LUND UNIVERSITY

Anisotropic Protein Interactions in Salt Solutions and at Interfaces: Coarse Grained Modeling

Kurut Sabanoglu, Anil

2014

Link to publication

Citation for published version (APA):
Kurut Sabanoglu, A. (2014). *Anisotropic Protein Interactions in Salt Solutions and at Interfaces: Coarse Grained Modeling*. [Doctoral Thesis (compilation), Computational Chemistry]. Theoretical Chemistry, Lund University.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Anisotropic Protein Interactions in Salt Solutions and at Interfaces: Coarse Grained Modeling

Anıl Kurut Şabanoğlu

Division of Theoretical Chemistry
Lund University, Sweden

LUNDS
UNIVERSITET

Doctoral Thesis

The thesis will be publicly defended on Friday 19^{th} of September 2014, at 13:15 in lecture hall B, Center for Chemistry and Chemical Engineering (Kemicentrum), Lund

The faculty opponent is Jens Erik Nielsen, Novozymes

The grading committee is Ilja Voets (Eindhoven University of Technology), Ingemar André (Lund University), Martin Malmsten (Uppsala University).

Front cover: Anisotropic protein interactions depicted by cookies
© Anıl Kurut Şabanoğlu, Korhan Şabanoğlu, Lena Tran, and Paula Leckius.

Anisotropic Protein Interactions in Salt Solutions & at Interfaces:
Coarse Grained Modeling

© Anıl Kurut Şabanoğlu
Doctoral Thesis 2014

Theoretical Chemistry
Center for Chemistry and Chemical Engineering
Lund University
P.O. Box 124
SE-221 00 Lund
Sweden
akurut@gmail.com

All rights reserved
ISBN 978-91-7422-365-1
Printed in Sweden by Media-Tryck, Lund University, Lund

Organization	Document name
LUND UNIVERSITY	DOCTORAL DISSERTATION
Center for Chemistry and Chemical Engineering P. O. Box 124 SE-221 00 Lund, Sweden	Date of issue 2014-08-26
Author(s)	Sponsoring organization
Anıl Kurut Şabanoğlu	Organizing Molecular Matter, OMM

Title and subtitle

Anisotropic Protein Interactions in Salt Solutions and at Interfaces: Coarse Grained Modeling

Anisotropic protein interactions have a strong orientation dependence resulting from an uneven distribution of charged and hydrophobic residues on the protein surface. They play an important role in protein behaviors such as protein association, surface adsorption and phase separation. In this thesis, we have studied the effect of anisotropic interactions on the behavior of various proteins mainly by focusing on electrostatic interactions. We have developed coarse grained models, specific to each system by considering their essential details and used Metropolis Monte Carlo method to simulate protein behaviors in salt solutions and at charged interfaces. We show that anisotropic dipolar interactions may overcome the net charge repulsion between similarly charged proteins and favor the protein association. The strong directionality of these interactions may reinforce specific protein orientations, required for protein activity. Note that hydrophobic anisotropy can also compete with the directionality of the dipolar interactions and may force the proteins into less favorable dipole orientations. We also show that the charge regulation effects and the specific Hofmeister ion binding can significantly alter the charge distribution of proteins, and thus they should not be overlooked in the studies of protein electrostatics. Our results indicate that to gain a comprehensive understanding of protein electrostatics, one needs to consider: (i) the higher order multipole interactions; (ii) the hydrophobic patchiness that can compete with the multipole interactions; (iii) the charge regulation effects; as well as (iv) the specific ion binding. The extent of these factors can roughly be estimated by examining the dipole moment, the locations of hydrophobic patches, the number of residues with acid dissociation constants around solution pH as well as the concentration of binding ions and the exposed area of their binding sites.

Key words

Anisotropic interactions, protein electrostatics, phase association, surface adsorption, coarse grained models, Monte Carlo simulations

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language English
ISSN and key title		ISBN 978-91-7422-365-1
Recipient´s notes	Number of pages 156	Price
	Security classification	

Distribution by (name and address)

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature A. Kurut Date August 12, 2014

Preface

Theoretical studies are excellent complimentary tools to experimental techniques and provide a deeper understanding on "*What molecules are experiencing in their small world*". The main goal of this thesis is to uncover the molecular details of experimentally observed protein behaviors using theoretical tools such as Monte Carlo simulations and coarse grained modeling.

The thesis is composed of six publications where various phenomena of protein systems are investigated by focusing on anisotropic interactions. The studied protein systems are considered in two main parts: *The first* covers solution behavior of a few globular proteins while *the second* deals with surface adsorption of intrinsically disordered proteins. The publications are referred to in the text by their roman numerals provided in the list of publications section.

In this thesis, you will first be introduced to the world of proteins and their implications in our lives. Then, we continue with the interactions that are crucial for operational proteins. In the *third* chapter, an overview of relevant theoretical and experimental tools will be given. Later, coarse grained modeling and the models used in the publications will be presented. The *last* chapter will provide an atlas of studied proteins together with the main findings of the publications. Finally, the thesis is concluded by an outlook including a future perspective to the presented studies.

I believe that a PhD thesis should not only govern the scientific discoveries but also deliver the experiences that are obtained during entire period of the PhD studies. Therefore, you will also find some instructions and guidelines for *how to develop coarse grained models* or *how to set up MC simulations* in the thesis.

Throughout the thesis, you will also encounter grey boxes, called *"box of imaginations"* where I imagine daily life metaphors for presented topics. I hope that you will find them both amusing and clarifying.

Although I have studied a tiny part of our lives connected to protein systems, I hope that the collection of these tiny bits will provide us a deeper awareness and understanding of ourselves and the world around us. I hope that the scientific achievements will eventually guide us to be more compatible with our planet and to protect its essentials.

List of Publications

I ***Dimerization of terminal domains in silk fibers: Hydrophobic and electrostatic anisotropy***
Kurut A., C. Dicko, and Lund M., *Manuscript*

II ***Anisotropic interactions in protein mixtures: Self assembly and phase behavior in aqueous solution.***
Kurut, A., Persson B.A., Åkesson, T., Forsman, J., and Lund, M., *J. Phys. Chem. Lett.*, 2012, 3(6), p731-734

III ***Solution electrostatics beyond pH: A coarse grained approach to ion specific interactions between macromolecules.***
Kurut A., and Lund M., *Faraday Discuss.*, 2013, 160, p271-278

IV ***Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk.***
Kurut, A., Henriques, J., Forsman, J., Skepö, M., and Lund, M., *Proteins*, 2014, 82(4), p657-667

V ***High expression of midkine in the airways of patients with cystic fibrosis.***
Nordin, S., Jovic, S., **Kurut, A.**, Andersson, C., Gela, A., Bjartell, A., Mörgelin, M., Olin, A., Lund, M., and Egesten, A., *Am. J. Resp. Cell. Mol.*, 2013, 49(6), p935-942

VI ***Adsorption of beta-casein to hydrophilic silica surfaces. Effect of pH and electrolyte.***
Svensson, O., **Kurut, A.**, and Skepö, M., *Food Hydrocolloids*, 2014, 36, p332-338

List of Contributions

I I have developed the SASA based hydrophobic model and performed all simulations and written the manuscript except the introduction part.

II I have co-developed the model and performed all simulations and written the manuscript.

III I have developed implicit ion binding model. I performed all simulations and prepared the scripts for plotting the figures. I was also involved in the revision of the manuscript.

IV I have modeled explicit zinc binding to histidines. I have performed the MC simulations and written the manuscript except the MD and DFT method part.

V I have performed the simulations and written the theoretical part of the manuscript.

VI I have designed and performed the experiments in 1, 2, and 8 mM calcium salt at pH 8.5; 140 mM sodium salt at pH 7, 20 and 90 mM sodium salt at pH 8.5. I have also contributed to the discussion and method parts of the manuscript.

Papers not Included in the Thesis

1. ***Faunus - a flexible framework for Monte Carlo simulation***.
Stenqvist, B., Thuresson, A., **Kurut, A.**, Vacha, R., and Lund, M., *Molecular Simulation*, 2013, 39(14-15), p1233-1239

Highlights of the Thesis

Proteins in Salt Solutions

- **Paper I**: Even in similarly net charged protein systems, attractive electrostatic interactions contribute to the binding affinity through charge anisotropy. Strongly directional dipolar interactions can drive the proteins into specific orientations and facilitate protein association.

- **Paper II**: Anisotropic electrostatics can still be significant in the oppositely charged protein mixtures where the driving force for phase separation is often considered to be the opposite net charge attraction. However, in lysozyme/α-lactalbumin mixtures, the dipolar interactions originating from the strong dipole moment of α-lactalbumin modulate the phase equilibrium. Suppressing its dipole moment diminishes the phase separation behavior.

- **Paper III**: Specific Hofmeister ion binding to hydrophobic motifs may significantly alter the titration curves and the hydrophobic surface exposures of proteins, leading to an inverted solution stability below and above the isoelectric point.

Proteins at Charged Interfaces

- **Paper IV**: Charge regulation effects become significant when the environmental pH matches the acid dissociation constant ($\mathrm{p}K_a$) of the most abundant residues. Therefore, at physiological conditions, the histidine content of proteins determines the extent of these effects. As a response to a surface electric field, histidine (HIS) residues can charge themselves up by proton uptake and provide an additional affinity to negatively charged surfaces. On the contrary, HIS can also adapt a neutral state in the

surfaces. On the contrary, HIS can also adapt a neutral state in the vicinity of positive surfaces and in non-polar environments where charges are unfavorable.

- **Paper V**: The titration of the membrane charged groups can also contribute to the pH response of protein adsorption in bacterial membrane. This contribution can roughly be included by estimating the surface charged density based on the solution pH and an effective membrane pK_a. However, this estimation will require a knowledge of membrane phospholipids and their pK_a values as well as their content in the membrane.

- **Paper VI**: Although the net charge of an amphiphilic disordered protein is similar to the surface charge, its surface adsorption can still be facilitated by relocating the oppositely charged residues close to the surface. The balance between overall repulsion and the local residue attractions may lead to a non-monotonic adsorption response to salt concentration as we have shown for β-casein adsorption to a silica surface. The adsorption is also affected by the interactions between the adsorbed proteins and their possible aggregation.

Acknowledgements

My journey to Sweden starts in a lunch meeting with a visiting swedish professor, **Ulf Olsson**, and my turkish supervisor **Seyda Bucak** who previously, convinced me to make a major in chemical engineering. In that meeting, Ulf's big smile and adorable, sun tanned, pink cheeks gave me the courage to speak about my craziest hypothesis about how his work on vesicle formation could be an explanation to the formation of first living cells. After that, I had been forced (!) by my turkish supervisor to take a walk with him around the university when we talked about my interest in theoretical studies and obviously, in biological systems. Towards the end of the walk, he asked me that would I like to work on a summer project in Sweden. At that time, the only thought in my mind was how tall he was. Being a member of a family with an average height of 162 cm, he was the tallest person I have ever stood side by side and he gave me the impression that visiting Sweden would be a *magical journey* as in the book *Guinever in Lands of Giants*. This didn't turn out so wrong after all due to having 2m of officemates in average. When I arrived in Sweden, Ulf handed me over to my current supervisor and mentor **Mikael Lund**. At the first glance, he was nothing like I imagined for a Swedish person and he wasn't Swedish either. He was more like one of my family members and during my journey in Sweden, he *indeed* became one of my family, he stood by me in all of my *tough fights* with simulations, swedish bureaucracies and *especially* my mom's cancer. Without these three people in my life, this book would *not* be composed. I will always be grateful to you to turn my life into a magical story.

I would also like to thank to my co-supervisior **Marie Skepö** for being a role model for me as a strong female researcher, showing me that good research and good parenting can be done simultaneously. I am also grateful to **Jan Forsman**, for the fruitful discussions about polymer adsorptions and his comments on my

manuscripts.

I own a special thank to **Björn The Great**. I could never imagine an office-mate better than you. Your warm welcome made me feel that I am belong to here. I still don't understand how could you possibly answer all of my infinitely many questions even when your were writing your thesis!

Ingrid, I don't even know how I could thank you enough for your support and help with all the troubles I encountered in Sweden. You are one of the most incredible person I have ever met who could help a total stranger without even asking a word "thanks" in return. **Paula**, thanks for helping me with printing. Kemicentrum is lucky to have such a caring person like you in the printing office.

I would like to thank all of my colleagues in the department for making this place such a nice working environment. I would also like to thank to all members of physical chemistry for the nice moments in the coffee room.

Sweden brought to me many friends that I have enjoyed each and every second of their presence with me. **Çelen**, I would not be able to survive in the cold swedish days without your warm friendship! Our after work studies were the most fun way of learning about proteins. I am looking forward to being a team with you again. **İrem**, whenever I see you I feel myself at home, thanks for turning the coffee breaks into home visits. You two made my home-sickness bearable.

Marie-Céline, thanks for being there for me all the time. Even though we are separated with 600 km, you are still much closer to me more than many. **Lena**, thanks for being such a caring and fun friend! No one would have support my crazy ideas and enjoy them, like you do. **Robert**, thanks for being such a fun mentor and a great friend at the same time. **Marta**, **Lisa** and **Korina**, I am going to miss so much our girls nights and scientific gossips. **Maxime**, thanks for all late-night office talks in your last year of PhD. **Martin Trulsson**, thanks for making all impossible mac installations fun for me. I have to admit that I might have failed some of them on purpose. **Behisht**, **Emna**, **Manu**, **Deepak**, and **Marianna**, thanks for showing me the pleasure of living in a corridor through our great evening chatings, singings, cookings, bakings and eatings. **Mehdi** and **Sara**, thanks for welcoming me into your family.

The **Thursday's Events Gang** deserves a special thank for making all thursdays nights unforgettable. **Paulius**, thanks for reminding me the existence of kindness in man-kind. We should still continue our traditional sunday brunches! **Svante**, thanks for the joy brought with your passion of competition to our thursday games. **Alexei**, thanks for consistently refusing my dance invitations without being annoyed. **Little Björn**, I really appreciate our hugs! They meant a

lot to me when I was trying to deal with my mom's sickness. **Steven**, thanks for voluntarily melting your brain to read my thesis and saving my life, my future childrens' lives and my future grandchildrens' lives. I will always remember how you calmed me down when I was almost burning the entire corridor. **Axel**, thanks for being there for all kinds of discussions and helping me with the endless mathematical expressions. **Samuel** and **Maja**, thanks for the many delightful parties in your place.

My dear chinese friends, **Fei**, **Hongduo** and **Jessie**, thanks for cheering me up with your limitless happiness. **Alice**, **Evelina** and **Narda** thanks for nice chats.

I also would like to thank to my family for their endless support, especially to **Korhan** who loves me with all of my personal-bugs. I am so lucky to have you beside me.

This work is dedicated to my co-supervisor **Torbjörn Åkesson**
who left us so unexpectedly.
He was the kindest and most enthusiastic teacher I have ever met.
I am deeply grateful that I had the chance to work under his guidance, and will always remember his joyful laugh and morning greetings.

Popular Scientific Summary

Proteins are molecular machines of the human body which regulate all life sustaining processes such as energy production, communication, transport of essential molecules, defense against microbes, etc. They are built by small molecular units called amino acids, bound together like pearls on a necklace. Each amino acid consists of a common part and a unique side chain. The former makes up the backbone of a protein chain and the latter determines the nature of the amino acid. There are 22 kinds of amino acid side chains which provide an enormous diversity to the protein chains. The amino acid side chains can have a water- (polar) or oil-like (hydrophobic) nature. Some polar amino acids may bear a positive or negative charge depending on the environmental conditions such as acidity and salt concentration. It is a well-known phenomenon that water tends to separate from oil. This also holds for an oil-like, hydrophobic amino acid where water tends to push these amino acids together to avoid contact with them. Due to this tendency, the protein chains fold into a globular shape so that most of the hydrophobic amino acids are located in the protein interior and the polar amino acids are on the surface. However, some proteins that lack hydrophobic amino acids do not fold into a globular shape and behave as flexible chains. These proteins are called *intrinsically disordered proteins*.

The nature of protein interactions is determined by the distribution of charged and hydrophobic amino acids on the surface, in case of globular proteins, or in the chain, in case of disordered proteins. The segregation of positive amino acids from the negative ones results in anisotropic interactions which resemble the interactions of two magnets where opposite poles attract and the similar ones repel each other. Thus, the nature of anisotropic interactions depends on the orientations of the proteins. These interactions can also originate from clustering of hydrophobic amino acids on protein surfaces, which creates sticky

surface patches.

In this thesis, we have studied the effect of hydrophobic and charge patchiness of protein surfaces on the protein-protein and protein-surface interactions in salt solutions. We used Monte Carlo simulations to mimic protein behaviors with the help of computers and played around with the acidity and salt concentration of the protein solutions to determine their impact. We have developed system specific protein models which represent proteins by collections of interacting spheres. In the model development, we used a **coarse graining** approach where we have only considered the details of proteins that are essential for the specific study.

We have shown that the hydrophobic and charge patchiness on a protein surface can be altered by binding of charged salt species on the protein surface – called the Hofmeister ion effect. Under correct solution acidity, the charge residues can adapt a charged or neutral state to maximize attraction with opposite charges or to minimize repulsion with similar charges. This phenomenon is called *charge regulation* which can also alter the charge patchiness of the protein surfaces. The resulting complex patch interactions may reinforce specific protein orientations, and may facilitate protein associations into functional machines.

Contents

CHAPTER 1

Proteins: Molecular Machinery

Proteins are molecular machines of living organisms which are involved in almost all vital biological processes. For instance, they are responsible for the energy production and storage in our body. Receptor and hormone proteins facilitate communication between organs working as signal transmitters and receivers. Immuno-proteins construct the defense system of our body against toxic materials and microorganisms. Transfer proteins regulate the logistics of essential molecules, such as oxygen, whereas membrane proteins control the passage of molecules and information in and out of the cell.

In a human body, all of these processes take place *continuously*, ensured by the accurate functioning of the hard working proteins. Despite of some exceptions, the protein activity relies on recognition of protein partners which resembles to a key-lock relation [1]. Most of the proteins have a unique 3D structure embedded with special features to ensure the key-lock recognition. Note that since proteins have some flexibility, the key and the lock are somehow softer and can slightly modify their shape. The protein features are activated in a narrow range of solution conditions which determines the optimum environment for their functioning. In order to understand how proteins operate in a human body, we need to investigate their solution dependent features as well as the interactions with their partners.

[1] A. Finkelstein and O. Ptitsyn, in *Protein Physics: A Course of Lectures* (Acedemic Press, San Diago, 2002), chap. Introduction.

1.1 What Does a Protein Look Like?

1.1.1 Building Blocks

Proteins are linear chains of 22 kinds of parts, the so called amino acids or residues, which are connected via strong covalent peptide bonds. All amino acids contain an amino group ($-NH_2$), a carboxyl group (–COOH) and a hydrogen bound central carbon atom (-C^{α}H-) which constitute the skeleton (backbone) of a protein chain – see Figure 1.1. Each amino acid also includes a specific side chain that determines its nature. Based on the differences in the side chains, the amino acids can be classified into four groups [2] where they are referred to with their three letters abbreviations such as ALA for alanine.

- **Nonpolar side chains** (hydrophobic = dislike water) usually contain bulky hydrocarbon groups such as ALA (alanine), ILE, LEU, MET, PHE, PRO, TRP, VAL. They tend to stick together in the presence of water to minimize water contacts.

- **Polar and uncharged side chains** (like water) usually contain an oxygen atom (or only -H) such as GLY, SER, THR, ASN, GLN, TYR

- **Ionizable side chains** (charged at certain conditions) contain acidic or basic groups that can obtain a negative (GLU and ASP) or positive charge (HIS, ARG, LYS). Their charges are strictly dependent on the environmental pH and the salt concentrations as well as nearby charged groups. The charge of a protein is therefore dependent on both the number of the ionizable groups and the environmental conditions.

- **Special side chains** includes PRO and CYS residues. The former forms kinks in the protein chains and usually referred to as a folding hindering residue. The latter provides covalent linkage through disulfide bonds, within and between protein chains.

Proteins can roughly be divided into two categories according to their structures; globular proteins and intrinsically disordered proteins (IDPs).

[2] P. Raven and G. Johnson, in *Biology*, edited by P. Reidy (Mc Graw Hill, Boston, 2002), chap. The Chemical Building Blocks of Life, 6th edn.

1.1.2 Globular Proteins

Globular proteins are folded proteins whose chains are tightly packed into a complex 3D shape. In the late 1950s, the first 3D structure of a globular protein was solved by Perutz [3] and Kendrew [4] who showed that each atom in a protein has a unique spatial location. Nowadays, with the help of X-ray crystallography and NMR spectroscopy both crystal and solution structures of hundreds of thousands proteins are readily available in protein databanks. The accumulated knowledge on protein structures has indicated that proteins consist of four structural levels:

Primary Structure

The sequence of amino acids in a protein chain is called the primary structure. Each amino acid sequence creats a unique *personality* due to the embedded features of amino acid side chains. In the early 1950s, Sanger's works on insulin protein showed that the residue order in a protein chain is protein specific and that each insulin has the same sequence [5]. The primary sequence of a protein starts with an amino group of the first residue and ends with the carboxyl of the last residue which are referred as N-terminal and C-terminal of the protein, respectively – see Figure 1.1. Note that N and C-terminals are also ionizable groups and contribute to the charge of a protein. A protein chain can contain up to thousands of amino acids and combination of 22 amino acid kinds provides a tremendous sequence diversity.

Secondary Structures

The secondary structures are regular, long periodic shapes formed by the local organization of a protein chain into a coil- or a sheet-like structure. They are stabilized by hydrogen bonds (H-bond) between the $-NH_2$ and $-COOH$ groups of the residues that lie in the protein backbone.

- *β-sheets* are formed by aligning distant parts of a protein chain parallel (or antiparallel) into a sheet like structure. Each distant part of the protein chain in a β-sheet are called β-strands that are constraint into an

[3] M. F. Perutz, Acta Crystallogra. **9**, 867 (Nov 1956).

[4] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, Nature **181**, 662 (Mar. 1958).

[5] F. Sanger and H. Tuppy, Biochem. J. **49**, 463 (1951).

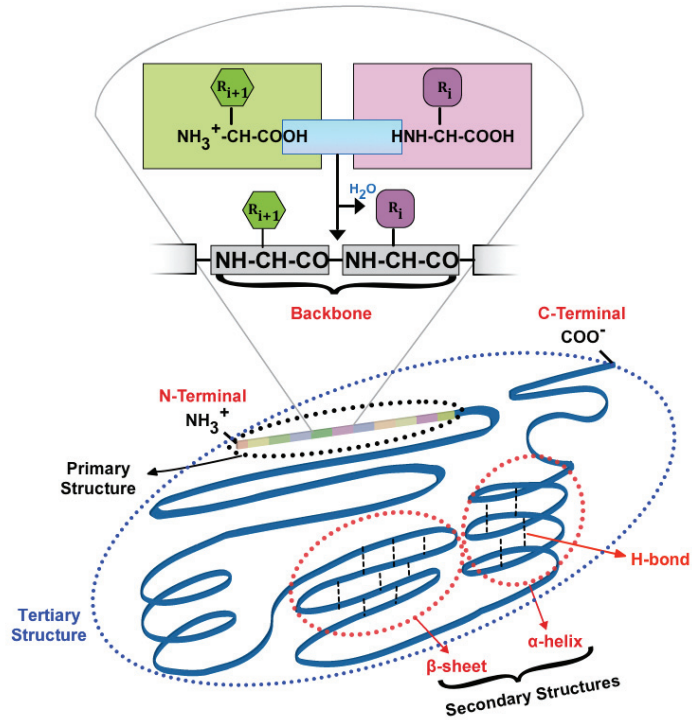


Figure 1.1: Structural levels of globular proteins (© Korhan Şabanoğlu).

extended, rod-like conformation. The sheet is reinforced by the H-bonds between the β-strands.

- *Helices* are spiral structures of the chain that are stabilized by the H-bonds between the residues i+4→i, i+3→i, and i+5→i, called α-helices (most common one), 3_{10}-helices and π-helices, respectively.

In a protein chain, these two main secondary structures are connected via other minor structures such as β-turns, U-turns and random coils. Amino acid sequences may show a tendency for a certain secondary structure which enables the knowledge based secondary structure predictions [6].

Motifs are the combinations of the secondary structures, observed in many proteins to accomplish a specific function. For instance, the helix-turn-helix motif, consisting of two helix bundles connected via a short amino acid sequence,

[6] D. Frishman and P. Argos, Proteins: Struct., Funct., Bioinf. **23**, 566 (1995).

is found in most of the DNA binding proteins to facilitate DNA attachment [7].

Tertiary Structure

A final spatial organization of the secondary structures and the motifs via side chain interactions defines the positions of each atom in space and forms a well-defined 3D shape, called tertiary structure. In this final folded state, hydrophobic residues are usually packed tightly in the interior part of the protein while the charged residues are restricted to the surface (unless required for a special functioning) resulting in a fairly rigid structure. The folded shapes of globular proteins are strictly dependent on the environmental conditions and are achieved in a narrow range of conditions. Even a small deviation from that range can cause protein unfolding (denaturation) resulting in a disruption of their functionality [1,2].

Note that protein functionality is mainly mediated by the surface of the proteins and therefore, the folded 3D structure acts as a skeleton to construct and reinforce the necessary surface for the protein activity.

Domains are the compact subglobular parts of a single large protein. Each of them behaves like a small globular protein and folds into their own 3D shapes independently. They often combine different functions into a single protein, such as one domain of an enzyme accommodates a co-factor and another binds to the substrate [2].

Quaternary Structure

Often complex machines need more than one protein chain to perform their functions. A few separate chains therefore, assemble into one single functioning unit where each chain is called a subunit. Hemoglobin is an example of such a complex protein which consists of four subunits: two α-chains and two β-chains – see Figure 1.2.

[7] I. B. Dodd and J. B. Egan, Nucleic Acids Res. **18**, 5019 (Sep. 1990).

[1] A. Finkelstein and O. Ptitsyn, in *Protein Physics: A Course of Lectures* (Acedemic Press, San Diago, 2002), chap. Introduction.

[2] P. Raven and G. Johnson, in *Biology*, edited by P. Reidy (Mc Graw Hill, Boston, 2002), chap. The Chemical Building Blocks of Life, 6th edn.

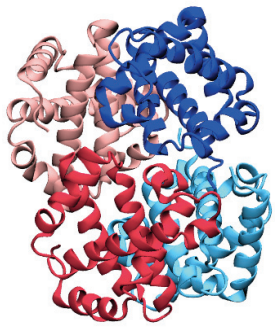


Figure 1.2: The tertiary structures of Hemoglobin subunits consisting of five helix bundles. The quaternary structure of Hemoglobin contains four subunits where two α-chains are depicted in blue and cyan and two β-chains are in red and purple.

1.1.3 Intrinsically Disordered Proteins

A well-defined shape is not always required for an active protein. There are intrinsically disordered and fully functional proteins which often consist of short peptide chains with only a few number of hydrophobic residues and/or a high content of kink-forming PRO residues [8,9]. They might however obtain a secondary and tertiary structure upon binding to their partners. Histatin5, for example, a histidine rich antibacterial salivary protein lacks structure in an aqueous solution but adapts a helical conformation in bacterial membranes [10].

1.2 What is Happening in Protein Solutions?

1.2.1 Protein Folding

Anfinsen's milestone experiment showed that an isolated, denatured (unfolded) ribonuclease A (an enzyme) is able to spontaneously fold into its well defined native structure and regain its enzymatic activity [11]. This proved that the

[8] V. N. Uversky, J. R. Gillespie, and A. L. Fink, Proteins **41**, 415 (Nov. 2000).

[9] A. K. Dunker, C. J. Oldfield, J. Meng, P. Romero, J. Y. Yang, J. W. Chen, V. Vacic, Z. Obradovic, and V. N. Uversky, BMC Genomics **9 Suppl 2**, S1 (Jan. 2008).

[10] P. A. Raj, M. Edgerton, and M. J. Levine, J. Biol. Chem. **265**, 3898 (Mar. 1990).

[11] C. B. Anfinsen, Science (New York, N.Y.) **181**, 223 (Jul. 1973).

unique 3D structure of a protein, which is required for the biological activity, is encoded in the primary sequence.

Protein folding is mainly guided by hydrophobic residues which tend to stick together and form the interior part of the proteins to avoid water contacts. Folding progresses in a stepwise manner, starting with a fast formation of secondary structures, followed by a slow organization into their unique 3D coordinates through the formation of intermediate states. In the cellular environment, a soup of proteins and molecules, these intermediate states are more prone to adhesion of unwanted molecules due to the exposure of sticky hydrophobic residues. This can lead to misfolding and jamming of the newly produced protein chains resulting in detrimental complications such as Cystic Fibrosis [2]. Therefore, protein folding in cells is sometimes assistant by proteins called Chaperones that protect unfolded parts from the unwanted adhesions [12,13].

Box of Imagination

The protein folding is an amazing process that is way beyond our technologic achievements in assembling machines. Imagine that we have this folding technology in car factories. This would imply that to build a car, we just need to attach all the car parts into a chain which will spontaneously form the car by folding all the parts into their correct positions when the right environmental conditions are provided.

1.2.2 Protein Binding

Protein function is coupled with the binding of proteins to their binding partners such as to other proteins [14] or small ligand molecules [15]. The binding processes occur through a molecular recognition of binding partners by the active surfaces of the proteins. Since ligands are small molecules, the active surface, where they bound to, is usually highly specific and consists of only a couple of residues. It is, therefore common to model a ligand binding process with mul-

[2] P. Raven and G. Johnson, in *Biology*, edited by P. Reidy (Mc Graw Hill, Boston, 2002), chap. The Chemical Building Blocks of Life, 6th edn.

[12] F. U. Hartl, Nature **381**, 571 (Jun. 1996).

[13] G. E. Karagöz, A. M. S. Duarte, E. Akoury, H. Ippel, J. Biernat, T. Morán Luengo, M. Radli, T. Didenko, B. a. Nordhues, D. B. Veprintsev, C. a. Dickey, E. Mandelkow, M. Zweckstetter, R. Boelens, T. Madl, and S. G. D. Rüdiger, Cell **156**, 963 (Feb. 2014).

[14] P. L. Kastritis and A. M. J. J. Bonvin, J. R. Soc. Interface **10**, 1 (2013).

[15] D. L. Mobley and K. A. Dill, Structure **17**, 489 (Apr. 2009).

tiple levels of details [16] by representing only the active surface of the protein and the ligand molecule in detailed.

However, the protein-protein binding can occur via a large binding surface which can easily contain up to a few dozen residues. Therefore, the nature of the *whole* protein surface becomes important in the binding process. The distribution of the charged and hydrophobic residues on the surface determines the surface characteristics and specific interactions with the binding partners. An uneven distribution of these residues results in orientational dependent, anisotropic interactions. For instance, segregation of sticky hydrophobic residues at one side of the protein can lead to binding via the hydrophobic surfaces to minimize their water contact. The effect of active surface nature can easily be seen in the binding of the similarly sized proteins, α-lactalbumin and lysozyme, to a horse shoe shaped protein CD14, (Figure 1.3). Although the shape and the size of both proteins allow binding to the gap of CD14, lysozyme prefers to bind at the side surface of CD14 due to the favorable surface interactions while α-lactalbumin binds to the gap of CD14.

1.2.3 Self-Assembly

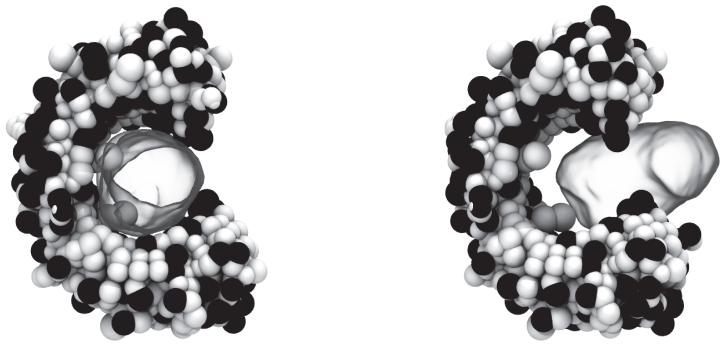
Certain functions in nature may require superior protein machines which can be built by an assembly of many proteins into one supramolecular structure. The assembly of the subunits occurs *spontaneously* without any external aid but their own binding sites and thus, it is called *self-assembly*. For instance, in a cell, a globular protein called *actin* self-assembles into long microfibers to build the cell infrastructure and to provide paths for the molecular transportation [17]. Silk proteins self-assemble into well structured and extremely stable fibers that make up the spider webs. [18]. Proteins do not only assemble into fibrillar structures. For instance, caseins, a family of milk proteins, form oblate aggregates which serve as both Ca^{2+} ion carriages and consistency regulators in dairy products [19].

[16] S. C. L. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, Annu. Rev. Phys. Chem. **62**, 41 (Jan. 2011).

[17] H. Lodish, A. Berk, and S. Zipursky, in *Molecular Cell Biology*, edited by W. Freeman (Mc Graw Hill, New York, 2000), chap. The Dynamics of Actin Self Assembly, 4th edn.

[18] F. Hagn, L. Eisoldt, J. G. Hardy, C. Vendrely, M. Coles, T. Scheibel, and H. Kessler, Nature **465**, 239 (May 2010).

[19] S. Ossowski, A. Jackson, M. Obiols-Rabasa, C. Holt, S. Lenton, L. Porcar, M. Paulsson, and T. Nylander, Langmuir **28**, 13577 (Sep. 2012).



(a) α-lactalbumin binding to CD14 (b) lysozyme binding to CD14

Figure 1.3: Preferred binding locations of α-lactalbumin and lysozyme on CD14, depicted by silver transparent bubble. The size and the active surface coded interactions of α-lactalbumin allow binding to the gap of the human milk protein CD14. However, similarly sized lysozyme favors attachment to the side surface instead of the gap although it could fit into the gap. (MC simulation study, unpublished)

Box of Imagination

Let's imagine that we are using this self-assembly technology in our car factory. This time, we don't even need to put the car parts into a chain. It is enough that we have all the parts in a container and wait for the car to construct itself!

Self-assembly of proteins does not always lead to useful tools. Various diseases are associated with undesired aggregations of proteins. For instance, it is thought that the accumulation of insoluble aggregates of unfolded β-amyloid proteins in brain tissues is related to a brain dementia called, "Alzheimer's disease" [20]. Moreover, the unwanted aggregation of hemoglobin proteins dramatically reduces oxygen carriage capacity of red blood cells and results in "Sickle Cell Anemia" [21].

[20] F. G. D. Felice, D. Wu, M. P. Lambert, S. J. Fernandez, P. T. Velasco, P. N. Lacor, E. H. Bigio, J. Jerecic, P. J. Acton, P. J. Shughrue, E. Chen-dodson, G. G. Kinney, and W. L. Klein, Neurobiol Aging. **29**, 1334 (2008).

[21] M. J. Stuart and R. L. Nagel, J.-Lancet **364**, 1343 (2004).

1.2.4 Protein Adsorption

Protein adsorption on bacterial membranes seems to be related to the function of antimicrobial proteins. The first step of any antimicrobial activity relies on the adsorption of proteins on negatively charged microbial membranes. Therefore, the protein adsorption on charged solid surfaces has been used for establishing an analogy to the antimicrobial activity. The killing activity may sometimes require further processes such as positioning of the adsorbed proteins in the membrane to disrupt cell integrity [22] or transportation through the membrane to target a specific inner organelle [23]. It is therefore not always adequate to directly relate the antimicrobial activity with the protein adsorption. However, adsorption behavior can still serve as a qualitative indicator of the activity since without adsorption it is *impossible* to perform the killing activity.

Similar to self-assembly, protein adsorption to surfaces is not always desired. It can cause problems especially for medical implants where protein accumulation might result in inflammations. Moreover, in the food industry, adsorbed proteins may promote bacterial growth and biofilm formation.

In order to get a better understanding of the complex protein behaviors and the possible implications of their malfunctioning, we need to investigate *"What kind of interactions present between the proteins?"*. This leads us to the second chapter.

[22] Y. Shai, Peptide Science **66**, 236 (2002).

[23] J. R. Luque-Ortega, W. van't Hof, E. C. I. Veerman, J. M. Saugar, and L. Rivas, FASEB J. **22**, 1817 (Jun. 2008).

CHAPTER 2

Interactions in the Protein World

As we have seen in the previous chapter, amino acids are linked through strong and stable chemical bonds and the primary sequence of a protein therefore remains intact unless targeted by enzymes. In addition, the thiol group (–SH) in cysteine residues can form strong disulfide bonds through an oxidizing process. This bond usually bridges two cysteine residues that are far apart in the primary sequence and directs the peptide chains in their folding paths.

Apart from these chemical bonds, non-bonded interactions also exist between amino acids which are the focus of this chapter. These interactions determine the favorable amino acid neighbors in the 3D structure as well as in their binding partners. The non-bonded interactions are weaker than chemical bonds and are sensitive to environmental conditions. They can therefore be manipulated by altering the conditions in the need of conformational changes and detachment from a binding partner.

2.1 Electrostatic Interactions

Electrostatic interactions are present between charged species as well as between polar and polarizable molecules. They cover long distances, making them effective even when proteins are far apart. They can be attractive or repulsive, depending on the net charges of the interacting proteins as well as the distribution of the charged residues.

Ion-ion Interactions

The electrostatic interactions of two charges *in vacuum* are described by the following Coulomb's law,

$$\mathcal{U}_{\text{vac}} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r} \tag{2.1}$$

where $\mathcal{U}$ is the interaction energy, ϵ_0 is the dielectric permittivity of vacuum, r is the distance between the charges, q is the charge. By convention, negative energy refers to an attractive interaction as between two oppositely charged particles.

When the charges interact in a medium for example, water and ethanol, the polarity of the surrounding molecules affects the interaction strength and reduces (screens) the energy proportional to the dielectric constant of the medium ($\epsilon_r = 80$ and 40 for water and ethanol respectively, at room temperature).

$$\begin{aligned} \mathcal{U}_{\text{medium}} &= \frac{1}{4\pi\epsilon_0}\frac{1}{\epsilon_r}\frac{q_1 q_2}{r} \\ \beta\mathcal{U}_{\text{medium}} &= \Big(\frac{e^2}{4\pi\epsilon_0}\frac{1}{\epsilon_r}\frac{1}{kT}\Big)\frac{z_1 z_2}{r} = l_B\frac{z_1 z_2}{r} \end{aligned} \tag{2.2}$$

$\mathcal{U}$ is the interaction energy (strictly speaking, interaction *free* energy), e is the electron elementary charge, z is the valency of the charges, $q = ze$ is the charge, l_B is the Bjerrum length, k is the Boltzmann constant, T is temperature and β equals to $1/kT$. The $1/r$ dependence of ion-ion interaction energy shows a slow decay with distance, providing a long range of action. For instance, in water and at room temperature, the electrostatic energy of two oppositely charged ions with a charge of $4e$ starts to exceed the thermal energy kT when their separation is around 100 Å. This means that they start to feel the electrostatic attraction even when they are 100 Å apart – see Figure 2.1.

When the medium contains salt species, the interaction of the two charges are further screened due to the interruption of their electric field by the ion cloud around them. Debye-Hückel theory accounts for the salt screening effect by introducing an exponential decay as a function of salt concentration.

$$\begin{aligned} \beta\mathcal{U}_{\text{salt solution}} &= l_B\frac{z_1 z_2}{r}\exp(-\kappa r) \\ \kappa^2 &= \frac{1}{\epsilon_0\epsilon_r kT}\sum q_i^2 c_i \end{aligned} \tag{2.3}$$

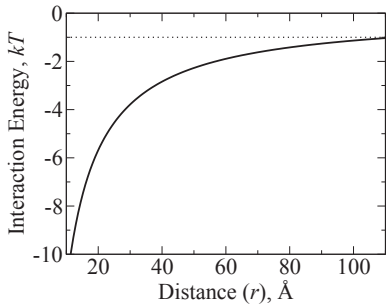


Figure 2.1: Interaction energy of two oppositely charged ions with a charge of $\pm 4e$ in water at room temperature.

where q_i is the charge of ionic specie i, and c_i is the concentration. Debye-Hückel theory is approximate and works best for monovalent ions due to weak ion correlation effects. It also overestimates the screening when the charges are at close contact due to continuum representation of salt ions [24].

Ion-dipole Interactions

The polarity of molecules originates from an uneven charge distribution which creates islands of positive and negative charges on their surfaces. For instance, the oxygen atom of a water molecule attracts electrons of hydrogen atoms and obtains a partially negative charge while leaving two partially positive hydrogen atoms. This charge separation creates a strong dipole on the water molecule which can respond to an external electric field, such as from an ion, by orienting itself according to the electric field – see Figure 2.2.
The interaction energy between a fixed point dipole (μ) at an angle θ and an ion (q) is given by

$$\mathcal{U}_{ion-dipole} = -\frac{1}{4\pi\epsilon_0}\frac{1}{\epsilon_r}\frac{q\mu\cos\theta}{r^2}. \tag{2.4}$$

where θ is the angle between the dipole and the vector connecting the center of dipole to the center of ion. $\mu = \sum_i \delta_i d_i$, δ is the partial charge and d is the distance from the origin. The $\cos\theta$ dependence of the energy shows the directional nature of dipolar interactions. When a dipole is directed towards

[24] J. N. Israelachvili, *Intermolecular and Surface Forces* (Acedemic Press, San Diego, 2011), 3rd edn.

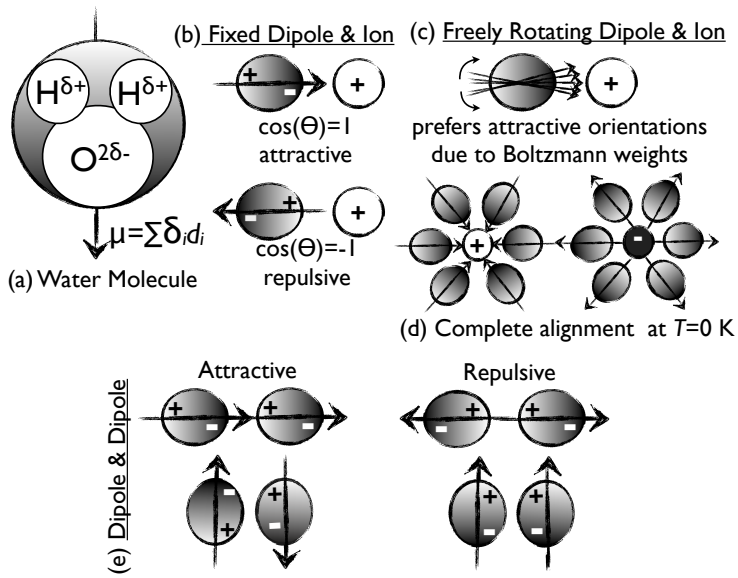


Figure 2.2: (a) Polarity of water molecules and (b) $\cos\theta$ dependence of their interactions with a positive ion. (c) Thermally averaged ion-freely rotating dipole interactions where dipoles prefer attractive orientations over repulsive ones due to Boltzmann weights. (d) Fully aligned dipoles around a positive and a negative ion at zero K which minimize the energy. (e) Orientational dependence of dipole-dipole interactions.

an ion, this will maximize attractive interactions ($\theta = 0°$) while the opposite direction ($\theta = 180°$) will lead to repulsive interactions – see Figure 2.2. The relative probability of observing these two orientations is given by the ratio of their Boltzmann weights, $\frac{w(\theta=0)}{w(\theta=180)} = \frac{e^{-\beta \mathcal{U}_{ion-dipole}(\theta=0)}}{e^{-\beta \mathcal{U}_{ion-dipole}(\theta=180)}}$. Therefore, a freely rotating dipole, like a water molecule in the liquid state, will always favor attractive orientations over repulsive ones – see Figure 2.2. Since the Boltzmann weights are temperature dependent ($\beta = 1/kT$) this preference is called the thermal averaging of ion-dipole interactions.

The dielectric screening of polar media comes from the ability of polar molecules to align themselves according to an external electric field. The alignment of the medium molecules reduces the electric field, and thus screens electrostatic interactions. Since the dipole moment of water is high, they show a strong response resulting in a high dielectric constant. Due to the Boltzmann weighing,

enhanced at lower temperatures when the thermal energy (kT) is low which manifest itself as a higher dielectric constant.

As seen in Equation 2.4, ion-dipole interactions have a $1/r^2$ distance dependence, shorter-ranged than ion-ion interactions.

Dipole-dipole interactions

The interaction energy between two point dipoles is given by

$$\mathcal{U}_{dip-dip} = \frac{1}{4\pi\epsilon_0}\frac{1}{\epsilon_r}\left(\frac{\boldsymbol{\mu_1}\boldsymbol{\mu_2}}{r^3} - 3\frac{\boldsymbol{\mu_1 r}\cdot\boldsymbol{\mu_2 r}}{r^5}\right) \tag{2.5}$$

where $\boldsymbol{\mu}$ is the dipole moment vector, $\boldsymbol{r}$ is the vector connecting the centers of the dipoles. $\boldsymbol{\mu\mu}$ and $\boldsymbol{\mu r}$ are the scalar product of the two vectors which is a function of the angle between them. Therefore, the dipole-dipole interactions also show orientational dependence. The most attractive interaction is achieved when the dipoles are aligned antiparallel leading to the shortest possible distance between the oppositely charged ends – see Figure 2.2. As seen, the energy has a distance dependence of $1/r^3$ which is shorter-ranged than ion-dipole interactions ($1/r^2$). The thermal averaging works exactly as for the ion-dipole interactions and favors attractive orientations over repulsive ones when the dipoles are free to rotate. Due to this Boltzmann weighing, the water molecules adopt certain orientations to maximize the attraction resulting in a special structuring.

Protonation of a Protein

The net charge of a protein is determined by the number of ionizable groups such as titratable (charged) side chains, N- and C-terminals and surface bound ions together with solution conditions.

An ionizable amino acid obtains charge via proton uptake (protonated state) or release (deprotonated state) which is controlled by its electrostatic surrounding and pH of the solution. The equilibrium between the protonated and deprotonated state ($\mathrm{HA} \rightleftharpoons \mathrm{H^+} + \mathrm{A^-}$) is defined by the equilibrium constant, $K_a = \frac{[H^+][A^-]}{[HA]}$ where [HA] is the concentration of the protonated state and [A] is the concentration of the deprotonated state. (Activity coefficients are equal to unity.) The pH is related to the logarithm of the equilibrium constant, called intrinsic acid dissociation constant $\mathrm{p}K_a$, as $\mathrm{pH} = \mathrm{p}K_a + \log\frac{[A^{-/0}]}{[HA^{0/+}]}$. The deprotonated state bears $-1e$ charge for ASP, GLU, TYR and CTR and is neutral for HIS, ARG, LYS, NTR.

pK_a measures the pH where the charged and the neutral states (protonated/deprotonated) of an ***isolated*** amino acid are equally probable. In other words, when pH = pK_a the average charge of an amino acid is $\pm 0.5e$ without considering any interactions ("+" applies for the positively charged and "−" for the negatively charged amino acids). Using the relation between pH and pK_a, the average charge of an isolated amino acid at a given pH can be calculated as,

$$\begin{aligned} \langle q^- \rangle &= \frac{-1}{1+10^{-(\mathrm{p}H-\mathrm{p}K_a)}} \quad \text{(for ASP, GLU, TYR, CTR)} \\ \langle q^+ \rangle &= \frac{+1}{1+10^{+(\mathrm{p}H-\mathrm{p}K_a)}} \quad \text{(for HIS, ARG, LYS, NTR).} \end{aligned} \tag{2.6}$$

Since pK_a reflects the charge of an ***isolated*** amino acid in aqueous solution, it does not include effect of electrostatic environment such as nearby charges or change of dielectric medium. However, an amino acid on a protein can have many charged neighbors and experience a different dielectric medium depending on where it is located such as on the surface or in the interior. Therefore, the apparent (effective) pK_a^* of an amino acid on a protein can be different than its intrinsic pK_a due to the interactions with the neighboring residues. For instance, if a positive residue is surrounded by many other positive charges, it will be costly to bear a positive charge and thus, its protonation will require higher proton concentration. This will shift its pK_a value to a lower pH, resulting in p$K_a^* <$ pK_a. Note that the Hofmeister ion binding on protein surfaces can also be treated as the proton binding equilibrium. The contribution of the bound ions to protein charge can be calculated with the same procedure once the equilibrium constant of the binding process is known.

The effect of electrostatic environment on a charged residue can be captured by considering electrostatic interactions between all charged amino acids of a protein and applying a self energy penalty for bearing a buried charge in the non-polar protein interior [24]. Although the former is easy to account for, the latter is challenging due to the difficulties in determining the dielectric constant and boundary of the protein interior. However, most charged amino acids are situated on the surface and thus, dielectric medium effects are usually insignificant compared to neighboring charge interactions – see Table 5.1 for the number of buried charges in studied proteins.

The pK_a^* also reflects the pH value at which the charge fluctuations of a residue is maximized due to the equal probability of the both protonation states

[24] J. N. Israelachvili, *Intermolecular and Surface Forces* (Acedemic Press, San Diego, 2011), 3rd edn.

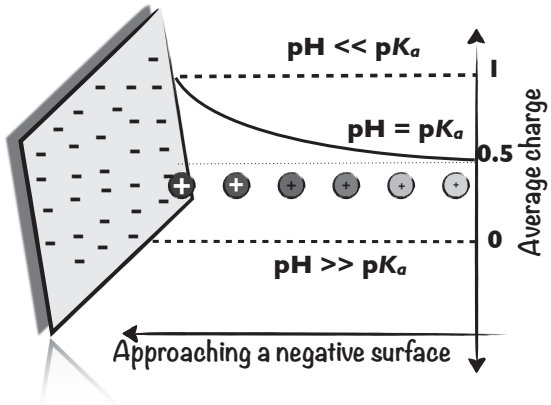


Figure 2.3: The charge regulation of a positive residue while approaching a negative surface at various pH.

[25]. This means that when pH≈pK_a^*, a residue can easily regulate its charge as a response to an external electric field by adopting the more favorable protonation state. For instance, when pH≈pK_a^*, a positively charged residue will favor the protonated state more and obtain higher average charge while approaching a negatively charged surface – see Figure 2.3. However, this charge regulation ability diminishes rapidly when the environmental pH deviates from the pK_a^* of the residue.

The distribution of charged amino acids determines the polarity of proteins. A high degree of anisotropy may result in a higher dipole moment and thus, stronger dipolar interactions. Note that proteins can bear both a net charge, a dipole moment and higher order multipoles at the same time resulting in non-zero ion-ion, ion-dipole, dipole-dipole and higher order multipole interactions. Once the average charges $\langle q_i \rangle$ of the amino acids are known, the molecular dipole moment of a protein can be calculated as $\boldsymbol{\mu} = \langle \sum_i q_i \mathbf{r}_i \rangle$ where $\mathbf{r}_i$ is the vector connecting the residue i to the protein center of mass.

2.1.1 Hydrogen Bonds

A hydrogen bond (H-bond) is present between an electronegative atom such as oxygen (O) and nitrogen (N) and a hydrogen atom bound to an electronega-

[25] M. Lund and B. Jönsson, Biochemistry **44**, 5722 (Apr. 2005).

tive atom (–HO, –HN). It has an electrostatic nature where the partial negative charge on the electronegative atom interacts with the partial positive charge on the H atom ($OH^+ \cdots O^-$).

H-bonds are also formed between -H and -OH groups of water molecules. In ice, the molecules are forced into a tetrahedral orientation to maximize their H-bonds resulting in four H-bonds per molecule. In liquid water, this number reduces to three due to the higher thermal motion of the water molecules [1].

In proteins, the amino groups (–NH) of the backbone are responsible for the construction of the H-bonds. Since these bonds are strictly dependent on the orientation of NH groups, they force the backbone to align in a particular way leading to secondary structure. In water, H-bonds can also be formed between the protein backbone and the water molecules. However, this is entropically opposed due to restricted configurations of the protein-attached water molecules.

2.1.2 Salt Bridges

Whenever two oppositely charged amino acid side chains are close enough in a protein or on a binding surface, they interact via so called salt bridges. It contributes to the stability of protein structures and assemblies by providing attractive electrostatic interactions. It usually involves a hydrogen bond between charged residues as well.

2.2 Van der Waals Interactions

Van der Waals (VDW) interactions consist of three components; Keesom, Debye and London dispersion interactions which are always attractive and have a $1/r^6$ distance dependence. Although they are short-ranged attractions, they exist between all atoms and their collective effect in macromolecules can enhance both the strength and the range of macro-molecular interactions.
Keesom interactions are the angle-averaged dipole-dipole interactions, present between two polar molecules with permanent dipoles.
Debye interactions are the angle-averaged dipole-induced dipole interactions, present between a non-polar but polarizable molecule and a permanent dipole.
London dispersion interactions are the induced dipole-induced dipole interactions and exist between all atoms regardless of their charges. They originate from the synchronization of instantaneous dipoles on atoms created by their

[1] A. Finkelstein and O. Ptitsyn, in *Protein Physics: A Course of Lectures* (Acedemic Press, San Diago, 2002), chap. Introduction.

asymmetrical distribution of electrons around nucleus. These dipoles and thus the dispersion interactions are of quantum mechanical nature.

VDW interactions are usually modeled with a $-C/r^6$ potential. C is dependent on the VDW radius of the atoms and their interaction strength – see Equation 2.7.

2.3 Exchange Repulsion

Exchange repulsion is a short-ranged and steep repulsion which determines the closest distance two atoms can approach. It originates from quantum mechanical effects and usually occurs at 2-3 Å separation that defines the impenetrable volumes of atoms. [1]. It can be modeled with a hard sphere ($U = \infty$ when $r \leqslant \sigma_{ij}$) or an r^{-12} potential where the latter provides a slightly softer repulsion.

The Lennard-Jones potential, is one of the most commonly used potentials for modeling VDW attraction and exchange repulsion due to its computational convenience ($r^{-12} = (r^{-6})^2$)[24].

$$\mathcal{U}_{LJ} = 4\epsilon\Big(\underbrace{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}}_{\text{ER}} - \underbrace{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}}_{\text{VDW}}\Big) \tag{2.7}$$

where ϵ is the strength, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ the average VDW radius of i and j, and r_{ij} is the distance between them.

2.4 Water Related Interactions

Water is a spectacular molecule with an unusually high melting (273 K) and boiling point (373 K) compared to other similarly sized molecules such as O_2 which melts at 54 K and boils 90 K. This is due to the strong dipolar interactions which need to be broken during evaporation/melting. The high energy cost of breaking these interactions results in elevated melting and boiling points. This polarity of water molecules provides as an excellent solvent medium for polar molecules but not for non-polar molecules resulting in a phenomenon called hydrophobic interactions.

[24] J. N. Israelachvili, *Intermolecular and Surface Forces* (Acedemic Press, San Diego, 2011), 3rd edn.

2.4.1 Hydrophobic Interactions

Since water molecules strongly interact via polar interactions, they prefer minimum contact with non-polar (hydrophobic) molecules that disrupt the interaction network. By clustering these molecules, water reduces the exposed hydrophobic surface area. This clustering results in an effective attraction between hydrophobic molecules that does not exist in a non-polar solvent or vacuum.

The hydrophobic attraction is the main driving force for the folding of non-polar amino acids towards the interior part of the globular proteins. Although most of the amino acids remain inside the protein, 15% (on average) solvent exposed hydrophobic surface is still commonly observed – see Table 5.2. As for charged residues, an uneven distribution of hydrophobic residues on the surface also leads to anisotropic interactions and may force the protein into specific orientations – see Chapter 5.

Hydrophobic interactions depend on solution conditions such as temperature, salt type and concentration. It has been shown that the temperature dependence is quite different for small and big molecules [26]. Around small molecules, the H-bonding network of water is maintained by restricting their orientations and thus by sacrificing the number of configurations. Therefore, the hydrophobic interactions for small molecules have an entropic nature, directly proportional to temperature [27]. However, around large hydrophobic aggregates, the H-bond network cannot be maintained any longer and results in an energy driven hydrophobic interactions, oppositely proportional to temperature [28].

2.4.2 Ion Specific Effects

Water also interacts differently with different types of ions, contributing to the so called Hofmeister effect. Small ions with a high charge density strongly attract water leading to a high degree of alignment. Since these well-hydrated ions are surrounded by many water molecules they tend to avoid non-polar surfaces resulting in enhanced hydrophobic attractions. This is called "salting out" due to clustering of proteins [29]. However, big ions with low surface charge den-

[26] D. Chandler, Nature **417**, 491 (2002).

[27] M. S. Moghaddam, S. Shimizu, and H. S. Chan, J. Am. Chem. Soc. **127**, 303 (Jan. 2005).

[28] D. Chandler, Nature **437**, 640 (Sep. 2005).

[29] M. Boström, F. W. Tavares, S. Finet, F. Skouri-Panet, a. Tardieu, and B. W. Ninham, Biophys. Chem. **117**, 217 (Oct. 2005).

sity are poorly hydrated due to the insufficient interaction energy with dipoles. These anions tend to behave like hydrophobic species and bind to non-polar side chains and backbone of proteins. This ion binding may stabilize proteins in solution by providing additional charges and blocking hydrophobic interactions [30].

2.5 Entropic Repulsion

Especially in the surface adsorption of a flexible protein, being close to a surface will substantially restrict its conformational freedom. The loss of accessible conformational states significantly diminishes chain entropy and results in an *effective* repulsion between the surface and the protein – see Section Statistical Mechanics. Note that there is no actual repulsion between the chain and the surface like in electrostatics. Hence, if we freeze the chain at one configuration, this entropic repulsion will be suppressed – see Figure 2.4.

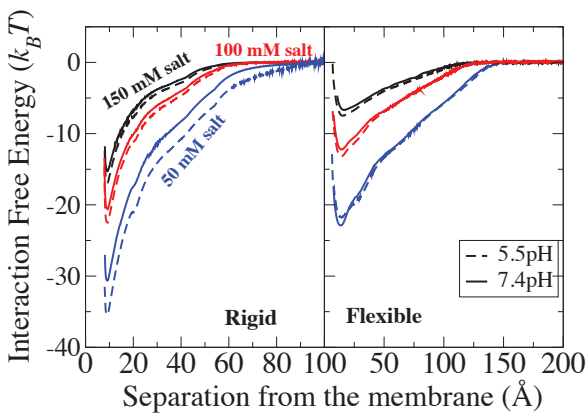


Figure 2.4: Interaction free energy of the positively charged Midkine with a negatively charged surface where Midkine is modeled as a rigid body (left) and a flexible chain (right). Since the rigid body does not have conformational freedom, being close to the surface is not as costly as for the flexible chain. (Unpublished data)

[30] E. A. Algaer and N. F. A. van der Vegt, J. Phys. Chem. B **115**, 13781 (Oct. 2011).

Box of Imagination

Interactions in a crowded bus

Imagine that you are standing in a crowded bus, next to a person wearing a heavy perfume and carrying many shopping bags.

- According to Pauli exclusion principle in buses, you cannot invade to the volume occupied by that person. Therefore, this part of the bus will be inaccessible.
- How well you synchronized your movements in case of sudden accelerations and breaks will provide a *van der Bus* attraction allowing closer distances without bumping into each other.
- Since you don't like the perfume, this will lead to an additional perfume-phobic interactions. You would stand even further away to avoid the perfume. Since the other perfume wearing people in the bus do not care about the heavy smell, they do not mind standing close to each other. This will result in segregation of perfume wearing people and cause an effective attraction between them.
- Approaching the shopping bags will reduce your balance and accessible standing positions and thus, it will result in an entropy loss. To minimize this loss, you would avoid the shopping bags by keeping your distance which leads to an effective repulsion.
- If the person is one of your close friends, your friendship will provide a strong electrostatic attraction. Therefore, you would prefer standing close despite the heavy perfume and shopping bags. In addition, since this attraction is long ranged, you would feel it as soon as you get on the bus and go to your friend directly.

All of these interactions will determine your relation with the perfume wearing person and define the preferred distance between you.

CHAPTER 3

Available Tools

3.1 Statistical Mechanics and Thermodynamics

Thermodynamics deal with equilibrium properties of macroscopic systems which represent averages of enormously many molecules (in the order of 10^{20}). Although it provides insight to experimental observations it is unable to explain molecular details behind them. Statistical mechanics, however, helps us to calculate thermodynamic properties from molecular ones and thus, brings molecular interpretations to experiments [31].

An experimental measurement is performed by taking a time average of an observable (M) such as pressure, energy...etc. If we want to calculate the pressure in a chamber using molecular properties, we first need to consider the tremendous number of molecules colliding with the wall and second, we need to do that over a *sufficiently long time period* to smear out the molecular fluctuations in the system. Gibbs suggested a practical method for solving this problem which constructs the first postulate of statical mechanics.

First Postulate: An *ensemble average* can be used instead of a *long time average* as long as the ensemble replicates thermodynamic properties of the macroscopic system. An ensemble here means a large collection of imaginary replicas of a macroscopic system where each replica can have different molecular details

[31] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover Publication Inc., New York, 1960).

(micro states). This postulate also establishes the basis of Monte Carlo simulations [32].

Box of Imagination

Imagine that you are a photographer working for National Geographic magazine and you want to find favorite places of a jaguar, based on your records and pictures.
In this case, you have two options.

1. You can either choose a dangerous method called *time averaging* where you attach cameras to a couple of jaguars to record where the jaguars are visiting over a long time period. Then you can calculate the probability of finding a jaguar based on the average time they spent in one location, or

2. You can choose an *ensemble averaging* where you place cameras to all possible locations that a jaguar can visit and take a large number of random pictures at each location. You then, can work out the relative probability of finding a jaguar at a given location by counting the number of pictures with a jaguar in it.

Ensembles can be classified according to represented macroscopic systems.

- **Microcanonical (NVU) ensemble**: replicates an isolated system where the number of particles (N), volume (V) and energy (U) are constant.

- **Canonical (NVT) ensemble**: replicates a closed, isothermal system where N, V, and temperature are constant.

- **Isobaric-isothermal (p_0VT) ensemble**: pressure, V, and T are constant.

All ensembles obey the first postulate and their averages are interchangeable. However, it is usually advantageous to use the one which represents the experimental conditions of interest.

In order to calculate an ensemble average, the relative number of occurrences (frequency, weight, probability) of micro states in the ensemble has to be known which is provided by the second postulate of statistical mechanics.

Second Postulate: When the energy of each micro state is equal, there is no preference of one to another. This means that in an NVU ensemble, each micro state is equally probable – *the principle of equal a priori probabilities*.

[32] D. Frenkel and B. Smit, *Understanding Molecular Simulations From Algorithms to Applications* (Acedemic Press, California, 2002).

Once the number of micro states (Ω) is known, the probability (P_i) of observing a state i in the NVU ensemble can be calculated as $P_i = 1/\Omega$. Hence, their relative occurrences (weights) will be equal and one (w_i=1, $\Omega = \sum_i w_i$). Ω_{NVU} is also called the **partition function** of the microcanonical (NVU) ensemble since it shows the partitioning (distribution) of micro states over the ensemble.

Now we have all the necessary information to calculate an ensemble average of an observable $\langle M \rangle$ in an NVU ensemble.

$$\langle M \rangle_{NVU} = \sum_i M_i P_i(NVU) = \sum_i M_i \frac{w_i(NVU)}{\sum_i w_i(NVU)} = \sum_i M_i \frac{1}{\Omega} \quad (3.1)$$

The number of accessible states is related to a thermodynamic property of the macroscopic system called *entropy*, *S*,

$$S = k \ln(\Omega(NVU)) \quad (3.2)$$

where k is Boltzmann's constant. Since there is always one accessible state for a system, the ground state, this equation explains *why the entropy of a system is always a positive number* – the second law of thermodynamics. Moreover, the entropy will be reduced by loss of accessible states which explains the origin of entropic repulsion.

In the **canonical ensemble** (NVT), the replicas are allowed to exchange heat and thus, they no longer have to have the same energy. In this case, the number of micro states (replicas) with the energy, U_i, is determined by the Boltzmann weight of that energy level, $w_i = \exp(-\beta U_i)$ where $\beta = 1/kT$. Thus, the partitioning of micro states over the ensemble will be $Q_{NVT} = \sum_i \exp(-\beta U_i)$.

If we imagine each energy level as a small microcanonical ensemble with Ω_E number of micro states, the canonical partition function can be expressed in terms of the microcanonical partition function as,

$$Q_{NVT} = \sum_U \Omega_{NVU} \exp(-\beta U(N, V)). \quad (3.3)$$

With this knowledge of the canonical partition function, we can calculate the average internal energy of the macroscopic system.

$$\langle U \rangle = \sum_U U P(U) = \sum_i U_i \frac{\exp(-\beta U_i)}{Q_{NVT}} = -\Big(\frac{\partial \ln Q_{NVT}}{\partial \beta}\Big)_{VN} \quad (3.4)$$

This is called *thermal averaging* since the weights of the states are temperature dependent. Using the Gibbs-Helmholtz equation from thermodynamics, we can

extract the relation between the canonical partition function and the Helmholtz free energy of the macroscopic system.

$$A = -kT \ln Q_{NVT} \tag{3.5}$$

From thermodynamics, we know that systems tend to minimize their Helmholtz free energy, equal to $A = U - TS$. This means that, a canonical system will minimize its internal energy and maximize its entropy.

Now, we can go on and define the partition functions of other ensembles with a similar approach. We can represent an **isobaric-isothermal system** by a canonical system with fluctuating volumes which does a volume work against an external pressure, p_0. Then, the partition function of an isobaric-isothermal ensemble can be written as,

$$\Delta_{Np_0T} = \sum_V Q_{NVT} \exp(-\beta p_0 V) \tag{3.6}$$

which is related to the Gibbs free energy,

$$G = -kT \ln \Delta_{Np_0T}. \tag{3.7}$$

3.1.1 Classical Statistical Mechanics

In the high-temperature limit of quantum mechanics, the discrete energy levels of micro states are very close to each other and can thus, be approximated as a continuum. The analogy of the sum in the discrete micro states is then taking an integral over all possible classical states [31].

$$\sum_{\substack{\text{micro}\\ \text{states}}} \exp(-\text{energy}/kT) \to \int_{\substack{\text{classical}\\ \text{states}}} \exp(-\text{energy}/kT) \tag{3.8}$$

The energy of the classical states is given by the system Hamiltonian, $\mathcal{H}$, where the kinetic ($\mathcal{U}_{kin}$) and the potential energy ($\mathcal{U}_{pot}$) are assumed separable.

$$\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N) = \underbrace{\frac{1}{2m}(p_{x_1}^2 + p_{y_1}^2 + p_{z_1}^2 \cdots + p_{z_N}^2)}_{\mathcal{U}_{kin}} + \mathcal{U}_{pot}(x_1, y_1, z_1 \cdots, z_N) \tag{3.9}$$

[31] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover Publication Inc., New York, 1960).

Here, $\mathbf{p}(p_x, p_y, p_z)$ is the momentum of a particle and $\mathbf{r}(x, y, z)$ are the spatial coordinates. Then, the classical canonical partition function can be written as;

$$\mathcal{Q}_{NVT}^{\text{class}} = \frac{1}{N!h^{3N}} \int_V \exp(-\beta \mathcal{U}_{kin}(\mathbf{p}^N)) \mathrm{d}\mathbf{p}^N \int_V \exp(-\beta \mathcal{U}_{pot}(\mathbf{r}^N)) \mathrm{d}\mathbf{r}^N \quad (3.10)$$

where h is Planck's constant. Since the micro states of the particles are no longer distinguishable, the denominator gets the $1/N!$ factor to avoid multiple counting. We can immediately perform the momentum integral since it is independent of the spatial coordinates. Then, equation 3.10 gives

$$\mathcal{Q}_{NVT}^{\text{class}} = \frac{1}{N!\Lambda^{3N}} \underbrace{\int_V \exp(-\beta \mathcal{U}_{pot}(\mathbf{r}^N)) \mathrm{d}\mathbf{r}^N}_{\text{configurational integral}} = \frac{\mathcal{Z}_N}{N!\Lambda^{3N}} \quad (3.11)$$

where $\Lambda = \frac{h}{\sqrt{2\pi m k T}}$ is the de Broglie thermal length and m is the mass of particles and $\mathcal{Z}_N$ is the *configurational integral*. [31] The average of an observable, M, can be calculated as follows,

$$\langle M \rangle_{NVT} = \frac{\int_V M(\mathbf{r}^N) \exp(-\beta \mathcal{U}_{pot}(\mathbf{r}^N)) \mathrm{d}\mathbf{r}^N}{\mathcal{Z}_N} \quad (3.12)$$

Now, the question of calculating the classical partition function is reduced to solving the configurational space integral which is dependent on the interactions between the particles. However, this integral is $3N$ dimensional and it is quite challenging to solve analytically (almost impossible when $N > 3$) Therefore, we need to use numerical solution tools such as Metropolis Monte Carlo simulations.

3.2 Monte Carlo Simulations

In order to calculate the equilibrium properties of the system of interest, we need to solve the configurational space integral given in Equation 3.11. The Monte Carlo (MC) method is one of the numerical tools that can be used to solve these kinds of multidimensional integrals. Since it is a stochastic approach based on a random sampling of an integration interval the sampling of the configurational space ($\mathbf{r}^N$) in Equation 3.11 is completely time independent. Therefore, MC provides *no information* about the dynamics of the system and is suitable for studies of equilibrium properties, only. [32].

[32] D. Frenkel and B. Smit, *Understanding Molecular Simulations From Algorithms to Applications* (Acedemic Press, California, 2002).

3.2.1 Metropolis Importance Sampling

Since the configurations in Equation 3.11 are Boltzmann weighted, most of them will have a negligible contribution to the integral due to their vanishing weights. Hence, generating configurations with a random distribution will be wasteful to evaluate the averages like in Equation 3.12. However, the sampling can be focused on the important region of configurational space by using the Metropolis algorithm (MA) [33]. MA samples the configurations directly from a Boltzmann distribution with a probability of finding configuration i, $\mathcal{P}(i) = \frac{\exp(-\beta U(\mathbf{r}_i^N))}{\mathcal{Z}_N}$.

At equilibrium, the transition probability from an old configuration, o, to all other new configurations, n, must be equal to the transition probability from all the configurations to the old one to maintain the stationary nature of the equilibrium. This detailed balance condition can be satisfied when

$$\mathcal{P}(o)\pi(o \to n) = \mathcal{P}(n)\pi(n \to o) \tag{3.13}$$

where $\pi(o \to n)$ is the transition probability from an old configuration to a new one. This gives a probability of accepting a trial move from an old to a new configuration as,

$$\mathcal{P}_{acc}(o \to n) = \frac{\mathcal{P}(n)}{\mathcal{P}(o)} = \exp(-\beta[\mathcal{U}(n) - \mathcal{U}(o)]) \tag{3.14}$$

Obviously the acceptance probability cannot exceed unity which yields,

$$\begin{aligned} \mathcal{P}_{acc}(o \to n) &= \exp(-\beta[\mathcal{U}(n) - \mathcal{U}(o)]) \quad \text{when } \mathcal{U}(o) < \mathcal{U}(n) \\ &= 1 \qquad\qquad\qquad\qquad\quad \text{when } \mathcal{U}(o) > \mathcal{U}(n) \end{aligned} \tag{3.15}$$

This acceptance criteria ensures that the sampled configurations are based on the Boltzmann distribution and thus an *unweighted* average of the sampled configurations will provide Boltzmann (thermal) averaged properties.

[33] N. Metropolis and S. Ulam, Journal of the American Statistical Association **44**, 335 (1949), PMID: 18139350.

Box of Imagination

Imagine that you are again a photographer working for National Geographic magazine and you have chosen the ensemble average method to find out the favorite place of jaguars in a rainforest.

Since you don't have an infinite number of cameras you want to choose wisely where to locate them.

- If you chose to spread the cameras evenly in the forest without considering the Boltzmann distribution of Jaguars, you would barely end up with a few pictures of a jaguar and your sampling will be insufficient to distinguish their favorite location.
- However, if you chose to locate them according to Boltzmann distribution which says that jaguars like swimming, you would locate more cameras around the river and less deep in the forest which will increase your chance to collect jaguar pictures and find out their favorite location.

3.2.2 Basic Monte Carlo Algorithm

Here are the basic steps of a translational Metropolis MC algorithm where the configurations are sampled from a *Boltzmann* distribution.

1. Generate a random initial configuration by placing all particles (molecules) at random locations in the simulation box.
2. Calculate the interaction energy of the initial configuration, $\mathcal{U}_{pot}(o)$.
3. Pick a random particle in the system
4. Generate a random number $\mathcal{R}_1$ between (-0.5 and 0.5)
5. Make a trial to move the particle to a random location with a displacement of $\mathcal{R}_1\mathcal{D}$, which provides a maximum displacement of $\mathcal{D}/2$. The multiplication with the random number ensures a random displacement.
6. Calculate the interaction energy of the new trial configuration, $\mathcal{U}_{pot}(n)$.
7. Accept the trial move if ($\mathcal{U}_{pot}(n) < \mathcal{U}_{pot}(o)$), otherwise
8. Calculate the energy difference between the two configurations, $\Delta\mathcal{U}_{pot} = \mathcal{U}_{pot}(n) - \mathcal{U}_{pot}(o)$.

9. Generate a second random number $\mathcal{R}_2$ between (0 and 1).

10. Accept the trial move if $\mathcal{R}_2 < \exp(-\Delta\mathcal{U}_{pot})$, otherwise

11. Reject the trial move and restore the old configuration.
    This ensures accepting configurations according to a Boltzmann distribution and eliminates unrealistic configurations such as overlapping particles.

3.2.3 Technical Details

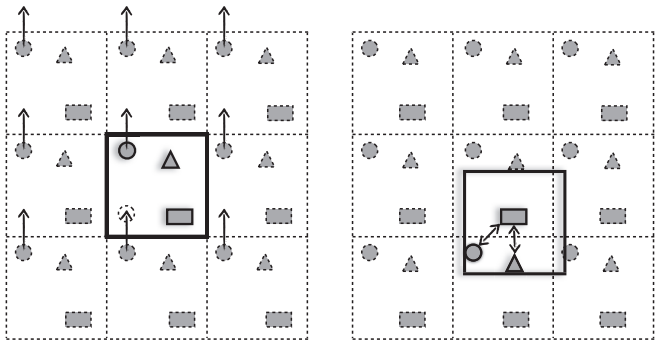
Technical details that are adapted to the system under study, can make the sampling much more efficient and thus can save appreciable computational time. In addition, proper settings are also required to eliminate artifacts and to ensure correct handling of the interactions.

Boundary Conditions

In simulations, the bulk properties of a macroscopic system are usually of interest. Yet, todays simulations are still far from handling billions of particles and the studied systems are limited to a few hundreds of thousands of particles. One problem of dealing with a small system is that most of the particles experience the surface of the simulation container instead of a bulk environment. In order to minimize boundary effects *periodic boundary conditions* can be used [32]. In this method, the actual simulation box is used as a unit cell to generate an infinite lattice of particles – see Figure 3.1. The replicated units are called the image of the simulation box. The periodicity of the box ensures that whenever a particle moves out of the container its replicated image enters from the opposite side of the container. Therefore, particles never feel the boundary of the container and only interact with their surrounding particles like in the bulk. However, when the size of the box is too small, this method can introduce unrealistic periodicity effects to the system.

Note that periodic boundaries cannot be used with a spherical container due to incontinuity of the lattice generated by a sphere. Therefore, non-periodic boundaries need to be applied with spherical containers. The same problem occurs for a radial axis of a cylindrical container as well.

[32] D. Frenkel and B. Smit, *Understanding Molecular Simulations From Algorithms to Applications* (Acedemic Press, California, 2002).



(a) Periodic boundaries (b) Minimum image convention

Figure 3.1: Graphical representations of periodic boundary conditions on a 2D system (left) and the minimum image convention for a cubic interaction cut-off (right).

Container Shape

This brings the question of what kind of shape is proper for the system. Let's say that we are interested in *two-body* interactions with an *implicit* salt and solvent. If the long-ranged electrostatic interactions quickly decays in the system a spherical container with hard boundaries can safely be used. However, the radius of the sphere must be big enough to cover all possible distances until the interactions die out to zero. When long-ranged interactions are involved, it is often better to use a cylindrical container with a periodic axial axis. Since we are only interested in the distance between two proteins, we can safely locate and move them only on the axial axis. The radius of the cylinder has to be big enough to ensure that the rotation of proteins will not be effected by its boundaries. Note that it is always safe to use a periodic box unless there is a charged surface at one of the container walls.

Minimum Image Convention

When long range interactions are weak and the box is big enough, the interactions can be truncated at the nearest image. In other words, interactions of a central particle will be calculated with the particles whose distances from the central one are smaller than the half box size – see Figure 3.1.

Monte Carlo Trial Moves

In order to generate new configurations and explore all possible configurational states, the particles must be able to visit the whole simulation box. Some basic moves such as **translation** and **rotation** update the position and the orientation of particles (or molecules) (section 3.2.2) and are almost always applied in an MC simulation unless it is a one-body bulk simulation.

All MC moves involving translation and rotation, have a displacement parameter ($\mathcal{D}$) which sets the maximum distance or angle that can be traveled at one trial. The displacement parameter should be determined by considering the box size, the density of the system and the strength of the interactions. An optimum displacement parameter maximizes the root mean square displacement during the simulation.

Conformational Moves

Flexible proteins can adapt many conformations and need moves that can update the relative positions and orientations of the residues within a chain. Here are four moves that are used in the studies of flexible proteins – see Figure 3.2.

- (a) *Pivot move* picks two random residues and defines a rotation axis according to the vector connecting them. Then, it rotates the residues that sit at one end of the protein around the rotation axis.
- (b) *Crankshaft move* defines the rotation axis as in the pivot move. Then, it rotates the residues sitting between the picked residues.
- (c) *Branch rotation* picks a random number of residues at one end of the chain and rotates them around a random axis.
- (d) *Reptation move* picks either the last or the first residue and translates it to a random position within the bond distance. Then, all the residues takes the locations of their neighboring residues where the movement propagates towards the unselected end.

Note that among these four, the reptation move is the only one that can modify bond lengths between the adjacent residues. If residues are connected with soft bonds which can fluctuate around an equilibrium distance, applying only crankshaft, pivot and branch rotation will be *insufficient* to sample conformations of the chain. Unfortunately, a reptation move can be slow to update distances between all residues since it requires (N/acceptance ratio) successive

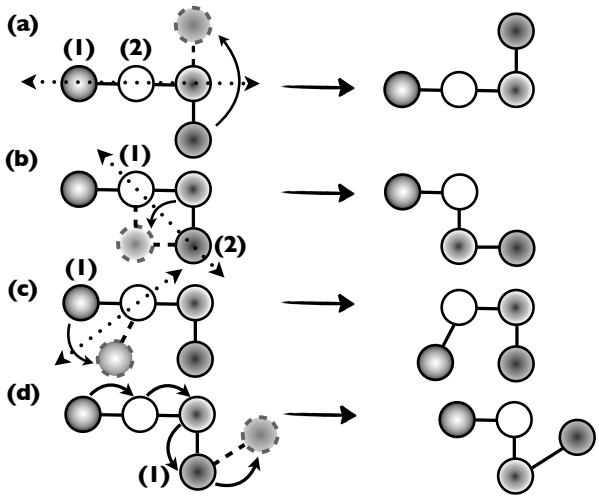


Figure 3.2: Illustration of conformational moves

moves on average to propagate along a chain. To improve the sampling efficiency, it can be used together with a single monomer translation which translates a random residue to a random location within slight deviations of the bond length.

Charge Swap Move

When charge regulation effects are pronounced in the system, it is important to allow charge fluctuations during the simulation. In order to switch charge of a residue on and off, a trial charge swap is applied to a randomly picked ionizable residue with an acceptance probability of $\mathcal{P}_{\text{acc}} = \min[1, \exp(-\beta\Delta\mathcal{U} \pm (\text{p}H - \text{p}K_a)\ln(10))]$ ($+$ is applied for protonation and $-$ applied for deprotonation).

Note that when pH is equal to the apparent $\text{p}K_a^*$ of a charged residue, the acceptance ratio of the charge swap move will be equal to 0.5, reflecting the equal probability of the protonated and the deprotonated states.

Isobaric Volume Move

In constant pressure simulations, the volume of the system is allowed to fluctuate using an isobaric volume move with an acceptance probability of $\mathcal{P}_{\text{acc}} = \min[1, \exp(-\beta(\Delta\mathcal{U}_{\text{pot}} + p_0\Delta V))]$. Whenever a volume move is applied, the

center of mass distances between the proteins are scaled according to the ratio between the old and the new box length. Note that only the inter-molecular interactions are modified by this scale. The particle distances within a protein are not scaled to maintain molecule integrity.

Replica Exchange

Replica exchange or parallel tempering, is an rare event sampling method to prevent getting trapped in local free energy minima. It is based on simultaneously simulating many replicas of a system at different thermodynamic states [34,35,36] and/or different Hamiltonians [37,38,39]. Each replica has easy access to a different region of the configurational space. Then, the idea is that sampling a broader space by exchanging configurations between the replicas at different regions. Note that the regions covered by replicas should overlap sufficiently to achieve acceptance. If there are $\mathcal{M}$ replicas to exchange configurations, it will be $\mathcal{M}$ times more computationally costly than a single simulation. However, the sampling in tempered simulations is more efficient than a single $\mathcal{M}$ times longer simulation by allowing access to the entire phase space [34,37,40] – see figure 3.3.

This method is especially useful in phase equilibrium studies where the free energy landscape is known to have at least two minima corresponding to gas and liquid phases. In order to construct a phase diagram from the equilibrium concentrations, ***both minima*** must be sampled in ***one*** simulation [32]. This can be done in an Np_0T simulation only when the pressure is close to the equilibrium pressure ($p \approx p_{eq}$) since low pressures tend to sample only the gas phase region (large box lengths) and high pressures tends to sample only the liquid phase – see Figure 3.3. However, determining the equilibrium pressure is impossible without sampling the two minima. With conventional sampling methods, the

[34] D. Bedrov and G. D. Smith, J. Chem. Phys. **115**, 1121 (2001).

[35] J. D. Chodera, W. C. Swope, F. Noé, J.-H. Prinz, M. R. Shirts, and V. S. Pande, J. Chem. Phys. **134**, 244107 (Jun. 2011).

[36] S. Patel, E. Vierling, and F. Tama, Biophys. J. **106**, 2644 (Jun. 2014).

[37] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).

[38] K. Wang, J. D. Chodera, Y. Yang, and M. R. Shirts, J. Comput.-Aided Mol. Des. **27**, 989 (Dec. 2013).

[39] M. Meli and G. Colombo, Int. J. Mol. Sci. **14**, 12157 (Jan. 2013).

[40] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (Dec. 2005).

[32] D. Frenkel and B. Smit, *Understanding Molecular Simulations From Algorithms to Applications* (Acedemic Press, California, 2002).

equilibrium pressure can only be obtain with many trial and error simulations at different pressures.

Parallel tempering allows for the sampling of the two minima without the knowledge of p_{eq}. The key is that the replicas should span a range of pressures which contain both higher and lower pressures than p_{eq}. This will ensure the sampling of both the gas and the liquid states. Hence, both minima will be accessible for each replica through configuration exchange. Apart from pressures, the Hamiltonian of the replicas can also be altered by setting different Debye screening lengths. In an electrostatically attractive system, high salt concentrations will favor the gas phase region which can drive a high pressure system towards the gas phase. Hence, a 2-dimensional replica exchange on both pressure and salt concentration can further improve the sampling and, as a bonus, provide the equilibrium pressure as a function of salt concentration.

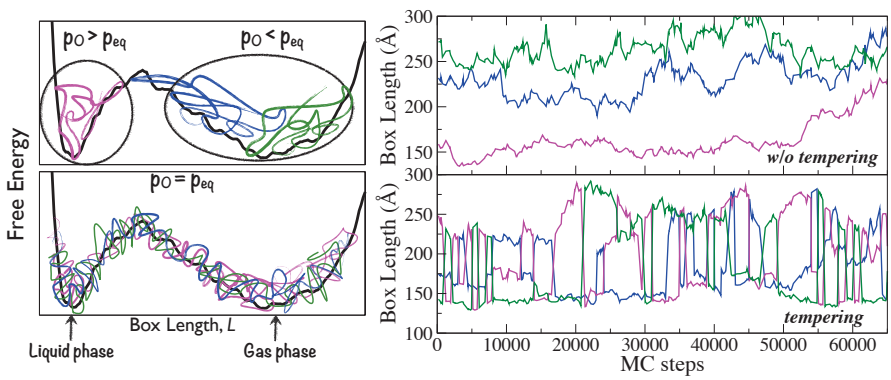


Figure 3.3: Left: Sampling of free energy landscape in an Np_0T simulation when the pressure is higher (pink) and lower (blue and green) than equilibrium pressure, with (bottom) and without (top) tempering.
Right: Sampled box lengths in three replicas as a function of MC steps with and without tempering.

In **Paper II**, we have applied such a 2D replica exchange where six replicas are simulated simultaneously and both configurations and volumes are swapped between the replicas of adjacent pressures. The swap move between replica $a(\kappa_a, p_{0a}, V_a, \mathbf{r}_a)$ and replica $b(\kappa_b, p_{0b}, V_b, \mathbf{r}_b)$ is attempted with the following

acceptance probability,

$$\begin{aligned}\mathcal{P}_{\text{acc}}[(a,b)\to(b,a)] = \min\Big\{1, \exp\Big(-\beta[&\mathcal{U}(\kappa_a,\mathbf{r}_b)\\ &+\mathcal{U}(\kappa_b,\mathbf{r}_a)-\mathcal{U}(\kappa_a,\mathbf{r}_a)-\mathcal{U}(\kappa_b,\mathbf{r}_b)\\ &+p_{0a}V_b+p_{0b}V_a-p_{0a}V_a-p_{0b}V_b]\Big)\Big\} \qquad (3.16)\\ = \min\Big\{1, \exp\Big(-\beta[&\Delta\mathcal{U}+\Delta p_0 V]\Big)\Big\}\end{aligned}$$

where κ is the inverse Debye screening length, V the volume, $\mathcal{U}$ is the energy of a given configuration. Figure 3.3 shows the box length sampling for three replicas with and without tempering.

3.3 Null Ellipsometry

Ellipsometry is an optical technique which provides information on the refractive index (n_f) and the thickness (d_f) of an adsorbed film at an interface as well as its adsorbed amount. The technique is based on detecting the deviations in the polarization of an incident beam after reflected at an interface [41].

A polarized light can be divided into two components, which are parallel (p) and perpendicular (s) to the plane of incidence – see Figure 3.4. The polarization state can be described by two ellipsometrical angles, Δ and Ψ, where the former defines the phase shift between the p and s-components and $\tan\Psi = X/Y$ defines the relative amplitude of the two components [42]. According to this definition, a linearly polarized light has an ellipsometrical angle ($\Delta = 0$ or π) while circularly polarized light has ($\Psi = \pi/4$ and $\Delta = \pm\pi/2$).

A basic set up for a null ellipsometer can be seen in Figure 3.5a. During a measurement, the light produced by a light source (LS) first travels through an analyzer (A) which polarizes the incoming light linearly. Then, it passes through a compensator (C, retarder) with a fixed azimuth angle ($\pm 45°$) to obtain an elliptical polarization with a quarter wave phase shift. After the compensator, the polarized light is shone on a sample (S) with a fixed incident angle, $\Phi_0 = 67.9°$.

[41] M. Malmsten, in *Biopolymers at interfaces*, edited by M. Malmsten (Marcel Dekker Inc., New York, 2003), chap. Ellipsometry and reflectometry for studying protein adsorption, pp. 539–545, 2nd edn.

[42] J. Humlíček, in *Handbook of ellipsometry*, edited by H. Tompkins and E. Irene (William Andrew, Norwich, 2005), chap. Polarized Light and Ellipsometry, pp. 3–14.

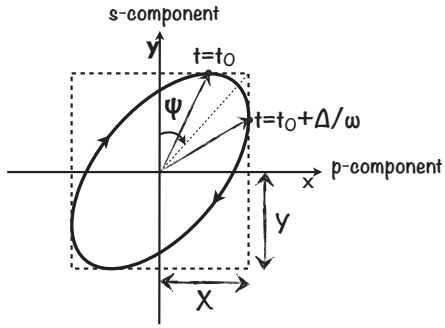


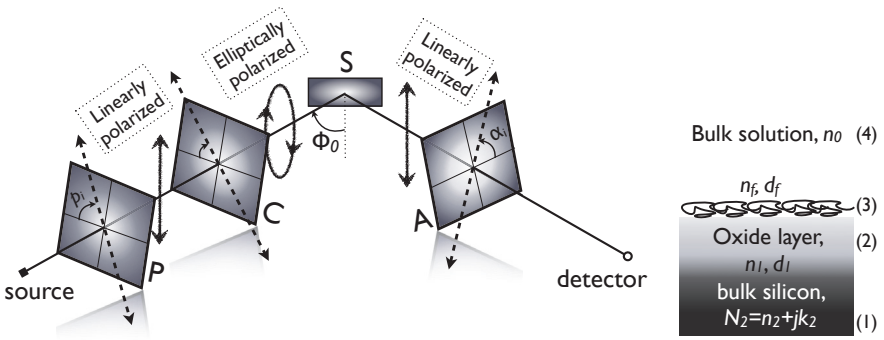
Figure 3.4: s and p component of elliptically polarized light described by Δ and Ψ. The light propagates in positive z-direction towards the reader. At $t = t_0$, the y component is at its maximum and after a time interval Δ/ω, the x component reaches its maximum.

At each measurement, the polarizer adjusts the incident light such that the elliptical polarization is inverted to a linear polarization after the reflection at the sample. The reflected light is then extinguished (or minimized) while passing through a suitably adjusted analyzer (A). Since the detector measures a null signal, this method is called null ellipsometry. The azimuths of polarizer and analyzer which extinguish the detected light, are called nulling angles [42]. For each compensator alignment, there are two sets of polarizer and analyzer nulling angles, referred to as nulling angles in zone one (p_1, a_1) and three (p_3, a_3) when the compensator is at $-45°$, and nulling angles in zone two (p_2, a_2) and four (p_4, a_4) when the compensator is at $+45°$. The relation between the nulling angles and Δ and Ψ can be found elsewhere [42]. The calibration is usually done by averaging the four zone measurements to determine the corrections for unavoidable errors at zero positions of A, P and C [43].

In **Paper VI**, oxidized silicon wafers were used as substrates due to their chemical inertness and excellent optical properties provided by the oxide layer [43,44]. When the thickness of oxide layer is larger than 100 Å it optimizes the optical contrast in the system and enables a separate determination of n_f and d_F. However, it also brings a multilayer substrate problem where the oxide and bulk silicon layers have to be treated separately in the optical model. The reflec-

[43] F. Tiberg and M. Landgren, Langmuir **9**, 927 (Apr. 1993).

[44] M. Landgren and B. Jönsson, J. Phys. Chem. **97**, 1656 (1993).



(a) Ellipsometer

(b) 4-layer optical model

Figure 3.5: Left: A basic set up for an ellipsometer in a polarizer, compensator, sample and analyzer (PCSA) arrangement. Right: Four layer optical model to determine adsorbed film properties.

tive index (n_1) and the thickness (d_1) of the oxide layer as well as the complex refractive index of bulk silicon ($N_2 = n_2 - jk_2$) must be determined in advanced. It has been shown that the optical properties of a multi-layered substrate can be obtained by performing measurements in multiple media [44].

Four unknown optical properties of a silicon wafer can be determined by two sets of Δ and Ψ which are obtained from measurements in air and buffer solution. After determining the substrate properties, the adsorbed film properties (n_f, d_f) are determined in an iterative way by using a four layer optical model (Figure 3.5b) and assuming a homogeneous film and a planar interface. Details of this iterative procedure can be found in reference [43]. The adsorbed amount, Γ, is calculated using the relation [45],

$$\Gamma = d_f \frac{n_f - n_0}{\frac{\mathrm{d}n}{\mathrm{d}c}} \tag{3.17}$$

where n_0 is the refractive index of buffer solution and $\frac{\mathrm{d}n}{\mathrm{d}c}$ is the refractive index increment of adsorbed substance. It assumes that the refractive index increment is constant up to concentrations obtained in the measurements. It is shown that $\frac{\mathrm{d}n}{\mathrm{d}c}$ lies around 0.18 mL/g and is fairly constant for protein solutions [45].

[44] M. Landgren and B. Jönsson, J. Phys. Chem. **97**, 1656 (1993).

[43] F. Tiberg and M. Landgren, Langmuir **9**, 927 (Apr. 1993).

[45] J. A. De Feijter, J. Benjamins, and F. A. Veer, Biopolymers **17**, 1759 (1978).

In this method, the determination of the adsorbed amount of proteins is more accurate than n_f and d_f since the errors in n_f and d_f cancel out to a large extent due to their interconnection. Therefore, the adsorbed amount can be determined within a 1% uncertainty, while 10% fluctuations can easily be observed in the thickness of the film.

CHAPTER 4

Getting Creative: Developing Coarse Grained Models

Coarse graining (CG) focuses on the essential physics of a system while neglecting or smoothing out (strictly speaking integrating over) the irrelevant details. This approach reduces the number of degrees of freedom in the system and allows large scale studies with a lower computational cost. Despite advances in atomistic force fields and computational techniques, CG models are still gaining popularity due to the conceptual benefits behind their simplicity [46]. They are widely used for studies of large scale protein systems such as protein association [47], protein aggregation [48] and membrane-protein interactions [49,50].

Defining essential details is crucial to develop *accurate* coarse grained models representing reality. For instance, modeling a disordered peptide as a rigid body at a given conformation will not represent its real behavior in solutions where it can adapt many conformations. In contrast, a rigid body model may be adequate for a globular protein with a well defined 3D shape with limited flexibility. Important details are not only system specific but they are also tightly related to the goal of the study. If you are interested in disordered peptide behavior in a β-sheet, you might as well constraint it to an extended conformation

[46] W. G. Noid, J. Chem. Phys. **139**, 090901 (Sep. 2013).

[47] H.-X. Zhou and P. A. Bates, Curr. Opin. Struct. Biol. **23**, 887 (Dec. 2013).

[48] C. Clementi, Curr. Opin. Struct. Biol. **18**, 10 (Feb. 2008).

[49] M. Baaden and S. J. Marrink, Curr. Opin. Struct. Biol. **23**, 878 (Dec. 2013).

[50] G. Illya and M. Deserno, Biophys. J. **95**, 4163 (Nov. 2008).

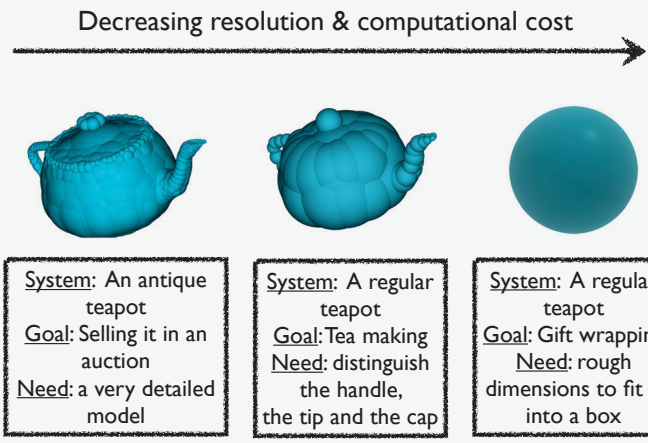
which will reflect its β-sheet confinement. This kind of conformational restrictions provides sampling of the *relevant* part of the configurational space while eliminating other conformational degrees of freedom.

Apart from restraining configuration space, CG modeling usually involves clustering of a number of atoms into a *superatom* (or *CG site*) which determines the resolution of the model. These sites interact through an effective potential that accounts for the collective effects of the clustered atoms. The parameters of the effective potential can be mapped from more detailed models (bottom-up approach, Paper II), knowledge-based statistical analysis (Paper IV) and macroscopic system properties such as binding constants and structure factors (top-down approach, Paper II, III, IV, V) [46].

Box of Imagination

Imagine that you want to model a teapot which needs to satisfy different user needs.

- If you need to sell this teapot in an auction you need to emphasis fine details of the pot which requires a model with many particles.
- If you want to make a gift wrap, then it is enough to know rough dimensions of the pot which can be modeled by a sphere.



Teapot figures are taken from the website: Bradshaw, G., *Sphere-Tree Construction Toolkit*, Feb. 2003, http://isg.cs.tcd.ie/spheretree/

[46] W. G. Noid, J. Chem. Phys. **139**, 090901 (Sep. 2013).

The resolution of the model has to be determined based on the system requirements. To study adsorption of disordered proteins, it is crucial to capture cooperation between the neighboring residues which requires at least an amino acid resolution (Paper IV and V). However, to study dipolar interactions on phase separation where hydrophobic interactions are fairly isotropic, it is sufficient to represent *charged residues* at an amino acid resolution and cluster the rest in one site (Paper III). Table 4.1 shows the CG models developed for each studied protein system and the requirements that are considered in the development can be found.

It is crucial to verify and validate a new model before predicting new outcomes. At a given set of conditions, the outcomes of independent simulations have to be consistent. After a consistency check, the model has to be validated by comparing with experimental observations which are not used during development. Once the model is validated, it can be used to predict new outcomes. The model development steps can be found below.

Steps of protein model development

1. Clearly define goals of your study.
2. Determine the essential physics in the system according to the goals of the study.
    - Make a preliminary bulk study on proteins to determine their net charge, charge capacitance and surface anisotropy.
    - Check if any conformational changes are involved for globular proteins.
3. Determine the required model resolution to capture the essentials.
4. Parameterize the effective interaction potentials according to the model resolution.
5. Select the proper simulation method such as *ensembles, boundary conditions, and MC moves* for the system of interest.
6. Verify your model/method by running independent simulations and comparing the outcomes with other simulation packages.

    If the outcomes are not consistent;

    - make system size dependence analysis if it is a many body problem,

- check the boundary conditions and the box sizes,
- apply more efficient sampling techniques and advanced MC moves.
- go back to second step.

7. Once the outcomes are consistent, validate the model by comparing with experimental observations.
    - If the outcomes are *inconsistent* with experiments, go back to second step.
8. Model is ready to predict new outcomes!

Weaknesses of the used models

For fairness, we will here briefly review the weaknesses of the models used in this study. In the amino acid level (AA) model, the side chain charges are assigned at the mass center of the titratable amino acids. Therefore, the sizes of the CG amino acid beads restrict the closest distance that can be achieved between two charged residues. However, on a real protein, the charges sit at the side chains which allows closer charge separations. The model therefore, underestimates charge-charge interactions at close contact. Up to some extend, this can be eliminated by introducing a two-bead model per amino acid which will allow closer separations due to the smaller sizes of the side chain beads. Note that the two-bead model will double the number of particles on a protein and will thus be computationally more costly.

The AA model is unable to capture H-bonding which requires explicit representation of –NH and –H groups of a protein backbone. It is therefore, inadequate to study H-bond driven protein behaviors such as assemblies of β-sheet forming peptides.

In the flexible protein model, the amino acids are attached by freely rotating bonds without any rotational restriction other than the size and the non-bonded interactions of the CG beads. Since the amino acids with bulky side chains are represented with bigger CG beads, the restricted rotations due to side chain sizes is captured in the model to some extent. However, it is known that the kink forming proline residues introduce an extra stiffness to the protein chain which is not incorporated into the flexible model. This means that, the flexibility of a protein chain with a high PRO content will be overestimated by the flexible protein model.

Implicit ion binding model only captures the "salting in" effect of the Hofmeister ions but not the "salting out" effects due to the surface tension changes.

Moreover, the binding affinity is not related to the solvent expose area of the hydrophobic side chains. It is also incapable of reproducing the possible bridging effects of the bound ions between two hydrophobic residues.

Table 4.1: CG models developed for each studied system along with the requirements considered in their development.

Paper I	**Protein**: N terminal domain of silk protein with a high dipole moment; globular **Goal**: Both electrostatic and hydrophobic anisotropy effects on the protein association **Essentials**: Distribution of hydrophobic patches and charged residues on the surface Response of the charged residues at close contact **Model**: *Resolution*: one CG site per amino acid located at their respective center of masses; *Rigid protein:* conformation from the crystal structure; *Fluctuating charges*: due to proton binding
Paper II	**Protein**: positively charged lysozyme and negatively charged α-lactalbumin with a high dipole moment (see Table 5.1) **Goal**: Effect of anisotropic charge distribution of α-lac on phase separation of the oppositely charged protein mixtures **Essentials**: Distribution of charges on the surface; Fairly isotropic distribution of hydrophobic residues; Globular proteins with negligible conformational changes **Model**: *Resolution*: one CG site per charged amino acids, one big CG site for the rest. *Rigid protein:* conformations from the crystal structures;
Paper III	**Protein**: γD-crystalline; globular **Goal**: Hofmeister anion binding effects on the protein charge and association **Essentials**: Anion binding to solvent exposed hydrophobic side chains and backbone with different strengths; Response of charged residues to the bound ions **Model**: *Resolution*: two CG sites per amino acid, one for backbone and one for side chain; *Rigid protein:* conformation from the crystal structure; *Fluctuating charges*: due to both anion and proton binding
Paper IV	**Protein**: Histatin 5; disordered protein; with a high histidine residue content **Goal**: Effect of histidine richness on the surface adsorption **Essentials**: All possible conformations due to protein flexibility; Cooperative effect of neighboring residues in the primary structure; Charge fluctuations on the Histidine residues; Zn^{2+} binding to de-protonated histidines, competing with protons; **Model**: *Resolution*: one CG site per amino acid connected with a spring; *Flexible protein:* allowing conformational modifications; *Fluctuating charges*: due to proton binding; *Explicit Zn^{2+} ion binding* to de-protonated histidines *Uniform and constant surface charge density*
Paper V	**Protein**: Midkine; disordered protein **Goal**: pH and salt concentration effect on surface adsorption **Essentials**: All possible conformations due to protein flexibility; Cooperative effect of neighboring residues in the primary structure; Charge fluctuations on the surface due to pH change **Model**: *Resolution*: one CG site per amino acid connected with a spring; *Flexible protein:* allowing conformational modifications; *Fluctuating charges*: due to proton binding; *Charged density of surface* is assigned according to pH

CHAPTER 5

Atlas of Studied Proteins

5.1 Electrostatic Anisotropy

The first approach to judge the importance of electrostatic interactions for a protein is to calculate the net charge and predict the repulsion between the proteins. Proteins with high net charges are expected to aggregate (precipitate/phase separate) upon addition of salt due to reduced electrostatic repulsion. However, N-terminal domain of spider silk protein (N-ter) with a high net charge (-4.1e, Table 5.1) shows the opposite behavior where salt stabilizes the monomers in solution [51].

Anisotropic Charge Distribution: Dipole moments (Paper I and II)
The protein net charge itself is not always sufficient to judge the effect of electrostatic interactions. In **Paper I**, we have shown that higher order multipoles such as dipoles can also have pronounced effects in the interactions. As seen in Table 5.1, N-ter has a high dipole moment which can lead to strong dipolar attractions. This attraction overcomes the net charge repulsion, resulting in a binding affinity in between the proteins that drives dimerization. The dimerization of N-ter therefore shows a contradictory salt response due to the screening of dipolar attraction. Since dipolar interactions are orientational dependent, they force an

[51] F. Hagn, C. Thamm, T. Scheibel, and H. Kessler, Angew. Chem., Int. Ed. Engl. **50**, 310 (Jan. 2011).

Table 5.1: Electrostatic properties of the studied proteins at pH 7 in 10 mM monovalent salt solution.

Structure	Protein	PDBid	Titrat. Resi.	Buried[a] Charges	Charge (e)	Capaci. (e^2)	Histidine Number	Dipole (eÅ)
Globular	Lysozyme	4LZT	32	1D, 1Y	7.1	0.24	1	29
	holo α-lac	1F6S	42	1D	-4.1	0.63	3	71
	apo α-lac	1F6R	42	1D	-5.6	0.88	3	88
	γD-crystalline	2KLJ	72	5C, 1Y	1.3	1.1	6	23
	N-ter compact	3LR2	13	0	-4.1	0.27	1	101
	N-ter loose	2LPJ	13	1K	-4.1	0.20	1	101
	C-ter	2KHM	9	1E	-0.2	0.17	0	32
IDP	Histatin5	–	20	–	4.7	0.55	7	–
	Midkine	–	45	–	18.5	0.30	0	–

[a] Titratable residues with solvent exposed side chain surfaces smaller than 1 Å ($\mathrm{SASA}_{sc} < 1$) are considered to be buried.

antiparallel subunit alignment in the dimer structure.

By simply analyzing the dipole moment of proteins together with charge-repulsion, the effect of dipolar interactions ca be anticipated. From Table 5.1, we can foresee that electrostatic (neither charge nor dipole) interactions will not be pronounced in solutions of silk C-terminal domain.

We also showed in **Paper II** that anisotropic dipolar interactions are significant for the phase separation of the oppositely charged proteins lysozyme/α-lactalbumin (α-lac) as is the net charge attractions. The high dipole moment of α-lac (Table 5.1) facilities alignment towards the positive lysozyme resulting in a structuring in concentrated solution. When we deliberately reduced the dipole moment of apo α-lac while maintaining its net charge, the phase separation is remarkably suppressed, indicating the importance of charge anisotropy.

Hofmeister Ion Binding (Paper III)

It is also important to consider the Hofmeister ion binding to hydrophobic surfaces, if present. In **Paper III**, we have shown that the anion binding effect becomes significant at fairly low ion concentrations (around 100 mM) and appreciably shifts pI towards lower pH values. When pH is above pI of the protein, anion binding *neutralizes* the protein and suppresses the repulsion. Hence the stability of protein solution is reduced in the presence of binding ions such as thiocyanate – the *reverse* Hofmeister series. In contrast, when pH is above pI, anion binding provides *additional* negative charges that *stabilizes* the protein solution – the *direct* Hofmeister series. Apart from providing extra charges, bound

ions also block hydrophobic interactions and facilitates further stabilization of protein solutions. Since Hofmeister ions contribute to the charge fluctuations, it may also promote charge regulation effects.

Interactions with Surfaces

Charge Regulation due to Protons (Paper IV)
Charge regulation is the ability to adopt the most favorable protonation state as a response to an external field such as originating from a charged surface or another highly charged protein. When environmental pH matches the pK_a^* of an ionizable residue, both protonated and deprotonated states are equally probable. This maximizes the charge fluctuation of the residue and facilitates the exchange of the protonation states under an external field. The *charge capacitance* which measures the net charge fluctuations on a protein, can be used to judge the extend of the regulation effects at a given pH. High charge fluctuations on many residues give rise to a high protein charge capacitance and thus, stronger regulation effects.

A brief estimation of the regulation effects can be done by counting the number of residues with p$K_a \approx$ pH. For instance, as shown in **Paper IV**, at physiologic conditions, the main contribution to the charge capacitance of a protein comes from the histidine residues and the N-terminal since their pK_a values are close to pH 7. Therefore, the number of histidines determines the charge capacitance and thus regulation effects at neutral pH – see Table 5.1. Note that this estimation is based on pK_a of the residues and thus it does not include the interactions with the electrostatic environment. Since the histidine residues in histatin5 are surrounded by many other positive residues, their proton uptake is restrained by the neighboring positive charges that shifts the pK_a^* to a lower pH (p$K_a^* < 6.3$). This shift increases the discrepancy between the pH and the pK_a^* resulting in a lower charge capacitance than γD-crystalline with a similar HIS content. A high histidine content of histatin5 provides appreciable surface induced charge regulation around pH 6 but not pH 8. At pH 6, histidines supply up to $+2e$ additional charges by proton uptake and promote surface adsorption. In contrast, at pH 8, an inhibited regulation ability of histidines lead to a lower surface affinity.

Multivalent Metal Ion Binding (Paper IV and VI)
Metal ion binding can also promote the adsorption to negative surfaces by providing extra positive charges to proteins and thus, additional electrostatic at-

traction – **Paper IV and VI**. Metal binding can facilitate cross-links both *within* (Histatin 5, Paper IV) and *between* the proteins (β-casein, Paper VI) leading to a reduced entropic cost of surface confinement as well as closer packing on the surface.

Chain Flexibility (Paper IV and VI)

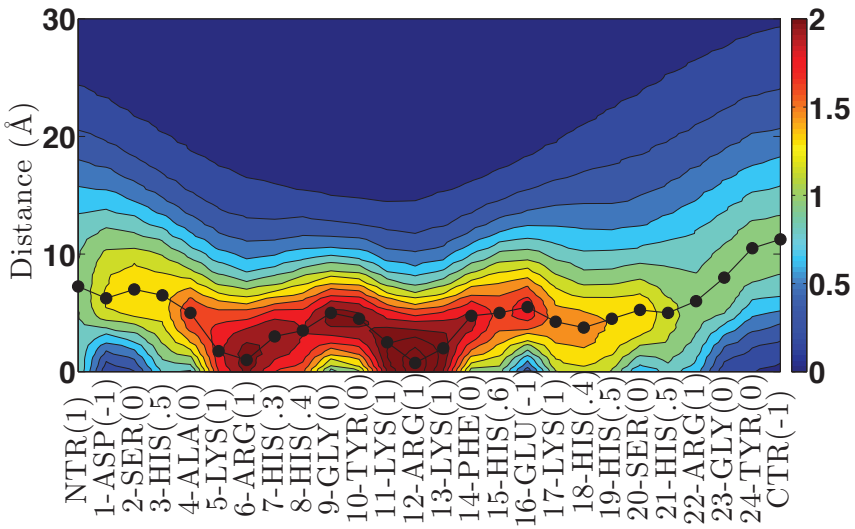


Figure 5.1: Residue concentration profile of Histatin5 as a function of distance from a negative surface. Positive ARG-LYS patches provide anchor points to the surface.

Flexibility of the protein chains brings another complexity into surface interactions. It allows relocation of residues in space which favors attractive interactions. Note that since the neighboring residues in the chain are bound to each other with a strong peptide bond, they still have to be located close to each other in the resulting conformation. For instance, if one negative and one positive residue is neighboring in the chain and interacting with a negative surface, the positive one will favor being in the vicinity of the surface, while the negative will be pulling it away. Therefore, the cooperation between the neighboring residues and thus, the anisotropy of the primary sequence, becomes remarkably pronounced in the interactions.

In **Paper IV**, we showed that cooperation between the two positive ARG-LYS blocks in the middle of histatin5 chain pulls the protein towards the negative

surface and provides two anchor points on the surface – see Figure 5.1. Once the chain is close to the surface, HIS residues charge themselves up as a response to the surface electric field. These extra charges overcome the entropic cost of locating N and C-terminal tails close to the surface and promotes adsorption.

In **Paper VI**, we showed that the chain anisotropy and flexibility of β-casein overcome the net charge repulsion between the surface and facilitates the surface adsorption. Since the chain consists of negatively and positively charged blocks, sufficiently separated by the non polar residues, the positive C-terminal block can easily direct the surface adsorption by locating itself in the vicinity of the surface.

5.2 Hydrophobic Anisotropy

At close separations, hydrophobic interactions can act as a glue between the proteins and stabilize the multimeric forms. The effect of these interactions can be estimated by considering the solvent exposed surface area (SASA) of the hydrophobic side chains. As seen in the Table 5.2, 15% of the protein surfaces, on average, consists of hydrophobic side chains which may be involved in the hydrophobic interactions.

The C-terminal domain of the silk protein (C-ter) has an unusually high hydrophobic surface exposure, compared to the other proteins. We also know from the electrostatic anisotropy analysis that its dipolar interactions are insufficient to form dimers. This implies that the dimerization of C-ter is driven by the hydrophobic interactions and possibly occurs during the secretion in the spider gland. We have shown in **Paper I** that the dimerization surface of the C-ter mainly consists of hydrophobic residues, supporting the hydrophobically driven dimerization. As seen in Table 5.2, lysozyme has the minimum hydrophobic surface exposure. Its small hydrophobic surface together with its high net charge at pH 7 implies that the hydrophobic interactions will be insignificant in its solution at physiological conditions.

Like charged residues, an uneven distribution of the hydrophobic side chains on the surface may lead to anisotropic interactions. We can estimate the effect of hydrophobic anisotropy by analyzing the hydrophobic patchiness of the protein surfaces. Table 5.3 shows the three largest hydrophobic patches of the studied proteins, their surface areas in percents of the total hydrophobic surface area and the number of residues in each patch. The C-ter and the loose conformation of the N-ter domains of the silk protein have a very large hydrophobic patch consisting 44% and 82% of their total hydrophobic surface, respectively. This

Table 5.2: Structural and surface properties of the studied proteins.

Structure	Protein	Residue Number	Hydrophobic Residue	SASA (Å^2)	Hydrophobic SASA[a]	R_{max} (Å)
Globular	Lysozyme	129	44	6631	7%	23
	holo α-lac	122	40	6781	15%	25
	apo α-lac	123	41	6548	16%	24
	γD-crystalline	174	51	8310	9%	27
	N-ter compact	127	57	6192	21%	23
	N-ter loose	127	57	6213	20%	23
	C-ter	108	51	6512	33%	25
IDP	Histatin5	24	2	–	–	42[b]
	Midkine	122	33	–	–	121[b]

[a]Hydrophobic SASA includes only solvent exposed side chain surfaces of hydrophobic residues. Exposed backbone surfaces are not considered.

[b]End-to-end distance is given for IDPs at pH 7 and in a salt free aqueous solution.

implies that hydrophobic interactions will be established through a large surface which may not be forcing a very specific orientation. However, for the other proteins, the hydrophobic patches are smaller and thus more localized which may favor a specific orientation to maximize the hydrophobic contacts.

Table 5.3: The top three largest hydrophobic patches of the studied proteins. Each residue within the distance of 6.5 Å are assigned to a patch.

	Patch 1		Patch 2		Patch 3	
Protein	Residue	Area	Residue	Area	Residue	Area
Lysozyme	7	24%	3	22%	8	17%
holo α-lactalbumin	4	22%	3	15%	9	13%
apo α-lactalbumin	4	22%	9	19%	8	18%
γD-crystalline	2	17%	3	11%	1	10%
N-ter compact	13	27%	9	24%	9	23%
N-ter loose[a]	42	82%	3	15%	1	3%
C-ter	19	44%	5	9%	4	9%

[a]Patch analysis has been done with a distance cut-off of 6 Å since the analysis with 6.5 Å is assigned all the residues into one patch.

As seen in Table 5.3, the area of the hydrophobic patches are not directly proportional with the number of residues in the patch. For instance, the largest patch of apo α-lactalbumin contains only four residues which make up 22% of the hydrophobic surface. However, its second largest patch contains *nine* residues (twice of the first patch) and it makes 19% of the hydrophobic surface. In this case, a pair-wise additive treatment of hydrophobic interactions, which

weights the patches according to the number of residues, will favor interactions with the second patch and will result in an ill-defined hydrophobic anisotropy. This may also cause further complications by competing with the electrostatic anisotropy due to favoring unrealistic orientations. This problem can be handled by using a SASA based potential instead of a pair-wise additive one. The strength of the interactions can be determined by the covered hydrophobic SASA upon approaching and an effective surface tension.

In **Paper I**, we used such a potential to capture the hydrophobic anisotropy of the N-ter. We showed that in the presence of hydrophobic interactions, the compact conformation of the N-ter forms dimers with a well-defined hydrophobic patch orientation as observed in its crystal structure. When the hydrophobic interactions are tuned off, the obtained dimeric structures lack patch correlation. This indicates that the SASA based hydrophobic model is able to realistically capture the hydrophobic anisotropy of the N-ter surface and the resulting patch correlations between the N-ter in the dimeric form. We have also showed that the loose conformation of N-ter is unable to set a specific hydrophobic patch orientation. Therefore, it no longer cooperates with the dipolar alignments resulting in suppressed dipolar attractions.

The extent of hydrophobic interactions can roughly be estimated by considering the maximum radius of the proteins. Let's assume that the residue with the longest distance from the protein centre is hydrophobic and two hydrophobic residues interact when their separation is shorter than a diameter of a water molecule (3 Å). Then, the hydrophobic interactions will rise when protein center of mass separation is smaller than $2R_{max} + d_{water}$.

CHAPTER 6

Conclusion and Outlook

In this thesis, we have looked into the molecular details of protein interactions with the help of coarse grained models. It is emphasized that anisotropic interactions, which are often overlooked, might be the underlying reason behind the observed protein behavior. We showed that a simple analysis of the protein electrostatics and hydrophobicity can be valuable to estimate the extent of the anisotropic effects.

While studying protein electrostatics it is important to keep in mind the specific ion binding effects. Since specifically bound ions can alter the protein charges significantly, they bring another complexity into the system. To eliminate this complication, the experimental electrolyte solutions should be chosen free of binding ions.

Our results show that a delicately designed coarse grained model can capture the significant physical details of the real proteins and provide insights into complex protein behaviors.

In scientific research, there is always room for improvements. This is also true for the studies presented here, and below, you can find some ideas for further studies:

- A Born energy type penalty can be applied to buried charged residues to account for the cost of charges in the protein interior.

- The SASA based hydrophobic potential should be parameterized in a sys-

tem where the electrostatics are insignificant and should be further investigated by comparing with explicit solvent simulations. It can also be generalized to capture ion specific "salting out" effects by incorporating the surface tension changes due to salt.

- The implicit ion binding model can be generalized to apply all kinds of ion binding processes including divalent ions. The model can further be improved by implementing an attractive potential between the occupied and the non-occupied binding sites to capture possible bridging effects of divalent ions. Moreover, the binding constant of ions can be related to the solvent exposed surface area of binding sites to incorporate the degree of solvent exposure in the model.
- Rotational restrictions can be incorporated to the freely-joint flexible protein model through dihedral angle constraints to capture possible chain stiffness due to proline residues.
- A limited degree of flexibility can be introduced to the rigid protein model by selecting protein conformations from a set of possible structures during simulations.
- Additional hydrogen bonding sites can be implemented in the amino acid model to capture the directionality of the hydrogen bonds.
- The adsorption of β-casein can further be investigated by the QCM- dissipation method to get a deeper understanding of the adsorbed layer characteristics and the effect of Ca^{2+} triggered aggregation on the surface.

Paper I

Dimerization of terminal domains in silk fibers: Hydrophobic and electrostatic anisotropy

Anıl Kurut,* Cedric Dicko, and Mikael Lund

Department of Theoretical Chemistry, Lund University, P.O.B. 124 SE-22100 Lund, Sweden

E-mail: anil.kurut@teokem.lu.se

*To whom correspondence should be addressed

Abstract

The well-tuned spinning technology from spiders has attracted many researchers with the promise of producing high performance, biocompatible and yet biodegradable fibres. So far, the intricate chemistry and rheology of spinning have eluded us. A break through was achieved recently, when the 3D structures of the N-, and C-terminal domains of spider dragline silk were resolved and their pH induced dimerization was revealed. To understand the terminal domains' dimerization mechanisms, we developed a protein model based on the crystallographic structures that reproduces charge and hydrophobic anisotropy of the complex protein surfaces. Monte Carlo simulations were used to study the N-terminal domain dimerization as a function of pH and ionic strength. We show that the hydrophobic and electrostatic anisotropy of the N-terminal domain cooperate constructively in the association process. The dipolar attractions at pH 6 lead to weakly bound dimers by forcing an antiparallel monomer orientation which is then, stabilized by hydrophobic locking at close separations. Elevated salt concentrations suppress the protein association due to screened electrostatic dipolar attraction. Moreover, the mutations on ionizable residues reveal a free energy binding, proportional to the dipole moment of the mutants. Its has previously been shown that the dimers, formed at pH 6, completely dissociate at pH 7 which is though to be due to the altered protein charges. In contrast, our pH study indicates that such pH increase has no influence on the charge distribution of the N-terminal domain. Instead, we find that the pH effect is not because of the electrostatics but because of the adapted structure at pH 7. In this structure, the hydrophobic patches are relocated on the surface to decouple hydrophobic patch and dipolar alignments, resulting in reduced dipolar and hydrophobic attractions.

Introduction

Silks have evolved to be some of nature's most impressive composite materials.[1,2] Silk fibres[3] are not only one of the toughest polymers known, but they have a number of other characteristics, such as biocompatibility and degradability, that make them an interesting as well as an important object for research in the general areas of biomaterials,[4] biomimetic and the coevolution of behavior,[5] morphology, and function.[2] Harnessing silk production and processing have however been proven difficult and elusive. A complete understanding of silk fibre formation is needed.

Spiders have the ability to control key aspects of silk production,[1,6] namely (i) silk protein composition, (ii) storage and (iii) fibre processing. Modern studies on mechanical strength of silks and spinning behavior[7–9] show a clear correlation between silk strength and relative elasticity with specific molecular architecture.[10,11] Genetic analysis,[12–14] and earlier amino acid composition,[15] of silk proteins sequences demonstrated a strong bias towards specific amino acid content dominated by alanine and glycine residues. Furthermore, silk gene molecular design show preferential motif arrangements (poly-alanine crystalline domains and glycine rich amorphous domains) that can be correlated to macroscopic properties of silk such as elasticity and strength.[14]

Inside the glands, the silk proteins can be stored for long periods of time in an apparently stable form.[16,17] This conformation seems to change just prior to fibre formation in order to allow the highly energy efficient conversion from their aqueous solution state to the insoluble fibrillar state.[6] For this the individual silk molecules must adopt shapes and conformations dictated by polar and non-polar moieties.[18] To achieve bulk orientations with optimal axial stiffness (which is important for mechanical properties of the fibres) the system must allow the molecules to self-organize in their extended configuration[19] followed by an efficient inter-molecular lock-in but with only a limited number of cross-links in order to maintain flow viscosity. However, β-strands (unlike α-helical structures) are unstable in solution as isolated secondary structures[20,21] and hence tend to engage and stabilize interactions with neighbouring strands leading to the construction of an intermolecular gelation network.[6]

To control such 'structurally reactive' silk proteins, progressive and judicious modifications

of solution conditions (pH, ionic strength, etc.) are necessary.[21] The detailed exploration of the chemical controls has highlighted the role of acidification of the silk proteins as a key step prior to fibre formation.[21–24] The exact role of this acidification was unclear until recently when the 'N-, and C-terminal domains' structures of dragline silk were resolved.[25,26] Specifically, the N-terminal domain plays a critical role in delaying aggregation at neutral pH. At pH values of around 6, the N-terminal domain is found as a *homodimer*, while at high salt concentration and neutral pH it is found as a *monomer*.[27–30] Several models have now been proposed to explain the role of the terminal domains in silk fibre formation[29,31,32] via micelles formation and activation of specific amino acids. The details of the interactions involved in the assembly of the N-terminal homo-dimer and its implication for spinning may not be resolved experimentally. Computational methods can provide new insights on the balance of forces and sequence of events leading to fibre formation.

In the present contribution, we modeled the N-, and C-terminal domains of silk protein at an amino acid resolution based on their available crystallographic structures. Using these models in Monte Carlo simulations, we studied the dimerization of the N-terminal domain as a function of pH and ionic strength. We also analyzed the distribution of hydrophobic patches and charged residues on the surface of both terminal domains to determine the surface anisotropy (patchiness). We investigated individual contributions of the hydrophobic and the electrostatic interactions to the binding affinity to understand their interplay in the dimerization mechanism.

Model and Method

Model

N- and C-terminal domains were modeled as rigid bodies built from a collection of beads, each representing an amino acid. The protein structures were taken from the Protein Data Bank with IDs 3LR2[25] and 2KHM,[26] respectively. The first 32 residues of C-terminal domain were truncated due to their ill-defined structure. Bead sizes were determined by using a common amino acid density of 1.0 g/L and molecular weights of the represented residues. Each bead was mapped onto the

atomistic structures of the proteins by locating them at the mass center of the represented residues. Protonation states of ionizable residues were allowed to switch during the simulations, by turning their charges on and off, as a response to the present ions and the solution pH. Water was treated as a dielectric continuum and interactions were modeled with three components (Equation 1):

1. Van der Waals and excluded volume interactions, between all beads, using Lennard-Jones potential,

2. electrostatic interactions with implicit salt ions, between charged beads, using Debye-Hückel potential, and

3. hydrophobic interactions, between hydrophobic beads, using a square well potential with a range of 3 Å (size of a water molecule) whose strength is based on an effective surface tension (γ_{eff}) and the change in the solvent accessible surface area (ΔA_{sas}) upon approaching two hydrophobic surfaces.

In addition to these, an intrinsic protonation energy was incorporated to the total energy (U_{tot}), following a protonation trial of an ionizable residue.

$$U_{tot} = \overbrace{4\varepsilon_{\text{LJ}} \sum_{i,j\in n} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right)}^{\text{VDW \& excluded volume energy}} + \overbrace{\left(\frac{1}{4\pi\varepsilon_0\varepsilon_r} \sum_{i>j\in n_c} \frac{q_i q_j}{r_{ij}} e^{-\kappa r_{ij}} \right)}^{\text{Electrostatic energy}} + \overbrace{\left(-\gamma_{\text{eff}} * \triangle A_{sas} \right)}^{\text{Hydrophobic energy}} + \underbrace{\left(k_B T \sum_{i\in n_p} (\text{p}K_{a,i} - \text{pH}) \ln 10 \right)}_{\text{Intrinsic protonation energy}} \tag{1}$$

Parameters of Equation 1 are given in Table 1.

To model hydrophobic interactions, a solvent accessible surface area (A_i) was assigned for each hydrophobic residue which was calculated with a probe radius of 1.5 Å using VMD software.[33] The surfaces of two residues were considered inaccessible when they were closer than 3 Å, resulting in $\Delta A_{sas} = \sum_{i\in n_{inacc}} A_i$. The effective surface tension were adjusted 2.97 dyne/cm in 10 mM NaCl solution which corresponds to 0.25 kcal/mol ($\sim$0.44 kT) attraction for a residue pair with

Table 1: Parameters in Equation 1

Parameter $=$ **value**	**Description**
n	all residues
r_{ij}	distance between residue i and j
e	electron unit charge
ε_0	permittivity of vacuum
$\varepsilon_r = 80$	water dielectric constant
n_t	ionizable residues
q_i	charge
$\kappa^{-1} = \sqrt{\frac{\sum_i z_i^2 c_i}{\varepsilon_0 \varepsilon_r k_B T}}$	Debye length
$T = 298$ K	temperature
k_B	Boltzmann's constant
n_p	protonated residues
pK_a = see Table 2	acid dissociation constant
γ_{eff} = see Table 3	effective surface tension

Table 2: Intrinsic acid dissociation constants (pK_a) of ionizable (titratable) residues including carboxyl (CTR) and amino (NTR) groups of the terminal residues of each domain.

		number	
Site	**pK_a**	**N-ter**	**C-ter**
ASP	4	2	1
GLU	4.4	5	1
HIS	6.3	1	0
TYR	9.6	0	2
LYS	10.4	2	0
CYS	10.8	0	1
ARG	12	1	2
CTR	2.6	1	1
NTR	7.5	1	1
total		**13**	**9**

an averaged sized A_{sas} of 30 Å^2. The residues ALA, ILE, LEU, MET, PHE, PRO, TRP, VAL are considered hydrophobic.

Table 3: Effective surface tension and the change in the macroscopic surface tension of water[34] ($\partial\gamma/\partial c_{\text{salt}}$) in NaCl salt solution. $\gamma_{\text{eff}} = \gamma_{\text{eff}}(0\text{M}) + c_{\text{salt}}\partial\gamma/\partial c_{\text{salt}}$

Salt	$\partial\gamma/\partial c_{\text{salt}}$ (mN/m)/(mol/L)	γ_{eff}(10 mM) (dyne/cm)	γ_{eff}(100 mM) (dyne/cm)
NaCl	1.70	2.97	3.12

Method

We used the Metropolis Monte Carlo (MC) algorithm in the canonical ensemble (NVT) at 298 K in a simulation box of 300x300x150 Å^3. To investigate bulk electrostatics, a protein monomer (chain A of the crystal dimer structure) was placed in a NaCl solution and protonation state swap moves were applied to each ionizable residues with a corresponding intrinsic protonation and electrostatic energies – see Equation 1. Statistical averages of the net charge $\langle Q\rangle = \langle\sum_{i\in n_c} q_i\rangle$ and the dipole moment $\langle\mu\rangle = \langle\sum_{i\in n_c} q_i r_i\rangle$ were sampled as a function of pH, where r_i is the distance of the residue i from the protein center of mass.

In the dimerization simulations, two monomers (chain A and B) were located in the simulation box and allowed to translate and rotate on the z-axis to explore all possible monomer-monomer separations (R_{AB}) and orientations with a Boltzmann probability weight of $\exp(-U_{tot}/kT)$. The charges on the ionizable residues were also allowed to fluctuate during the simulations. The thermally averaged probability distribution function of monomer separations ($P(R_{AB})$) was determined by constructing a histogram of the visited separations. Using the probability distribution function, the free energy of interactions was calculated as $F = -kT\ln\frac{P(R_{AB})}{P(R_{AB,max})}$.

During the dimerization simulations, the statistical averages of the three energy components VDW, hydrophobic and electrostatic, were sampled as a function of protein-protein separation. In addition, the electrostatic energy was decomposed into ion-ion, ion-dipole, ion-quadrupole, and dipole-dipole energy to determine their separate contributions–see Table 4. The relative orientations of the dipole moments (dipole-dipole alignment) was calculated by taking the scalar product

between the dipole moment vectors of the two monomers. The dipole alignment due to the net charge of the neighboring monomer (ion-dipole alignment) was determined by averaging the z-component of the dipole moment.

Table 4: 3^{rd} degree electrostatic multipole expansion of protein-protein interactions where proteins are allowed to move only on the z-axis. Note that this expansion is valid when the protein-protein separation is bigger than the size of the protein, $R_{AB} > \mathscr{R}_{max} = 23$ Å. Note that salt screening effect is not incorporated in the multipole expansion.

component	**energy expressions**
multipole-multipole	$\sum_i^{N_A}\sum_j^{N_B}\frac{q_i q_j}{r_{ij}}$
ion-ion	$\frac{\sum_i^{N_A} q_i \sum_j^{N_B} q_j}{R_{AB}}$
ion-dipole	$-\frac{\sum_i^{N_A} q_i \sum_j^{N_B} q_j z_j}{R_{AB}^2} + \frac{\sum_i^{N_A} q_i z_i \sum_j^{N_B} q_j}{R_{AB}^2}$
ion-quadrupole	$\frac{\sum_i^{N_A} q_i \sum_j^{N_B} q_j (z_j^2 - 1/2y_j^2 - 1/2x_j^2)}{R_{AB}^3} + \frac{\sum_j^{N_B} q_j \sum_i^{N_A} q_i (z_i^2 - 1/2y_i^2 - 1/2x_i^2)}{R_{AB}^3}$
dipole-dipole	$-2\frac{\sum_i^{N_A} q_i z_i \sum_j^{N_B} q_j z_j}{R_{AB}^3} + \frac{\sum_i^{N_A} q_i y_i \sum_j^{N_B} q_j y_j}{R_{AB}^3} + \frac{\sum_i^{N_A} q_i x_i \sum_j^{N_B} q_j x_j}{R_{AB}^3}$

To determine the surface patchiness of the proteins, a separate hydrophobic patch analysis was performed on the monomers where each surface exposed hydrophobic residue within the distance of 6.5 Å (two averaged sized beads at contact) is assigned to a patch. As seen in Figure 2, each patch was designated with a vector (h_i), pointing to its surface center. During dimerization simulations, the scalar products ($h_{ij} = h_i \cdot h_j$) were calculated between the patch vectors of the monomer A (h_i) and monomer B (h_j) to determine the correlations of the hydrophobic patches in the dimers as well as to set an effective surface tension.

The binding affinity, ΔG, was calculated from the simulated interaction free energies (F),

$$\Delta G = -RT \ln K_{\text{dimer}}, \text{ where } K_{\text{dimer}} = \int_{2R_{hs}}^{\infty} (\exp^{-F(R_{AB})/kT} - 1) 4\pi R_{AB}^2 dR_{AB} \tag{2}$$

where K_{dimer} is the association constant of the homo-dimers and R_{hs} is the closest distance between the two monomers obtained in simulations. The difference in the binding affinities accompanied by a mutation is calculated as $\Delta\Delta G = \Delta G^{mut} - \Delta G^{wild}(\text{pH}\,6)$ with respect to the affinity of the wild

type at pH 6 and 10 mM ionic strength.

In all simulations, configurations (a given charged state, monomer-monomer separation and orientation) were generated by applying translation, rotation and charge swapping moves to the monomers with a frequency of 1:1:n_t. The statistical averages were taken over $45 \cdot 10^6$ configurations after 10^5 configurations of equilibration.

Results and Discussions

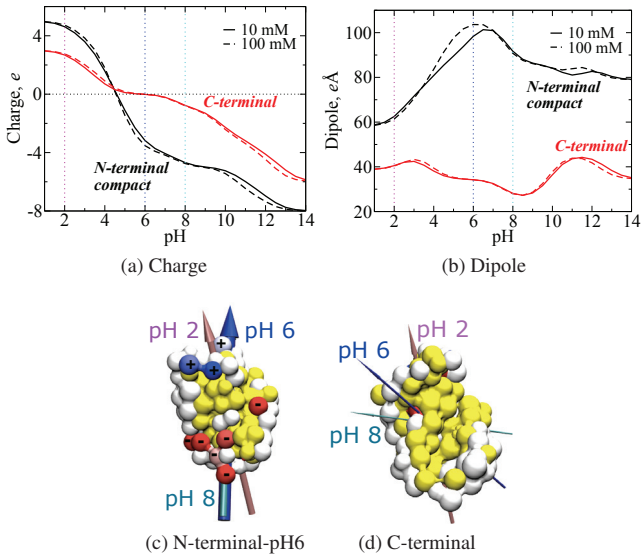
Structure analysis of the terminal domains

We first investigated the surface anisotropy of the N- and C- terminal domains of the silk protein resulting from the uneven distribution of the charged and hydrophobic residues on their surfaces.

Electrostatic anisotropy of the monomers

The anisotropic distribution of charged residues were quantified via calculating the dipole moment of the proteins. The net charge of the protein was calculated as a function of pH and salt concentrations to estimate the effect of net charge repulsions between the monomers. As seen in Figure 1, the charge of N-terminal domain decreases rapidly from +4.5e to -4e when solution pH is elevated from 2 to 6. This is accompanied by a dramatic increase of the dipole moment with a peak at pH 6. The charge distribution of N-terminal domain shows a remarkable anisotropy formed by clusters of GLU and ASP residues at the negative N-terminal end and HIS, ARG and LYS at the positive N-terminal end. This provides a significant dipole moment with a well-defined direction (Figure 1c). These clustered residues yield a maximized dipole moment around pH 6 where dimerization is favored.[27,35]

Elevated salt concentrations allow higher charges on the clustered residues by screening the repulsion between them (Table 5). Therefore, proteins can achieve the same level of de-protonation state at a slightly lower pH value. This influence is magnified in the dipole moment due to the collective effect of the clustered charges. Note that, the charge of LYS and ARG in the clusters are



(a) Charge (b) Dipole

(c) N-terminal-pH6 (d) C-terminal

Figure 1: Net charge (a) and dipole moment (b) of terminal domains as a function of pH and at different NaCl concentration. (c) and (d) show the dimerization surface where hydrophobic residues are depicted in yellow and charged residues at pH 6 and 10 mM salt are depicted in red and blue. Arrows show the dipole moment directions of N and C terminals where the arrow thicknesses are proportional to the strength of the dipoles.

already fully protonated even at low salt concentrations, thus their protonations are not effected by salt. In contrast, the C-terminal domain has an isotropic distribution of a few charges resulting in a weak dipole moment, insensitive to salt.

Hydrophobic anisotropy of the monomers

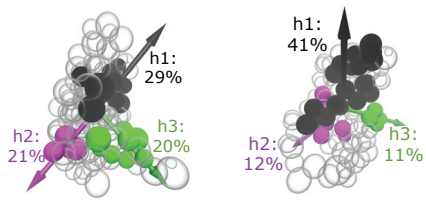
The hydrophobic anisotropy of a protein is caused by the clustering of solvent exposed hydrophobic residues on the surface. In order to quantify the hydrophobic anisotropy, we examined the solvent accessible surface area (A_{sas}) of both terminal domains and determined the area and the location of their three largest hydrophobic patches. We show that both domains have large hydrophobic surfaces that are reduced upon dimerization – see Table 6, Figure 1c and 1d.

In the C-terminal domain, 40% of the surface is hydrophobic which is roughly twice of that

Table 5: The effect of salt screening on the clustered charged residues at pH 6

	Residue	10 mM	100 mM
Positive end	HIS6	0.151	0.173
	LYS65	1.000	1.000
	ARG60	1.000	1.000
Negative end	ASP39	-0.921	-0.960
	ASP40	-0.876	-0.943
	GLU79	-0.858	-0.932
	GLU84	-0.770	-0.896
	GLU85	-0.887	-0.943

similarly sized globular proteins such as lysozyme and α-lactalbumin, hinting at a high preference to a hydrophobic association. As seen in Figure 2b, the C-terminal domain has three large patches



(a) N-terminal (b) C-terminal

Figure 2: Top three largest hydrophobic patches of terminals and the vectors point to surface centers of the patches.

(named h1 to h3 from the largest to smallest) in which the largest one (h1) includes more than 40% of the hydrophobic surface. N-terminal has a slightly lower degree of hydrophobic anisotropy where the largest patch constitutes 29% of the hydrophobic surface.

Anisotropic correlations in the crystal dimers

After defining the dipole moment and the hydrophobic patch vectors of the terminal domain monomers, we investigated the orientations of these vectors in the dimer configurations, resolved by the crystallography experiments. Figure 3 shows the alignment of dipole moment and hydrophobic patch vectors in the *crystal* structures. In order to quantify the monomer orientations in the dimers, we calculated the scalar products of the patch and dipole vectors – see Figure 3c.

Table 6: Surface properties of the crystal structures of terminal domains. Solvent accessible surface area was calculated with a 1.5 Å probe, including both side chain and backbone of the solvent exposed residues.

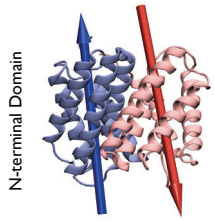
	N-terminal Domain		C-terminal Domain	
	monomer	**dimer**	**monomer**	**dimer**
Residue number, n	125	250	108	116
number of hydrophobic residues	57	114	51	52
Total surface area, A_{tot}, Å^2	6192	9812	6512	7918
Hydrophobic surface in % of A_{tot}	28%	24%	38%	29%
Binding surface per monomer, Å^2	1475	–	2548	–
Hydrophobic % of the binding surface	41%	–	55%	–

These scalar products are used as a reference state to compare the simulated dimer structures with the crystal structures.

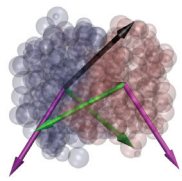
The monomer orientation in the N-terminal domain dimers ensures an antiparallel dipole alignment, yielding a scalar product of dipole moments, (d-d ≈ -1). However, this orientation suppresses ion-dipole interactions, (i-d ≈ 0). Moreover, the antiparallel monomers are located sufficiently close (17.7 Å) to enable close contact salt bridge interactions between the ASP residues of one monomer and LYS and ARG residues of the other monomer on the binding surface.

The monomer orientation in the crystal structure of the C-terminal domain's dimer also provides antiparallel dipole moments at pH 6, (d-d ~ -1). However, as we have previously shown that the C-terminal domain dipole is insufficient to achieve this alignment, implying a cooperation with the hydrophobic anisotropy. This configuration, indeed, maximizes the hydrophobic interactions by providing a full correlation between the first and the third largest patches (h11$\sim$ h33 $\sim$ -1) resulting in exclusion of these patches from the surface upon dimerization.

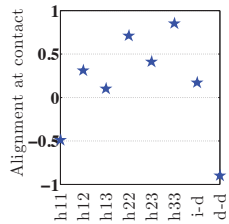
It has been shown that C-terminal dimerization involves a disulfide bond between the monomers and a tweezer like interactions between the helix 1 of chain A and helix 5 of the chain B.[26] Since the tweezer like entanglement is inaccessible without unfolding the monomer our rigid protein model is inadequate to study its dimerization and thereby, we only present a dimerization study on the N-terminal domain.



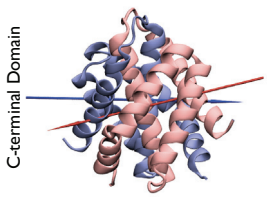
(a) Dipole alignment at pH 6 in crystal structure



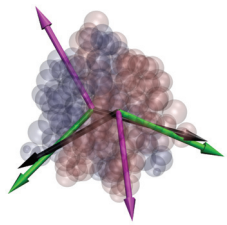
(b) Patch alignment in crystal structure



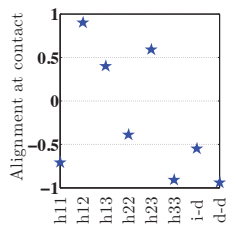
(c) Patch and dipole alignment in crystal structure



(d) Dipole alignment at pH 6 in crystal structure



(e) Patch alignment in crystal structure



(f) Patch and dipole alignment in crystal structure

Figure 3: Hydrophobic and dipole moment alignment of dimers in the crystal structures. Chain A and B are shown in blue and pink, respectively. h_{ij} is the scalar product of i^{th} patch vector of chain A and j^{th} patch vector of chain B. i-d is ion-dipole alignment and d-d is dipole-dipole alignment.

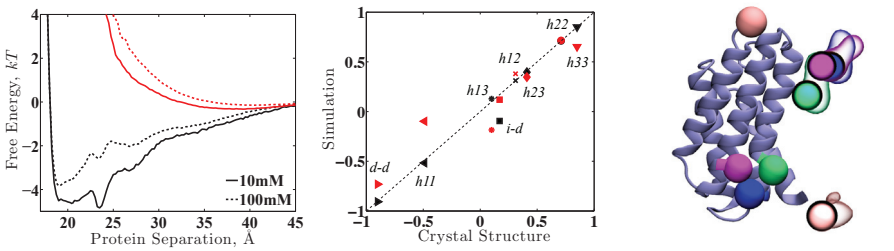
Dimerization study of the N-terminal Domain in salt solution

Effect of hydrophobic interactions

Experiments show that the N-terminal domain forms a stable dimer at pH 6 and low salt concentrations which is diminished by elevated pH and salt,[27,35] hinting the importance of electrostatics. We first performed dimerization simulations at pH 6 in 10 mM salt solution by considering only electrostatic and VDW interactions. We find that the affinity without hydrophobic interactions is insufficient to form stable dimers. The dimers obtained in the simulation show a random hydrophobic patch correlation, $h_{ij} \approx 0$ (Figure 4). However, with an additional short range attraction, the N-terminal domain forms strongly bound dimers where the monomer orientation *completely*

matches the crystal structure within the thermal fluctuations – see Figure 4b and c. This indicates that the hydrophobic and electrostatic interactions cooperate constructively to form stable dimers.

This is also confirmed by the experiments that the mutation of a centrally located hydrophobic residue on the binding surface to a negatively charged ARG (A72R), diminishes dimer formation.[30]



(a) Free energy (b) Patch vector alignment (c) Isodensity with hydrophobic interactions

Figure 4: (a) The interaction free energy of the N-terminal domain with (black) and without (red) hydrophobic attraction. (b) The alignment of hydrophobic patch and dipole vectors at pH 6 and in 10 mM NaCl solution with (black) and without (red) hydrophobic attraction. (c) Preferred locations of HIS6 (pink transparent bubble) ASP40 (purple transparent bubble), GLU79 (green transparent bubble) and GLU84 (blue transparent bubble) of chain B in the simulated dimers and in the crystal structure (shaded spheres with corresponding colors). The chain A in the crystal and in the simulated dimers are located on top of each other and depicted by a cartoon and solid ball representations, respectively.

Effect of salt concentration and pH

In line with the experimental observations,[27,35] the self association of the N-terminal domain is predicted at pH 6 which is then reduced by elevated pH and salt concentrations – see Table 7. Moreover, the dimer formation is diminished at pH 2 due to the protonation of negative ASP and GLU residues.

Figure 5 shows that free energy of interactions have two minima at low salt concentrations, revealing two modes of dimerization. The minimum at 23 Å monomer separation maximizes the electrostatic attraction while the one at 18 Å separation maximizes the hydrophobic and VDW at-

Table 7: Difference in the calculated binding free energies of the N-terminal domain in kcal/mol as a function of salt and pH, with respect to the affinity at pH 6 and 10 mM ($\Delta G - \Delta G_{\mathrm{pH6,10mM}}$).

Salt	pH 6	pH 8
10 mM	0	1.26
100 mM	0.69	1.16

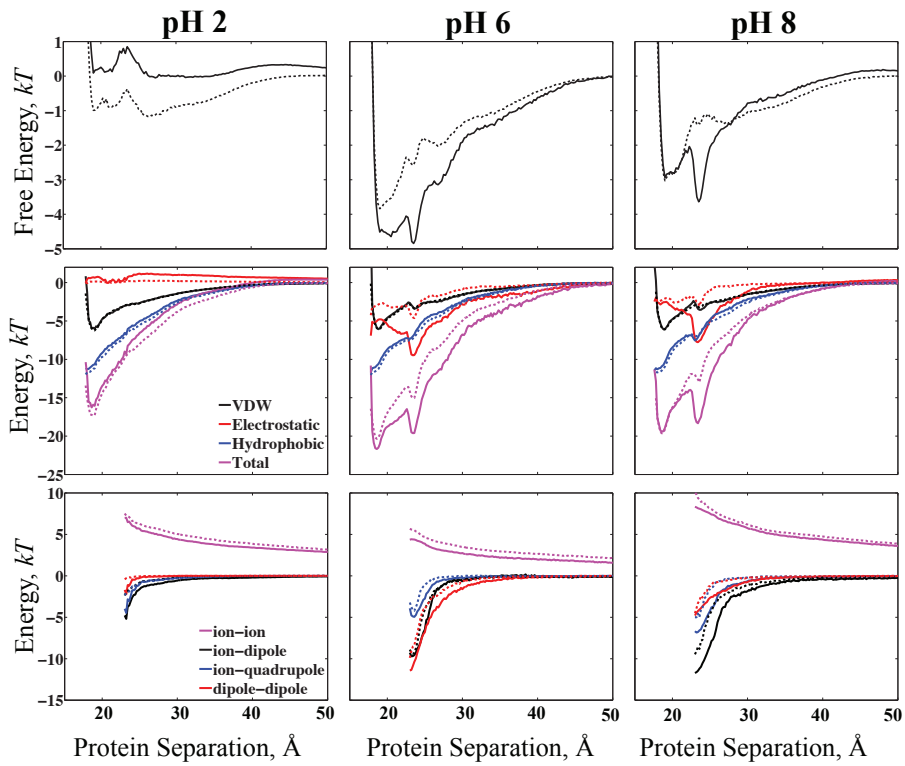


Figure 5: Association of the N-terminal domain as a function of monomer-monomer separation in 10 (solid lines) and 100 mM (dotted lines) salt solutions.

tractions. However, at high salt concentration, hydrophobic interactions dominate over the screened electrostatics and diminish the minimum originated from electrostatics. Note that even for the minimum at 23 Å monomer separation, electrostatic and hydrophobic interactions cooperate together and makes almost equal contributions to the binding energy.

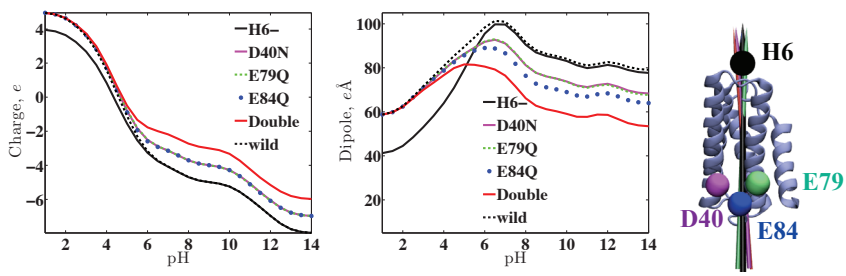
As seen in Figure 5 (row3), increasing pH from 6 to 8, mainly affects the dipole-dipole inter-

actions and together with the enhanced net charge repulsion leads to a reduced binding affinity. We have shown that the dipole moments and the net charges at pH 6 and pH 7 are similar. This implies that the pH effect on the charge distribution will not be pronounced in the association of the N-terminal domain.

Effect of mutations

In order to quantify the dipole moment effect on the binding affinity, we simulated dimerization of three single, D40N, E79Q, E84Q; one double D40NE84Q; and one deletion mutation H6-. As seen in the Figure 6, all mutations, except deletion of H6, reduce the dipole moment of the N-terminal domain at pH 6 while preserving its direction. The most dramatic reduction (25%) is caused by the double mutation D40NE84Q due to the simultaneous elimination of the two charges from the negative end. Calculated dipole moments of the mutants at pH 6 are $\mu^{wt} \approx \mu^{H6} > \mu^{E79} = \mu^{D40N} > \mu^{E84} > \mu^{D40NE84Q}$.

The calculated dipole moments correlate well with the calculated binding constants, $K^{wt}_{\text{dimer}} > K^{E79}_{\text{dimer}} > K^{E84}_{\text{dimer}} > K^{D40N}_{\text{dimer}} = K^{H6}_{\text{dimer}} > K^{D40NE84Q}_{\text{dimer}}$, revealing a direct proportionality between the dipole moment and the binding affinity (Table 8, Figure 7). Moreover, the predicted binding constants are in a *semi-quantitatively* agreement with experimental observations.[29,35]



(a) Charge (b) Dipole (c) Mutated residues and their dipoles

Figure 6: Electrostatic properties of mutated N-terminal domains as a function of pH at 10 mM salt concentration. Each mutation is depicted with a color that are seen in the legends and on the residues.

Table 8: Difference in the binding constants of mutated N-terminal domains in kcal/mol with respect to the wild type affinity at corresponding pH ($\Delta\Delta G = \Delta G^{mut} - \Delta G^{wt}$).

	pH	H6-	D40N	E79Q	E84Q	Double	wild
Simulation	**6**	0.35	0.35	0.19	0.21	0.85	0
	8	-0.04	0.51	0.05	0.24	0.65	0
Experiment[a 29]	**6**	1.0±0.2[b]	n.a	0.3±0.2	0.6±0.1	n.a.	0

[a]Experiments are performed at 50 mM ionic strength [b] HIS6 replaced with ALA residue instead of deletion. n.a. not available

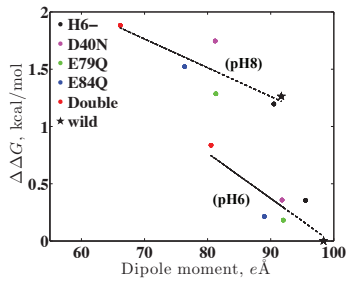
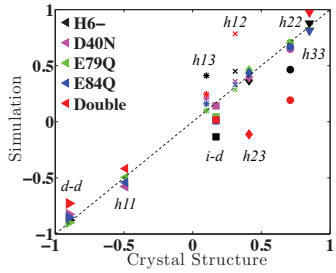


Figure 7: $\Delta\Delta G$ as a function of dipole moment of N-terminals with various mutations at 10 mM salt concentration.

The double mutation D40NE84Q which has the weakest dipole moment causes the most dramatic reduction in the electrostatic attraction (Figure 9), resulting in a significantly suppressed binding constant. Note that the mutation on ASP residue (D40N) has a more pronounced effect on the binding affinity than the mutation on GLU79 due to the broken salt bridge attractions between the ASP40 and ARG60/LYS65. Although the dipole moment seems insensitive to H6- mutation when pH is higher than 5, an affinity reduction, similar to D40N mutation, is observed at pH 6 but not at pH 8 – see Figure 9 and Table 8. This is because H6 has an ability to charge itself up as a response of negative charges at the close vicinity when solution pH *matches* its pK_a (≈ 6). Hence, it contributes to the dipole moment and provides an additional attraction only at pH 6 and in the vicinity of negative charges but neither at pH 8 nor in the bulk alone without any nearby negative charges.

In order to determine the monomer orientations in the mutated dimers, we analyzed the align-



(a) Alignments

Figure 8: (a) Hydrophobic and dipole correlations of the N-terminal domains' mutants at pH 6 and 10 mM salt concentration compared with the crystal structure of the wild type. Mutations are depicted by colors and the vectors are depicted with symbols.

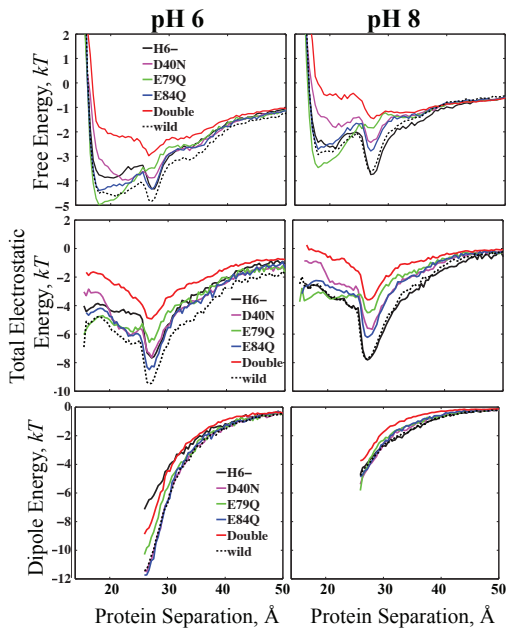


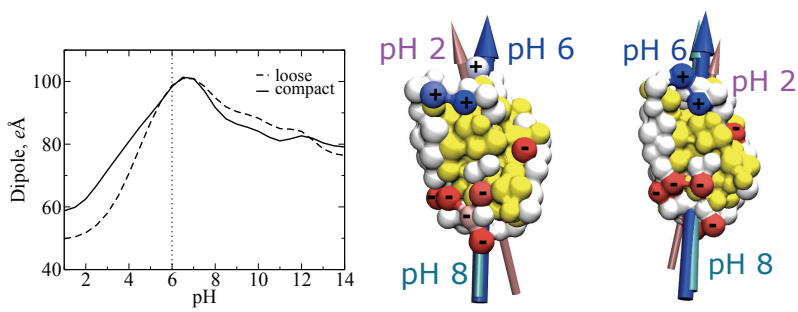
Figure 9: Association of the N-terminals domains mutants at 10 mM salt concentration.

ment of the hydrophobic patch vectors when two monomers were at contact – see Figure 8. A significant deviation from the wild type dimer structure is observed for the double mutation where

only the $h11$ alignment on the binding surface is preserved. This indicates a conserved binding surface with weakly defined monomer orientations due to the diminished dipolar interactions and the salt bridges. The monomer orientation in the N-terminal domain dimers is insensitive to single mutations, except H6, suggesting that charge regulation effect of HIS6 is important for achieving orientations as in the crystal structure. Note that our rigid protein model neglects any possible conformational changes in the wild type caused by the mutations and hence reflects *purely* electrostatic effects caused by the mutations.

Effect of conformational change

It has been recently shown that the N-terminal domain adopts two different structures depending on the environmental pH.[36] At pH 6, it obtains a compact conformation which is identical to the crystal structure. However, at pH 7, it prefers a loose conformation by pushing a helix bundle on the dimerization site outwards. This distorts the binding surface in a way that the helix bundle separates some of the hydrophobic residues from the charged ones.(Figure 10). Since this change



(a) Dipole moment (b) compact structure (c) loose structure

Figure 10: a) The dipole moment of the compact and loose conformation of N-terminal domain. b) and c) are the compact and loose conformation of the protein, respectively.

has no significant influence on the charge distribution, it maintains the direction and the strength of the dipole moment. To determine the effect of this conformational change on the binding affinity of the N-terminal domain, we investigated the dimerization of the loose conformation and compared

it with the compact one. We chose to simulate the dimerization of the loose conformation at pH 6 and low salt concentrations to ensure that environmental conditions are right for a possible dimerization.

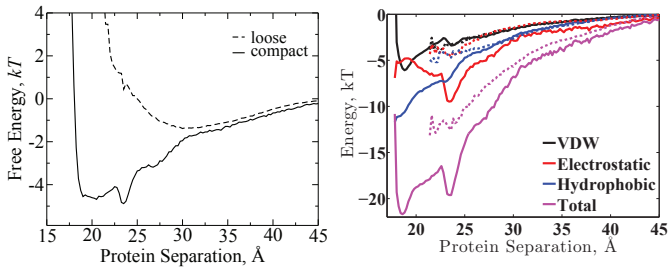


Figure 11: Free energy of interactions with compact (solid) and loose structure (dashed) of N-terminal domain at pH 6.

As seen in Figure 11, the attraction between the monomers is significantly suppressed when the loose conformation is adopted, resulting in a lower binding affinity, $\Delta\Delta G = 1.68$ kcal/mol. This suppression originates from the reduction of both hydrophobic and electrostatic energies, implying a destructive cooperation of the hydrophobic and electrostatic anisotropies. Note that our model is unable to capture free energy cost of the conformational changes since it models both conformations as a separate rigid structure without any flexibility and it does not allow to switch between the conformations during a simulation.

Conclusion

We have here, presented a dimerization study of terminal domains of silk proteins to understand the balance between the electrostatic and hydrophobic interactions as a function of pH and salt concentration. We have shown that in the dimerization process of the N-terminal domain, the electrostatic and hydrophobic interactions cooperate together where the dimerization proceeds via electrostatic dipole interactions, forcing an antiparallel monomer orientation, followed by specific hydrophobic interactions induced locking of the dimer. In line with the experimental observations,

the strongest monomer binding is predicted at pH 6 and low salt concentrations. Elevated salt concentrations and pH suppress the electrostatic dipole attraction and lead to a weaker binding. The mutation study on the dimerization have shown that the binding constant is proportional to the dipole moment. However, we have also predicted that the pH effect on the dipole moment of the N-terminal domain is insufficient to prevent the dimer formation at pH 7. This indicates that the hindered dimerization at pH 7 is not because of the altered charge distribution but because of the adopted loose conformation. We have shown that this loose conformation destroys the constructive cooperation between the electrostatic and hydrophobic interactions and significantly reduces the binding affinity. Interestingly, the analysis on the surface properties of the C-terminal domain monomer has indicated a strong influence of hydrophobic interactions on the assembly rather than electrostatics. The large hydrophobic surface of the C-terminal domain and the possible disulphide bond between the monomers support the fact that the C-terminal is dimerized during protein secretion in the spider gland.

References

(1) Knight, D.; Vollrath, F. *Naturwissenschaften* **2001**, *88*, 179–182.

(2) Craig, C. L. *Spiderwebs and silks: tracing evolution from molecules to genes to phenotypes*; Oxford University Press: New York, 2003.

(3) Porter, D.; Guan, J.; Vollrath, F. *Adv. Mater.* **2013**, *25*, 1275–9.

(4) Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L. *Biomaterials* **2003**, *24*, 401–16.

(5) Vollrath, F.; Selden, P. *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 819–846.

(6) Vollrath, F.; Porter, D.; Holland, C. *MRS Bull.* **2013**, *38*, 73–80.

(7) Vollrath, F. *Rev. Mol. Biotechnol.* **2000**, *74*, 67–83.

(8) Vollrath, F.; Porter, D. *Soft Matter* **2006**, *2*, 377.

(9) Vollrath, F.; Porter, D.; Holland, C. *Soft Matter* **2011**, *7*, 9595.

(10) Riekel, C.; Madsen, B.; Knight, D.; Vollrath, F. *Biomacromolecules* **2000**, *1*, 622–6.

(11) van Beek, J. D.; Hess, S.; Vollrath, F.; Meier, B. H. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10266–71.

(12) Xu, M.; Lewis, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 7120–4.

(13) Guerette, P. A.; Ginzinger, D. G.; Weber, B. H.; Gosline, J. M. *Science* **1996**, *272*, 112–115.

(14) Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. *Int. J. Biol. Macromol.* **1999**, *24*, 271–5.

(15) Andersen, S. O. *Comp. Biochem. Physiol.* **1970**, *35*, 705–711.

(16) Knight, D.; Vollrath, F. *Tissue cell* **1999**, *31*, 617–20.

(17) Inoue, S.; Tanaka, K.; Arisaka, F.; Kimura, S.; Ohtomo, K.; Mizuno, S. *J. Biol. Chem.* **2000**, *275*, 40517–28.

(18) Jin, H.-J.; Kaplan, D. L. *Nature* **2003**, *424*, 1057–61.

(19) Donald, A. M.; Windle, A. H.; Hanna, S. *Liquid crystalline polymers*, 2nd ed.; Cambridge Universit Press: Cambridge, 1992.

(20) Nesloney, C. L.; Kelly, J. W. *Bioorg. Med. Chem.* **1996**, *4*, 739–766.

(21) Dicko, C.; Kenney, J. M.; Vollrath, F. *Adv. Protein Chem.* **2006**, *73*, 17–53.

(22) Vollrath, F.; Knight, D. P.; Hu, X. W. *Proc. R. Soc. Lond. B* **1998**, *265*, 817–820.

(23) Dicko, C.; Kenney, J. M.; Knight, D.; Vollrath, F. *Biochemistry* **2004**, *43*, 14080–7.

(24) Dicko, C.; Vollrath, F.; Kenney, J. M. *Biomacromolecules* **2004**, *5*, 704–10.

(25) Askarieh, G.; Hedhammar, M.; Nordling, K.; Saenz, A.; Casals, C.; Rising, A.; Johansson, J.; Knight, S. D. *Nature* **2010**, *465*, 236–8.

(26) Hagn, F.; Eisoldt, L.; Hardy, J. G.; Vendrely, C.; Coles, M.; Scheibel, T.; Kessler, H. *Nature* **2010**, *465*, 239–42.

(27) Hagn, F.; Thamm, C.; Scheibel, T.; Kessler, H. *Angew. Chem., Int. Ed.* **2011**, *50*, 310–3.

(28) Wallace, J.; Shen, J. *J. Phys. Cham. Lett.* **2012**, *3*, 658–662.

(29) Schwarze, S.; Zwettler, F. U.; Johnson, C. M.; Neuweiler, H. *Nat. Commun.* **2013**, *4*, 2815.

(30) Kronqvist, N. et al. *Nat. Commun.* **2014**, *5*, 3254.

(31) Silvers, R.; Buhr, F.; Schwalbe, H. *Angew. Chem. Int. Ed.* **2010**, *49*, 5410–2.

(32) Eisoldt, L.; Thamm, C.; Scheibel, T. *Biopolymers* **2012**, *97*, 355–61.

(33) Humphrey, W.; Dalke, a.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–8, 27–8.

(34) Bostro, M.; Williams, D. R. M.; Ninham, B. W. *Langmuir* **2001**, 4475–4478.

(35) Landreh, M.; Askarieh, G.; Nordling, K.; Hedhammar, M.; Rising, A.; Casals, C.; Astorga-Wells, J.; Alvelius, G.; Knight, S. D.; Johansson, J.; Jörnvall, H.; Bergman, T. *J. Mol. Biol.* **2010**, *404*, 328–36.

(36) Jaudzems, K.; Askarieh, G.; Landreh, M.; Nordling, K.; Hedhammar, M.; Jörnvall, H.; Rising, A.; Knight, S. D.; Johansson, J. *J. Mol. Biol.* **2012**, *422*, 477–87.

Paper II

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

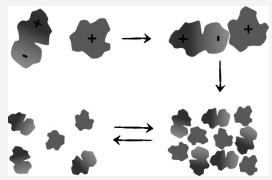
Anisotropic Interactions in Protein Mixtures: Self Assembly and Phase Behavior in Aqueous Solution

Anıl Kurut,* Björn A. Persson, Torbjörn Åkesson, Jan Forsman, and Mikael Lund*

Department of Theoretical Chemistry, Lund University, POB 124 SE-22100 Lund, Sweden

Supporting Information

ABSTRACT: Recent experimental studies show that oppositely charged proteins can self-assemble to form seemingly stable microspheres in aqueous salt solutions. We here use parallel tempering Monte Carlo simulations to study protein phase separation of lysozyme/α-lactalbumin mixtures and show that anisotropic electrostatic interactions are important for driving protein self-assembly. In both dilute and concentrated protein phases, the proteins strongly align according to their charge distribution. While this alignment can be greatly diminished by a *single* point mutation, phase separation is completely suppressed when neglecting electrostatic anisotropy. The results highlight the importance of subtle electrostatic interactions even in crowded biomolecular environments where other short-ranged forces are often thought to dominate.



SECTION: Biophysical Chemistry

In a set of experimental studies, it has been shown that 1:1 mixtures of lysozyme (lys) and α-lactalbumin (α-lac) self-assemble into well-defined micrometer-sized spheres.[1−3] This process is highly dependent on the salt concentration, and microspheres are formed only for the calcium depleted *apo* form of α-lac at salt concentrations below 100 mM, while for the calcium loaded *holo* form of α-lac, the self-assembly vanishes. Contrary to phase separation of similarly charged proteins,[4,5] lys and apo α-lac phase separate due to attractive electrostatic interactions between the two oppositely charged proteins. Interestingly, this electrostatic interaction is far from isotropic, and it has recently been shown that the anisotropic charge distribution of α-lac, combined with the high positive net charge of lys gives rise to strong directional protein−protein interactions.[6,7] The remaining question is how such long-ranged, noncentrosymmetric interactions from charged residues influence protein assembly and phase transition? We here present a mesoscopic simulation study of aqueous protein mixtures where we investigate the contribution of anisotropic protein−protein interactions on phase transition of proteins as well as on microstructuring in dense protein environments.

Previous studies on anisotropic protein−protein interactions[8−14] typically describe proteins as patchy spherical particles with attractive sites interacting via angularly dependent square well potentials. These models rely mainly on ambiguously placed attractive patches and not on anisotropy owing to nonuniform surface charge distributions of real proteins. One exception is the embedded discrete charge model[9,11,15] where charged residues are projected onto a hard sphere. Although approximately accounting for anisotropic electrostatics, this model treats the excluded volume as spherical symmetric. To remedy this, here we develop a protein model that captures anisotropic interactions due to both excluded volume and electrostatic interactions, while at the same time being fast enough for simulation studies of phase transitions involving *many* protein molecules.

Simulating phase behavior at an atomistic level is unfeasible due to the vast number of solvent and solute molecules in the system. We therefore gradually coarse-grained the proteins (see illustration in Figure 1) while maintaining (i) the salt-dependent potential of mean force between lys/α-lac and (ii) the distance-dependent average alignment of α-lac's dipole moment with respect to lys. The latter unitless property ranges between +1 and −1, where +1 corresponds to a fully aligned dipole. The final model, (c), maintains all charged residues as in the crystal structure, while the remaining part is represented by a central particle. As shown in the Supporting Information, this model closely reproduces the well-tested reference model (a).

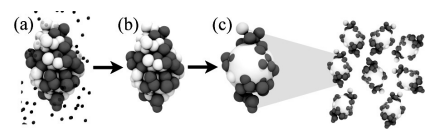


Figure 1. Coarse graining starting from an implicit solvent/explicit salt model where amino acids are represented by spheres (a).[6,16] From (a)→(b) Debye−Hückel electrostatics; (b)→(c) neutral residues are collected into a central sphere. See Supporting Information for details.

Metropolis Monte Carlo simulations in the canonical ensemble, *NVT*, were used for two-protein interactions when coarse-graining model (a), while simulations in the isobaric−isothermal ensemble, *NpT*, were used to study many-protein

Received: December 22, 2011
Accepted: February 21, 2012

 © XXXX American Chemical Society

interactions and phase behavior with model (c). The system energy for a given configuration was $U = \sum_{i<j} A\varepsilon_{LJ}[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^{6}] + (e^2 z_i z_j/4\pi\varepsilon_0\varepsilon_r r_{ij})\zeta$, where ε_{LJ} is the depth of the Lennard-Jones potential, σ_i is the diameter, and $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. r_{ij} is the separation between particle i and j, z_i is the charge valency, e is the electron unit charge, and $\varepsilon_0\varepsilon_r$ is the dielectric permittivity of water. For explicit salt, $\zeta = 1$; for implicit salt, $\zeta = e^{-\kappa r_{ij}}$, where κ is the inverse Debye screening length. For NpT simulations, the histogram method was used to sample volume probability distributions, $P(V)$, related to the constrained Gibbs free energy, $G = -kT \ln P(V)$ where k is Boltzmann's constant.[17] Minima in G correspond to (meta-)stable phases, and to obtain coexistence concentrations, G was reweighted, $G' = G \pm \Delta pV$, such that two minima had equal free energies.[18] The Helmholtz free energy is $A = G' - pV$. We used parallel tempering in both p and κ to efficiently sample configurational space at low salt concentrations where strongly attached clusters form. For simulation details, see Supporting Information.

Using model (c), we performed NpT simulations with 40 proteins to investigate the Helmholtz free energy of bulk 1:1 protein mixtures containing lys and four different forms of α-lac: apo, holo, smeared, and mutated apo (see Table 1 and Figure 2). Phase separation, indicated by a common tangent of two points in the Helmholtz free energy curve, was observed only for mixtures containing the apo forms and at low salt concentrations. This is in perfect agreement with experimental observations.[2] Increasing the salt concentration leads to reduced electrostatic interactions between the oppositely charged lys and α-lac, and eventually suppresses phase separation. The same is the case for holo α-lac, where the bound calcium ion reduces the net charge of α-lac.

To identify the contribution of anisotropic electrostatics to phase separation, we artificially reduced the α-lac dipole moment from 78 to 34 D by swapping merely two charged amino acids (Table 1). As seen in Figure 3, this mutation, which does not alter the net charge, drastically decreases the critical salt concentration at coexistence and widens the immiscibility gap of lys/apo α-lac. We also approached isotropic electrostatics by smearing out the total charge of apo α-lac on *all* titratable residues and, as seen in Figure 2, lack of directional interactions completely obstructs phase separation in the investigated salt concentration range.

Table 1. Model Details for Lys and α-Lac[a]

	α-lac				
protein	apo	holo	mutated[b]	smeared[c]	lys
PDB entry	1F6R	1F6S	1F6R	1F6R	4LZT
$N_{residues}$	124	125	124	124	130
net charge, pH 7.5	−5.9	−4.5	−5.9	−5.9	7.0
dipole moment[d] (D)	78	68	34	3	20

[a]For all particles, ε_{LJ} =0.05 kT in models (a) and (b), ε_{LJ} = 0.075 kT in model (c). σ_i = 30 Å for the central sphere (see ref 6 for σ of other particles). [b]Residues D79 and K99 are swapped. [c]Net charge is uniformly smeared over all titratable sites. [d]Calculated with respect to the protein mass center.

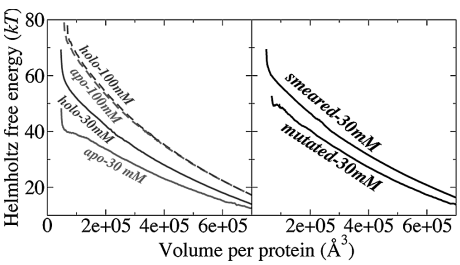


Figure 2. Helmholtz free energy for 1:1 mixtures of lys with apo, holo, mutated, and smeared α-lac at 30 mM (solid lines) and 100 mM (dashed lines) salt concentrations.

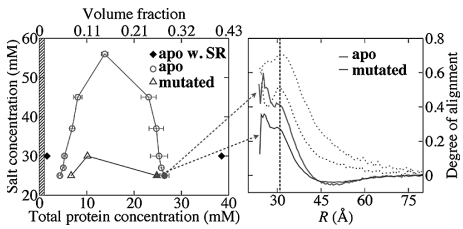


Figure 3. Left: Fluid−fluid coexistence curve of lys and α-lac in different forms and with extra short-ranged attraction (SR). The shaded area shows the experimentally observed concentration range of the dilute phase; error bars reflect the standard deviation of five independent Monte Carlo runs. Right: α-lac dipole moment alignment with respect to lys (25 mM salt) as a function of protein separation in the dense phases (solid) and at infinite dilution (dotted); the vertical line indicates the position of the $w(r)$ minima.

It is instructive to follow the alignment of α-lac's dipole moment with respect to the mass center of lys (see Figure 3, right): at infinite protein dilution, the apo form of α-lac is aligned by more than 70% when in contact with lys and then levels off to zero when the proteins separate. As expected, by diminishing the dipole moment with a point mutation, this alignment is reduced. A similar behavior is seen in the dense protein phases but to a lesser degree due to the surrounding proteins. Still, the apo form of α-lac is aligned by ∼40%, underlining that anisotropic electrostatics are involved in the dense phase structuring.

Experimentally, microsphere formation occurs at low salt and at total protein concentrations between 0.075 and 1.05 mM,[2] implying that the equilibrium concentration of the dilute phase is in the sub-millimolar range. Our model predicts a dilute phase concentration 5−70 times higher, and thus underestimates the width of the coexistence curve. Similar issues have been reported for lys.[11,19] One reason could be the lack of charge regulation in our model, although this is probably of minor importance.[6,20] It has also been proposed that apo α-lac undergoes conformational changes into a semifolded molten globule state upon binding to lys, leading to exposure of hydrophobic residues.[3] Such a mechanism, which will be temperature dependent,[8,19] likely leads to a short-range attraction unaccounted for by our rigid protein model. Although quantitative agreement is not the chief goal of this work, we attempted to remedy the above discrepancy by pragmatically incorporating a square well potential (depth 1 kT,

width 3.1 Å, corresponding to a water layer) between the neutral spheres in model (c). This spherically symmetric potential is justified by lack of angular correlations in the hydrophobic interactions for the present system (see Supporting Information). Figure 3 shows that this simple addition significantly improves the agreement with the experimental dilute phase density. Further, the simulated two-body potential of mean force, $w(R)$, is related to the thermodynamic dissociation constant, $K_d^{-1} \approx 4\pi \int_{\text{contact}}^{\infty} (e^{-w(R)/kT} - 1) R^2 \, dR$.[21] Without short ranged attraction, K_d = 737 μM at 39 mM salt and pH 7.5, merely 32% off the experimental value of 490 ± 70 μM.[7] Including the above square well potential, we obtained K_d = 446 μM. Thus, the short-ranged attraction improves agreement with the experimental phase behavior as well as consistency at the two-protein level. Note that our reference model (a) was originally used to reproduce lys–lys osmotic second virial coefficients[16] and that we have made no efforts to fit it to experimental data for the current lys/α-lac system.

We have shown that anisotropic electrostatic interactions resulting from a nonuniform distribution of amino acid charges play an important role in protein self-assembly. Due to the high charge anisotropy of α-lac, this protein aligns in the electric field of lys, both in dilute and concentrated protein environments, showing that directional electrostatics are influential even at crowded conditions. Phase separation is sensitive to subtle charge changes, and a single amino acid point mutation that minimizes the protein dipole moment of α-lac, while retaining the net charge, is enough to significantly shift the immiscibility gap to lower salt concentrations. This dipole moment reduction changes the close contact alignment and, consequently, the structure of the dense phase. Artificially smearing out the charge over the protein surface, and thus approaching isotropic electrostatics, completely inhibits phase separation. These results show that phase behavior of aqueous protein mixtures is sensitive to anisotropic electrostatics, both qualitatively and quantitatively, and that inclusion of directionality in theoretical models is important for accurate prediction of coexistence curves.

■ ASSOCIATED CONTENT

Supporting Information

Simulation details and model development. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: anil.kurut@teokem.lu.se (A.K.); mikael.lund@teokem.lu.se (M.L.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Linnaeus Center Organizing Molecular Matter and the eSSENCE program at Lund University. Computational resources were provided by LUNARC.

■ REFERENCES

(1) Nigen, M.; Croguennec, T.; Renard, D.; Bouhallab, S. Temperature Affects the Supramolecular Structures Resulting from α-Lactalbumin–Lysozyme Interaction. *Biochemistry* **2007**, *46*, 1248–1255.

(2) Nigen, M.; Croguennec, T.; Bouhallab, S. Formation and Stability of α-Lactalbumin–Lysozyme Spherical Particles: Involvement of Electrostatic Forces. *Food Hydrocolloids* **2009**, *23*, 510–518.

(3) Nigen, M.; Le Tilly, V.; Croguennec, T.; Drouin-Kucma, D.; Bouhallab, S. Molecular Interaction between Apo or Holo α-Lactalbumin and Lysozyme: Formation of Heterodimers as Assessed by Fluorescence Measurements. *Biochim. Biophys. Acta* **2009**, *1794*, 709–715.

(4) Taratuta, V. G.; Holschbach, A.; Thurston, G. M.; Blankschtein, D.; Benedek, G. B. Liquid–Liquid Phase Separation of Aqueous Lysozyme Solutlons: Effects of pH and Salt Identity. *J . Phys. Chem.* **1990**, *94*, 2140–2144.

(5) Shukla, A.; Mylonas, E.; Di Cola, E.; Finet, S.; Timmins, P.; Narayanan, T.; Svergun, D. I. Absence of Equilibrium Cluster Phase in Concentrated Lysozyme Solutions. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5075–80.

(6) Persson, B. A.; Lund, M. Association and Electrostatic Steering of α-Lactalbumin and Lysozyme Heterodimers. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8879–8885.

(7) Salvatore, D. B.; Duraffourg, N.; Favier, A.; Persson, B. A.; Lund, M.; Delage, M.-M.; Silvers, R.; Schwalbe, H.; Croguennec, T.; Bouhallab, S.; Forge, V. Investigation at Residue Level of the Early Steps during the Assembly of Two Proteins into Supramolecular Objects. *Biomacromolecules* **2011**, *12*, 2200–2210.

(8) Lomakin, A.; Asherie, N.; Benedek, G. B. Aleolotopic Interactions of Globular Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9465–9468.

(9) Carlsson, F.; Malmsten, M.; Linse, P. Monte Carlo Simulations of Lysozyme Self-Association in Aqueous Solution. *J. Phys. Chem. B* **2001**, *105*, 12189–12195.

(10) Kern, N.; Frenkel, D. Fluid–Fluid Coexistence in Colloidal Systems with Short-Ranged Strongly Directional Attraction. *J. Chem. Phys.* **2003**, *118*, 9882–9889.

(11) Rosch, T. W.; Errington, J. R. Investigation of the Phase Behavior of an Embedded Charge Protein Model through Molecular Simulation. *J. Phys. Chem. B* **2007**, *111*, 12591–12598.

(12) Liu, H.; Kumar, S. K.; Sciortino, F. Vapor–Liquid Coexistence of Patchy Models: Relevance to Protein Phase Behavior. *J. Chem. Phys.* **2007**, *127*, 084902–1–084902–5.

(13) Wentzel, N.; Gunton, J. D. Effect of Solvent on the Phase Diagram of a Simple Anisotropic Model of Globular Proteins. *J. Phys. Chem.* **2008**, *112*, 7803–7809.

(14) Gögelein, C.; Nägele, G.; Tuinier, R.; Gibaud, T.; Stradner, A.; Schurtenberger, P. A Simple Patchy Colloid Model for the Phase Behavior of Lysozyme Dispersions. *J. Chem. Phys.* **2008**, *129*, 085102-1–085102-12.

(15) Abramo, M. C.; Caccamo, C.; Costa, D.; Pellicane, G.; Ruberto, R. Molecular Dynamics of an Embedded-Charge Model of Lysozyme Aqueous Solutions. *J. Phys. Chem. B* **2010**, *114*, 9109–9118.

(16) Lund, M.; Jönsson, B. A Mesoscopic Model for Protein–Protein Interactions in Solution. *Biophys. J.* **2003**, *85*, 2940–2947.

(17) Hunter, J. E.; Reinhardt, W. P. Finite-Size Scaling Behavior of the Free Energy Barrier between Coexisting Phases: Determination of the Critical Temperature and Interfacial Tension of the Lennard-Jones Fluid. *J. Chem. Phys.* **1995**, *103*, 8627–8637.

(18) Brukhno, A. V.; Åkesson, T.; Jönsson, B. Phase Behavior in Suspensions of Highly Charged Colloids. *J. Chem. Phys.* **2009**, *113*, 6766–6774.

(19) Biesheuvel, P. M.; Lindhoud, S.; de Vries, R.; Cohen Stuart, M. A. Phase Behavior of Mixtures of Oppositely Charged Nanoparticles: Heterogeneous Poisson–Boltzmann Cell Model Applied to Lysozyme and Succinylated Lysozyme. *Langmuir* **2006**, *22*, 1291–300.

(20) Biesheuvel, P. M.; Lindhoud, S.; Cohen Stuart, M.; de Vries, R. Phase Behavior of Mixtures of Oppositely Charged Protein Nanoparticles at Asymmetric Charge Ratios. *Phys. Rev. E* **2006**, *73*, 041408.

(21) Wennerström, H. *Statistical Mechanical Description of Surfactant Self-assmbly. In Organized Solutions: Surfactants in Science and Technology*; Marcel Dekker: New York, 1992.

Paper III

Solution electrostatics beyond pH: a coarse grained approach to ion specific interactions between macromolecules

Anıl Kurut* and Mikael Lund*

Received 13th April 2012, Accepted 10th May 2012
DOI: 10.1039/c2fd20073b

Oblivious to ion specificity, pH has been a key parameter for macromolecular solutions for little more than a century. We here widen the concept by describing the ionization of macromolecules not only *via* pH, but also pX where X are other binding species. Using binding constants, measured by NMR, of chloride and thiocyanate to amino acid motifs on γ-crystallin, we calculate i) titration curves as a function of pH and pX and ii) estimate second virial coefficients using both approximate theory and computer simulations. In agreement with experiment, a Hofmeister reversal for protein–protein interactions is observed when crossing iso-electric conditions. Thiocyanate binding further leads to large charge fluctuations that may trigger intermolecular charge regulation interactions.

1 Introduction

For nearly a century, chemists and other natural scientists have been brought up with concepts such as pH, Brønsted acids and bases, auto-protolysis, and pK_a-values. This terminology focuses on thermodynamic equilibrium processes of merely a single ionic species, namely *protons*. pH is hence a key parameter throughout a number of scientific disciplines and is commonly used as a convenient handle to control inter- and intra-molecular electrostatic interactions. Protons are, however, not alone. Plenty of other ions may be present, yet our standard terminology is unable to account for those. During the last two decades there has been a large development in the molecular understanding of ion specific effects; the reader may refer to recent reviews for a full account.[1–3] It has for example long been known that proteins can be titrated with anions and that this changes the thermodynamic properties,[4–6] yet only recently have the underlying molecular mechanisms started to unravel. Experimental and theoretical works both show that small and large anions distinctly distribute on molecular surfaces; large anions bind to non-polar motifs[7–11] and thiocyanate, for example, binds to the backbone of model peptides.[12] The binding affinities to molecular motifs are currently being systematically quantified[12,13] and an obvious question is if specific ion binding can be treated at the same level as protons have been for so long. That is, through the chemical potential (*cf.* pH) and binding equilibrium constant to molecular motifs (*cf.* pK_a values) we may reduce the microscopic ion-motif potential of mean force to a two state binding model:

$$MX \rightleftharpoons M + X \tag{1}$$

where M is the motif and X is the ion in question. Arguably, protons distinguish themselves from other ions in that they are always present in water and their binding

Department of Theoretical Chemistry, Lund University, P.O.B. 124, SE-22100 Lund, Sweden. E-mail: anil.kurut@teokem.lu.se; mikael.lund@teokem.lu.se; Tel: +46 462 223167

This journal is © The Royal Society of Chemistry 2013

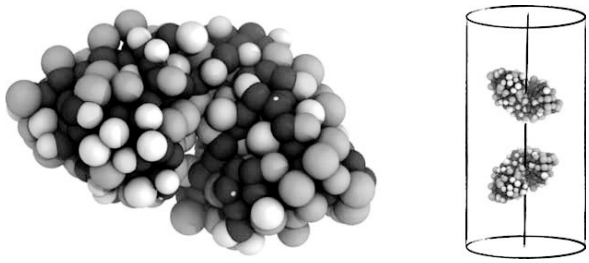


Fig. 1 Left: protein model used in the Monte Carlo simulations. Protons can bind to acidic and basic sites (turquoise); thiocyanate may bind to the backbone (black) as well as to hydrophobic side chains (yellow). Right: two-body simulation model in a cylindrical container – solvent and salt are handled implicitly.

is of covalent nature. This is fitting for a two-state model as the configurational space may readily be divided into a well defined bound and unbound state. This (arbitrary) division becomes more difficult for weakly bound ions, yet formally a thermodynamic binding constant can always be defined, provided that the ion-motif potential of mean force has a distinct minimum and decays with distance, r, as $1/r^3$ of faster.[14] The mere fact that a binding constant can be reliably measured, preferably with different experimental techniques, of course also supports a two-state model.

In this work we present a coarse grained model for macromolecules in solution, taking into account specific ion binding to dedicated molecular motifs. The model is based on continuum electrostatics, combined with two-state binding of protons and an arbitrary number of other ions. Ultimately this allows for studies of ion specific effects at large length scales hitherto unattainable with existing models.

2 Model and theory

2.1 Simulation model

Solvent and salt particles are treated at the Debye–Hückel level while the rigid proteins are described in mesoscopic detail where each amino acid is represented by two soft spheres: one for the backbone and another for the side chain – see Fig. 1. These spheres are placed at their respective mass centers as found in the atomistic NMR or crystal structure and their radii are determined from the atomic weight, assuming a density of 1 g ml^{-1}. To capture specific ion binding to the macromolecular surface we specify binding sites (Table 1) where the free energy difference between the bound and free state is $k_BT\ln 10(\mathrm{pX} - \mathrm{p}K_d)$. Analogous to pH, pX is the minus logarithm of the activity† of the binding ion, X, and K_d is the dissociation constant for the process. In addition to electrostatic and van der Waals pair interactions, hydrophobic side-chains interact with a square-well potential and the final system Hamiltonian becomes:

$$\beta U = \sum_{i\neq j}^{N} \frac{\lambda_B z_i z_j e^{-\kappa r_{ij}}}{r_{ij}} + \sum_{i\neq j}^{N} 4\beta\varepsilon_{LJ}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i\neq j}^{N_{\mathrm{hydroph.}}} \beta u_{sw} + \sum_{i\neq j}^{N_{\mathrm{bound}}} (\mathrm{pX}_i - \mathrm{p}K_{d,i})\ln 10 \quad (2)$$

where the first two sums run over all particles, the third over hydrophobic side chains and the last term over sites with bound ions, only. $\lambda_B = 7.1$ Å is the Bjerrum length

† In this work we assume that the activity is equal to the concentration.

 This journal is © The Royal Society of Chemistry 2013

Table 1 Amino acid properties and binding sites of γD-crystallin (PDB: 2KLJ[16]). Only particles with solvent accessible surface area (SASA) larger than 30 Å^2 are included as ion binding sites; this is indicated by parenthesis in the *N* column. When ions bind to a hydrophobic site, this is rendered non-hydrophobic. A probe with radius 1.4 Å is used in SASA measurements

Atom type	N	σ/Å	pK_{d,H^+}	$pK_{d,\ SCN^-}$ [12]	Hydrophobic
Asp	12 (10)	4.8	4.0		no
Glu	10 (9)	5.2	4.4		no
His	6 (4)	5.4	6.3		no
Tyr	14 (7)	5.9	9.6		no
Lys	1 (1)	5.2	10.4		no
Cys	6 (0)	4.5	10.8		no
Arg	21 (21)	5.8	12.0		no
Ctr	1 (1)	4.7	2.6		no
Ntr	1 (1)	4.7	7.5		no
Ala	4 (0)	3.1		0.82	yes/no
Ile	6 (0)	4.8		0.82	-//-
Leu	17 (3)	4.8		0.82	-//-
Met	4 (3)	5.2		0.82	-//-
Phe	6 (1)	5.6		0.82	-//-
Pro	4 (4)	4.3		0.82	-//-
Trp	4 (0)	6.3		0.82	-//-
Val	6 (0)	4.3		0.82	-//-
Ser	17	3.9			no
Thr	4	4.4			no
Asn	7	5.6			no
Gln	11	5.2			no
Gly	14	1.2			no
backbone	156 (5)	4.7		0.60	no
backbone (Pro, Gly)	18 (3)	4.7		1.30	no

for water, z are particle charge numbers, r_{ij} the distance between the i'th and j'th particle, $1/\kappa = 3.04/\sqrt{C_s}$ Å is the Debye screening length,[15] C_s the molar 1 : 1 salt concentration, $\beta = 1/k_B T$ is the inverse thermal energy, and $\beta u_{sw} = -0.8$ for surface separations less than 1.4 Å – zero otherwise. Finally, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ is the Lennard–Jones diameter and $\beta\varepsilon_{LJ} = 0.088$.

Using Metropolis Monte Carlo (MC) simulations[17] we simulate either one or two of the proteins in a cylindrical container and average over i) mass center positions (fixed along the main axis of the cylinder), ii) orientations, and iii) ion binding states. Translations and rotations for two-body simulations, only, are performed in a combined MC move where a random displacement vector and angle is used to generate configurations in the canonical ensemble. Averaging over ion binding states is done by swap moves where the state of randomly chosen binding sites – see Table 1 – is alternated between bound and unbound. "Bound" simply means that the site charge is changed by $+1e$ or $-1e$ for proton and anion binding, respectively. Ion binding to a hydrophobic group renders the group non-hydrophobic, thus excluding it from hydrophobic interaction. All simulations were performed using the Faunus project.[18]

2.2 Charge regulation

Charge binding sites on molecular surfaces give rise to a fluctuating molecular charge distribution that can be perturbed by an external electric potential. This leads to a correlation effect, known as charge regulation or "fluctuation forces", that lowers the system free energy.[19,20] The mechanism is well known for proton binding

sites, yet for other ions it is largely unexplored and we therefore give a brief overview.

Consider an unperturbed macromolecule with a fluctuating charge distribution with a charge valency, Z, that occurs with the intrinsic probability $P(Z)$. Exposing the molecule to an external electric potential, φ, due to the surrounding chemical environment, the charge ensemble average becomes,

$$\langle Z\rangle = \frac{\int P(Z)Z\exp(-\beta\varphi eZ)\mathrm{d}Z}{\int P(Z)\exp(-\beta\varphi eZ)\mathrm{d}Z}. \tag{3}$$

It then follows that,

$$-\frac{\partial\langle Z\rangle}{\beta e\partial\varphi} = \langle Z^2\rangle - \langle Z\rangle^2 \equiv C. \tag{4}$$

where we have introduced the charge capacitance, C, which is simply the variance of the mean charge.[21] If φ stems from another fluctuating charge distribution, $P(Z')$, the two distributions have the interaction free energy,

$$\beta w(R) = -\ln\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(Z)P(Z')\exp\left(-\lambda_B ZZ'/R\right)\mathrm{d}Z\mathrm{d}Z' \approx \frac{\lambda_B\langle Z\rangle\langle Z'\rangle}{R} - \frac{\lambda_B^2}{2R^2}\left(C\langle Z'\rangle^2 + C'\langle Z\rangle^2 + CC'\right) \tag{5}$$

where R is the center–center distance between the two distributions and in the last step we have assumed that P, P' are normal distributions. The same result can be obtained using a multipole expansion[20–22] and the first term is clearly the monopole–monopole interaction term, while the remaining attractive terms are monopole-induced and induced-induced monopole interactions. For the sake of simplicity we do not discuss angular dependent terms and further note that these decay faster than the above terms – $1/R^4$ and $1/R^6$ for monopole–dipole and dipole–dipole interactions, respectively.[23]

Thus, in the intermolecular interaction energy given by eqn (5), electrostatics are condensed to two molecular parameters, i) the average net charge, $\langle Z\rangle$ and ii) the charge capacitance, C. The latter property is central for describing charge regulation phenomena[21] and entails the overall fluctuation effect of all binding ions and thus depends on solution conditions, structure and protein sequence.

2.3 Osmotic second virial coefficient

Through eqn (5), the protein net charge and capacitance can be used to estimate the osmotic second virial coefficient of protein solutions using,

$$B_2 = \frac{2\pi\sigma^3}{3} - 2\pi\int_{\sigma}^{\infty}\left(e^{-\beta}w(R) - 1\right)R^2\mathrm{d}R \tag{6}$$

where the first term, B_2^{HS}, is a simple hard sphere contribution for σ smaller than protein–protein contact, $\sigma = 25$ Å. To take into account the effect of salt screening, λ_B in eqn (5) is replaced with[15]

$$\lambda_B\exp[-\kappa(R-\sigma)]/(1+\kappa\sigma) \tag{7}$$

where we note that the quadratic charge regulation terms are subject to stronger salt screening than the direct monopole–monopole term.

 This journal is © The Royal Society of Chemistry 2013

3 Results and discussion

3.1 Binding of protons, chloride, and thiocyanate

The overall electrostatic properties of bio-molecules are traditionally classified from the pH titration curve, *i.e.* the net charge as a function of the proton activity (pH). To describe how these properties are influenced by additional ions, further dimensions must be added to the titration curve: one for each additional binding species. In the following we assume that the salt cation does not bind to the protein surface, although this could easily be included by expanding Table 1; the computational cost would be negligible.

Fig. 2, top, shows the net charge of γ-crystallin in the presence of chloride and thiocyanate, respectively. Due to weak binding at moderate salt concentrations[5,12,24] we shall assume that chloride interacts only through double layer forces, that is through the Debye screening length. The iso-electric point, pI where the net-charge is zero, thus remains constant over a large span of salt concentrations and the classical one-dimensional pH picture thus gives a good description of the electrostatic properties. For pH ≠ pI the net charge does however change with salt concentration due to screening of intramolecular electrostatic interactions. At high ionic strengths the protein can hence accommodate a higher net-charge, while if $\langle Z \rangle = 0$, screening has little effect since the internal electrostatic energy, roughly proportional to $\langle Z \rangle^2$, is near a minimum.

In contrast, the strongly bound thiocyanate ion behaves very differently. Changing the salt concentration from 30 mM to 300 mM causes a drop in the iso-electric point from nine to five because the increased salt activity drives anion adsorption to the backbone and side-chain binding sites. This process is further aided by the positive net-charge of the protein at low pH, although at sufficiently high salt concentrations the binding sites saturate while at the same time the favorable electrostatic interactions diminish due to screening.

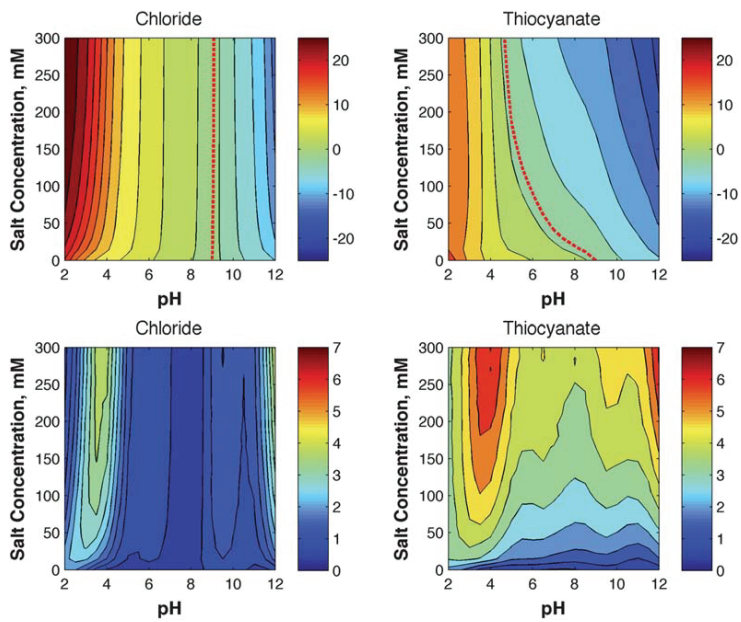


Fig. 2 Protein net charge (top) and charge capacitance (bottom) for γ-crystallin in NaCl (left) and NaSCN (right). The dashed lines on the charge plots represent the iso-electric conditions.

The charge capacitance, C, describes the protein's ability for charge regulation. In the chloride solution, the capacitance is influenced by salt only through the ionic strength; the proton concentration causes maxima at high and low pH, corresponding to the pK_d of acidic and basic sites. As the salt concentration increases the electrostatic interactions between sites are screened and whereby they become independent of each other. In the limit of infinite salt, the capacitance approaches an ideal value, given by[21]

$$C^{\text{ideal}} = \sum_i \frac{10^{pX_i - pK_{d,i}}}{(1 + 10^{pX_i - pK_{d,i}})^2} \tag{8}$$

where the sum runs over all binding sites. Compared to the capacitance where sites interact, the ideal curve tends to be larger in magnitude and hence less broad.

Thiocyanate also here displays a remarkably different picture compared to chloride and, overall, the capacitance is higher in magnitude for all salt and pH conditions. This is trivially due to the added number of binding sites that allows for more charge fluctuations. Increasing the thiocyanate concentration leads to higher capacitances that approach the ideal value as given by eqn (8) and may potentially result in strong charge regulation interactions with other molecular matter. However, as already mentioned, the quadratic charge regulation terms in eqn (5) are screened with $\sim e^{-2\kappa r}$ compared to $\sim e^{-\kappa r}$ for the direct charge–charge term and hence diminish rapidly with increasing salt. Still, thiocyanate causes capacitances several times higher than chloride even at low salt concentrations and the addition of binding ions may hence be a useful tool to tune intermolecular interaction mechanisms, charge regulation in particular.

3.2 Protein–protein interactions

3.2.1 Virial coefficients. In the previous section we discussed the electric properties of γ-crystallin using the average charge, $\langle Z \rangle$, and capacitance, C, while considering both protons and binding anions. Using these isotropic terms we now estimate the osmotic second virial coefficient, B_2, using eqn (5) and eqn (6). As shown in Fig. 3, B_2 for both SCN^- and Cl^- reaches minima for pH close to the iso-electric points, while at extreme pH the net charge of the protein causes a high (repulsive) virial coefficient. Note that B_2 here does not become negative due to the dominance of the repulsive first term in eqn (5) and that we have neglected attractive short ranged van der Waals contributions.

It is instructive to investigate the B_2 *difference* between the NaSCN and NaCl solutions – see Fig. 3, right. At near iso-electric conditions, ΔB_2 is zero and thus marks a reversal of the Hofmeister series. That is, when the protein has

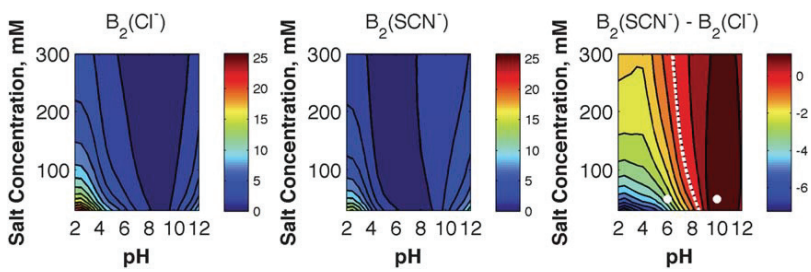


Fig. 3 Left and middle: estimated virial coefficients for γ-crystallin in NaCl and NaSCN, respectively using eqn (5) and eqn (6). Right: difference in second virial coefficient between NaSCN and NaCl. The dashed white line represents zero (Hofmeister reversal), while the two white dots mark the conditions used for the two-body simulations in Fig. 4. Note that the B_2 values are normalized with B_2^{HS}.

 This journal is © The Royal Society of Chemistry 2013

a net positive charge, bound thiocyanate anions lower the protein–protein repulsion more than chloride; the *reverse* Hofmeister series.[25] For a net negatively charged protein, binding of thiocyanate causes more repulsion than chloride, *i.e.* the *direct* Hofmeister series. This is exactly what is found in SAXS measurements for γ-crystallin as well as other proteins.[26] Experimentally, the inversion for γ-crystallin occurs at pH 4.5 and at 500 mM salt[26] which is lower than the iso-electric point in dilute solution (pI 6.7–8.9)[26,27] and thus fits the picture that addition of binding salts down shifts pI, *cf.* Fig. 2. While the low salt Hofmeister reversal predicted here also agrees with previous theoretical observations[8,28,30] the current work is distinguished in that ion binding adheres to the complex surface topology of the protein, while at the same time no explicit ions are required.

3.2.2 Two-body simulations. In the previous section we used perturbation theory to estimate the second virial coefficient of γ-crystallin solutions and crudely neglected anisotropy due to electrostatics, van der Waals and excluded volume. This can be remedied by instead explicitly simulating two coarse grained proteins (Fig. 1, right) and from the radial distribution function, $g(R)$, extracting the potential of mean force, $\beta w(R) = -\ln g(R)$. As shown in Fig. 4 the protein–protein attraction is larger for thiocyanate anions than chloride anions at pH 6, while the reverse is true at pH 10. This result is in qualitative agreement with the simplistic multipole calculations in Fig. 3, and more elaborate computer simulations with explicit ions,[8] as well as with experimental data.[26] Note also, that there are now distinct minima in the interaction free energies, stemming mostly from van der Waals interactions and, to a lesser extent, hydrophobic interactions *cf.* eqn (2).

4 Conclusion

Using a simple theoretical model that combines ion binding equilibria to macromolecular motifs with continuum electrostatics, we have investigated ion specific charge properties of the protein γ-crystallin. In particular we note that the iso-electric "point" of the protein in thiocyanate solution varies strongly with the salt concentration and is thus a function of both pH and pSCN. Osmotic second virial coefficients – calculated both from a multipole expansion and from two body computer simulations – show a reverse salting-in Hofmeister series for the cationic protein while the direct Hofmeister series is followed under anionic conditions. This is in agreement with SAXS measurements as well as with other theoretical work. While our model accounts for salting in effects, salting out due to highly solvated ions is

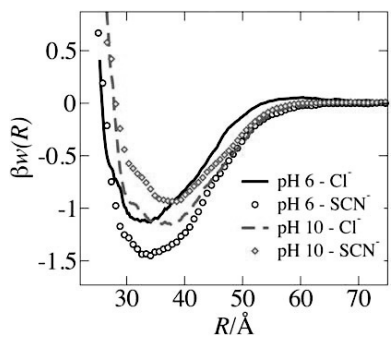


Fig. 4 Simulated potential of mean force, $w(R)$, along the protein–protein mass center coordinate, R, at different pH and anionic species (50 mM). Note the Hofmeister reversal going from low to high pH.

neglected. This may be remedied by introducing salt sensitive hydrophobic interactions;[29] a simple surface tension argument could be a viable option.

With multiple binding sites on the protein surface, thiocyanate brings about large fluctuations in the molecular charge distribution. This may lead to significant charge regulation phenomena when interacting with other molecular matter and specific ion effects may thus provide yet another handle to control intermolecular interactions.

Acknowledgements

We thank Paul Cremer and Pavel Jungwirth for useful discussions. For financial support we thank the eSSENCE@LU strategic program and the Linneaus Center of Excellence "Organizing Molecular Matter", Lund, Sweden.

References

1 P. Jungwirth and B. Winter, *Annu. Rev. Phys. Chem.*, 2008, **59**, 343–66.
2 Y. Zhang and P. S. Cremer, *Annu. Rev. Phys. Chem.*, 2010, **61**, 63–83.
3 P. Lo Nostro and B. W. Ninham, *Chem. Rev.*, 2012, **112**, 2286–2322.
4 E. Volkin, *Journal of Biological Chemistry*, 1948, **175**, 675–681.
5 P. Retailleau, M. Riès-Kautt and A. Ducruix, *Biophys. J.*, 1997, **73**, 2156–63.
6 K. D. Collins, *Methods (San Diego, Calif.)*, 2004, **34**, 300–11.
7 R. L. Baldwin, *Biophys. J.*, 1996, **71**, 2056–63.
8 M. Lund and P. Jungwirth, *J. Phys.: Condens. Matter*, 2008, **20**, 494218.
9 H. Krienke, V. Vlachy, G. Ahn-Ercan and I. Bako, *J. Phys. Chem. B*, 2009, **113**, 4360–4371.
10 C. Gibb, *J. Am. Chem. Soc.*, 2011, **133**, 7344–7347.
11 E. A. Algaer and N. F. A. van der Vegt, *J. Chem. Phys. B*, 2011, 13781–13787.
12 K. Rembert, J. Paterova, J. Heyda, C. Hilty, P. Jungwirth and P. Cremer, *J. Am. Chem. Soc.*, 2012, **134**, 10039–10046.
13 L. M. Pegram and M. T. Record, *J. Phys. Chem. B*, 2008, **112**, 9428–36.
14 D. F. Evans and H. Wennerström, *The Colloidal Domain - Where Physics, Chemistry, Biology and Technology Meet*, VCH Publishers, New York, 1994.
15 T. L. Hill, *An Introduction to Statistical Thermodynamics*, Dover Publications Inc., New York, 1986.
16 J. Wang, X. Zuo, P. Yu, I.-J. L. Byeon, J. Jung, X. Wang, M. Dyba, S. Seifert, C. D. Schwieters, J. Qin, A. M. Gronenborn and Y.-X. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 10507–10515.
17 N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller and Others, *J. Chem. Phys.*, 1953, **21**, 1087.
18 M. Lund, M. Trulsson and B. Persson, *Source Code Biol. Med.*, 2008, **3**, 1.
19 K. Linderstrøm-Lang, *Compt. Rend. trav. lab. Carlsberg*, 1924, **15**, 1–29.
20 J. G. Kirkwood and J. B. Shumaker, *Proc. Natl. Acad. Sci. U. S. A.*, 1952, **38**, 855–62.
21 M. Lund and B. Jönsson, *Biochemistry*, 2005, **44**, 5722–7.
22 M. L. Grant, *J. Phys. Chem. B*, 2001, **105**, 2858–2863.
23 D. Bratko, A. Striolo, J. Z. Wu, H. W. Blanch and J. M. Prausnitz, *J. Phys. Chem. B*, 2002, **106**, 2714–2720.
24 X. Chen, T. Yang, S. Kataoka and P. Cremer, *J. Am. Chem. Soc.*, 2007, **129**, 12272–12279.
25 Y. Zhang and P. S. Cremer, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15249–53.
26 S. Finet, F. Skouri-Panet, M. Casselyn, F. Bonneté and A. Tardieu, *Curr. Opin. Colloid Interface Sci.*, 2004, **9**, 112–116.
27 S. Zigman, J. Schultz and T. Yulo, *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.*, 1983, **75**, 425427.
28 M. Boström, F. W. Tavares, S. Finet, F. Skouri-Panet, A. Tardieu and B. W. Ninham, *Biophys. Chem.*, 2005, **117**, 217–24.
29 M. Jönsson, M. Skepö and P. Linse, *J. Phys. Chem. B*, 2006, **110**, 8782–8.
30 M. Boström, D. F. Parsons, A. Salis, B. W. Ninham and M. Monduzzi, *Langmuir*, 2011, **27**, 9504–9511.

 This journal is © The Royal Society of Chemistry 2013

Paper IV

PROTEINS

Role of histidine for charge regulation of unstructured peptides at interfaces and in bulk

Anıl Kurut,* João Henriques, Jan Forsman, Marie Skepö, and Mikael Lund

Division of Theoretical Chemistry, Lund University, P.O. Box 124, SE-22100 Lund, Sweden

ABSTRACT

Histidine-rich, unstructured peptides adsorb to charged interfaces such as mineral surfaces and microbial cell membranes. At a molecular level, we investigate the adsorption mechanism as a function of pH, salt, and multivalent ions showing that (1) proton charge fluctuations are—in contrast to the majority of proteins—optimal at neutral pH, promoting electrostatic interactions with anionic surfaces through charge regulation and (2) specific zinc(II)-histidine binding competes with protons and ensures an unusually constant charge distribution over a broad pH interval. In turn, this further enhances surface adsorption. Our analysis is based on atomistic molecular dynamics simulations, coarse grained Metropolis Monte Carlo, and classical polymer density functional theory. This multiscale modeling provides a consistent picture in good agreement with experimental data on Histatin 5, an antimicrobial salivary peptide. Biological function is discussed and we suggest that charge regulation is a significant driving force for the remarkably robust activity of histidine-rich antimicrobial peptides.

Proteins 2014; 82:657–667.
© 2013 Wiley Periodicals, Inc.

Key words: histidine richness; antimicrobial activity; adsorption of unstructured proteins; Monte Carlo simulations; charge regulation; specific metal binding.

INTRODUCTION

Nature has produced a number of biomolecules rich in histidines (HIS). These include not only globular proteins such as hisactophilin and histidine-rich glycoprotein, but also intrinsically disordered peptides (IDPs) such as the histatins. The latter are characterized by lack of stable tertiary structure when the protein exists as an isolated polypeptide under *in vitro* physiological conditions. It has recently been shown that approximately 30% of all proteins in eukaryotic organisms belong to this group and that IDPs are involved in a large number of central biological processes and diseases.[1] This discovery challenges the traditional protein structure paradigm, which states that a specific well-defined structure is required for correct function. Biochemical evidence has since shown that IDPs are functional and that lack of folded structure is in fact related to function.[2,3]

Histatins are Salivary peptides with multifunctionality. They not only protect the oral environment against microbial attack, but are also involved in the formation of protective films on solids, which is important for the maintenance of oral health and surface integrity. The formation of the acquired tooth enamel pellicle involves the selective adsorption of salivary macromolecules onto the enamel surface. This conditioning film acts as a selective permeability barrier between the tooth surface and the oral environment and regulates the defence against enamel demineralization. It is hence crucial to understand how lack of structure relates to function when adsorbed to exposed biological interfaces. Although we in this article focus on the antimicrobial activity of Histatin 5 (His5), we emphasize that our arguments can be generally applied to biomolecular surface adsorption.

His5 consists of 24 amino acids and is the most active member of the histatins against pathogenic yeast *Candida albicans*.[4] Its antimicrobial activity and possible killing mechanisms are therefore subject to much scientific attention.[5–9] Contrary to many antimicrobial peptides, His5 penetrates the cell membrane without causing

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: eSSENCE strategic program, Sweden's Innovation Agency "Vinnova", the Linneaus Center of Excellence "Organizing Molecular Matter", and LUNARC, Lund, Sweden.

*Correspondence to: Anil Kurut; Division of Theoretical Chemistry, Lund University, P.O. Box 124, SE-22100 Lund, Sweden. E-mail: anil.kurut@teokem.lu.se

Received 11 July 2013; Revised 12 September 2013; Accepted 26 September 2013
Published online 10 October 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24445

© 2013 Wiley Periodicals, Inc.

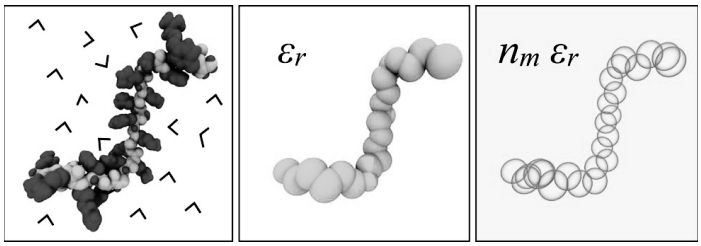


Figure 1

Models in decreasing order of detail, left to right: Atomistic (MD), soft sphere amino acid bead model (MC), and hard sphere amino acid model (DFT). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

membrane rupture, which is attributed to its histidine richness.[10,11] Further, HIS form complexes with zinc ions,[12,13] which promote adsorption to bilayer vesicles[6] and its killing activity.[9] It has also been shown that the 16 C-terminal amino acids with a 25% of HIS content are the active segment of His5,[14] supporting the importance of HIS in the activity.

Accepting that both microbial cell membranes and tooth enamel represent negatively charged interfaces, we here investigate the role of histidine on the anionic surface adsorption of a flexible peptide, exemplified by His5. Our theoretical analysis is based on three levels of detail entailing (i) atomistic bulk simulations, (ii) a flexible amino acid bead model with Debye-Hückel electrostatics, and (iii) classical polymer density functional theory (DFT). Common for all models is that pH is kept *constant* meaning that the peptide protonation state can fluctuate and respond to the local chemical environment. This gives rise to charge regulation mechanisms[15–17] that may contribute to intermolecular interactions in biomolecular systems—for a recent review, see Ref. 18. As will be shown, charge regulation becomes important for histidine-rich biomolecules close to neutral pH. We further investigate how binding metal ions such as zinc influence the adsorption free energy to charged surfaces using an empirical ion-binding potential.

The article is organized as follows: First, we present the three theoretical models and proceed to study bulk electrostatic and structural properties of His5. We then study the adsorption mechanism and free energy as a function of pH and salt concentration. Finally, we discuss the effect of multivalent ions, including specific metal binding and possible implications for biological activity.

PROTEIN MODELS AND METHODS

We have used three models with different levels of detail (Fig. 1), one atomistic and two coarse grained (CG), with a common feature of fluctuating charges on basic and acidic residues. That is, the chemical potential of protons is kept constant during the simulation, upgrading most simulation work where partial charges are fixed. Atomistic level molecular dynamics (MD) simulations are used to investigate bulk properties of His5 and to verify that the CG models capture the essential physical features. Monte Carlo (MC) simulations and an amino acid bead model are used to investigate both bulk properties and surface affinities. Because atomistic MD requires substantial computational resources, we have used this method only to scan a limited parameter range while fine scanning has been performed using a much faster CG model. Finally, we have tested hard-sphere amino acid bead model using classical polymer density functional theory, which are well suited for further adsorption studies in concentrated solutions.

CG Monte Carlo simulations

Model

His5 is modeled as a flexible chain where each residue is CG into a sphere and connected with harmonic bonds[19]—see Figure 1. The amino acid radii are determined from their molecular weight by assuming a common density of 0.9 g/mL. During the simulations, the charges of basic and acidic amino acids and C and N terminals are allowed to fluctuate according to pH, when $\xi_{reg}=1$ in Eq. (1). An average bulk charge is assigned to each site when $\xi_{reg}=0$. Water and monovalent ions are treated implicitly, whereas divalent ions are explicitly represented by spheres with a 2 Å radius.[20]

In addition to proton binding, we also mimic specific zinc coordination to deprotonated HIS using a short range (SR) square-well attraction, adjusted to $1.2\times10^5\,M^{-1}$, the experimental binding constant of zinc to isolated imidazole groups[13] —see Table I and Eq. (1). The attraction range of 2.4 Å is the average distance between Zn^{2+} and HIS in three different zinc binding proteins (PDBs 2XQV, 2XY4, 2WB0, 2IIM).

The system energy for a given configuration is,

$$U_{tot} = \overbrace{\left(4\varepsilon\sum_{i>j}^{n}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)}^{\text{Lennard-Jones (LJ)}} + \underbrace{\left(\frac{1}{4\pi\varepsilon_0\varepsilon_r}\sum_{i>j}^{n_t}\frac{q_i q_j}{r_{ij}}e^{-\kappa r_{ij}}\right)}_{\text{Debye-Hückel}} + \underbrace{\left(\xi_{reg}k_B T\sum_{1}^{n_p}(pK_a - pH)\ln(10)\right)}_{\text{Intrinsic titration energy}} + \underbrace{\left(\xi_{surf}\sum_{i}^{n_t}q_i\Phi_0 e^{-\kappa d_i}\right)}_{\text{Gouy-Chapman (G-C)}} + \underbrace{\left(\sum_{i}^{n_b}\frac{1}{2}k_h(R_i - Re)^2\right)}_{\text{Harmonic Bond}} + \underbrace{\left(\sum_{1}^{n_H} - \varepsilon_{Zn}(d_{Zn-H} < 2.4)\right)}_{\text{short range attraction (SR)}} \quad (1)$$

where all the parameters are given in Table I.

Table I
Parameters in Eq. (1)

Parameter = value	Description
n=26	All residues
ε=0.05 k_BT	Strength of LJ
$\sigma_{ij}=(\sigma_i+\sigma_j)/2$ Å	LJ diameter
=range (4, 8.2)Å	6.2 Å in average
r_{ij}	Center of mass distance
e	Electron unit charge
ε_0	Permittivity of vacuum
ε_r=80	Water dielectric constant
n_t=20	Titratable residues
q_i	Charge
κ^{-1}	Debye length
T=298 K	Temperature
k_B	Boltzmann's constant
N_A	Avogadro's number
ξ_{reg}=1 or 0 (on/off)	Charge regulation switch
n_p	Protonated residues
pK_a= see Table II	Acid dissociation constant
ξ_{surf}=1 or 0 (on/off)	G-C potential switch
$\Phi_0=\frac{2k_BT}{q}\sinh^{-1}\frac{\rho}{\sqrt{(8k_BT\varepsilon_0\varepsilon_r)}}$	Potential at the surface
$\rho=-0.053$ C/m^2	Surface charge density
d_i	Distance to the surface
k_h=0.76 k_BT	Spring constant
R_i	Bond length
R_e=4.9 Å	Equilibrium bond length
$n_b=n-1$	Number of bonds
ε_{Zn}=4.57 k_BT	Strength of SR
n_H	Number of HIS
d_{Zn-H}=2.4 Å	Bound Zn(II)-HIS distance

Method

A single peptide is simulated using the Metropolis MC algorithm in the canonical ensemble (*NVT*) at 298 K in a volume of 300 × 300 × 300 Å^3 and 300 × 300 × 150 Å^3 for bulk and surface adsorption studies, respectively. In bulk studies (ξ_{surf}=0), periodic boundaries are applied in all directions, whereas in adsorption simulations (ξ_{surf}=1), a uniform surface charge density is assigned to the $+xy$ wall and hard boundaries are applied in the z-direction. Chain conformations are sampled using crankshaft, pivot, branch rotations, reptation, and single monomer MC moves together with whole chain rotation and translation. All MC simulations are performed using the Faunus framework.[21]

Atomistic constant-pH MD simulations

Model

A modified version of the GROMOS 54A7 force field[22] is used for the atomistic description of His5. All simulations are started with a linear structure of His5, built using MacPyMOL.[23] The charges of 20 sites, including N and C terminals, are updated during the simulations—see next section.

Method

All constant-pH MD simulations are performed using a stochastic titration method,[24–31] which relies on three steps. First, the protonation states of residues are determined using Poisson-Boltzmann/Monte Carlo (PB/MC) calculations. Second, the solvent is relaxed around the frozen peptide using a short MD simulation. The third step is an MD simulation of the unrestraint system. All simulations are performed in the isobaric-isothermal ensemble (*NPT*), at 300 K and 1 bar with the GROMACS 4.0.7 distribution.[32–35] PB and MC calculations are performed with MEAD[36] and PETIT[37,38] software, respectively. The simulations are run for 20 ns in triplicate, and the last 15 ns of each replica is used for the analysis. To confirm the convergence of the structural properties, two independent replica-exchange MD (REMD) simulations in temperature are performed with 64 replicas, each in a salt-free solution at pH 7—see Supporting Information.

Density functional theory

Classical polymer density functional theory (DFT) was originally developed by Woodward[39] and has seen subsequent developments and extensions. We shall refrain from a complete description here and instead refer to a recent work,[20] where a rather general exposition is provided. In this work, we have used a version of the implicit solvent approach, denoted "model II" in Ref. 40.

However, we will here also (approximately) account for differences in size between monomers (coarse-grained amino acids) and solvent particles (water). Given the minor role of dispersion (Lennard-Jones) interactions in our system, these are neglected in the DFT approach. Tests have confirmed that the Lennard-Jones interactions have a negligible influence on the results. Moreover, the CG amino acid monomers have a common (hard-sphere) diameter of $d_m = 6.2$ Å, which is the average value in the soft sphere CG model. The neighboring monomers are connected by stiff bonds with a fixed length of 5 Å. The solvent particles enter explicitly via a net incompressibility constraint (see Ref. 40: $n_t = n_m \gamma d_m^3 + n_s d_s^3$ where n_m and n_s are solvent and monomer densities, and we have simply set $n_t = 1$ corresponding to simple cubic close packing. The parameter γ accounts for the reduction in excluded volume brought about by bonding (connected spheres exclude less volume than free ones). According to Hall et al.,[41] this factor can be estimated to about 0.63, for a pearl-necklace chain. Woodward, on the other hand, found an empirical value of 0.83. In our case, with fused hard-sphere chains, this would be reduced somewhat. Given that we work with a rather crude model anyway, with excluded volume interactions playing a minor role, we chose to simply set $\gamma = 0.65$, which might be a reasonable compromise of all these considerations. We have made tests to ensure that our findings and conclusions are insensitive to the specific value of γ.

Note that the polymer DFT is *exact* for the case of noninteracting monomers. To treat titrations properly, the two different states of the monomers (charged/uncharged) must be properly accounted for, according to the prescription described in Ref. 40.

Analysis

We have analyzed statistical mechanical averages of the residual charges $\langle q_i \rangle$, the net charge of His5, $Q = \Sigma_i \langle q_i \rangle$, the end-to-end distance $R_{e2e}^2 = \langle r_{1,last}^2 \rangle$, radius of gyration, $R_g^2 = \frac{\Sigma_i m_i \langle r_{i,cm}^2 \rangle}{\Sigma_i m_i}$ where $r_{i,cm}$ is the distance between residue i and the mass center of His5 as a function of the center of mass distance from the surface. The charge capacitance is analyzed using the fluctuations of the net charge, $C = \langle Q^2 \rangle - \langle Q \rangle^2 = -\frac{\partial Q k_B T}{\partial e \Phi}$ where Φ is the external electric potential from the surface. Probability density functions with respect to distance from the surface are sampled using the histogram method and normalized such that the integral yields the average monomer concentration times the z-box length ($26 \times 150/(300 \times 300 \times 150)$, Å^{-2}).

The analyses of the MD simulations have been performed using GROMACS 4.0.7 software package[32–35] and in-house tools. The calculations of correlation-corrected errors for averages over a single simulation replicate have been computed using standard methods.[42] The final errors of the averages over replicates have been computed using the law of total variance.

RESULTS AND DISCUSSION

Bulk properties

The bulk properties of His5 in dilute solutions have been investigated as a function of monovalent salt concentration and pH. To mimic biological conditions, ionic strengths of 80 and 150 mM were chosen, whereas for completeness, we also include the salt-free case.

Electrostatic properties

As shown in Figure 2 (left), the isoelectric point (pI) of His5 is predicted to 10.5 by both MC and MD simulations, coinciding with previous theoretical studies.[11,43] Calculated net charges using both models are in a good agreement over the whole pH range. At extreme acidic and basic conditions, the net charge reaches $+15e$ and $-5e$, respectively, and contrary to most proteins and peptides, His5 is cationic over a wide pH range. At intermediate and high ionic strengths, the reduced electrostatic repulsion due to salt screening enables protonation of nearby basic residues and allows for a higher net charge.

Charge fluctuations result in a high charge capacitance [see Eq. (1)] and hence, a high charge regulation ability. That is, the molecular charge may readily respond to an external electric field by taking up or releasing protons. As shown in Figure 2, the capacitance of His5 has four maxima and three minima. In this study, we focus on the biological pH range. At salt-free conditions, the peak and the minimum are located at pH 5 and 7, respectively, meaning that the charge distribution can be easily perturbed at pH 5 but not at pH 7. Although the capacitance extrema predicted by MD are at slightly lower pH, both models agree on the shift of these extrema to higher pH with increasing ionic strength and on the capacitance minimum around pH 8.

To determine, which residues contribute to the capacitance peak, we investigate the residual stoichiometric acid dissociation constant (pK_a^*) obtained from the titration curve of each residue. At pH=pK_a^*, a residue has high charge fluctuations due to equally probable protonation states, leading to a capacitance peak. Table II shows that the peak around pH 6 originates from HIS residues with pK_a^* ranging from 5.4 to 6.2. At pH 8, only the N-terminal contributes slightly to the capacitance, thus leading to a minimum.

Structural properties

Although atomistic and CG models agree well on the electrostatic properties, they differ slightly on the structure predictions. As seen in Figure 2, the atomistic model favors more compact conformations than the CG model,

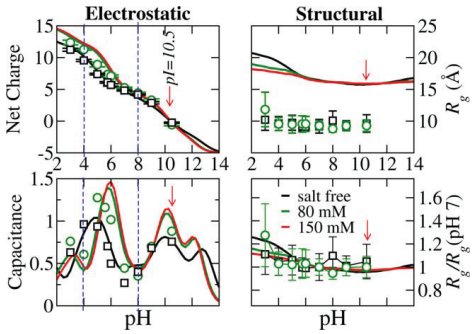


Figure 2

Electrostatic and structural properties of His5 as a function of pH and 1:1 salt concentration calculated using MC (lines) and MD (symbols) simulations. The interval fenced by the dashed lines shows saliva conditions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

resulting in a smaller gyration radius, R_g. This can be due to smaller excluded atom volumes combined with hydrogen bonds. In the CG flexible chain models, the high entropic cost of compact conformations combined with lack of hydrogen bonds favors random coil conformations, and thus larger R_g. Experimental studies[11,14,44] have shown that in addition to β−turns, His5 lacks secondary structure in aqueous solution, thus supporting the CG models. Nevertheless, the true R_g is probably in between and both models still predict the same relative changes in R_g induced by salt and pH—see Figure 2.

Interaction with negative surfaces

Effect of charge regulation

Although most proteins have capacitance peaks around pH 4,[45] we here show that the capacitance peak of His5 is at pH 6 coinciding with normal salivary conditions. This implies that HIS provides a significant ability to adjust the net charge in the vicinity of a negative microbial membrane. Figure 3 shows that His5 becomes more positively charged while approaching a negative surface at pH 6. Because this cationic gain depends on the number of residues with pK_a^*=pH, the abundance of HIS provides an additional charging capacity up to $3e$ and enhances the affinity to the surface (see Fig. 3, bottom). In contrast, the lack of residues with $pK_a^* \approx 8$ results in a slight net charge increase and the surface affinity is thus unaffected.

Note that the charge regulation may also lead to neutralization of the peptide. His5 penetrates cells without deforming the membrane integrity[7,10,46] and bearing charges, while passing the apolar membrane interior may be costly. With its high capacitance, HIS may lower this free-energy cost by adapting the deprotonated state, thus avoiding pore formation on the membrane.

Effect of pH

The high positive charge at pH 6 combined with charge regulation results in a strong adsorption to the negative surface—see Figure 3. Because of the release of protons, this adsorption is reduced at alkaline conditions, but at low ionic strengths, the surface adsorption is still significant. In contrast, the interaction free energy at low pH reaches $8.5\,k_BT$ with a free energy minimum closer to the surface hinting at irreversible adsorption with a more compact adsorbed layer.

Effect of monovalent salt

As seen in Figure 3, the surface affinity of His5 decreases with increasing monovalent salt concentration because of the screening of electrostatic attraction. Although salt is also reducing the repulsion within the peptide, thus allowing higher net charges, this effect is insufficient to compensate for the screened surface attraction. Our model predicts a clear correlation

Table II
Stoichiometric (pK_a^*) and Intrinsic (pK_a) Acid Dissociation Constants of Residues at 0, 80, and 150 mM 1:1 Salt

Residue id		1	2	3	4	5	6	7	8	9	10	11	12
Residue	NTR[a]	ASP	SER	HIS	ALA	LYS	ARG	HIS	HIS	GLY	TYR	LYS	ARG
pK_a	7.5	4.0	—	6.3	—	10.4	12.0	6.3	6.3	—	9.6	10.4	12.0
pK_a(0 mM)	7.3	1.7	—	**5.2**	—	10.0	12.5	**4.0**	**4.5**	—	8.1	10.3	12.8
pK_a^*(80 mM)	7.6	3.2	—	**6.0**	—	10.2	12.2	**5.4**	**5.7**	—	8.9	10.4	12.3
pK_a^*(150 mM)	7.7	3.4	—	**6.1**	—	10.2	12.1	**5.6**	**5.8**	—	9.1	10.4	12.2
Residue id	13	14	15	16	17	18	19	20	21	22	23	24	
Residue	LYS	PHE	HIS	GLU	LYS	HIS	HIS	SER	HIS	ARG	GLY	TYR	CTR[a]
pK_a	10.4	—	6.3	4.4	10.4	6.3	6.3	—	6.3	12.0	—	9.6	2.6
pK_a^*(0 mM)	9.9	—	**5.1**	1.3	11.0	**4.5**	**5.0**	—	**5.0**	13.0	—	9.5	1.0
pK_a^*(80 mM)	10.0	—	**6.0**	3.2	10.7	**5.7**	**5.9**	—	**5.8**	12.3	—	9.7	2.2
pK_a^*(150 mM)	10.1	—	**6.2**	3.4	10.7	**5.8**	**5.9**	—	**5.9**	12.3	—	9.8	2.3

[a]NTR and CTR are abbreviations for N and C terminals, respectively.

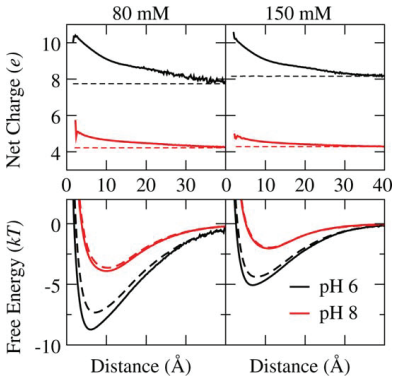


Figure 3

Net charge and interaction free energy of His5 as a function of distance from a surface with a charge density $\rho = -300\,\text{Å}^2/e$ with (solid lines) and without (dashed lines) charge regulation at pH 6 and pH 8 where the capacitance is maximized and minimized, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

between ionic strength and surface affinity and is thus consistent with experiments showing that the biological activity of His5 diminishes with increasing ionic strength.[11,47]

Adsorption mechanism

To determine the adsorption mechanism, we have investigated the gyration radius, R_g, and the distribution of individual residues while His5 approaches the surface. As seen in Figure 4, the electric field from the surface induces conformational changes on His5 when the separation is smaller than 40 Å. The z-component of R_g first increases indicating a chain elongation toward the surface, followed by a contraction due to xy-stretching of the adsorbed chain on the surface. The elongation is reflected as a depletion of some residues from the bulk—see Figure 4. These residues, containing two positive ARG-LYS patches, drive the peptide to the surface and mediate the adsorption by providing two grafting regions. The primary ARG-LYS patch consists of the 11th to 13th residues and provides the strongest attachment.

Conformational entropy drives the N (NTR) and C (CTR) terminal domains further from the surface. However, the energetic gain of positioning the NTR and the nearby secondary ARG-LYS patch close to the surface, partly compensates for this entropy and the NTR consequently adsorbs stronger than the CTR.

Many antimicrobial salivary peptides, including histatins, are expressed by multiple gene families.[48] Diversities in these families cause polymorphism, which may influence the microbial activity.[49] For instance, a rare variant of proline-rich glycoprotein does not bind to bacteria.[50] We have tested the sensitivity of the primary grafting patch to a point mutation by replacing the 12th ARG with a negative TYR. This mutation eliminates the grafting to the surface (Fig. 6) resulting in a 40% reduction in the overall surface affinity—see Figure 5. In addition, mutating the positive primary ARG-LYS patch to a neutral GLY completely prevents adsorption of this region and supports the possible suppression of interactions with microbial membranes due to protein polymorphism.

Because of HIS attraction, the adsorption of CTR persists even when the primary patch is completely desorbed (Fig. 6), indicating the prominent role played by HIS. An experimental study of His5 fragments[14] showed that the HIS residues in the CTR domain and the primary ARG-LYS patch are the active parts. This supports our predictions and the importance of electrostatics for antimicrobial activity.

Effect of pH

As seen in the residue concentration profiles (Fig. 7), increasing pH from 6 to 7, markedly reduces the surface affinity of HIS and results in a weaker adsorption of both ARG-LYS patches and the CTR. This is more pronounced at pH 8 where the CTR completely desorbs from the surface due to lack of charge regulation combined with released protons. Although the ARG-LYS patches are still positive at this pH, their affinity to the surface is insufficient to compensate conformational entropy loss of the terminals, thus leading to desorption of both terminals and the secondary ARG-LYS patch.

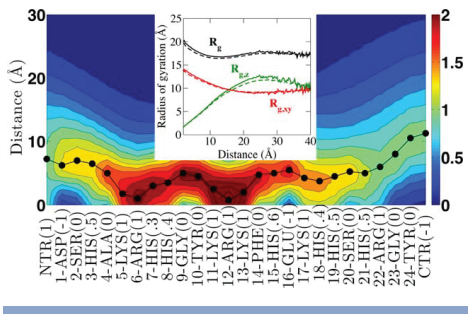


Figure 4

Residual concentration profile in mM as a function of the distance from the surface at pH 6 and 80 mM 1:1 salt. *Inset*: Radius of gyration and its components with (solid) and without (dashed) charge regulation. Circles indicate the most probable location of each residue relative to the surface and average bulk charges are shown in parentheses. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

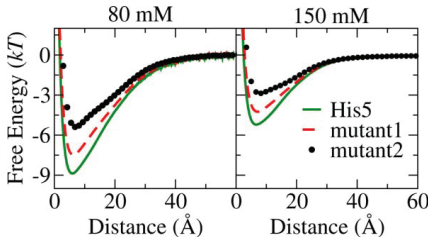


Figure 5

Interaction-free energy of His5 (solid lines), mutant1 (dashed lines, 12-ARG is replaced by a negative TYR) and mutant2 (circles, primary ARG-LYS patch is mutated to a neutral GLY) at pH 6 and 80 mM 1:1 salt solution. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

This gradual desorption can be seen in Figure 8 where we have plotted the most probable residue positions relative to the surface.

Previous reports show that the activity of His5 is independent of pH.[47,51] However, in these studies, the pH dependence was investigated at low ionic strength (~10 mM) where arginine and lysine dominate the surface attraction. We here show that even at 80 mM ionic strength—see Figure 3, His5 has appreciable affinity to the surface, independent of the studied acidic conditions. This suggests that His5 is active against microbes once its surface affinity is above a critical value. Furthermore, our results indicate that the surface affinity of His5 is sensitive to pH changes in environments where the ionic strength is relatively high (such as saliva). As will be shown, this statement is valid only for zinc-free conditions.

Note that the surface is modeled independently of solution conditions with a constant surface charge density. In reality, the phosphate groups of microbial membranes may change their protonation state according to pH. Our model may therefore slightly underestimate the surface affinity with increasing pH.

Effect of monovalent salt

The residue concentration profiles in Figure 7 show that adsorption of all residues is reduced with increasing salt concentration. The most pronounced salt screening effect is observed on the adsorption of the CTR domain at pH 7 and 8 where charge regulation effects are of little importance—see Figure 8. In contrast, at pH 6, the cooperation between positive patches and the charge regulation of HIS partly compensates salt screening and provides sufficient attraction to preserve the peptide conformation at the surface. These adsorption features are also captured by DFT as shown in the Supporting Information.

Effect of charge regulation

To estimate the importance of charge regulation, we have investigated the residue profiles with and without charge regulation. The latter is performed by assigning bulk charge to each residue corresponding to a certain pH. As seen in Figure 9, all profiles except for HIS are mostly unaffected by charge regulation. The surface affinity of HIS without additional regulated charges is insufficient to compensate for the conformational entropy loss, resulting in weaker adsorption of CTR and the secondary ARG-LYS patch—see also Figure 7. This emphasizes the importance of charge regulation on the CTR adsorption, the active part of His5, and potentially on biological activity. In the present example, the charge regulation mechanism is important within a nanometer from the surface, whereas direct electrostatic attraction with arginine and lysine are of longer range (first order vs. second order multipole electrostatics)[15]. Thus, LYS and ARG cause a long range attraction, whereas charge regulation of HIS enhances the adsorption at short separations and at low pH. This is consistent with recent experimental work mapping the roles played by LYS and ARG.[52]

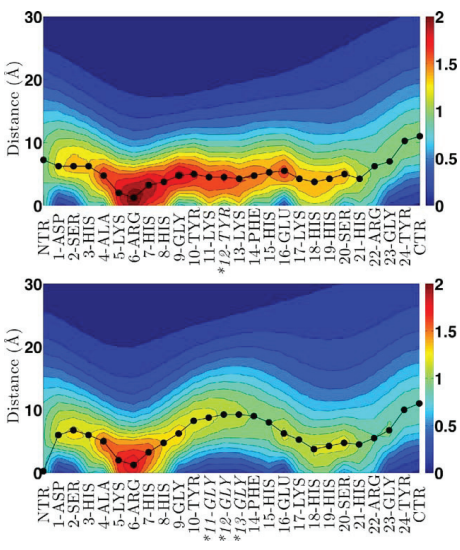


Figure 6

Concentration profiles of mutant1 (*top*) and mutant2 (*bottom*) at pH 6 and 80 mM 1:1 salt solution as a function of the distance from the surface. Mutated amino acids are indicated by *. Circles show the most probable positions of residues relative to the surface. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

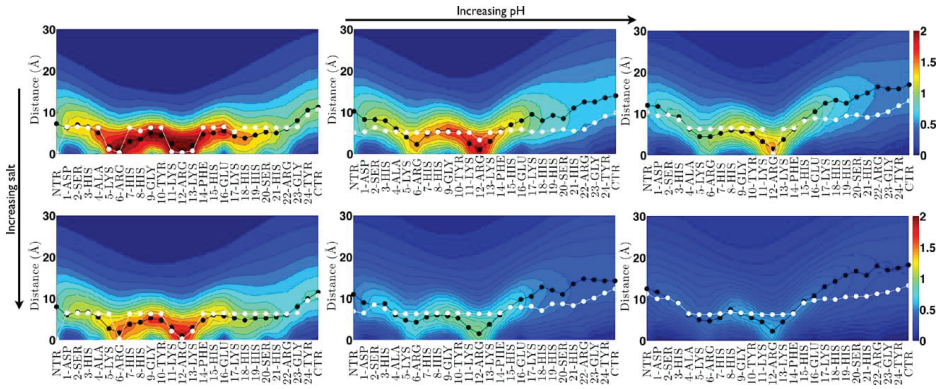


Figure 7

Residual concentration profile in mM as a function of distance from the surface at pH 6, 7, and 8; in 80 m*M* and 150 mM 1:1 salt. White and black circles show the most probable positions of residues relative to the surface predicted by MC and DFT, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Effect of divalent ions

Nonbinding ions

In addition to monovalent ions, biological environments contain divalent metal ions such as magnesium, zinc, and calcium.[48] Taking saliva as an example, Ca^{2+} is the most abundant divalent ion[48] and its concentration depends on the oral health state.[53,54] Because high levels of divalent ions may lead to ion correlations,[55] we investigated how Ca^{2+} affects the surface affinity of His5. Note that calcium does not bind *specifically* to His5 and will here be representative of any generic divalent ion. Simulations are performed at 6.15 mM of Ca^{2+} corresponding to an increased saliva level[48] with *explicit* divalent and implicit monovalent ions. This hybrid model has previously been shown to correctly reproduce electrostatics in strongly coupled systems.[20] For comparison with the monovalent salt studies in previous sections, the total ionic strength is kept constant at 80 mM. As seen in Figure 10, the adsorption and charge regulation of His5 are unaffected by Ca^{2+} ions at 80 mM ionic strength indicating that calcium induces no or only weak correlation effects and behaves chiefly as an ionic screening agent like monovalent ions.

Binding ions

Contrary to calcium, zinc(II) selectively coordinates to the imidazole group of HIS[12,13] with one Zn^{2+} ion

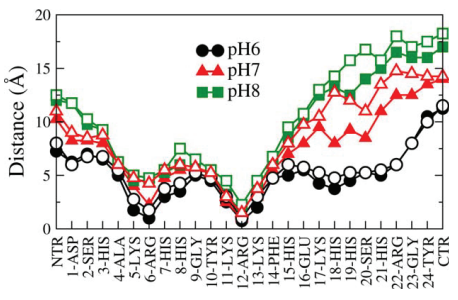


Figure 8

Most probable positions of residues relative to the surface at 80 mM (full symbols) and 150 mM (open symbols) [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

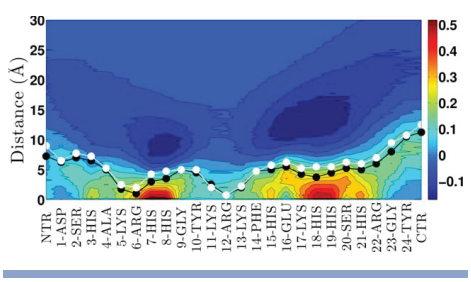


Figure 9

Difference between the residual concentration profiles with and without charge regulation at pH 6 and 80 mM 1:1 salt solution. The most probable positions of residues with and without charge regulation are shown by black and white circles, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

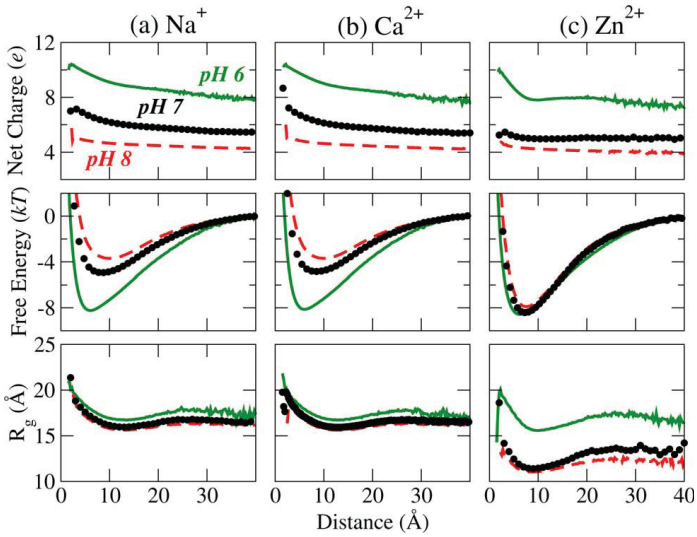


Figure 10

Properties of His5 as a function of the surface distance at 80 mM total ionic strength with and without divalent ions. See Table 3 for the concentration of divalent ions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

chelating up to four HIS.[56] It has been shown that this binding promotes the activity of His5 by facilitating the adsorption to microbial membranes and inducing fusion.[6,9] To mimic zinc binding, we have investigated the surface affinity of His5 in the presence of explicit zinc ions with a specific short-range attraction to HIS as detailed in the model section. Simulations are performed with two Zn^{2+} ions corresponding to 0.25 mM to allow all possible

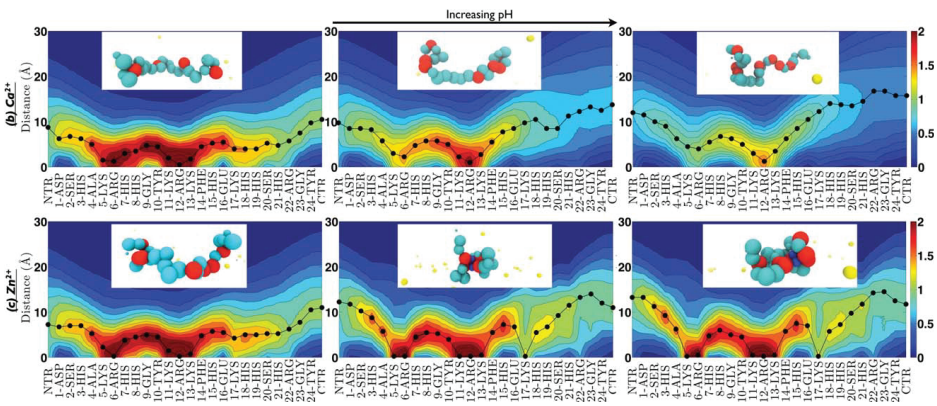


Figure 11

Concentration profiles of residues with explicit divalent ions at pH 6, 7, and 8 and 80 mM total ionic strength. Deprotonated HIS (red), calcium (yellow), and zinc (blue) ions are shown in the snapshots of His5 conformation on the surface.

Table 3
Concentrations (mM) of Explicit Divalent and Implicit Monovalent Ions and Corresponding Debye Lengths, $1/\kappa$

	Na^+	Cl^-	Ca^{2+}	Zn^{2+}	κ^{-1}(Å)
(a)	80	80	0	0	10.8
(b)	61.55	73.85	6.15	0	11.7
(c)	61.55	74.35	6.15	0.25	11.7

zinc coordinations. Although this concentration is higher than salivary levels (μM),[57] it is still relevant to biological applications, including usage of dental products with zinc salts, which may bring the oral zinc level up to 0.5 mM.[58] As seen in Figure 10, the surface affinity of His5 is enhanced by zinc at pH > 6 where protons are unable to compete for zinc-binding sites. In contrast, at pH 6 or lower, this competition is won by protons, rendering the charge distribution oblivious to the presence of zinc. Our results are in line with experimental studies showing that His5 forms strong metal–peptide complexes at neutral and basic conditions.[59] Finally, Figure 10 shows a reduction in gyration radius on adsorption, suggests that zinc complexation induces compact conformations, which are also observed experimentally,[59] via bridging HIS—see snapshots in Figure 11. Interestingly, zinc complexation leads to a strong adsorption independent of pH by providing more grafting points on the surface and restoring positive charges lost from proton release.

CONCLUSION

Focusing on electrostatics, we have investigated bulk properties of an unstructured histidine-rich peptide, His5, and its adsorption to a uniformly charged surface to determine the role of HIS. In bulk solution, CG and atomistic models have been explored using MC and MD simulations. All models quantitatively agree on bulk electrostatics indicating that a sufficient level of detail is included in the CG models. Qualitative agreement is obtained on structural properties by predicting the same pH and salt dependence, suggesting that the effect of pH and salt can be captured without full atomic detail.

We show that His5 has an uncommonly high charge capacitance at pH 6 due to a high histidine content. Matching capacitance peak conditions activates a charge regulation mechanism and turns HIS into positive charge providers under an external electric field. This mechanism can be generalized to any titratable sites of any proteins, regardless of size and structure, with matching pK_a and pH. Although charge regulation promotes adsorption at pH 6, increasing pH causes proton release and consequently desorption of His5. However, when pH > 6, deprotonated HISs act as zinc chelators that restore the positive charges lost from proton release. In effect, the pH sensitivity is eliminated and zinc thus ensures a robust, pH-independent surface affinity. Finally, charge regulation may neutralize the peptide, thereby preventing unfavorable desolvation interactions while passing bilayers. This mechanism may highlight why His5 can penetrate microbial cells without rupturing their membranes.

ACKNOWLEDGMENTS

We thank Chris H.J. Evers for piloting this study.

REFERENCES

1. Wright PE, Dyson H. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999; 293:321–331.
2. Ward J, Sodhi J, McGuffin L, Buxton B, Jones D. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.
3. Liu J, Faeder JR, Camacho CJ. Toward a quantitative theory of intrinsically disordered proteins and their function. Proc Natl Acad Sci USA 2009;106:19819–19823.
4. Kavanagh K, Dowd S. Histatins: antimicrobial peptides with therapeutic potential. J Pharm Pharmacol 2004;56:285–289.
5. Tsai H, Bobek L. Human salivary histatins: promising anti-fungal therapeutic agents. Crit Rev Oral Biol Med 1998;9:480–497.
6. Melino S, Rufini S, Sette M, Morero R, Grottesi A, Paci M, Petruzzelli R. Zn(2+) ions selectively induce antimicrobial salivary peptide histatin-5 to fuse negatively charged vesicles. Identification and characterization of a zinc-binding motif present in the functional domain. Biochemistry 1999;38:9626–9633.
7. Den Hertog A, Van Marle J, Van Veen H, Van't Hof W, Bolscher J, Veerman E, Amerongen A. Candidacidal effects of two antimicrobial peptides: histatin 5 causes small membrane defects, but ll-37 causes massive disruption of the cell membrane. Biochem J 2005;388:689.
8. Dong J, Vylkova S, Li X, Edgerton M. Calcium blocks fungicidal activity of human salivary histatin 5 through disruption of binding with *Candida albicans*. J Dent Res 2003;82:748–752.
9. Rydengård V, Andersson Nordahl E, Schmidtchen A. Zinc potentiates the antibacterial effects of histidine-rich peptides against *Enterococcus Faecalis*. FEBS J 2006;273:2399–2406.
10. Luque-Ortega JR, van't Hof W, Veerman ECI, Saugar JM, Rivas L. Human antimicrobial peptide histatin 5 is a cell-penetrating peptide targeting mitochondrial ATP synthesis in Leishmania. FASEB J 2008;22:1817–1828.
11. Helmerhorst EJ, van't Hof W, Breeuwer P, Veerman EC, Abee T, Troxler RF, Amerongen AV, Oppenheim FG. Characterization of histatin 5 with respect to amphipathicity, hydrophobicity, and effects on cell and mitochondrial membrane integrity excludes a candidacidal mechanism of pore formation. J Biol Chem 2001;276:5643–5649.
12. Grogan J, McKnight CJ, Troxler RF, Oppenheim FG. Zinc and copper bind to unique sites of histatin 5. FEBS Lett 2001;491:76–80.
13. Gusman H, Lendenmann U, Grogan J, Troxler RF, Oppenheim FG. Is salivary histatin 5 a metallopeptide? Biochem Biophys Acta 2001; 1545:86–95.
14. Raj PA, Edgerton M, Levine MJ. Salivary histatin 5: dependence of sequence, chain length, and helical conformation for candidacidal activity. J Biol Chem 1990;265:3898–3905.
15. Kirkwood JG, Shumaker JB. Forces between protein molecules in solution arising from fluctuations in proton charge and configuration. PNAS 1952;38:863–871.
16. Wyman J, Gill SJ. Binding and linkage: functional chemistry of biological macromolecules, p 330, University Science Books: Mill Valley, 1990.
17. Kao Y-H, Fitch CA, Bhattacharya S, Sarkisian CJ, Lecomte JT, Garcia-Moreno EB. Salt effects on ionization equilibria of histidines in myoglobin. Biophys J 2000;79:1637–1654.

18. Lund M, Jönsson B. Charge regulation in biomolecular solution. Q Rev Biophys 2013;46:265–281.
19. Evers CHJ, Andersson T, Lund M, Skepö M. Adsorption of unstructured protein β-casein to hydrophobic and charged surfaces. Langmuir 2012;28:11843–11849.
20. Forsman J, Nordholm S. Polyelectrolyte mediated interactions in colloidal dispersions: hierarchical screening, simulations, and a new classical density functional theory. Langmuir 2012;28:4069–4079.
21. Lund M, Trulsson M, Persson B. Faunus: an object oriented framework for molecular simulation. Source Code Biol Med 2008;3:1.
22. Schmid N, Eichenberger A, Choutko A, Riniker S, Winger M, Mark A, Van Gunsteren W. Definition and testing of the GROMOS forcefield versions 54A7 and 54B7. Eur Biophys J 2011;40:843–856.
23. Schrödinger. MacPyMOL: PyMOL Enhanced for Mac OS X; 2012.
24. Baptista A, Teixeira V, Soares C. Constant-pH molecular dynamics using stochastic titration. J Chem Phys 2002;117:4184–4200.
25. Machuqueiro M, Baptista A. Constant-pH molecular dynamics with ionic strength effects: protonation-conformation coupling in decalysine. J Phys Chem B 2006;110:2927–2933.
26. Machuqueiro M, Baptista A. The pH-dependent conformational states of kyotorphin: a constant-pH molecular dynamics study. Biophys J 2007;92:1836–1845.
27. Machuqueiro M, Baptista A. Acidic range titration of HEWL using a constant-pH molecular dynamics method. Proteins 2008;72:289–298.
28. Machuqueiro M, Baptista A. Molecular dynamics at constant pH and reduction potential: application to cytochrome C3. J Am Chem Soc 2009;131:12586–12594.
29. Campos S, Machuqueiro M, Baptista A. Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein. J Phys Chem B 2010;114:12692–12700.
30. Machuqueiro M, Baptista A. Is the prediction of pKa values by constant-pH molecular dynamics being hindered by inherited problems? Proteins 2011;79:3437–3447.
31. Vila-Viçosa D, Campos S, Baptista A, Machuqueiro M. Reversibility of prion misfolding: insights from constant-pH molecular dynamics simulations. J Phys Chem B 2012;116:8812–8821.
32. Berendsen H, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 1995;91:43–56.
33. Lindahl E, Hess B, Van Der Spoel D. GROMACS 3.0:a package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–317.
34. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A, Berendsen H. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–1718.
35. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–447.
36. Bashford D, Gerwert K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. J Mol Biol 1992;224:473–486.
37. Teixeira V, Cunha C, Machuqueiro M, Oliveira A, Victor B, Soares C, Baptista A. On the use of different dielectric constants for computing individual and pairwise terms in Poisson-Boltzmann studies of protein ionization equilibrium. J Phys Chem B 2005;109:14691–14706.
38. Baptista A, Martel P, Soares C. Simulation of electron-proton coupling with a Monte Carlo method: application to cytochrome c3 using continuum electrostatics. Biophys J 1999;76:2978–2998.
39. Woodward CE. Density functional theory for inhomogeneous polymer solutions. J Chem Phys 1991;94:3183.
40. Xie F, Woodward CE, Forsman J. Fluid-fluid transitions at supercritical conditions. Langmuir 2013;29:2659–2666.
41. Dickman R, Hall CK. Equation of state for chain molecules: continuous-space analog of Flory theory. J Chem Phys 1986;85:4108.
42. Allen M, Tildesley D. Computer simulation of liquids, Vol. 18. New York: Oxford university press; 1989.
43. Nikawa H, Fukushima H, Makihira S, Hamada T, Samaranayake LP. Fungicidal effect of three new synthetic cationic peptides against *Candida Albicans*. Oral Dis 2004;10:221–228.
44. Raj PA, Marcus E, Sukumaran DK. Structure of human salivary histatin 5 in aqueous and nonaqueous solutions. Biopolymers 1998;45:51–67.
45. Lund M, Jönsson B. On the charge regulation of proteins. Biochemistry 2005;44:5722–5727.
46. Den Hertog AL, Wong Fong Sang HW, Kraayenhof R, Bolscher JGM, Van't Hof W, Veerman ECI, Nieuw Amerongen AV. Interactions of histatin 5 and histatin 5-derived peptides with liposome membranes: surface effects, translocation and permeabilization. Biochem J 2004;379:665–672.
47. Xu T, Levitz SM, Diamond RD, Oppenheim FG. Anticandidal activity of major human salivary histatins. Infect Immun 1991;59:2549–2554.
48. Edgar M, Dawes C, O'Mullane D, editors. Saliva and Oral Health, 3rd ed. London: British Dental Association; 2004.
49. Rudney J. Does variability in salivary protein concentrations influence oral microbial ecology and oral health? Crit Rev Oral Biol Med 1995;6:343–367.
50. Azen EA. Genetics of salivary protein polymorphisms. Crit Rev Oral Biol Med 1993;4:479–485.
51. Kacprzyk L, Rydengård V, Mörgelin M, Davoudi M, Pasupuleti M, Malmsten M, Schmidtchen A. Antimicrobial activity of histidine-rich peptides is dependent on acidic conditions. Biochim Biophys Acta 2007;1768:2667–2080.
52. Kacprzyk L, Rydengård V, Mörgelin M, Davoudi M, Pasupuleti M, Malmsten M, Schmidtchen A. Antimicrobial activity of histidine-rich peptides is dependent on acidic conditions. BBA - Biomembranes 2007;1768:2667–2680.
53. Sah N, More S, Bhutani H. Estimation and comparision of salivary calcium levels in healthy subjects and patients with gingivitis and periodontitis: a cross-sectional biochemical study. AOSR 2012;2:13–16.
54. Hassan SHA, Al-Sandook TA. Salivary calcium concentration in patients with high incidence of calculus formation. Al–Rafidain Dent J 2005;5:88–90.
55. Guldbrand L, Jönsson B, Wennerström H, Linse P. Electric double layer forces. A Monte Carlo study. J Chem Phys 1984;80:2221–2228.
56. Edsall JT, Felsenfeld G, Goodman DS, Gurd FRN. The association of imidazole with the ions of zinc and cupric copper. J Am Chem Soc 1954;76:3054–3061.
57. Bales C, Freeland-Graves J, Askey N, Behmardi F, Pobocik R, Fickel JJ, Greenlee P. Zinc, magnesium, copper, and protein concentrations in human saliva: age- and sex-related differences. Am J Clin Nutr 1990;51:462–469.
58. Harrap G, Best J, Saxton C. Human oral retention of zinc from mouthwashes containing zinc salts and its relevance to dental plaque control. Arch Oral Biol 1984;29:87–91.
59. Brewer D, Lajoie G. Evaluation of the metal binding properties of the histidine-rich antimicrobial peptides histatin 3 and 5 by electrospray ionization mass spectrometry. Rapid Commun Mass Sp 2000;14:1736–1745.

Paper V

High Expression of Midkine in the Airways of Patients with Cystic Fibrosis

Sara L. Nordin[1]*, Sandra Jovic[1]*, Anil Kurut[3], Cecilia Andersson[1], Anele Gela[1], Anders Bjartell[4], Matthias Mörgelin[2], Anders I. Olin[2], Mikael Lund[3], and Arne Egesten[1]

[1]Section of Respiratory Medicine and Allergology, and [2]Section of Infection Medicine, Department of Clinical Sciences Lund, Lund University, Skåne University Hospital, Lund, Sweden; [3]Division of Theoretical Chemistry, Lund University, Lund, Sweden; and [4]Division of Urological Cancers, Department of Clinical Sciences Malmö, Lund University, Skåne University Hospital, Malmö, Sweden

Mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene result in impaired host defense during cystic fibrosis (CF), where *Pseudomonas aeruginosa* becomes a key pathogen. We investigated the expression pattern of the antibacterial growth factor midkine (MK) in CF and the possible interference with its activity by the altered airway microenvironment. High MK expression was found in CF lung tissue compared with control samples, involving epithelia of the large and small airways, alveoli, and cells of the submucosa (i.e., neutrophils and mast cells). In CF sputum, MK was present at 100-fold higher levels, but was also subject to increased degradation, compared with MK in sputum from healthy control subjects. MK exerted a bactericidal effect on *P. aeruginosa*, but increasing salt concentrations and low pH impaired this activity. Molecular modeling suggested that the effects of salt and pH were attributable to electrostatic screening and a charge-neutralization of the membrane, respectively. Both the neutrophil elastase and elastase of *P. aeruginosa* cleaved MK to smaller fragments, resulting in impaired bactericidal activity. Thus, MK is highly expressed in CF, but its bactericidal properties may be impaired by the altered microenvironment, as reflected by the *in vitro* conditions used in this study.

Keywords: cystic fibrosis; midkine; *Pseudomonas aeruginosa*; host defense

Host defense functions of the airways are impaired in cystic fibrosis (CF), resulting in long-term colonization with bacterial pathogens, and especially *Pseudomonas aeruginosa*. The presence of *P. aeruginosa* causes bouts of inflammation, and in most cases disease progression (1). Host defense is affected at several levels in CF, including defects in mucociliary clearance, dysregulated inflammation, and the accumulation of inflammatory effector cells (in particular, neutrophils) (2). CF originates from mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Abnormal electrolyte transport across airway epithelia, resulting in increased salt concentrations and/or a reduced depth of the airway surface liquid (ASL), has been hypothesized to cause the initial CF host-defense defect (3–6). Such physical alterations of ASL may affect the antibacterial activity of host defense proteins (HDPs), as exemplified by the bactericidal β defensins (hBD-1 and hBD-2), which are salt-sensitive and thus possibly inactivated during CF (7, 8). Recently, a lack of *CFTR* in a porcine model was not found to increase transepithelial Na^+ or liquid absorption, or to reduce the depth of ASL (9). Instead, the lower pH of ASL in *CFTR* knockout pigs resulted in a reduced antibacterial activity of innate antibiotics (10).

Recently, we showed that midkine (MK) and pleiotrophin, the two members of a family of heparin-binding growth factors, demonstrate strong antibacterial activity *in vitro* (11, 12). In addition, MK is constitutively expressed in the epithelia of large airways, and may serve as an HDP in this context (13). In this study, we investigated MK expression in the airways during CF, and we investigated how the bactericidal activity of this molecule is affected by salt, pH, and the proteolysis brought about by the elastases of neutrophils and by *P. aeruginosa*.

CLINICAL RELEVANCE

This is the first report on the expression of the antibacterial growth factor midkine in cystic fibrosis. Midkine exerts bactericidal activity against *Pseudomonas aeruginosa*, but the activity is impaired by increasing salt concentrations and lowered pH. In addition, elastase of *P. aeruginosa* degrades midkine, resulting in a lower antibacterial activity. These findings can provide targets for therapeutic intervention.

MATERIALS AND METHODS

Special Reagents

Rabbit antibodies to MK, preimmune rabbit IgG, and recombinant human MK were obtained from Peprotech (London, UK). Elastase from *P. aeruginosa* was obtained from Merck Millipore (Solna, Sweden), and neutrophil elastase was obtained from Sigma-Aldrich (St. Louis, MO).

In Situ Hybridization

Lung tissue was obtained from individuals undergoing lung transplantation for endstage CF and we also acquired excess lung tissue from previously healthy donors at transplantation, as approved by the Ethics Committee of Lund (LU412-03). The *in situ* hybridization procedure is described in the online supplement.

Immunohistochemistry

A staining protocol (EnVision K5007; DAKO, Glostrup, Denmark) was used to visualize bound MK antibodies in thin sections. In the case of double staining, MK immunoreactivity was visualized with a goat anti-rabbit Alexa-Fluor 555–conjugated antibody (diluted 1:200; Molecular

(*Received in original form March 7, 2013 and in final form June 6, 2013*)

*These authors contributed equally to this study.

This work was supported by Swedish Research Council grant 2010-4224, Swedish Heart and Lung Foundation grant 20100164, the Medical Faculty of Lund University, Swedish Government Funds for Clinical Research, the Bergh Foundation, the Greta and Johan Kock Foundation, and the Alfred Österlund Foundation. Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) through Lunarc, the Center for Scientific and Technical Computing at Lund University. The work was performed within the framework of the Swedish national strategic e-science research program eSSENCE.

Correspondence and requests for reprints should be addressed to Arne Egesten, M.D., Ph.D., Section of Respiratory Medicine and Allergology, Department of Clinical Sciences Lund, Lund University, Skåne University Hospital, Tornavägen 10, BMC B14, SE-22184 Lund, Sweden. E-mail: Arne.Egesten@med.lu.se

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Cell Mol Biol Vol 49, Iss. 6, pp 935–942, Dec 2013
Copyright © 2013 by the American Thoracic Society
Originally Published in Press as DOI: 10.1165/rcmb.2013-0106OC on July 1, 2013
Internet address: www.atsjournals.org

Probes, Eugene, OR). Mouse monoclonal antibodies to surfactant protein–A (SP-A), myeloperoxidase, or chymase were visualized using an Alexa-Fluor 488–conjugated goat anti-mouse antibody (1:200; Molecular Probes). Nuclei were visualized using Hoechst 33342 stain. No staining was observed in sections using isotype-matched control antibodies.

Bacteria

P. aeruginosa strain PA01 and the clinical CF strains have been described previously (14–16). All strains included nonmucoid phenotypes, to facilitate the viable count assay.

Viable Count Assay

Viable counts were performed as described elsewhere (13). Bacteria grown to midlogarithmic phase were diluted in a buffer containg Tris-HCl (10 mM) and glucose (5 mM), pH 7.4. To assess the antibacterial activity of cleaved MK by *P. aeruginosa* elastase or elastase from neutrophils, MK was preincubated with *P. aeruginosa* elastase or the elastase of neutrophils for different times before the viable count assay.

Electron Microscopy

Details of electron microscopy are described in the online supplement.

Flow Cytometry and Entrance of Vital Dyes

Details of flow cytometry and the entrance of vital dyes are described in the online supplement.

Metropolis Monte Carlo Simulations

Metropolis Monte Carlo simulations were used to determine the net charge of MK as a function of pH, and also the interaction-free energy of MK with a charged surface (i.e., a bacterial cell membrane) (17–19). The MK molecule was described as a flexible bead model, each bead representing an amino acid (*see* Figure E1 in the online supplement). The canonical ensemble of peptide configurations and protonation states was explored. and the total energy for a given configuration was calculated as described elsewhere (20).

Sputum Samples

Sputum samples were obtained from four patients (20–54 years of age) suffering from CF and colonized with *P. aeruginosa*. Induced sputum from three healthy individuals (32–54 years of age) was obtained after the inhalation of 5 ml nebulized sodium chloride (30 mg/ml). All subjects gave their written, informed consent to participate in the study, which was approved by the Ethics Committee in Lund (2011-434).

SDS-PAGE and Western Blotting

MK (5 μg) was incubated with *P. aeruginosa* elastase (0.5 μg), neutrophil elastase (0.5 μg), conditioned medium from overnight cultures with *P. aeruginosa* (strain PA01), or PBS alone for 1 hour or 18 hours at 37°C. The proteins were separated on Tris-tricine gels, and visualized by Coomassie staining. For Western blotting, the peptides were first transferred onto a PVDF membrane (Merck Millipore) using a blotting system, and subsequent procedures were performed according to the instructions of the manufacturer (Bio-Rad, Hercules, CA).

Mass Spectrometry

Details of mass spectrometry are described in the online supplement.

RESULTS

Increased Expression of MK in the Airways during CF

Lung tissue obtained from individuals undergoing lung transplantation for endstage CF was compared with excess lung tissue from previously healthy donors at transplantation (described in Table 1). *In situ* hybridization showed *MK* gene expression in the epithelium of both the large and small airways, and to a lesser extent in alveoli (Figures 1A–1C). At the protein level, a similar pattern was observed, with the detection of MK protein in epithelial cells of the large and small airways (Figures 1E–1G). A morphometric analysis of the immunohistochemical staining showed a significantly higher level of detection at all sites investigated, namely, in the airway wall (Figure 1I), in the subepithelial compartment (Figure 1J), and in the epithelium (Figure 1K). In alveoli, double staining for MK and SP-A revealed the presence of MK in Type 2 pneumocytes (Figure 1L). In scattered subepithelial cells of the large and small airways, double staining also revealed the presence of MK in neutrophils (Figure 1M) and mast cells (Figure 1N).

TABLE 1. CHARACTERISTICS OF PATIENTS WITH CF AND CONTROL SUBJECTS

	Control Subjects	Patients with CF
Sex, M/F, *n*	2/6, 8	2/3, 5
Age, yr	63 (33–76)	30 (23–38)
Current smokers, y/n	0	0
Ex-smokers, y/n	0	0
Inhaled GCS, y/n	0	5/0
Oral GCS, y/n	0	3/2
β_2 agonist, y/n	0	5/0
Lung function		
FEV_1 percent predicted*	110 (82–141)	31 (22–45)
FEV_1/(F)VC*	86 (66–121)	50 (33–84)
VC percent predicted*	104 (82–126)	54 (46–70)
TLC percent predicted*	ND	104 (79–129)
RV percent predicted*	ND	215 (113–318)
Percent TLCO SB	ND	60 (50–71)

Definition of abbreviations: CF, cystic fibrosis; F, female; FEV_1, forced expiratory volume in 1 second; (F)VC, (forced) vital capacity; GCS, glucocorticosteroid; M, male; ND, not determined; RV, residual volume; SB, single-breath; TLC, total lung capacity; TLCO, diffusing capacity for carbon monoxide; VC, vital capacity; y/n, yes/no.
*Data are presented as mean (range).

MK Demonstrates Antibacterial Activity against *P. aeruginosa*

Using a viable count assay to investigate bactericidal activity, we found that MK demonstrated dose-dependent antibacterial activity against *P. aeruginosa* (Figures 2A–2D). The different strains of *P. aeruginosa* used in this study were sensitive to MK at somewhat varying concentrations. The antibacterial activities of MK against the most commonly used laboratory strain of *P. aeruginosa*, PA01, and a clinical isolate from a patient with CF (strain 032) were very similar. The antibacterial effects of 0.1 μM MK for these strains were approximately 90%, and complete bacterial killing was seen when the concentration of MK reached 0.3 μM (Figures 2A and 2D). In the case of the other two clinical isolates obtained from patients with CF (strains 335A and 022A), higher concentrations of MK were required. Only 10% bacterial killing was seen at an MK concentration of 0.1 μM, and for complete bacterial killing, an MK concentration of 1 μM was needed (Figures 2B and 2C). MK showed no cytotoxicity against bronchial epithelial cells in this range of concentrations, as detected by the release of lactate dehydrogenase (data not shown).

Both scanning electron microscopy and negative staining (transmission electron microscopy) were performed to visualize morphological changes of the bacteria (strain PA01) after incubation with MK (Figure 2E). After exposure, the bacteria exhibited a lytic appearance, including blebs and the release of intracellular contents, indicating a loss of membrane integrity.

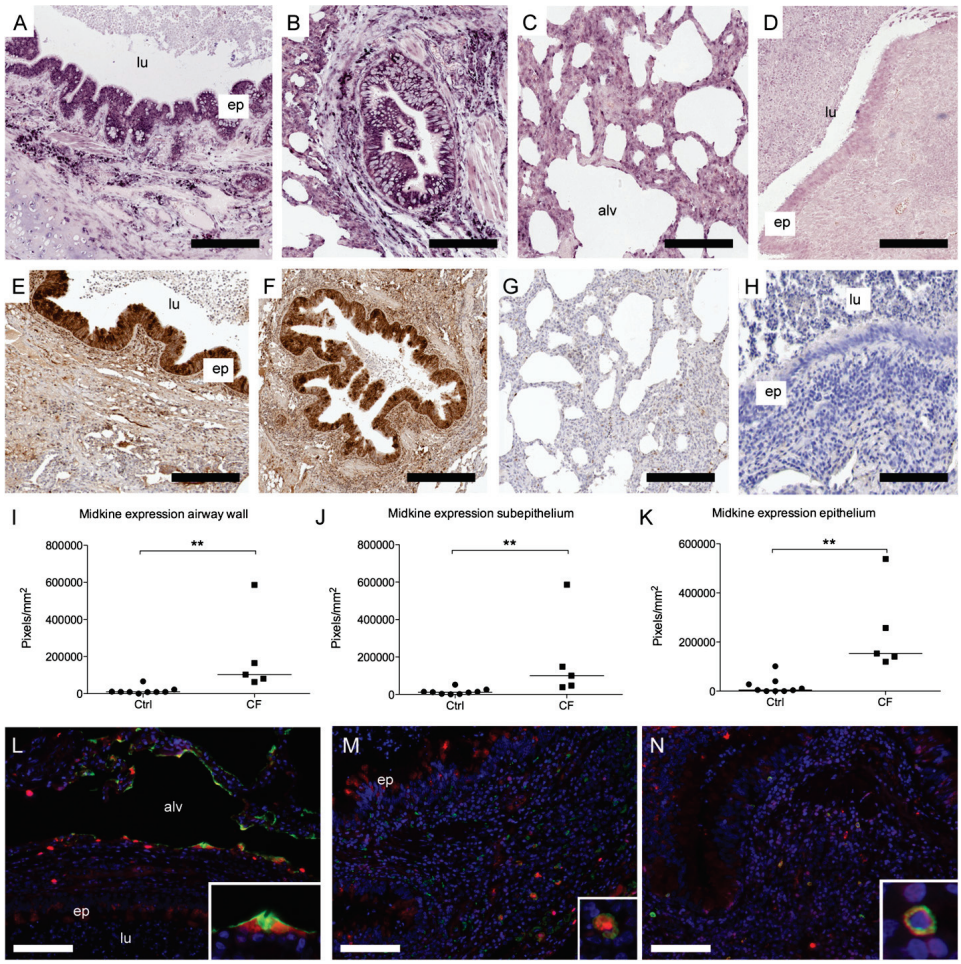


Figure 1. Midkine (MK) expression in the airways during cystic fibrosis (CF). *In situ* hybridization showed a high expression of MK in the epithelium (ep) of both the large (*A*) and small (*B*) airways. (*C*) Hybridization appears as *violet* staining. Abundant neutrophils are seen in the lumen (lu). Only weak expression was detected in alveoli (alv). (*D*) Hybridization with a scrambled control oligonucleotide resulted in a loss of labeling. Immunohistochemical staining, where the presence of MK appears as a *brownish* color, was used to detect MK in the central airways (*E*), small airways (*F*), and alveolar parenchyma (*G*). The negative control, obtained by replacing the primary antibody with preimmune IgG, is shown in *H*. The density of MK immunoreactivity, expressed as positive pixels per square millimeter in the entire airway wall (*I*), and subdivided into the subepithelium (*J*) and epithelium (*K*), was measured and showed significant differences at these levels (**$P < 0.01$, as assessed by Student *t* test for paired observations with unequal variance). Triple fluorescent immunohistochemical staining, where MK (*red*, Alexa-Fluor 555; Molecular Probes) colocalized with surfactant protein–A (*L*), neutrophils (*M*), myeloperoxidase, and mast cells (*N*) (tryptase), is shown in *green* (Alexa-Fluor 488). The DNA of nuclei is stained with 4′6-diamidino-2-phenylindole (*blue*). *Scale bars* in *A*, *B*, *D–G*, and *L–N*, 200 μm; *scale bars* in *C* and *H*, 100 μm. *Inserts* depict a type 2 pneumocyte (*L*), a neutrophil (*M*), and a mast cell (*N*) respectively. Ctrl, control.

Membrane integrity can be assessed by the entry of vital dyes, and propidium iodide (PI) was previously used to study membrane damage in bacteria (21). Thus, flow cytometry was used to detect the entrance of PI, and after 1 hour of incubation with MK (1 μM), the membrane integrity of bacteria was lost, allowing for the internalization of PI (Figure 2F).

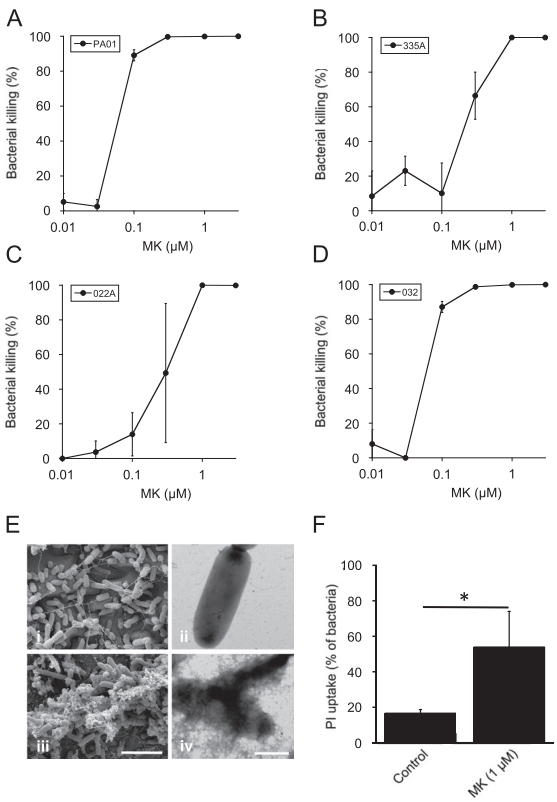


Figure 2. Bactericidal activity of MK against *Pseudomonas aeruginosa.* The bactericidal activity of MK against the *P. aeruginosa* strains PA01 (*A*), 335A (*B*), 022A (*C*), and 032 (*D*) was determined according to a viable count assay. Bacteria were incubated with MK at the concentrations indicated for 1 hour, followed by plating on agar dishes. The resultant colony counts were compared with the number after incubation in buffer alone, and are expressed as percentage of killing. The data represent the mean (± SD) from three separate experiments. (*E*) Scanning electron micrographs (*i* and *iii*) and negative staining, followed by transmission electron microscopy (*ii* and *iv*), show the morphological appearance of *P. aeruginosa* (strain PA01) after incubation in buffer alone, where the bacteria appeared intact (*i* and *ii*), and after 1 hour with MK (1 μM), where blebbing and the release of intracellular contents were seen, indicating membrane disruption (*iii* and *iv*). *Scale bars* in *i* and *iii*, 2 μm. *Scale bars* in *ii* and *iv*, 500 nm. (*F*) *P. aeruginosa* (strain PA01) was incubated in buffer or in the presence of MK (1 μM), followed by incubation with the vital dye propidium iodide (PI), and flow cytometry was then performed to detect permeabilization. Bacteria incubated with MK showed permeabilization, as indicated by the increased internalization of PI. The data represent the mean (± SD) from three separate experiments. *$P < 0.05$, as assessed by Student *t* test.

Effects of Salt and pH on the Antibacterial Properties of MK

The antibacterial properties of MK at different pH values and salt concentrations were investigated using a viable count assay. A decrease in bacterial killing was observed as the pH decreased and as the salt concentration increased. Inconsistent results have been obtained regarding Na^+ concentrations in the ASL of healthy lung tissue, ranging from 50 mM to 140 mM, with a pH of around 7.4 (3–6). In a porcine model of CF, the ASL demonstrated Na^+ at 140 mM, whereas the pH was around 7.0 or less (10). Interestingly, the bactericidal activity of MK showed both a pH-dependent and salt-dependent decrease (Figure 3A).

As seen in Figure 3C, the modeled net charge of MK is largely constant in a pH range of 5–8, suggesting that pH-driven changes in killing activity are not attributable to changes in the protonation state of MK. We therefore investigated possible changes of the bacterial membrane with changes in pH. Although the conditions for calculating the net charge of MK in bulk solution are well determined (pK_a, pH, salt, sequence), the determination of the pH-dependence of the membrane surface charge density remains more hypothetical. This is because the exact lipid composition of *P. aeruginosa* is unknown, as are the stoichiometric phosphate pK_a values in the complex membrane environment. Previous studies of bacterial membranes have shown surface charge densities ranging from -0.00022 to -0.076 $e/Å^2$, depending on the lipid composition and on physiological conditions (22–25). For instance, increasing the pH from 3.5 to 7 resulted in a charge inversion from positive to negative of lipid films composed of phosphatidylserine and phosphatidylcholine (28). Because of this arbitrariness, we assigned the charge densities -0.0016, -0.0025, and -0.0033 $e/Å^2$ to pH 6, 6.5, and 7.4 by assuming a head group density of 60 $Å^2$/lipid (26), 20% charged groups, and an effective pK_a of 6 (22). This allowed us to qualitatively predict the pH-dependent peptide–surface interaction energy, whereas for a quantitative result, a more detailed knowledge of membrane characteristics must be obtained. Because the assigned charge densities are sufficiently low (meaning that the charged groups are far from each other and thus are not interacting), the protonation of these groups is assumed to be independent of salt concentration. As seen in Figure 3B, by increasing the salt concentration, the binding of MK to the membrane is dramatically reduced because of electrostatic screening by the added salt ions. This reduction correlates well with the decrease in antimicrobial activity, as seen in Figure 3A. Increasing the surface charge density, thus mimicking a pH increase, results in stronger surface affinities because of

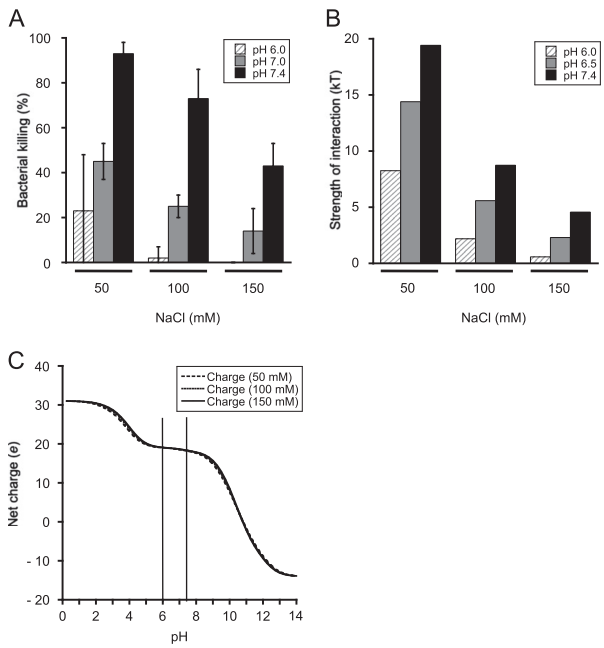


Figure 3. The influence of salt and pH on the bactericidal activity of MK against *P. aeruginosa*. (*A*) *P. aeruginosa* (strain PA01) was incubated with MK (1 μM) at the pH and sodium chloride (NaCl) concentrations indicated for 1 hour, and thereafter tested by a viable count assay. Both pH and sodium chloride caused concentration-dependent reductions of bacterial killing ($P < 0.01$, two-way ANOVA, followed by the Holm-Sidak *post hoc* test; $n = 4$ for each combination of pH and NaCl). The data represent the mean (± SD) of four separate experiments. (*B*) The strength of the electrostatic interaction between MK and the bacterial plasma membrane was predicted using Monte Carlo simulations. The results paralleled the findings from the viable count assay in *A*. (*C*) The net charge of MK was calculated at different sodium chloride concentrations and pH values. At the sodium chloride concentrations and in the pH range used for the viable count experiments, the net charge of MK was unaffected, suggesting that changes in net charge at the bacterial surface were the main determinants of bactericidal activity.

the stronger electrostatic attraction. Again, this correlates well with experimental observations in the viable count assay.

Detection of MK in Sputum and Degradation by Elastases

In sputum from patients with CF, MK was detected at 9.2 ng/ml (0.7 nM; range, 1.7–26 ng/ml, or 0.1–1.9 nM; $n = 6$). In induced sputum from healthy individuals, mean MK levels of 105 pg/ml (0.8 pM) have previously been reported (range, 95–116 pg/ml, or 0.7–0.9 pM) (13). In the case of induced sputum, inhaled nebulized sodium chloride could contribute to a diluting effect.

To characterize MK in sputum further, SDS-PAGE followed by Western blotting was performed on sputum from five patients with CF colonized with *P. aeruginosa*, in parallel with induced sputum from three healthy control subjects (Figure 4A). Faint bands corresponding to the recombinant MK holoprotein was detected in the sputum of control subjects, whereas a band corresponding to a slightly smaller amount of MK was seen in the sputum of both patients with CF and healthy individuals. In addition, sputum from patients with CF contained additional bands of lower molecular weight, indicating an increased degradation of MK.

MK is likely to encounter the elastase of both neutrophils and *P. aeruginosa* in CF airways. Thus, MK was incubated with these elastases, followed by separation using SDS-PAGE (Figure 4B). Similar fragments were obtained when MK was cleaved with elastase from neutrophils or from *P. aeruginosa*. However, neutrophil elastase cleaved MK more efficiently than elastase from *P. aeruginosa*. After 1 hour of incubation of MK with neutrophilic elastase, the holoprotein was degraded into three fragments, and after 18 hours of incubation, one single band was detected. Most of the holoprotein was observed after 1 hour of incubation of MK with *P. aeruginosa* elastase, and after 18 hours of incubation, most of the MK was degraded into smaller fragments (Figure 4B). To investigate whether the degradation impaired bactericidal activity, aliquots of the digests were incubated with *P. aeruginosa* (strain PA01) in the viable count assay (Figure 4C). The degradation of MK caused by the elastase of neutrophils showed the most profound decrease in antibacterial activity. Moreover, the antibacterial activity of MK was completely lost when an 18-hour digest of MK with elastase from neutrophils was performed. This correlates very well with the protein-stained gel, where after 1 hour of incubation, the holoprotein was lost and only fragments were detected, and after the 18-hour incubation, only one faint fragment was seen, indicating the almost complete degradation of MK. After a 1-hour digest of MK with *P. aeruginosa* elastase, no decrease in antibacterial activity was observed. However, the antibacterial activity of MK decreased after an 18-hour digest of MK with *P. aeruginosa* elastase (Figure 4C). This also correlates well with the protein-stained gel, where the holoprotein was seen after 1 hour, and after 18 hours of incubation, fragments were seen as well as a faint band of the MK holoprotein.

The three fragments obtained after 18 hours of the cleavage of MK with *P. aeruginosa* elastase were analyzed by mass spectrometry and N-terminal protein sequencing (Figure 4D). The MK molecule was most prone to degradation at its C-terminal tail, which is responsible for most of the antibacterial activity (13). In addition, the N-terminus was also degraded (Figure 4D).

The incubation of MK with conditioned medium (CM) from *P. aeruginosa* (strain PA01) produced some degradation of MK after 3 hours of incubation (Figure 4E). However, after 18 hours of incubation of MK with the CM, only faint bands were

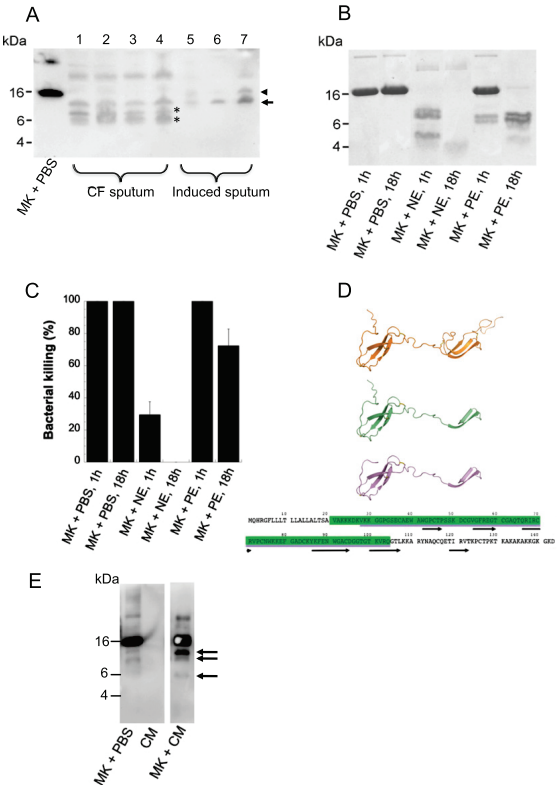


Figure 4. Detection of MK in the sputum of patients with CF, and degradation by human neutrophil elastase and the elastase of *P. aeruginosa in vitro*, impairing its bactericidal activity. (*A*) Sputum from five patients with CF (*lanes 1–5*) colonized with *P. aeruginosa* and induced sputum from three healthy control subjects (*lanes 5–7*) were separated by SDS-PAGE and subjected to Western blotting to detect MK. Faint bands corresponding to the recombinant MK holoprotein is seen in the sputum of control subjects (*arrowhead*), whereas a slightly smaller band is seen in both healthy individuals and patients with CF (*arrow*). In the sputum of patients with CF, additional bands of lower molecular weight indicate increased degradation in MK (*asterisks*). (*B*) MK was incubated in buffer alone (PBS), with neutrophil elastase (NE) or with the elastase of *P. aeruginosa* (PE) for 3 or 18 hours, followed by separation on SDS-PAGE and visualization by Coomassie staining. Both proteases caused degradation to three distinct forms, which were analyzed by N-terminal sequencing and mass spectrometry, where the sequences on both sides of the two domains were degraded (*D*). One representative experiment out of three separate experiments is shown. After 18 hours, neutrophil elastase had completely degraded MK. The upper model shows the intact midkine holoprotein (*orange*), the *green* model (*middle*) and the *violet* model (*lower*) depicts the two fragments resulting from degradation by elastase of *P. aeruginosa*. The corresponding sequences are indicated by color in the protein sequence. (*C*) The bactericidal activity remaining after degradation was investigated using a viable count assay with *P. aeruginosa* (strain PA01), resulting in a loss of bactericidal activity, whereas in the case of *P. aeruginosa* elastase, much of the bactericidal activity was retained. Both the neutrophil elastase and elastase of *P. aeruginosa* demonstrated a significant time-dependent reduction of bacterial killing ($P < 0.01$, two-way ANOVA, followed by the Holm-Sidak *post hoc* test; $n = 3$ for each elastase). (*E*) Conditioned medium from *P. aeruginosa* (strain PA01), incubated with MK, resulted in three bands of similar size after 3 hours (*D*, *arrows*).

detected, indicating that MK had been almost completely degraded by proteases released from the bacteria.

DISCUSSION

In this study, we showed that MK is expressed in the airways of patients with CF at all levels, in contrast to healthy individuals, in whom a lower expression was found, mainly in the epithelium of the large airways (13). The bactericidal activity of MK was diminished both by increasing the salt concentration and by decreasing pH, counteracting its role as an innate antibiotic in CF. In addition, neutrophil elastase (and to some extent, the elastase of *P. aeruginosa*) degraded MK, resulting in impaired bactericidal activity.

Retinoic acid is an important factor for the maintenance of airway integrity, and it also promotes the expression of MK in many tissues during health (11). The increased expression of MK in CF may be explained by several factors, including hypoxia (hypoxia-inducible factor–1α), the activation of NF-κB by proinflammatory cytokines, and oxidative stress (27–29). All these prerequisites for the increased expression of MK are likely to be present, especially in advanced CF disease, which often involves hypoxia. A high level of chronic inflammation involving NF-κB is seen in CF, and this transcription factor, which activates the transcription of proinflammatory mediators, has been suggested as a pharmaceutical target to treat the disease (30). The pyocyanin of *P. aeruginosa* produces reactive oxygen species (ROS) that cause the expression of proinflammatory cytokines in epithelial cells of the airway (31). In addition, a high accumulation of activated neutrophils occurs, producing ROS in CF airways (32).

The concentration and function of antimicrobial proteins are likely to be most important in the thin liquid layer constituting the ASL on the epithelial cell surface (9). In a previous study, we determined the MK concentration in ASL, using an air–liquid interface model, and found a concentration of 0.7 μM, corresponding to a bactericidal level (13). The composition of sputum differs in many ways from that of ASL. In particular, sputum contains a high content of anionic mucins. Thus, the MK content of sputum likely reflects events occurring at the epithelial cell surface. However, bactericidal activity is less likely to be conducted in the context of sputum, and more likely to be conducted in the ASL.

The antibacterial activity of many HDPs decreases with the presence of salt, a feature long believed to explain part of the impaired host defense in CF (7, 8). Recently, the antibacterial activity of lactoferrin and lysozyme, two major antibacterial proteins of ASL, was shown to become reduced at lower

pH (10, 33). In the case of MK, our results show that the net charge of the molecule is mostly unaffected by pH values in the physiological range. Instead, the charge on the bacterial membrane may become neutralized, thus weakening the disruptive properties of MK. In addition to its bacteriolytic properties, lactoferrin mediates bacteriostatic effects through the sequestration of iron, and lysozyme demonstrates muraminidase activity, resulting in the degradation of the bacterial cell wall (33). However, MK mediates its bactericidal activity through membrane disruption, and the pH-dependent level of activity is likely a result of changes in electrostatic attraction with the bacterial membrane for all three proteins (i.e., lactoferrin, lysozyme, and MK) (12). Because most HDPs kill bacteria via membrane disruption, the protonation of the bacterial membrane is likely a general effect of their antibacterial activity. Taken together, the effects of salt and pH are attributable to electrostatic screening and a charge-neutralization of the membrane, respectively. *P. aeruginosa* is also known to alter the composition of its lipid membrane, resulting in a reduced anionic charge, increasing its protection against HDPs (34).

MK may contribute to airway remodeling, which begins early in CF and becomes extensive with disease progression (35). In an animal model, MK was found to contribute to vascular remodeling during hypoxia, and it may therefore play an important role in this aspect of CF as well (27).

During CF, inflammation results in a high accumulation of neutrophils, with the release of serine proteases (e.g., elastase in the airways) (36). The elastase of *P. aeruginosa* is also likely to play a role in interfering with the function of HDPs such as SP-A and SP-D (37). Both elastases showed a similar pattern, degrading the N-terminus and the C-terminus of MK, with the C-terminus exhibiting high antibacterial activity (13). Containing antiparallel β-sheets, the two domains seem more resistant to proteolysis. However, when discussing proteolytic activity, it is important to bear in mind that sputum is a mucin-rich environment with many necrotic cells, which may not reflect the conditions that prevail during critical events at the epithelial cell surface.

In conclusion, MK is highly expressed in CF, but its bactericidal properties may be impaired by the altered microenvironment, as reflected by the mimicked conditions *in vitro* used in this study.

Author disclosures are available with the text of this article at www.atsjournals.org.

***Acknowledgments*:** The authors are grateful to Pia Andersson, Maria Baumgarten, and Christina Möller for skillful technical assistance, and to Dr. Oonagh Shannon for help with flow cytometry.

References

1. O'Sullivan BP, Freedman SD. Cystic fibrosis. *Lancet* 2009;373:1891–1904.
2. Hartl D, Gaggar A, Bruscia E, Hector A, Marcos V, Jung A, Greene C, McElvaney G, Mall M, Döring G. Innate immunity in cystic fibrosis lung disease. *J Cyst Fibros* 2012;11:363–382.
3. Joris L, Dab I, Quinton PM. Elemental composition of human airway surface fluid in healthy and diseased airways. *Am Rev Respir Dis* 1993;148:1633–1637.
4. Smith JJ, Travis SM, Greenberg EP, Welsh MJ. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* 1996;85:229–236.
5. Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW, Boucher RC. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 1998;95:1005–1015.
6. Guggino WB. Cystic fibrosis and the salt controversy. *Cell* 1999;96:607–610.
7. Goldman MJ, Anderson GM, Stolzenberg ED, Kari UP, Zasloff M, Wilson JM. Human beta–defensin–1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell* 1997;88:553–560.
8. Bals R, Wang X, Wu Z, Freeman T, Bafna V, Zasloff M, Wilson JM. Human beta–defensin 2 is a salt-sensitive peptide antibiotic expressed in human lung. *J Clin Invest* 1998;102:874–880.
9. Chen JH, Stoltz DA, Karp PH, Ernst SE, Pezzulo AA, Moninger TO, Rector MV, Reznikov LR, Launspach JL, Chaloner K, *et al*. Loss of anion transport without increased sodium absorption characterizes newborn porcine cystic fibrosis airway epithelia. *Cell* 2010;143:911–923.
10. Pezzulo AA, Tang XX, Hoegger MJ, Alaiwa MH, Ramachandran S, Moninger TO, Karp PH, Wohlford-Lenane CL, Haagsman HP, van Eijk M, *et al*. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nature* 2012;487:109–113.
11. Muramatsu T. Midkine and pleiotrophin: two related proteins involved in development, survival, inflammation and tumorigenesis. *J Biochem* 2002;132:359–371.
12. Svensson SL, Pasupuleti M, Walse B, Malmsten M, Mörgelin M, Sjögren C, Olin AI, Collin M, Schmidtchen A, Palmer R, *et al*. Midkine and pleiotrophin have bactericidal properties: preserved antibacterial activity in a family of heparin-binding growth factors during evolution. *J Biol Chem* 2010;285:16105–16115.
13. Nordin SL, Andersson C, Bjermer L, Bjartell A, Mörgelin M, Egesten A. Midkine is part of the antibacterial activity released at the surface of differentiated bronchial epithelial cells. *J Innate Immun* 2013;5:519–530.
14. Holloway BW. Genetic recombination in *Pseudomonas aeruginosa*. *J Gen Microbiol* 1955;13:572–581.
15. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, *et al*. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 2000;406:959–964.
16. Carlsson M, Shukla S, Petersson AC, Segelmark M, Hellmark T. *Pseudomonas aeruginosa* in cystic fibrosis: pyocyanin negative strains are associated with BPI-ANCA and progressive lung disease. *J Cyst Fibros* 2011;10:265–271.
17. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087–1092.
18. Evers CH, Andersson T, Lund M, Skepö M. Adsorption of unstructured protein β-casein to hydrophobic and charged surfaces. *Langmuir* 2012;28:11843–11849.
19. Zhou H. A Gaussian chain model for treating residual charge–charge interactions in the unfolded state of proteins. *Proc Natl Acad Sci USA* 2002;99:3569–3574.
20. Lund M, Persson B, Trulsson M. Faunus: a framework for molecular simulation. *Source Code Biol Med* 2008;3:1.
21. Gant VA, Warnes G, Phillips I, Savidge GF. The application of flow cytometry to the study of bacterial responses to antibiotics. *J Med Microbiol* 1993;39:147–154.
22. Müller DJ, Engel A. The height of biomolecules measured with the atomic force microscope depends on electrostatic interactions. *Biophys J* 1997;73:1633–1644.
23. Butt HJ. Measuring local surface charge densities in electrolyte solutions with a scanning force microscope. *Biophys J* 1992;63:578–582.
24. Wasserman E, Felmy AR. Computation of the electrical double layer properties of semipermeable membranes in multicomponent electrolytes. *Appl Environ Microbiol* 1998;64:2295–2300.
25. van der Wal A, Norde W, Zehnder AJB, Lyklema J. Determination of the total charge in the cell walls of Gram-positive bacteria. *Colloids Surf B Biointerfaces* 1997;9:81–100.
26. Moncelli MR, Becucci L, Guidelli R. The intrinsic pKa values for phosphatidylcholine, phosphatidylethanolamine, and phosphatidylserine in monolayers deposited on mercury electrodes. *Biophys J* 1994;66:1969–1980.
27. Reynolds PR, Mucenski ML, Le Cras TD, Nichols WC, Whitsett JA. Midkine is regulated by hypoxia and causes pulmonary vascular remodeling. *J Biol Chem* 2004;279:37124–37132.
28. You Z, Dong Y, Kong X, Beckett LA, Gandour-Edwards R, Melamed J. Midkine is a NF-kappaB–inducible gene that supports prostate cancer cell survival. *BMC Med Genomics* 2008;1:6.

29. Hobo A, Yuzawa Y, Kosugi T, Kato N, Asai N, Sato W, Maruyama S, Ito Y, Kobori H, Ikematsu S, *et al*. The growth factor midkine regulates the renin–angiotensin system in mice. *J Clin Invest* 2009;119:1616–1625.
30. Bodas M, Vij N. The NF-kappaB signaling in cystic fibrosis lung disease: pathophysiology and therapeutic potential. *Discov Med* 2010;9:346–356.
31. Rada B, Gardina P, Myers TG, Leto TL. Reactive oxygen species mediate inflammatory cytokine release and EGFR-dependent mucin secretion in airway epithelial cells exposed to *Pseudomonas pyocyanin*. *Mucosal Immunol* 2011;4:158–171.
32. Kolpen M, Hansen CR, Bjarnsholt T, Moser C, Christensen LD, van Gennip M, Ciofu O, Mandsberg L, Kharazmi A, Döring G, *et al*. Polymorphonuclear leucocytes consume oxygen in sputum from chronic *Pseudomonas aeruginosa* pneumonia in cystic fibrosis. *Thorax* 2010;65:57–62.
33. Singh PK, Tack BF, McCray PB Jr, Welsh MJ. Synergistic and additive killing by antimicrobial factors found in human airway surface liquid. *Am J Physiol Lung Cell Mol Physiol* 2000;279:L799–L805.
34. Roy H, Dare K, Ibba M. Adaptation of the bacterial membrane to changing environments using aminoacylated phospholipids. *Mol Microbiol* 2009;71:547–550.
35. Regamey N, Jeffery PK, Alton EW, Bush A, Davies JC. Airway remodelling and its relationship to inflammation in cystic fibrosis. *Thorax* 2011;66:624–629.
36. Sagel SD, Wagner BD, Anthony MM, Emmett P, Zemanick ET. Sputum biomarkers of inflammation and lung function decline in children with cystic fibrosis. *Am J Respir Crit Care Med* 2012;186:857–865.
37. Mariencheck WI, Alcorn JF, Palmer SM, Wright JR. *Pseudomonas aeruginosa* elastase degrades surfactant proteins A and D. *Am J Respir Cell Mol Biol* 2003;28:528–537.

Paper VI

Food Hydrocolloids 36 (2014) 332–338

Contents lists available at ScienceDirect

Food Hydrocolloids

journal homepage: www.elsevier.com/locate/foodhyd

Adsorption of β-casein to hydrophilic silica surfaces. Effect of pH and electrolyte

Olof Svensson, Anıl Kurut, Marie Skepö*

Division of Theoretical Chemistry, Lund University, P.O.B. 124, SE-221 00 Lund, Sweden

ARTICLE INFO

Article history:
Received 30 June 2013
Accepted 15 September 2013

Keywords:
Milk protein
Casein
Surfaces
Hydrophilic silica
Adsorption

ABSTRACT

Adsorption of β-casein to hydrophilic silica surfaces has been studied as an effect of pH and electrolyte, in the latter concentration, valency, and specificity (calcium or magnesium) have been considered. The used protein concentration has been an order of magnitude below the critical aggregation concentration, which implies that the protein is in monomeric form. By varying the salt concentration, the pH, and the concentration of divalent ions as calcium and magnesium, it is clearly shown that electrostatic interactions are of importance for adsorption of β-casein to silica surfaces and tunes the adsorbed amount and saturation of the surface. Our results show that there is counterbalance between: (i) electrostatic repulsion between the surface and the protein, (ii) electrostatic attraction between positively charged amino acids in the protein and the surface, and (iii) electrostatic repulsion and excluded volumes between adsorbed proteins at the surface, and that the positively charged amino acids serve as anchoring points.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

β-casein (together with α_{S1}-casein) is the most abundant milk protein and belongs to the Ca-sensitive phosphoproteins. The primary structure of β-casein is proline-rich and consists of 209 amino acid residues, including five phosphorylated serines. The molecular mass is approx. 24 kD. β-casein belongs to the family of intrinsically disordered proteins (IDPs), (Dunker et al, 2001) and the protein has an amphiphilic character with one hydrophilic and one more hydrophobic domain at neutral pH. The N-terminal region of the sequence is rich in polar and negatively charged amino residues, including all five phosphorylated serines located at position 15, 17, 18, 19, and 35 while the main part of the hydrophobic and positive residues are located in the C-terminus. The isoelectric point is estimated to 5.2 (Hipp, Groves, & McMeekin, 1952). β-caseins have a tendency to form aggregates at an critical concentration of 0.5 mg/ml in aqueous solutions, and this process is dependent on the concentration and the temperature (Leclerc and Calmettes, 1997a; Schmidt and Payens, 1972a).

Numerous studies have been devoted to β-casein adsorption to hydrophobic surfaces. At neutral pH, β-casein adsorbs with the hydrophobic C-terminal anchored to hydrophobic surfaces, while the hydrophilic N-terminus protrudes into the solution and forms a brush-like structure (Atkinson, Dickinson, Horne, & Richardson, 1995; Brooksbank, Davidson, Horne, & Leaver, 1993; Dalgleish & Leaver, 1991; Dickinson, Horne, Phipps, & Richardson, 1993; Fragneto, Su, Lu, Thomas, & Rennie, 2000; Mackie, Mingins, & North, 1991; Murray & Cros, 1998). For hydrophobic surfaces it has also been shown that the adsorbed amount increases as pH is lowered towards the isoelectric point, and that the presence of calcium ions modifies the structure of the β-casein monolayer, reducing the hydrophilic layer thickness and the adsorbed amount (Atkinson et al., 1995). Theoretical studies using self-consistent field theory qualitatively confirm this structure, and significant effects of ionic strength and pH are found (Dickinson, Pinfield, Horne, & Leermakers, 1997; Leermakers, Atkinson, Dickinson, & Horne, 1996) in line with experimental results (Kull, Nylander, Tiberg, & Wahlgren, 1997a, 1997b; Lee, Park, Chung, & Kim, 2004; Velev, Campbell, & Borwankar, 1998). However, for hydrophilic surfaces, such as negatively charged silica, the few available studies show that adsorption can be both strengthened (Kull et al., 1997a, 1997b) and weakened (Lundin, Elofsson, Blomberg, & Rutland, 2010) by increasing ionic strength. Nevertheless, the influence of ionic strength indicates the importance of electrostatic interactions for β-casein adsorption.

Moreover, Monte Carlo simulations of the adsorption of *one* β-casein molecule (Evers, Andersson, Lund, & Skepö, 2012) to different kind of surfaces have shown that it acts as an amphiphilic chameleon, changing properties according to the chemical

* Corresponding author.
E-mail addresses: Marie.Skepo@teokem.lu.se, marie.skepo@teokem.lu.se (M. Skepö).

0268-005X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodhyd.2013.09.006

environments, and adsorbs to any kinds of surfaces not only due to direct electrostatic and hydrophobic interactions but also due to conformational arrangements on the surface and charge regulation. Further, theoretical studies of unstructured saliva proteins also show that adsorption to charged surfaces can be mediated by conformational arrangements where attractive domains are located close to the charged surface (Skepö, Linse, & Arnebrant, 2006).

In this study we have used β-casein as a model protein to study how electrostatic interactions impact the adsorption of unstructured phosphoproteins. Since the silica surface charge is dependent on pH- and salt concentration, (Bolt, 1957) the experiments are conducted under thoroughly controlled conditions. To be able to trace the effect of charge regulation, the pH regime 7–11 is studied where several amino acids have their intrinsic pKa. The effect of divalent ions has been studied by adding $CaCl_2$ and $MgCl_2$. The effects of protein purity and the addition of urea are also presented.

2. Materials and methods

2.1. Materials

Throughout the study deionized MQ water has been used (Milli-Q Gadient A10, Millipore Corp. USA). Tris(hydroxymethyl)aminomethane (TRIS) was obtained from Affymetrix Inc., USA (Cat. No. 22675), sodium chloride from Scharlau S.L., Spain (Prod. No. SO0227005P), ethylenediaminetetra-acetic acid (EDTA) from MP Biomedicals Inc., USA (Cat. No. 194822), calcium chloride from Lancaster Inc., USA, and urea (Prod. No. 51465) and sodium hydroxide (Prod. No. 38210) from Sigma Corp., USA. Moreover, magnesium chloride hexahydrate (Prod. No. 105833) and hydrochloric (Cat. No. 100317) acid was purchased from Merck Chemicals KGaA, Germany. β-casein from bovine milk with a purity ≥98% was obtained from Sigma Corp., USA (Prod. No. C6905).

The freeze-dried protein was mixed with the appropriate solution and the pH was adjusted to approximately 8 with 0.1 M NaOH to ensure complete dissolution. The solution was subsequently sterile filtered (pore size 20 μm) to remove protein aggregates and large size impurities. To reduce low molecular weight impurities including divalent ions from the freeze-dried sample, a concentration cell was employed two times (Vivaspin 20, MWCO 10 kDa, Prod. No. VS2002, Sartorius Stedim Biotech GmbH, Germany). This involved two consecutive centrifugations (2500 g at 18 °C, 1.5 h), which removed 95% of the initial solvent. The protein sample was then fractionated by size exclusion chromatography using a column with a separation range of 3–70 kDa (Superdex 75 10/300 GL, GE Healthcare, Sweden). The chromatography was performed with a running column solution of 50 mM NaCl at pH 10, adjusted by 1 M NaOH. The purification was started when the column pH reached to 8.5, which is then increased to 9.5 at the end of purification. As seen in Fig. 1, two peaks were detected in the elution profile and the main peak associated with β-casein was collected and stored in frozen aliquots for the ellipsometric investigation. Protein concentrations were determined using absorbance measurements at 280 nm and an extinction coefficient of 11,500 $M^{-1}cm^{-1}$ (Swaisgood, 1982).

Silicon wafers with an oxide layer of approximately 30 nm were obtained from Semiconductor Wafer Inc., Taiwan. The wafers were cut into appropriate sizes for the ellipsometric investigation and subjected to a cleaning procedure developed at the RCA laboratories (Kern, 1970). First, the surfaces were gently boiled in an acidic solution, containing HCl (37%), H_2O_2 (30%) and water (1:1:5 by volume), for 5 min and rinsed three times in water. Then they were gently boiled in an alkaline solution, composed of NH_4OH (25%), H_2O_2 (30%) and water (1:1:5 by volume), for 5 min. Finally, the surfaces were rinsed in water three times and then in ethanol twice and stored in ethanol. Directly before use the surfaces were washed in ethanol and water, dried with nitrogen and plasma treated in a low pressure residual air glow discharge unit for 5 min (PDC-32 G, Harrick Scientific Corp., USA). By using this procedure, the surface area per negative charge on silica at pH 7 is reported to range between 900 and 500 $Å^2$ for aqueous solutions containing 0.01 and 0.1 M NaCl, respectively (Eigel et al., 1984) and a water contact angle of 5° (Malmsten, Burns, & Veide, 1998). Lateral friction microscopy has shown that the surface roughness is on the Ångström scale, (Donose, Taran, Vakarelski, Shinto, & Higashitani, 2006) i.e. the surface can be considered smooth in comparison with the size of the protein.

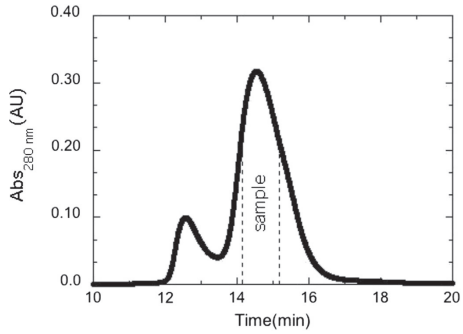


Fig. 1. Elution profile from size exclusion chromatography. The collected fraction used in experiments is indicated by two vertical dashed lines.

2.2. Method

Ellipsometry is an optical technique that is used to detect changes in the polarization of light upon reflection (Azzam & Bashara, 1977). By applying an appropriate optical model these changes can provide information on, for example, the refractive index and thickness of thin films. In this work, the changes in polarization were detected by null ellipsometry (Cuypers, 1976). The instrument used was a Rudolph thin film ellipsometer (type 43603-200E, Rudolph Research Corp., USA) automated according to Landgren and Jönsson (1993). Light emitted from a xenon arc lamp was detected at 401.5 nm and the angle of incidence was set to 67.9°. The 5 mL trapezoid cuvette made of optical glass (Hellma Optik GmbH Jena, Germany) was equipped with a magnetic stirrer and maintained at a constant temperature by a thermostat. Rinsing was performed *in situ* at a flow rate of 16 mL/min without exposing the silica surface to the ambient atmosphere.

A clean silica surface was mounted to a 5 mL ellipsometric cuvette and optical parameters were determined in air and in the liquid to calculate the complex refractive index of the silicon and the thickness and refractive index of the silicon oxide layer (Landgren & Jonsson, 1993). To minimize instrumental systematic errors four zone measurements were conducted both in air and in the liquid phase. The incubation was initiated by adding an aliquot (~250 μL) of the purified β-casein sample to the cuvette containing the appropriate solution. Due to β-casein adsorption, an extra layer was introduced to the optical model, and the refractive index and thickness of this layer were determined from the recorded optical parameters. The adsorbed amount was calculated

assuming a linear increase of the refractive index with the concentration using a increment of 0.18 mL/g, (Schmidt & Payens, 1972a, 1972b) hence notice that reported adsorption amounts in this study depend directly on the assumed value of the refractive index increment.

10 mM TRIS was used as a standard buffer and pH was adjusted by titration with 1 M HCl. The TRIS buffer has an effective buffer capacity in range of approximately 7–9 (pKa = 8.1). Above this range, pH was adjusted by small aliquots of sodium hydroxide and the solution pH, measured directly before rinsing, is presented. The buffer may influence protein adsorption (Kurrat, Prenosil, & Ramsden, 1997) therefore control experiments were conducted without TRIS where the pH was adjusted with sodium hydroxide, see supplementary material. The same qualitative results were obtained indicating that the adsorption is unaffected by TRIS. Apart from TRIS and sodium chloride, EDTA, urea, magnesium chloride, and calcium chloride were also used as indicated in the figures. All measurements were performed at 25 °C with a 10 min rinsing time. β-Casein was added at time zero in the presented figures with a concentration of 0.05 mg/ml.

To verify the reproducibility, four measurements were conducted in 140 mM NaCl and 10 mM TRIS solution at pH 8.5. An average adsorbed amount (estimated from the plateau values) of 0.96 mg/m^2 with a standard deviation of 0.06 mg/m^2 was obtained (6%). The corresponding values for the layer thickness was 5.8 nm with a standard deviation of 0.1 nm.

3. Results and discussion

The adsorption of β-casein to negatively charged, hydrophilic silica surfaces has been studied as a function of (i) salt concentration, (ii) pH, (iii) divalent ions, and (iv) solvent conditions (urea addition). Since β-casein is an intrinsically disordered protein, it can be considered as a flexible chain of amino acids, (Dumas, Brignon, Grosclaude, & Mercier, 1972) which can adapt many conformations while approaching to a surface. In addition, the amino acid sequence of the chain determines physicochemical properties of β-casein, for instance, at pH 7, β-casein has an amphiphilic character with a negative hydrophilic N-terminal domain and a more hydrophobic C-terminal domain containing three positive regions (Evers et al., 2012). Due to this character, β-casein aggregates when its concentration is above a critical aggregation concentration, ~0.5 mg/ml (Schmidt & Payens, 1972a, 1972b; Leclerc & Calmettes, 1997a, 1997b) The experiments are therefore conducted at 0.05 mg/ml of β-casein to avoid aggregates in solution and to be able to study adsorption of monomeric form only.

In the studied pH regime, i.e. pH 7–11, five histidines (pKa = 6.5), four tyrosines (pKa = 10.1), and 11 lysines (pKa = 10.8) have their intrinsic pKa values and thereby can act as a proton donor/acceptor and contribute to the charge regulation mechanism. However, this mechanism is insignificant at pH 8.5 due to lack of residues with pKa around 8.5. To investigate the effect of electrostatics, a reference system was selected at pH 8.5 due to the simplicity of non-fluctuating charges and at 150 mM ionic strength (10 mM TRIS and 140 mM NaCl) due to the relevance to physiological conditions.

3.1. Effect of monovalent salt

Fig. 2 shows the amount of adsorbed β-casein after 30 min, at pH 8.5, as a function of total ionic strength including monovalent salt (NaCl) and 10 mM TRIS. At low ionic strength and pH 8.5, β-casein has an ideal theoretical charge of −20*e* and the bare silica has a surface charge density of approximately −2.5 μC/cm^2, where the latter is strongly dependent on the presence of salt and proteins

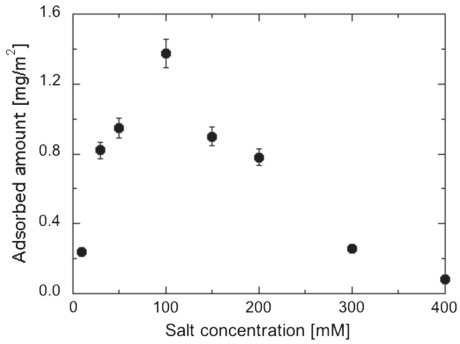


Fig. 2. Plateau values of adsorbed amount of 0.05 mg/ml β-casein at pH 8.5 as a function of NaCl-concentration in mM (including 10 mM Tris).

(Samoshina, Nylander, Shubin, Bauer, & Eskilsson, 2005). The surface charge density of silica is shown to decrease by a factor of three to −7.5 μC/cm^2 when 100 mM KCl is added to the solution (Samoshina et al., 2005). Our measurements show that at low salt concentrations, only a small amount of protein adsorbs, whereas an increased salt concentration induces adsorption, and a maximum in adsorbed amount is reached at 100 mM salt. A further increase of salt results in less adsorption. By assuming that the surface is a smooth lattice, the adsorbed amount gives a surface area per adsorbed molecule of 3150 Å^2 (56 Å*56 Å) corresponding to a sphere with a radius of 28 Å. This value implies that the molecules are in relatively close contact and that the intermolecular interactions between the adsorbed molecules at the surface *cannot* be neglected.

By varying the salt concentration, it is clearly shown that electrostatic interactions are of importance for adsorption of β-casein to silica surfaces and tunes the adsorbed amount. The most important interactions are: (i) electrostatic repulsion between the surface and the protein, (ii) electrostatic attraction between positively charged amino acids in the protein and surface, and (iii) electrostatic repulsion between adsorbed proteins at the surface. At low salt concentrations (cf. 10 mM) the electrostatic repulsion in (i) and (iii) dominates, which results in low adsorption. An increase of the salt concentration to 100 mM screens the repulsion in (i) and (iii) which enhances the adsorption. Further increase of salt, above 100 mM, screens the electrostatic attraction between the surface and positively charged amino acids in the protein which results in less adsorbed protein. Hence, the positively charged amino acids serve as anchoring points. This mechanism has been verified to play an important role for adsorption of PRP1 to negatively charged surfaces. PRP1 has a net charge of -9*e* and the charge distribution and character resembles that of β-casein (Skepö et al., 2006).

Our results are in contrast to the study by Lundin et al. (2010) who have done equivalent experiments, and shown that β-casein has *high* affinity to hydrophilic silica at *low* ionic strength and pH 8.5.

3.2. Effect of pH

As seen in Fig. 3a, the highest adsorption is obtained at pH 7 whereas the adsorption is completely diminished when pH > 10.

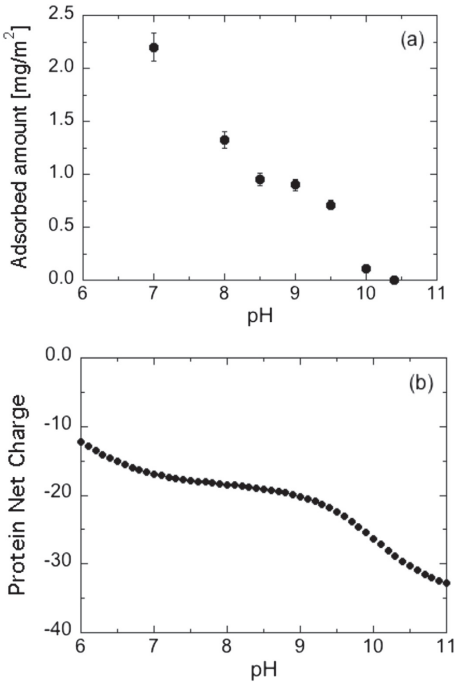


Fig. 3. (a) Adsorbed amount of 0.05 mg/ml β-casein in 10 mM Tris and 140 mM NaCl, and (b) the theoretical net charge of the protein as a function of pH.

This can partly be explained from the theoretical titration curve for the protein given in Fig. 3b, where an increase in pH from 7 to 10.5 implies a change in net charge from $-16e$ to $-30e$, which increases the electrostatic repulsion between the surface and the protein as well as between the adsorbed protein molecules.[1]

Due to deprotonation of the silanol groups on bare silica, the surface charge density is affected by pH. The results obtained by Samoshina et al. (2005) show that silica surface charge was measurable at pH ≈ 5, and decreased progressively with the following approximate numbers pH 7 (-2.5 μC/cm^2), pH 8 (-5 μC/cm^2), pH 8.5 (-7.5 μC/cm^2), and pH 9 (-12 μC/cm^2) at an ionic strength of 100 mM. The decrease in the surface and protein net charge due to the released protons in combination with the charge regulation abilities of the protein also contributes to the repulsion between the surface and the protein as well as between the adsorbed proteins, resulting in a reduced adsorbed amount with increasing pH. Reliable values on the thickness of the protein layer were obtained above an adsorbed amount of 0.5 mg/m^2 and these results (data not shown) indicate a thickness of ≈6 nm independent of the solution pH.

[1] Notice, that the theoretical titration curve does not take into account intra- and intermolecular interactions and hence charge regulation effects is omitted.

3.3. Addition of divalent ions

The hydrophilic moiety of β-casein in the N-terminal domain contains five phosphorylated serine amino acid residues, which are known to bind divalent cations as magnesium and calcium, with a higher affinity for the latter (Baumy & Brule, 1988; Follows, Holt, Nylander, Thomas, & Tiberg, 2004; Holt & Sawyer, 1988; Holt, Wahlgren, & Drakenberg, 1996). The phosphorylated amino acids are responsible for a significant part of the protein charge and Baumy & Brule (1988) have shown that β-casein has a binding capacity of approximately seven Ca^{2+}- or Mg^2-ions for pH > 7.00 and low ionic strengths, and that the binding capacity decreases with lowering pH and increasing salt concentration. At an ionic strength of 100 mM, β-casein is expected to bind 2.7 and 3.3 Ca^{2+}- and Mg^{2+}-ions respectively.

The effect of divalent ions has been studied by adding $CaCl_2$ and $MgCl_2$ both to the reference system, and to system where monovalent salt is omitted. In this part of the study, the surface charge density of silica in the reference system can be regarded as constant due to the insignificant variations of ionic strength upon addition of divalent ions. Fig. 4 shows the adsorbed amount after 30 min as a function of calcium-concentration. Notice that already an addition of 1 mM Ca^{2+}-ions results in a pronounced increase in plateau values of adsorbed amount (c.f. 2 mM in milk). At an addition of 8 mM $CaCl_2$ an adsorbed amount of 3 mg/m (Hipp et al., 1952) was obtained, which corresponds to a monolayer of β-casein at a saturated hydrophobic surface, see ref (Nylander, Tiberg, Su, Lu, & Thomas, 2001) and references therein. For calcium concentration below 8 mM, the added calcium reduces the adsorbed layer thickness from 6 nm to 4.5 nm, implying a more compact adsorbed layer, whereas for concentrations above 8 mM, the layer thickness increases from 4.5 nm to 7 nm. The latter imply that either a second layer of adsorbed proteins is formed due to protein aggregation or a more brush-liked structured layer is created on the surface due to crowding. In comparison, the thickness of the adsorbed layer of β-casein to *hydrophobized* silica surfaces decreases or stays constant upon addition of calcium (Atkinson et al., 1995; Atkinson, Dickinson, Horne, Leermakers, & Richardson, 1996; Follows et al., 2004; Kull et al., 1997a, 1997b; Velev et al., 1998).

The effect of calcium ions on the adsorption at low and high ionic strength was also investigated by comparing two systems:

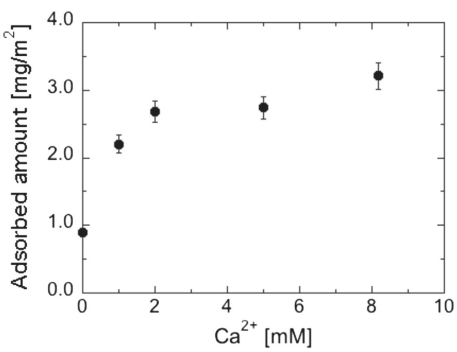


Fig. 4. Adsorbed amount in mg/m (Hipp et al., 1952) of 0.05 mg/ml β-casein in 10 mM Tris at pH 8.5 as a function of calcium-concentration in mM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

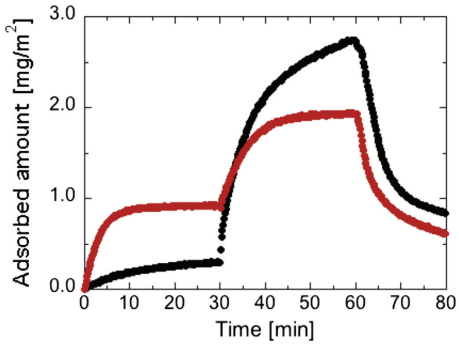


Fig. 5. Adsorption isotherms for 0.05 mg/ml β-casein. After 30 min $CaCl_2$ was added to the solution, which gave rise to a final $CaCl_2$ concentration of 5 mM in the cuvette, after 60 min rinsing with sample buffer. Solution conditions, black curve 10 mM Tris at pH 8.5, and red curve 10 mM Tris at pH 8.5 supplemented with 140 mM NaCl. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

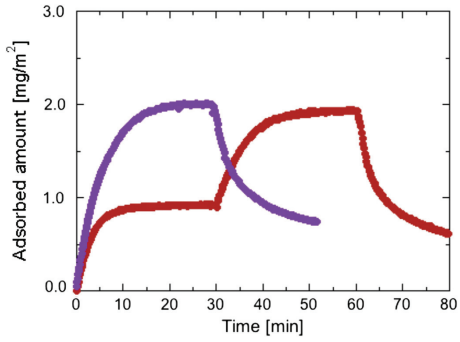


Fig. 6. Adsorption isotherms for 0.05 mg/ml β-casein in 10 mM Tris supplemented with 140 mM NaCl at pH 8.5 for (a) a surface equilibrated in 5 mM $CaCl_2$ 30 min before addition of β-casein (purple curve), and (b) addition of $CaCl_2$ 30 min after measurement initiation. (red curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(I) 0.05 mg/ml protein in 10 mM Tris, pH 8.5 and (II) 0.05 mg/ml protein in 10 mM Tris, pH 8.5, supplemented with 140 mM NaCl. After 30 min 5 mM $CaCl_2$ was added to the cuvettes, see Fig. 5.

At salt free conditions (black curve) the adsorption is increasing upon addition of Ca^{2+}, from ≈0.25 to 2.75 mg/m^2. For the latter value, the plateau value is not obtained within the time limit of the experiment. The adsorption at 150 mM ionic strength (red curve (in web version)) increases from ≈0.9 mg/m^2 to 1.8 mg/m^2. The plateau values of the adsorption isotherms, with respect to ionic strength, indicate that this is not an ordinary screening effect. Rinsing with sample buffer was performed after 60 min of measurement and the reversibility of the protein adsorption is similar in both systems, approximately 70%. Recalculation of the net charges of β-casein by applying the binding capacities of calcium at low and high ionic strength results in a protein net charge of ≈ $-4e$ and ≈ $-12e$ respectively. Applying the new net charges on the adsorbed amounts results clearly indicate that it is the electrostatic repulsion between the adsorbed proteins that is controlling the saturation of the surface.

The effect of calcium has been further evaluated by applying the reversed process i.e. the surface was equilibrated in $CaCl_2$ in 30 min before addition of the protein i.e. calcium acts as counterions to the surface. As shown in Fig. 6, neither the adsorption, the kinetics, the plateau values, nor the reversibility upon rinsing, were affected. These results also indicate that it is the electrostatic repulsion between the adsorbed proteins that determine the amount of adsorbed proteins on the surface.

Fig. 7 shows the adsorption kinetics for the reference system when 5 mM $CaCl_2$ (red curve) and 5 mM $MgCl_2$ (blue curve) were added after 30 min. As expected, before the addition of divalent salt, the adsorption data is reproducible; thereafter a deviation is visible, where the adsorbed amount of β-casein is 25% lower with Mg^{2+}-ions. After rinsing with sample buffer (60 min) the adsorbed amounts are converging to the same level (0.6 mg/m^2) as before the addition, hence the increased adsorbed amount due to the addition of divalent ions is reversible adsorbed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Our study indicates that the adsorbed amount of β-casein to negatively charged hydrophilic silica surfaces increases as NaCl (0.9 mg/m^2) < $MgCl_2$ (1.5 mg/m^2) < $CaCl_2$ (1.9 mg/m^2) where the adsorbed amount are given within the parenthesis. Regarding the adsorbed layer thickness, 5 nm was obtained when NaCl was added, whereas 7 nm in thickness was reached upon addition of divalent ions. Since this study is performed ten times below the critical aggregation concentration and since the thicknesses are equal, the deviance in adsorbed amount is attributed to ion affinity effects and not due to adsorption of induced aggregates.

3.4. Hydrophobic interactions and hydrogen bonding

Urea, a small hydrophilic molecule, is a widely used protein denaturant in *in vitro* unfolding/refolding experiments (Hua, Zhou, Thirumalai, & Berne, 2008; Zangi, Zhou, & Berne, 2009). Fig. 8a depicts the isotherms for the reference system (black curve) and the reference system supplemented with 1 M Urea (green curve) (in web version), and Fig. 8b the corresponding system without addition of monovalent salt. The adsorption kinetics depict similar behaviour although the plateau values deviate, where an addition of urea decreases the maximum adsorbed amount from 1 to 0.6 mg/m^2. Regarding reversibility, both systems display similar trends where approximately 50% is desorbed after rinsing. From this we can conclude that when the electrostatic interactions are suppressed - an addition of urea will clearly affect the adsorption and already an addition of 1 M urea the adsorption is decreased by 40%. The latter indicates that hydrophobic interactions and/or hydrogen bonding between the protein and the surface influence the adsorption. The mechanisms behind possible hydrophobic interactions are unknown to the authors. The silica surface contains two different groups, −OH and −O−, where the latter can be considered more hydrophobic than the former. Before alkaline treatment the silica surface has a water contact angle of 50°and afterwards 5°, hence on the macroscopic scale, the surface can be considered hydrophilic and without hydrophobic patches. A plausible explanation is that the reduced hydrophobic interactions between the adsorbed proteins result in larger surface area coverage per protein; hence, the surface becomes saturated with fewer

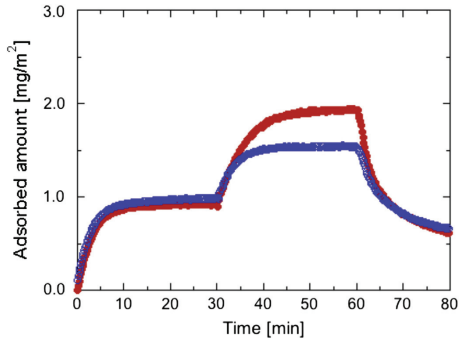


Fig. 7. Adsorption isotherms for 0.05 mg/ml β-casein in 10 mM Tris, pH 8.5, supplemented with 140 mM NaCl. After 30 min, 5 mM Ca2+ (red curve) and 5 mM Mg2+ (blue curve) were added to the cuvette, the experiments were running for an additional 30 min, and then rinsing with sample buffer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

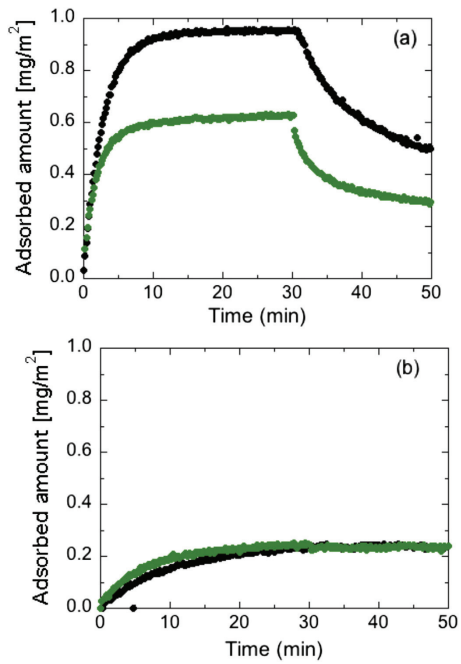


Fig. 8. Adsorption isotherms of 0.05 mg/ml β-casein at pH 8.5 (a) in 10 mM Tris/140 mM NaCl (black curve), and 10 mM Tris/140 mM NaCl/1 M Urea (green curve), and (b) in 10 mM Tris (black curve), and 10 mM Tris/1 M Urea (green curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proteins i.e. excluded volume effects are of importance. Notice also that Urea does disrupt hydrogen bonds hence hydrogen bonding seems plausible.

4. Conclusions

In this study, we show that adsorption of *monomeric* β-casein to *hydrophilic* silica surfaces is dependent on electrostatic interactions and the amphiphilic character of the protein. The results indicate that there is counterbalance between: (i) electrostatic repulsion between the surface and the protein, (ii) electrostatic attraction between positively charged amino acids in the protein and surface, and (iii) electrostatic repulsion between the adsorbed proteins at the surface. The fact that the charge of the protein and the silica surface varies as a function of pH and salt concentration and that the binding capacity of divalent ions varies as a function of pH and salt makes it a delicate task to interpret the results. Moreover, our results suggest that hydrophobic interactions and hydrogen bonding within the protein, or hydrogen bonding between the protein and the surface are of importance.

Acknowledgements

We acknowledge Vinnova, the Vinnmer program, and The Royal Physiographic Society in Lund, Per-Eric and Ulla Schybergs Foundation, for financial support, Prof. T. Nylander for useful discussions F. Xie, and S. Ouhajji for the experimental support.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.foodhyd.2013.09.006.

References

Bolt, G. H. (1957). *Journal of Physical Chemistry, 61*(9), 1166–1169.
Cuypers, P. A. (1976). *Dynamic ellipsometry: Biochemical and biomedical applications*. Rijksuniversiteit Limburg.
Dunker, A. K., et al. (2001). *Journal of Molecular Graphics and Modelling, 19*, 26–59.
Kern, W. (1970). *Rca Review, 31*(2), 234-&.
Atkinson, P. J., Dickinson, E., Horne, D. S., Leermakers, F. A. M., & Richardson, R. M. (1996). *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics, 100*(6), 994–998.
Atkinson, P. J., Dickinson, E., Horne, D. S., & Richardson, R. M. (1995). *Journal of the Chemical Society Faraday Transactions, 91*(17), 2847–2854.
Azzam, R. M. A., & Bashara, N. M. (1977). *Ellipsometry and polarized light*. North-Holland Pub. Co.
Baumy, J. J., & Brule, G. (1988). *Lait, 68*(4), 409–417.
Brooksbank, D. V., Davidson, C. M., Horne, D. S., & Leaver, J. (1993). *Journal of the Chemical Society Faraday Transactions, 89*(18), 3419–3425.
Dalgleish, D. G., & Leaver, J. (1991). *Journal of Colloid and Interface Science, 141*(1), 288–294.
Dickinson, E., Horne, D. S., Phipps, J. S., & Richardson, R. M. (1993). *Langmuir, 9*(1), 242–248.
Dickinson, E., Pinfield, V. J., Horne, D. S., & Leermakers, F. A. M. (1997). *Journal of the Chemical Society Faraday Transactions, 93*(9), 1785–1790.
Donose, B. C., Taran, E., Vakarelski, I. U., Shinto, H., & Higashitani, K. (2006). *Journal of Colloid and Interface Science, 299*(1), 233–237.
Dumas, B. R., Brignon, G., Grosclaude, F., & Mercier, J.-C. (1972). *European Journal of Biochemistry, 25*(3), 505–514.
Eigel, W. N., Butler, J. E., Ernstrom, C. A., Farrell, H. M., Harwalkar, V. R., Jenness, R., et al. (1984). *Journal of Dairy Science, 67*(8), 1599–1631.
Evers, C. H. J., Andersson, T., Lund, M., & Skepö, M. (2012). *Langmuir, 28*, 11843–11849.
Follows, D., Holt, C., Nylander, T., Thomas, R. K., & Tiberg, F. (2004). *Biomacromolecules, 5*(2), 319–325.
Fragneto, G., Su, T. J., Lu, J. R., Thomas, R. K., & Rennie, A. R. (2000). *Physical Chemistry Chemical Physics, 2*(22), 5214–5221.
Hipp, N. J., Groves, M. L., & McMeekin, T. L. (1952). *Journal of the American Chemical Society, 74*(19), 4822–4826.
Holt, C., & Sawyer, L. (1988). *Protein Engineering, 2*(4), 251–259.
Holt, C., Wahlgren, N. M., & Drakenberg, T. (1996). *Biochemical Journal, 314*, 1035–1039.
Hua, L., Zhou, R. H., Thirumalai, D., & Berne, B. J. (2008). *Proceedings of the National Academy of Sciences of the United States of America, 105*(44), 16928–16933.

Kull, T., Nylander, T., Tiberg, F., & Wahlgren, N. M. (1997a). *Langmuir: The ACS Journal of Surfaces and Colloids, 13*(19), 5141–5147.
Kull, T., Nylander, T., Tiberg, F., & Wahlgren, N. M. (1997b). *Langmuir, 13*(19), 5141–5147.
Kurrat, R., Prenosil, J. E., & Ramsden, J. J. (1997). *Journal of Colloid and Interface Science, 185*(1), 1–8.
Landgren, M., & Jonsson, B. (1993). *Journal of Physical Chemistry, 97*(8), 1656–1660.
Leclerc, E., & Calmettes, P. (1997a). *Physica B, 241–243*, 1141–1143.
Leclerc, E., & Calmettes, P. (1997b). *Physical Review Letters, 78*(1), 150–153.
Lee, M., Park, S. K., Chung, C., & Kim, H. (2004). *Bulletin of the Korean Chemical Society, 25*(7), 1031–1035.
Leermakers, F. A. M., Atkinson, P. J., Dickinson, E., & Horne, D. S. (1996). *Journal of Colloid and Interface Science, 178*(2), 681–693.
Lundin, M., Elofsson, U. M., Blomberg, E., & Rutland, M. W. (2010). *Colloids and Surfaces B: Biointerfaces, 77*(1), 1–11.
Mackie, A. R., Mingins, J., & North, A. N. (1991). *Journal of the Chemical Society Faraday Transactions, 87*(18), 3043–3049.
Malmsten, M., Burns, N., & Veide, A. (1998). *Journal of Colloid and Interface Science, 204*(1), 104–111.
Murray, B. S., & Cros, L. (1998). *Colloids and Surfaces B-Biointerfaces, 10*(4), 227–241.
Nylander, T., Tiberg, F., Su, T. J., Lu, J. R., & Thomas, R. K. (2001). *Biomacromolecules, 2*(1), 278–287.
Samoshina, Y., Nylander, T., Shubin, V., Bauer, R., & Eskilsson, K. (2005). *Langmuir, 21*(13), 5872–5881.
Schmidt, D. G., & Payens, T. A. J. (1972a). *Journal of Colloid and Interface Science, 39*(3), 655–662.
Schmidt, D. G., & Payens, T. A. J. (1972b). *Journal of Colloid and Interface Science, 39*(3), 655-&.
Skepö, M., Linse, P., & Arnebrant, T. (2006). *The Journal of Physical Chemistry, 110*(24), 12141–12148.
Swaisgood, H. E. (1982). *Developments in Dairy Chemistry, 1*, 1–59.
Velev, O. D., Campbell, B. E., & Borwankar, R. P. (1998). *Langmuir, 14*(15), 4122–4130.
Zangi, R., Zhou, R. H., & Berne, B. J. (2009). *Journal of the American Chemical Society, 131*(4), 1535–1541.