can make an important difference in the titration behavior of weak polyelectrolytes, despite otherwise sharing the same chemical composition.

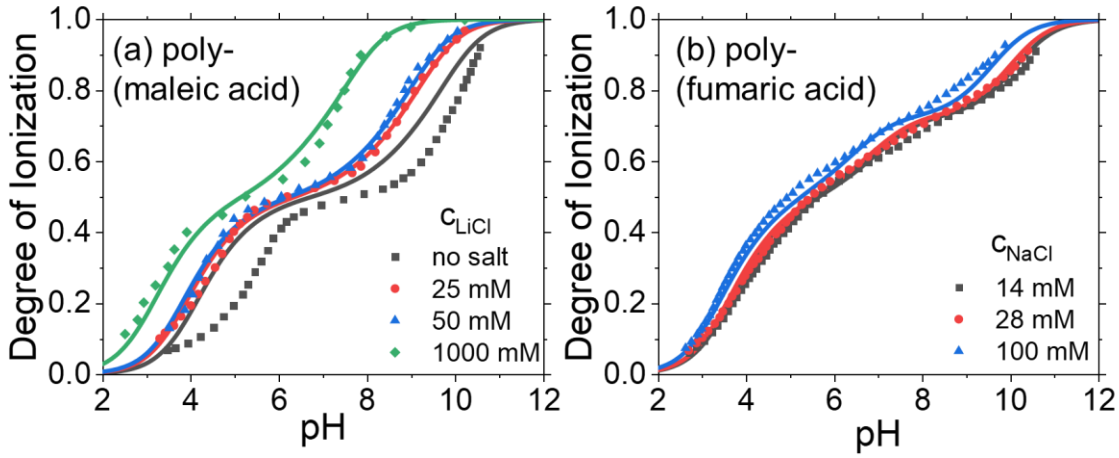

Figure 5.3 The degree of ionization for (a) poly(maleic acid) in pure water and in three lithium chloride solutions and (b) poly(fumaric acid) in three sodium chloride solutions according to experiments[11, 44] (symbols) and theory (lines).

Figure 3 compares the theoretical and experimental titration curves for PMA (MW=17.6 kg/mol[44]) and PFA (MW=6.6 kg/mol[11]) in pure water and in aqueous solutions of LiCl and NaCl. Our thermodynamic model performs quite well at moderate to high salt concentrations. Interestingly, the ionization of PMA and PFA extends over 9 pH units owing to the strong intra-chain repulsion. Due to the difference in stereochemistry between the two polymers, there is a noticeable contrast between the titration curves. While PMA shows a short plateau in the titration curve when it is half ionized, no such feature is observed in the titration curve for PFA. The difference between the two titration curves can be understood by the larger proportion of the weakly repulsive mesomeric bonds in PFA compared to that in PMA. Two charged sites connected by a racemic bond in PFA show more significant repulsion than that of PMA, leading to an extended titration curve in the late stage of ionization.

As was seen for PAA, the truncation of the intrachain correlations to only those between nearest neighbors becomes problematic at low to dilute salt conditions. Recently, an efficient approach was demonstrated by Garcés and coworkers using the Gibbs-Bogoliubov variational principle to incorporate intra-chain correlations via an effective one-body potential[49, 50]. The agreement between the exact results and this new approach was quantitative when applied to weak polyelectrolytes with rod-like configurations where the long-range coulomb interaction was described by a DH potential. While this new method is promising, its application to realistic systems has not yet been achieved. For example, consider the no salt condition for the titration of PMA shown in Figure 3a. The initial ionization of the polymer is significantly delayed, but otherwise the curve appears quite similar to the 25 mM case. This indicates that while the long-range interaction is relevant at low degrees of ionization, these interactions will be screened out by the salt ions that are attracted to the accumulated charge of the polymer. This cannot be accounted for through a DH potential which predicts a significantly extended titration curve at dilute conditions since the long-range interactions maintain relevancy.

Alternatively, Ghasemi and Larson utilized the random phase approximation (RPA) to account for the intramolecular electrostatic correlations of polymers with a specified structure (e.g., coil-like)[51]. The RPA method incorporates the intramolecular correlations through an effective one-body potential which utilizes the average behavior of the ionizable segments. Thus, unlike our model employed here, it cannot capture the step-like titration since the explicit charge status of sites (viz. paired charge states of neighbors) is not considered. Nonetheless, this method is a key step towards a universal description

of weak polyelectrolytes by accounting for contributions beyond nearest neighbors in a convenient way. Unfortunately, the use of a fixed structure of the polymer is problematic as the conformation is dependent upon the charge of the polymer. The coupling of conformation and ionization is not easily described with analytical models because both electrostatic and non-electrostatic (e.g., hydrophobic) interactions determine the polymer configuration. While the polymer tends to expand upon ionization to reduce the repulsive Coulomb interactions, the hydrophobic interactions cause the polymer to collapse upon itself. Future work in these directions could lead to a more accurate description of weak polyelectrolytes across all solution conditions.

The choice of counterions is of particular importance for the ionization of weak polyelectrolytes with a high linear charge density. We show the titration curve of poly(maleic acid) in three different 50 mM salt solutions in Figure 4. While the charging of the polymer with pH is relatively independent of the salt type at a low degree of ionization, the choice of salt cation plays a major role above 50% ionization where nearest-neighbor interactions become important. While lithium chloride shows only a small plateau when it is half charged, tetramethylammonium (TMA) chloride exhibits a plateau in ionization spanning across nearly 5 pH units. Interestingly, even at pH 11, the polymer is still only 75% charged when TMA is used as the counterion. To model the corresponding titration curves, we used the pK_a and distance parameters for the racemic and mesomeric bond determined previously for PMA in lithium chloride solution and adjusted the racemic and mesomeric non-electrostatic energies to account for the salt type effect.

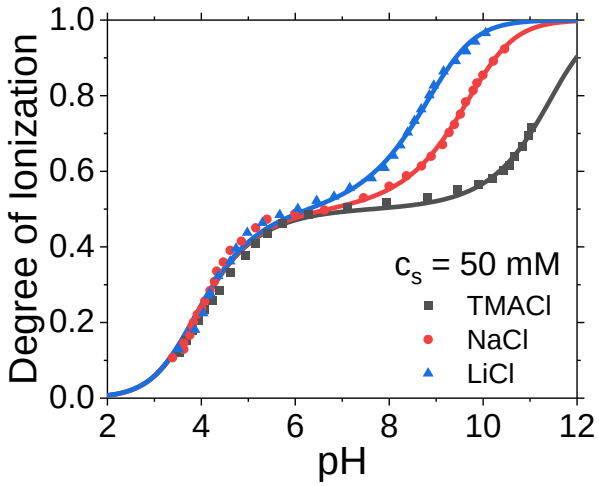

Figure 5.4 The ionization behavior of poly(maleic acid) in tetramethylammonium chloride, sodium chloride, and lithium chloride at 100 mM salt concentration according to experiments[11, 44] (symbols) and theory (lines).

While some variation in the distance between adjacent sites may exist for different salt types as was seen previously for PAA, we expect that the influence of salt type is mostly non-electrostatic in origin. This has been suggested previously by the conventional site-binding model which found the pair interaction for TMACl to unexpectedly increase with salt concentration for both the racemic and mesomeric bonds[38]. For TMACl, NaCl, and LiCl, the interaction energies were determined to be -1.90, -3.72, and -4.53 k_BT, respectively, for the racemic bond. For NaCl and TMACl, it was necessary to include a non-electrostatic energy of +1.99 k_BT and +4.86 k_BT between the adjacent sites connected by a mesomeric bond as well. As expected from the titration curves, the repulsion between neighboring sites increases as the counterion is changed from lithium to tetramethylammonium. This can be understood by the differences in the hydration of the counterions since lithium is well solvated while TMA is poorly solvated, leading to a

modulation of the local water structure around the polymer. The importance of salt type in the two-body interaction reflects the need for a more comprehensive understanding of the solvation effects.

5.3.3 Weak polyelectrolytes with branched architecture

Branched polyelectrolytes behave quite differently from their linear counterparts because of additional ionizable neighbors at the branching points. For example, a linear poly(ethyleneimine) chain is fully charged below pH 2 but a substantial number of amine groups remain uncharged when the polymer has a branched architecture[27]. While the titration behavior is affected by the chemical nature of the amine groups with different coordination numbers, such effects are relatively minor in comparison with that due to the topological arrangement of the ionizable sites[8].

In comparison with linear chains of similar molecular weight, dendrimers have a more compact internal structure thereby more interactions among ionizable sites in the same polymer. A dendrimer is typically described as an onion with each shell from the center containing double the number of sites within the previous shell. The total number of shells is denoted as the generation, *G*. If all sites are ionizable, the total number of ionizable groups is 2^{G+2}-2, and the number of ionizable groups in the outermost shell is 2^{G+1}. The ionization of a dendrimer proceeds first through the odd shells, which constitute two-thirds of the total number of ionizable groups, then followed by the ionization of the even shells. However, to ionize the even shells is more difficult than in their linear counterparts because there are three pair interactions in the dendrimer case versus the two pair interactions in the linear case. Previous investigations based on the site-binding model,

MC simulation and SCFT calculations demonstrated that the ionization of groups in the outer shells is easier than that for those groups in the inner shells due to the increased density of ionizable sites[19, 52]. To better understand the ionization behavior of weak polyelectrolyte dendrimers, we must be able to account for not only the topological effects but also the influence of solution conditions.

To demonstrate the application of our molecular thermodynamic model to weak polyelectrolytes with branched architecture, we consider the acid-base equilibrium of 1,4-diaminobutane poly(propylene imine) (PPI) dendrimers at different salt concentrations. The titration curves for PPI dendrimers of generation one to five were studied by van Duijvenbode et al. in three sodium chloride solutions of 0.1 M, 0.5 M, and 1.0 M using experimental techniques and a site-binding model[39]. The original site-binding model employed four equilibrium constants and three pair interaction energies to account for different locations of amine groups (i.e., the shell position) within the dendrimer. The large number of parameters are required because of significant differences in the chemical environment from the outer shell to the inner shell of the dendrimer.

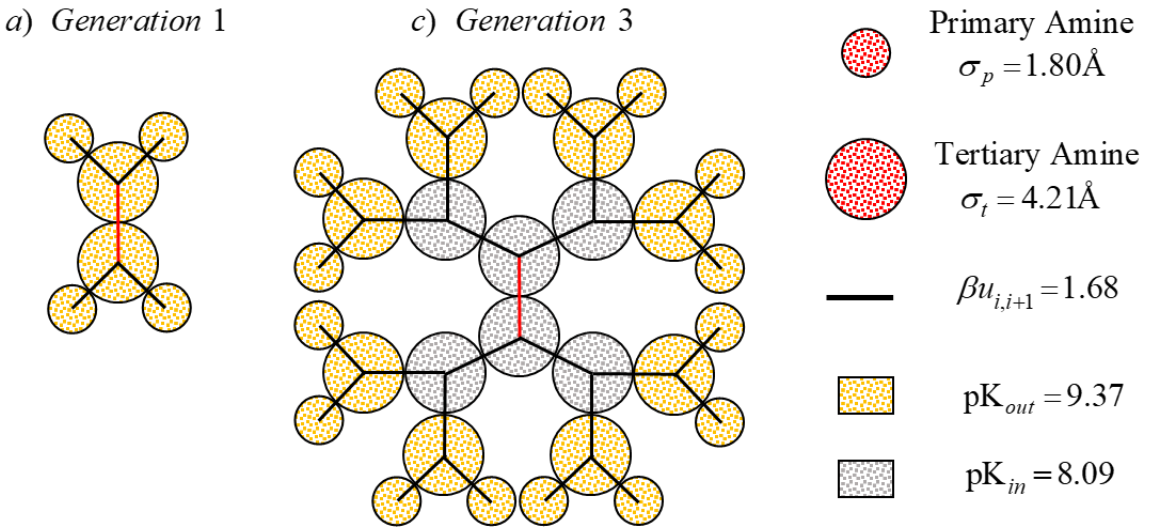

Figure 5.5 Coarse-grained dendrimer of different generations employed in our molecular thermodynamic model. The equilibrium constant for the two outermost shells of any dendrimer is given by pK_{out} and all other (i.e., inner) shells are given by pK_{in}. For generation 1, both shells are given by pK_{out}. All consecutive charged amines, besides the two in the center, experience a non-electrostatic repulsion of $\beta u_{i,i+1}$.

We show the coarse-grained model for dendrimers considered in our work in Figure 6. We distinguish two different size parameters: one for the primary amine (those in outermost shell) and one for the tertiary amines (all others). They are given by $\sigma_p = 1.80 \pm 0.22$Å and $\sigma_t = 4.21 \pm 0.47$Å, respectively. There is a non-electrostatic repulsion of $\beta u_{i,i+1} = 1.68 \pm 0.13$Å between all amines (black lines) besides between the two center amines (red line). For a generation 1 dendrimer, the equilibrium constant of all amines is given by $pK_{out} = 9.37 \pm 0.01$. To consider higher generation dendrimers, we find that all shells besides the two outermost can be described by $pK_{in} = 8.09 \pm 0.06$. As shown in Figure 5b, a generation three dendrimer is represented by two inner shells and two outer shells. The reason for the different equilibrium constants for the amine groups reflects the

significant change in local chemical environment of the dendrimer in its inner region from its outer region. We see that our coarse-grained model greatly simplifies the depiction of a dendrimer employed in the original site-binding model and provides physical meaning to the correlated parameters.

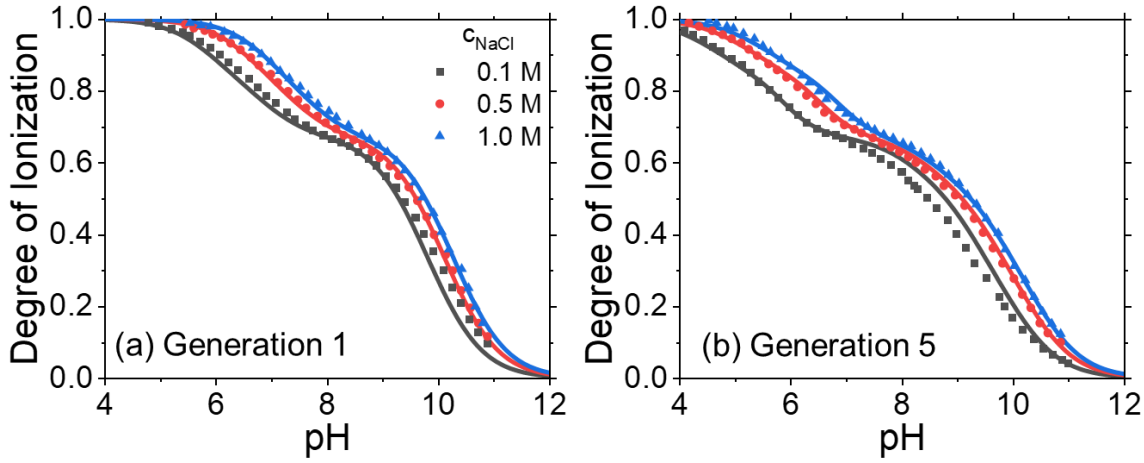

Figure 5.6 The titration curves for 1,4-diaminobutane poly(propylene imine) (PPI) dendrimers in aqueous solutions at three concentrations of sodium chloride according to experiments[39] (symbols) and theory (lines).

Overall, the titration behavior of dendrimers is well accounted for through the coupling of our thermodynamic model with the site-binding description of ionization in weak polyelectrolytes. Figure 6 compares the theoretical and experimental titration curves for PPI dendrimers of generations one and five at 0.1, 0.5, and 1.0 M sodium chloride concentrations. The titration results for dendrimers of generations two, three, and four can be found in Figure S6. In contrast to a polyacid, a polybase reduces its charge as the pH increases because it is charged in the protonated form. The increase in salt concentration shifts the titration curve to higher pH because of electrostatic screening. In the

conventional site-binding model, the parameters must vary with the solution condition (e.g., salt concentration) in order to reproduce the experimental titration data. However, with the assistance of the thermodynamic model, we use only the size of the titratable group (i.e., primary and tertiary amine) to account for the effects of the solution condition on the apparent equilibrium constant and pair interaction energies. We fit the theoretical results to the titration curves for the different dendrimer generations at 0.5 and 1.0 M sodium chloride concentration. The interactions with non-neighboring segments seem to be relevant at 0.1 M as indicated by the overestimation in the ionization when the generation five is weakly charged. The excellent agreement between the theoretical results and experimental data indicates that the physics of weak dendrimer ionization is well described by incorporating the electrostatic inter- and intra-correlations as well as the influence of changes in local electric potential into the site-binding model.

The theoretical model shows deviation from the experimental data at low degree of ionization while it is more quantitative in the later stages of ionization. In the early stage of charging, the interactions of segments with non-neighboring segments are of more significance as there are not sufficient ions drawn into the dendrimer to screen these interactions at lower salt concentrations. This is why our results overestimate the ionization at 0.1 M for the generation 5 dendrimer since only the correlation with neighboring segments is considered explicitly. For lower generation dendrimers, the discrepancy between experiments and theory is reduced as the dendrimer is less compact so the importance of non-neighboring segments is diminished. One exception is the generation two dendrimer (Figure S6b) which shows a consistent overestimation of the

ionization for all three salt concentrations beyond when the polymer is two-thirds ionized. The reason for this is that the equilibrium constant of the middle (i.e., second outermost) shell is not well represented by pK_{out}, but by a value ~0.3 units lower (viz. pK = 9.1). This likely reflects generation two as a transition from the small molecule (G=1) where the local environment is the relatively the same for all amines to a large molecule ($G \geq 3$) which shows two clear regions (viz. inner and outer) which are reflected by their own pK values.

5.4 Conclusion

We developed a molecular thermodynamic model for predicting ionization of weak polyelectrolytes by incorporating the site-binding model with classical density functional theory (cDFT). The new model accounts for the local thermodynamic condition for individual ionizable sites through the separation of the apparent equilibrium constant into two components: a chemical contribution, i.e., the thermodynamic equilibrium constant for proton binding/dissociation of ionizable functional groups, and physical contributions due to non-bonding interactions (e.g., electrostatics) with all chemical species in the solution. The influence of solution conditions on the one-body term is affiliated with the electrostatic correlations as determined by the mean spherical approximation (MSA) and with the change in the local electric potential upon ionization of the monomer as determined by cDFT. In addition, we account for the non-ideality due to electrostatic correlations between neighboring monomers within the polymer. The electrostatic work to form a pair of charged sites can be related to the negative natural log of the radial distribution function which is determined from the reference charged hard-sphere system. The incorporation of the thermodynamic non-ideality into the site-binding model allows us to describe the

titration behavior of weak polyelectrolytes with different arrangements of ionizable sites and configurations.

The molecular thermodynamic model provides a theoretical basis to estimate the parameters employed in previous applications of the site-binding model. While quantities like the intrinsic binding constants (viz., pK_a values) must be determined by fitting with experimental data, they maintain physically realistic values and are invariant with the solution conditions. We demonstrate the *predictive* application of our thermodynamic model by extensive comparison to experimental titration behavior of weak polyelectrolytes across different solution conditions and polymer architectures. Specifically, we first consider the linear weak polyacid, poly(acrylic acid) (PAA), in various solutions of alkali chloride ranging in salt concentration from 5 mM to 1 M. It is found that the titration behavior at moderate to high salt concentration (i.e., greater than or equal to 20 mM) can be described through two parameters: the thermodynamic equilibrium constant and the size of the monomer. The monomer size varies with the types of alkali cations present in solution likely owing to differences in their hydration behavior. At the lowest salt concentration considered in this work, the importance of intra-chain correlations beyond nearest neighbors cannot be ignored and noticeable deviation is seen between the theoretical values and experimental data. A similar trend was found for the linear weak polyacids, poly(maleic acid) (PMA) and poly(fumaric acid) (PFA), which have a much greater density of carboxyl groups along their backbones in comparison with PAA. Despite the strong correlations between neighboring monomers in PMA and PFA, our model agrees well with the experimental titration data at moderate to high salt concentrations. As a final

example for the application our model to branched weak polyelectrolytes, we considered the titration behavior of poly(propylene imine) dendrimers. In spite of the complex structure of dendrimers, the influence of solution conditions on the degree of ionization was quantitatively captured by our thermodynamic model. Thus, the inclusion of inter- and intra-chain correlations into the one- and two-body terms of the conventional site-binding model, respectively, provides a feasible method to account for the salt concentration effect on the titration behavior of different weak polyelectrolytes.

According to the site-binding model, the influence of chain length on the ionization behavior of weak polyelectrolytes is significant when the end-effects are relevant (e.g., M<20). However, recent experiments (e.g., Laguecir et al.[18]) indicate that the chain length is important even for long polymers; although the chain length effect is most significant at dilute conditions. Such effect was not captured in this work because our calculations are based on the site-binding model but with a predictive one-body energy and the two-body interaction parameter. Indeed, the description of weak polyelectrolytes using our thermodynamic model provides only a first step towards a fully predictive model for weak polyelectrolyte ionization. A major shortcoming of the site binding model is the truncation of the multi-body expansion of the free energy to only include up to the nearest neighbor two-body term. While previous applications of the site-binding model with nearest neighbor interactions have been used successfully to describe the titration behavior of weak polyelectrolytes at low salt concentrations, our model demonstrates that correlations beyond adjacent monomers do become relevant at such conditions. The success of the conventional method must be contributed to the weak correlation between

non-consecutive sites. While further improvements can be done systematically by accounting for contributions beyond nearest-neighbor interactions, the theoretical description of these interactions is not well understood, and incorporation of multi-body correlations are computationally more challenging. Future work incorporating such correlations may glean useful insights from the charging behavior of weak polyelectrolytes in dilute aqueous solutions. Moving beyond the site-binding model, polymer density functional theory (PDFT) provides a convenient avenue to analyze the microscopic structure of polymers in solution and near surfaces. We are currently pursuing further development along these lines.

Supporting Information

S5.1. Change in excess chemical potential upon ionization of a monomer

1) Electrostatic correlation

The electrostatic correlations are accounted by using the mean spherical approximation (MSA)[33].

$$\mu_i^{el} = -l_B\left[\frac{Z_i^2\Gamma}{1+\Gamma\sigma_i}+\eta\sigma_i\left(\frac{2Z_i-\eta\sigma_i^2}{1+\Gamma\sigma_i}+\frac{\eta\sigma_i^2}{3}\right)\right] \tag{S5.1}$$

where $l_B = \beta e^2/4\pi\varepsilon_0\varepsilon_r$ is the Bjerrum length (7.14 Å for water at room temperature), $\beta = 1/k_BT$ is the inverse of the Boltzmann factor, k_B, and the absolute temperature T, e is the elementary charge, and ε_0 and ε_r are the vacuum permittivity and the relative dielectric constant of the solution, respectively, Z_i is the valence of the species, and Γ is the MSA screening parameter which is given by

$$\Gamma^2 = \pi l_B\sum_j \rho_j\left(Z_j^{eff}\right)^2 \tag{S5.2}$$

$$Z_j^{eff} = \frac{Z_j-\eta\sigma_j^2}{1+\Gamma\sigma_j} \tag{S5.3}$$

$$\eta = \frac{1}{H(\Gamma)}\sum_j \rho_j\frac{Z_j\sigma_j}{1+\Gamma\sigma_j} \tag{S5.4}$$

$$H(\Gamma) = \sum_j \rho_j\frac{\sigma_j^3}{1+\Gamma\sigma_j}+\frac{2}{\pi}\left(1-\frac{\pi}{6}\sum_j\rho_j\sigma_j^3\right) \tag{S5.5}$$

2) Charging potential

Because the binding site model requires an analytical expression for the surface potential of the fixed particle, we fit the cDFT predictions of the local electric potential with a DH-like expression over the relevant parameter space ($1\text{Å} < \sigma < 8\text{Å}$ and $c_s < 1$ M):

$$\beta\mu_i^{charge} = Z_i^2 \frac{A_\sigma \sqrt{I}}{1 + B_\sigma \sqrt{I}} \tag{S5.6}$$

where A_σ and B_σ are fitting parameters that vary with the diameter of the monomer in angstroms and I is the ionic strength, $I = \sum_j c_j Z_j^2 / 2$ where c_j is the concentration of species j in the bulk solution. These also will be dependent upon the size of the salt ions present in the solution; however, the sensitivity to the ion size is low at the concentrations considered. The fitting parameters are given by the following equations

$$A_\sigma = 2.27769 + 0.55745 \exp\left[-\sigma / 2.32949\right] \tag{S5.7}$$

$$B_\sigma = -2.25294 + 2.99137 \exp\left[\sigma / 25.82945\right] \tag{S5.8}$$

Figure S1a shows the variation of the excess chemical potential with the cesium chloride concentration for a charged hard sphere of diameter 3.57 Å. We see significant changes in both contributions at low salt concentration because they both vary approximately with the square root of the ion concentration. As expected, the increase in salt concentration results in a more negative value, i.e., it shifts the chemical equilibrium towards the ionized state. The combination of these two contributions provides a convenient method to account for the thermodynamic non-ideality in the one-body term.

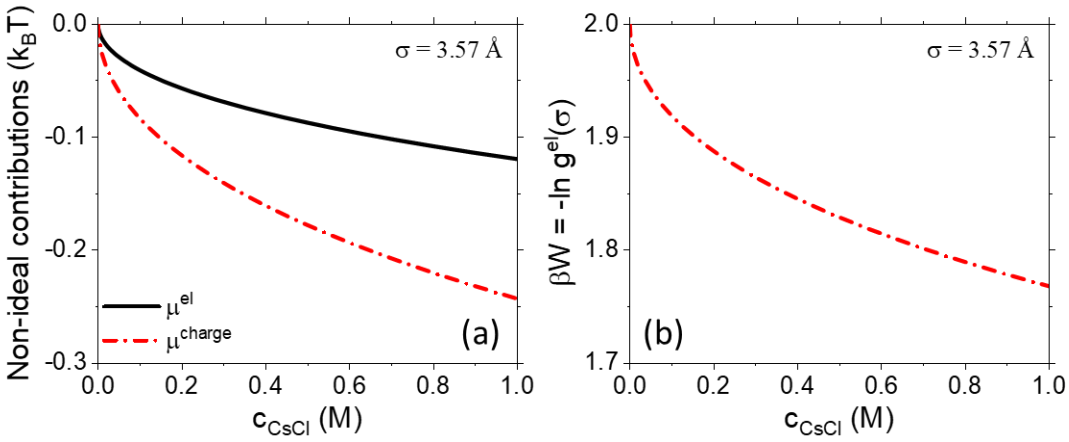

Figure S 5.1 (a) The influence of electrostatic correlations and charging potential on the excess chemical potential for a monomer with a hard-sphere diameter of 3.57 Å. (b) The work to form a bond between two adjacent positive charges at contact from MSA.

2. Site-binding description of poly(acrylic acid)

To demonstrate the power of the nearest-neighbor site-binding model, we consider the titration behavior of poly(acrylic acid), PAA, in a cesium chloride aqueous solution. PAA is a linear chain with low line charge density, i.e., the neighboring sites are not too close, which does not exhibit stereochemistry-dependent titration properties like those of the higher line charge density polyacids. We show a comparison of the nearest-neighbor model to the experimental data for poly(acrylic acid) in alkali chloride solutions in Figure S2[20].

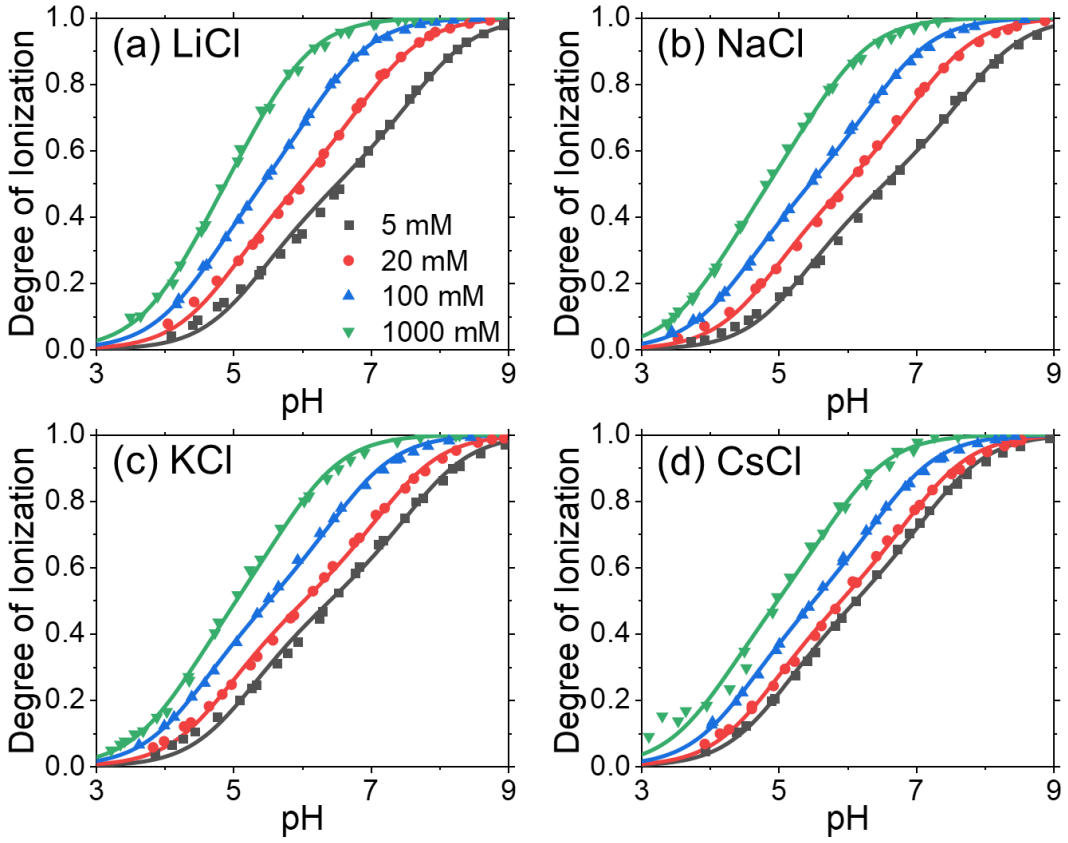

Figure S 5.2 The titration behavior of poly(acrylic acid) in (a) lithium chloride, (b) sodium chloride, (c) potassium chloride and (d) cesium chloride aqueous solution from experiment (symbols) and nearest-neighbor site-binding model (lines).

The agreement between the two is excellent giving credence to the simplification of the multi-body correlations to that of adjacent monomers at moderate to high salt concentrations. It is seen that as the salt concentration is decreased, there is a greater discrepancy between the site-binding model and the experimental results when the polymer is only partially ionized. This can likely be attributed to the long-range interactions not being sufficiently screened. It can be expected that as the ionization of the polymer increases, more counterions will be present in the vicinity of the polymer providing additional screening than compared to the bulk concentration of ions. Such an increase of

ions near the polymer has been demonstrated previously through cylindrical PB calculations[20] and by explicit ion MC simulations[14]. The increased presence of ions near the surface is likely why the nearest-neighbor model is still sufficient even at salt concentrations of 1 mM where it would be expected that interactions beyond the neighboring monomers will be present based on bulk salt concentrations. To compensate, the nearest-neighbor model predicts a larger pK′ which misses the initial ionization of PAA. For each solution condition (i.e., salt concentration and cation choice), the apparent dissociation constant and nearest-neighbor pair interaction are determined, and we show these values in Table S1-S4.

Table S 5.1 The correlated parameters for the nearest-neighbor site-binding model to describe the titration of poly(acrylic acid) in lithium chloride aqueous solution.

c_{LiCl} (mM)	pK'	W^p
5	5.66	1.97
20	5.23	1.67
100	4.85	1.31
1000	4.55	0.78

Table S 5.2 The correlated parameters for the nearest-neighbor site-binding model to describe the titration of poly(acrylic acid) in sodium chloride aqueous solution.

c_{NaCl} (mM)	pK'	W^p
5	5.64	2.07
20	5.18	1.84
100	4.75	1.63
1000	4.37	1.12

Table S 5.3 The correlated parameters for the nearest-neighbor site-binding model to describe the titration of poly(acrylic acid) in potassium chloride aqueous solution.

c_{KCl} (mM)	pK'	W^p
5	5.50	2.08
20	5.16	2.00
100	4.77	1.72
1000	4.53	1.18

Table S 5.4 The correlated parameters for the nearest-neighbor site-binding model to describe the titration of poly(acrylic acid) in cesium chloride aqueous solution.

c_{CsCl} (mM)	pK'	W^p
5	5.33	1.83
20	5.13	1.80
100	4.79	1.65
1000	4.41	1.32

Based off the fitted parameters, we see that the apparent protonation constant and the nearest-neighbor pair interaction both decrease with increase in salt concentration. This can be understood intuitively as resulting from the electrostatic screening due to the increased number of ions present in the solution. As a result, the ionization (i.e., a decrease in the degree of protonation for a polyacid) increases due to the lower free energy barrier to charge a site. Understandably, at salt concentrations lower than those considered, the importance of interactions beyond nearest neighbor become relevant and the model would not perform well. A major difficulty when employing the nearest neighbor site-binding model is that sufficient experimental data must be available to correlate experimental data and parameters must be determined for every set of solution conditions. Thus, there is no predictive capability to this model which limits its applicability. To improve upon the current approach, we must account for the thermodynamic non-idealities that govern the titration behavior of weak polyelectrolytes.

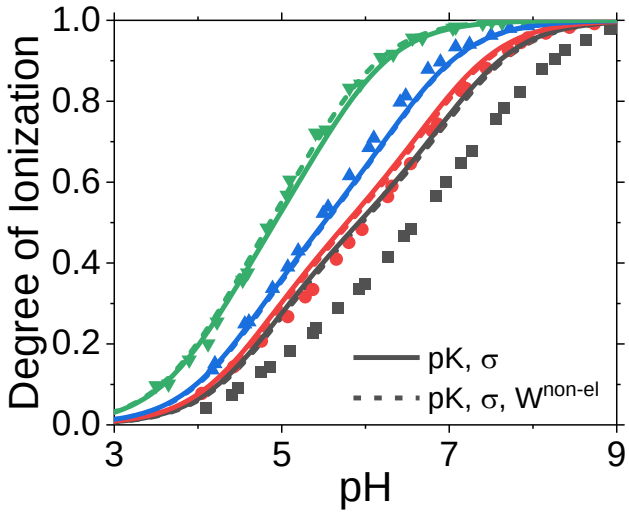

Figure S 5.3 Titration curves for poly(acrylic acid) in aqueous solution of lithium chloride from experiment (symbols) and theoretical correlations (lines). The theoretical lines are fit using the thermodynamic equilibrium constant, size of monomer, and a possible third parameter: the non-electrostatic pair energy.

The titration curve of PAA is only weakly influenced by the incorporation of a non-electrostatic interaction between two consecutive charge sites. However, we find that this leads to an overfitting of the experimental data. The fitting procedure introduces an attractive non-electrostatic energy of -0.59 k_BT to compensate a larger repulsive contribution due to electrostatics for the decreased monomer size (2.73 Å versus 3.56 Å when non-electrostatic interaction was not included). This provides a greater variation with salt concentration which allows a better fit of the 20 mM case. However, the worse agreement of the theoretical results should be attributed to the interactions beyond nearest neighbors. Thus, non-electrostatics are of minor significance to PAA as can be expected from the large separation between sites.

Bibliography

1. Priya James, H., et al., *Smart polymers for the controlled delivery of drugs – a concise overview.* Acta Pharmaceutica Sinica B, 2014. **4**(2): p. 120-127.

2. Jaganathan, S., *Bioresorbable polyelectrolytes for smuggling drugs into cells.* Artificial Cells, Nanomedicine, and Biotechnology, 2016. **44**(4): p. 1080-1097.

3. Bolto, B. and J. Gregory, *Organic polyelectrolytes in water treatment.* Water Research, 2007. **41**(11): p. 2301-2324.

4. Seitz, S. and H. Ajiro, *Self-assembling weak polyelectrolytes for the layer-by-layer encapsulation of paraffin-type phase change material icosane.* Solar Energy Materials and Solar Cells, 2019. **190**: p. 57-64.

5. Ng, J., et al., *Stimuli-Responsive Hybrid Polymer for Enhanced Solid–Liquid Separation of Industrial Effluents.* Environmental Science & Technology, 2019. **53**(11): p. 6436-6443.

6. Kobayashi, M., M. Terada, and A. Takahara, *Reversible adhesive-free nanoscale adhesion utilizing oppositely charged polyelectrolyte brushes.* Soft Matter, 2011. **7**(12): p. 5717-5722.

7. Li, C., Y. Gu, and N.S. Zacharia, *Tuning Wet Adhesion of Weak Polyelectrolyte Multilayers.* ACS Applied Materials & Interfaces, 2018. **10**(8): p. 7401-7412.

8. Koper, G.J.M. and M. Borkovec, *Proton binding by linear, branched, and hyperbranched polyelectrolytes.* Polymer, 2010. **51**(24): p. 5649-5662.

9. Nekrasova, T.N., et al., *Potentiometric titration of polyacrylic acid, polymethacrylic acid and poly-l-glutamic acid.* Polymer Science U.S.S.R., 1965. **7**(5): p. 1008-1018.

10. Ulrich, S., M. Seijo, and S. Stoll, *A Monte Carlo Study of Weak Polyampholytes: Stiffness and Primary Structure Influences on Titration Curves and Chain Conformations.* The Journal of Physical Chemistry B, 2007. **111**(29): p. 8459-8467.

11. Kitano, T., et al., *Dissociation behavior of poly(fumaric acid) and poly(maleic acid). 1. Potentiometric titration and intrinsic viscosity.* Macromolecules, 1987. **20**(7): p. 1598-1606.

12. Gonzalez Solveyra, E., et al., *Theoretical Modeling of Chemical Equilibrium in Weak Polyelectrolyte Layers on Curved Nanosystems.* Polymers, 2020. **12**(10): p. 2282.

13. Landsgesell, J., et al., *Simulations of ionization equilibria in weak polyelectrolyte solutions and gels.* Soft Matter, 2019. **15**(6): p. 1155-1185.

14. Rathee, V.S., et al., *Explicit Ion Effects on the Charge and Conformation of Weak Polyelectrolytes.* Polymers, 2019. **11**(1): p. 183.

15. Hofzumahaus, C., P. Hebbeker, and S. Schneider, *Monte Carlo simulations of weak polyelectrolyte microgels: pH-dependence of conformation and ionization.* Soft Matter, 2018. **14**(20): p. 4087-4100.

16. Rathee, V.S., et al., *Weak polyelectrolyte complexation driven by associative charging.* The Journal of Chemical Physics, 2018. **148**(11): p. 114901.

17. Carnal, F. and S. Stoll, *Adsorption of Weak Polyelectrolytes on Charged Nanoparticles. Impact of Salt Valency, pH, and Nanoparticle Charge Density. Monte Carlo Simulations.* The Journal of Physical Chemistry B, 2011. **115**(42): p. 12007-12018.

18. Laguecir, A., et al., *Size and pH effect on electrical and conformational behavior of poly(acrylic acid): Simulation and experiment.* European Polymer Journal, 2006. **42**(5): p. 1135-1144.

19. Uhlík, F., et al., *Modeling of Ionization and Conformations of Starlike Weak Polyelectrolytes.* Macromolecules, 2014. **47**(12): p. 4004-4016.

20. Sadeghpour, A., et al., *Influence of alkali metal counterions on the charging behavior of poly(acrylic acid).* Polymer, 2009. **50**(16): p. 3950-3954.

21. Witte, K.N., S. Kim, and Y.-Y. Won, *Self-Consistent Field Theory Study of the Effect of Grafting Density on the Height of a Weak Polyelectrolyte Brush.* The Journal of Physical Chemistry B, 2009. **113**(32): p. 11076-11084.

22. Lützenkirchen, J., et al., *Comparison of Various Models to Describe the Charge−pH Dependence of Poly(acrylic acid).* Journal of Chemical & Engineering Data, 2011. **56**(4): p. 1602-1612.

23. Borkovec, M., G.J.M. Koper, and C. Piguet, *Ion binding to polyelectrolytes.* Current Opinion in Colloid & Interface Science, 2006. **11**(5): p. 280-289.

24. Borkovec, M. and G.J.M. Koper, *Ising Models of Polyprotic Acids and Bases.* The Journal of Physical Chemistry, 1994. **98**(23): p. 6038-6045.

25. Smits, R.G., G.J.M. Koper, and M. Mandel, *The influence of nearest- and next-nearest-neighbor interactions on the potentiometric titration of linear poly(ethylenimine).* The Journal of Physical Chemistry, 1993. **97**(21): p. 5745-5751.

26. Chandler, D., *Introduction to modern statistical.* Mechanics. Oxford University Press, Oxford, UK, 1987. **40**.

27. Borkovec, M. and G.J.M. Koper, *Proton Binding Characteristics of Branched Polyelectrolytes.* Macromolecules, 1997. **30**(7): p. 2151-2158.

28. Simonin, J.-P., L. Blum, and P. Turq, *Real Ionic Solutions in the Mean Spherical Approximation. 1. Simple Salts in the Primitive Model.* The Journal of Physical Chemistry, 1996. **100**(18): p. 7704-7709.

29. Gómez-Bombarelli, R., et al., *Computational Calculation of Equilibrium Constants: Addition to Carbonyl Compounds.* The Journal of Physical Chemistry A, 2009. **113**(42): p. 11423-11428.

30. Mishra, A.K. and J.C. Ahluwalia, *Apparent molal volumes of amino acids, N-acetylamino acids, and peptides in aqueous solutions.* The Journal of Physical Chemistry, 1984. **88**(1): p. 86-92.

31. Gallegos, A. and J. Wu, *Charge Regulation of Natural Amino Acids in Aqueous Solutions.* Journal of Chemical & Engineering Data, 2020. **65**(12): p. 5630-5642.

32. Yu, Y.-X., J. Wu, and G.-H. Gao, *Density-functional theory of spherical electric double layers and ζ potentials of colloidal particles in restricted-primitive-model electrolyte solutions.* The Journal of Chemical Physics, 2004. **120**(15): p. 7223-7233.

33. Blum, L., *Mean spherical model for asymmetric electrolytes.* Molecular Physics, 1975. **30**(5): p. 1529-1535.

34. Maribo-Mogensen, B., G.M. Kontogeorgis, and K. Thomsen, *Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions.* Industrial & Engineering Chemistry Research, 2012. **51**(14): p. 5353-5363.

35. Kawaguchi, S., T. Kitano, and K. Ito, *Dissociation behavior of poly (fumaric acid) and poly (maleic acid). 3. Infrared and ultraviolet spectroscopy.* Macromolecules, 1992. **25**(4): p. 1294-1299.

36. Jiang, J., et al., *A molecular-thermodynamic model for polyelectrolyte solutions.* The Journal of Chemical Physics, 1998. **108**(2): p. 780-784.

37. Xu, Q., et al., *Improved radial distribution functions for Coulomb charged fluid based on first-order mean spherical approximation.* The Journal of Chemical Physics, 2008. **128**(21): p. 214508.

38. de Groot, J., et al., *Dissociation Behavior of Poly(maleic acid): Potentiometric Titrations, Viscometry, Pulsed Field Gradient NMR, and Model Calculations.* Macromolecules, 1998. **31**(13): p. 4182-4188.

39. van Duijvenbode, R.C., M. Borkovec, and G.J.M. Koper, *Acid-base properties of poly(propylene imine)dendrimers.* Polymer, 1998. **39**(12): p. 2657-2664.

40. Lide, D.R., *CRC handbook of chemistry and physics*. Vol. 85. 2004: CRC press.

41. Dong, H., et al., *The Effects of Chemical Substitution and Polymerization on the pKa Values of Sulfonic Acids.* The Journal of Physical Chemistry B, 2009. **113**(43): p. 14094-14101.

42. Dong, H., H. Du, and X. Qian, *Prediction of pKa Values for Oligo-methacrylic Acids Using Combined Classical and Quantum Approaches.* The Journal of Physical Chemistry B, 2009. **113**(39): p. 12857-12859.

43. Reijenga, J., et al., *Development of Methods for the Determination of pKa Values.* Analytical Chemistry Insights, 2013. **8**: p. ACI.S12304.

44. Kawaguchi, S., et al., *Dissociation behavior of poly(fumaric acid) and poly(maleic acid). II. Model calculation.* Macromolecules, 1990. **23**(3): p. 731-738.

45. de Groot, J., J.G. Hollander, and J. de Bleijser, *Stereochemical Configuration of Poly(maleic acid) as Studied by 13C NMR.* Macromolecules, 1997. **30**(22): p. 6884-6887.

46. Henderson, D. and W.R. Smith, *Exact analytical formulas for the distribution functions of charged hard spheres in the mean spherical approximation.* Journal of Statistical Physics, 1978. **19**(2): p. 191-200.

47. Edward, J.T., P.G. Farrell, and J.L. Job, *Re-examination of the Kirkwood-Westheimer Theory of Electrostatic Effects. I. Calculation of the ``Effective'' Dielectric Constant.* The Journal of Chemical Physics, 1972. **57**(12): p. 5251-5256.

48. Kawaguchi, S., T. Kitano, and K. Ito, *Infrared and ultraviolet spectroscopic studies on intramolecular hydrogen bonding in an alternating copolymer of isobutylene and maleic acid.* Macromolecules, 1991. **24**(22): p. 6030-6036.

49. Blanco, P.M., et al., *Coupling of Charge Regulation and Conformational Equilibria in Linear Weak Polyelectrolytes: Treatment of Long-Range Interactions via Effective Short-Ranged and pH-Dependent Interaction Parameters.* Polymers, 2018. **10**(8): p. 811.

50. Garcés, J.L., et al., *Dealing with long-range interactions in the determination of polyelectrolyte ionization properties. Extension of the transfer matrix formalism to the full range of ionic strengths.* Journal of Polymer Science Part B: Polymer Physics, 2017. **55**(3): p. 275-284.

51. Ghasemi, M. and R.G. Larson, *Role of electrostatic interactions in charge regulation of weakly dissociating polyacids.* Progress in Polymer Science, 2021. **112**: p. 101322.

52. Klein Wolterink, J., et al., *Annealed Star-Branched Polyelectrolytes in Solution.* Macromolecules, 2002. **35**(24): p. 9176-9190.

Chapter 6. Ising density functional theory for weak polyelectrolytes with strong coupling of ionization and intrachain correlations

We report a theoretical framework for weak polyelectrolytes by combining the polymer density functional theory with the Ising model for charge regulation. The so-called Ising density functional theory (iDFT) provides an accurate description of the effects of polymer conformation on the ionization of individual segments and is able to account for both the intra- and inter-chain correlations due to the excluded-volume effects, chain connectivity and electrostatic interactions. Theoretical predictions of the titration behavior and microscopic structure of ionizable polymers are found in excellent agreement with experiment.

6.1 Introduction

The structure and thermodynamic properties of weak polyelectrolytes have been extensively studied[1-3]. However, it remains theoretically challenging to predict the complex interplay of monomeric ionization with the polymer conformation in various electrolyte solutions. The incomplete knowledge has been a major hurdle in understanding a wide variety of natural phenomena occurring in biological systems and in diverse technological applications of weak polyelectrolytes including bioadhesion and gene delivery[3-6]. A number of theoretical methods have been proposed for understanding the titration behavior of weak polyelectrolyte systems, including those based on the Poisson-Boltzmann equation[7, 8], polymer scaling laws[9, 10], self-consistent-field theory[11-13], and coarse-grained molecular models[14-16]. Existing methods have been quite successful

in describing the ionization and conformation of weak polyelectrolyte systems in direct comparison with experimental data. For example, polymers like poly(acrylic acid) whose ionizable sites are weakly correlated can be well described by various forms of site-binding models[17-19]. The molecular theory by Szleifer and coworkers predicts the brush height of poly(acrylic acid) at different solution conditions (viz. pH and salt concentration) as well as different polymer lengths and grafting densities in quantitative agreement with experimental data[15]. Zhulina and Birshtein identified several scaling regimes for weak polyelectrolyte brushes such as the osmotic brush, salted brush, and the Pincus brush that have been validated by experiments[11]. More recently, Müller and coworkers demonstrated that the single-chain-in-mean-field simulation approach is able to capture the main features of weak polyelectrolyte brushes in response to the solution pH, salt concentration, and grafting density[20]. Whereas the site-binding models are mostly focused on the ionization of individual segments without explicitly considering the polymer conformations, the self-consistent polymer theories and scaling analysis are typically concerned with the coupling between the chain conformation and the average ionization of weak polyelectrolytes. The existing methods are not able to address the correlated ionization of neighboring sites at different polymer conformations.

While a mean-field approach is useful to describe certain aspects of weakly charged polyelectrolyte systems, it is in general not be able to account for the polymer conformation effects on ionization and the distribution of charges along the polymer chains from a molecular perspective. In particular, mean-field methods are problematic for weak polyelectrolytes when the ionization of individual segments is strongly correlated with both

inter- and intra- interactions. Herein we present a new theoretical framework for weak polyelectrolytes by integrating the Ising model for charge regulation with the polymer density functional theory to account for the excluded-volume effects, long-range electrostatic interactions, and intramolecular correlations. By comparison with experimental data for the ionization of weak polyelectrolyte brushes, we demonstrate that the so-called *Ising* Density Functional Theory (iDFT) provides an accurate description of ionization-conformation coupling in diverse weak polyelectrolyte systems.

6.2 Thermodynamic models and methods

We consider a generic model of weak polyelectrolyte systems where each polymeric species is represented by tangent hard-sphere chains of ionizable segments, monomeric ions by charged hard spheres each with a fixed valence, and the solvent by a dielectric continuum. A similar model was proposed before by Müller and coworkers within the context of single-chain in self-consistent-field theory[20]. For a polymer chain with *M* spherical segments, the chain conformation is fully specified by a multidimensional vector, $\mathbf{R}_k = (\mathbf{r}_{k,1}, \mathbf{r}_{k,2}, ..., \mathbf{r}_{k,M_k})$, where $\mathbf{r}_{k,i}$ denotes the segment position, and the ionization of each polymer chain can be specified by vector $\mathbf{S}_k = (s_{k,1}, s_{k,2}, ..., s_{k,M_k})$, where $s_{k,i} = \pm 1$ or 0 is the charge number (viz., valence) of segment i in polymer chain *k*. A unique feature of weak polyelectrolyte systems is that the polymer charge is a dynamic variable coupled with the conformational changes. The dynamic coupling is responsible for the correlated ionization of neighboring segments as well as the long-ranged, many-body correlations that defy conventional mean-field descriptions.

According to the polymer density functional theory (DFT)[21, 22], the structure and thermodynamic properties of a weak polyelectrolyte system can be uniquely determined from the grand potential, which according to our model is a functional of the polymer and ionic density profiles, $r_k(\mathbf{X}_k)$ and $r_a(\mathbf{r})$, respectively,

$$\begin{aligned}\Omega[\{\rho_k(\mathbf{X}_k)\},\{\rho_\alpha(\mathbf{r})\}] &= \sum_\alpha \int d\mathbf{r}'\rho_\alpha(\mathbf{r}')\left\{k_BT[\ln\rho_\alpha(\mathbf{r}')-1]+V_\alpha^{ext}(\mathbf{r}')-\mu_\alpha\right\} \\ &+\sum_k \int d\mathbf{X}'_k\rho_k(\mathbf{X}'_k)\begin{Bmatrix}k_BT[\ln\rho_k(\mathbf{X}'_k)-1]+V_k^B(\mathbf{X}'_k)+V_k^{ext}(\mathbf{X}'_k)] \\ +\mu_k^\infty(\mathbf{X}'_k)-\mu_k+\mu_k^H(\mathbf{S}'_k)\end{Bmatrix} \\ &+F^{ex}[\{\rho_k(\mathbf{X}_k)\},\{\rho_\alpha(\mathbf{r})\}]\end{aligned} \tag{6.1}$$

where $\int d\mathbf{X}'_k = \sum_{\{s'_{k,1}\}}\int d\mathbf{r}'_{k,1}\sum_{\{s'_{k,2}\}}\int d\mathbf{r}'_{k,2}\cdots\sum_{\{s'_{k,M_k}\}}\int d\mathbf{r}'_{k,M_k}$, k_B is the Boltzmann constant, T is the absolute temperature, $V_k^B(\mathbf{X}_k)$ is the bond potential for polymer k, m_k and m_a are, respectively, the chemical potentials of polymeric species k in its fully uncharged state [Eq. (S56)] and monomeric ion a, $V_k^{ext}(\mathbf{X}_k)$and $V_a^{ext}(\mathbf{r})$ are external potentials. The excess Helmholtz energy $F^{ex}$is also a functional of the polymer and ionic density profiles, $r_k(\mathbf{X}_k)$ and $r_a(\mathbf{r})$; it accounts for the thermodynamic non-ideality arising from interactions between different chemical species and non-bonded intramolecular interactions.

In comparison with a conventional polymer DFT, Eq.(6.1) contains two addition terms. The first term, $\mu_k^\infty(\mathbf{X}_k)$, accounts for the contribution due to intramolecular electrostatic interactions for the polymer in configuration $\mathbf{X}_k$. As shown in Eq.(S39), this term is equivalent to the electrostatic binding energy between charged segments at infinite

dilution. The second term, $\mu_k^H(\mathbf{S}_k)$, is affiliated with the ionization of individual segments in each weak polyelectrolyte chain

$$\mu_k^H(\mathbf{S}_k) = \sum_{i=1}^{M_k} \Delta\mu_{k,i}^H(s_{k,i}) \tag{6.2}$$

where $\Delta\mu_{k,i}^H(s_{k,i})$ is the change in the grand potential due to deprotonation/protonation of the ionizable segments

$$\Delta\mu_{k,i}^H(s_{k,i}) = -\, s_{k,i} \ln 10 k_B T (\mathrm{pK}_{k,i} - \mathrm{pH}). \tag{6.3}$$

In Eq.(6.3), $s_{k,i} = 0$ means that the polymer segment is at its neutral state, $s_{k,i} = -1$ stands for the deprotonation of a weak acidic segment, and $s_{k,i} = +1$ for the protonation of a weak basic segment. The equilibrium constant of the deprotonation or protonation reaction, $\mathrm{K}_{k,i}$, depends on the identity of the ionizable site and system temperature but not on the solution composition. Other values of $s_{k,i}$ are also possible if the polymer segment contains multiple ionizable sites.

Given an explicit expression for the excess Helmholtz energy F^{ex}, the polymer and ionic density profiles can be obtained by minimizing the grand potential [Eq.(6.1)] with respect to the density profiles of polymers and monomeric ions, leading to a set of the Euler-Lagrange equations,

$$\rho_k(\mathbf{X}_k) = \exp\left\{-\beta[V_k^B(\mathbf{X}_k) + V_k^{ext}(\mathbf{X}_k) + \mu_k^{ex}(\mathbf{X}_k) + \mu_k^{\infty}(\mathbf{X}_k) + \mu_k^H(\mathbf{S}_k) - \mu_k]\right\}, \tag{6.4}$$

$$r_a(\mathbf{r}) = \exp\left\{-\beta[V_a^{ext}(\mathbf{r}) + \mu_a^{ex}(\mathbf{r}) - \mu_a]\right\} \tag{6.5}$$

where $\beta = 1/(k_B T)$, $\mu_k^{ex}(\mathbf{X}_k) = dF^{ex}/d\rho_k(\mathbf{X}_k)$ and $\mu_a^{ex}(\mathbf{r}) = dF^{ex}/d\rho_a(\mathbf{r})$ are the local excess chemical potentials of polymers and free ions, respectively. For tangent hard-sphere chains, the bond potential can be expressed in terms of the one-dimensional Dirac-delta function $\delta(r)$:

$$\exp[-\beta V_k^B(\mathbf{X}_k)] \sim \sum_{i=1}^{M_k-1} \frac{\delta[|\mathbf{r}_{k,i+1} - \mathbf{r}_{k,i}| - \sigma_k(s_{k,i}, s_{k,i+1})]}{4\pi\sigma_k^2(s_{k,i}, s_{k,i+1})} \tag{6.6}$$

where $\sigma_k(s_{k,i}, s_{k,i+1}) = [\sigma_k(s_{k,i}) + \sigma_k(s_{k,i+1})]/2$ is the center-to-center distance between neighboring segments of hard-sphere diameter $\sigma_k(s_{k,i})$ and $\sigma_k(s_{k,i+1})$, and the proportionality constant can be fixed from the normalization condition for the probability distribution of polymer conformations. Due to the solvation effect, the hard-sphere diameter of a polymer segment may depend on its charge status. From Eqs. (6.4) and (6.6), we can determine the density profiles of individual polymer segments

$$\rho_{k,i}(\mathbf{r}, s_{k,i}) = \int d\mathbf{X}'_k \delta(\mathbf{r} - \mathbf{r}'_{k,i}) \delta_{s_{k,i} s'_{k,i}} \rho_k(\mathbf{X}'_k) \tag{6.7}$$

where $\delta(\mathbf{r})$ and $\delta_{ss'}$ represent the three-dimensional Dirac-delta function and the Kronecker delta function, respectively. Subsequently, $\rho_{k,i}(\mathbf{r}, s_{k,i})$ can be used to calculate the fractions of each polymer segment at different charge states and the degree of ionization for the entire polymer chain.

Within the framework of the coarse-grained model considered in this work, the above procedure is formally exact, applicable to weak polyelectrolytes at uniform as well as inhomogeneous conditions. The main difference between iDFT proposed in this work

and previous adoptions of the two-state model (e.g., [12, 23]) is that iDFT accounts for segment position and ionization state on an equal footing. By using a multi-dimensional density $\rho_i(\mathbf{r},s)$, iDFT is able to capture the correlated ionization of individual segments and solution effects explicitly. On the one hand, iDFT extends the conventional theories of weak-polyelectrolyte titration (*viz*., the site-binding model and various modifications) by an explicit consideration of the solution conditions as well as the statistics of polymer conformations and its dependence on intersegment interactions. On the other hand, different from the various forms of polymer mean-field methods, iDFT accounts for the local charge status of individual segments instead of only the average degree of ionization. Thus, iDFT provides a systematic description of the strong coupling between polymer conformation and segment-level ionization from the ensemble perspective. While a large number of samples must be generated to represent weak polyelectrolyte chains with different assignments of segment charges in molecular simulations, iDFT circumvents the explicit enumeration of all possible arrangement of polymer charges by using segment densities at different charge states $\rho_{k,i}(\mathbf{r},s_{k,i})$, which can be calculated by an iterative procedure. From the density profiles of individual segments at different charge states, we can calculate the degree of ionization for *each* segment and thermodynamic properties of the system as detailed in Supporting Information.

Like most DFT calculations, approximations are inevitable in formulating the excess Helmholtz energy functional in order to get numerical results. For weak polyelectrolyte systems represented by the tangent hard-sphere chains of ionizable

segments and charged hard spheres for monomeric ions, the excess Helmholtz energy is conventionally formulated in the context of the thermodynamic perturbation theory (TPT)[22]

$$F^{ex} \approx F_m^{ex}\left[\left\{r_{k,i}(\mathbf{r},s_{k,i})\right\},\left\{r_a(\mathbf{r})\right\}\right] - \sum_k \int d\mathbf{X}_k r_k(\mathbf{X}_k) k_B T \ln y_k(\mathbf{X}_k) \tag{6.8}$$

where $F_m^{ex}[\{r_{k,i}(\mathbf{r},s_{k,i})\},\{r_a(\mathbf{r})\}]$ denotes the excess Helmholtz energy of the corresponding monomeric system, and $y_k(\mathbf{X})$ is the multibody cavity correlation function[24]. The former takes into account contributions due to hard-sphere repulsion (viz., excluded volume effects) as described by the modified fundamental measure theory[25], the mean electrostatic potential, and electrostatic correlations[26]. In the first-order approximation (TPT1)[27], the multibody cavity function is represented by a superposition of the two-body correlation functions of the monomeric reference system, $\ln y_k(\mathbf{X}_k) \simeq \sum_{i=1}^{M_k-1} \ln y_k(\mathbf{x}_{k,i},\mathbf{x}_{k,i+1})$, where $y_k(\mathbf{x}_{k,i},\mathbf{x}_{k,i+1})$ can be estimated from the mean-spherical approximation (MSA)[28]. The relevant details are given in Supporting Information.

At a given charge state, iDFT predicts that the segment density $r_{k,i}(\mathbf{r},s_{k,i})$ is given by [see Eqs. (S44-S51)]

$$\begin{aligned}\rho_{k,i}\left(\mathbf{r},s_{k,i}\right) = \exp\left\{-\beta\left[\lambda_{k,i}^{ex}\left(\mathbf{r},s_{k,i}\right)-\mu_k\right]\right\} \\ \times \sum_{\{s'_{k,i-1}\}} G_{k,i}^{L}\left(\mathbf{r},s'_{k,i-1},s_{k,i}\right) \sum_{\{s'_{k,i+1}\}} G_{k,i}^{R}\left(\mathbf{r},s_{k,i},s'_{k,i+1}\right)\end{aligned} \tag{6.9}$$

where $/_{k,i}^{ex}(\mathbf{r},s_{k,i})$ is the one-body potential for the segment in its given charge state $s_{k,i}$ (see Eq. S51). The propagator functions, $G_{k,i}^{L}$ and $G_{k,i}^{R}$, depend upon the charge state of segment *i* and of the neighboring segment through the two-body cavity correlation function, $y_k(\mathbf{x}_{k,i},\mathbf{x}_{k,i+1})$ (see Eq. S52 and S53). The degree of ionization for segment *i* in polymer *k* at position **r** is given by

$$\begin{aligned}\alpha_{k,i}\left(\mathbf{r},s_{k,i}\right) &= \frac{\rho_{k,i}\left(\mathbf{r},s_{k,i}\right)}{\sum_{\{s'_{k,i}\}}\rho_{k,i}\left(\mathbf{r},s'_{k,i}\right)} \\ &= \frac{\exp\left\{-\beta\left[\lambda_{k,i}^{ex}\left(\mathbf{r},s_{k,i}\right)\right]\right\}\sum_{\{s_{k,i-1}\}}G_{k,i}^{L}\left(\mathbf{r},s_{k,i-1},s_{k,i}\right)\sum_{\{s_{k,i+1}\}}G_{k,i}^{R}\left(\mathbf{r},s_{k,i},s_{k,i+1}\right)}{\sum_{\{s'_{k,i}\}}\exp\left\{-\beta\left[\lambda_{k,i}^{ex}\left(\mathbf{r},s'_{k,i}\right)\right]\right\}\sum_{\{s_{k,i-1}\}}G_{k,i}^{L}\left(\mathbf{r},s_{k,i-1},s'_{k,i}\right)\sum_{\{s_{k,i+1}\}}G_{k,i}^{R}\left(\mathbf{r},s'_{k,i},s_{k,i+1}\right)}\end{aligned} \quad . \quad (6.10)$$

Eq.(10) indicates that the degree of ionization for each polymer segment is directly correlated with the charge states of adjacent segments. By contrast, such correlation is neglected in previous mean-field methods that typically invoke a chemical equilibrium expression that involves the local proton concentration to calculate the local degree of ionization. Far from the surface (i.e., the bulk solution), this expression reduces to the Henderson–Hasselbalch equation.

The coupled effects of chain connectivity and ionization lead to long-range correlations that cannot be captured by conventional weak polyelectrolyte theories. For example, if the intramolecular correlations are ignored such as that in a typical polymer mean-field theory (PMFT), the propagator functions would no longer depend upon the

charge state of segment *i* (i.e., $G_{k,i}^{R}\left(\mathbf{r},s_{k,i},s_{k,i+1}\right)=G_{k,i}^{R}\left(\mathbf{r},s_{k,i+1}\right)$). In that case, the degree of ionization for segment *i* becomes

$$\alpha_{k,i}^{MF}\left(\mathbf{r},s_{k,i}\right)=\frac{\exp\left\{-b\left[I_{k,i}^{ex}\left(\mathbf{r},s_{k,i}\right)\right]\right\}}{\sum_{\left\{s'_{k,i}\right\}}\exp\left\{-b\left[I_{k,i}^{ex}\left(\mathbf{r},s'_{k,i}\right)\right]\right\}}. \tag{6.11}$$

Eq.(11) predicts that the ionization of different segments in a polymer chain is coupled only through the changes in the one-body potential.

Typically, PMFT also ignores both electrostatic and intramolecular correlations. The one-body potential often includes a simplified volume-exclusion term (e.g., incompressibility[29]) in addition to the mean electrostatic potential. For the sake of comparison with iDFT, we performed PMFT calculations by neglecting only the intramolecular correlations (i.e., $y(\mathbf{X})$=1). An improvement over PMFT is achieved by using the polymer density functional theory (PDFT)[30-32], which incorporates $y(\mathbf{X})$ but only at the mean-field level. In this case, the two-body correlation term is approximated by the weighted average contribution of segment *i*+1 in its different charge states given that segment *i* is in state $s_{k,i}$:

$$\ln y_k\left(\mathbf{x}_{k,i},\mathbf{x}_{k,i+1}\right)\simeq\sum_{\{s'_{k,i+1}\}}\alpha_{k,i+1}^{MF}\left(\mathbf{x}'_{k,i+1}\right)\ln y_k\left(\mathbf{x}_{k,i},\mathbf{x}'_{k,i+1}\right) \tag{6.12}$$

where $\alpha_{k,i+1}^{MF}(\mathbf{x}\text{¢}_{k,i+1})$ stands the average degree of ionization for segment *i*+1 at position $\mathbf{r}_{k,i+1}$ and charge state $s_{k,i+1}$. The incorporation of intramolecular correlations in this manner reduces the computational burden in enumerating over all neighboring pair states.

However, it ignores segment-level details unique for weak polyelectrolyte systems and fails to predict subtle effects such as the plateau in the ionization of strongly correlated polymers that can be captured through iDFT.

6.3 Results and discussion

To demonstrate the numerical performance of iDFT in comparison with PMFT and PDFT, we consider two representative weak polyelectrolyte systems: one is a weakly correlated brush of poly(acrylic acid), PAA; and the other is a strongly correlated brush of poly(maleic acid), PMA. The molecular parameters for polymers and the ionic species in the solution were determined previously from the titration of both weak polyelectrolytes in bulk solutions[33, 34]. Specifically, each PAA segment has a hard-sphere diameter of 3.37 Å and an intrinsic deprotonation constant of $pK_i = 5.22$. Whereas in PAA the charge correlation between neighboring segments is relatively weak, PMA exhibits much stronger correlation among adjacent ionizable sites as evidenced in a plateau (or step-like behavior) in the titration curve[35]. Because of the strong intrachain correlation effects, the half-charged polymer exists predominantly in consecutives pairs of one charged and one neutral monomers. Beyond this point, further ionization requires a larger driving force (*viz*. higher pH) to overcome strong intramolecular repulsion arising from the immediate neighbors. An accurate description of this polymer requires consideration of not only the intrachain repulsion between neighboring sites, but also its stereochemical configuration as demonstrated by the site-binding rotational isomeric model[36]. For simplicity, we ignore the stereochemical effects and consider the monomer segments in PMA to have a hard-sphere diameter of 1.95 Å (viz. the distance between the mesomeric bond in PMA). In

addition, we include a non-electrostatic repulsion between consecutive charges (i.e., $e_{i,i+1}\left(-1,-1\right)=2.5k_BT$) representative of a loss in intramolecular hydrogen bonding between adjacent charges sites[35]. The inclusion of a non-electrostatic interaction is consistent with our molecular model for weak polyelectrolytes in bulk solution[34]. The intrinsic deprotonation constant for a PMA monomer is set to $\mathrm{pK}_i=4.44$ in agreement with our previous study[34]. The parameters for the salt ions, proton, and hydroxyl ions are also the same as those used for bulk systems. The appropriate expressions for the excess chemical potential, $\hat{m}_i^{ex,b}$, as well as the self-consistent determination of the concentrations of all species in the bulk solution (viz. salt ions, protons, and hydroxyl ions) and their molecular parameters can be found in the SI or a previous work[33]. It should be noted that the solvent is considered implicitly through a dielectric background and thus may miss some solvent-induced behavior.

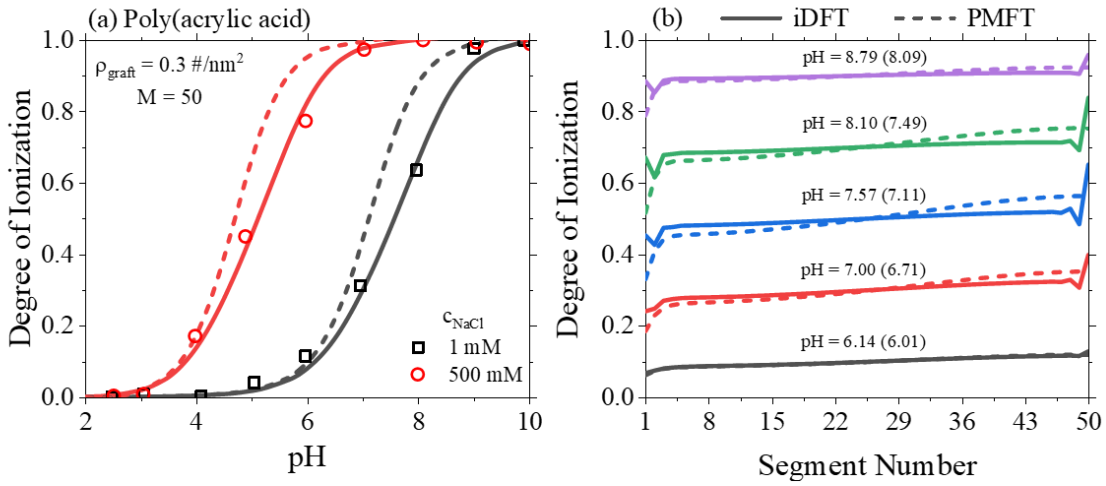

Figure 6.1 (a) Degree of ionization as a function of pH for a brush of poly(acrylic acid), PAA, at two salt concentrations. The grafting density of the brush is 0.3 #/nm^2 and each chain has 50 segments. The solid and dashed lines are calculated by iDFT and PMFT, respectively, while the symbols are from experiments[37]. (b) The segment-level degree of ionization at different average ionization of the polymer (α_{av}=0.1, 0.3, 0.5, 0.7, and 0.9) which correspond to the pH values given by iDFT (or PMFT) and a sodium chloride concentration of 1 mM.

Figure 1a shows a comparison of the theoretical results with experiment for the titration curves of the weak polyelectrolyte brush at 1 mM and 500 mM sodium chloride solutions[37]. In the DFT calculations, we assume that each polymer chain has 50 segments and the grafting density was estimated to be 0.3 #/nm^2 (viz. a moderately dense brush such that the brush height is greater than the distance between grafting points). While the polymer charge depends on the grafting density[14], it is relatively insensitive to the chain length, except at dilute conditions[38]. While iDFT is able to reproduce the experimental results at both salt concentrations, PMFT shows noticeable discrepancy because it ignores intrachain correlation due to the ionization of neighboring segments. In this case, the predictions by PDFT (not shown) are nearly indistinguishable from those by the iDFT indicating that the ionizable sites are only weakly correlated thus the weighted average of

the interaction with its neighbors is a reasonable approximation. The difference between iDFT and PMFT predictions remains significant even at high salt concentration, suggesting that the coupled effect of ionization and intrachain correlation cannot be accounted for only by the electrostatic interaction between neighboring segments. It should be noted that lateral inhomogeneity is not accounted for in our numerical implementation of different DFT methods. While such effects may be relevant to the properties of brush systems at low grafting densities, we expect that the importance of ionization and intrachain correlations will not be altered when the lateral inhomogeneity is explicitly considered.

Figure 1b presents the segment-level ionization of poly(acrylic acid) in 1 mM sodium chloride and a grafting density of 0.3 #/nm^2 for different degrees of average polymer ionization (α_{av}=0.1, 0.3, 0.5, 0.7, and 0.9, which are shown as different pH values predicted by iDFT and PMFT). Such information is not accessible through experiment, although some advancements have been made recently to discern the ionization degree of different regions within a brush[39]. The segment-level detail is important as the distribution of charged segments within a brush determines the brush properties (e.g., surface binding). Qualitatively, PMFT provides a correct description that the segments in the exterior of a brush are more likely to be charged than those located in the interior. However, iDFT predicts a more gradual increase in the segment ionization from the interior to exterior of the brush compared to PMFT, which is reminiscent to the segment-level details predicted by Monte Carlo simulations and analytical theories for bulk weak polyelectrolytes[40-42]. In particular, PMFT fails to capture the sharp increase in the end-segment degree of ionization. Previous Monte Carlo simulations have shown that the end-

segment ionization is much larger than those in the middle of a polymer chain[40]. The enhanced ionization of end segments is intuitive because a terminal segment experiences less intrachain repulsion by having only one nearest neighbor. Because the end segments are more likely to be ionized, their adjacent segments have lower degrees of ionization, leading to a small oscillation in the degree of ionization for segments near the end of the polymer brush. Clearly, PMFT is not able to describe the ionization behavior of the polymer brush at the segment-level because it neglects the coupling between ionization and intrachain correlations. Note that PMFT employed in this work is based on the tangent-chain model thus is different from typical self-consistent mean-field treatments which may utilize a different polymer model or sample conformations of a self-avoiding polymer[15]. Because different ways for representing the polymer backbone are mostly manifested in terms of the short-range interactions and segment-level details, we expect that the effects of ionization-coupling and the long-range correlations discussed in this work are equally applicable to other types of mean-field methods.

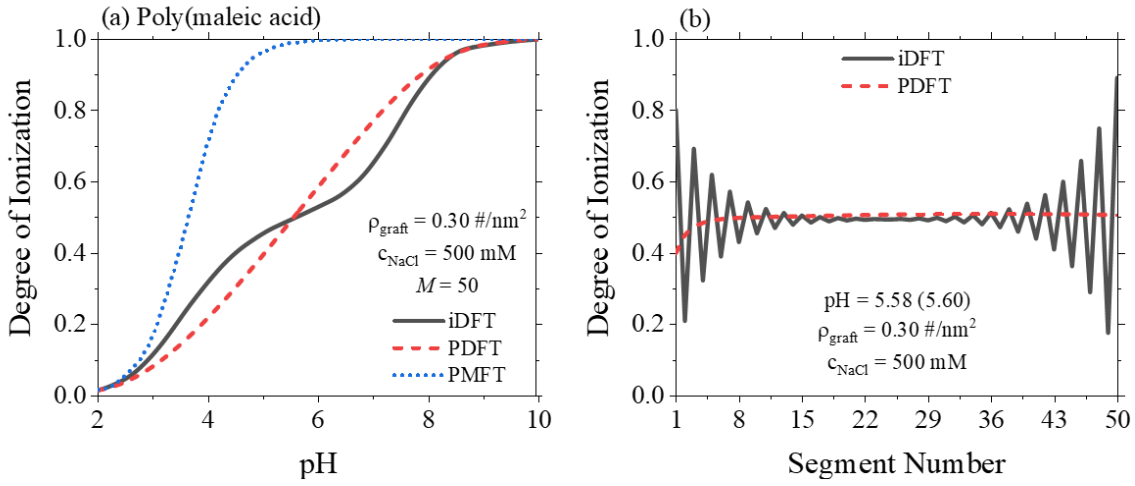

Figure 6.2 (a) The degree of ionization for a poly(maleic acid) (PMA) brush as a function of pH predicted by iDFT, PDFT, and PMFT as shown by the black solid, red dashed, and blue dotted lines, respectively. (b) The segment level degrees of ionization for PMA when the polymer is half charged on average from iDFT (solid black line) and PDFT (dashed red line). The pH corresponds to the iDFT (or PDFT) result for an average polymer degree of ionization of 0.50. The grafting density is 0.3 #/nm^2 and the sodium chloride concentration is 500 mM. The chain length of the polymer brush is 50.

The coupled effects of ionization and intrachain correlation would be more significant in a poly(maleic acid) (PMA) brush where the ionization of individual segments are strongly correlated. As mentioned previously, the titration of PMA in bulk solutions exhibits a step-like feature due to strong intrachain repulsion. This behavior can be described by a theoretical method that explicitly accounts for the charge status of neighboring segments. In Figure 2a, we show the degree of ionization of a PMA brush as a function of pH predicted by iDFT, PDFT and PMFT. As expected, both PMFT and PDFT fail to describe the ionization behavior of a strongly correlated polymer because PMFT neglects the intramolecular correlations while PDFT treats them only at the mean-field level (i.e., neglect the charge status of neighboring sites). Conversely, iDFT is able to describe the plateau in the titration curve in agreement with previous studies for strongly

correlated weak polyelectrolyte systems[18]. While PMFT clearly deviates from the iDFT predictions, PDFT captures at least the resistance to charging due to intrachain correlations, which are accounted for at the mean-field level [see Eq. (12)]. However, the mean-field approximation leads to noticeable disagreement with iDFT in terms of the degree of polymer ionization. Compared to iDFT, PDFT underpredicts the polymer ionization below the middle point but overpredicts at a higher degree of ionization. The erroneous behavior can be attributed to inaccuracy in treating the contribution of different charge states of the neighboring segments through a weighted contribution, particularly when the charges of the neighboring segments are strongly correlated. As shown in Figure 2b, the correlated ionization effect is exhibited by the segment-level degrees of ionization. While PAA shows only a small oscillation of ionization extending from the end segments (see Figure 1b), PMA exhibits a strong correlation over the entire polymer chain. The strong correlation between neighboring segments leads to the end segment being over five times more likely to be ionized than its neighbor. By incorporating the multiple states of each polymer segments and the intrachain correlations, iDFT provides a faithful description of both weakly and strongly correlated ionization in weak polyelectrolytes. It is worth noting that the ionization of maleic acid as a molecule is different from that as a monomer in PMA due to the loss in the double bond which leads to a different pKa value. The PMA titration proceeds in a two-step fashion because of the strong intramolecular correlation between the two -COOH sites which resists ionization of the second -COOH site. The two pKa values reported in some previous work might be attributed to the neglect of this nearest-neighbor correlations. In this work, we use only a single pKa value for the -COOH site

because our model accounts for nearest-neighbor interactions and the difference between racemic and mesomeric bonds. Our pKa value for PMA is similar to that used in a previous description of the PMA titration in the bulk with a site-binding model[36].

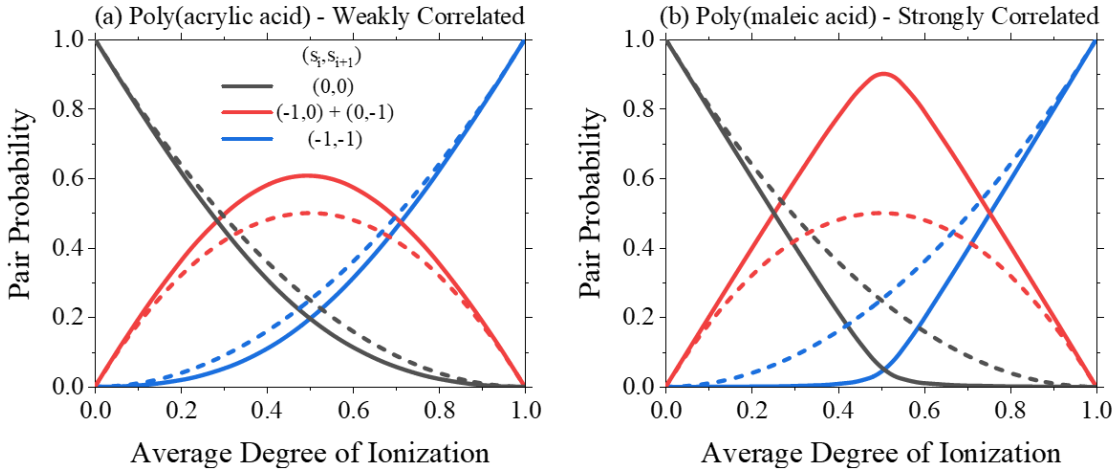

Figure 6.3 Probability of two consecutive segments in a polyacid chain being in different charge states (as described by different pairs of s_i and s_{i+1}) as a function of the average degree of ionization for the polymer in (a) a PAA brush and (b) a PMA brush. The solid lines are calculated from iDFT while the dashed lines are from PDFT. Here the sodium chloride concentration is 500 mM.

The coupling between ionization and intrachain correlation is also evident in the probability distribution of consecutive paired segments at different charge states. In a mean-field theory, the correlation between the charge states of consecutive segments is neglected. As a result, the probability of finding two consecutive sites in states s_i and s_{i+1} is simply given by $\alpha_i^{MF}(s_i)\times\alpha_{i+1}^{MF}(s_{i+1})$. As shown in Figure 3, PMFT predicts that the probability of segments *i* and *i*+1 with only one charge occurs when half of the polymer segments are charged (*viz.* 50% probability of occurring when both segments are on average half-charged). On the other hand, intrachain correlation results in the alternating

distribution of charges to avoid unfavorable electrostatic interactions (e.g., two consecutive negatively-charged sites). In comparison with PMFT, iDFT predicts a lower probability of the paired state where both monomers have the same charge, but a higher probability of the paired state where only one monomer is charged. Unsurprisingly, the pair probabilities for PAA predicted by iDFT do not differ significantly from the mean-field description because of the weakly correlated ionization. Nonetheless, the intrachain repulsion does result in a different charge distribution of the weak polyelectrolyte. When there is strong correlation between ionization of consecutive sites (e.g., PMA), the distribution of the paired states heavily favors those with only one charged monomer in the pair. Interestingly, at the point in which the polymer is half-charged, the polymer contains essentially no consecutive neutral (or charged) sites. The correlated ionization explains the plateau in the titration curve shown in Figure 2a. Beyond this point, further charging of the polymer must overcome a stronger intrachain repulsion from its two neighboring sites.

Figure 3b also explains why PDFT fails to capture the titration behavior of PMA. At the low degree of polymer ionization, iDFT predicts that only non-consecutive monomers will ionize to avoid the strong intrachain repulsion. However, PDFT cannot account for this behavior because the coupling between ionization of consecutive segments and intrachain correlation is not included. Instead, the intrachain repulsion depends on the ionization of segment *i* in the mean-field approximation for the ionization degree of the neighboring segments. PDFT underpredicts the ionization of PMA below the middle point because it overpredicts the intrachain repulsion. The intersection of the iDFT and PDFT titration curves near 50% ionization results from a similarity in intrachain repulsion felt by

each monomer (despite the otherwise incorrect description of the physics in PDFT). When the neighboring monomers are 50% ionized, PDFT predicts that a charged monomer will experience two half-contributions for the intrachain repulsion due to the weighted average approximation [see Eq. (12)]. On the other hand, iDFT predicts the transition from no immediate intrachain repulsion when a monomer is ionized below 50% of chain ionization to instead repulsions from two immediate neighbors above 50% chain ionization. Thus, there is a transition point at 50% ionization in which the monomer experiences on average one intrachain repulsion equivalent to that predicted by PDFT. The intersection of the two curves when the polymer is half-charged is in agreement with the conventional nearest-neighboring site-binding model and its mean-field approximation[18]. Beyond this point, iDFT predicts a lower degree of ionization compared to PDFT since further charging is limited by repulsions from two immediately neighboring charged segments while PDFT accounts for only an average repulsion.

6.4 Conclusion

In summary, iDFT provides a theoretical framework to account for conformation-ionization coupling in weak polyelectrolyte systems that cannot be captured with conventional theoretical methods. It shows an accurate description of both intrachain correlation and thermodynamic effects on ionization of individual segments. Such effects are important for understanding charge regulation in hetero-weak polyelectrolytes in particular for zwitterionic polymers and polypeptides that will be studied in our future work. While the discussion above is focused on a coarse-grained model, similar procedures can also be established for other weak-polyelectrolyte models including those with

atomistic details by incorporating the Ising degree of freedom to account for the charge states of ionizable sites. Because of its versatility to account for the multiple charge states of individual segments and polymer configuration self-consistently, we expect that iDFT will be useful for more realistic models of weak polyelectrolyte systems including those containing stereochemistry-dependent polymers.

Supporting Information

S6.1 *Numerical implementation for weak polyelectrolyte brush*

For systems with the density distributions changing only in the z-direction, we can integrate out the x- and y-direction to greatly relieve the computational burden. The density profile of the monomers in a polymer chain can then be expressed as

$$\rho_{k,i}(z,s_{k,i},s_{k,i+1}) = \exp\left[\beta\mu_k - \beta\lambda_{k,i}\left(z,s_{k,i}\right)\right] \times\left[\sum_{\{s_{k,i-1}\}} G^L_{k,i}\left(z,s_{k,i-1},s_{k,i}\right)\right] G^R_{k,i}\left(z,s_{k,i},s_{k,i+1}\right) \tag{S6.1}$$

and

$$\rho_{k,i+1}(z,s_{k,i},s_{k,i+1}) = \exp\left[\beta\mu_k - \beta\lambda_{k,i+1}\left(z,s_{k,i+1}\right)\right] \times G^L_{k,i+1}\left(z,s_{k,i},s_{k,i+1}\right)\left[\sum_{\{s_{k,i+2}\}} G^R_{k,i+1}\left(z,s_{k,i+1},s_{k,i+2}\right)\right] . \tag{S6.2}$$

The propagator functions are given as

$$G^L_{k,i}\left(z,s_{k,i-1},s_i\right) = \exp\left[-\beta\tau_{k,i}\left(z,s_{k,i-1},s_{k,i}\right)/2 + \beta u^{\infty}_k\left(s_{k,i-1},s_{k,i}\right)\right] \times\int \exp\left[-\beta\lambda_{k,i-1}\left(z',s_{k,i-1}\right)\right]\frac{\theta\left[|z-z'|-\sigma_k\left(s_{k,i-1},s_{k,i}\right)\right]}{2\sigma_k\left(s_{k,i-1},s_{k,i}\right)} \times\exp\left[-\beta\tau_{k,i-1}\left(z',s_{k,i-1},s_{k,i}\right)/2\right]\sum_{\{s_{k,i-2}\}} G^L_{k,i-1}\left(z',s_{k,i-2},s_{k,i-1}\right)dz' \tag{S6.3}$$

for i=3, …, M_k with

$$
\begin{aligned}
G_{k,2}^{L}\left(z,s_{k,i-1},s_{i}\right) = & \exp\left[-\beta\tau_{k,i}\left(z,s_{k,i-1},s_{k,i}\right)/2\right] \\
& \times\exp\left[-\beta\tau_{k,i-1}\left(\sigma_{k}\left(s_{k,i-1}\right)/2,s_{k,i-1},s_{k,i}\right)/2\right] \\
& \times\exp\left[\beta u_{k}^{\infty}\left(s_{k,i-1},s_{k,i}\right)\right] \\
& \times\exp\left[-\beta\lambda_{k,i-1}\left(\sigma_{k}\left(s_{k,i-1}\right)/2,s_{k,i-1}\right)\right] \\
& \times\frac{\theta\left[|\,z-\sigma_{k}\left(s_{k,i-1}\right)/2\,|-\sigma_{k}\left(s_{k,i-1},s_{k,i}\right)\right]}{2\sigma_{k}\left(s_{k,i-1},s_{k,i}\right)}
\end{aligned}
\tag{S6.4}
$$

and

$$
\begin{aligned}
G_{k,i}^{R}\left(z,s_{k,i},s_{k,i+1}\right) = & \exp\left[-\beta\tau_{k,i}\left(z,s_{k,i},s_{k,i+1}\right)/2+\beta u_{k}^{\infty}\left(s_{k,i},s_{k,i+1}\right)\right] \\
& \times\int\exp\left[-\beta\lambda_{k,i+1}\left(z',s_{k,i+1}\right)\right]\frac{\theta\left[|\,z-z'\,|-\sigma_{k}\left(s_{k,i},s_{k,i+1}\right)\right]}{2\sigma_{k}\left(s_{k,i},s_{k,i+1}\right)} \\
& \times\exp\left[-\beta\tau_{k,i+1}\left(z',s_{k,i},s_{k,i+1}\right)/2\right]\sum_{\left\{s_{k,i+2}\right\}}G_{k,i+1}^{R}\left(z',s_{k,i+1},s_{k,i+2}\right)dz'
\end{aligned}
\tag{S6.5}
$$

for i=1,…,M_k-1.

The density distribution of the first polymer segment in the brush (i.e., the segment attached to the wall) is given by

$$
\rho_{k,1}\left(z\right)=\sum_{\left\{s_{k,1}\right\}}\rho_{k,1}\left(z,s_{k,1}\right)\delta\left[z-\frac{\sigma_{k}\left(s_{k,1}\right)}{2}\right] \tag{S6.6}
$$

where $\sum_{\{s_{k,1}\}}\rho_{k,1}(\sigma_k(s_{k,1})/2,s_{k,1})=\rho_{graft}$. The constant $\exp(\beta\mu_k)$ in Eq. (S56) and (S57) must be determined by the normalization condition applied to any segment, i.e., $\int\rho_{k,i}(z)dz=\rho_{graft}$ [43]. The density profile for free ions is given by

$$
\rho_{\alpha}(z)=\exp\left\{-\beta[V_{\alpha}^{ext}(z)+\mu_{\alpha}^{ex}(z)-\mu_{\alpha}]\right\} \tag{S6.7}
$$

where the external potential (applicable to both free ions and polymer segments) is given by

$$\beta V_i^{ext}(z) = \begin{cases} \infty & z < \sigma_i / 2 \\ 0 & else \end{cases} . \tag{S6.8}$$

The mean electrostatic potential for a planar charged wall is given by[44]

$$\beta e\psi(z) = \beta e\psi_s + 4\pi l_B z \int_z^\infty \sum_i \rho_i(z') Z_i dz' + 4\pi l_B \int_0^z z' \sum_i \rho_i(z') Z_i dz' \tag{S6.9}$$

where ψ_s is the surface electrostatic potential. However, in the system of interest, we consider the surface charge to be fixed (*viz.* $Q = 0$) and not the surface potential. This can be self-consistently determined by choosing ψ_s to satisfy the electroneutrality condition:

$$\int_0^\infty \sum_i \rho_i(z) Z_i e \, dz + Q = 0 . \tag{S6.10}$$

Once the density distributions of all polymer segments and ions is known, we can determine the degree of ionization for segment *i* in the polymer brush *k* by

$$\alpha_{k,i}(s_{k,i}) = \frac{\int_0^\infty \rho_{k,i}(z, s_{k,i}) dz}{\rho_{graft}} . \tag{S6.11}$$

The probability of consecutive monomers in the polymer existing in the paired state *s* and s' is given by

$$P_k(s, s') = \frac{\sum_{i=1}^{M_k - 1} \int_0^\infty \rho_{k,i}(z, s, s') dz}{(M_k - 1)\rho_{graft}} . \tag{S6.12}$$

For a weak polyacid, the possible paired states of s and s' are [-1,-1], [-1,0], [0,-1], and [0,0]. Lastly, the brush height can be determined from[45]

$$H = \frac{2\sum_{i=1}^{M_k}\int_0^\infty z\rho_{k,i}(z)\,dz}{M_k\rho_{graft}}. \qquad \text{(S6.13)}$$

S6.2. Additional Technical Details and Results:

We employ the primitive model for polyelectrolyte solutions which treat monomers and free ions as charged hard spheres each with their own unique diameter and the solvent as a dielectric continuum. The implicit treatment of the solvent can miss unique solvent-induced behavior; however, for the present work, this treatment was sufficient. If necessary, the solvent could be treated explicitly like in many self-consistent field theories. Besides, the parameters employed here account for solvent-mediated interactions as demonstrated by comparison to experimental results for the activity coefficients in bulk solutions. The ion sizes used in this work were determined previously by comparison to experimental data for activity coefficients (viz. 3.22 Å and 3.62 Å for sodium and chloride ion, respectively). Going beyond the primitive model, we could consider ion-specific non-electrostatic interactions; although this was beyond the scope of current work. Both protons and hydroxyl ions were considered explicitly (which are represented by charged hard spheres of diameter 5.0 Å and 3.22 Å, respectively) whose bulk concentration values are determined based on the relation $\mathrm{pH} = -\log a_{H^+} = -\log c^b_{H^+} - \beta\mu^{ex,b}_{H^+}/\ln 10$ and $\mathrm{pOH} = -\log a_{OH^-} = -\log c^b_{OH^-} - \beta\mu^{ex,b}_{OH^-}/\ln 10$, where $\mu_i^{ex,b}$ is the excess chemical potential of each in the bulk.

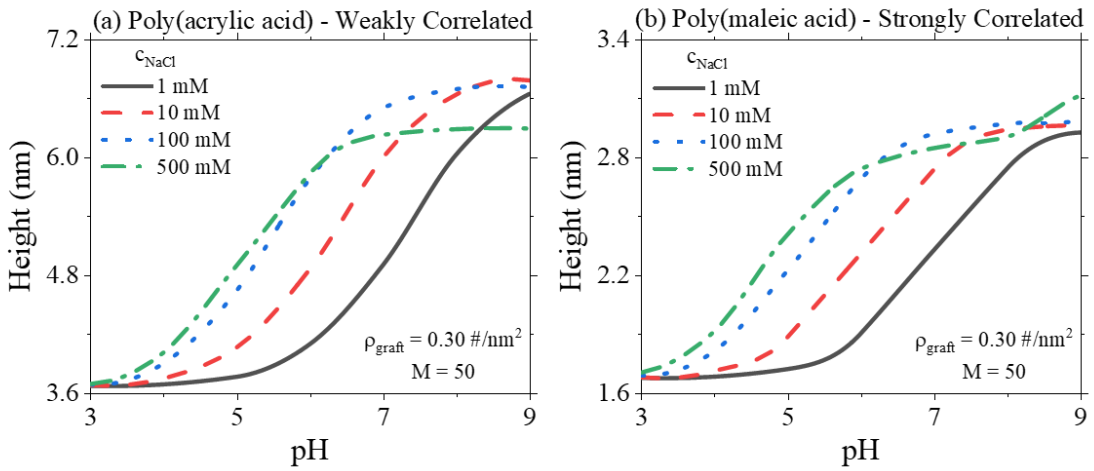

Figure S 6.1 The influence of pH on the height of the (a) PAA brush and (b) PMA brush at different salt concentrations for M=50 and $\rho_{graft} = 0.30\#/nm^2$.

Figure S1shows the how the height of the polyacid brush is influenced by the pH (or degree of ionization. For PAA, we see that as the pH increases, the brush height will also increase. This is reasonable since the degree of ionization will also increase with pH and the brush should extend due to the repulsive coulomb interactions. Since the ionization occurs earlier at higher salt concentrations, we see that the height of the brush increases at a lower pH as the salt concentration increases. When the weak polyelectrolyte is fully charged (viz. high pH), the height decreases with salt concentration as expected. The PMA brush shows a plateau in its height similar to what is seen for its ionization (see Figure 2a). Since the ionization of the brush changes little in this region, it is not surprising that the brush height remains constant since the charge is approximately constant.

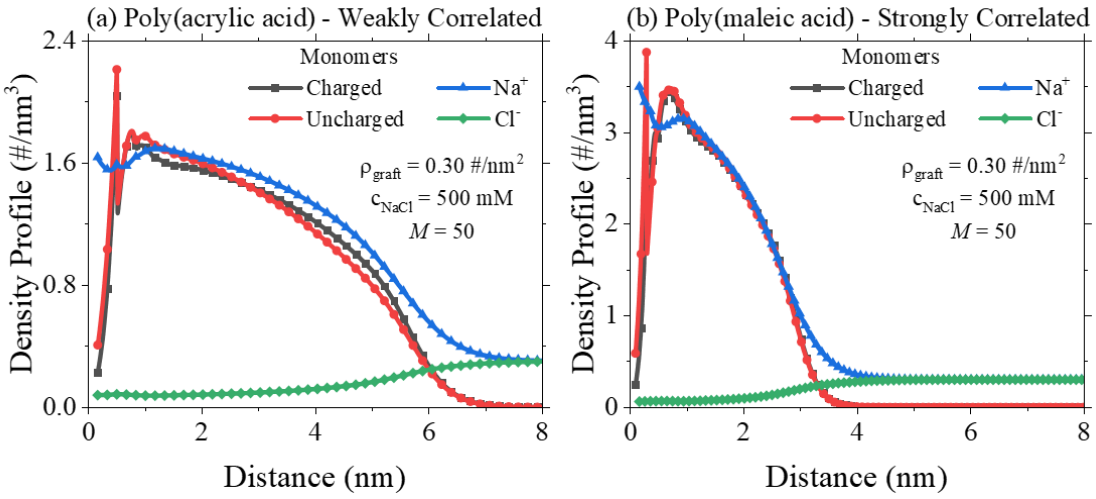

Figure S 6.2 The density profile at $\rho_{graft} = 0.30\,\#/\,nm^2$, c_{NaCl} *= 500 mM, and M=50 for the (a) PAA brush and (b) PMA brush systems at 50% average degree of ionization (pH 5.11 and 5.58, respectively).*

We next consider the density profile of the half-charged PAA (or PMA) brush and the salt ions at $\rho_{graft} = 0.30\,\#/\,nm^2$ and c_{NaCl} = 500 mM in Figure S2. The brush shows an extended structure due to volume exclusion effects and interchain coulomb interactions. In addition, the distribution of the charged and uncharged segments is relatively similar. We see that there is an accumulation of counterions near the surface and maintain a significant presence throughout the brush profile. However, the charge of the brush is not fully compensated by the counterions as evident by the higher presence of counterions beyond the brush region. In addition, the coions are not fully excluded from the inner regions of the brush likely because of favorable correlations with the counterions present. While some ion pairing or ion condensation is likely present at such conditions, it appears that the counterions are unable to satisfy electroneutrality within the brush layer.

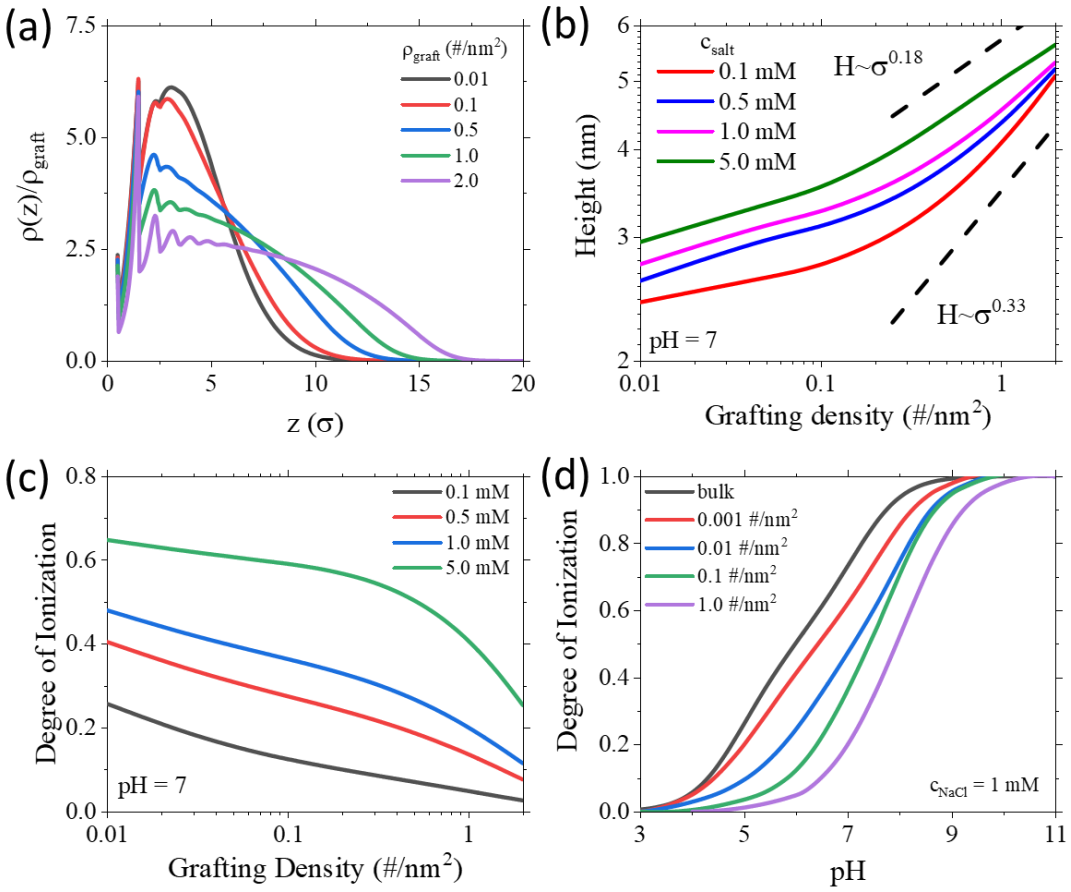

Figure S 6.3 (a) The influence of grafting density on the PAA brush at different salt concentrations for pH=7. (b) The normalized density profile of the PAA brush for different grafting densities for c_{NaCl} = 1 mM and pH = 7. (c) The degree of ionization for PAA as a function of grafting density for different salt concentrations at pH = 7. (d) The degree of ionization for PAA as a function of pH for different grafting densities at c_{NaCl} = 1 mM. In this case, the number of segments is 25 in agreement with Hollingsworth et al.[46]

We show in Figure S3a the brush density profile as a function of distance at different grafting densities when the pH=7 and the salt concentration is 1 mM. Since we consider excluded volume effects through our hard-sphere model using MFMT, we also find these strong oscillations in the brush profile resulting from the dense packing of the monomer segments. As the grafting density increases, the volume available to a single chain decreases and so the chain must extend further away from the surface. As a result,

the height will increase with grafting density despite the lower charge of the brush due to packing constraints. When the grafting density is low (e.g., 0.01 #/nm^2), the brush exhibits the typical parabolic profile expected from previous scaling analysis[11]. As the grafting density is increased, the oscillations in the density profile exhibit a stronger amplitude and a slower decay (i.e., the excluded volume effects are still of significance even beyond the first few monomers). Without the hard-sphere repulsion, these strong oscillations would not be present, and the brush would decrease in height with grafting density as a result of the decrease in its ionization.

We show in Figure S3b the scaling relation of the brush height with grafting density at different salt concentrations in the osmotic regime at pH=7. The conventional methods predict either a decrease or no change in brush height at these conditions because the polymer ionization decreases with grafting density and the excluded volume effects are neglected or weakly accounted for (e.g., incompressibility of the fluid)[15]. On the other hand, iDFT agrees with experiments that the brush height scales with grafting density by a factor ranging from 0.18 to 0.33 depending on the salt concentration[46]. To understand why the scaling exponent tends to decrease with salt concentration, we first note that at a fixed grafting density, the brush will always be more charged when the salt concentration is increased and is therefore also greater in height. However, when the brush becomes sufficiently dense (i.e., ρ_{graft} =2 #/nm^2), it will behave similar to a neutral brush because it is only weakly ionized. Therefore, the salt concentration is of little significance as seen by the lines collapsing together in Figure S2b. Since the brush charge is reduced at lower salt concentrations, a further decrease in ionization resulting from an increase in grafting

density will be of minor importance. On the other hand, at higher salt concentrations, the decrease in ionization with grafting density will play a contributing role in determining the brush height since there will be a corresponding reduction in repulsive electrostatic interactions. Since the decrease in ionization will limit the increase in height resulting from excluded volume effects, the scaling factor will decrease with salt concentration. When excluded volume effects are not well accounted for, the charge of the polymer determines its brush height and thus will lead to the collapse of the brush with increase in grafting density. Thus, an accurate description of weak polyelectrolyte brushes requires the correct treatment of excluded volume effects.

The degree of ionization is a key factor in determining the height of the PAA brush. We show how the grafting density influences the degree of ionization at different salt concentrations at pH=7 in Figure S3c. As can be expected, increasing the grafting density results in a decrease in the ionization of the polymer. This is due to an increase in interchain interactions (i.e., the electrostatic interactions between other chains). In addition, it becomes more difficult for the counterions to compensate the local charge. While salt concentration plays a noticeable role at low grafting density, at high grafting density the lines approach one another. This is because the ionization of the polymer becomes insignificant (viz. neutral chain) so the presence of salt ions is no longer relevant. We show how grafting density effects the entire titration curve of the PAA brush in Figure S4d. The grafting density plays a noticeable role in the early stage of charging in the brush. As the grafting density increases, the pH at which the polymer brush reaches 10% charged noticeably shifts. Further ionization shows a similar change in charge with pH at different

grafting densities. The reason for the influence of grafting density during the initial charging process is that the charge density is significant even at the initial stage and there are not sufficient salt ions to screen these interactions. This effect is less noticeable as the bulk salt concentration is increased since these initial long-range interactions are more easily screened.

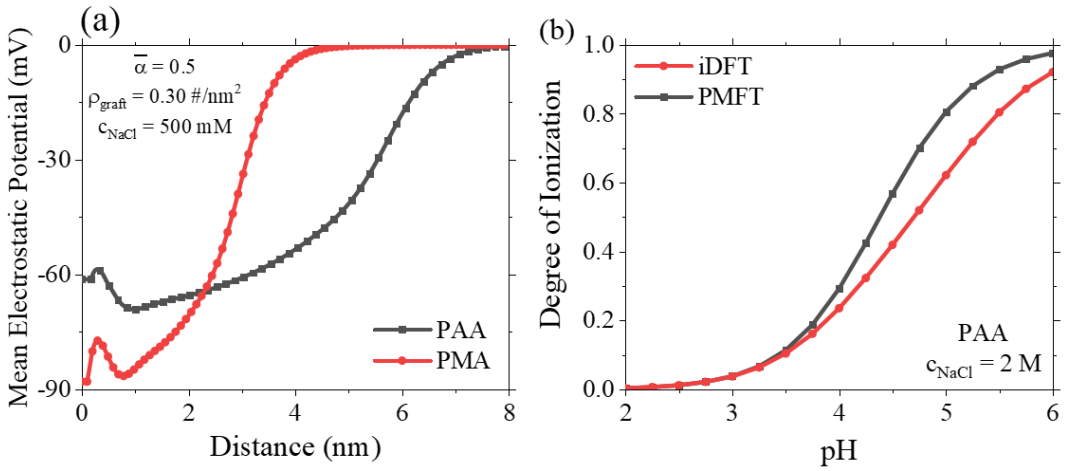

Figure S 6.4 (a) The mean electrostatic potential for the 50% charged PAA and PMA brush at c_{NaCl} = 500 mM and $\rho_{graft} = 0.30\,\#/\,nm^2$. The corresponding pH to 50% charged is 5.11 and 5.58, respectively. (b) The degree of ionization predicted by iDFT and PMFT as a function of pH at 2 M sodium chloride concentration.

In Figure S4a, we show the mean electrostatic potential as a function of distance for the PAA and PMA brush when half-charged at 500 mM salt concentration and $\rho_{graft} = 0.10\,\#/\,nm^2$. The surface potential is larger for the PMA brush than PAA brush due to the larger charge density of PMA since the monomers are half the diameter of PAA's monomer. The large draw of counterions to the brush layer leads to a quicker decay of the electrostatic potential for PMA compared to PAA. Despite the strong oscillations (i.e., zig-zag pattern) shown in the segment-level ionization of PMA, the electrostatic potential

remains smooth. This is because the segments are not localized to a discrete region and neither are the counterions so the net charge remains consistently negatively charged until after the brush layer where the excess counterions compensate the charge. The nearest-neighbor segment interaction decays slowly at high salt concentration [$\ln y \sim 1/(1+\Gamma\sigma)^2$; see Eq.(S41)] leading to a sufficient interaction between neighboring segments even at 2M salt concentration as shown in Figure S4b.

Bibliography

1. Gonzalez Solveyra, E., et al., *Theoretical Modeling of Chemical Equilibrium in Weak Polyelectrolyte Layers on Curved Nanosystems.* Polymers, 2020. **12**(10): p. 2282.

2. Landsgesell, J., et al., *Simulations of ionization equilibria in weak polyelectrolyte solutions and gels.* Soft Matter, 2019. **15**(6): p. 1155-1185.

3. Yuan, W., et al., *Weak polyelectrolyte-based multilayers via layer-by-layer assembly: Approaches, properties, and applications.* Advances in Colloid and Interface Science, 2020. **282**: p. 102200.

4. Yap, H.P., et al., *Colloid Surface Engineering via Deposition of Multilayered Thin Films from Polyelectrolyte Blend Solutions.* Langmuir, 2005. **21**(10): p. 4328-4333.

5. Tong, W., C. Gao, and H. Möhwald, *Stable Weak Polyelectrolyte Microcapsules with pH-Responsive Permeability.* Macromolecules, 2006. **39**(1): p. 335-340.

6. Ilyas, S., et al., *Weak polyelectrolyte multilayers as tunable membranes for solvent resistant nanofiltration.* Journal of Membrane Science, 2016. **514**: p. 322-331.

7. Biesheuvel, P.M., M. van der Veen, and W. Norde, *A Modified Poisson−Boltzmann Model Including Charge Regulation for the Adsorption of Ionizable Polyelectrolytes to Charged Interfaces, Applied to Lysozyme Adsorption on Silica.* The Journal of Physical Chemistry B, 2005. **109**(9): p. 4172-4180.

8. Zhulina, E.B. and O.V. Borisov, *Poisson–Boltzmann Theory of pH-Sensitive (Annealing) Polyelectrolyte Brush.* Langmuir, 2011. **27**(17): p. 10615-10633.

9. Zhulina, E.B., T.M. Birshtein, and O.V. Borisov, *Theory of Ionizable Polymer Brushes.* Macromolecules, 1995. **28**(5): p. 1491-1499.

10. Israels, R., et al., *Charged Polymeric Brushes: Structure and Scaling Relations.* Macromolecules, 1994. **27**(12): p. 3249-3261.

11. Lyatskaya, Y.V., et al., *Analytical self-consistent-field model of weak polyacid brushes.* Macromolecules, 1995. **28**(10): p. 3562-3569.

12. Willott, J.D., et al., *Behavior of Weak Polyelectrolyte Brushes in Mixed Salt Solutions.* Macromolecules, 2018. **51**(3): p. 1198-1206.

13. Witte, K.N., S. Kim, and Y.-Y. Won, *Self-Consistent Field Theory Study of the Effect of Grafting Density on the Height of a Weak Polyelectrolyte Brush.* The Journal of Physical Chemistry B, 2009. **113**(32): p. 11076-11084.

14. Gong, P., J. Genzer, and I. Szleifer, *Phase Behavior and Charge Regulation of Weak Polyelectrolyte Grafted Layers.* Physical Review Letters, 2007. **98**(1): p. 018302.

15. Gong, P., et al., *Behavior of Surface-Anchored Poly(acrylic acid) Brushes with Grafting Density Gradients on Solid Substrates: 2. Theory.* Macromolecules, 2007. **40**(24): p. 8765-8773.

16. Tagliazucchi, M., et al., *Self-Organized Polyelectrolyte End-Grafted Layers Under Nanoconfinement.* ACS Nano, 2014. **8**(10): p. 9998-10008.

17. Sadeghpour, A., et al., *Influence of alkali metal counterions on the charging behavior of poly(acrylic acid).* Polymer, 2009. **50**(16): p. 3950-3954.

18. Koper, G.J.M. and M. Borkovec, *Proton binding by linear, branched, and hyperbranched polyelectrolytes.* Polymer, 2010. **51**(24): p. 5649-5662.

19. Lützenkirchen, J., et al., *Comparison of Various Models to Describe the Charge−pH Dependence of Poly(acrylic acid).* Journal of Chemical & Engineering Data, 2011. **56**(4): p. 1602-1612.

20. Léonforte, F., U. Welling, and M. Müller, *Single-chain-in-mean-field simulations of weak polyelectrolyte brushes.* The Journal of Chemical Physics, 2016. **145**(22): p. 224902.

21. Forsman, J. and C.E. Woodward, *Classical Density Functional Theory of Polymer Fluids*, in *Variational Methods in Molecular Modeling*, J. Wu, Editor. 2017, Springer Singapore : Imprint: Springer,: Singapore. p. 101-136.

22. Wu, J. and Z. Li, *Density-functional theory for complex fluids.* Annu. Rev. Phys. Chem., 2007. **58**: p. 85-112.

23. Nakamura, I. and Z.-G. Wang, *Salt-doped block copolymers: ion distribution, domain spacing and effective χ parameter.* Soft Matter, 2012. **8**(36): p. 9356-9367.

24. Wertheim, M.S., *Thermodynamic perturbation theory of polymerization.* The Journal of Chemical Physics, 1987. **87**(12): p. 7323-7331.

25. Yu, Y.-X. and J. Wu, *A modified fundamental measure theory for spherical particles in microchannels.* The Journal of Chemical Physics, 2003. **119**(4): p. 2288-2295.

26. Gillespie, D., M. Valiskó, and D. Boda, *Density functional theory of the electrical double layer: the RFD functional.* Journal of Physics: Condensed Matter, 2005. **17**(42): p. 6609-6626.

27. Kierlik, E. and M.L. Rosinberg, *Perturbation density functional theory for polyatomic fluids. III. Application to hard chain molecules in slitlike pores.* The Journal of Chemical Physics, 1994. **100**(2): p. 1716-1730.

28. Li, Z. and J. Wu, *Density Functional Theory for Planar Electric Double Layers: Closing the Gap between Simple and Polyelectrolytes.* The Journal of Physical Chemistry B, 2006. **110**(14): p. 7473-7484.

29. Okrugin, B.M., et al., *Electroresponsive Polyelectrolyte Brushes Studied by Self-Consistent Field Theory*. Polymers, 2020. **12**(4): p. 898.

30. Jiang, J., V.V. Ginzburg, and Z.-G. Wang, *Density functional theory for charged fluids.* Soft Matter, 2018. **14**(28): p. 5878-5887.

31. Jiang, T. and J. Wu, *Self-organization of multivalent counterions in polyelectrolyte brushes.* The Journal of Chemical Physics, 2008. **129**(8): p. 084903.

32. Borówko, M., et al., *Adsorption of ions on surfaces modified with brushes of polyampholytes.* The Journal of Chemical Physics, 2012. **137**(7): p. 074707.

33. Gallegos, A. and J. Wu, *Charge Regulation of Natural Amino Acids in Aqueous Solutions.* Journal of Chemical & Engineering Data, 2020. **65**(12): p. 5630-5642.

34. Gallegos, A., G.M.C. Ong, and J. Wu, *Thermodynamic non-ideality in charge regulation of weak polyelectrolytes.* Soft Matter, 2021. **17**: p. 9221–9234.

35. Kawaguchi, S., et al., *Dissociation behavior of poly (fumaric acid) and poly (maleic acid). II. Model calculation.* Macromolecules, 1990. **23**(3): p. 731-738.

36. de Groot, J., et al., *Dissociation Behavior of Poly(maleic acid): Potentiometric Titrations, Viscometry, Pulsed Field Gradient NMR, and Model Calculations.* Macromolecules, 1998. **31**(13): p. 4182-4188.

37. Ferrand-Drake del Castillo, G., R.L.N. Hailes, and A. Dahlin, *Large Changes in Protonation of Weak Polyelectrolyte Brushes with Salt Concentration—Implications for Protein Immobilization.* The Journal of Physical Chemistry Letters, 2020. **11**(13): p. 5212-5218.

38. Laguecir, A., et al., *Size and pH effect on electrical and conformational behavior of poly(acrylic acid): Simulation and experiment.* European Polymer Journal, 2006. **42**(5): p. 1135-1144.

39. Ehtiati, K., et al., *How Dissociation of Carboxylic Acid Groups in a Weak Polyelectrolyte Brush Depend on Their Distance from the Substrate*. Langmuir, 2020. **36**(9): p. 2339-2348.

40. Uhlík, F., et al., *Modeling of Ionization and Conformations of Starlike Weak Polyelectrolytes.* Macromolecules, 2014. **47**(12): p. 4004-4016.

41. Berghold, G., P. van der Schoot, and C. Seidel, *Equilibrium charge distribution on weak polyelectrolytes.* The Journal of Chemical Physics, 1997. **107**(19): p. 8083-8088.

42. Castelnovo, M., P. Sens, and J.-F. Joanny, *Charge distribution on annealed polyelectrolytes.* The European Physical Journal E, 2000. **1**(2-3): p. 115-125.

43. Wei, Z., et al., *Density Functional Theory of Polymer Structure and Conformations.* Polymers, 2016. **8**(4): p. 121.

44. Boda, D. and D. Gillespie, *Calculating the electrostatic potential profiles of double layers from simulation ion density profiles.* Hungarian Journal of Industry and Chemistry, 2013. **41**(2): p. 123-130.

45. Jiang, T., Z. Li, and J. Wu, *Structure and Swelling of Grafted Polyelectrolytes: Predictions from a Nonlocal Density Functional Theory.* Macromolecules, 2007. **40**(2): p. 334-343.

46. Hollingsworth, N.R., S.I. Wilkanowicz, and R.G. Larson, *Salt- and pH-induced swelling of a poly(acrylic acid) brush via quartz crystal microbalance w/dissipation (QCM-D).* Soft Matter, 2019. **15**(39): p. 7838-7851.

Chapter 7. Molecular thermodynamics for polypeptide adsorption at inorganic surfaces

A faithful description of polypeptide adsorption remains a theoretical challenge from a molecular perspective due to the strong coupling of local thermodynamic non-ideality and inhomogeneous ionization of the adsorbate and substrate sensitive to the solution condition such as pH, ion valence, and salt concentration. Building upon a recently developed coarse-grained model for natural amino acids in bulk electrolyte solutions, here we report a molecular theory to predict polypeptide adsorption on ionizable inorganic surfaces over a broad range of solution conditions. Our thermodynamic model is able to account for amino-acid sequence effects on physical binding as well as surface associations such as hydrogen bonding or bidentate coordination. It is applicable to polypeptides containing any type of residues regardless of the solution pH, salt type and concentration. The theoretical predictions have been validated by extensive comparison with experimental data for the adsorption of polypeptides at titanium surfaces.

7.1 Introduction

Understanding the thermodynamic properties of polypeptides at interface is of critical importance to a myriad of real-world problems and technological applications such as surface fouling in biomedical devices and food systems[1-3], adhesion in wet environments[4, 5], and controlled drug delivery[6-8]. A theoretical description of polypeptides is complicated because of the diversity in possible amino-acid sequences and the flexibility in microscopic structure and charge behavior sensitive to the local environment. Under inhomogeneous conditions, each residue exhibits unique interactions

with other residues and chemical species in the solution as well as with the interface. To account for the sequence effect, a theoretical model must distinguish segment-level interactions including different charge states of ionizable residues. For adsorption at inorganic surfaces, the model must incorporate possible ionization of the substrate and the inherent coupling between the chemical equilibrium and local ion distributions. To complicate the matter further, the behavior of both the polypeptide and the surface is sensitive to the solution conditions such as pH, ion valence, and salt concentration.

Experimental investigations of polypeptides at interface provide a useful guide in the development of a satisfactory thermodynamic model. In particular, a wide variety of techniques have been used to study the interfacial behavior of polypeptides at inorganic surfaces[9, 10]. For example, phage display[11], quartz crystal microbalance measurement[12], ellipsometry[13], and surface plasmon resonance (SPR)[14] are a few methods that are commonly used to determine the affinity of peptides binding with specific surfaces. Alternatively, single molecule force spectroscopy (SFMS) provides insight about the interaction between each peptide and an inorganic surface at the single molecular level, thereby allowing for a detailed study of molecular interactions not available with other techniques[15]. Maity et al. took advantage of SFMS to identify the sequence effect by replacing a proline residue in a polypeptide chain with an alanine residue and found that the mutation had a detrimental impact on the binding capability with a mica surface[16]. The reduced binding was attributed to differences in the conformational freedom of the polypeptide when the residue was changed from proline to alanine. The distinctive cyclic structure of the proline residue leads to a high conformational rigidity in the polypeptide

chain compared to that with the alanine residue. The experimental results thus unequivocally demonstrate the importance of segment-level details in determining polypeptide adsorption. The importance of intramolecular interactions in polypeptides and the resulting multibody correlations makes polypeptide adsorption distinctively different from their monomeric counterparts (viz. amino acids). Another technique that is commonly utilized to evaluate the affinity of peptides with inorganic surfaces is through peptide arrays. The combinatory procedure provides a quick evaluation of the surface affinity by measuring the change in the color of the solution due to the peptide binding[17]. The resulting color changes can then be mapped into binding scores to unravel the mechanisms underlying the adsorption (or lack thereof) of the polypeptide to the surface.

The interfacial behavior of polypeptides can also be measured indirectly by monitoring its concentration in the bulk phase upon interaction with a surface. The adsorption isotherm, i.e., the amount of polypeptide attached to the surface as a function of the *equilibrium* concentration in the bulk solution, can be determined from the reduction of the polypeptide concentration in the bulk solution[18-23]. By utilizing semi-empirical models such as the Langmuir equation, one can gain valuable insight into the strength of surface binding. Unfortunately, the semi-empirical models often have little predictive capability. They are mostly used only as a benchmark for comparing different polypeptides as well as different surfaces and solution conditions. The experimental data are valuable for the calibration of various thermodynamic models and validation of their hypotheses to account for the key physics necessary to describe different aspects of surface-polypeptide interactions.

A number of theoretical methods have been proposed to describe the adsorption and interfacial behavior of polypeptides at inorganic surfaces[24, 25] [26]. These methods offer valuable insights to bridge the knowledge gap between microscopic and macroscopic observations. For example, Kurut et al. utilized molecular simulation and a polymer density functional theory (DFT) to study the charge regulation of histidine in unstructured peptides at interfaces[27]. They found that the adsorption of the peptide composed of histidine to be weaker at pH 8 than 6 because the histidine residues are less likely to be charged. In addition, added salt lowered the adsorption amount due to the screening effect on electrostatic-driven adsorption. Biesheuvel et al. studied the adsorption of ionizable polyelectrolytes such as lysozyme on a silica surface using a modified Poisson-Boltzmann model[28]. The mean-field method accounted for the charge regulation of the surface and the protein resulting from the intermolecular electrostatic interactions between the surface, ions, and protein. In general, the adsorption of lysozyme exhibited a maximum in adsorption with pH at a pH less than the isoelectric point. The increase in the salt concentration reduced the adsorption at high pH when the adsorption was electrostatic-driven, whereas the added salt would increase the adsorption at low pH when the electrostatic interactions between the surface and the lysozyme were less important. A different approach by Leermakers et al. utilized a polymer self-consistent-field theory to study the adsorption of β-Casein[29, 30]. They found that the protein formed a dense layer near the surface and that the adsorption tended to increase with pH and bulk protein concentration.

Molecular simulation methods have also been utilized for predicting polypeptide adsorption with different levels of microscopic details, ranging from atomistic to coarse-grained models. For example, Verde et al. employed an all-atom molecular dynamics (MD) simulation to study peptide adsorption on a gold surface[31]. It was found that the dynamics of peptide adsorption is dictated by a free-energy barrier that depends on the chain flexibility. The key to overcoming the barrier to adsorption was that the peptide must have high conformational flexibility but low conformational stability, i.e., the polypeptide should be able to freely adapt to the surface/local environment and does not get trapped in a single conformation. Xie et al. investigated the adsorption of neuromedin-B peptide, a decapeptide originally isolated from porcine spinal cord, at different self-assembled monolayers (SAM) using all-atom MD simulation[32]. They found that the surface induces the peptide conformational change compared to that in the bulk solution and that the variation was greatest for a hydrophobic SAM. Alternatively, polypeptide adsorption has been investigated with coarse-grained models[33, 34]. In comparison with atomistic methods, a coarse-grained model offers a significant decrease in the computational burden thus is suitable for the systematic study of more realistic systems. Hyltegren et al. investigated the adsorption of the saliva protein histatin 5 (24 residues) on silica surface using a bead-spring model with each bead representing an amino-acid residue[35]. Their work highlights the importance of the peptide and surface charge in predicting the peptide adsorption. Further details can be included in the coarse-grained description of polypeptide adsorption. For example, Qiu et al. used the Martini model that maps each residue into one or more beads in order to capture its intrinsic characteristics such as polarity and shape.[36]

They identified the important role of the adsorbed water in driving the polypeptide adsorption to a neutral surface through van der Waals forces.

A major difficulty in modeling polypeptide adsorption lies in the accurate representation of solution conditions (e.g., pH, ion valence, and salt concentration). The environmental effects are important not only for polymer-surface interactions but also for both the inter- and intramolecular interactions that dictate the peptide conformation. In particular, the solution condition directly influences the charge behavior of the polypeptide and substrate that is often utilized for the dynamic control of the polypeptide properties in practical applications. Recently, Grünewald et al. demonstrated a convenient way to incorporate titratable beads into the Martini coarse-grained model[37]. Conversely, the charge regulation can be implemented with thermodynamic models that explicitly take into account the chemical equilibrium. The theoretical calculations are significantly faster than the simulation counterparts. One example of the theoretical approach is by incorporating charge regulation through the mean-field approximation, i.e., the average ionization of each monomer or surface site is assumed uncorrelated with other ionizable groups in the system[38]. The mean-field approximation allows for the consideration of multi-chain systems near charge regulated surfaces under realistic conditions, opening up the possibility of computational screening and materials design[10, 39, 40]. A similar mean-field approach to charge regulation was used in the work by Kurut et al. discussed earlier[27]; however, both these methods fail to account for the influence of intramolecular correlations on the ionization behavior of individual residues.

In a previous work,[41] we developed a coarse-grained thermodynamic model for describing the adsorption of natural amino acids onto inorganic surfaces in various aqueous solutions. The thermodynamic model provides a faithful description of the charge behavior of the amino acids as a function of pH and solution conditions. We also demonstrated that the charge regulation of various ionizable surfaces is well described by an explicit consideration of the inhomogeneous ion distributions and protonation/deprotonation reactions[42]. Herein, we extend the coarse-grained models to polypeptides and examine their adsorption behavior at the titanium surface utilizing the recently developed Ising density functional theory (iDFT)[43]. The molecular thermodynamic model allows us to predict the adsorption of various polypeptides in response to the changes in solution conditions. This predictive model fills the gap between phenomenological models to describe polypeptide adsorption and various molecular simulation methods. Importantly, the theoretical procedure will provide a foundation for the future studies of polypeptide systems including polymer-mediated adhesion.

7.2 Thermodynamic models and methods

7.2.1 A coarse-grained model for polypeptide adsorption

In this work, we are interested in developing a coarse-grained model applicable to polypeptide adsorption at an ionizable surface over a broad range of solution conditions. Toward that end, we employ an augmented primitive model (APM) that takes into account the molecular excluded volume, short-range interactions, and the ionization states of each amino-acid residue. In APM, each monomer or polymer segment has a unique hard-sphere diameter σ_i, and may exist in different charge states as described by "Ising" parameter s_i.

Except the chain connectivity, the model is similar to our previous studies for the adsorption of amino acids and their titration behavior in bulk solutions[41, 44]. The equilibrium constant for the ionization of each amino acid side chain is assumed to be the same as that for the monomer[44].

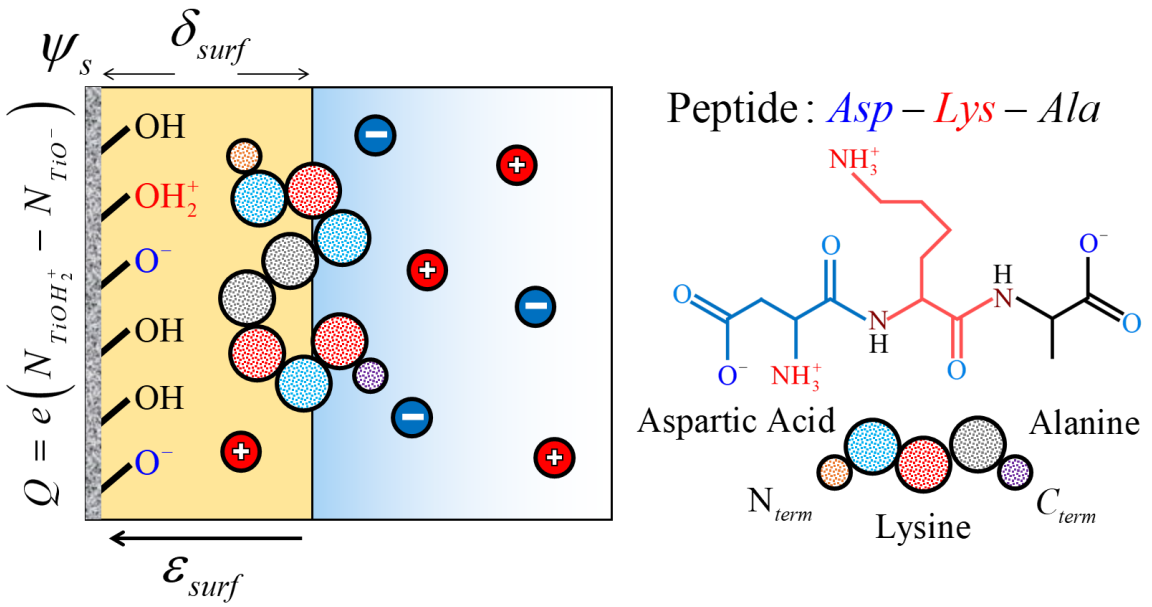

Figure 7.1 Schematic of a coarse-grained model for polypeptide adsorption on a titanium surface. The polypeptide molecule is represented by a tangent-hard-sphere chain where each segment corresponds to an amino-acid residue. In addition to electrostatic interactions due to the charge of ionizable residues and the terminal groups, the polypeptide can interact with the surface through a square-well potential of strength ε_{surf} *and width* δ_{surf} *representing non-electrostatic interactions*

Figure 1 shows a schematic representation of a polypeptide chain near an inorganic surface as described by our coarse-grained model. The polypeptide molecules are considered as tangent- hard-sphere chains where each monomer represents one amino-acid residue. These monomers can be acidic, basic, or neutral depending on the functional groups in the side chain as well as on the local environment. In the absence of the polypeptide, the coarse-grained model reduces to the primitive model of aqueous

electrolyte solutions, i.e., the ions are represented by charged hard spheres and the solvent by a dielectric continuum of relative permittivity $\varepsilon_r = 78$, which corresponds to that for liquid water at ambient conditions. The augmented aspect of the primitive model refers to, in part, the consideration of ionization equilibrium for each amino-acid residue and the inclusion of various well-recognized but poorly understood water-mediated interactions that dictate short-range interactions. For simplicity, we consider such interactions empirically through the square-well (SW) model with SW width fixed to $\delta_{surf} = 0.4\,\text{nm}$, approximating the same as the diameter of a water molecule[45]. While the phenomenological model neglects atomic details and the anisotropic nature of individual residues, it provides a flexible framework to quantify the interfacial behavior of polypeptides by treating the square-well attraction energy ε_{surf} as an adjustable parameter. Approximately, this parameter characterizes the short-range interaction between an amino-acid with a specific surface thus may be assumed independent of the solution conditions. We demonstrated in our previous work that the augmented primitive model can well describe the thermodynamic properties of amino acids in bulk solutions as well as their adsorption to inorganic surfaces and the theoretical predictions were found in good agreement with experimental observations[41, 44]. Further details on the molecular model can be found in our previous work.

Polypeptide adsorption at an inorganic surface is driven by either electrostatic attraction or surface binding (non-electrostatic). While the electrostatic interactions are

coupled with ionic species in the solution, the surface binding is represented by a constant energy as represented by a square-well potential for each residue

$$V_i^{ext}(z) = \begin{cases} \infty & z < \sigma_i / 2 \\ -\varepsilon_{surf,i} & \sigma_i / 2 \le z \le \delta_{surf} \\ 0 & else \end{cases}. \tag{7.1}$$

Spruijt et al. proposed that the binding strength of polymers containing carbonyl groups to a silica surface is dependent upon the number of protonated surface hydroxyl groups[46]. The surface groups will deprotonate as the pH is increased, which results in fewer binding sites for the polymer thereby a weaker average binding energy. In this work, we also find that the attraction energy is related to the available surface sites (*viz*. surface functional groups) that the polymer can bind and such effects can be represented by the empirical relation:

$$\varepsilon_{surf,i} = \varepsilon_{surf,i}^{0}\left[1-\left(\frac{N_{TiO^-}}{N_{sites}}\right)^2\right] \tag{7.2}$$

where $\varepsilon_{surf,i}^{0}$ is the surface binding energy at an uncharged surface, and N_{TiO^-}/N_{sites} corresponds to the fraction of negatively charged surface sites.

Lastly, we consider the charge regulation of the inorganic surface through the deprotonation and protonation of its surface hydroxyl sites. For a titanium dioxide surface, the deprotonation and protonation reactions are given by

$$\text{-TiOH} \rightarrow \text{-TiO}^- + \text{H}^+, \tag{7.3}$$

$$\text{-TiOH} + \text{H}^+ \rightarrow \text{-TiOH}_2^+. \tag{7.4}$$

These reactions are governed by the equilibrium constants, K_D and K_P, respectively, which are related to the solution $\mathrm{pH} = -\log a_{H^+}$ and the activities of the surface sites:

$$K_D = \frac{N_{TiO^-}}{N_{TiOH}} \frac{\gamma_{TiO^-}}{\gamma_{TiOH}} a_{H^+}, \tag{7.5}$$

$$K_P = \frac{N_{TiOH_2^+}}{N_{TiOH}} \frac{\gamma_{TiOH_2^+}}{\gamma_{TiOH}} \frac{1}{a_{H^+}}, \tag{7.6}$$

where N_i refers to the number of surface site i per unit area, and γ_i is the activity coefficient of the corresponding surface site. The activity coefficient of each surface group accounts for its physical interactions with the local environment[47]. In this work, we assume that the surface activity coefficients are mainly determined by the electrostatic interactions between surface sites, ions, and the polypeptides. As a result, we can express the ratio of activity coefficients for the uncharged to charged states of the surface functional groups by the mean electrostatic potential (viz. $\gamma_{TiOH} / \gamma_{TiO^-} = \exp[\beta e \psi_s]$ and $\gamma_{TiOH} / \gamma_{TiOH_2^+} = \exp[-\beta e \psi_s]$. The equilibrium constant and the ratio of activity coefficients can be combined into an *apparent* equilibrium constant: $K'_D = K_D \gamma_{TiOH} / \gamma_{TiO^-}$ and $K'_P = K_P \gamma_{TiOH} / \gamma_{TiOH_2^+}$ which correspond to the experimentally measured equilibrium constants. An explicit expression for the surface charge density can be derived in terms of the total number of available sites per unit area, the apparent equilibrium constants, and the proton activity a_{H^+}:

$$Q = -e\left(N_{TiO^-} - N_{TiOH_2^+}\right) = -e\, N_{sites} \frac{K'_D / a_{H^+} - K'_P\, a_{H^+}}{1 + K'_D / a_{H^+} + K'_P\, a_{H^+}} \tag{7.7}$$

where the total number of ionizable sites, $N_{sites} = N_{TiOH} + N_{TiO^-} + N_{TiOH_2^+}$, can be determined from experimental data or estimated from the crystalline structure. Since the surface charge is related to the mean electrostatic potential through Gauss' equation, there is a coupling between the surface charge and the polypeptide adsorption. A comprehensive description of the surface charge regulation can be found in our previous publications[41, 42].

7.2.2 Ising density functional theory

Ising density functional theory (iDFT) provides a convenient avenue to describe the ionization of weak polyelectrolytes at inhomogeneous conditions[43]. For a polypeptide represented by *M* ionizable segments, the chain conformation is fully specified by a multidimensional vector, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_M)$, where $\mathbf{r}_i$ denotes the segment position, and the ionization state of the entire polymer chain can be specified by the vector $\mathbf{S}_k = (s_1, s_2, ..., s_M)$, where $s_i = \pm 1$ or 0 is the charge number (*viz.*, the valence of segment *i* in the polypeptide chain). The density profile of the polypeptide in configuration $\mathbf{X} = (\mathbf{R}, \mathbf{S})$ is given by

$$\rho(\mathbf{X}) = \exp\left\{-\beta[V^B(\mathbf{X}) + V^{ext}(\mathbf{X}) + \mu^{ex}(\mathbf{X}) + \mu^{\infty}(\mathbf{X}) + \mu^H(\mathbf{S}) - \mu]\right\} \tag{7.8}$$

where $V^B(\mathbf{X})$ is the bond potential, μ is the chemical potential of the polymer in its fully uncharged state, V^{ext} is the non-electrostatic external potential for the entire polymer chain (that for each segment is given by Eq. 1), $\mu^{ex}(\mathbf{X}) = \delta F^{ex} / \delta\rho(\mathbf{X})$ is the local excess chemical potential to account for the thermodynamic non-ideality due to inter- and intra-molecular interactions, $\mu^{\infty}(\mathbf{X})$ is the contribution to the potential energy due to the

intramolecular electrostatic interactions in configuration **X**, and $\mu^H(\mathbf{S})$ is affiliated with the ionization of individual segments, i.e.,

$$\mu^H(\mathbf{S}) = \sum_i^M \Delta\mu_i^H = -\ln 10 k_B T \sum_i^M s_{k,i}(\mathrm{pK}_{k,i} - \mathrm{pH}). \tag{7.9}$$

The ionization energy accounts for the change in the grand potential due to deprotonation/protonation of the ionizable segments. The equilibrium constant of the deprotonation or protonation reaction, K_i, depends on the identity of the ionizable site and system temperature but not on the solution composition.

According to the thermodynamic perturbation theory, the local excess chemical potential can be decomposed into one-body and two-body contributions:

$$\mu^{ex}(\mathbf{X}) = \sum\left[\mu_i^{mon}(\mathbf{x}_i) + \mu_i^{ch}(\mathbf{x}_i)\right] + \sum \ln y(\mathbf{x}_i, \mathbf{x}_{i+1}) \tag{7.10}$$

where μ_i^{mon} is the excess chemical potential for the monomer system, μ_i^{ch} is excess chemical potential due to chain connectivity which describes the free energy due to intrachain correlation by segment i in state s_i, and $\ln y$ is the natural log of the cavity correlation function. The first summation accounts for the thermodynamic non-ideality due to monomeric interactions, and the second summation accounts for the intramolecular correlations[48]. In this work, we consider the excluded-volume effects through the modified fundamental measure theory[49], the electrostatic correlations by the reference fluid method using the mean spherical approximation [50], and the square-well attractions through a mean-field method[45].

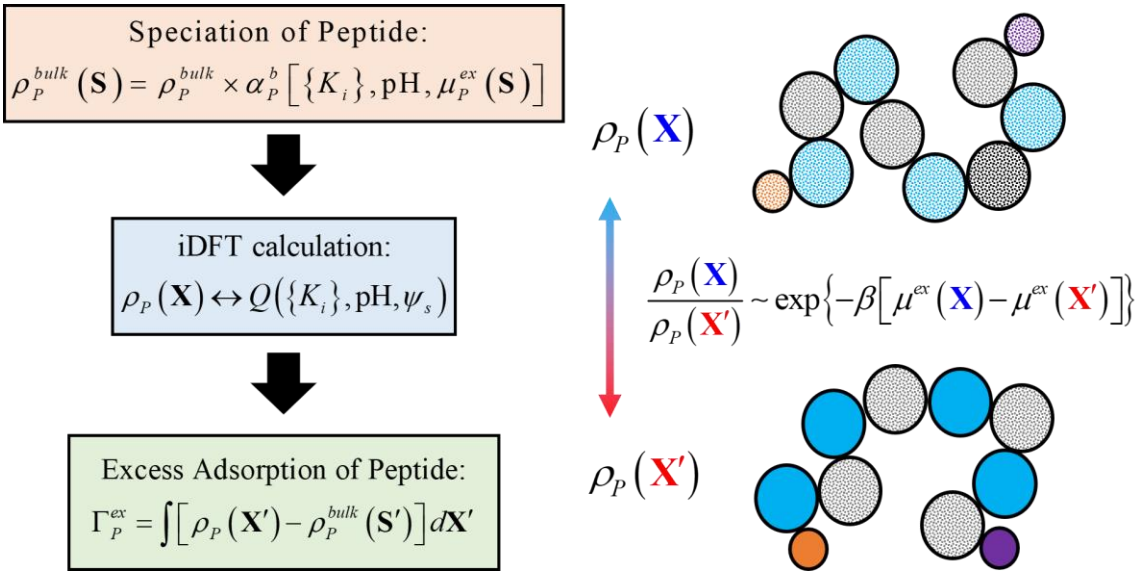

Figure 7.2 A flowchart for determining the adsorption of peptides on an inorganic surface through the Ising density functional theory (iDFT).

In Figure 2, we show a flowchart to determine the adsorption of polypeptides on an inorganic surface. In numerical implementation of iDFT, we first determine the speciation of the polypeptide in the bulk based on the inter- and intra- molecular correlations with other peptides and ions present in the solution:

$$\rho_P^{bulk}(\mathbf{S}) = \rho_P^{bulk} \times \frac{\exp\left[-\beta\left\{\mu^{ex}(\mathbf{S}) + \mu^{\infty}(\mathbf{S}) + \mu^{H}(\mathbf{S})\right\}\right]}{\sum_{\{\mathbf{S}'\}} \exp\left[-\beta\left\{\mu^{ex}(\mathbf{S}') + \mu^{\infty}(\mathbf{S}') + \mu^{H}(\mathbf{S}')\right\}\right]}. \tag{7.11}$$

where $\mu^{ex}(\mathbf{S})$ and $\mu^{\infty}(\mathbf{S})$ are the bulk counterparts of $\mu^{ex}(\mathbf{X})$ and $\mu^{\infty}(\mathbf{X})$, respectively, since the fluid is uniform in the bulk solution and the conformation **R** does not influence the inter- and intramolecular correlations. Since we evaluate the intramolecular interactions at the nearest-neighbor level (i.e., interaction between *i* and *i*+1 only), Eq. (10) can be solved using the standard practices employed in the nearest-neighbor site-binding

model[51]. In the bulk solution, the inter- and intra-molecular correlations are dependent upon the average concentrations of ions and polypeptide segments. Because the excess chemical potential of the polypeptide depends on its valence vector S, we must self-consistently solve for the speciation of the polypeptide in the bulk solution. Next, we solve for the density profiles of the polypeptide and ions near the inorganic surface which has its own surface charge density dependent upon the local environment (viz. mean electrostatic potential). Different from alternative theoretical approaches that consider only the intermolecular interactions on the distribution of polypeptides and ions[28], iDFT also accounts for the intramolecular correlations between the polypeptide residues that directly influence the polypeptide charge and chain conformation. The intrachain contribution is particularly important for polypeptides with different types of charged residues in which the amino-acid sequence can play an important role in the adsorption process. Lastly, we calculate the excess adsorption of the polypeptide residues in their given charge state from

$$\Gamma_{P,i}^{ex}\left(s_i\right)=\int\left[\rho_i\left(\mathbf{r},s_i\right)-\rho_i^{bulk}\left(s_i\right)\right]d\mathbf{r} \tag{7.12}$$

where the density profile of a residue in its charge state is given by

$$\rho_i\left(\mathbf{r},s_i\right)=\exp\left\{-\beta[V^{ext}(\mathbf{r},s_i)+\mu_i^{mon}(\mathbf{r},s_i)+\Delta\mu_i^{H}(s_i)-\mu]\right\}G_i^{L}\left(\mathbf{r},s_i\right)G_i^{R}\left(\mathbf{r},s_i\right). \tag{7.13}$$

Here, G_i^L and G_i^R are the propagator functions that account for the contribution from the rest of the polymer given that segment is at position **r** in state s_i. Additional details on iDFT can be found in our previous publication[43].

7.3 Results and discussion

In our previous work, we demonstrated that classical density functional theory (cDFT) can well capture amino acid adsorption and the influence of pH and salt concentration on the ionization behavior of amino acids in the bulk and near inorganic surfaces[42]. In the following, we provide 3 case studies to illustrate the application of our coarse-grained model to describe the pH effect on the adsorption of oligopeptides at a planar surface of titanium oxide. The oligopeptide is composed of some or all of the following residues: alanine, lysine, and aspartic acid. The equilibrium constants for two ionizable residues are taken from our previous work[44]: $\mathrm{pK}_{Lys} = 10.95$ and $\mathrm{pK}_{Asp} = 3.97$. The experimental data used for the validation of our theoretical results are from Imamura et al.[52]. Unfortunately, no titration data exists for the titanium dioxide surface used in their work; therefore, we have chosen to use the following parameters to describe the protonation and deprotonation equilibrium constants of the hydroxyl groups at the solid surface, $\mathrm{pK}_P = -4.1$ and $\mathrm{pK}_D = 6.5$, respectively, and the number of the surface sites to be 1.0 #/nm^2. The calibrated equilibrium constants were determined from a titanium surface different than the one considered here[41]. We expect that the equilibrium constants should not differ much from those of the surface studied by Imamura et al. since both are titanium surface functional groups. However, due to differences in surface preparation which can change the number of surface hydroxyl sites[53], we correlated the surface site density to give the best agreement with the experimental data.

In addition, we determine a binding energy for each residue based off correlation with the experimental data: $\varepsilon^0_{surf,Asp} = 5.50k_BT$, $\varepsilon^0_{surf,Ala} = 3.75k_BT$, and $\varepsilon^0_{surf,Lys} = 1.90k_BT$. The non-electrostatic binding energy is affiliated with the hydrogen bonding by the carbonyl group in the backbone for all three residues and an additional binding capability by the carboxyl group in aspartic acid. The energy values are weaker than that found for the adsorption of amino acids to titanium because of the absence of the amine and carboxyl groups in the residues (viz. loss in hydrogen bonding groups)[41]. The binding strength of lysine is lower than the alanine residue even though they both contain one carbonyl group which may be indicative of the extended conformation of the lysine structure that favors the side chain orientation towards the surface and therefore, the backbone carbonyl group will be less likely to interact with the surface.

7.3.1 Adsorption of oligopeptide with acidic and neutral residues

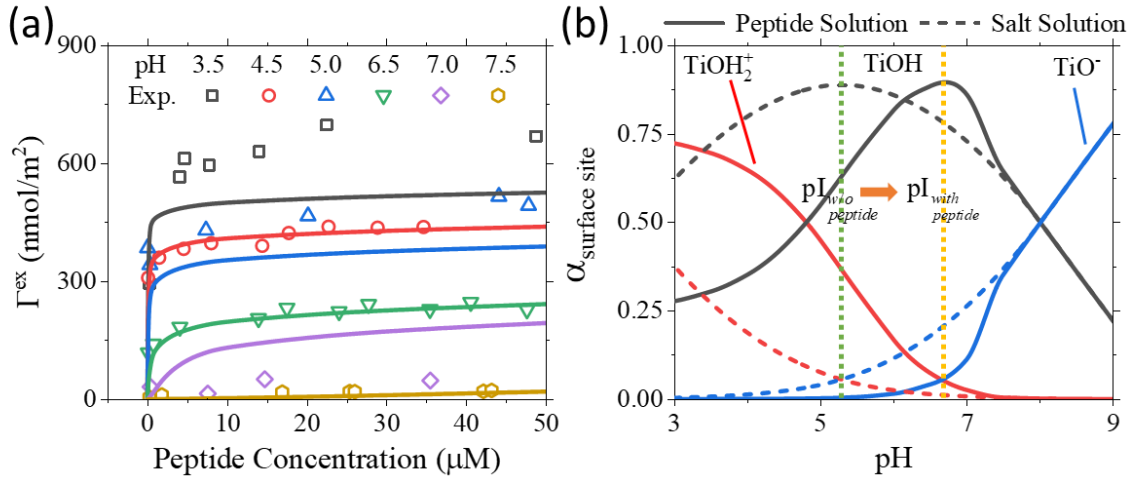

Figure 7.3 The adsorption isotherms for the polypeptide DADADADA on the titanium surface in a 100 mM potassium nitrate solution. The lines are the theoretical predictions and the symbols are experimental data[52]. (b) The fractions of surface sites on the titanium surface with and without the polypeptide (at 50 μM) in 100 mM KNO_3 solution. The surface sites $TiOH_2^+$, $TiOH$, and TiO^- correspond to the red, black, and blue lines, respectively.

We first consider the adsorption of the oligopeptide DADADADA on a titanium surface as investigated experimentally by Imamura et al.[52]. The peptide is composed of four acidic residues (viz. aspartic acid, *D*) and four neutral residues (viz. alanine, *A*). The adsorption isotherm in the range of 0 to 50 μM peptide concentrations at 6 pH values in a 100 mM potassium nitrate solution. Figure 3a shows a comparison between the theoretical and experimental results. In general, the decrease in pH results in an increased adsorption of the acidic polypeptide. The adsorption versus pH trend is expected because the oligopeptide is predominately negatively charged (due to the deprotonation of the carboxyl side chain in the aspartic acid residues) while the titanium surface increases in charge with a decrease in pH. The increased adsorption at low pH is thus driven by more favorable

electrostatic interactions between the polypeptide and the inorganic surface. In addition, the adsorption is favored by non-electrostatic surface binding as evidenced by the excess adsorption of the peptide at the surface up to pH 7.5 when the titanium surface is negatively charged (isoelectric point of 5.3)[41]. Overall, our theoretical predictions agree with the experimental results at a semi-quantitative level indicating that the coarse-grained model for polypeptide adsorption captures the ionizations of of the aspartic acid residues and the titanium surface reasonably well.

Figure 3b shows the influence of the polypeptide adsorption on the dissociation of the surface hydroxyl sites at the titanium surface. When there is no polypeptide present, the surface sites will be net positively charged at pH below the isoelectric point (pI=5.3) and net negatively charged at pH above the isoelectric point. In the presence of the polypeptide ($\rho_p^{bulk} = 50\mu\text{M}$), the surface composition is significantly different at low pH. The large change in the composition of surface sites can be attributed to the strong attraction (both electrostatically and non-electrostatically) of the aspartic acid and alanine residues in the polypeptide to the surface. The isoelectric point of the titanium surface is shifted from 5.3 to approximately 6.7 as shown by the vertical dotted lines in Figure 3b. The negatively charged residues in the polypeptide promote the presence of positively charged sites on the surface and inhibit the formation of negatively charged sites, thereby shifting the isoelectric point of the surface to a higher pH. As the pH is further decreased, the favorable electrostatic interaction between the surface and the aspartic acid residues leads to a larger positive surface charge compared to the salt only solution. At low pH, the aspartic acid residues start to lose their negative charge, which explains why the surface

does not dissociate to the $-TiOH_2^+$ group as quickly with a decrease in pH as it did at slightly higher pH since there are less favorable charge-charge interactions. At high pH, there is no difference in the titration of the titanium surface whether the polypeptide is present or not in the bulk solution since the polypeptide will no longer adsorb on the surface. Thus, we see that the correct treatment of the thermodynamic non-ideality in the charge regulation of both the residues and the surface is important to an accurate description of polypeptide adsorption to inorganic surfaces.

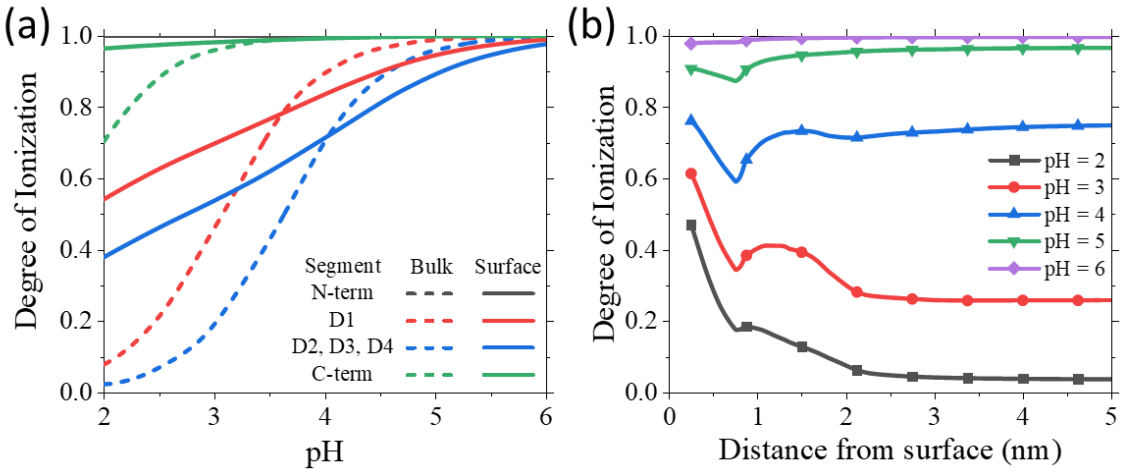

Figure 7.4 (a) The degree of ionization for aspartic-acid residues from DADADADA in the bulk and that adsorbed on the titanium surface. The lines for D1, D2, and D3 are nearly indistinguishable from one another. (b) The average degree of ionization for aspartic acid residues as a function of the distance from the surface for different pH. The peptide concentration is 50 μM and the potassium chloride concentration is 100 mM.

The charge regulation of the different residues and terminal groups is key to the adsorption of the polypeptide to the titanium surface. We show in Figure 4a the degree of ionization for aspartic acid residues and the terminus groups in the bulk solution and adsorbed to the surface. As a result of the intrachain correlations and the spatial

conformation of the polypeptide, the acidic residues may not necessarily be at the same average degree of ionization near the interface. We find that the ionization degree for D1 is significantly greater than the degree of ionization for the other three residues (D2, D3, and D4) whether in the bulk or adsorbed to the surface. The similarity in the ionization for the other three residues indicates that the intrachain correlations of the polypeptide play an important role in determining the charge status of the residues. While D2, D3, and D4 are surrounded by neutral alanine residues, D1 is neighbored by a positively charged N-terminus group (which is always charged below pH 9 as shown in Figure 4a) and therefore D1's ionization is promoted due to the favorable charge-charge interaction. Since the adsorbed polypeptides mostly lay flat on the surface, there is not a significant difference between the latter three aspartic acid residues. A mean-field method (i.e., one that does not incorporate intramolecular interactions such as the modified Poisson Boltzmann method[28]) would miss the different ionization behavior of D1 from the other aspartic acid residues.

The ionization of polypeptides near the interface is different from that in the bulk due to the changes in the local environment as well as surface-induced behavior (i.e., charge regulation). At high pH, the polypeptide charge is unaffected by its local environment since the driving force due to the chemical driving force is large. As the pH is lowered, the degree of ionization for those residues at the surface is initially lower than that in the bulk solution. Since there is a strong adsorption of the polypeptide even at pH 5 where the surface is nearly neutral, the repulsion between charged aspartic acid residues at the surface will lead to the ionization of the residue being less favorable than that in the

dilute bulk solution. However, as the pH is further decreased, the surface becomes sufficiently charged such that there is a favorable electrostatic interaction with the surface that overcomes the repulsion between the aspartic acid residues at the surface. As a result, the average ionization of the residues at the surface decreases slower with pH compared to the residues when they are in the bulk solution. Our theoretical method improves upon conventional approaches by capturing the charge regulation of the polypeptides due to inter- and intra-chain correlations important for the polypeptide adsorption to ionizable inorganic surfaces.

Figure 4b shows the average degree of ionization for the aspartic acid residues as a function of distance from the surface. When the polypeptide is greater than 3 nm from the surface, the behavior of the polypeptide is equivalent to its bulk counterparts since the surface-induced effects are relatively minor (i.e., salt ions have sufficiently screened the long-range electrostatic interaction). At very short distances (i.e., within square-well width), the ionization of the aspartic acid is typically increased relative to those in the bulk due to the favorable electrostatic interaction between to the surface and aspartic acid residues. For example, whereas the aspartic acid residues are essentially neutral at pH 2, about half the aspartic residues are charged near the titanium surface. Interestingly, within the surface region, there is a significant drop in the ionization between those residues at contact with the surface and at the end of the adsorption region. We attribute the large change in ionization within the short distance to the strong screening of the positively charged surface sites by the negatively charged aspartic residues. At low pH, the surface charge is not fully compensated by the oppositely charged amino-acid residues, and

therefore, an increased presence of counterions near the surface explains the gradual decrease in ionization until it reaches that of the bulk. On the other hand, at higher pH such as pH 4, the surface ionization increases beyond the adsorption region since there is not enough favorable electrostatic interactions with the surface to compensate the high accumulation of aspartic residues which inhibits the ionization to the same extent as lower pH conditions.

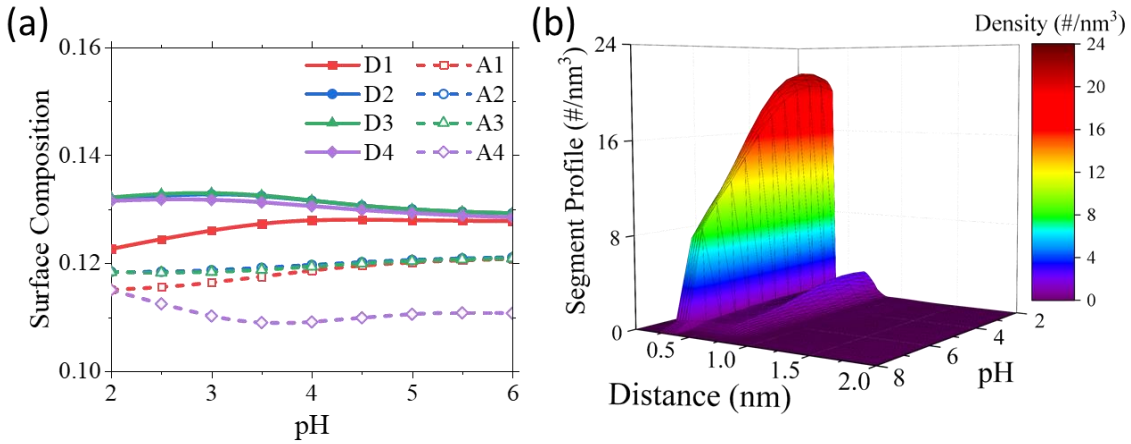

Figure 7.5 (a) The surface composition (defined as the fraction of species within the square-well width) of different residues in the oligopeptide DADADADA as a function of pH in a 100 mM KNO_3 solution. (b) The segment profile of the oligopeptide as a function of distance from the surface and solution pH. The concentration of polypeptide and salt in solution is 50 μM and 100 mM KNO_3.

We next investigate the polypeptide at the surface through the composition at the surface (defined as the fraction of the different residues within the square-well width). The C- and N-terminus groups are depleted from the surface (viz. less than 0.02 for surface composition). As shown in Figure 5a, the aspartic acid residues are present at the surface to a higher degree than alanine residues. The difference can be attributed to its higher non-electrostatic binding energy and the favorable electrostatic interaction between the residue and the surface. The alanine residues are also adsorbed because the polypeptide mostly

lays flat on the surface at low concentration. Interestingly, the aspartic acid residue on the N-terminus side of the oligopeptide, D1, is present at a lesser degree compared to the other three aspartic acid residues (D2, D3, D4). A similar behavior is found for the A4 residue which is connected to the C-terminus group. Thus, the ionization of N- and C-terminus groups has a significant influence on the adsorption of the neighboring residues in the polypeptide chain. Specifically, the positively charged N-terminus group will be repulsed by the positively charged surface which will reduce the likelihood of the neighboring aspartic acid to be within the surface when the surface is positively charged at low pH. As the pH increases, the difference in the surface composition between the residue D1 and the other residues is negligible at pH greater than 6 because the surface is mostly neutral and therefore the repulsion due to the N-terminus group is no longer relevant. On the other end, the C-terminus group is negatively charged and therefore it is attracted to the surface when it is positively charged at low pH. However, the surface must be sufficiently positively charged in order for the C-terminus group to compete for space with the aspartic acid residues. As a result, the A4 residue is generally further from the surface because it is not surrounded by the strongly interacting aspartic residues that the other alanine residues are surrounded by. Thus, the inclusion of the terminal groups is an important component to understanding the conformation of polypeptides at interfaces due to their intrachain correlations with neighboring monomers and electrostatic interactions with surface.

In addition, we consider the total segment distribution as a function of distance from the surface and pH in Figure 5b. We see a noticeable increase in the density of polypeptide at the surface as the pH is decreased which follows from the adsorption trend in Figure 3a.

The polypeptide mostly exists at the surface due to the strong non-electrostatic binding energy within a short-range of the surface. Beyond the square-well width, there is a sharp drop off in the density and then a formation of a second peak in the density profile as the pH is decreased. The second peak results from segments (mostly alanine) that are pushed out of the surface layer in favor of other residues (viz. aspartic acid). We find that essentially no polypeptides are found in excess beyond the surface region, thus demonstrating that there is not a sufficient driving force for the adsorption of polypeptides to the surface through electrostatics alone.

7.3.2 Adsorption of oligopeptide with basic and neutral residues

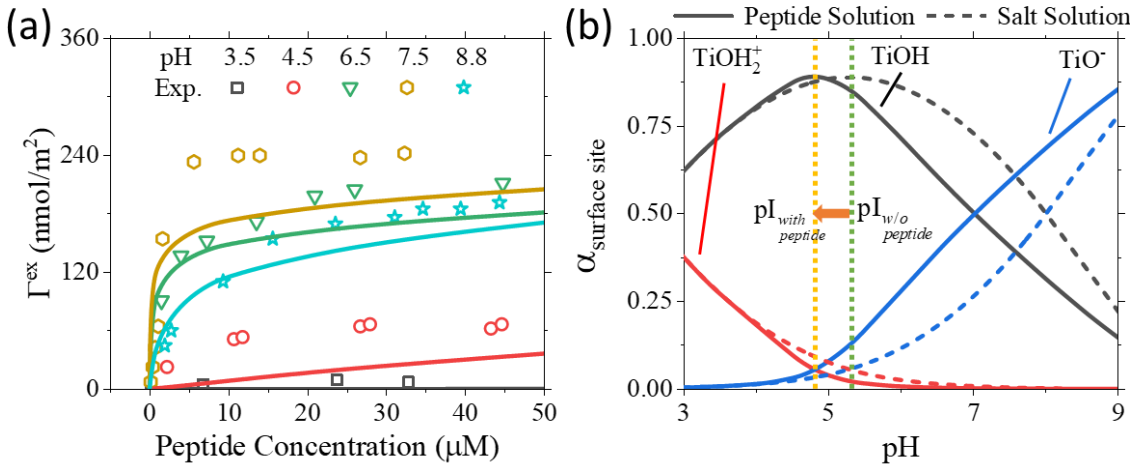

Figure 7.6 (a) The pH-dependence of adsorption isotherms for polypeptide KAKAKAKA on a titanium surface in a 100 mM potassium nitrate solution. The lines are the theoretical predictions and the symbols are experimental data[52]. (b) The fractions of different surface sites at the titanium surface with and without the polypeptide (at 50 μM) in 100 mM KNO_3 solution. (See comments in the caption of figure 3)

We next consider the adsorption of polypeptide KAKAKAKA composed of four basic lysine residues and four neutral alanine residues to the titanium surface. The polypeptide is fully ionized in the pH range considered in this work with a net charge of

+4 due to the four lysine residues. As shown in Figure 6a, the polypeptide interacts favorably with the surface when it is negatively charged. The binding energies due to non-electrostatic interactions of the basic residues with the surface is much weaker than those of the acidic residues (viz., 1.90 k_BT vs. 5.50 k_BT), which leads to less adsorption compared to the acidic polypeptide (see Figure 3a). Interestingly, the polypeptide adsorption increases with pH up to a maximum at pH 7.5 before decreasing as shown by the experimental results for pH 8.8. It would be expected that the adsorption of the basic polypeptide would continue to increase as the surface charge increases negatively (i.e., an increase in pH) since there are favorable charge-charge interactions and lysine residues maintain their charge even at pH>10. To explain the decrease in adsorption, we focus our attention on the non-electrostatic interaction between the surface and the carbonyl group which is facilitated by a hydrogen bond to the surface $-TiOH$ group (viz., we model this interaction through the square-well potential). However, when the surface deprotonates, the peptide can no longer hydrogen bond to the surface and therefore it loses its binding capability. We model the loss in hydrogen bond capability through Eq.(2) by decreasing the binding energy as the surface hydroxyl sites depronate.. As a result, it becomes less favorable for the peptide to adsorb to the surface at high pH (viz. a decrease in the adsorption) since the adsorption is facilitated mostly through the electrostatic interaction between the positively charged monomers and the negatively charged surface sites.

Figure 6b shows the effect of the polypeptide KAKAKAKA adsorption on the speciation of the titanium surface. Similar to the findings for polypeptide DADADADA, the speciation of the titanium surface is influenced by the adsorption of the basic

polypeptide; although to a lesser degree since the binding energy due to non-electrostatic interactions is much weaker (viz., 1.90 k_BT vs. 5.50 k_BT). There is a shift in the isoelectric point from 5.3 to 4.8 since the positively charged residues promote the deprotonation of the surface hydroxyl groups and inhibit the protonation. Since the polypeptide adsorption takes place in the pH range from 6 to 10, the surface behaves as it would in a solution without the polypeptide outside this pH range. As expected, the surface carries a greater negative charge at pH 7 when the peptide is adsorbed than it would have in the salt-only solution. Overall, our model performs well to capture the adsorption of the basic polypeptide to the titanium surface since we account for the change in non-electrostatic binding resulting from the loss in possible binding sites on the surface.

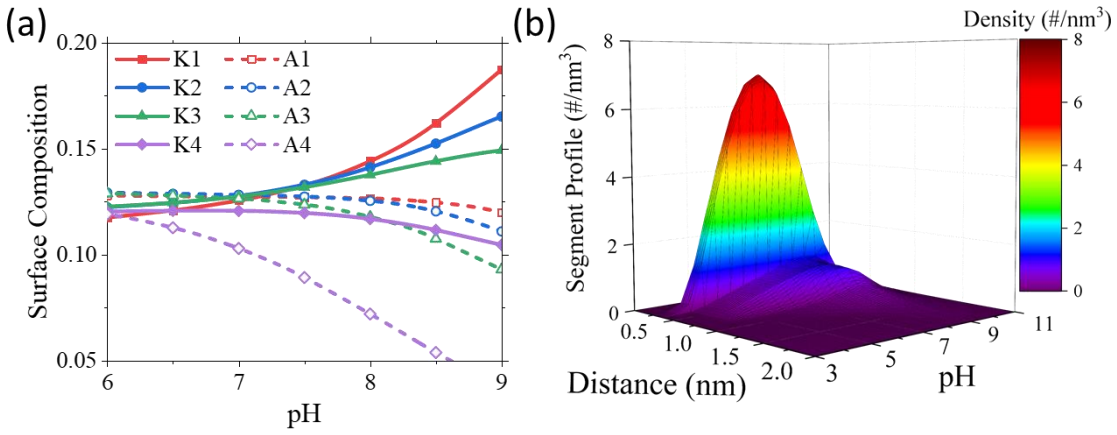

Figure 7.7 (a) The surface composition (defined as the fraction of species within the square-well width) of the residues in the oligopeptide KAKAKAKA as a function of pH in a 100 mM KNO_3 solution. (b) The segment profile of the oligopeptide as a function of distance from the surface and solution pH. The concentration of polypeptide and salt in solution is 50 μM and 100 mM KNO_3.

Figure 7a shows the surface composition of the oligopeptide KAKAKAKA at different pH values. Unlike the surface composition for the oligopeptide DADADADA,

there is significant variation in the surface composition for the basic polypeptide in the segment distribution with pH. The greater change can be attributed to the fact that the basic polypeptide does not interact with the surface non-electrostatically as strongly as the acidic polypeptide did. Therefore, the electrostatics play a more important role in determining the conformation of the polypeptide. As the pH increases, K1 and A1 are the most likely to be present at the surface out of the lysine residues and alanine residues, respectively. We can attribute this to the N-terminus group, which is positively charged in the pH range considered, and neighbors the K1 leading to two consecutive positive charges that can interact favorably with the negatively charged surface. Since the A1 residue neighbors with the K1 residue, which is generally in contact with the surface, it is less likely for the A1 residue to orient away from the surface. On the other end, the lysine residue K4 is followed by the neutral A4 and then the negatively charged C-terminus group. The C-terminus is repulsed by the negatively charged surface and therefore prefers to orient away from the surface. In addition, the A4 group can bind non-electrostatically to the surface, but it is outcompeted by the lysine residues and other alanine residues as indicated by Figure 7a. Therefore, we expect that the K4, A4, and C-terminus tend to form a tail region away from the surface. The surface composition of the C- and N-terminus groups are less than 0.04. Clearly, the segment-level detail of our coarse-grained model allows for unique insight into the polypeptide's behavior at the surface.

We present in Figure 7b the total segment distribution as a function of distance from the surface and pH. The accumulation of the polypeptide at the surface follows a bell-shape with pH in the range of 3 to 11. Unlike the acidic polypeptide, the basic

polypeptide shows a noticeable presence of peptide in the following layers after the surface, particularly around pH 8. The profile away from the surface agrees with the notion that the peptide has a tail region away from the surface which is present with K4, A4, and C-terminus groups. The weaker non-electrostatic binding results in more conformational flexibility for the polypeptide as it no longer needs to lay flat along the surface to maximize the attractive forces.

7.3.3 Adsorption of oligopeptide with acidic and basic residues

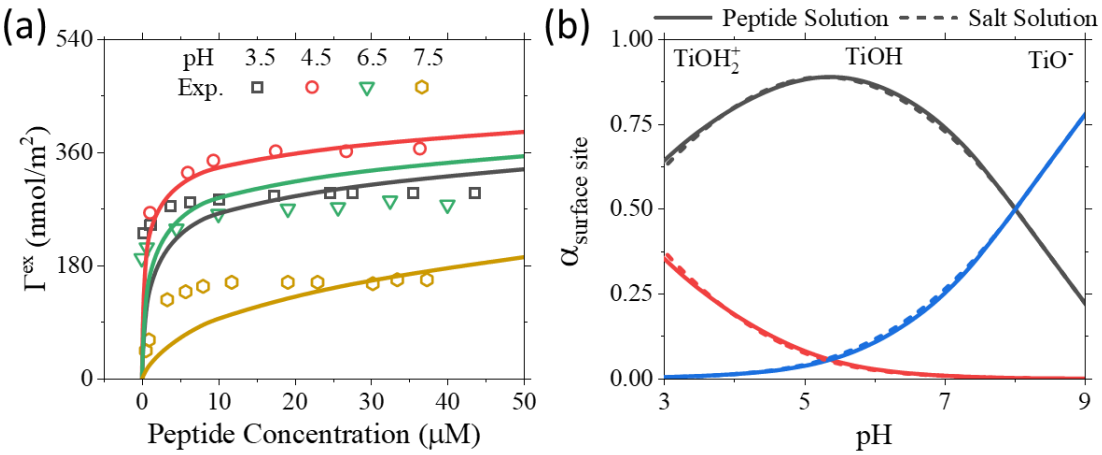

Figure 7.8 (a) The pH-dependence of adsorption isotherms for amphiphilic polypeptide DKDKDKDK on the titanium surface in a 100 mM potassium nitrate solution. The lines are the theoretical predictions and the symbols are experimental data[52]. (b) The fractions of surface sites at the titanium surface with and without the polypeptide (at 50 μM) in 100 mM KNO_3 solution.

The last case we consider is the adsorption of the polyampholyte DKDKDKDK to the titanium surface. Figure 8a shows the adsorption isotherms at different solution pH values. The alternating sequence of the polypeptide leads to attractive intrachain interactions between all neighboring residues; therefore, all residues maintain full ionization in the pH range considered. As a result, the polypeptide is neutral; we expect

that the polypeptide adsorption is less sensitive to the change in surface charge with pH. Indeed, we find that the polypeptide shows relatively similar adsorption for the pH 3.5, 4.5, and 6.5, but a noticeable drop in adsorption when the pH is 7.5. The latter can be explained by the weaker non-electrostatic binding due to the deprotonation of the surface sites as the surface charges negatively. Thus, we find that the polyampholyte represents a middle ground between the strong binding by the aspartic acid residues and the weaker binding by the lysine residues. We show the influence of the polypeptide on the surface speciation in Figure 8b. As expected from its "neutral" nature, the polypeptide should have a minimal impact on the speciation of the titanium surface as it does not contribute to the net charge of the fluid.

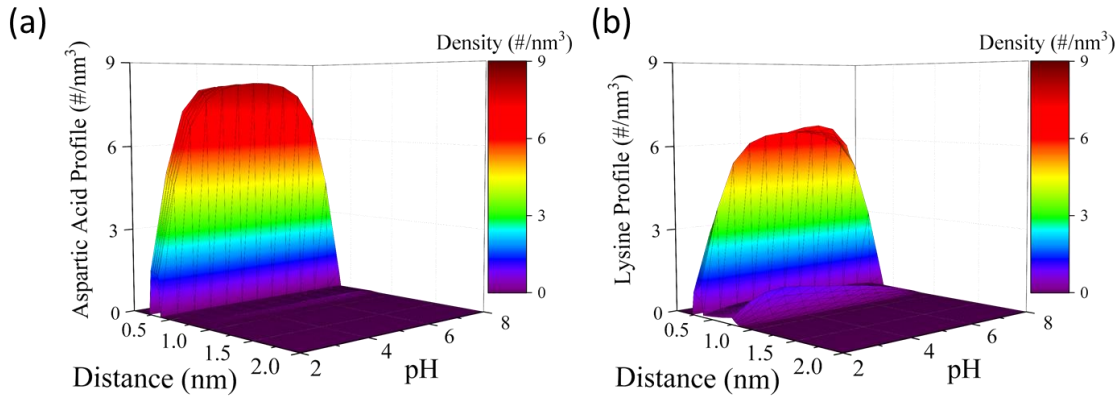

Figure 7.9 The segment profile for the (a) aspartic acid residues and (b) lysine residues of the oligopeptide as a function of distance from the surface and solution pH. The concentrations of polypeptide and KNO_3 salt in the bulk solution are 50 μM and 100 mM, respectively.

Lastly, Figure 9 shows the segment distribution as a function of distance and pH for the two different residues in the oligopeptide DKDKDKDK. The polyampholyte shows adsorption to the surface in the range from 2 to 8 after which the high charge of the surface

leads to the counterions outcompeting the “neutral” polypeptide in the surface region. By comparing Figure 9a and 9b, we see that the aspartic acid residues dominate the surface region and show little presence in the following layers. In contrast, lysine residues show a noticeable presence in the layer following the surface; particularly at the lower pH. Since aspartic acid residues bind strongly to the surface through the square-well potential, they maintain a relatively consistent occupancy of the surface region in the pH range that the polypeptide is adsorbed to the interface. On the other hand, lysine residues do not bind as strongly to the surface and therefore are more susceptible to the surface charge. When the surface is positively charged (low pH), it becomes more preferable for the lysine residues to not be in the surface layer. At high pH, the favorable electrostatic interaction between the positively charged lysine residues and negative charged surface sites lead to more lysine in the surface region and therefore the presence of lysine in the second layer is reduced. At low and high pH, the surface becomes highly charged and counterions will occupy the surface region instead of the polypeptide. Thus, we see that despite the neutral nature of the polymer, the polypeptide does adopt unique conformations at the surface to maximize the attractive forces between it and the surface.

7.4 Conclusion

There have been substantial interests in understanding the interaction of polypeptides and proteins with inorganic surfaces due to their applications to the design of bioadhesives and the remediation of biofouling. In the present study, we have initiated a theoretical work for such possibilities by employing a coarse-grained model for polypeptides that captures their adsorption at an inorganic surface under different solution

conditions. By leveraging our previous work to account for the key physics governing charge regulation and the interaction of amino acids with ionic species in aqueous solutions, we developed a molecular-thermodynamic framework that is able to predict the adsorption of polypeptides on inorganic surfaces driven by electrostatic binding or surface complexation. The thermodynamic model integrates ionization equilibrium for both polypeptide and inorganic surfaces with the Ising density functional theory (iDFT) that facilitates an accurate description of the inhomogeneous distributions of polypeptides and ionic species near inorganic surfaces. Importantly, this model accounts for both chemical and physical interactions between polypeptides and an inorganic surface, which is key to capturing the changes in adsorption due to variation in solution conditions.

To demonstrate the effectiveness of this molecular thermodynamic model, we compare theoretical predictions directly with experimental adsorption data for different oligopeptides featuring acidic, basic, and neutral monomers on a titanium surface. We then investigated the adsorption behavior of different oligopeptides at the titanium surface in aqueous solutions that are varied in pH.. We found that polypeptide adsorption typically shows a maximum in pH that is dependent upon the type of residues in the peptide. For acidic residues, the non-electrostatic surface binding makes a significant contribution to the adsorption even when the electrostatic charges of the acidic residue and the surface are of the same sign. On the other hand, a basic residue like lysine shows a much weaker dependence on non-electrostatic surface binding and the adsorption is mostly driven by electrostatic attraction from the surface. A maximum adsorption takes place at high pH where the surface sites are mostly deprotonated and as a result the non-electrostatic

interaction is diminished due to the lack of hydrogen bonding sites for the carbonyl group in the backbone.

An accurate description of the charge regulation for both polypeptides and the underlying surface in a highly inhomogeneous environment plays an important role in understanding the adsorption behavior of polypeptides, particularly when the peptide sequence contains acidic and basic residues, to inorganic surfaces. The interaction between polypeptides and the surface also affects equilibrium between different charged states leading to the speciation of the polymer significantly different from that in the bulk solution. Because of the shift in speciation, the peptide adsorbed at a highly charged surface may exist in an ionized state of opposite charge to the surface while those in the bulk solution are entirely neutral. Although the molecular-thermodynamic model employs a number of semi-empirical parameters, it provides a predictive description of thermodynamic non-idealities that are relevant to describe the environmental effects on both polypeptide adsorption and chemical equilibrium. In the future, we plan to extend this molecular-thermodynamic framework to describing the adsorption of flexible proteins which are of keen interest for practical applications such as in bioadhesives.

Acknowledgements

This work is financially supported by the NSF-DFG Lead Agency Activity in Chemistry and Transport in Confined Spaces under Grant No. NSF 2234013the National Science Foundation Harnessing the Data Revolution Big Idea under Grant No. NSF 1940118. Additional support is provided by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1326120. The computational work used resources of the

National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy, under Contract DE-AC02-05CH11231.

Bibliography

1. Lim, Y.P. and A.W. Mohammad, *Influence of pH and ionic strength during food protein ultrafiltration: Elucidation of permeate flux behavior, fouling resistance, and mechanism.* Separation Science and Technology, 2012. **47**(3): p. 446-454.

2. Barish, J.A. and J.M. Goddard, *Anti-fouling surface modified stainless steel for food processing.* Food and Bioproducts processing, 2013. **91**(4): p. 352-361.

3. Mérian, T. and J.M. Goddard, *Advances in nonfouling materials: perspectives for the food industry.* Journal of agricultural and food chemistry, 2012. **60**(12): p. 2943-2957.

4. Waite, J.H., *Mussel adhesion – essential footwork.* Journal of Experimental Biology, 2017. **220**(4): p. 517-530.

5. Stewart, R.J., et al., *The role of coacervation and phase transitions in the sandcastle worm adhesive system.* Advances in colloid and interface science, 2017. **239**: p. 88-96.

6. MacEwan, S.R. and A. Chilkoti, *Applications of elastin-like polypeptides in drug delivery.* Journal of Controlled Release, 2014. **190**: p. 314-330.

7. González-Aramundiz, J.V., et al., *Polypeptides and polyaminoacids in drug delivery.* Expert Opinion on Drug Delivery, 2012. **9**(2): p. 183-201.

8. Talmadge, J.E., *The pharmaceutics and delivery of therapeutic polypeptides and proteins.* Advanced drug delivery reviews, 1993. **10**(2-3): p. 247-299.

9. Costa, D., L. Savio, and C.-M. Pradier, *Adsorption of amino acids and peptides on metal and oxide surfaces in water environment: a synthetic and prospective review.* The Journal of Physical Chemistry B, 2016. **120**(29): p. 7039-7052.

10. Costa, D., P.A. Garrain, and M. Baaden, *Understanding small biomolecule-biomaterial interactions: A review of fundamental theoretical and experimental approaches for biomolecule interactions with inorganic surfaces.* Journal of Biomedical Materials Research Part A, 2013. **101**(4): p. 1210-1222.

11. Petrenko, V., *Evolution of phage display: from bioactive peptides to bioselective nanomaterials.* Expert opinion on drug delivery, 2008. **5**(8): p. 825-836.

12. Ogi, H., *Wireless-electrodeless quartz-crystal-microbalance biosensors for studying interactions among biomolecules: A review.* Proceedings of the Japan Academy, Series B, 2013. **89**(9): p. 401-417.

13. Van Noort, D., et al., *Monitoring specific interaction of low molecular weight biomolecules on oxidized porous silicon using ellipsometry.* Biosensors and Bioelectronics, 1998. **13**(3-4): p. 439-449.

14. Nguyen, H.H., et al., *Surface plasmon resonance: a versatile technique for biosensor applications.* Sensors, 2015. **15**(5): p. 10481-10510.

15. Landoulsi, J. and V. Dupres, *Probing Peptide–Inorganic Surface Interaction at the Single Molecule Level Using Force Spectroscopy.* ChemPhysChem, 2011. **12**(7): p. 1310-1316.

16. Maity, S., et al., *Elucidating the mechanism of interaction between peptides and inorganic surfaces.* Physical Chemistry Chemical Physics, 2015. **17**(23): p. 15305-15315.

17. Kuboyama, M., et al., *Screening for silver nanoparticle-binding peptides by using a peptide array.* Biochemical engineering journal, 2012. **66**: p. 73-77.

18. Imamura, K., et al., *Adsorption characteristics of various proteins to a titanium surface.* Journal of bioscience and bioengineering, 2008. **106**(3): p. 273-278.

19. Lambert, J.F., *Adsorption and polymerization of amino acids on mineral surfaces: a review.* Orig Life Evol Biosph, 2008. **38**(3): p. 211-42.

20. Begonja, S., et al., *Adsorption of cysteine on TiO2 at different pH values: Surface complexes characterization by FTIR-ATR and Langmuir isotherms analysis.* Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2012. **403**: p. 114-120.

21. El Shafei, G.M.S. and N.A. Moussa, *Adsorption of Some Essential Amino Acids on Hydroxyapatite.* Journal of Colloid and Interface Science, 2001. **238**(1): p. 160-166.

22. Imamura, K., et al., *Adsorption Behavior of Amino Acids on a Stainless Steel Surface.* Journal of Colloid and Interface Science, 2000. **229**(1): p. 237-246.

23. O'Connor, A.J., et al., *Amino acid adsorption onto mesoporous silica molecular sieves.* Separation and Purification Technology, 2006. **48**(2): p. 197-201.

24. Gama, M.d.S., A.G. Barreto, and F.W. Tavares, *The binding interaction of protein on a charged surface using Poisson–Boltzmann equation: lysozyme adsorption onto SBA-15.* Adsorption, 2021. **27**(7): p. 1137-1148.

25. Boubeta, F.M., G.J.A.A. Soler-Illia, and M. Tagliazucchi, *Electrostatically Driven Protein Adsorption: Charge Patches versus Charge Regulation.* Langmuir, 2018. **34**(51): p. 15727-15738.

26. Biesheuvel, P.M., F.A.M. Leermakers, and M.A.C. Stuart, *Self-consistent field theory of protein adsorption in a non-Gaussian polyelectrolyte brush.* Physical Review E, 2006. **73**(1): p. 011802.

27. Kurut, A., et al., *Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk.* Proteins: Structure, Function, and Bioinformatics, 2014. **82**(4): p. 657-667.

28. Biesheuvel, P.M., M. van der Veen, and W. Norde, *A Modified Poisson−Boltzmann Model Including Charge Regulation for the Adsorption of Ionizable Polyelectrolytes to Charged Interfaces, Applied to Lysozyme Adsorption on Silica.* The Journal of Physical Chemistry B, 2005. **109**(9): p. 4172-4180.

29. Leermakers, F.A.M., et al., *Self-Consistent-Field Modeling of Adsorbed β-Casein: Effects of pH and Ionic Strength on Surface Coverage and Density Profile.* Journal of Colloid and Interface Science, 1996. **178**(2): p. 681-693.

30. Dickinson, E., et al., *Self-consistent-field modelling of casein adsorption Comparison of results for αs1-casein and β-casein.* Journal of the Chemical Society, Faraday Transactions, 1997. **93**(3): p. 425-432.

31. Verde, A.V., J.M. Acres, and J.K. Maranas, *Investigating the specificity of peptide adsorption on gold using molecular dynamics simulations.* Biomacromolecules, 2009. **10**(8): p. 2118-2128.

32. Xie, Y., M. Liu, and J. Zhou, *Molecular dynamics simulations of peptide adsorption on self-assembled monolayers.* Applied Surface Science, 2012. **258**(20): p. 8153-8159.

33. Wei, S. and T.A. Knotts IV, *A coarse grain model for protein-surface interactions.* The Journal of chemical physics, 2013. **139**(9): p. 09B631_1.

34. Kmiecik, S., et al., *Coarse-grained protein models and their applications.* Chemical reviews, 2016. **116**(14): p. 7898-7936.

35. Hyltegren, K., et al., *Adsorption of the intrinsically disordered saliva protein histatin 5 to silica surfaces. A Monte Carlo simulation and ellipsometry study.* Journal of Colloid and Interface Science, 2016. **467**: p. 280-290.

36. Qiu, R., J. Xiao, and X.D. Chen, *Further understanding of the biased diffusion for peptide adsorption on uncharged solid surfaces that strongly interact with water molecules.* Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2017. **518**: p. 197-207.

37. Grünewald, F., et al., *Titratable Martini model for constant pH simulations.* The Journal of Chemical Physics, 2020. **153**(2): p. 024118.

38. Longo, G.S., M. Olvera de la Cruz, and I. Szleifer, *Equilibrium Adsorption of Hexahistidine on pH-Responsive Hydrogel Nanofilms.* Langmuir, 2014. **30**(50): p. 15335-15344.

39. Wagner, A.M., M.P. Gran, and N.A. Peppas, *Designing the new generation of intelligent biocompatible carriers for protein and peptide delivery.* Acta pharmaceutica sinica B, 2018. **8**(2): p. 147-164.

40. Dai, X. and Y. Chen, *Computational Biomaterials: Computational Simulations for Biomedicine.* Advanced Materials, 2022: p. 2204798.

41. Gallegos, A. and J. Wu, *Molecular thermodynamics for amino-acid adsorption at inorganic surfaces.* AIChE Journal, 2021: p. e17432.

42. Ong, G.M.C., A. Gallegos, and J. Wu, *Modeling Surface Charge Regulation of Colloidal Particles in Aqueous Solutions.* Langmuir, 2020.

43. Gallegos, A., G.M.C. Ong, and J. Wu, *Ising density functional theory for weak polyelectrolytes with strong coupling of ionization and intrachain correlations.* The Journal of Chemical Physics, 2021. **155**(24): p. 241102.

44. Gallegos, A. and J. Wu, *Charge Regulation of Natural Amino Acids in Aqueous Solutions.* Journal of Chemical & Engineering Data, 2020.

45. Jin, Z., Y. Tang, and J. Wu, *A perturbative density functional theory for square-well fluids.* The Journal of Chemical Physics, 2011. **134**(17): p. 174702.

46. Spruijt, E., P.M. Biesheuvel, and W.M. de Vos, *Adsorption of charged and neutral polymer chains on silica surfaces: The role of electrostatics, volume exclusion, and hydrogen bonding.* Physical Review E, 2015. **91**(1): p. 012601.

47. Loux, N.T., *Extending the diffuse layer model of surface acidity behaviour: III. Estimating bound site activity coefficients.* Chemical Speciation & Bioavailability, 2009. **21**(4): p. 233-244.

48. Wertheim, M., *Thermodynamic perturbation theory of polymerization.* The Journal of chemical physics, 1987. **87**(12): p. 7323-7331.

49. Yu, Y.-X. and J. Wu, *Structures of hard-sphere fluids from a modified fundamental-measure theory.* The Journal of Chemical Physics, 2002. **117**(22): p. 10156-10164.

50. Voukadinova, A., M. Valiskó, and D. Gillespie, *Assessing the accuracy of three classical density functional theories of the electrical double layer.* Physical Review E, 2018. **98**(1): p. 012116.

51. Gallegos, A., G.M.C. Ong, and J. Wu, *Thermodynamic non-ideality in charge regulation of weak polyelectrolytes.* Soft Matter, 2021.

52. Imamura, K., et al., *Adsorption characteristics of oligopeptides composed of acidic and basic amino acids on titanium surface.* Journal of Bioscience and Bioengineering, 2007. **103**(1): p. 7-12.

53. Zhuravlev, L., *Concentration of hydroxyl groups on the surface of amorphous silicas.* Langmuir, 1987. **3**(3): p. 316-318.

Chapter 8. A hierarchical model of weak polyelectrolytes with ionization and conformation consistency

Conventional models of weak polyelectrolytes are mostly concerned either with the polymer conformation in terms of an average degree of ionization or with the correlated ionization of individual segments assuming a rigid structure. While the decoupling of segment-level ionization and polymer conformation is beneficial from a computational perspective, it misses long-range intra-chain correlations (i.e., interactions beyond adjacent monomers) important for systems at low polymer concentration with strong electrostatic interactions. In this work, we propose a molecular theory that incorporates the classical density functional theory and the site-binding model to account for the effects of the local ionic environment on both inter- and intra-chain correlations. An effective persistence length is introduced to describe the coupling of polymer conformation with monomer ionization and long-range electrostatic interactions. Through an extensive comparison of the theoretical predictions with experimental titration curves for poly(acrylic acid) in different alkali chloride solutions, we demonstrate that the thermodynamic model is able to quantify the charging behavior of weak polyelectrolytes over a broad range of solution conditions.

8.1 Introduction

Understanding the charging behavior of weak polyelectrolytes is critical to technological applications such as bioadhesion[1], drug delivery[2], wastewater treatment[3] and thermal energy storage[4]. In particular, the performance of weak polyelectrolytes depends on the polymer charge and conformation and their responses to

subtle variations in the local chemical environment. Due to the long-range nature of electrostatic interactions and intra-chain correlation, it remains a theoretical challenge to accurately describe weak polyelectrolytes in dilute solutions owing to the intrinsic coupling between ionization and polymer conformation. A poor description of the long-range correlation effects may lead to erroneous predictions of "smart" polymeric systems which achieve their specific functionality such as targeted drug delivery and controlled release taking advantage of the changes in the ionization and structure of the polymer backbone in response to the pH of the intracellular fluids[2, 5].

The ionization of a weak polyelectrolyte is dependent on its conformation (and vice versa). While the charge-conformation coupling is relatively insignificant at high polymer or salt concentration due to the electrostatic screening effects, it can be significant in dilute solutions as often encountered in practical applications. The long-range intra-chain correlation is difficult to describe from a theoretical perspective. For weak polyelectrolytes at low polymer concentration, the site-binding model and its modifications have achieved great success for describing the titration of individual polymer chains by representing each polymer as a one-dimensional lattice of ionizable sites that interact only through nearest neighbors[6-8]. However, the lattice model typically neglects the charge effects on polymer conformation and long-range intra-chain interactions that are important at low electrolyte concentration. In many cases, the good performance is likely fortuitous because adjustable parameters are used to fit experimental data.

We recently demonstrated that, coupled with a thermodynamic model to account for the solution non-ideality, a nearest-neighbor site-binding model (nnSB) can accurately

capture weak polyelectrolyte titration at moderate to high salt concentrations, but deviations become more noticeable as the salt concentration is reduced[9]. The failure at low salt concentration highlights the importance of long-range intra-chain interactions to achieve a successful description of the charging behavior of dilute weak polyelectrolytes at different solution conditions. While our model reasonably describes the thermodynamic non-ideality due to the intermolecular interactions, the intramolecular interactions were truncated to the nearest-neighbor level just as in the conventional site-binding methods and, as such, the conformation of the polymer was immaterial to the ionization states of individual segments. The deficiency at low salt concentration is expected because the coupling between the polymer conformation and long-range interactions is most important for weak polyelectrolytes in dilute salt solutions. The purpose of this work is to develop a comprehensive model that can quantify both the conformation and ionization of the polymer across various salt concentrations.

Conventional site-binding models are not applicable to weak polyelectrolytes at a finite polymer concentration. For such systems, the theoretical calculations are mostly based on mean-field methods that account for the average charge density or, equivalently, the average degree of polymer ionization. The mean-field methods have received considerable attention in recent years leading to significant advancements in modeling weak polyelectrolytes in the bulk and inhomogeneous fluids[10-13]. For example, the self-consistent field theory (SCFT) was used to investigate the ionization and conformation of linear and star-like weak polyelectrolytes[14]. The mean-field approach provides a reasonable prediction of the polymer conformation, but shows significant deviation from

the experimental and simulation results for ionization in dilute electrolyte solutions. The poor performance is understandable because, like many other mean-field methods, SCFT ignores long-range intramolecular correlations. Such effects are magnified for weak polyelectrolytes at low salt concentration due to the strong Coulomb interactions. The mean-field methods become more problematic at low polymer concentration because, in that case, both the conformation and ionization behavior of weak polyelectrolytes are dominated by the intramolecular correlations.

Recently, we developed Ising density functional theory (iDFT) that treats the polymer conformation and ionization on an equal standing through a multidimensional vector specifying both the polymer ionization state and spatial configuration[15]. While iDFT shows noticeable improvements over conventional mean-field methods for describing intra-chain correlations, it is not applicable to dilute solutions of weak polyelectrolytes due to the theoretical and numerical difficulties in evaluating multibody intrachain correlations. To carry out calculations for weak polyelectrolytes, the intrachain correlations were truncated to the nearest-neighbor level resulting in a version of iDFT referred to as iDFT2. Because the excess Helmholtz energy is formulated by truncating the intramolecular correlations at the nearest-neighbor level, iDFT2 becomes inaccurate at dilute electrolyte conditions when long-range intramolecular interactions are relevant. In principle, Monte Carlo (MC) simulations are able to handle the coupled effects of the polymer ionization and conformation because they can treat both inter- and intra-molecular correlations exactly[11, 16, 17]. For example, Laguecir and coworkers showed that the MC results agreed well with the experimental data for the degree of ionization of

poly(acrylic acid) using a freely-jointed hard-sphere chain model with the electrostatic interactions described through the screened Debye-Hückel (DH) potential[18]. However, the DH potential is problematic at high salt concentration and the computational burden to sample the ensemble space increases greatly with the polymer size. The simulation is in particular computationally demanding when salt ions as well as the protons and hydroxyl ions are explicitly considered.

In this work, we develop a liquid-state method to describe the titration of weak polyelectrolytes by incorporating classical density functional theory (cDFT) with the site-binding model. Unlike previous methods, our thermodynamic model is able to account for both short- and long-range interactions within a single polymer chain explicitly. By considering ionization and chain conformation on an equal footing, we are able to describe the free energy of ionization for individual segments in a local solution environment. Besides, cDFT allows us to accurately describe the effects of solution conditions on the ionization of individual segments. Meanwhile, the site-binding model accounts for long-range intra-chain correlations and conformation of the polymer that are ignored in the conventional mean-field methods.

In the remainder of this article, we will present the theoretical details and numerical results as follows. We first outline the site-binding model and show how the long-range intra-chain correlations can be incorporated into an effective one-body potential in an accurate and computationally efficient manner. Next, we discuss how the polymer conformation can be accounted for using a free energy model. Then we show how the solution conditions can be accounted for using a molecular thermodynamic model to

describe the inter- and intramolecular interactions and correlations governing the ionization and conformation of the polymer. The combination of these different methods is referred as the hierarchical site-binding model (hSB) as we account for the behavior of the polymer at the microscopic level (viz. interactions and ionization of individual segments) to the macroscopic charge and conformation behavior of the polymer. After we have outlined are methodology, the numerical results are discussed by direct comparison with experimental titration curves for linear poly(acrylic acid) at different concentrations of alkali chloride solutions. Lastly, we investigate further the ionization and conformation of the polymer at dilute and high salt conditions.

8.2 Thermodynamic models and methods

In this section, we first recapitulate the site-binding model for weak polyelectrolytes that is conventionally used to describe the ionization of polymer segments. Next, we introduce a theoretical procedure for its extension to include long-range intra-chain interactions based on the Gibbs-Bogoliubov variational principle. A worm-like chain model is then used to describe the polymer configuration (*viz*. the distance between monomers in the chain) through an effective persistence length parameter that allows for the evaluation of the long-range interactions. Lastly, we present our molecular thermodynamic model to account for the influence of solution conditions on both the inter- and intrachain correlations by employing a tangent hard-sphere chain model for weak polyelectrolytes.

8.2.1 Site-binding model

The conventional site-binding models for describing the ionization of weak polyelectrolytes are built upon the one-dimensional Ising model[19, 20]. With the assumption that the polymer takes a rigid conformation, each ionizable segment exists in one of two charge states (*viz*. charged or uncharged), reflecting the protonation of a basic group or deprotonation of an acidic group in a polymer chain[21]. The free energy of ionization depends on the solution pH, the intrinsic binding energy of the proton to the monomer, and the energy change from the uncharged monomer to charged monomer due to nearest-neighbor interactions. In later modifications of the site-binding model[22], the intramolecular interactions can be calculated beyond the nearest-neighbor level but the chain conformation is typically assumed rigid (i.e., the distance between sites is known).

Within the site-binding method, a weak polyelectrolyte chain is represented as a collection of M ionizable sites with a specific architecture (*e.g.*, linear or branched) describing the connection (i.e., bonds) between the polymer segments. The ionization state of a polymer chain is fully specified by multi-dimensional vector $\mathbf{S} = (s_1, s_2, ..., s_M)$, where $s_i = +1$ or 0 for a basic monomer (or $s_i = -1$ or 0 for an acidic monomer) denotes the charge number (*viz*., valence $Z_i(s_i)$ of segment i). In general, the chain conformation depends on the positions of all segment, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_M)$, where $\mathbf{r}_i$ denotes segment position. For uniform systems, the multi-dimensional vector can be described in terms of the distance between the ionizable sites (*viz*. $r_{i,j} = |\mathbf{r}_i - \mathbf{r}_j|$). For a bulk system, the inter- and intra-molecular interactions are independent of the segment position $\mathbf{r}_i$. A unique

feature of a weak polyelectrolyte system is that the polymer charge is a dynamic variable coupled with the conformation.

For a weak polyelectrolyte chain consisting of only one type of ionizable groups (*i.e.*, a polyacid or polybase) in a given spatial configuration **R** specified by the set of $r_{i,j}$, the degree of ionization for the polymer (and more specifically, for each segment) can be evaluated from the semi-grand canonical partition function

$$\Xi(\mathbf{R}) = \sum_{\mathbf{S}'} \exp\left[-\beta \Gamma(\mathbf{R}, \mathbf{S}')\right] \tag{8.1}$$

where $\beta^{-1} = k_B T$, k_B and T are the Boltzmann constant and absolute temperature, respectively. Within the site-binding model, the Hamiltonian is typically expressed in terms of the one and two-body potentials:

$$\Gamma(\mathbf{R}, \mathbf{S}) = \sum_{i=1}^{M} \lambda_i \mid s_i \mid + \sum_{i=1}^{M-1} W_{i,i+1} \mid s_i s_{i+1} \mid + \sum_{i=1}^{M-2} \sum_{j>i+1}^{M} \Psi_{i,j}\left(r_{i,j}\right) \mid s_i s_j \mid \tag{8.2}$$

where we have taken the absolute value of the segment state(s) to treat acidic or basic monomers on an equal footing. The first summation on the right side of Eq. (2) accounts for the free energy change due to the binding (or release) of a proton to (or from) a basic (or acidic) monomer as well as the difference in the excess chemical potential of the segment between charged and neutral states. Note that this term is directly dependent of the ionization state of any other monomers in the polymer chain; it only feels other monomers and ions through the background mean-field. The second summation is over the intramolecular interactions between all neighboring pairs in which there is an additional energy contribution when both monomers in the pair are charged. If no further terms were

considered in Eq. (2), the Hamiltonian would be equivalent to the nearest-neighbor site-binding model where λ_i and $W_{i,i+1}$ are parameters that are correlated with experimental titration data to great success at moderate to high salt conditions[23]. For the nearest-neighbor case, the degree of ionization for segment *i* can be solved analytically and thus serves as a convenient tool to describe the ionization of weak polyelectrolytes. The simple model is not applicable to dilute conditions where long-range interactions are not sufficiently screened. In this work, we consider the long-range interaction between non-neighboring monomers that is missed by the conventional nearest-neighbor model which is given by the third summation on the right side of Eq. (2). The long-range interactions depend on the ionization state of segments *i* and *j* as well as the distance $r_{i,j}$ between the two sites. The latter is known exactly in the case of a rigid weak polyelectrolyte (e.g., a rigid rod: $r_{ij} = (j-i)\sigma$, where σ is the distance between two sites); otherwise, it must be determined in a self-consistent manner along with a theoretical procedure for the evaluation of the long-range interactions. We will discuss how we can use the worm-like chain model to determine this distance in Section 2.3.

8.2.2 Incorporating long-range interactions into the nearest-neighbor site-binding model

The added complexity introduced by long-range interactions makes the analytical evaluation of the site-binding model unrealistic. However, such a description is necessary when considering weak polyelectrolytes at conditions where the long-range interactions maintain relevancy. Even at moderate to high salt conditions, it is unlikely that the long-range interactions between monomers of short separation (*e.g.*, next nearest neighbor) are

completely screened as is assumed in the conventional nearest-neighbor site-binding model. Thus, we desire a method that allows for the efficient and accurate consideration of the long-range interactions for weak polyelectrolyte systems.

To incorporate interactions with non-neighboring monomers, we follow the so-called local effective interaction parameters (LEIP) method developed by Blanco and coworkers[22, 24-26]. In this approach, the Hamiltonian is split into a short-range (*i.e.*, nearest-neighbor) and long-range component,

$$\Gamma(\mathbf{R},\mathbf{S}) = \Gamma^0(\mathbf{S};\{w_\gamma\}) + \Delta\Gamma(\mathbf{R},\mathbf{S};\{w_\gamma\}),$$

where

$$\Gamma^0\left(\mathbf{S};\left\{w_\gamma\right\}\right) = \sum_{i=1}^{M}\left(\lambda_i + w_i\right)|\, s_i\,| + \sum_{i=1}^{M-1} W_{i,i+1}\,|\, s_i s_{i+1}\,|, \tag{8.3}$$

$$\Delta\Gamma\left(\mathbf{R},\mathbf{S};\left\{w_\gamma\right\}\right) = \sum_{i=1}^{M-2}\sum_{j>i+1}^{M}\Psi_{i,j}\left(r_{i,j}\right)|\, s_i s_j\,| - \sum_{i=1}^{M} w_i\,|\, s_i\,|, \tag{8.4}$$

and $\{w_\gamma\}$ represent a set of one-body potentials yet to be determined. Importantly, the reference Hamiltonian Γ^0 is only dependent on the one-body and nearest-neighbor interactions. The partition function corresponding to this Hamiltonian can be solved exactly within the conventional nearest-neighbor model. Next, we utilize the Gibbs-Bogoliubov inequality[27]

$$\Omega \le \tilde{\Omega} \equiv \Omega^0\left(\left\{w_\gamma\right\}\right) + \left\langle \Delta\Gamma\left(\mathbf{R},\mathbf{S};\left\{w_\gamma\right\}\right)\right\rangle_0 + F^{ent}\left(\left\{r_{i,j}\right\}\right) \tag{8.5}$$

to determine the optimal values of w_γ, where $\Omega^0(\{w_\gamma\}) = -\ln \Xi^0(\{w_\gamma\})$ and $\langle \cdots \rangle_0$ represents the free energy and the thermal average corresponding to Γ^0 (i.e., the nearest-neighbor Hamiltonian), respectively. An additional term F^{ent} is present upon taking the thermal average of $\Delta\Gamma(\mathbf{R},\mathbf{S})$ because we are also averaging over the possible conformations of the polymer. As a result, we have decoupled the direct connection between conformation and ionization; however, there is still an implicit connection due to the self-consistent relation between the distance of sites and the degree of ionization resulting from the long-range interactions. The thermal average of $\Delta\Gamma$ as considered by the unperturbed system is

$$\left\langle \Delta\Gamma\left(\mathbf{R},\mathbf{S};\{w_\gamma\}\right),\{r_{i,j}\}\right\rangle_0 = \Phi^0 - \sum_{i=1}^{M} w_i \alpha_i^0 , \tag{8.6}$$

where α_i^0 is the monomer's average degree of ionization in the reference system, and Φ^0 is the long-range interaction energy averaged over the unperturbed free energy Γ^0:

$$\Phi^0 = \left\langle \sum_{i=1}^{M-2} \sum_{j>i+1}^{M} \Psi_{i,j}\left(r_{i,j}\right) | s_i s_j | \right\rangle_0 = \sum_{i=1}^{M-2} \sum_{j>i+1}^{M} \Psi_{i,j}\left(r_{i,j}^0\right) h_{i,j}^0 , \tag{8.7}$$

where the thermal average of the distance between sites $r_{i,j}^0 = \sqrt{\langle r_{i,j}^2 \rangle_0}$ is determined based on the choice of F^{ent} and the correlation function $h_{i,j}^0$ can be evaluated analytically through the nearest-neighbor framework (see SI). The parameters $\{w_\gamma\}$ can then be determined through a minimization scheme of Eq. (5). Because the parameter w_i is affiliated with the change in free energy due to long-range intrachain interactions resulting from the

ionization of segment *i*, the variational principle implies (given that optimal values for $\{w_i\}$ is used) that all thermal averages (e.g., degree of ionization, correlation functions, etc.) can be evaluated by replacing the thermal average $\langle\cdots\rangle$ by $\langle\cdots\rangle_0$. As a result, the long-range interactions can be conveniently incorporated into the nearest-neighbor framework.

While the LEIP method was employed previously by Blanco and coworkers[22, 24-26], it did not account for the solution condition effect on the terms λ_i and $W_{i,i+1}$. In addition, the polymer conformation was evaluated through the rotational isomeric state (RIS) model[24] or by simulating the stretching of the chain[25]. In this work, we aim to account for the solution conditions on the inter- and intramolecular interactions through a molecular thermodynamic model. We also account for the polymer conformation $r_{i,j}^0$ through a worm-like chain model that allows for an analytical evaluation of its properties.

8.2.3 Spatial configuration of the weak polyelectrolyte

The spatial configuration of the weak polyelectrolyte is directly related to the polymer's ionization and vice versa. Thus, it is critical for the conformation of the polymer to be considered when evaluating the thermodynamic properties of weak polyelectrolyte systems. Whereas a comprehensive picture of the polymer includes explicitly *both* the spatial configuration **R** and state configuration **S** when evaluating the thermal average, such calculation is cumbersome due to the inherent multibody integration required. Simulation techniques can be employed to estimate the ionization and conformation behavior, but such methods are in general time consuming[11]. Instead, we take an alternative approach that focuses on the macroscopic feature of the chain (*viz*. the end-to-end distance) which decouples the spatial and state configuration of the polymer, but still

accounts for the relation between the ionization and conformation implicitly in a self-consistent manner.

According to the worm-like chain model proposed by Shen and Wang[28, 29], the average distance between segments *i* and *j* in a polymer chain is given by

$$\left(r_{i,j}^{0}\right)^{2}=\left\langle r_{i,j}^{2}\right\rangle_{0}=2\ell^{eff}\left(j-i\right)\sigma\left\{1-\frac{\ell^{eff}}{\left(j-i\right)\sigma}\left[1-\exp\left(-\frac{\left(j-i\right)\sigma}{\ell^{eff}}\right)\right]\right\} \tag{8.8}$$

where ℓ^{eff} is the effective persistence length, and σ is the bond length. For a worm-like chain, the radius of gyration is given by

$$R_{g}^{2}=\frac{M\ell^{eff}\sigma}{3}-\left(\ell^{eff}\right)^{2}\left\{1-\frac{2\ell^{eff}}{M\sigma}+2\left(\frac{\ell^{eff}}{M\sigma}\right)^{2}\left[1-\exp\left(\frac{M\sigma}{\ell^{eff}}\right)\right]\right\}. \tag{8.9}$$

This expression results in the correct limiting behavior for the radius of gyration of an ideal gaussian chain and rod-like chain (viz. $R_{g}^{ideal}=\sigma\sqrt{M/6}$ and $R_{g}^{rod}=\sigma M/\sqrt{12}$, respectively). The free energy is given by

$$F^{WLC}=F^{ent}+\Phi^{0} \tag{8.10}$$

where Φ^{0} is the long-range interaction energy defined previously in Eq. (7) and

$$F^{ent}=-\frac{3}{2}M\ln\left(1-\frac{\upsilon^{2}}{M}\right)-3\ln\upsilon. \tag{8.11}$$

In Eq.(11), the chain expansion is defined by

$$\upsilon^{2}=\frac{\left\langle r_{1,M}^{2}\right\rangle}{\left(M-1\right)\sigma^{2}}=2\frac{\ell^{eff}}{\sigma}\left\{1-\frac{\ell^{eff}}{\left(M-1\right)\sigma}\left[1-\exp\left(-\frac{\left(M-1\right)\sigma}{\ell^{eff}}\right)\right]\right\}, \tag{8.12}$$

which is the end-to-end distance of the worm-like polymer normalized by the end-to-end distance of an ideal chain, and υ is a function of ℓ^{eff}. The first term on the right side of Eq.(12) represents the entropic contribution due to the compression of the chain; the expression is an interpolation between the elastic free energies obtained from integrating the worm-like chain and the freely-jointed chain force-extension relationships through a finite extensibility approximation[30, 31]. The second contribution arises from the resistance to chain compression; it was first deduced by Flory[32, 33]. Since the long-range interaction energy depends on the distance between the monomers, it is explicitly related to the parameter ℓ^{eff} through $r_{i,j}^{0}$. Thus, we must determine numerically the effective persistence length that minimizes the free energy of the worm-like chain F^{WLC}. The energy contribution from the long-range component is also dependent on the ionization state of the polymer; therefore, the distance parameter is also indirectly related to the charge regulation of the polymer. As a result, we are able to provide insight into the chain expansion as a function of solution conditions and ionization degree of the polymer.

8.2.4 Molecular thermodynamic model for weak polyelectrolytes

In the prior sections, we have introduced a framework to describe the ionization and conformation of weak polyelectrolytes applicable to all solution conditions. However, to accurately capture the behavior of the polymer, we must be able to quantify how the solution conditions effects the inter- and intra-chain interactions. Here, we introduce a molecular thermodynamic model for weak polyelectrolytes that provides a means to account for the thermodynamic non-ideality resulting from the solution conditions.

We employ the primitive model for polyelectrolyte solutions to account for electrostatic interactions and short-range interactions. As shown in Figure 1, each polymer is represented by a tangent hard-sphere chain where each segment can be either charged or uncharged depending on its ionization state. For simplicity, we assume that the hard-sphere diameter σ_i is not influenced by the ionization state or salt concentration, but may depend on the salt type. Unlike typical coarse-grained models used in MC simulations, the salt ions are explicitly considered as charged hard spheres. The hydrated diameter for the salt ions was taken from Simonin et al.[34] (*viz.* $\sigma_{Li^+} = 4.76\,\text{Å}$, $\sigma_{Na^+} = 3.90\,\text{Å}$, $\sigma_{K^+} = 3.34\,\text{Å}$, $\sigma_{Cs^+} = 1.89\,\text{Å}$, and $\sigma_{Cl^-} = 3.62\,\text{Å}$),. The polymer and salt ions are immersed in a continuum medium with a dielectric constant representative of liquid water at room temperature ($\varepsilon_r = 78.4$).

Thermodynamic non-ideality plays an important role in both the one- and two-body potentials employed in the site-binding model. In the conventional nearest-neighbor model, λ_i and $W_{i,i+1}$ change with the solution condition. Here, we show that these terms can be described using analytical expressions derived from the molecular thermodynamic model which allows for an accurate depiction of the inter- and intra-chain interactions. The one-body potential depends on the difference in intermolecular interactions between the charged and uncharged state of the monomer as well as a contribution due to the binding energy of the proton:

$$\lambda_i = -s_i k_B T \left(\mathrm{pK}_i - \mathrm{pH}\right) \ln 10 + \Delta\mu_i^{mon} + \Delta\mu_i^{ch} \tag{8.13}$$

where the first term on the right is affiliated with the change in free energy due to the binding of a proton, and the middle term results from change in intermolecular interactions of segment *i* with the other monomers and ions in the system when charged versus uncharged, and the last term is the change in excess chemical potential due to chain connectivity which describes the change in intrachain correlation energy for all chains due to segment *i* when it is charged versus uncharged. Typically, the monomeric contribution μ_i^{mon} can be decomposed into an excluded volume term $\Delta\mu_i^{hs}$ and an electrostatic correlations term $\Delta\mu_i^{el}$.

In Eq. (13), pK_i is the negative logarithm of the equilibrium constant for the deprotonation or protonation reaction. This term depends on the identity of the ionizable site and temperature, but not on the solution composition. As mentioned above, $s_i = +1$ corresponds to the protonation of a weak basic segment, and $s_i = -1$ for the deprotonation of a weak acidic segment. In this work, we use the pK_i for the acrylic acid monomer (viz., pK_i = 4.756[35]). The hard-sphere component $\Delta\mu_i^{hs}$ is only relevant when the hard-sphere diameter of the monomer varies with the ionization state (e.g., due to the change in hydration). In general, this term is important only in highly concentrated electrolyte solutions and therefore, for simplicity, we assume the monomer's diameter is independent of its ionization state (i.e., $\Delta\mu_i^{hs} = 0$)[36]. The electrostatic correlations term, $\Delta\mu_i^{el}$, arises from the non-uniform distribution of free ions and other polymer segments around a charged monomer.

We utilize the mean-spherical approximation (MSA) to describe the electrostatic correlations[37]. MSA captures the finite-size effect of ions on the electrostatic interactions while reducing to the Debye-Hückel (DH) limiting law at dilute electrolyte concentrations[38]. According to MSA, the change in excess chemical potential due to electrostatic correlations is expressed as

$$\beta\Delta\mu_i^{el} = -l_B\left[\frac{\Gamma^{MSA}}{1+\Gamma^{MSA}\sigma_i} + \frac{2Z_i\left(s_i\right)\eta\sigma_i}{1+\Gamma^{MSA}\sigma_i}\right] \tag{8.14}$$

where $l_B = \beta e^2 / 4\pi\varepsilon_0\varepsilon_r$ is the Bjerrum length (7.14 Å for liquid water at room temperature), e is the elementary charge; ε_0 and ε_r are the vacuum permittivity and dielectric constant of the solvent, respectively; σ_i and $Z_i(s_i)$ are the hard-sphere diameter and valence of the monomer in state s_i; η is a parameter related to the asymmetry of the system (i.e., differences in size and valence of the ionic species); and Γ^{MSA} is the MSA screening parameter. Approximately, Γ^{MSA} is proportional to the square root of the ionic strength similar to that for the Debye screening parameter. As a result of this square root dependency, the importance of electrostatics becomes relevant even at dilute conditions (which will be seen in the two-body terms as well).

We next focus our attention towards the nearest-neighbor energy term, $W_{i,i+1}$, used in the site-binding model. This term accounts for the free energy to ionize a segment given that its neighbor is already ionized; thus, it is affiliated with the intrachain interactions. By contrast, the intermolecular interactions are accounted for by the one-body term λ_i. As we showed in our previous work[9], this term can be decomposed into two contributions:

$W_{i,i+1} = -k_B T \ln g^{el}_{i,i+1}(\sigma_{i,i+1}) + u_{i,i+1}$ where the first term accounts for the electrostatic work to bring to charged segments from infinity to contact (i.e., to form an ionic bond), and the second term is non-electrostatic in origin and accounts for specific interactions (e.g., breakdown of intramolecular hydrogen bonding[39]). While a quantitative prediction of the non-electrostatic energy is difficult to achieve from a molecular perspective, we expect that the short-range interaction is relatively insensitive to solution conditions. In this work, the non-electrostatic component $u_{i,i+1}$ is estimated by fitting the experimental data across different solution conditions and the value determined is used as the non-electrostatic energy between all paired segments.

The electrostatic component in the nearest-neighbor term can be estimated from the radial distribution function for the two segments at contact using a reference system of charged hard spheres described by MSA:[40]

$$g^{el}_{i,i+1} = \exp\left[-\frac{l_B}{\sigma} \frac{Z_i - \eta\sigma^2}{1+\Gamma^{MSA}\sigma} \frac{Z_{i+1} - \eta\sigma^2}{1+\Gamma^{MSA}\sigma} \right]. \tag{8.15}$$

At infinite dilution, the radial distribution function reduces to the exponential of the negative contact value of the electrostatic pair potential (*i.e.*, $-\ln g^{el}_{i,i+1} = l_B Z_i Z_{i+1} / \sigma_{i,i+1}$). Since $\Gamma^{MSA} \sim \sqrt{I}$, the electrostatic work will always reduce in magnitude with added salt due to electrostatic screening. Therefore, added salt can promote ionization if the neighboring segments have the same charge sign or inhibit ionization if they have opposite charge signs. In addition, due to the square root dependence of Γ^{MSA} on the salt concentration, the electrostatic work will vary significantly even in dilute salt solutions.

While the nearest-neighbor term has shown to be sufficient at capturing the titration behavior of polyacids in moderate to high salt concentrations, it misses the interactions with segments beyond those that are adjacent to it. As a result, the titration behavior is noticeably overpredicted in dilute solutions because the additional resistance to ionization due to the long-range interactions is not accounted for in the Hamiltonian.

In order to capture the long-range interactions, an additional term is necessary in the Hamiltonian compared to the conventional site-binding model [see Eq. (2)]: $\Psi_{i,j}$. The long-range contribution depends on the distance between monomers *i* and *j*, $r_{i,j}$, and their charge states s_i and s_j. The explicit coupling of ionization and conformation has been a major difficulty in analytical methods due to the large computational burden that is required to consider all polymer configurations. Using the Gibbs-Bogoliubov variational principle shown in Section 2.2, we are able to incorporate the long-range correlations in the nearest-neighbor framework through a straightforward, yet efficient manner. In addition, we can account for the spatial configuration of the polymer (i.e., $r_{i,j}$) by minimizing the free energy of the worm-like chain model outlined in Section 2.3.

We now focus our attention on the expression used for long-range potential. The usual choice for the electrostatic interaction between charges in solution follows from the Debye-Hückel potential:

$$\beta \Psi_{i,j}\left(r_{i,j}\right) = \frac{l_B}{r_{i,j}} Z_i Z_j \exp\left[-\kappa r_{i,j}\right]; \tag{8.16}$$

however, such potential is only valid for dilute and uncorrelated systems[18]. Nonetheless, the DH potential conveniently captures both the long-range nature of coulomb interactions and the screening effect resulting from salt ions through the DH parameter κ. Since the interaction strength only depends on the salt concentration and distance between the sites, the DH potential is unable to discriminate between the salt-specific effects that result in differences in the titration behavior of weak polymers in different electrolyte solutions[41]. In addition, we can intuitively expect that there is an accumulation of counterions within the vicinity of the charged polymer that leads to a stronger electrostatic screening than expected from the bulk salt concentration. The local concentration of ions near the polymer depends on the degree to which the polymer is charged and the strength of electrostatic interactions between the polymer segments and the ions. The local electric potential describes these two factors; unfortunately, such information is not easily determined without expensive computational procedures such as iDFT[15]. Instead, we approximate the local electric potential of the weak polyelectrolyte (WPE) by

$$\phi_{WPE}^{local} \approx \phi_{mon}^{local} \times \sqrt{\alpha_{WPE}}\,, \tag{8.17}$$

where ϕ_{mon}^{local} is the electrostatic potential at the surface of an isolated charged monomer (i.e., not in a polymer chain) in the background salt solution and α_{WPE} is the average degree of ionization for the polymer. The local electric potential is related to the degree of ionization by its square-root based on preliminary fitting as well as the fact that the counterions will quickly accumulate in the vicinity of the polymer upon charging. In principle, the local electric potential is segment-dependent (i.e., the electric potential is not

constant along the polymer chain); however, such considerations would be cumbersome from a computational perspective. We can determine the electrostatic potential at the surface of the isolated charged monomer through classical density functional theory (cDFT) by fixing a particle of absolute charge *e* and hard-sphere diameter σ_i at the origin. cDFT provides an accurate description of the inhomogeneous distribution of ions in an aqueous environment[42]. Additional details on the electrostatic potential for the monomer can be found in the SI. With the local electric potential for the polymer now determined, we can find the local concentration for each ion by

$$c_i^{local} = c_i \exp(-\beta e Z_i \phi_{WPE}^{local}) \tag{8.18}$$

where c_i and Z_i are the bulk concentration and valence of ion *i*. Some general trends to note is that when the polymer is uncharged, the local counterion concentration is the same as that in the bulk. As the polymer charges, the counterions will accumulate at the surface at a rate dependent upon the electrostatic driving force (i.e., ϕ_{mon}^{local}). The strength of the driving force is greatest at dilute salt concentrations because the electrostatic screening is weakest. Thus, at high salt concentration, the electrostatic potential for the monomer will be small and the local concentration of ions will be only slightly greater than the bulk solution. As a result of this modification, the DH parameter κ is replaced by a local DH parameter $\kappa^{local} = \sqrt{4\pi l_B \sum_i c_i^{local}}$.

Besides accounting for the local concentration of ions in the vicinity of the polymer, we must also account for the salt-dependent nature of the long-range interaction. Experimental titration data for poly(acrylic acid) indicates that the choice of counterion

results in a difference in the ionization behavior of the polymer at low salt concentration. In order to capture this effect, we introduce an empirical parameter τ_{cat^+} that modifies the strength of screening by the counterion. We find that the electrostatic screening is weaker for more hydrated ions (viz. lithium and sodium) than its less hydrated counterparts (viz. potassium and cesium). It seems that the charge compensation by the counterion is lessened by the presence of a large hydration shell which may be attributed to an effective screening by the dipole nature of the solvent (viz. water).

The two-body potential can be derived by an analytical approximation of the radial distribution function of charged spheres from MSA reported by Henderson & Smith[43]. With the proposed modifications of c_s^{local} and τ_{cat^+}, we obtain

$$\beta\Psi_{i,j}\left(r_{i,j}\right)=\frac{l_B}{r_{i,j}}\frac{Z_i-\eta\sigma^2}{1+\Gamma^{MSA}\sigma}\frac{Z_j-\eta\sigma^2}{1+\Gamma^{MSA}\sigma}\exp\left[-\tau_{cat^+}\kappa^{local}\left(r_{i,j}-\sigma\right)\right], \qquad (8.19)$$

which is equivalent to the nearest-neighbor potential [see Eq. (15)] when $r_{i,j}=\sigma$. Using Eq. (22) to calculate the long-range potential, we are able to account for the intrachain interactions between charged monomers at dilute salt conditions. Note that our equation does not reduce to the DH limit for large *r* due to the τ_{cat^+} present in the exponential and the use of the local concentration in κ^{local}. However, as discussed previously, the DH potential is not valid for strongly correlated systems and thus, we expect that our potential captures the long-range intrachain correlations due to the specific screening effect of salt counterions.

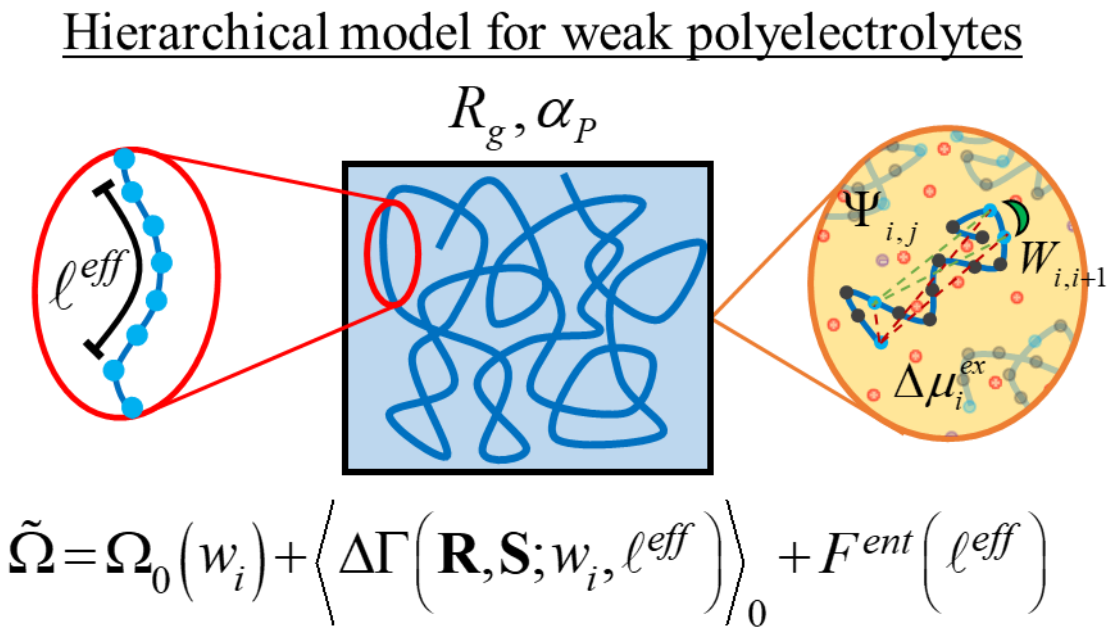

Figure 8.1 A schematic of the hierarchical model used to determine the charge regulation and conformation behavior of weak polyelectrolytes in an aqueous salt solution.

We show a summary of our theoretical procedure to determine the ionization and conformation behavior of weak polyelectrolytes in aqueous salt solution in Figure 1. Our method employs the site-binding model to describe the Hamiltonian of weak polyelectrolytes in their different ionization states due to the inter- and intramolecular interactions. Unlike the conventional nearest-neighbor site-binding model, we consider an additional contribution due to the long-range interactions between sites. While usually the inclusion of non-neighboring interactions is computationally expensive, we take advantage of the Gibbs-Bogoliubov inequality to approximate the long-range correlations into an effective one-body term (i.e., the interaction of a segment with all other segments is described by the average behavior). In addition, the conformation of the polymer is directly coupled to the ionization of the polymer which can also be costly to consider. By

employing a free energy description of the chain conformation (viz. worm-like chain model) in terms of an effective persistence length parameter ℓ^{eff}, we are able to self-consistently couple the ionization and conformation of the weak polyelectrolyte since the long-range interactions depend on ℓ^{eff}. Lastly, we can account for the solution condition effects on the one-body term, λ_i, and the two-body terms, $W_{i,i+1}$ and $\Psi_{i,j}$, through our molecular thermodynamic model. Specifically, we introduce a long-range potential that accounts for the accumulation of counterions around the polymer and the salt-specific screening effects. By combination of these approaches, we can determine the ionization and conformation behavior of weak polyelectrolytes at all solution conditions.

8.3 Results and discussion

In the following, we illustrate the long-range correlation effects on the conformation and the ionization behavior of weak polyelectrolytes. For the calibration of theoretical results, our discussion is focused on the aqueous solutions of poly(acrylic acid). PAA is a linear polyacid with a relatively low density of ionizable groups. At medium to high salt concentrations, the weak interaction between monomers in poly(acrylic acid) leads to a titration curve that exhibits a smooth transition from the fully uncharged state to fully charged state. Unlike that for strongly correlated weak polymers such as poly(maleic acid)[39], PAA ionization lacks a step-like feature in the titration curve. Its titration behavior has been well studied using different experimental[41, 44-46] and computational methods[47-49]. The experimental results can be reasonably described with various mean-field approximations owing to the weak interactions between neighboring sites[21, 41, 50]. However, as was shown by our previous implementation[9], the truncation of intramolecular

correlations to the nearest-neighbor (or completely neglecting the correlation effects like in some mean-field theories) results in a poor description of the physics at low salt conditions[15]. Here, we demonstrate a comprehensive model that fully captures the ionization and conformation behavior of weak polymers at all solution conditions.

Table 8.1 Coarse-grained parameters used to describe the ionization and conformation of poly(acrylic acid) in different salt solutions.

Poly(acrylic acid)	Known values	Salt-dependent parameters	τ_{cat^+}	$u_{i,i+1}$ (k_BT)
pK_i	4.756	LiCl	0.168	0.062
σ	4.874 Å	NaCl	0.182	0.075
		KCl	0.315	0.471
		CsCl	0.617	0.556

Our thermodynamic model employs two fitting parameters to match the experimental data: τ_{cat^+} and $u_{i,i+1}$ which are the effective screening parameter and the non-electrostatic pair energy, respectively. We show in Table 1 the value for the parameters used to capture the experimental data of poly(acrylic acid) in different salt solutions. We consider two additional parameters, pK_i and σ_i, which are the thermodynamic equilibrium constant and the hard-sphere diameter, respectively, to be the same for all monomers in the poly(acrylic acid) chain (i.e., no difference in the chemical structure or hydration of the monomers). Unlike the parameters employed in the conventional nearest-neighbor model, the parameters determined here are intrinsic to the polymer and they are independent of the solution conditions (viz. pH and salt concentration). As a result, the thermodynamic equilibrium constant and monomer size should be the same for all alkali metals because it is an intrinsic property of the polymer (i.e., independent of the specific salt present in the

solution). We chose the pK_i value to be the same as that of the free monomer acrylic acid (viz. $pK_i = 4.756$[35]); thus, our model reduces to the single monomer limit. In our previous work using a nearest-neighbor framework, the pK_i value needed to be fit and was significantly larger (viz. $pK_i = 5.22$)[9]. The larger value of our previous model is not surprising since the long-range interactions were not included and these interactions will inhibit the ionization of the polymer. Therefore, to a zeroth-order approximation, the larger pK_i in the nearest-neighbor model accounted for this. The excluded volume (i.e., hard-sphere diameter) of a monomer is an important term as this governs the distance between neighboring sites in our model. In the nearest-neighbor framework, it was necessary to fit σ_i for PAA in different salt solutions (i.e, LiCl vs CsCl) in order to capture the experimental data. In this work, we determine the hard-sphere diameter based off the excluded volume calculated by the molecular model of PAA using the Molinspiration Cheminformatics 2018 software package (http://www.molinspiration.com). The estimated volume for a monomer in PAA was 60.63 $Å^3$ which corresponds to a hard-sphere diameter of $\sigma_i = 4.874\,Å$. Thus, our model is free of fitting parameters for the intrinsic properties of the polymer (i.e., size and pK_i).

We find that the non-electrostatic energy and the effective screening parameter to be dependent on the specific salt cation present in the solution. Since our coarse-grained model does not account for specific chemical effects due to water, we expect that these parameters incorporate such effects in an approximate manner. In general, the hydration of cations is highly dependent upon the Pauling radius of the ion ($\sigma_{hyd} \sim 2/\sigma_{Pauling}$)[51]. Thus, the adsorption of the cations to the vicinity of the polymer will impact the water

structure at the surface of the polymer due to the difference in hydration of the cations. The local modulation of the water structure could contribute to the repulsive non-electrostatic energy between neighboring sites which we find to be more repulsive when the cation is less hydrated. The non-electrostatic energy $u_{i,i+1}$ was determined to be -0.062, 0.075, 0.471, and 0.556 k_BT for PAA in solution of LiCl, NaCl, KCl, and CsCl, respectively. There is a significant difference in $u_{i,i+1}$ between Na^+ and K^+ while the values are similar for K^+ and Cs^+ which follows from more hydrated nature of Na^+ whereas K^+ and Cs^+ are minimally hydrated.

Besides the non-electrostatic energy, the effective screening parameter is expected to be dependent on the salt type, but not on the salt concentration, since it is related to the ability of the counterions to screen charges between monomers. We find that the screening parameter increases in value (i.e., stronger electrostatic screening) in the order of $Li^+<Na^+<K^+<Cs^+$. Since the counterions each have their own unique hydration structure, we expect that the hydration layer of the ion must be the reason for the different effective screening of the ions. While the larger hydration of the ion leads to a larger hard-sphere diameter, excluded volume effects are relatively insignificant at the concentration ranges considered. Thus, we expect that the weaker screening by Li^+ can be attributed to the dipole of water molecules reducing the contribution of a bare charge to screen out charge-charge interactions. We find the screening parameter τ_{cat^+} to be 0.168, 0.182, 0.315, and 0.617 for Li^+, Na^+, K^+, and Cs^+, respectively.

8.3.1 Titration of poly(acrylic acid) in different alkali chloride solutions

Experiments have demonstrated that the choice of alkali metal influences the charging behavior of poly(acrylic acid); in particular, at low salt concentrations where long-range interactions play a key role in the ionization of the polymer[41]. Since previous methods have relied on the DH potential to describe long-range interactions[22], they are unable to differentiate between the different salt solutions. In this work, we have modified the long-range interaction term [Eq. (21)] to include an effective screening parameter τ_{cat^+} that accounts for the effect of the alkali cation on the charging behavior. We propose that this term originates from dipole water molecules in the hydration layer of the alkali cation reducing the effective charge of the cation. As a result, the more hydrated ions like Li^+ screen less effectively than the less hydrated cations such as Cs^+ (viz. τ_{cat^+} is 0.168 and 0.617 for Li^+ and Cs^+, respectively).

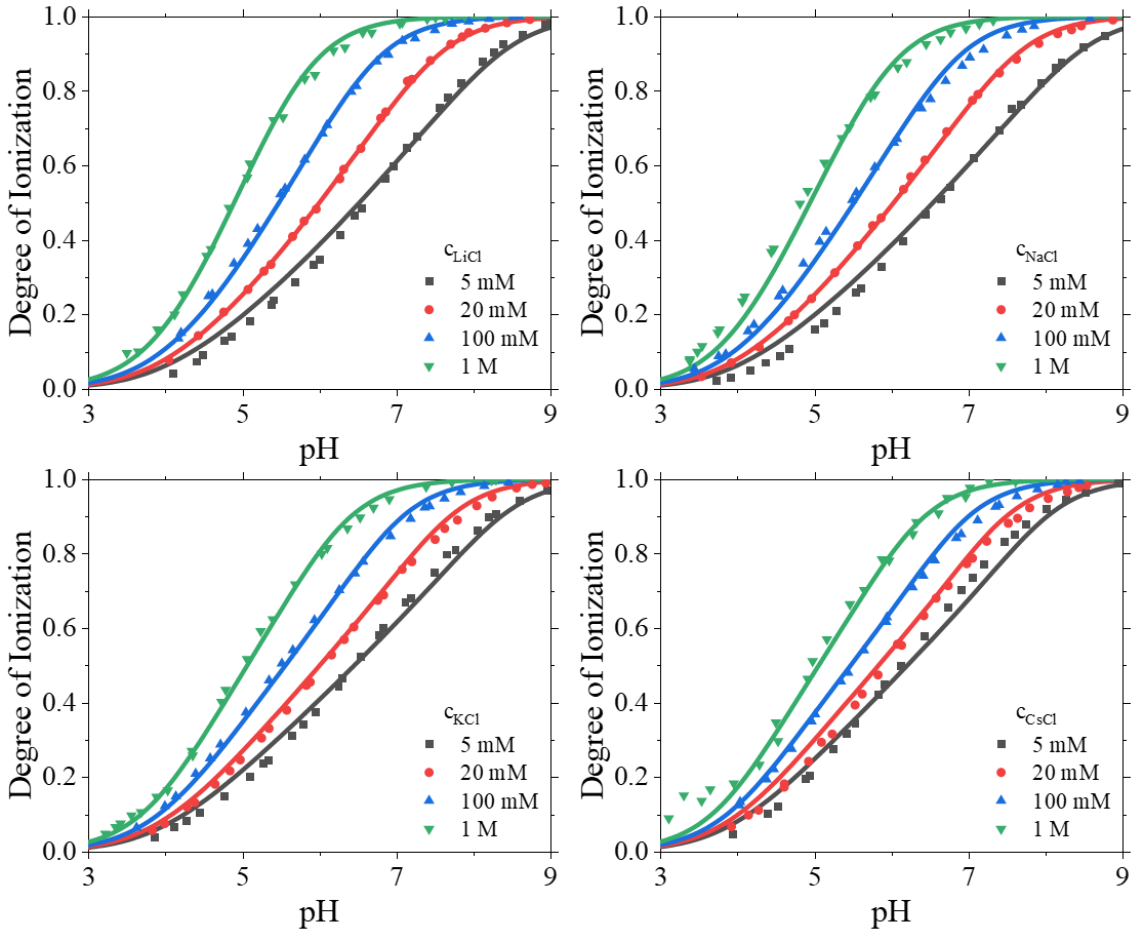

Figure 8.2 Titration curves for poly(acrylic acid) at 25 °C in the aqueous solutions of (a) lithium chloride, (b) sodium chloride, (c) potassium chloride, and (d) cesium chloride solutions from experiment[41] (symbols) and theoretical correlations (lines).

We show in Figure 2 the titration curves of PAA (MW=88 kg/mol; *M*~1200) in four types of alkali chloride aqueous solutions (lithium, sodium, potassium, and cesium) at different salt concentrations[41]. The theoretical predictions by the hierarchical site-binding model outlined in this work is given by the solid lines and the experimental data by the symbols. The ionization of the acidic polymer increases when the pH or salt concentration is increased. As the salt concentration is increased, the long-range

interactions between monomer segments in the polymer chain weaken and ultimately are mostly screened out leading to an increase in the charge of the polymer. Interestingly, the salt counterion (viz., cation for the case of the polyacid) influences the charging behavior of the polymer. Our previous work using the nearest-neighbor site-binding model relied on different size parameters for the acrylic acid monomer in order to capture the salt-dependent ionization. However, the influence of salt type is most evident as the salt concentration is decreased indicating that it is the long-range intrachain correlations that lead to the difference in polymer charge with different salt cations. Thus, we fixed the monomer size based off volumetric data for the acrylic acid monomer and employed a long-range potential that accounts for the salt-type through the empirical parameter τ_{cat^+}. For $\tau_{cat^+} < 1$, this potential decays slower than predicted by the DH potential which means that the intrachain energy maintains its significance at larger separations. The long-range correlation between monomers in the polymer chain is reasonable since the cumulative charge in the vicinity of the polymer is significant and should behave different from its free (i.e., not in a chain) counterparts. Our model can well capture the titration behavior of poly(acrylic acid) across a myriad of solution conditions owing to the successful description of both the long-range interactions and the self-consistent conformation of the polymer with ionization.

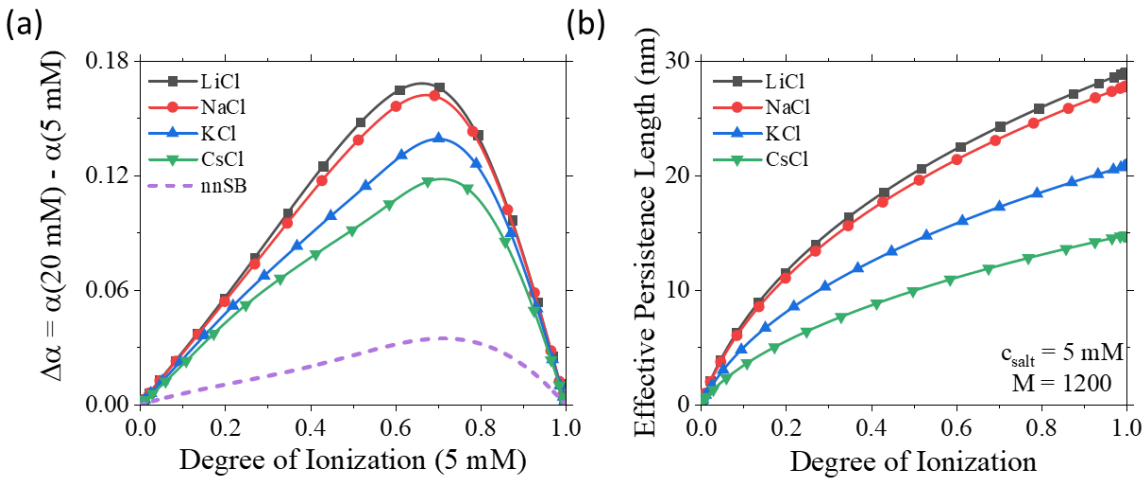

Figure 8.3 (a) The difference in degree of ionization of poly(acrylic acid) at 20 mM and 5 mM lithium chloride solution in four different salt solutions (viz. LiCl, NaCl, KCl, and CsCl). The dashed line is the prediction of the nearest-neighbor model for LiCl. (b) The effective persistence length of poly(acrylic acid) in a 5 mM salt solution.

We show the difference in degree of ionization between 5 mM and 20 mM for different alkali chloride solutions as determined from our model calculations in Figure 3a. In addition, the prediction from the nearest-neighbor site-binding model is also shown. The figure clearly demonstrates that the ionization behavior differs significantly between the different alkali metals. The reason for the larger change in ionization of PAA in LiCl compared to the other alkali metals originates from the weaker screening effect of Li^+ ions on the long-range interaction as described through our model by τ_{cat^+}. As the hydration layer decreases in size around the alkali ion, the electrostatic screening by the ions increases and therefore the long-range potential is less important which leads to less change in charge with salt concentration.

Next, we consider the conformation behavior of the polymer in the different alkali chloride solutions at a salt concentration of 5 mM in Figure 3b. Since the long-range

interactions are stronger in LiCl than CsCl, it is therefore expected that the expansion of the polymer will be greater in LiCl. We find this to be true as the effective persistence length, which corresponds to the expansion of the polymer [υ in Eq. (15)], decreases from LiCl to CsCl. Our findings agree with a previous experimental study that found the radius of gyration of PAA to decrease from LiCl to CsCl[45]. We see that the polymer expands quickly at low degree of ionization and then increases in size further at a slower rate. The reason for the slower expansion is the entropic penalty that resists the polymer extending due to the charge-charge interactions. Thus, there is a self-consistent nature in the polymer ionization and conformation as the polymer expands to decreases the repulsive long-range interactions. If we assumed a rigid model for the weak polyelectrolyte, e.g., an ideal gaussian chain or a rod-like, we would miss the coupling of ionization and conformation in the charge regulation of polymers and find the agreement with experiments to be less satisfactory.

8.3.2 Titration of poly(acrylic acid) in lithium chloride solution

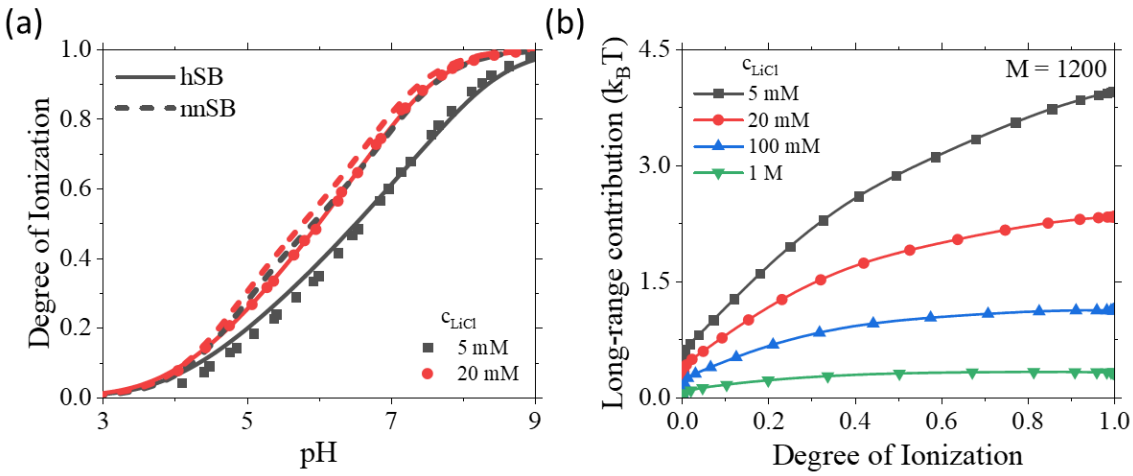

Figure 8.4 (a) The degree of ionization for poly(acrylic acid) in a 5 mM and 20 mM lithium chloride solution. The solid and dashed lines are the correlated results for our hierarchical site-binding model (hSB) and the nearest-neighbor site-binding model (nnSB) using optimized parameters from our previous work[9], respectively. The symbols are experimental data for PAA (MW=88 kg/mol; M~1200). (b) The long-range contribution (w_i) to the free energy of ionization of a single segment as a function of degree of ionization in four different lithium chloride solutions.

To better understand the importance of long-range interactions, we compare the theoretical predictions for the hierarchical site-binding model (hSB) and nearest-neighbor site-binding model (nnSB) for the titration behavior of poly(acrylic acid) (MW=88 kg/mol; *M*~1200) immersed in a lithium chloride aqueous solution at 5 mM and 20 mM in Figure 4a. We show the theoretical predictions for the nearest-neighbor model and our self-consistent approach that accounts for the ionization and conformation behavior due to long-range interactions. As expected, the nearest-neighbor model completely misses the ionization behavior as the salt concentration is decreased. On the other hand, the hierarchical model can correctly capture the ionization of the polymer at 5 mM and 20 mM LiCl indicating that the long-range interactions are properly accounted for. When they are

not included, such as in the nearest-neighbor framework, the change in ionization with salt concentration becomes minimal since the change in the intermolecular and nearest-neighbor intrachain interactions are small. However, the long-range intrachain interactions are significantly affected by the change in the salt concentration due to the exponential factor related to the salt concentration through the local Debye parameter (see Eq. [21]). Thus, the new model correctly predicts a noticeable drop in ionization when the LiCl concentration is reduced from 20 mM to 5 mM.

We show how the long-range contribution changes with degree of ionization at different salt concentrations in Figure 4b. In general, the long-range contribution increases steadily with ionization due to the higher number of charge-charge interactions. However, the increase in salt concentration will weaken the long-range correlations resulting in a noticeable drop in the energy contribution at a given degree of ionization. At 1 M LiCl, the contribution due to long-range interactions is negligible and the difference between the nearest-neighbor model and our self-consistent approach would be approximately equivalent because the charge-charge interaction is effectively screened. At the lowest concentration considered here (viz. 5 mM), the repulsive long-range contribution is around 2.5 k_BT at complete ionization which is equivalent to reducing the pH by 1 unit. Thus, the driving force for ionization is significantly impacted by the long-range interactions present between charged monomers in the polymer chain. Clearly any method that does not incorporate these important long-range interactions is missing the key physics governing polymer ionization.

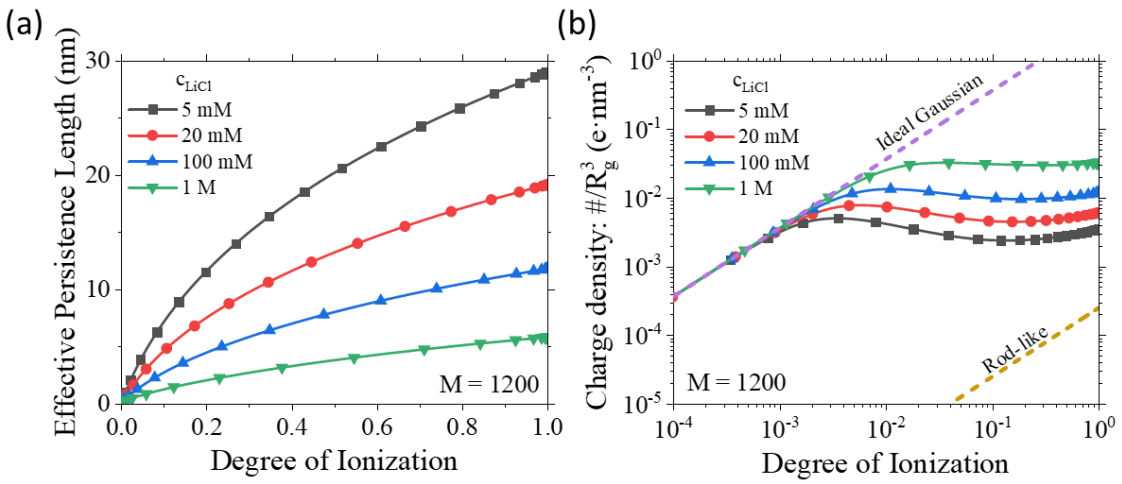

Figure 8.5 (a) The effective persistence length ℓ^{eff} and (b) the charge density for poly(acrylic acid) as determined through our self-consistent model as a function of degree of ionization at four different lithium chloride solutions.

The ionization of the polymer leads to a corresponding change in the polymer conformation due to the self-consistent nature of our coarse-grained model for PAA, which is also applicable to other weak polyelectrolytes. The polymer will expand with ionization due to the strong repulsion between the same-charged monomers in the polymer chain. Thus, it can be expected that salt will play a major role in determining the size of the polymer chain since the long-range interactions are heavily dependent on the salt concentration through the Debye parameter κ. We show the variation in effective persistence length of the polymer with degree of ionization at different lithium chloride concentrations in Figure 5a. Here we see clearly that the ionization of the polymer directly leads to a corresponding increase in the polymer size and the increase in salt concentration leads to a large drop in ℓ^{eff}. It can be expected that further increase in the salt concentration would lead to the effective persistence length changing very little with ionization because the long-range interactions are fully screened. It is at this point that the nearest-neighbor

model and long-range model converge, and the polymer behaves like an ideal gaussian chain (i.e., random-walk) where $\langle r_{i,j}^2 \rangle = (j-i)\sigma^2$. On the other hand, further decrease of the salt concentration would lead to a stronger driving force for the expansion of the polymer due to the long-range repulsive interactions. In this case, the polymer is approaching the rod-like limit where $\langle r_{i,j}^2 \rangle = (j-i)^2\sigma^2$. Thus, our self-consistent approach interpolates between these two limits by accounting for the entropic and electrostatic contributions for chain expansion.

To further examine the self-consistent determination of the polymer conformation, we show the charge density for PAA at different lithium chloride concentrations in Figure 5b. For any rigid conformation (e.g., ideal gaussian or rod-like), charging of the polymer will lead to a linear increase in the charge density with ionization. However, when the conformation is self-consistently coupled to the ionization, it can be expected that the polymer will expand to minimize the repulsive interactions. As a result, the charge density of our model polymer will be between the ideal gaussian and rod-like limit. We expect that the polymer will behave like an ideal gaussian chain when it is uncharged and then deviate from this limit as the ionization increases. Our model predicts a rapid expansion of the polymer at low degree of ionization as was seen through the effective persistence length after which ℓ^{eff} increases at a lesser rate with ionization. The expansion of the polymer leads to a maximum in the charge density with ionization which is more noticeable at lower salt concentrations. In general, the charge density of the polymer depends on the salt concentration since its expansion is related to the strength of the intramolecular repulsion. Neglecting the intramolecular interactions like in mean-field theories and the

nearest-neighbor site-binding model would miss the physics that leads to the expansion of the polymer at low salt conditions.

8.3.3 Effect of chain length on ionization and conformation properties of PAA

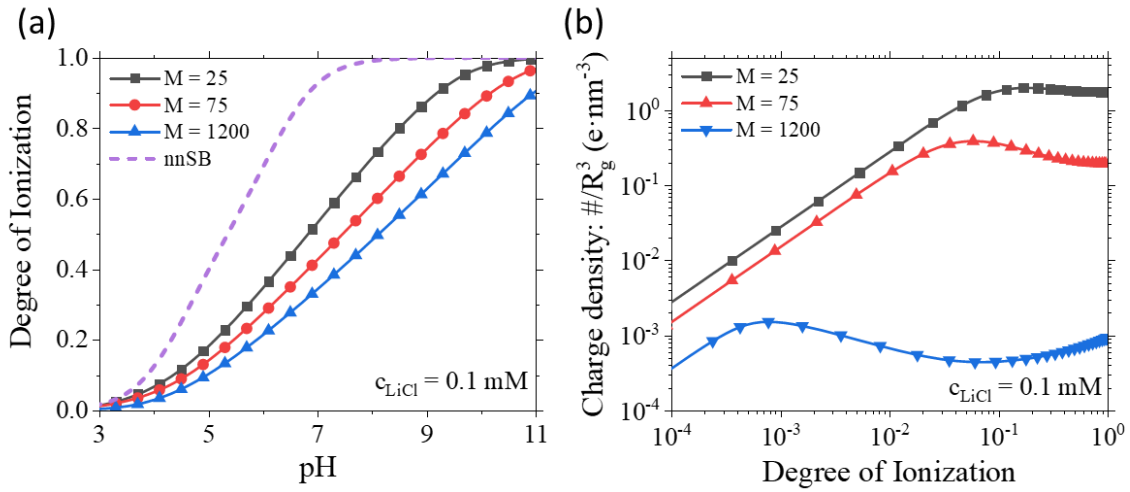

Figure 8.6 (a) The degree of ionization as a function of pH for three different chain lengths (M=25, 75, and 1200) in a 0.1 mM lithium chloride solution. The dashed lines are the nearest-neighbor site-binding model predictions. (b) The charge density of the polymer as a function of degree of ionization for three different chain lengths.

Lastly, we consider the ionization and conformation behavior of poly(acrylic acid) at different chain lengths. In the original site-binding model, it was found that the increase in chain length beyond 20 monomers led to no significant difference in ionization[21]. However, experimental results have demonstrated that the chain length can have a substantial dependence on the polymer's titration properties[18]. The reason for the failure of the nearest-neighbor model is due to its neglection of the long-range interactions between charged monomers in the chain. We show the ionization behavior of PAA in a 0.1 mM lithium chloride solution at three different chain lengths (viz. *M* = 25, 75, and 1200) in Figure 6a. Our model predicts that increasing the chain length does lead to a noticeable decrease in the ionization of the polymer at these solution conditions. The chain

length effect is most noticeable at higher degrees of ionization; however, even at low degrees of ionization there is a noticeable increase in resistance to ionization as the chain length increases. The reason for this is that even at the same degree of ionization, a longer chain will have a higher number of repulsive interactions due to there being more charged monomers ($n_{charged} = \alpha M$). Since the nearest-neighbor model does not account for these long-range interactions, the chain length effect predicted by nnSB becomes irrelevant past $M > 20$ for the model poly(acrylic acid).

It has been previously demonstrated that the salt concentration has a substantial influence on the role that chain length has on the ionization of the polymer[18]. The relevancy of salt concentration for the chain length effect is because the salt concentration limits the distance that the long-range interactions are relevant. Thus, if the salt concentration is high, a monomer only feels (i.e., has a significant interaction with) the first few monomers that are adjacent to it. On the other hand, if the salt concentration is low, the monomer will feel other charged monomers that are separated by tens to hundreds of monomers along the chain. Since the polymer has a finite chain length, the range of relevant long-range interactions can exceed the distance between the monomers in the chain; thus, a longer chain would lead to more inhibition on the ionization due to the additional repulsive contribution of these added monomers. We show the ionization behavior of PAA for different chain lengths at 100 mM in Figure S1. As expected, the chain length effect becomes less relevant as the salt concentration increases.

The charge density of the polymer is also highly dependent on the polymer's chain length. We show the charge density of PAA as a function of degree of ionization at three

different chain lengths in Figure 6b. We find that the charge density decreases with chain length since the number of charges scales with *M* at the same degree of ionization and the radius of gyration scales with $M^{0.5}$ when the polymer is mostly uncharged. Thus, the charge density approximately scales with $M^{-0.5}$ at low degree of ionization. Similar to Figure 3b, the charge density curve increases linearly with degree of ionization as it follows the same behavior as an ideal gaussian chain. However, once the polymer reaches a sufficient charge, the polymer size expands quickly before stabilizing to an approximate linear increase in the volume ($\sim R_g^3$) with degree of ionization (i.e., charge density curve does not change with further ionization). While the long chain polymer ($M = 1200$) shows a noticeable maximum in charge density with degree of ionization, the maximum is relatively insignificant for the short chain polymer ($M = 25$). This is due to the smaller entropic penalty that occurs for the expansion of the polymer as the length increases. Besides the difference in relative maximum for different chain length, the position of the maximum occurs at lower degrees of ionization as the chain length increases since the expansion is related to the number of charged monomers ($n_{charged}$). We find that the maximum occurs when ~4 monomers are charged in the polymer chain for *M=25* and *75*, but approximately *1* for *M=1200*. The expansion of the polymer chain when more than one monomer is charged for the long chain can be attributed to the ease in the long chain polymer to expand. Clearly, our model is able to capture the self-consistent nature of the ionization and conformation behavior of poly(acrylic acid) in aqueous solutions.

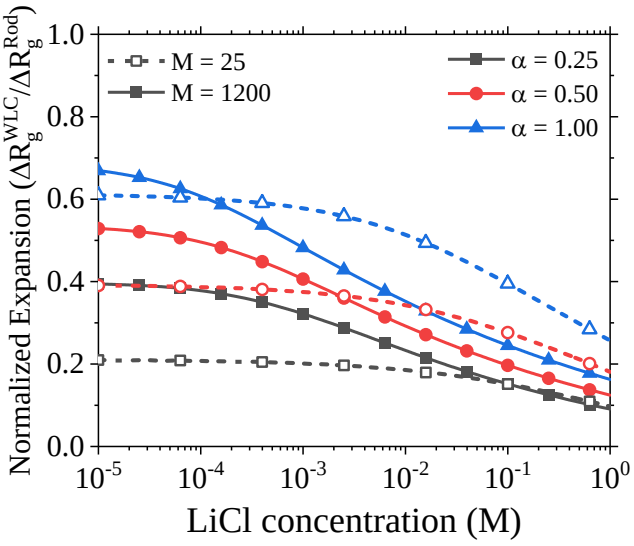

Figure 8.7 The normalized expansion defined as the ratio of the difference in radius of gyration between the worm-like chain (WLC) and the ideal gaussian chain ($R_g^{WLC} - R_g^{ideal}$) over the difference of the rod-like chain and the ideal gaussian chain ($R_g^{Rod} - R_g^{ideal}$) at different lithium chloride concentrations for three different degrees of ionization and two different chain lengths.

To better understand the difference in conformation of the short chain ($M = 25$) and long chain ($M = 1200$) polymer, we show the normalized expansion $\Delta R_g^{WLC} / \Delta R_g^{Rod}$, where $\Delta R_g^{***} = R_g^{***} - R_g^{ideal}$, in Figure 7a. When the normalized expansion is close to zero, the polymer conformation is similar to that of an ideal chain whereas a value close to unity means that the polymer is significantly stretched in only one-direction. As expected, an increase in ionization leads to an increase in the normalized expansion (i.e., towards to the rod-like limit) due to the increase in intramolecular repulsion resulting from the charged sites. On the other hand, the increase in salt concentration decreases the normalized expansion (i.e., towards the ideal-chain limit). As salt is added to the solution, the long-range interactions are further screened which leads to a smaller driving force for the

expansion of the polymer. It should be noted that as the salt concentration is decreased, the pH must be increased to maintain the same degree of ionization. As a result, counterions (i.e., lithium ions) must be added to maintain charge neutrality with the protons and hydroxyl ions ($c_{OH^-} >> c_{H^+}$) in solution; therefore, the actual concentration of ions is much higher than shown here. To clarify this, we have included an additional figure that shows the results for Figure 7 when the pH effect on the concentration of ions is considered as well but still plotted in terms of the background lithium chloride concentration. In general, the normalized expansion approaches a limit at dilute salt concentrations due to the tradeoff between the expansion of the polymer due to long-range repulsive interactions and the resistance to expansion due to the entropic penalty resulting from the limiting number of possible conformations. Increasing the chain length generally results in further expansion at dilute salt concentrations since there is more charged monomers to interact with; however, there is a limit since the repulsion between the monomers at large separation may not be enough to compete with the entropic penalty for further expansion. A higher line charge density (i.e., a closer spacing between charged sites in the backbone of the polymer) polymer would have a conformation closer to the rod-limit than a lower line charge density polymer since the intrachain repulsive is greater.

We find that the behavior of the normalized expansion for the short chain and long chain is highly dependent on the salt concentration. The long chain expands quicker with charge than its short chain counterpart at dilute salt conditions; however, at high salt conditions, the long chain polymer remains collapsed while the short chain polymer still expands to a significant degree. The reason for the opposing trends can be attributed to the

range and strength of the long-range interactions as well as the greater conformational freedom when the chain length is increased. When the salt concentration is dilute, the long-range interactions are only somewhat screened and therefore they maintain relevancy even between monomers that are significantly separated. The resulting repulsion between monomers will promote the expansion of the polymer; and in the case of the long-chain polymer, the expansion occurs with little entropic penalty. On the other hand, the short-chain polymer faces a greater entropic penalty when expanding and therefore a larger driving force (i.e., further ionization of the polymer) is necessary for the expansion to be similar to the long chain polymer. However, at high salt concentration, the long-range interactions quickly decay due to the shortened screening length (κ^{-1}) and therefore these long-range interactions are not as relevant. As shown in Figure 6b, at the same degree of ionization, the charge density of the polymer decreases with an increase in chain length. Therefore, the smaller expansion of the long chain polymer at high salt concentration can be attributed to the fact that the charged sites are further apart from one another than in the short-chain case. In fact, the greater separation between charged sites is also the reason that the normalized expansion begins to decrease at an earlier salt concentration for the long chain than the short chain. A similar logic can be applied as to why the salt concentration shows less influence on the short chain expansion since the charged sites are closer to one another and therefore the interactions cannot be sufficiently screened. Thus, a comprehensive model that accounts for the entropic and coulombic contributions to the self-consistent determination of the polymer conformation is vital to the correct depiction of weak polyelectrolytes in aqueous solution.

8.4 Conclusion

We have developed a thermodynamic model to account for the ionization and conformational behavior of weak polyelectrolytes in aqueous solution through a self-consistent treatment of the inter- and intramolecular correlations using a worm-like tangent chain model. This model improves upon conventional methods by incorporating the conformation and long-range correlations (i.e., between non-neighboring segments) that are typically neglected by approximating these correlations through an effective one-body potential determined via the Gibbs-Bogoliubov variational principle. This allows for a convenient yet accurate description of weak polymer ionization even though the long-range interactions are treated implicitly. In addition, we find that any consideration of a rigid conformation (i.e., one that does not depend on ionization) leads to a poor agreement between experimental results and model predictions.

Besides the self-consistent nature of our model, we also correctly capture the influence of solution conditions on the inter- and intramolecular correlations of the polymer through our thermodynamic model. The correct description of non-ideality effects is key to the accurate depiction of weak polyelectrolytes in different salt solutions as well as at different salt concentrations. It also allows for a predictive capability to our model by using physically reasonable values that can be determined via correlation with experimental data or through some other geometrical means. Whereas the nearest-neighbor framework works well at moderate to high salt conditions, our model maintains its accuracy even at low salt concentrations since we account for the long-range correlations. In particular, we include an empirical correction to the long-range potential through an effective screening

parameter that accounts for the salt cations ability to screen the long-range potential due to its hydration layer and by accounting for an elevated concentration of counterions in the vicinity of the polymer. Neglecting either of these corrections results in a form similar to the screened Debye-Huckel potential typically used by others and would lead to unsatisfactory results.

In this work, we have focused our attention towards poly(acrylic acid) which is a linear weak polyacid with a low density of charged sites along the backbone (i.e., weakly correlated). Poly(acrylic acid) serves as a reasonable starting point to investigate other systems of interest including strongly correlated polymers like poly(maleic acid) that also exhibit stereochemical-dependent features as well as hetero-weak polyelectrolytes such as zwitterionic polymers. In particular, the long-range correlations are important for the latter since conventional methods that truncate at the nearest-neighbor level will miss the attractive and repulsive interactions between non-adjacent sites. Lunkad et al. demonstrated the importance of considering long-range interactions by showing the similarity in titration behavior of a block copolymer with one block containing acidic monomers and the other block containing basic monomers versus a polymer with alternating acidic and basic monomers[52]. Molecular architectures besides the linear polymer such as star-like polymers or dendrimers are also of importance due to their practical interest[7, 14]. While this model is a significant advancement over conventional methods, the explicit coupling of long-range interactions and spatial configuration of the polymer is only implicitly accounted for through the effective persistence length parameter. We are currently pursuing further development along these directions.

Bibliography

1. Kord Forooshani, P. and B.P. Lee, *Recent approaches in designing bioadhesive materials inspired by mussel adhesive protein.* Journal of polymer science. Part A, Polymer chemistry, 2017. **55**(1): p. 9-33.

2. Priya James, H., et al., *Smart polymers for the controlled delivery of drugs – a concise overview.* Acta Pharmaceutica Sinica B, 2014. **4**(2): p. 120-127.

3. Bolto, B. and J. Gregory, *Organic polyelectrolytes in water treatment.* Water Research, 2007. **41**(11): p. 2301-2324.

4. Seitz, S. and H. Ajiro, *Self-assembling weak polyelectrolytes for the layer-by-layer encapsulation of paraffin-type phase change material icosane.* Solar Energy Materials and Solar Cells, 2019. **190**: p. 57-64.

5. Jaganathan, S., *Bioresorbable polyelectrolytes for smuggling drugs into cells.* Artificial Cells, Nanomedicine, and Biotechnology, 2016. **44**(4): p. 1080-1097.

6. de Groot, J., et al., *Dissociation Behavior of Poly(maleic acid): Potentiometric Titrations, Viscometry, Pulsed Field Gradient NMR, and Model Calculations.* Macromolecules, 1998. **31**(13): p. 4182-4188.

7. van Duijvenbode, R.C., M. Borkovec, and G.J.M. Koper, *Acid-base properties of poly(propylene imine)dendrimers.* Polymer, 1998. **39**(12): p. 2657-2664.

8. Borkovec, M., G.J.M. Koper, and C. Piguet, *Ion binding to polyelectrolytes.* Current Opinion in Colloid & Interface Science, 2006. **11**(5): p. 280-289.

9. Gallegos, A., G.M.C. Ong, and J. Wu, *Thermodynamic non-ideality in charge regulation of weak polyelectrolytes.* Soft Matter, 2021.

10. Gonzalez Solveyra, E., et al., *Theoretical Modeling of Chemical Equilibrium in Weak Polyelectrolyte Layers on Curved Nanosystems.* Polymers, 2020. **12**(10): p. 2282.

11. Landsgesell, J., et al., *Simulations of ionization equilibria in weak polyelectrolyte solutions and gels.* Soft Matter, 2019. **15**(6): p. 1155-1185.

12. Nap, R.J., S.H. Park, and I. Szleifer, *Competitive calcium ion binding to end-tethered weak polyelectrolytes.* Soft Matter, 2018. **14**(12): p. 2365-2378.

13. Rathee, V.S., et al., *Weak polyelectrolyte complexation driven by associative charging.* The Journal of Chemical Physics, 2018. **148**(11): p. 114901.

14. Uhlík, F., et al., *Modeling of Ionization and Conformations of Starlike Weak Polyelectrolytes.* Macromolecules, 2014. **47**(12): p. 4004-4016.

15. Gallegos, A., G.M.C. Ong, and J. Wu, *Ising density functional theory for weak polyelectrolytes with strong coupling of ionization and intrachain correlations.* The Journal of Chemical Physics, 2021. **155**(24): p. 241102.

16. Hyltegren, K. and M. Skepö, *Adsorption of polyelectrolyte-like proteins to silica surfaces and the impact of pH on the response to ionic strength. A Monte Carlo simulation and ellipsometry study.* Journal of Colloid and Interface Science, 2017. **494**: p. 266-273.

17. Carnal, F. and S. Stoll, *Adsorption of Weak Polyelectrolytes on Charged Nanoparticles. Impact of Salt Valency, pH, and Nanoparticle Charge Density. Monte Carlo Simulations.* The Journal of Physical Chemistry B, 2011. **115**(42): p. 12007-12018.

18. Laguecir, A., et al., *Size and pH effect on electrical and conformational behavior of poly(acrylic acid): Simulation and experiment.* European Polymer Journal, 2006. **42**(5): p. 1135-1144.

19. Brush, S.G., *History of the Lenz-Ising model.* Reviews of modern physics, 1967. **39**(4): p. 883.

20. Cipra, B.A., *An introduction to the Ising model.* The American Mathematical Monthly, 1987. **94**(10): p. 937-959.

21. Koper, G.J.M. and M. Borkovec, *Proton binding by linear, branched, and hyperbranched polyelectrolytes.* Polymer, 2010. **51**(24): p. 5649-5662.

22. Garcés, J.L., et al., *Dealing with long-range interactions in the determination of polyelectrolyte ionization properties. Extension of the transfer matrix formalism to the full range of ionic strengths.* Journal of Polymer Science Part B: Polymer Physics, 2017. **55**(3): p. 275-284.

23. Smits, R.G., G.J.M. Koper, and M. Mandel, *The influence of nearest- and next-nearest-neighbor interactions on the potentiometric titration of linear poly(ethylenimine).* The Journal of Physical Chemistry, 1993. **97**(21): p. 5745-5751.

24. Blanco, P.M., et al., *Coupling of charge regulation and conformational equilibria in linear weak polyelectrolytes: Treatment of long-range interactions via effective short-ranged and pH-dependent interaction parameters.* Polymers, 2018. **10**(8): p. 811.

25. Blanco, P.M., et al., *Effect of charge regulation and conformational equilibria in the stretching properties of weak polyelectrolytes.* Macromolecules, 2019. **52**(21): p. 8017-8031.

26. Blanco, P.M., et al., *Role of charge regulation and fluctuations in the conformational and mechanical properties of weak flexible polyelectrolytes.* Polymers, 2019. **11**(12): p. 1962.

27. Chandler, D., *Introduction to modern statistical.* Mechanics. Oxford University Press, Oxford, UK, 1987. **5**.

28. Shen, K. and Z.-G. Wang, *Electrostatic correlations and the polyelectrolyte self energy.* The Journal of chemical physics, 2017. **146**(8): p. 084901.

29. Shen, K. and Z.-G. Wang, *Polyelectrolyte chain structure and solution phase behavior.* Macromolecules, 2018. **51**(5): p. 1706-1717.

30. Marko, J.F. and E.D. Siggia, *Stretching dna.* Macromolecules, 1995. **28**(26): p. 8759-8770.

31. Jha, P.K., et al., *Nonlinear effects in the nanophase segregation of polyelectrolyte gels.* Macromolecules, 2009. **42**(16): p. 6284-6289.

32. Flory, P.J., *Principles of polymer chemistry*. 1953: Cornell University Press.

33. De Gennes, P.d., *Collapse of a polymer chain in poor solvents.* Journal de Physique Lettres, 1975. **36**(3): p. 55-57.

34. Simonin, J.-P., L. Blum, and P. Turq, *Real Ionic Solutions in the Mean Spherical Approximation. 1. Simple Salts in the Primitive Model.* The Journal of Physical Chemistry, 1996. **100**(18): p. 7704-7709.

35. Lide, D.R., *CRC handbook of chemistry and physics*. Vol. 85. 2004: CRC press.

36. Gallegos, A. and J. Wu, *Charge Regulation of Natural Amino Acids in Aqueous Solutions.* Journal of Chemical & Engineering Data, 2020.

37. Blum, L., *Mean spherical model for asymmetric electrolytes.* Molecular Physics, 1975. **30**(5): p. 1529-1535.

38. Maribo-Mogensen, B., G.M. Kontogeorgis, and K. Thomsen, *Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions.* Industrial & Engineering Chemistry Research, 2012. **51**(14): p. 5353-5363.

39. Kawaguchi, S., et al., *Dissociation behavior of poly (fumaric acid) and poly (maleic acid). II. Model calculation.* Macromolecules, 1990. **23**(3): p. 731-738.

40. Zhou, Y., S. Yeh, and G. Stell, *Criticality of charged systems. I. The restricted primitive model.* The Journal of Chemical Physics, 1995. **102**(14): p. 5785-5795.

41. Sadeghpour, A., et al., *Influence of alkali metal counterions on the charging behavior of poly(acrylic acid).* Polymer, 2009. **50**(16): p. 3950-3954.

42. Yu, Y.-X., J. Wu, and G.-H. Gao, *Density-functional theory of spherical electric double layers and ζ potentials of colloidal particles in restricted-primitive-model electrolyte solutions.* The Journal of chemical physics, 2004. **120**(15): p. 7223-7233.

43. Henderson, D. and W.R. Smith, *Exact analytical formulas for the distribution functions of charged hard spheres in the mean spherical approximation.* Journal of Statistical Physics, 1978. **19**(2): p. 191-200.

44. Blaakmeer, J., et al., *Adsorption of weak polyelectrolytes on highly charged surfaces. Poly (acrylic acid) on polystyrene latex with strong cationic groups.* Macromolecules, 1990. **23**(8): p. 2301-2309.

45. Kim, H.G., et al., *Dissociation Behavior of Surface-Grafted Poly(Acrylic Acid): Effects of Surface Density and Counterion Size.* Journal of Colloid and Interface Science, 1993. **157**(1): p. 82-87.

46. Wu, T., et al., *Behavior of surface-anchored poly (acrylic acid) brushes with grafting density gradients on solid substrates: 1. Experiment.* Macromolecules, 2007. **40**(24): p. 8756-8764.

47. Léonforte, F., U. Welling, and M. Müller, *Single-chain-in-mean-field simulations of weak polyelectrolyte brushes.* The Journal of Chemical Physics, 2016. **145**(22): p. 224902.

48. Gong, P., et al., *Behavior of surface-anchored poly (acrylic acid) brushes with grafting density gradients on solid substrates: 2. Theory.* Macromolecules, 2007. **40**(24): p. 8765-8773.

49. Lützenkirchen, J., et al., *Comparison of Various Models to Describe the Charge−pH Dependence of Poly(acrylic acid).* Journal of Chemical & Engineering Data, 2011. **56**(4): p. 1602-1612.

50. Kitano, T., et al., *Dissociation behavior of poly(fumaric acid) and poly(maleic acid). 1. Potentiometric titration and intrinsic viscosity.* Macromolecules, 1987. **20**(7): p. 1598-1606.

51. Marcus, Y., *Thermodynamics of solvation of ions. Part 5.—Gibbs free energy of hydration at 298.15 K.* Journal of the Chemical Society, Faraday Transactions, 1991. **87**(18): p. 2995-2999.

52. Lunkad, R., et al., *Role of pKA in Charge Regulation and Conformation of Various Peptide Sequences.* Polymers, 2021. **13**(2): p. 214.

Chapter 9. Single-chain-in-density-functional-theory simulations for weak polyelectrolytes

Conventional theories of weak polyelectrolytes are either computationally prohibitive to account for multidimensional inhomogeneity or oversimplistic in describing the coupling effects of ion-explicit interactions with long-range inter- and intrachain correlations. To bridge this gap, we implement the Ising density functional theory for ionizable systems with the single-chain-in-mean-field algorithm. The single-chain-in-density-functional-theory (scDFT) shows significant improvements over conventional mean-field methods by incorporating both ion-explicit interactions and long-range correlations. With an explicit consideration of position-dependent polymer configurations and the ionization states of individual segments, scDFT provides a faithful description of the structure and thermodynamic properties of weak polyelectrolytes thereby opening up new opportunities for engineering design and applications.

9.1 Introduction

Weak polyelectrolytes play a critical role in a wide variety of natural phenomena occurring in biological systems and are broadly used in many technological applications such as water treatment, bioadhesion and gene delivery[1-4]. The diverse functionality of weak polyelectrolytes is largely attributed to the adaptability of the electrostatic charge and polymer conformation to external stimuli and local solution environment. Numerous studies have been devoted to understanding the structure and thermodynamic properties of weak polyelectrolytes[5-8]. However, it remains a theoretical challenge to predict both short-range and long-range correlation effects underlying the complex interplay of

monomeric ionization with polymer conformation. Existing theoretical methods, including the Poisson-Boltzmann equation[9, 10], polymer scaling laws[11, 12], self-consistent-field theory[13-15], and coarse-grained molecular models[16-18], typically describe the coupling effects between the average ionization and the conformation of weak polyelectrolyte chains without an explicit consideration of small ions. Meanwhile, simulation methods are often limited to short polymer chains without an explicit consideration of small ions due to the computational cost[19-21]. While the mean-field methods are quite successful in capturing certain aspects of weak polyelectrolyte systems, they are not able to account for ion-explicit electrostatic effects and long-range correlations underpinning polymer ionization at the segment-level and therefore have limited capability in predicting the functionality.

Recently, a new theoretical framework for weak polyelectrolytes was developed by combination of the Ising model with the polymer density functional theory[22]. In principle, the Ising density functional theory (iDFT) provides an exact procedure to describing the equilibrium properties of ionizable systems by treating the charge state and the position of individual segments on an equal footing. An accurate free-energy functional can be formulated to account for both short and long-range interactions including intrachain correlations. While the theoretical procedure is applicable to both atomic and coarse-grained models of ionizable systems, the numerical difficulty in solving the density profiles of polymer conformations limits its application to systems with only one-dimensional inhomogeneity and nearest-neighbor interactions (hereby referred to as 1D-iDFT2). In this work, we demonstrate that this caveat can be corrected by implementing

iDFT within the framework of the single-chain-in-mean-field (SCMF) algorithm[23]. The SCMF algorithm temporarily approximates the explicit interactions between ions, solvent, and polymer chains in terms of effective one-body interactions (*viz*., external fields). A system of independent polymer chains is simulated in these external fields in order to capture the conformational distributions of polymer chains (i.e., through Monte Carlo sampling of the single-chain conformations). The simulation results, in turn, are used to calibrate the external potentials of individual species by utilizing the polymer density functional theory. The iterative procedure eschews the direct simulation of the multi-chain system while capturing both intermolecular interactions and intra-chain correlations explicitly (on long wavelengths). The quasi-instantaneous field approximation[24] invoked by the SCMF algorithm is also able to capture the dynamic behavior of polymer systems if the polymer conformations vary slowly on the time scale applied to updating the external fields. A previous implementation of the SCMF algorithm for weak polyelectrolyte systems did not consider explicitly the ionization state of individual segments or the intrachain electrostatic correlations[23]. In combination with iDFT, the SCMF algorithm will be able to capture both short-range and long-range correlations. We refer to our approach as single-chain-in-density-functional-theory (scDFT) to distinguish from 1D-iDFT2 in which the long-range intrachain correlation effects were ignored. We will demonstrate that scDFT improves upon the conventional mean-field approaches by accounting for the segment-level ionization behavior of the weak polyelectrolytes and the conformations of individual chains due to its sequence of charged segments.

9.2 Thermodynamic models and methods

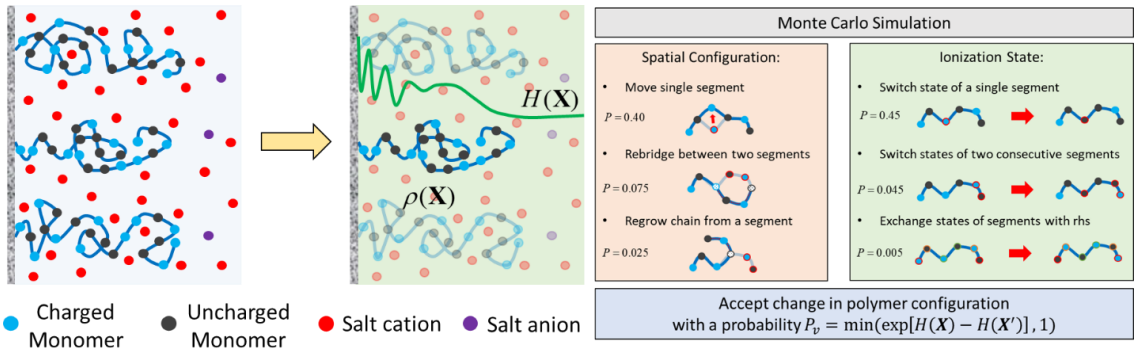

Figure 9.1 Schematic of single-chain simulation in the presence of an effective potential (viz., external field) that accounts for inter- and intramolecular correlations. The simulated polymer may change its conformation or the ionization state of individual segments through various Metropolis Monte Carlo moves.

To demonstrate the numerical procedure, we consider a generic model for a weak polyelectrolyte brush as shown schematically in Figure 1. Each polymeric species is represented by tangent hard-sphere chains of ionizable segments, monomeric ions by charged hard spheres each with a fixed valence, and the solvent by a dielectric continuum. The coarse-grained model explicitly accounts for protons and hydroxyl ions, which are directly related to the pH of the system (viz. the bulk densities are proportional to $\rho^b_{H^+} \sim 10^{-\mathrm{pH}}$ and $\rho^b_{OH^-} \sim 10^{\mathrm{pH}-14}$). For a polymer chain with *M* segments, the conformation is fully specified by the multidimensional vector, $\mathbf{R}_k = (\mathbf{r}_{k,1}, \mathbf{r}_{k,2}, ..., \mathbf{r}_{k,M_k})$, where $\mathbf{r}_{k,i}$ denotes the position of segment i in polymer chain *k*, and the ionization of each chain can be specified by vector $\mathbf{S}_k = (s_{k,1}, s_{k,2}, ..., s_{k,M_k})$, where $s_{k,i} = \pm 1$ or 0 is the charge number [viz., valence $Z_k(s_{k,i})$]. A single-chain configuration is fully specified by the multidimensional vector $\mathbf{X}_k = (\mathbf{R}_k, \mathbf{S}_k)$ where the position and ionization state of segment

i is $\mathbf{x}_{k,i} = (\mathbf{r}_{k,i}, s_{k,i})$. A unique feature of weak polyelectrolyte systems is that the polymer charge is a dynamic variable coupled with the conformation. The dynamic coupling is responsible for the correlated ionization of neighboring segments as well as the long-ranged intrachain correlations.

The single-chain-in-mean-field (SCMF) simulation is concerned with an ensemble of independent polymer chains under an effective single-chain Hamiltonian $H(\mathbf{X})$ that accounts for the potential of mean force for each polymer chain due to interactions with the environment. The single-chain Hamiltonian includes contributions from the inter- and intrachain correlations as well as the free energy of deprotonation/protonation of individual segments and polymer interactions between other ionic species in the solution. Note that the effective single-chain Hamiltonian, $H(\mathbf{X})$, depends on $\mathbf{X}$ since the position and state of all other segments in the polymer chain are needed to determine the long-range intrachain interactions. The exact contributions to this field will be discussed later in the iDFT section.

In implementing SCMF simulation, we first generate a set of n (~1000) self-avoiding polymer chains that are independent of one another where the first segment (left-hand side) is connected tangentially to the surface and all other segments cannot intersect with the wall (i.e., hard wall). In addition, each polymer chain satisfies the bonding constraint set by the freely jointed tangent chain model (viz., $|\mathbf{r}_i - \mathbf{r}_{i+1}| = \sigma$) where σ is the hard-sphere diameter of a monomer). The instantaneous density profiles of the polymer segments in state s_i at position $\mathbf{r}$ can be determined from

$$\rho_i(\mathbf{r}, s_i) = \frac{1}{\Delta_{bin}^3} \sum_k^n \theta\left(\frac{\Delta_{bin}}{2} - | x - x'_{k,i} | \right) \times \theta\left(\frac{\Delta_{bin}}{2} - | y - y'_{k,i} | \right) \theta\left(\frac{\Delta_{bin}}{2} - | z - z'_{k,i} | \right) \delta_{s_i s'_{k,i}} \tag{1}$$

where Δ_{bin} is the width of the bin (~0.05σ), $\theta(r)$ is the step function, i.e., it is unity when r>0 and zero otherwise, and $\delta_{s_i s'_{k,i}}$ is the Kronecker delta function. During Monte Carlo simulation, we count all segments i with state s_i that are within the box cell $\mathbf{r} \pm \Delta_{bin}/2$ and divide by the box volume Δ_{bin}^3 to determine the density at position $\mathbf{r}$. The effective single-chain Hamiltonian $H(\mathbf{X})$ is then calculated from the ion distribution determined by the one-dimensional iDFT calculation in which the long-range intrachain contribution is accounted for (hereby referred to as 1D-iDFTm). Note that the ion distribution depends on the configuration of the polymer brushes due to the intermolecular interactions between the ions and the brush segments.

The effective single-chain Hamiltonian of the polymer chain *k* due to its inter- and intramolecular interactions can be decomposed into one- and two-body terms, respectively, as

$$H_k(\mathbf{X}_k) = V_B(\mathbf{R}_k) + \sum_i^M \left[\lambda_{k,i}(\mathbf{x}_{k,i}) + \Delta\mu_i^H(s_{k,i}) \right] + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \left[u_k^\infty(\mathbf{x}_{k,i}, \mathbf{x}_{k,j}) - \ln y_k(\mathbf{x}_{k,i}, \mathbf{x}_{k,j}) \right] \tag{2}$$

where $V_B(\mathbf{R}_k)$ is the bond potential, $\lambda_{k,i}(\mathbf{x}_{k,i})$ is the contribution due to intermolecular interactions with other species in solution and the surface for the segment i in state $s_{k,i}$ at

position $\mathbf{r}_{k,i}$, $\Delta\mu_{k,i}^{H}$ is the chemical contribution due to the protonation/deprotonation of the monomers, u_k^{∞} accounts for the intermolecular interactions between two monomers at infinite dilution, and y_k is the two-body cavity correlation function (CCF) that accounts for the intrachain correlation energy between two monomers. While the bond potential dictates the chain connectivity and ensures a self-avoiding freely-jointed tangent chain conformation, the one- and two-body terms serve as analogs of the fluctuating external fields and the bonded interactions in SCMF simulations of *bare* particle-based Hamiltonians. In Eq.(2), the first summand is composed of various contributions to the local excess chemical potential for segment *i*

$$\lambda_{k,i}\left(\mathbf{r}_{k,i}, s_{k,i}\right) = \mu_i^{mon}\left(\mathbf{r}_{k,i}, s_{k,i}\right) + \mu_{k,i}^{ch}\left(\mathbf{r}_{k,i}, s_{k,i}\right) + V_{k,i}^{ext}\left(\mathbf{r}_{k,i}, s_{k,i}\right) + e\psi\left(\mathbf{r}\right) Z_k\left(s_{k,i}\right). \quad (3)$$

Here, the first term results from the intermolecular interactions of segment *i* at state s_i with the other monomers and ions in the system (i.e., the monomer views the system as a collection of monomers not in a chain as well as ions in the solution); the monomeric contribution typically includes hard-sphere and electrostatic correlations determined by the fundamental measure theory (FMT)[25] and mean-spherical approximation (MSA)[26], respectively. Note that the local correlations (e.g., packing effects of the hard spheres) cannot be captured by the SCMF formalism using the *bare* interactions of the underlying particle-based model because the intermolecular interactions are *not* smooth and slowly varying in space and time. Instead, we utilize iDFT to account for these local correlations. The second term μ_i^{ch} is the excess chemical potential due to chain connectivity. This term describes the free energy due to intrachain correlation by segment *i* in state s_i. The third

term is the non-electrostatic external potential due to the surface which, in this case, it is given by a hard-wall potential. The last term in Eq. (3) is the mean electrostatic potential ψ which accounts for direct coulomb interactions among charged monomers and ions as well as the surface.

In addition, Eq.(2) includes a one-body potential which results from the deprotonation/protonation of ionizable sites:

$$\Delta\mu_{k,i}^{H}(s_{k,i}) = -s_{k,i}k_BT(\mathrm{pK}_{k,i} - \mathrm{pH})\ln 10 \tag{4}$$

where k_B and T are the Boltzmann constant and absolute temperature, respectively. In Eq. (4), $s_{k,i} = 0$ means that the polymer segment is at its neutral state, $s_{k,i} = -1$ stands for the deprotonation of a weak acidic segment, and $s_{k,i} = +1$ for the protonation of a weak basic segment. The equilibrium constant of the deprotonation or protonation reaction, $\mathrm{K}_{k,i}$, depends on the identity of the ionizable site and system temperature but not on the solution composition. Other values of $s_{k,i}$ are also possible if the polymer segment contains multiple ionizable sites.

The second summand in the decomposition of the effective single-chain Hamiltonian $H(\mathbf{X})$ contains the intrachain interactions $u_k^{\infty}(\mathbf{x}_i, \mathbf{x}_j)$ and the correlation effects (i.e., the two-body CCF) $y_k(\mathbf{x}_i, \mathbf{x}_j)$ in configuration **X**. The former term, u_k^{∞}, accounts for the intramolecular interactions between the polymer segments in configuration **X** at infinite dilution. Note that this term is contained within the chemical potential for the polymer in state **X** and that it is present due to an ideal chain reference

system (viz., a fully uncharged hard-sphere chain with no other interactions in conformation **R**). Thus, the chemical potential for a polymer in configuration **X** is given by $\mu_k(\mathbf{X}_k) = \mu_k^{ideal}(\mathbf{R}_k) + \Delta\mu_k^H(\mathbf{S}_k) + \mu_k^\infty(\mathbf{X}_k)$ where $\mu_k^\infty(\mathbf{X}_k) = \sum_{i=1}\sum_{j>i} u_k^\infty(\mathbf{x}_i, \mathbf{x}_j)$. Meanwhile, the direct interaction between polymer segments u_k^∞ can be separated into electrostatic and the short-range non-electrostatic contributions:

$$u_k^\infty\left(\mathbf{x}_{k,i}, \mathbf{x}_{k,j}\right) = \frac{l_B}{r_{i,j}} Z_k\left(s_{k,i}\right) Z_k\left(s_{k,j}\right) + u_k^{sr}\left(\mathbf{x}_{k,i}, \mathbf{x}_{k,j}\right), \tag{5}$$

where $l_B = e^2 / 4\pi\varepsilon_0\varepsilon_r k_B T$ is the Bjerrum length (7.14 Å for liquid water at room temperature), e is the elementary charge; ε_0 and ε_r are the vacuum permittivity and dielectric constant of the solvent, respectively. The first term is the electrostatic pair potential u_k^C for two charges separated at a distance $r_{i,j} = |\mathbf{r}_{k,i} - \mathbf{r}_{k,j}|$, and the second term is the non-electrostatic contribution such as van der Waals interactions or hydrogen bonding. In this work, we consider $u_k^{sr}(\mathbf{x}_{k,i}, \mathbf{x}_{k,j}) = 0$ for simplicity.

While the CCF is typically truncated after the nearest-neighbor interactions as was done in the 1D-iDFT2, SCMF allows us to consider all two-body correlations between the polymer segments. Although an exact formulation of the CCF at inhomogeneous conditions is unknown, we can estimate the CCF by the local density approximation:

$$\begin{aligned} y_k\left(\mathbf{x}_{k,i}, \mathbf{x}_{k,j}\right) \approx & \sqrt{y_k^{MSA}\left[\left\{n_{\alpha,t}\left(\mathbf{r}_{k,i}\right)\right\}, s_{k,i}, s_{k,j}; r_{i,j}\right]} \\ & \times \sqrt{y_k^{MSA}\left[\left\{n_{\alpha,j}\left(\mathbf{r}_{k,j}\right)\right\}, s_{k,i}, s_{k,j}; r_{i,j}\right]} \end{aligned} \tag{6}$$

where y_k^{MSA} accounts for the charge states of segments *i* and *j*, and $\{n_{\alpha,t}\}$ are weighted densities for all monomers and ions as defined in FMT[25]. For a bulk system, the two-body CCF can be evaluated from MSA[27, 28]:

$$\ln y_k^{MSA}\left(\left\{n_{\alpha,t}\right\}, s_{k,i}, s_{k,j}; r_{i,j}\right) = \ln g^{HS}\left(r_{i,j}\right) + \frac{l_B}{r_{i,j}}\left[Z_k\left(s_{k,i}\right)Z_k\left(s_{k,j}\right) - Z_k'\left(s_{k,i}\right)Z_k'\left(s_{k,j}\right)\right]\exp\left[-\kappa\left(r_{i,j}-\sigma_{i,j}\right)\right] \quad (7)$$

where the effective valence is $Z_k'(s_{k,i}) = [Z_k(s_{k,i}) - \eta\sigma_k(s_{k,i})^2]/[1+\Gamma\sigma_k(s_{k,i})]$, $\sigma_k(s_{k,i})$ is the diameter of the segment *i* in its respective state $s_{k,i}$, and $\sigma_{i,j} = [\sigma_k(s_i) + \sigma_k(s_j)]/2$. κ, Γ, and η are, respectively, the Debye screening parameter, MSA screening parameter and a charge asymmetry term defined in terms of weighted densities $n_{\alpha,t}$. In Eq.(7), the first term accounts for the two-body correlation energy due to the hard-sphere excluded-volume effect, and the second-term arises from electrostatic and hard-sphere coupling effects. The hard-sphere term is approximately zero except when the two segments are at contact (i.e., $r_{i,j} = \sigma_{i,j}$).

The self-consistent determination of the effective external field requires a determination of the ion distribution resulting from the intermolecular interactions between one another, the polymer brush segments, and the surface. The grand potential of the system [given by Eq.(S1)] is minimized utilizing Picard iteration to determine the density profiles of the ions which satisfies

$$\rho_t(\mathbf{r}) = \exp\left\{-\beta\left[\mu_t^{mon}(\mathbf{r}) + \mu_t^{ch}(\mathbf{r}) + V_t^{ext}(\mathbf{r}) + e\psi(\mathbf{r})Z_t - \mu_t\right]\right\} \quad (8)$$

where Z_t and μ_t is the valence and chemical potential of the ion *t*. Once the effective external fields are determined from 1D-iDFTm, we run Monte Carlo (MC) simulation in which each polymer segment is, on average, given the opportunity to change its ionization state from charged to uncharged (or vice versa) as well as displace its position. The possible MC moves for changing the conformation and ionization of the polymer are shown on the right-hand side of Figure 1. Additional details on the simulation protocol can be found in the Supporting Information. We accept a change in the polymer's configuration with the Metropolis algorithm, i.e., with a probability $P_{acc} = \min\{\exp[\beta H(\mathbf{X}_{old}) - \beta H(\mathbf{X}_{new})], 1\}$ where $\mathbf{X}_{new}$ and $\mathbf{X}_{old}$ refer to the new and old configuration of the polymer, respectively, and $\beta = 1/k_B T$. With the new configurations of the polymer chains determined from the MC moves, the instantaneous densities of the polymer segments can then be updated for generating a new set of external fields calculated from 1D-iDFTm.

9.3 Results and discussion

We first apply scDFT to a polymer brush of weak polyacids of length *M=25* at two grafting densities ρ_{graft} = 0.10 and 1.00 #/nm^2. The parameters chosen for the size $\sigma = 5.0\,\text{Å}$ and $\text{pK} = 5.0$ are representative of a monomer in poly(acrylic acid). The parameters for the salt ions, proton, and hydroxyl ions are also the same as those used for bulk systems[29]. The appropriate expressions for the excess chemical potential, $\mu_i^{ex,b}$, as well as the self-consistent determination of the concentrations of all species in the bulk solution (viz. salt ions, protons, and hydroxyl ions) and their molecular parameters can be

found in a previous work[30]. The solvent is considered implicitly through a dielectric background (viz. $\varepsilon_r = 78.4$). At high grafting density, the system is considered to be homogeneous in the xy-plane (parallel to surface) and thus the ion distribution varies only in the z-direction (perpendicular to the surface). At low grafting density, the assumption of lateral homogeneity breaks down because the brushes no longer overlap with one another. The polymer segment profiles can be calculated from $\rho_i(z, s_i) = \frac{\rho_{graft}}{n\Delta_{bin}} \sum_k^n \theta\left(\Delta_{bin}/2 - |z - z'_{k,i}|\right) \delta_{s_i s'_{k,i}}$. Note that the density profiles are laterally averaged whereas the polymer conformation are in three-dimensional space; therefore, the long-wavelength fluctuation effects are not considered in the present work. The chain is simulated in the quasi-instantaneous external field until the equilibrium morphology is obtained based on the convergence of the average properties including density profiles, brush thickness, and degree of ionization.

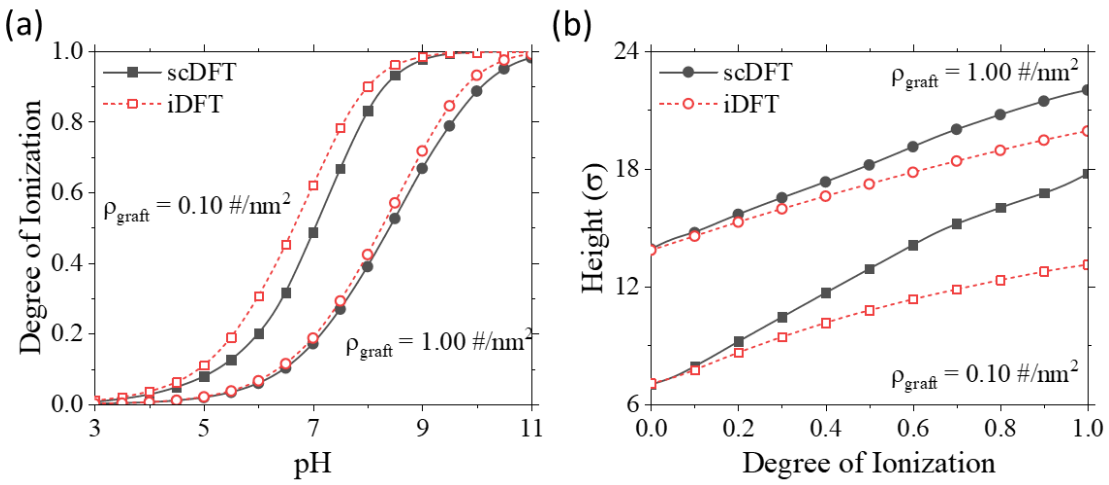

Figure 9.2 (a) Degree of ionization as a function of pH and (b) the brush height as a function of degree of ionization for the model weak polyacid brush predicted by scDFT (solid line) and 1D-iDFT2 (dashed line). The chain length is 25 and the salt concentration is 10 mM. We consider two different grafting densities: 0.10 #/nm^2 and 1.00 #/nm^2.

Figure 2 compares scDFT and 1D-iDFT2 for the degree of ionization and height of a weak polyacid brush. We see that the 1D-iDFT2 calculation predicts the earlier ionization of the polymer than predicted by scDFT since 1D-iDFT2 neglects the long-range interactions that are considered in scDFT. The difference between the two methods is less noticeable when the grafting density is increased from 0.10 #/nm^2 to 1.00 #/nm^2. Since the long-range correlations between the charged monomers are heavily influenced by the local salt concentration, the increase in grafting density of the polymer brush (and therefore charge density in the brush region) drives more counterions to adsorb in this region and these counterions screen the intrachain electrostatic correlation (see Figure S5c). It can therefore be expected that further decrease in the salt concentration in the bulk solution will cause a greater deviation between the 1D-iDFT2 and scDFT predictions for the weak polyacid brush. In addition, we expect a similar situation when the brush grafting density

is decreased below 0.10 #/nm^2. However, our method assumes the ion distribution to be homogeneous in the lateral directions which may not be valid when the brush is sufficiently dilute due to a heightened presence of counterions near the brush and a lower presence in between the brushes. Interestingly, at dilute grafting densities, the density distribution of ions is approximately the same as that in the bulk solution (see Figure S5a) since there is limited intermolecular interactions with the polymer and the surface is a hard wall. In this case, the intramolecular correlations dictate the polymer conformation since the intermolecular correlations are approximately uniform like that in the bulk. Thus, we expect that scDFT can accurately predict the properties of weak polymer brushes across a wide range of solution conditions because we account for both the inter- and intrachain correlations important for determining the polymer conformation and ionization.

Besides the degree of ionization, another point of comparison between the two methods is the height of the weak polyacid brush shown in Figure 2b. As one expects from prior studies on weak polyelectrolyte brushes[13-15, 23, 31], the brush height increases when the grafting density is increased or the polymer gains more charge (i.e., an increase in the pH for a weak polyacid). The scDFT and 1D-iDFT2 predictions are nearly identical at low pH because long-range interactions are not relevant for an uncharged polymer. However, scDFT predicts a stronger dependence of the brush height on the degree of ionization than the 1D-iDFT2 due to the inclusion of long-range intrachain correlations. The difference in brush height between the two methods is less noticeable at higher grafting densities as was the case for the low degree of ionization since intermolecular correlations (viz. hard-sphere correlations and mean electrostatic potential) primarily dictate the

polymer conformation. The 1D-iDFT2 and other polymer mean-field theories like the molecular theory[32] all overestimate the ionization of the polymer and underestimate the expansion of the polymer by neglecting the long-range intramolecular interactions between the charged monomers in the polymer chain. scDFT marks a substantial improvement over existing methods by taking advantage of the single-chain simulation and coupling with the iDFT to treat the multibody state and conformation on equal standing.

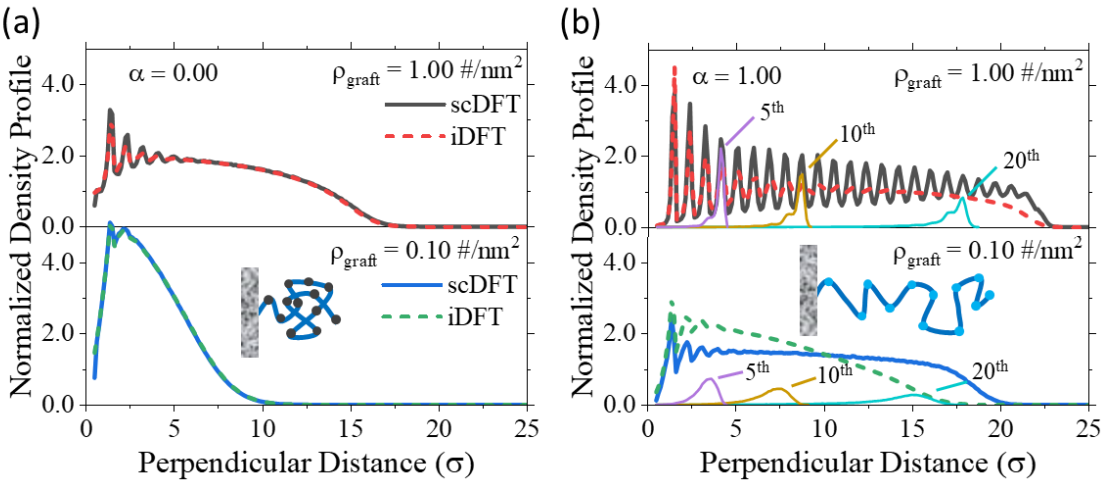

Figure 9.3 The normalized density profile (defined as $\rho(z)/\rho_{graft}$) for a polyacid brush when the polymer is (a) fully uncharged and (b) fully charged. The predictions by scDFT and 1D-iDFT2 are given by the solid and dashed lines, respectively, at two different grafting densities: 0.10 #/nm^2 and 1.00 #/nm^2. The chain length is 25 and the salt concentration is 10 mM. Panel (b) also includes the density profile for segment numbers 5, 10, and 20 as predicted by scDFT.

Figure 3 shows the density profiles of the weak polyacid brush at a grafting density of 0.10 #/nm^2 and 1.00 #/nm^2 when the brush is either neutral or fully charged. As discussed earlier, the density profiles for the uncharged polymer predicted by scDFT and 1D-iDFT2 are approximately the same. The importance of intrachain excluded volume included by scDFT is effectively screened by the interchain excluded volume due to other

tethered chains and ions. Nonetheless, the density profile determined by scDFT is inherently different from that determined by the 1D-iDFT2 calculation since intrachain excluded volume is explicitly considered and therefore the difference in the distribution of segments, although relatively minor in this case, is present. The 1D-iDFT2 calculation, like other mean-field theories[32], becomes inaccurate at low grafting densities due to the inability to account for intrachain excluded volume. At high grafting density, the hard-sphere packing results in a strong oscillation in the density of segments perpendicular to the surface as seen previously in 1D-iDFT2 calculations[33]. This behavior is not well captured by alternative theories because they utilize an incompressibility condition[32]. The difference in scDFT and 1D-iDFT2 is more noticeable for the charged polymer case. In general, scDFT predicts the polymer segments are distributed further away from the surface than that predicted by 1D-iDFT2 since the intrachain electrostatic interactions will result in stretched conformations being favored. At high grafting density, both methods exhibit almost discrete peaks for the segment profile, which indicates that the brush is significantly stretched (i.e., rod-like). Further evidence for the stretched conformation of the brush can be seen by the segment profile for segment numbers 5, 10, and 20 in Figure 3b which show a narrow region in which the segment is present. In contrast to the segment profile for the high grafting density, the segment profile at the low grafting density is broad since the segments are not localized to a specific region. A benefit of scDFT is that directly sampling the polymer conformation in the external field allows us to capture the rod-like behavior; such behavior would be difficult to quantify using a method that generates the conformations beforehand like in the molecular theory[16]. In the case of scDFT, the

extended structure of the polymer persists further than that predicted by the 1D-iDFT2 calculations because long-range intrachain interactions favor the more rod-like conformation.

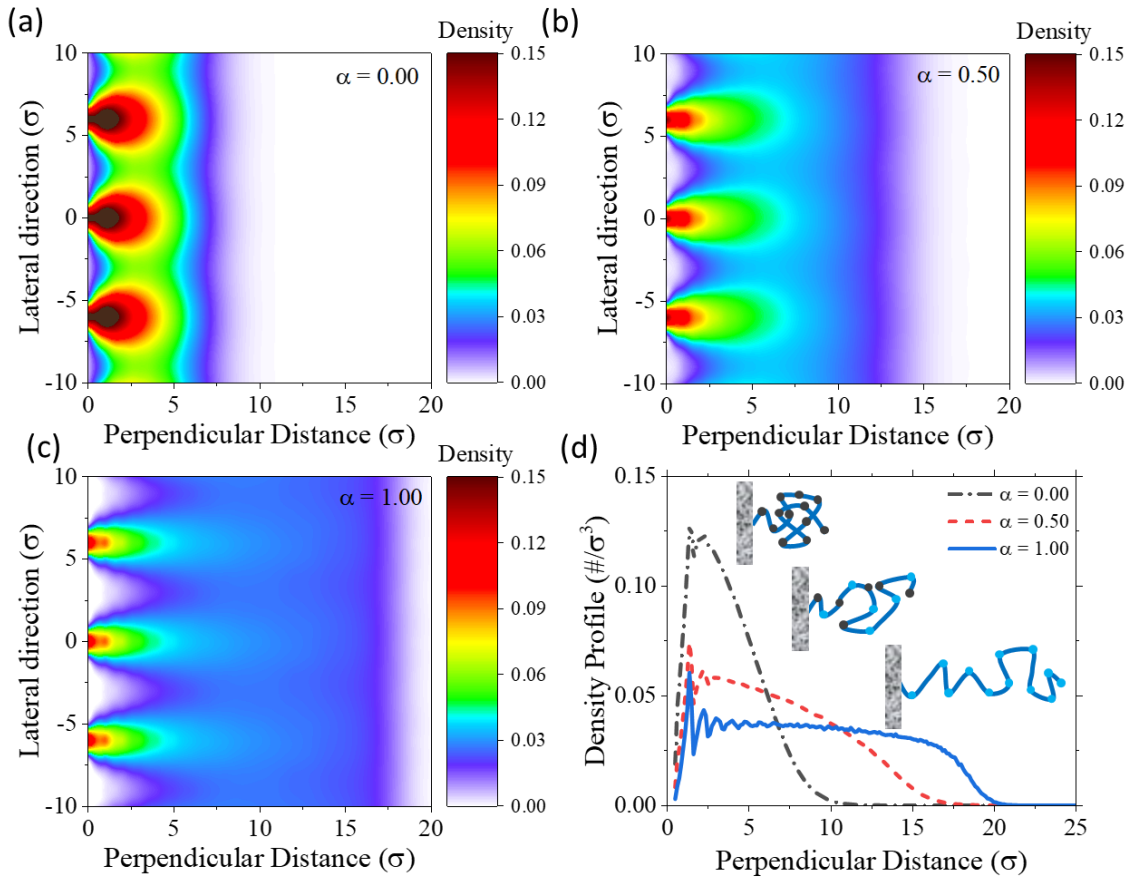

Figure 9.4 Two-dimensional view of the weak polyacid brush along the surface at three different degrees of ionization: (a) 0.00, (b) 0.50, and (c) 1.00. The chain length is 25 and the salt concentration is 10 mM. The grafting density of the brush is 0.10 #/nm^2. (d) The density profile in the perpendicular direction from the surface at the three respective degrees of ionization.

In comparison to the one-dimensional implementation of iDFT, one major advantage of scDFT is that it accounts for the lateral correlations in the polymer conformation resulting from the intrachain correlations. In Figure 4, we show a two-

dimensional view of the density profiles of polymer segments along the surface when the polyacid brush is uncharged, half-charged, and fully charged at an approximate grafting density of 0.10 #/nm^2 (viz. ~6 units spacing between adjacent tethered points on a lattice surface). As mentioned previously, the lateral inhomogeneity in the distribution of ions should be accounted for when neighboring brushes do not overlap. We see that, at ρ_{graft} = 0.10 #/nm^2, there is still significant overlap between the brushes when the chain length is 25. Further decrease in the grafting density would result in an inhomogeneous distribution of the ions between the surfaces, which would affect the polymer conformation and ionization behavior of the brush. When the polymer is fully uncharged (Figure 4a), the brush is collapsed and spreads across the surface. There is significant overlap between neighboring brushes as evident by the increased density between the two grafting points. It should be noted that the brush segments from one brush are allowed to overlap with brush segments from a different brush since we consider the chains as independent and interact only through the self-consistently determined single-chain Hamiltonian. At low grafting densities, the spatial arrangement of other brushes could be relevant to the conformation and ionization of a brush, but in general, the intermolecular correlations will reasonably account for this behavior.

Figure 4b shows the conformation behavior of the polymer when the brush is half-charged. In this case, the intermolecular correlations are mostly due to Coulomb interactions as indicated by the brush extending much further from the surface than it does in the uncharged case. The presence of charged segments within the brush results in electrostatic repulsion within the same polymer and with segments in other tethered chains.

As a result, the brush swells because the polymer chains extend in the direction perpendicular from the surface. For a neutral brush, the polymer density decays sharply as it extends further from the surface since the repulsive forces is not strong enough to stretch the polymer chains. On the other hand, when the polymer is fully ionized, the brush is stretched to about 75% of its contour length. In this case, the charge-charge interactions between intrachain segments result in the extended conformation of the polymer. Due to the strong repulsion from other grafted chains, the tethered polymer is mostly extended perpendicular to the surface with very low density of segments between neighboring chains, particularly within the first 5 units from the surface (Figure 4c).

Lastly, Figure 4d shows the density profiles of the polymer segments in the perpendicular direction from the surface. As expected from the prior analysis, the brush transitions from a collapsed state to a stretched state as it goes from uncharged to fully charged. In the fully charged case, the oscillatory density profile in the inner region of the brush indicates that the first few segments of the brush are mostly perpendicular to the surface. This conformation follows from the lack of segments between the two brushes as seen through Figure 4c. Such insight into the conformational behavior of the polymer cannot be well described through 1D-iDFT2 calculations since the conformation is dictated not only by intermolecular interactions, but also intrachain interactions as described by scDFT.

9.4 Conclusion

In summary, scDFT provides a new theoretical procedure to account for the ionization-conformation coupling in weak polyelectrolyte systems that cannot be captured with conventional methods. It is numerically efficient compared to explicit particle-based Monte Carlo simulation by approximating the inter- and intramolecular interactions through an effective single-chain Hamiltonian determined via the Ising density functional theory. As a result, it can be easily parallelized for GPU computing allowing for quick determination of realistic polymeric systems. In addition, the inclusion of intrachain correlations beyond the nearest-neighbor level is important for understanding the charge regulation and conformation behavior in hetero-weak polyelectrolytes such as zwitterionic polymers and polypeptides. While the discussion above is focused on a coarse-grained model, similar procedures can be established for other weak-polyelectrolyte models including those with atomistic details by incorporating the Ising degree of freedom to account for the charge states of ionizable sites.

Acknowledgements: This work is financially supported by the NSF-DFG Lead Agency Activity in Chemistry and Transport in Confined Spaces under Grant No. NSF 2234013 and Mu1674/18. Additional support is provided by the NSF Graduate Research Fellowship under Grant No. DGE-1326120.

Bibliography

1. Yap, H.P., et al., *Colloid Surface Engineering via Deposition of Multilayered Thin Films from Polyelectrolyte Blend Solutions.* Langmuir, 2005. **21**(10): p. 4328-4333.

2. Yuan, W., et al., *Weak polyelectrolyte-based multilayers via layer-by-layer assembly: Approaches, properties, and applications.* Advances in Colloid and Interface Science, 2020. **282**: p. 102200.

3. Tong, W., C. Gao, and H. Möhwald, *Stable Weak Polyelectrolyte Microcapsules with pH-Responsive Permeability.* Macromolecules, 2006. **39**(1): p. 335-340.

4. Ilyas, S., et al., *Weak polyelectrolyte multilayers as tunable membranes for solvent resistant nanofiltration.* Journal of Membrane Science, 2016. **514**: p. 322-331.

5. Muthukumar, M., *50th Anniversary Perspective: A Perspective on Polyelectrolyte Solutions.* Macromolecules, 2017. **50**(24): p. 9528-9560.

6. Boroudjerdi, H., et al., *Statics and dynamics of strongly charged soft matter.* Physics reports, 2005. **416**(3-4): p. 129-199.

7. Gonzalez Solveyra, E., et al., *Theoretical modeling of chemical equilibrium in weak polyelectrolyte layers on curved nanosystems.* Polymers, 2020. **12**(10): p. 2282.

8. Landsgesell, J., et al., *Simulations of ionization equilibria in weak polyelectrolyte solutions and gels.* Soft Matter, 2019. **15**(6): p. 1155-1185.

9. Biesheuvel, P.M., M. van der Veen, and W. Norde, *A Modified Poisson−Boltzmann Model Including Charge Regulation for the Adsorption of Ionizable Polyelectrolytes to Charged Interfaces, Applied to Lysozyme Adsorption on Silica.* The Journal of Physical Chemistry B, 2005. **109**(9): p. 4172-4180.

10. Zhulina, E.B. and O.V. Borisov, *Poisson–Boltzmann Theory of pH-Sensitive (Annealing) Polyelectrolyte Brush.* Langmuir, 2011. **27**(17): p. 10615-10633.

11. Zhulina, E.B., T.M. Birshtein, and O.V. Borisov, *Theory of Ionizable Polymer Brushes.* Macromolecules, 1995. **28**(5): p. 1491-1499.

12. Israels, R., et al., *Charged Polymeric Brushes: Structure and Scaling Relations.* Macromolecules, 1994. **27**(12): p. 3249-3261.

13. Lyatskaya, Y.V., et al., *Analytical self-consistent-field model of weak polyacid brushes.* Macromolecules, 1995. **28**(10): p. 3562-3569.

14. Willott, J.D., et al., *Behavior of Weak Polyelectrolyte Brushes in Mixed Salt Solutions.* Macromolecules, 2018. **51**(3): p. 1198-1206.

15. Witte, K.N., S. Kim, and Y.-Y. Won, *Self-Consistent Field Theory Study of the Effect of Grafting Density on the Height of a Weak Polyelectrolyte Brush.* The Journal of Physical Chemistry B, 2009. **113**(32): p. 11076-11084.

16. Gong, P., J. Genzer, and I. Szleifer, *Phase Behavior and Charge Regulation of Weak Polyelectrolyte Grafted Layers.* Physical Review Letters, 2007. **98**(1): p. 018302.

17. Gong, P., et al., *Behavior of Surface-Anchored Poly(acrylic acid) Brushes with Grafting Density Gradients on Solid Substrates: 2. Theory.* Macromolecules, 2007. **40**(24): p. 8765-8773.

18. Tagliazucchi, M., et al., *Self-Organized Polyelectrolyte End-Grafted Layers Under Nanoconfinement.* ACS Nano, 2014. **8**(10): p. 9998-10008.

19. Blanco, P.M., et al., *Influence of macromolecular crowding on the charge regulation of intrinsically disordered proteins.* Soft Matter, 2021. **17**(3): p. 655-669.

20. Settanni, G., et al., *pH-Dependent Behavior of Ionizable Cationic Lipids in mRNA-Carrying Lipoplexes Investigated by Molecular Dynamics Simulations.* Macromolecular Rapid Communications, 2022. **43**(12): p. 2100683.

21. Staňo, R., et al., *Electrostatically Cross-Linked Reversible Gels—Effects of pH and Ionic Strength.* Macromolecules, 2021. **54**(10): p. 4769-4781.

22. Gallegos, A., G.M.C. Ong, and J. Wu, *Ising density functional theory for weak polyelectrolytes with strong coupling of ionization and intrachain correlations.* The Journal of Chemical Physics, 2021. **155**(24): p. 241102.

23. Léonforte, F., U. Welling, and M. Müller, *Single-chain-in-mean-field simulations of weak polyelectrolyte brushes.* The Journal of Chemical Physics, 2016. **145**(22): p. 224902.

24. Daoulas, K.C. and M. Müller, *Single chain in mean field simulations: Quasi-instantaneous field approximation and quantitative comparison with Monte Carlo simulations.* The Journal of Chemical Physics, 2006. **125**(18): p. 184904.

25. Yu, Y.-X. and J. Wu, *Structures of hard-sphere fluids from a modified fundamental-measure theory.* The Journal of Chemical Physics, 2002. **117**(22): p. 10156-10164.

26. Gillespie, D., M. Valiskó, and D. Boda, *Density functional theory of the electrical double layer: the RFD functional.* Journal of Physics: Condensed Matter, 2005. **17**(42): p. 6609-6626.

27. Jiang, J., et al., *A molecular-thermodynamic model for polyelectrolyte solutions.* The Journal of chemical physics, 1998. **108**(2): p. 780-784.

28. Henderson, D. and W.R. Smith, *Exact analytical formulas for the distribution functions of charged hard spheres in the mean spherical approximation.* Journal of Statistical Physics, 1978. **19**(2): p. 191-200.

29. Gallegos, A., G.M.C. Ong, and J. Wu, *Thermodynamic non-ideality in charge regulation of weak polyelectrolytes.* Soft Matter, 2021.

30. Gallegos, A. and J. Wu, *Charge Regulation of Natural Amino Acids in Aqueous Solutions.* Journal of Chemical & Engineering Data, 2020. **65**(12): p. 5630-5642.

31. Willott, J.D., et al., *Critical Salt Effects in the Swelling Behavior of a Weak Polybasic Brush.* Langmuir, 2014. **30**(7): p. 1827-1836.

32. Gong, P., et al., *Behavior of surface-anchored poly (acrylic acid) brushes with grafting density gradients on solid substrates: 2. Theory.* Macromolecules, 2007. **40**(24): p. 8765-8773.

33. Jiang, T., Z. Li, and J. Wu, *Structure and Swelling of Grafted Polyelectrolytes: Predictions from a Nonlocal Density Functional Theory.* Macromolecules, 2007. **40**(2): p. 334-343.

Chapter 10. Conclusions and Outlook

In this dissertation, we have developed coarse-grained models and theoretical tools to characterize the charge behaviors of monomers, polymers, and surfaces. We demonstrate that thermodynamic non-ideality resulting from the interactions between the ionizable species and the rest of the solution can be quantitatively captured using molecular thermodynamic models. As a result, we are able to reproduce experimental data using physically realistic models and molecular parameters. New theoretical tools developed in this thesis, such as the Ising density functional theory (iDFT), bridge the gap between existing methods and open up new possibilities to describe diverse ionizable systems including weak polyelectrolyte. In particular, the combination of iDFT and the single-chain-in-mean-field algorithm represents the first molecular theory for weak polyelectrolytes in non-uniform fluids that incorporates long-range intrachain correlations. Lastly, we demonstrate that these new models and tools can be utilized to describe the adsorption and interfacial phenomena of ionizable systems based on comparison with experimental data.

Following the introduction and methodology chapters (Chapter 1 and Chapter 2), Chapter 3 and 4 discuss the development of a coarse-grained model for amino acids in bulk solutions and near an interface using classical density functional theory. The key conclusions are:

1. Our coarse-grained model accurately predicts both the charge regulation and activity coefficients of amino acids in bulk solutions;

2. We can predict the thermodynamic equilibrium constants for all amino acids by extrapolation to the ideal limit;

3. Adsorption of amino acids to surfaces is governed by the coupling between electrostatic interactions and short-range attractions to the surface;

4. The dominant charge state of an amino acid is often different from that in the bulk, particularly when the surface is highly charged.

Chapter 5 incorporates a molecular thermodynamic model into the site-binding model for weak polyelectrolytes. The key findings are:

1. Thermodynamic non-ideality resulting from the inter- and intramolecular interactions can be captured through our molecular thermodynamic model;
2. Nearest-neighbor interactions are sufficient at moderate to high salt concentration; however, the model performs poorly at low to dilute salt concentrations as the ionization is dominated by long-range intrachain correlations;
3. The type of salt counterion plays a key role in determining the ionization behavior of the weak polymer; particularly at low salt concentrations.

Chapter 6 outlines the new and state of art approach to describing weak polyelectrolytes. Ising density functional theory bridges the gap between polymer density functional theory and the site-binding model. Key results include:

1. Ising DFT accounts for the polymer conformation and the polymer charge on an equal footing (i.e., the state and position of all monomer segments in the polymer chain);

2. The incorporation of intrachain correlations is an improvement over conventional methods that consider only the intermolecular interactions since the ionization state of other monomers in the chain influence the likelihood of a monomer to ionize.

Chapter 7 demonstrates an application of iDFT by considering the oligopeptide adsorption to a titanium surface. The key findings are:

1. Polypeptide adsorption can be semi-quantitatively captured using our coarse-grained model and iDFT;
2. iDFT captures the importance of nearest-neighbor interactions on the ionization of segments due to the intrachain attraction or repulsion with opposite- or like-charged residues, respectively;
3. The coupling between surface charge and the ionization of the polypeptide results in unique behavior different from when they are immersed in only a salt solution.

Chapter 8 utilizes a variational principle to incorporate long-range correlations into the nearest-neighbor site-binding model in which we use classical density functional theory to describe the inter- and intramolecular interactions. We find the following conclusions:

1. The conformation and ionization behavior of weak polymers is dictated by the intrachain correlations in the uniform (i.e., bulk) fluid, particularly as the salt concentration is decreased;
2. The thermodynamic properties of weak polymers are sensitive to the solution conditions (viz. salt type, salt concentration, and pH);

3. Variational principle provides a convenient, yet accurate, description of weak polyelectrolytes in bulk solutions.

Chapter 9 presents a state-of-the-art approach to weak polyelectrolytes in uniform fluids by combining iDFT with the single-chain-in-mean-field algorithm. The key conclusions are:

1. The inclusion of long-range intrachain correlations into an inhomogeneous theory provides new insights into weak polymers near interfaces;
2. The new method allows for consideration of weak polyelectrolytes in any solution conditions; i.e., it is no longer restricted by the nearest-neighbor framework used in the original application of iDFT;
3. It captures the coupling of ionization and conformation in the bulk and near the interface due to the external potential.

In summary, this dissertation demonstrates how the charge regulation of monomers, polymers, and surfaces can be well characterized using the developed coarse-grained models and new theoretical methods. The accomplishment from this dissertation provides valuable computational infrastructure for the molecular modeling of diverse ionizable systems. The new theoretical methods should allow for unique insights into these systems that include long-range correlations between monomers in weak polyelectrolyte chains.

There is still plenty of research work that needs to be done to fully understand the charge regulation of monomers, polymers, and surfaces. On the monomeric side, a more comprehensive model that accounts for the dipolar nature of neutral amino acids would be invaluable. In particular, a more faithful description of the short-range interaction of

amnio acids with salt ions in a non-uniform fluid would provide unique insight into the interfacial phenomena. On the polymer side, the framework of iDFT coupled with the single-chain-in-mean-field algorithm allows for efficient and accurate description of weak polyelectrolytes. The development of more realistic models as well as more accurate coarse-grained potentials will be extremely valuable to future applications that utilize ionizable polymers. Lastly, throughout this work the solid surfaces are treated at the mean-field level, despite obvious simplifications that are involved. Similar to the conventional polymer mean-field theories, the use of a mean-field description has been sufficient; yet clearly important physics is missed by washing the charge over the entire surface instead of on specific sites. Research along these directions will continue to pave the path towards a comprehensive understanding of charge regulation from small molecules to polymers and surfaces.