

PHASE EQUILIBRIUM IN PROTEIN MIXTURES: MESOSCOPIC STUDY OF
LYSOZYME/ALPHA-LACTALBUMIN

by
Anil Kurut

Submitted to the Institute of Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science
In Chemical Engineering

Yeditepe University
2010

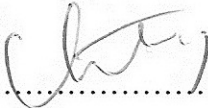
PHASE EQUILIBRIUM IN PROTEIN MIXTURES: MESOSCOPIC STUDY OF
LYSOZYME/ALPHA-LACTALBUMIN

APPROVED BY:

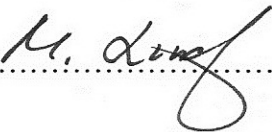
Assist. Prof. Seyda Bucak
(Supervisor)


.....

Assist. Prof. Nihat Baysal


.....

Assist. Prof. Mikael Lund


.....

DATE OF APPROVAL: 25/06/2010

ACKNOWLEDGEMENT

I am so thankful to my dear thesis supervisor Assist. Prof. Seyda Bucak and unofficial co-supervisor Assist. Prof Mikael Lund for not only their technical support but also their valuable efforts for correcting my thesis.

In addition, I want to thank so much to Björn Persson for his unlimited patients of my infinitely many questions about running simulations and clusters.

I also like to thank the committee members of my thesis for their illuminative suggestions and critics. And also I wish to thank all the members of Chemical engineering departments and Theoretical Chemistry department in Lund University for their contribution on my background.

My deepest thank is to my boy friend Korhan Şabanoğlu for being beside me at any moment and his great contributions to figures of the thesis.

ABSTRACT

PHASE EQUILIBRIUM IN PROTEIN MIXTURES: MESOSCOPIC STUDY OF LYSOZYME/ α -LACTALBUMIN

The phase separation of food proteins resulting from protein self assembly is the key to manipulative control of food quality and to develop novel methods of designing food products. Thus, understanding factors like solution condition behind the phase transition of these proteins is of considerable interest. In this project, the phase transition in the aqueous solution of two oppositely charged food proteins, lysozyme (LYS) and two forms of α -lactalbumin (α -LA) has been studied under various ionic strengths. The solution has been described at a mesoscopic level. Two novel coarse grained approaches, the Surface Charge (SC) and the Twin Sphere (TS) models, have been developed to represent the structure of the proteins. The interactions between proteins including both electrostatic and van der Waals terms have been investigated using Monte Carlo simulations. The solvent has been treated as a structureless dielectric continuum and salt particles have been presented either explicitly or implicitly. The phase transition characteristics of LYS/ α -LA have been captured in excellent agreement with experimental observations by the SC model in which the surface charge distribution of proteins is maintained. As expected, the system which contains LYS and calcium loaded form of α -LA (holo α -LA) does not form any phase transition regardless of salt concentration. While LYS/calcium depleted apo α -LA shows two phase region under low salt concentrations and the formation of aggregates is completely disappeared at 100 mM salt. In addition, the effect of α -LA alignment with respect to LYS on self assembly has been investigated using the modified SC model. The modified model is constructed by redistributing surface charges of α -LA randomly, thereby drastically reducing its dipole moment. Although phase transition is observed when the dipole moment of α -LA is reduced, the dilute phase becomes more favorable resulting from the decrease in the attraction between LYS and α -LA.

ÖZET

PROTEIN KARIŞIMLARINDAKİ FAZ DENGELERİ: LYSOZYME/ALPHA-LACTALBUMİN'İN MESOSKOPIK DÜZEYDE CALIŞMASI

Proteinlerin kendiliğinden kümeleşmeye olan eğilimleri sonucu ortaya çıkan faz değişimleri, gıda kalitesi yönetiminde ve yeni gıda ürünleri geliştirmede büyük bir öneme sahiptir. Bu çalışmada, iki gıda protein olan Lysozyme (LYS) ve α -lactalbumin'in (α -LA), sulu tuz çözeltisi içerisindeki kümeleşmeden kaynaklı faz değişimi, Monte Carlo simülasyonları kullanılarak incelenmiştir. Protein yapılarını basitleştirmek adına iki yeni indirgenmiş model geliştirilmiştir; Yüzey Yükü (SC), İkiz Küreler (TS). Yüzey yükü modeliyle elde edilen faz değişimi karakteristikleri deneysel çalışmalarla birebir örtüşmektedir. Beklenildiği üzere, LYS ve kalsiyum bağlanmış (holo) α -LA içeren karışımda, tuz konsantrasyonuna bağlı olmaksızın faz değişimi gözlenmemiş, LYS/kalsiyumsuz (apo) α -LA karışımında ise sadece düşük tuz konsantrasyonlarında iki fazın denge hali gözlemlenebilmiştir. Ayrıca, α -LA'nın taşıdığı ikiz kutup (dipol) vektörün LYS'ye yönelmesinden dolayı oluşan yapılanmanın kümeleşmeye olan etkisini incelemek için, α -LA'nın SC modeli mutasyona uğratarak ikiz kutup vektörü yok edilmiştir. LYS/mutasyonlu α -LA karışımında faz değişimi elde edilmiş olmasına rağmen, LYS ve α -LA arasındaki elektriksel çekimin azalmasından kaynaklı, kümeleşmeye olan eğilimde azalma gözlenmiştir.

TABLE OF CONTENTS

| | |
|---|------|
| ACKNOWLEDGMENTS | iii |
| ABSTRACT | iv |
| ÖZET..... | v |
| TABLE OF CONTENTS | vi |
| LIST OF FIGURES | ix |
| LIST OF TABLES | xii |
| LIST OF SYMBOLS/ABBREVIATIONS | xiii |
| 1. INTRODUCTION..... | 1 |
| 2. THEORETICAL BACKGROUND..... | 3 |
| 2.1. AMINO ACIDS..... | 3 |
| 2.1.1. Properties of Amino Acids..... | 4 |
| 2.2. PROTEINS | 5 |
| 2.2.1. Self-assembly of Proteins | 6 |
| 2.2.2. Lysozyme C and α -Lactalbumin..... | 6 |
| 2.3. INTERMOLECULAR INTERACTIONS..... | 9 |
| 2.3.1. Lennard-Jones (L-J) Potential..... | 9 |
| 2.3.2. Electrostatic Interactions..... | 10 |
| 2.3.2.1. <i>Coulomb interaction in vacuum</i> | 10 |
| 2.3.2.2. <i>Coulomb interaction in a dielectric medium</i> | 11 |
| 2.3.2.3. <i>Coulomb interaction in salt solution (Debye-Hückel Theory)</i> .. | 12 |
| 2.3.2.2. <i>Correction of the Debye-Hückel theory</i> | 13 |
| 2.4. BASIS OF MONTE CARLO SIMULATIONS | 14 |
| 2.4.1. Statistical Mechanics and Ensembles | 14 |
| 2.4.1.1. <i>Micro-canonical Ensemble (NVE)</i> | 14 |
| 2.4.1.2. <i>Canonical Ensemble (NVT)</i> | 15 |
| 2.4.1.3. <i>Isobaric-isothermal Ensemble (NPT)</i> | 16 |
| 2.4.1.4. <i>Grand-canonical Ensemble (μVT)</i> | 16 |
| 2.4.2. Free Energy Calculations..... | 17 |
| 2.4.3. The Metropolis Monte Carlo Method..... | 18 |
| 2.4.4. Periodic Boundary Conditions..... | 20 |

| | |
|---|----|
| 2.4.5. Minimum Image Convention..... | 21 |
| 3. METHODOLOGY AND MODELING..... | 22 |
| 3.1. COARSE GRAINING OF THE SOLUTION | 22 |
| 3.1.1. Implicit Solvent and Explicit Salt..... | 22 |
| 3.1.2. Implicit Solvent and Implicit Salt..... | 23 |
| 3.2. COARSE GRAINING OF THE PROTEIN STRUCTURE..... | 23 |
| 3.2.1. Amino Acid Model (AA) | 24 |
| 3.2.2. Surface Charge Model (SC) | 24 |
| 3.2.3. Twin Sphere Model (TS)..... | 24 |
| 3.3. SEPARATION DEPENDENT COARSE GRAINING..... | 26 |
| 3.4. USED PROGRAMS OF FAUNUS | 27 |
| 3.4.1. Isobaric | 27 |
| 3.4.2. Twobody..... | 29 |
| 3.4.3. Grand-canonical Titration | 30 |
| 3.5. SAMPLING EFFICIENCY IMPROVEMENT..... | 31 |
| 3.5.1. Displacement Parameter Optimization..... | 31 |
| 3.5.2. Combined Translational and Rotational Moves..... | 32 |
| 3.5.3. Cluster Moves..... | 32 |
| 4. RESULTS AND DISCUSSIONS | 34 |
| 4.1. ONE BODY SIMULATIONS..... | 34 |
| 4.1.1. pH Dependence of Net-Charges of LYS and α -LA | 34 |
| 4.1.2. Salt Dependence of Net-Charges of LYS and α -LA | 35 |
| 4.2. TWO-BODY SIMULATIONS | 36 |
| 4.2.1. Implicit Salt Treatment..... | 37 |
| 4.2.2. Surface Charge Model..... | 38 |
| 4.2.3. Twin Sphere Model | 40 |
| 4.3. MANY-BODY SIMULATIONS..... | 42 |
| 4.3.1. Size Dependency | 42 |
| 4.3.2. Improvement of Sampling..... | 44 |
| 4.3.3. Reproduction of Experimental Observations | 49 |
| 5. CONCLUSION AND FUTURE WORK..... | 51 |
| 5.1. CONCLUSION | 51 |
| 6.2. FUTURE WORK | 52 |

| | |
|--|----|
| APPENDIX A. CODED AMINO ACIDS | 54 |
| APPENDIX B. NET CHARGES OF LYS AND ALPHA-LA..... | 55 |
| REFERENCES..... | 56 |

LIST OF FIGURES

| | |
|---|----|
| Figure 2.1. The Basic Structure of α -amino acids..... | 3 |
| Figure 2.2. Change in the net charge of an amino acid in a function of pH | 4 |
| Figure 2.3. Molecular Structure of Lysozyme | 7 |
| Figure 2.4. Crystal structure of Apo (PDB ID 1F6R, left) and Holo (PDB ID 1F6S, right) Bovine α -LA..... | 8 |
| Figure 2.5. Lennard-Jones “12-6” Potential..... | 10 |
| Figure 2.6. Two dimensional periodic system | 21 |
| Figure 3.1. Coarse Graining Steps | 25 |
| Figure 3.2. Separation dependent coarse graining and the cut-off radius | 27 |
| Figure 3.3. Alignment of α -LA (red) with respect to LYS (blue)..... | 30 |
| Figure 3.4. Displacement parameter optimization | 32 |
| Figure 4.1. The Net Charge on Lysozyme, Apo and holo α -Lactalbumin at various ... pH in a 30 mM salt solution..... | 35 |
| Figure 4.2. The Net Charge on Lysozyme, Apo and holo α -Lactalbumin at various ... salt concentrations | 36 |
| Figure 4.3. The pmf and the degree of alignment between LYS and the two forms ... of α -LA in the AA model at various salt concentrations for explicit and .. implicit salt treatments | 37 |

| | | |
|--------------|--|----|
| Figure 4.4. | The radius dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and SC models at 5 mM and 100 mM salt | 39 |
| Figure 4.5. | ϵ_{LJ} dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and SC models at 5 mM and 100 mM salt..... | 39 |
| Figure 4.6. | The radius dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and TS models at 5 mM and 100 mM salt. | 40 |
| Figure 4.7. | ϵ_{LJ} dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and TS models at 5 mM and 100 mM salt | 41 |
| Figure 4.8. | The potential of mean force and the degree of alignment between LYS and apo α -LA for the employed coarse graining models at 5mM and 100 mM salt | 41 |
| Figure 4.9. | The free energy of LYS and apo α -LA for various numbers of proteins .. at 5 and 25 mM salt concentration with the SC model..... | 43 |
| Figure 4.10. | The free energy of LYS and apo α -LA for various numbers of proteins .. at 5 and 25 mM salt with TS model..... | 43 |
| Figure 4.11. | Divided window simulation for 30 ideal particles..... | 45 |
| Figure 4.12. | The divided window simulations performed for 20 proteins at 55mM salt | 45 |
| Figure 4.13. | Free energy difference for 20 proteins at 25mM salt and 0.6 mM pressure. | 46 |
| Figure 4.14. | The free energy of 20 proteins system at 25 mM salt concentration with and without displacement parameter scaling at 0.6 (left) and | |

| | |
|---|----|
| 0.7 mM pressure (right) | 47 |
| Figure 4.15. Free energy of system containing 40 proteins at 30 mM salt concentration simulated in 3 windows (solid lines) and 1 window (dashed line)..... | 48 |
| Figure 4.16. Free energy of LYS and α -LA system containing 40 proteins at 30 mM . salt concentration and 0.7 mM pressure..... | 48 |
| Figure 4.17. Free energy differences of LYS- apo (left) and holo (right) α -LA system containing 40 proteins at 30 (black) and 100 mM (red) salt concentration and 0.7 mM pressure | 49 |
| Figure 4.18. The free energy difference of LYS and modified apo α -LA system containing 40 proteins at 30 and 25 mM salt concentration and 0.7 mM pressure | 50 |
| Figure 5.1. The pmf and the degree of alignment between LYS and apo α -LA with a cut off radius 40 Å (red) and without cut off (black)..... | 53 |
| Figure 5.2. Free energy (as a function of box length, left) and radial distribution function (as a function of protein separation, right) of the system containing 10 real (red) and simple (black) proteins at 30 mM salt and..... 0.7 mM pressure with (circles) and without (line) dipole approximation.... | 53 |

LIST OF TABLES

| | |
|--|----|
| Table 3.1. D and κ in a function of 1:1 salt concentration at 298 K | 23 |
| Table 3.2. Number of particles in the examined coarse grained protein structures | 26 |
| Table A.1. Amino acids and their basic properties | 54 |
| Table B.1. Net charges of LYS and α -LA in a function of pH and salt concentration .. | 55 |

LIST OF SYMBOLS / ABBREVIATIONS

| | |
|-----------|--|
| A | Lennard-Jones Repulsion Constant |
| a | contact distance between two charged particles |
| a^* | activity |
| B | Lennard-Jones Attraction Constant |
| D | inverse Debye screening length |
| dp | displacement percent |
| E | energy of subsystems |
| F | mechanical property |
| $g(r)$ | radial distribution function |
| I | ionic strength |
| k | Boltzmann's Constant |
| K | number of Monte Carlo moves |
| K_0 | intrinsic dissociation constant |
| L | length of simulation box |
| l_B | Bjerrum length |
| L-J | Lennard-Jones |
| M | number of subsystems |
| N | number of particles |
| NPT | isothermal-isobaric ensemble |
| NVT | canonical ensemble |
| P | pressure |
| P_i | probability of microstate i |
| p_{ij} | pair-wise probability |
| Q | partition function |
| q_l | charge of molecule |
| Q_{NVT} | canonical partition function |
| R | amino acid side chain |
| $ranf$ | random number between 0 and 1 |
| r_{cut} | cut-off radius |

| | |
|---------------------|--|
| r | separation of the center of masses |
| r^* | separation of the center of masses at contact |
| T | temperature |
| $U(r)$ | interaction energy |
| U_{ij} | potential energy between particle i and j |
| V | volume of simulation box |
| z_j | valency of ion j |
| β | inverse thermal energy, |
| ϵ_0 | dielectric permittivity of vacuum |
| ϵ_{LJ} | depth of L-J potential |
| $\epsilon_{LJ,ave}$ | average depth of L-J potential |
| ϵ_r | dielectric permittivity of a medium |
| θ | angle between dipole vector and z axis |
| κ | inverse Debye screening length |
| μ | chemical potential |
| $\vec{\mu}_z$ | z component of unit dipole vector |
| μVT | grand canonical ensemble |
| ξ_1 | random number between -1 and 1 |
| ξ_2 | random number between 0 and 1 |
| ρ_j | number density of ion j |
| σ | separation between the particles at zero L-J potential |
| σ_{ave} | average separation between the particles at zero L-J potential |
| ψ | potential around central ion |
| $\omega(r)$ | potential of mean force |
| $\Delta\xi$ | maximum allowed displacement |
| Ω_{NVE} | number of available microstate |
| AA | amino acid |
| CG | Coarse grained |
| CPU | Central Processing Unit |
| DH | Debye-Hückel |
| GCA | generalized geometric cluster algorithm |

| | |
|--------------|------------------------------|
| LPB | linearized Poisson-Boltzmann |
| LYS | Lysozyme |
| MC | Monte Carlo |
| PDB | Protein Data Base |
| <i>pmf</i> | potential of mean force |
| SC | surface charge |
| SM | statistical mechanics |
| TS | twin sphere |
| vdW | van der Waals |
| α -LA | α -Lactalbumin |

1. INTRODUCTION

Food systems are colloidal systems that consist of water, biopolymers (proteins and polysaccharides), low molecular weight ingredients (salt, sugar, surfactant, etc.) and colloidal particles (oil droplets or air bubbles). All these components have two main functions; nutritional and structure-forming which play significant role in the quality of food [1].

Natural food products (milk, whey, egg white) contain various proteins whose formation of microstructure through self-assembly determines the texture, mechanical stability, consistency, appearance and taste of food products [1-2]. Thus, studying protein-protein interactions, which mediate self-assembly, have received considerable attention in modern-day research. Most of these researches often focus on irreversible assembly of proteins formed by protein denaturation [3]. In contrast, self-assembly of food proteins into different microstructures without denaturation is still a less studied phenomena [2]. Protein self-assembly strongly depends on physicochemical conditions such as temperature, pH, salt concentration, salt type, and solvent [1-5].

In this work, the self-assembly of two globular, oppositely charged food proteins, hen egg white Lysozyme (LYS) and bovine α -Lactalbumin (α -LA) is computationally investigated under various salt concentrations in order to understand the protein-protein interactions leading to the formation of microstructures. Since Monte Carlo simulations is a powerful computational technique to study phase transitions and critical phenomena [6], it is used to investigate the phase behavior of LYS/ α -LA mixture in aqueous salt solutions.

Although the available computational power has increased tremendously during the years, performing simulations based on a fully atomistic treatment of proteins in an aqueous salt solution is still beyond the present technology [7]. Thus, the atomistic system of interest is gradually coarse grained in order to reduce number of particles.

In this work, coarse graining models are developed to represent solvent, salt particles as well as the structure of proteins. Water molecules are treated by a structureless dielectric

continuum and salt particles are represented at the Debye-Hückel (DH) level by modifying the pair potential between charged particles. Moreover, two different coarse grained (CG) models are developed to describe the protein structure: Surface Charge (SC) and Twin Sphere (TS) models.

In this thesis, the basic information about amino acids and proteins, intermolecular interactions as well as statistical mechanics and Monte Carlo simulations are described in the theoretical background section. In the methodology and models section, the developed coarse grained models are explained comprehensively and details of simulations including the corresponding software and implemented features for the improvement of statistics are described. The results of the simulations are classified according to the simulated number of proteins in the system and results of system containing more than two proteins are presented under many body simulations subsection in the results and discussion section. And finally, the thesis conclusions and potential directions for this work are summarized in the conclusion and future work section.

2. THEORETICAL BACKGROUND

2.1. AMINO ACIDS

There are more than 700 amino acids discovered in nature [8]. Most of these are α -amino acids which have a carboxyl group and an amino group bound to each end of the middle carbon atom, the so called α -carbon [9]. The basic structure of α -amino acids is given below.

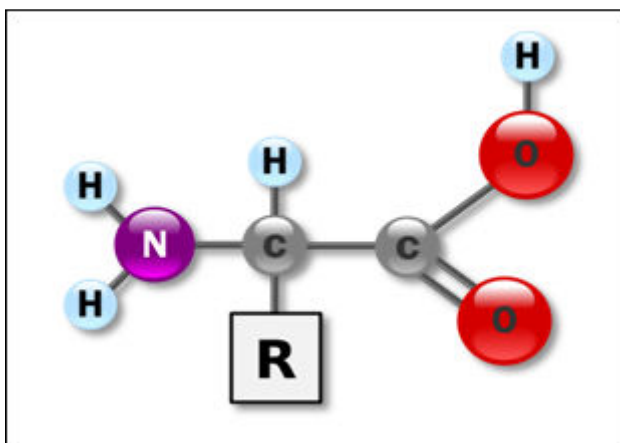


Figure 2.1. The Basic Structure of α -amino acids [10]

Only 20 of these amino acids (19 α -amino acids and one α -imino acid proline) are associated with the synthesis of proteins in living cells [11] and they are generally referred as “coded” or essential amino acids [8].

The amino acids differ by their side chains, R. The simplest amino acid, glycine has a hydrogen atom as its side chain and since the α -carbon carries only three different ligands, it is not chiral. In other words, the mirror images of glycine cannot be superimposed on one another unlike the other amino acids [9, 11]. The basic properties of 20 coded amino acids, including proline, can be seen in Appendix A.

2.1.1. Properties of Amino Acids

The protonation state of amino acids is a function of pH due to the presence of ionizable side chains. Since even the simplest amino acids carry α -carboxylic acid group (α -COOH) and α -amino group (α -NH₂) they have at least two ionizable sites [12]. The α -NH₂ and α -COOH groups have pK_a values around 9-10 and 2-2.5, respectively. Therefore at neutral pH, amino acids carry both negative and positive charges and exist in zwitterionic (dipolar ionic) form [11-12]. At low pH, both of these groups are protonated yielding a net positive charge on the amino acid. At high pH, both groups are deprotonated resulting in a net negative charge. An illustration of this behavior is given in Figure 2.2.

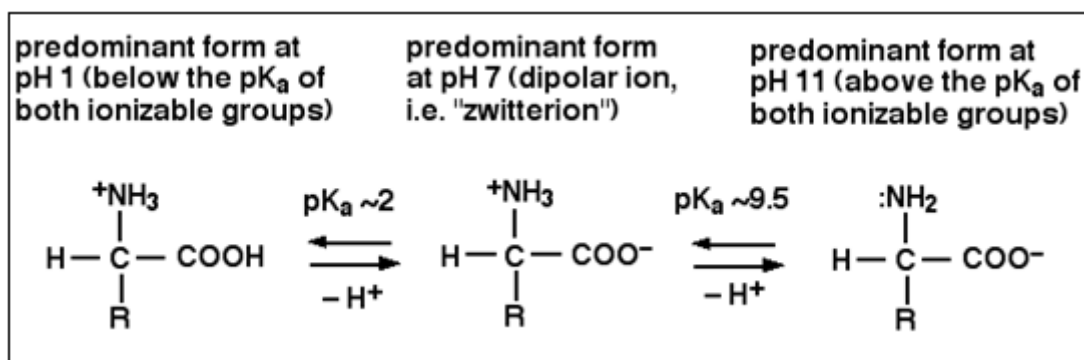


Figure 2.2. Change in the net charge of an amino acid in a function of pH (When the pH increases (from left to right) amino acid transforms from protonated state to deprotonated state and its charge becomes negative) [13]

The pH at which the net charge of an amino acid is zero, is called the isoelectric point (pI). This isoelectric point depends on the number of ionizable (titratable) groups and on their electrostatic interactions [12]. In a hypothetical amino acid containing two α -COOH and two α -NH₂ groups; even if the four titratable sites are distributed symmetrically in the molecule, the pK_a of the first deprotonated group will always be lower than the second one. The pK_a difference is due to intramolecular interactions within the molecule; removing a second proton will be more difficult than the first one because of the decrease in the net positive charge.

2.2. PROTEINS

Proteins are macromolecules which contain up to several thousand amino acids. Like carbohydrates, they contain hydrogen, oxygen and carbon but also nitrogen and sulfur [14]. The amino acid in proteins are linked together by a peptide bond and the sequence of amino acids are determined by the gene coding for the specific peptide chain [15].

The sequence of amino acids in a protein determines its primary structure. This sequence always starts with an N-terminal and ends with a C-Terminal. The hydrogen bonds between amino and keto groups of the peptide bonds stabilize the secondary structure of proteins. The main elements of secondary structure are α -helixes formed by the wounding of backbone (polypeptide chain) around an imaginary axis due to hydrogen bonds and β -strands formed by stretched polypeptide chains. The tertiary structure of a protein describes the conformation of its secondary structure in 3D space determined by ionic and hydrophobic interactions as well as hydrogen and van der Waals attractions [9]. The formation of tertiary structure from amino acid chains is called the “folding process”. The quaternary structure is determined by the arrangement of several polypeptides due non-covalent interactions [9, 15]. This final structure is called the native conformation of protein [16]. Any external influence (heat, certain chemicals) which may lead to deteriorate the 3D structure of the protein is called denaturation [15].

Proteins are generally classified according to their shapes as fibrous or globular proteins. Fibrous proteins are quite elongated and have relatively simple structures. On the other hand, globular proteins have more complex shapes formed by tight folding of the amino acid chain. The diameters of these proteins are usually in the range of 50-100 Å [15].

Proteins have many ionizable groups in the amino acid side chains as well as in the N and C terminals. The protonation state of these groups determines the net charge of the protein according to pH of the environment.

2.2.1. Self-assembly of Proteins

Molecular assembly, which is the spontaneous association of molecules into stable well-defined aggregates due to non-covalent bonds, is the milestone of biological systems [17]. Globular proteins are capable of self-assembly into various supramolecular structures such as aggregates, fibres, and nanotubes. The weak non-covalent attractive interactions are the main driving forces of supramolecular aggregates [1]. In addition, mixtures of oppositely charged globular proteins self assemble into micro-sized spherical aggregates under specific physicochemical conditions. It is shown that the microspheres is a universal form of self assembly of binary-mixtures of proteins when the size and charge compensations between proteins are ensured [4].

The formation of aggregates can be interpreted theoretically in terms of a gas-liquid phase separation. The liquid dilute in protein is referred as gas phase and the dense liquid rich in protein and aggregates is referred as liquid phase [18].

2.2.2 Lysozyme C and α -Lactalbumin

Lysozyme (LYS) is an enzyme that exists in animal tissue and fluids. Its richest source is egg white. The first attempt to extract LYS from egg white was made by Wolff in 1922 but it cannot be considered successful as he obtained LYS without any nitrogen or sulphur atoms. In the following year, Meyer and his colleagues managed to extract LYS from egg white by precipitating it with alcohol [19].

LYS is a small globular protein with 129 residues. Its molecular mass is 16.2 kDa [20]. It is generally known as hen egg white protein and its Protein Data Base (PDB) ID is 4LZT [21]. Since it has a PI value around 11 it is positively charged at moderate pH and its dipole moment is negligible. Due to its almost globular shape and thermodynamic stability LYS is often used for research studies. The molecular structure of LYS can be seen in Figure 2.3 [22].

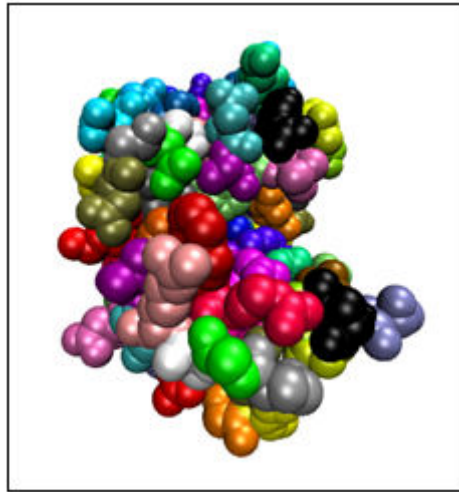


Figure 2.3. Molecular Structure of Lysozyme [22]
(Each color denotes different residue)

α -Lactalbumin is a major whey protein in milk. The production of this protein takes place in the lactating mammary gland of a number of mammals [23] and is responsible for the synthesis of lactose, milk sugar. Understanding the role of α -LA in milk has great importance to enhance the efficiency of milk production in domestic farm species [24] and to design novel food products [1].

Two different forms of α -Lactalbumin are studied in this project; the apo and holo form which contain 123 and 122 residues, respectively. Their molecular weights are around 85 kDa. The holo form contains bound calcium, while apo α -Lactalbumin does not. The bound calcium has an important role in the formation of disulfide bonds during the folding process (transformation into the functionalized 3D structure, *native structure*) of denatured protein. The absence of calcium leads to transconformation into molten globular state in which protein does not have a certain native structure. The bound calcium consequently increases the stability of the native state relative to molten globular state and accelerates the refolding process of the denatured protein (with intact disulfide bonds) at least two orders of magnitude [25]. Although calcium increases the rate of folding process, it is not required for the refolding because both apo and holo forms shows the same fold and activity [25].

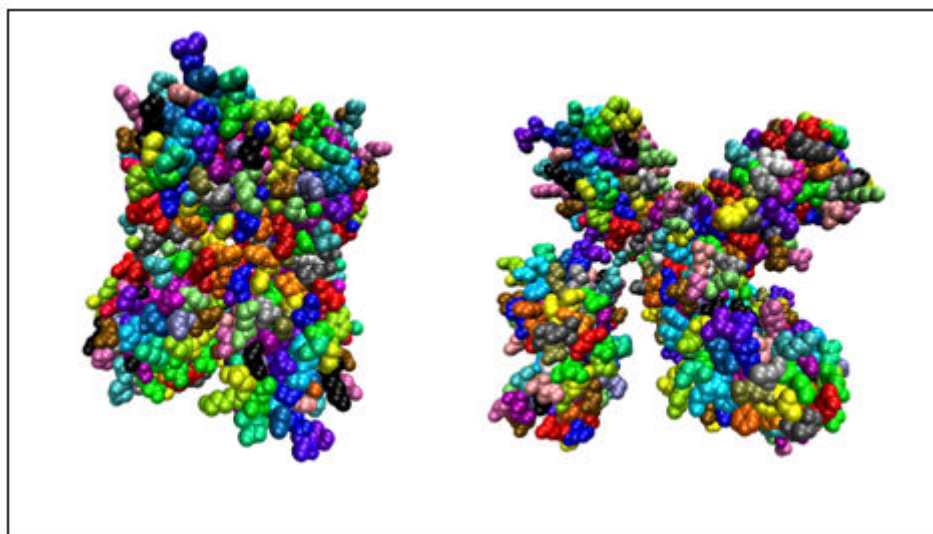


Figure 2.4. Crystal structure of Apo (PDB ID 1F6R, left) and Holo (PDB ID 1F6S, right) Bovine α -LA (Each color denotes different residue) [25]

Although both α -LA forms are negatively charged at neutral pH, the apo form is more negative than the holo form and the dipole moment of the apo form is higher than the holo form. The charges and dipole moments of apo and holo form at 7.5 pH and 30 mM 1:1 salt concentration are $6.2e$ and $-4.6e$; $80.9e\text{\AA}$ and $69.5e\text{\AA}$, respectively. The pI values of apo and holo form are 4.2 and 4, respectively.

Nigen and colleagues have studied the formation of supramolecular aggregates resulting from self-assembly between LYS and α -LA under different physico-chemical conditions [2, 4-5, 26-28]. They showed that LYS is able to form microspheres with the calcium depleted form of α -LA but unable to do so with the holo form [5]. The formation of microspheres occurs successively by reorganization of small spherical aggregates into large microspheres [27]. It is also reported that, the shape of the aggregates is temperature dependent. At 10°C the aggregates are amorphous, while at 45°C they have a well defined spherical shape [5]. It is also known that, the ability to form aggregates is inversely proportional to the ionic strength at pH 7.5 and self assembly of LYS and α -LA is not observed at 100 mM salt [2]. In addition, both LYS and α -LA are found in equimolar ratios within the microspheres and they are uniformly distributed in the overall supramolecular structure [26, 28].

2.3. INTERMOLECULAR INTERACTIONS

Intermolecular interactions are the non-covalent interactions which hold molecules together and thus, determine bulk properties of solids, liquids and gaseous. Electromagnetic interactions are the sources of the intermolecular interactions. Although, important long-range intermolecular interactions exist between big colloidal particles, their range of action rarely exceeds 100 nm. The first successful expression for intermolecular interaction which includes both attractive and repulsive term was proposed by Mie in 1903 [29].

2.3.1. Lennard-Jones (L-J) Potential

Mie's proposal constituted the basis of Lennard-Jones Theory [30]. His work was adopted by Lennard Jones in 1920s; it has an attractive term, $-B/r^m$, and a repulsive term, A/r^n , where A and B are constants and r is the separation between particles and $n > m$ [30-31]. The power of the attractive term is six which reflects *London dispersion* in van der Waals attractive (vdW) interactions. The reason for this interaction is fluctuating dipoles in the molecules mutually induced by one on the other. The repulsive term is due to overlapping of electron clouds at very short separations, the *Pauli exchange repulsion* [31]. In general, an exact expression for this repulsion is impossible to determine in a classical framework. The power of 12 is used to approximate the repulsive term due to its computational convenience and reasonably success [30]. The final form of the L-J potential hence becomes,

$$U_{LJ}(r) = 4\varepsilon_{LJ} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (2.1)$$

where ε_{LJ} is the depth of the potential at $r = r^*$ (separation of the center of masses at contact) and σ is the separation between the particles at which the potential is zero, as shown in Figure 2.5. In this work, the parameters for the L-J Potential are extracted from second virial coefficient measurements of proteins [31-32]. When the interacting molecules are different then it is common to use the geometric average for ε_{LJ} and arithmetic average for σ [30].

$$\varepsilon_{LJ,ave} = \sqrt{\varepsilon_{LJ,1}\varepsilon_{LJ,2}} \quad (2.2)$$

$$\sigma_{ave} = \frac{\sigma_1 + \sigma_2}{2} \quad (2.3)$$

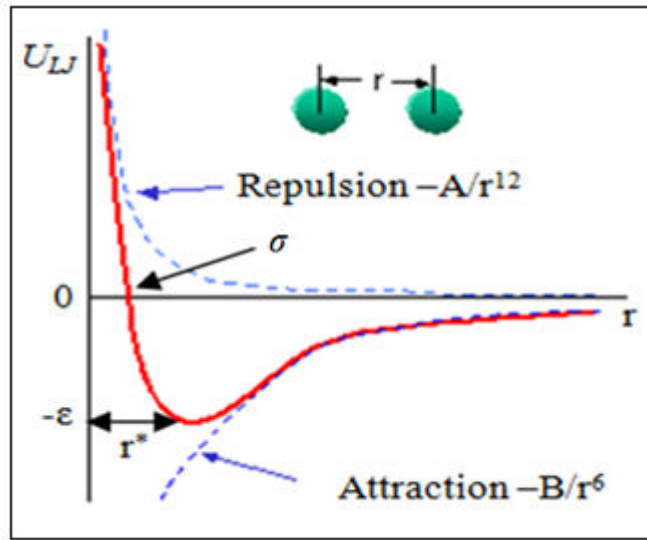


Figure 2.5. Lennard-Jones “12-6” Potential

2.3.2. Electrostatic Interactions

2.3.2.1. Coulomb interaction in vacuum

Two charges q_1 and q_2 , separated by a distance r , interact via the Coulomb interaction,

$$U(r) = \frac{q_1 q_2}{4\pi\varepsilon_0 r} \quad (2.4.)$$

where ε_0 is the dielectric permittivity of vacuum. Electrostatic interactions in vacuum are very long ranged due to the slow decay with the separation between the charges.

2.3.2.2. Coulomb interaction in a dielectric medium

When the Coulomb interaction takes place in a dielectric medium the interaction strength will be reduced because of the presence of many nearby medium (solvent) molecules. The reduction of the electrostatic interactions is given by the dielectric constant of the medium (ϵ_r) [29].

$$U(r, T) = \frac{q_1 q_2}{4\pi\epsilon_0\epsilon_r r} \quad (2.5)$$

At room temperature (298 K), the dielectric constant of water is $\epsilon_r = 78.5$ which means that the interaction strength between two charged particles in water is reduced by almost two orders of magnitude, compared to vacuum. This reduction is referred to as *dielectric screening*. The presence of permanent or induced dipoles (water molecules in an aqueous solution) around the charges diminishes the electric field coming from charged particle, resulting in a reduction of electrostatic interactions.

It is convenient to calculate the interaction energy as dimensionless energy, $U(r)/kT$, where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant and T is the absolute temperature. The solvent mediated Coulomb potential then reaches,

$$\frac{U(r, T)}{kT} = \frac{q_1 q_2}{4\pi\epsilon_0\epsilon_r r} \frac{1}{kT} = \frac{e^2}{4\pi\epsilon_0\epsilon_r kT} \frac{z_1 z_2}{r} = l_B \frac{z_1 z_2}{r} \quad (\text{dimensionless}) \quad (2.6)$$

where l_B is the Bjerrum length, $e = 1.6 \times 10^{-19}$ C, is the elementary charge and z is the valence of the charged particles. In water at room temperature ($\epsilon_r \approx 80$) the Bjerrum length is equal to 7 Å.

It is important to notice that an enormous simplification has been made in this expression by assuming that the solvent is a structureless dielectric continuum. Instead of considering all pair-wise interactions between solvent molecules and two charged particles, the solvent is simply included by introducing a dielectric constant in the expression for the interaction energy. The interaction consequently corresponds to an averaged interaction over all water molecule configurations.

2.3.2.3. Coulomb interaction in salt solution (Debye-Hückel Theory)

The Debye-Hückel theory considers the electrostatic interaction between ions in a salt solution. Since ions in the solution have a preference to be closer to the oppositely charged ions, each has an ionic atmosphere around [33].

The Debye-Hückel theory expresses the mean electrostatic potential around the central ion as a function of separation from the center of the ion (r). It assumes that the ions are immersed in a uniform dielectric continuum. It builds the potential based on the solution of the linearized Poisson-Boltzmann (LPB) equation at very low concentration (limiting law) given below [31],

$$\frac{1}{r^2} \frac{d}{dr} \left[r^2 \frac{d\psi(r)}{dr} \right] = -\frac{e}{\epsilon_0 \epsilon_r} \sum_j z_j \rho_j \left[1 - \frac{z_j e \psi(r)}{kT} \right] \quad (2.7)$$

where ψ is the potential around central ion, z_j is the valence of ion j in the ionic atmosphere, ρ_j is the number density of ion j . Because of the electro-neutrality condition of the salt solution, the first term cancels out and it yields [31]

$$\frac{1}{r^2} \frac{d}{dr} \left[r^2 \frac{d\psi(r)}{dr} \right] = \frac{e^2}{\epsilon_0 \epsilon_r kT} \sum_j \rho_j z_j^2 \psi(r) = \kappa^2 \psi(r) \quad (2.8)$$

where $\kappa = \frac{e^2}{\epsilon_0 \epsilon_r kT} \sum_j \rho_j z_j^2$, is the inverse Debye screening length (D) and the ionic strength of the solution is $I = \frac{\sum_j \rho_j z_j^2}{2}$ M. In an aqueous 1:1 electrolyte solution at 298 K the Debye screening length becomes

$$D = \frac{1}{\kappa} = \frac{3.04}{\sqrt{I}} \text{ (\AA)} \quad (2.9)$$

The solution for the potential is

$$\psi(r) = \frac{ze \exp(-\kappa(r - a))}{4\pi \epsilon_0 \epsilon_r r (1 + \kappa a)} \quad (2.10)$$

where a is the contact distance between two charged particles. The free energy of interaction between two charges is,

$$U(r) = \psi_1 z_2 = \frac{z_1 z_2 e^2 \exp(-\kappa(r-a))}{4\pi\epsilon_0\epsilon_r r(1+\kappa a)} \quad (2.11)$$

At dilute salt concentrations, $\kappa a \rightarrow 0$ and

$$U(r) = \frac{z_1 z_2 e^2 \exp(-\kappa r)}{4\pi\epsilon_0\epsilon_r r} \quad (2.12)$$

$$\frac{U(r)}{kT} = l_B \frac{z_1 z_2}{r} \exp(-\kappa r) \quad (2.13)$$

The Debye-Hückel theory introduces an exponential decay to the interaction energy as a function of increasing κ which is proportional to the square root of the ionic strength. Note that the Debye-Hückel theory estimates the screening due to all ions at a mean field level. Indeed, this theory has its own validity restrictions. It is valid for dilute monovalent salt solutions. When the valence of the ions increases, the accuracy of the approximation drastically decreases [29].

2.3.2.4. Correction of the Debye-Hückel theory

The effect of the size of the colloidal particle is taken into account and it is considered with a uniform surface charge density. The salt ions are treated as point charges. In this case, the interaction potential between two identical particles (effective pair potential) with diameter a is given as follow,

$$\beta U(r) = l_B \frac{z_1 z_2}{r} \frac{\sinh^2(\kappa a)}{\kappa^2 a^2} \exp(-\kappa r) \quad (2.14)$$

where β is the inverse thermal energy, $1/kT$. When the sizes of the particles differ, the expression takes the form,

$$\beta U(r) = l_B \frac{z_1 z_2}{r} \frac{\sinh(\kappa a_1) \sinh(\kappa a_2)}{\kappa^2 a_1 a_2} \exp(-\kappa r) \quad (2.15)$$

when the size of the particles are relatively small with the Debye length, $\kappa a \rightarrow 0$ and $\frac{\sinh(\kappa a)}{\kappa a} \rightarrow 1$ and eventually, Equation 2.15 turns into Equation 2.13 [34].

2.4. BASIS OF MONTE CARLO SIMULATIONS

2.4.1. Statistical Mechanics and Ensembles

Statistical mechanics (SM) aim to calculate the thermodynamic properties such as free energy, entropy, and pressure of a macroscopic system starting from a molecular level. In classical mechanics, time averaging over a sufficiently long period is used to calculate a thermodynamic property [31]. Instead of time averaging, SM averages over an infinite number of instantaneous snapshots of the real system; the so-called *ensemble average* [35]. An ensemble is a hypothetical collection of very large number of subsystems (M) which have identical thermodynamic properties with the macroscopic system but differ at microscopic (molecular) level [31].

In SM, it is important to choose appropriate ensemble for the system of interest. The ensemble generally mimics the conditions in the corresponding experimental setup. For instance, if temperature (T), pressure (P) and the amount of material are kept constant in the experimental system, a convenient choice for the ensemble is often the isothermal-isobaric ensemble (NPT) in which temperature, pressure and the number of particles (N) are fixed. This choice is not obligatory because the calculated averages are the same regardless of the ensemble, the only change will be in the calculation method and it is possible to switch one ensemble to another using Legendre transformations [35].

2.4.1.1. *Micro-canonical Ensemble (NVE)*

In this ensemble, number of particles (N), volume (V) and the total energy of all subsystems (E) are identical and constant. This kind of system corresponds to a closed

insulated system. The probability of finding each possible state is equal. Thus, the probability of any microstate P_i is

$$P_i = \frac{1}{\Omega_{NVE}} \quad (2.16)$$

where Ω_{NVE} is the number of available microstates. Thus, the average of a mechanical property, F , is a simple arithmetic average [35]

$$\langle F \rangle = \frac{1}{\Omega_{NVE}} \sum_{i=1}^{\Omega_{NVE}} F_i \quad (2.17)$$

2.4.1.2. Canonical Ensemble (NVT)

This ensemble has constant N , V and T and corresponds to a system which is in thermal equilibrium with its surroundings. Working with this ensemble is more convenient because in practice, a system has always some physical contact between real bodies and thus, it is always in thermal equilibrium. Since the internal energies of the subsystems vary, the probability of finding any microstate is no longer constant but is given by the well known Boltzmann distribution [35],

$$P_i = \frac{\exp(-\beta E_i)}{\sum_j \exp(-\beta E_j)} = \frac{\exp(-\beta E_i)}{Q_{NVT}} \quad (2.18)$$

where E_i is the energy of micro state i and Q_{NVT} is called the canonical *partition function*, and analogous to Ω_{NVE} in the micro-canonical ensemble. The sum is taken over all microstates. Since the probability of any microstate is known, the average of a mechanical property, F , can be calculated as [31]

$$\langle F \rangle = \frac{\sum F_i \exp(-\beta E_i)}{Q_{NVT}} \quad (2.19)$$

2.4.1.3. *Isobaric-isothermal Ensemble (NPT)*

If the actual thermodynamic system under consideration has constant temperature and pressure, like aqueous solutions of proteins, exposed to the atmosphere in a temperature controlled heat bath, the *NPT* ensemble is often the appropriate choice to find the average of thermodynamic properties. The *NPT* partition function is [35],

$$Q_{NPT} = \int_0^{\infty} Q_{NVT} \exp(-\beta PV) dV \quad (2.20)$$

Then the average can be calculated as

$$\langle F \rangle = \frac{\int_0^{\infty} F_i Q_{NVT} \exp(-\beta PV) dV}{Q_{NPT}} \quad (2.21)$$

This ensemble is convenient to simulate phase transitions because the system is free to transform into the lowest Gibbs free energy state at constant pressure. If the system tends to phase separate into two bulk phases at different densities *NPT* ensemble provides prediction of densities at which the both phases exist by fluctuating the volume hence the density [36].

2.4.1.4. *Grand-canonical Ensemble (μVT)*

If the number of particles in the system is not constant and the main concern of the simulation is the average number of particles in the system the grand-canonical ensemble may be the optimal choice because in this ensemble, instead of number of particles, the chemical potential of the species is fixed. For example, if the adsorbed gas molecules are in equilibrium with the gas in the reservoir, the chemical potential and the temperature of the gas inside the adsorbent and in the reservoir must be equal. It means that with known chemical potential and temperature, the equilibrium concentration can be calculated by using grand-canonical ensemble [36]. The partition function and the averages of the grand-canonical ensemble are,

$$Q_{\mu VT} = \sum_N Q_{NVT} \exp(\beta \mu N) \quad (2.22)$$

$$\langle F \rangle = \frac{\sum_N F Q_{NVT} \exp(\beta \mu N)}{Q_{\mu VT}} \quad (2.23)$$

2.4.2. Free Energy Calculations

In statistical mechanics the free energy associated with the microstate is expressed in terms of the partition function of the corresponding ensemble,

$$\text{free energy} = kT \ln(Q) \quad (2.24)$$

where the Q is the partition function of the ensemble.

In the canonical ensemble, the free energy corresponds to the Helmholtz free energy because the independent variables are N , V , T . If the ensemble is isobaric-isothermal the free energy corresponds to Gibbs free energy as a function of N , P , T [31]. In an NPT ensemble the probability of finding a particular volume i can be written as

$$P(V_i) = \frac{Q(V_i)}{Q_{NPT}} \quad (2.25)$$

Then, the free energy change associated with going from volume i to j can be expressed as

$$\begin{aligned} \Delta A &= kT [\ln Q(V_j) - \ln Q(V_i)] = kT \ln \left[\frac{Q(V_j)}{Q(V_i)} \right] \\ &= kT \ln \left[\frac{Q(V_j)/Q}{Q(V_i)/Q} \right] = kT \ln \left[\frac{P(V_j)}{P(V_i)} \right] \end{aligned} \quad (2.26)$$

where $\frac{P(V_j)}{P(V_i)}$ is the relative probability of observing volume j with respect to volume i [35].

It is also convenient to calculate the free energy of interaction (potential of mean force, pmf , $w(r)$) using radial distribution function $g(r)$ which is the deviation of local density at r ($\rho(r)$) from the bulk density.

$$g(r) = \frac{\rho(r)}{\rho_{bulk}} \quad (2.27)$$

$$w(r) = -kT \ln (g(r)) \quad (2.28)$$

The radial distribution depends on temperature, bulk density and the interaction between the particles. It can be obtained indirectly from x-ray diffraction measurements [31].

2.4.3. The Metropolis Monte Carlo Method

In the *unbiased* Monte Carlo (MC) method, each particle in the system is located randomly in the configuration space by assigning a random point to each particle. Then, the potential energy of the configuration is calculated as [37],

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N U_{ij}(r_{ij}), \quad i \neq j \quad (2.29)$$

where U_{ij} is the potential energy between particle i and j and r_{ij} is the minimum distance between those. If the NVT ensemble is considered, this configuration will have a weight $\exp(-E/kT)$ in the average calculation of the desired quantity. Since volume and number of particles are fixed in an NVT ensemble, the new configuration can be generated by either displacing or rotating of a random particle (multiple particle moves are also possible) and then, according to the total potential energy of the system, assign a Boltzmann weight to the configuration. Generating new configurations are continues until the equilibrium is reached by means of sampling the entire configurational space adequately [36]. This is an inefficient way of calculating averages, especially for close packed configurations, since most of the generated configurations will have very small Boltzmann weights and hence negligible contributions to the average value. In other words, most of the computational time will be spend adding *nearly zero* to the average.

In the Metropolis MC method, instead of choosing configurations randomly, they are chosen with a probability $\exp(-E/kT)$ and then weighted equally in the average (arithmetic average). The steps of the Metropolis algorithm are given below [37],

- Place N particles in any configuration and calculate the system energy.
- Move one particle in each direction according to $x_{new} = x_{old} + \xi_1 \Delta x$ where ξ_1 is a random number between -1 and 1, and Δx is the maximum allowed displacement.
- Calculate the energy change for the move and $\Delta E = E_{new} - E_{old}$
- If $\Delta E < 0$ accept the move.
- If $\Delta E > 0$ accept the move with the probability $\exp(-\Delta E/kT)$ as follows,
 - Take a random number, ξ_2 , between 0 and 1
 - If $\xi_2 < \exp(-\Delta E/kT)$ accept the move.
 - If $\xi_2 > \exp(-\Delta E/kT)$ reject the move and put the particle back to its old position.

Whether the move is accepted or rejected, the final configuration is used for calculating ensemble averages. For instance, if M Monte Carlo moves are attempted, the average of some property, F , is

$$\bar{F} = \frac{1}{M} \sum_{j=1}^M F_j \quad (2.30)$$

The advantage of this ‘‘importance’’ sampling is that the selection of the configurations is focused on regions of configurational space in which the Boltzmann weights are high [37]. Since the acceptance rule must ensure that the occurrence frequency of the configurations obeys the desired probability distribution it differs for each ensemble [36]. The acceptance rule of ensembles used in this project is given below;

- NVT ensemble: $\xi_2 < \exp(-\beta\Delta E)$ accept the trial move
- NPT ensemble: $\xi_2 < \exp\left(-\beta\Delta E + P\Delta V - \frac{N+1}{\beta} \ln\left(\frac{V_{new}}{V_{old}}\right)\right)$ accept the trial move of the box length, L or the particle position.

- μVT ensemble:
 - $\xi_2 < \exp \left(\frac{V}{\Lambda^3(N+1)} \exp (\beta\mu - \beta[E(N+1) - E(N)]) \right)$ accept insertion of a particle.
 - $\xi_2 < \exp \left(\frac{\Lambda^3 N}{V} \exp (-\beta\mu + \beta[E(N-1) - E(N)]) \right)$ accept removal of a particle [36].

2.4.4. Periodic Boundary Conditions

In computer simulations, it is essential to reproduce bulk properties of the studied thermodynamic system. However, the number of particles (N) which can be handled in a simulation is restricted by the computational speed and available storage memory of the computer. In addition, the computation time of a simulation is proportional to N^2 . Therefore computer simulations are usually performed for a relatively small number of particles [38]. On the other hand, simulating only a small number of particles brings about boundary effects since a large number of particles borders the surface of the simulation box. For instance, in a simple cubic 10 nm x 10 nm x 10 nm box, almost 30 % of 1000 particles stay within 10 percent of the box length of the cube face [36]. In order to overcome boundary effects, and mimic infinite bulk surrounding, periodic boundary conditions are usually implemented [36, 38].

In the periodic boundary approach, the cubic simulation box is replicated throughout space to form a hypothetical infinite lattice so that if a particle leaves the simulation box will enter the box through the opposite face, see Figure 2.6. This is ensured by moving each replica of the original particle simultaneously with identical displacement as the original one. The boundary effect is eliminated by simply removing the rigid walls around the simulation box.

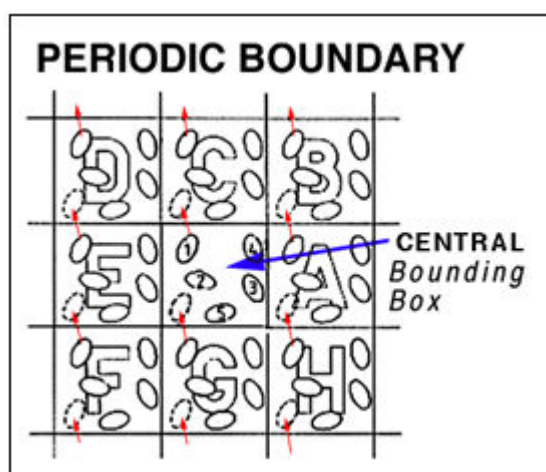


Figure 2.6. Two dimensional periodic system
(Central box is the original one and labeled boxes are the replicas of it. Particle 1 is leaving the central box and its periodic image in box G is entering to central box) [39]

2.4.5. Minimum Image Convention

The periodic boundary conditions may create infinitely many particles which a particle may interact with. Since calculating infinitely many pair-wise interaction potentials is impossible, the interaction must be truncated at some distance. In the minimum image method, only interactions with the nearest image are considered which simply implies cubic cut-off at half of the box length away from the central particle [38] and in practice only this central box is truncated in the simulation.

This truncation method simply neglects the long-range part of the interaction. Therefore, it introduces under estimations when the long range interactions are important. There are some truncation schemes for handling of long-range interactions such as, Ewald summation, fast multipole method and particle-mesh-based techniques. Ewald summation is the most widely used and the computational time for system containing N particles scales as order of $N^{3/2}$, $O(N^{3/2})$ [36]. However it becomes more costly for large systems. In the Ewald summation, the long range interactions are calculated by using rapidly converging Fourier summation in Fourier space rather than summation in real space [40].

3. METHODOLOGY AND MODELING

3.1. COARSE GRAINING OF THE SOLUTION

In this work, it is aimed to investigate the interaction between LYS and α -LA which leads to the formation of spherical supramolecular aggregates in aqueous salt solution [2]. However, simulating many proteins as well as solvent and salt particles is computationally costly and requires a reduction of the number of particles. Fortunately, an enormous reduction can be achieved by simply eliminating the solvent and salt particles from the simulation box with the help of statistical mechanical theory. In order to coarse grain the solution; the interaction potential between the proteins is modified to treat the solvent and salt particles implicitly. This is explained in detail in the following two sections.

3.1.1. Implicit Solvent and Explicit Salt

Water molecules are treated implicitly using the dielectric continuum approach as described by Equation 2.6. Since the simulations are performed at constant temperature, 298 K, ϵ_r is 78.5 equal to that of pure water and the Bjerrum length is hence 7.12 Å. The van der Waals attraction and short range repulsion parts of the potential are described by the Lennard-Jones 12-6 potential. Thus, the effective pair interaction between particle i and j is given by,

$$\beta u_{ij} = l_B \frac{z_i z_j}{r} + \frac{4\epsilon_{LJ}}{kT} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right] \quad (3.1)$$

σ_{ij} is calculated using Equation 2.3 and ϵ_{LJ} differs according to the model used to represent the structure of the protein, as will be described in Section 3.2. The salt particles are treated explicitly and each is represented by a sphere of radius 4 Å [41]. The system energy is calculated by summing all pair-wise interactions.

Notice that the dielectric continuum approach provides enormous reduction of the number of particles, $O(10^3)$, by simply eliminating water particles from the simulation box,

resulting in a reduction of computational time by $O(10^6)$, as seen in Figure 3.1. Since the water molecules are treated as a structureless dielectric continuum, and no explicit water molecules are present. A weakness of this approach is that hydrophobic interactions between proteins are not considered. However, the hydrophobic interactions of LYS and α -LA are negligible due to their high solubilities in water and as well as their high charges which suggest that electrostatic interactions dominate.

3.1.2. Implicit Solvent and Implicit Salt

In order to eliminate salt particles from the simulation, the electrostatic part of the pair potential is now treated at the Debye-Hückel level, as shown in Equation 3.2.

$$\beta u_{ij} = l_B \frac{z_i z_j}{r} \exp(-\kappa r) + \frac{4\epsilon_{LJ}}{kT} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right] \quad (3.2)$$

where κ implicitly represents salt particles in solution. κ is calculated using Equation 2.9 and some values of κ and D are given in Table 3.1.

Table 3.1. D and κ in a function of 1:1 salt concentration at 298 K

| Salt Concentration (mM) | D ($1/\kappa$) (Å) | κ (Å ⁻¹) |
|----------------------------|---------------------------|--------------------------------|
| 5 | 42.99 | 0.023 |
| 25 | 19.22 | 0.052 |
| 30 | 17.55 | 0.057 |
| 40 | 15.20 | 0.066 |
| 50 | 13.59 | 0.074 |
| 100 | 9.61 | 0.104 |

3.2. COARSE GRAINING OF THE PROTEIN STRUCTURE

In order to study phase transition of LYS and α -LA, it is crucial to deal with an adequate number of proteins which may lead to the phase separation. Since simulating

many proteins requires further reduction of the number of particles, a coarse graining approach is also applied to the structures of the proteins and two coarse graining models for the protein structures are developed. These models are explained from the detailed one to the primitive one in the following sections.

3.2.1. Amino Acid Model (AA)

This model treats each amino acid as a single sphere, located at the center of mass of the corresponding amino acid, obtained from the PDB. The radius of each sphere ($\sigma \approx 7\text{\AA}$) is determined by assuming a uniform density of 0.9 g/mL throughout the protein and using the molecular mass of each amino acid [41]. This assumption is reasonable if one considers that the densities of most organic compounds are between 0.7 and 1.25 g/mL [42]. Average charges of titratable sites are determined by a separate titration Monte Carlo simulation in which a single protein is titrated according to pH, salt concentration, and pKa of the particular amino acid [43].

3.2.2. Surface Charge Model (SC)

In this model, developed during this project, charged amino acids are retained in their original locations (as in the AA model), but all neutral amino acids are clustered into *one* big sphere located in the center of mass of the entire protein. The radius of the sphere is used as a fitting parameter in the potential of mean force (pmf, free energy of interaction) between LYS and α -LA and three different particle radii; 15 \AA , 20 \AA , and 25 \AA are tried by considering the actual volume of the proteins. 15 \AA is selected due to its excellent agreement with the pmf of the AA model. ϵ_{LJ} is adjusted by comparing with the AA model potentials and set to 0.3, as will be seen Section 4.2.2.

3.2.3. Twin Sphere Model (TS)

The entire protein is represented by two oppositely charged spheres. The spheres are located at the mass center of positive and negative amino acids, respectively. The radii of the spheres are determined using the same method as in the SC and 15 \AA is finally chosen as the radii.

All coarse graining steps mentioned so far are summarized in Figure 3.1. It is started with detailed systems in which solvent and salt particles, and atoms in the protein structure are treated explicitly; and ended with the reduced system which merely two spheres for each protein. Table 3.2 shows the number of particles required for each coarse grained model.

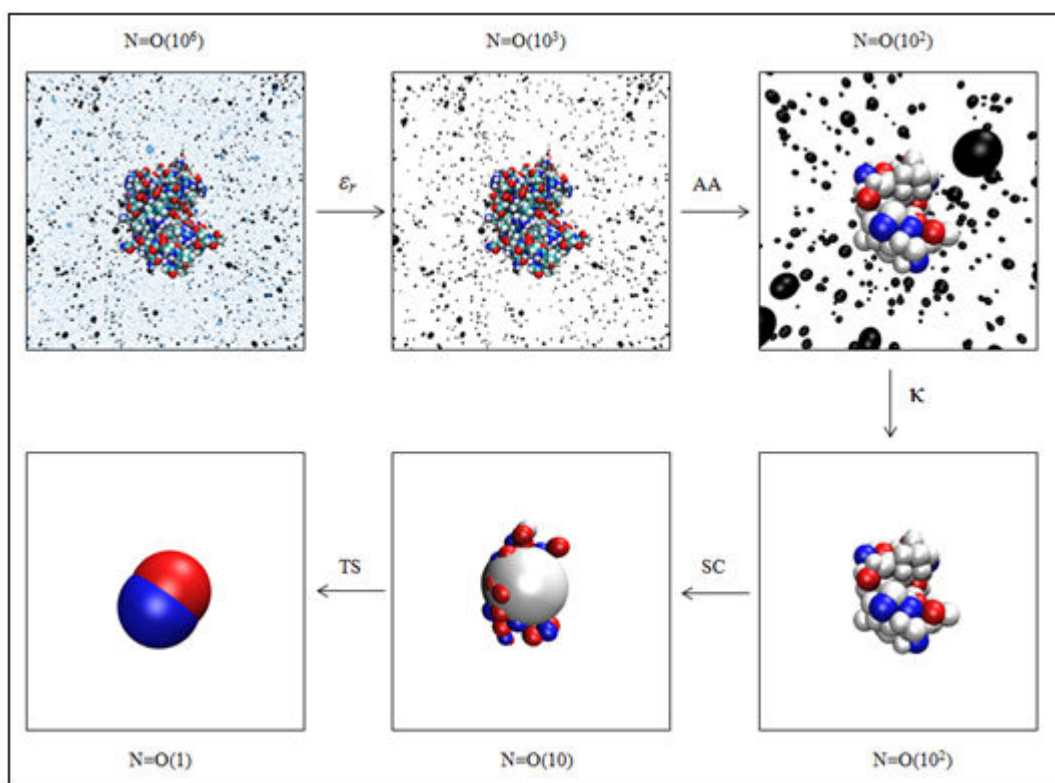


Figure 3.1. Coarse Graining Steps

(Each particle in the system are represented by spheres with different colors; light blue-water, black-salt, 1st and 2nd systems in 1st row; each atom in protein has different colors, in the rest of systems; blue-negatively charged residue, red-positively charged residue, white-neutral residues. The number of particles in each system is given above its corresponding figure and the arrows indicates the models or variables introduced into the system in each coarse graining step)

Table 3.2. Number of particles in the examined coarse grained protein structures

| Protein Model | Number of particles | | |
|---------------|---------------------|------------------|-------------------|
| | LYS | Apo α -LA | Holo α -LA |
| Atomistic | 1183 | 6006 | 5945 |
| AA | 129 | 123 | 122 |
| SC | 33 | 42 | 44 |
| TS | 2 | 2 | 2 |

3.3. SEPARATION DEPENDENT COARSE GRAINING

In this model, the structure of the proteins is represented by either the SC or TS model depending on the distance between two proteins. When the separation between two molecules increases, the distance between charges on the molecules becomes negligible, relatively with the distance between the two molecules. Thus, the two molecules feel each other as charged dipoles. By considering this fact, the protein structures are represented by the TS model when they are far away from each other. While, the more detailed structure model, SC, is used to capture orientational ordering when they are close to each other.

A spherical cut-off r_{cut} is used to determine when to use which model (TS or SC). When $r > r_{cut}$ the TS model is used otherwise the SC model is used to represent the proteins, see in Figure 3.2. The cut-off value is determined according to the Debye length of the solution. Since the electrostatic interactions is extremely weak when r exceeds a few Debye lengths [31], the r_{cut} is set to five times D in order not to underestimate the interaction energy.

The twin spheres have an appreciable size (half of the entire volume of a protein). Thus using the DH expression introduces an error into the system since this expression neglects the size of the spheres. In order to take into account the excluded volume of the spheres, the corrected DH expression given in Equation 2.14 is used to evaluate electrostatic part of the interaction energy. The effective pair interaction energy between particles becomes,

$$\beta u_{ij} = l_B \frac{z_i z_j \sinh(\kappa a_i) \sinh(\kappa a_j)}{r} + 4\epsilon_{LJ} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right] \quad (3.3)$$

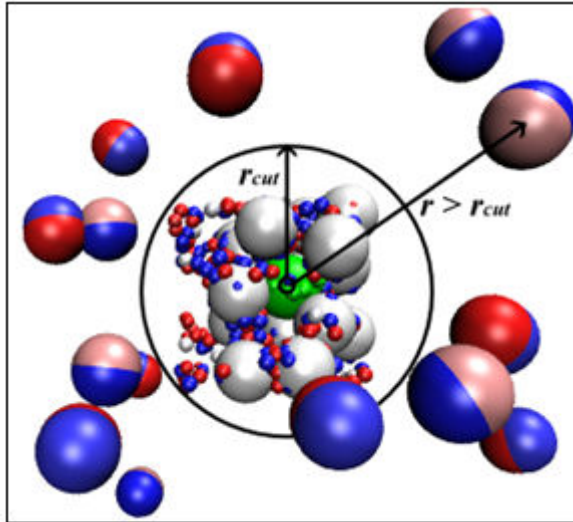


Figure 3.2. Separation dependent coarse graining and the cut-off radius

3.4. USED PROGRAMS OF FAUNUS

Faunus is an object oriented frame work written using the C++ programming language. It is constructed in order to easily create molecular simulation programs. The current programs of Faunus focus on Metropolis Monte Carlo simulations of charged macro-molecules in aqueous salt solution. Since Faunus is an open source, its newest release can always be downloaded freely via the subversion repository at Sourceforge [44].

3.4.1. Isobaric

The many body simulations are performed with the program “*Isobaric*” in order to investigate phase transition of LYS- α -LA mixtures in the aqueous 1:1 salt solution. By considering experimental studies of LYS- α -LA mixture and the possibility of the coexistence of two phases, the *NPT* ensemble is chosen to simulate this system. The Metropolis MC method is used for efficient sampling of configurations. Periodic boundary conditions are applied to eliminate boundary effect from the system and the minimum

image convention is implemented in the interaction energy calculations in which solvent particles are treated implicitly using the dielectric continuum approach while salt particles are described at the Debye-Hückel level. The total interaction energy of a generated configuration is calculated by summing all pair-wise interaction energies, as described in Equation 3.2. Both SC and TS models are used to represent the proteins. The simulations are performed at 298 K and at pH 7.5. Two different osmotic pressure values, 0.6 and 0.7 mM, are used to investigate the effect of the pressure on the phase transition and to make dense phase (aggregates) favorable in the system.

During the simulation, the sampled of box lengths are stored in a histogram to obtain relative probability of the distribution of box lengths. This is required for the calculation of the Gibbs energy of the system according to Equation 2.26 [35]. The simulations are started at random configuration and random box size. Therefore, before collecting statistical output, a short initialization run is performed to eliminate initial bias from the statistics.

In order to generate new configurations, mainly, three trial moves are implemented: Translation, rotational and volume moves. Translational moves are performed by adding random numbers between $-\Delta/2$ and $+\Delta/2$ to the x, y, z coordinates of the molecular center of mass;

$$x_{new} = x_{old} + \Delta(ranf - 0.5) \quad (3.4)$$

where *ranf* is random number between [0 and 1] and Δ is the displacement parameter. Note that, $\Delta/2$ is the maximum displacement of a particle. In addition, rotational moves are performed to change orientation of proteins. The rotation of a protein is specified by the rotation of a random unit vector with random angular displacement. Translational and rotational moves are performed on one particle at a time.

The volume move is applied by changing the box length from L_{old} to L_{new} $L_{new} = L_{old} * \Delta L$ where ΔL is a random number between $[-\Delta L_{max}, +\Delta L_{max}]$. After a volume move, the mass center positions of each protein are rescaled according to percent change in the box length and all pair interactions are recalculated to determine the energy of the new

configuration. Since the computational cost of volume move is equal to the cost of N translational or rotational moves, the frequency of attempting a volume move was $1/N$ times the frequency of translation or rotation move. The other implemented moves into the Isobaric program are mentioned in Section 3.5.

3.4.2. Twobody

The *Twobody* program is used to evaluate the potential of mean force (pmf) between LYS and α -LA and the orientation of α -LA according to LYS. Since only two molecules are simulated in this program, the distance and the orientation between these two determine the energy of the generated configuration. Thus, one dimensional translation and rotation moves are sufficient to explore the configurational space of the system. The NVT ensemble is used to simulate this system. Since there is no risk of population of molecules at the surface, a spherical simulation cell with rigid boundaries is chosen. At the initial configuration, the two proteins are placed on either side of the z axis with the same distance from the origin. Translational moves are performed on z axis and both proteins are moved at the same time to opposite directions with random displacement. The acceptance rule for translation or rotation explained in Section 2.4.2 is used to accept or reject a move.

The solvent is described as a structureless continuum and salt particles are treated either explicitly or implicitly. The interaction between proteins is calculated using Equation 3.1 and 3.2 for explicit and implicit salt particles, respectively.

The free energy of the two body interaction (pmf) is calculated by sampling the probability function of separation between the two proteins. The alignment of α -LA with respect to LYS is determined by the z component of the unit dipole vector of α -LA,

$$\text{degree of alignment} = \langle \vec{\mu}_z \rangle = \cos(\theta) \quad (3.5)$$

where θ is the angle between dipole vector and z axis, as shown in Figure 3.3. The degree of alignment equal to 1 ($\theta=0$) indicates full alignment of α -LA and 0 indicates random alignment.

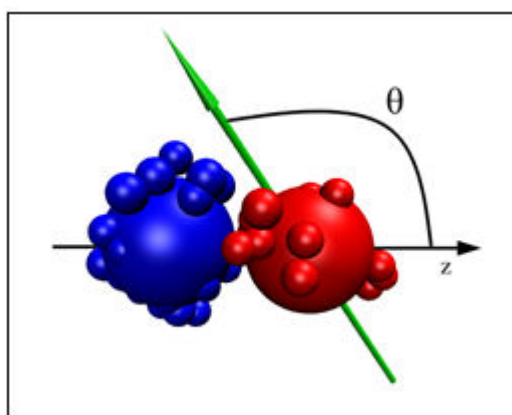


Figure 3.3. Alignment of α -LA (red) with respect to LYS (blue)

Note that none of these programs mentioned above can handle charge regulation on the proteins due to induction from the other proteins. Although the charges on the proteins are determined by titrating according to pH, the effect of the nearby charged protein is not considered [29]. Since the effect of induced charges is important only when pH is close to pI values of one of the proteins [45], it is not expected to have a large impact on the LYS/ α -LA system at pH 7.5.

3.4.3. Grand-canonical Titration

This program is used as a tool to obtain the net charges of each protein and to include the effect of salt concentration and pH on the protein charge distribution. Each protein is titrated at different monovalent salt concentrations and pH using the *Grand-canonical Titration* program. Since titrating a protein is nothing else than inserting or removing protons from the titratable groups of the protein, using the μVT ensemble is the most convenient way of simulating this system. The protein structure is treated at the amino acid level with explicit titratable sites. The solvent is described with dielectric continuum approach and ions are mimicked by charged hard spheres (interaction energies are infinity when they are overlapping, otherwise zero) [46].

The titration can be written as $HA \rightleftharpoons A^- + H^+$. The energy changes associated with protonation and deprotonation is given by,

$$\beta\Delta E = \beta\Delta U_{el} \pm \ln 10(pH - pK_0) \quad (3.6)$$

where $K_0 = \frac{a_{MO^-} a_{H^+}}{a_{MOH}}$ is the intrinsic dissociation constant of titratable site, ΔU_{el} is the change in electrostatic free energy, a is the activity. In order to maintain constant pH condition, deprotonation of a site is performed by assigning a negative charge to it instead of transferring a proton from site to the bulk. When a site is deprotonated the electroneutrality is satisfied by removing or adding counter ions. The acceptance and rejections of the deprotonations and protonations of titratable sites determine the charges on the proteins [46].

3.5. SAMPLING EFFICIENCY IMPROVEMENT

3.5.1. Displacement Parameter Optimization

The choice of Δ in translational moves is important for efficient sampling of configurational space. Efficiency of sampling can be defined as the number of independent configurations visited in a given amount of computational time. Since the number of independent configurations is inversely proportional to mean-square error in the observable, efficiency also corresponds to obtaining lowest statistical error in a given amount of CPU (Central Processing Unit) time. The measure of efficiency in a MC simulation is [36]

$$efficiency = \frac{\sum(\text{displacement of all accepted translation moves})^2}{CPU\ time} \quad (3.7)$$

Since sampling of independent configurations is difficult in the dense phase, Δ optimization is implemented in order to increase the efficiency of sampling in small box lengths. Δ is optimized by maximizing the average displacement in NVT simulations performed at 100 Å box length for ten proteins and Δ is scaled with the square of the box length,

$$\Delta = dp * (\text{box length})^2 \quad (3.8)$$

where dp is the displacement percent. The optimum dp is 0.08 % where the maximum average displacement is 1.44 Å, as shown in Figure 3.4.

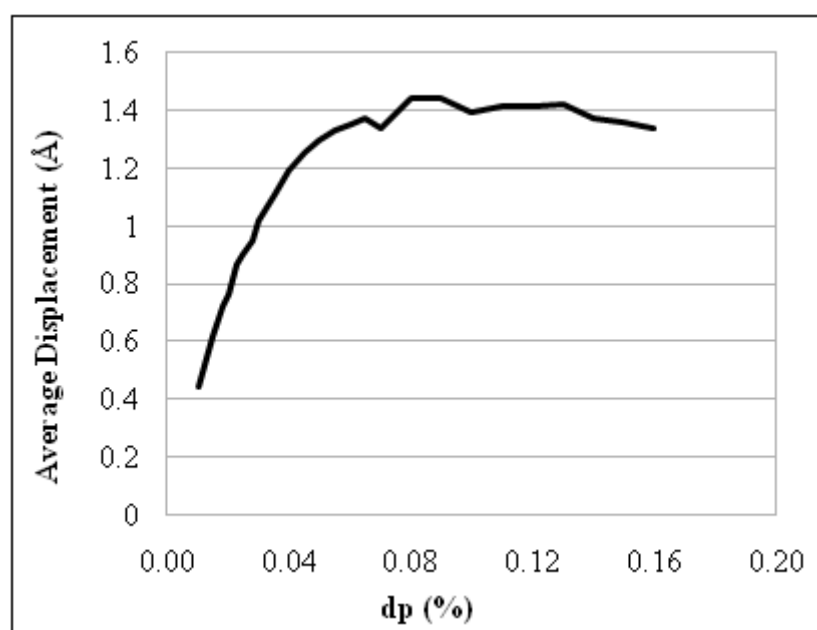


Figure 3.4. Displacement parameter optimization

3.5.2. Combined Translational and Rotational moves

In order to avoid calculating intermolecular interaction for each translation and rotation moves separately, a combined translation and rotation move is implemented into the isobaric program. In addition, in every tenth of the translational / rotational moves, the one with very long displacement is performed to disassemble possible clusters of proteins in the dense phase.

3.5.3. Cluster Move

The cluster move used in the *Isobaric* program is an adapted method of generalized geometric cluster algorithm (GCA) [47-48]. Instead of evaluating an energy change at one particle with the rest, the Boltzmann weight is applied pair-wise. This is achieved by finding a global transformation that, if applied to all particles, regenerates the initial configuration. This transformation is implemented in a form of a translation, $\Delta\vec{r}$, in Faunus.

When a random particle, i , is move; all remaining particles, j , are than considered to move along with a pair-wise probability

$$p_{ij} = \max(1 - e^{\beta\Delta U_{ij}}, 0) \quad (3.9)$$

where $\beta\Delta U_{ij} = U_{i^{new},j} - U_{i^{old},j}$. If a second particle move is accepted, all remaining particles are then reconsidered with a similar probability. The chain continues until no more particles are moved. This rejection free algorithm exploits pair-wise Boltzmann weight to facilitate large rearrangements and more uncorrelated sampling.

4. RESULTS AND DISCUSSIONS

4.1. ONE-BODY SIMULATIONS

One-body simulations are performed in order to determine the net charges on the proteins with respect to salt concentration and pH. The proteins are described at the amino acid level (AA model) and the charge on each amino acid is determined by titrating the proteins using a Grand-canonical titration scheme. The charges obtained from these simulations are then used to develop the other coarse grained models, SC and TS.

4.1.1. pH Dependence of Net-Charges of LYS and α -LA

The pH dependence of the protein net charges at 30 mM 1:1 salt concentration is given in Figure 4.1. The titrations are performed for each protein at pH 2, 7.5 and 12. As expected, the net charges on the proteins are decreasing with increasing pH, due to the transformation of titratable site from protonated state to deprotonated states. The net charges obtained are in good agreement with experimental [49] and other simulation results [41].

Since pH 7.5 is below pI of LYS and above the pI of α -LA, it is chosen to study pmf and phase separation behavior of LYS and α -LA. Both two-body and many-body simulations are performed at 7.5 pH at which LYS and α -LA are almost equally but oppositely charged; $6.8e$, $-6.2e$ respectively.

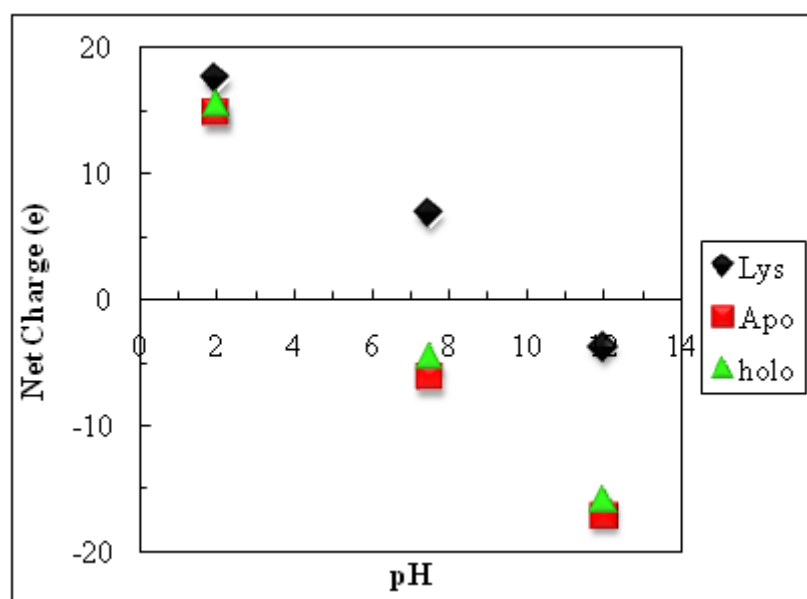


Figure 4.1. The Net Charge on Lysozyme, Apo and holo α -Lactalbumin at various pH in a 30 mM salt solution

4.1.2. Salt Dependency of Net-Charges of LYS and α -LA

The titrations of proteins are performed at pH 7.5 and various 1:1 salt concentrations from 5 mM to 100 mM. The net charge of the proteins as a function of salt is shown in Figure 4.2. Below the pI of the protein, the net charge increases with increasing salt concentration, above pI value it is vice versa. For example, removing a second proton from the negatively charged protein will be easier in high salt because attraction between proton and negatively charged protein will be screened due to the salt particles. However, the effect of salt concentration and also electrostatic screening on protein charge is not as pronounced as the effect of pH. Thus, the net charges differ only slightly with salt concentration as expected [45].

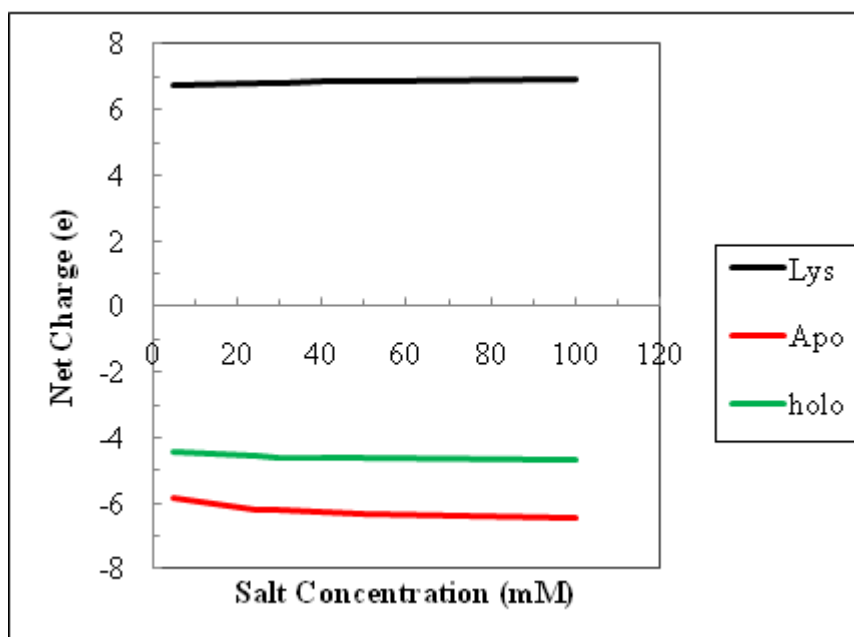


Figure 4.2. The Net Charge on Lysozyme, Apo and holo α -Lactalbumin at various salt concentrations

The data obtained from one-body simulations are given in Appendix B.

4.2. TWO-BODY SIMULATIONS

The potential of mean force between LYS and α -LA is investigated in order to develop coarse grained models. In coarse graining models it is important to maintain the potential of mean force by means of the depth of the potential, the location of the minimum free energy and the range of the attraction (tail of the potential curve). In addition, the alignment of α -LA with respect to LYS is evaluated in coarse graining procedure. In previous studies, the AA model has been shown to accurately represent the surface topology and surface charge density of the full atomistic structure [7, 41]. Therefore, in this work, the relatively detailed AA model is taken as reference point in the comparison of pmf and alignment obtained by the developed coarse graining models.

4.2.1. Implicit Salt Treatment

In the previous study of Persson and Lund, the potential of mean force (angular averaged free energy of the two-body interaction) and the degree of alignment between LYS and α -LA have been investigated by embedding salt particles and proteins into a structureless dielectric medium [41]. In this work, further coarse graining of solution is performed and the salt particles is treated *implicitly* using the Debye-Hückel approximation. The comparison of the explicit salt treatment with new implicit one is shown in Figure 4.3.

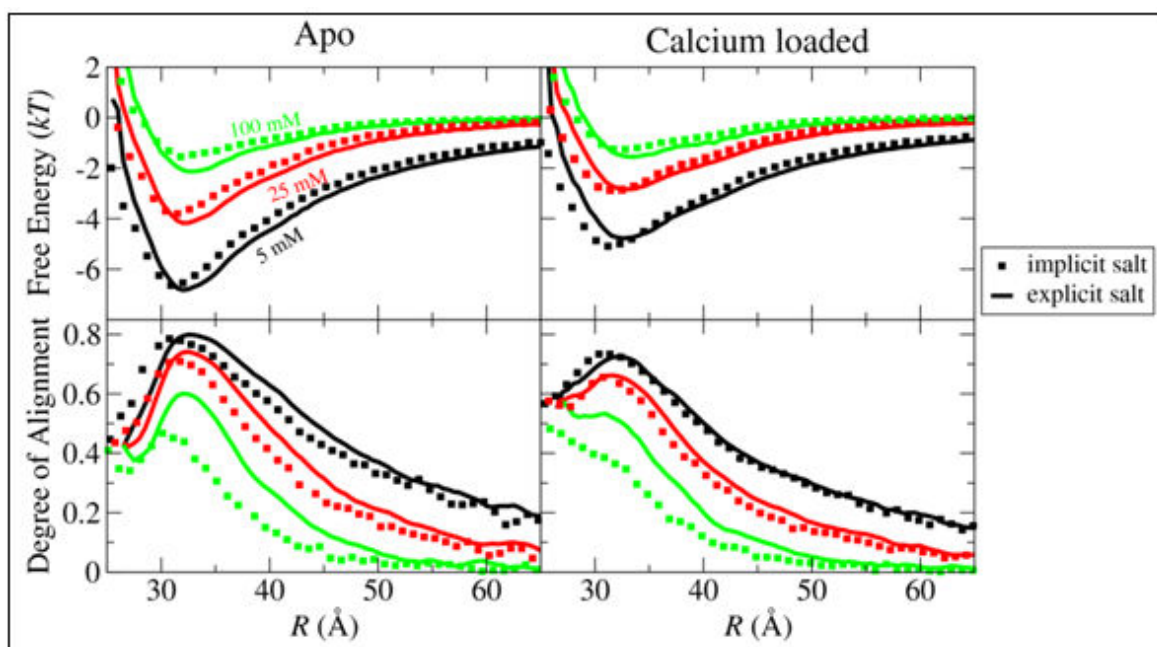


Figure 4.3. The pmf and the degree of alignment between LYS and the two forms of α -LA in the AA model at various salt concentrations for explicit and implicit salt treatments

At low salt concentrations, excellent agreement in both pmf and degree of alignment is obtained. A slight deviation from the explicit salt treatment is obtained at 100 mM because of the DH approximation, as explained in Section 2.3.2. Since the self assembly of LYS and α -LA has been observed at low salt concentrations [2], the implicit salt treatment seems applicable. This treatment brings about a reduction of the number of particles up to $O(10^3)$ for big box sizes at relatively high salt concentrations (e.g. $300\text{\AA} \times 300\text{\AA} \times 300\text{\AA}$ cubic box at 100mM).

The effect of electrostatic screening due to salt particles can clearly be seen in the dramatic difference between the pmf at 5 mM and 100 mM salt concentrations. At 5 mM salt concentration the attraction between Lys and apo α -LA is 3 times stronger than the attraction at 100 mM salt.

It has experimentally been shown that the binding constant of LYS-apo α -LA and LYS-holo α -LA (calcium loaded) have the same order of magnitude [28]. This is also interpreted from the Figure 4.3 where the minimum energy in pmf of LYS-apo α -LA is slightly deeper than the pmf of LYS-holo α -LA.

4.2.2. Surface Charge Model

In the SC model, all neutral amino acids are replaced with one big sphere and the radius of this sphere needs to be adjusted to give a pmf as in the AA model. The value of R is determined according to the actual volume of the proteins and followed by fine fitting to the minimum free energy separation (distance between centers of masses of proteins at contact) of the AA model, as shown in Figure 4.4. 15 Å is chosen for the final radius due to its excellent agreement with the free energy curve.

After determining the radius of the big sphere, the strength of the attraction (the depth of the pmf curve) between LYS and α -LA is fitted by adjusting ϵ_{LJ} . Although ϵ_{LJ} equal to 0.2 gives satisfactory results for the AA model it seems incapable of reflecting vdW interactions for the SC model. Thus, two-body simulations are performed at various ϵ_{LJ} and at 15 Å radius using SC model, as shown in Figure 4.5. Excellent agreement with AA model is achieved for $\epsilon_{LJ} = 0.3$ kT. Equally good agreement is observed for the calcium loaded holo form of α -LA.

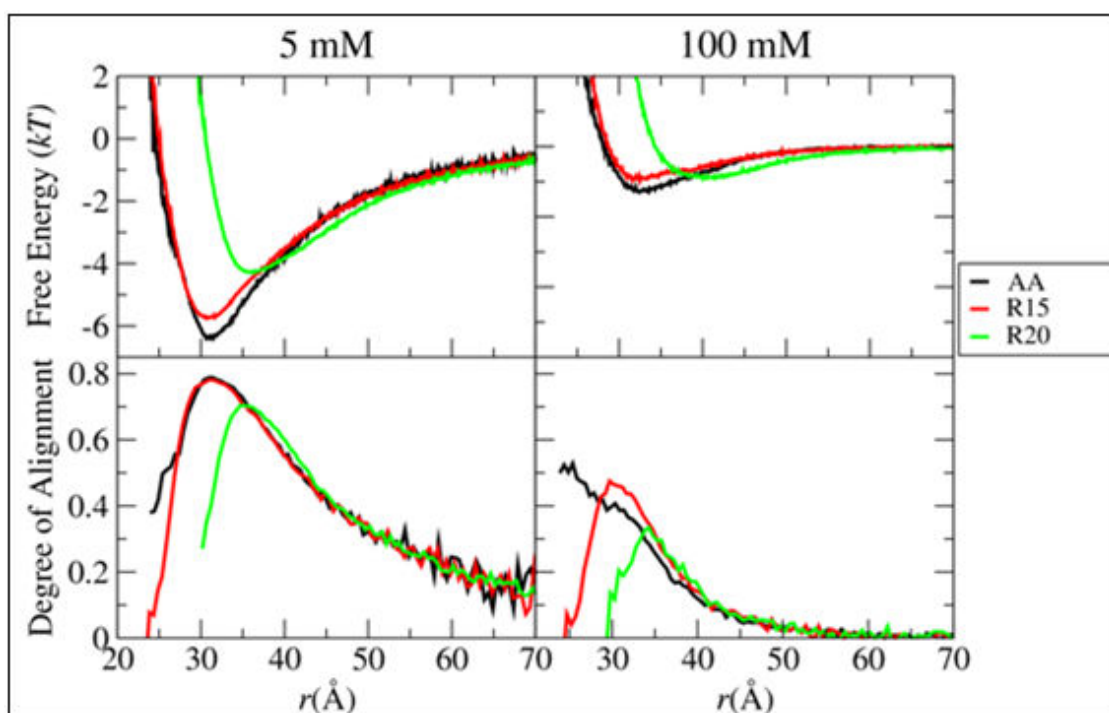


Figure 4.4. The radius dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and SC models at 5 mM and 100 mM 1:1 salt concentrations

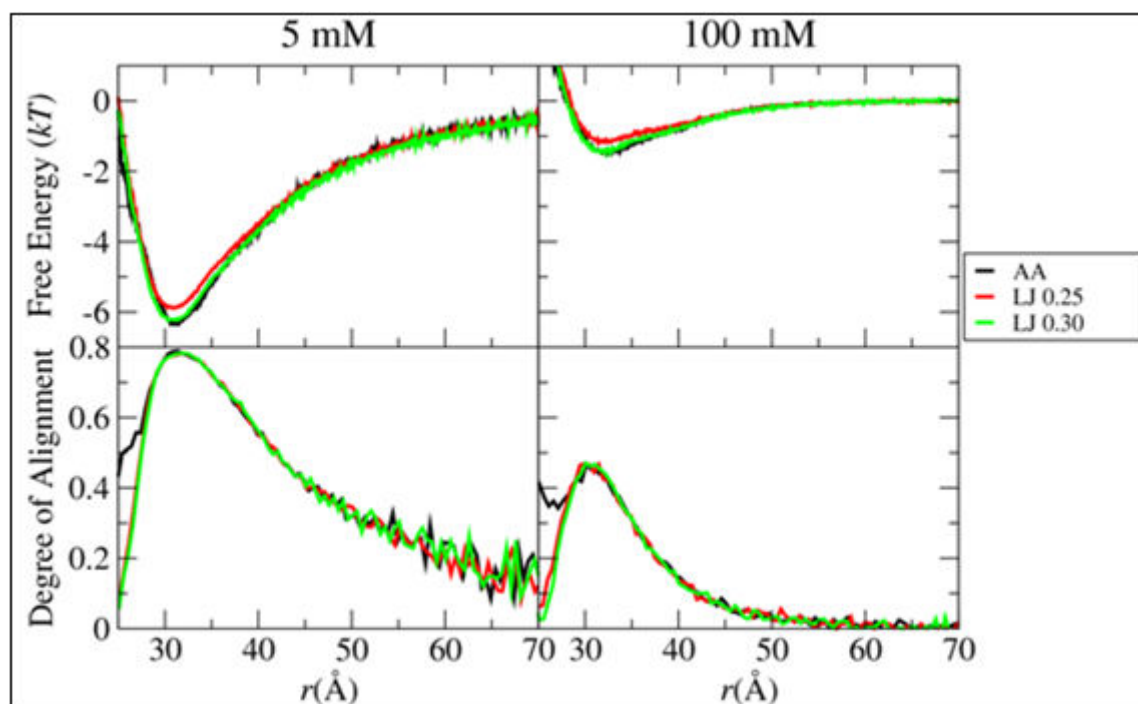


Figure 4.5. ϵ_{LJ} dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and SC models at 5 mM and 100 mM 1:1 salt concentrations

4.2.3. Twin Sphere Model

In this model, described in Section 3.2.3, the radii of the two spheres are adjusted according to the contact separation in the pmf curve of the AA model. Since this model is an extremely simplified model, an excellent agreement with the AA model pmf and alignment is not expected. 15Å which gives satisfactory strength of interaction and contact distance is chosen as radius of spheres, as shown in Figure 4.6. Since changing ϵ_{LJ} does not make a big difference in the strength of the interaction, as shown in Figure 4.7, 0.3 is used for the depth of the Lennard-Jones interaction, ϵ_{LJ} .

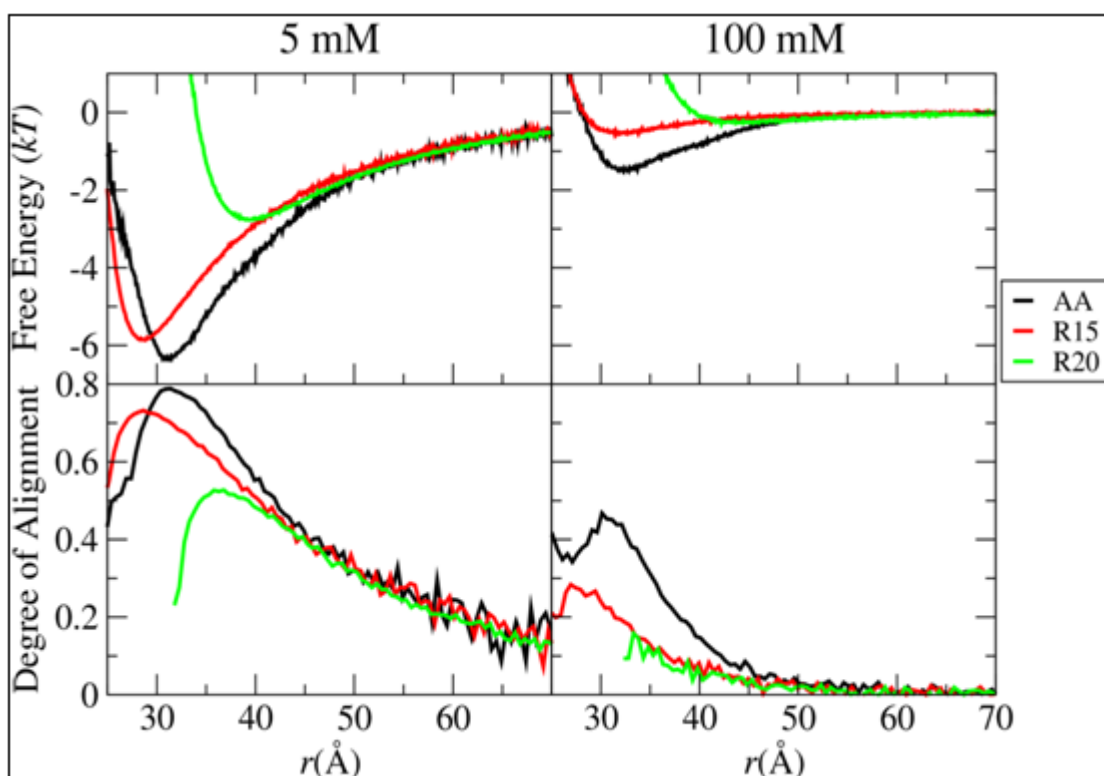


Figure 4.6. The radius (R) dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and TS models at 5 mM and 100 mM 1:1 salt concentrations

The potential of mean force and the degree of alignment for the coarse graining models from the most detailed one to the simplest one are shown altogether in Figure 4.8.

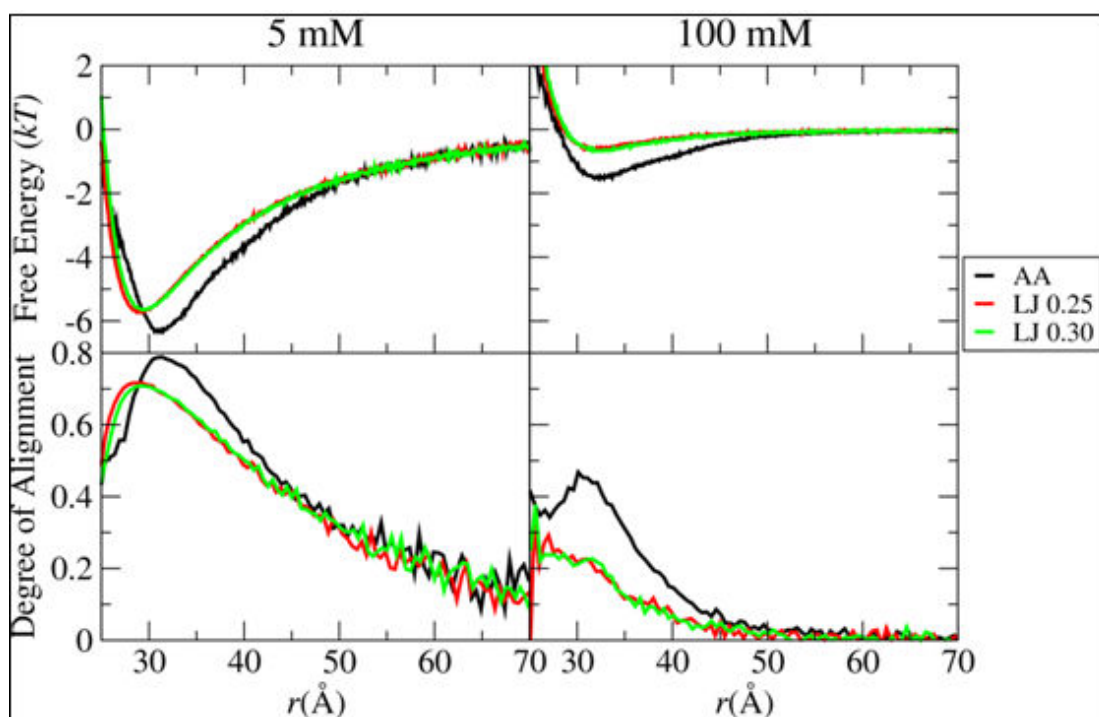


Figure 4.7. ϵ_{LJ} dependency of the pmf and the degree of alignment between LYS and the apo α -LA in AA and TS models at 5 mM and 100 mM salt concentrations

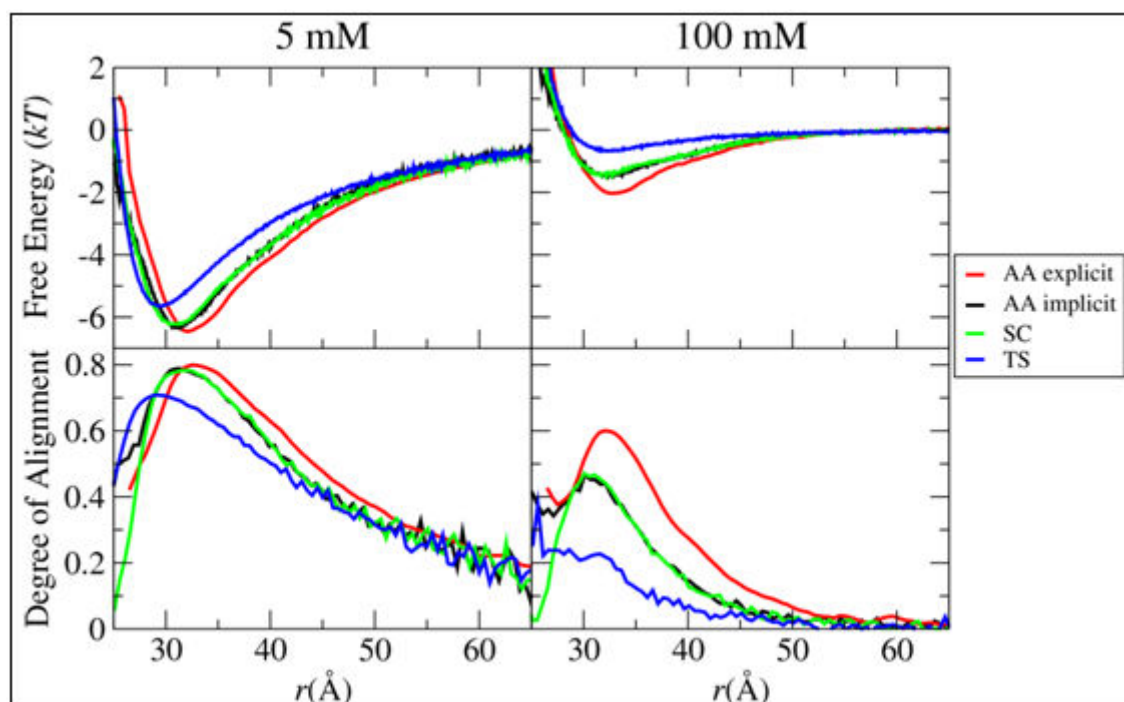


Figure 4.8. The potential of mean force and the degree of alignment between LYS and apo α -LA for the employed coarse graining models at 5mM and 100 mM salt concentration

4.3. MANY BODY SIMULATIONS

Since a major part of this work is to investigate the phase behavior of aqueous LYS/ α -LA mixtures, it is necessary to simulate more than two proteins. However, at this point, it is essential to know the limitations of the system such as, the minimum number of proteins sufficient to simulate phase separation and the minimum salt concentration that can be simulated without having box size effects on the truncated potentials, as described in Section 2.4.5. In order to determine these limits, a dependency analysis is performed and the results are given in the following section.

It is important to notice that all many body simulations are performed for an equal number of LYS and α -LA molecules. The salt used in all simulations is monovalent and the temperature is 298 K. In addition, since the volume of the simulation box is fluctuating in the many body simulations the free energy of the system is given with respect to box length unless the systems containing different numbers of proteins are compared. If such systems are considered, the free energy per particle with respect to protein concentrations is given in order to make these systems comparable.

4.3.1. Size Dependency

The first attempt for the size dependency analysis of the system is performed with the SC model and the many body simulations are performed for 10, 20 and 40 number of proteins at 5 and 25 mM with SC model, as shown in Figure 4.9.

Although the free energy curves obtained from these simulations are supposed to be similar, the resulting curves are quite different from each other. However these differences cannot be interpreted as size dependency of the system because of the insufficient sampling of the configurations and hence poor convergence.

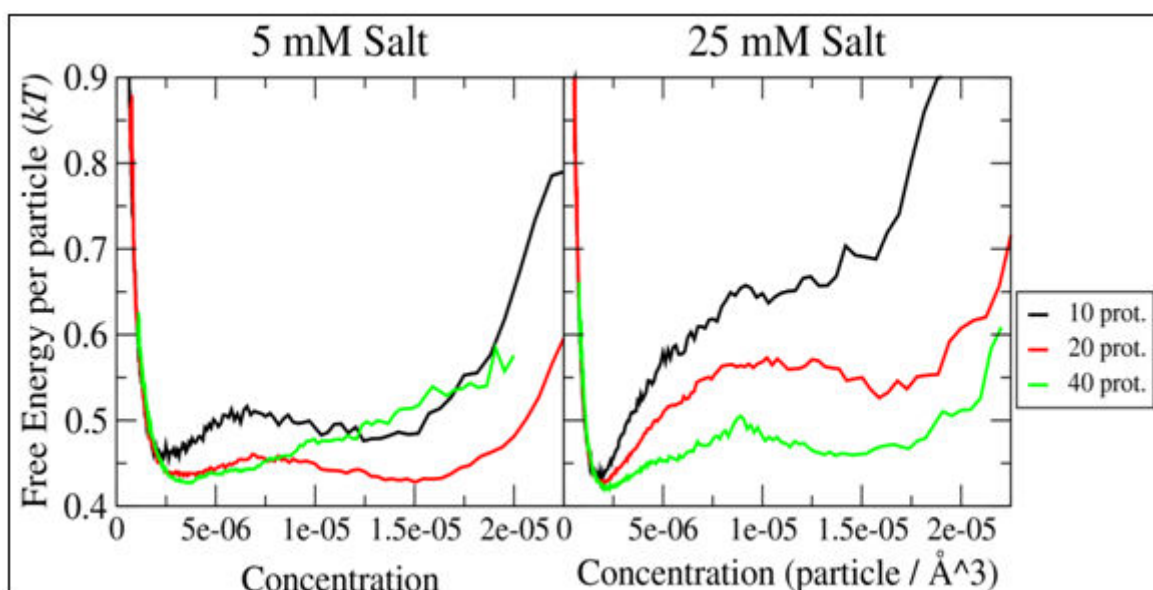


Figure 4.9. The free energy of LYS and apo α -LA for various numbers of proteins at 5 and 25 mM salt concentration with the SC model

The second attempt for size dependency analysis is performed with TS model because it ensures complete convergence of the curves demanding less CPU time, as shown in Figure 4.10.

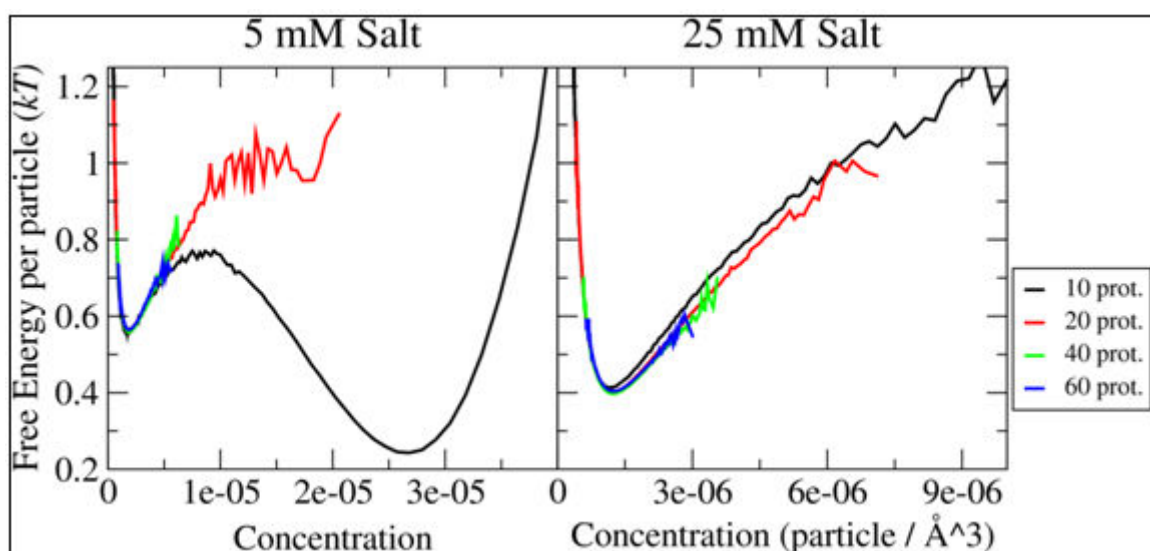


Figure 4.10. The free energy of LYS and apo α -LA for various numbers of proteins at 5 and 25 mM salt concentration with TS model

When the results obtained with the TS model are considered, the simulation with 10 proteins at 5mM salt does definitely not denote the same characteristic with the others. It is concluded that the range of interaction energy in a system with 10 proteins at 5 mM salt is beyond the simulated box size and therefore the minimum image truncation of interaction energy introduces artifacts. Thus, phase transition simulations should be performed with 20 proteins or more and no lower than 25 mM salt. In addition, it is clear that the very simple TS model is incapable of predicting the phase transition in the system. Thus, it can be concluded that the surface charge density and the surface topology have crucial effect on the phase transition and hence, self assembly of LYS and α -LA.

4.3.2. Improvement of Sampling

Since a difficulty in obtaining converged curves is encountered in the analysis of size dependency, the simulations are performed dividing the box length into small windows with minimum and maximum box length restrictions in which the free energy differences in the sampled configurations do not exceed five kT . Since the lower energy states have priority in the sampling of the configurations, a free energy barrier higher than five kT between two phases may cause sticking in one of those free energy minima and insufficient sampling of configuration space.

At first, the isobaric program is tested for neutral hard spherical particles (interaction energy is infinity when they are overlapping and zero otherwise) without any salt in the system, and the results are presented in Figure 4.11. Since an excellent agreement with the restricted box length simulations and the unrestricted one is obtained, this method is used for simulation of the protein system.

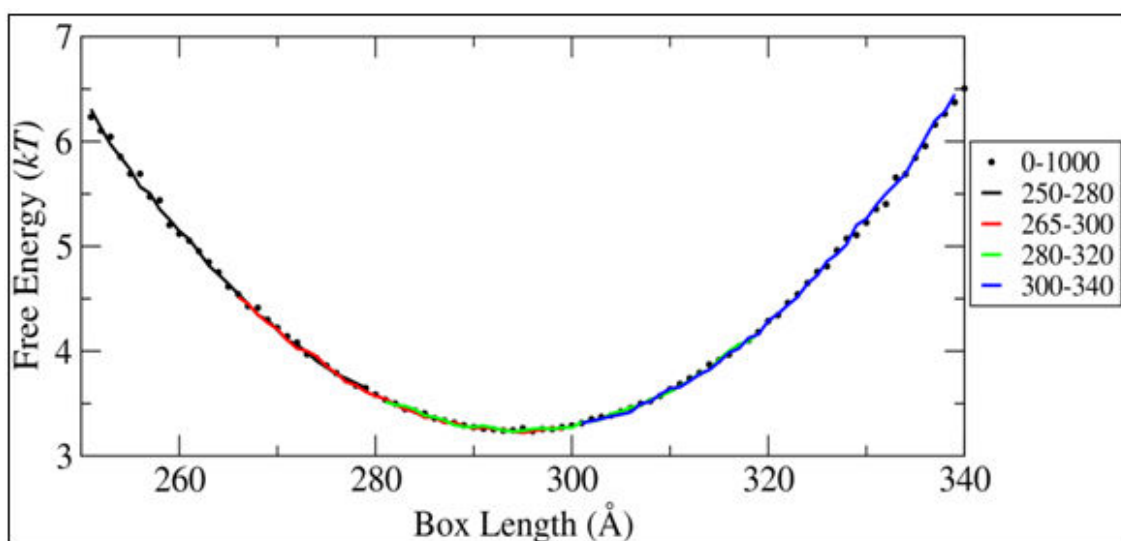


Figure 4.11. Divided window simulation for 30 ideal particles (Circles denote the unrestricted box length simulation while the curve labels correspond to minimum and maximum box lengths allowed in the simulations)

The simulation of protein systems in divided windows has been started for high salt concentrations case in order to test the method for relatively simple one phase system. Therefore, the system which contains 20 proteins at 55mM salt concentration is simulated in divided windows and they are compared with the unrestricted one in Figure 4.12.

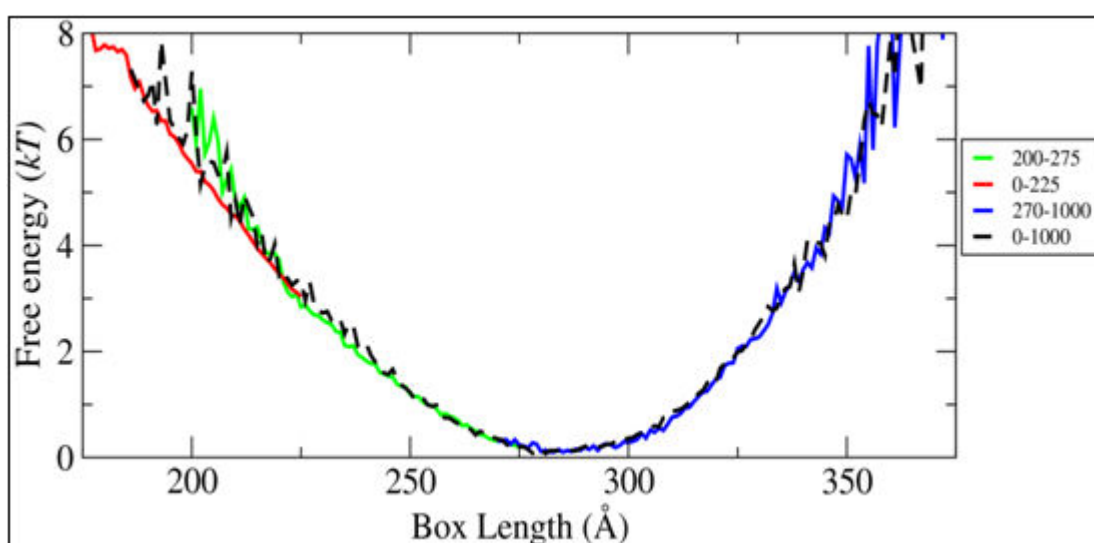


Figure 4.12. The divided window simulations performed for 20 proteins at 55mM salt concentration

After testing the program with one phase protein system, the divided window method is applied to systems where phase transition is expected. The simulation is performed for 20 proteins at 25 mM salt concentration and 0.6 mM pressure, as shown in Figure 4.13. Due to the insufficient dense phase sampling, the free energy curve obtained from small box and the medium box size cannot be combined properly. Since the choice of displacement parameter in the translation moves is important to achieve sufficient sampling of configuration space and fast convergence [37], the sampling in the dense phase configurations is improved by implementing a displacement parameter as 0.08 % of the square of box length, as described in Section 3.5. The scaling of displacement parameter according to volume or higher order power of box length is not preferred in order to avoid translations longer than the box size [36].

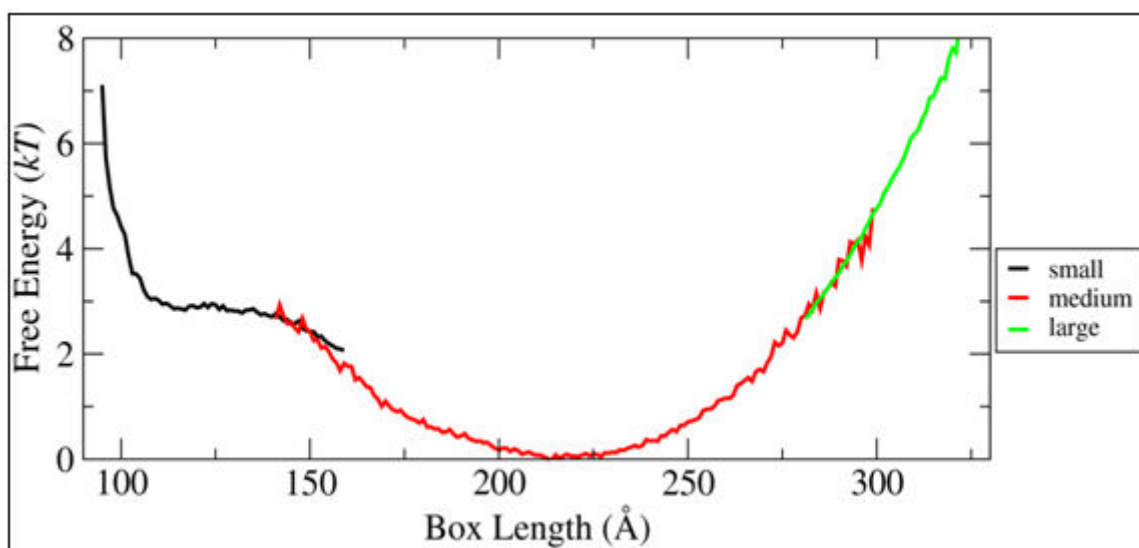


Figure 4.13. Free energy difference for 20 proteins at 25mM salt concentration and 0.6 mM pressure

The free energy curves with displacement parameter scaling are given in Figure 4.14. The pressure in the simulation is increased in order to force the system into the aggregated phase and to see the improvement in the dense phase sampling clearly.

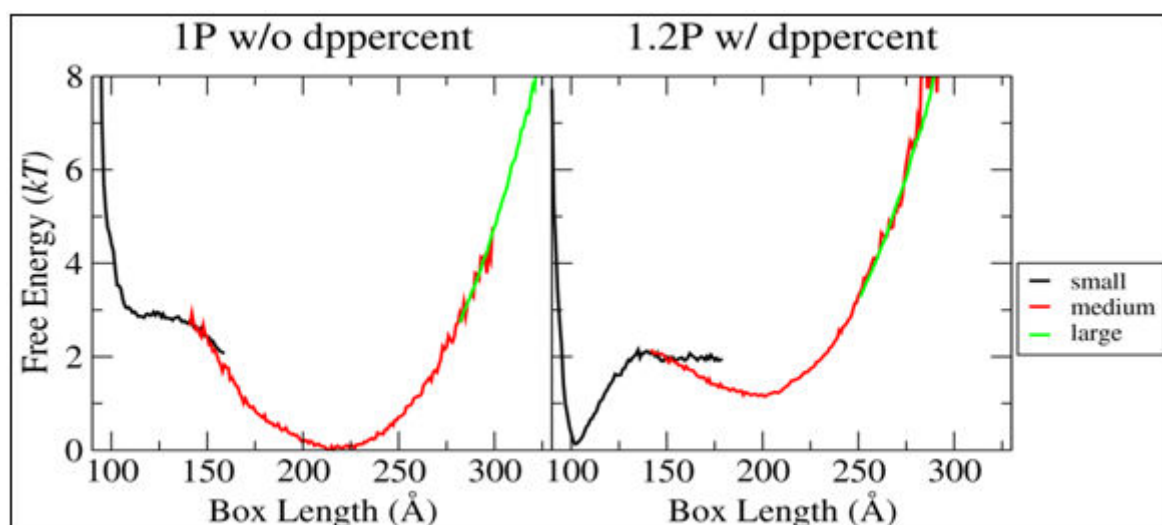


Figure 4.14. The free energy of 20 proteins system at 25 mM salt concentration with and without displacement parameter scaling at 0.6 (left) and 0.7 mM pressure (right)

Although less statistical error in the dense phase curve is an indication of sampling improvement in the small box sizes the combination problem is still present. A possible explanation in a hysteresis effect (strong pair attraction): since disintegration of the clusters in the dense phase is difficult, this may diminish the generation of independent configurations and results in improper sampling of configurational space. The system is sticking to one local minimum, depending on the initial configuration. In order to avoid hysteresis in the system three different moves are implemented into the program: Combined translation and rotation, long displacement translation and a cluster move, as shown in Figure 4.15. The perfect combination of windows is achieved using these three moves and the combined one shows good agreement with the one window simulation.

In the mean time, the execution of the simulations is speeded up by improving parallelization of execution and by the help of available computers in new cluster (computer farm) “Platon” in order to achieve satisfactory convergence in the unrestricted box size simulations. In addition, the speeding the program up enables to increase the simulated number of proteins and hence, possibility of generating independent configurations.

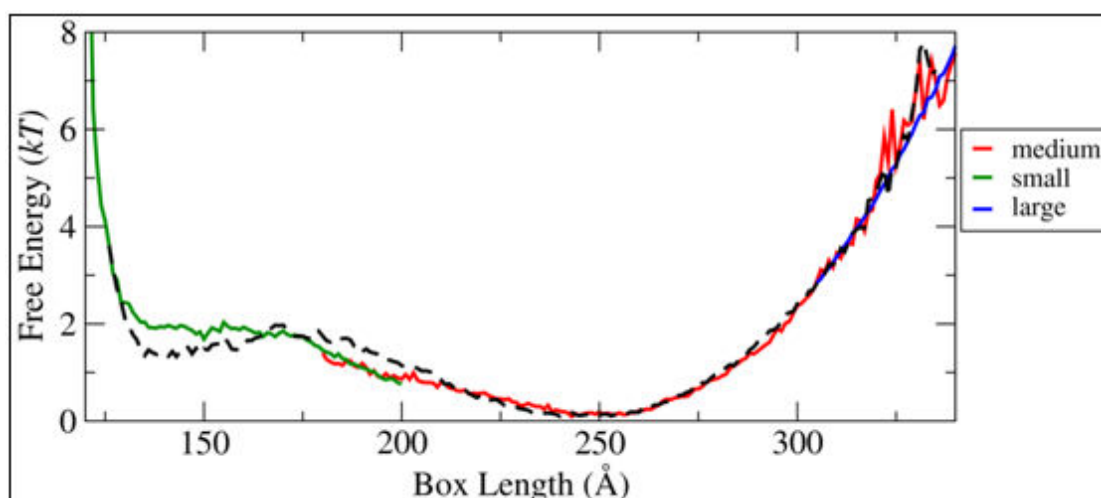


Figure 4.15. Free energy of system containing 40 proteins at 30 mM salt concentration simulated in 3 windows (solid lines) and 1 window (dashed line)

The salt concentration is increased to reduce the strong electrostatic attraction in the possible clusters thereby enhancing the sampling in the dense phase region. The implemented moves and speed up definitely improve the sampling efficiency of the system and allow reproducible free energy curves for the unrestricted box size simulations, as shown in Figure 4.16.

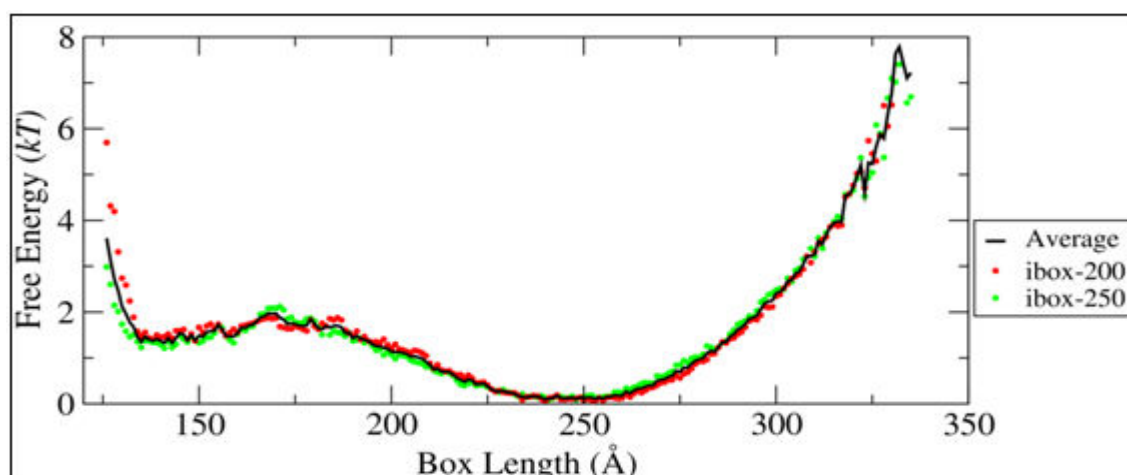


Figure 4.16. Free energy of LYS and α -LA system containing 40 proteins at 30 mM salt concentration and 0.7 mM pressure

(The dots denotes the simulations of the same system starting from different initial configurations and the line denotes average of these two)

4.3.3. Reproduction of Experimental Observations

After the sufficiently sampled and converged free energy curves are achieved the phase transition characteristic of LYS and α -LA system is studied by mimicking the experimental physico-chemical conditions. Nigen and colleagues reported that LYS has ability to self assemble into various supramolecular structure with only apo- α -LA, while unable to do so with holo α -LA [28]. It has been shown that the ability to form aggregates decreases with increasing salt concentration and aggregation completely disappears at 100 mM salt [2]. These characteristics of the self-assembly of LYS and α -LA are reflected excellently in the simulation results, as shown in Figure 4.17.

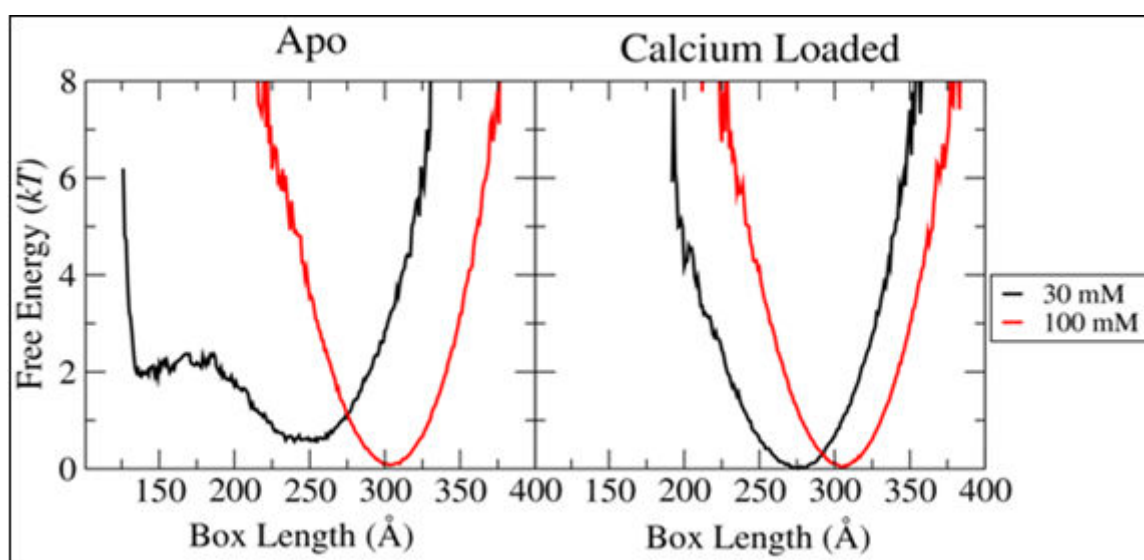


Figure 4.17. Free energy differences of LYS- apo (left) and holo (right) α -LA system containing 40 proteins at 30 (black) and 100 mM (red) salt concentration and 0.7 mM pressure

LYS-holo α -LA system does not show any two phase region at all. However, the phase transition in LYS-apo α -LA system is observed at low salt concentration and it is not obtained at 100 mM salt concentration, as also observed experimentally. At this point, it should be stressed that *no* effort has been made to fit our input parameters in order to obtain this result.

In order to investigate the uneven charge distribution and hence degree of alignment effect on the formation of aggregates, the charges on the SC model of α -LA has been modified so that its dipole moment is been almost abolished while maintaining the net charge. The simulation is performed using this modified apo α -LA structure at 30 mM salt concentration, as shown in Figure 4.18. Since the obtained two phase region in the free energy is not very distinct as in original one, obviously the uneven charge distribution of α -LA does play role in the formation of the aggregates. However it is not the main reason of the aggregation. In addition, decreasing dipole moment of α -LA causes reduction in the attraction and consequently shifts the phases towards to the big box sizes. It can be concluded that although the alignment of α -LA has an impact on the phase separation, it is not as pronounced as the electrostatic attraction between two oppositely proteins as expected [4].

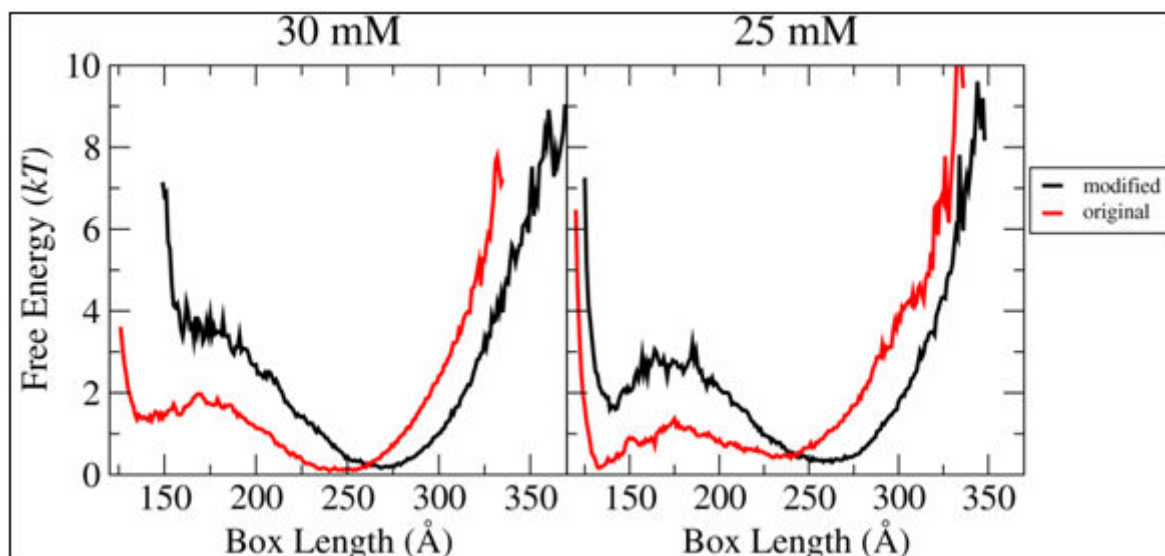


Figure 4.18. The free energy difference of LYS and modified apo α -LA system containing 40 proteins at 30 and 25 mM salt concentration and 0.7 mM pressure.

5. CONCLUSION AND FUTURE WORK

5.1. CONCLUSION

In this project, a mesoscopic study of the self assembly of LYS- α -LA in aqueous salt solution is presented. Traditional metropolis MC simulations with periodic boundary conditions and minimum image convention are performed to investigate phase transition behavior of LYS- α -LA. Water is treated as a structureless dielectric continuum, and the salt particles are represented both explicitly and implicitly (DH level representation) in solution.

In order to describe proteins, two different coarse grained models have been developed. In the already established amino acid model (AA), each amino acid in the protein is represented by a sphere and the pH dependent average charges of the amino acids are placed at the center of masses of each sphere. The newly developed surface charge model maintains charged amino acids at their original locations, while neutral part of the protein is replaced with one big sphere. The last model developed is “twin sphere” (TS) model in which protein is treated a cluster of merely two oppositely charged spheres. The net charges of the proteins are determined using a grand canonical titration scheme in which the proteins are allowed to titrate according to pH and salt concentration.

In order to determine parameters of the new coarse grained protein models, two-body simulations are performed using the *NVT* ensemble and the detailed AA model is taken as a reference point. The SC model is in excellent agreement with the AA model both for alignment and intermolecular free energies.

The phase transition of LYS and α -LA is investigated by performing many body simulations using the isobaric-isothermal ensemble. In order to improve sampling of configuration space, three different MC moves are implemented: Combined translation and rotation, long displacement translation and cluster move. In addition, displacement parameter is scaled with according to the square of the box length to increase the efficiency of sampling in the dense phase.

The results obtained from the many body simulations show excellent agreement with the experimental observations. The coexistence of two phases is obtained only for LYS-apo α -LA at low salt concentrations. The dense phase, which corresponds to formation of aggregates, is completely absent when the salt concentration increases or the calcium bound holo form of α -LA is used. In addition, in order to determine the effect of dipole moment of α -LA on formation of aggregates, SC model of α -LA is modified so that its dipole moment is diminished and it is found that the uneven charge distribution on α -LA has significant influence on the formation of aggregates but not as dominant as electrostatic attraction.

The importance of a mesoscopic description is evident from the fact that small point mutations or the insertion of calcium can change the aggregation behavior significantly. To the best of our knowledge, this study is the first theoretical work that addresses protein phase transition at a near atomistic level.

5.2. FUTURE WORK

In order to further speed up the simulations the separation dependent coarse graining is started to be implement, as described in Section 3.3. The potential of mean force between LYS and α -LA is obtained without losing any information using 40 Å cut off radius, as shown in Figure 5.1. The model is tested for simple molecules containing 4 charged particles and for LYS- α -LA system, as seen in Figure 5.2.

In the following study, this separation dependent coarse graining model will be improved and applied for the LYS and α -LA systems containing many proteins. When the phase transition characteristics of well studied LYS and α -LA system will be predicted using this model, the systems containing more complex proteins will also be investigated.

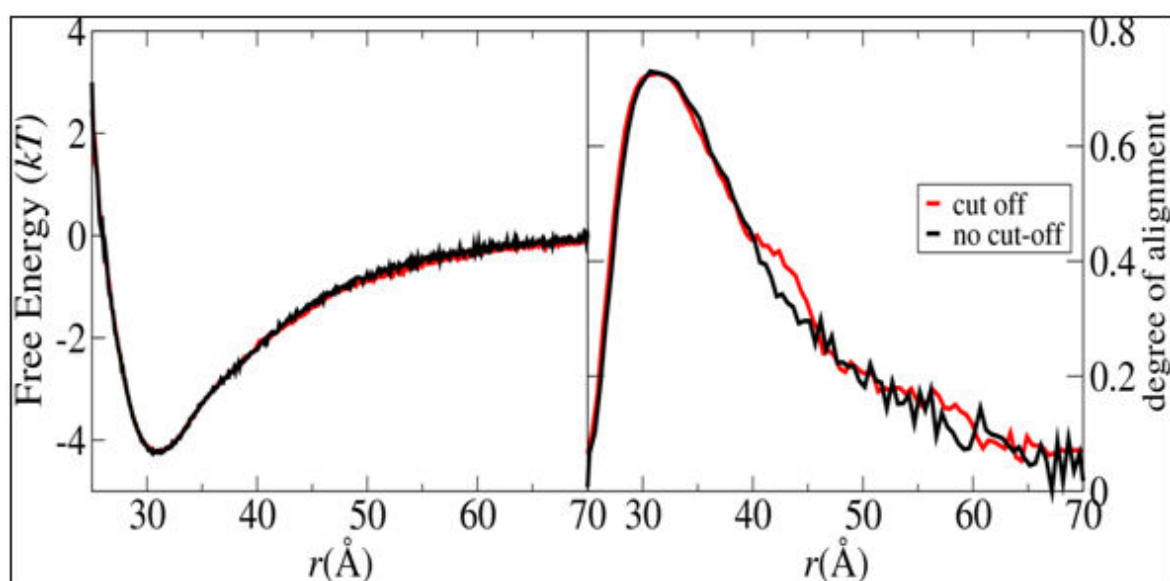


Figure 5.1. The pmf and the degree of alignment between LYS and apo α -LA with a cut off radius 40 \AA (red) and without cut off (black)

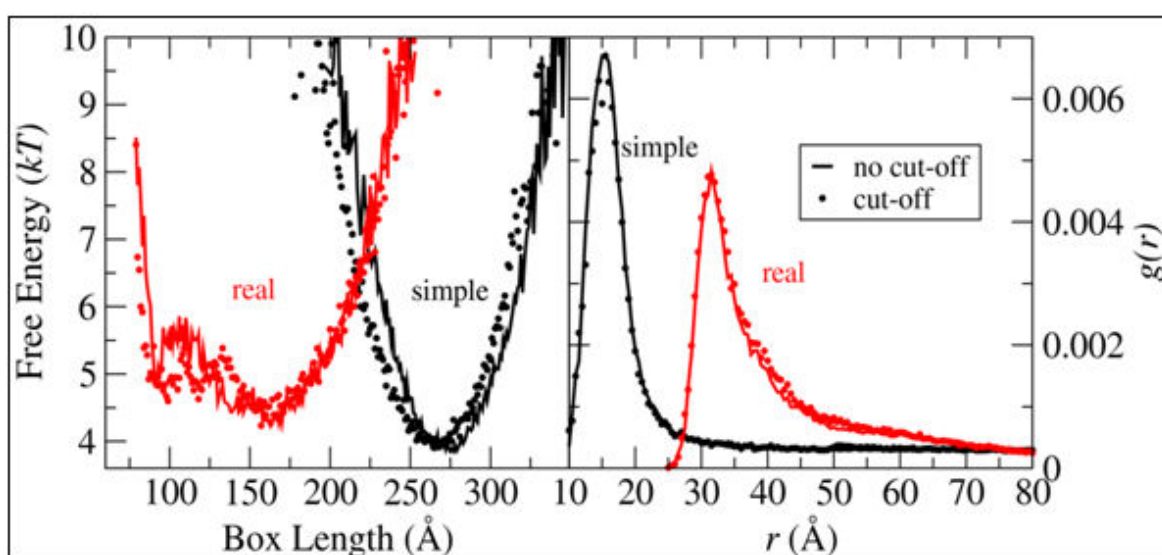


Figure 5.2. Free energy (as a function of box length, left) and radial distribution function (as a function of protein separation, right) of the system containing 10 real (red) and simple (black) proteins at 30 mM salt concentration and 0.7 mM pressure with (circles) and without (line) dipole approximation.

APPENDIX A: CODED AMINO ACIDS

Table A.1. Amino acids and their basic properties [50]

| Amino Acid Name | 3-letter Code | 1-letter Code | Mass | Volume | pK _a | pI | Solubility |
|-----------------|---------------|---------------|--------|----------------|-----------------|---------|----------------|
| | | | Da | Å ³ | Side Chain | at 25°C | g/100g at 25°C |
| Alanine | Ala | A | 71.09 | 88.6 | - | 6.107 | 16.65 |
| Arginine | Arg | R | 156.19 | 173.4 | ~12 | 10.76 | 15 |
| Aspartic Acid | Asp | D | 115.09 | 111.1 | 4.5 | 2.98 | 0.778 |
| Asparagine | Asn | N | 114.11 | 114.1 | - | - | 3.53 |
| Cysteine | Cys | C | 103.15 | 108.5 | 9.1-9.5 | 5.02 | very high |
| Glutamic Acid | Glu | E | 129.12 | 138.4 | 4.6 | 3.08 | 0.864 |
| Glutamine | Gln | Q | 128.14 | 143.8 | - | - | 2.5 |
| Glycine | Gly | G | 57.05 | 60.1 | - | 6.064 | 24.99 |
| Histidine | His | H | 137.14 | 153.2 | 6.2 | 7.64 | 4.19 |
| Isoleucine | Ile | I | 113.16 | 166.7 | - | 6.038 | 4.117 |
| Leucine | Leu | L | 113.16 | 166.7 | - | 6.036 | 2.426 |
| Lysine | Lys | K | 128.17 | 168.6 | 10.4 | 9.47 | very high |
| Methionine | Met | M | 131.19 | 162.9 | - | 5.74 | 3.381 |
| Phenylalanine | Phe | F | 147.18 | 189.9 | - | 5.91 | 2.965 |
| Proline | Pro | P | 97.12 | 112.7 | - | 6.3 | 162.3 |
| Serine | Ser | S | 87.08 | 89 | - | 5.68 | 5.023 |
| Threonine | Thr | T | 101.11 | 116.1 | - | - | very high |
| Tryptophan | Trp | W | 186.12 | 227.8 | - | 5.88 | 1.136 |
| Tyrosine | Tyr | Y | 163.18 | 193.6 | 9.7 | 5.63 | 0.0453 |
| Valine | Val | V | 99.14 | 140 | - | 6.002 | 8.85 |

APPENDIX B: NET CHARGES OF LYS AND ALPHA-LA

Table B.1. Net charges of LYS and α -LA in a function of pH and salt concentration

| 2 pH | Charges (<i>e</i>) | | | Dipoles (<i>e</i>Å) | | |
|------------------|---------------------------|------------|-------------|----------------------------|------------|-------------|
| Salt (mM) | LYS | apo | holo | LYS | apo | holo |
| 5 | 16.7 | 13.8 | 14.2 | 29.8 | 38.6 | 37.6 |
| 25 | 17.4 | 14.6 | 15.2 | 32.0 | 39.7 | 38.5 |
| 30 | 17.5 | 14.7 | 15.4 | 32.2 | 39.7 | 38.7 |
| 40 | 17.6 | 14.8 | 15.5 | 32.6 | 40.1 | 38.9 |
| 50 | 17.7 | 14.9 | 15.7 | 32.9 | 40.3 | 39.1 |
| 100 | 18.0 | 15.3 | 16.1 | 34.1 | 41.0 | 39.8 |
| | | | | | | |
| 7.5 pH | Charges (<i>e</i>) | | | Dipoles (<i>e</i>Å) | | |
| Salt (mM) | LYS | apo | holo | LYS | apo | holo |
| 5 | 6.7 | -5.8 | -4.4 | 29.1 | 79.6 | 69.6 |
| 25 | 6.8 | -6.2 | -4.6 | 29.3 | 80.7 | 69.5 |
| 30 | 6.8 | -6.2 | -4.6 | 29.3 | 80.9 | 69.5 |
| 40 | 6.8 | -6.3 | -4.6 | 29.3 | 81.1 | 69.5 |
| 50 | 6.9 | -6.3 | -4.6 | 29.3 | 81.2 | 69.5 |
| 100 | 6.9 | -6.5 | -4.7 | 29.4 | 81.6 | 69.5 |
| | | | | | | |
| 12 pH | Charges (<i>e</i>) | | | Dipoles (<i>e</i>Å) | | |
| Salt (mM) | LYS | apo | holo | LYS | apo | holo |
| 5 | -3.2 | -15.0 | -13.8 | 27.7 | 66.3 | 55.0 |
| 25 | -3.7 | -16.9 | -15.6 | 28.4 | 64.8 | 53.4 |
| 30 | -3.8 | -17.1 | -15.8 | 28.5 | 64.6 | 53.2 |
| 40 | -3.9 | -17.5 | -16.1 | 28.7 | 64.3 | 53.0 |
| 50 | -4.0 | -17.8 | -16.5 | 28.8 | 64.2 | 52.7 |
| 100 | -4.4 | -18.7 | -17.3 | 29.4 | 63.9 | 52.3 |

REFERENCES

1. Semenova, M.G., "Thermodynamic analysis of the impact of molecular interactions on the functionality of food biopolymers in solution and in colloidal systems", *Food Hydrocolloids*, Vol. 21, pp. 23-45, 2007.
2. Nigen, M., T. Croguennec, and S. Bouhallab, "Formation and stability of α -Lactalbumin-lysozyme spherical particles: Involment of electrostatic forces", *Food Hydrocolloids*, Vol. 23, pp. 510-518, 2009.
3. van der Linden, E. and P. Venema, "Self-assembly and aggregation of proteins", *Current Opinion in Colloid and Interface Science*, Vol. 12, no 4-5, pp. 158-165, 2007.
4. Desfouge`res, Y., et al., "Charge and Size Drive Spontaneous Self-Assembly of Oppositely Charged Globular Proteins into Microspheres", *J. Phys.Chem*, Vol. 114, pp. 4138-4144, 2010.
5. Nigen, M., et al., "Temperature Affects the Supramolecular Structures Resulting from α -Lactalbumin–Lysozyme Interaction", *Biochemistry : a weekly publication of the American Chemical Society*, Vol. 46, no 5, pp. 1248-1255, 2007.
6. Wang, F. and A.P. Landau, "Determining density of states for classical statistical models: A random walk algorithm to produce a flat histogram", *Physical Review E*, Vol. 46, no. 5, pp. 1248-1255, 2001.
7. Lund, M. and B. Jönsson, "A mesoscopic Model for Protein-Protein Interactions in Solution", *Biophysical Journal*, Vol. 85, pp. 2940-2947, 2003.
8. G. C. Barrette and D.T. Elmore, *Amino Acids and Peptides*, First ed., Cambridge University Press, Cambridge, 2004.

9. Buxbaum, E., *Fundamentals of Protein Structure and Function*, First ed., Springer, Roseau, 2007.
10. Larsen, D., University of California, *The Complete Chemistry Textbook*, 2008, <http://chemwiki.ucdavis.edu/@api/deki/files/6196/>
11. Doonan, S., "Peptides and Proteins (Tutorial Chemistry Texts)", First ed, The Royal Society of Chemistry (RSC), Cambridge, 2002.
12. Stenesh, J., *Biochemistry*. Vol. 2., Plenum Press, New York, 1998.
13. Arendall, B., "Amino Acids", *bioc462 Lecture Notes*, The University of Arizona, 2000.
http://www.biochem.arizona.edu/classes/bioc462/462a/NOTES/Amino_Acids/
14. Shipman, J., J.D. Wilson, and A. Todd, *An Introduction to Physical Science*. 7 ed., D. C. Heath, Boston, 1993.
15. Walsh, G., *Protein Biochemistry and Biotechnology*, John Willy and Sons, London, 2004.
16. Pelesko, J.A., *SELF ASSEMBLY: The science of Things that Put Themselves Together*, Taylor & Francis, Boca Raton, 2007.
17. Whitesides, G., J. Mathias, and C. Seto, "Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures", *American Association for the Advancement of Science*, Vol. 254, no. 5036, pp. 1312-1319, 1991.
18. Dumetz, A.C., et al., "Protein Phase Behavior in Aqueous Solutions: Crystallization, Liquid-Liquid Phase Separation, Gels, and Aggregates", *Biophysical Journal*, Vol. 94(January), pp. 570–583, 2008.

19. Abraham, E.P., "Some Properties of Egg-White Protein", *Biochemical Journal*, Vol. 33, no. 4, pp. 622-630, 1939.
20. "UniProtKB/Swiss-Prot P00698 (LYSC_CHICK)", *UniProtKB/Swiss-Prot Database*, 2010.
21. Walsh, M.A., et al., "Refinement of triclinic hen egg-white lysozyme at atomic resolution", *Acta Crystallogr.*, Vol. D, no. 54, pp. 552-546, 1998.
22. Goodsell, D.S., "Molecule of the Mouth", *RSCS Protein Data Bank*, 2000.
23. Qasba, P.K., S. Kumar, and K. Brew, "Molecular Divergence of Lysozymes and -Lactalbumin", *Critical Reviews in Biochemistry and Molecular Biology*, Vol. 32, no. 4, pp. 255-306, 1997.
24. Noble, M.S., M.B. Wheeler, and W.L. Hurley, " α -Lactalbumin: A Limiting Factor in Milk Production", *Illini DairyNet Papers of University of Illinois*, 2000. <http://www.livestocktrail.uiuc.edu/dairyNet/paperDisplay.cfm?ContentID=323>.
25. Chrysina, E.D., K. Brew, and K.R. Acharya, "Crystal structures of apo- and holo-bovine alpha-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions", *J.Biol.Chem*, Vol. 275, pp. 37021-37029, 2000.
26. Nigen, M., et al., "Apo alpha-lactalbumin and lysozyme are colocalized in their subsequently formed spherical supramolecular assembly", *The FEBS Journal*, Vol. 274, no. 23, pp. 6085-6093, 2007.
27. Nigen, M., et al., "Dynamic and supramolecular organisation of alpha-lactalbumin/lysozyme microspheres", *Biophysical Chemistry*, Vol. 146, pp. 30-35, 2010.

28. Nigen, M., et al., "Molecular interaction between apo and holo alpha-lactalbumin and lysozyme: Formation of heterodimers as assessed by fluorescence measurements", *BBA Protein and Proteomics*, Vol. 1794, pp. 709-715, 2009.
29. Israelachvili, J.N., *Intermolecular and Surface Forces with Applications to Colloidal and Biological Systems*, second ed., Academic Press, 1985.
30. Stone, A.J., *The Theory of Intermolecular Forces*, Oxford University Press, Oxford, 2000.
31. Hill, T.L., *An Introduction to Statistical Thermodynamics*, Dover Publications Inc., New York, 1960.
32. Whalley, E. and W.G. Schneider, "Intermolecular Interactions of Argon Krypton and Xenon", *The Journal of Chemical Physics*, Vol. 23, no. 9, 1955.
33. Lyklema, J., *Fundamentals of Interface and Colloid Science*, Vol. 1, Academic Press, San Diego, 2000.
34. Bresford-Smith, B., "Some Aspects of Strongly Interacting Colloidal Dispersions", *Ph.D. Thesis at Australian National University*, 1985.
35. Broukhno, A.V., *Free Energy and Surface Forces in Polymer Systems Monte Carlo Simulation Studies*, Lund University, Lund, 2003.
36. Frenkel, D. and B. Smit, *Understanding Molecular Simulation From Algorithms to Applications: Computational Science Series*, Academic Press, San Diego, 2001.
37. Metropolis, N., A.W. Rosenbluth, and E. Teller, "Equation of State Calculations by Fast Computing Machines", *The Journal of Chemical Physics*, Vol. 21, no. 6, pp. 1087-1092, 1953.

38. Allen, M.P. and D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Bristol, 1987.
39. Kriz, R.D., MSE2094 class NoteBook, Virginia Tech. College of Engineering, 1999.
http://www.sv.vt.edu/classes/MSE2094_NoteBook/MolecDyn/periodic.gif.
40. Sangster, M.J.L. and M. Dixon, "Interionic potentials in alkali halides and their use in simulations of the molten salts", *Advances in Physics*, Vol. 25, no. 3, pp. 247-343, 1976.
41. Persson, B.A. and M. Lund, "Association and electrostatic steering of α -Lactalbumin-Lysozyme heterodimers", *Physical Chemistry Chemical Physics*, Vol. 11, no. 39, pp. 8879-85, 2009.
42. Gmelin, L. and H. Watts, *Hand Book of Chemistry*, Vol. 7, University of Michigan Library, 1848.
43. Pellicane, G., G. Smith, and L. Sarkisov, "Molecular Dynamics Characterization of Protein Crystal Contacts in Aqueous Solution", *Physical Review Letter*, Vol. 101, no. 24, 2008.
44. Lund, M., M. Trulsson, and B. Persson, "Faunus: An object oriented framework for molecular simulation", *Source Code for Biology and Medicine*, Vol. 3, no. 1, 2008.
45. Lund, M. and B. Jonsson, "On the Charge Regulation of Proteins", *Biochemistry*, Vol. 44, pp. 5722-5727, 2005.
46. Labbez, C. and B. Jönsson, "A New Monte Carlo Method for the Titration of Molecules and Minerals in Applied Parallel Computing", *State of the Art in Scientific Computing*, pp. 66-72, Springer, Berlin, 2007.
47. Dress, C. and W. Krauth, "Cluster algorithm for hard spheres and related systems", *Journal of Physics A: Math. and Gen.*, Vol. 28, no. 23, pp. L597-L601, 1995.

48. Liu, J. and E. Luijten, "Rejection-free Geometric Cluster Algorithm for Complex Fluids", *Physical Review Letters*, Vol. 92, no. 3, pp. 035504(1)-035504(4), 2004.
49. Teske, C.A., H.W. Blanch, and J.M. Prausnitz, "Chromatographic measurement of interactions between unlike proteins", *Fluid Phase Equilibria*, Vol. 219, no. 2, pp. 139-148, 2004.
50. Sühnel, J., *The Amino Acid Repository*, Leibniz Institute for Age Research - Fritz Lipmann Institute, 2005.