2010

Modeling Hydrophobic Effects at different lengthscales

Jihang Wang
Virginia Commonwealth University

Modeling Hydrophobic Effects at different lengthscales

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Jihang Wang
B. Sc, Nanjing University, China, 2003

Director: Dr. Alenka Luzar
Professor, Deparment of Chemistry

Virginia Commonwealth University
Richmond, Virginia
December, 2010

Acknowledgment

This work is dedicated to my advisor, Professor Alenka Luzar, nothing would be possible without her kind support and mentorship. I would also like to give my special thanks to Professor Dusan Bratko, for his help and support during writing and preparing all the materials for this dissertation, as well as the everyday work I have been doing over the past five and a half years; and Dr. Chris Daub, my former lab mate, who gave me tremendous help in different research projects. I cannot forget to thank my former research advisor Dr. Darius Kuciaskas and former lab mates for teaching me basic skills about scientific research. In addition, Virginia Commonwealth University and other agencies (National Science Foundation and Altria Group) are gratefully appreciated for their generous financial support throughout my long graduate school career. The supports from my family and dear friends are very important for me as well. Last but not least, I would like to thank my committee members and all the readers for taking your time to read into this dissertation, I hope I have made my little piece of contribution to the science community.

Table of Contents

List of Figures

List of Tables

Abstract

MODELING HYDROPHOBIC EFFECTS AT DIFFERENT LENTHSCALES

By Jihang Wang, Ph.D

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2010.

Major Director: Alenka Luzar
Ph.D, Professor, Department of Chemistry

Understanding hydrophobic effects at different length scales is relevant to many complex and poorly understood everyday phenomena in materials science and biology. In this thesis, a variety of theory/computational methods in statistical physics and statistical mechanics are used to address three separated, but interconnected problems: (1) How solvation free energy scales with a partical size that is charged? This problem has never been attempted to solve before despite its immense importance in colloidal and protein solutions (J. Wang, D. Bratko, K. Leung and A. Luzar, Hydrophobic hydration at different length-scales: manipulating the crossover by charges, to be submitted to *J. Stat. Phys*. (Special issue on Water and Associated Liquids)); (2) Can onset to capillary evaporation, seen in some protein complexes with large hydrophobic areas be predicted in a simple way? A simple coarse-grained model of water/protein system, which is developed here shows the conditions for wet

and dry hydrophobic protein cavities, and is able to reproduce all atom simulation results. The method should serve as an intermediate step between the initial screening of protein hydrophobic cavities and expensive molecular simulations (J. Wang, S. Kudesia, and A. Luzar, Computational probe of dewetting events in protein systems, in preparation for submission to *J. Phys. Chem. B)*; (3) How to predicts hydrophobicity of a mixed surface from the knowledge of its pure constituents? To this end, wetting free energy on synthetic and biological heterogeneous surfaces is studied. Two distinct mechanisms responsible for their non-additivity have been identified in each case (J. Wang, D. Bratko and A. Luzar, Probing surface tension additivity on heterogenous surfaces: a molecular approach, *Proc. Natl. Acad. Sci*, under review).

Chapter 1. Introduction

Hydrophobic effect pertains to hydrophobic hydration and hydrophobic interaction. Scheraga and coworkers[1-2] made a useful distinction between the two phenomena: hydrophobic hydration refers to the structure of water in the immediate vicinity of the apolar solute molecules, and the thermodynamic properties of very dilute solutions. Hydrophobic interaction refers to the solvent induced interaction between two or more apolar solute molecules. It implies an association of solute molecules by solvent mediated force at finite concentrations. The free energy change, *W(R)*, of bringing the two apolar solutes together from infinity to a separation *R*, termed the potential of mean force (PMF), is related to the radial distribution function of the two apolar solutes, $g_{AA}(R)$, through the relation, $W(R) = -k_B T \ln g_{AA}(R)$, where k_B is the Boltzmann constant and T is the absolute temperature.

Because of its importance in everyday life, the hydrophobic effect has been well documented in many review papers[3-10]. However, a more general problem initially proposed by Stillinger[11] of different lengthscales and effects associated with hydrophobicity has not attracted much attention until the last decade[12-15]. Hydrophobic hydration and interaction both depend on length scale, but in different ways.

In the case of hydrophobic hydration, water surrounding an apolar solute acts differently due to the different size of the solute. For a small apolar solute in aqueous environment, the hydrogen bond network is not perturbed significantly, and water can

manage to form hydrogen bonds around the solute. Each water molecule is still able to participate in up to four hydrogen bonds with its neighbors (Fig. 1a), same as that in bulk environment. However, if the apolar solute is sufficiently large, the water molecules near its surface can no longer maintain the complete hydrogen bond network, and water tends to sacrifice one hydrogen bond per molecule on average to minimize the energy loss[11, 16-19] (Fig. 1b). There is a crossover of the solvation free energy from small apolar solute, where it is entropy dominated, to large apolar solute, where it is enthalpy dominated.[6-7, 20]

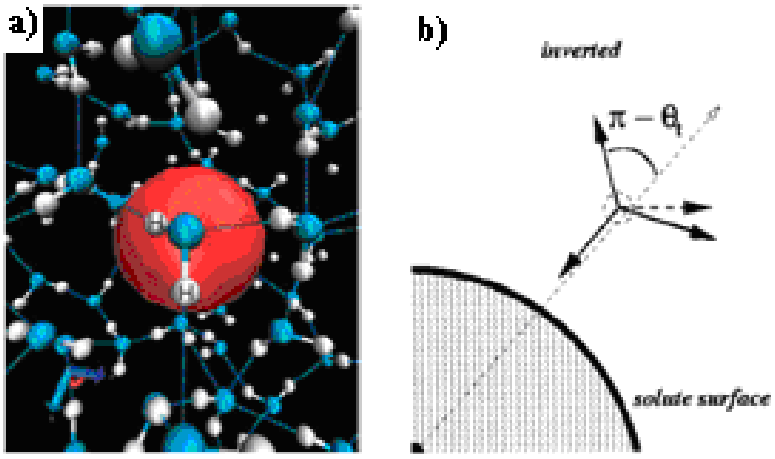


Figure 1. Hydrogen-bonding networks near small solute (a) (Ref 6), water molecule can make up to four hydrogen bonds to the neighbors. The blue and white particles represent the oxygen (O) and hydrogen (H) atoms of water molecules, respectively, red particle is the methane-like hydrophobic solute. (b) Water sacrifices one hydrogen bond per molecule near the surface of large solute. (Ref 21)

1.1 Crossover lengthscale studies of hydrophobic hydration

Recent theoretical[15, 20] and computer simulation studies[7, 22] have revealed that for a small hydrophobic solute in water, the solvation free energy scales with the solute volume, and for a large hydrophobic solute, it scales with the solute surface area. For neutral apolar solutes, the crossover is predicted to be at the sub-nanometer scale,[6-7, 15, 20] which is the correlation length of the hydrogen bonds in water. I will introduce our own work about the crossover lengthscale in hydrophobic hydration and how electrostatic interactions may affect the crossover in current section. The detailed results and discussions will be presented in Chapter 3.

With a simple mean field (MF) model, Luzar *et al*[16-17] successfully predicted the 25% reduction of water hydrogen bonds next to a planar hydrophobic surface by pure geometric restrictions, in good agreement with molecular simulations[18] and Second Harmonic Generation (SHG) experiments[19]. The same MF model[16-17] also predicted the hydrogen bonds' contribution on the solvation free energy density of a planar hydrophobic surface. To seek the physical understanding of the predicted crossover[6] in hydrophobic hydration, we extended the model to spherical surfaces to predict the hydrogen bonds reduction near surfaces with different curvatures, and calculated the contribution from hydrogen bonds to the solvation free energy of spherical hydrophobic solutes. Based on the same idea of pure geometric restrictions in hydrogen bonding, we further developed a generalized model for hydrogen bonds reduction near surfaces of different geometries. In addition to the hydrogen bonding effect, we adapted the methods from Stewart and Evans[23] to calculate the solvation

free energy from Van der Waals interactions between solvation water and hydrophobic solutes. By adding these two parts together, we obtained the mean field prediction of solvation free energy as a function of solute radii, in good agreement with all atom simulations.[7] Our approach uses an efficient mean field model to catch the basic physics of the problem: Crossover happens when the range of interactions becomes small compared to the size of the solutes. The crossover lengthscale depends on the perturbation range of water at solute surfaces.

So far, only the crossover lengthscale for the solvation free energy of neutral solutes have been considered in literature, however, water behavior and the crossover lengthscale of solvation free energy could be different when the solute is charged. By putting charges on the solute, the system will be more interesting as a model for biological scientists. In aqueous solution of neutral solutes, the relevant length-scale is the hydrogen bond correlation length, around 0.5 nm. In the case of aqueous solution of charged solute, water molecules and counter-ions around the solute screen the charges on solute based on Debye-Hückel theory and the relevant lengthscale could be related to the "Debye screening length"[24], which is determined by ionic strength in the system. The screening effects may push the crossover to a larger length scale. (Fig. 2) We should note that when we charge a hydrophobic solute, usually the hydration process becomes favorable (solvation free energy becomes negative), but there should still be a crossover in this context.

Our studies about the crossover lengthscales involving electrostatic interactions would have an important consequence for biological modeling community: In

modeling of biomolecular systems, the solvation free energy is usually calculated by $\Delta G = \gamma \cdot A$,[25-27] where γ is the surface tension and A is the surface area. The solvation free energy scales with the surface area in this relation. The size of a typical biological system is usually above the nanometer scale, which is always larger than the predicted crossover for neutral solutes. Therefore, the solvation free energy should always scale with the surface area, and the current treatment is adequate. However, most biomolecules have charges/partial charges on their atoms. The electrostatic interactions could possibly change the crossover to a different lengthscale (e.g. screening length is ~1 μm in pure water, and ~1 nm in physiological solution).

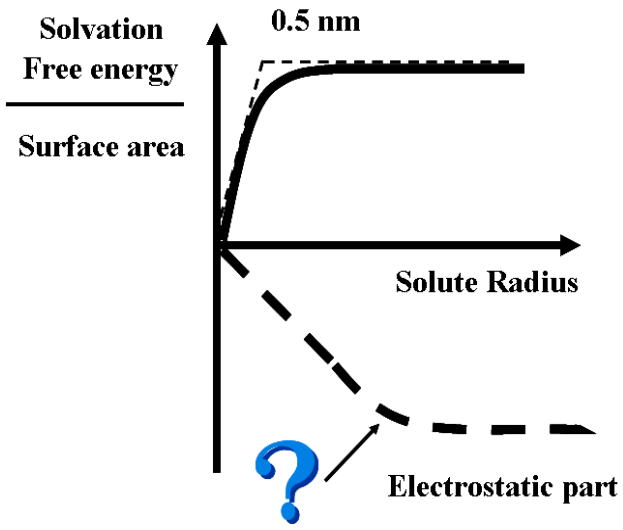


Figure 2. Solvation free energy (normalized by the surface area) crossover of neutral solutes (solid line) and our proposed crossover of charged solutes (dashed line).

If that is the case, proteins could often fall in the regime where solvation free energy scales with volume. For example, recent studies from Baker's group[28] suggested that the term scaling with volume (pV, where p is the solvent pressure) in solvation free energy should be taken into account in the calculations of solvation forces with implicit solvent models in biological systems.

In our studies, we use molecular dynamics (MD) simulations to calculate the solvation free energies of charged solutes with different radii but with the same charge densities on the surface. We vary the Debye screening length[24] by adding different amount of mono-charged cations and anions into the solvation systems. However, due to the lack of screening charges in our reference state, the electrostatic part of the solvation free energy calculated in our model depends only weakly on the screening length. Furthermore, at normal condition of charge density and salt concentrations, the crossover is still in the same order of magnitude as that predicted in non-polar systems, just slightly over the 1 nm mark. Our findings suggest that it should be reasonable approximation to use the relation $\Delta G = \gamma \cdot A$ in biological systems at large lengthscales.

1.2 Computational probe of dewetting events in protein systems

In the case of hydrophobic interaction between small apolar solutes, solute-water distribution functions show apolar solutes separated by a layer of water (i. e. PMF solvent separated minimum is larger compared to solvent-contact minimum)[29]. In the case of water confined between large apolar solutes (large here means larger than the correlation length of the solvent), i.e. extended hydrophobic surfaces, the modified Kelvin's Equation[30],

$$D_C \sim 2\Delta\gamma / (\rho\Delta\mu + b\gamma / L), \qquad (1\text{-}1)$$

predicts the critical distance D_c, where vapor state becomes more stable than liquid state. In the equation, ρ is the number density, $\Delta\mu$ is the difference in chemical potential of water between the liquid and vapor phases, γ is liquid/vapor surface tension, L is the lateral size of the surface and b is a geometrical parameter (of order 1). Young's equation $\Delta\gamma \equiv \gamma_{wl} - \gamma_{wv} = -\gamma\cos\theta_C$ gives $\Delta\gamma$, where γ_{wl} and γ_{wv} denote wall/liquid and wall/vapor surface tension, respectively, and θ_c is the contact angle on the apolar confinement surface.

Hydrophobic confinement effects on the phase behavior of water between extended hydrophobic surfaces have been studied extensively in simulations and theory.[8, 30-42] If the separation between two nanoscale hydrophobic walls is smaller than the critical distance D_c, the water confined between the walls will favor the vapor state, and may experience a phase transition from liquid state to vapor state (so-called cavitation or capillary evaporation).[30, 43-45]

Discussions of confinement-induced evaporation require one to make an important distinction between different criteria used to characterize hydrophobicity.[33] For cavitations in any system to take place, the contact angle of water on the surfaces must be above 90°. When only short-ranged interactions are present, this requirement is equivalent to having the ratio of the surface/water and water/water attraction, below 0.5.[33] In literature, a less stringent or weaker condition of hydrophobicity for surfaces to attract water less than water attracts itself is often invoked.[46-47] This condition suffices to promote hydrophobic association but not for capillary evaporation. Recently, the capillary evaporation events attract more and more attentions due to their increasingly important roles in biological systems, for example, during the last stage of protein folding, the protein-ligand binding, protein aggregation, as well as biological self-assembling processes.[9, 48-50] Note that it is very hard to identify a biological surface to fulfill the first criterion of hydrophobicity due to the nature of biological systems, although there are an abundance of surfaces that fulfill the second. In current section, I introduce a coarse graining approach we developed as a computational probe of dewetting (capillary evaporation) events in biological systems (focusing on the confinements in proteins). The complete model along with the results and discussions will be presented in Chapter 4.

Water evaporation from hydrophobic confinements were successfully predicted by atomistic simulations,[8, 32, 41-42, 51-52] as well as implicit models and coarse grained models of solvent,[30-31, 34, 38-40] consistent with the macroscopic picture. Water depletion on nano-scale hydrophobic surfaces has also been directly observed in

Atomic force microscopy (AFM) in the tapping mode (AFM).[53-59] The biological systems such as proteins are far more complicated than model hydrophobic solutes, although there are several attempts to study water behavior in the confinements of biological complex,[60-61] discussing the last few layers of water getting out of the confinement in the final stage of protein folding by expulsion versus spontaneous evaporation, the details about the water evaporation events in biological systems were not well understood.

Berne and coworkers made a series of studies with all atom computer simulations using empirical force fields to investigate water mediated hydrophobic collapse of proteins[36, 51-52, 62]. They first reported direct observation of the capillary evaporation of water in the confinement of melittin tetramer[52], but no evaporation in some other protein systems with large hydrophobic patches consisting of hydrophobic amino acid residues (e.g. the BphC enzyme)[61]. The evaporation transition is attributed to both the presence of large connected areas of surface hydrophobic amino acid residues and the confined geometry in the melittin tetramer system. Employing a simple scoring function based on the distributions of hydrophobic residues on the surfaces forming the confinements,[51] the same research group screened over 400 dimers, tetramers and multi-domain proteins and conducted atomistic simulations on the 50 top candidates to investigate the nanoscale capillary evaporation events in these protein complex, and reported direct observations of evaporation transition in only six protein systems,[8, 51-52] but not in most other candidate systems even though higher hydrophobic scores were achieved from their screening function.[51]

The capillary evaporation events in protein-ligand binding processes have been studied with a similar criteria:[62] more than 1,000 protein-ligand systems were screened and 15 top candidates meeting the evaporation standards were chosen for all atom simulations, again only 6 out of the 15 top candidates exhibited evaporation transitions when putting water molecules in places of ligands in those complex, including the bovine β-lacto globulin, reported earlier by Qvist *et al*[25] to have a large dry cavity in its solvated form, based on a combination study of NMR experiments and computer simulations. In biological complex like proteins, the hydrophobicity of patches on their surfaces are not the only factors governing the drying transition of water in confinements, especially when they are usually surrounding by a lot of highly hydrophilic patches. The geometry/narrowness of the confinement also plays an important role to induce the evaporation transition of water.

Different empirical models of water molecules and different treatments of long-ranged electrostatic interactions were tested in Berne group's studies and evaporation transitions in those systems were reported to be robust, i.e. not model dependent.[51] It indicates that the capillary evaporation events in protein systems are likely controlled only by the thermodynamic properties of the confinements. While all-atom molecular simulations are powerful and informative, they are also quite computationally expensive, more direct and efficient models and methods are preferred in the event when only thermodynamics (e.g. phase transitions) are important. The implicit solvent models and coarse graining approach have gained successes in predicting the confined water behavior between extended hydrophobic

surfaces, and also provided qualitative dynamic information about the processes.[30-31, 34, 38-40, 63] However, there are no direct studies of the capillary evaporation events for confined water available in literature to bridge between the model surfaces and more complicated biological systems, despite the fact that coarse grained models are widely accepted in protein folding community.[64-65]

In our study, we take a step further to construct a coarse-grained model for both the confinements (proteins) and solvent (water) in the investigation of capillary evaporation inside protein complexes with large hydrophobic area, which are believed to be crucial during the final stage of protein folding and protein-ligand binding processes. We focused our studies on the protein systems that have exhibited water evaporation transition in detailed all-atom simulations. With this very computationally efficient model, we are able to predict capillary evaporations in same systems as identified by much more sophisticated methods in a considerably short amount of time. We can conclude that the capillary evaporation is controlled by macroscopic thermodynamics which can be captured by the coarse grained model taking both the shape of proteins and protein-water interactions into account. The molecular details such as the empirical force fields are not required in the model.

From recent studies, we have learned that large matched hydrophobic areas in two corresponding surfaces in protein systems are required but not sufficient to ensure the evaporation transition in the confined region. There should be two important types of effects inducing the water evaporation in melittin tetramer system: chemistry and geometry. From the geometric point of view, the concave shape of the confinement

may affect the hydrogen bonds network more than the slab geometry, so a more confined geometry such as that in the melittin tetramer system would induce the evaporation easier than other geometries. On the other hand, Giovambattista *et al.*[66] attempted to separate the chemistry part of the melittin tetramer system from the topological effects. They flattened both melittin dimer surfaces and studied the water inside confined region of the relatively flat surfaces, and found out the evaporation only occurred in a very small region close to the center of the hydrophobic patches on the surfaces. As stated previously, water evaporation process can only take place when the macroscopic contact angle of water on the surfaces exceeds 90^{o}. We use a similar approach to flatten the melittin dimer surface and measure the microscopic analog of the macroscopic contact angle[67] on the flattened surface. We indeed find a hydrophobic water contact angle (~113^{o}) for the central region of the melittin dimer, and the capillary evaporation condition is satisfied. In this study, we reveal the physical reason of the capillary evaporation inside a confined protein system.

1.3 Wetting behavior of heterogeneous surfaces

The protein surfaces discussed in the previous section are taken from the Protein Data Bank (PDB), and they consist of different types of amino acid residues, with large disparities in hydrophobicity. This kind of heterogeneity is commonly present in surfaces involved in both materials and biology. Researchers are seeking to accurately predict the hydrophobicity of mixed heterogeneous surfaces given the information about their pure constituents. Therefore, it is crucial to understand the additivity or nonadditivity of surface free energies on heterogeneous surfaces. In current section, I will introduce our studies in probing surface free energy additivity on chemically heterogeneous surfaces using microscopic analogue of macroscopic contact angles, with surface heterogeneities in two different lengthscales (molecular scale heterogeneity and nanoscale heterogeneity). The complete results and discussions of this project will be presented in Chapter 5.

Wetting phenomena on chemically heterogeneous surfaces are important in materials and biology, examples ranging from inkjet printing to protein hydration.[68-69] Conventional metrics of surface interactions, designed for homogeneous systems, can often be applied to mixed surfaces characterized by averaged properties of multiple ingredients. The design of composite surface materials, and characterization of biosurfaces, benefit from combining rules predicting the interfacial free-energy change of wetting, $\Delta\gamma$, from the knowledge about individual constituents and surface composition. In view of Young equation, $\Delta\gamma=-\gamma\cos\theta_c$, the strength of inter-solute adhesion, $W_a \sim -2(\Delta\gamma+\gamma_{sg})$, relates to contact angle θ_c; here γ and γ_{sg} denote surface

tensions of the solvent and dry solute, respectively.[70] Contact angles on macroscopic heterogeneous surfaces are commonly estimated by Cassie equation[71-72]

$$\cos\theta_c = f_A r_A \cos\theta_A + f_B r_B \cos\theta_B \quad (1\text{-}2)$$

developed by assuming linear additivity of the wetting free energies, $\Delta\gamma$. Here f_α is the projected fractional area occupied by component α, θ_α the contact angle on a homogeneous surface of type α, and r_α the roughness factor defined as the ratio of solvent-exposed areas of patch of type α in the mixture and that of a surface fragment of equal projected area on a pure α surface. On macroscopically mixed surfaces, the roughness of surface components is often equal as on homogeneous surfaces ($r_\alpha \sim 1$), suggesting linear interpolation

$$\cos\theta_c = f_A \cos\theta_A + f_B \cos\theta_B \quad (1\text{-}3)$$

Presuming additivity of molecular polarizabilities, dipoles, and charges, Israelachvili and Gee [73] proposed an alternative expression for the contact angle of surfaces mixed at the molecular level

$$(1+\cos\theta_c)^2 = f_A(1+\cos\theta_A)^2 + f_B(1+\cos\theta_B)^2 \quad (1\text{-}4)$$

Measurements and simulation studies in a broad range of systems, from Lennard-Jones model surfaces to self-assembled monolayers (SAMs), showed mixing relations for $\cos\theta_c$ to break down upon addition of hydrophilic surface islands. Measured deviations from the additivity predictions have different signs, resulting in more hydrophilic (positive deviation in $\cos\theta_c$) or more hydrophobic surfaces (negative deviation).[74-78] Simulation studies have so far not addressed systems with negative deviations. Positive deviations were attributed to differences between averaged

surface properties and those under the droplet perimeter [79-80], drop size effects [81], solvent depletion at the solid/liquid interface [78], and patch size dependence [82-83], an observation reinforced by recent adhesion force measurements.[78]

Despite helpful insights from molecular simulations[35, 37, 78, 81, 83-84], systematic studies of heterogeneous-surface free energy at the molecular level are still lacking, and the knowledge of underlying principles behind observed deviations of *both* signs from additivity remains incomplete. Among the various properties discussed as measures of wettability on molecular scale[37, 84], the increase in the local compressibility of the solvent[37, 85-86] has been proposed as a viable measure of hydrophobicity[87-88] particularly useful in studies at molecular resolution[87]. Our calculations on molecularly mixed surfaces, reported in Chapter 5 (Section 5.3.2), capture compressibility changes consistent with contact angle variation that inherently averages over large areas. However, it is impossible to address surface free energy additivity from the compressibility perspective and contact angle remains the key quantity to study. We use molecular dynamics (MD) to measure the microscopic analogue of macroscopic contact angles[67] on a variety of molecularly mixed surfaces as a function of composition. We consider functionalized synthetic substrates and biomimetic surfaces comprised of protein fragments to test predictions for surface free energies of heterogeneous surfaces and their generalizations to patterned surfaces mixed at molecular and fragment levels.

Thermodynamics predictions for patterned surfaces are impeded by the lack of a rigorous measure of surface composition. For molecularly mixed surfaces, projected

fractional areas, f_α, are only approximate descriptors as the actual exposure of individual moieties can depend on local environment (see e.g. Fig. 3 in ref [78]). In analogy with the inclusion of macroscopic roughness in Eq. 1-2, our study addresses the modified exposure of distinct types of surface groups in the mixture. For molecularly mixed synthetic surfaces we show that *changes* in the solvent-accessible surface (SAS) of mixture components provide a unified, molecular-level explanation of *both* positive and negative deviations from the linear additivity of $\cos\theta_c$. Our results show linear additivity, analogous to Cassie equation, prevails on surfaces with small patches and moderate polarity variations when SAS-based fractional areas determine surface composition.

Heterogeneity is a generic feature of natural materials. For example, a spectrum of hydrophilic/hydrophobic surface compositions is spanned as a newly synthesized protein approaches native conformation. In addition to calculations on model synthetic surfaces, we are the first to study the microscopic analogue of contact angles on protein-like biological surfaces. We demonstrate and interpret prominent deviations from additivity on prototypical biological surfaces where stronger polarity variations come into play. Here, the origins of deviations from additivity in $\cos\theta_c$ are different; they include anti-cooperative interactions between a water molecule and multiple polar sites on the surface, preferential wetting of polar patches, and changes in the comparatively loose connection between the liquid interface[89] and hydrophobic domains. These features conform with the stronger influence of scarce hydrophilic moieties, critical to tuning of biomolecular solubility and function in aqueous solution[84,

[87]. Our work should assist in predictions of protein interactions and abatement of unwanted association that interferes with protein refolding in biotechnology[90].

The rest of the dissertation is organized as follows: I describe general models and methods used in our studies in Chapter 2, and they are also discussed in details when modified and actually applied in each research project in the following chapters. The three major research projects are presented in the order of their appearances in the introduction, in Chapters 3, 4, 5, respectively. Additional information and technical details are documented in "Appendices" section towards the end of the dissertation.

Chapter 2. Models & Methods

I have applied various models and methods of statistical physics and statistical mechanics [91-93] on both molecular and mesoscopic scales, all of them listed in Fig. 3 under different categories. I will have an overview for each of these standard models and methods in current chapter, based on different lengthscales.

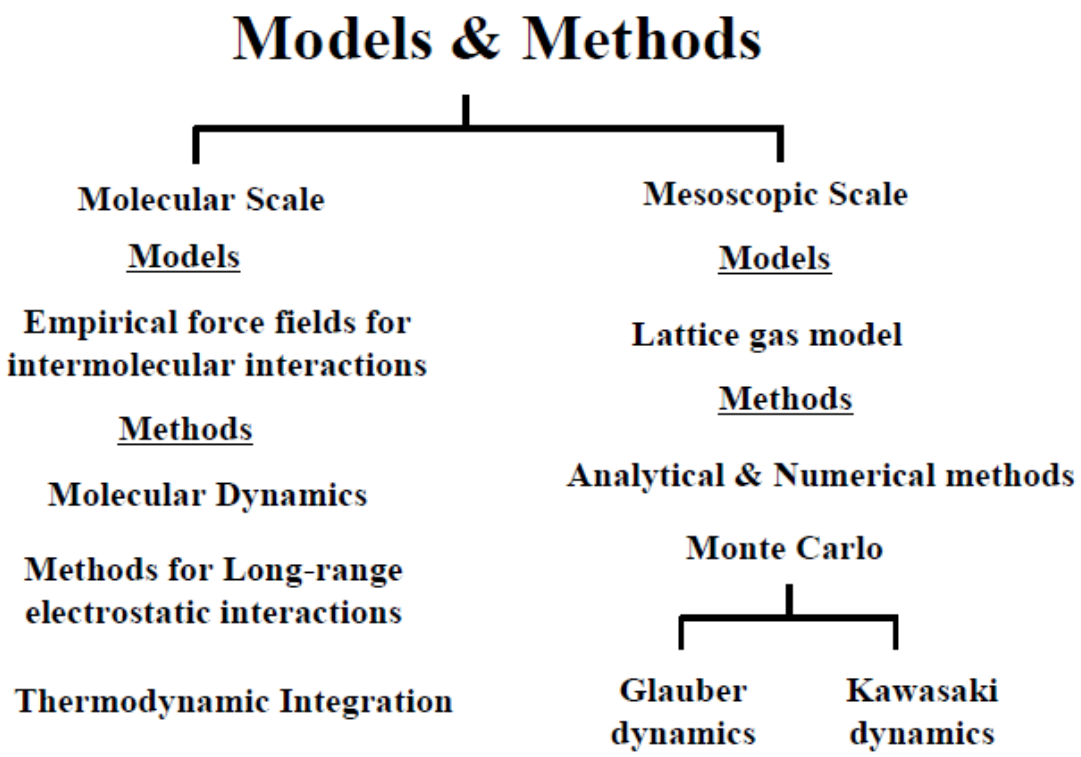


Figure 3. Standard methods and models used in this dissertation.

2.1 Molecular Scale

Models

2.1.1 Empirical force fields for intermolecular interactions

In classical molecular simulations, the intermolecular and intra-molecular interactions are modeled by the empirical force fields with the mathematical energy functions. The intra-molecular interactions include energies from bonds, angles, torsions, dihedral angles and so on, while the intermolecular non-bonded interactions include the short-ranged Lennard-Jones (LJ) energy[94] and the long-ranged electrostatic energy between molecules. The parameters in empirical force fields for all types of interactions are carefully parameterized to reproduce both experimental results and selected properties in high-level quantum mechanical calculations.

There are many types of force fields developed by different groups and they have been widely used in all kinds of systems in computer simulations, ranging from both small molecules and macromolecules to biological systems such as proteins and DNAs. Among them, I have used the AMBER (Assisted Model Building and Energy Refinement) force fields[95], originally developed by Peter Kollman's group at the University of California, San Francisco; the CHARMM (Chemistry at HARvard Molecular Mechanics) force fields[96] originally developed at Harvard University; and the OPLS (Optimized Potential for Liquid Simulations) force fields[97] developed by William L. Jorgensen's group at Yale University. The water model is also important in studying the hydrophobic effects, I use a generally accepted three-site rigid and non-polarizable water model in Extended Simple Point Charge (SPC/E)[98] model for

its efficiency and reproducibility of many structural and thermodynamic properties of liquid water at ambient condition, such as effective pair potential, density, radial distribution functions, and diffusion constant. In different projects, I have also tested other water models, including the transferable intermolecular potential 3 point (TIP3P)[99] model and the Simple Point Charge (SPC)[100] model, and both of them give similar results as the SPC/E model in my studies.

The sketch of SPC/E water model is shown in Fig 4, in addition to the partial charges on all of the three sites. Oxygen site has Lennard-Jones[94] parameters of $\varepsilon = 0.6502 kJ / mol$, and $\sigma = 3.166$ Å.

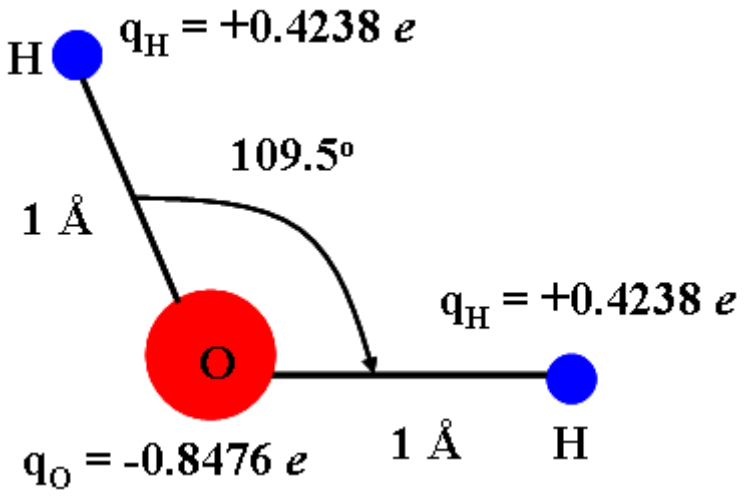


Figure 4. Parameters for Extended Simple Point Charge (SPC/E)[98] water model

I have mostly been using the rigid models in molecular simulations part of my study. Therefore, I will focus on the intermolecular interactions in this section. In classical molecular simulations, the interactions are assumed to be pairwise additive. The short ranged interactions of a system are normally represented by Lennard-Jones (LJ) energies as shown in equation 2-1,

$$E_{LJ} = \sum_{i \neq j} \varepsilon_{ij} [(\frac{R_{\min,ij}}{r_{ij}})^{12} - 2(\frac{R_{\min,ij}}{r_{ij}})^{6}], \qquad (2\text{-}1)$$

in which i, j represent the atoms in the system, $\varepsilon_{i,j}$ is the LJ well depth of the pair, taken as the geometric mean of the ε of each atom, r_{ij} is the distance between atoms i and j, $R_{min,ij}$ is the LJ radius for the pair, usually calculated by the arithmetic mean of atomic radii of the atoms, another common representation is $R_{\min,ij} = \sigma_{ij} \cdot 2^{\frac{1}{6}}$. The first term in equation 2-1 is the repulsive part of the interaction, and the second term is the attractive part. Total LJ energy represents the nature of Van der Waals interaction between atoms.

In addition, if there are charges present in the system, the long-ranged electrostatic interactions are included in the potential energy. The electrostatic energies are calculated by the Coulomb's law in Equation 2-2,

$$E_{Coulomb} = \sum_{i \neq j} \frac{q_i q_j}{4\pi D r_{ij}}, \qquad (2\text{-}2)$$

where q_i and q_j are the partial charges on atoms i and j respectively, and D is the dielectric constant.

The short-ranged and long-ranged interactions are treated differently in molecular simulations, details will be given in the following sections.

Methods

2.1.2 Molecular Dynamics Simulation

Molecular Dynamics (MD) is one of major methods in computer simulations. The idea is to model the many body systems similarly as in real experiments, but at the molecular level and much shorter time scales as complements to conventional experiments. At the same time, we also want our system to be able to represent a macroscopic sample to calculate the thermodynamic properties. The periodic boundary condition is usually enabled to achieve this goal. Part of a two-dimensional projection of an infinite sample system is sketched in Fig. 5, the simulation box is the in the center. In such system, the atom moving out from one boundary of the simulation box enters from the opposite side at the same time.

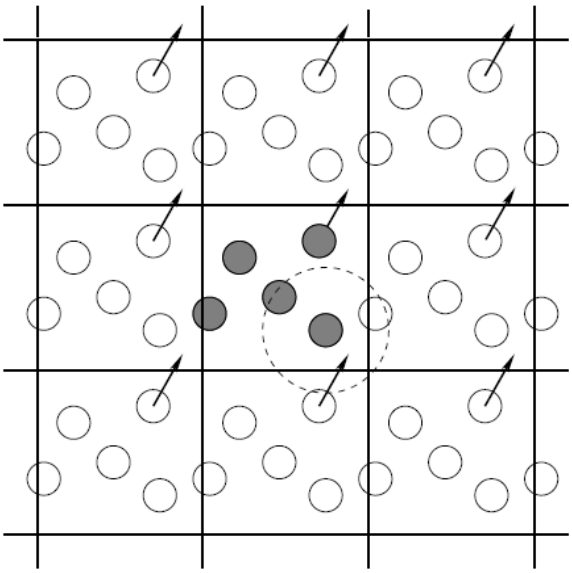


Figure 5. Periodic Boundary Condition in Molecular Dynamics simulation (projected on two dimensions), the box in the middle (with filled spheres) represents the simulation box. The arrow indicates the direction of the movement of that particular atom.

To start a MD simulation, we first need to initiate all the atomic coordinates of the system, create initial velocities on the atoms following a Gaussian distribution at the simulation temperature, and calculate the forces (as derivatives of the potential energy) based on empirical force fields discussed in Section 2.1.1. The total potential energy of particles in a three-dimensional system is

$$U_{total} = \frac{1}{2} \sum_{i,j,\vec{n}}{}' U(|\vec{r_{ij}} + \vec{n}L) \qquad (2\text{-}3)$$

where $\vec{r_{ij}}$ is the vector between each pair in the simulation box and $\vec{n}$ is an arbitrary vector with three integers and L is the box size of the simulation box. The prime in the sum indicates the self interactions (when $i=j$ and $\vec{n}=(0,0,0)$) is excluded.

The interactions between atoms are assumed to be pairwise additive in equation 2-3, however, when the periodic boundary condition is applied, the system size is infinite, and it is not feasible to evaluate the interactions for each pair in an infinite system. Practically, we can only deal with short-ranged interactions (discussed below in this section), and mathematical transformations are needed for the long-ranged electrostatic interactions (discussed in Section 2.1.3).

We can afford to truncate short-ranged interactions (i.e. the Lennard Jones interactions) beyond a certain cutoff distance r_c. There are different ways to truncate the interactions, including simple truncation, truncation and shift, and minimum image convention. To avoid discontinuities in the potential energies, the "truncation and shift" method is applied in our study. In this treatment, the interaction gradually vanishes at the cutoff radius, as described in equation (2-4).

$$U'(r) = 0 \text{, when } r > r_c \text{;} \quad U'(r) = U(r) - U(r_c) \quad \text{when } r \le r_c \text{.} \qquad (2\text{-}4)$$

The force on each atom is then calculated by the derivative of the potential energy, $f_i = -\frac{\partial U_i}{\partial r_i}$. And by solving the Newton's equations of motion $m_i \ddot{r}_i = f_i$, we are able to derive the positions and velocities of the atoms at next time step (usually at the scale of femto-seconds). Typically the Verlet algorithm[101] is employed for integrating the equations of motion,

$$\vec{r}(t+\Delta t) = 2\vec{r}(t) - \vec{r}(t-\Delta t) + \Delta t^2 \vec{a}(t) \qquad \text{(2-5a)}$$

$$\vec{v}(t) = \frac{\vec{r}(t+\Delta t) - \vec{r}(t-\Delta t)}{2\Delta t} \qquad \text{(2-5b)}$$

$\vec{r}(t)$ is the coordinates at time *t*, $\vec{a}(t) = f(t)/m$ is the acceleration at time *t*, where *m* represents the mass and *f (t)* is the force at time *t* and $\vec{v}(t)$ is the velocity at time *t*.

In equilibrium, we can calculate thermodynamic properties of the system by averaging them over all the time steps. We can also get the dynamics of the system during the simulation as the system evolves by itself. The molecular dynamics simulations can be complicated in practice, especially when we are dealing with systems at large scales, fortunately, there are many open source computational packages available, which are usually free for research purposes. I have used the LAMMPS[102] ("Large-scale Atomic/Molecular Massively Parallel Simulator") developed and maintained by a group in Sandia National Lab, and the DL_POLY simulation package[103] developed at Daresbury Laboratory by W. Smith and coworkers. In both cases, I tested the simulation packages for our systems and made modifications in the source codes inside these packages to make them suitable for my studies. (A typical subroutine additional to the standard package is attached in

Appendix C)

2.1.3 Long-range electrostatic interactions

The long-range electrostatic interactions are always difficult to compute in computer simulations. In molecular simulations, the electrostatic interaction (Coulomb interaction, shown in equation 2-2, falls off at the rate of r^{-1}) is a typical type of the long-ranged interactions. As illustrated in Fig. 6, the infinite simulation box can be built up by roughly spherical layers with periodic boundary conditions. However, due to the long-ranged nature of these interactions, the truncation treatment described in previous section may cause large deviations in simulations. The brute force solution to such problem is also impractical even with modern computers, and researchers have been questing and developing efficient alternatives over the decades.

The most common treatment for the Coulomb interactions is the Ewald summation method[104], and its extensions such as SPME (Smooth Particle Mesh Ewald)[105]. The idea is to split the electrostatic interactions into two parts in both the error function $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ and the complimentary error function $erfc(x) = 1 - erf(x)$. The complimentary error function vanishes to zero with increasing x quickly, this part becomes short-ranged and can be summed in real space similarly by the truncation method. The other part containing the error function is Fourier transformed into a reciprocal space (so called k space), and the sum converges quickly there by increasing the k vectors. The final results are given in equation 2-6,

where V is the volume of simulation box, k is the vector in the reciprocal space, α is the coupling parameter, q_i is the charge on atom i, and r_i is the coordinates of atom i.

$$U_{electrostatic} = \frac{1}{V}\sum_{k\neq 0}\frac{4\pi}{k^2}|\sum_{i=1}^{N} q_i \exp(i\vec{k}\cdot\vec{r}_i)|^2 \exp(-k^2/4\alpha)$$
$$-(\alpha/\pi)^{\frac{1}{2}}\sum_{i=1}^{N} q_i^2 + \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j erfc(\sqrt{\alpha} r_{ij})}{r_{ij}} \qquad (2\text{-}6)$$

Although the Ewald summation methods are commonly accepted in computing long-ranged interactions in molecular simulations, they are still computationally expensive, especially in some large simulation systems we are studying. There are alternative methods developed by different groups recently, including the "Gaussian Truncation (GT)" model from Weeks group[106] in University of Maryland and the "Force Matching (FM)" method from Voth group[107] in University of Utah. I will assess these methods in details in Chapter 3.

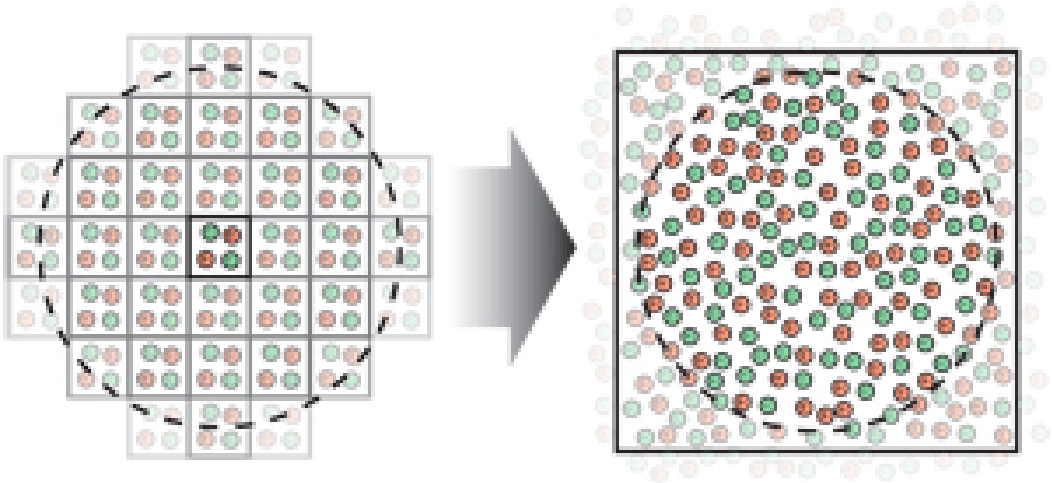


Figure 6. Building up the infinite simulation box in a roughly spherical shape in Ewald summation method.[92] The system is simplified to contain only two pairs of ions (red for anions and green for cations).

2.1.4 Thermodynamic Integration for Free energy calculation

Free energy is a useful measurement of the preferred direction of different kind of reaction, and in most cases the free energy difference between states is most easily calculated and also most interesting. One way to calculate the free energy is by thermodynamic integration. We use this method to calculate the solvation free energy of a solute in water (in Chapter 3).

We denote the total solvation energy of the solute as $V_{sol} = \sum_{i=1}^{N_w} V_0(r_i)$, the interactions between the solute and all the water molecules in the system. To get the solvation free energy, we increase the interactions in the system gradually by a coupling parameter λ, ranging from 0 to 1, λ=0 when there's no water-solute interaction in the system, and λ=1 when the water-solute interactions are fully charged, typically 10 steps are used between the charging process (λ=0, 0.1, 0.2 … 1), and the free energy calculated by thermodynamic integration process is given in equation 2-7.

$$\Delta G_{solvation} = \int_{\lambda_0}^{\lambda_1} d\lambda \langle \frac{\partial V_{sol}}{\partial \lambda} \rangle_{\lambda} \qquad (2\text{-}7)$$

2.2 Mesoscopic Scale

Models

At the mesoscopic scale, the molecular simulations with all atomistic details may not be efficient and in most cases not needed if we are only interested in the thermodynamic quantities. Coarse graining procedures are among the choices to study large systems, e.g. proteins. Just as the name implies, the idea of coarse graining is to use cruder models for atoms or atomic groups, and averaged interactions between these groups to reduce the degrees of freedom in the system, hence reduce the computational burden in simulation. Our coarse grained method is based on the lattice gas model.

2.2.1 Lattice gas model

Lattice gas model is equivalent to the Ising model[91], only with different representations. Instead of having spins which are only allowed to be at two states (up or down, represented by $S = \pm 1$) in Ising model, each site in a lattice gas system is either empty or occupied by one liquid molecule, and no two molecules can be at the same lattice site, the variable transformation is $n = (S+1)/2$, so n can only take the value of 0 and 1. Only molecules on the nearest neighboring lattice sites have interactions between each other with a coupling energy (ε in equation 2-8), otherwise the molecules are independent.

In a three-dimensional lattice gas model, the system is modeled by a three dimensional box of size $L_x \times L_y \times L_z$ with cubic lattice sites, and each site can be occupied either by liquid or vapor. The Hamiltonian is

$$H = -\varepsilon \sum_{<i,j>\in bulk} n_i n_j - \mu \sum_i n_i \quad , \tag{2-8}$$

the first term on the right hand side represents the water-water interactions, $n_{i/j} = 0$ when the site *i/j* is occupied by vapor, and $n_{i/j} = 1$ when the site is occupied by liquid, ε is the interaction parameter between neighboring liquid water occupied sites, μ is the chemical potential.

In the lattice system we are using for confined water in proteins, certain lattice sites are fixed to represent the protein surfaces. The modified model and parameters are discussed in details in Chapter 4.

Methods

The analytical and numerical methods applied in the mean field studies for hydrophobic hydration are discussed in details in Chapter 3 and Appendices. I will focus on introduction of the Monte Carlo simulations in this section.

2.2.2 Monte Carlo Simulation

The Monte Carlo method was originally developed at the end of the Second World War in studying the diffusion of neutrons in fissionable material by a group of scientists in Los Alamos National Laboratory. Since then, it has been widely applied in statistical physics, especially as an alternative and complementing method to the Molecular Dynamics (MD) simulation discussed in Section 2.1.2. The name of Monte Carlo (MC) method comes from the fact that the random numbers are extensively used in the method, and the City of Monte Carlo is famous for its casino. The random numbers are generated uniformly in the interval from 0 to 1, to avoid biases in the sampling. One advantage of the MC method compared to the MD simulation is that it could sample a broad range of the phase space more effectively, by making some "unphysical moves" in the system, and help the system to reach equilibrium fast, the shortcoming is that there would be no straightforward way to extract the dynamic information of a system from MC.

The steps in a basic MC simulation of a system with N particles at temperature T are described briefly as following: First, a particle is selected at random, and the energy involving this particle is calculated as $U(r)$; Then give the particle a random displacement $r'=r+\Delta$, and calculate the new energy $U(r')$; The difference in energy of

such a trial move can be computed and the move from *r* (old configuration) to *r'* (new configuration) is accepted with the probability:

$$acc(o \to n) = \min(1, \exp\{-\beta[U(r') - U(r)]\}, \qquad (2\text{-}9)$$

where acc denotes the probability of acceptance, min() takes the smaller number in the bracket and $\beta=1/kT$ is the Boltzmann factor, with k= 1.3806504(24)×10^{-23} J/K^{-1}. This algorithm is first documented by Metropolis *et al.* in 1953,[108] and sometimes referred to Metropolis Monte Carlo method, which satisfies the detailed balance of the system: $P(old)P(old \to new) = P(new)P(new \to old)$, and alternatively,

$$\frac{P(old \to new)}{P(new \to old)} = \frac{P(new)}{P(old)} = e^{-(E_{new} - E_{old})/kT} \qquad (2\text{-}10)$$

At equilibrium, the thermodynamic quantities could be taken by averaging over all the configurations during trial moves. The techniques such as periodic boundary condition, truncation method for short-ranged interactions and methods for long-ranged interactions discussed in Section 2.1 also apply to the MC simulations.

2.2.3 Glauber dynamics & Kawasaki dynamics

There are two types of dynamics to study the lattice gas system described in Section 2.2.1 with Monte Carlo (MC) simulation. In Glauber dynamics[109], a liquid occupied sites in lattice gas can evaporate or appear anywhere in the system with temperature dependent probabilities during a trial move, the number of particles in the system is not conserved, it uses Grand Canonical ensemble (the chemical potential μ, volume V and temperature T are constant during the moves). The procedure of Glauber dynamics of an Ising system is described as following: One spin out of an

N-spin system is chosen at random, and the energy of flipping that spin is calculated at ΔE, and from detailed balance, the probability of accepting a spin flip attempt is $1/(1+e^{\Delta E/kT})$. One unit of time in is N spin-flip attempts.

Although the Glauber dynamics MC method is qualitatively informative, it does not give a physical time scale in the system, since there is no mass transport associating with the nonphysical moves. To incorporate the effect of mass transport, the Kawasaki dynamics[110] was developed a little later, where solvent (water) occupied sites can only jump with temperature dependent probabilities from one place to an empty neighboring site, the total number of particles is conserved during trial moves (the number of particles N, volume V and temperature T are constant during the moves). The acceptance probabilities of both methods are determined by Metropolis criterion (equation 2-9).[108]

Chapter 3. Crossover length scale studies of hydrophobic hydration

As introduced in Section 1.1, the crossover lengthscale of hydrophobic hydration is of interest in both theoretical and bio-modeling community. In current chapter, I will first discuss the mean field model based only on the geometric restrictions to predict hydrogen bonds reductions near hydrophobic surfaces, and further the solvation free energy of spherical hydrophobic solutes. Then our molecular simulation studies of electrostatic contributions to the hydration of spherical solutes will be presented in the second part of this chapter.

3.1 Mean Field Model

The mean field model is inherited from Luzar *et al.*'s earlier work[16-17], which has successfully predicted that each water molecule sacrifices one hydrogen bond on average at planar hydrophobic surface. We have extended the same model to the spherically curved surfaces to make it suitable to study the solvation of hydrophobic solutes in water. Furthermore, a generalized model with similar idea has been developed during the process.

We assume in bulk conditions, each water molecule can form up to four hydrogen bonds with the neighboring water molecules in a tetrahedral network. The distance between the pairs should be equal to the length of a pair O-H…O, d = 0.3 *nm*, and the angle between the bonds is the tetrahedral angle $\alpha_T = 109.5^o$.

3.1.1 Convex Surfaces

For the convex surfaces in water (a spherical solute etc.), let the radius of the curvature be R, and the distance from central water molecule to the center of the curvature be $(R+x_0)$. The four neighboring positions of possible hydrogen bonded water molecules are controlled by two angles (depicted in Fig. 7): $\theta \in [0,\pi]$ is the angle between the surface normal and the central axis of the tetrahedron; $\gamma \in [0,2\pi]$ is the rotation angle of the central axis of the tetrahedron. If we make a tangent plane to the surface, the distance of the four possible hydrogen bonding positions to the plane can be described by the parameters in equation (3-1), derived from geometric relations with the auxiliary chart in Fig. 8.

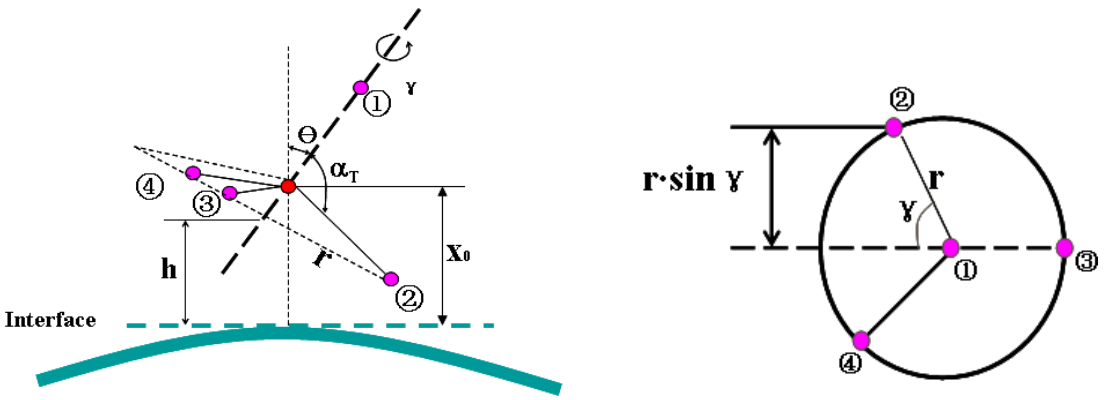


Figure 7. Model used to represent the water molecules near the convex interface. (left) side view(right) bird's-eye view from the rotation axis. The notations are introduced in the text.

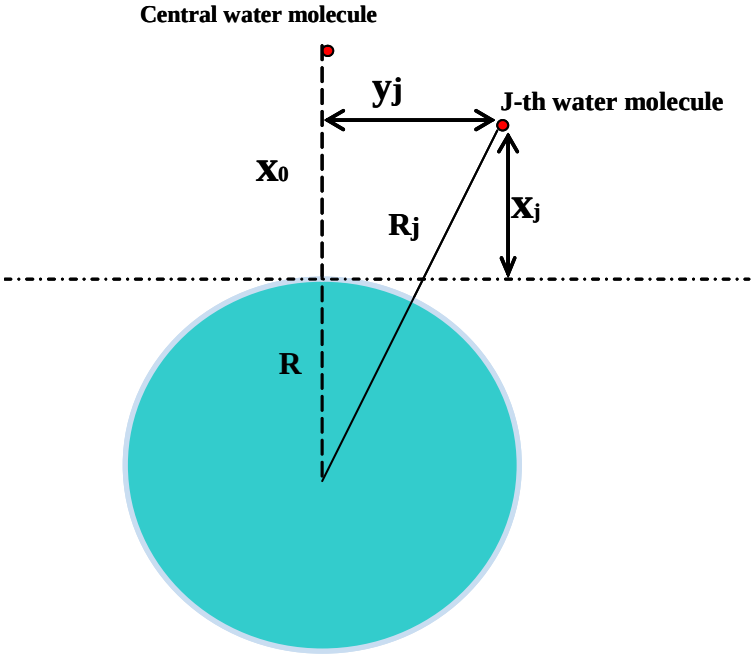


Figure 8. Geometrical calculations for convex surface.

$$x_1 = x_0 + d\cos\theta$$

$$x_j = h + r\sin\gamma_j \sin\theta (j = 2, 3, 4),$$

$$\gamma_2 = \gamma, \gamma_3 = \gamma + \frac{2\pi}{3}, \gamma_4 = \gamma + \frac{4\pi}{3} \quad , \qquad (3\text{-}1)$$

where $r = d \cdot \sin(\pi - \alpha_T)$, and $h = x_0 - \cos\theta \cdot \sqrt{d^2 - r^2}$.

Therefore, the distance from the *i-th (i=1,2,3,4)* water molecule to the center of curvature becomes

$$y_i^2 = d^2 - (x_0 - x_i)^2$$

$$R_i^2 = y_i^2 + (R + x_i)^2 \Rightarrow R_i^2 = d^2 - (x_0 - x_i)^2 + (R + x_i)^2 \qquad (3\text{-}2)$$

The correlation length in the model is 0.3 nm, so the surface layer thickness (the maximum distance of the central water molecule from the surface, x_0 is 0.3 nm. Within this model, we only consider the geometrical restriction for the hydrogen bonds between water molecules. If the distance from a certain tetrahedron vertex position to the center of curvature is larger than the curvature radius ($R_j > R$), we assume this hydrogen bond is allowed, and should be counted in the following calculations. On the other hand, if $R_j < R$, it is not possible to form a hydrogen bond from this position to the central water molecule, and this bond should be excluded. The possible bonding condition requires the position of x_j satisfies the relation:

$$d^2 - {x_0}^2 + 2x_0 \cdot x_j + 2R \cdot x_j > 0 \qquad (3\text{-}3)$$

We may define an equation to represent the possible bonds of a center water molecule which is x_0 nm away from the surface and with certain orientation (θ,γ). $\Gamma_i(x_0,\theta,\gamma) = 1$ if the number of possible hydrogen bonds equals to i, otherwise, $\Gamma_i(x_0,\theta,\gamma) = 0$. The ratio of the possible hydrogen bonds each water molecule can make at certain distance x_0 away from the interface, including all possible orientations, is calculated by integration[16]:

$$Y_i(x_0) = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} \Gamma_i(x_0,\theta,\gamma)\, d\theta\, d\gamma \qquad (3\text{-}4)$$

When only the geometrical restriction is considered, in bulk condition, no interface exists, and each water molecule can make four hydrogen bonds to its neighbors. Let the uniform number density of bulk water be ρ_w, then the average density of possible hydrogen bonds in bulk water should be $2\rho_w$. When we bring in the interface, the average hydrogen bond density is reduced due to the geometrical restriction we calculated above, and it becomes

$$\rho(x_0) = \frac{1}{2}\rho_w \sum_{i=0}^{4} i \cdot Y_i(x_0), \qquad x_0 \in [0, d]. \tag{3-5}$$

The difference between the average number of hydrogen bonds at the interface and in the bulk system is the integral of equation (3-4) over the interfacial region,

$$I = \int_0^d [2\rho_w - \rho(x_0)]dx_0 \tag{3-6}$$

We can monitor the reduction of the hydrogen bonds near the surfaces of hydrophobic solutes at different radius due to pure geometric restrictions. The $Y_i(x)$ functions defined in Equation (3-4) reflect the ability of water to make i hydrogen bonds at distance x_0 away from the solute surfaces. At the small lengthscale shown in Fig. 9(a), the function $Y_4(x_0)$ (maroon curve) becomes non-zero very close to the solute surface (x_0 < 0.1 nm), and $Y_1(x_0)$ (red curve) is always negligible. At the larger length scale shown in Fig. 9(b), the function $Y_1(x_0)$ (red curve) is non-zero very close to the solute surface (x_0 < 0.05 nm), and $Y_4(x_0)$ (maroon curve) is zero until (x_0 > 0.1 nm). That means the water always tends to make more hydrogen bonds near a small solute than near a large solute, as expected.

After integrating the $Y_i(x_0)$ functions over the surface region $0 \le x_0 \le 0.3nm$, we

get the surface-to-bulk hydrogen bonds ratio plotted in Fig. 10. It is the ratio between the numbers of possible hydrogen bonds each water molecule can form at the solute surface to the bulk water (we assume each water molecule can make four hydrogen bonds in bulk). As solute surface becomes sufficiently large, the reduction in hydrogen bonds reaches 25%, in agreement with the results at planar surface.[11, 16-19]

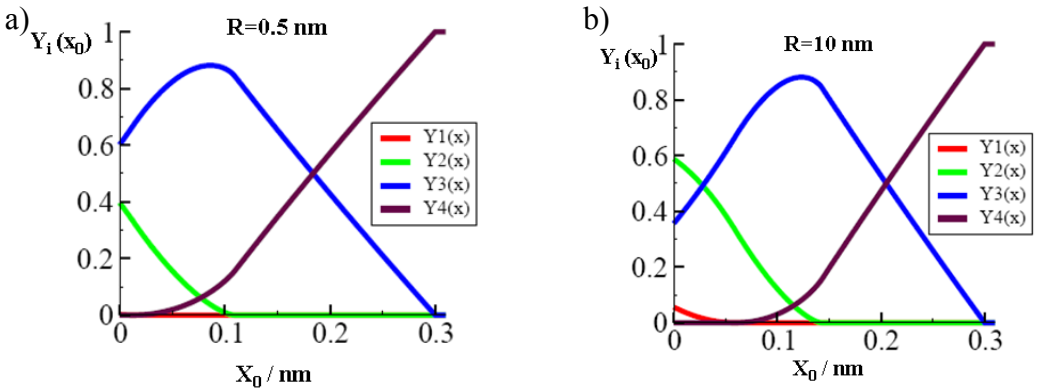


Figure 9. The fractions of water molecules able to form at most i bonds as a function of the distance from the solute surface. The functions $Y_i(x_0)$ are defined in Equation 3-4. a) water molecules near small solute (R=0.5 nm) b) water molecules near large solute (R=10 nm)

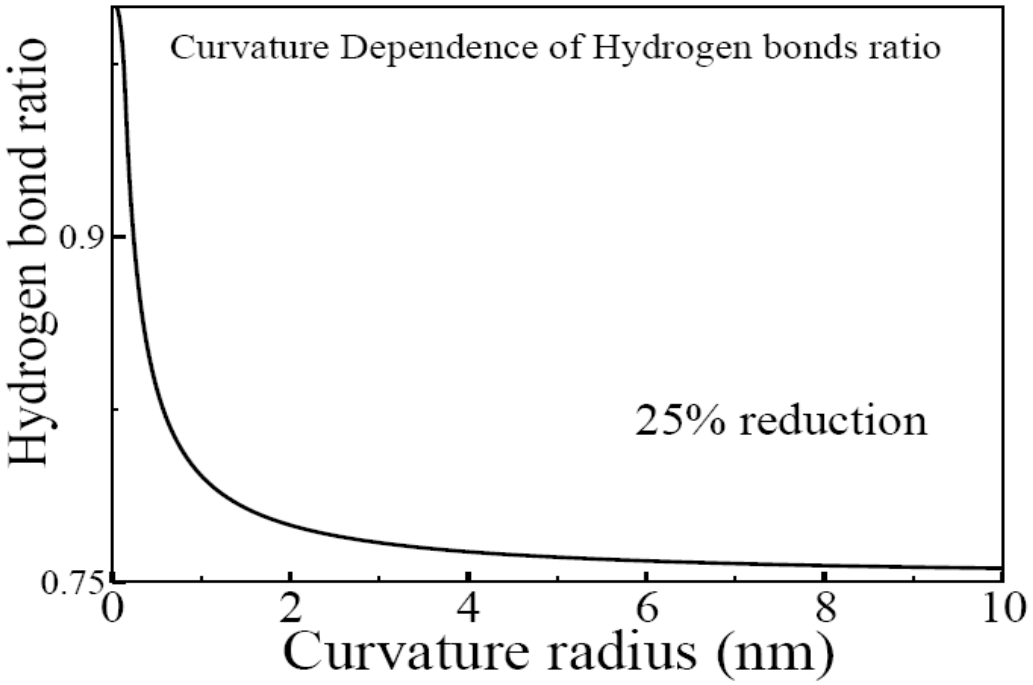


Figure 10. The surface-to-bulk hydrogen bonds ratio (described in text) dependence on the radius of the solutes (Convex curvature).

3.1.2 Concave Surfaces

Similar to the convex curvature described in the previous section, the possible bonding condition of water molecules near a concave curvature can also be derived based on geometric calculations.

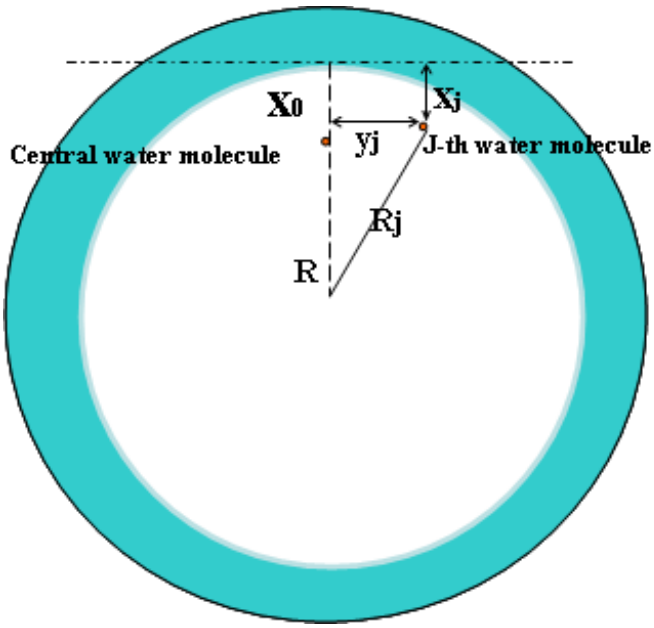


Figure 11. Geometrical calculations for convex surface.

Here, instead of equation 3-3, the position of all the possible bonding water positions x_j must satisfy the relation,

$$d^2 - {x_0}^2 + 2x_0 \cdot x_j - 2R \cdot x_j < 0 \quad , \qquad (3\text{-}7)$$

to be able to form hydrogen bonds with the center molecule. We can apply the equations 3-4 and 3-5 to water molecules near the concave surfaces as well to calculate the reduction of surface hydrogen bonds compared to bulk, and the curve is shown in Fig 12, as the hydrogen bond ratio increases with radii of the curvature, and reaches the plateau value of 75% when the radius is sufficiently large, in good agreement with the results at planar surface.[11, 16-19]

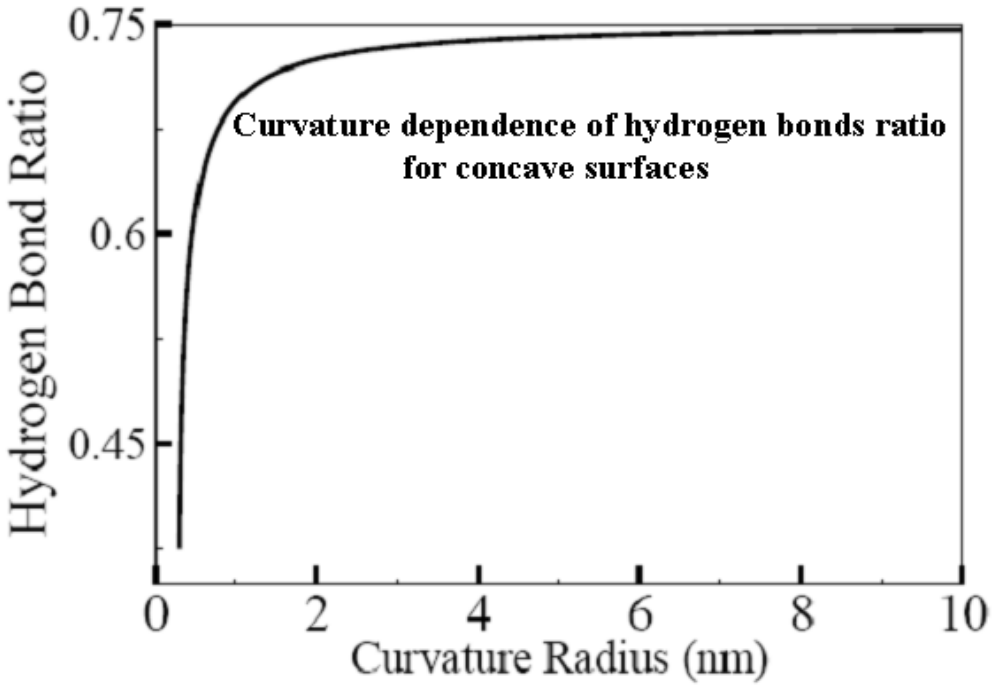


Figure 12. The surface-to-bulk hydrogen bonds ratio dependence on the radius of the concave curvature.

3.1.3 Generalized Model Surfaces

The mean field model can further be extended to all kinds of model surfaces in three-dimensional space. Instead of two angles for the tetrahedron rotations near the spherical surfaces, three angles similar to the rotational Euler angles in the space are needed in the generalized model, as shown in Fig. 13. α_T is the tetrahedral angle, 109.5^o, the central water molecule is depicted as a red sphere and surrounding possible hydrogen bonding positions with three angles γ, ψ, θ are in magenta color (following the tetrahedral geometry).

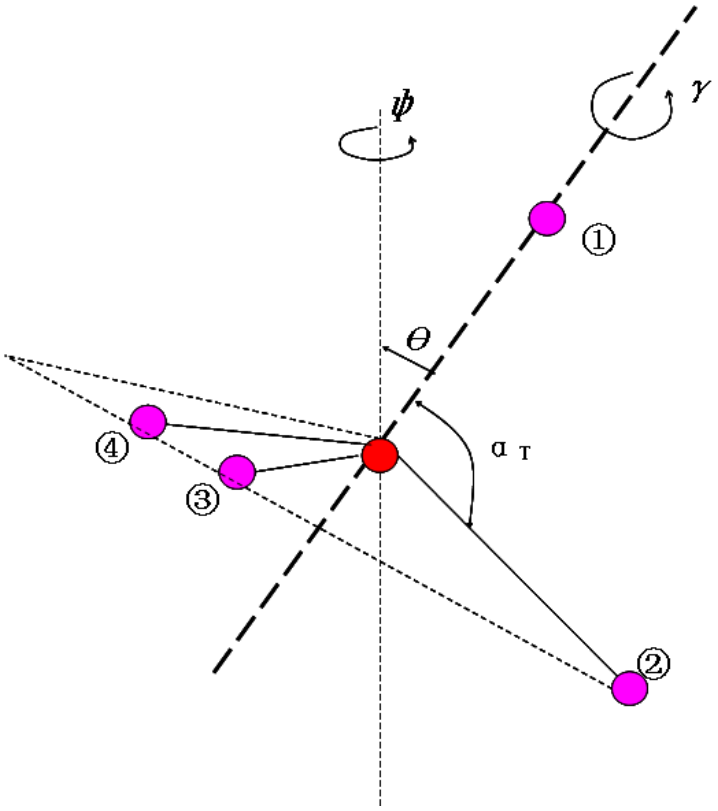


Figure 13. Representation of a tetrahedral system with three angles (described in text).

The *x, y, z* coordinates of the four possible hydrogen bonding positions can be derived from geometric calculations, although they are more complicated than the ones shown in previous sections: let *d* be the correlation length of the mean field model, 0.3 nm, and (x_0, y_0, z_0) be the coordinates of central water molecule. The possible bonding position #1 has the coordinates:

$$x_1 = x_0 + d * \sin\theta\cos\varphi \qquad (3\text{-}8a)$$

$$y_1 = y_0 - d * \sin\theta\sin\varphi \qquad (3\text{-}8b)$$

$$z_1 = z_0 - d * \cos\varphi \qquad (3\text{-}8c)$$

Let $r = d * (\pi - \alpha_T)$, and

$$x_h = x_0 - \frac{d}{3} * \sin\theta\cos\varphi \qquad (3\text{-}9a)$$

$$y_h = y_0 + \frac{d}{3} * \sin\theta\sin\varphi \qquad (3\text{-}9b)$$

$$z_h = z_0 - \frac{d}{3} * \cos\theta \qquad (3\text{-}9c)$$

$$\gamma_i = \gamma + 2\pi / 3 * (i-1) \quad \text{for i=2,3,4.} \qquad (3\text{-}9d)$$

then the rest three possible positions for water are:

$$x_i = x_h - r * [(\cos\theta\sin\gamma_i) * \cos\varphi + \cos\gamma_i * \sin\varphi)] \qquad (3\text{-}10a)$$

$$y_i = y_h + r * [(\cos\theta\sin\gamma_i) * \sin\varphi - \cos\gamma_i * \cos\varphi)] \qquad (3\text{-}10b)$$

$$z_i = z_h + r * \sin\theta\sin\gamma \qquad (3\text{-}10c)$$

where *i=2,3,4*. The coordinates (x_i, y_i, z_i) will be used in tests to satisfy various criterions needed for certain geometries of model surfaces.

In principle, with the generalized Mean Field model, it is possible to estimate the hydrogen bonds' reduction based on pure geometric restriction near a model hydrophobic surface with any shape, given the mathematical functions or coordinates

of the surface geometry. Unfortunately, the correlation length in the model is always the hydrogen bonds length (0.3 nm), and there are no long-ranged interactions included in our model. Therefore, any structure correlation length larger than 0.6 nm will be considered the same in the model, which limits the application of the model to a wider range of practical problems. However, the generalized mean field model completes the geometric study of water hydrogen bonds near model hydrophobic surfaces and we are looking forward to using these results in future studies.

3.1.4 Mean Field model estimation of solvation free energy

Hydrogen bonds' contribution to the surface free energy of a model hydrophobic surface is studied by a model originally developed by Luzar and coworkers.[16] By integrating over the interfacial region, the surface free energy density based on the model is related to the reduction of surface hydrogen bonds:

$$\frac{\partial F}{\partial A} = -kT \ln[(1+\alpha)/(e^{-\beta E}+\alpha)] \cdot I \,. \qquad (3\text{-}11)$$

A is the surface area, and I is defined in equation 3-6.

The parameters α (the ratio between two sub volumes of bound and unbound O…OH pairs) and E (the energy of hydrogen bonding) are obtained by fitting the results from spectroscopy experiments[111-112]. We use $\alpha=28\pm8$ and $E=-13.4\pm0.8$ *kJ/mol*, in line with the hydrogen bonding energy in literature.[113]

In the studies of solvation free energy of hydrophobic solutes, we only need to consider the convex curvatures, and the contribution from reduction of hydrogen bonds of interfacial water can be calculated from results discussed in section 3.1.1

(Fig. 10). Solvation free energies contributed by hydrogen bonding reduction are normalized by surface area, and plotted as a function of solute radii as the black curve in Fig 14. The solvation free energy density (solvation free energy over the solute surface area) is equivalent to the surface tension.

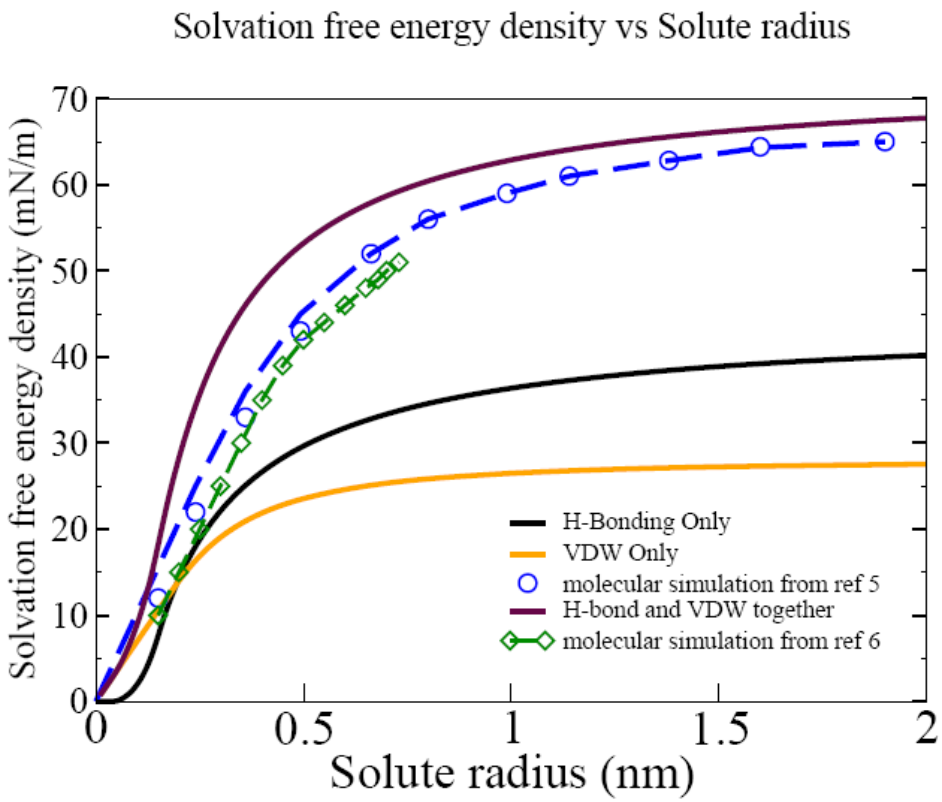


Figure 14. Comparison of solvation free energy (normalized by surface area) between molecular simulations and mean field model predictions. The black curve only counts the hydrogen bonding effects, the orange curve only counts Van der Waals interactions, and maroon curve counts both effects. The blue open circles and green open diamonds are data from two computer simulations (ref [6] and [7] respectively), the dashed lines are guides to the eye.

In addition to hydrogen bonds effects, the Van der Waals interaction effects on the solvation free energy have been studied over the past century. We have mostly adapted the results from Fowler's original work[114] in 1936 and more recent work from Stewart and Evans.[23] The details and derivations are shown in Appendix A. We use the Lennard-Jones parameters of the SPC/E (Extended Simple Point Charge)[98] water model and equations (A-14) and (A-17) (in Appendix A) to calculate to the surface free energy from the Van der Waals contribution (orange curve in Fig 15).

Adding both parts together, we get the complete solvation free energy from mean field model prediction (maroon curve in Fig 14) and it can be directly compared to the computer simulation results[6-7], which have included all the details about the water molecules (open circles and open diamonds in Fig 14). As shown in Fig. 14, the small-to-large crossover occurs at a similar radius (around 0.5 nm) based on our mean field model. This is the range of interactions (hydrogen bonds and Van der Waals interactions) in the solvation system of neutral solutes.

3.2 Molecular simulation studies for charged solutes

In previous section, we only considered the solvation systems of neutral apolar solutes. However, the systems with charged solutes are of more interests for studies of biological systems. As we discussed in Section I, the crossover lengthscale of charged solutes between the two regimes (solvation free energy scaling with volume and scaling with surface area) is important in biological studies.[25-28]

In charged systems, the mean field model involves numerically solving the Poisson Boltzmann equation[115] (Described in Appendix B). There are both slowly varying background field and instantaneously varying field in our solvation systems. The Mean field model averages them together. The model works fine for the potential, since the positive and negative parts in potential cancels out. However, when it comes to the dielectric energy, the Laplace operator of the field does not cancel (as shown in equation 3-12), in this case, the mean field model has natural difficulty in predicting the dielectric energy.

$$E_{Dielectric} = \frac{\varepsilon}{2}\int \nabla^2 \psi dV \qquad (3\text{-}12a)$$

$$\left\langle \int \nabla^2 \psi \right\rangle dV \neq \int \nabla^2 < \psi > dV \qquad (3\text{-}12b)$$

Therefore, the solvation free energies in charged solute systems are studied in molecular simulations. The model used in the calculation is similar to that reported by Dzubiella and Hansen[116]. The water molecules are simulated with SPC/E model[98], and the solute is represented as a hard sphere, which interacts with heavy atoms (water oxygen atoms and counter-ions) as

$$V_0(r) = \Phi(r - R_0)^{-12} \qquad (3\text{-}13)$$

where the parameter Φ is fit to make the repulsive energy equal to k_BT at the distance $r - R_0 = 1\,\text{Å}$, and the radius of the solute can be defined as $R = R_0 + 1\,\text{Å}$.

There are two steps in the thermodynamic integration to calculate solvation free energy. The first step is to grow the neutral solute (without charge) to the desired size, and the coupling parameter λ in equation (2-7) (Chapter 2) is naturally chosen to be the radius of the solute. The second step is to charge the solute gradually with a coupling parameter λ, where $Q(\lambda) = \lambda \cdot Q$.

For the first step, we were able to reproduce the results from ref [116], shown in Fig 15, the crossover occurs at ~0.5 *nm*.

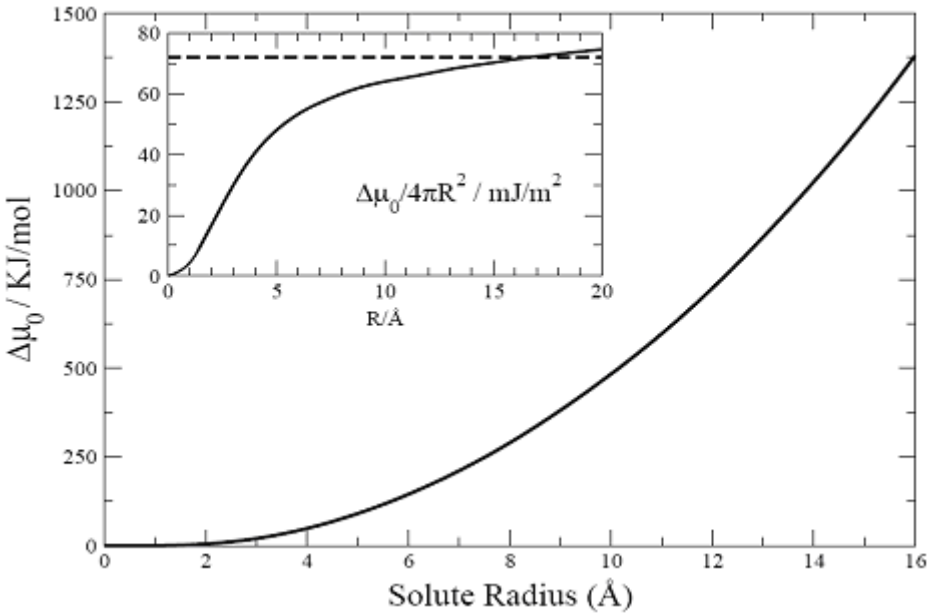


Figure 15. Solvation free energy as a function of solute radii for neutral solutes. (inset) Solvation free energy density (normalized by surface area) as a function of solute radii, the dashed line indicates the liquid/vapor surface tension (72 *mN/m*).

In the second step, the system we are studying systems consist of one macro-ion (solute), counter-ions and sufficient water molecules solvating the macroion. The solvation free energy is the energy difference between two states (in vacuum and in water solution, as in equation 3-14) sketched in Fig. 16. In the second step of thermodynamic integration, we charge the solute and counter ions simultaneously from $\lambda = 0$ to 1. The total interactions contain three parts: (1) water-water interactions, this should not contribute to the free energy part since it is not directly related to λ; (2) ion-water interactions, this part is proportional to λ; (3**)** ion-ion interactions, this part is proportional to λ^2. These three parts of the energy are calculated separately in simulation (in source code provided in Appendix C).

$$\Delta G_{Solvation} = G_{Liquid}(R) - G_{gas}(R). \qquad (3\text{-}14)$$

When we charge the solute, the electrostatic interactions overwhelm the Van der Waals interactions between the solute and water molecules. Therefore, the solvation free energy becomes negative, and it indicates the solvated state is more favorable for a charged solute. For different sizes of solutes, we keep the surface charge density of the solute fixed for all sizes of solutes. In the preliminary work, the surface charge density is chosen as $Q/4\pi R^2 = 0.02e/4\pi \text{Å}^2$, similar to the common surface charge densities in biological systems.

We modified the DL_POLY molecular dynamics package[103] to carry the molecular dynamics simulations in *NPT* ensembles with periodic boundary conditions. The time step is usually 2 *fs*. There are multiple modules having been modified to be suitable for our research purposes, and a major supplemental module I have written

(with Fortran 90 computer language) in addition to the DL_POLY standard package is attached in Appendix C.

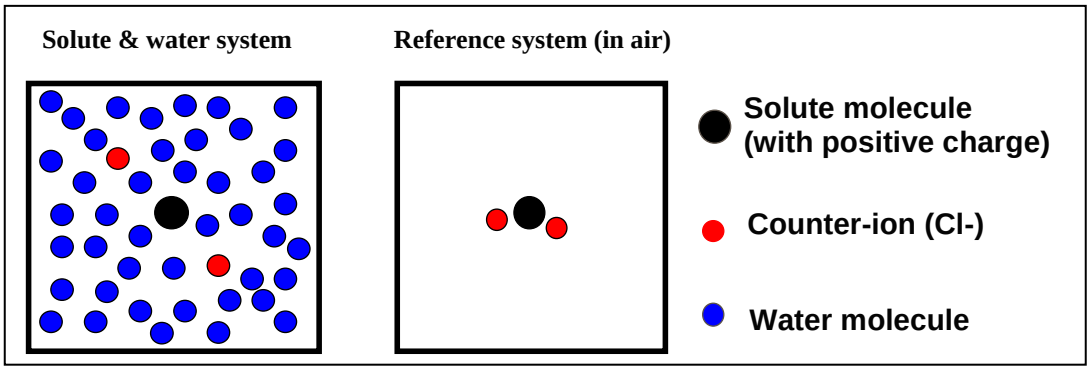


Figure 16. Representation of the model system.

The common treatment for the Coulomb interactions is the Ewald summation method[104], described in Section 2.1.3 and its extensions such as SPME (Smooth Particle Mesh Ewald)[105], which have been included in the distributed DL_POLY package. However, these methods are still quite computationally expensive. We tested a couple of efficient methods in calculating the electrostatic potential to reduce the computational cost in the following.

First, we used the "Gaussian Truncation" method developed by Rogers and Weeks[117], which separates the coulomb interactions *(1/r)* into short-ranged and long-range parts similarly as in Ewald sum treatment. In homogeneous systems, Rogers and Weeks suggested that we may not need the long-ranged part because that will be averaged out for a uniform background, and all charges should be treated equally if one wants to get that nice result of a Poisson's equation defined with a smoothed charge distribution[118]. We treated all the electrostatic interactions (water-water, water-ion and ion-ion) in their short-ranged form, the results were in good agreement with standard calculations in pure water systems and even concentrated aqueous salt solutions. However, when introducing a macro-ion (e.g. colloid particle or protein) in solution, using this method became very challenging. Due to the non-homogeneous background in electrostatic interactions, the long-ranged interactions must be taken into account, and the cost of calculating that part is comparable to the Ewald summation treatment.

Voth and coworkers[107] have developed another alternative method for Ewald summation to treat the long-range electrostatic interaction with a short-range effective potential, they have also successfully extended this method into various homogeneously disordered condensed phase systems. Their approach is to fit the long-range force from standard simulations of a bulk water system by short-ranged functions (usually polynomial functions), using so-called "Force Matching" (FM) algorithm, represented in equation 3-15.

$$F_{FM}(r \geq r_{core}) = \frac{1}{r^2} + \sum_{i=0}^{N} a_i r^i \qquad (3\text{-}15a)$$

$$F_{FM}(r < r_{core}) = \frac{1}{r^2} - \frac{1}{r_{core}^2} + F_{FM}(r_{core}) , \qquad (3\text{-}15b)$$

r is the distance between atoms, and $r_{core} = 0.148\text{nm}$ is the smallest distance in the fitting process. The second term in Equation 3-15a is the polynomial function fitting to the long range forces, the parameters for a cutoff distance of 1 nm are listed in Table 1.

i	a_i
0	−0.165477570871E-03
1	0.288823451703E-03
2	−0.122247561247E-03
3	0.963712701767E-05
4	0.251954672874E-06
5	−0.735796273353E-07
6	0.353601771929E-08
7	−0.525765995765E-10

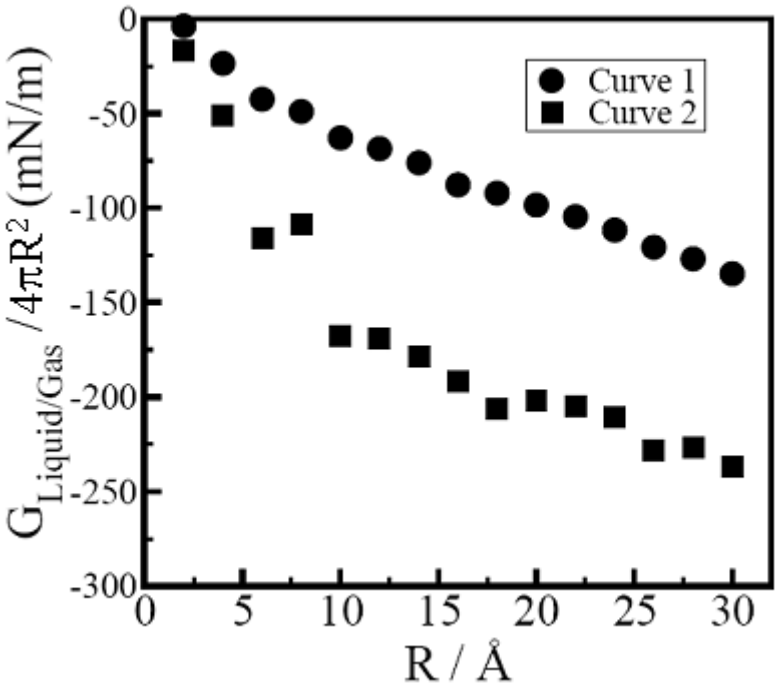
Table 1. Force Matching (FM) parameters for 1 nm cut off (based on the data provided in ref [107]

With the shortened cutoff distance, the force matching methods can save up to 80% of computational effort compared to the conventional Ewald summation algorithm. However, with a large macro ion, we were not able to sustain the short cutoff which is essential for its efficiency. With much longer cutoff distance, not only the parameters in the Force Matching algorithm need to be carefully re-parameterized, the efficiency of the method also declines significantly, and it would eventually be comparable to the efforts of Ewald summation.

Overall, there are difficulties to extend both alternatives into the simulation with macro ions, due to the small cutoff requirements in both methods. Although we have implemented both algorithms into the DL_POLY package, we still chose to use the built-in particle mesh Ewald summation algorithm in our studies.

The results of electrostatic part of solvation free energy have been summarized in Fig. 17. Fig. 17(a) shows the free energy in two states, curve 1 is the free energy density in the reference gas phase state and curve 2 is the free energy in aqueous solution, which represents the complete solvated state. In Fig. 17(b), curve 3 represents the solvation free energy from charging the neutral solute, which is the difference between the two curves in Fig. 17(a). In curve 3, we can notice that the crossover occurs beyond the 1 *nm* mark, showing that charges on the solute do move the crossover up to a larger lengthscale, confirming our predictions. However, the increment is not large enough to change the empirical views of free energy estimation in biological systems.

a)



b)

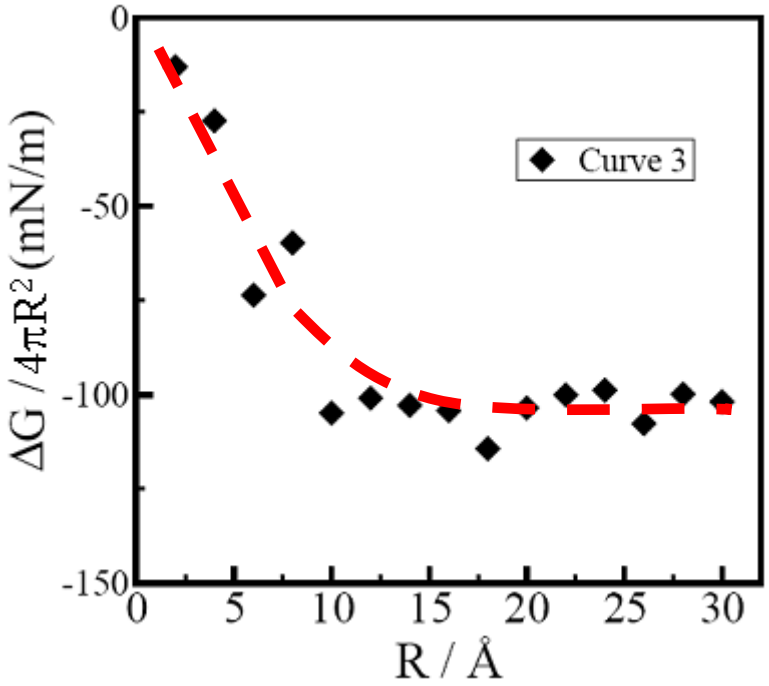


Figure 17. (a) Curve 1 is for the free energy density in gas phase $G_{Gas}(R)$**,** Curve 2 is for the free energy density in aqueous solution $G_{Liquid}(R)$. (b) Curve 3 is the solvation free energy density, the free energy difference between two states. The red dashed line is a guide to the eye.

In these calculations, we kept the thickness of water layers outside the solute as constant (20 Å), to have enough solvation shells of each solute. In addition, we manage to add in pairs of salt ions to control the Debye screening length of the system, which is calculated by

$$\kappa^{-1} = \sqrt{\frac{\varepsilon_r \varepsilon_0 k_B T}{2 N_A e^2 I}} \qquad (3\text{-}16)$$

where I is the ionic strength of the electrolyte, and here the unit should be mole/m^3, ε_0 is the permittivity of free space, ε_r is the dielectric constant of the medium, k_B is the Boltzmann constant, T is the absolute temperature, N_A is the Avogadro number and e is the elementary charge.

The Lennard-Jones parameters of ions in the system (cations are modeled as Na^+ and anions are modeled as Cl^-) are listed in Table 2, and the number of ions, salt ion pairs and water molecules for two different series of systems are listed in Tables 3 and 4.

	ϵ (10^{-23} J)	σ (Å)
O–O	107.95	3.169
O–Na^+	86.60	2.876
O–Cl^-	86.60	3.250
Na^+–Na^+	59.37	2.73
Na^+–Cl^-	28.32	3.87
Cl^-–Cl^-	27.87	4.86

Table 2. Lennard Jones parameters[119] for the ions in molecular simulations

Solute Radii (Å)	$N_{counter\ ion}$	Charge/ counter ion (e)	$N_{salts\ (NaCl)}$	Salt charge(e)	N_{water}
2	1	0.08	1	0.78	92
4	1	0.32	2	0.78	195
6	1	0.72	2	1	356
8	2	0.64	4	0.9	588
10	2	1	5	1	903
12	3	0.96	6	1	1315
14	4	0.98	10	1	1835
16	5	1.024	13	1	2478
18	6	1.08	18	1	3255
20	8	1	22	1	4181
22	10	0.968	28	1	5266
24	12	0.96	36	1	6525
26	13	1.04	44	1	7971
28	16	0.98	53	1	9615
30	18	1	65	1	11471

Table 3. The model systems used in simulation to control the Debye screening length at 5±0.5 Å, and the simulation box size is (2R+10) Å, R is the solute radius.

Solute Radii (Å)	$N_{counter\ ion}$	Charge/ counter ion (e)	$N_{salts\ (NaCl)}$	Salt charge(e)	N_{water}
2	1	0.08	1	0.45	355
4	1	0.32	1	0.54	579
6	1	0.72	1	0.51	873
8	2	0.64	1	0.59	1243
10	2	1	0	0	1420
12	3	0.96	0	0	1898
14	4	0.98	0	0	3088
16	5	1.024	0	0	3361
18	6	1.08	0	0	5056
20	8	1	0	0	6103
22	10	0.968	0	0	7275
24	12	0.96	0	0	8579
26	13	1.04	0	0	10021
28	16	0.98	0	0	11606
30	18	1	0	0	13341

Table 4. The model systems used in simulation to control the Debye screening length at 17±1 Å, and the simulation box size is (2R+18) Å, R is the solute radius.

In Tables 3 and 4, we varied the Debye screening length by more than three times for two series of systems. However, the solvation free energies are almost identical as plotted in Fig. 17 (points overlapped on each other). The reason could be that the number of salt ions in the system is much less than the number of water molecules (usually smaller than a ratio of 1:100), the majority of the system is made up by water molecules and salt ions are not expected to make sizable impact on the solvation free energy. On the other hand, no Debye screening factors are considered in the reference state (gas phase) of our model, and the free energy in reference states are significant compared to the final solvated state (aqueous solutions). Therefore, the macroscopic Debye screening length does not alter the crossover lengthscale of the solvation free energy (per unit area) at the molecular level as originally proposed. With the simulation results, we can conclude that the electrostatic interactions do not significantly alter the crossover lengthscales in the solvation processes for our model system, and the conventional treatment of the free energy in biological (macroscopic) systems is valid.

Chapter 4. Hydration of and solvent-induced interaction between protein surfaces

As introduced in Section 1.2, the capillary evaporation events in protein systems have attracted attentions due to its possible impact in biological processes, such as protein folding, protein-ligand binding and protein aggregations. There are two requirements for the confinement to induce the evaporation: The surface hydrophobicity has to be equivalent to that of a flat surface with water contact angle larger than 90 degrees; The grand potential of vapor phase must be smaller than the grand potential of liquid phase in confinements (the critical distance for slab geometry is described in equation 1-1 in Chapter 1, and the requirements for other geometries are demonstrated in Appendix D). However, in protein systems, the surfaces are always heterogeneous and the shapes of the surfaces are complicated. There is no simple solution to describe the confinements in proteins as those in Appendix D.

Due to the large sizes and complicated structures, biological systems are usually expensive to study in molecular simulations with all the atomistic details. If we are only interested in the surface hydrophobic properties on the protein surfaces and drying transition of water in the confinements, the coarse grained model can be much less detailed. There are amino acid residues exposed on the surface of proteins, some of them are hydrophobic while others are hydrophilic, and the hydrophobicity of the protein surfaces are mostly attributed to these amino acid residue types. In current chapter, we develop a coarse graining approach for both the shape of the proteins and the residue-water interactions incorporating a previously developed implicit water

model[31, 34] in a lattice gas system to study the water evaporation in confined protein systems. Our model and parameters are described in section 4.1, results and discussions follow in section 4.2 and 4.3, respectively.

4.1. Coarse graining approach

Various coarse grained models have been proposed over the years for different types of systems, and have gained success in predicting thermodynamic properties as well as in estimating dynamic information in large scale model surfaces and biological systems.[34, 40, 64-65] The recent computer simulations have also suggested different choices of empirical force fields did not affect the observation of capillary evaporation, indicating atomistic details may not be required in predicting evaporation events. With these inspirations, we develop a coarse graining model for both protein and solvent (water) specialized in probing the capillary evaporation events in the confinements of proteins. Obviously, the coarse grained model cannot give detailed information at the molecular level (such as single amino acid mutation), but we are able to capture the basic physics governing the evaporation process in protein systems, and the model is useful as a intermediate step between basic bioinformatics tools and studying molecular behaviors of the system.

4.1.1 Lattice gas model of the confined protein systems

The protein-confined water system is modeled by a three dimensional lattice gas system on cubic lattices of size $L_x \times L_y \times L_z$. similar to the general lattice gas model

described in section 2.2.1. However, in our system, some sites are fixed to represent the surface (protein), and other sites can be occupied either by the liquid or vapor of water. The Hamiltonian becomes

$$H = -\varepsilon \sum_{<i,j>\in bulk} n_i n_j - \sum_{i\in surface} \varepsilon_{s,i} n_j - \mu \sum_i n_i \qquad (4\text{-}1)$$

where the first term in the right hand side represents the water-water interactions $n_{i/j}$ = 0 when the site *i/j* is occupied by vapor, and $n_{i/j}$ = 1 when the site is occupied by liquid water, ε is the interaction parameter between neighboring liquid water occupied sites, μ is the chemical potential. The second term at the right hand side represents the surface interaction, the interaction between the surface and neighboring solvent sites coupling parameter $\varepsilon_{s,i}$ is in effect when there is a liquid water occupied site *j* adjacent to that surface site *i*.

The lattice constant and coupling parameters for the solvent (water) are carefully tuned[34] to reproduce the thermodynamic properties of water at ambient conditions[111, 113]. At room temperature T = 300 K, the isothermal compressibility $\rho k_B T \kappa_T$ =0.062, pressure P = 1 *atm* and surface tension of water γ=70 *mN/m*, if we adopt the reasonable zero temperature approximation, the surface tension was calculated at $\gamma \sim \varepsilon / 2a^2$. We are using $\varepsilon = 1.2646 k_B T$, lattice size a = 0.193 nm and chemical potential $\mu = -6\varepsilon/2 + 1.84247 \times 10^{-4} k_B T$ to ensure that the corresponding lattice-gas is at close proximity to liquid-gas coexistence.[34] k_B is the Boltzmann constant and T is the absolute temperature.

4.1.2 Coupling parameters for proteins

The protein structure files (*.pdb) are downloaded from the Protein Data Bank (PDB) (http://www.rcsb.org/). We take all the heavy atom coordinates of each amino acid residue from the pdb files and represent them by the corresponding lattice sites in the three-dimensional lattice gas system (a two dimensional projection of the system is sketched in Fig. 18). The shapes of the protein surfaces are conserved with this treatment. To be consistent with the solvent model described in previous section, the lattice size is set to be 1.93 Å, it is also similar to the resolution of most coordinates files in the PDB. Each heavy atom occupies 8 lattice sites (a 2*2*2 cubic) closest to its central position, at approximately 58 $Å^3$ in size and the interactions between the protein occupied sites and adjacent solvent available sites are converted to lattice parameter ε_s in equation (4-2) below, derived from the open, removal and close procedures of neighboring lattice positions.[24]

ΔG represents the hydration energy for an entire amino acid (the energy of moving it out from water into its condensed vapor), taken from the hydrophobicity scale of Kyte & Doolittle.[120] It is proportional to the energy of "remove an amino acid site and close the cavity" in the lattice gas model. z is coordination number, A is the exposed surface area for a single amino acid residue[121], L_c is the lattice constant[34], ε is the water-water interaction, and ε_s is the water-surface interaction. The energy cost to remove one amino acid site in the system is $\Delta U_{remove} = -z \cdot \varepsilon_s$, and the energy to close this cavity by a water site is $\Delta U_{close} = \frac{z}{2} \cdot \varepsilon$, for each amino acid, the removal and closure cost can also be written as $\Delta U_{remove+close} = \Delta G \cdot \frac{z \cdot Lc^2}{A}$, and it is equivalent to

$\Delta U_{remove} + \Delta U_{close}$, by equalizing these two representations, we can arrive at the equation (4-2), to get the surface-water interaction parameter.

$$\varepsilon_s = \varepsilon / 2 - \Delta G \cdot L_C{}^2 / A \qquad (4\text{-}2)$$

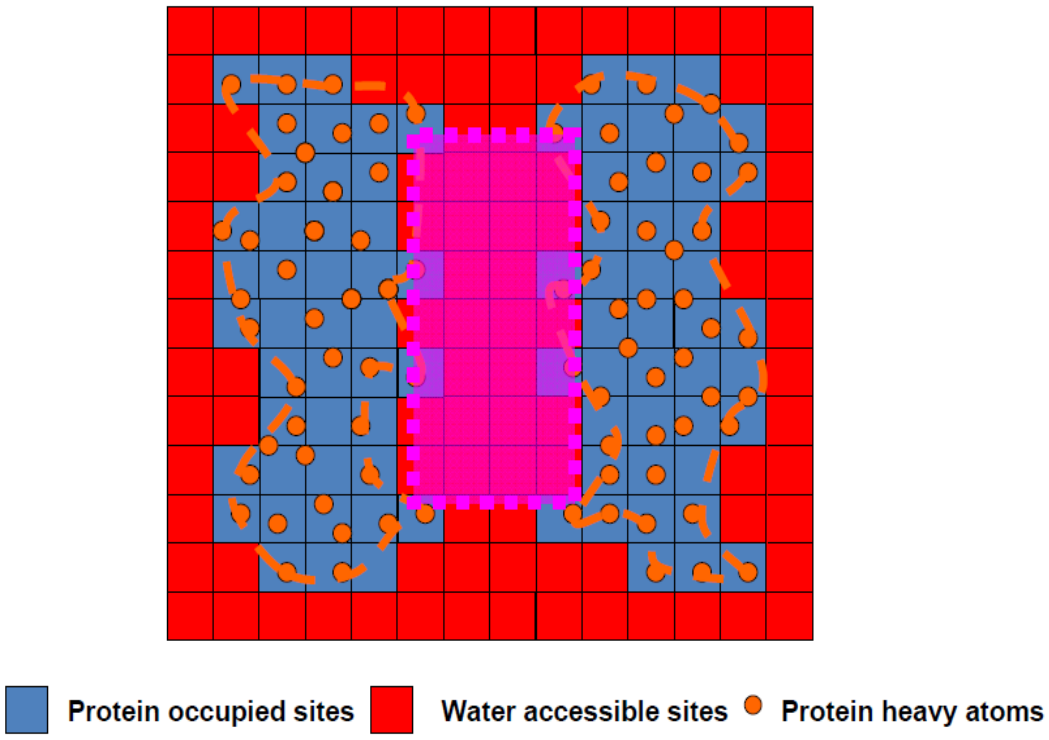


Figure 18. Coarse grained model for water confined between protein surfaces in a lattice gas system (projected to two dimensions), the magenta rectangular region represents the confined region being studied.

From Eq. 4-2, we can calculate the ratio between surface-water interaction and water-water interaction $\varepsilon_{s,i}/\varepsilon$ for all the protein occupied sites, based on which amino acid residues each site belong to (Table 5). From the relationship $\varepsilon_{s,i} = (\frac{1+\cos\theta_{c,i}}{2})\varepsilon$, $\theta_{c,i}$ represents the macroscopic contact angle on the surface compromising only one component *i*, we find that only those atoms from five out of twenty types of amino acids have an $\varepsilon_{s,i}/\varepsilon$ value less than 0.5, capable of inducing the evaporation transition (macroscopic contact angle of water on the surface is larger than 90°).[30, 33] Therefore, we would observe the evaporation transition only if there are these five types of amino acid residues together on the protein surface to form "hydrophobic patches".

As described in Section 1.2, when capillary evaporation is concerned, the "hydrophobic" term requires a more stringent definition than the one found in some literature.[24, 46] The arithmetic mean of all the values of ε_s is assigned to a site if it is within the 2*2*2 cubic region of more than one heavy atom in the system. In this treatment, if one hydrophobic residue is adjacent to hydrophilic residues, the lattice sites at the boundary of both residues would take the arithmetic mean of the ε_s of all the residues involved, and most likely will become hydrophilic ($\varepsilon_s/\varepsilon > 0.5$) It is consistent with the idea that hydrophilic site in a hydrophobic sea could turn the surface to be hydrophilic and the surface free energy in biological systems is non-additive. The hydrophobicity of different types of patches are actually anti-cooperative. [87, 122-123]

Amino acid	$\varepsilon_s/\varepsilon$	Amino acid	$\varepsilon_s/\varepsilon$
Asp	0.836	Pro	0.608
Asn	0.790	Trp	0.595
Glu	0.759	Cys	0.523
His	0.751	Met	0.515
Gln	0.736	Gly	0.5
Arg	0.728	**Phe**	**0.495**
Lys	0.690	**Leu**	**0.412**
Ser	0.690	**Ile**	**0.410**
Thr	0.647	**Val**	**0.410**
Tyr	0.614	**Ala**	**0.395**

Table 5. Surface interactions between solvent and protein occupied sites of different amino acids (the amino acid residues have an $\varepsilon_s/\varepsilon$ value less than 0.5 are in bold)

4.1.3 Monte Carlo simulations with Glauber & Kawasaki dynamics

As introduced in section 2.1, there are two types of dynamics to evolve the lattice gas system. The liquid water occupied sites in lattice gas can either jump with temperature dependent probabilities from one place to an empty neighboring site (Kawasaki dynamics[110], canonical ensemble), or evaporate and appear with certain probabilities (Glauber dynamics[109], grand canonical ensemble). The Kawasaki dynamics can be used to extract kinetic information from the systems, but due to its significantly slower speed, in our studies, we use Glauber dynamics only to determine the wettability of the confined region.

However, we are also interested in the qualitative predictions of the dynamics of the capillary evaporation events inside the protein systems at times, then a combination of Glauber and Kawasaki dynamics is needed in such studies to make close estimations of Monte Carlo simulation cycles to the diffusion constants of water. The Glauber dynamics is applied at the interface, where the lattice sites can interchange with the external reservoir, and the system is still open to bulk water and the pressure is moderate, while Kawasaki dynamics is applied away from the interface (within the system) to make mass transportations of water molecules in the confinements. The probability of acceptance follows the Metropolis Monte Carlo method[108]. We adapt an algorithm developed by Luzar and Leung[31, 34] that combines Kawasaki and Glauber dynamics together. By combining of Glauber and Kawasaki dynamics together, we could take advantage of both methods. The Glauber dynamics is applied at the interface (top part of Fig. 19), as the molecules will interchange with

the external reservoir, while Kawasaki dynamics is applied away from the interface (bottom part of Fig. 19). The numbers of Monte Carlo cycles are correlated with water diffusion time in bulk system and qualitative estimations of the timescales of evaporations events are discussed in section 4.2.5.

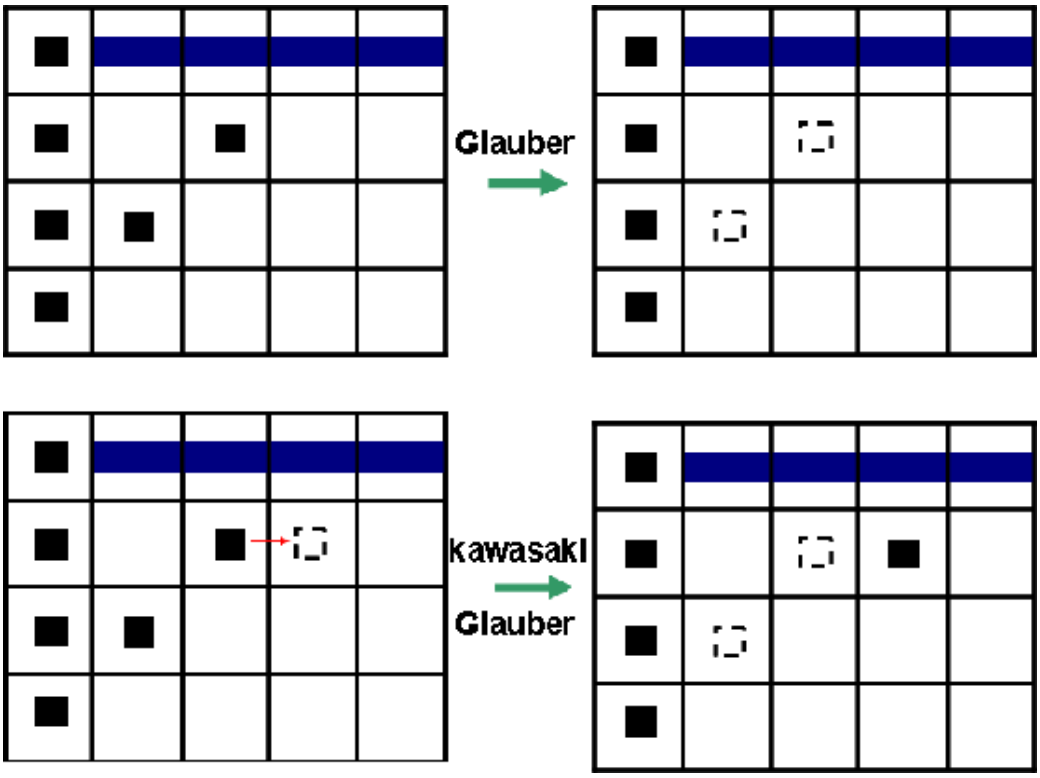


Figure 19. Lattice Gas System, (Top) Glauber dynamics and (Bottom) Kawasaki & Glauber dynamics. The blue stripes are fixed sites representing confinement. Filled squares are liquid sites and empty squares are vapor sites. The red arrow indicates the direction of a Kawasaki move.

4.2Results & Discussions

4.2.1 Melittin tetramer (pdb id: 2mlt)

Melittin is extensively studied by researchers, especially in biophysical community due to the presence of large hydrophobic patches in both of its dimers[21, 51-52, 66, 124], and it is one of the first natural proteins with capillary evaporation in its confinement demonstrated by Berne and coworkers in computer simulations.[52] In our study, the two dimers are extended by a distance of 4 Å along their centers of mass, and the coarse graining approach described in section 4.1 is applied on the system by transforming the coordinates of the pdb file into coordinates within a three dimensional lattice gas system. All water accessible sites in simulation box were filled initially with liquid water before starting the simulation. Firstly, the Monte Carlo (MC) method with Glauber dynamics is applied to the system and we record the number of water occupied sites during simulation and the ratio of average number of the water occupied sites in the confinement to total water accessible sites in the same region at the end of the simulation. The system is illustrated in Fig. 20. There is a strong drying transition observed in the confinement of melittin tetramer.

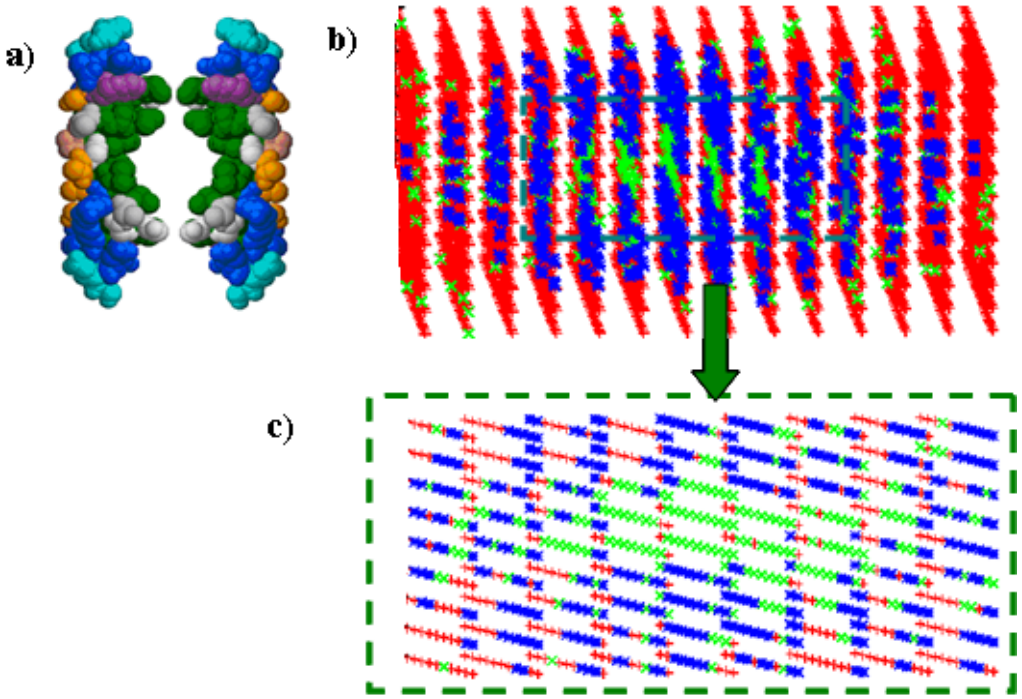


Figure 20. Evaporation in melittin tetramer system: (a) two melittin dimers (2MLT), green represents atoms from hydrophobic residues and other colors represent atoms from hydrophilic residues ; (b) the whole simulation box; water occupied sites are in red, protein occupied sites are in blue and empty sites are in green for visualization purpose. (c) The confined region highlighted in (b).

The ratio between the number of water occupied sites during the simulation and total number of water accessible sites is plotted in Fig. 21, during each Monte Carlo cycle, on average, every water accessible site is picked once for a trial move (either accepted or rejected). We should notice that in our model, we only take the nearest neighbor interaction into account (Eq. 4-1), by design, we would not observe a complete drying cavity due to the lack of long range interactions. There are always water occupied sites staying around the hydrophilic protein sites ($\varepsilon_s/\varepsilon > 0.5$), and we cannot exclude those sites while counting the number of water occupied sites in that particular region. Nevertheless, we are able to observe a substantial ratio of water occupied sites evaporated during the simulation, and we can conclude there is a capillary evaporation occurring in this confinement from the observation. In melittin tetramer system, we record a ratio of remaining water occupied sites to the total water available sites at around 25%, it indeed demonstrates a strong drying transition in that system. Normally, if the ratio is less than or around 50%, we can already observe the capillary evaporation, it is an arbitrary threshold that the water density inside the confinement reduces to about half of the bulk density.

Due to the resolution of the coarse grained model, it is hard to conduct the studies at molecular scales as in MD simulations. For example, we cannot evaluate the effects of single amino acid mutations to capillary evaporation in the melittin tetramer system, which have been successfully studied in the MD simulation.[52]

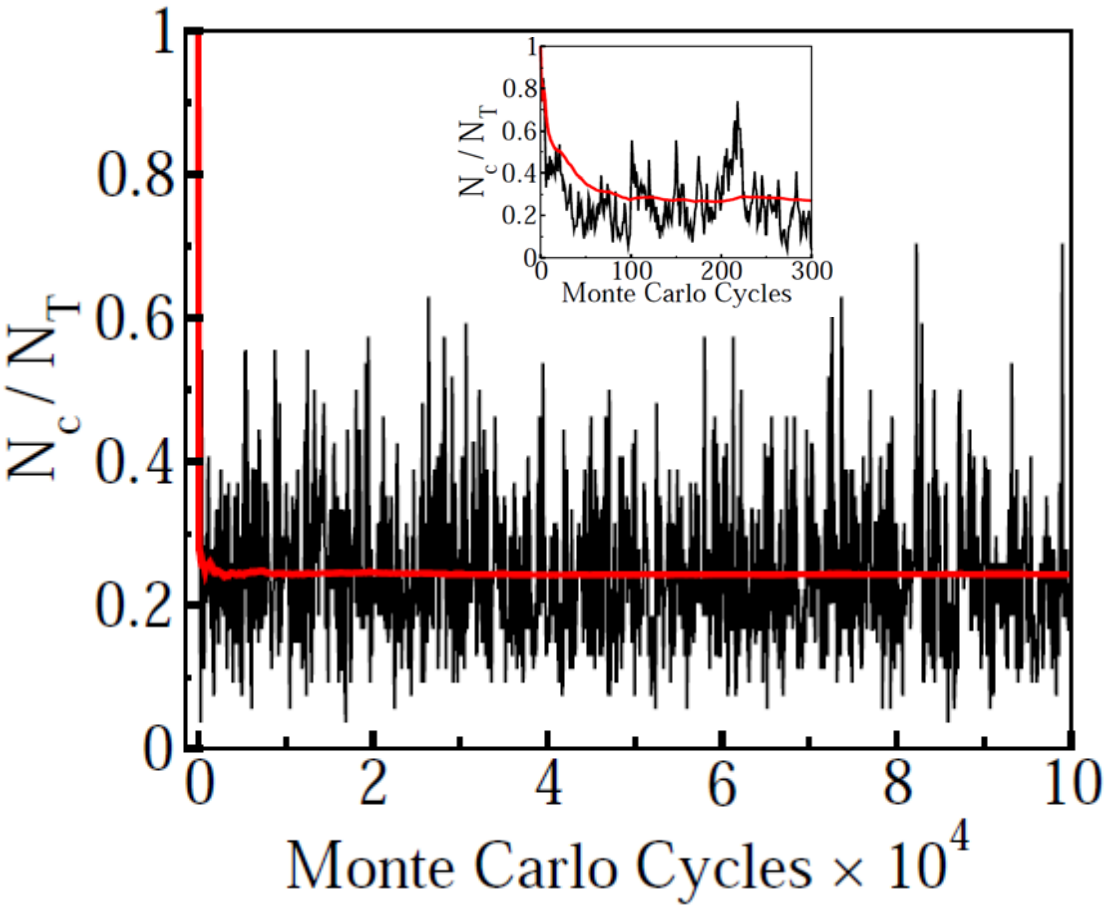


Figure 21. The ratio of number of water occupied sites (N_c) to total number of water accessible sites (N_T) in the confinement of melittin tetramer during the MC simulations. Red line represents the average ratio and the black line represents the instant ratio during the simulation. (Inset) The same ratio during the first 300 MC cycles.

The melittin tetramer is unique among many proteins with large hydrophobic surfaces. There is a strong drying transition in the confined region of the system. The relatively flat shape of the melittin dimers which form the confinement also facilitate the further studies to this particular protein.[66] With a similar procedure of Giovambattista *et al*[66], we were able to flatten the melittin dimer surface to study the chemistry part of the melittin dimer surface, the flattening procedures and contact angle measurements will be described in details in Chapter 5. Despite the fact that the entire flattened melittin dimer has a water contact angle at ~20^{o}, well below 90^{o}, indicating a very hydrophilic surface overall, we demonstrated the same dimer surface indeed has a central portion with water contact angles ~113^{o}, above 90^{o}, which is a necessary condition for water evaporation.[122] Our virtual contact angle measurements unveiled that both surface chemistry (hydrophobicity) and shape of the proteins are important in motivating the evaporation inside the melittin tetramer, especially, the evaporation transition only becomes possible when there is certain part of the surface that forms the confinement with water contact angle larger than 90 degrees, in agreement with the theoretical definition of capillary evaporation.[33] A snapshot of a sessile water nanodrop on top of a surface constructed by the central hydrophobic regions of melittin dimer surface is shown in Fig. 22.

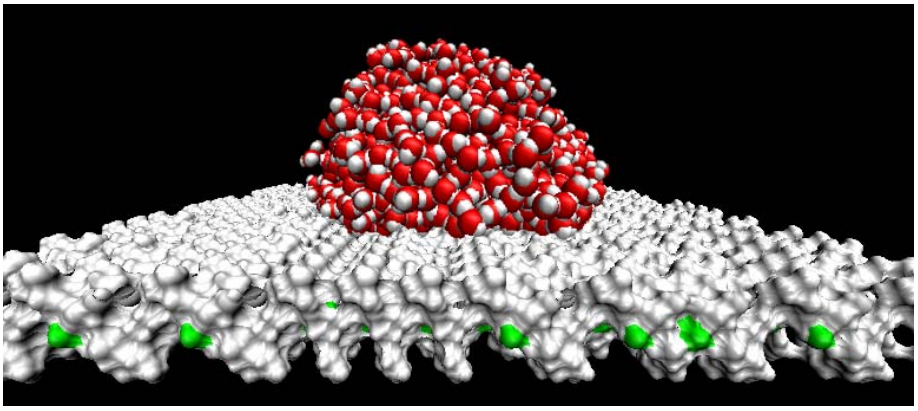


Figure 22. Nanodrop geometry used in water contact angle calculations on a surface constructed by central region of melittin dimer surface (details in text and in Chapter 5). The grey color represents surface from hydrophobic residues, green represents surface from hydrophobic residues. In the drop, red spheres represent water oxygen atoms and white spheres represent water hydrogen atoms.

4.2.2 Additional dimer, tetramer and multi-domain proteins

Additional proteins with confined region constructed by matched large and connected hydrophobic patches, according to the screening method reported by Hua *et al.*[51] are addressed by our methods in this section. Similarly, we chose the top candidates in each category including 10 tetramers, 20 dimers and 20 multi-domain proteins, and separate the two oligomers/domains which form the confinement by an additional 4~6 Å from their centers of mass distances. The results for the proteins identified to have evaporation transitions are summarized in Table 6.

Our model picked out the same set of proteins having demonstrated evaporation transitions in all atom simulations. Coarse-graining of protein and solvent qualitatively reproduces results from atomistic simulation. In Table 6, we have included several protein systems (pdb IDs: 1g5y, 1fe6, 1k2e, 1m4i) with high scores in the screening functions from Hua *et al*[51], but did not show capillary evaporations in their computer simulations. We recorded the remaining water ratios well above the 50% threshold, which means no evaporation with our model either, consistent with the simulation studies. The α_2D protein (pdb ID: 1QP6) was evaluated by Huang *et al*[125] in an earlier simulation study to show capillary evaporation with certain distances between its two polypeptide chains, we recorded a remaining water ratio ~54%, it is on the limit of evaporation in our model. A more complete list of protein systems we have studied are documented in Appendix E.

PDB ID	Category	N_i	N_r	$\frac{< N^2 > - < N >^2}{< N >}$	N_r/N_i%
1j2w	**tetramer**	**45**	**21.2**	**2.64**	**47**
1j3q	**dimer**	**49**	**23.7**	**1.00**	**48**
1f4n	**dimer**	**23**	**10.8**	**0.74**	**47**
1g6u	**dimer**	**28**	**15.4**	**1.18**	**55**
1d1g	**dimer**	**23**	**9.3**	**1.31**	**40**
1fsz	**Multi-domain**	**38**	**17.2**	**1.57**	**45**
1g5y	tetramer	57	53	1.77	93
1fe6	tetramer	41	31.8	0.21	78
1QP6 [125]	**dimer**	**38**	**20.4**	**1.33**	**54**
1k2e	dimer	44	28	0.76	64
1m4i	dimer	43	34.8	0.28	81

Table 6. Additional dimer, tetramer and multi-domain proteins identified for capillary evaporation. N_i represents the initial number of water occupied sites in the confinement, N_r represents the remaining number of water occupied sites in the confinement. The proteins shown capillary evaporation in computer simulations are in bold fonts. The fifth column represents a column similar to compressibility of water inside the confinement. The rows with bold fonts represent the protein systems identified to exhibit capillary evaporation in molecular simulations

4.2.3 Protein cavities and dry ligand binding sites

Ligand binding sites in protein complexes are related to the local hydrophobicity of the proteins. Ligands usually replace the water in the active binding sites, and if the binding sites are predetermined to be dry, the binding affinity will be large. Bovine β-lacto-globulin (BLG) was identified to have a dry cavity for ligand binding with both simulation and experimental proofs.[126] The ligand binding calyx is made up of 12 aliphatic residues and one aromatic residue.[127] The calyx has a sufficient amount of hydrophobic patches. By identifying the calyx in the coordinates, we used our coarse graining method to transform the protein into lattice sites, assigned all water occupied sites with liquid water inside the cavity initially, and evolved the system with the Monte Carlo method. The number of water occupied sites inside the cavities is tracked during the simulation. (shown in Fig. 23) In the end of the simulation, we record a ~30% of the water accessible sites are occupied by water, which in our model demonstrates a strong evaporation transition.

The dry ligand binding cavity does not only exist in Bovine β-lacto-globulin complex. Young *et al*[62] employed similar criteria as Hua *et al* [51] to screen ~1,800 protein-ligand binding systems and tested the top 15 candidates in all atom simulations. They demonstrated that six out of the fifteen proteins have a drying cavity in addition to the previous defined Bovine β-lacto-globulin.

We apply the same coarse-grained procedures as described in previous section to study the top candidates of the protein-ligand binding systems. The results for these

systems are summarized in Table 7. Our coarse grained model again identified the same proteins having dry cavities as in all atom simulations.

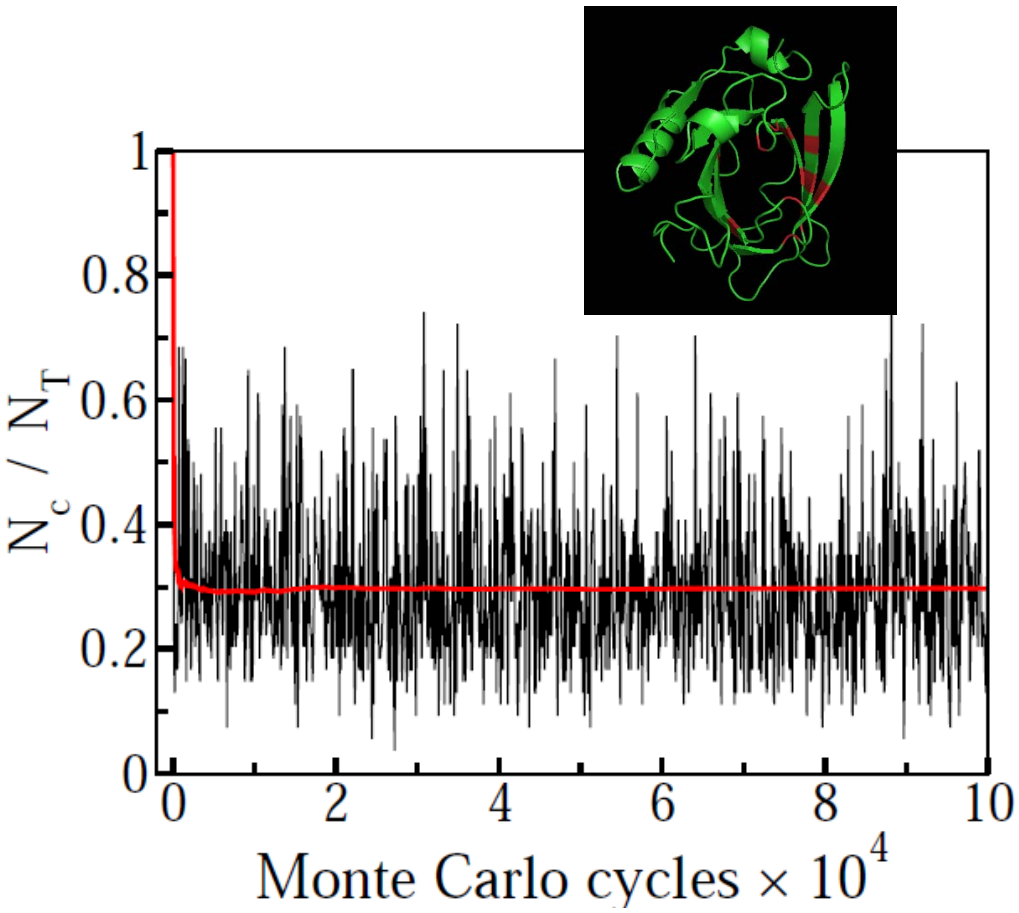


Figure 23. Number of water occupied sites in the confinement of Bovine β-lacto-globulin during the MC simulations. The ratio of number of water occupied sites (N_c) to total number of water accessible sites (N_T) in the confinement of melittin tetramer during the MC simulations. Red line represents the average ratio and the black line represents the instant ratio during the simulation. (Inset: ribbon representations of Bovine β-lacto-globulin protein, red color strips indicate the position of the cavity.)

PDB ID	Ligand	N_i	N_r	$\frac{< N^2 > - < N >^2}{< N >}$	N_r/N_i%
1e7g	**MYR_A1008**	**25**	**10.7**	**1.61**	**43**
1y9l	**UND_150**	**34**	**16.7**	**2.50**	**49**
1wbe	**DKA_A1001**	**24**	**11.9**	**1.49**	**50**
1wub	**OTP_1001**	**67**	**24.9**	**1.67**	**37**
1rbp	**RTL**	**49**	**26.1**	**2.88**	**53**
1lid	OLA_132	55	43.5	0.68	79
1cvu	**ACD**	**46**	**20.6**	**2.24**	**45**
1dbj	AE2	37	31.9	0.30	86
1ure	PLM_132	20	13.5	0.71	68
1g74	OLA_132	48	36.9	0.82	77

Table 7. Additional protein-ligand systems tested in coarse-grained model. N_i represents the initial number of water occupied sites in the confinement, N_r represents the remaining number of water occupied sites in the confinement. The proteins shown capillary evaporation in computer simulations are in bold fonts.

4.2.4 Dynamics information

Within our lattice gas model, we are able to obtain qualitative dynamic information of the drying transition from an algorithm developed by Luzar and Leung[34] to combine Glauber[109] and Kawasaki[110] dynamics in our simulation.

To study the relaxation time of fluctuations at equilibrium, we determine the density time correlation functions,

$$C(t) = \frac{< \delta N(0) \cdot \delta N(t) >}{< \delta N(0)^2 >} \qquad (4\text{-}3)$$

where *N(t)* is the number of water occupied sites inside the confined region at time t (represent by the number of Monte Carlo cycles). The density time correlation function (Eq. 4-3) with both Glauber and Kawasaki dynamics for the melittin dimer confined region after equilibration is plotted in Fig. 24.

By calibrating the number of Monte Carlo passes in Kawasaki dynamics with the water diffusion constant[34], with the relation $6Dt = |\Delta R^2|$ at long time, where D is the diffusion constant of water, and ΔR^2 is the mean square displacement during the time t, we extract a rough estimation of the time scales of the fluctuation. During the short time scale, the relaxation is not linear in the semi-log plot (Fig. 24 top), we extracted the dynamics by integrating the function C(t) for the first 1,000 MC cycles, and the relaxation time is ~100 Monte Carlo cycles in Kawasaki dynamics, which is approximately equivalent to ~5 ps in real time in this system.

On the other hand, the MC simulation with Glauber dynamics gives a much faster decay (Fig 24 bottom) in ~4 MC cycles by integrating the C(t) for first 1,000 MC cycles. Here, we compare the two dynamics and find the Kawasaki dynamics runs

much slower in timescales than the Glauber dynamics, for the small system, $25 * \tau_{Glauber} \sim \tau_{Kawasaki}$ according to the relaxation of density time correlation functions calculated in two methods, coincides with the previous studies by Leung and Luzar[31, 34].

It is actually impractical to use Kawasaki dynamics only when all of the sites are filled in with water initially. Therefore, we studied the actual capillary evaporation time in Glauber dynamics only, and the relation between the time scales in Kawasaki dynamics and Glauber dynamics has been applied. From Fig. 21, the estimated evaporation time of confined water is ~200 MC cycles with Glauber dynamics, is equivalent to ~5,000 MC cycles with Kawasaki dynamics, and ~250 ps in real time, which is of the same magnitude as that determined by all-atom Molecular Dynamic simulations.[52]

We also tried to relate the "pseudo mean square displacements" in Glauber dynamics directly with the water diffusion. We counted the number of Monte Carlo passes needed for one accepted deletion and one sequential accepted addition in the simulation as the time t, and recorded the distance between the deletion site and the addition site as the displacement ΔR^2, and from the diffusion equation, interestingly, we get a similar time scale as we determined from a combination of Kawasaki and Glauber dynamics together.

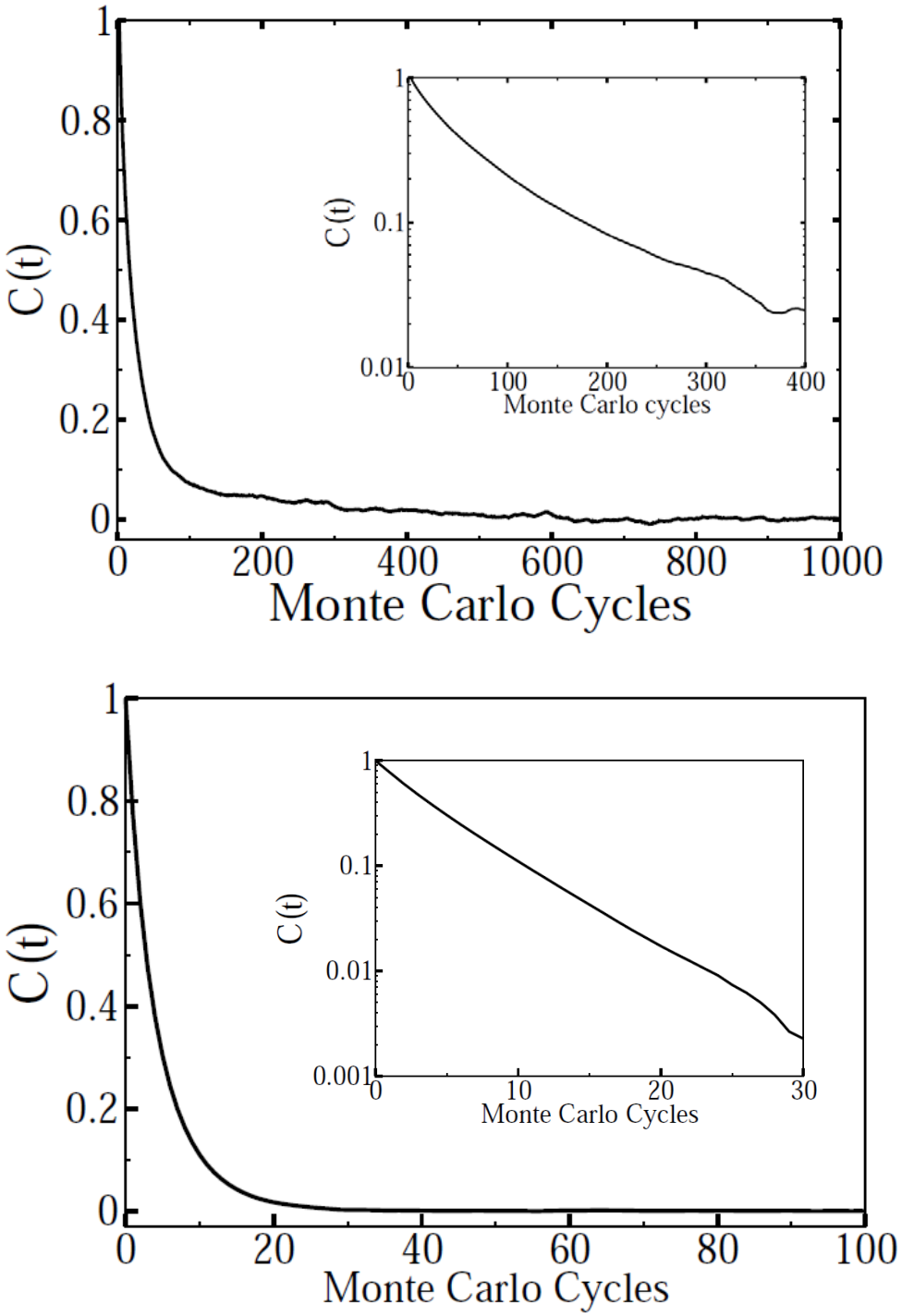


Figure 24. Density time correlation function of water occupied sites inside the confinement of melittin tetramer (pdb id: 2mlt) with Kawasaki dynamics (top) and Glauber dynamics (bottom). (Insets: A semi-log plot for the correlation function during the short timescales)

Furthermore, the density time correlation function for water occupied sites in the binding sites of Bovine β-lacto-globulin calculated in MC simulation with Glauber dynamics is plotted in Fig. 25. The relaxation time is ~10 MC cycles by integrating the first 200 MC cycles, and it corresponds to a real time less than ~10 ps in this particular system, coincides with the findings in all-atom simulation[126] that the water molecules came out of the region within the equilibration stage (if they were initially put inside the cavity).

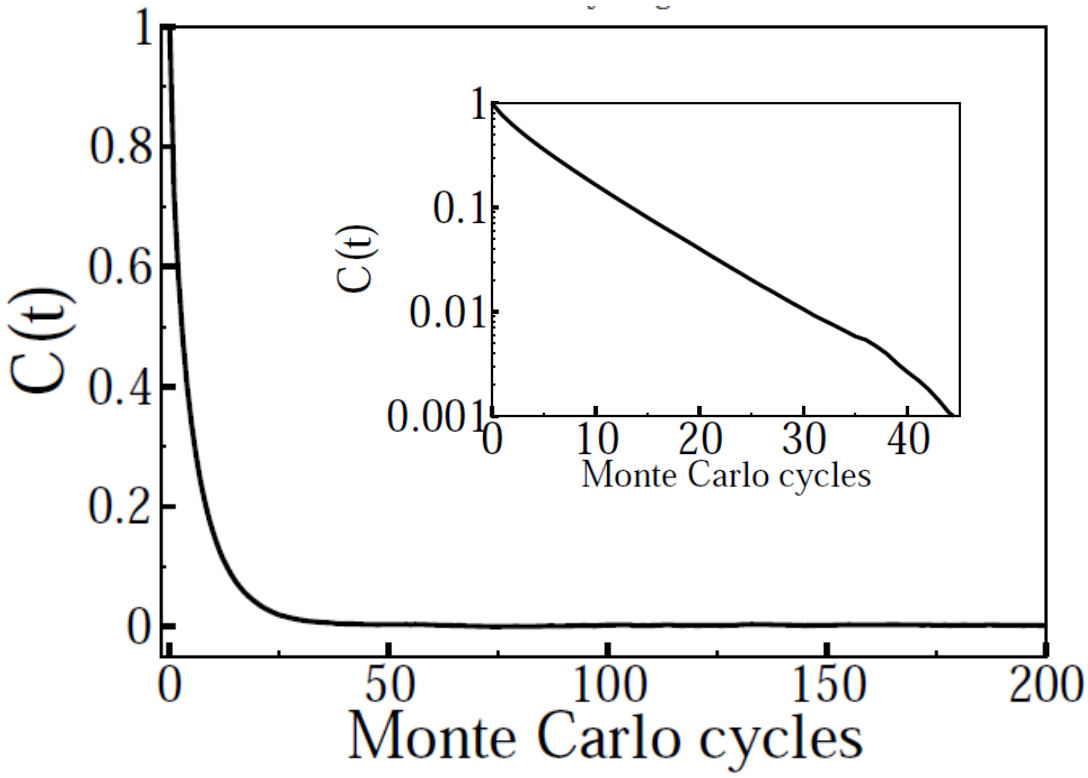


Fig. 25 Number correlation function of water occupied sites inside the cavity of Bovine β-lacto-globulin (pdb id: 3blg). (Inset: the semi-log plot for the correlation function at short times.)

4.3 Conclusions

Coarse graining procedures are among the choices to study large systems, e.g. proteins. The idea is to use cruder models for atoms or atomic groups, and use averaged interactions between these groups. Previous successes in coarse grained model predictions of water behavior in the confinements of model hydrophobic surfaces [30-31, 34], hydrophobic tubes[40] imply the thermodynamics of water can be captured without detailed atomistic information (such as empirical force fields). In this study, we were able to predict thermodynamic behavior of water confined in protein systems (dry or wet), as well as qualitative information about the dynamics of the drying transition when it happens, using a computational-efficient coarse grained model. Although these are qualitative estimations for the timescale of the drying transition, we were able to relate our Monte Carlo (MC) simulation results with the all-atom Molecular Dynamics (MD) simulation results at a much less computational expensive way. We also interpreted the origin of the evaporation inside the melittin tetramer system in terms of the microscopic analog of macroscopic contact angles[37]. We can conclude that for evaporation transition, what really matters is the protein surface topology and the macroscopic thermodynamic properties of the solvent, the atomistic details such as empirical force fields are not crucial in the model.

As pointed out previously by different research groups,[8, 66] the water drying transition is not merely determined by the chemistry or the geometry of the protein system, but the combination of these two factors. While the simple screening method [51], based on the hydrophobicity of amino acid residues exposed on the surfaces, left

out the geometric factor, we used a slightly more detailed model but incorporated both the protein shape and the protein-water interactions. We demonstrated that we can predict the dry/wet confinements and cavities in protein systems when only the thermodynamics are important. In addition, we were able to obtain the rough estimation for the dynamic information as well by combining two techniques in the Monte Carlo simulations and relate the Monte Carlo steps to water diffusion time.

Our model is especially useful in a quick test of a protein system to determine whether the capillary evaporation is possible. Combined with the screening method developed by Berne and coworkers[51], our method can serve as an intermediate step between the initial screening and extensive studies for the molecular details (i.e. single mutation studies of amino acid residues[52]). The coarse graining model is sufficient when one is only interested in the thermodynamics property of water inside the confinement system. It captured the basic physics of the water evaporation processes in these confined or cavity regions.

Chapter 5. Wetting behavior of heterogeneous surfaces

Surface free energy of a chemically heterogeneous surface is often treated as an approximately additive quantity through Cassie equation.[71] However, deviations from additivity are common and molecular interpretations are still lacking. In current chapter, we use molecular simulations to measure the microscopic analogue of contact angle, θ_c, of aqueous nanodrops on heterogeneous synthetic and natural surfaces as a function of surface composition. The synthetic surfaces are comprised of graphene functionalized with prototypical nonpolar and polar head group: methyl, amino and nitrile. We demonstrate positive as well as negative deviations from the linear additivity. We show the deviations reflect the uneven exposure of mixture components to the solvent and the linear relation is recovered if fractions of solvent accessible surface are used as the measure of composition.

As polarity fluctuations on the surface intensify, the linear relation can no more be obtained. Protein surfaces represent such natural patterned surfaces, also characterized by larger patches and roughness. Our calculations reveal strong deviations from linear additivity on a prototypical surface comprising surface fragments of melittin dimer. The deviations reflect the anti-cooperative influence of polar patches, their preferential wetting, and changes in the position of the liquid interface above hydrophobic patches. Since solvent-induced contribution to the free energy of surface association grows as $\cos\theta_c$, deviations of $\cos\theta_c$ from the linear relation represent a direct measure of nonadditivity of surface biointeractions.

This chapter is organized as follows: in the first two sections we describe model systems with patches sizes at molecular scale (synthetic surfaces) and nanoscale (natural protein-like surfaces), and present results for wetting surface free energies as a function of surface composition. In Discussion we show the impact of solvent accessible surface areas replacing the mole surface fractions on results presented in previous sections. The conclusions follow.

5.1 Synthetic surfaces with patches at molecular scale

5.1.1 Model surfaces

We start with surfaces constructed by planting molecular groups on molecularly smooth surfaces. The model surfaces are designed as functionalized graphene sheets with surface groups of different polarities, previously considered in studies of wetting on mixed SAMs. [74-77, 83-84, 128] We choose this rigid substrate to focus on the influence of mixed surface chemistry, decoupled from topography changes involved in wetting of flexible SAM deposits. Carbon atoms, packed into hexagonal graphene lattice, are characterized by Lennard-Jones parameters (summarized in Table 8) adjusted following Werder *et al* [129] to give contact angle of ~ 108°, close to experimental values of graphene [130-131] and hydrocarbon. Prototypical surface groups, $-CH_3$, $-NH_2$ and -CN, are planted on the lattice with density (~21 $Å^2$ per group) comparable to that achieved in SAMs experiments. [74-77]

While we make no explicit assumptions about the pH of the drop, amino groups are presumed to remain nonionized. For a small, approximately 2000 water molecule

nanodrop, ionization of a single NH2 group under the drop via NH_2+H_2O -> NH_3^+ + OH^- elevates pH from initial value of around 7 to more than12, way above the group pK. Ionization is therefore not significant at given conditions. Force fields we use are given in Table 8. We compared results obtained by using Lennard-Jones parameters and partial charges of surface groups from three different force fields, CHARMM v27[96], Amber Parm-94[95] and OPLS-AA[97].

Atom	σ_{A-O} (Å)	ε_{A-O} (kcal/mol)	Partial charge (e)
C(Graphene)[129]	3.214	0.0361	0
OPLS-AA[97]			
C(-CH_3)	3.5	0.066	-0.18
H(-CH_3)	2.5	0.03	0.06
N(-NH_2)	3.25	0.17	-0.7
H(-NH_2)	-	-	0.35
C(-CN)	3.3	0.066	0.46
N(-CN)	3.2	0.17	-0.46
Amber[95]			
C(-CH_3)	3.816	0.1094	-0.18
H(-CH_3)	2.974	0.0157	0.06
N(-NH_2)	3.648	0.17	-0.84
H(-NH_2)	1.2	0.0157	0.42
CHARMM[96]			
C(-CH_3)	3.670	0.08	-0.27
H(-CH_3)	2.352	0.022	0.09
N(-NH_2)	3.296	0.2	-0.80
H(-NH_2)	0.4	0.046	0.40

Table 8. Force fields used for surfaces head groups

5.1.2 Water Contact Angle Calculations

To estimate the deviations from linear additivity, we study wetting free energies, quantified in terms of microscopic analogues of water contact angle[67]. We use water drops comprising 2,000 SPC/E[98] water molecules to measure the microscopic analog of the macroscopic contact angles on the surface. The computer simulations are carried out by the LAMMPS molecular dynamics package[102] in NVT ensemble with Nose-Hoover thermostat.[132] We use periodic boundary conditions and extra space in z direction (300 Å) to avoid interact ions with system's images in z direction. The long-range electrostatic interactions are treated by SPME (Smooth Particle Mesh Ewald Sums) method.[133] Each system is equilibrated for at least 500 ps, and typically run by additional 5 ns to secure contact angle convergence. For very hydrophilic surfaces, production runs of up to 10 ns were made. We record the trajectories with 1 ps interval. The contact angles are computed from circular drop contours extrapolated to substrate surface. We adopt the drop analysis technique of de Ruijter *et al*[134] further developed by different groups[129, 135]

In each recorded MD simulation configuration, cylindrical binning has been introduced to get the water drop isochors, and the surface normal through the center of mass of the droplet is used as reference axis. The bins have a height of 0.5 Å and the radial bin boundaries are located at $r_i = \sqrt{i\delta A/\pi}$ for i = 1,…, N_{bin} with a base area per bin of $\delta A = 80\,\text{Å}^2$, so that all bins contain the same volume. From such a profile, the equimolar dividing surface is determined within every horizontal layer of the binned drop, and a circular fit through these points is extrapolated to the reference plane to

measure the contact angle. The z axis position of the reference plane is determined by averaging the effective heights of each atomic group.

For hydrophilic surfaces, the drop becomes asymmetric, and the spherical shape of the drop may not be guaranteed. We modify our technique by adapting the method developed by Toxvaerd and coworkers[136]， where we determine the center of mass of each horizontal layer individually, and average over all the configurations first, then calculate the equimolar dividing surface accordingly. The rest of the procedures stay the same as described above. However, this treatment gives the contact angle of our hydrophilic surfaces within the error bars of the results from original method. To be consistent, we keep the traditional method for all surfaces. Typical drop contour profiles of different types of surfaces are shown in section 5.2.1. (Fig. 36)

To check for possible droplet size dependence, we also performed test runs with water drops containing 4000 water molecules on a four times larger surface, and found no significant differences in contact angles. Further, for methyl-covered substrate, we calculated the wetting free energy of semi-infinite surfaces using planar confinement geometry with lateral periodicity. We used Grand Canonical Monte Carlo simulations to determine the amount of water between the interfaces. Thermodynamic integration was carried out to obtain the wetting free energy $\Delta\gamma$, and water surface tension, γ, was obtained by the pressure tensor method[137]. The Young contact angle $\cos^{-1}(-\Delta\gamma/\gamma)$ agreed within 3-5^{o} with the value obtained from the nanodroplet geometry. The details of these calculations are discussed in Appendix F.

5.1.3 Simulation Results

We list the calculated contact angles on pure surfaces (functionalized by groups of a single type) in Table 9; we also include experimental water contact angles on SAMs with the same end groups.[74, 77, 138, 139] In separate calculations, SPC[100] and TIP3P[99] water models revealed no significant differences in contact angles from those observed with SPC/E[98] water model.

The water contact angles on the hydrophobic surfaces (-CH_3) are robust with various force fields, while contact angles on hydrophilic (-NH_2) surfaces appear sensitive to the force field choice. In particular, with OPLS-AA force fields, the contact angle of -NH_2 surface is too low compared to the experimental results. This is due to the lack of Lennard-Jones parameters on the hydrogen atoms of -NH_2 groups, allowing water molecules to approach -NH_2 groups too closely.

In addition, the line tension effects are evaluated for the nanoscopic droplets to compare to the macroscopic observed contact angles θ_∞ based on the empirical relation

$$\cos\theta = \cos\theta_\infty - \frac{\tau}{\gamma}\frac{1}{r_B} \qquad (5\text{-}1)$$

where τ is the line tension, γ is the water/air interfacial tension and r_B is the contact radius of the droplet. The line tension is predicted to be at the value of 10^{-10} N in literature[69, 83, 129], the hydrophobic contact angles are overestimated by ~5 degrees and the hydrophilic contact angles are underestimated by ~10 degrees from the molecular simulation, which coincides with our observed contact angles comparing to the those reported in experiments (Table 9). [74, 77, 138, 139]

	OPLS-AA	CHARMM	Amber	Experimental (refs)
$-CH_3$	109 °	106 °	109 °	107° [138]/112° [139]
$-NH_2$	18°	32 °	36°	43° [138] / 50° [77]
-CN	50°	-	-	61° [74] /63° [139]

Table 9. Contact angles on homogeneous surfaces

In the rest of this chapter, we focus on the results with Amber force fields when $-NH_2$ groups are involved, because they reproduced best the experiment on $-NH_2$ functionalized surfaces. Parameters for the nitrile group are available only in OPLS-AA force fields and we use these when -CN moieties are involved.

We mix the small groups on the surfaces with a bias toward alternations, minimizing the number of contacts between surface groups of the same type. The cosines of contact angles on mixed methyl/nitrile surfaces shown in Fig. 26 Left reveal positive deviations from the linear dependence on the fractional area of methyl groups, f_{CH3}, similar to those observed in experiments with methyl/nitrile terminated SAMs deposits[74]. The plot of cosine of contact angle vs. f_{CH3} on methyl/amino surfaces (Fig. 26 Right) reveals a distinctly different composition-dependence. Here,

the cosine of contact angle deviates from linear additivity in the negative direction. When f_{CH3} exceeds ~30%, the surfaces are more hydrophobic than predicted by the Cassie equation, Equation 1-3.

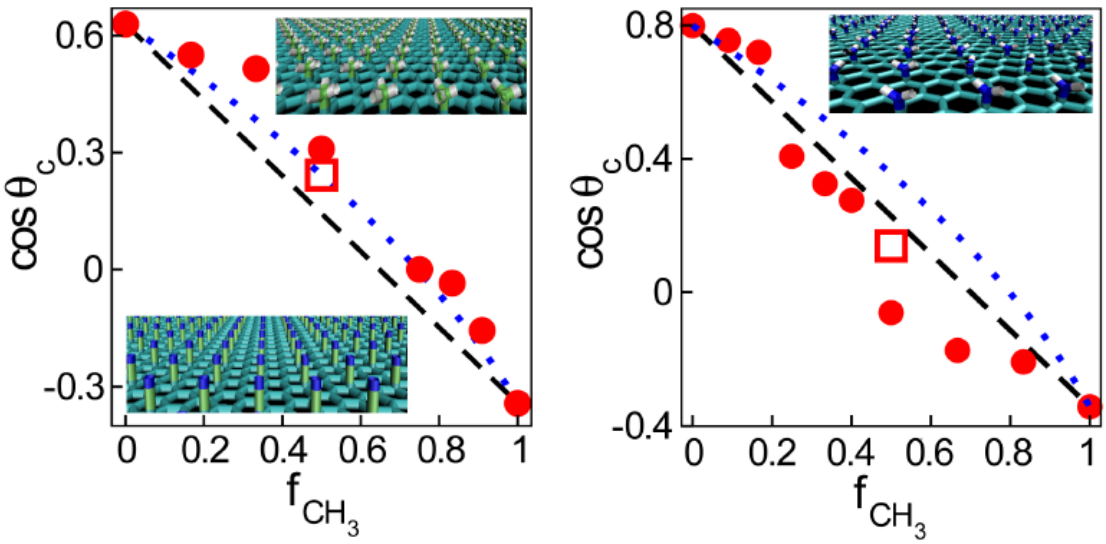


Figure 26. Cosine of contact angle on mixed -CH_3/-CN (left) and -CH_3/-NH_2 (right) surfaces as a function of fractional area of methyl groups. Solid circles: simulation results, Open squares: results with larger patches (4x4 head groups) black dashed line: linear additivity approximation due to Cassie (Eq. 1-3), blue dotted line: approximation from Israelachvili and Gee. (Eq. 1-4) [73] Error bars are comparable to the size of the symbols. Insets: Model surfaces covered with -CH_3 groups (upper left), -CN groups (lower left), and -NH_2 groups (upper right) The underlying graphene surfaces are shown in cyan color, C atoms in small groups are lime, N atoms are blue, and H atoms are white.

Actually, notwithstanding the differences in parameters of different empirical force fields, for all three force fields we find similar *deviations* from linear additivity of $cos\theta_c$ on mixed -NH_2/-CH_3 surfaces, showing that the interfacial free energy additivity itself is virtually insensitive to force field choice (Fig. 27).

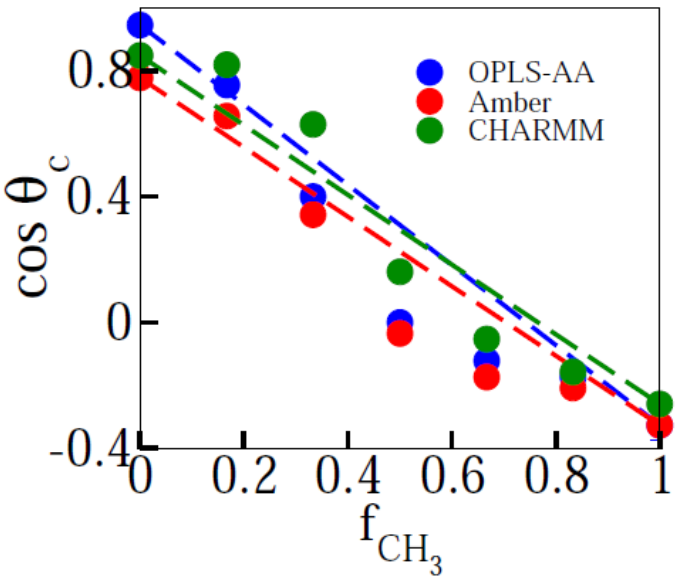


Figure 27. Cosine of contact angles on mixed -CH_3/-NH_2 surfaces as a function of the mole fraction of substrate surface covered by -CH_3. The dashed lines are the Cassie prediction[71] with all the force fields we tested in our studies.

Experiments on methyl/amino terminated SAMs deposits featured nearly linear dependence with slight positive deviations in $\cos\theta_c$ [77]. We attribute the excessive hydrophobicity on functionalized graphene to partial shielding of amino groups by bulkier methyl groups. In SAM deposits, on the other hand, chain flexibility generally facilitates exposure of polar moieties[83]. For example, *dry* OH-terminated SAM chains extend to a lower height than methyl-terminated ones, however, under the drop they outstretch their nonpolar counterparts to facilitate wetting [83]. Israelachvili and Gee [73] approximation predicts positive deviations. As illustrated in Fig. 26, this approximation shows a qualitative agreement with experiment and simulation on methyl/nitrile surfaces while by design it cannot capture negative deviations.

Negative deviations of $\cos\theta_c$ from linear dependence have also been observed in experiments with SAMs of octadecylphosphonic acid[82], however, the lack of microscopic insight into experimental surfaces precluded molecular interpretations. We take advantage of our simulation setup to look into the details of water structure next to model surfaces to unveil essential differences between mixed and homogeneous systems (end-points in Fig. 26).

In Fig. 28, we present the running coordination number per unit area, $N_c(z)$, $N_C(z) = \rho_b \int_0^z g_w(z')dz'$ for water molecules next to solvated surfaces; here, ρ_b is the bulk number density of water, z is the distance from the substrate carbon atom layer and $g_w(z)$ the wall/water distribution function. We consider pure surfaces and equimolar mixtures (f_{CH3}=50%). On pure hydrophilic surfaces, a fraction of water molecules is shown to penetrate partially between -NH_2 or -CN groups, approaching

the underlying carbon layer. Substitution of only half of -NH_2 or -CN groups by -CH_3 groups suffices to exclude water molecules from the first solvation layer. As shown in the inset of Fig 28, the net exclusion from the -NH_2/-CH_3 mixture is about twice bigger than from the -CN/-CH_3 one.

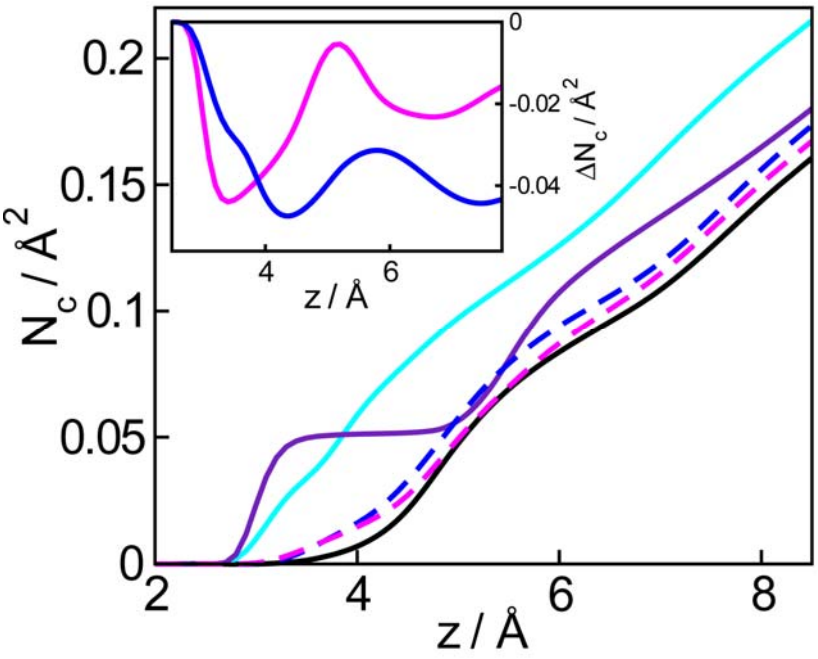


Figure 28. Running coordination numbers of water on surfaces with different groups. Inset: the differences in running coordination numbers between mixed -NH_2/-CH_3 surface and pure -NH_2 surface (blue), and between mixed -CN/-CH_3 surface and pure -CN surface (magenta).

Figs. 29 and 30 illustrate the profiles of the average number of hydrogen bonds per water molecule, $n_{HB}(z)$, and the average orientation of water dipole, $\mu_{\perp}(z)$, along z direction.

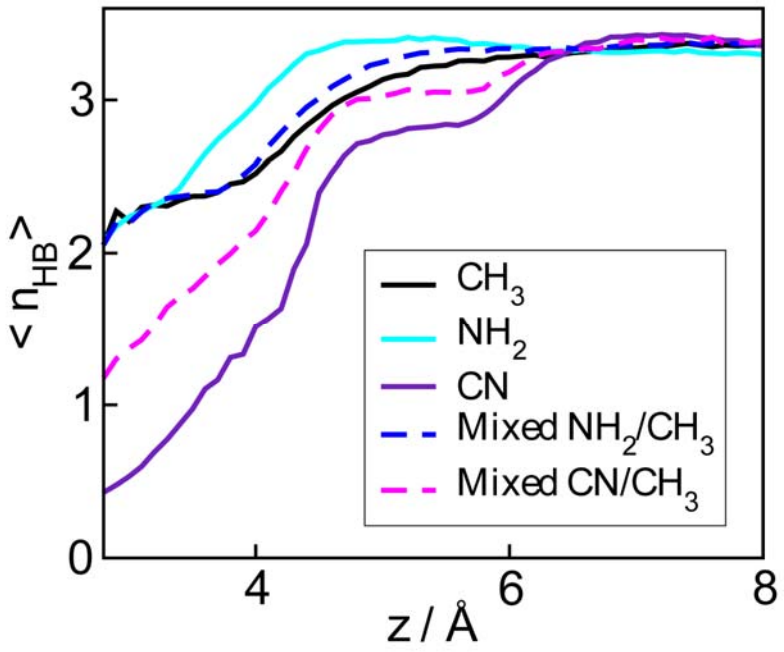


Figure 29. Average number of hydrogen bonds per water molecule as a function of the distance from the functionalized graphene surface.

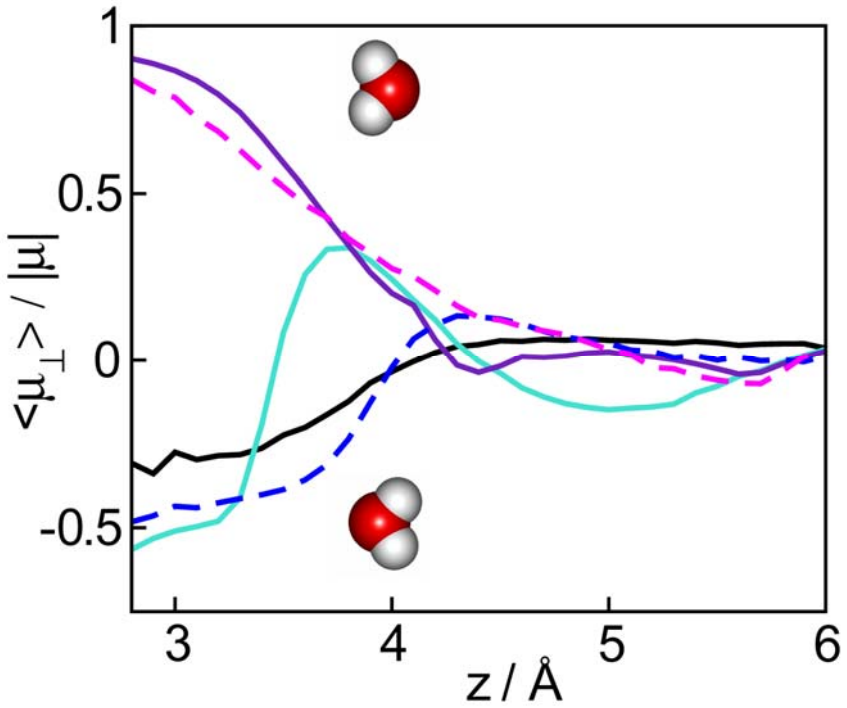


Figure 30. Averaged dipole orientations of water molecules near the functionalized substrate surface. (The surface is on the left) The sketches of water molecules show preferred molecular orientations near pure -CN (top) and -NH_2 (bottom) surfaces.

A significant reduction in $n_{HB}(z)$ is observed near the pure -CN surface (Fig. 29) as orientational ordering of water dipoles by strongly polar -CN groups interferes with hydrogen bonding (Fig 30). A weaker influence is induced by $-NH_2$ groups. When $-CH_3$ groups are planted on the $-NH_2$ surface, profiles $n_{HB}(z)$, and $\mu_{\perp}(z)$ approach the profiles characteristic of pure methyl-covered surfaces. At equimolar $-NH_2/-CH_3$ composition, the change in both quantities is nearly completed. At 50% methyl/nitrile surface, on the other hand, water properties remain under strong influence of nitrile groups. This picture is reinforced by angular distribution functions of water dipoles shown in Fig. 31. The dipole distribution of water molecules near hydrophobic pure $-CH_3$ surface is in good agreement with simulations on SAMs surface with the same head group[140], and the polarity of groups -CN and $-NH_2$ are large enough to affect the water dipole orientations near these surfaces.

The distribution of water dipole orientations next to the surface covered by equimolar mixture of $-CH_3$ and $-NH_2$ groups is very close to that of pure $-CH_3$ surface and different from that at pure $-NH_2$ surface. While the peaks at the mixed $-CN/-CH_3$ surface are shifted away from the peak position of the pure -CN surface, the average orientation appears to be similar in both cases, in contrast with the behavior of the $-NH_2/-CH_3$ mixed surfaces.

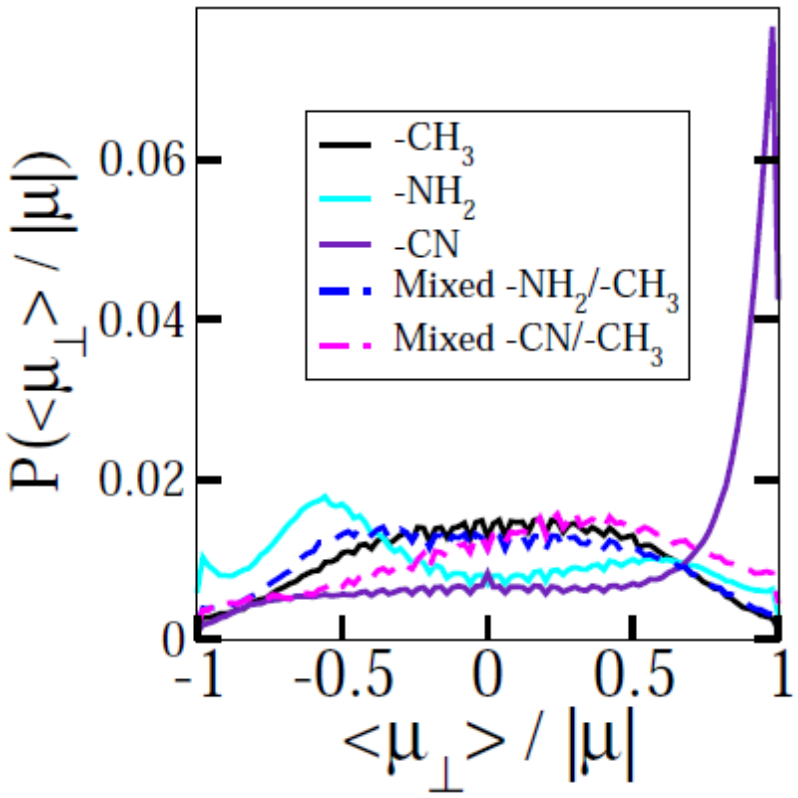


Figure 31. Distribution of dipole orientation at the graphite surfaces functionalized by different head-groups.

In the bulk, water forms a roughly tetrahedral hydrogen bonding network, with each water molecule averaging about 3.4 hydrogen bonds to its neighbors. At the interface, the hydrogen bonding network is significantly perturbed.[11, 16, 18] Concomitant disruption of tetrahedral coordination at our model surfaces is described in Fig. 32.

The perturbations of the hydrogen bonding network can be quantified in terms of the ability of water molecules to maintain tetrahedral coordination. In Fig. 32, we plot the distributions of the O-O-O angles of water triplets near various surfaces.

The distribution is defined by Equation 5-1, similar as in ref [141].

$$P\{\cos\theta\} = a\sum_{i=1}^{n_i}\sum_{j=1}^{n_{ij}}\sum_{k<j}\delta\left(\frac{r_{ij}^{\,2} + r_{ik}^{\,2} - r_{jk}^{\,2}}{2\,|\,r_{ij}r_{jk}\,|} - \cos\theta\right) \qquad (5\text{-}1)$$

P is the probability, a is the normalization factor, $r_{\alpha\beta}$ is the distance between water molecules α and β.

Here, the sum goes over all water oxygen triplets located within the interface and satisfying O-O distances that permit H-bond formation. In pure water, the distribution P($cos\theta_{o\text{-}o\text{-}o}$) features a broad peak around $\cos\theta_{o-o-o} \sim -0.3$ when water molecules are forming tetrahedral networks, while a small peak at half lower angle corresponds to interstitial water molecules not in tetrahedral formation[142]. At the surfaces, the tetrahedral network is partially disrupted with shifts of the large peaks to lower angles (less negative $\cos\theta_{o\text{-}o\text{-}o}$); at the same time the interstitial water peaks rise, indicating more water molecules lack tetrahedral coordination.

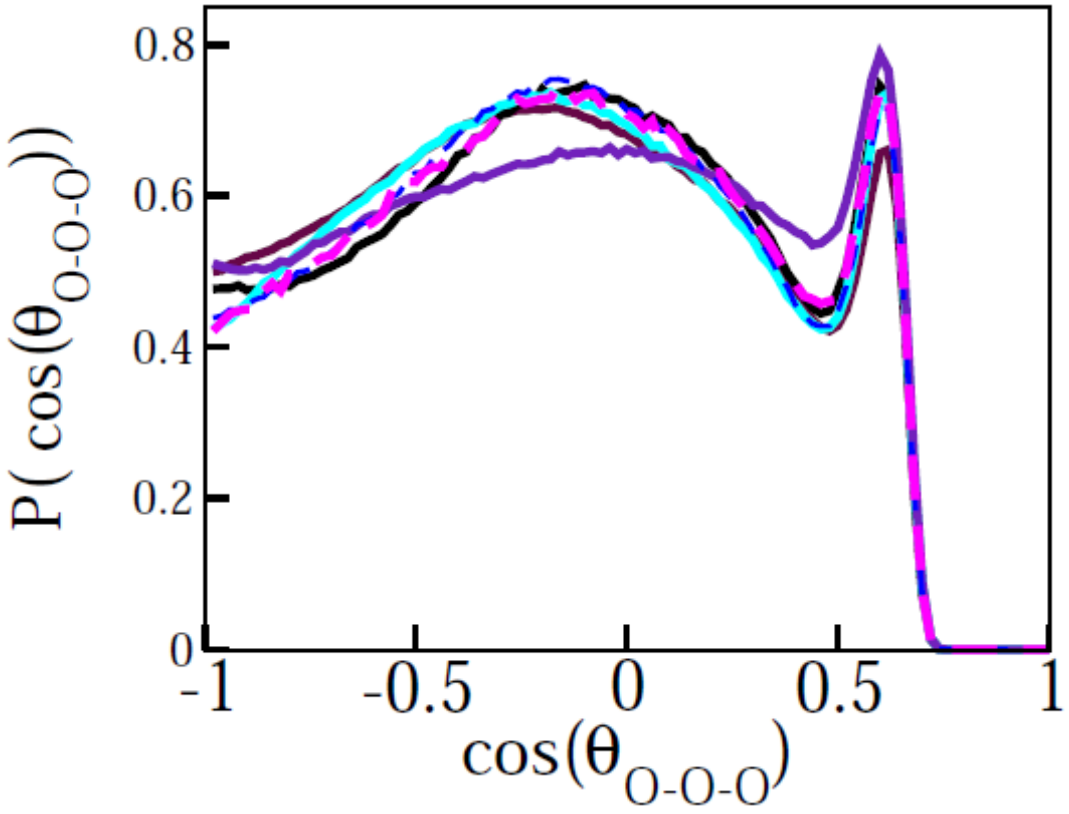


Figure 32. Distribution of O-O-O angles in triplets of water molecules (defined in equation 5-1).

Among all types of surfaces we study, the O-O-O angles' peak shift is most pronounced at the surface covered by -CN groups (indigo curve in Fig. 32), consistent with our observation of the reduction in the average number of hydrogen bonds of water molecules near that surface (Fig. 29). We note that, by limiting the statistics to water triplets with O-O distances within the hydrogen bond length of up to 3.5 Å, a larger fraction of water molecules outside the hydrogen bonded network remain unaccounted for. Inclusion of these water molecules would enhance the measured extent of the disruption of tetrahedral coordination in interfacial water.

Overall, the rapid change in surface character, observed when methyl groups are replacing amino groups, is consistent with reduced exposure of smaller $-NH_2$ groups as they become surrounded by $-CH_3$ ones. The taller -CN groups, on the other hand, protrude above surrounding methyl groups. The enhanced role of methyl groups in $-CH_3/-NH_2$ mixtures, and that of nitrile groups on $-CN/-CH_3$ surfaces, are consistent with observed deviations in the cosine of the contact angle on mixed surfaces.

5.2 Natural surfaces with patches at nanoscale

Biological surfaces comprise ingredients with widely varied hydration affinities. Protein surfaces present a well-known example despite preferred exposure of hydrophilic groups. The overall surface energetics will only approximately follow the sum of all component contributions[35, 123]. As with molecularly mixed surfaces, we examine the (non)additivity of wetting free energy on heterogeneous protein surfaces by calculating microscopic analogues of water contact angle[67, 135] as function of surface composition. The repeating units of the surfaces are two different surface patches of melittin dimer, a well-characterized protein[8, 66] with regions of contrasting polarities. The crystal structure is provided from the Protein Data Bank (PDB ID: 2MLT)[124].

The patch of type A is a bigger nanosized area of the dimer comprised of amino-acid residues 2-20 of both monomers (Fig. 33), which includes a hydrophobic pocket flanked by predominantly polar residues. As such, surface A is representative of typical water-soluble proteins. The second surface type, B, mimics the nanosized hydrophobic pocket carved from the central region of fragment A. The comparatively flat fragment comprises atoms from four residues, including two complete LEU13 residues on both monomers and represents the most hydrophobic area on melittin[66]; it is situated in the central region between adjacent melittin dimers forming a tetramer. The sizes of patches A and B are approximately 1.4x2.5 nm^2, and 0.7x0.8 nm^2, respectively. The third patch type, C, is a hydrophilic fragment of size equal to that of type B, carved out of patch A to enable studies of patch size effects.

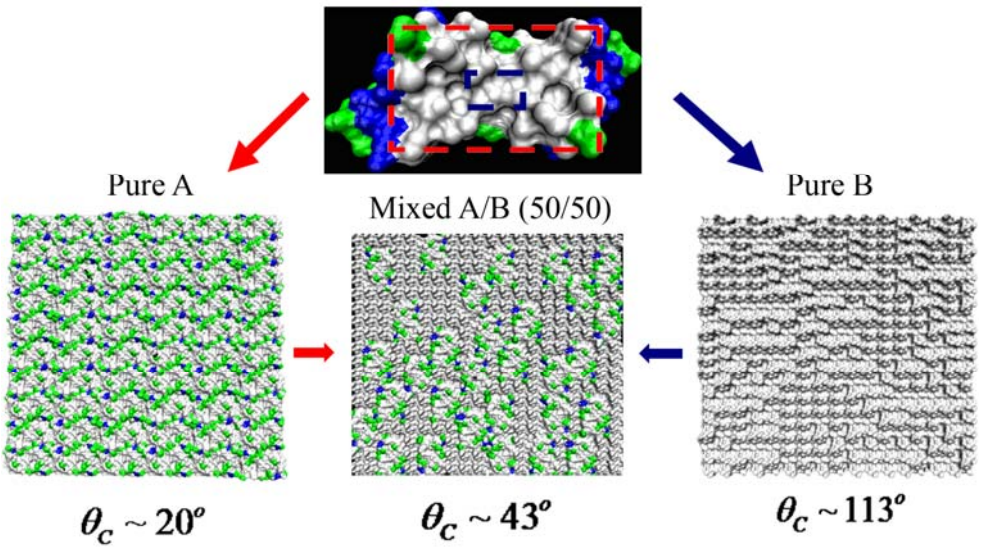


Figure 33. Melittin-based surfaces comprising protein fragments of chosen type: (top) 2MLT (melittin dimer): the part inside the red rectangle represents a patch of type A; the part inside blue rectangle represents a patch of type B. (bottom left) Randomized hydrophilic surface comprising flattened and replicated patches of type A. (bottom right) Randomized hydrophobic surface prepared by replicating patches of type B. (bottom center) Randomized mixed A/B (50/50) surface prepared by mixing patches of type A and equal-size domains of six smaller patches B. Grey color represents hydrophobic residues; other colors represent hydrophilic residues. Among the latter, green color denotes neutral and hydrophilic and blue means basic and hydrophilic.

To characterize the wettability of selected surface fragments in terms of contact angle, we unravel the native "cupped" configuration of the fragment. We follow the procedure developed by Giovambattista *et al*.[66] in which the protein interface was geometrically modified by shifting residues along the inter-dimer direction so that the contact interface between dimers became flat, while preserving the characteristic chemistry. After we download the pdb file of melittin dimer (2MLT), we take all the coordinates of C_α position of the surface hydrophobic residues in 2MLT (melittin dimer) from the PDB, and fit all these positions into a plane. Then we find a rotation matrix to rotate the plane parallel to the *xy* plane of the coordinate system, for every amino acid residue, we measure the closest distance from the atoms within the residue to the plane (now parallel to the *xy* plane, so the distance equals to the distance along the *z*-axis). Then move each atom within the residue the same distance toward the plane. As a result, the closest atom will be on the plane, we can get a roughly planar surface. (See in Fig 34.)

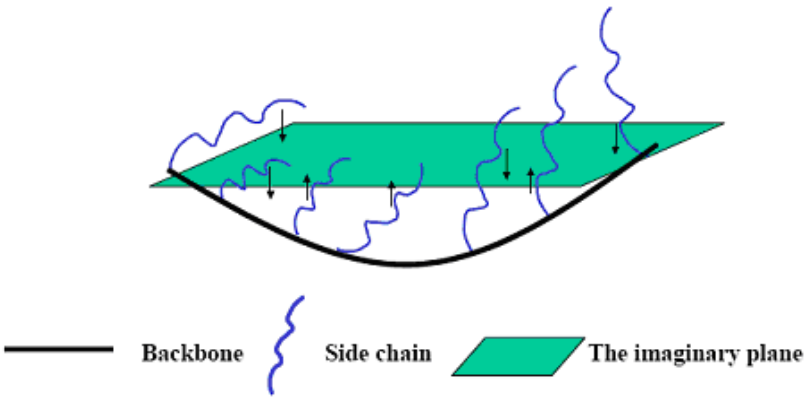


Figure 34. Flatten the melittin dimer surface

We use the CHARMM v27[96] force field which is widely accepted in reproducing the bonding and non-bonding properties of biological systems. The electrostatic potential maps in Fig. 35 were calculated with APBS program[143], in analogy with ref[66].

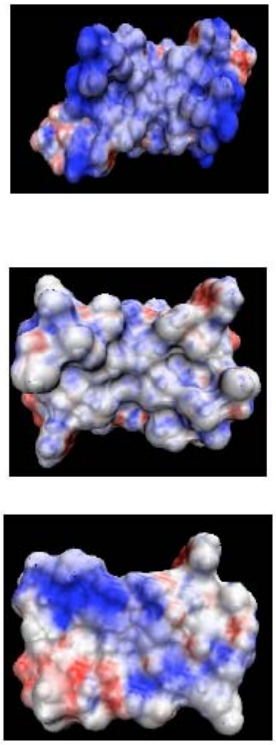


Figure 35. Solvent accessible surface area (SASA) colored by electrostatic potential for (top) original melittin dimer (2MLT); (middle) the truncated melittin dimer and (bottom) flattened, truncated melittin dimer The coloring gradient ranges from -5k_BT/e (red) to 5k_BT/e (blue). The electrostatic potentials at each grid are calculated in APBS program and the figures are produced by VMD software.

We use patches of types A, B, or C as building blocks of larger, approximately square-shaped surfaces designed by patch replication. The final surfaces of side length ~14 nm are sufficiently big to accommodate nanosized droplets (containing around 2000 water molecules) and enable contact angle calculations (Fig. 36). Surfaces were prepared in two ways, by periodic replication, where all patches possess identical orientation, or randomly, by allowing random 180^{o} rotations of individual patches, however, the measured contact angles proved virtually insensitive to replication method (Table 10). Randomized patterns of surfaces of types A and B, and mixed AB surfaces are shown in Fig 33. Droplet contours from Molecular Dynamics simulations are presented in Fig. 36.

	Periodic	Randomized
Type A	22 ^{o}±3 ^{o}	20^{o}±4^{o}
Type B	115^{o}±1 ^{o}	113^{o}±1 ^{o}
Type C	20^{o}±3^{o}	18^{o}±2^{o}

Table 10. Contact angles on flattened protein surfaces prepared by periodic or random replication of patches of type A, B and C

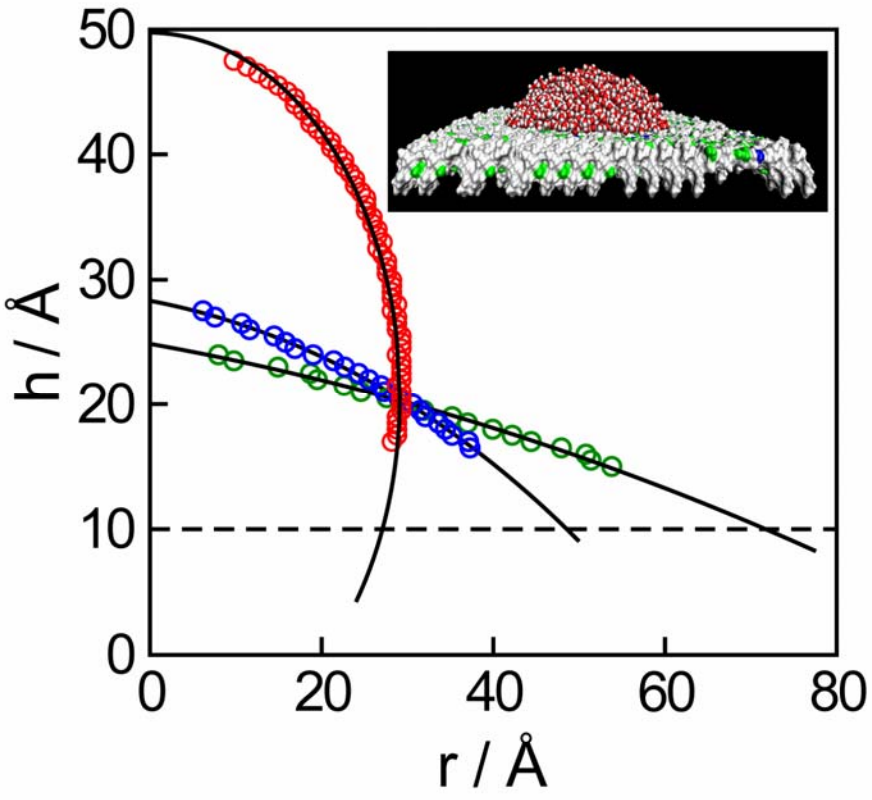


Figure 36. Typical drop profiles for several types of protein surfaces, the circles represent the data points for surface types A (green), B (red) and mixed A/B at 50% (blue). Black solid lines are fitted to simulated data, and the dashed line denotes the height of flattened protein surfaces. Inset: Nanodrop geometry used in water contact angle calculations on a mixed A/B surface.

The wetting coefficient, $cos\theta_c$, displays large positive deviations from predictions of Cassie equation, Eq. 1-3. The Israelachvili and Gee[73] approximation presents an apparent improvement over Cassie's equation, however, the agreement may be spurious as this prediction has been derived to describe effects of mixing at the molecular scale. As the solvent-induced contribution to the adhesion free energy equals $2\gamma cos\theta_c$, these deviations quantify the nonadditivity of surface biointeractions.

In Fig. 37 Bottom we present the results for surfaces with smaller hydrophilic patches of type C (θ_c~18^o) mixed with patches of type B. Patch sizes of both types equal 0.7x0.8 nm^2. Compared to mixtures of bigger patches shown in Fig. 37 Top, the present system features even stronger deviations in cosine of contact angles from both Cassie[71-72] and Israelachvili & Gee[73] predictions.

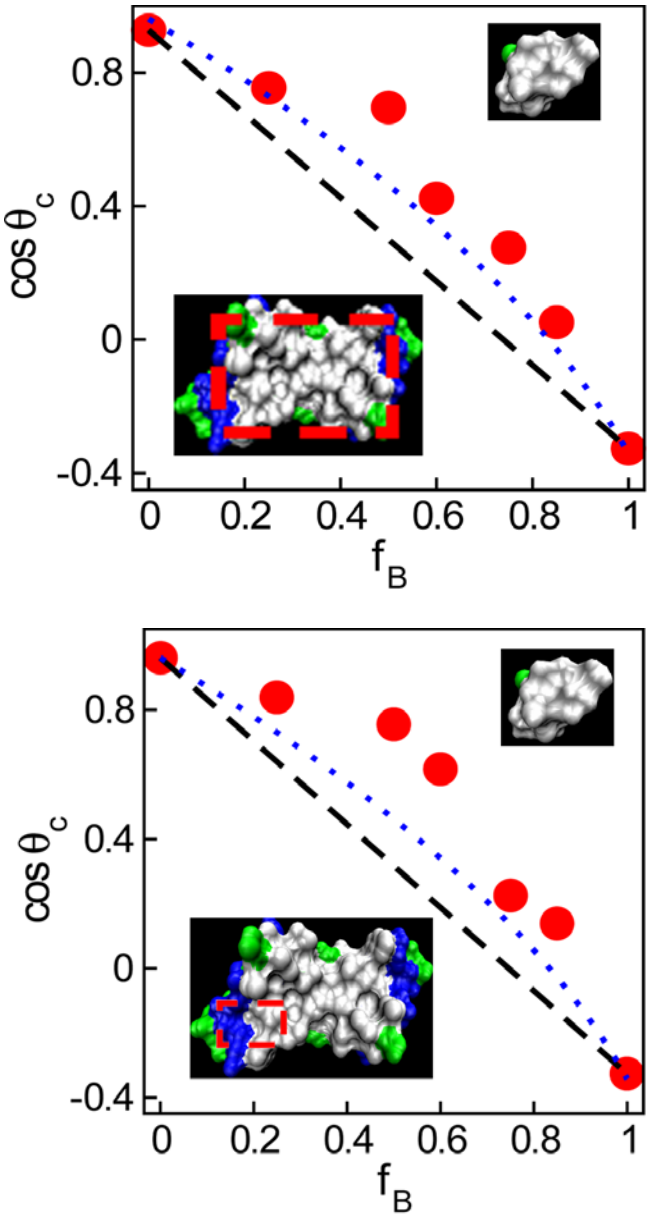


Figure 37. Top: Representative regions of melittin surface A (θ_c~20°, left, lower inset) and hydrophobic fragment B (θ_c~113°, upper inset). Symbols: $\cos\theta_c$ as a function of the hydrophobic surface fraction, f_B, in mixtures of types A and B. Black dashed line: additivity approximation[71-72]. Blue dotted line: approximation from Israelachvili and Gee,[73] Same color notations as in Fig 33. Bottom: $\cos\theta_c$ as a function of the hydrophobic surface fraction in mixtures of fragments B and C (θ_c~20°, red rectangle in the right lower inset). Patch areas are 3.5 nm^2 in system A/B and 1/6 of that in system C/B.

5.3 Discussion

5.3.1 Solvent-accessible area

The original Cassie equation (Eq. 1-2) included approximate accounts for both the chemical heterogeneity and variable roughness on a mixed surface. In subsequent applications of Cassie relation, the roughness of individual surface components was often included implicitly [74, 76-77, 81, 83], by using the values of $\cos\theta_\alpha$ of pure components, which already reflected any roughness *on the homogeneous surface*. For this approach to be valid, the roughnesses and concomitant exposures of distinct surface components to the solvent should be insensitive to mixing. This condition is usually met when combining sizeable surface patches as done e.g. on our protein-like model surfaces. On molecularly mixed surfaces, however, the exposure of taller or bulkier surface moieties (-CN > $-CH_3$ > $-NH_2$) increases while shorter groups become effectively shielded, and their solvent accessible surface (SAS) [144-145] reduced below the value observed on a homogeneous surface. The share of the solvent accessible surface of each group type (α), $f(\mathrm{SAS})_\alpha$, rather than its stoichiometric fraction, f_α, therefore approximately determines the group's contribution to surface properties. This assumption is valid as long as the total exposed area, unlike component shares, remains approximately invariant.

In Fig. 38 we compare the two measures of surface composition, the fractions of hydrophobic surface (area under $-CH_3$ groups on synthetic surfaces, or type B fragments on protein-like surfaces) calculated in terms of SAS, or from projected areas. SAS areas were calculated using the procedure from ref[145]. First, we generate a

spherical mesh of test points to represent the total accessible surface of each atom on the surface, the radius of the sphere is the σ between the atom and the molecular probe (water oxygen in SPC/E model) of the Lennard-Jones parameters from corresponding force fields. The test points on the mesh sphere are evenly distributed at the best approximation by Golden Section Spiral distribution. Each test point has been assigned a small portion of the accessible surface of the atom. If the test points of one certain atom are outside the spheres generated by any other atoms (no overlap), the portions of the accessible surface assigned to these points are assigned to that atom as accessible surface area. We note that each surface has two sides in principle, but we only count the upper side (facing up along the z axis).

Our results show $-CH_3$ groups to be overrepresented in their share of total SAS on $-CH_3$/$-NH_2$ surface, while the opposite is true in the $-CH_3$/-CN mixture.

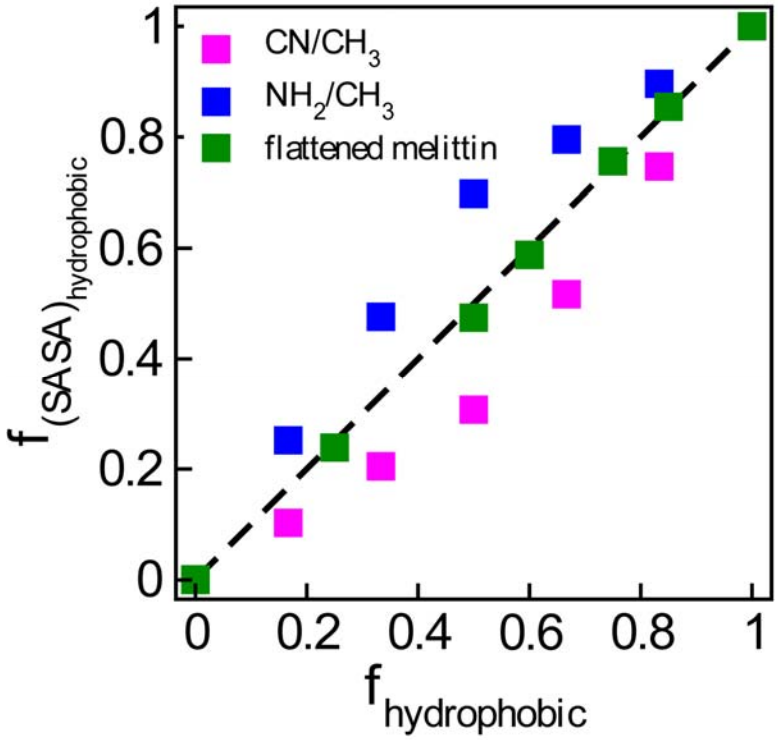


Figure 38. Composition of mixed hydrophobic/hydrophilic surfaces comprised of -CN/-CH_3 (magenta), -NH_2/-CH_3 (blue) and flattened melittin surfaces (green) calculated from SAS (*y* axis), or from projected surfaces (*x* axis).

The changes in solvent accessible surfaces conform to our results for a set of physical properties on synthetic surfaces (Figs. 28, 29, 30), all of which show a disproportionate influence of the bigger species in the mixture. To account for the uneven exposure of moieties of different types on a mixed surface, in Fig. 39 we present the modified additivity plots of the simulated cosine of contact angle as a function of the fraction of SAS of methyl groups, $f(\mathrm{SAS})_{\mathrm{CH3}}$. For both -$CH_3$/-$NH_2$ and -CH_3/-CN mixtures, in this representation, the majority of the points agree with linear additivity prediction. The modified additivity plots for the mixed -CH_3/-NH_2 with different types of force fields are plotted in Fig. 40, which demonstrates that the steric effects are not force fields dependent.

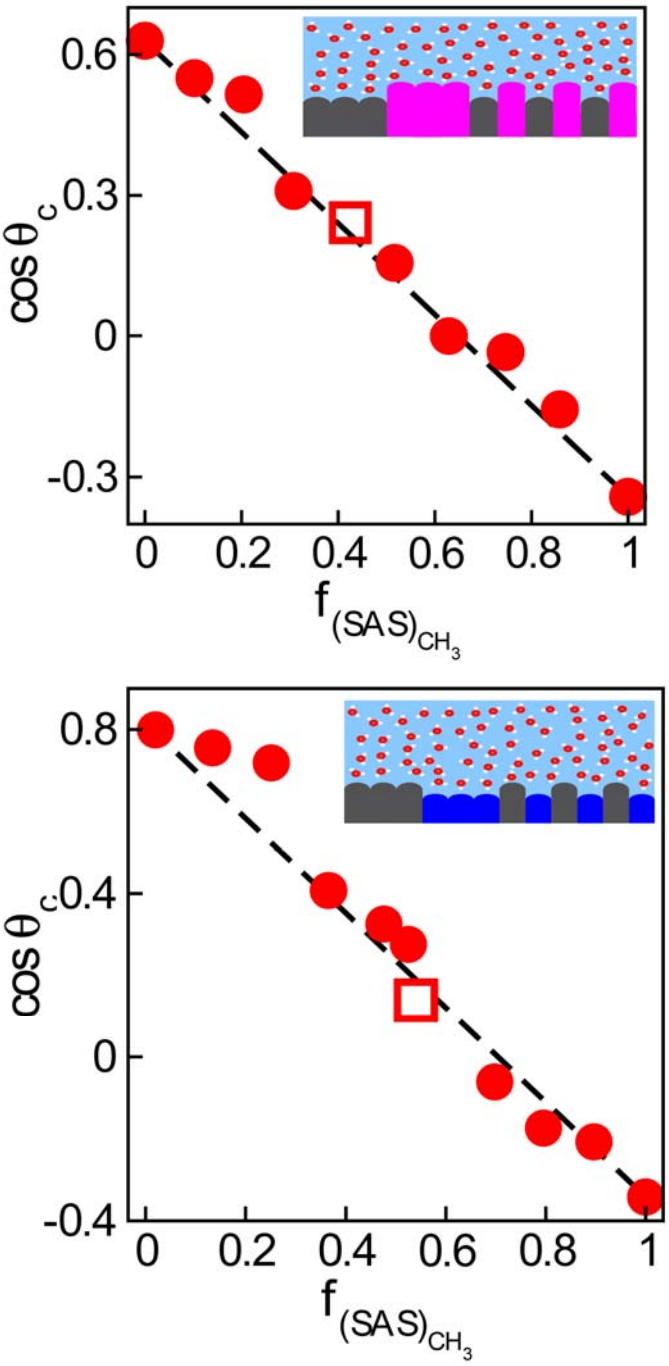


Figure 39. (top) Cosine of contact angle on mixed -CN/-CH_3 and (bottom) -NH_2/-CH_3 surfaces as a function of the fractional solvent accessible area $f_{(SAS)_{CH3}}$ (see text). The dashed lines are the Cassie predictions. Open squares represent the cosine of contact angle with larger (4X4 groups) patches. Insets are sketches of wet molecularly rough -CN/-CH_3 and -NH_2/-CH_3 surfaces, mixed at f_{CH3}=0.5.

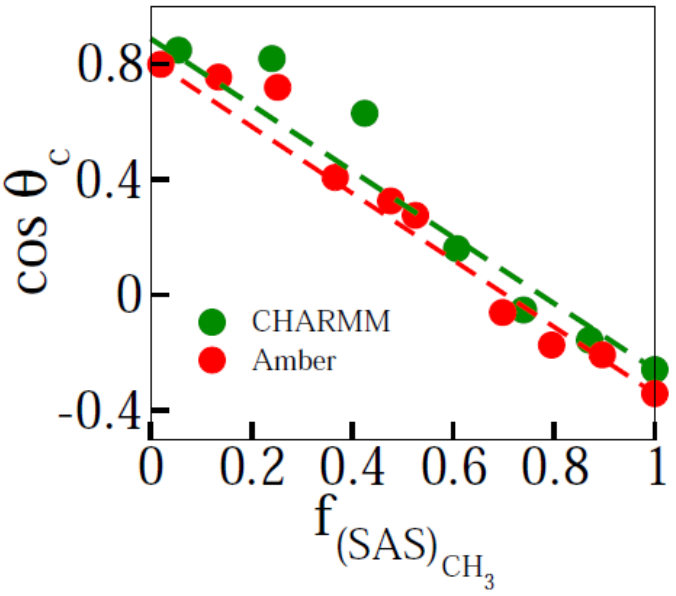


Figure 40. Cosine of contact angles on mixed -CH_3/-NH_2 surfaces as a function of the fraction of surface defined by solvent accessible surface, SAS, covered by -CH_3, with both CHARMM and Amber force fields.

According to Fig. 38, the solvent accessible surfaces of hydrophobic and hydrophilic protein components are virtually insensitive to mixing at the patch level, and the fractions of net SAS remain very close to the fractions of projected surface areas for the two patch types.

There is a different treatment for molecular surface area in literature[146], as shown in Fig. 41, the molecular surface area (MSA) was developed by Connolly[146], with similar idea of the SASA described in the previous section. The MSA takes the concave curvature of the molecular surfaces into account, and is reported to perform better in the predictions of hydrophobic interaction additivity with implicit solvent

models. The Molecular surface areas are calculated by using Connolly molecular surface package.[147]

However, as we are only interested in the fractions of the areas on different types of patches on the molecular surfaces, the fractional areas of protein surfaces we calculated by Molecular Surface Area (MSA)[146] are indistinguishable from f(SAS). Steric effects, critical on molecularly mixed synthetic surfaces, are therefore not a significant source of deviations of $\cos\theta_c$ from additivity on mixed protein-like surfaces.

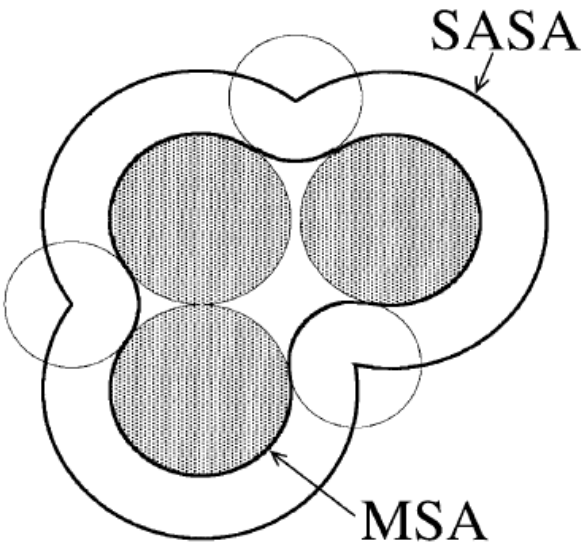


Figure 41. Definitions of solvent-accessible surface area (SASA)[145] and of molecular surface area (MSA) [146] are illustrated by a configuration of three spherical molecules (shaded circles). Each open circle here is a spherical probe SASA of the three-methane configuration is traced out by the center of the probe. MSA corresponds to a part of the probe surface facing the three-molecular configuration.

On the other hand, the original roughnesses of the surfaces are initially included in Cassie's linear additive equation (Equation 1-1, $cos\theta_c = f_A r_A cos\theta_A + f_B r_B cos\theta_B$). At the molecular level, we are able to estimate the production of the roughness factor and the areal fraction assuming the Wenzel-like relation[148], and the pre factors including the roughness of the surface, are calculated as $r_\alpha = A_{SAS\ Patch} / A_{SAS_{Total}}$ per patch, where α denotes the surface type and the A_{SAS} is the surface area calculated by SAS.

The contact angle results from above descriptions for the mixed systems are plotted in Fig. 42. The simulation results (red spheres) qualitatively agree with predictions from the additivity relation using the complete formula. Cassie equation still holds at the molecular level, when the molecular roughness is taken into account.

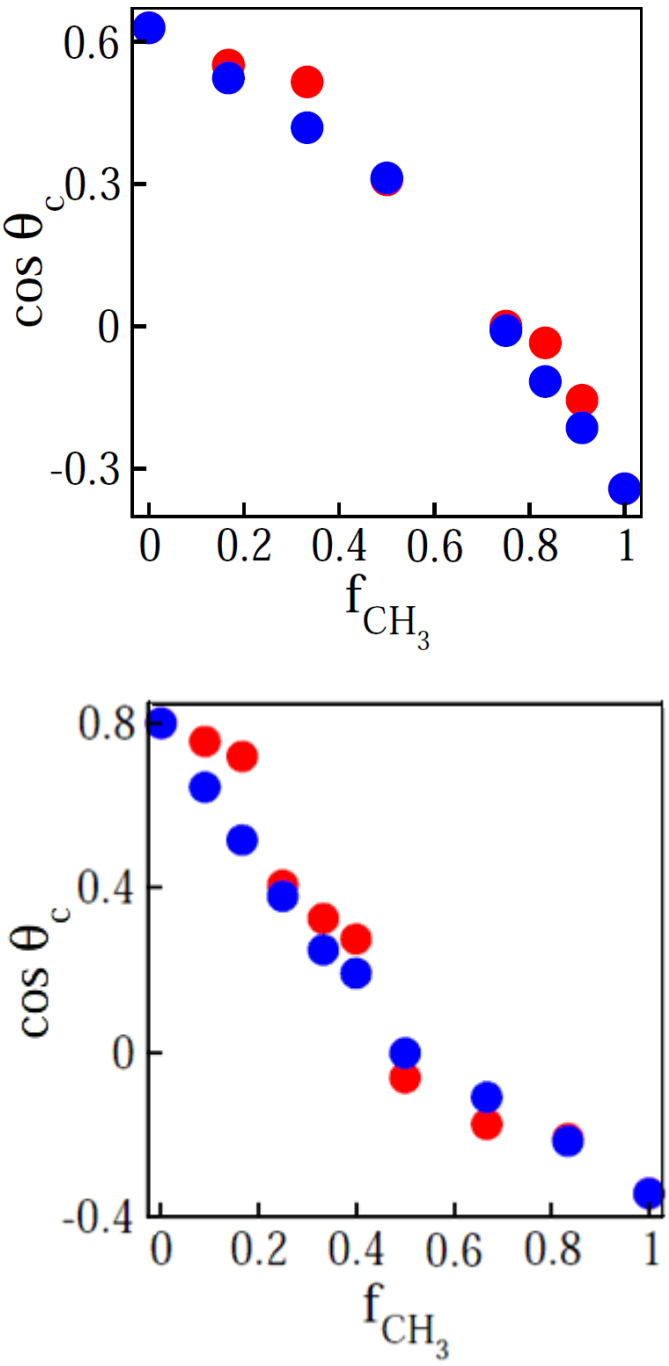


Figure 42. Red spheres (top) Cosine of contact angle on mixed -CN/-CH_3 and (bottom) -NH_2/-CH_3 surfaces as a function of the fractional area f_{CH3} (see text). The blue spheres are the Cassie predictions with relative roughness included (Equation 1-1, see text).

5.3.2 Local water compressibility

To complement the contact angle calculations, we also monitored density fluctuations of interfacial water above our model surfaces. Local water compressibility, κ, can be obtained from density fluctuations:

$$\rho kT\kappa = \frac{< N^2 > - < N >^2}{< N >} \qquad (5\text{-}2)$$

Here, ρ is the number density and N the number of water molecules in specified volume, k is Boltzmann constant, T is temperature, and the angular brackets denote the ensemble average. To establish a connection with ref[87], we determined density fluctuations within the solvation layer under the droplets at varied surface compositions. The calculations are done with at least four 400 ps trajectories on each type of the surfaces (mixed -NH_2 and -CH_3 synthetic surface, using AMBER Parm-94 force field[95]). Only water molecules within 4 Å distance from any surface heavy atom are considered. To discard the effect of fluctuations at the three-phase droplet contour, we considered only the central region of droplet base within 20 Å from the projection of the center of mass of the entire water drop on *xy* plane. Water compressibility for the whole spectrum of compositions is presented in Fig 43. Water near hydrophobic surfaces features higher compressibility than near hydrophilic ones. In agreement with ref [141] (see also Fig. 2 in ref[149]), the compressibility variation with surface composition is especially rapid on hydrophobic surfaces, explaining the increased slope of κ at high fractional solvent accessible surface, $f(SAS)_{CH3}$, in Fig 43(b). The observed compressibility dependence on $f(SAS)_{CH3}$ reaffirms our findings pertaining to extended synthetic surfaces with mixed composition.

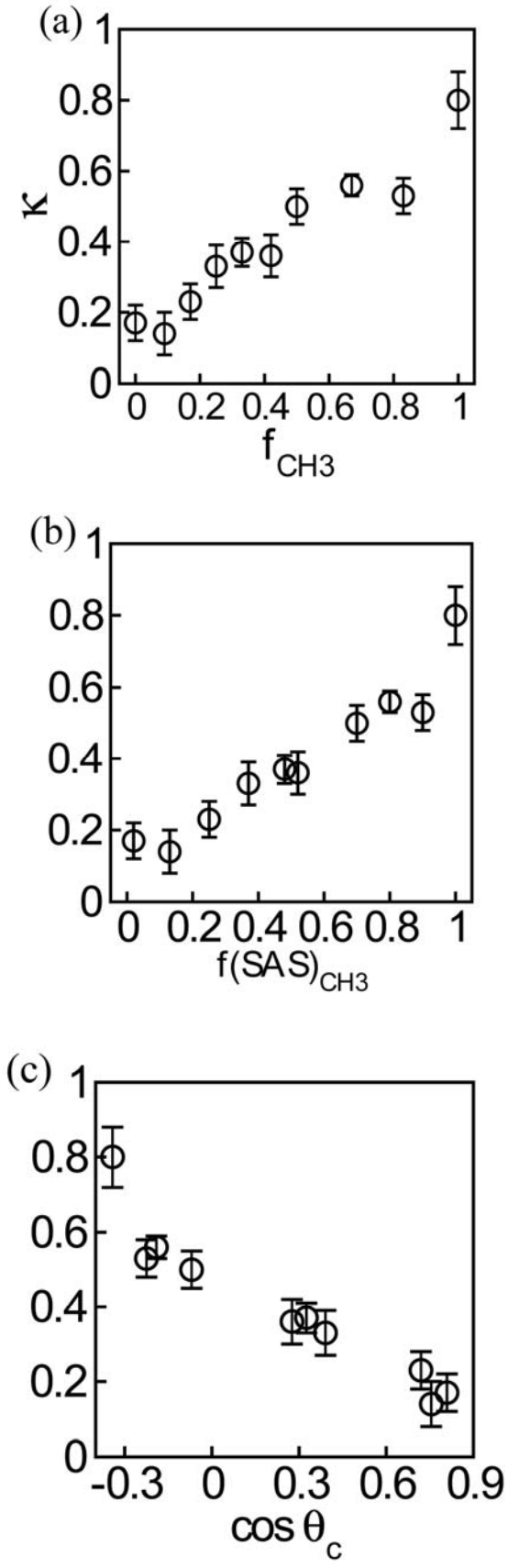


Figure 43. Water compressibility (in arbitrary units) next to surfaces with different fraction of hydrophobic groups (y axis); x axis represents the mole fraction of $-CH_3$ groups on mixed $-NH_2/-CH_3$ surfaces in a); hydrophobic fractional area in terms of SAS in b); and cosine of contact angle in c). Error bars are obtained as standard error in the mean.

5.3.3 Non-additive character of polar surface sites

The few outliers observed in Fig. 39 at low $f(\mathrm{SAS})_{\mathrm{CH3}}$ can be explained in terms of reported asymmetry in the effect of polar heterogeneities: the effect of adding or removing a polar group is much weaker when the background is already hydrophilic. This is attributed[87] to increased interface softness[89] atop *apolar* domains; such interfaces are much more susceptible to surface perturbations[34, 87]. The asymmetry can be increased by competition among orienting fields acting on water dipoles in the presence of multiple polar sites.

These anti-cooperative effects are especially prominent on protein-like surfaces characterized by pronounced polarity variations. Note that hydrophilic protein patches (Fig. 37, patch A, θ_c~20^o) are themselves quite heterogeneous, comprising both hydrophobic and highly polar groups, which strongly influence the surface even as a minority component. In the presence of these extremely polar sites, the asymmetric influence of surface heterogeneities[87] dominates the dependence of the contact angle on surface composition. The substantially stronger impact of polar groups on the hydrophobic background (f_B in Fig. 37 close to unity) explains strong positive deviations from linear dependence of $\cos\theta_c$ on f_B.

5.3.4 Patch size effects

On surfaces dominated by polar groups, or on strongly segregated surfaces, there exist islands of predominant hydrophilic character and lengthscale comparable to the drop size. In these situations, water bias for large hydrophilic domains generally reduces contact angles [81, 83].

Partial segregation, resulting in moderate-size patches can affect the contact angle in two different ways. On molecularly mixed synthetic surfaces, strongly influenced by the changes in the solvent accessible surface, the smaller, partially shielded moieties regain their exposure to water upon demixing. This reduces the contact angle of -CH_3/-NH_2 mixtures, but increases it in mixtures of -CH_3/-CN. To estimate the magnitude of the effect, we performed additional simulations with larger patches (4x4 groups in a patch) in equimolar mixtures (f_{CH3}=0.5) devoid of large segregated domains. Results are given as open square points in Figs. 1 and 6. In the mixture -CH_3/-NH_2, the increase in the patch size rises the fractional solvent accessible area f $(SAS)_{NH2}$ from ~0.31 to ~0.47, lowering contact angle by ~10°. In contrast, segregation reduces exposure of taller -CN groups in the -CH_3/-CN mixture, lowering f $(SAS)_{CN}$ ~0.7 to ~0.6, rising θ_c by ~3°. This agrees with calculations on mixed -CH_3/-CH_2-OH surfaces where an increase in patch size lead to higher contact angle[83]. As shown in Fig. 39, the use of simulated SAS fractional areas correctly accounts for these changes.

On the protein-like surfaces, on the other hand, increased patch size emphasizes the anti-cooperative character of water/surface interactions. While patch size has no

effect on pure surfaces, transition from small to moderate-size patches can result in higher contact angles in the mixtures[81, 83]. This is corroborated by the comparison of the results for two different patch sizes in Fig 37, showing increased deviations of $\cos\theta_c$ from the Cassie line upon six-fold reduction in the area of the patches. The observed increase is consistent with predictions from a recent MD study of water structure and potential of mean force between heterogeneous platelets with varied surface pattern [35].

5.4 Concluding Remarks

We identify two distinct mechanisms responsible for the non-additivity of wetting free energies on heterogeneous surfaces. On molecularly mixed synthetic surfaces with moderately polar ingredients steric effect dominate, and the observed positive *and negative* deviations from Cassie equation are explained in terms of changes in solvent accessible areas of mixture components. Prototypical biological surfaces, characterized by strongly contrasting polarities, on the other hand, generally show positive deviations in $\cos\theta_c$, which can be rationalized by non-additive interactions between water and highly polar surface sites.

An additional source of deviations from the Cassie behavior can be traced to nonuniform wetting of polar and apolar domains under the droplet perimeter. When adjacent surface areas have very different polarities, fluctuations of nanodroplet base [83, 150] favor inclusion of polar patches. Simulated distribution of water atop a heterogeneous surface comprised of melittin fragments, (Fig. 44), confirms preferential wetting of hydrophilic areas. The bias will be strongest in the area beneath the drop perimeter, reducing the contact angle[150].

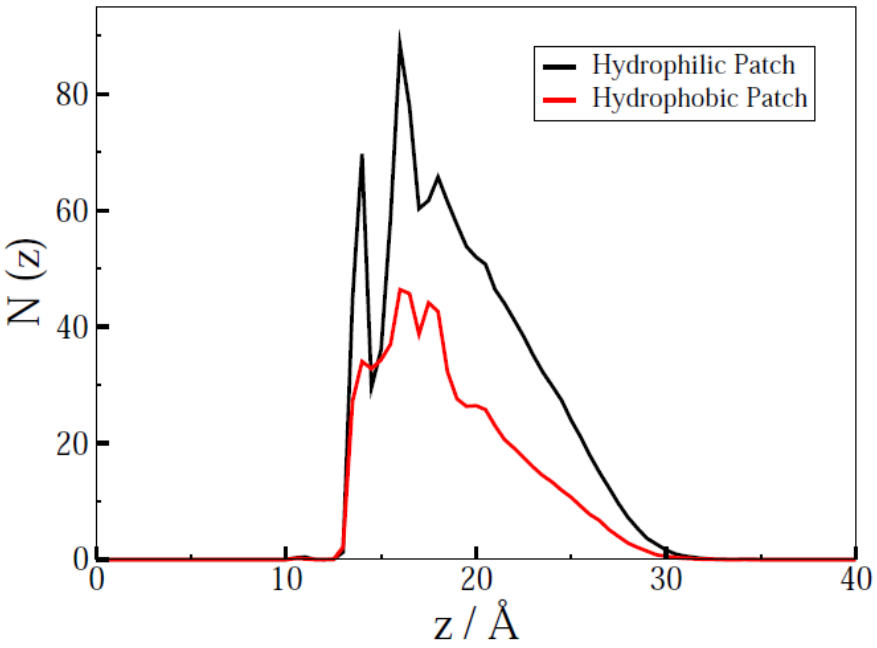


Figure 44. Distribution of water molecules near symmetrically mixed protein surfaces with patches of type A (hydrophilic) and type B (hydrophobic) indicate preferential wetting of hydrophilic patches. N(z) is the number of water molecules in slabs z ± 0.5Å located above hydrophilic (black) or hydrophobic (red) patches. Note also the comparatively loose liquid interface on hydrophobic domains.

Further, the liquid interface above a neat hydrophobic domain can be slightly detached, but will adhere closely when the interface is pinned by adjacent hydrophilic patches[30, 34, 78, 89, 151]. Effective roughness, and substrate/solvent interaction on hydrophobic domains can therefore depend on the *environment*, introducing additional uncertainty in the predictions of Cassie equation.

Qualitatively, the above molecular mechanisms rationalize the large positive deviation in cosine of contact angles on mixed protein-like surfaces. The intricate influence of anti-cooperative hydration, nonuniform wetting, and partial detachment of the liquid/hydrophobe interface remain to be integrated into a predictive theory for wetting free energy of heterogeneous biosurfaces, a challenge to be addressed in future studies.

In future studies, we are also planning to evaluate the averaged surface forces on these mixed surfaces, which we will be able to compare to the experiments that directly measure the forces in hydrophobic systems (Surface Forces Apparatus (SFA) and Atomic Force Microscope (AFM)).[46] The surface forces on heterogeneous surfaces should have the same additivity as the contact angles (linearly additive or non-additive).

References

1. Hildebra.Jh, Nemethy G, Scheraga HA, Kauzmann W (1968) A criticism of term hydrophobic bond. *J. Phys. Chem.* 72(5):1841.
2. Nemethy G, Scheraga HA (1962) The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. *J. Phys. Chem.* 66:1773.
3. Dill KA (1990) Dominant Forces in Protein Folding. *Biochemistry* 29(31):7133-7155.
4. Hummer G, Pratt LR, Garcia AF (1996) Free Energy of Ionic Hydration. *J. Phys. Chem.* 100(4):1206-1215.
5. Widom B, Bhimalapuram P, Koga K (2003) The hydrophobic effect. *PCCP* 5(15):3085-3093.
6. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* 437(7059):640-647.
7. Rajamani S, Truskett TM, Garde S (2005) Hydrophobic hydration from small to large lengthscales: Understanding and manipulating the crossover. *Proc. Natl. Acad. Sci. USA* 102(27):9475-9480.
8. Berne BJ, Weeks JD, Zhou RH (2009) Dewetting and Hydrophobic Interaction in Physical and Biological Systems. *Annu. Rev. Phys. Chem.* 60:85-103.
9. Ball P (2008) Water as an active constituent in cell biology. *Chem. Rev.* 108(1):74-108.
10. Baron R, Setny P, McCammon JA (2010) Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* 132(34):12091-12097.
11. Stillinger FH (1973) Structure in Aqueous Solutions of Nonpolar Solutes from the Standpoint of Scaled-Particle Theory. *J. Solution Chem.* 2(2-3):141-158.
12. Berne BJ (1996) Inferring the hydrophobic interaction from the properties of neat water. *Proc. Natl. Acad. Sci. USA* 93(17):8800-8803.
13. Luzar A (1996) Water hydrogen-bond dynamics close to hydrophobic and hydrophilic groups. *Faraday Discuss.* 103:29-40.
14. Lum K, Chandler D, Weeks JD (1999) Hydrophobicity at small and large length scales. *J. Phys. Chem. B* 103(22):4570-4577.
15. Varilly P, Patel AJ, Chandler D (2010) An improved coarse-grained model of solvation and the hydrophobic effect. *Los Alamos Archive* (http://arxiv.org/abs/1010.5750v1).
16. Luzar A, Svetina S, Zeks B (1983) The Contribution of Hydrogen-bonds to the Surface-tension of Water. *Chem. Phys. Lett.* 96(4):485-490.
17. Luzar A, Svetina S, Zeks B (1985) Consideration of the spontaneous polarization of water at the solid liquid interface. *J. Chem. Phys.* 82(11):5146-5154.
18. Lee CY, McCammon JA, Rossky PJ (1984) The Structure of Liquid Water at An Extended Hydrophobic Surface. *J. Chem. Phys.* 80(9):4448-4455.

19. Du Q, Superfine R, Freysz E, Shen YR (1993) Vibrational Spectroscopy of Water at the Vapor Water Interface. *Phys. Rev. Lett.* 70(15):2313-2316.
20. Huang DM, Geissler PL, Chandler D (2001) Scaling of Hydrophobic Solvation Free Energies. *J. Phys. Chem. B* 105(28):6704-6709.
21. Cheng YK, Rossky PJ (1998) Surface topography dependence of biomolecular hydrophobic hydration. *Nature* 392(6677):696-699.
22. Wang L, Friesner RA, Berne BJ (2010) Hydrophobic interactions in model enclosures from small to large length scales: non-additivity in explicit and implicit solvent models. *Faraday Discuss.* 146:247-262.
23. Stewart MC, Evans R (2005) Wetting and drying at a curved substrate: Long-ranged forces. *Phys. Rev. E* 71(1).
24. Bromberg KADaS (2003) *Molecular Driving Forces* (Garland Science, New York).
25. Chothia C (1974) HYDROPHOBIC BONDING AND ACCESSIBLE SURFACE-AREA IN PROTEINS. *Nature* 248(5446):338-339.
26. Eisenberg D, McLachlan AD (1986) SOLVATION ENERGY IN PROTEIN FOLDING AND BINDING. *Nature* 319(6050):199-203.
27. Sharp KA, Nicholls A, Fine RF, Honig B (1991) RECONCILING THE MAGNITUDE OF THE MICROSCOPIC AND MACROSCOPIC HYDROPHOBIC EFFECTS. *Science* 252(5002):106-109.
28. Dong F, Wagoner JA, Baker NA (2008) Assessing the performance of implicit solvation models at a nucleic acid surface. *PCCP* 10(32):4889-4902.
29. Pratt LR, Chandler D (1977) Theory of hydrophobic effect. *J. Chem. Phys.* 67(8):3683-3704.
30. Lum K, Luzar A (1997) Pathway to surface-induced phase transition of a confined fluid. *Phys. Rev. E* 56(6):R6283-R6286.
31. Leung K, Luzar A (2000) Dynamics of capillary evaporation. II. Free energy barriers. *J. Chem. Phys.* 113(14):5845-5852.
32. Leung K, Luzar A, Bratko D (2003) Dynamics of capillary drying in water. *Phys. Rev. Lett.* 90(6).
33. Luzar A (2004) Activation barrier scaling for the spontaneous evaporation of confined water. *J. Phys. Chem. B* 108(51):19859-19866.
34. Luzar A, Leung K (2000) Dynamics of capillary evaporation. I. Effect of morphology of hydrophobic surfaces. *J. Chem. Phys.* 113(14):5836-5844.
35. Hua L, Zangi R, Berne BJ (2009) Hydrophobic Interactions and Dewetting between Plates with Hydrophobic and Hydrophilic Domains. *J. Phys. Chem. C* 113(13):5244-5253.
36. Huang XH, Zhou RH, Berne BJ (2005) Drying and hydrophobic collapse of paraffin plates. *J. Phys. Chem. B* 109(8):3546-3552.
37. Giovambattista N, Debenedetti PG, Rossky PJ (2007) Hydration behavior under confinement by nanoscale surfaces with patterned hydrophobicity and hydrophilicity. *J. Phys. Chem. C* 111(3):1323-1332.

38. Dzubiella J (2010) Explicit and implicit modeling of nanobubbles in hydrophobic confinement. *Anais Da Academia Brasileira De Ciencias* 82(1):3-12.
39. Miller TF, Vanden-Eijnden E, Chandler D (2007) Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. *Proc. Natl. Acad. Sci. USA* 104(37):14559-14564.
40. ten Wolde PR, Chandler D (2002) Drying-induced hydrophobic polymer collapse. *Proc. Natl. Acad. Sci. USA* 99(10):6539-6543.
41. Hummer G, Rasaiah JC, Noworyta JP (2001) Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* 414(6860):188-190.
42. Jensen MO, Mouritsen OG, Peters GH (2004) The hydrophobic effect: Molecular dynamics simulations of water confined between extended hydrophobic and hydrophilic surfaces. *J. Chem. Phys.* 120(20):9729-9744.
43. Luzar A, Bratko D, Blum L (1987) Monte-Carlo simulation of hydrophobic interaction. *J. Chem. Phys.* 86(5):2955-2959.
44. Shelley JC, Patey GN (1996) Boundary condition effects in simulations of water confined between planar walls. *Mol. Phys.* 88(2):385-398.
45. Wallqvist A, Berne BJ (1995) Molecular Dynamics Study of the Dependence of water solvation free energy on solute curvature and surface area. *J. Phys. Chem.* 99(9):2885-2892.
46. Leckband D, Israelachvili J (2001) Intermolecular forces in biology. *Q. Rev. Biophys.* 34(2):105-267.
47. Southall NT, Dill KA, Haymet ADJ (2002) A view of the hydrophobic effect. *J. Phys. Chem. B* 106(3):521-533.
48. Hummer G (2010) Molecular bonding under water's influence. *Nature Chemistry* 2(11):906-907.
49. Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* 130(9):2817-2831.
50. Eun C, Berkowitz ML (2010) Fluctuations in Number of Water Molecules Confined between Nanoparticles. *J. Phys. Chem. B* 114(42):13410-13414.
51. Hua L, Huang XH, Liu P, Zhou RH, Berne BJ (2007) Nanoscale dewetting transition in protein complex folding. *J. Phys. Chem. B* 111(30):9069-9077.
52. Liu P, Huang XH, Zhou RH, Berne BJ (2005) Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* 437(7055):159-162.
53. Israelachvili JN (1992) Adhesion forces between surfaces in liquids and condensable vapors. *Surf. Sci. Rep.* 14(3):109-159.
54. Christenson HK, Claesson PM (2001) Direct measurements of the force between hydrophobic surfaces in water. *Adv. Colloid Interface Sci.* 91(3):391-436.
55. Ball P (2003) Chemical physics: How to keep dry in water. *Nature* 423(6935):25-26.
56. Jensen TR, *et al.* (2003) Water in contact with extended hydrophobic surfaces: Direct evidence of weak dewetting. *Phys. Rev. Lett.* 90(8).

57. Klosgen BM, *et al.* (2003) On the deposition of model membranes onto hydrophobic substrates. *Biophys. J.* 84(2):188A-188A.
58. Tyrrell JWG, Attard P (2001) Images of nanobubbles on hydrophobic surfaces and their interactions. *Phys. Rev. Lett.* 87(17).
59. Yakubov GE, Butt HJ, Vinogradova OI (2000) Interaction forces between hydrophobic surfaces. Attractive jump as an indication of formation of "stable" submicrocavities. *J. Phys. Chem. B* 104(15):3407-3410.
60. Levy Y, Onuchic JN (2006) Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* 35:389-415.
61. Zhou RH, Huang XH, Margulis CJ, Berne BJ (2004) Hydrophobic collapse in multidomain protein folding. *Science* 305(5690):1605-1609.
62. Young T, *et al.* (2010) Dewetting transitions in protein cavities. *Proteins-Structure Function and Bioinformatics* 78(8):1856-1869.
63. Willard AP, Chandler D (2008) The role of solvent fluctuations in hydrophobic assembly. (Translated from English) *J. Phys. Chem. B* 112(19):6187-6192 (in English).
64. Bratko D, Blanch HW (2001) Competition between protein folding and aggregation: A three-dimensional lattice-model simulation. *J. Chem. Phys.* 114(1):561-569.
65. Teplow DB, *et al.* (2006) Elucidating amyloid beta-protein folding and assembly: A multidisciplinary approach. *Acc. Chem. Res.* 39(9):635-645.
66. Giovambattista N, Lopez CF, Rossky PJ, Debenedetti PG (2008) Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc. Natl. Acad. Sci. USA* 105(7):2274-2279.
67. Hautman J, Klein ML (1991) Microscopic Wetting Phenomena. *Phys. Rev. Lett.* 67(13):1763-1766.
68. Leopoldes J, Dupuis A, Bucknall DG, Yeomans JM (2003) Jetting micron-scale droplets onto chemically heterogeneous surfaces. *Langmuir* 19(23):9818-9822.
69. Rossky PJ (2010) Exploring nanoscale hydrophobic hydration. *Faraday Discuss.* 146:13.
70. Rowlinson JS, Widom B (2002) *Molecular Theory of Capillarity* (Dover, New York).
71. Cassie ABD (1948) Contact Angles. *Discuss. Faraday Soc.* 3:11-16.
72. Cassie ABD, Baxter S (1944) Wettability of Porous surfaces. *Trans. Faraday Soc.* 40:546.
73. Israelachvili JN, Gee ML (1989) Contact Angles on Chemically Heterogeneous Surfaces. *Langmuir* 5(1):288-289.
74. Bain CD, Evall J, Whitesides GM (1989) Formation of Monolayers by the Coadsorption of Thiols on Gold - Variation in the Head Group, Tail Group, and Solvent. *J. Am. Chem. Soc.* 111(18):7155-7164.
75. Bain CD, *et al.* (1989) Formation of Monolayer Films by the Spontaneous Assembly of Organic Thiols from Solution onto Gold. *J. Am. Chem. Soc.* 111(1):321-335.

76. Bain CD, Whitesides GM (1989) Formation of Monolayers by the Coadsorption of Thiols on Gold - Variation in the Length of the Alkyl Chain. *J. Am. Chem. Soc.* 111(18):7164-7175.
77. Arima Y, Iwata H (2007) Effect of wettability and surface functional groups on protein adsorption and cell adhesion using well-defined mixed self-assembled monolayers. *Biomaterials* 28(20):3074-3082.
78. Kuna JJ, *et al.* (2009) The effect of nanometre-scale structure on interfacial energy. *Nat. Mater.* 8(10):837-842.
79. Gao LC, McCarthy TJ (2009) An Attempt to Correct the Faulty Intuition Perpetuated by the Wenzel and Cassie "Laws". *Langmuir* 25(13):7249-7255.
80. McHale G (2007) Cassie and Wenzel: Were they really so wrong? *Langmuir* 23(15):8200-8205.
81. Lundgren M, Allan NL, Cosgrove T (2007) Modeling of wetting: A study of nanowetting at rough and heterogeneous surfaces. *Langmuir* 23(3):1187-1194.
82. Woodward JT, Gwin H, Schwartz DK (2000) Contact angles on surfaces with mesoscopic chemical heterogeneity. *Langmuir* 16(6):2957-2961.
83. Halverson JD, Maldarelli C, Couzis A, Koplik J (2010) Atomistic simulations of the wetting behavior of nanodroplets of water on homogeneous and phase separated self-assembled monolayers. *Soft Matter* 6:1297-1307.
84. Godawat R, Jamadagni SN, Garde S (2009) Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proc. Natl. Acad. Sci. USA* 106(36):15119-15124.
85. Mittal J, Hummer G (2008) Static and dynamic correlations in water at hydrophobic interfaces. *Proc. Natl. Acad. Sci. USA* 105(51):20130-20135.
86. Bratko D, Curtis RA, Blanch HW, Prausnitz JM (2001) Interaction between hydrophobic surfaces with metastable intervening liquid. *J. Chem. Phys.* 115(8):3873-3877.
87. Acharya H, Ranganathan S, Jamadagni SN, Garde S (2010) Mapping Hydrophobicity at the Nanoscale: Applications to Heterogeneous Surfaces and Proteins. *Faraday Discuss.* 146:353.
88. Bratko D, Daub CD, Luzar A (2009) Water-mediated ordering of nanoparticles in an electric field. *Faraday Discuss.* 141:55-66.
89. Huang DM, Chandler D (2002) The hydrophobic effect and the influence of solute-solvent attractions. *J. Phys. Chem. B* 106(8):2047-2053.
90. Cellmer T, Bratko D, Prausnitz JM, Blanch H (2005) Protein-folding landscapes in multichain systems. *Proc. Natl. Acad. Sci. USA* 102(33):11692-11697.
91. Chandler D (1987) *Introduction to Modern Statistical Mechanics* (Oxford, New York).
92. Allen MP, Tildesley DJ (1987) *Computer Simulation of Liquids* (Oxford University Press, New York).
93. Frenkel D, Smit B (2002) *Understanding molecular simulation, from algorithms to applications* (Academic Press, San Diego).

94. Lennard-Jones JE (1924) On the Determination of Molecular Fields. *Proc. R. Soc. Lond. A* 106(738):463-477.
95. Cornell WD, *et al.* (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* 117(19):5179-5197.
96. MacKerell AD, *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102(18):3586-3616.
97. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118(45):11225-11236.
98. Berendsen HJC, Grigera JR, Straatsma TP (1987) The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* 91(24):6269-6271.
99. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79(2):926-935.
100. H. J. C. Berendsen JPMP, W. F. van Gunsteren and J. Hermans (1981) *Intermolecular Forces* ed Pullman B (Reidel, Dordrecht), p 331.
101. Verlet L (1967) Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159(1):98.
102. Plimpton S (1995) Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *Journal of Computational Physics* 117(1):1-19.
103. Smith W, Forester TR (1996) DL_POLY_2.0: A general-purpose parallel molecular dynamics simulation package. *J. Mol. Graphics* 14(3):136-141.
104. Deleeuw SW, Perram JW, Smith ER (1980) Simulation of Electrostatic Systems in Periodic Boundary-Conditions. 1. Lattice Sums and Dielectric-Constants. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 373(1752):27-56.
105. Essmann U, *et al.* (1995) A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 103(19):8577-8593.
106. Chen YG, Weeks JD (2006) Local molecular field theory for effective attractions between like charged objects in systems with strong Coulomb interactions. *Proc. Natl. Acad. Sci. USA* 103(20):7560-7565.
107. Izvekov S, Swanson JMJ, Voth GA (2008) Coarse-graining in interaction space: A systematic approach for replacing long-range electrostatics with short-range Potentials. *J. Phys. Chem. B* 112(15):4711-4724.
108. N. Metropolis AWR, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087.
109. Glauber RJ (1963) Time Dependent Statistics of the Ising Model. *Math. Phys.* 4:294.
110. Kawasaki K (1966) Diffusion constants near critical point for time dependent ising models. I. *Phys. Rev.* 145(1):224.

111. Walrafen G (1972) *Water, a comprehensive treatise*, ed Franks F (Plenum Press, New York), Vol 1, p 151.
112. Luck WAP (1967) Spectroscopic studies concerning structure and thermodynamic behaviour of H2O CH3OH and C2H5OH. *Discuss. Faraday Soc.* (43):115-&.
113. Eisenburg D, Kauzmann W (1969) *The structure and properties of water* (Oxford Univ. Press, London).
114. Fowler RH (1937) A Tentative Statistical Theory of Macleod's Equation for Surface Tension, and the Parachor. *Proc. R. Soc. Lond. A* 159:229-246.
115. Bratko D, Luzar A, Chen SH (1988) Electrostatic Model for Protein Reverse Micelle Complexation. (Translated from English) *J. Chem. Phys.* 89(1):545-550 (in English).
116. Dzubiella J, Hansen JP (2004) Competition of hydrophobic and Coulombic interactions between nanosized solutes. *J. Chem. Phys.* 121(11):5514-5530.
117. Rodgers JM, Weeks JD (2008) Interplay of local hydrogen-bonding and long-ranged dipolar forces in simulations of confined water. *Proc. Natl. Acad. Sci. USA* 105(49):19136-19141.
118. Rodgers JM, Weeks JD (2008) Local molecular field theory for the treatment of electrostatics. *Journal of Physics-Condensed Matter* 20(49).
119. Spohr E (1999) Molecular simulation of the electrochemical double layer. *Electrochim. Acta* 44(11):1697-1705.
120. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157(1):105-132.
121. Harano Y, Imai T, Kovalenko A, Kinoshita M, Hirata F (2001) Theoretical study for partial molar volume of amino acids and polypeptides by the three-dimensional reference interaction site model. *J. Chem. Phys.* 114(21):9506-9511.
122. Wang J, Bratko D, Luzar A (2010) Probing surface tension additivity on chemically heterogeneous surfaces: A molecular approach. *under review in Proc. Natl. Acad. Sci. USA*
123. Wang L, Friesner RA, Berne BJ (2010) Hydrophobic interactions in model enclosures from small to large length scales: non-additivity in explicit and implicit solvent models. *Faraday Discuss.* 146:247.
124. Terwilliger TC, Eisenberg D (1982) The Structure of Melittin .1. Structure Determination and Partial Refinement. *J. Biol. Chem.* 257(11):6010-6015.
125. Huang Q, Ding SW, Hua CY, Yang HC, Chen CL (2004) A computer simulation study of water drying at the interface of protein chains. *J. Chem. Phys.* 121(4):1969-1977.
126. Qvist J, Davidovic M, Hamelberg D, Halle B (2008) A dry ligand-binding cavity in a solvated protein. *Proc. Natl. Acad. Sci. USA* 105(17):6296-6301.
127. Qin BY, *et al.* (1998) Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry* 37(40):14014-14023.

128. Szori M, Tobias DJ, Roeselova M (2009) Microscopic Wetting of Mixed Self-assembled Monolayers: A Molecular Dynamics Study. *J. Phys. Chem. B* 113(13):4161-4169.
129. Werder T, Walther JH, Jaffe RL, Halicioglu T, Koumoutsakos P (2003) On the water-carbon interaction for use in molecular dynamics simulations of graphite and carbon nanotubes. *J. Phys. Chem. B* 107(6):1345-1352.
130. Wang SR, Zhang Y, Abidi N, Cabrales L (2009) Wettability and Surface Free Energy of Graphene Films. *Langmuir* 25(18):11078-11081.
131. Shin YJ, *et al.* (2010) Surface-Energy Engineering of Graphene. *Langmuir* 26(6):3798-3802.
132. Evans DJ, Holian BL (1985) The Nose-Hoover Thermostat. *J. Chem. Phys.* 83(8):4069-4074.
133. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 98(12):10089-10092.
134. de Ruijter MJ, Blake TD, De Coninck J (1999) Dynamic wetting studied by molecular modeling simulations of droplet spreading. *Langmuir* 15(22):7836-7847.
135. Daub CD, Bratko D, Leung K, Luzar A (2007) Electrowetting at the nanoscale. *J. Phys. Chem. C* 111(2):505-509.
136. Ingebrigtsen T, Toxvaerd S (2007) Contact angles of Lennard-Jones liquids and droplets on planar surfaces. *J. Phys. Chem. C* 111(24):8518-8523.
137. Bratko D, Daub CD, Leung K, Luzar A (2007) Effect of field direction on electrowetting in a nanopore. *J. Am. Chem. Soc.* 129(9):2504-2510.
138. Keselowsky BG, Collard DM, Garcia AJ (2003) Surface chemistry modulates fibronectin conformation and directs integrin binding and specificity to control cell adhesion. (Translated from English) *Journal of Biomedical Materials Research Part A* 66A(2):247-259 (in English).
139. Sigal GB, Mrksich M, Whitesides GM (1998) Effect of surface wettability on the adsorption of proteins and detergents. *J. Am. Chem. Soc.* 120(14):3464-3473.
140. Mezger M, *et al.* (2010) On the Origin of the Hydrophobic Water Gap: An X-ray Reflectivity and MD Simulation Study. *J. Am. Chem. Soc.* 132(19):6735-6741.
141. Daub CD, Bratko D, Ali T, Luzar A (2009) Microscopic Dynamics of the Orientation of a Hydrated Nanoparticle in an Electric Field. *Phys. Rev. Lett.* 103(20).
142. Soper AK, Castner EW, Luzar A (2003) Impact of urea on water structure: a clue to its properties as a denaturant? *Biophys. Chem.* 105(2-3):649-666.
143. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98(18):10037-10041.

144. Lee B, Richards FM (1971) Interpretation of Protein Structures - Estimation of Static Accessibility. (Translated from English) *J. Mol. Biol.* 55(3):379-& (in English).
145. Shrake A, Rupley JA (1973) Environment and Exposure to Solvent of Protein Atoms - Lysozyme and Insulin. *J. Mol. Biol.* 79(2):351-371.
146. Connolly ML (1983) Solvent-Accessible Surfaces of Protein and Nucleic-Acids. *Science* 221(4612):709-713.
147. Connolly ML (2006) The molecular surface package, version 3.9.3.
148. Wenzel RN (1936) Rresistance of Solid Surface to Wetting by water. *Ind. Eng. Chem.* 28:988.
149. Anonymous (2010) General Discussion. *Faraday Discuss.* 146:367-393.
150. Daub CD, Wang J, Kudesia S, Bratko D, Luzar A (2010) The influence of molecular-scale roughness on the surface spreading of an aqueous nanodrop. *Faraday Discuss.* 146:67.
151. Willard AP, Chandler D (2009) Coarse-grained modeling of the interface between water and heterogeneous surfaces. *Faraday Discuss.* 141:209-220.

Appendices

Appendix A. Derivations on Van der Waals effects on surface free energy near the curved surfaces

Fowler's Equation represents the surface tension of liquid at a planar interface.

$$\gamma = -\pi \frac{N^2}{V^2} \int_0^\infty dz \int_z^\infty dj \int_f^\infty f df \int_f^\infty g(r)(-\frac{\partial E(r)}{\partial r})dr \qquad \text{(A-1)}$$

Equation (A-1) is derived by the integration over the volume shown in Fig. A1a, where $r > \sigma \Rightarrow g(r) = 1$, and $r < \sigma \Rightarrow g(r) = 0$, *N/V* is the number density of the liquid phase.

If we use the Lennard-Jones potential $E(r) = 4\varepsilon(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6})$ with SPC/E (extended Simple Point Charge) water model parameters $\sigma = 3.166\,\text{Å}$, $E = 0.6502\text{kJ/mol}$, the Van der Waals interactions' contribution to the surface tension of a planar surface is ~28.23 mN/m.

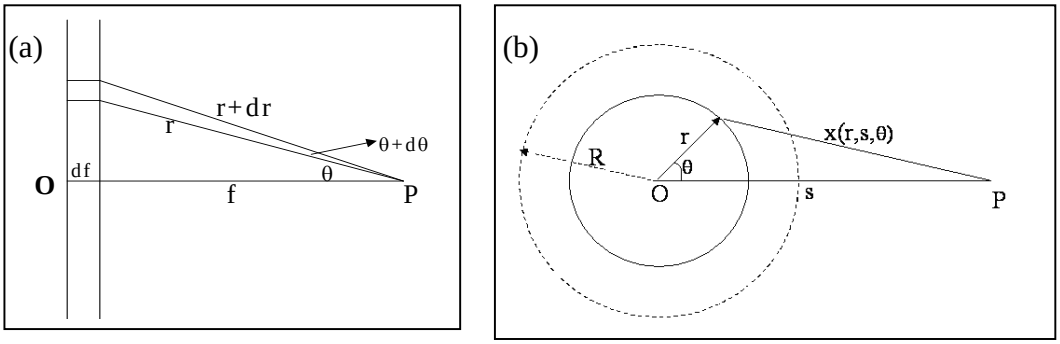


Fig A1. Sketch for the derivation of Fowler's equation (a) and its extension for curved surface (b).

The full analytical derivation of the Van der Waals interaction contribution on the solvation free energy of curved hydrophobic surface follows:

As shown in Fig. A1b,

$$x^2(r,s,\theta) = r^2 + s^2 - 2r \cdot s \cdot \cos\theta \tag{A-2}$$

$$\int_0^R 4\pi r^2 dr \int_R^\infty 2\pi s^2 ds \int_0^\pi \sin\theta \bullet \Phi[x(r,s,\theta)] d\theta \tag{A-3}$$

$$x < \sigma \Rightarrow \Phi(x) = 0$$

$$x > \sigma \Rightarrow \Phi(x) = 4\varepsilon\left(\frac{\sigma^{12}}{x^{12}} - \frac{\sigma^6}{x^6}\right) \tag{A-4}$$

Δρ is the difference in number density of the two phases at the interface, then

$$\gamma(R) = -\frac{\Delta\rho^2 \cdot \pi\varepsilon}{R^2} \int_0^R r^2 dr \int_R^\infty s^2 ds \int_{Max[(s-r),\sigma]}^{(s+r)} \frac{-4}{2rs}\left[\frac{\sigma^{12}}{x^{12}} - \frac{\sigma^6}{x^6}\right] dx \tag{A-5}$$

$$\gamma(R) = -\frac{\Delta\rho^2 \cdot 2\pi\varepsilon}{R^2} \int_0^R r dr \int_R^\infty s ds \int_{MAX[(s-r)^2,\sigma^2]}^{(s+r)^2} \left[\frac{\sigma^{12}}{y^6} - \frac{\sigma^6}{y^3}\right] dy \tag{A-6}$$

$$y = x^2 = s^2 + r^2 - 2sr\cos\theta$$

$$\gamma(R) = \frac{\Delta\rho^2 \cdot 2\pi\varepsilon\sigma^2}{R^2} \int_0^R r dr \int_R^\infty s ds \bullet \left[\frac{\sigma^{10}}{5x^5} - \frac{\sigma^4}{2x^4}\right]\Big|_{x=MAX[(s-r)^2,\sigma^2]}^{x=(s+r)^2} \tag{A-7}$$

$$\gamma(R) = \frac{\Delta\rho^2 \cdot 2\pi\varepsilon\sigma^2}{R^2} \int_0^R r dr \int_R^\infty s ds \bullet \left[\frac{\sigma^{10}}{5(s+r)^{10}} - \frac{\sigma^4}{2(s+r)^4}\right]$$
$$- \frac{\Delta\rho^2 \cdot 2\pi\varepsilon\sigma^2}{R^2} \int_0^R r dr \int_R^\infty s ds \bullet \left[\frac{\sigma^{10}}{5Max[(s-r)^{10},\sigma^{10}]} - \frac{\sigma^4}{2Max[(s-r)^4,\sigma^4]}\right] \tag{A-8}$$

$$\int_0^R r dr \int_R^\infty s ds \bullet \left[\frac{\sigma^{10}}{5(s+r)^{10}} - \frac{\sigma^4}{2(s+r)^4}\right] = \frac{\sigma^4}{720}\left[30\ln 2 - 15/2 - 41\sigma^6/256R^6\right] \tag{A-9}$$

① For $R > \sigma$, separate the integrals into several parts:

1. $s - r > \sigma \Rightarrow s > r + \sigma, r < s - \sigma$

$$\int_0^{R-\sigma} rdr \int_R^{\infty} sds \bullet [\frac{\sigma^{10}}{5(s-r)^{10}} - \frac{\sigma^4}{2(s-r)^4}]$$
$$= \frac{\sigma^4}{720}[30\ln(R/\sigma) - \frac{1}{6}(1-(\frac{\sigma}{R})^6) + 29\frac{R}{\sigma} - 29(\frac{R}{\sigma})^2] \quad \text{(A-11)}$$

$$\int_{R-\sigma}^{R} rdr \int_{r+\sigma}^{\infty} sds \bullet [\frac{\sigma^{10}}{5(s-r)^{10}} - \frac{\sigma^4}{2(s-r)^4}]$$
$$= \frac{\sigma^4}{720}[\frac{139}{6} - 29\frac{R}{\sigma} - 52(R/\sigma)^2] \quad \text{(A-12)}$$

2. $s - r < \sigma \Rightarrow s < r + \sigma$

$$\int_{R-\sigma}^{R} rdr \int_R^{r+\sigma} sds \bullet [\frac{\sigma^{10}}{5\sigma^{10}} - \frac{\sigma^4}{2\sigma^4}] = \frac{\sigma^4}{720}[\frac{9}{2} - 54(\frac{R}{\sigma})^2] \quad \text{(A-13)}$$

$$\gamma(R) = \pi\varepsilon\sigma^4\Delta\rho^2[\frac{3}{4} - \frac{\sigma^2}{18R^2} \bullet (2 + 3\ln\frac{2R}{\sigma}) - \frac{\sigma^8}{27648R^8}] \quad \text{(A-14)}$$

② For $R < \sigma$, Do it similarly:

1. $s - r > \sigma \Rightarrow s > r + \sigma, r < s - \sigma$

$$\int_0^{R} rdr \int_{r+\sigma}^{\infty} sds \bullet [\frac{\sigma^{10}}{5(s-r)^{10}} - \frac{\sigma^4}{2(s-r)^4}] = \frac{9}{80}\sigma^2R^2 + \frac{13}{270}\sigma R^3 \quad \text{(A-15)}$$

2. $s - r < \sigma \Rightarrow s < r + \sigma$

$$\int_0^{R} rdr \int_R^{r+\sigma} sds \bullet [\frac{\sigma^{10}}{5\sigma^{10}} - \frac{\sigma^4}{2\sigma^4}] = -\frac{3}{80}R^4 - \frac{\sigma R^3}{10} - \frac{3}{40}\sigma^2R^2 \quad \text{(A-16)}$$

$$\gamma(R) = \pi\varepsilon\sigma^4\Delta\rho^2[\frac{3}{80} + \frac{\sigma^2}{48R^2} \bullet (4\ln 2 - 1) - \frac{\sigma^8}{92160R^8} - \frac{3R^2}{40\sigma^2} - \frac{7R}{135\sigma}] \quad \text{(A-17)}$$

Appendix B. Non-Linear Poisson–Boltzmann equation

The Poisson–Boltzmann (PB) equation is a differential equation that describes electrostatic interactions in ionic solutions:

$$\frac{1}{r^2}\frac{d}{dr}(r^2\frac{d\Psi}{dr}) = -\frac{e_0}{\varepsilon}[n_+^0 \exp(-\frac{e_0\Psi}{kT}) - n_-^0 \exp(+\frac{e_0\Psi}{kT})] \qquad \text{(B-1)}$$

r is the radial coordinate, Ψ is the electrostatic potential, e_0 is the elementary charge, k is the Boltzmann constant, T is the absolute temperature, and n_+^0 and n_-^0 are the number densities of simple ions in the reference point of the electrostatic potential.

The system consists of a macro-ion with charge z_i and radius R_i, in a spherical cell with a radius R_p, the boundary conditions are set as

$$(\frac{d\Psi}{dr})_{r=R_i} = \frac{z_i e_0}{4\pi\varepsilon R_i^2} \quad \text{and} \quad (\frac{d\Psi}{dr})_{r=R_p} = 0 \qquad \text{(B-2)}$$

The equation B-1 is solved numerically in both aqueous and gas phase systems (Fig. 17 in main text). The reference concentration n_+^0 and n_-^0 were not know in advance, their values were determined by a trial and error procedure to satisfy the boundary conditions and yield the desired ion concentration c_i as

$$c_i = \frac{n_i^0}{VN_A}\int_V \exp(\frac{-e_0\Psi + u_{iP}}{kT})dV \qquad \text{(B-3)}$$

where V is the volume of the system, N_A is Avogadro number, u_{iP} is the hard core potential $u_{iP} = \infty$ when $r_{iP} < R_i$ and zero otherwise.

The values of free energies of aqueous system is quite small compared to the gas

phase system, and when we take the difference between them, we get back the results predicted by the simplified Born equation:

$$\Delta G = (1 - \frac{1}{\varepsilon}) \frac{(ze_0)^2}{2R_i} \qquad \text{(B-4)}$$

The mean field method failed to give satisfied predictions of solvation free energies due to the difficulties discussed in main text (Section 3.2)

Appendix C. Supplemental module to calculate solvation free energy in DL_POLY package

Module tielec_module

```
c*********************************************************************
c       Fortran 90 format
CC      dl_poly module for calculating electrostatic energy for thermodynamic
integration!
CC    Using Ewald Sum, similar to A-T book
CC      Jihang Wang
c       2008/04/12
c*********************************************************************
        use setup_module
        use config_module
        Contains
C*********************************************************************
        Subroutine ti_driver (natms,Etotal,Eion,Ewater)
c*********************************************************************
c       dl_poly subroutine for thermodynamic integration quantity
c       for Electrostatic interaction only, Ewald Summation Method
c*********************************************************************
        Implicit None
        Integer nsol, ii,i,j,k,l,m,n,natms
        Double Precision Ewater, Etotal, Eion
        Double Precision EwaterR, EtotalR, EionR
        Double Precision EwaterK, EtotalK, EionK
        Double Precision KAPPA,boxl,fact1, rcut
```

```
      Double Precision FUNCTION ERFC
      Integer natms
      Parameter (fact1=1389.35386)
      Integer Nw,Ni
      nsol=0
      Ewater=0.0D0
      Etotal=0.0D0
      Eion=0.0D0
      EwaterK=0.0D0
      EtotalK=0.0D0
      EionK=0.0D0
      EwaterR=0.0D0
      EtotalR=0.0D0
      EionR=0.0D0
      rcut=9.0
      Do ii=1,natms
      If(atmnam(ii).EQ."OW".OR.atmnam(ii).EQ."HW"
     & .OR.atmnam(ii).EQ."SP".OR.atmnam(ii).EQ."SN") Then
      nsol=nsol+1
      Endif
      Enddo
      Nw=nsol
      Ni=nsol+1
      boxl=cell(1)
      KAPPA=5.0/boxl
CCCC    water-water interaction
        Call RWALD (EwaterR,1,Nw,Nw)
```

```
            Call KWALD (1,Nw,EwaterK)
            Ewater=(EwaterR+EwaterK)*fact1
CCCC    Ion-Ion interaction
            Call RWALD (EionR,Ni,natms,Nw)

            Call KWALD (Ni,natms,EionK)
            Eion=(EionR+EionK)*fact1
CCCC    Total Electrostatic Energy
           Call RWALD (EtotalR,1,natms,Nw)
           Call KWALD (1,natms,EtotalK)
           Etotal=(EtotalR+EtotalK)*fact1
          Write(111,*) Nw,Ni,natms
           return
           CONTAINS
C*****************************************************************
SUBROUTINE KWALD (N1,N2,VK)
            Implicit None
            Double Precision VD,twopi,Ksq
            Double Precision Cumcos,Cumsin
            Double Precision RXI,RYI,RZI,ZI
            Double Precision RSQPI, VS, VK, coeff,kri
            Integer N1,N2,II,I,J,m,i1,j1,m1,ix,ij,iz,iy
            Double Precision kx(-8:8), ky(-8:8), kz(-8:8)
            Parameter(twopi=6.2831852, RSQPI=0.56419)
            VD=0.0D0
            VK=0.0D0
            VS=0.D0
```

```
      Do i1=-8,8
      kx(i1)=DBLE(i1)*twopi/boxl
      ky(i1)=DBLE(i1)*twopi/boxl
      kz(i1)=DBLE(i1)*twopi/boxl
      Enddo
       coeff=0.0D0
       Do ix=-8,8
            Do iy=-8,8
               Do iz=-8,8
      If(ix.EQ.0.AND.iy.EQ.0.AND.iz.EQ.0) Goto 100
      Ksq=kx(ix)*kx(ix)+ky(iy)*ky(iy)+kz(iz)*kz(iz)
      coeff=twopi/Ksq*exp(-Ksq/(4.*KAPPA*KAPPA))
       Cumcos=0.D0
       Cumsin=0.D0
       DO 123 I = N1, N2
            RXI = xxx(I)
            RYI = yyy(I)
            RZI = zzz(I)
            ZI    = chge(I)
      kri=kx(ix)*RXI+ky(iy)*RYI+kz(iz)*RZI
      Cumcos=Cumcos+ZI*COS(kri)
      Cumsin=Cumsin+ZI*SIN(kri)
123      Continue
      VD=VD+coeff*(Cumcos*Cumcos+Cumsin*Cumsin)
100      Continue
                Enddo
               Enddo
```

```
          Enddo
        VD=VD/boxl**3
         DO 25 II = N1, N2
             VS = VS + chge(II) * chge(II)
25       CONTINUE
         VS = RSQPI * KAPPA * VS
         VK = VD - VS
        Return
         End Subroutine KWALD

C*****************************************************************
SUBROUTINE RWALD ( VR,N1,N2,Nw)
    Implicit None
         INTEGER    N1,N2, I, J,Nw, Nmi,Nmj
         Double Precision    VR,ROX, ROY, ROZ
         Double Precision    RXI, RYI, RZI, ZI, RXIJ, RYIJ, RZIJ
         Double Precision    RIJSQ, RIJ, KRIJ,    VIJ
         VR = 0.0D0
         DO 100 I = N1, N2-1
          Nmi=0
         If (I.LE.Nw) Nmi=INT((I-1)/3)+1
             RXI = xxx(I)
             RYI = yyy(I)
             RZI = zzz(I)
             ZI    = chge(I)
             DO 99 J = I+1, N2
            Nmj=0
```

```
            If (J.LE.Nw) Nmj=INT((J-1)/3)+1
            If (Nmi.EQ.Nmj.AND.Nmi.NE.0) Goto 99
                  RXIJ = RXI - xxx(J)
                  RYIJ = RYI - yyy(J)
                  RZIJ = RZI - zzz(J)
               ROX=RXIJ
               ROY=RYIJ
               ROZ=RZIJ
                  RXIJ = RXIJ - boxl*ANINT(ROX/boxl)
                  RYIJ = RYIJ - boxl*ANINT(ROY/boxl)
                  RZIJ = RZIJ - boxl*ANINT(ROZ/boxl)
                  RIJSQ = RXIJ * RXIJ + RYIJ * RYIJ + RZIJ * RZIJ
                  RIJ     = SQRT(RIJSQ)
              If (RIJ.LT.rcut) Then
                  KRIJ    = KAPPA * RIJ
                  VIJ     = ZI*chge(J) * ERFC(KRIJ)/RIJ
                  VR      = VR + VIJ
               Endif
99             CONTINUE
100       CONTINUE
            RETURN
            END Subroutine RWALD
C*********************************************************************
         End Subroutine ti_driver
C*********************************************************************
Double Precision FUNCTION ERFC ( X )
*********************************************************************
```

```
** APPROXIMATION TO THE COMPLEMENTARY ERROR FUNCTION
C       ** REFERENCE:
C       ** ABRAMOWITZ AND STEGUN, HANDBOOK OF MATHEMATICAL
FUNCTIONS,       **
C       **       NATIONAL BUREAU OF STANDARDS, FORMULA 7.1.26
*****************************************************************
            Double Precision      A1, A2, A3, A4, A5, P
            PARAMETER ( A1 = 0.254829592, A2 = -0.284496736 )
            PARAMETER ( A3 = 1.421413741, A4 = -1.453152027 )
            PARAMETER ( A5 = 1.061405429, P    =    0.3275911      )
            Double Precision      T, X, XSQ, TP
 *****************************************************************
            T    = 1.0 / ( 1.0 + P * X )
            XSQ = X * X
            TP = T * ( A1 + T * ( A2 + T * ( A3 + T * ( A4 + T * A5 ) ) ) )
            ERFC = TP * EXP ( -XSQ )
            RETURN
            END Function ERFC
CC*****************************************************************
Double Precision FUNCTION VothF ( RR )
*****************************************************************
C       Voth's idea for Ewald alternative for coulomb Force!
C** REFERENCE:      JPCB 112, 4711 (2008)
*****************************************************************
      Implicit None
            Double Precision    A0, A1, A2, A3, A4, A5, A6, A7, RR,R
            PARAMETER (A0=-0.165477570871D-3,A1=0.288823451703D-3)
```

```
      PARAMETER (A2=-0.122247561247D-3,A3=0.963712701767D-5 )
      PARAMETER (A4=0.251954672874D-6, A5=-0.735796273353D-7)
      PARAMETER (A6=0.353601771929D-8, A7=-0.525765995765D-10)
      Double Precision      C1, R2,R3,R4,R5,R6,R7
      parameter (C1=5.29177249D-1)
C     have to convert length into atomic units (C1)
      VothF=0.D0
      R=RR/C1
      R2=R*R
      R3=R2*R
      R4=R2*R2
      R5=R4*R
      R6=R3*R3
      R7=R6*R
      VothF=A0+A1*R+A2*R2+A3*R3+A4*R4+A5*R5+A6*R6+A7*R7
      Return
          END Function VothF
CC*****************************************************************
      Double Precision FUNCTION VothE ( RR )
   *****************************************************************
C     Voth's idea for Ewald alternative For Coulomb Energy!
C** REFERENCE:     JPCB 112, 4711 (2008)
*****************************************************************
      Implicit None
          Double Precision    A0, A1, A2, A3, A4, A5, A6, A7, R,RR
          PARAMETER (A0=-0.165477570871D-3,A1=0.288823451703D-3)
          PARAMETER (A2=-0.122247561247D-3,A3=0.963712701767D-5 )
```

```
      PARAMETER (A4=0.251954672874D-6, A5=-0.735796273353D-7)
      PARAMETER (A6=0.353601771929D-8, A7=-0.525765995765D-10)
      Double Precision      C1,R2,R3,R4,R5,R6,R7,R8
      parameter (C1=5.29177249D-1)
C     have to convert length into atomic units (C1) and converted back later (C2)
      VothE=0.D0
      R=RR/C1
      R2=R*R/2.D0
      R3=R2*R/3.D0
      R4=R2*R2/4.D0
      R5=R4*R/5.D0
      R6=R3*R3/6.D0
      R7=R6*R/7.D0
      R8=R4*R4/8.D0
      VothE=A0*R+A1*R2+A2*R3+A3*R4+A4*R5+A5*R6+A6*R7+A7*R8
      VothE=VothE*C1
      Return
      END Function VothE
CC*****************************************************************
      END module tielec_module
```

Appendix D. Generalized Kelvin Equation calculations of different geometries

Confinement affects the phase behavior of liquid water. If the separation of two finite size hydrophobic walls (contact angle greater than 90^o) is smaller than the critical distance predicted by the Kelvin Equation, the water will spontaneously evaporate.

Grand Potentials for liquid and vapor phases:

Liquid phase: $\Omega_L \sim -PV + A_W \gamma_{wl}$ (C-1)

Vapor phase: $\Omega_V \sim -P_V V + A_W \gamma_{wv} + A_\gamma$ (C-2)

P is the pressure, A is the coexistence area, and A_w is the confined wall surface area, V is the volume of the confined region

Surface tension: $\Delta\gamma \equiv \gamma_{wl} - \gamma_{wv} = -\gamma \cos\theta_C$ (C-3)

θ_c is the contact angle of water on the surface

$$P - P_V = \rho\Delta\mu \sim 1atm \quad \text{(C-4)}$$

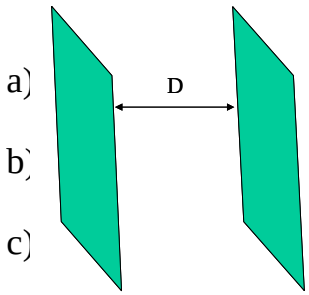
ρ is the number density, and μ is the chemical potential

Capillary evaporation requires $\Omega_V < \Omega_L$

$$V \cdot \rho\Delta\mu + A_W \gamma \cos\theta_c + A\gamma < 0 \quad \text{(C-5)}$$

Let D_c be the critical distance for the capillary evaporation:

Appendix D1. The confinement between two planar surfaces:



a)

b)

c)

a)

$$A_w = 2 \cdot L^2$$
$$A = 4L \cdot D$$
$$V = L^2 \cdot D$$

L is the lateral size of plates, and D is the distance between plates, from equation C-5,

$$D_c < \frac{-2\gamma \cos\theta_c}{\rho\Delta\mu + 4\gamma / L} \qquad \text{(C-6)}$$

a) Contact angle has to be larger than 90^o, $\cos\theta_C < 0$, to make D_c positive.

b) When $L \to \infty$, Critical distance $D_c \propto \cos\theta_c$

c) When lateral size of the plates L is small (several nanometers or less)

Critical distance $D_C \propto L \cdot \cos\theta_C$,

If $L = 2\ nm$ (similar to the mellittin dimer size),

and $\cos\theta_C \sim -0.4$ (contact angle ~ 113^o)

The critical distance D_c is ~4 *Å*

Appendix D2. The confinement inside a hemisphere surface:

$$A_W = 2\pi R^2$$
$$A = \pi R^2$$
$$V = \frac{2}{3}\pi R^3$$

R is the radius of the hemisphere

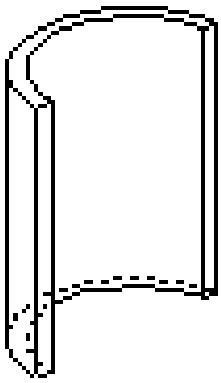
$$R_c < \frac{-3(1+2\cos\theta_C)\gamma}{2\rho\Delta\mu} \qquad \text{(C-7)}$$

Contact angle has to be larger than 120°, $\cos\theta_c < -\frac{1}{2}$, to make *R* positive

At the extreme condition $\cos\theta_c = -1$ $(\theta_C = 180^o)$

Critical radius R_c ~ 1 μm

Appendix D3. The confinement in a half-cylindrical surface:



$$V = \frac{1}{2}\pi R^2 \cdot L$$

$$A_W = \pi R \cdot L$$

$$A = 2R \cdot L$$

L is the length of the cylinder, and *R* is the radius

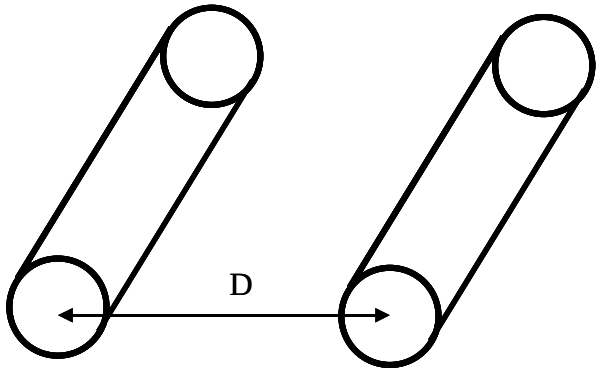
$$R_c < \frac{-2\gamma(2 + \pi\cos\theta_c)}{\pi\rho\Delta\mu} \qquad \text{(C-8)}$$

Contact angle has to be larger than 129.5°, $\cos\theta_c < -\frac{2}{\pi}$

At the extreme condition $\cos\theta_C = -1 \quad (\theta_C = 180^o)$

Critical radius R_c ~ 0.5 μm

Appendix D4. The confinement between two cylinders



$$V = (2D \cdot R - \pi R^2) \cdot L$$
$$A = 2D \cdot L + 2 \cdot (2D \cdot L - \pi R^2)$$
$$A_W = 2\pi R \cdot L$$

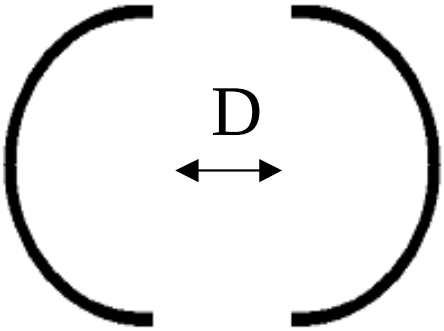
D is the distance between two cylinders, *R* is the radius, and *L* is the length of each cylinder.

If $L \to \infty$, $$D < \frac{\pi R^2 \rho \Delta\mu - 2\pi R\gamma \cos\theta_c}{2\rho\Delta\mu \cdot R + 2\gamma} \qquad \text{(C-9)}$$

If the contact angle is ~113°, $\cos\theta_C \sim -0.4$

the size of the cylinder tube ~ 5 Å, Critical distance D_c~4.4 Å

Appendix D5. The confinement between two hemispheres:



D is the distance between two centers

R is the radius

$$V = \frac{4}{3}\pi R^3 + \pi R^2 \cdot D$$

$$A = 2\pi R \cdot D$$

$$A_W = 4\pi R^2$$

$$D < -\frac{4R\gamma\cos\theta_C + \frac{4}{3}R^2\rho\Delta\mu}{2\gamma + R\rho\Delta\mu} \qquad \text{(C-10)}$$

If R is in the nanometer range, $D_C \propto -R\cos\theta_C$

For example, if the contact angle is ~113^o, $\cos\theta_C$ ~ –0.4

And the radius of the hemispheres ~ 10 Å

Critical distance D_c~ 8 Å

Appendix E. Additional protein systems in capillary evaporation studies

PDB ID	Category	N_i	N_r	$\frac{< N^2 > - < N >^2}{< N >}$	N_r/N_i%
1hsi	dimer	115	98.6	0.28	86
1J30	dimer	135	115.7	0.25	86
1i4s	dimer	107	83.8	0.66	78
1jvl	dimer	126	109.4	0.24	87
1cmb	dimer	102	67.5	0.65	66
1jr8	dimer	111	81.6	0.73	74
1hul	dimer	108	94.6	0.26	88
1ipi	dimer	109	75.7	0.65	69
1bja	dimer	108	97.8	0.15	91
1k94	dimer	113	91.0	0.46	81
1bj3	dimer	112	101.0	0.18	90
1gfw	dimer	131	73.8	1.91	56
1bbh	dimer	115	86.4	0.74	75
1eyv	dimer	109	85.8	0.58	79
1ub3	tetramer	161	113.0	1.22	70
1tvx	tetramer	129	116.8	0.21	91
1xz4	tetramer	122	105.0	0.00	86
1tlf	tetramer	119	112.9	0.08	95
1plf	tetramer	116	101.8	0.29	88
4aah	tetramer	152	130.7	0.30	86

PDB ID	Category	N_i	N_r	$\frac{< N^2 > - < N >^2}{< N >}$	N_r/N_i%
1han	multi-domain	114	73.7	1.97	65
1plq	multi-domain	114	73.8	1.94	65
5ldh	multi-domain	117	87.1	0.90	74
1a5z	multi-domain	114	94.6	0.62	83
1mdr	multi-domain	110	80.5	0.79	73
1pgs	multi-domain	118	97.9	0.38	83
1boh	multi-domain	120	89.7	0.90	75
1bg5	multi-domain	111	89.3	0.53	80
1akl	multi-domain	128	102.3	0.43	80
1cne	multi-domain	127	108.7	0.33	86
1hyt	multi-domain	121	98.8	0.45	82
1clc	multi-domain	124	96.9	1.01	78
1aco	multi-domain	133	115.1	0.33	87
1dhy	multi-domain	120	72.0	2.17	60
1cpo	multi-domain	113	85.3	2.05	75
1pkp	multi-domain	118	86.4	0.62	73
1ldm	multi-domain	113	71.4	1.82	63
1bli	multi-domain	115	88.6	0.39	77
2mbr	multi-domain	112	99.4	0.23	89

Appendix F. Grand Canonical Monte Carlo (GCMC) simulations on surface tension

In grand canonical ensemble, the system volume, temperature and chemical potentials of the species are kept constant by allowing particle insertions and deletions. The acceptance probabilities for attempted additions and deletions are given by .

$$P_{N\to N+1} = \min\{1, \frac{<N>}{(N+1)} e^{\beta(\mu^{ex}-\Delta U)}\}$$
$$P_{N\to N-1} = \min\{1, \frac{N}{<N>} e^{\beta(-\mu^{ex}-\Delta U)}\} \qquad \text{(F-1)}$$

In the equation above, μ_{ex} is the excess chemical potential of a molecule, ΔU is the energy change upon addition or deletion of a molecule, and $<N>$ is the average number of molecules in the system, $\beta = kT$ and k is the Boltzmann constant, and T is the absolute temperature.

We use the GCMC method to study the water surface tension on different types of surfaces described in Chapter 5 with thermodynamic integration (See section 2.1.4). In our simulation, the water molecules are represented by the SPC/E model, chemical potential μ is carefully tuned to be -12.52kT with a simulation box of 19.75*19.75*19.75 $Å^3$ in size.

The surface tension is calculated in equation (F-2), where S is the surface area, and U is the energy between water and surface.

$$\Delta\gamma = \frac{1}{S}\int d\lambda \left(\frac{\partial U}{\partial \lambda}\right)_{\mu VT} \qquad \text{(F-2)}$$

The reference system is set to be a smooth hydrophobic surface as described in (Bratko *et al*, *J. Am. Chem. Soc.* 129, 2504, **2007**), the interaction parameters of atoms on the surface are gradually turned on by the coupling parameter λ, while the interactions parameter of the smooth surface is gradually turned off by the conjugated factor $(1-\lambda)$ simultaneously.

We record wetting surface free energy σ of the graphene-like surface (described in Chapter 5) to be 19.3 mN/m, and from the relation between the contact angle and surface tension $\theta_c \sim -\cos^{-1}(\sigma / \gamma_{lv})$, the contact angle on this surface is ~110^o, in good agreement with our molecular simulation studies.

Vita

Jihang Wang was born on Nov, 9th 1981, in Changzhou, Jiangsu Province, China.

Education:

Graduate student **(Ph.D** expected in *December, 2010***)**, Chemistry
Department of Chemistry, Virginia Commonwealth University, Richmond, VA
Advisor: Professor Alenka Luzar

Bachelor of Science, Chemistry, *Nanjing University, Nanjing, China – July, 2003*

Experience:

Teaching Assistant, *Department of Chemistry, VCU, 08/2003 - 12/2009*

Research Assistant, *Department of Chemistry, VCU, 01/2004-07/2005*
- Studies in ultrafast (femto-second) nonlinear spectroscopy, focusing on three-pulse photo echo experiments in membranes.

Advisor: Dr. Darius Kuciauskas

Research Assistant, *Department of Chemistry, VCU, 08/2005 - present*
- Studies in statistical mechanics, theoretical and computational chemistry, focusing on modelling hydrophobic effects at different lengthscales

Advisor**:** Dr. Alenka Luzar (support: National Science Foundation)

Awards:

✧ Recipient of the *Altria (Philip Morris) Research Fellowship*, *Spring, 2010*.

✧ Graduate fellowship in *Virginia Commonwealth University*, *Fall, 2010*.

Skills:

- Molecular dynamics and Monte Carlo simulations to study thermodynamic and structural aspects of condensed phase systems.

- Information technology skills, including: Microsoft Office, FORTRAN and C programming, mathematical software including Mathematica, Matlab, SAS, molecular simulations packages DL_POLY, LAMMPS, and molecular visualization program VMD (Visualize Molecular Dynamics).

- Operating ultrafast (femtosecond) laser apparatus

Scientific contributions:

Publications

C. D. Daub, **J. Wang**, S. Kudesia, D. Bratko and A. Luzar, The influence of molecular-scale roughness on the surface spreading of an aqueous nanodrop, *Faraday Discussion* 2010, **146**, 67-77

J. Wang, D. Bratko and A. Luzar, Probing surface tension additivity on chemically heterogeneous surfaces: a molecular approach, under review in *Proceedings of the National Academy of Science, U.S.A.*

J. Wang, S. Kudesia, D. Bratko and A. Luzar, Computational probe of dewetting events in protein systems, in preparation

J. Wang, D. Bratko and A. Luzar, Hydrogen bonds contribution to crossover lengthscales of hydrophobic hydration, in preparation

M. von Domaros, **J. Wang**, D. Bratko and A. Luzar, Dynamics at a Janus Interface, in preparation.

J. Seyed Yazdi, **J. Wang**, D. Bratko and A. Luzar, Switchable Nano-Wetting Dynamics, in preparation.

J. Ritchie, **J. Wang**, D. Bratko and A. Luzar, What Determines Contact Angle of a Nanodrop on a Heterogenous Surface?, in preparation.

Presentations (presenting author first)

A. Luzar, **J. Wang**, C. D. Daub and D. Bratko "*Modeling of hydrophobic interactions*", American Chemical Society Colloid and Surface Science Symposium, Columbia University, NY, June 14-19, 2009, invited talk, keynote

J. Wang, D. Bratko and A. Luzar, "*Probing surface tension additivity on chemically heterogeneous surfaces: a molecular approach"*, Division of Colloid and Surface Chemistry, 239th American Chemical Society Meeting, San Francisco, CA, March 21-25, 2010, contributed talk

J. Wang, D. Bratko and A. Luzar, "*Wetting Free Energy on Heterogeneous Surfaces: from Synthetic to Biological*", poster presentation, Faraday Discussion 146, in Richmond, VA, April 11-14, 2010.