LUND UNIVERSITY

Structure and Phase Behavior of Protein Solutions

Persson, Björn

2011

Link to publication

Citation for published version (APA):
Persson, B. (2011). *Structure and Phase Behavior of Protein Solutions*. [Doctoral Thesis (compilation), Computational Chemistry]. Department of 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

Structure and Phase Behavior of Protein Solutions

Björn Persson

Department of Theoretical Chemistry
Lund University, Sweden

LUND UNIVERSITY

Doctoral Thesis

The thesis will be publicly defended on Friday 11th of November 2011, 10.30 in lecture hall B, Center for Chemistry and Chemical Engineering, Lund

The faculty opponent is Marjolein Dijkstra,
Utrecht University, Utrecht, The Netherlands

Front cover: Snapshot from a simulation of human Lactoferrin at pH 7 and 25 mM of salt, see Paper IX.

© Björn Persson 2011
Doctoral Thesis

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

All rights reserved

ISBN 978-91-7422-282-1
Printed by Media-Tryck, Lund University, Lund

Organization
LUND UNIVERSITY

Document name
DOCTORAL DISSERTATION

Date of issue
November 11, 2011

Author(s)
Björn Persson

Sponsoring organization

Title and subtitle

Structure and Phase Behavior of Protein Solutions

Abstract

The complexity of proteins is reflected in the structure of their fluids. This is seen from strongly inhomogeneous distributions about the single protein. Using molecular simulations of protein models, based on experimental structures, the present work probes the directional forces between proteins. Emphasis is made on the mutual orientation of proteins and the interplay of electrostatic and short ranged van der Waals forces. We show that this may lead to non intuitive mutual amplification of the interactions - resulting in the formation of stereospecific aggregates - as well as having importance for the phase behavior of oppositely charged proteins.

Key words
Statistical Mechanics, Simulations, Protein Interactions, Phase Separation, Aggregation, Mutual Amplification

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-7422-282-1

Recipient´s notes

Number of pages

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 _____ Date _____

'Trying is the first step towards failure' – H. Simpson

List of Papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals. The papers are appended at the end of the thesis.

I **Faunus: An Object Oriented Framework for Molecular Simulation**
Lund, Trulsson and Persson.
Source Code Biol Med., 3, 1 (2008)

II **Enhanced Protein Steering: Cooperative Electrostatic and van der Waals Forces in Antigen-Antibody Complexes**
Persson, Jönsson and Lund
J. Phys. Chem. B, 113, 10459 (2009)

III **Association and Electrostatic Steering of αLactalbumin-Lysozyme Heterodimers**
Persson and Lund
Phys. Chem. Chem. Phys., 11, 8879 (2009)

IV **Molecular Evidence of Stereo-specific Lactoferrin Dimers in Solution**
Persson, Lund, Forsman, Chatterton and Åkesson
Biophys. Chem., 11, 187 (2010)

V **Protein Adsorption and Complement Activation for Di-block Copolymer Nanoparticles**
Vauthier, Persson, Lindner and Cabane
Biomater., 32, 1646 (2011)

VI **Investigation at Residue Level of the Early Steps during the Assembly of Two Proteins into Supramolecular Objects**
Salvatore, Duraffourg, Favier, Persson, Lund, Delage, Silvers, Schwalbe, Croguennec, Bouhallab and Forge
BioMarcomol., 12, 2200 (2011)

VII **Anisotropic Interactions in Protein Mixtures: Self Assembly and Phase Behavior**
Kurut, Persson, Åkesson, Forsman and Lund
Submitted

VIII **A Theoretical Study of the Optical *pH* Nanosensor Glu**[3]
Persson, Lund, Forsman, Leiding, Peterson Årsköld, Åkesson
Manuscript

IX **Multibody Protein Interactions and Genetic Variants**
Persson, Lund, Forsman and Åkesson
Manuscript

List of Contributions

All papers employed own in-house computer codes/programs. These were developed together with the help and support from Mikael Lund and Chris Evers.

I I have taken part in the majority of the software development. I have also taken part in the design of the framework. The paper was written by Lund and I participated in the discussion of it.

II I conducted most of the calculations and participated in the writing of the article.

III I performed most of the simulations and participated in the writing and discussion of the manuscript.

IV I did all calculations and had the major responsibility in the production of the article.

V I performed data analysis for Figure 4 and 8, and participated in the discussion of the manuscript.

VI I took part in the discussion of the theoretical contribution. The calculations where preformed by Lund, based on our previous calculations in Paper III. I also took part in commenting and finalizing the paper

VII I participated in initiating and starting the project. I was partly responsible in the development of the simplified model and supervision of the project. I developed a large part of the program used and participated in the refinement of the paper. Kurut performed the calculations.

VIII I developed the model and performed the calculations. I had the major responsibility for writing the articel.

IX I perfomed all calculations and had major responsibility for writing the paper together with Lund.

Contents

Populärvetenskaplig sammanfattning på svenska 1

Brief summary 3

1 **Introduction** 5

2 **Statistical Mechanics and Thermodynamics** 7
2.1 Statistical Thermodynamics 7
2.2 Thermodynamics 9
2.2.1 Adsorption 10
2.2.2 Equilibrium Constants and Titration 10

3 **Intermolecular Interactions** 13
3.1 Coulombic Interactions 14
3.2 Short ranged Interactions 14
3.3 McMillan-Mayer Theory 15

4 **Monte Carlo Simulations** 17
4.1 Geometries and Boundary Conditions 17
4.2 Evolution of the Chain 18
4.2.1 Non-rejective Markov Steps 19
4.2.2 Parallel Tempering and Phase coexistence 20

5 **Summary of Results and Concluding Remarks** 23
5.1 Anisotropic Protein Interactions 23
5.1.1 Mutual Amplification 23
5.1.2 Orientational Effects 24
5.2 Protein Adsorption 24
5.3 pH-detection in Nano Compartments 24
5.4 Concluding Remarks 24

Acknowledgements 27

Appendix 29
Paper I: Faunus: An object oriented framework for molecular simulation 29

Paper II: Enhanced Protein Steering: Cooperative Electrostatic and van der Waals Forces in Antigen-Antibody Complexes 39
Paper III: Association and electrostatic steering of alpha-lactalbumin-lysozyme heterodimers 47
Paper IV: Molecular evidence of stereo-specific lactoferrin dimers in solution 57
Paper V: Protein adsorption and complement activation for di-block copolymer nanoparticles 63
Paper VI: Investigation at Residue Level of the Early Steps during the Assembly of Two Proteins into Supramolecular Objects 77
Paper VII: Anisotropic Interactions in Protein Mixtures: Self Assembly and Phase Behavior 91
Paper VIII: A theoretical study of the optical pH nanosensor Glu^3 . . 97
Paper IX: Multibody Protein Interactions and Genetic Variants 111

Populärvetenskaplig sammanfattning på svenska

Även om vi inte ser molekyler dagligdags har vi sedan länge vant oss vid att sakerna kring oss innehåller *något*. När vi satt ett glas över ett brinnande ljus och det slocknat så är det *något* som tagit slut. Kemi rör beskrivningen av dessa sällan direkt synliga förlopp där molekyler omlagras och ger upphov till de strukturer vi finner världen uppbyggd av: gaser, vätskor och solida material. Dessa processer kan försiggå inne på laboratorier, men även i vår närhet finns gott om tillsynes enkla skeenden, till exempel en sockerbits upplösning på botten av en kaffekopp, som innehåller en fantastisk massa kemi. Då kemi är ett brett ämne kan det vara värt att nämna att denna avhandling speglar proteinlösningars egenskaper betraktat genom den fysikaliska kemins ögon. Fokus ligger därför främst på strukturen i proteinlösningar: hur fördelar proteiner sig i lösningen, bildar proteinerna aggregat och under vilka betingelser fasseparerar lösningarna? För att kunna beskriva dessa system måste vi konstruera modeller av proteinerna och det övriga innehållet i lösningarna. Sedan utnyttjar vi oss av statistisk mekanik, en teori som möjliggör beräkningar av jämviktsegenskaper och därmed kan utvärdera modellbeskrivningar. Beräkningar görs med Metropolis Monte Carlo-metod, en simuleringsteknik som är anpassad till den statistisk mekaniska problemställningen. Huvuddelen av denna avhandling handlar om växelverkan mellan proteiner. Arbetet undersöker modeller av proteinlösningar för att studera hur proteiners oregelbundna strukturer påverkar deras beteende i vattenlösning. Proteinerna som ingår i studierna är sådana som vi bland annat finner i mjölk och ägg. Avhandlingen visar att det kan finnas en kraftig förstärkning av växelverkan mellan proteiner. Det framgår att de så kallade elektrostatiska och van der Waals bidragen kan ge upphov till en ömsesidigt ökad attraktion. I avhandlingen påvisas det även att laddningsfördelningen på proteiner kan spela en avgörande roll för proteinlösningarnas fasbeteende. Resultaten pekar på att det samspel som noteras mellan distributionen av laddningar över proteiner och proteiners form har ett stort inflytande på såväl huruvida proteiner bildar aggregat som aggregatens egenskaper.

Brief summary

- **Paper I:** Here we present a programming framework and suggest an object oriented design of Metropolis Monte Carlo packages. All code used in the thesis is developed in *Faunus* and is publicly available.
- **Paper II:** In this paper we evaluate the interaction of Lysozyme with the antigen HyHEL5 and its constituents. We introduce orientational concepts like alignment and use them to quantify anisotropicity.
- **Paper III:** The study focuses on the dimer formation between Lysozyme and αLactalbumin. The latter exists in two forms, one with bound calcium and one *apo*-form. The result shows good qualitative agreement with experiments and suggests that the aggregate is highly ordered.
- **Paper IV:** Lactoferrin is shown to form stereospecific dimers. This is supported with experimental observations. We show that an intricate charge matching across the binding surface is responsible for the attraction at short separations.
- **Paper V:** The concept of using diblock copolymer particles as drug delivery vessels is examined experimentally. It is found that although the adsorption of small proteins may be decreased with a thick polymer brush, it can not be eliminated. Even though adsorption can not be completely prevented, the results show that immune reaction is significantly decreased.
- **Paper VI:** The structure of the Lysozyme-αLactalbumin dimer is determined experimentally and compared with simulation results. An excellent agreement between calculations and experiments is obtained.
- **Paper VII:** We continue the work in Paper III to study the phase behavior of Lysozyme-αLactalbumin. The results agree well with the experimental observations and discriminate between the *apo* and *holo* form accordingly. They show that the charge distribution is crucial for the aggregation and hence necessary to include in modeling.
- **Paper VIII:** It is shown how one can obtain reliable models of titrating macromolecules. We are here able to compare calculations of an optical pH-meter with experimental results. The agreement is good and it is also shown

how the model can be extended to study adsorption at zwitterionic membranes.

- **Paper IX:** The attraction between Lactoferrin molecules is dissected and shown to be due to a mutual amplification of electrostatic and van der Waals forces. We show that a similar attraction exists between four different sources of Lactoferrin. It is found that the aggregation of Lactoferrin increases as the salt concentration increases, in accordance with experimental observations.

Chapter 1

Introduction

In many respects proteins are rather unique structures and a signature of living organisms. They form a large part of ourselves and the living entities we have around us. Many proteins are absolutely necessary for our bodies to function[1], while some are disastrous for our life-sustaining inner machinery [2]. A protein is defined by its particular amino acid sequence. However, this is not the single determinant of their structure and properties. The folding of proteins is a very complex subject and their shape is highly dependent upon solution conditions such as solvent, pH, temperature, electrolyte type and concentration as well as presence of other proteins and solutes [1]. This variation is also true for their interactions. Proteins display a very rich palette of properties: catalyse chemical reactions, transmit signals and self assemble into a variety of structures. This promotes them as favorite templates from which we may gain understanding of very complex interactions and studies of proteins and protein solutions are therefore relevant for the entire scope of science and technology: from obtaining understanding about the chemical processes of cells and organisms, gain knowledge to construct new materials, to obtain intricate theoretical insight into macromolecular and colloidal problems. The work presented here will focus on the latter. What are the implications of the various sources of forces between proteins and how do these forces interfere with each other? Although proteins are very complicated structures they are often modeled as rather simple objects like sticky hard spheres[3], augmented with a homogeneous surface charge density[4] or regularly spaced attractive patches[5]. Other models are much more detailed with an atomistic description of the proteins[6] or

[1] T. Creighton, *Proteins - Structures and Molecular Properties* (1993).

[2] M. Moayeri and S. H. Leppla, Molecular Aspects of Medicine **30**, 439 (2009).

[3] P. R. ten Wolde and D. Frenkel, Science **277**, 1975 (1997).

[4] T. Gibaud, F. Cardinaux, J. Bergenholtz, A. Stradner, and P. Schurtenberger, Soft Matter **7**, 857 (2011).

[5] N. Kern and D. Frenkel, J Chem Phys **118**, 9882 (2003).

[6] A. H. Elcock, D. Sept, and J. A. McCammon, J. Phys. Chem. B **105**, 1504 (2001).

even the entire solution[7]. The former approach has often led to a general and qualitative understanding of phase behavior while the latter brought insight to specific protein system, but at a high computational cost which limits studies of phase behavior. Much of the work in this thesis is devoted to evaluating a structurally detailed model, that is computationally tractable in order to obtain explicit understanding of particular protein solutions based on experimental protein structures in a well converged manner[8]. How is the interaction between two proteins affected by its shape? Proteins are often found to regulate processes and display recognition in terms of well defined aggregates. Which forces, if any particular, may be ascribed to this and how can we understand these to then possibly template the design? It is in this light I will present the following work and attempt to broaden and clarify the understanding of the mutual interactions between proteins. In this thesis I will in general be brief and make reference to more extensive work regarding the general understanding of the subject and methods I have used, unless the discussion is of more explicit nature. The results and detailed descriptions of methods of calculation will be appended in terms of papers and manuscripts.

[7] K. Kubiak-Ossowska and P. A. Mulheran, J. Phys. Chem. B 115, 8891 (2011).
[8] M. Lund and B. Jönsson, Biophys J 85, 2940 (2003).

Chapter 2

Statistical Mechanics and Thermodynamics

For the entire part of modern science thermodynamics and statistical mechanics have been a corner stone in our understanding of transfer of matter and energy, from the principles behind the conversion of thermal and mechanical work to the structure of fluids and solids. Even though much of the foundation of thermodynamics dates back to 19th century[9,10] it is still an indispensable tool. The drawback is that it lacks the ability to address questions on a molecular level. Why does water freeze at 273 K and argon at 84 K? What properties account for this difference between the substances? To answer such questions one has to resort to statistical thermodynamics. The following chapter will give a brief overview of some of the fundamentals of statistical mechanics and thermodynamics. For many of the details I refer the reader to references[11,12,13].

2.1 Statistical Thermodynamics

It is common to start the derivation with the assumption of *ergodicity*, i.e. that the observable long time average of property $\mathscr{M}$ is the same as the *ensemble* average[13]. That is, one assumes that any trajectory will visit all parts of phase space. In an isolated system where the energy (U), the volume (V) and number of particles (N) are conserved the entropy, (S), is defined as,

$$S = k_{\mathrm{B}} \ln W_{U,V,N}. \tag{2.1}$$

[9] J. W. Gibbs, *The Scientific Papers of J Willard Gibbs* (Dover Publications Inc., 1961).
[10] J. R. Partington, *A short history of chemistry* (Dover Publications Inc., 1989), 3rd edn.
[11] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, 1985), 2nd edn.
[12] D. A. McQuarrie, *Statistical Mechanics* (Harper Row, 1973).
[13] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover Publications Inc., 1986).

W is all possible enumerations that satisfiy the thermodynamic constraints of the system and k_B is Boltzmann's constant. We assume that there is no reason to prefer any state, of the same energy, over another, *equal a priori probability*. Adopting this postulate, we may write the entropy as

$$S = -k_B \sum_i \rho_i \ln \rho_i, \tag{2.2}$$

where ρ is the probability associated with a state. In fact, this result is true regardless of the constraints to the system[14]. W is often referred to as the micro canonical partition function. The different ensembles in statistical mechanics have different partition functions, each accounting for the particular constraints that are imposed. The more common are the canonical

$$Q(N,V,T) = \sum_i W_i e^{-\beta U_i}, \tag{2.3}$$

grand canonical

$$\Xi(\mu,V,T) = \sum_N Q(N,V,T) e^{\beta \mu N} \tag{2.4}$$

and the isobaric partition function

$$\Delta(N,P,T) = \sum_V Q(N,V,T) e^{-\beta P V}. \tag{2.5}$$

T denotes the temperature, P the pressure, μ the chemical potential and β is the inverse thermal energy, $(k_B T)^{-1}$. From the partition function we can easily obtain the distribution of states $\{\rho_i\}$,

$$\rho_i = \frac{e^{-\beta U_i}}{Q(N,V,T)}, \tag{2.6}$$

and define thermal averages of any observable $\mathcal{M}$,

$$\langle \mathcal{M} \rangle = \sum_i \mathcal{M}_i \rho_i = \frac{\sum_i \mathcal{M}_i e^{-\beta U_i}}{Q(N,V,T)}, \tag{2.7}$$

here exemplified with the canonical ensemble. Note that the summation in 2.7 is over all states, rather than energy levels as in 2.3. According to the first postulate this average corresponds to experimentally measured quantities. In statistical mechanics one often work with quite arbitrary distributions: of energy levels, energy states, particle separations etc. These distributions are directly related to free energy differences,

$$\beta(\Phi_i - \Phi_j) = \ln \frac{\rho_j}{\rho_i}, \tag{2.8}$$

[14] E. T. Jaynes, The Physical Review **106**, 620 (1957).

where Φ is an arbitrary potential. The probability distributions of any ensemble may actually be obtained in a general manner without reference to the postulates above if Equation 2.2 is regarded as a unique measure of uncertainty [14]. To pass over to a continuum description one proceeds by replacing the sum of states with an integral over all particles degrees of freedom, momenta and spatial coordinates (phase space). Since the kinetic contribution to the energy is a quadratic function of their momenta and, classically, independent of the interactions and momenta of the other particles, one may solve the kinetic contribution independently. For example, the canonical partition function of a monatomic fluid will then become

$$Q(N,V,T) = \frac{1}{N!\Lambda^{3N}} \int_V e^{-\beta U(\mathbf{\Gamma})} d\mathbf{\Gamma} \tag{2.9}$$

where Λ is the de Broglie length, U now refers to the spatially dependent interaction energy only and $\mathbf{\Gamma}$ is a convenient notation for the spatial coordinates of all particles. The factor of $N!$ takes into account that the molecules are indistinguishable. Some times it is conventional to distinguish the full partition function from the spatially dependent factor in Equation 2.9, often denoted as the *configuration integral*[13]. To avoid double notation I will refrain from this since the present work only considers situations where the kinetic contribution does not affect the distributions. I leave it to the reader to keep this and any possible factor of Λ^{-3N} in mind.

2.2 Thermodynamics

The theory of thermodynamics is built on a few sets of laws that where derived over some hundred years of studies of energy and mass transfer at equilibrium or between states at equilibrium[11]. For this reason it is a phenomenological theory. However, most of these laws appear very reasonable, e.g. that there exists no *perpetum mobili*. Thermodynamics, in the energy representation, prescribes that there exists a potential which is minimum for any type of constraint. The energy, U, is such a potential,

$$U(S,V,N) = TS - PV + \mu N, \tag{2.10}$$

with independent variables S, V and N. T, P and μ are linear response functions of U. Due to the properties of Equation 2.10 it follows that one may change the independent variables by a Legendre transformation[11]. Helmholtz free energy, A, serves as a good example where T plays the role of independent variable rather than S

$$A(T,V,N) = U - \frac{\partial U}{\partial S} S = U - TS. \tag{2.11}$$

In this manner the theory may be adapted to the desired experimental conditions. These potentials are referred to as fundamental equations while response functions like $\mu(S,V,N)$ are called equations of states. Knowledge of the fundamental equations completes the thermodynamic description of any system in difference to the response functions which only contain partial information.

When one discusses phase coexistence in thermodynamics it is important to distinguish between field- and density variables[15]. If two fluids coexist there are some properties that differ between them, density variables, while some must be invariant due to equilibrium, field variables. Examples of the former are the molar Helmholtz free energy and the density and the latter are chemical potential and temperature. The choice of parameters effects how the phase diagram is depicted: as coexistence lines or miscibility regions and how tie lines are drawn. If one can compute the molar Helmholtz free energy as function of molar volume, for a fixed amount and temperature, it is straightforward to identify the densities at coexistence. Since the pressure,

$$P = -\frac{\partial A}{\partial V}, \tag{2.12}$$

is a field variable the coexisting points must have the same derivative. Further, since they must also have identical chemical potential it follows from Equation 2.10, where Euler's theorem has been used, that this is only achieved if the points also have a common tangent.

2.2.1 Adsorption

In many instances it is desirable to study the adsorption of molecules to interfaces. One of the most well used and long lived isotherms is that derived by Langmuir,

$$s = \frac{c\chi}{1 + c\chi}. \tag{2.13}$$

Here s is the coverage of the surface or interface, c is the concentration of the adsorbent in bulk and χ is a parameter that describes the affinity to the surface. The reason for its long survival is that it is simple while capturing the single most distinct properties of adsorption, namely saturation. However, one should also bare in mind that the assumptions underneath are rather severe. The isotherm is the result of an adsorption where all regions of the surface have the same affinity and the adsorbents are treated independently of each other, except that they exclude volume at the surface[13,16]. Other isotherms exist which allow for consideration of multilayers and interactions between adsorbents like the BET isotherm and Bragg-Williams theory [13].

2.2.2 Equilibrium Constants and Titration

In thermodynamics one is often interested in finding equilibrium constants. The equilibrium constant of a reversible process

$$nA + mB \rightleftharpoons iC \tag{2.14}$$

[15] R. B. Griffiths and J. C. Wheeler, Phys. Rev. A 2, 1047 (1970).
[16] D. Evans and H. Wennerström, *The Colloidal Domain* (Wiley-Vch, 1999), 2nd edn.

is given as

$$K = \frac{a_C^i}{a_A^n a_B^m} = \frac{\gamma_C^i}{\gamma_A^n \gamma_B^m} \frac{c_C^i}{c_A^n c_B^m}, \tag{2.15}$$

where a refers to activities, γ to activity coefficients and c to concentrations. The equilibrium constant is truly a constant, which only depends on the temperature. Often, one only has experimental access to concentrations. This in turn can give rise to a stochiometric, concentration dependent, equilibrium constant due to the left factor on right hand side of Equation 2.15. At low concentrations this does not pose any problem since the activity coefficients tend to unity and one can determine thermodynamic equilibrium constants that may be exploited in simulations. Consider a small monoprotic carboxylic acid. These typically have pK_a values of 4.8 at low concentration and in pure water[17], p stands for $-\log$ and K_a is the dissociation constant of the acid. Since the acid is neutral while the dissociation pair is composed of one carboxyl ion and one hydrogen ion, their activity coefficients will change differently with salt concentration, due to electrostatic screening. Figure 2.1 displays

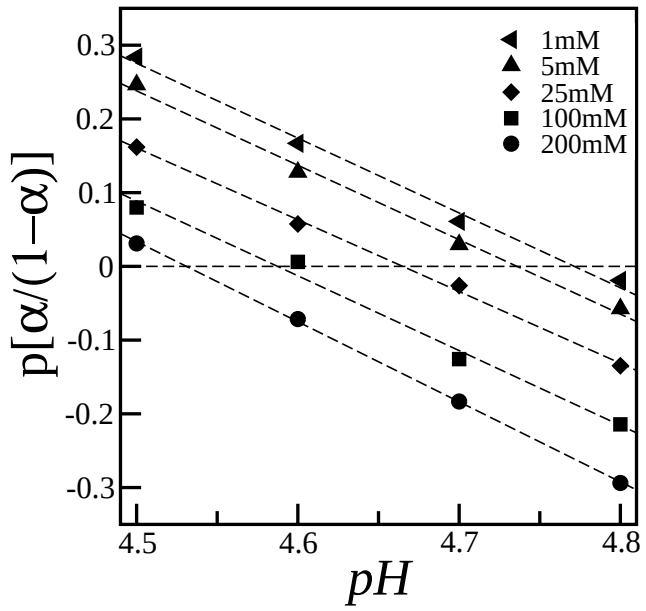


Figure 2.1: Fitted lines give effective pK_a of 4.77 (1 mM), 4.74 (5 mM), 4.67 (25 mM), 4.59 (100 mM) and 4.53 (200 mM).

the result of simulations of one titrating amino acid with a pK_a value of 4.8 at different salt concentrations of a 1:1 electrolyte for different pH-values (pH is $-\log a_{H^+}$). Here α is the degree of dissociation which is sampled from a simulation, performed

[17] M. Smith, *Critical Stability Constants*, vol. 3 (Plenum Press, 1977).

with the primitive model in an identical fashion to Papers II and III. Effective pK_a values are read off where the fitted lines go through zero. 100 mM of monovalent salt gives rise to downshift of -0.2 pK units. This is in very good agreement with experimental observations of salt effects on monoprotic caboxylic acids which show a shift of -0.2 regardless of pK_a-value in pure water[17]. The example demonstrates how one may use thermodynamic equilibrium constants to include free energy differences that arise from bond breakages and solvation effects that are often complicated and expensive to include explicitly.

Chapter 3

Intermolecular Interactions

The interaction of two molecules in any given configuration is governed by their instantaneous mutual and internal orientations and separations. It is in principle possible to write down a formal solution for the energy. The task would then be to solve the Schrödinger equation[18]. However, it is for practical reasons not possible to address this quantum mechanical problem, even with approximate methods, if one wishes to obtain answers to questions with respect to the structure of protein solutions at thermal equilibrium. Special areas do exist where this is feasible[19], but hardly for the problems in scope of the present work. In order to take a step forward, it is possible to characterize the interactions of atoms and molecules into different contributions. These contributions may individually be estimated and taken in account for in modeling. Usually these contributions are divided into electrostatic, exchange repulsion and dispersion interactions. The latter is often augmented with thermally averaged dipole-dipole and dipole-induced dipole interactions, although they are of electrostatic nature. The reason for this joint treatment is that they have the same distance dependence. This joint term is then referred to as a van der Waals interaction. One commonly ignores that some of these contributions are responses to the total electrostatic field, hence are not pair wise additive. In general it is a popular strategy to assume pair wise additivity and see how far it holds. Despite this strongly pragmatic and apparently ignorant treatment, these simplifications of the matter do not pose any greater limitations to obtain reasonable descriptions of molecules nor to explain the structure of fluids[20,16]. Also, work to further develop these simpler descriptions is an active field in the scientific community [21].

[18] B. H. Bransden and C. J. Joachain, *Physics of Atoms and Molecuels* (Prentice Hall, 2003).

[19] A. Öhrn, *Development and Application of a First Principle Molecular Model for Solvent Effects*, Ph.D. thesis, Lund University (2007).

[20] J. Israelachvili, *Intermolecular and Surface Forces* (1992).

[21] A. Holt, *Modelling of polarization by molecular force fields: Further development of the NEMO potential*, Ph.D. thesis, Lund University (2009).

3.1 Coulombic Interactions

Charged molecules interact over large distances because of the very long ranged electric field they emit [20]. The interaction between two charges, separated by a distance r_{ij} is

$$u(r) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \tag{3.1}$$

q are their charges and ϵ_0 is the dielectric permitivity of vacuum. The coulomb interaction has been verified over huge length scales, from 10^{-17} to 10^{7} meters [22]. As will be formalized in the latter part of the chapter, the effective interaction between two molecules is often strongly influenced by the surrounding. The interaction between two charges in a polar solvent is screened by the surrounding solvent molecules since they preferentially will align their dipoles such that the total emitted field is dampened. The strength of this screening is determined by the dipolar density of the solution. The dampening gives rise to a modified interaction between two charges,

$$u(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}. \tag{3.2}$$

ϵ_r is the dielectric constant of the solvent. The effect of the solvent simply scales the coulomb interaction with ϵ_r^{-1}. The most common dipolar fluid, and relevant for this work, is water with a dielectric constant of about 80 at room temperature. The dielectric constant varies with temperature, which underlines that Equation 3.2 is a free energy. Even though the interaction is dampened it still retains the long ranged character. The decay of the interaction between two charges changes dramatically if one considers a solution of ions in a dipolar solvent. The interaction then goes over to Equation 3.3,

$$u(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} e^{-\varkappa r_{ij}}, \tag{3.3}$$

where

$$\varkappa^2 = \frac{1}{\epsilon_0\epsilon_r k_B T} \sum_i c_i q_i^2. \tag{3.4}$$

This is often called the Debye-Hückel interaction. Actually, the exponential decay, on top of the r_{ij}^{-1}, arises due to the long range nature of the field.

3.2 Short ranged Interactions

It is well known that the correlation between electrons of atoms gives rise to a short ranged attraction that goes as r^{-6}. Also, all molecules have 'size' and can not occupy the same space. This is called the exchange repulsion. This latter contribution can

[22] B. Halle and B. Jönsson, *Klassisk Elektrostatik: En fälthandbok* (Kemicentrum Lund, 1985).

only be accurately addressed with quantum mechanics. Still, what is always true is that this repulsion is very short ranged. Two common descriptions of the short ranged interaction between atoms are either as impenetrable spheres or with an interaction that decays as r^{-12}, where the distance is scaled by the mean diameter of the atoms, σ_{ij}. These repulsive contributions are often augmented with an attractive term that goes as r^{-6}. The repulsive term that goes as r^{-12} is simply chosen of mathematical convenience.

$$u(r) = 4\epsilon \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \tag{3.5}$$

$$u(r) = \begin{cases} -\frac{C_{vdW}}{r_{ij}^6} & r_{ij} > \sigma_{ij} \\ \infty & r_{ij} < \sigma_{ij} \end{cases} \tag{3.6}$$

The particular combination or powers in Equation 3.5 is called a Lennard-Jones potential. Sometimes the attractive term in Equation 3.6 is neglected, this is then simply referred to as a hard sphere potential. The fact that these phenomenological descriptions are very much in use, yet old, is that they tend to capture the fundamental physics.

3.3 McMillan-Mayer Theory

One could ask the question, well we can estimate the interaction of e.g. ions in water as above, but what sort of properties would we compute from those interactions as we neglect the explicit presence of the solvent? Will an estimated pressure or a pair distribution really have physical meaning? Luckily they do. One major result from statistical mechanics is the so called McMillan-Mayer theory[13]. Consider the two component partition function, with one of the components at constant number density while the other has constant chemical potential,

$$Z(N_1, T, V, \mu_2) = \sum_{N_2} \frac{e^{\beta \mu_2 N_2}}{N_2! \Lambda_2^{3N_2}} \frac{1}{N_1! \Lambda_1^{3N_1}} \int d\mathbf{\Gamma}_1 d\mathbf{\Gamma}_2 e^{-\beta U(\mathbf{\Gamma}_1, \mathbf{\Gamma}_2)}. \tag{3.7}$$

One can then specify an effective interaction, $\mathscr{W}$, between all particles N_1 in the presence of the second component,

$$e^{-\beta \mathscr{W}(\mathbf{\Gamma}_1)} = \sum_{N_2} \frac{e^{\beta \mu_2 N_2}}{N_2! \Lambda_2^{3N_2}} \int d\mathbf{\Gamma}_2 e^{-\beta U(\mathbf{\Gamma}_1, \mathbf{\Gamma}_2)}. \tag{3.8}$$

If one is able to estimate the influence of the second component it is possible to formulate the problem as an one component problem rather than two. This is highly desirable since it opens the door to significantly decrease the computational effort. If one divides Equation 3.7 with the partition function of a system with no particles of

component 1 and derives the pressure one finds that this is nothing but the osmotic pressure. It also follows that the distribution functions are properly accounted for as well as other properties[13]. Although formally very promising, $\mathscr{W}(\mathbf{\Gamma}_1)$ is very difficult to compute from first principles. In practice one often approximates it as a sum of pair interactions, w_{ij},

$$\mathscr{W}(\mathbf{\Gamma}_1) \approx \sum_{i<j}^{N_1} w_{ij}. \tag{3.9}$$

Combinations of Equation 3.2 or 3.3 with either Equation 3.5 or 3.6 are examples of w_{ij} that are often used in this context. McMillan-Mayer theory is very important since it clearly states that the problems related to the liquid state can be simplified and this is also the starting point for considering coarse grained models.

Chapter 4

Monte Carlo Simulations

The Metropolis Monte Carlo (MC) algorithm is an efficient way to calculate thermal averages and generate distributions from which free energy differences are computable [23,24,25]. Multidimensional integrals like Equation 2.9 span over a huge coordinate space and to solve them numerically one has to settle with evaluating just a few points. The strength of the MC method is that it, ideally, picks those points which give the largest contribution to the integral. The algorithm evolves from one configuration to another where each step only depends on the previous, a so called Markov chain, through a random walk. However, the result is far from random. Instead the algorithm is designed such that each visited configuration is given the correct statistical weight. Thus, it is more appropriate to talk about a random walk within the equilibrium distribution. One should always keep in mind that although the method is often successful there are no guaranties that a simulation will represent the true equilibrium state. This depends on how efficient one can generate configurations, cross possible barriers and that the calculation is sufficiently long. In practice, the only way to assure that the equilibrium state has been sampled is to start the simulation from several different initial conditions and see that all calculations mutually converge. Equilibrium in an MC simulation always has to be ascertained with proper criteria for the particular system: '*Operationally, a system is in an equilibrium state if its properties are consistently described by thermodynamic theory!*' [11].

4.1 Geometries and Boundary Conditions

Since a molecular simulation usually is restricted to something between a few or hundred thousands molecules one has to be cautious if to compare with experimen-

[23] N. A. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953).

[24] M. Allen and D. Tildesley, *Computer Simulation of Liquids* (1989).

[25] D. Frenkel and B. Smith, *Understanding Molecular Simulations* (Academic Press, 2002).

tal measurements of thermodynamic properties. The latter are performed on 'infinitely' large systems, with a negligible contribution from surfaces and interfaces in general. Therefore it is usually desirable to construct bulk like models. A number of choices exist: cell models[26], periodic boundary conditions[27] and closed hyper geometries[28]. In a cell model[29] one restricts the attention to the behavior of a central part of a cavity. This inner region is supposed to resemble bulk conditions. Surface effects might still be found at the boundaries but this region is not sampled or the effects may be corrected for with an external potential[30,31]. One way to remove surfaces is to impose periodic boundary conditions. This is usually done in cubic geometry. One may truncate at the nearest neighbor level, leading to the minimum image approximation[25], or include an infinite lattice sum, the Ewald summation[32]. Although the periodic boundaries remove the surface, the method has the intrinsic property of not being isotropic. This may be addressed with a spherical cut off and long range corrections [27,25]. In the end, any choice should give the same result as the size of system becomes very large. The limit is usually reached at different rates and at different computational costs. The choice ought to be made from practical convenience and with caution as to include all necessary length scales: range of interaction, aggregate sizes etc., that the particular system exhibits. For comparisons I refer to the review by Linse[33]. External corrections to inhomogeneous systems also exist, like the charged sheet method[34] or different Ewald techniques[35,36].

4.2 Evolution of the Chain

In order to obtain a large representative number of configurations it is necessary to somehow generate new configurations. This is achieved by different schemes that are more or less efficient, depending on which system that is studied. The simplest way is to generate random configurations. Such procedures have the advantage that any proposed configuration will be uncorrelated with the previous one. However, if the probability density is such that it is vanishingly small over a large region of configuration space, it will be very unlikely to generate a new acceptable configuration and the chain will progress very slowly. A second approach is to change the configuration a bit at the time, e.g. one may try to perturb the location or orientation of one particle. Such a variation will of course be easier to accept, but at the same

[26] C. Woodward and B. Svensson, J. Phys. Chem. **95**, 7471 (1991).
[27] B. Smith and C. P. Williams, J. Phys – Condens Mat. **2**, 4281 (1990).
[28] M. Trulsson, J. Chem. Phys. **133**, 174105 (2010).
[29] J. A. Barker, Proceedings of the Royal Society of London. A **230**, 390 (1955).
[30] C. Tanford and J. G. Kirkwood, J Am Chem Soc **79**, 5333 (1957).
[31] H. L. Friedman, Molec. Phys. **29**, 1533 (1975).
[32] P. Ewald, Ann. Phys. **369**, 253 (1921).
[33] P. Linse, Adv. Polym. Sci. **185**, 112 (2005).
[34] J. P. Valleau and G. M. Torrie, J Chem Phys **76**, 4623 (1982).
[35] A. H. Widmann and D. B. Adolf, Comp. Phys. Comm. **107**, 167 (1997).
[36] I.-C. Yeh and M. L. Berkowitz, J Chem Phys **111**, 3155 (1999).

time the new configuration will be highly similar of the previous. The probability to accept a new configuration depends on its relative weight compared to the previous. An equilibrium state is by definition stationary, hence the mean flux from and to all configurations must be zero, a criterion called *balance*. This is hard to implement and the firmer condition of *detailed balance*, or some times even stronger ones[25], are usually enforced. This means that the probabilities of two states must be balanced by the probabilities to exchange between them,

$$\rho_1 \pi(1 \to 2) = \rho_2 \pi(2 \to 1). \tag{4.1}$$

$\pi(i \to j)$ is the transition probability between two states. In the Metropolis algorithm the acceptance criterion for a change that is consistent with the constraints of the canonical ensemble is

$$p(1 \to 2) = \min[1, \exp(-\beta(U_1 - U_2)]. \tag{4.2}$$

Additional factors arise in the last condition of Equation 4.2 if one allows for mass and volume transfer etc. due to contraction and expansion of phase space[25].

4.2.1 Non-rejective Markov Steps

Consider the following situation. A system tends to form aggregates of a finite size i.e. the particles attract each other at short range but repel each other at longer separations. Then we might expect a situation where small aggregates, dimers or trimers are formed rather quickly, if we update the coordinates one particle at the time. But in order to form larger structures the small aggregates have to merge. With the proposed procedure this might take a rather long time if the attraction between the particles is strong. We could augment the move with some condition, e.g. if particles are within a certain distance we consider them to be an aggregate and we apply the proposed change to all members of the aggregate. This will reduce otherwise long lived correlations between clusters, but would force us to reject any change of the aggregation number, in order to achive detailed balance. One can instead device an algorithm that works with pairwise Boltzmann factors of all particles[37]. The requirement is that one can find a coordinate transformation which, when performed on all particles, regenerates the initial configuration under all conditions. In the work related to this thesis, this has been implemented as a translation in a cubic periodic symmetry. The algorithm progresses in an iterative fashion, where a random particle is picked and translated. The probability for any of the remaining particles to 'move along' is

$$p_{ij} = \max\left[1 - \exp(-\beta \Delta u_{ij}), 0\right], \tag{4.3}$$

where Δu_{ij} is the energy difference between the moved and remaining particle before and after the translation. The algorithm now progresses to evaluate the probability of the particles which followed the first one with the remaining, and the chain

[37] J. Liu and E. Luijten, Phys Rev Lett **92**, 035504 (2004).

continues until no more are moved. Compared to a sequential update of single particle coordinates, this algorithm is potentially much more expensive. Contrary to the sequential scheme there is no final choice between a new and old configuration.

4.2.2 Parallel Tempering and Phase coexistence

As indicated above it is very desirable to generate uncorrelated configurations. One appealing strategy would be to exchange all particle coordinates of two systems. Let's consider the case of two systems i and j where the Hamiltonian depends on some parameter λ. Also, these systems are exerted to some external pressure P. Then the acceptance probability to interchange the coordinates between the systems, $acc[(i,j) \rightarrow (j,i)]$, and exchange λ and P is

$$\min[1, \exp(-\beta\Delta(U(\lambda) + PV)] \tag{4.4}$$

where $\Delta(U+PV) = U_i(\lambda_j) - U_i(\lambda_i) + (P_j - P_i)V_i + U_j(\lambda_i) - U_j(\lambda_j) + (P_i - P_j)V_j$ [38]. This is often referred to as parallel tempering. Parallel since one has to keep a number of simulations running in parallel and tempering since it modulates the effective interactions. If we chose to vary the temperature Equation 4.4 would look slightly different[38]. In the presence of widely separated free energy minima this method is often an efficient way to pass over barriers by varying the Hamiltonian such that the interactions are decreased. To obtain sufficient acceptance, the variation in the effective interaction can not be too large between two systems. Therefore one uses several systems in a ladder of λ. This approach is most appealing if one is interested in all values of λ. One straightforward way to establish weather or not a particular system has a coexistence region is by inspecting the Helmholtz free energy as function of volume. This is possible by exploiting a constrained free energy [39]. The distribution of volumes in the NPT ensemble is proportional to

$$\rho(V) \propto e^{-\beta PV} \int_V e^{-\beta U(\mathbf{\Gamma})} d\mathbf{\Gamma}. \tag{4.5}$$

The constrained free energy, $\widetilde{G}$, associated with the probability distribution is then given as

$$-\ln[\rho(V)] = \beta\widetilde{G} = \beta A(V; P_i) + \beta PV \tag{4.6}$$

where P_i denotes the true internal pressure $\partial A/\partial V$. This function has minima or minimum for volumes where $P_i = P$ but there is nothing that stops us from estimating the constrained free energy over the entire span of volumes. Further, we may use Equation 4.6 to compute the Helmholtz free energy as a function of volume since

$$A = \widetilde{G} - PV. \tag{4.7}$$

[38] M. Doxastakis, V. G. Macrantzas, and D. N. Theodorou, J Chem Phys 115, 11352 (2001).
[39] J. E. Hunter and W. P. Reinhardt, J Chem Phys 103, 8627 (1995).

Most fluids exhibit a miscibility gap where a dense and a dilute liquid coexist. At conditions of coexistence $\widetilde{G}$ has two equally deep minima, corresponding to the coexisting densities. It is not even necessary to simulate at the coexistence pressure, since the constrained free energy may simply be transformed to any pressure. For practical considerations, it is desirable to perform calculations close to conditions of coexistence to improve sampling. For a macroscopic system, the coexistence points may be found by balancing the applied pressure such that the probability density of the points are equal. For small systems such estimate of these densities may suffer in precision due to very different fluctuations about the coexistence points. This difficulty may be addressed by balancing an integrated probability density about the minima. Nevertheless, the method will most certainly reveal if any coexistence exists. In comparison to the Gibbs ensemble technique[40] this approach allows one to use roughly twice the number of particles. With the latter technique one usually tries to pass through the immiscibility gap such that half the number of particles are in a dense/dilute regime while in the constrained approach all particles contribute to the sampling of all densities. The drawback of the constrained free energy approach is that as the size of the system increases, so does the barrier which has to be crossed. Since a reliable $\widetilde{G}$ relies on a proper sampling of all $Q(N, V_i, T)/Q(N, V_j, T)$ this may only be achieved with a biased sampling for large systems. The Gibbs ensemble, on the other hand, completely removes the problem of passing barriers and is probably more efficient for large systems.

[40] B. Smith, P. de Smedt, and D. Frenkel, Molec. Phys. **68**, 931 (1989).

Chapter 5

Summary of Results and Concluding Remarks

This chapter summarizes the result in the papers which are appended to the thesis.

5.1 Anisotropic Protein Interactions

Using a description of proteins based on the experimentally determined structure is important. The interactions are complex and we can hardly expect a good description of the phase behavior based on very simple models. Protein structures are often not well described as simple geometric objects. This goes for their charge distribution as well as their spatial extension. Some consequences of these complicated structures are addressed in Paper II, III, IV, VI, VII and IX. Although many details in these papers are interesting, I wish to highlight two broad themes of general importance.

5.1.1 Mutual Amplification

Paper II, IV and IX describe how the electrostatic and van der Waals components enhance each other for certain proteins. This is examined for two different systems, Lactoferrin and Lysozyme-HyHEL5. It seems like the shape and distribution of charges of *Bovine* Lactoferrin give rise to, spatially, common minima, of the components in the pair potential. It is this apparent design that is responsible for the formation of a stereospecfic aggregate. The explicit demonstration of amplification in Paper IX shows that this effect can be profound and the aggregation can not even be qualitatively described by an overlap approximation. This clearly casts doubts on prevalent use of DLVO-like representations of proteins and suggests that more detailed protein models are important to assure theoretical predictions of proteins interactions.

5.1.2 Orientational Effects

The density distribution of proteins about each other is highly inhomogeneous in many cases. This is demonstrated in Paper II, III, IV and VI. In some particular situations, like the systems of Lysozyme-HyHEL5 and Lysozyme-αLactalbumin, this can be well understood from a multipole expansion of the charge distribution. In the discussion of Paper III it is hypothesized that the strong alignment of αLactalbumin is a real feature of its heterodimer with Lysozyme. This is verified experimentally in Paper IV. We also suggest, in Paper III, that the asymmetric charge distribution of αLactalbumin may be important for the formation of macroscopic aggregates. In Paper VII we examined this in a study of the phase behavior of αLactalbumin-Lysozyme mixtures. The findings clearly show how the complex charge distribution is crucial for the location or even existence of a two phase region. The experimental observations are well represented by the model, while neglect of charge asymmetry does not capture the physics of the system.

5.2 Protein Adsorption

The adsorption of Bovine Serum Albumin (BSA) onto polymer dressed nanoparticles is the main subject of Paper V. It is found that BSA manages to adsorb even though the particle surface is heavily coated with monomers. The paper evaluates a few simple concepts to quantify the adsorption and to predict the interaction of nanoparticles with the immune system. In order to design drug delivery vessels that are resistant to immune response we found that it is important to control the properties of the polymer brush.

5.3 pH-detection in Nano Compartments

Paper XIII describes a model of a molecular pH-meter, Glu^3. The paper shows how one can use a model, that is principally the same as the protein model, to capture the behavior of titrating molecules. The determination of pH with Glu^3 is highly dependent on electrolyte conditions. With the model we are able to show that the probe response to pH is independent of counterion concentration and that the response is not perturbed by the interactions of a zwitterionic membrane, even though the probe itself might adsorb. Hence, the model suggests that the probe has broad applicability.

5.4 Concluding Remarks

From the result presented in this thesis it is clear that the shape of proteins, in the broadest sense, is elementary for their aggregation. This is a vital property that is highly relevant for the understanding of complex fluids. Previous results have shown that for isotropic interactions the phase diagram follows a corresponding-state law,

where the reduced second virial coefficient plays the role of reduced temperature[41, 42]. However, if directional forces are introduced the state diagram no longer follows such universal behavior[5]. This underlines the importance of numerical simulations of molecular models to determine the liquid state of protein solutions. It is intriguing that a simple model, as in this work, is able to capture such complex behavior as well defined dimer formation. Though the model is advanced in comparison to others[43,44,45,46,47] it is fair to ask how the results compare with a detailed model of the amino acids, also including solvent explicitly. From practical considerations it does not appear feasible to use a fully atomistic description. Rather, the most likely way forward is to be found in a similar way, as the long ranged interactions are handled in Paper IX. It would be highly desirable to devise models with a resolution of the interactions that varies with separation, all the way to distances where the amino acids are only separated by a few water molecules. Better descriptions of the amino acid's effective interactions might also be obtained from a more stringent coarse graining scheme[48,49]. Despite my claim of new and extended understanding of protein-protein interactions there are still lots to examine within the subject. Finaly I hope the reader has found the work in this booklet interesting and the details in the papers inspiring for their future research and wondering.

[41] G. A. Vliegenthart and H. N. W. Lekkerkerker, J. Chem. Phys. **112**, 5364 (2000).
[42] P. Katsonis, S. Brandon, and P. G. Vekilov, J. Phys. Chem. B **110**, 17638 (2006).
[43] W. R. Bowen, X. Cao, and P. M. Williams, Proc. R. Soc. Lond. A **455**, 2933 (1999).
[44] F. Carlsson, P. Linse, and M. Malmsten, The journal of physical chemistry B, Condensed matter, materials, surfaces, interfaces and biophysical **105**, 9040 (2001).
[45] T. W. Rosch and J. R. Errington, J. Phys. Chem. B **111**, 12591 (2007).
[46] M. C. Abramo, C. Caccamo, D. Costa, G. Pellicane, and R. Ruberto, J. Phys. Chem. B **114**, 9109 (2010).
[47] Y. Liu, L. Porcar, J. Chen, W.-R. Chen, P. Falus, A. Faraone, E. Fratini, K. Hong, and P. Baglioni, J. Phys. Chem. B **115**, 7238 (2011).
[48] K. Maksimiak, S. Rodziewicz-Motowidlo, C. Czaplewski, A. Liwo, and H. A. Scheraga, J. Phys. Chem. B **107**, 13496 (2003).
[49] S. A. Hassan, J. Phys. Chem. B **108**, 19501 (2004).

Acknowledgements

Så är det dags att ge de människor som möjliggjort denna avhandling vederbörlig uppskattning. Att lära mig en massa nya ting och därefter göra något vettigt av det hela är en konst som kräver sin domptör. Det har varit fruktansvärt roligt att ha er, Torbjörn, Mikael och Janne, som handledare under dessa år. Jag har haft både nytta och nöje av den här tiden med er. Jag måste rikta ett stort tack till de tre musketörerna Martin, Martin och Maxime. Var helst våra upptåg tagit plats, från Ven till Amsterdam, har jag haft det otroligt roligt. Ni har även varit ovärdeliga när jag tvivlat på programmen och letat med ljus och lykta efter smarta lösningar. Bosse och Gunnar, ni ska ha tack för många tips och funderingar blandat med regelrätta sågningar i allmänt gemyt. Magnus ska ha en stor eloge för arbetet med att hålla räknemaskinerna välsmorda, utan vilka jag vore hjälpslös. Till Anıl vill jag säga att det varit otroligt roligt att dela kontor med dig. Det känns trist att lämna dig ensam med hibiskusen, men jag är säker på att du klarar dig utmärkt. Måste också passa på att tacka Asbjørn som, förutom det vetenskapliga pratet, har öppnat dörren till en norska utöver det vanliga. Jonas och Samuel, det har varit mycket trevligt med alla fikastunder och lycka till med skrivandet. Sedan ska jag passa på att tacka Lennart för alla fantastisk föreläsningar om kemi i alla dess former, det betydde mycket. Så vill jag även rikta en liten tacksam tanke till Björn som så plirigt gav en första tjuvtitt av den teoretiska sidan av kemin för tio år sedan.

Till sist vill jag lämna några ord till dem som verkligen gjort denna avhandling möjlig. Ett enormt tack till Mamma och Pappa. Exakt vad ni gjort för att jag skulle falla för kemin har jag ingen aning om, men utan er skulle jag knappast vara här.

Så har jag spart det bästa, Sandra. Jag vet inte riktigt var eller hur jag skall börja, du har stöttat i vått och torrt och på alla vis. Inte minst med och tillsammans med Tage, vår fantastiska glädjespridare. Tack.

Paper I

Lund, Trulsson and Persson
Source Code Biol Med, 3, 1 (2008)
©2008 Lund et al.

Source Code for Biology and Medicine

BioMed Central

Software review

Open Access

Faunus: An object oriented framework for molecular simulation

Mikael Lund*[1], Martin Trulsson[2] and Björn Persson[2]

Address: [1]Institute of Organic Chemistry and Biochemistry, The Academy of Sciences of the Czech Republic, Flemingovo nam.2, CZ-16610 Prague 6, Czech Republic and [2]Department of Theoretical Chemistry, University of Lund, P.O.B 124 SE-22100 Lund, Sweden

Email: Mikael Lund* - mlund@mac.com; Martin Trulsson - martin.trulsson@teokem.lu.se; Björn Persson - bjorn.persson@teokem.lu.se

* Corresponding author

Published: 1 February 2008

Source Code for Biology and Medicine 2008, **3**:1 doi:10.1186/1751-0473-3-1

Received: 18 December 2007
Accepted: 1 February 2008

This article is available from: http://www.scfbm.org/content/3/1/1

© 2008 Lund et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: We present a C++ class library for Monte Carlo simulation of molecular systems, including proteins in solution. The design is generic and highly modular, enabling multiple developers to easily implement additional features. The statistical mechanical methods are documented by extensive use of code comments that – subsequently – are collected to automatically build a web-based manual.

Results: We show how an object oriented design can be used to create an intuitively appealing coding framework for molecular simulation. This is exemplified in a *minimalistic* C++ program that can calculate protein protonation states. We further discuss performance issues related to high level coding abstraction.

Conclusion: C++ and the Standard Template Library (STL) provide a high-performance platform for generic molecular modeling. Automatic generation of code documentation from inline comments has proven particularly useful in that no separate manual needs to be maintained.

Background

Molecular simulation has become a standard tool for investigating molecular systems such as proteins, polymer solutions and other colloidal particles. It is safe to say that for biological applications Molecular Dynamics (MD) is by far the most popular method as it provides both static and dynamic properties of the system. Metropolis Monte Carlo (MC) simulation [1], on the other hand, is less utilized and relatively few software packages exists [2-4]. One advantage of MC simulation is that it allows "unphysical" particle moves, enabling a more creative sampling of the configurational space [5]. The tradeoff for this freedom to move particles is the loss of all dynamic information and, in addition, MC programs tend to become less general. However, if one is interested in equilibrium properties only – binding constants, free energy changes, pK_a values etc. – MC simulation may be a good option.

Using a standard, pre-compiled software package should require no prior knowledge of programming and as such can be a fast and practical approach for solving a specific scientific problem. On the other hand, the underlying physical theory is somewhat hidden and there is always a risk that the application is regarded as a "black box" producing numbers. It becomes even worse if new features are to be implemented. The alternative is for researchers to create their own programs. This approach of course requires some programming skills and writing an advanced simulation program from scratch may be an overwhelming -and likely error prone – task. Instead the programmer may resort to existing libraries, thus

Source Code for Biology and Medicine 2008, **3**:1 http://www.scfbm.org/content/3/1/1

approaching the black box situation described above. However, the abstraction level will typically be lower which has several advantages that allows the researcher to (i) have a high level of control, and (ii) experience high performance due a minimalistic design. In this text we present a modular C++ [6] framework or class library that can be used to construct MC simulation programs in an expeditious manner. Other C/C++ libraries for molecular simulation do exist: MDAPI [7], OOMPAA [8], Glotzilla, for example, albeit none of these target Monte Carlo simulation specifically. Due to the common language, faunus can easily interweave these libraries to broaden the intrinsic feature set with additional well-proven code. A successful example of incorporating features from an external library, Gromacs GMX [9], is presented in the text.

Implementation

Object oriented design

The object oriented capabilities of C++ have enabled us to create a more appealing interface than traditional procedural approaches. For example, the handling of particles – a key undertaking of all classical simulations – is provided by a class hierarchy:

```
class point {

public:

   double x, y, z;

   double dist(point &);

   ...

};

class particle : public point {

public:

   double charge, radius;

   ...

};
```

The Standard Template Library (STL) is subsequently used to construct a vector of particles, **vector<particle>**, that allows for easy access and manipulation. For instance, **p[i].radius** will return the size of the *i*'th particle.

Polymorphic classes – Virtual functions

One of the unique features of C++ is *polymorph classes* that allows for very generic and intuitively appealing code. To demonstrate this, we outline the design of our framework for handling the simulation container – see Figure 1. Essentially, the end programmer will want to select among different geometries -a box, sphere, cylinder etc. For each geometry we need functions that can calculate the volume, generate a random point or decide whether a given point falls within the boundaries. We now construct a polymorph class, **container**, that defines the unimplemented *virtual functions*. Derived classes – **box, cylinder** etc. -then implement specialized versions of the functions and the **container** class hence acts as an interface to the various geometries. This means that we can construct functions that accept any geometry derived from the **container** class. For example:

```
double concentration(container &c)

{ return N/c.volume(); }
```

Due to a large overhead, virtual functions may, however, negatively impact performance and are generally avoided in critical, inner loops.

Performance aspects

Function inlining via templates

The most computationally demanding step in most molecular simulations is the evaluation of configurational energies. Hence the applied pair potential must be highly optimized and preferably inlined in all inner loops. This is accomplished by passing a pair potential *class* as a template parameter that creates a local instance inside the inner loop template,

```
class coulomb {

    float energy(particle &a, particle &b)

    {

      return a.charge*b.charge/a.dist(b);

   }

};

template<class T_pairpot>

class innerloop {

   T_pairpot pair;

   float sum(vector<particle> &p) {

      for (i = 0; i<N-1; i++)

         for (j = i+1; j<N; j++)
```

Source Code for Biology and Medicine 2008, **3**:1 http://www.scfbm.org/content/3/1/1

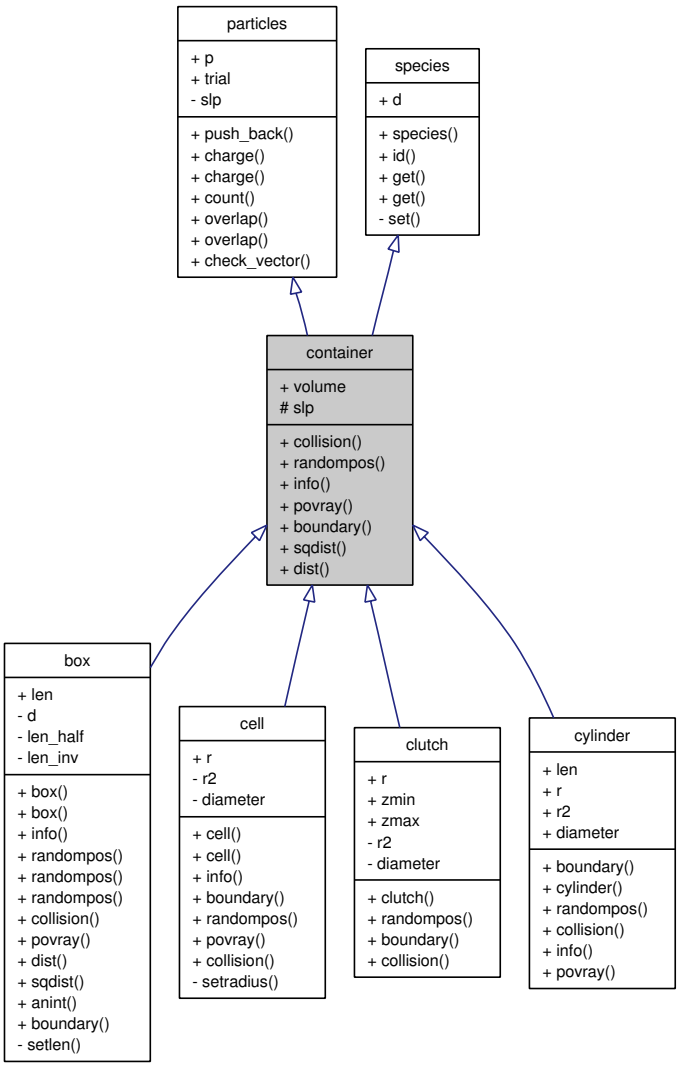


Figure 1
Graphical class hierarchy. Schematic representation of class inheritance used for the container class. Intuitive inheritance is used whenever possible. For example, a container contains particles, it can have a shape etc. The graphical representation is produced using Doxygen.

```
            u = u+pair.energy(p [i], p
[j]);

      ...

};
```

This so-called Expression Template technique [10] enables the programmer to arbitrarily invoke various pair potential functions, re-cycling the inner loop implementation,

```
innerloop<coulomb> elec;

innerloop<lennardjones> lj;

elec.sum(p);

...
```

Passing arguments as references

In C++ standard argument passing is done by creating a new copy of the object. Working with large, aggregate structures such as particle vectors, this will negatively impact performance. To circumvent this we pass all complex objects as references, for example: **void function(someclass &)**.

Minimize memory consumption

Computer simulations of classical mechanical systems are usually not memory intensive and by minimizing the memory requirements there is a good chance that the executing code will stay in the local cache. Code re-cycling via class enheritage is extensively utilized to reduce the memory imprint as well as assist efficient development. In this regard C++ templates are a concern since code will be generated for each template type. We therefore stride to avoid extensive use of multiple template types – for example it would seem silly to instantiate both a **float** and a **double** version of an elaborate template class.

Only for systems with tens of thousands of atoms the particle coordinate vector may extend beyond the local cache – for example, one megabyte of cache can encompass $1024^2/8/3 \approx 43700$ double precision three dimensional particle coordinates. List methods and additional single particle information may decrease this number and, while not yet implemented in faunus, sophisticated methods do exist for cache efficient bookkeeping of many particle systems [11].

Parallelization

In systems that equilibrate fast, Monte Carlo simulations can be linearly parallelized using the "embarrassingly simple technique" – that is start several independent runs with different random seeds, combining the results afterwards. Tightly coupled parallelization is incorporated in parts of the code by threading the energy evaluation into two processes: before and after a trial move. For systems with particles in the order of hundreds, this scales well on dual-core computers, whereas the overhead becomes unacceptable for small systems. To enable threading, the appropriate compiler flag for OpenMP [12] must be set; as of writing the GNU, Intel and IBM C++ compilers all support OpenMP.

Code documentation

We provide a class library and as such need to describe both what the classes do as well as how to use them. This can be conveniently achieved using a code documentation system – here we have chosen **Doxygen**[13] since (i) the documentation appears as normal code comments and (ii) the output is highly configurable, allowing LaTeX equations to be inserted etc. In Figure 2 we show how commented code is used to construct a web based manual of the available classes and functions in the code library. Another very useful feature of Doxygen is the ability to generate a graphical view of the class hierarchy. This enables the end programmer to visually see how a class is constructed as shown in Figure 1.

Results and discussion

General features

The class library provides simulation routines for ions, macromolecules and polymers in solution with a strong focus on electrostatic interactions using the primitive model of electrolytes where the solvent is treated as a structureless dielectric continuum [14]. It is, however, completely possible to expand the library to other systems, include explicit solvent etc. The routines have been developed over several years in connection with a number of scientific investigations, including proteins in solution [15]. As of writing, the code library contains general classes for the following,

• Explicit treatment of ions, including ion-ion correlation effects.

• Macromolecules – Proteins, flexible chains, charged surfaces.

• Charge regulation of molecules [16].

• Particle distribution functions and other statistical mechanical averages.

• Standard file formats are supported: PQR, Gromacs (GRO, XTC), Povray, XYZ.

Source Code for Biology and Medicine 2008, **3**:1 http://www.scfbm.org/content/3/1/1

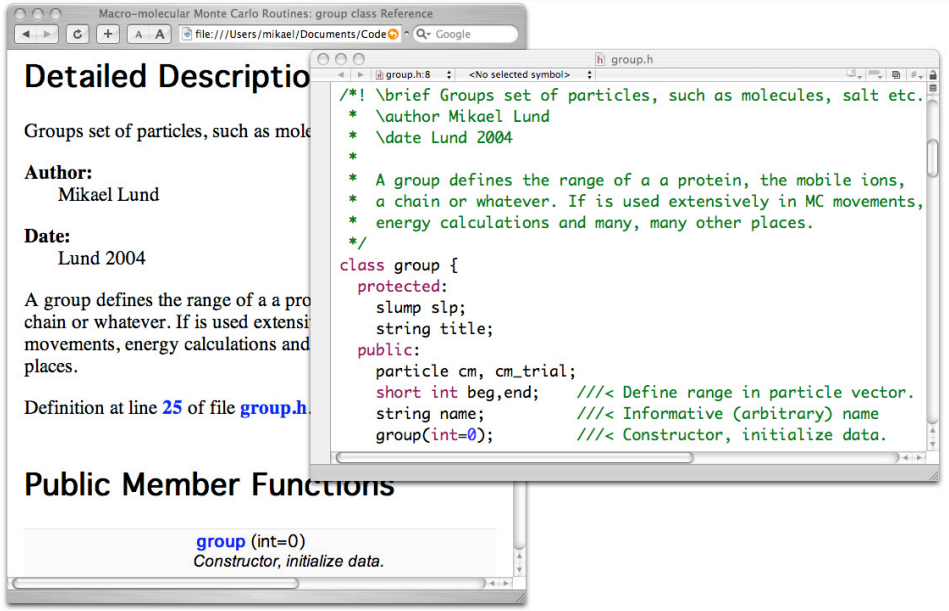


Figure 2
Source code manual. Code documentation through code comments. All code is commented with special keywords that are eventually collected into a web based manual.

An example of protein ionization will be presented later in the text. We stress that the project is under ongoing development and encourage interested users and developers to contribute.

"Trajectory" output
Molecular Dynamics simulation packages often save the time propagated particle trajectory to disk which is subsequently analyzed. In order to adopt this strategy we include an export routine that can write the simulated particle configurations to a compressed Gromacs XTC file [9]. This (sizable) file can be analyzed using the extensive set of tools provided in the Gromacs package, or visualized using VMD [17], for example. As an example of the latter, we have simulated lysozyme interacting with a fab-H fragment and, using VMD, plotted the spatial mass center distribution as shown in Figure 3.

An example: Proton titration
Figure 4 shows – in 50 lines of code – a complete MC program for simulating the protonation state of a protein in a salt solution at a given pH value. Experimentally this corresponds to a standard potentiometric titration experiment where the net-charge is measured as a function of pH [18]. We will not go through all the lines in the code as the comments should be more or less self-explanatory. The overall program structure is

1. Set up the simulation cell (line 12)

2. Add protein(s) and ions (line 17–25)

3. Main loop with salt- and proton moves (line 32–49)

4. Print results and (line 50)

Results and comparisons with experimental data for such calculations can be found in a recent article [19]. The particles in the systems are clustered into groups and derived classes; there is a general group class (line 23) and a class for macromolecules (line 17). Note that we have also incorporated a general polymorphic class for markov moves and data analysis so that all derived classes have a common interface. For example, both the salt move class

Source Code for Biology and Medicine 2008, **3**:1 http://www.scfbm.org/content/3/1/1

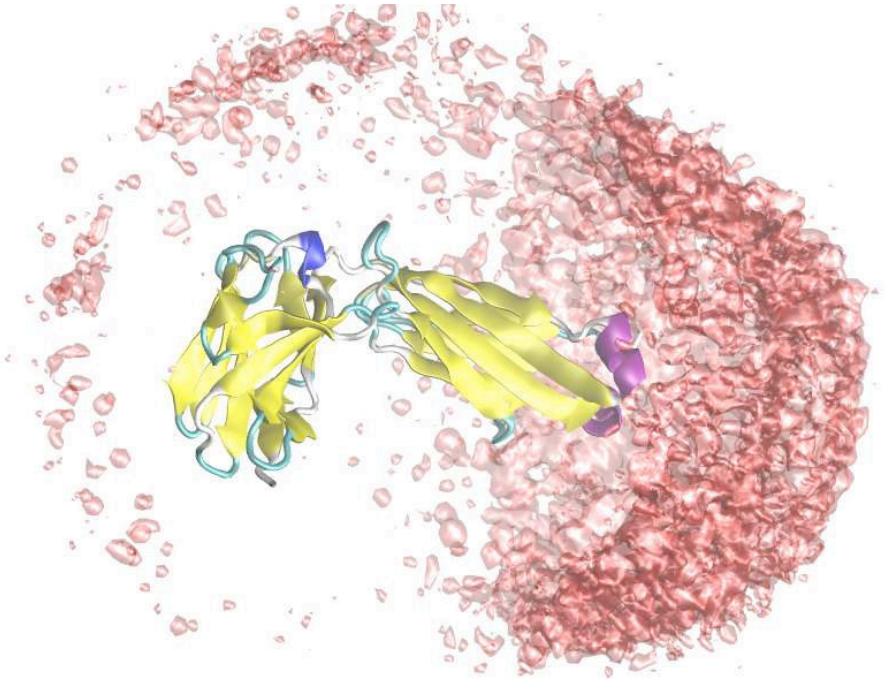


Figure 3
Graphical analysis. Lysozyme interacting with a fab-fragment – a simulation containing more than 340 amino acid residues as well as salt particles. The probability of finding lysozyme's mass-center around the the fab fragment is illustrated by the pink iso-surface. VMD [17] and Povray [20] was used to visualize the generated output from Faunus.

(line 36) and titration class (line 39) will store information about energy changes, if the move was a success etc. Data and analysis about each type of move is automatically shown by calling the respective information functions (line 51).

The source code for this and other examples are included with the class library and should serve as good starting points for developing new programs.

Conclusion

A general class library for (macro-)molecular simulation is presented. We focus on Monte Carlo methods and the primitive model of electrolytes, although we see no technical limitations in expanding the project to cover other methods and molecular levels. The software design is object oriented, meaning that the code is extensively recycled which has several advantages:

- Programs can be developed in a modular manner (à la "Lego" bricks).
- Development and debugging is reduced.
- Memory requirements are minimized.

The class library is documented through extensive use of inline code comments. These comments are subsequently collected by a third party program (Doxygen) that will automatically construct a code manual and, hence, obsolete a separately maintained instruction book. In 50 lines of C++ code we demonstrate how to construct a complete MC program that can simulate protein protonation states in an aqueous salt solution. High performance in inner loops is established using Expression Templates, completely compatible with the flexibility and intuitive appeal of an object oriented design.

```
#include <iostream>
#include "analysis.h"
#include "container.h"
#include "potentials.h"
#include "countdown.h"
typedef pot_coulomb T_pairpot;          // Specify pair potential
#include "markovmove.h"

using namespace std;

int main() {
  cell::cell con(100.);                 // Use a spherical container
  canonical nvt;                        // Use the canonical ensemble
  pot_setup cfg;                        // Setup pair potential (default)
  interaction<T_pairpot> pot(cfg);      // Functions for interactions
  countdown<int> clock(10);             // Estimate simulation time
  macromolecule protein;                // Group for the protein
  ioaam aam(con);                       // Protein input file format is AAM
  protein.add( con, aam.load(
        "calbindin.aam" ) );            // Load protein from disk
  protein.move(con, -protein.cm);       // ..translate it to origo (0,0,0)
  protein.accept(con);                  // ..accept translation
  group salt;                           // Group for salt and counter ions
  salt.add( con, particle::NA, 34+19); //   Insert sodium ions
  salt.add( con, particle::CL, 34 );    //   Insert chloride ions
  saltmove sm(nvt, con, pot);           // Class for salt movements
  aam.load(con, "confout.aam");         // Load old config (if present)
  chargereg tit(nvt,con,pot,salt,7.6); // Prepare titration. pH 7.6
  systemenergy sys(pot.energy(con.p)); // System energy analysis
  cout << con.info() << tit.info();     // Some information

  for (int macro=1; macro<=10; macro++) {   // Markov chain
    for (int micro=1; micro<=1e3; micro++) {
      switch (rand() % 2) {                 // Randomly chose move
        case 0:
          sys+=sm.move(salt);               // Displace salt particles
          break;
        case 1:
          sys+=tit.titrateall();            // Titrate protein sites
          protein.charge(con.p);            // Re-calc. protein charge
          protein.dipole(con.p);            // Re-calc. dipole moment
          break;
      }
    }                                       // END of micro loop
    sys.update(pot.energy(con.p));          // Update system energy
    aam.save("confout.aam", con.p);         // Save config. to disk
    cout << "Macro step " << macro
         << " completed. ETA: " << clock.eta(macro);
  }                                         // END of macro loop
  cout << sys.info() << sm.info()           // Print results
       << tit.info() << salt.info() << protein.info();
}
```

Figure 4
Source code example. Example of a Monte Carlo simulation program to calculate protein ionization states in an aqueous salt solution using explicit ions and the detailed three-dimensional protein structure.

Availability and requirements

Project name: faunus

Project home page: http://faunus.sourceforge.net

Operating systems: MacOS X, Linux, "Cygwin"

Programming language: C++

Other requirements: C++, Doxygen (optional)

License: GNU GPL

Restrictions to use by non-academics: GNU GPL

The latest version can be downloaded using the versioning control system "subversion" (SVN). On most UNIX type operating systems this is done by invoking the following shell command,

```
$ svn checkout

http://faunus.svn.sourceforge.net/

svnroot/faunus/trunk faunus
```

Prospective developers are welcome to contact the authors for write access to the online code repository, currently hosted by Sourceforge, Inc.

Authors' contributions

ML did the Faunus software design and wrote the paper. BP and MT contributed to the implementation. All authors have read and approved the final manuscript.

Acknowledgements

The authors would like to thank the Research School of Pharmaceutical Sciences, Sweden, the Linnaeus Center of Excellence on Organizing Molecular Matter, Sweden, and the European Molecular Biology Organization. We also thank Pavel Jungwirth for useful discussions and Sourceforge, Inc. for hosting the project.

References

1. Metropolis NA, Rosenbluth AW, Rosenbluth MN, Teller A, Teller E: **Equation of State Calculations by Fast Computing Machines.** *J Chem Phys* 1953, **21:**1087-1097.
2. Kamberaj H, Helms V: **Monte Carlo simulation of biomolecular systems with BIOMCSIM.** *Computer Physics Communications* 2001, **141(3):**375-402.
3. Carlsson F, Malmsten M, Linse P: **Monte Carlo Simulations of Lysozyme Self-Association in Aqueous Solution.** *J Phys Chem* 2001, **105:**12189-12195.
4. Hu J, Ma A, Dinner AR: **Monte Carlo simulations of biomolecules: The MC module in CHARMM.** *J Comp Chem* 2006, **27(2):**203-216.
5. Frenkel D, Smit B: *Understanding Molecular Simulation* San Diego: Academic Press; 1996.
6. Stroustrup B: *The C++ Programming Language* 3rd edition. Boston: Addison-Wesley; 1997.
7. **MDAPI** [http://www.ks.uiuc.edu/Development/MDTools/mdapi/]
8. **Object-Oriented Model for Probing Assemblages of Atoms** [http://mccammon.ucsd.edu/~oompaa]
9. Berendsen H, Spoel D, Drunen R: **GROMACS: A message passing parallel molecular dynamics implementation.** *Comp Phys Comm* 1995, **91:**43-56.
10. Veldhuizen T: **Expression Templates.** *C++ Report* 1995, **7:**26-31.
11. Meloni S, Rosati M, Colombo L: **Efficient particle labeling in atomistic simulations.** *J Chem Phys* 2007, **126:**121102.
12. Dagum L, Menon R: **OpenMP: An Industry-Standard API for Shared-Memory Programming.** *IEEE Computational Science and Engineering* 1998, **05:**46-55.
13. **Doxygen** [http://www.stack.nl/~dimitri/doxygen/index.html]
14. Hill TL: *An Introduction to Statistical Thermodynamics* New York: Dover Publications Inc; 1986.
15. Lund M: *PhD Thesis: Electrostatic Interactions in and between bio-molecules* Lund, Sweden: Lund University; 2006.
16. Lund M, Jönsson B: **On the charge regulation of proteins.** *Biochemistry* 2005, **44(15):**5722-5727.
17. Humphrey W, Dalke A, Schulten K: **VMD – Visual Molecular Dynamics.** *J Mol Graphics* 1996, **14(1):**27-8-33-8.
18. Tanford C, Roxby R: **Interpretation of protein titration curves. Application to lysozyme.** *Biochemistry* 1972, **11:**2192-2198.

19. Lund M, Jonsson B, Woodward CE: **Implications of a high dielectric constant in proteins.** *J Chem Phys* 2007, **126:**225103.
20. **POV-Ray – The Persistence of Vision Raytracer** [http://www.povray.org]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral

Paper II

Persson , Jönsson and Lund
J. Phys. Chem. B 113, 10459 (2009) ©2010 American Chemical Society. Reprinted with permission.

J. Phys. Chem. B **2009,** *113,* 10459–10464 **10459**

Enhanced Protein Steering: Cooperative Electrostatic and van der Waals Forces in Antigen−Antibody Complexes

Björn A. Persson, Bo Jönsson, and Mikael Lund*

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

Received: May 15, 2009; Revised Manuscript Received: June 9, 2009

We study the association of the cationic protein lysozyme with several almost neutral protein fragments but with highly uneven charge distributions. Using mesoscopic protein models, we show how electrostatic interactions can align or steer protein complexes into specific constellations dictated by the specific charge distributions of the interacting biomolecules. Including van der Waals forces significantly amplifies the electrostatically induced orientational steering at physiological solution conditions, demonstrating that different intermolecular interactions can work in a cooperative way in order to optimize specific biochemical mechanisms. Individually, the electrostatic and van der Waals interactions lead only to a relatively weak intermolecular alignment, but when combined, the effect increases significantly.

Introduction

The mutual interaction between biomolecules is of importance not only for biological processes in living cells but also for a range of technical applications, including enzymatic reactions, protein recognition, aggregation, and crystallization. How proteins interact with other solutes is dictated by their molecular structures as well as the solution conditions such as salt, pH, and temperature.[1,2] A classic example is the interaction between the superoxide ion with the enzyme superoxide dismutase where both species have the same net charge but the subtle charge distribution determines the enzymatic rate of conversion.[3,4] More recently, a number of publications have dealt with the computationally more involved problem of several interacting proteins, with methods spanning from Poisson−Boltzmann continuum electrostatics,[2,5−9] to explicit salt models,[10−12] to atomistic solvent molecular dynamics simulations.[13,14]

Measured thermodynamic properties of the protein association process—binding constants and osmotic second virial coefficients, for example—represent time averages over *all* microscopic configurations of the solution. That is, averages over all possible protein−protein orientations and conformations as well as salt and solvent positions. This of course involves an enormous number of possible configurations, and calculating thermodynamic properties is hence a demanding task. Approximations are therefore commonly applied, and in this work, we treat the proteins as rigid entities, water molecules are replaced by a structureless dielectric continuum, and salt particles are described as charged, hard spheres (Figure 1). This model simplification allows us to calculate the free energy of interaction of two proteins in a solution at arbitrary salt and pH conditions.

For two equally charged proteins, the free energy of interaction is usually dominated by a strong net repulsion of electrostatic origin, which is modulated by the addition of salt and it becomes substantially screened at physiological conditions. While the free energy of binding is often attributed to more specific short-ranged interactions,[15−18] we here show that an uneven charge distribution, caused by ionized residues, may lead to a considerable orientational ordering of the protein molecules, effectively steering them into specific constellations. This is exemplified for the interaction between hen-egg white lysozyme with the fragments of fab (HyHEL-5 antigene−antibody complex) using Monte Carlo simulations. While steering interactions between proteins are well-known,[6,8,19−21] we here show that, when combined with anisotropic van der Waals (vdW) forces, protein steering can be significantly enhanced due to nonlinear coupling of angular dependent interactions.

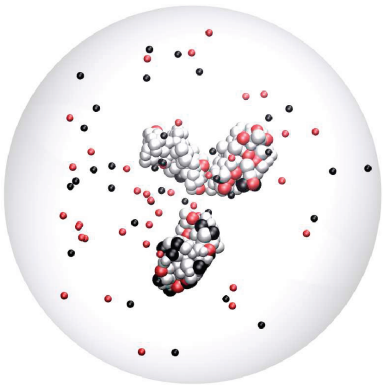


Figure 1. Snapshot of a simulation of a fab fragment and lysozyme in an aqueous salt solution. The proteins and salt particles are colored according to their charges: gray/red = negative, black = positive. Befittingly, the white background illustrates the structureless continuum solvent.

Methodology

Monte Carlo Simulation. Mobile ions and residues in the rigid proteins are mimicked by charged or neutral hard spheres, while the solvent is treated as a dielectric continuum. Each sphere, representing an amino acid, is placed at the center of mass of the residue with respect to the internal coordinates taken

* To whom correspondence should be addressed. E-mail: mikael.lund@teokem.lu.se.

10.1021/jp904541g CCC: $40.75 © 2009 American Chemical Society
Published on Web 07/07/2009

from PDB entry 3HYL containing the complex between the lysozyme and the complete antibody.[22] The radius of an amino acid sphere is determined so as to give the residue the same molecular volume as in the atomic structure, which has been shown to give a good representation of the protein surface.[12] The system, enclosed in a spherical cell with a diameter of 200 Å, contains two proteins as well as counterions and salt (see Figure 1), and we sample configurations of the canonical ensemble distribution using the traditional Metropolis Monte Carlo (MC) algorithm,[23] averaging over all protein angles, charge states, and ion configurations. The total number of particles in the simulations varies with salt concentration, and in the largest system treated here, the complete fab molecule plus lysozyme at 100 mM salt concentration amounts to 1200 ions and amino acid spheres. The attempted configurations consist of random moves of all counterions and salt particles (radii 2 Å) as well as random moves and rotations of the two proteins. In addition, the charge states of the N_{tit} titrating amino acids—Glu, Asp, His, Lys, Arg, and N- and C-terminal groups—are allowed to vary in the simulation.[24] The configurational energy is based on the following Hamiltonian

$$U_{\mathrm{tot}} = \sum_{i\neq j}^{N_{\mathrm{all}}} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} - \sum_{i}^{N_{\mathrm{a}}}\sum_{j}^{N_{\mathrm{b}}} \frac{C_{\mathrm{vdW}} kT}{r_{ij}^{\,6}} \pm \sum_{i}^{N_{\mathrm{tit}}} (\mathrm{pH} - \mathrm{p}K_{\mathrm{a},i}) \tag{1}$$

where r_{ij} is the distance between particles i and j, q their charge, ε_0 the permittivity of vacuum, $\varepsilon_r = 78.7$ is the relative dielectric constant of water, $T = 298$ K is the temperature, k is Boltzmann's constant, and pK_{a} is the intrinsic acidity constants of the ionizable amino acid groups.[24] The first sum is performed over all particles (salt ions and amino acid residues) in the simulation cell, while the second is over the amino acids in the two proteins a and b, only. This corresponds to the primitive model of electrolytes[25] but with an additional van der Waals (vdW) term acting between amino acid residues. The vdW parameter, C_{vdW}, is set to 25000 Å^6, which is reasonable for the interaction across water.[12,26] The model does not explicitly account for specific short-range hydrophobic interactions,[27−29] yet for the overall, more long-ranged electrostatic interactions, the solvent mediated Coulomb potential performs remarkably well, as reflected in comparisons with experimentally determined second virial coefficients for protein complexes.[7,10,12,30] All MC simulations were performed using the freely available coding framework, Faunus,[31] while the molecular graphics were created using the VMD package.[32] Note that Monte Carlo simulations can be used to study static or equilibrium properties only and that these, within the same model, are *identical* to results obtained using molecular or Brownian dynamics.[33]

Analysis. The protein−protein radial distribution function, $g(R)$, is sampled and subsequently used to calculate the potential of mean force, or Helmholtz free energy of interaction,

$$A(R) = -kT \ln g(R) + \mathrm{const} \tag{2}$$

where R is the mass center separation of the two proteins. We also perform a detailed analysis of the interprotein angular properties: The two protein mass centers are connected by the system z-axis, and the degree of dipolar alignment of a single protein can therefore be probed by averaging the z-component of the molecular dipole moment unit vector, $\langle \mu_z \rangle$; see the Appendix. The dipole moment is not uniquely defined for a charged molecule and depends on the choice of origin for the coordinate system. Here, we calculate the dipole moment with respect to the center of mass which is compatible with commonly used multipole expansions.[26,34] The direction of the dipole moment varies of course during the simulation due to the random rotations of the proteins, but it varies also in both direction and magnitude due to the titration of the acidic and basic residues.[35] If $\langle \mu_z \rangle = \pm 1$, then the dipole is fully (anti)aligned, while if the protein can, on average, rotate freely it is *zero*. We use μ_z to gain insight into two distinct characteristics of protein−protein interactions, namely, the *specificity* by studying the probability function, $P(\mu_z)$, at a fixed interprotein separation (typically the free energy minimum) and the *range* of the alignment by following $\langle \mu_z \rangle$ as a function of the protein−protein separation. The probability function, $P(\mu_z)$, further allows us to estimate the alignment free energy difference between discrete projections of the dipole, μ_z and μ_z',

$$\Delta A_{\mu_z \to \mu_z'} = -kT \ln \frac{P(\mu_z')}{P(\mu_z)} \tag{3}$$

Finally, let us briefly discuss when a molecular dipole moment can be considered to be "large". The answer depends on the size of the molecule and the type of interaction. For example, using a multipole expansion,[26] the ion−dipolar free energy of interaction is proportional to μ^2/R^4, while the dipole−dipole free energy is proportional to μ^4/R^6. Reformulating the energy in terms of the molecular size, $\propto V^{1/3} \propto N^{1/3}$, we can obtain a more comparative measure of the dipolar magnitude, $\mu^2/N^{4/3}$, for the ion−dipole term at contact where N is proportional to the number of atoms (or residues) in the protein.

TABLE 1: Properties of the Four Proteins Obtained from MC Simulations at pH 7.6 and at a Salt Concentration of 100 mM[a]

protein	N_{residues}	pI	Z	μ (D)	R_{min} (Å)	$\mu^2/N_{\mathrm{residues}}^{4/3}$
lysozyme	129	10.9	6.8	140		30
L-chain	214	7.6	−0.1	740	57	428
H-chain	225	7.0	−1.3	595	49	259
HL complex	439	7.1	−1.4	230	54	16

[a] The isoelectric points, pI, have been obtained from a set of simulations at variable pH. R_{min} is the separation in the free energy minimum when interacting with lysozyme (vdW and electrostatic interactions included), and the last column shows the effective dipole strength; see text. The dipole moment is given in debyes (D).

Results and Discussion

Overview of the Studied Proteins. The electric properties of the individual proteins studied are shown in Table 1. At pH 7.6, lysozyme carries a net charge, $Z = \sum_i q_i/e$, of close to +7 and an insignificant electrical dipole moment (defined relative to the center of mass). At the same condition, both the heavy (H) fab and the combined complex (HL) carries a small negative charge, while the light (L) fragment lacks a monopole, since the pH coincides with the iso-electric point, pI. Even though the net charges of H, HL, and L are small or zero, they all have an appreciable number of basic and acidic residues. For H and L, these give rise to large dipolar moments and naively we would expect their interaction with lysozyme—with a high net charge—to resemble that of an ion interacting with a dipole.[34,36] This implies an orientational alignment or "steering" of the dipolar protein. Making use of our previously derived "effective dipole moment", i.e., $\langle \mu^2 \rangle / N^{4/3}$, we find that the dipolar character

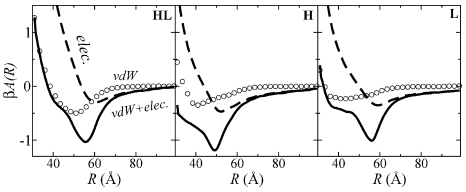


Figure 2. Free energy of interaction of lysozyme with the complete HL protein and the H and L fragments. Simulated at 17 mM salt and with the following protein Hamiltonians ("force fields"): circles, only vdW; dashed, only electrostatics; full drawn, electrostatic and vdW interactions. ($\beta = 1/kT$ is the inverse thermal energy.)

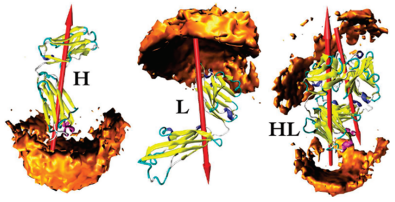


Figure 3. Simulated lysozyme mass center distribution (orange iso-density surfaces) around the heavy (H), light (L), and heavy−light (HL) molecules at pH 7.6 and 17 mM salt. The red arrows represent the dipole moment directions for the H and L fragments. The surfaces represent average iso-densities (8 Å^3 resolution) of 20 times the bulk value and have been sampled from MC generated configurations using the VMD package.[32]

of H and L is roughly 1 order of magnitude larger than that for lysozyme and the HL fragment. We will investigate the dipolar steering in the simulations by calculating orientational correlations between lysozyme and the three antibody fragments H, L, and HL.

Free Energy of Interaction. Figure 2 shows the Helmholtz free energy of interaction, averaged over all protein orientations, charge states, and salt positions, between lysozyme and the antibody fragments. All three effective pair potentials display minima at short separations, and in the presence of 17 mM salt, the depth is roughly 1 kT for all three pairs. In order to dissect the physical contributions to the total free energy, we have performed simulations with three different protein Hamitonians: (i) only electrostatic interactions, i.e., $C_{vdW} = 0$, (ii) only van der Waals interactions, and (iii) both electrostatic and vdW interactions. Hamiltonians i and ii give rise to weak free energy minima at significantly different positions, while the combination of electrostatic and vdW interactions in Hamiltonian iii produces much deeper minima for all three systems. For the H and L fragments, the positions of the total free energy minima coincide with those found when only electrostatic interactions are considered. Thus, there is an apparent nonlinear coupling between electrostatic and vdW interactions that tends to significantly lower the free energy of the protein association.

In light of our previous discussion of the uneven charge distributions of the H and L fragments, we now investigate the angularly dependent binding of lysozyme to these proteins. While the *radial* distribution function, $g(r) = \exp[-w(r)]$, incorporates an angular average over all protein orientations, we can also use MC simulations to calculate an angularly dependent *spatial* distribution function. In Figure 3, we have visualized the preferential protein−protein interaction by plotting the spatial distribution of lysozyme around the three target

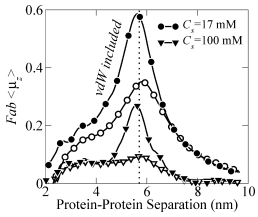


Figure 4. Average z-component of the L fragment dipole moment vector when interacting with lysozyme at pH 7.6 at high and low salt concentrations with (closed symbols) and without (open symbols) vdW interactions. The vertical, dashed line indicates the free energy minimum separation.

molecules. These iso-density surfaces reveal that lysozyme is indeed steered toward the negative poles of the free H and L fragments. As for the complete HL antibody, the two constituting fragments, H and L, are positioned such that their dipoles are antiparallel (Figure 3, right), thus reducing the overall dipole moment of HL to 230 D. While this design in itself demonstrates an interesting mechanism for stabilizing the HL complex, it significantly impacts the interaction with lysozyme. In particular, we note that the latter binds to both ends of HL, with a preference for the negative poles of the constituting fragments. This clearly hints that the lysozyme−HL association has the character not of an ion−dipole interaction but rather of higher order electrical moments.

Instructively, while the effective pair potentials shown in Figure 2 exhibit modest free energy minima of about −1 kT, they camouflage the strong preferential binding of lysozyme toward H and L. In the next section, we show that the alignment free energy can be several kT.

Dipolar Correlations. In the previous section, we demonstrated, using spatial distribution functions, how lysozyme preferentially binds to the negative ends of the H and L fragments and that the binding sites revolve around the poles of the molecular dipole moments, μ. Hence, the dipole moment is an adequate quantity for probing orientational alignments and in this section we track μ-correlations when two proteins interact.

Figure 4 shows the average z-component of the dipole moment of the L fragment as a function of distance to lysozyme. As discussed in the Methodology section, a value of unity means that the dipole is perfectly aligned, while if zero it can rotate freely. It is clear that, as the dipolar fab fragment approaches the cationic lysozyme molecule, the former is aligned or steered in the generated electric field. As vdW interactions are switched on, the dipolar ordering is drastically enhanced from 35 to 60% for 17 mM salt and from 9 to 26% for 100 mM salt. Thus, the vdW-induced amplification of angular ordering is, in relative terms, more pronounced at physiological salt conditions with an enhancement by a factor of 3. It can be attributed to an increment of close contacts that counteract the entropic penalty of aligning the molecules. This result demonstrates a biologically relevant example of a strong coupling between intermolecular forces that leads to an amplification of electrically induced molecular ordering.

To probe the specificity of the protein−protein interaction, we investigate the probability, $P(\mu_z)$, of finding the dipole in a specific projection at a fixed interprotein separation, namely, that of the minimum free energy (see Table 1). As shown in Figure 5, the distributions of H and L are narrow, indicating that lysozyme indeed concentrates around the negative pole of these molecules. Using eq 3, we estimate the alignment free

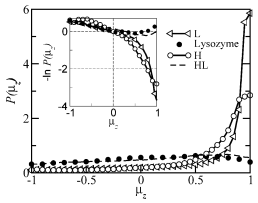


Figure 5. Probability distribution for the dipole moment z-components when lysozyme interacts with H, L, and HL at pH 7.6 and at a salt concentration of 17 mM at a fixed interprotein separation, corresponding to the minimum in free energy. The inset shows the alignment free energy change, cf. eq 3.

TABLE 2: Electrostatic Energy Components for Discrete Configurations[a]

component	energy
system	$\sum_{i\neq j}^{N} q_i q_j / r_{ij}$
protein−protein (pp)	$\sum_i^{N_a} \sum_j^{N_b} q_i q_j / r_{ij}$
ion−ion (ii)	$\sum_i^{N_a} q_i \sum_j^{N_b} q_j / R$
ion−dipole (dd)	$-\sum_i^{N_a} q_i \sum_j^{N_b} q_j z_j / R^2 + \sum_j^{N_b} q_j \sum_i^{N_a} q_i z_i / R^2$
ion−quadrupole (iq)	$\sum_i^{N_a} q_i \sum_j^{N_b} q_j (z_j^2 - {}^1/_2 y_j^2 - {}^1/_2 x_j^2) / R^3$ $+\sum_j^{N_b} q_j \sum_i^{N_a} q_i (z_i^2 - {}^1/_2 y_i^2 - {}^1/_2 x_i^2) / R^3$

[a] Note that the multipole expressions are valid for molecules connected along the z-axis only (see the Appendix) and that we have omitted the $(4\pi\varepsilon_0\varepsilon_r)^{-1}$ prefactor.

energy of L to be -4 kT, or that it is 50 times more likely to find lysozyme near the negative end of L than at the positive end at 17 mM salt. At 100 mM salt, the alignment free energy is reduced to -2 kT (not shown) which is still a significant number.

Thus, even though solvent mediated electrostatic interactions may be screened at physiological conditions and consequently contribute little to the angularly averaged free energy, they still provide effective guidance for steering proteins into specific orientations. In general terms, the electric field of a charged molecule—here lysozyme—aligns another dipolar protein (H or L), thereby creating a distance dependent interaction "funnel". This mechanism can be qualitatively understood from a few intrinsic protein properties, such as the dipole moment and net charge. That is, projecting the complex charge distribution within a protein onto a multipole one is indeed able to qualitatively predict the essential effects of certain protein−protein interactions. We also note that the angularly averaged potential of mean force essentially conceals the role played by electrostatic interactions and, instead, considerably more insight is gained by studying dipolar correlations as presented here.

Energy Decomposition. In order to decipher which interaction mechanisms control the steering of lysozyme toward H, L, and HL, it proves useful to decompose the energetics of these systems. During the simulation, we average the electrostatic energy of the whole system, the interaction energy between the proteins, as well as interprotein ion−ion, ion−dipole, and ion−quadrupole energies; see Table 2. The protein−protein (pp) energy accounts for the direct interaction between all charged groups on the two proteins, and as seen in Figure 6, minima in U_{pp} are observed for all complexes. Summing all possible electric multipolar terms should, ideally, add up to give U_{pp}, but we here restrict our analysis to include only ion−ion, ion−dipole, and ion−quadrupole terms. Note that in all cases the dipole−dipole term is negligible. For the H and L complexes, the sudden drop in electrostatic energy at separations corresponding to the free energy minima is caused by orientational correlations of the dipole. This is seen by the fact that most of U_{pp} is picked up by the ion−dipole and ion−ion terms where the latter is monotonically decreasing and causes merely the curve to slope. The ion−quadrupole term generally contributes only to a minor extent, except for HL where it dominates the alignment. That is, the ion−dipole term is negligible and the minimum is caused solely by the quadrupole. This is in agreement with our previous discussion and is also evident from Figure 3. Thus, decomposing the electrostatic pair interaction into a few of the first multipolar terms, we can understand the underlying driving force of the system. This seems possible even though the interprotein separation is short, compared with the size of the proteins and at conditions where the screening length is shorter than the length scale of the proteins.

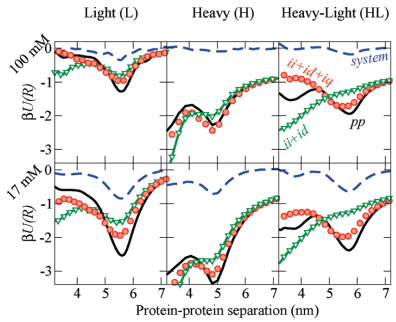


Figure 6. Decomposition of the electrostatic components of the thermally averaged interaction energy between lysozyme and H, L, and HL. Dashed lines: System electrostatic energy including proteins and salt particles (shifted to zero). Triangles: Ion−ion (ii), ion−dipole (id) energy. Circles: Ion−ion, ion−dipole, and ion−quadrupole (iq) energy. Full drawn lines: Protein−protein (pp) electrostatic energy calculated by explicitly summing over all charged groups. The energies represent ensemble averages over all protein angles and salt positions.

The vdW/electrostatic co-operativity is manifested also in the protein−protein energy components. Figure 7 shows the average protein−protein electrostatic and vdW energy, respectively, when either vdW or electrostatic interactions are switched on or off. With vdW interactions included, the electrostatic contribution is significantly enhanced and *vice versa*. The strength of the amplification of either interaction type is around $1-2$ kT.

At high ionic strength, we note that the total electrostatic energy (U_{sys})—defined as the sum of *all* interparticle Coulomb interactions, including salt—hardly varies with the protein−protein separation. At the same time, the direct electrostatic interaction between the two proteins, U_{pp}, varies quite a lot. That is, the electrostatic interactions between the two proteins are effectively screened out by the salt. However, even if the total electrostatic interactions only give a negligible contribution to the overall interaction energy, they still have a profound effect on the orientational arrangement of the complex. At low salt concentration and weak screening, we do see an effect of the total electrostatic interactions, but the vdW term is dominating the magnitude of the free energy minimum.

Biological Significance. In the experimentally determined crystal structure of the complex between lysozyme and the combined HL complex,[37] lysozyme is located in one end of the elongated antibody. Our study indeed suggests that there is a steering mechanism of lysozyme toward the poles of HL, as illustrated in Figure 3, but we see no large preference for one

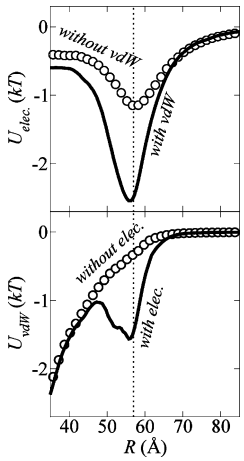


Figure 7. Average protein−protein energies for the lysozyme−L complex at 17 mM salt as a function of mass center separation, R. Top: Electrostatic energy, U_{pp}, when vdW interactions are switched on or off. Bottom: vdW energy when electrostatics are switched on or off. The vertical dashed line indicates the separation at the free energy minimum.

end over the other. Obviously, the crystal environment causes a preference for one end, which could be due to short-ranged specific interactions in the binding site[15−18] not included in the present model. The strength of the presented model is to capture the solvent mediated interactions taking place at intermediate and long-range protein separations. While the steering of lysozyme toward the combined HL antibody is moderate, the interaction with the individual fragments is stronger. We can only speculate if this has any biological relevance, but it is interesting to note that the dipole moment of both H and L peaks close to physiological pH. In any case, the present study demonstrates that significant biomolecular organization can be induced by cooperative electrostatic and van der Waals interactions.

Conclusion

We have studied the free energy of interaction as well as orientational steering mechanisms between mesoscopic protein pairs in aqueous salt solutions. In the case where one protein is highly charged and the other approximately neutral, but with an uneven distribution of surface charges, significant molecular alignment takes place. We demonstrate that the alignment mechanism in some cases can be seen as an ion−dipole or ion−quadrupole interaction. Electrostatic and van der Waals interactions can couple to form a cooperative effect that leads to a strong amplification of molecular alignment at physiological salt conditions. The solvent mediated electrostatic contribution to the angularly averaged free energy or "potential of mean force" is generally in the order of kT or less. However, the alignment free energy or preference for specific binding sites can be several kT, even at physiological salt conditions. Hence, the angularly averaged potential of mean force essentially conceals the role played by electrostatic interactions and, instead, considerably more insight is gained by studying angular correlations.

Acknowledgment. For financial support, the authors would like to thank the Swedish Research Council, the Research School of Pharmaceutical Sciences (FLÄK), the Linnaeus Center of Excellence on Organizing Molecular Matter, Sweden, the European Molecular Biology Organization (EMBO), and the Royal Physiographic Society in Lund.

Appendix

Insight into the electrostatic interaction between two arbitrary charge distributions (ρ,N) and (ρ',N') can be gained with an electric multipole expansion. This well established procedure reveals the leading term and the asymptotic behavior of the interaction of the electrical moments and provides information on how the distributions are distorted from spherical symmetry. That is, the NN' pair interactions decompose into a series of terms with increasing distance dependency. If the coordinates of the two distributions are given relative to two points P and P', the distance between two charges i and j can always be expressed as $|\vec{r}_{ij}| = |\vec{R} - \vec{r}_i + \vec{r}_j|$; see Figure 8. Omitting trivial constants and denoting the magnitude of charges by q, the total interaction energy is given by

$$U = \sum_{i=1}^{N} \sum_{j=1}^{N'} \frac{q_i q_j}{r_{ij}} \tag{4}$$

The multipole expansion is a Taylor series of this sum with derivatives evaluated at $\vec{r}_i = \vec{r}_j = \vec{0}$

$$U(R) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{x_i \partial}{\partial x_i} + \frac{y_i \partial}{\partial y_i} + \frac{z_i \partial}{\partial z_i} + \frac{x_j \partial}{\partial x_j} + \frac{y_j \partial}{\partial y_j} + \frac{z_j \partial}{\partial z_j} \right)^n \sum_{i=1}^{N} \sum_{j=1}^{N'} \frac{q_i q_j}{r_{ij}} \tag{5}$$

Note that, while the total interaction is independent of the points of expansion, the electrical moments are not. The series converges given that

$$\left| \frac{(\vec{r}_j - \vec{r}_j) \cdot \vec{R}}{R^2} + \frac{|r_j - r_i|^2}{R^2} \right| < 1 \tag{6}$$

is satisfied[38] and the result can be collected into a set of terms, $U_{\alpha\beta}$, that refers to different electrical moments of ρ and ρ'. For example, the second order term in eq 5 contains the ion−quadrupole, dipole−dipole, and quadrupole−ion contributions to the energy. Note that, if the terms are arranged in a table as below, one can immediately identify terms of equal distance dependency as they appear on the same off-diagonal. The ion−ion term decays as R^{-1} and each successive off-diagonal as R^{-2}, R^{-3}, and so forth.

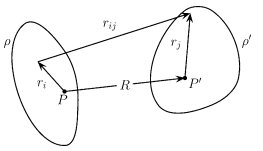


Figure 8. Schematics of two charge distributions and their metrics.

$$\begin{cases} U_{II} & U_{ID} & U_{IQ} & \ldots \\ U_{DI} & U_{DD} & \ldots & \\ U_{QI} & \ldots & & \\ \ldots & & & \end{cases} \tag{7}$$

Often, this is performed as a Laplace expansion, since the expressions become more compact in spherical polar coordinates, as compared to Cartesian. However, if the coordinate system is chosen such that the points of expansion lie on the z-axis, the expression for any $U_{\alpha\beta}$ simplifies significantly. For example, for the ion−quadrupole term, we get

$$U_{IQ} = \frac{1}{2}\sum_i^N q_i \sum_j^{N'} q_j \left[x_j^2 \frac{3R_x^2 - R^2}{R^5} + 6y_j z_j \frac{R_y R_z}{R^5} + y_j^2 \frac{3R_y^2 - R^2}{R^5} + 6z_j x_j \frac{R_y R_z}{R^5} + z_j^2 \frac{3R_z^2 - R^2}{R^5} + 6x_j y_j \frac{R_y R_z}{R^5} \right] \tag{8}$$

that simplifies to

$$U_{IQ} = \sum_i^N q_i \sum_j^{N'} q_j (z_j^2 - 1/2y_j^2 - 1/2x_j^2)/R^3) \tag{9}$$

with the particular choice $\overrightarrow{OP'} - \overrightarrow{OP} = (0,0,\pm R)$. Note that all even ($\alpha + \beta$ even) sets $U_{\alpha\beta}$ and $U_{\beta\alpha}$ are symmetric with respect to an interchange of i and j, while all odd sets are antisymmetric.

References and Notes

(1) Gibas, C. J.; Subramaniam, S.; McCammon, J. A.; Braden, B. C.; Poljak, R. J. *Biochemistry* **1997**, *36*, 15599–15614.
(2) Elcock, A. H.; McCammon, J. A. *Biophys. J.* **2001**, *80*, 613–625.
(3) Head-Gordon, T.; Brooks, C. L. *J. Phys. Chem.* **1987**, *91*, 3342–3349.
(4) Getzoff, E. D.; Cabelli, D. E.; Fisher, C. L.; Parge, H. E.; Viezzoli, M. S.; Banci, L.; Hallewell, R. A. *Nature* **1992**, *358*, 347–351.
(5) Asthagiri, D.; Neal, B.; Lenhoff, A. M. *Biophys. Chem.* **1999**, *78*, 219–231.
(6) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001**, *306*, 1139–1155.
(7) Carlsson, F.; Malmsten, M.; Linse, P. *J. Phys. Chem. B* **2001**, *105*, 12189–12195.
(8) Spaar, A.; Dammer, C.; Gabdoulline, R. R.; Wade, R. C.; Helms, V. *Biophys. J.* **2006**, *90*, 1913–1924.
(9) McGuffee, S. R.; Elcock, A. H. *J. Am. Chem. Soc.* **2006**, *128*, 12098–110.
(10) Allahyarov, E.; Löwen, H.; Hansen, J. P.; Louis, A. A. *Europhys. Lett.* **2002**, *57*, 731–737.
(11) Striolo, A.; Bratko, D.; Wu, J. Z.; Elvassore, N.; Blanch, H. W.; Prausnitz, J. M. *J. Chem. Phys.* **2002**, *116*, 7733–7743.
(12) Lund, M.; Jönsson, B. *Biophys. J.* **2003**, *85*, 2940–2947.
(13) Cuendet, M. A.; Michielin, O. *Biophys. J.* **2008**, *95*, 3575−3590.
(14) Pellicane, G.; Smith, G.; Sarkisov, L. *Phys. Rev. Lett.* **2008**, *101*, 248102–248105.
(15) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1089–1093.
(16) Novotny, J.; Bruccoleri, R. E.; Saul, F. A. *Biochemistry* **1989**, *28*, 4735–4749.
(17) Adachi, M.; Kurihara, Y.; Nojima, H.; Takeda-Shitaka, M.; Kamiya, K.; Umeyama, H. *Protein Sci.* **2003**, *12*, 2125–2131.
(18) Li, Y.; Urrutia, M.; Smith-Gill, S. J.; Mariuzza, R. A. *Biochemistry* **2003**, *42*, 11–22.
(19) Tan, R. C.; Truong, T. N.; McCammon, A. J.; Sussman, J. L. *Biochemistry* **1993**, *32*, 401–403.
(20) Song, X. *Mol. Simul.* **2003**, *29*, 643–647.
(21) Krushelnitsky, A. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2117–2128.
(22) Cohen, G.; Sheriff, S.; Davies, D. *Acta Crystallogr., Sect. D* **1996**, *52*, 315–326.
(23) Metropolis, N. A.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1097.
(24) Lund, M.; Jönsson, B. *Biochemistry* **2005**, *44*, 5722–5727.
(25) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications: New York, 1987.
(26) Israelachvili, J. *Intermolecular and Surface Forces*, 2nd ed.; Academic Press: London, 1991.
(27) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1999**, *291*, 149–162.
(28) Cerutti, D.; Ten-Eyck, L. F.; McCammon, J. A. *J. Chem. Theory Comput.* **2005**, *1*, 143–152.
(29) Lu, B.; Cheng, X.; Huang, J.; McCammon, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19314–9.
(30) Lund, M.; Jungwirth, P. *J. Phys.: Condens. Matter* **2008**, *20*, 494218–494221.
(31) Lund, M.; Trulsson, M.; Persson, B. *Source Code Biol. Med.* **2008**, *3*, No. 1.
(32) Humphrey, W. F.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 31–38.
(33) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Computational Science Series, Vol. 1; Academic Press: London, 2001.
(34) Bratko, D.; Striolo, A.; Wu, J. Z.; Blanch, H. W.; Prausnitz, J. M. *J. Phys. Chem. B* **2002**, *106*, 2714–2720.
(35) Kirkwood, J. G.; Shumaker, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **1952**, *38*, 855–862.
(36) Piñero, J.; Bhuiyan, L. B.; Bratko, D. *J. Chem. Phys.* **2004**, *120*, 11941–11947.
(37) Hibbits, K. A.; Gill, D. S.; Willson, R. C. *Biochemistry* **1994**, *33*, 3584–3590.
(38) Steiger, D.; Glaser, R. *J. Comput. Chem.* **2001**, *22*, 208–215.

JP904541G

Paper III

Persson and Lund
Phys. Chem. Chem. Phys. 11, 8879 (2009) ©.

View Online

PAPER www.rsc.org/pccp | Physical Chemistry Chemical Physics

Association and electrostatic steering of α-lactalbumin–lysozyme heterodimers†

Björn A. Persson and Mikael Lund*

Received 8th May 2009, Accepted 3rd July 2009
First published as an Advance Article on the web 3rd August 2009
DOI: 10.1039/b909179c

The salt and pH dependent association of hen egg white lysozyme with α-lactalbumin whey proteins has been studied using molecular level Monte Carlo simulations. A highly uneven charge distribution of α-lactalbumin leads to strongly ordered heterodimers that may facilitate the formation of structured, mesoscopic aggregates. This electrostatic steering gives rise to 80% alignment at 5 mM 1 : 1 salt which, due to screening, diminishes to 60% at 100 mM salt. The free energy of interaction minima, dominated by electrostatics, ranges between $-9\,kT$ at 1 mM salt to $-2\,kT$ at 100 mM (neutral pH). Calculated osmotic second virial cross coefficients indicate complexation in the pH interval 6–10. Multivalent ions are found to effectively destabilize the protein complex and, at constant ionic strength, the order is $La^{3+} > Ca^{2+} > Mg^{2+} > Na^{+}$. Upon binding of calcium to α-lactalbumin both the interaction and orientational alignment with lysozyme are reduced due to induced changes in the whey protein charge distribution. This potentially explains the experimentally observed absence of supramolecular structuring for the calcium loaded holo α-lactalbumin. Where available, good agreement is found with experimental data.

Downloaded by Lund University on 23 June 2011
Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C

Introduction

Whey proteins such as α-lactalbumin, β-lactoglobulin and serum albumin are by-products in the making of a variety of dairy products including cheese. These proteins have a high biological value, are relatively inexpensive and produced in vast amounts. Besides pure academic attention, there is also a considerable interest in using them for nutrition, as anti-bacterial agents and in technical applications.[1,2] Protein–protein interactions and protein assembly are governed by a range of intermolecular interactions that to a large extent can be controlled by changing solution conditions such as pH, temperature, solvent and salt concentration as well as type. Accounting for these effects in a theoretical framework requires a firm understanding of the underlying physical mechanisms and a number of computational approaches have been applied to capture protein–protein interactions in aqueous salt solutions. These include models based on Poisson–Boltzmann (PB) electrostatics,[3,4] explicit salt/implicit solvent simulations[5,6] and also more involved simulations where both proteins, salt and the solvent are treated in atomistic detail.[7,8] Since implicit solvent models describe the solvent as a *structureless* dielectric continuum, these inherently do not capture hydrophobic interactions. For the hydrophilic proteins studied here, this restriction is however not crucial; explicit solvent molecular dynamics (MD) simulations[7] are on a par with implicit solvent Monte Carlo (MC) simulation studies[5,6] that again correctly reproduce measured osmotic second virial coefficients. Common for the mentioned models is that they are rooted in statistical thermodynamics, *i.e* for a given system energy function or physical force field, they provide *ensemble averages* and *free* energies. This contrasts with common docking procedures that employ empirical relations or mere energy minimization schemes.

A number of experimental studies[9–11] have focused on the heterogeneous assembly of hen egg white lysozyme and various forms of α-lactalbumin (α-Lac). In particular, Nigen *et al.*[12] showed that these proteins self-assemble into supramolecular structures and that solution conditions such as ionic strength and temperature have a large impact on the protein assembly. Recently the same authors measured the association constants between lysozyme and α-Lac and demonstrated, in accord with other work, the presence of heterodimers.[1,13] A rich salt and temperature dependent phase behaviour has likewise been observed for heterogeneous mixtures of oppositely charged lysozyme variants.[14]

In this work we study, on a microscopic level, the association of lysozyme and α-Lac, focusing on electrostatic interactions. In particular, we explore a hitherto neglected aspect of the complexation process, namely the orientational ordering of the heterocomplex. This electrostatic "steering",[15–19] as we will show, is controlled by the uneven charge distribution on the whey proteins and may be an important mechanism for building supramolecular structures.

Methodology

Molecular simulations

All protein simulations are performed using Metropolis Monte Carlo simulations[21] in a closed, iso-thermal ensemble (*NVT*)—see details in Table 1. One *or* two proteins, counter-ions

Department of Theoretical Chemistry, Chemical Center, POB 124, S-221 00 Lund, Sweden. E-mail: mikael.lund@teokem.lu.se; Fax: +46 (0)46 222 4543; Tel: +46 (0)46 222 4501

† Electronic supplementary information (ESI) available: Links to view 3D visualisations of structures using FirstGlance. See DOI: 10.1039/b909179c

View Online

Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C
Downloaded by Lund University on 23 June 2011

and salt particles are immersed in a spherical simulation cell (see Fig. 1) containing a continuum solvent described by the dielectric constant of water, $\varepsilon_r = 80$. While the solvent extends to infinity beyond the cell boundary, ions and proteins are not allowed to penetrate the spherical cell wall. The mass centres of each protein are connected by the system z-axis, along which the two proteins are randomly translated during the simulation—salt particles are randomly displaced in any direction. All inter-protein angles are explored by randomly rotating the proteins around their mass centres. The proteins are constructed from their crystal structures by treating each amino acid residue as a single sphere ($\sigma \approx 7$ Å), located at the center of mass of that residue. Amino acid radii are defined by the molal volumes of the original, atomistic residues. This coarse grained representation has been shown to give almost identical results as a full, atomistic model because the surface topology and charge distribution of the protein are maintained.[6] During the simulation, charges on acidic or basic amino acid residues can fluctuate between the protein and the bulk solution. The trial energy for such an exchange is calculated according to:

$$\Delta U = \Delta U_{el} \pm (\mathrm{pH} - \mathrm{p}K_0)\cdot \ln 10 \qquad (1)$$

where $\mathrm{p}K_0$ is the intrinsic acid dissociation constant of the isolated amino acid residue and ΔU_{el} is the electrostatic energy change. Thus, the protein charge distribution is—as is also the case in a real, pH buffered solution—not fixed but will respond to changes in the environment.[22–24]

All particles in the system are treated as (charged) hard spheres of diameter σ_i and the system energy, U, is evaluated as a sum of all pair-wise interactions,

$$U = \sum_{i,j\in p,s}\left(\frac{e^2 z_i z_j}{4\pi\varepsilon_0\varepsilon_r} + u_{ij}^{hs}\right) - kT\sum_{i,j\in p}\frac{C_{vdW}}{r_{ij}^6} \quad i\neq j \qquad (2)$$

where p and s denotes the proteins and salt particles, respectively. The first term is the solvent screened Coulomb potential, u_{ij}^{hs} is the hard sphere contribution equal to $+\infty$ if $r_{ij} < (\sigma_i + \sigma_j)/2$, zero otherwise. The last term takes into account short ranged van der Waals interactions between amino-acids on the two proteins. The parameter C_{vdW} is set equal to 25 000 Å^6 which is appropriate for proteins in aqueous solutions.[6,25] The Monte Carlo moves described above are accepted if the total energy change, $\Delta U = U_{new} - U_{old}$, is smaller than zero. If $\Delta U > 0$ the move is accepted with the probability $\exp[-\Delta U/kT]$ as prescribed in the classic Metropolis algorithm.[21] All salt particles have a diameter of 4 Å and can be either monovalent, divalent or trivalent. The distinction between Mg^{2+} and Ca^{2+} is incorporated implicitly *via* the ability of the latter to bind to α-Lac. That is, in solutions containing Mg^{2+} we use the apo structure of α-Lac (PDB entry 1F6R)† while for Ca^{2+} solutions we use the calcium loaded holo form (PDB entry 1F6S).† As detailed in the results section, the holo and apo forms have different structures as well as charge distributions which influence their interaction with lysozyme.

Table 1 Monte Carlo simulation details

Property	
T/K	298.15
Cell radius/Å	100–300
Number of proteins	1–2
Salt pairs	$\mathcal{O}(10^2)$
Configurations	$\mathcal{O}(10^9)$
Software[20]	http://faunus.sourceforge.net

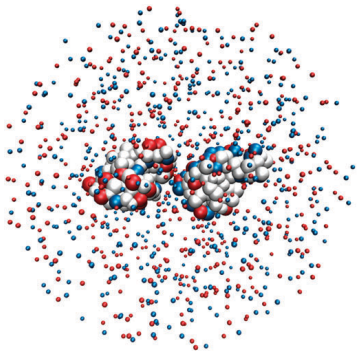


Fig. 1 Illustration of the model used for simulating two proteins in a salt solution. The solvent is treated as a dielectric continuum while salt particles and proteins are described by (clusters of) hard, charged spheres. For clarity the salt size has been reduced.

During the simulation, performed using the Faunus programming framework,[20] we sample the probability distribution, $g(R)$, of the protein mass-center separation, R, which is directly related to the free energy of interaction or potential of mean force *via* $w(R) = -kT \ln g(R)$. In cases where the protein–protein interaction is very attractive we use umbrella sampling to improve statistics at large protein separations. The microscopic potential of mean force can be condensed to a thermodynamic property, the osmotic second virial coefficient,[26] by integrating over the mass-center separation, R

$$B_2 = 2\pi\sigma_{hs}^3/3 - 2\pi\int_{\sigma_{hs}}^{\infty}(e^{-w(R)/kT} - 1)R^2\,\mathrm{d}R. \qquad (3)$$

Here the first term is the hard-sphere contribution, B_2^{hs}, where σ_{hs} is set to the minimum observed mass-center separation between the two proteins (~25.5 Å). The virial coefficient provides information about the overall two-body interaction: if positive the proteins repel, while if negative they attract each other. Moreover, it can be obtained experimentally by scattering techniques[27] or chromatography[28,29] and is thus a useful property for connecting theory and measurements.

Dielectric response of the proteins

The presented model assumes a high and *uniform* dielectric response throughout the medium, including the proteins, and thus neglects electric polarization effects from the non-polar protein interior. However, for solvated surface charges on single protein molecules it has previously been shown[30–32] that this approximation is very reasonable in that the effect of the dielectric boundary is small—typically plus or minus a few

 This journal is © the Owner Societies 2009

View Online

Downloaded by Lund University on 23 June 2011
Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C

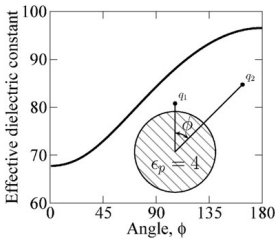


Fig. 2 Two solvated charges, q_1 and q_2, respectively located one and ten Ångströms away from the surface of a low dielectric, spherical cavity of radius ten Ångströms. The effective dielectric constant is calculated by matching the Coulomb energy, $q_1 q_2 / \epsilon_r^{\text{eff}} r_{12}$, to the exact interaction energy obtained from the numerical solution to the electrostatic problem[32] at different angles, ϕ.

tenths of a kT.[32] This can be illustrated by solving the electrostatic problem of a low dielectric sphere in a high dielectric medium—see Fig. 2. The effective dielectric constant oscillates around the bulk value for different charge constellations which, partly, explains why a uniform, high dielectric constant is a good approximation. Perhaps counter-intuitively, the effective solvent screening can be *higher* than in bulk water even when a low dielectric cavity is in between two charges (ϕ = 180, Fig. 2). This is due to a subtle balance between attractive and repulsive induced ion-dipole interactions,[32] demonstrating that a low dielectric interior does not always enhance electrostatic interactions. In absolute energies the polarization effect is, however, small and even more so at finite salt concentrations. Given these considerations and the fact that the proteins are described on a mesoscopic level, a uniform dielectric model seems quite appropriate.

Results and discussion

Properties of the studied proteins

We study the pair interactions between lysozyme and two variants of α-Lac: the apo form and the holo form with one calcium ion bound. The properties of the proteins are given in Table 2 as well as in Fig. 3 and, as can be seen, at neutral pH lysozyme is positively charged, while both forms of α-Lac are negatively charged. Since the proteins have similar sizes we can directly compare their molecular dipolar moments, which for α-Lac is much larger than for lysozyme over a broad pH range.

Effect of salt concentration and pH

First, we investigate the free energy of interaction, $w(R)$, between lysozyme and α-Lac in salt solutions at pH 7.5—see

Table 2 Number of residues (N_{res}), total charges (Z), dipole moments (μ) and PDB† entries for the studied proteins (pH 7.5)

	N_{res}	Z/e	μ/D	PDB
Lysozyme	129	6.8	139	4LZT
Holo α-lactalbumin (Ca^{2+})	122	−4.6	331	1F6S
Apo α-lactalbumin	123	−6.0	403	1F6R

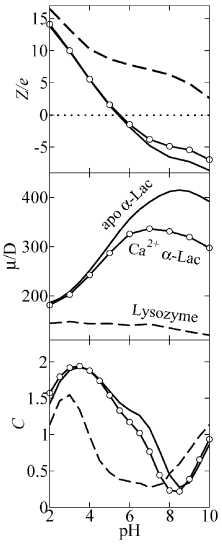


Fig. 3 Calculated electrostatic properties of the studied proteins in 5 mM 1 : 1 salt solutions. Top: average molecular net-charge, Z. Middle: molecular dipole moment, μ, evaluated around the protein center of mass. Bottom: charge capacitance, $C \propto \partial Z/\partial \text{pH}$ which reflects the proteins' propensity for charge regulation.[22,24]

Fig. 4. Due to the high opposite charges on the two proteins we see a strong attraction for all protein combinations. The calcium-free apo form exhibits the strongest interactions due to its large negative charge and dipolar moment (see Table 2). With calcium bound the α-lactalbumin charge increases while at the same time the dipole moment decreases by 18%. This leads to the least attractive combination for all salt concentrations. As expected, increasing the 1 : 1 salt concentration the electrostatic interactions are screened and weaker protein–protein association is observed. Still, at 100 mM salt we observe a significant free energy minimum of $-2\,kT$ (-5 kJ mol^{-1}) and we thus anticipate pairs to form even at intermediate ionic strengths. Increasing the salt concentration from 39 mM to 124 mM causes the experimental dissociation constant to increase by one order of magnitude.[13] This corresponds to a free energy difference of roughly 2 kT which is in good agreement with simulated free energy minima differences obtained from Fig. 4.

The solution pH plays a crucial role for the protein interactions as it determines the charge distribution of the participating macromolecules—see Fig. 3. To illustrate this we calculate the potential of mean force at various pH and condense the results into osmotic second virial coefficients according to eqn (3). As shown in Fig. 5 the attraction ($B_2 < 0$) between lysozyme and α-Lac is largest around neutral pH where both proteins have large, opposite net-charges. As pH is either decreased or increased the proteins go towards their fully protonated or de-protonated states and

View Online

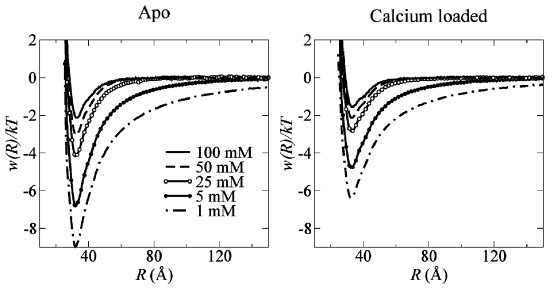


Fig. 4 Potential of mean force, $w(r)$, for lysozyme with apo-lactalbumin (left) and calcium loaded holo α-lactalbumin (right) at various 1 : 1 salt concentrations and at pH 7.5.

Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C
Downloaded by Lund University on 23 June 2011

hence start to repel each other ($B_2 > 0$). As can be expected from our discussion above, when calcium is bound the cross virial coefficient increases. The effect on B_2 is particularly large since long-ranged, attractive interactions give a relative large contribution to B_2.

Orientational alignment

It is worthwhile to look closer at the spatial charge distribution or "patchiness" of the studied proteins. The apo form of α-Lac has an impressive electric dipole moment of 400 Debye (D), almost three times larger than that of lysozyme (Table 2). From this basic observation we anticipate that the high charge on lysozyme may align α-Lac upon binding—*i.e.* that the complexation process has the characteristics of an ion–dipole interaction.[33] To follow the alignment, we sample the ensemble average of the molecular dipole moment z-component, $\langle \mu_z \rangle$, which provides direct information about the ordering. In our simulations, the system z-axis connects the two protein mass-centers (see the Methodology section) meaning that if the average dipole moment z-component is $+1$ the dipole is fully aligned, while if -1 it is fully anti-aligned. Fig. 6 illustrates the orientational alignment of α-Lac as it approaches lysozyme. In the free energy minimum, α-Lac is firmly aligned by 80% at 5 mM 1 : 1 salt concentration. This alignment of the whey-protein is observed in both the apo and holo form, with the latter to a lesser extent as would be predicted due to weaker

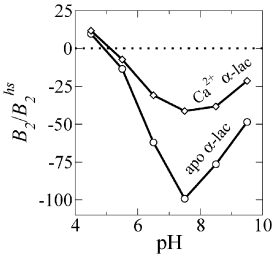


Fig. 5 Osmotic second virial cross coefficients, B_2/B_2^{hs}, for lysozyme interacting with apo α-Lac (circles) and calcium loaded holo α-Lac (diamonds) at various pH and at 5 mM 1 : 1 salt.

interactions. Upon addition of salt, the electrostatic interactions are screened and the range of alignment decreases. Still, even at 100 mM salt we see an appreciable alignment of 60% and that random orientations are not observed until the proteins are separated by 60 Å or more. Thus, lysozyme and α-Lac seem to form both *well-defined* and stable dimers. In this study we investigate the interaction between two proteins only, and can therefore not predict the formation of higher order multimers. We will pursue this issue in a later study.

We now investigate the alignment when the proteins are close together at the free energy minimum and in Fig. 7 we show the z-component probability, $P(\mu_z)dz$. Both the apo form and the holo form (not shown) of α-Lac show narrow distributions with $P(\mu_z)$ maxima close to $\mu_z = +1$. This implies that the molecular dipole, when in contact with lysozyme, is almost fully aligned. Lysozyme on the other hand has $P(\mu_z)$ maxima closer to $+0.5$, indicating that this molecule is rotating more freely. These observations are to be expected from the large differences in dipolar moments between lysozyme and the α-Lac variants. We can convert the probability distribution, $P(\mu_z)$, into a "free energy of alignment",

$$\Delta A(\mu_1 \rightarrow \mu_2) = -kT \ln[P(\mu_2)/P(\mu_1)] \qquad (4)$$

That is, we can calculate the free energy difference between having lysozyme at, say, the negative pole and at the positive pole of α-Lac. For the apo form, ΔA for this configuration change amounts to as much as $5\,kT$ (12 kJ mol^{-1})—see inset of Fig. 7. In other words: it is $e^5 = 150$ times more likely to find lysozyme near the negative end of α-Lac than at the positive end.

Describing the intricate charge pattern of proteins *via* their molecular dipole moments is of course an approximation. In order to test this representation we now follow the inter-protein alignment in a more rigorous manner: during our MC simulations we sample the probability of finding one protein in small cubic volume elements around the other—*i.e.* we calculate the *spatial* distribution function, $g(\mathbf{r})$. Fig. 8 shows iso-density plots of α-Lac and lysozyme, respectively. Clearly, lysozyme binds preferentially to one end of α-Lac which roughly coincides with the negative pole of the molecular dipole moment. A similar picture is obtained for the apo form. Thus, as a first approximation the protein dipole moment

 This journal is © the Owner Societies 2009

View Online

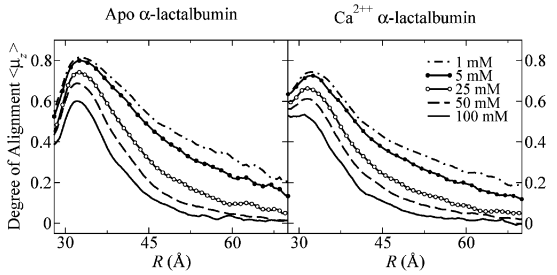


Fig. 6 Degree of dipole alignment of α-Lac as a function of the protein–protein mass-center separation. Plotted for the apo (left) and holo (right) form at various 1 : 1 salt concentration and pH 7.5. $\langle \mu_z \rangle = 1.0$ means that the molecular dipole is fully aligned with respect to lysozyme, while zero means that it is, on average, freely rotating.

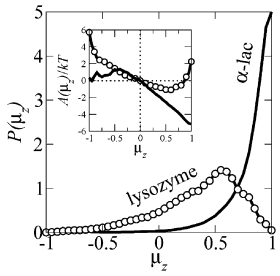


Fig. 7 Probability distribution of the protein dipole moments in the free energy minimum (R = 32 Å). +1 means that the dipole is fully aligned, while if −1 it is anti-aligned. The inset shows the corresponding alignment free energy, $A(\mu_z) = -kT \ln P(\mu_z)$, chosen to zero at $\mu_z = 0$. Calculated for the interaction between lysozyme and apo α-Lac at 5 mM salt.

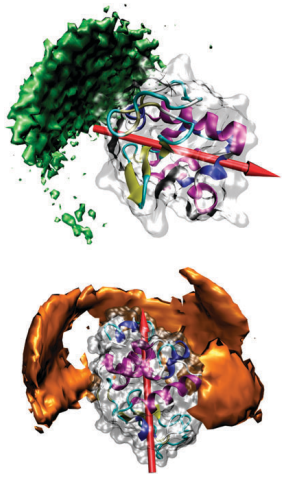


Fig. 8 Top: lysozyme mass-center distribution (green iso-surface) around calcium loaded holo α-lactalbumin (transparent) at pH 7.5 and 5 mM 1 : 1 salt. The red arrow shows the direction of the α-Lac molecular dipole moment. Bottom: the reverse situation with lysozyme in the center and α-Lac around it (orange iso-surface). Graphics produced using VMD[34] and Faunus.[20]

seems to be an excellent indicator for the orientational alignment upon association with (charged) proteins. In contrast, the distribution of α-Lac around the egg-shaped lysozyme is more diffuse but with a preference for the equator. This can be attributed to the fact that (1) the dipole moment of lysozyme is relatively small and (2) an equatorial approach gives the minimum mass-center separation which maximizes the ion–ion, ion–dipole and van der Waals attractions between lysozyme and α-Lac.

Note that the presented iso-density surface is directly related to the iso-free energy surface *via* the relation $A(\mathbf{r}) = -kT \ln g(\mathbf{r})$. The results shown in Fig. 8 then imply that lysozyme can—with the same free energy—bind to a range of different positions on α-Lac. This is merely a consequence of the entropy present in a liquid aqueous solution. In a crystal structure this entropic contribution is of course not present and hence one observes only one, unique inter-protein configuration. While early "best visual fit" modelling of lysozyme and whey protein association led to the conclusion that no specific interaction exist between lysozyme and α-Lac,[9] the present study underpins that weak interactions can indeed lead to specificity, although with a somewhat wide distribution of possible binding sites.

Multivalent ions

We have already considered the effect of binding calcium to α-Lac but so far we have presented results for mono-valent 1 : 1 salt only. In our MC simulations we treat ions explicitly which allows us to correctly describe the effect of multivalent ions. This is a possibility generally not present in models based on the linearized Poisson–Boltzmann approximation—Debye–Hückel theory, for example—which neglect ion–ion correlations, important in solutions of strongly interacting electrolytes.[35] It has been shown experimentally[12] that even small amounts of $CaCl_2$ destabilize aggregates of lysozyme

Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C
Downloaded by Lund University on 23 June 2011

View Online

Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C
Downloaded by Lund University on 23 June 2011

Table 3 Free energy minima, $w(R)_{min}$, for the interaction between lysozyme and apo α-Lac in various salt solutions of identical ionic strength, I. NaCl = 1 : 1, $MgCl_2$ = 2 : 1, $LaCl_3$ = 3 : 1. Note that $CaCl_2$ = 2 : 1, simulated with calcium loaded holo α-Lac

	I/mM	C_s/mM	$w(R)_{min}/kT$
NaCl	5	5	−6.7
$MgCl_2$	5	1.6	−5.9
$CaCl_2$	5	1.6	−4.5
$LaCl_3$	5	0.8	−4.0

and α-Lac, while $MgCl_2$ does so too, but to a lesser extent. The binding constant of calcium to α-Lac is five orders of magnitude larger than that for magnesium[36] and—as shown in the previous section—the net charge and dipole moment of α-Lac change significantly in the presence of calcium. Besides ion-binding the salt screening effect of high valency ions is stronger than for mono-valent ions. That is, less 2 : 1 than 1 : 1 salt is needed to reach the same ionic strength, $I = \frac{1}{2}\Sigma z_i^2 c_i$. Instead of presenting the full potential of mean force, Table 3 shows merely the simulated free energy minima for the interaction between lysozyme and apo α-Lac in solutions of 1 : 1, 2 : 1 and 3 : 1 salts. For a fixed ionic strength, I = 5 mM, high valency cations very effectively reduce the electrostatic attraction between the proteins. Trivalent ions are found to destabilize more effectively than calcium, even though we have not considered specific M^{3+} binding to the protein. The findings presented in Table 3 are in excellent agreement with experiment where the destabilising order is found to be $Ca^{2+} > Mg^{2+} > Na^{+}$.[12] Our results suggest that trivalent cations such as La(III) and Fe(III) will have even stronger effects than calcium at equal ionic strengths, but this still needs to be verified experimentally. The fact that different salts with identical ionic strengths screen differently is, partly, a manifestation of the strong ion–ion correlations present. It is also immediately apparent that this effect cannot be captured with classical Debye–Hückel theory where the salt effect is described solely by the Debye screening length (related to the ionic strength).

Conclusion

Using molecular simulations we have provided a microscopic description of the salt dependent complexation of lysozyme and two variants of α-lactalbumin whey proteins. Due to the large and opposite charges at neutral pH the proteins attract each other with interaction free energies ranging between −1 to −9 kT. Although monovalent salts, in quantitative accord with experiment,[13] screen the attractive electrostatic interactions, free energy minima are observed even at 100 mM salt. Calculated second virial cross coefficients indicate that attractive interactions dominate in the pH range 6–10 while at lower or higher pH the proteins dissociate due to electrostatic repulsion.

The uneven charge distribution of α-Lac leads to a strong orientational alignment when cationic lysozyme molecules approach. The process resembles the interaction between an ion and a dipole and lysozyme has a clear preference for the negative end of α-Lac which coincides with the negative pole of the molecular dipole moment. The degree of alignment in the free energy minima ranges between 50–80% for 1 : 1 salt concentrations (1 to 100 mM). When apo α-Lac binds calcium the dipole moment decreases from 400 to 330 Debye. This, in combination with a loss of negative charge, de-stabilizes the heterocomplex with lysozyme as is also observed experimentally.[12]

Multi-valent ions have a strong de-stabilizing effect on the protein complexes and we show, at constant ionic strength, that the cation order is $La^{3+} > Ca^{2+} > Mg^{2+} > Na^{+}$. This ordering arises due to a combination of ion–ion correlations and changes in the protein charge distribution upon binding of calcium. While the tri-valent case has not yet been investigated experimentally, the remainder of the series is in excellent agreement with measurements.

In summary we have shown—on a microscopic level—that lysozyme and α-Lac interact strongly and in *specific* constellations due to a highly uneven charge distribution on the latter. Over a broad range of pH and salt conditions, the protein complex is stable as well as well-defined and it is plausible that the observed orientational correlations may facilitate molecular ordering, involving multiple proteins.

Acknowledgements

For financial support the authors would like to thank the Swedish Research Council, the Royal Physiographic Society in Lund and the Linnaeus Center of Excellence on Organizing Molecular Matter, Lund University.

References

1 H. R. Ibrahim, N. Taniyama and T. Aoki, *Lett. Drug Des. Discovery*, 2004, **1**, 101–109.
2 I. Lopezexposito and I. Recio, *Int. Dairy J.*, 2006, **16**, 1294–1305.
3 A. H. Elcock, R. R. Gabdoulline, R. C. Wade and J. A. McCammon, *J. Mol. Biol.*, 1999, **291**, 149–162.
4 F. Carlsson, M. Malmsten and P. Linse, *J. Phys. Chem.*, 2001, **105**, 12189–12195.
5 E. Allahyarov, H. Löwen, J. P. Hansen and A. A. Louis, *Europhys. Lett.*, 2002, **57**, 731–737.
6 M. Lund and B. Jönsson, *Biophys. J.*, 2003, **85**, 2940–2947.
7 G. Pellicane, G. Smith and L. Sarkisov, *Phys. Rev. Lett.*, 2008, **101**, 248102–248105.
8 M. A. Cuendet and O. Michielin, *Biophys. J.*, 2008, **95**, 3575–3590.
9 N. K. Howell, N. A. Yeboah and D. F. V. Lewis, *Int. J. Food Sci. Technol.*, 1995, **30**, 813–824.
10 N. Howell and E. Li-Chan, *Int. J. Food Sci. Technol.*, 1996, **31**, 439–451.
11 M. Nigen, T. Croguennec, D. Renard and S. Bouhallab, *Biochemistry*, 2007, **46**, 1248–1255.
12 M. Nigen, T. Croguennec and S. Bouhallab, *Food Hydrocoll.*, 2009, **23**, 510–518.
13 M. Nigen, V. Tilly, T. Croguennec, D. Drouinkucma and S. Bouhallab, *Biochim. Biophys. Acta, Proteins Proteomics*, 2009, **1794**, 709–715.
14 P. M. Biesheuvel, S. Lindhoud, R. de Vries and M. A. Cohen-Stuart, *Langmuir*, 2006, **22**, 1291–1300.
15 R. C. Tan, T. N. Truong, A. J. McCammon and J. L. Sussman, *Biochemistry*, 1993, **32**, 401–403.
16 R. E. Kozack, M. J. D'mello and S. Subramaniam, *Biophys. J.*, 1995, **68**, 807–814.
17 A. Spaar, C. Dammer, R. R. Gabdoulline, R. C. Wade and V. Helms, *Biophys. J.*, 2006, **90**, 1913–1924.
18 A. Krushelnitsky, *Phys. Chem. Chem. Phys.*, 2006, **8**, 2117–2128.

 This journal is © the Owner Societies 2009

View Online

19 B. Persson, B. Jönsson and M. Lund, *J. Phys. Chem. B*, 2009, **113**(30), 10459–10464.
20 M. Lund, M. Trulsson and B. Persson, *Source Code Biol. Med.*, 2008, **3**, DOI: 10.1186/1751-0473-3-1, No. 1.
21 N. A. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1097.
22 J. Kirkwood and J. Shumaker, *Proc. Natl. Acad. Sci. U. S. A.*, 1952, **38**, 863–71.
23 A. H. Elcock and J. A. McCammon, *Biophys. J.*, 2001, **80**, 613–625.
24 M. Lund and B. Jönsson, *Biochemistry*, 2005, **44**, 5722–5727.
25 J. N. Israelachvili, *Intermolecular and Surface Forces, Second Edition: With Applications to Colloidal and Biological Systems (Colloid Science)*, Academic Press, 1992.
26 T. L. Hill, *An Introduction to Statistical Thermodynamics*, Dover Publications, 1987.
27 O. D. Velev, E. W. Kaler and A. M. Lenhoff, *Biophys. J.*, 1998, **75**, 2682–2697.
28 J. Bloustine, V. Berejnov and S. Fraden, *Biophys. J.*, 2003, **85**, 2619–2623.
29 P. M. Tessier, H. R. Johnson, R. Pazhianur, B. W. Berger, J. L. Prentice, B. J. Bahnson, S. I. Sandler and A. M. Lenhoff, *Proteins: Struct., Funct., Genet.*, 2003, **50**, 303–311.
30 A. Warshel, S. T. Russel and A. K. Churg, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 4785–4789.
31 R. Penfold, J. Warwicker and B. Jönsson, *J. Phys. Chem. B*, 1998, **102**, 8599–8610.
32 M. Lund, B. Jönsson and C. E. Woodward, *J. Chem. Phys.*, 2007, **126**, 225103–225110.
33 D. Bratko, A. Striolo, J. Z. Wu, H. W. Blanch and J. M. Prausnitz, *J. Phys. Chem. B*, 2002, **106**, 2714–2720.
34 W. F. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 31–38.
35 B. Jönsson and H. Wennerström, *Electrostatic Effects in Soft Matter and Biophysics*, Dordrecht, Netherlands, 2001.
36 E. Permyakov, *FEBS Lett.*, 2000, **473**, 269–274.

Published on 03 August 2009 on http://pubs.rsc.org | doi:10.1039/B909179C
Downloaded by Lund University on 23 June 2011

Paper IV

Björn A Persson, Mikael Lund, Jan Forsman, Dereck E.W. Chatterton and Torbjörn Åkesson
Biophys. Chem. 11 187 (2010)

Biophysical Chemistry 151 (2010) 187–189

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Molecular evidence of stereo-specific lactoferrin dimers in solution

Björn A Persson [a,*], Mikael Lund [a], Jan Forsman [a], Dereck E.W. Chatterton [b], Torbjörn Åkesson [a]

[a] Department of Theoretical Chemistry, Lund University, POB 124, 210 00 Lund, Sweden
[b] Dairy Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

ARTICLE INFO

Article history:
Received 4 May 2010
Received in revised form 15 June 2010
Accepted 18 June 2010
Available online 27 June 2010

Keywords:
Protein–protein interaction
Lactoferrin
Self-assembly
Stereospecific

ABSTRACT

Gathering experimental evidence suggests that bovine as well as human lactoferrin self-associate in aqueous solution. Still, a molecular level explanation is unavailable. Using force field based molecular modeling of the protein–protein interaction free energy we demonstrate (1) that lactoferrin forms highly stereo-specific dimers at neutral pH and (2) that the self-association is driven by a high charge complementarity across the contact surface of the proteins. Our theoretical predictions of dimer formation are verified by electrophoretic mobility and N-terminal sequence analysis on bovine lactoferrin.

© 2010 Elsevier B.V. All rights reserved.

1. Main

This combined theoretical and experimental study of the pair interaction of bovine lactoferrin (LF) shows that the protein self-assembles into dimers. Free energy calculations reveal a well defined, stereo-specific complex that is stabilized by a region of complementary amino acid residues. Historically, LF was first isolated from bovine milk but the glycoprotein is found in all mammals. It is secreted in various exocrine fluids such as bile, saliva, pancreatic fluid and tears [1]. Inflammatory reactions and viral infections are often followed by increased plasma concentrations of LF and the protein is considered part of the acute-phase immune response. More specifically, LF has antibacterial, antiviral and anti-inflammatory activity, underlining that the protein is an important part of the immune system [2]. LF is also interesting from a technical perspective, for instance, cationic protein-stabilized emulsions have been prepared of the protein and LF is known to be a fouling component of contact lenses [3,4]. LF has also received attention as a nutritional additive in milk formulas [5]. The protein consists of two major lobes and approximately 700 amino acids. The macromolecule has the shape of a dumbbell and is well described by a bi-axial ellipsoid with half-axis of 47 Å and 26 Å. Several experimental studies imply that the protein self-assembles: dimers of bovine LF have been misstaken for IgG2, that has a molecular weight twice of the LF monomer and sedimentation equilibrium has revealed a concentration-dependent average molecular weight [6,7]. Further, LF is found to elute with molecular weights corresponding to monomers, dimers and trimers, according to chromatographic analysis of bovine milk [8,9]. Scattering studies have also revealed large aggregates of human LF and that these dissolve at high ionic strength [10]. These observations all suggest that LF is able to self-assemble into larger structures, although no molecular level explanation has been presented. Using Metropolis Monte Carlo simulations we investigate the angularly averaged free energy of interaction between two LF molecules in an aqueous salt solution using the Faunus open source software package [11,12]. This software package has been applied with success for computation of virial coefficients of lysozyme and binding constants of lysozyme and *apo*-α-lactalbumin [13,14]. The simulation procedure is described in detail in the Appendix as well as elsewhere and we shall give here only a brief overview [13,15]: each protein is treated as a rigid body consisting of spherical amino acid beads that may be either charged or neutral, depending on the type and protonation state; PDB entry 1BLF was used to generate the coarse grained *holo*-Lactoferrin structure. A rigid model is reasonable since the structure of the *holo*-form is considered stiff and is resistant to proteolytic degradation[1,16]. The solvent is treated as a dielectric continuum and particles interact via a combined Lennard–Jones (LJ) and salt screened Debye–Hückel potential, where the total energy for each configuration, U, is given by the sum of the amino acid pair interactions.

$$U = \sum_{i<j}^{N} \left[\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} e^{-\kappa r_{ij}} + 4\epsilon \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \right] \quad (1)$$

r_{ij} is the separation between amino acids, q is the charge of an amino acid and κ is the inverse Debye length, related to the ionic strength. ϵ_0 is the permittivity of vaccum and ϵ_r the dielectric constant of water,

* Corresponding author. Tel.: +46 46 222 8241; fax: +46 46 222 4413.
E-mail address: bjorn.persson@teokem.lu.se (B.A. Persson).
URL: http://www.teokem.lu.se/~bpe/ (B.A. Persson).

0301-4622/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2010.06.005

set to 80 in this work. σ_{ij} is the mean diameter of the amino acids and ϵ is the LJ interaction parameter (0.05 kT). k is Boltzmann's constant and T is the temperature. During the MC simulation, proteins are translated as well as rotated to explore configurations in the canonical ensemble at 298 K. The free energy of interaction (or potential of mean force, PMF) is obtained by sampling the protein–protein radial distribution function as a function of the mass center separation, R. In contrast to typical docking methods, the present approach correctly takes into account thermal motion and hence the entropy of the protein–protein association process. Fig. 1A shows the calculated lactoferrin–lactoferrin free energy of interaction at pH 5, 7 and 9 and at an ionic strength of 5 mM ($\kappa = 0.023$ Å^{-1}). At these pH values the LF net charge is $+23e$, $+14e$ and $+4e$. Titration simulations estimate pI to 9.4, slightly above the experimental range, 8–9 [17–19]. At pH 5 the interaction is purely repulsive due to a large net charge of the proteins. However, as pH is increased, a deep narrow minimum appears at approximately 47 Å. The attraction broadens at pH 9, close to isoelectric conditions, but the minimum is, interestingly, very distinct. We now demonstrate that the narrow free energy minimum is due to a few highly specific residue–residue interactions that lock the entire complex into a unique constellation. Firstly, by inspecting the MC generated configurations (see animation in Appendix) it is clear that when the proteins are in contact, only few orientations are observed. Secondly, we identify a close to orthogonal set of vectors, spanned from the proteins mass center to the mass centers of E385 and E66 and correlate the E385 angles with the E66 dihedrals (inset of Fig. 2). This allows us to estimate the orientational free energy and, as evident from Fig. 2, the two proteins are aligned in a strongly stereo-specific manner, dominated by a narrow set of protein–protein orientations. The choice of E385 was motivated by the fact that it is parallel to the line connecting the protein mass centers at the free energy minimum. The total interaction between the two proteins is influenced by both electrostatic and van der Waals interactions. Due to the non-zero molecular net charge, any attractive electrostatic contribution must be due to charge patches on the protein surface. To qualitatively investigate the effect of this patchiness, we identify a few amino acids with high inter-protein charge-complementarity. These are selected by inspecting the protein–protein contact surface of different configurations at the free energy minimum. Two such candidates are glutamate E535 and lysine K498. In Fig. 1B we present the PMF between mutated forms of LF as well as for the wildtype at pH 8, where the charge of the wildtype is $+11e$. Reverting the negative charge of E535 has dramatic effects on the potential of mean force: this single point mutation leads to a complete loss of the attractive free energy minimum. The mutation increases the net charge of the protein by two units and an increased repulsion can be expected. Conversely, the protein net charge is maintained by swapping charges

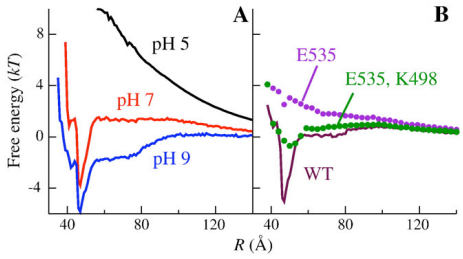


Fig. 1. Free energy of interaction (PMF) between lactoferrin molecules as a function of the mass center separation, R, at (A) different pH values for the wildtype, (B) different mutations and wildtype at pH 8. The ionic strength is 5 mM in all cases. Note, one kT equals 2.48 kJ/mol at 298 K. The maximum error is about 0.2 kT.

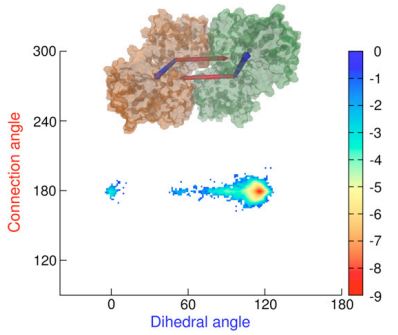


Fig. 2. Angular free energy (kT) of the connection angle (red vectors) and dihedral (blue vectors) at pH 8 and at 5 mM ionic strength. Sampling is performed within a 60 mass center separation. The inset shows the two proteins and the reference frames. The scale is relative to a free orientation.

of K498 with E535 — both of which are in the contact region. This also significantly diminishes the free energy minimum, underlining that the contact surface is highly tailored to facilitate a dimeric complex, assembled via complementary electrostatic interactions. Coordinates of a typical dimer configuration is supplied in the online material. Lastly, we provide experimental data to back our theoretical predictions. LF was separated using 12% Tris–Glycine SDS-PAGE under non-reducing conditions followed by staining with Coomassie blue R-250. Results indicate a primary band at 80 kDa, consistent with the molecular weight of the protein in its monomeric form. Two distinct bands, consistent with the molecular weight of dimeric and trimeric LF are also clearly visible at 160 kDa and 240 kDa (Fig. 3). To

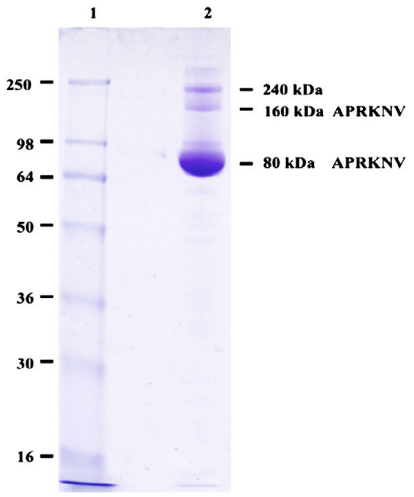


Fig. 3. SDS PAGE analysis of LF. Proteins were separated using 12% Tris–Glycine SDS PAGE under non-reducing conditions and pH 8.3 following by staining with Coomassie blue R-250. Lane 1: Molecular weight marker; Lane 2: LF. N-terminal sequence analysis of the electroblotted bands is indicated.

Table 1
Properties of coarse grained amino acids.

Particle	Radius (Å)	pKa
Ala	3.1	–
Arg	4.0	12.0
Asp	3.6	4.0
Gly	2.9	–
Glu	3.8	4.4
His	3.9	6.3
Lys	3.7	10.4

PVDF membranes (Invitrogen). The 80 kDa and 160 kDa electroblotted bands were excised and amino acid sequence analyses were performed on an Applied Biosystems PROCISE HT protein sequencer with on-line identification of phenylthiohydantoin-derivatives using standard programmes. Lactoferrin was identified in both bands.

While experimental data and the present simulations suggest the formation of LF dimer, it is desirable to further investigate special orientation of the aggregates by experimental scattering techniques. It would also be of importance to evaluate if the model is able to predict trimers. We intend to continue with such studies and to further scrutinize the driving force in the simulated system. Initial examination of more detailed models shows consistent results with this work.

Acknowledgment

We would like to express thanks to the Royal Physiographic Society in Lund, Organizing Molecular Matter (Lineaus Center of Excellence), and Esben S. Sørensen, University of Århus for the sequence analysis.

Appendix A

To coarse grain the protein we assume that each amino acid has a density of 0.9 g/ml and use the molecular weight to compute the radius. The sphere is centered at the mass center in the atomistic amino acid structure. It has been shown that this provides a good topological description of proteins and reproduces atomistic results [13]. To evaluate the average charge of each amino acid we perform Monte Carlo simulations with single proteins. In such calculations, each amino acid is allowed to titrate in a Markov chain, according to a Boltzmann weight that is modified by the pKa-value and pH of the solution. This provides us with an estimate of the charge distribution of the proteins that are being used as input when we evaluate the pair-interaction. A complete description of simulation details and parameters can be found in the reference [13] and [15]. The radii of some of the different amino acids, along with pKa-values used to generate average charges, are provided in Table 1.

Lactoferrin is here modeled in its *holo*-form. This is achieved by placing a unit charge at the site of each ferric ion in the original coordinates, approximating a ferric ion synergistically bound with a carbonate. The magnitude of the Lennard–Jones interaction parameter is obtained from a Hamaker argument. Since the densities of proteins are fairly constant one may estimate the contribution of the interactions across a water medium on an amino acid level. The magnitude in this work is chosen so it reproduces that of a previously published model [13]. Since the range of interaction depends on the size of the amino acid, large residues will effectively attract each other more than smaller which is reasonable. This does not capture any sort of specific hydrophobic interactions. What is being modeled is rather an averaged dispersion–hydrophobic interaction and the geometrical influences of the experimental structure. In the simulations we restrict the sampling to distances shorter than 200 Å and the radial distribution function is sampled with a bin width of 1 Å. Each PMF is generated with approximately 10^7 configurations, which was found sufficient to produce static results with low noise. The animation provided displays a representative trajectory of the Monte Carlo simulation at pH 8 and clearly shows how the two proteins bind in a specific orientation. Red colored particles represent negatively charged amino acids while blue colored are positively charged.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bpc.2010.06.005.

References

[1] L. Adlerova, A. Bartoskova, M. Faldyna, Lactoferrin: a review, Vetetinarni Midicinia 53 (2008) 457–468.
[2] J.H. Brock, The physiology of lactoferrin, Biochem. Cell Biol. 80 (2002) 1–6.
[3] A. Sarkar, K.K.T. Goh, H. Singh, Colloidal stability and interactions of milk-protein-stabilized emulsions in an artificial saliva, Food Hydrocolloids 23 (2009) 1270–1278.
[4] L. Meagher, H.J. Griesser, Interaction between adsorbed lactoferrin layers measured directly with the atomic force microscope, Colloids and Surfaces B 23 (2002).
[5] M. Tomita, H. Wakabayashi, Kouichirou, K. Yamauchi, T. Yaeshima, K. Iwatsuki, Twenty-five years of research on bovine lactoferrin applications, Biochimie 91 (2009) 52–57.
[6] F.L. Schanbacher, K.L. Smith, L.C. Ferguson, The similarity of bovine lactoferrin dimer to igg2, Fed. Proc. 30 (1971) 532.
[7] F.J. Castellino, W.W. Fish, K.G. Mann, Structural studies on bovine lactoferrin, J. Biol. Chem. 245 (1970) 4269–4275.
[8] R.J. Harmon, F.L. Schanbacher, L.C. Fergusson, K. Larry, Changes in lactoferrin, immunoglobulin g, bovine serum albumin and alpha-lactalbumin during acute experimental and natural coliform mastitis in cows, Infect. Immun. 13 (1976) 533–542.
[9] H. Wang, W.L. Hurley, Identification of lactoferrin complexes in bovine mammary secretions during mammary gland involution, J. Dairy Sci. 81 (1998) 1896–1903.
[10] S.E. Babina, F.V. Tuzikov, N.A. Tuzikova, V.N. Buneva, G.A. Nevinskii, Effect of nucleotides on the oligomeric state of human lactoferrin, Mol. Biol. 40 (2006) 121–131.
[11] N.A. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1097.
[12] M. Lund, M. Trulsson, B. Persson, Faunus: An object oriented framework for molecular simulation, Source Code Biol. Med. 3 (2008) 3:1.
[13] M. Lund, B. Jönsson, A mesoscopic model for protein–protein interactions in solution, Biophys. J. 85 (2003) 2940–2947.
[14] B.A. Persson, M. Lund, Association and electrostatic steering of alpha-lactalbumin–lysozyme heterodimers, Phys. Chem. Chem. Phys. 11 (2009) 8879–8885.
[15] M. Lund, B. Jönsson, On the charge regulation of proteins, Biochemistry 44 (2005) 5722–5727.
[16] E.N. Baker, H.M. Baker, A structural framework for understanding the multifunctional character of lactoferrin, Biochimie 91 (2009) 3–10.
[17] S. Belegrinou, I. Mannelli, P. Lisboa, F. Bertagnol, A. Valsesia, G. Ceccone, P. Colpo, H. Rauscher, F. Rossi, ph-dependent immobilization of proteins on surfaces functionalized by plasma-enhanced chemical vapor depostion of poly(acrylic acid)- and poly(ethylene oxide)-like films, Langmuir 24 (2008) 7251–7261.
[18] M.C. Wahlgren, T. Arnebrant, M.A. Paulsson, The adsorption from solutions of beta-lactoglobulin mixed with lactoferrin or lysozyme onot silica and methylated silica surfaces, J. Coll. Interface Sci. 158 (1993) 46–53.
[19] F. Superti, R. Siciliano, B. Rega, F. Giansanti, P. Valenti, G. Antonini, Involvement of bovine lactoferrin metal saturation, sialic acid and protein fragments in the inhibition of rotavirus infection, Biochim. Biophys. Acta 1528 (2001) 107–115.

Paper V

Vauthier, Persson, Lindner and Cabane
Biomater. 32, 1646-56 (2011) ©

Biomaterials 32 (2011) 1646–1656

Contents lists available at ScienceDirect

Biomaterials

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

Protein adsorption and complement activation for di-block copolymer nanoparticles

Christine Vauthier [a,b,*], Bjorn Persson [c], Peter Lindner [d], Bernard Cabane [e]

[a] Univ Paris Sud, Physico-chimie Pharmacotechnie Biopharmacie, UMR CNRS 8612, F-92296 Chatenay-Malabry, France
[b] CNRS, Chatenay-Malabry, F-92296, France
[c] Theoretical Chemistry, Chemical Center, Lund University, SE 221 00 Lund, Sweden
[d] ILL, BP 156, 38042 Grenoble Cedex, France
[e] PMMH, CNRS UMR 7636, ESPCI, 10 Rue Vauquelin, 75231 Paris Cedex 05, France

ARTICLE INFO

Article history:
Received 14 July 2010
Accepted 13 October 2010
Available online 18 November 2010

Keywords:
Core-diffuse shell nanoparticles
Mesh size
BSA
Fibrinogen
C3
Adsorption isotherm

ABSTRACT

Four types of nanoparticles with core-diffuse shell structures have been synthesized through self-assembly of PICBA-Dextran block copolymers. These nanoparticles are designed to carry pharmaceutically active molecules into the human body through injection into the blood stream. In this work, we have determined how the characteristics of the diffuse shell influence the adsorption of three types of proteins: Bovine Serum Albumin (BSA), fibrinogen, and a protein from the complement system that triggers recognition and elimination by macrophages. We have determined the structural characteristics of the diffuse shells using Nuclear Magnetic Resonance (NMR), Small Angle Neutron Scattering (SANS) and Quasi-Elastic Light Scattering (QELS). We have measured the adsorption of Bovine Serum Albumin (BSA) through Immuno-diffusion methods, and found that it adsorbed in substantial amounts even when the distance between dextran chains at the core-diffuse shell interface is quite short. We have observed the aggregation of the nanoparticles induced by fibrinogen, and found that it was prevented when the density of dextran chains protruding from the core surface was sufficiently high. Finally we have measured the activation of the complement system by the nanoparticles, and found that it was also limited by the surface density of dextran chains that protrude from the core and by their mesh size within the diffuse shell.
© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Colloidal particles are often used to carry molecules that are pharmaceutically active from the point of administration to a target site in the body. The requirements for such a particle are threefold (a) it should carry a sufficient number of active molecules; (b) it should avoid the body's mechanisms for recognition of foreign particles; (c) it should be able to deliver the active molecules to the target. In order to perform these functions, most colloidal carriers rely on a *core-diffuse shell* structure. Typically, the core is made of a hydrophobic polymer that dissolves or traps the active molecules. The diffuse shell (also called corona) is made of hydrophilic polymers that "hide" this hydrophobic core [1].

Colloidal carriers that are injected into the blood stream find themselves in a concentrated protein solution. The most abundant protein, serum albumin, is present in blood at a concentration of 4%. Consequently, the first event that takes place when a particle is injected in the blood is the adsorption of proteins, mainly serum albumin. Subsequently, other proteins may either displace the bound serum albumins or bind to the absorbed serum albumin layers [2–4]. Recognition and elimination of these foreign particles by the body defense mechanisms including macrophages may then occur through the absorption of specific proteins, e.g. the C3 protein of the complement system [5]. Upon binding, the C3 protein will change conformation, expose a reactive site, and release a signaling molecule that triggers the chain of bio-chemical events called the complement activation cascade [6]. This will lead to the elimination of the particles by the macrophages [7].

There are indications that some chemical characteristics of the diffuse shell have an effect on complement activation. Indeed a chain length effect has been found for diffuse shells made of PEG or polysaccharides [8,9]. Also, an effect of the grafting density of hydrophilic chains on the particle surfaces has been found in the case of PEG chains [8]. The nature of the hydrophilic chains may also be important. These effects may be caused by interactions of C3 with albumins that are already adsorbed [10]. Indeed, since the absorption of the C3 complement protein takes place after that of serum albumin, it will be influenced by the amount of bound serum

* Corresponding author. Univ Paris Sud, Physico-chimie Pharmacotechnie Biopharmacie, UMR CNRS 8612, F-92296 Chatenay-Malabry, France. Fax: +33 1 46 83 53 12.
E-mail address: Christine.vauthier@u-psud.fr (C. Vauthier).

0142-9612/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biomaterials.2010.10.026

albumins, their configurations at the surface, and their accessibility. Thus, in order to design the diffuse shell for avoiding complement activation, we need to know more about the adsorption of albumin.

Recently we have synthesized a collection of nanometric-sized drug carriers through self-assembly of block copolymers. As a hydrophobic block we chose poly(isobutylcyanoacrylate) (PIBCA), which is a bioerodible and bioeliminable polymer that has already been used to deliver in vivo a number of drug molecules [11,12]. As a hydrophilic block, we chose dextran, which is a polysaccharide known to restrict protein adsorption [13–15]. The self-assembly of these copolymers does yield particles with a core-diffuse shell structure that is determined by the respective lengths of the hydrophobic and hydrophilic blocks [9,16]. By changing the conditions of polymerization, we have produced two types of particles that differ only by the core size and by the density and extension of the diffuse shell. Remarkably, these two types have very different effects on complement activation. This opens the possibility of finding how the characteristics of the diffuse shell control protein adsorption and complement activation.

In this paper we report characterizations of these polymer particles by Small Angle Neutron Scattering (SANS), measurements of the adsorption of the abundant blood proteins, serum albumin, through rocket-immunoelectrophoresis [17] and radial immunodiffusion [18], measurements of fibrinogen-induced aggregation through SANS, and measurements of the activation of the complement system, through 2D immunoelectrophoresis [9,19]. We attempt to answer the following questions:

(a) Does the diffuse shell limit protein adsorption, what is the location of the adsorbed proteins, and what is their configuration?
(b) What is the relation between the number of hydrophilic chains that protrude out of the core and the number of proteins that can be adsorbed?
(c) Is there a relation between the number or the configuration of adsorbed proteins and the activation of the complement system?

2. Materials and methods

2.1. Preparation of nanoparticles

PIBCA nanoparticles were prepared by redox radical emulsion polymerization (RREP) according to the method of Bertholon et al. [9,16]. The polymerization initiated on dextran produces bloc-copolymers which self-assemble into nanoparticles with a *core-diffuse shell* structure.

Two dispersions were prepared named A and B, differing by the amount of initiator (dextran-cerium). Nanoparticles A were obtained in 8 mL of nitric acid 0.2 M containing 0.137 g of dextran (molar mass 69300 Da). The solution was placed under vigorous magnetic agitation at 40 °C and purged with nitrogen for 10 min. Then, 2 mL of a solution of cerium ammonium nitrate (0.08 M in nitric acid 0.2 M) and 0.5 mL of isobutylcyanoacrylate (IBCA) were added successively and the polymerization was allowed to proceed for 1 h at 40 °C. For nanoparticles B, the volume of nitric acid 0.2 M was 9.3 mL, the amount of dextran was 0.0502 g and the volume of cerium ammonium nitrate was 0.7 mL. All other conditions were the same.

Immediately after polymerization, the nanoparticle dispersions were cooled down to room temperature using an ice bath. They were purified by dialysis (Spectra/Por® membrane 100 000 g/mol molecular weight cut off) against 1 L of water 4 times to remove unreacted reagents and to raise the pH to a neutral value. Further purification included a first centrifugation at 720-× *g* for 10 min followed by a second centrifugation at 17,000-× *g* for 15 min to remove any aggregates. After the second centrifugation, the lower part of the supernatant (340 μl) was collected as fractions A_{middle} and B_{middle} respectively and the upper part of the supernatant (600 μl) as fractions A_{top} and B_{top} respectively. The volume fractions of the dispersions were around 1%.

2.2. Particle sizes and structures

2.2.1. Quasi-elastic light scattering (QELS)

The hydrodynamic radii of the nanoparticles were measured through QELS using a Zetasizer ZS90 instrument (Malvern, France). Measurements were performed at 25 °C at a scattering angle of 90°. The intensity correlation function was fitted through the instrument software to yield a volume-weighted distribution of particle sizes. From this distribution we calculated the total volume and total surface area of the nanoparticles, and the radius that describes the corresponding hydrodynamic volume/surface ratio. This radius is identical to the z-averaged radius of the particles defined by Eq. (1).

$$R_z = \sum N_i R_i^3 \Big/ \sum N_i R_i^2 \quad (1)$$

Dispersions were diluted in Milli® Q water filtered over a 0.22 μm membrane prior to the measurement.

2.2.2. Scanning electron microscopy (SEM)

Images of freeze-dried nanoparticle dispersions were obtained using a scanning electron microscope LEO 1530 (LEICA) equipped with a Gemini column. Prior to observation, freeze-dried samples were mounted on supports and coated with a 2 nm Pt/Pd layer using a Cressington sputter coater 208HR apparatus (Cressington, France) operating under argon atmosphere. The sizes of more than 300 individual nanoparticles were measured to determine the mean radius and the parameters of the size distribution.

2.2.3. Small Angle Neutron Scattering (SANS)

Scattering patterns from liquid nanoparticle dispersions were obtained using the instrument D11 at the ILL. The experiments yielded scattering patterns consisting of intensities as a function of scattering vector $\boldsymbol{q}$. Since the dispersions were isotropic, the scattering patterns were averaged over all directions of $\boldsymbol{q}$, yielding spectra of intensity I vs. magnitude q of the scattering vector, which is related to wavelength λ and scattering angle θ by $q = (4\pi/\lambda)\sin(\theta/2)$. The range of q values was 6×10^{-3} to 9×10^{-2} nm^{-1}. In this range, the interference patterns probe variations of the density of scattering length $\boldsymbol{\rho}(r)$ with a resolution of 10 nm. Scattering is then produced by the difference between the scattering density of a particle, $\boldsymbol{\rho}(r)$, and that of the solvent, $\boldsymbol{\rho}_s$. For the characterization of the particles, the dispersions were made in H_2O, as the difference of scattering density between the particles and H_2O was sufficient to produce strong scattering [20]. Spectra were fitted with a calculated curve for a lognormal distribution of spheres. The z-averaged value (R_z) was calculated using Eq. (1). Other experiments were made with dispersions made in H_2O-D_2O mixtures that matched the scattering density of the PIBCA-dextran nanoparticles, in order to enhance the contrast of the adsorbed proteins [20].

2.3. Composition of the nanoparticles

The monomer compositions of the PIBCA-dextran copolymers were determined through (^{1}H) NMR spectroscopy. The copolymers were dissolved at 40 °C in dimethyl-d_6-sulfoxide (DMSO, d_6) (Carlo Erba) at the concentration of 20 mg/mL. Spectra were then recorded on a 300 MHz Brucker instrument (Brucker, France). The composition was deduced from the spectra by calculating the ratio between the integral of dextran peaks (4.3–5 ppm) and that of the CH peak of PIBCA (1.8–2.0 ppm) [21]. The molecular weight of the PIBCA part of the copolymer was calculated from the composition in dextran and PIBCA deduced from the NMR spectra and the known molecular weight of dextran (69 300 g/mol), assuming that the copolymer was composed of one chain of PIBCA and one chain of dextran as shown by Bertholon et al. [16].

2.4. BSA adsorption through immunomethods

Nanoparticles at a volume fraction of 0.5% were incubated with different concentrations of BSA in 10 mM phosphate buffer for 3 h at 37 °C. After incubation, the amount of non-adsorbed BSA was determined using immunochemical methods including immunoelectrophoresis [17] and radial immunodiffusion [18] as previously described [20].

In both cases, agarose gel plates (12 cm × 8.5 cm) were prepared on Gelbond® films for agarose gel (GE healthcare, BioScience, Sweden). The agarose solution (13 mL) at a concentration of 1% contained 0.3 mL of the bovine serum albumin antisera (Sigma, France) diluted at 1/5 in a saline phosphate buffer (10 mM, NaCl 140 mM, KCl 25 mM, pH 7.5). For the radial immunodiffusion technique, the gel was prepared in the saline phosphate buffer (10 mM, NaCl 140 mM, KCl 25 mM, pH 7.5) whereas tricine buffer (lactate calcium 1 mM, Tris 63 mM, Tricine 27 mM, pH 8.6) was used to prepare gels for the immunoelectrophoresis method. Wells holding 5 μL samples were formed using a 2 mm diameter punch on the gel plates. After deposition of the samples in the wells (5 μL) including standard solutions of BSA and nanoparticles incubated with the different concentrations in BSA, gels were placed in a humid chamber and let to diffuse in a quiet place for 48 h at room temperature in the case of the radial immunodiffusion technique. The conditions for the immunoelectrophoresis were 18 h, 230 V, 12 mA (Electrophoresis power supply EPS 600, Amersham, Pharmacia Biotech, Orsay, France) using Tricine buffer as the running buffer for the electrophoresis performed in a Pharmacia LKB Multiphor II apparatus (Amersham, Pharmacia Biotech, Orsay, France). At the end of both types of experiments, the gels were dried using Watmann® filter papers and stained with coomassie blue to reveal of the formation of BSA-antibody immunoprecipitate. Each gel contained a set of 5–6 standard solutions of BSA of concentrations ranging from 20 to 900 μg/mL for internal calibration purpose. In the case of the radial

immunodiffusion technique, the calibration curve is a linear relationship between the square of the diameter of the immunoprecipitation ring and BSA concentration in the well for BSA concentrations ranging from 100 to 900 μg/mL. In the case of the immunoelectrophoresis, the calibration curve is a linear relationship between the height of the immunoprecipitation peak and the concentration in BSA in the well in the range of BSA concentrations comprised between 20 and 200 μg/mL. In both cases, calibration curves with correlation coefficients above 0.99 were obtained with a good reproducibility.

2.5. Fibrinogen-induced aggregation of the nanoparticles

Phosphate buffer 100 mM, pH 7.4 containing different concentrations of fibrinogen (human fibrinogen from Sigma) was added to nanoparticle dispersions at a volume ratio of 1/10. Final concentrations of fibrinogen ranged from 0 to 2 mg/mL. The behavior of the dispersion during the addition of the fibrinogen was observed for eventual precipitation. The microscopic aspects of the dispersions containing 2 mg/mL fibrinogen were observed with an optical microscope using the phase contrast mode (Olympus BH2). Dispersions that appeared homogeneous by microscopic observations were characterized by QELS and SANS.

2.6. Activation of the complement system

Activation of the complement system induced by the nanoparticles was evaluated using a 2D immunoelectrophoresis method as described in previous publications [9,19].

Briefly, the nanoparticles (400 cm^2 in 100 μl, as calculated from the hydrodynamic radius of the nanoparticles) were incubated with 50 μl human serum and 50 μl veronal buffer containing 0.15 mM calcium chloride and 0.5 mM magnesium chloride for 1 h at 37 °C. After incubation, 5 μL of each sample was subjected to a first electrophoresis on a 1% agarose gel. The second-dimension electrophoresis was carried out on Gelbond® films in 1% agarose gel plates containing a polyclonal antibody to human C3 (Complement C3 antiserum rose in goat, Sigma, France), recognizing both C3 and C3b. The films were finally dried and stained with Coomassie blue to reveal the presence of C3 and C3b which have reacted with the antibody (Sigma).

The complement activation factor was calculated from the ratio of the area of the peak attributed to C3b over the sum of the areas of the peaks attributed to C3b and to C3. This ratio was then normalized on a scale ranging from 0 to 100 in which 100 indicated total activation and 0 the spontaneous activation measured in absence of nanoparticles.

3. Results

3.1. Characterization of the nanoparticles

In order to understand what characteristics of the nanoparticles determined protein adsorption, we determined their geometric dimensions and their compositions. From these data we calculated, for each type of nanoparticle dispersion, the average thickness of the diffuse shell and the surface density of chains in the diffuse shell at the core surface.

3.1.1. Particle radii according to SEM, SANS and QELS

SEM images of the particles show that they are globular and polydisperse in sizes (Fig. 1). In the case of the B particles, we were able to measure individual particle radii, and to obtain the distribution of these radii (Table 1). In the case of type A particles the smaller particles were not imaged in SEM.

SANS spectra of the dispersions in H_2O measure the *core radii* of the nanoparticles. SANS spectra of dispersions A_{middle} A_{top} and B_{middle} are shown in Fig. 2. For type A dispersions, the scattered intensity follows Guinier's law at small q values and then continues with the q^{-4} decay predicted by Porod's law for dense particles, without any of the oscillations that are expected for monodisperse spheres. These features are the usual behavior when there is a single population of nanoparticles with sizes broadly distributed around a mean value, since each spectrum is the sum of intensities scattered by all particles in the dispersion [22]. Accordingly we fitted each spectrum with the calculated scattering curves for spheres with lognormal distributions of radii. From these distributions we calculated the specific surface area, for comparison with protein adsorption measurements, and the corresponding radius, R_z, for comparison with the hydrodynamic radius determined through QELS (Table 1). As expected, nanoparticles isolated from the lower part of the supernatant (fractions A_{middle}) were of larger size than those isolated from the top part (A_{top}).

The spectrum of the B_{middle} nanoparticles is different: it has a power-law decay at low q values which is characteristic of a population of aggregated particles. Accordingly, we fitted this spectrum with the scattering curve calculated for a lognormal distribution of spheres plus a population of larger particles (Table 1). This was consistent with QELS measurements presented next.

QELS measurements yield the *distribution of hydrodynamic radii* of nanoparticles (Fig. 3). For A_{middle}, A_{top} and B_{top},, QELS shows a single population of nanoparticles with a lognormal distribution, from which we calculated the z-averaged hydrodynamic radii (Table 1). These radii characterize the Brownian motions of particles with their diffuse shells and the solvent that is trapped by the diffuse shells. Hence the hydrodynamic thickness of the diffuse shell can be calculated as a difference between the average hydrodynamic radius and the average core radius. However, since the populations of nanoparticles are polydispersed, the results from different methods can be compared only if they make the same average of the distribution of radii. Thus, we compared the z-averaged diffuse shell radius calculated from the distribution of particle sizes measured through QELS with the z-averaged core radius calculated from the distribution that fits the SANS spectrum. The difference of these radii was taken as the diffuse shell thickness (Table 1).

We found that this thickness was larger than the end-to-end distance of dextran polymers that have the same molar mass M_w as the dextran block of the copolymers used here. Indeed, extensive light scattering measurements on dilute aqueous dextran solutions yield the law given in Eq. (2) for the radius of gyration R_g, in nm, as a function of the molar mass M_w in Da [23].

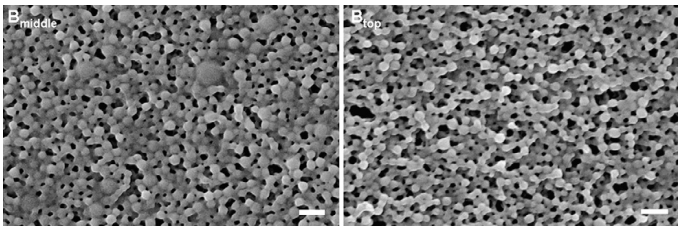


Fig. 1. Scanning electron micrographs of Type B nanoparticles. Lyophilized powder coated with a 2 nm thick Pt/Pd layer. Scale bar = 200 nm.

Table 1
Radii of the nanoparticles as determined by QELS, SEM and SANS (R_z calculated with Eq. (1)). Diffuse shell thicknesses were deduced from comparing QELS with either SEM (*) or SANS (**). The experimental error was estimated as ± 5 nm for QELS measurements and as ± 2 nm for SANS measurements.

	QELS		SEM*		SANS**		Diffuse shell thickness (nm)
	R_z (nm)	STD (nm)	R_z* (nm)	STD (nm)	R_z** (nm)	STD (nm)	
A_{middle}	112	32			63	20	49**
A_{top}	47	24			26	8	21**
B_{middle}	88	40	37		40	20	48**
B_{top}	49	12	32				17*

$$R_g = 0.4(M_w)^{0.3} \quad (2)$$

The low value of the exponent is due to the fact that dextran polymers are branched. According to this law, the radius of gyration of the dextran polymers used here should be 11 nm. For a Gaussian chain, the corresponding end-to-end distance would be 28 nm, and for a branched polymer it would still be about the same. As shown in Table 1, the calculated value of the diffuse shell thickness of A_{middle} nanoparticles (49 nm) is larger than this length, indicating that the dextran chains were stretched as a result of their confinement in the diffuse shell. In contrast, for the smaller nanoparticules (Atop and Btop), the diffuse shell thickness (respectively 21 and 17 nm) is smaller than this value. In this case the small thickness results from the high curvature of the nanoparticle surfaces [24,25].

In the case of Bmiddle nanoparticles, the distribution of hydrodynamic radii was bimodal with a population of large aggregates (2–4 μm) (Fig. 3). These aggregates contributed a very small fraction of the surface area but to a larger fraction of the light scattering intensity. Consequently, the precision of surface area determinations can be considered to be comparable to that of the other measurements.

3.1.2. Composition of the particles

The compositions in dextran and PIBCA of the copolymer forming the nanoparticles were determined by NMR. NMR spectra from the A and B-type dispersions were similar to those obtained in previous work [16]. The composition of each copolymer was deduced from the integration of peaks due to the characteristic protons of dextran and PIBCA (Table 2). The fraction of PIBCA found in type A copolymers was much lower than that found in type B copolymers. This result agrees with the fact that type A copolymers were prepared with a higher concentration of dextran in the original solution. In addition, there was also a smaller difference between the middle and top fractions of each dispersion.

In a previous work [16], we have shown that the copolymer was composed of one chain of dextran and one chain of PIBCA. Knowing the molecular weight of dextran (69,300 g/mol), and the composition of the copolymer we calculate the degree of polymerization and the molecular weight of the PIBCA moiety. Results summarized in Table 2 showed that the Mw of the PIBCA part of the type B copolymers were much larger than those in type A copolymers (Table 2).

We can then calculate the number N of PIBCA chains in the core of a nanoparticle according to the core volume V_{core} and the PIBCA chain volume $v_{P-chain}$ or the mass per chain and the specific mass of PIBCA $d = 1300$ kg/m^3:

$$N = \frac{V_{core}}{v_{P-chain}} = \frac{4\pi}{3}\frac{R^3}{v_{P-chain}} \quad (3)$$

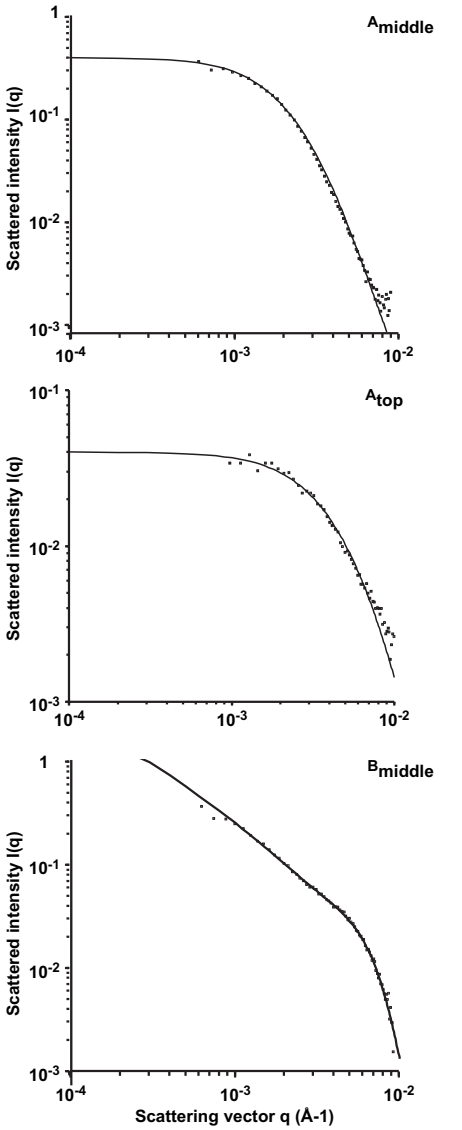


Fig. 2. SANS spectrum of the dispersions A_{middle}, A_{top}, B_{middle} in H_2O (volume fraction 1% in Phosphate buffer 10 mM, pH 7.4). For A_{middle} and A_{top} the fits are the scattering curves calculated for dispersions of dense homogeneous spheres with lognormal distributions of radii. For B_{middle} the fit is obtained with a similar distribution plus a population of larger particles.

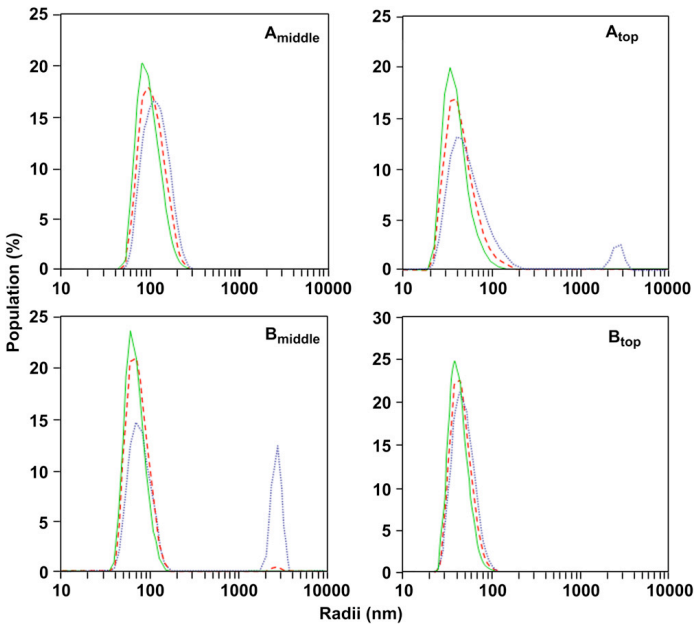


Fig. 3. Distributions of hydrodynamic radii for the nanoparticles according to QELS. Horizontal axis: hydrodynamic radii R_i. Vertical axis: populations of particles $P(R_i)$, normalized to 100%. In each graph, the blue (short doted line) curve is the volume-weighted distribution, obtained from the instrument software. The red (long doted line) curve is the surface-weighted distribution, and the green (solid line) curve is the distribution of volume/surface ratios, both calculated from the volume-weighted distribution. The average radius calculated from the ratio of total volume/total surface is the z-average hydrodynamic radius R_z, which is then compared with the core radius R_z^* extracted from TEM or R_z^{**} extracted from SANS.

This can be combined with the core surface area to yield the area per chain at the core surface, or its inverse, the surface density of dextran chains that are tied to the core:

$$a_{P-chain} = \frac{S_{core}}{N} = \frac{3v_{P-chain}}{R} \quad (4)$$

In this equation, the radius R is that corresponding to the specific surface area, i.e. the ratio of the total volume of all the cores to the total surface area of all the cores. This radius is given as "R_z" in Table 1. The resulting values of the surface area per dextran chain are given in Table 2. These values show that the B-type particles have a very large surface area per chain, due to the very high molar mass of their PIBCA chains.

Finally we calculate the volume fraction of dextran chains in the diffuse shell according to the volume of one dextran chain:

$$\phi_{dextran} = \frac{Nv_{D-chain}}{V_{corona}} \quad (5)$$

We can also calculate the average mesh size ξ of the diffuse shell according to the scaling law [26]

$$\frac{\xi}{b} = (\phi_{dextran})^{\left(\frac{-1}{d-d_f}\right)} \quad (6)$$

where b is the length of the statistical segment of the dextran chain ($b = 1$ nm) [27], d is the space dimension ($d = 3$) and d_f is the fractal dimension of the dextran polymers (approximately $d_f = 2$) [23]. The values of $\varphi_{dextran}$ and ξ are also listed in Table 2. In the case of the A_{top} particles, the calculated values of $\phi_{dextran}$ and ξ were excessively high, due to the very low degree of polymerization of the PIBCA block. Since this degree of polymerization is an average, it is

Table 2
Summary of the characteristics of the copolymers composing the nanoparticles and calculation of the characteristics of the diffuse shell of dextran: area per dextran chain at the surface of the core of the nanoparticles, volume fraction of dextran in the diffuse shell and mesh size formed by dextran chains in the diffuse shell.

Sample	%PIBCA	Mw PIBCA[a]	PIBCA degree of polymerization	Area per chain at the core surface (nm^2)	Volume fraction of dextran in the diffuse shell	Mesh size of diffuse shell (nm)
A_{middle}	48 ± 1	63969	418	3.9 ± 0.6	0.24 ± 0.01	4 ± 0.2
A_{top}	22 ± 0.5	19 546	128	2.8 ± 0.6		
B_{middle}	83 ± 2	338 347	2211	32 ± 5	0.02 ± 0.001	47 ± 2
B_{top}	75 ± 2	207 900	1359	25 ± 5	0.13 ± 0.01	8 ± 0.3

[a] Calculated from the composition of the copolymer.

likely that copolymers with the smaller PIBCA blocks remained in the aqueous phase, and that the A_{top} nanoparticles contained less dextran than assumed by equations/3/and/5/.

3.2. Adsorption of BSA

3.2.1. Adsorption isotherms through radial immunodiffusion and rocket-immunoelectrophoresis

The adsorption of BSA on the nanoparticles at a volume fraction of 0.5% in aqueous dispersion was determined through two methods: radial immunodiffusion and rocket immunoeletrophoresis. Both methods measure the concentration of BSA that remains in the dispersion upon equilibrium with the nanoparticles. Although the analysis is performed without previous separation of the nanoparticles from the incubation medium containing BSA, only the free BSA will be able to penetrate in the gel during either radial diffusion or electrophoresis and will be quantified by the two methods. As shown in previous work [20,28], adsorption equilibrium was reached after 1 h of incubation and both methods yield very similar adsorption isotherms. The amount of adsorbed BSA was calculated for each initial concentration of BSA in the solution, and then the results were converted into adsorption isotherms. The isotherms given by each type of nanoparticles have a similar shape (Fig. 4) with a first section of high affinity binding (i.e. the amount adsorbed equals the initial amount) followed by a plateau that indicates saturation of the nanoparticle surfaces (i.e. the adsorbed amount saturates regardless of the total amount of BSA). It was found that the concentration of the adsorbed BSA at the beginning of the plateau varied slightly for the different nanoparticles considered in this work (Table 3). These values were of the same order of magnitude than that found in a previous work considering the same type of nanoparticles prepared by another polymerization method [20,28]. They were close to those given in the literature for different types of surfaces, excepted for those which were coated with a dense brush of polyethylene glycol [29–32].

3.2.2. Structural effects of adsorbed BSA

The amounts of adsorbed BSA were too small (in comparison with the polymer mass per nanoparticle) to be determined through SANS. Indeed, attempts were made to measure the scattering of the adsorbed BSA using the contrast match technique, as in previous work [20]. SANS experiments were performed with dispersions in blends of $H_2O + D_2O$. The isotopic compositions were close to the average scattering density of the nanoparticles. The shapes of spectra obtained at the different contrasts with and without BSA were similar. The composition of the solvent that matched the nanoparticles was 29.3% D_2O, 70.7% H_2O in absence of BSA and 29.7% D_2O, 70.3% H_2O in presence of BSA. The slight increase of the average scattering density of the nanoparticles with BSA can be attributed to the contribution of the protein. However the precision of this determination was not nearly as good as that of the immunodiffusion methods.

On the other hand, SANS at full contrast and QELS were used to evidence structural effects such as aggregation or dissociation of particles that could be caused by the adsorption of BSA. The main effect was found with the dispersion B_{middle}, which was initially aggregated (Figs. 2 and 3). Upon addition of BSA, the QELS data indicated that the population of very large aggregates (radii 2–4 μm) was strongly reduced, and the population of nanoparticles shifted to lower sizes (Fig. 5). Similarly, the average radius measured through SANS at full contrast also shifted to lower values. For all other dispersions (A_{middle}, A_{top} and B_{top}), there were no effects of BSA adsorption: the distributions of hydrodynamic radii remained the same, and the spectra from dispersions with BSA were nearly identical to those of the initial nanoparticles.

3.3. Interactions of the nanoparticles with fibrinogen

Addition of fibrinogen in type B dispersions induced a strong precipitation of the nanoparticles. In contrast, type A dispersions remained homogeneous during and after addition of fibrinogen. Observations through optical microscopy confirmed that large precipitates were formed in type B dispersions while type A dispersions remained homogeneous at the macro- and microscopic scales (Fig. 6). The analysis of type A dispersions by QELS revealed only a slight increase in the hydrodynamic radius of the nanoparticles. After the addition of 2 mg/mL of fibrinogen, the hydrodynamic radius of dispersion A_{middle} increased from 112 to 123 nm

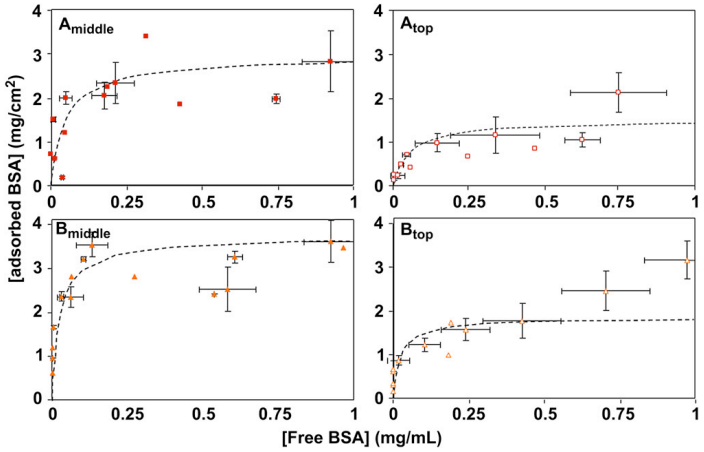


Fig. 4. Adsorption isotherms of BSA on the different preparations of nanoparticles after incubation in 10 mM phosphate buffer (pH 7.4) for 3 h at 37 °C. The nanoparticles were used at a volume fraction of 0.5%. Doted line: simulation of the amount of BSA adsorbed on the nanoparticle surface calculated from the Langmuir adsorption equation [7].

Table 3
BSA adsorbed at the plateau of saturation deduced from the adsorption isotherms and from simulations performed with a lognormal distribution of the particle size and various thickness of the adsorbed layer. The monolayer adsorption is calculated for BSA molecules in the "edge-on" configuration (see Discussion).

Sample	Adsorbed amount Γ mg/m^2		Langmuir fit parameters			Experimental value
	Beginning of plateau	End of plateau	Affinity χ mL/μg	Monolayer Γ_0 mg/m^2	Fraction Unavailable α	Fraction unavailable
A_{middle}	2.1 ± 0.2	3.0 ± 0.2	0.02	3.66	0.2	0.18
A_{top}	1.0 ± 0.1	2.4 ± 0.3	0.02	3.66	0.7	0.34
B_{middle}	2.7 ± 0.4	3.4 ± 0.2	0.04	3.66	0	0.07
B_{top}	1.4 ± 0.3	2.8 ± 0.5	0.04	3.66	0.5	0.23

while that of dispersion A_{top} increased from 47 to 52 nm. SANS also detected a slight aggregation of A_{top} (up to 5 particles per aggregate) but not in the case of A_{middle}.

3.4. Activation of the complement system

Finally, the capacity of the different nanoparticles to activate the complement system was measured by analyzing the C3 protein of the complement in serum by 2D immunoelectrophoresis. The two fractions of nanoparticles A induced a similar and weak activation of the complement system shown on the gel by the larger peak on the left side which corresponds to the native C3 protein. The complement activation factor was around 30.1 ± 0.4 and 27 ± 3 for nanoparticles A_{middle} and A_{top} respectively (Fig. 7). The two fractions of nanoparticles B have a very different capacity to activate the complement system. Fraction B_{middle} induced an almost complete activation of the complement shown by the almost absence of the peak of C3 protein on the left side of the gel and the presence of a large peak on the right side corresponding to the activated form of the protein C3, i.e. C3B. The complement activation factor (95 ± 6) was closed to the maximum value (100) (Fig. 7). The presence of two peaks of proteins on the electrophoregram of the serum incubated with nanoparticles B_{top} indicated that the activation of the complement system produced by these nanoparticles was much lower than that found for nanoparticles B_{middle}. The complement activation factor was much lower (41 ± 7) but remained significantly higher than those obtained with nanoparticles A_{middle} and A_{top}.

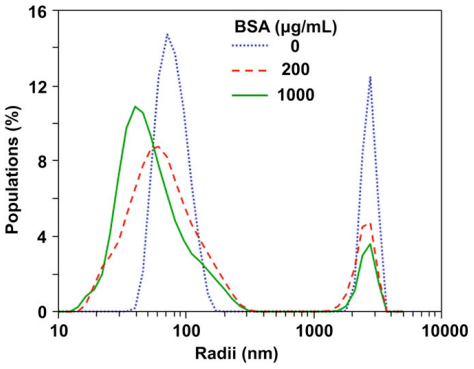


Fig. 5. Effect of added BSA on the population of aggregates within the B_{middle} dispersion, according to QELS. The concentrations of BSA are indicated in the Figure. Horizontal axis: hydrodynamic radii R_i. Vertical axis: populations of particles $P(R_i)$, normalized to 100%. Note the reduction in magnitude of the population of very large aggregates and the shift to lower radi of the main population. For all other dispersions, the addition of BSA had no effect on the measured radii (see the text).

4. Discussion

The aim of this discussion is to deal with the questions raised in the introduction, regarding the effects of the diffuse shell on the adsorption of small proteins (BSA), fibrinogen, and the larger proteins that trigger the activation of the complement system. The discussion is based on the characteristics of the diffuse shells (Table 2), the characteristics of the proteins (Table 4 below) [33–39], the measured adsorption of BSA (Table 3), the occurrence of fibrinogen-induced aggregation and the measurements of complement activation.

4.1. Does the diffuse shell limit the adsorption of small proteins such as BSA?

We have measured the adsorption of BSA on nanoparticles that have very different diffuse shells: A-type diffuse shells have a very high surface density of dextran chains (spaced by about 2 nm); B-type diffuse shells have a low surface density of dextran chains (spaced by about 6 nm). Yet all these particles adsorb a substantial amount of BSA, comparable to the amounts adsorbed on other surfaces [20,31,32,40–43]. Hence these diffuse shells do not prevent the adsorption of small proteins. Still, there are differences in the adsorption isotherms of the 4 types of nanoparticles. In order to interpret these differences, it is instructive to use a thermodynamic model for the adsorption. The simplest one is the Langmuir model, which assumes equilibrium between a solution of free proteins and a collection of independent adsorption sites at the nanoparticle surfaces. The equilibrium adsorption is ruled by two parameters: the affinity χ of the protein for the surface sites, which has the dimensions of an inverse concentration, and the saturation value of the adsorption, Γ_0 (1-α), where Γ_0 is the adsorbed amount for a full protein monolayer, and α is the fraction of the surface area that is unavailable for adsorption because of the dextran chains. The model yields the following equation for the adsorbed amount, Γ, as a function of the concentration ρ, of free protein in the solution (Eq. (7)).

$$\Gamma = \Gamma_0(1-\alpha)\frac{\rho\chi}{1+\rho\chi} \quad (7)$$

In order to use this equation, we must determine Γ_0, and for this we must make an assumption regarding the configuration of the adsorbed proteins. BSA can be approximated as a prism-shaped object, with dimensions 8.4 × 8.4 × 3.8 nm^3 [33]. Adsorption in a "flat" configuration requires a free area of 70 nm^2, which is certainly not available at the surfaces of the A-type particles, and rather unlikely at the surfaces of the B-type particles. Adsorption in an "edge" configuration requires a free area of 26 nm^2, which may be more easily found if the adsorption site need not be circular (see below). In the following, we assume that BSA is adsorbed on all particles in the "edge" configuration. A full monolayer in this configuration would contain $\Gamma_0 = 3.66$ mg/m^2 of adsorbed protein. With this choice, the Langmuir model parameters are as given in Table 3 above. Accordingly, the fraction (α) of the core surface that

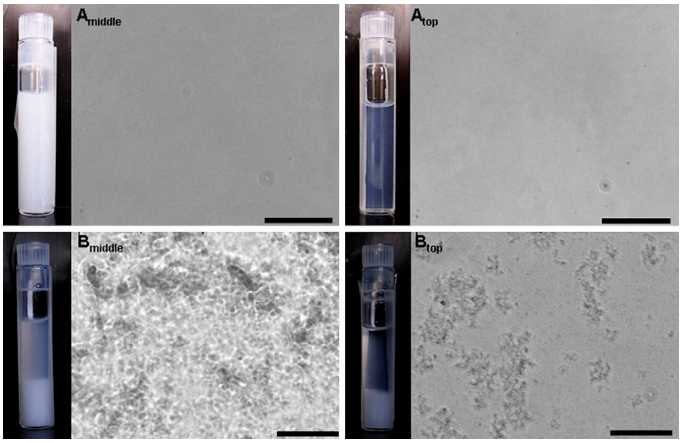


Fig. 6. Macroscopic and microscopic images of the nanoparticle dispersions as they appeared right after the addition of fibrinogen at a concentration of 2 mg/mL to the different dispersions of nanoparticles. Scale bar: 50 μm.

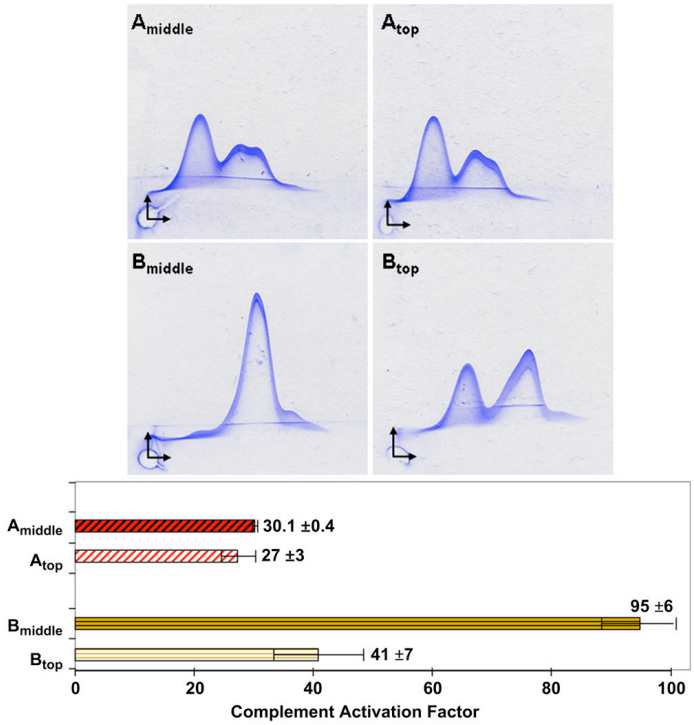
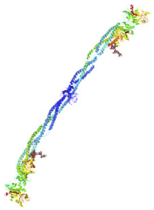


Fig. 7. Activation of the complement system evaluated from 2D immunoelectrophoresis experiments after incubation of the different nanoparticles with human serum in veronal buffer supplemented with calcium chloride and magnesuim chloride. The concentration in nanoparticles was adjusted to obtain an equivalent surface area of 2000 cm^2/mL in the test tube. Upper part: immunoelectrophoregram shown by the analysis of the C3 proteins after incubation of the serum with the different nanoparticles. Lower part: complement activation factor deduced from the analysis of the immunoelectrophoregrams.

Table 4
Physical characteristics of the proteins.

Protein	BSA [33]	Fibrinogen [34,35]	Complement C3 [5]
MM	67 000	340 000 Dimer consisting of three pairs of polypeptide chains (Aα, Bβ and γ)$_2$	184 342* [36] Chain α 110 000 Chain β 75 000 1 disulfide bond between chains
Specific density		1.38	
PI	4.7	5.8	5.7
Shape	Prism/hard shape	Cylinder	Elongated curve shape
Dimensions (nm)	8 × 8 × 3.8	diameter 3.2, length 47.5	15.2 × 9 × 8.4
Plasma concentration mg/mL	40	2.6–3	1.2
PBD ID	HSA : 1E78 [37] Not find yet BSA	3GHG [38]	2A73 [39]
			

* calculated from the aminoacid sequence.

is unavailable for adsorption is only 0.2 in the case of A_{middle}, and 0.7 in the case of A_{top}.

How do the BSA molecules manage to adsorb on surfaces that already have a high density of hydrophilic chains? The first step in protein adsorption is the diffusion of the protein through the diffuse shell to the surface. This diffusion may be prevented if the overall dimensions of the protein are much larger than the mesh size of the diffuse shell. Thus the dimensions of BSA (Table 4) must be compared with the average mesh sizes of the diffuse shells (Table 2). For type A particles, the average mesh size is $\xi = 4$ nm, and there are of course considerable fluctuations around this value. Therefore diffusion of BSA through the mesh is difficult but certainly not impossible. For type B particles, the average mesh sizes are $\xi = 47$ nm and 8 nm for B_{middle} and B_{top} respectively, and therefore diffusion of BSA through these diffuse shells is not restricted. Furthermore, in case of type B_{middle} particles, the mesh size (47 nm) of the dextran brush is of the same order of magnitude as the thickness of the diffuse shell (48 nm). Completely free path from the end of the diffuse shell to the surface of the core may exist on the surface of this nanoparticle which can give almost free access of the core surface to proteins.

The next step is the binding of the protein to the core surface. At first sight, it appears that the surface is too crowded with dextran chains for this to occur. For A_{middle} particles, the area per chain at the core surface is only 3.9 nm^2 and for A_{top} it is 2.8 nm^2 (Table 2). The required area for a BSA macromolecule adsorbed in the "edge" configuration is much larger, equal to 26 nm^2. A simple statistical calculation indicates that a BSA macromolecule is very unlikely to find an area of this magnitude available for adsorption at the core surface. Consider the insertion of an object with area S_1 in to a system with area S having N *randomly* distributed "hard" points. The density of points on the surface, ρ, is given in Eq. (8).

$$\rho = \frac{N}{S} \tag{8}$$

The probability of a successful insertion, P_{insert} (where the object does not overlap any of the points) is given by Eq. (9)

$$P_{insert} = \left(1 - \frac{S_1}{S}\right)^N = \left(1 - \frac{\rho S_1}{N}\right)^N \tag{9}$$

For a large surface with many "hard points" ($N \to \infty$) this yields:

$$P_{insert} = \exp(-\rho S_1) \tag{10}$$

The dimensionless argument of the exponential is the mean number of points per inserted disc. The free energy difference gives a measure of the process of inserting macromolecules of different area in the presence of these points. Using the experimental numbers we find a probability of insertion equal to 1×10^{-3} in the case of A_{middle} particles and even lower for A_{top}. (Table 5).

A similar mismatch has already been observed for other particles with hydrophilic diffuse shells [29,30,44]. However, this mismatch could be explained in different ways:

(i) The BSA macromolecules can change conformations, which make it possible for them to fit into crowded areas. Indeed it is well known that BSA does not retain its native conformation upon adsorption [10,32,45–47].
(ii) The surfaces of albumin macromolecules have "patches" of positive electrical charges that are attracted to negative charges at the nanoparticle core surface [37].
(iii) The BSA macromolecules can take conformations that make it possible for them to bind to the diffuse shells of dextran.

We conclude that the diffuse shells of dextran are not able to prevent the adsorption of *small flexible proteins*, even when the density of chains within the diffuse shell is quite high.

4.2. Does the diffuse shell limit the adsorption of larger proteins?

Fibrinogen is a large protein that binds to "foreign" surfaces and causes them to aggregate through a bridging mechanism [48]. Hence, resistance to aggregation by fibrinogen is a good predictor of

Table 5
Comparison of the diffuse shell mesh sizes with the protein dimensions. Insertion probabilities of the proteins at the core surfaces, based on the area per dextran chain at the core surface (see Table 2).

Sample	diffuse shell mesh size/BSA	diffuse shell mesh size/fibrinogen	diffuse shell mesh size/C3	P_{insert} for BSA	P_{insert} for Fibrinogen	P_{insert} for C3
A_{middle}	1.1	1.3	0.5	1×10^{-3}	6×10^{-2}	4×10^{-9}
A_{top}				1×10^{-4}	2×10^{-2}	1×10^{-12}
B_{middle}	12.4	14.7	5.6	4×10^{-1}	7×10^{-1}	1×10^{-1}
B_{top}	2.0	2.4	0.9	3×10^{-1}	6×10^{-1}	5×10^{-2}

resistance to adsorption of large proteins. In this respect, the two particle types are quite dissimilar. Type B particles were strongly aggregated by fibrinogen, whereas A_{top} are only weakly aggregated, and A_{middle} not at all. These differences are in line with the protection of the particle cores by dextran chains.

First, the diffusion of fibrinogen through the diffuse shell of dextran depends on its dimensions (a thin cylinder with diameter 3.2 nm and length 47 nm) (Table 4). The diameter is comparable or smaller than the mesh size of the diffuse shell for all particles, but the length is much larger. Hence, diffusion of fibrinogen through the diffuse shell is not restricted if it diffuses along its length. Next, the binding of a fibrinogen macromolecule requires an area of about 10 nm^2, compared with an area per dextran chain of 3–4 nm^2 for A particles and 30 nm^2 per dextran in B particles (Table 2). From the statistical argument given above, we find an insertion probability of 0.06 for A_{middle} particles and 0.7 on B_{middle} particles (Table 5). Hence the adsorption of fibrinogen is indeed prevented by high surface density of dextran chains in the case of type A particles, but not in the case of type B particles. This is in good agreement with the observation of fibrinogen-induced aggregation in the case of B particles only.

Complement C3 protein is also a large protein (molar mass 180000) with dimensions $15.2 \times 9 \times 8.4$ nm^3. The adsorption of C3 is the first event that triggers the activation of the complement cascade [5,6]. The dimensions of C3 are larger than the mesh size of type A diffuse shells, but smaller than those of type B diffuse shells. As in the case of BSA, diffusion through the diffuse shells of type A particles may be difficult, but not impossible due to the fluctuations of dextran chain conformations. However, the binding of C3 requires an area of 80 nm^2. The probability of finding an area of this magnitude at the nanoparticle surfaces is practically zero for type A particles (Table 5). On the other hand, this insertion probability at the core surface is 0.1 for B_{middle}, and 0.05 for B_{top} (Table 5).

Thus we have two factors that may explain the differences in complement activation between the nanoparticles. Both yield a satisfactory correlation. Fig. 8a presents the correlation between diffuse shell mesh size and complement activation, while Fig. 8b presents the correlation between the insertion probability at the core surface and complement activation. The correlation appears surprisingly good for the mesh size, and worse for the insertion probability at the core surface. A possible interpretation for this effect is that the C3 protein does not need to reach the actual core surface in order to trigger complement activation. Indeed, the dextran diffuse shell may contain albumin macromolecules that have changed conformations in order to bind to the core surface. These bound albumins may be recognized by the C3 proteins and in this way trigger the complement cascade.

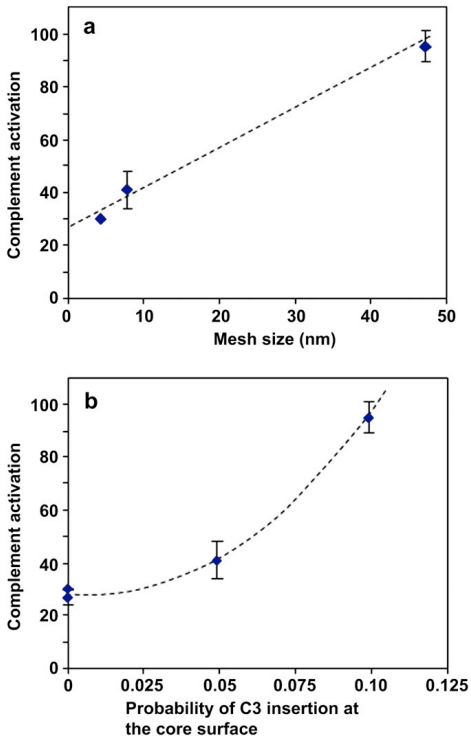


Fig. 8. Relation between complement activation induced by the nanoparticles and the mesh size of the diffuse shell (a) and the insertion probability of the C3 protein at the core surface (b).

5. Conclusion

As presumed, the characteristics of the diffuse shells determine the interactions of the core-diffuse shell nanoparticles with serum proteins. The three investigated proteins are very different, and their behaviors are not at all the same. BSA diffuses through the diffuse shell and binds at the core surface even when the density of hydrophilic chains is quite high. This may be due to conformational flexibility of the protein and to attractions between opposite charges located at the protein surface and at the core surface. Fibrinogen manages to diffuse through the diffuse shell because it has the shape of a long cylinder. However, its adsorption at the core surface is restricted by the density of hydrophilic chains. The binding of fibrinogen to neighboring nanoparticles leads to the formation of aggregates of the nanoparticles. We also measured the activation of the complement cascade which is triggered by adsorption of protein C3. From these measurements, we inferred that C3 is repelled by diffuse shells that have a dense mesh of hydrophilic chains. If it manages to diffuse through this mesh, then it may bind to albumins that are already bound to the core surface or to the diffuse shell. In this case, the complement cascade is activated. In order to prevent aggregation caused by fibrinogen and

recognition by the C3 protein of the complement system, it is necessary to keep these two proteins away from the core surface and from small proteins which adsorbed first during the opsonization process. This requires diffuse shells that have a very high density of hydrophilic chains.

Acknowledgments

The authors wish to thank Henkel Biomedical for constant support in giving the monomer used in this study, and Nicolas Tsapis at the University of Paris South, UMR CNRS 8612 for his assistance in scanning electron microscopy (Plateforme ICMPE, CNRS, Thiais, France). This work was supported by ILL proposal 9-12-176.

Appendix

Figures with essential colour discrimination. Certain figures in this article, particularly Figs. 3–8, are difficult to interpret in black and white. The full colour images can be found in the on-line version, at doi:10.1016/j.biomaterials.2010.10.026.

References

[1] Vauthier C, Bouchemal K. Methods for the preparation and manufacture of polymeric nanoparticles. Pharm Res 2009;26:1025–58.
[2] Olivier JC, Vauthier C, Taverna M, Puisieux F, Ferrier D, Couvreur P. Stability of orosomucoid-coated polyisobutylcyanoacrylate nanoparticles in the presence of serum. J Control Release 1996;40:157–68.
[3] Cedervall T, Lynch I, Foy M, Berggård T, Donnelly SC, Cagney G, et al. Detailed identification of plasma proteins adsorbed on copolymer nanoparticles. Angew Chem Int Ed Engl 2007;46:5754–6.
[4] Noh H, Vogler EA. Volumetric interpretation of protein adsorption: competition from mixtures and the Vroman effect. Biomaterials 2007;28:405–22.
[5] Sahu A, Lambris JD. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. Immunol Rev 2001;180:35–48.
[6] Law SK, Dodds AW. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. Protein Sci 1997;6:263–74.
[7] Vonarbourg A, Passirani C, Saulnier P, Benoit JP. Parameters influencing the stealthiness of colloidal drug delivery systems. Biomaterials 2006;27:4356–73.
[8] Vittaz M, Bazile D, Spenlehauer G, Verrecchia T, Veillard M, Puisieux F, et al. Effect of PEO surface density on long-circulating PLA-PEO nanoparticles which are very low complement activators. Biomaterials 1996;17:1575–81.
[9] Bertholon I, Vauthier C, Labarre D. Complement activation by core-shell poly (isobutylcyanoacrylate)-polysaccharide nanoparticles: influences of surface morphology, length, and type of polysaccharide. Pharm Res 2006;23:1313–23.
[10] Andersson J, Nilsson Ekdahla K, Lambris JD, Nilsson B. Binding of C3 fragments on top of adsorbed plasma proteins during complement activation on a model biomaterial surface. Biomaterials 2005;26:1477–85.
[11] Vauthier C, Ponchel G, Labarre D. Design aspects of poly(alkylcyanoacrylate) nanoparticles for drug delivery. J Drug Target 2007;15:641–63.
[12] Nicolas J, Couvreur P. Synthesis of poly(alkylcyanoacrylate)-based colloidal nanomedicines, 1. Wiley Interdiscip Rev Nanomed Nanobiotechnol; 2009. p.111–127.
[13] Osterberg E, Bergström K, Holmberg K, Schuman TP, Riggs JA, Burns NL, et al. Protein-rejecting ability of surface-bound dextran in end-on and side-on configurations: comparison to PEG. J Biomed Mater Res 1995;29:741–7.
[14] Monchaux E, Vermette P. Development of dextran-derivative arrays to identify physicochemical properties involved in biofouling from serum. Langmuir 2007;23:3290–7.
[15] Perrino C, Lee S, Choi SW, Maruyama A, Spencer ND. A biomimetic alternative to poly(ethylene glycol) as an antifouling coating: resistance to nonspecific protein adsorption of poly(l-lysine)-graft-dextran. Langmuir 2008;24:8850–6.
[16] Bertholon I, Lesieur S, Labarre D, Besnard M, Vauthier C. Characterization of dextran–poly(isobutylcyanoacrylate) copolymers obtained by redox radical and anionic emulsion polymerization. Macromolecules 2006;39:3559–67.
[17] Laurell CB. Quantitative estimation of proteins by electrophoresis in agarose gel containing antibodies. Anal Biochem 1966;15:45–52.
[18] Mancini G, Carbonara AO, Heremans JF. Immunochemical quantitation of antigens by single radial immunodiffusion. Immunochemistry 1965;2:235–54.
[19] Boackle RJ, Caughman GB, Vesely J, Medgyesi G, Fudenberg HH. Potentiation of factor H by heparin: a rate-limiting mechanism for inhibition of the alternative complement pathway. Mol Immunol 1983;20:1157–64.
[20] Vauthier C, Lindner P, Cabane B. Configuration of bovine serum albumin adsorbed on polymer particles with grafted dextran corona. Colloids Surf B Biointerfaces 2009;69:207–15.
[21] Pirker S, Kruse J, Noe C, Langer K, Zimmer A, Kreuter J. Characterization of polybutyleyanoacrylate nanoparticles. part II: determination of polymer content by NMR-analysis. Int J Pharm 1996;128:189–95.
[22] Lindner P, Neutrons Zemb Th. X-rays and light: scattering methods applied to soft condensed matter. Amsterdam: Elsevier; 2002. – North Holland Delta Series.
[23] Ioan CE, Aberle T, Burchard W. Structure properties of dextran. 2. dilute solution. Macromolecules 2000;33:5730–9.
[24] Garvey MJ, Tadros TF, Vincent B. A comparison of the adsorbed layer thickness obtained by several techniques of various molecular weight fractions of poly (vinyl alcohol) on aqueous polystyrene latex particles. J Colloid Interface Sci 1976;55:440–53.
[25] Lafuma F, Wong K, Cabane B. Bridging of colloidal particles through adsorbed polymers. J Colloid Interface Sci 1991;143:9–21.
[26] Cabane B, Dubois M, Lefaucheux F, Robert MC. Mesh size of aqueous TMOS gels. J Non Cryst Solids 1990;119:121–31.
[27] Guizard C, Chanzy H, Sarko A. Molecular and crystal structure of dextrans: a combined electron and X-ray diffraction study. 1. the anhydrous high-temperature polymorph. Macromolecules 1984;17:100–7.
[28] Olivier JC, Vauthier C, Taverna M, Ferrier D, Couvreur P. Preparation and characterization of biodegradable poly(isobutylcyanoacrylate) nanoparticles with the surface modified by the adsorption of proteins. Colloids Surf B Biointerfaces 1995;4:349–56.
[29] Gaucher G, Asahina K, Wang J, Leroux JC. Effect of poly(N-vinyl-pyrrolidone)-block-poly(D, L-lactide) as coating agent on the opsonization, phagocytosis, and pharmacokinetics of biodegradable nanoparticles. Biomacromolecules 2009;10:408–16.
[30] Bosker WTE, Iakovlev PA, Norde W, Cohen Stuart MA. BSA adsorption on bimodal PEO brushes. J Colloid Interface Sci 2005;286:496–503.
[31] Roach P, Farrar D, Perry CC. Surface tailoring for controlled protein adsorption: effect of topography at the nanometer scale and chemistry. J Am Chem Soc 2006;128:3939–45.
[32] Norde W, Giacomelli CE. BSA structural changes during homomolecular exchange between the adsorbed and the dissolved states. J Biotechnol 2000;79:259–68.
[33] Yampolskaya G, Platikanov D. Proteins at fluid interfaces: adsorption layers and thin liquid films. Adv Colloid Interface Sci 2006;128-130:159–83.
[34] Wasilewska M, Adamczyk Z, Jachimska B. Structure of fibrinogen in electrolyte solutions derived from dynamic light scattering (DLS) and viscosity measurements. Langmuir 2009;25:3698–704.
[35] Rosenfeld MA, Leonova VB, Konstantinova ML, Razumovskii SD. Self-assembly of fibrin monomers and fibrinogen aggregation during ozone oxidation. Biochemistry (Mosc) 2009;74:41–6.
[36] Fredslund F, Jenner L, Husted LB, Nyborg J, Andersen GR, Sottrup-Jensen L. The structure of bovine complement component 3 reveals the basis for thioester function. J Mol Biol 2006;361:115–27.
[37] Pbd Id 1E78 Bhattacharya AA, Curry S, Franks NP. Binding of the general anesthetics propofol and halothane to human serum albumin high resolution crystal structures. the protein data bank. J Biol Chem 2000;275:38731–8.
[38] Pbd Id 3GHG Kollman JM, Pandi L, Sawaya MR, Riley M, Doolittle RF. Crystal structure of human fibrinogen. the protein data bank. Biochemistry 2009;48:3877–86.
[39] Pbd Id 2A73 Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, et al. Structures of complement component C3 provide insights into the function and evolution of immunity. the protein data bank. Nature 2005;437:505–11.
[40] Norde W, Lyklema J. The adsorption of human plasma albumin and bovine pancreas ribonuclease at negatively charged polystyrene surfaces: 1 adsorption isotherms: effects of charge, ionic strength, and temperature. J Colloid Interface Sci 1978;66:257–65.
[41] Forcini D, Hamilton WA. Surface enrichment of proteins at quartz/water interfaces: a neutron reflectivity study. J Colloid Interface Sci 2005; 285:458–68.
[42] Amadeu do Serro APV, la Catarino Fernandes A, de Jesus Vieira Saramago B, Norde W. Bovine serum albumin adsorption on titania surfaces and its relation to wettability aspects. J Biomed Mater Res 1999;46:376–81.
[43] Verrecchia T, Huve P, Bazile D, Veillard M, Spenlehauer G, Couvreur P. Adsorption/desorption of human serum albumin at the surface of poly(lactic acid) nanoparticles prepared by a solvent evaporation process. J Biomed Mater Res 1993;27:1019–28.
[44] Oksuka H, Nagasaki Y, Kataoka K. Surface characterization of functionalized polylactide through the coating with heterobifunctional poly(ethylene glycol)/polylactide block copolymers. Biomacromolecules 2000;1:39–48.
[45] Roach P, Farrar D, Perry CC. Interpretation of protein adsorption: surface-induced conformational changes. J Am Chem Soc 2005;127:8168–73.
[46] Shang L, Wang Y, Jiang J, Dong S. pH-dependent protein conformational changes in albumin: gold nanoparticle bioconjugates: a spectroscopic study. Langmuir 2007;23:2714–21.
[47] Wangoo N, Suri R, Shekhawat G. Interaction of gold nanoparticles with protein: a spectroscopic study to monitor protein conformational changes. Appl Phys Lett 2008;92. 133104–133111-133104-3.
[48] Sit PS, Marchand RE. Surface-dependent differences in fibrin assembly visualized by atomic force microscopy. Surf Sci 2001;491:421–32.

Paper VI

Salvatore, Duraffourg, Favier, Persson, Lund, Delage, Silvers, Schwalbe,
Croguennec, Bouhallab and Forge
BioMarcomol (2011) ©

ARTICLE

pubs.acs.org/Biomac

Investigation at Residue Level of the Early Steps during the Assembly of Two Proteins into Supramolecular Objects

Delphine B. Salvatore,[†,‡] Nicolas Duraffourg,[†] Adrien Favier,[§] Björn A. Persson,[∥] Mikael Lund,[∥] Marie-Madeleine Delage,[‡] Robert Silvers,[⊥] Harald Schwalbe,[⊥] Thomas Croguennec,[‡] Saïd Bouhallab,[‡] and Vincent Forge*[,†]

[†]Laboratoire de Chimie et Biologie des Métaux (UMR 5249), CEA−Grenoble, 17, rue des Martyrs, F-38057 Grenoble, France

[‡]UMR 1253, INRA, Agrocampus Ouest, Science et Technologie du Lait et de l'Oeuf, 65 rue de Saint-Brieuc, F-35000 Rennes, France

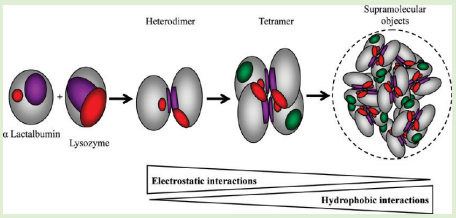
[§]Laboratoire de Résonance Magnétique Nucléaire, Institut de Biologie Structurale Jean-Pierre Ebel, 41, rue Jules Horowitz, F-38027 Grenoble, France

[∥]Department of Theoretical Chemistry, University of Lund, POB 124, S-22100 Lund, Sweden

[⊥]Institute for Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe University, Max-von-Laue-Strasse 7, D-60438 Frankfurt am Main, Germany

Supporting Information

ABSTRACT: Understanding the driving forces governing protein assembly requires the characterization of interactions at molecular level. We focus on two homologous oppositely charged proteins, lysozyme and α-lactalbumin, which can assemble into microspheres. The assembly early steps were characterized through the identification of interacting surfaces monitored at residue level by NMR chemical shift perturbations by titrating one ^{15}N-labeled protein with its unlabeled partner. While α-lactalbumin has a narrow interacting site, lysozyme has interacting sites scattered on a broad surface. The further assembly of these rather unspecific heterodimers into tetramers leads to the establishment of well-defined interaction sites. Within the tetramers, most of the electrostatic charge patches on the protein surfaces are shielded. Then, hydrophobic interactions, which are possible because α-lactalbumin is in a partially folded state, become preponderant, leading to the formation of larger oligomers. This approach will be particularly useful for rationalizing the design of protein assemblies as nanoscale devices.



INTRODUCTION

Interactions between proteins are ubiquitous in biological processes, and their understanding remains one of the central purposes of biology, biochemistry, and biophysics. Protein−protein interactions can lead to self-assembly within ordered aggregates. Nowadays, the use of protein self-assembly for the design of nanoscale devices receives an increasing interest.[1] Supramolecular objects resulting from protein assembly have been intensively characterized at macroscale.[2−5] Their morphology can be controlled by monitoring the physicochemical conditions. Linear and fibrillar assembly, such as amyloïd fibrils, are favored at pHs far from the protein isoelectric point (pI) and at low ionic strength when electrostatic repulsions are high. Contrarily, spherical objects are obtained by incubation of proteins at a pH close to their pI and at ionic strength favoring electrostatic interactions.[6−8] The assembling properties of proteins are controlled by changing not only their physicochemical properties, but also their conformational state. Indeed, it is well-known that partial denaturation of proteins is a key parameter favoring protein assembly.[9] In some cases, structural changes, such as intermolecular β-sheets formation in amyloid fibrils, are also involved; intermolecular β-sheets formation is likely to be responsible for protein assembly in a preferential direction leading to linear organization. Although plenty of studies have been performed to characterize the resulting supramolecular objects at the macroscale, few studies are available to understand how information contained at the molecular level spreads on the macroscale. In this study we propose to get insight into interactions at the molecular scale driving the organization of proteins into spherical supramolecular objects.

We focus on the interactions and self-assembly in a binary protein system involving bovine α-lactalbumin (LAC) and hen egg white lysozyme (LYS), two globular proteins of about 14 kDa and composed of 123 and 129 amino acids, respectively. They

Received: March 1, 2011
Revised: May 5, 2011
Published: May 05, 2011

ACS Publications © 2011 American Chemical Society

share a sequence identity of 35−40%, and their three-dimensional structures are similar, both composed of two domains separated by a cleft. Their α-domains contain four α-helices and the β-domains are composed of an antiparallel three-stranded β-sheet and a flexible loop. Their structures are stabilized by four disulfide bridges. Despite these similarities, LAC and LYS exhibit distinct biological function:[10] LAC is the regulatory subunit of the enzyme lactose synthase and LYS is a catalytic enzyme involved in the hydrolysis of Gram-positive bacterial cell wall peptidoglycans. These proteins differ also in their metal ion binding ability as well as their molecular stability. Egg white LYS belongs to the C-type family without a specific metal binding site, whereas LAC contains a calcium binding site located between the two domains. The binding constant is 10^{-7}−10^{-8} M and binding of calcium stabilizes the tertiary structure of LAC.[11,12] The denaturating temperature is 64 °C for the calcium-loaded form, holo α-lactalbumin (hLAC), and 27 °C for the calcium-depleted form, apo α-lactalbumin (aLAC).[13] Above this melting point, aLAC adopts a molten globule state characterized by the presence of native secondary structure but no well-defined tertiary structure due to high side-chain flexibility.[14−16] The stability of LYS is relatively high, with no significant unfolding at acidic pH, and its denaturing temperature is 74 °C.[17,18] A further important difference is that LAC is an acidic protein with a pI of 4.5, whereas LYS is a basic protein with a pI of 10.7. As a consequence, their charge is opposite at neutral pH. Their coprecipitation has been reported and electrostatic interactions have been evidenced.[19,20]

Scheme 1. Reaction Scheme of Dimer and Tetramer Formation with the Expression of the Corresponding Dissociation Constants (K_{ddimer} and $K_{dtetramer}$)

$$\text{aLAC} + \text{LYS} \underset{}{\overset{\text{Kd}_{\text{DIMER}}}{\rightleftharpoons}} \text{aLAC-LYS} \qquad Kd_{DIMER} = \frac{[aLAC].[LYS]}{[aLAC-LYS]}$$

$$2\ \text{aLAC-LYS} \underset{}{\overset{\text{Kd}_{\text{TETRAMER}}}{\rightleftharpoons}} (\text{aLAC-LYS})_2 \qquad Kd_{TETRAMER} = \frac{[aLAC-LYS]^2}{[(aLAC-LYS)_2]}$$

Electrostatic interactions occur between LYS and both forms of LAC.[20−23] With hLAC, only heterodimers are formed. By contrast, the interactions between LYS and aLAC lead to the formation of either amorphous aggregates or supramolecular ordered objects, depending on protein concentration, pH, ionic strength, and temperature. Densely packed microspheres of 1−5 μm diameter, in which both proteins are homogenously distributed in an equimolar ratio, are formed at low ionic strength, pH 7.5, and 45 °C.[24] The aLAC-LYS heterodimer has been proposed to constitute the building block for such self-assembly.

Although the assembly resulting from the interactions between aLAC and LYS has been investigated on a macroscale, no detail on protein−protein interface is available yet. The purpose of the present study is to determine the amino acids involved in the interactions at the early steps of the assembly mechanism leading to microspheres. The interface residues were determined according to the chemical shift perturbations (CSPs) monitored by nuclear magnetic resonance (NMR) during titration of one ^{15}N-labeled protein by its unlabeled partner. This approach has been widely adopted to establish the three-dimensional structure of biological complexes[25−28] and to screen for high affinity protein ligands.[29−31] At the temperature favorable to microsphere formation (45 °C), aLAC adopts a molten globule state that prevents the use of NMR to determine its interacting surface.

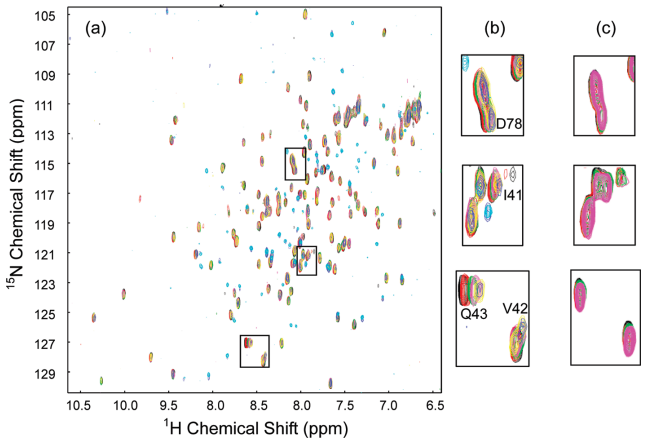


Figure 1. (a) Overlay of ^{1}H−^{15}N HMQC spectra recorded during the titration of 10 μM ^{15}N-hLAC by unlabeled LYS, in 30 mM Tris-HCl 15 mM NaCl at pH 7.5 and 45 °C. The reference spectrum of ^{15}N-hLAC is colored in black and the spectra obtained at protein ratio [unlabeled LYS]/[^{15}N-hLAC] equal to 5 in red, 20 in green, 30 in magenta, 50 in yellow, and 100 in blue. The ^{1}H−^{15}N HMQC spectrum of unlabeled LYS (light blue) was recorded through detection of ^{15}N natural abundance. (b) Highlights of spectral regions with shifting crosspeaks. (c) Highlights of similar spectral regions during the control titration. The reference spectrum of ^{15}N-hLAC is colored in black and the spectra obtained at protein ratio [unlabeled hLAC]/[^{15}N-hLAC] equal to 1 in red, 15 in green, and 30 in magenta.

In the case of the aLAC-LYS dimer, only the interacting region of LYS could be monitored. To get insight into the interacting region of LAC, the interfaces within hLAC-LYS have also been investigated, although these dimers do not further assemble. Mapping of interacting residues on protein surfaces reveals that hLAC interacts with LYS through a narrow surface composed of amino acids located in the β-domain. The interacting surface of LYS is much more scattered, with amino acids belonging to both α- and β-domains. These findings correlate with predictions by molecular simulations and provide experimental evidence for an orientational effect due to electrostatic interactions. The interacting region of LYS with aLAC is similar to that obtained with hLAC. However, additional residues are involved in a cooperative manner. This is evidence of a second binding event. We interpret this effect as the further assembly of heterodimers into tetramers. Because most of the electrostatic charge patches are shielded within the tetramers, the hydrophobic interactions, which are possible because aLAC is in a partially folded state, become preponderant and lead to the formation of larger oligomers.

MATERIALS AND METHODS

Protein Purification. Commercial lysozyme (LYS) was purchased from Ovonor. Holo α-lactalbumin (hLAC) was purified from bovine whey according to the method described by Caussin et al.[32] Apo α-lactalbumin (aLAC) was obtained by removing the calcium by dialysis against water at pH 3.[24] Recombinant uniformly ^{15}N-labeled LYS and ^{15}N-labeled hLAC were produced and purified as previously reported.[33,34]

NMR Experiments. All NMR spectra were obtained using a VARIAN vnmrs 800 spectrometer equipped with a triple-resonance (^{1}H, ^{13}C, ^{15}N) cold probe. Two-dimensional ^{1}H−^{15}N heteronuclear multiple quantum coherence (HMQC) spectra were acquired using the band-Selective Optimized Flip-Angle Short-Transient (SOFAST) pulse sequence,[35] using 128 scans and a recovery delay of 200 ms, leading to an experiment duration of 54 min. For the ^{1}H and ^{15}N dimensions, 537 and 50 complex points and spectral widths of 13440 and 2400 Hz were used, respectively. Zero fillings were applied in both dimensions giving rise to 1536 × 200 points matrices. All spectra were processed using NMRPipe[36] and analyzed using NMRView.[37]

Stock solutions of ^{15}N-labeled proteins (0.1 mM) and unlabeled proteins (2 mM) were prepared by solubilization of protein powder in 30 mM Tris-HCl buffer containing 15 mM NaCl and 10% (v/v) D_2O, pH 7.5. They were filtered through a 0.2 μm membrane. The protein concentration was determined by measuring the absorbance at 280 nm using extinction coefficients of 2.64 $L\cdot g^{-1}\cdot cm^{-1}$ for LYS and 2.01 $L\cdot g^{-1}\cdot cm^{-1}$ for hLAC and aLAC.

The titrations experiments were carried out at 45 °C. Once the ^{1}H−^{15}N HMQC reference spectrum of the ^{15}N-labeled protein was recorded, the progress of the titration was monitored by recording two-dimensional ^{1}H−^{15}N HMQC spectrum after each addition of its unlabeled partner. The interaction surface of hLAC with LYS was determined by titrating 10 μM of ^{15}N-labeled hLAC with increasing amount of unlabeled LYS giving protein molar ratios LYS/^{15}N-hLAC of 1, 5, 10, 20, 30, 50, 75, and 100. The reverse titrations of ^{15}N-labeled LYS by unlabeled hLAC (titration points corresponding to hLAC /^{15}N-LYS of 1, 5, 10, 20, 30, 40, 50, 60, and 70) and ^{15}N-labeled LYS by unlabeled aLAC (aLAC /^{15}N-LYS of 1, 5, 10, 20, 30, 40, and 50) were performed to determine the interaction surface of LYS with hLAC and aLAC, respectively. To assess nonspecific CSPs, the control titrations of ^{15}N-hLAC by unlabeled hLAC and ^{15}N-LYS by unlabeled LYS were performed in the same way for few titration points (unlabeled protein/^{15}N-labeled protein of 1, 15, and 30).

Data Analysis. For each protein molar ratio, unlabeled protein/^{15}N-labeled protein, a combined chemical shift $\Delta\delta$ (ppm) was calculated per amino acid according to eq 1:[38]

$$\Delta\delta(\text{ppm}) = \sqrt{(\Delta\delta^{1}\text{H})^2 + \left(\frac{\gamma_N}{\gamma_H}(\Delta\delta^{15}\text{N})\right)^2} \quad (1)$$

where $\Delta\delta^{1}$H and $\Delta\delta^{15}$N are the variations of chemical shift compared to the reference spectrum in ^{1}H and ^{15}N dimensions, respectively, and γ_N and γ_H are the magnetogyric ratio (rad/T·s) of ^{1}H and ^{15}N atoms, respectively.

A simple equilibrium between two globular proteins in solution and a heterodimer was assumed for calculating the dissociation constant (K_d) and the combined chemical shift at saturation ($\Delta\delta_{max}$) of logarithmic dependencies with eq 2.[39]

$$\Delta\delta(\text{ppm}) = \left(\frac{\Delta\delta_{max}}{2[P]_t}\right)([P]_t + [L]_t + K_d - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}) \quad (2)$$

where $[P]_t$ and $[L]_t$ are the total concentration of ^{15}N-labeled protein and unlabeled protein partner, respectively.

When the evolution of $\Delta\delta$ obeys to a sigmoid behavior, the Hill equation[40] (eq 3) was used for a first estimation of the apparent K_d and the $\Delta\delta_{max}$

$$\Delta\delta(\text{ppm}) = \frac{\Delta\delta_{max}[L]^n}{K_d{}^n + [L]^n} \quad (3)$$

where n is the Hill coefficient.

When considering the association of two heterodimers into tetramer (Scheme 1), the logarithmic and sigmoid dependencies were fitted with eqs 4 and 5, respectively:

$$\Delta\delta(\text{ppm}) = \frac{[\text{PL}] + 2[(\text{PL})_2]}{[P]_t}.\Delta\delta_{max} \quad (4)$$

$$\Delta\delta(\text{ppm}) = \frac{2[(\text{PL})_2]}{[P]_t}.\Delta\delta_{max} \quad (5)$$

where [PL] and $[(PL)_2]$ are the concentrations of dimer and tetramer, respectively, expressed as a function of the corresponding K_ds (Kd_{DIMER} and $Kd_{TETRAMER}$):

$$[\text{PL}] = \frac{[P].[L]}{Kd_{DIMER}} \text{ and } [(\text{PL})_2] = \frac{[\text{PL}]^2}{Kd_{TETRAMER}} = \frac{[P]^2.[L]^2}{Kd_{DIMER}{}^2.Kd_{TETRAMER}}$$

where [P] and [L] are the free concentrations of ^{15}N-protein and unlabeled protein partner at equilibrium, respectively. $[P]$ is expressed by eq 6, assuming that the amount of bound unlabeled protein is negligible as compared to its total concentration (i.e., $[L]_t = [L]$):

$$[P] = \left(\frac{Kd_{DIMER}{}^2.Kd_{TETRAMER}}{2[L]_t{}^2}\right).\left(-\left(1 + \frac{[L]_t}{Kd_{DIMER}}\right) + \sqrt{\left[\left(1 + \frac{[L]_t}{Kd_{DIMER}}\right)^2 + 4[P]_t.\frac{[L]_t{}^2}{Kd_{DIMER}{}^2.Kd_{TETRAMER}}\right]}\right) \quad (6)$$

Molecular Simulation. Mesoscopic Monte Carlo simulations in the canonical ensemble were carried out using the Faunus software framework.[41] The mutual spatial distribution function, directly related to the interaction free energy, was sampled for a pair of LYS and aLAC/hLAC model proteins at pH 7.5 and a salt concentration of 39 mM. Each protein is coarse grained to the amino acid level, based on experimentally

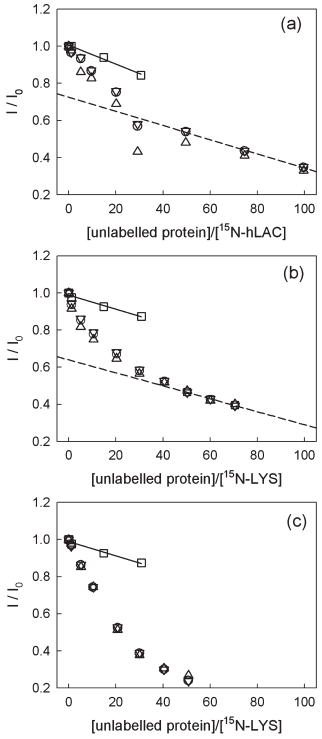


Figure 2. Total peak intensity changes during the titrations of ^{15}N-hLAC by LYS (a), ^{15}N-LYS by hLAC (b), and ^{15}N-LYS by aLAC (c). The total peak intensity (circle) was normalized to reference spectra and split into two groups: the total peak intensity of shifting (triangle up) and nonaffected (triangle down) peaks. The total peak intensity changes during the control titrations (square) of ^{15}N-hLAC by hLAC (a) and ^{15}N-LYS by LYS (b,c) are also depicted. Continuous lines correspond to linear regressions.

determined crystal structures using the pdb entries 4LZT (LYS), 1F6S (hLAC), and 1F6R (aLAC). Proteins, counterions, and salt particles are immersed in a spherical simulation cell containing a continuum solvent described by the dielectric constant of water, $\varepsilon_r = 80$.

RESULTS AND DISCUSSION

Chemical Shift Perturbation Approach. Interacting surfaces were identified according to CSPs measured during the titration of one ^{15}N-labeled protein by its unlabeled partner. The use of band-Selective Optimized Flip-Angle Short-Transient (SOFAST) pulse sequence[35] enabled us to obtain well-resolved spectra in a reasonable time, even when working at 10 μM protein concentration in order to limit protein assembly. Only, the interaction surface of aLAC with LYS can not be determined using the CSP method. Indeed, the spectrum of ^{15}N-labeled aLAC shows a weak dispersion and broad resonances which are characteristic of the

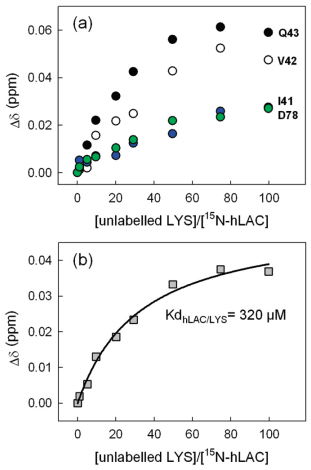


Figure 3. (a) Combined chemical shift ($\Delta\delta$) of ^{15}N-hLAC significant residues: Q43 (black), V42 (white), I41 (blue), and D78 (green), at various ratios of unlabeled LYS. (b). Averaged $\Delta\delta$ changes (gray squares): curve fitted with eq 2 (see text; continuous line). The residues used for the averaging are listed in Table 1.

molten globule state at a temperature of 45 °C (Figure S1).[42] The reference ^{1}H−^{15}N HMQC spectra of ^{15}N-labeled hLAC and ^{15}N-labeled LYS are well-resolved and show approximately 100 backbone amide cross-peaks, which can be used to follow the titrations in detail (Figure 1). Changes in cross-peak chemical shifts reflect changes in the local environment of the ^{15}N-labeled protein backbone amides and therefore identify residues involved in the interaction with the unlabeled partner.[29]

The approach can be illustrated with the titration of ^{15}N-labeled hLAC by unlabeled LYS. Figure 1a shows an overlay of the full spectra obtained after different additions of unlabeled LYS. The NMR spectra obtained during the titrations of ^{15}N-labeled LYS by both forms of LAC are depicted in the Supporting Information (Figures S2 and S3). Examples of significantly changed resonances are depicted in Figure 1b. One can note that the crosspeak assigned to V42 is superimposed with another crosspeak within the reference spectrum and diverges upon titration. The observed chemical shift changes during titration can be attributed to the interactions between hLAC and LYS. Indeed, no significant shift is observed during the control titration, that is, ^{15}N-labeled protein titrated by its unlabeled equivalent (Figure 1c), indicating that no homodimerization occurs. High concentrations of unlabeled protein are added at the end of the titration and, at that stage, cross peaks of this protein appears due to the natural abundance of ^{15}N (Figure 1a; blue lines). These additional cross peaks overlap sometime with those of the titrated ^{15}N-protein and induces "false positives". The uncertain cross peaks are removed from the study (Table S1). The amino acids that are affected upon titration are characterized by one single cross peak, which is shifting together with an intensity decrease.

Table 1. Dissociation Constants (K_d) and Standard Errors Estimated by Fitting the Titration Curves Resulting from the Averaged $\Delta\delta$ Changes of Listed Residues with Either Logarithmic or Sigmoid Dependencies (see Text)[a]

dimer	observed protein	residues	K_d (μM)
hLAC-LYS	^{15}N-hLAC	T4, T38, V42, Q43, D78	$K_{dhLAC/LYS}$ = 320 ± 50
	^{15}N-LYS	A32, S36, N77, A122	$K_d^1{}_{LYS/hLAC}$ = 293 ± 52
		G4, A9, I98, N103	$K_d^2{}_{LYS/hLAC}$ = 725 ± 185
aLAC-LYS	^{15}N-LYS	A9, F34, N37, N77	$K_d^L{}_{LYS/aLAC}$ = 390 ± 90
		E35, R125, C127, L129	$K_d^S{}_{LYS/aLAC}$ = 196 ± 30 Hill coefficient = 2

[a] $K_{dhLAC/LYS}$ and $K_{dLYS/hLAC}$ were obtained by fitting the data obtained by titrating ^{15}N-hLAC by unlabelled LYS and ^{15}N-LYS by unlabelled LAC, respectively. In the latter, two distinct dependencies were observed (K_d^1 and K_d^2). $K_d^L{}_{LYS/aLAC}$ and $K_d^S{}_{LYS/aLAC}$ by fitting the titration curves obtained by titrating ^{15}N-LYS by unlabelled aLAC for either logarithmic (L) or sigmoid (S) dependencies.

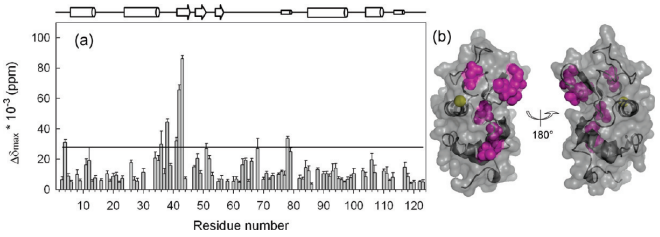


Figure 4. (a) Combined chemical shift changes at saturation ($\Delta\delta_{max}$) as a function of residue rank in the primary structure of hLAC. The $\Delta\delta_{max}$ values were estimated by fitting with eq 2 and $K_{dhLAC/LYS}$ imposed at 320 μM, the $\Delta\delta$ changes monitored during the titration of ^{15}N-hLAC by unlabeled LYS for every considered residue. The error bars correspond to the standard error from the fit. The threshold value (horizontal line) corresponds to $\Delta\delta$ calculated with $\Delta\delta^1$H = 0.02 ppm and $\Delta\delta^{15}$N = 0.2 ppm and is equal to 28 × 10^{-3} ppm. The secondary structures regions of hLAC are depicted schematically above the panel. (b) Localization of residues in interaction with LYS in a ribbon representation of hLAC (generated with Pymol). The residues with $\Delta\delta_{max}$ above the threshold are colored in purple; the calcium ion is depicted as a yellow hard sphere. The structures are represented after a 180° rotation around the vertical axis.

There are two possible reasons for the intensity decrease because of signal broadening; either the exchange between dimer and monomers is on an intermediate time-scale for NMR or the relaxation time changes due to an increase in the apparent protein weight upon dimer formation.[43,44] The total spectrum intensities, that is, the sum of all peak intensities, normalized to that of the reference spectrum, are reported as a function of the protein ratio in Figure 2. While only a linear decrease due the non-negligible dilution of ^{15}N-labeled proteins (from 10 to 8 μM) upon addition of its unlabeled partner is detected for the control experiments, a much larger decrease occurs during the heterodimer titration. Considering the titrations involving hLAC and LYS (Figure 2a,b), half of the initial peak intensity is lost at a protein ratio around 50. However, at a higher protein ratio, the dependence becomes linear, with a slope similar to that of the control experiment. For the three types of titration, there is no difference in intensity decrease when considering shifting and nonaffected cross-peaks independently; the peak intensity decrease is not specific to interacting residues, but affects the whole protein. This differs from observations made during the early steps of amelogenin's supramolecular assembly.[45] In that case, the dimer formation on an intermediate time-scale for NMR, and the cross peak intensities decrease only for the amino acids involved in the interactions. In our case, the signal broadening results from molecular weight increase of the formed complexes, rather than from the time-scale of the heterodimer formation. The dimer formation induces mostly a chemical shift change for the residues involved in the interaction, without additional signal broadening compared to that induced by the apparent weight increase. As a consequence, one can reasonably assume that the heterodimer formation is fast on a NMR time scale, that is, faster than the millisecond order of magnitude.[43]

One can notice that the intensity decrease was found to be much more pronounced for the titration involving LYS and the apo form of LAC (Figure 2c). About 80% of the initial intensity was lost with a protein ratio of 20. Because of the substantial loss of NMR signal, the titration of ^{15}N-LYS by aLAC was not performed for protein ratios higher than 50. This additional decrease of intensity reflects the appearance of larger oligomers upon the formation of dimers.

HOLOLAC-LYS Dimer. Only few cross-peaks are perturbed with increasing protein ratio LYS/^{15}N-hLAC (Figure 1). The variations of their combined chemical shifts ($\Delta\delta$ calculated according to eq 1, see Materials and Methods) are shown in Figure 3a for selected examples. Despite rather small chemical shift changes (Figure 1b), which suggest weak interactions,[46] clear tendencies are observed on the $\Delta\delta$ dependencies with the protein ratio. A plateau is reached for a protein ratio larger than 100; this corresponds also to the boundary of the nonlinear decrease of the peak intensities (Figure 2a), which is attributed to the heterodimer formation (see above). $\Delta\delta$ of significantly shifting amino acids (Table 1) were averaged, and a dissociation constant ($K_{dhLAC/LYS}$) of 320 ± 50 μM was determined by fitting the resulting titration curve with eq 2 (see Materials and Methods; Figure 3b). This K_d obtained by the titration of 10 μM ^{15}N-labeled protein is around 20-fold higher than the one

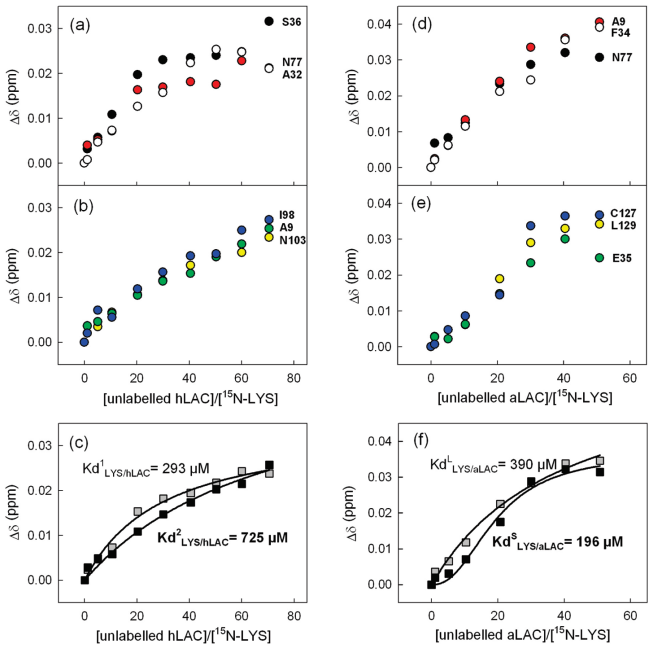


Figure 5. Combined chemical shift ($\Delta\delta$) changes of ^{15}N-LYS significant residues at various ratio of unlabeled hLAC (left panels) and unlabeled aLAC (right panels): (a) S36 (black), N77 (white), and A32 (red); (b) I98 (blue), A9 (green), and N103 (yellow). (c) Averaged $\Delta\delta$ changes, curve fitted with eq 2 (see text; continuous lines): (d) A9 (red), N77 (black), and F34 (white); (e) C127 (blue), L129 (yellow), and E35 (green). (f) Averaged $\Delta\delta$ changes of the residues with clear logarithmic behaviors (see Table 1; gray squares), averaged $\Delta\delta$ changes of the residues with clear sigmoid behaviors (black squares). The continuous lines correspond to the results of the fitting with either a logarithmic or a sigmoid law (see text).

previously determined in the same conditions by fluorescence anisotropy by titrating 1 μM protein.[22] This apparent discrepancy is likely to be due to the fact that the fluorescence anisotropy method does not provide access to dissociation constants of magnitude orders larger than the protein concentration giving rise to a nonspecific signal, that is, 10 μM. At higher protein concentration, artifacts occur that impair the fluorescence measurements.[22] Furthermore, the hydrophobic fluorescent probe covalently bound to the proteins may increase protein affinity through nonspecific hydrophobic interactions. Overall, both features of the dimers, weak interaction and fast exchange, strongly suggest the transient character of the complexes.[47]

It is noteworthy that, because of weak affinity, about 68% of ^{15}N-hLAC is bound to LYS at the end of the titration and, thus, $\Delta\delta$ of the fully bound state is not observed. Therefore, the combined chemical shift at saturation ($\Delta\delta_{max}$) of every residue was estimated by fitting their titration curves with eq 2 and $K_{dhLAC/LYS}$ fixed at 320 μM. Figure 4a represents the pattern of $\Delta\delta_{max}$ changes as a function of the amino acid sequence of hLAC. Note that missing values correspond to nonassigned amino acids and false positives. The amino acids exhibiting a $\Delta\delta_{max}$ larger than a threshold value arbitrarily chosen for highlighting the most significantly changed amino acids (see legend of Figure 4a) are listed in Table S2. The same threshold will be used throughout this study. The amino acids of hLAC sensitive to the interaction with LYS are mainly located within in the N-terminal part of the β-sheet. A few others are found at the N-terminal part of the protein and within the long loop of the β-domain. As illustrated in Figure 4b on the crystal structure of hLAC, these residues are located on one side of the protein. One can notice here that, although most of the significantly shifted residues are solvent accessible (Table S2), some of them are not exposed at the protein surface. However, their environment must be influenced by local conformational changes because of their proximity with exposed amino acids involved in the interaction.[48]

During the reverse titration, $\Delta\delta$ changes of significantly shifting residues with increasing ratio hLAC/^{15}N-LYS are shown in Figure 5a,b. In that case, two classes of interacting amino acids can be distinguished. Some residues, as illustrated with A32, S36, and N77 (Figure 5a), exhibit steeper titration curves as compared to others, such as A9, I98, and N103 (Figure 5b). Average dependencies have been calculated from each data set (Figure 5c and Table 1) and two dissociation constants ($K_d^1{}_{LYS/hLAC}$ and $K_d^2{}_{LYS/hLAC}$) have been determined. The dissociation constant corresponding to the steeper dependence, $K_d^1{}_{LYS/hLAC}$ (293 $\pm$ 52 μM), is similar to

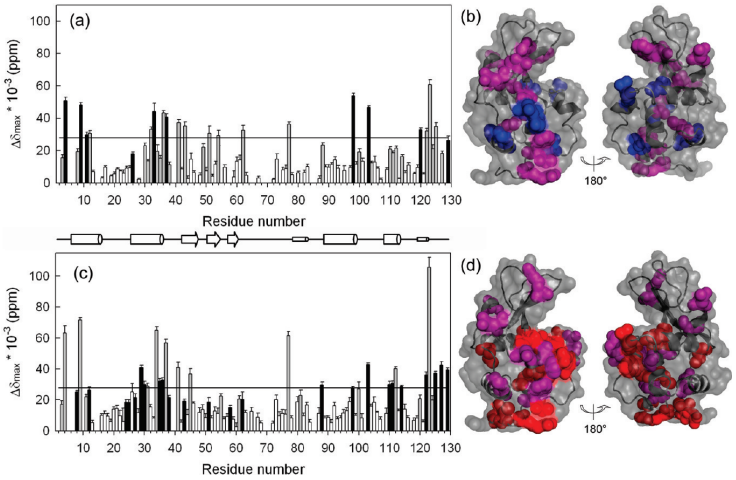


Figure 6. (a) Combined chemical shift changes at saturation ($\Delta\delta_{max}$) as a function of residue rank in the primary structure of LYS, upon titration by hLAC. The $\Delta\delta_{max}$ values were estimated by fitting the $\Delta\delta$ changes with eq 2 and K_d imposed to either 293 μM ($K_d^1{}_{LYS/hLAC}$; gray bars) or 725 μM ($K_d^2{}_{LYS/hLAC}$; black bars). When the data quality was not high enough to choose between the two possible Kds, the curve fitting was arbitrarily performed with $K_d^1{}_{LYS/hLAC}$ (white bars). (b) Localization of residues in interaction with hLAC in a ribbon representation of LYS. The residues with $\Delta\delta_{max}$ above the threshold (28×10^{-3} ppm) are colored either in purple or in blue when their changes obey to $K_d^1{}_{LYS/hLAC}$ or $K_d^2{}_{LYS/hLAC}$, respectively The structures are represented after a 180° rotation around the vertical axis. (c) Combined chemical shift changes at saturation ($\Delta\delta_{max}$) as a function of residue rank in the primary structure of LYS, upon titration by aLAC. The $\Delta\delta_{max}$ values were estimated by fitting the $\Delta\delta$ changes with either a logarithmic law or a sigmoid law and K_d imposed to 390 μM ($K_d^L{}_{LYS/aLAC}$; black bars) or 196 μM ($K_d^S{}_{LYS/aLAC}$ and $n = 2$; gray bars). When the data quality was not high enough to choose between the two possible K_ds, the curve fitting was arbitrarily performed with $K_d^L{}_{LYS/aLAC}$ (white bars). (d) Localization of residues in interaction with aLAC in a ribbon representation of LYS. The residues with $\Delta\delta_{max}$ above the threshold are colored either in purple or in red when their changes obey to a logarithmic or a sigmoid law, respectively.

$K_{dhLAC/LYS}$, whereas $K_d^2{}_{LYS/hLAC}$ (725 ± 185 μM) is larger. Figure 6a shows the pattern of $\Delta\delta_{max}$ changes as function of the amino acid sequence of LYS. The values of $\Delta\delta_{max}$ were estimated by fitting the chemical shift change of each amino acid with both values of K_d and the values provided by the best fit (i.e., giving the higher correlation coefficient) were reported: gray bars with $K_d^1{}_{LYS/hLAC}$ and black bars with $K_d^2{}_{LYS/hLAC}$. When compared to the pattern of hLAC (Figure 4a), the amino acids with significant chemical shift changes are scattered all along the sequence and the largest $\Delta\delta_{max}$ is about half. One can notice here that the extent of chemical shift change seems related to the broadness of the distribution of interaction sites. As illustrated in Figure 6b, the interacting residues of LYS (Table S2) are scattered on a larger surface than is the case for hLAC (Figure 4b). Notice, however, that the residues exhibiting a higher K_d value (colored in blue) are all located in the α-domain.

According to previous studies, electrostatic interactions are preponderant for the heterodimer formation.[20,23] Positively charged patches on the LYS surface are expected to interact with negatively charged patches on the LAC surface. LYS possesses mostly positive charges which are rather homogenously distributed on its surface (Figure 7a). The surface electric charge distribution of LAC is uneven (Figure 7b), with a face containing negative charges only and another one with a mix of negative and positive charges. This gives LAC a strong dipolar moment (400 D), whereas the dipole moment of LYS is 2−3 times weaker (140 D).[23] The probability to find one protein around the other was estimated by Monte Carlo simulations at pH 7.5 and a salt concentration of 39 mM, with the same approach as previously used for investigating the electrostatic interactions between LAC and LYS at lower ionic strength.[23] The two proteins and salt particles are placed within a simulation cell containing a continuum solvent described by the dielectric constant of water ($\varepsilon = 80$). The model explicitly incorporates van der Waals interactions as well as continuum electrostatic interactions between ionized residues while allowing for a continuum exploration of Boltzmann averaged interprotein orientations. The isodensity surfaces represent the spatial distribution, that is, all the possible locations, of protein residues when interacting with its partner during the trajectories drawn from Monte Carlo simulations (Figure 8). The green isodensity surface around hLAC shows that LYS is steered toward the negative pole of the dipolar moment of hLAC (Figure 8a). This area matches very well with the interacting surface of hLAC determined experimentally (Figure 8a, purple side chains). The distribution of hLAC residues around LYS (orange isodensity surface) is more scattered with a preference for the equator of the egg-shape LYS (Figure 8b). This distribution correlates very well with the more diffuse interacting surface of LYS with hLAC determined by NMR (Figure 8b, purple and blue side chains). These results

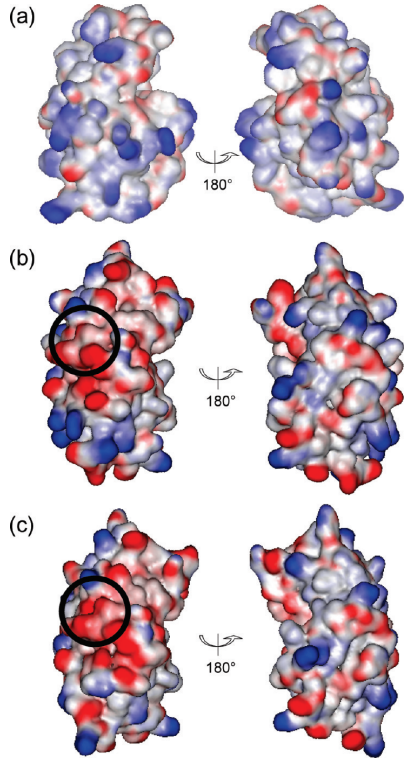


Figure 7. Electrostatic surfaces of LYS (a), hLAC (b), and aLAC (c). The electrostatic potential was calculated with Delphi at 39 mM ionic strength and pH 7.5. Negative and positive surface patches are represented in red and blue, respectively. Proteins are represented in the same orientation as in Figures 4 and 6 and after rotation of 180° around the vertical axis. The calcium binding site of LAC is surrounded by the black circle. The pictures were generated with Accelrys.

evidence that the extent of preferential orientation of protein is related to the dipolar moment strength and that heterodimers exhibit a relatively low specificity of binding. The residues on LYS surface exhibiting a higher K_d value are probably involved in the less specific contacts among all the possible configurations of the heterodimers.

$_{APO}$LAC-LYS Dimer. The interactions between LYS and the apo form of LAC lead to the formation of heterodimers and then to larger oligomers. This further oligomerization results in the formation of supramolecular structures.[22,24] Under the conditions propitious to microspheres formation (45 °C), aLAC is in a partially folded state, the so-called molten globule state. As a consequence, it is not possible to monitor its interacting surface. Nevertheless, the spectra obtained during the titration performed for few unlabeled LYS/^{15}N-aLAC ratios do not become more resolved (Figure S1). Therefore, aLAC remains in its molten globule state and its interaction with LYS does not lead to stabilization of its tertiary structure.

During the titration of ^{15}N-LYS by aLAC, two different types of $\Delta\delta$ dependencies with protein ratio are observed: a classical logarithmic one (Figure 5d) and a sigmoid dependence with a lag-phase at low protein ratio (Figure 5e). The sigmoid dependencies denote a cooperative binding. These residues are involved in the interaction with aLAC in a sequential manner; they are not involved in the first binding event but in a second one. As a first approach, no assumption is made about the reaction scheme of the events monitored during this titration; averages of each data set (Figure 5f) were fitted with either eq 2, which corresponds to a simple binding, or the Hill equation (eq 3 in Materials and Methods) for the sigmoid dependencies, and two dissociation constants ($K_d^{L}{}_{LYS/aLAC}$ and $K_d^{S}{}_{LYS/aLAC}$) were calculated (Table 1). The value of $K_d^{L}{}_{LYS/aLAC}$ (390 ± 90 μM) is identical within the error limit to that obtained for the hLAC-LYS dimer ($K_{dhLAC/LYS}$ and $K_d^{1}{}_{LYS/hLAC}$). For the sigmoid dependencies, $K_d^{S}{}_{LYS/aLAC}$ is lower, with a value equal to 196 ± 30 μM and a Hill coefficient around 2. The pattern of $\Delta\delta_{max}$ is shown in Figure 6c: gray and black bars when best fitted with $K_d^{L}{}_{LYS/aLAC}$ and $K_d^{S}{}_{LYS/aLAC}$, respectively. This pattern shows some similarities with that obtained in the presence of hLAC (Figure 6a). However, larger chemical shift changes are observed and the majority of the residues with $\Delta\delta_{max}$ larger than the defined threshold belong to the α-domain (Figure 6c and Table S2). All the residues with a sigmoid behavior are localized within the α-domain (Figure 6d, colored in red) and they define a region around an interaction site defined by residue with logarithmic behavior (Figure 6d, colored in purple). One can notice that the $\Delta\delta_{max}$ estimated with the sigmoid equation (Figure 6c; black bars) are on average half those obtained with the logarithmic equation (Figure 6c; gray bars).

The cooperative dependencies of several residues imply that at least two events of the interactions between aLAC and LYS are monitored, these residues being sensitive to the second event only. A straightforward possibility is that the heterodimers further assemble into tetramers. This can be described by the sequential reaction (Scheme 1). The residues involved in the interaction as the first binding event begins, that is, dimer formation, show a logarithmic-like behavior (Figure 5d,f) and those only involved in the second event, that is, the further assembly of dimer into tetramer, have a sigmoid behavior (Figure 5e,f). By considering these two equilibriums, the averaged titration curves (Figure 5f) are well fitted with the eqs 4 and 5 (Materials and Methods; Figure 9). From these results, the dimer dissociation constant (K_{dDIMER}) is estimated to 490 ± 70 μM and the one for tetramer formation ($K_{dTETRAMER}$) to 5.0 ± 2.5 μM. Despite the quite high uncertainty on the fit parameter values, it is possible to reproduce the general features of the dependencies. The behavior of residues with logarithmic dependencies (Figure 9; gray squares) can be fitted to the evolution of the amount of LYS involved in oligomers (without distinction between dimer and tetramer). The general features of this dependency are sensitive mainly to Kd_{DIMER} and weakly to $Kd_{TETRAMER}$ and, as a consequence, Kd_{DIMER} is close to $K_d^{L}{}_{LYS/aLAC}$ (Figure 5f). The value of $Kd_{TETRAMER}$ is about 2 orders of magnitude lower than that of Kd_{DIMER}. This is not surprising because the equilibrium between dimer and tetramer is controlled by the maximal dimer concentration which is limited by the LYS concentration (i.e., 10 μM in our experiments). Therefore, the dissociation

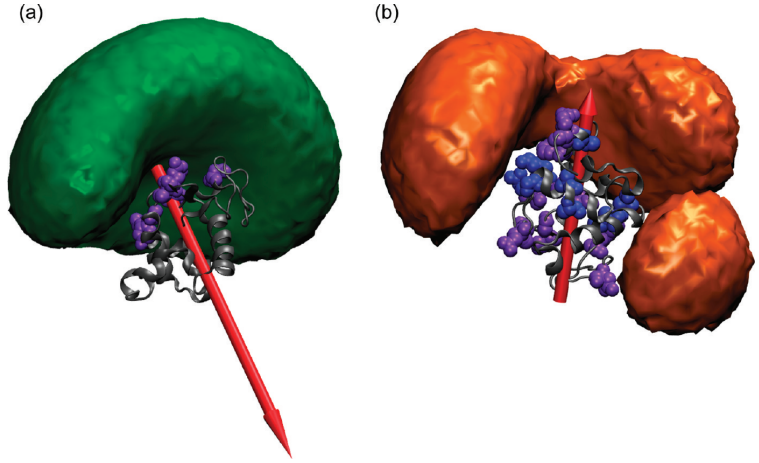


Figure 8. Combination of computational and experimental results. (a) LYS distribution around hLAC. The green surface is the LYS isodensity surface around hLAC. Interacting residues of hLAC determined experimentally are represented as purple hard spheres. (b) hLAC distribution around LYS. The orange surface is the hLAC isodensity surface around LYS. Interacting residues of LYS determined experimentally are represented as hard spheres colored either in purple ($K_d^1{}_{LYS/hLAC}$) or blue ($K_d^2{}_{LYS/hLAC}$). The red arrow represents the electric dipole moment equal to 400 and 140 D for hLAC and LYS, respectively. Graphics produced using VMD.

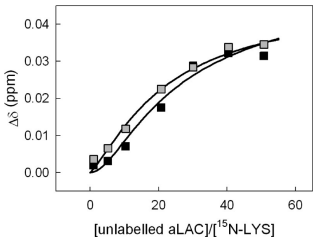
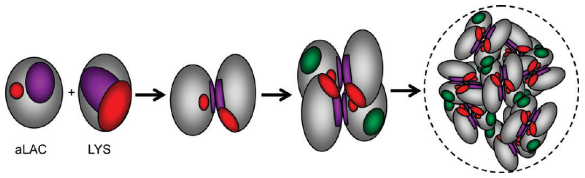


Figure 9. Evolutions of $\Delta\delta$ obtained during the titration of ^{15}N-LYS by aLAC fitted by assuming a sequential reaction described in Scheme 1. The dissociation constants of dimer (Kd_{DIMER}) and tetramer ($Kd_{TETRAMER}$) were evaluated by fitting the logarithmic (gray squares) and sigmoid (black squares) dependencies with eq 4 and 5, respectively, giving Kd_{DIMER} = 450 ± 70 μM, $\Delta\delta_{maxDIMER}$ = 0.05 ± 0.005 ppm and $Kd_{TETRAMER}$ = 5.0 ± 2.5 μM, $\Delta\delta_{maxTETRAMER}$ = 0.08 ± 0.002 ppm. The continuous lines correspond to the best fits.

constant must be at maximum of the same order of magnitude as the LYS concentration in order to have significant tetramer formation. Therefore, our titration data can be interpreted on the basis of the further association of heterodimers into tetramers (Scheme 1). Moreover, the appearance of tetramers is drastically favored once the aLAC-LYS dimers are formed. One can notice here that, according to the value of $Kd_{TETRAMER}$, only 50% of the LYS can be involved in tetramer. Going back to our measurements, this implies that the chemical shift variation of residues with sigmoid behavior is, at the end of the titration, only about half of what it should be if LYS was fully involved in tetramer. This can explain the difference of $\Delta\delta_{max}$ between residues with logarithmic behavior and those with sigmoid behavior (Figure 6c).

Based on our observations, we can propose a mechanism for the tetramer formation and their assembly into supramolecular objects (Scheme 2). Although we do not have experimental data on the aLAC interacting surface, the topology of the protein in its molten globule state remains native-like, with a significant amount of native contacts between side chains.[49] Therefore, it is reasonable to assume that aLAC is interacting with LYS to form a dimer through a similar region to that evidenced on hLAC surface. This hypothesis is also supported by the fact that the interacting surfaces of LYS with both forms of LAC are located in the same region of LYS. The tetramer formation requires further interactions between aLAC and LYS and involves additional LYS residues, that is, those with sigmoid dependencies. The absence of bound calcium on aLAC induces the presence of another negatively charged patch on the surface (Figure 7c), which may constitute a second site for interacting with LYS. Indeed, LYS remains in a compact state, that is, without hydrophobic patch on its surface, and it is reasonable to assume that these additional interactions required for tetramer formation are still predominantly electrostatic. These secondary interactions between aLAC and LYS lead to narrow the ensemble of possible heterodimer conformations among the numerous possibilities illustrated in Figure 8 and results in a tighter binding. Furthermore, aLAC is in a molten globule state, that is, with hydrophobic patches on the surface because of destabilized tertiary structure. Upon tetramer formation, most of the charges on the protein surfaces are shielded, and the interactions through the hydrophobic patches present on aLAC surface become preponderant facilitating the further aLAC-LYS self-assembly into larger oligomers. However, the factors which control the formation of microspheres instead of amorphous aggregates are still unknown.

Scheme 2. Schematic Representation of aLAC-LYS Dimer and Tetramer Formation[a]



[a] The LYS interacting surface determined experimentally is represented in purple ($K_d^{L}{}_{LYS/aLAC}$) and the residues interacting in a cooperative manner ($K_d^{S}{}_{LYS/aLAC}$) are represented in red. Residues of aLAC involved in dimer formation are supposed to be located in a similar region of that determined experimentally for hLAC (purple) and additional binding residues correspond to those near the calcium binding site (red). The aLAC is slightly larger than LYS because the molten globule state of LAC is known to be less compact than the native state.[50,46] A first binding event involves purple patches, whereas the second binding event for tetramer formation involves red patches. Oligomerization into supramolecular objects is mainly driven by hydrophobic associations favored by the exposition of hydrophobic patches (green) on the surface of molten globule of aLAC.

CONCLUSION

In summary, the work described here provides insights into the assembly mechanisms of ordered aggregates by two proteins. The initial formation of heterodimers is driven by electrostatic interactions. Despite the rather unspecific and weak interactions between the two proteins, the NMR approach leads to the identification of the interaction surfaces. One of the proteins has a narrow interacting site, while the second has interacting sites scattered on a broad surface. These findings agree with a preferential alignment between proteins resulting from the strength of their dipole moments. The further assembly of these rather unspecific heterodimers into tetramers leads to the tighter interactions between both proteins. Within the tetramers, most of the electrostatic charge patches on of the protein surfaces are shielded. As consequence, the hydrophobic interactions, which are possible because one the proteins is in a partially folded state, become preponderant and lead to the formation of larger oligomers. We are convinced that such an approach will be particularly useful to understand the formation of ordered aggregates by proteins. It should be extended to other mixed protein systems in order to unravel their self-assembly mechanisms and rationalize their use in the design of nanoscale devices.

ASSOCIATED CONTENT

Supporting Information. Detailed Materials and Methods and supporting Figures S1−S3 and Tables S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*Phone: +33 4 38 78 94 05. Fax: +33 4 38 78 54 87. E-mail: vincent.forge@cea.fr.

ACKNOWLEDGMENT

We are very grateful to C. Bougault for fruitful discussions. We thank the French NMR platform at IBS (Grenoble, France) for access to the 800 MHz NMR spectrometer (TGE RMN THC Fr3050). This work was supported in part by the French National Research Agency (Grant ANR PNRA LACLYS), the INRA (Rennes, France), and the CEA (Grenoble, France). B.A.P. and M.L. thank the Swedish Linneaus Center of Excellence "Organizing Molecular Matter" for financial support. R.S. is supported by the Polytechnische Stiftung and H.S. is a member of the DFG-funded cluster of excellence: macromolecular complexes. BMRZ is funded by the state of Hesse.

REFERENCES

(1) Zhang, S. G. *Nat. Biotechnol.* **2003**, *21* (10), 1171–1178.
(2) van der Linden, E.; Venema, P. *Curr. Opin. Colloid Interface Sci.* **2007**, *12* (4−5), 158–165.
(3) Goers, J.; Permyakov, S. E.; Permyakov, E. A.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2002**, *41* (41), 12546–12551.
(4) Krebs, M. R. H.; Wilkins, D. K.; Chung, E. W.; Pitkeathly, M. C.; Chamberlain, A. K.; Zurdo, J.; Robinson, C. V.; Dobson, C. M. *J. Mol. Biol.* **2000**, *300* (3), 541–549.
(5) Sagis, L. M. C.; Veerman, C.; van der Linden, E. *Langmuir* **2004**, *20* (3), 924–927.
(6) Desfougeres, Y.; Croguennec, T.; Lechevalier, V.; Bouhallab, S.; Nau, F. *J. Phys. Chem. B* **2010**, *114* (12), 4138–4144.
(7) Krebs, M. R. H.; Devlin, G. L.; Donald, A. M. *Biophys. J.* **2007**, *92* (4), 1336–1342.
(8) Mossa, S.; Sciortino, F.; Tartaglia, P.; Zaccarelli, E. *Langmuir* **2004**, *20* (24), 10756–10763.
(9) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
(10) Qasba, P. K.; Kumar, S. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32* (4), 255–306.
(11) Kronman, M. J.; Sinha, S. K.; Brew, K. *J. Biol. Chem.* **1981**, *256* (16), 8582–8587.
(12) Bushmarina, N. A.; Blanchet, C. E.; Vernier, G.; Forge, V. *Protein Sci.* **2006**, *15* (4), 659–671.
(13) Hendrix, T.; Griko, Y. V.; Privalov, P. L. *Biophys. Chem.* **2000**, *84* (1), 27–34.
(14) Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S. Y.; Ptitsyn, O. B. *FEBS Lett.* **1981**, *136* (2), 311–315.
(15) Kuwajima, K. *FASEB J.* **1996**, *10* (1), 102–109.
(16) Troullier, A.; Reinstadler, D.; Dupont, Y.; Naumann, D.; Forge, V. *Nat. Struct. Biol.* **2000**, *7* (1), 78–86.
(17) Fink, A. L.; Calciano, L. J.; Goto, Y.; Kurotsu, T.; Palleros, D. R. *Biochemistry* **1994**, 33 (41), 12504–12511.
(18) Shih, P.; Holland, D. R.; Kirsch, J. F. *Protein Sci.* **1995**, *4* (10), 2050–2062.
(19) Howell, N. K.; Yeboah, N. A.; Lewis, D. F. V. *Int. J. Food Sci. Technol.* **1995**, *30* (6), 813–824.
(20) Nigen, M.; Croguennec, T.; Bouhallab, S. *Food Hydrocolloids* **2009**, *23* (2), 510–518.

(21) Nigen, M.; Croguennec, T.; Renard, D.; Bouhallab, S. *Biochemistry* **2007**, *46* (5), 1248–1255.
(22) Nigen, M.; Le Tilly, V.; Croguennec, T.; Drouin-Kucma, D.; Bouhallab, S. *Biochim. Biophys. Acta* **2009**, *1794* (4), 709–715.
(23) Persson, B. A.; Lund, M. *Phys. Chem. Chem. Phys.* **2009**, *11* (39), 8879–8885.
(24) Nigen, M.; Croguennec, T.; Madec, M. N.; Bouhallab, S. *FEBS J.* **2007**, *274*, 6085–6093.
(25) Akerud, T.; Thulin, E.; Van Etten, R. L.; Akke, M. *J. Mol. Biol.* **2002**, *322* (1), 137–152.
(26) Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1997**, *36* (15), 4393–4398.
(27) McKenna, S.; Moraes, T.; Pastushok, L.; Ptak, C.; Xiao, W.; Spyracopoulos, L.; Ellison, M. J. *J. Biol. Chem.* **2003**, *278* (15), 13151–13158.
(28) Van Nuland, N. A. J.; Kroon, G. J. A.; Dijkstra, K.; Wolters, G. K.; Scheek, R. M.; Robillard, G. T. *FEBS Lett.* **1993**, *315* (1), 11–15.
(29) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274* (5292), 1531–1534.
(30) Lepre, C. A.; Moore, J. M.; Peng, J. W. *Chem. Rev.* **2004**, *104* (8), 3641–3675.
(31) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B. L.; Wendt, M. D.; Zhang, H. C.; Fesik, S. W.; Rosenberg, S. H. *Nature* **2005**, *435* (7042), 677–681.
(32) Caussin, F.; Famelart, M. H.; Maubois, J. L.; Bouhallab, S. *Lait* **2003**, *83* (5), 1–12.
(33) Schlörb, C.; Ackermann, K.; Richter, C.; Wirmer, J.; Schwalbe, H. *J. Biomol. NMR* **2005**, *33* (2), 95–104.
(34) Grobler, J. A.; Wang, M.; Pike, A. C. W.; Brew, K. *J. Biol. Chem.* **1994**, *269* (7), 5106–5114.
(35) Schanda, P.; Brutscher, B. *J. Am. Chem. Soc.* **2005**, *127* (22), 8014–8015.
(36) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6* (3), 277–293.
(37) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4* (5), 603–614.
(38) Schumann, F.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. *J. Biomol. NMR* **2007**, *39* (4), 275–289.
(39) Fielding, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51* (4), 219–242.
(40) Forge, V.; Mintz, E.; Guillain, F. *J. Biol. Chem.* **1993**, *268* (15), 10953–10960.
(41) Lund, M.; Trulsson, M.; Persson, B. *Source Code Biol. Med.* **2008**, *3*, 1–8.
(42) Alexandrescu, A. T.; Evans, P. A.; Pitkeathly, M.; Baum, J.; Dobson, C. M. *Biochemistry* **1993**, *32* (7), 1707–1718.
(43) Roberts, G. C. K. *The Determination of Equilibrium Dissociation Constants of Protein−Ligand Complexes by NMR*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2003; pp 309−319.
(44) Rule, G.; Hitchens, T. *Fundamentals of Protein NMR Spectroscopy*; Springer: New York, 2006.
(45) Buchko, G. W.; Tarasevich, B. J.; Bekhazi, J.; Snead, M. L.; Shaw, W. J. *Biochemistry* **2008**, *47* (50), 13215–13222.
(46) Vaynberg, J.; Qin, J. *Trends Biotechnol.* **2006**, *24* (1), 22–27.
(47) Crowley, P. B.; Ubbink, M. *Acc. Chem. Res.* **2003**, *36* (10), 723–730.
(48) Zuiderweg, E. R. P. *Biochemistry* **2002**, *41* (1), 1–7.
(49) Balbach, J.; Forge, V.; Lau, W. S.; Jones, J. A.; VanNuland, N. A. J.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (14), 7182–7185.
(50) Kataoka, M.; Tokunaga, F.; Kuwajima, K.; Goto, Y. *Protein Sci.* **1997**, *6* (2), 422–430.

Paper VII

Kurut, Persson, Åkesson, Forsman and Lund
Submitted

Anisotropic Interactions in Protein Mixtures: Self Assembly and Phase Behavior

Anıl Kurut,* Björn A. Persson, Torbjörn Åkesson, Jan Forsman, and Mikael Lund†
Department of Theoretical Chemistry,Lund University, P.O.B. 124 SE-22100 Lund, Sweden‡
(Dated: September 15, 2011)

The effect of anisotropic protein-protein interactions on the phase behavior of mixtures of lysozyme and α-lactalbumin in aqueous salt solutions is investigated, using parallel tempering Monte Carlo simulations. We present a coarse grained protein model which captures the non-uniform distribution of charged amino acids, responsible for strong directional protein interactions. In semi-quantitative agreement with recent experimental data our model predicts that calcium binding or elevated salt concentrations retard micro-sphere formation in lysozyme/α-lactalbumin mixtures. It is further shown that decreasing the charge anisotropy on α-lactalbumin by a *single* point mutation causes a large shift in the critical point of the fluid-fluid coexistence curve towards lower salt concentrations. Minimizing the electrostatic anisotropy completely suppresses phase separation at the corresponding experimental conditions.

There have been several theoretical studies on how anisotropic interactions influence self association and phase behavior of proteins [1–7]. Proteins are typically described as patchy spherical particles with a number of attractive sites interacting via angular dependent square well potentials. These models rely mainly on ambiguously placed attractive patches and do not focus on anisotropy due to non-uniform surface charge distributions of real proteins. One exception is the embedded discreet charge model by Carlsson *et al.* [2, 4, 8] where charged residues are projected onto a hard sphere. While accounting for anisotropic electrostatics, this model treats the excluded volume as spherically symmetric. We here present a simulation study of protein mixtures using a mesoscopic protein model sensitive to pH, salt concentration, point mutations as well as anisotropic interactions as derived from the protein structure.

In a recent experimental study of the phase behavior of lysozyme (lys) and α-lactalbumin (α-lac) mixtures, Nigen *et al.* [9–11] observed self-assembly of lys and the calcium depleted (apo) form of α-lac into well defined micro-spherical aggregates at low salt concentrations (< 100 mM). For higher salt concentrations as well as for the calcium loaded (holo) form of α-lac, self-assembly does not occur [11]. Due to the anisotropic nature of the charge distribution of α-lac, combined with the high positive net charge of lysozyme, strong directional protein-protein interactions have been observed both theoretically and experimentally [12]. We here describe – at a microscopic level – how such long ranged non-centrosymmetric interaction mechanisms influence many-body assembly and phase transition of protein mixtures.

Simulating phase behavior at an atomistic level is unfeasible and we therefore coarse grained to the following three levels, all with a continuum solvent – see illustration in Fig. 1:

(a) Proteins were described as rigid bodies, built up by spheres. Each sphere represents an amino acid with mass centers located as in the experimental crystal structure. Salt particles were *explicitly* included [13].

(b) As (a) but with *implicit* salt particles described at the Debye-Hückel level.

(c) As (b) but neutral amino acids were clustered into *one* larger sphere, located in the mass center of the protein.

Metropolis Monte Carlo (MC) simulations in the canonical ensemble, NVT, was used to study two-body interactions while the isobaric-isothermal ensemble, NpT, was used to study many-body interactions and the associated phase behavior – see Table I. The system energy, U, for a given configuration is

TABLE I: MC simulation details. Unless otherwise stated all simulations are performed at 298 K and at pH 7.5. The source code is available through the Faunus project [14].

Ensemble	NVT	NpT
Simulation geometry	Spherical	Cubic
Boundaries	Hard	Periodic/MI
Models	(a) / (b)	(c)
N_{salt}	212-848/0	0
N_{proteins}	2	40[a]

[a] 30–100 proteins were tested for size dependencies – see appendix.

$$U = \sum_{i<j} 4\epsilon_{\text{LJ}} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + \frac{e^2 z_i z_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \zeta \quad (1)$$

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 is the

*Electronic address: anil.kurut@teokem.lu.se
†Electronic address: mikael.lund@teokem.lu.se
‡URL: http://www.teokem.lu.se

inter-particle separation, z the charge valency, e the electron unit charge, and $\epsilon_0\epsilon_r$ is the dielectric permittivity of water. For explicit salt simulations $\zeta = 1$ while 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, $G = -k_BT \ln P(V)$ [15]. Minima in G correspond to (meta-) stable phases and to obtain coexistence concentrations, G was re-weighted, $G' = G \pm \Delta pV$ to give two minima with equal free energies [16]. The Helmholtz free energy was calculated as $A = G' - pV$ and in order to efficiently sample configurational space at low salt concentrations (where strong cluster formation occurs) we used parallel tempering in both p and κ – see appendix.

Model (a) was used as reference model since it captures the original protein topology as well as anisotropic interactions [12, 17]. Using two-body simulations of lys interacting with α-lac we heuristically coarse grained model (a) while maintaining (i) the salt dependent potential of mean force (pmf) and (ii) the distance dependent alignment of α-lac's dipole moment with respect to lys. The latter property ranges between ± 1 where $+1$ corresponds to a fully aligned dipole. As seen in Fig. 1, deviations from (a) is due mainly to the introduction of implicit salt while protein coarse graining, (b)→(c), influences neither the pmf nor the anisotropy. Calcium binding to α-lac decreases the net-charge and hence lowers the attraction to lys. An overview of the models are shown in Table II.

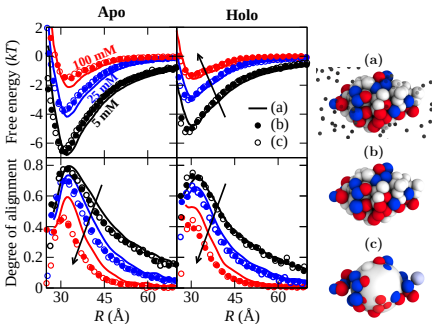


FIG. 1: Interaction between lys with α-lac at three different coarse graining levels and at various 1:1 salt concentrations (arrows indicate increasing salt). *Top graphs:* Angularly averaged potential of mean force, $w(R)$, as a function of the protein-protein mass center separation, R. *Bottom graphs:* Dipolar alignment of α-lac.

Using model (c), we performed NpT simulations with 40 proteins to investigate the Helmholtz free energy of bulk protein 1:1 mixtures containing lys and four different forms of α-lac: apo, holo, smeared, and mutated apo – see Fig. 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 agreement with experimental observations [10]. Increasing the salt concentration leads to reduced electrostatic interactions between the oppositely charged lys and α-lac. The same is the case for holo α-lac, where the added calcium ion reduces the net charge of α-lac.

TABLE II: Details of coarse grained models for lysozyme and α-lactalbumin. For model (a) and (b) $\epsilon_{\rm LJ} = 0.05\ k_BT$ and for model (c), $\epsilon_{\rm LJ} = 0.075\ k_BT$ and $\sigma_i = 15$ Å for the central sphere.

Protein	α-lac				lys
	apo	holo	mutated[a]	smeared[b]	
PDB entry	1F6R	1F6S	1F6R	1F6R	4LZT
$N_{\rm particles}$ – (a)/(b)	124	125	124	124	130
$N_{\rm particles}$ – (c)	42	44	42	42	33
Net charge, pH 7.5	-5.9	-4.5	-5.9	-5.9	7.0
Dipole moment[c] (D)	78	68	34	3	20

[a]Residues D79 and K99 are swapped.
[b]Net charge is smeared out on all titratable sites.
[c]Calculated with respect to the protein mass center

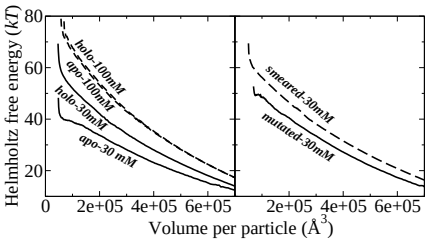


FIG. 2: Helmholtz free energy for 1:1 mixtures of lys with: apo, holo, mutated and, smeared α-lac at 30 mM and 100 mM salt concentrations.

In order to investigate the contribution from the anisotropic charge distribution of apo α-lac we reduced the dipole moment from 78 to 34 Debye by swapping only two charged amino acids (Table II). As seen in Fig. 3 this mutation – which does not alter the net charge – drastically decreased the critical salt concentration at coexistence and widens the immiscibility gap of lys/apo α-lac. By smearing out the total charge of apo α-lac on *all* titratable residues we approach isotropic electrostatics and as seen in Fig. 2 this completely prevents phase separation in the investigated salt concentration range.

Experimentally, micro-sphere formation occurs at low salt and at a total protein concentration ranging from 0.075 mM to 1.05 mM [10] implying that the equilibrium concentration of the dilute phase is in the sub millimolar range. Our model predicts a dilute phase concentration 5 to 70 times higher, and thus underestimates the width of the coexistence curve. Similar issues have

been reported for lysozyme [4, 18]. One reason could be lack of charge regulation in our model, although this is probably of small importance [13, 19]. It has also been proposed that apo α-lac undergoes conformational changes into a semi-folded molten globule state upon binding to lysozyme, leading to exposure of hydrophobic residues [11]. Such a mechanism – which would be temperature dependent [1, 18] – most likely leads to short-range attraction unaccounted for by rigid body models. To remedy this, we pragmatically incorporated a square well potential – depth one k_BT, width 3.1 Å, corresponding to a water layer – between the neutral spheres in model (c). As seen in Fig. 3, this 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, $K_d^{-1} \approx 4\pi \int_{\text{contact}}^{\infty} (e^{-w(R)/k_BT} - 1)R^2 dR$ [20]. Without short ranged attraction, $K_d = 691$ μM at 39 mM salt and pH 7.5 which is 25% off the experimental value of 490 ± 70 μM [12]. Including the above square well potential, we obtained $K_d = 480$ μM. Thus, the short ranged attraction improves not only the agreement with the experimental phase behavior, but also consistency at the twobody level. Note that our reference model (a) was originally parameterized to reproduce lys-lys osmotic second virial coefficients [17] and that we have made no efforts to fit it to experimental data for the current lys/α-lac system.

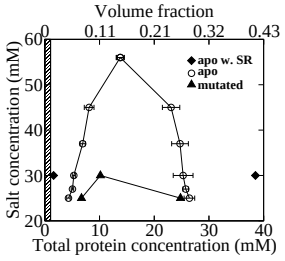


FIG. 3: The fluid-fluid coexistence curve of lys and α-lac in the form of apo, mutated, and with extra short ranged attraction (SR) – see text. The shaded area shows the experimentally observed concentration range of the dilute phase.

We have developed a coarse grained model which describes proteins as rigid bodies consisting of one centrally located non-polar sphere and several discrete charges at the original positions of titratable amino acids. This model captures anisotropic inter-protein interactions due to excluded volume as well as surface charges and accurately predicts the experimentally observed phase behavior of lys/α-lac with respect to salt concentration, pH and binding of divalent ions. On one hand, calcium binding lowers the net charge of holo α-lac and, while the orientation with respect to lys is unchanged, the decreased protein-protein interaction is sufficient to suppress phase separation. On the other hand, a single amino acid point mutation that minimizes the apo α-lac dipole moment – while retaining the net charge – is enough to significantly shift the immiscibility gap to lower salt concentrations. Artificially smearing out the charge over the protein surface completely inhibits phase separation. These results show that the phase behavior of dense aqueous protein mixtures is highly sensitive to anisotropic electrostatics – both qualitatively and quantitatively – and that their incorporation in theoretical models is crucial for accurate predictions of coexistence curves.

This work was supported by the Linnaeus center Organizing Molecular Matter, Lund University and computational resources was provided by LUNARC.

Appendix. Parallel tempering in κ and p using six replicas was used in the NpT simulations and configuration/volume exchange between replicas, i and j, was attempted with the following acceptance criterion [21],

$$\begin{aligned} acc[(i,j) \to (j,i)] &= \min\Big\{1, \exp\Big(-\beta[U(\kappa_i, \mathbf{r}_j) \\ &+U(\kappa_j, \mathbf{r}_i) - U(\kappa_i, \mathbf{r}_i) - U(\kappa_j, \mathbf{r}_j) \\ &+p_i V_j + p_j V_i - p_i V_i - p_j V_j]\Big)\Big\} \\ &= \min\Big\{1, \exp\Big(-\beta[\Delta U + \Delta pV]\Big)\Big\} \end{aligned} \quad \text{(A.2)}$$

where $\beta = 1/k_BT$. Each replica was equilibrated for 0.2 million MC steps while averages were collected over 2.5 million steps. Combined protein translation/rotation, volume re-scaling and replica exchange moves were randomly performed with the ratio $N : N^{-1} : 0.001N^{-1}$ where N is the total number of proteins. The resulting free energy as a function of volume and replica exchange events are shown in Fig. 4.

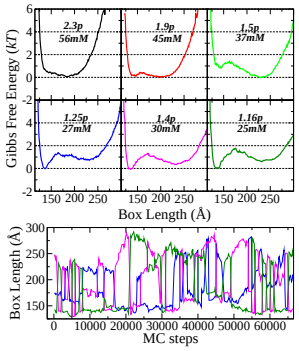


FIG. 4: *Top*: Constrainted Gibbs free energy of a system containing 20 lys and 20 apo α-lac molecules as a function of simulation box size obtained by tempering in osmotic pressure and salt concentration. (p refers to 10^{-6} proteins per Å^3) *Bottom*: Replica exchange events as a function of MC steps.

System size dependence was checked by NVT simula-

tions at densities corresponding to the dense and dilute phases at coexistence. As seen in Fig. 5, varying the total number of proteins from 30–100 at 30 mM salt does not cause any appreciable change in the protein-protein potential of mean force.

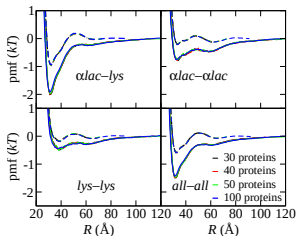


FIG. 5: Inter-protein potential of mean force as a function of the protein-protein mass center separation, R, in a dense (solid lines) and dilute phase (dashed lines). Calculated at 30 mM salt using model (c).

Salt density differences between dense and dilute phases were checked using model (a) with grand canonical salt. At a salt chemical potential of 30 mM, we observed salt concentrations of 22.2 and 28.7 mM in the dense and dilute phases, respectively. This corresponds to 12% difference in κ, stemming mainly from excluded volume, justifies usage of a constant κ in the NpT ensemble. Charges in model (a) was set using a separate MC simulation where protons fluctuate between the protein and bulk with a ratio dependent on pH, pK_a as well as intramolecular interactions [22]. ϵ_{LJ} was estimated to 0.05 k_BT based on a previously developed r^{-6} potential, known to reproduce experimental second virial coefficients for lysozyme [17].

[1] A. Lomakin, N. Asherie, and G. B. Benedek, Proc. Natl. Acad. Sci. **96**, 9465 (1999).
[2] F. Carlsson, M. Malmsten, and P. Linse, J. Phys. Chem. B **105**, 12189 (2001).
[3] N. Kern and D. Frenkel, J. of Chem. Phys. **118**, 9882 (2003).
[4] T. W. Rosch and J. R. Errington, J. Phys. Chem. B **111**, 12591 (2007).
[5] H. Liu, S. K. Kumar, and F. Sciortino, J. Chem. Phys. **127**, 084902 (2007).
[6] N. Wentzel and J. D. Gunton, J. Phys. Chem **112**, 7803 (2008).
[7] C. Gögelein, G. Nägele, R. Tuinier, T. Gibaud, A. Stradner, and P. Schurtenberger, J. Chem. Phys. **129**, 085102 (2008).
[8] M. C. Abramo, C. Caccamo, D. Costa, G. Pellicane, and R. Ruberto, J. Phys. Chem. B **114**, 9109 (2010).
[9] M. Nigen, T. Croguennec, D. Renard, and S. Bouhallab, Biochemistry **46**, 1248 (2007).
[10] M. Nigen, T. Croguennec, and S. Bouhallab, Food Hydrocolloids **23**, 510 (2009).
[11] M. Nigen, V. Le Tilly, T. Croguennec, D. Drouin-Kucma, and S. Bouhallab, Biochim. Biophys. Acta **1794**, 709 (2009).
[12] D. B. Salvatore, N. Duraffourg, A. Favier, B. A. Persson, M. Lund, M.-M. Delage, R. Silvers, H. Schwalbe, T. Croguennec, S. Bouhallab, et al., Biomacromolecules **12**, 2200 (2011).
[13] B. A. Persson and M. Lund, Phys. Chem. Chem. Phys. **11**, 8879 (2009).
[14] M. Lund, B. Trulsson, and B. Persson, Source Code Biol. Med. **3:1** (2008).
[15] J. E. Hunter and W. P. Reinhardt, J. Chem. Phys. **103**, 8627 (1995).
[16] A. V. Brukhno, T. Åkesson, and B. Jönsson, J. Chem. Phys. **113**, 6766 (2009).
[17] M. Lund and B. Jönsson, Biophysical Journal **85**, 2940 (2003).
[18] P. M. Biesheuvel, S. Lindhoud, R. de Vries, and M. A. Cohen Stuart, Langmuir **22**, 1291 (2006).
[19] P. M. Biesheuvel, S. Lindhoud, M. A. Cohen Stuart, and R. de Vries, Physical Review E **73**, 041408 (2006).
[20] H. Wennerström, *H. Statistical mechanical description of surfactant self-assmbly. In Organized solutions: surfactants in science and technology* (Marcel Dekker, New York, 1992).
[21] M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, J. Chem. Phys. **115**, 11352 (2001).
[22] A. A. R. Teixeira, M. Lund, and F. L. B. da Silva, J. Chem. Theory Comput. **6**, 3259 (2010).

Paper VIII

Persson, Lund, Forsman, Leiding, Peterson Årsköld, Åkesson
Manuscript

A theoretical study of the optical pH nanosensor Glu^3

Björn A Persson,[*,†] Mikael Lund,[†] Jan Forsman,[†] Thom Leiding,[‡] Sindra Peterson Årsköld,[‡] and Torbjörn Åkesson[†]

Department of Theoretical Chemistry, Lund University, POB 124, 221 00 Lund, Sweden, and Department of Biochemistry, Lund University, POB 124, 221 00 Lund, Sweden

E-mail: bjorn.persson@teokem.lu.se

Abstract

We propose a model for the optical *pH* sensor Glu^3. This coarse grained model is used to evaluate the influence of confinement and salt effects. The calculation of stoichiometric pK_a values shows a good agreement with experiments. It is shown that confinement has little effect on the *pH* report of the sensor. We also demonstrate the usability of the model by extending the study to investigate the importance of zwitterionic membranes on the titration of the porphyrin core.

Introduction

Optical nanosensors are an increasingly important tool in analyzing the state of biological samples, providing a means to noninvasively monitor local conditions.[1,2] Applications span both microscopic imaging and spectroscopic investigations. One important application is the detection of ion concentration gradients, such as *pH* differences across lipid membranes. Such transmembrane electrochemical gradients are prominent in biology and medicine; examples include bioenergetics and photosynthesis, apoptosis, muscle contraction, multidrug resistance and viral infectivity. The dynamics and characteristics of *pH* gradients can be investigated using isolated cells or organelles, or in experimental model systems such as artificially assembled liposomes, in combination with *pH* nanosensors, i.e. optically active molecules that change their spectral signature in response to their

[*]To whom correspondence should be addressed
[†]Lund University
[‡]Lund University

protonation state. Glu^3 (generation 3 polyglutamic porphyrin dendrimer)[3–5] is a novel porphyrin-based *pH* nanosensor which is particularly well suited for such measurements, as it will not diffuse across lipid membranes. Once enclosed in a microcompartment, Glu^3 will remain there and report on the *pH* in that volume. The potency of this probe was demonstrated by the investigation of the influenza drug target A/M2.[6] Due to the impermeable nature and favorable spectral qualities of the probe, quantitative data on proton and K^+ passage through A/M2 were obtained, leading to a new understanding of the functionality of this protein during viral infection. The *pH* nanosensor Glu^3 consists of a porphyrin core, which displays distinctly different absorption and fluorescence spectra depending on the porphyrin protonation state. It undergoes an acid-base equilibrium between the neutral and the dicationic form. As with many porphyrins, Glu^3 does not reveal any noticeable amount of the intermediate, singly protonated form. Attached to the porphyrin core are four polyglutamic dendrons. These dendrons, which together hold up to 32 negative charges, serve to make Glu^3 highly water soluble and disinclined to pass through lipid bilayers or membranes. They also stabilize the fully protonated state of the porphyrin, bringing up the naturally low *pK* of the porphyrin protonations to the physiologically relevant regime of around *pH* 6.[3–5] The procedure is complicated by the fact that *pH* is related to a stoichiometric equilibrium constant, rather than a thermodynamic equilibrium constant. The stochiometric constant varies with different electrolyte solutions due to the charged nature of Glu^3. As the charged dendrites interact with ions in solution, a certain salt sensitivity of the apparent *pK* of the probe Glu^3 is detected, and by large variations of the ionic strength, the apparent equilibrium constant of Glu^3 was tunable to between 6.7 and 5.3.[5] The reliable *pH* read-outs obtained with Glu^3 were achieved by rigorous calibration of the probe against the relevant sample conditions in conjunction with experiments.[5,7] This paper introduces a model of Glu^3 to better understand these issues, which can prove useful for interpreting results as well as optimizing experiment design. Further, we will demonstrate how the model may be used to evaluate influences of confinement and the presence of surfaces, both important in the study of small cavities. Also, we will briefly discuss the concept of Hill coefficients.

Model

The equilibrium constant of an arbitrary acid-base reaction is well defined in terms of the concentrations, c, and excess chemical potentials, $\beta\mu^{ex}$, of the reactants and products as i protons are released

$$K_a = \frac{c^i_{H^+} c_{A^{i-}}}{c_{AH_i}} \frac{e^{i\beta\mu^{ex}_{H^+}} e^{\beta\mu^{ex}_{A^{i-}}}}{e^{\beta\mu^{ex}_{AH_i}}}. \tag{1}$$

By rewriting Equation 1 we can express the ratio of the two states as

$$\frac{c_{A^{i-}}}{c_{AH_i}} = e^{\ln(10)(i\cdot pH - pK_a) - \beta(\mu^{ex}_{A^{i-}} - \mu^{ex}_{AH_i})}. \tag{2}$$

The degree of dissociation, α, is defined as,

$$\alpha = \frac{c_{A^{i-}}/c_{AH_i}}{1 + c_{A^{i-}}/c_{AH_i}}. \tag{3}$$

Using Equation 2 in 3 we obtain

$$\alpha = \frac{e^{\ln(10)(i\cdot pH - pK_a) - \beta(\mu^{ex}_{A^{i-}} - \mu^{ex}_{AH_i})}}{1 + e^{\ln(10)(i\cdot pH - pK_a) - \beta(\mu^{ex}_{A^{i-}} - \mu^{ex}_{AH_i})}}. \tag{4}$$

Data from spectroscopic measurements of the dissociation are on the other hand usually fitted to the so called Hill-equation,[8]

$$\alpha = \frac{10^{\widetilde{n}(pH - \widetilde{pK_a})}}{1 + 10^{\widetilde{n}(pH - \widetilde{pK_a})}}, \tag{5}$$

where $\widetilde{n}$ is called the Hill coefficient and $\widetilde{pK_a}$ is an effective pK_a. The connection between Equation 4 and 5 is by no means obvious. We will return to this later. This paper now presents a molecular model and evaluates its applicability to the experimental situation. We consider a model of the dendrimer solution based on the primitive model, where molecular water is replaced with a dielectric continuum.[9] The dendrimer is built from atomistic detail with the *Avogadro* software.[10] After assembling the dendrimer from porphyrin and glutamate templates, the final structure is stripped from hydrogens. The porphyrin core is kept at this level. The side chains are simplified in the following way: Carbonyl groups of the amino acids as well as side chains are replaced with a single atom, as is done for the hydrocarbon part of the side chain. In this manner, each glutamate subunit is modeled as four beads. In total, the dendrimer consists of 164 atoms, 112 dendrite beads and 52 porphyrin atoms. The dendrimer model is evaluated with the open source Monte Carlo package *Faunus*.[11] The porphyrin core is treated as a rigid body while the bonds connecting glutamate particles are harmonic. Excluded volume interactions are implemented as repulsive r^{-12} potentials. All charged particles interact with a solvent screened Coulomb potential. The configurational energy is given as

$$U = \sum_{i<j}^{N} \left(\frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} + \varepsilon \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \right) + \sum_{ij}^{N_B} k(r_{ij} - l_{eq})^2, \tag{6}$$

where the first sum runs over all particles and the latter over all bonds. q is the charge of each particle, ε_0 is the dielectric permittivity of vacuum, ε_r is the dielectric constant of the medium and r_{ij} is the inter particle distance. ε is the magnitude of the exchange repulsion, σ is the mean diameter of the particles, k is the force constant and l_{eq} is the equilibrium distance of the harmonic bonds. Salt particles and the chain monomers are displaced and new configuations are accepted with the standard Metropolis criterion for the canonical ensemble.[12] Further, salt particles stand in equilibrium with a bulk at fixed chemical potential. The chemical potentials are determined from canonical simulations of neat salt solutions. Carboxyl groups and the dendrimer core are allowed to titrate during the simulation. This is performed with a grand canonical titration scheme.[13] By specifying pH and a pK_a-value for acid-base pairs this allows the calculation of free energy costs of dissociating or binding protons. The routine exploits the possibility to exchange protons with monovalent cations while retaining a correct description of the chemical equilibrium. All carboxyl groups are titrated individually but the porphyrin core is handled as a double titration. Note that it is only necessary to probe the fully occupied and dissociated porphyrin configurations since the intermediate does not contribute to the visual spectra of Glu^{3-}.[5] All carboxylic acids are as-

signed the same pK_a-value, $pK_{a,G}$, and the porphyrin core is given the value $pK_{a,P}$. Simulations of bulk solutions are performed in a spherical cell with one porphyrin immersed in the center of the sphere and surrounded by salt particles. The degree of dissociation of the porphyrin core is sampled during the simulation, after an initial equilibration period. Dissociation curves are constructed for varying solution conditions and effective $\widetilde{pK_a}$ and $\widetilde{n}$ are determined by fitting the data to Equation 5. To investigate the spatial distribution of the dendrimer inside a lipid vesicle and the possible influence of surface interactions, simulations are also performed in a slab geometry with periodic boundary conditions in $x-y$ directions, using the minimum image convention. Hard walls are introduced in z direction. One wall is covered with grafted chains to mimic a phospholipid monolayer. Similar models have previously been used to describe lipid membranes.[14] Phosphatidylcholine head-groups (POPC) are modeled as follows: a neutral bead, constrained to the $x-y$-plane of the wall is connected to a negatively charged bead, which in turn is connected to a positively charged bead. The bonds of the grafted chains are treated in the same way as those of the dendrites. A meanfield correction for an infinite phospholipid layer, including a solute free salt solution was included with the charged sheet method.[15] As the dendrimer is rotated and translated in the z-direction the probability distribution of dendrimer wall separation, $g(r)$, is sampled. The dendrimer charge and degree of dissociation are monitored and the potential of mean force, $w(r)$, is obtained as $w(r)/k_BT = -\ln(g(r))$. Model parameters are summarized in Table 1. pK_a-values and structural parameters of the dendrimer are expanded upon and motivated below. Similar models have been successfully applied to charge regulation of proteins[16] and to estimate pK-shifts of calcium binding constants of calmodulin.[17–19]

Table 1: Parameter values used if nothing else is explicitly stated.

T	298 K
ε_r	78
$pK_{a,G}$	5.0
$pK_{a,P}$	10.5
l_{eq}	4 Å
k/k_BT	1Å^{-2}
ε/k_BT	1

Results and discussion

Parametrisation

The model is adjusted to experimental results at 2 mM KCl ($\widetilde{pK_a} \approx 6.7$ and $\widetilde{n} \approx 1.0$) and then used to study varying solution conditions without further adaptation. The variation of $\widetilde{pK_a}$ and $\widetilde{n}$ with respect to $pK_{a,G}$, k and l_{eq} are presented in Table 2. Results are presented for $pK_{a,G}$-values of 4.0 and 5.0, the former being typical for the glutamate side chain and the latter approximately the upper bound of small mono-protic carboxylic acids.[20] $\widetilde{pK_a}$ does not vary much within the considered range of $pK_{a,G}$ for any particular description of the harmonic bonds of the dendrites. This shows that the value of the dissociation constant of the phorphyrin core can not be varied significantly

Table 2: $\widetilde{pKa}$ and $\widetilde{n}$ for different parameter settings. $pK_{a,P}$ is set to 10.5 and 2 mM KCl. Dendrimer concentration is 2μM and the salt is 2 mM KCl.

$pK_{a,G}$	4.0				5.0			
k/l_{eq}	3.0/1.0	4.0/1.0	4.0/0.5	5.0/0.5	3.0/1.0	4.0/1.0	4.0/0.5	5.0/0.5
$\widetilde{pK_a}$	7.01	6.92	6.85	6.73	6.75	6.67	6.58	6.47
$\widetilde{n}$	1.9	1.8	2.0	1.9	1.2	1.3	1.3	1.3

from 10.5. This value was obtained from a preliminary search of the parameter space. The dependency of the estimated effective dissociation constant and Hill coefficient on the harmonic bonds is also weak. The small systematic trend with slightly higher dissociation constants with shorter and firmer bonds is an effect of bringing the dendrite charges closer to the porphoryn core and thereby increasing the stability of the positively charged acid form of the dendrimer. Contrary to $\widetilde{pK_a}$, $\widetilde{n}$ has a substantial variation with $pK_{a,G}$. Since the model is in significantly better agreement using the higher pK_a-value for the glutamates, this was adopted.

In the context of ratiometric pH-meters like Glu^3 it is insightful to scrutinize the concept of Hill coefficients and its relation to thermodynamics. From Equation 4 one can infer that if the difference in excess chemical potentials of the dissociated and associated states are constant, as a function of pH, the Hill coefficient will be 2 for the reaction under consideration. To display Hill coefficients $\widetilde{n} \neq i$, $\beta(\mu^{ex}_{A^{i-}} - \mu^{ex}_{AH_i}) = \beta\Delta\mu^{ex}$ must vary with pH. The success of Equation 5 can be understood if one considers the situation when $\beta\Delta\mu^{ex}$ varies fairly slowly around the half point. A series expansion of the difference in excess chemical potential around the half point leads to

$$\beta\Delta\mu^{ex} \approx \beta\Delta\mu^{ex}_{1/2} + \lambda(pH - pH_{1/2})\ln(10) + ... \tag{7}$$

where λ is the linear Taylor coefficient. Neglecting terms of higher order than linear in pH, the terms in Equation 2 may be rewritten as $\widetilde{pK_a} = \frac{1}{i}(pK_a + \frac{\beta}{\ln(10)}\Delta\mu^{ex}_{1/2})$ and $\widetilde{n} = i - \lambda$. Clearly we don't just expect that ideal systems are described by Equation 5 but also interacting systems, with a pH dependent difference in excess chemical potential of the solute. In the present model deviations in $\widetilde{n}$ from 2 are due to changes in $\Delta\beta\mu^{ex}$ associated with binding of protons to the dendrites. Finally we conclude that the model is rather insensitive to the choice of k and l_{eq}. Since each harmonic bond is approximately equal to two carbon-carbon, carbon-nitrogen or carbon-oxygen bonds a reasonable choice is 4 $\AA$ and the force constant was set to 1.0 $kT/\AA^2$. The Hamiltonian, Equation 6, contains a trivial self term for the core, that gives rise to an energy corresponding to -0.75 pK units. This should be added to $pK_{a,P}$ to compare the parameter fit with experimental values. We find a value of 9.75, or an average pK of 4.9. Uncharged tetraarylporphyrins have been reported to have pK values per proton in the range of 2-4.[4] It should be noted that dendritic modification of such porphyrin has been found to induce large changes in dissociation constants by structural effects even of neutral dendrites.[7] Taking this into consideration, we find our value of the dissociation constant for the porphyrin core to be reasonable.

Solute concentration dependency

We will now present calculations for different dendrimer concentrations as well as salt types and salt concentrations. Solute concentration effects have previously been observed in binding of calcium to proteins and explained in terms of an increased counterion concentration and the increased screening associated to the latter. These studies concluded a decreased affinity due to increased electrostatic screening which would imply a downshift of $\widetilde{pK_a}$.[17,19] Figure 1 shows results of titration simulations over the same concentration range, spanning two orders of magnitude. To exaggerate the counterion effect the simulations are performed with 4 mM KCl. Values obtained from

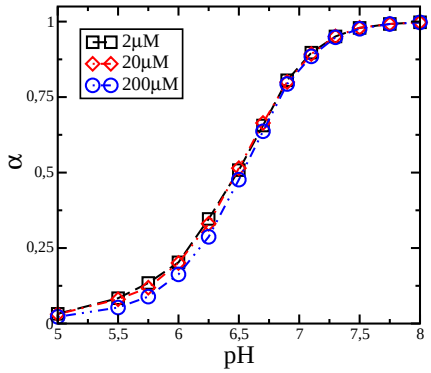


Figure 1: Glu^3 degree of dissociation, simulated at different Glu^3 concentrations and 4 mM KCl.

Table 3: Solute concentration dependency: results from fitting Equation 5 to data in Figure 1.

$[Glu^3]$ (μM)	$\widetilde{pK_P}$	$\widetilde{n}$
2	6.46	1.31
20	6.46	1.34
200	6.51	1.45

fitting to Equation 5 are summarized in Table 3. All changes are small. This implies that the properties of the probe will not be influenced by the volume of the probed cavity, which is important for a broader applicability.

Salt dependency

Dissociation curves are computed for two types of electrolytes, KCl between 2.2 to 135 mM as well as $MgCl_2$ between 0 to 8 mM in addition to 2.2 mM KCl. In order to computationally cope with the increasing number of particles with salt concentration, KCl was simulated at porphyrin concentrations of 200 μM and $MgCl_2$ at 20 μM. This will have negligible effect on the results, as seen above. In Figure 2 we compare the $\widetilde{pK_a}$ with the experimental results, as a function of the ionic

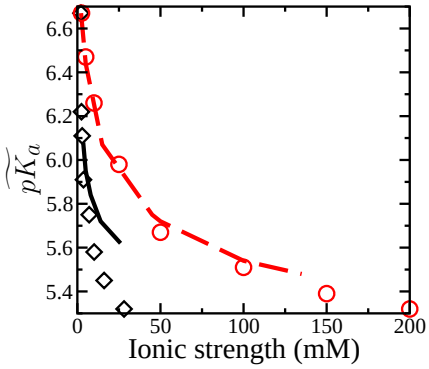


Figure 2: $\widetilde{pK_a}$ as function of ionic strength for KCl [2.2, 5, 15, 45, 135] mM (circles and dashed line) and $MgCl_2$ [0, 1, 2, 4, 8] mM with 2.2 mM KCl (diamonds and full line). Lines are simulation results and symbols are experimental results.

strength. In general, we find good agreement between the simulations and the experimental results. The comparison with KCl is exellent. Since Mg^{2+} is a divalent ion, electrostatic correlations tend to increase the screening. Although the simulations clearly capture this trend, it is not to the full extent of the experimental system and we note an overestimate of $\widetilde{pK_a}$ for Mg^{2+}. $\widetilde{n}$ displays a strong dependency upon salt concentration. As the ionic strength is increased, the Hill coefficient approaches the ideal value of 2 for both salt types, see Figure 3A. This is inconsistent with the

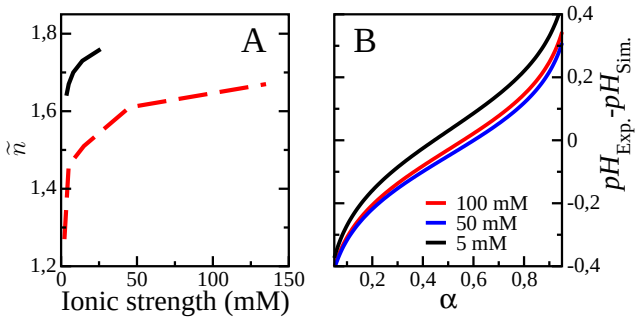


Figure 3: A: $\widetilde{n}$ as function of ionic strength. Full black line is $MgCl_2$ and dashed red line is KCl. B: *pH* discrepancy between model and experiments as function of degree of dissociation for three different concentrations of KCl.

experimental results. In fact, the measured data show little variation of $\widetilde{n}$ with salt concentration, ranging from 1.0 to 1.1 between 2 and 160 mM K. For Mg it is 1.2 over the entire range. According to the analysis above, the dissociation constant of the porphyrin core is likely to change with *pH* for

further reasons than the electrostatic interactions between the titrating dendrites and salt solution. These may be a changing solvent structure around the porphyrin core or conformational changes that affect the binding energy. The model's lack of such responses may explain the need to use a slightly higher dissociation constant of the glutamic residues to obtain a low Hill coefficient. The discrepancy in $\widetilde{n}$ shows that one can not directly compare the degree of dissociation of the experimental and model system. Figure 3B shows the difference in *pH* between measurements and calculations as a function of α. For α between 0.1 and 0.9 this difference is between -0.3 and 0.3 with a rather low dependency on salt concentration for 5 to 100 mM KCl.

Inside vesicles

Enclosing dendrimers inside vesicles expose them to a surface of phospholipids. We will now demonstrate how one can use the model to investigate interactions with surfaces and their influence on the dissociation reaction. One dendrimer was immersed in a cube with sides of 250 Å with periodic boundary conditions in $x-y$ direction as described above. The phospholipid layer was built up with 892 POPC chains to obtain a reasonable head group area of 70 Å^2. A KCl concentration of 30 mM was employed. Figure 4 shows the potential of mean force between the center of mass of Glu^3 and the grafting plane of the phospholipid head groups. The head groups extend

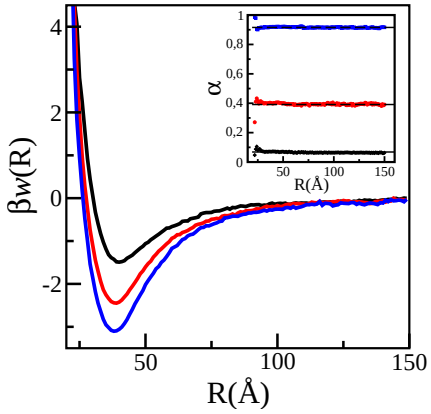


Figure 4: $\beta w(R)$ between Glu^3 and the grafting plane of the phospholipid wall at *pH* 5.00, 5.75 and 6.00 with 30 mM KCl. The inset shows the degree of dissociation as function of separation. Dashed lines indicate bulk values.

approximately 10 Å perpendicular to the surface and the Glu^3 model has a radius of gyration of roughly 18 Å. This corresponds reasonably well to the minima at 38 Å. As *pH* increases so does the magnitude of the dendrimers charge and the adsorption. The distribution of Glu^3 inside a vesicle is therefore expected to vary over the course of a titration experiment. However, the degree of dissociation is independent of distance to the phospholipid wall (see inset). The potential of

the zwitterionic membrane is strong enough to impose a significant concentration gradient but too weak to affect the core of the dendrimer.

Conclusion

A proposed porphyrin model has been evaluated and compared with experimental results. The model captures the change of $\widetilde{pK_a}$ with ionic strength well. A molecular, but numerically tractable model may prove useful to interpret measurements in small cavities and pores, especially for probes like Glu^3 which is sensitive to both pH, salt concentration and salt type. To demonstrate the usability of the model, calculations have been performed outside a phospholipid membrane and the influence of the membrane has been determined. At high pH the dendrimer adsorbs strongly while the spectral properties are uneffected. These results are in agreement with experimental observations. The model suggests that the carboxylic acids of the glutamate based dendrites are less acidic than those of an isolated glutamate. Since an acid-base constant contains a contribution from the solvation, the crowded environment of the porphyrin may explain this discrepancy. Also, the solvation may be dependent upon the degree of ionization of the dendrimer. This could implicate a varying pK_a of the glutamates as function of pH and possibly explain the difficulty to reproduce the experimentally determined Hill coefficients. The model concludes that the dependency upon solute concentration is very small in the concentration ranges that are covered within the paper. We have also demonstrated that the Hill coefficient may be strongly influenced by the chemistry of the binding ligand in a very broad sense, in this paper exemplified with the release of protons which decreases the electrostatic potential on the porphyrin and increases its proton affinity. Although the general agreement with experiments is good, we find discrepancy with respect to the Hill coefficient. This may possibly be understood in terms of a changing intrinsic binding of protons to the porphyrin core or changing solvent structure as function of pH.

Acknowledgement

We would like to express thanks to the Royal Physiographic Society in Lund and Organizing Molecular Matter (Lineaus Center of Excellence) for funding and Lunarc for computational resources.

Supporting Information

In order to obtain a tractable model the phospholipid head groups and glutamate residues were given a simple representation. Figure 5 illustrates the coarse grained models of glutamate units and the phospholipid head groups. In the simulation salt stand in equilibrium with a reservoir. This is achieved by adding and deleting electroneutral pairs of cations and anions A and B. The following acceptance criteria ensure detailed balance where the former is for deletion and the latter

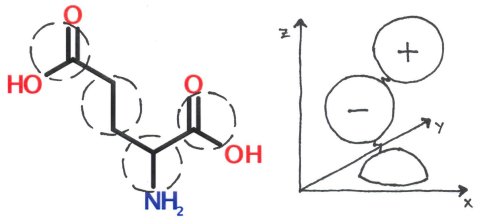


Figure 5: Left: Illustration of coarse grained glutamine. Right: Illustration of phosphatidylcholine head group model.

for addition of ions,[21]

$$P = \min\left[1, \frac{N_A!N_B!}{(N_A-n_A)!(N_B-n_B)!V^{(n_A+n_B)}}\exp\left(-\beta\Delta U-(n_A\beta\mu_A+n_B\beta\mu_B)\right)\right]$$
$$P = \min\left[1, \frac{N_A!N_B!V^{(n_A+n_B)}}{(N_A+n_A)!(N_B+n_B)!}\exp\left(-\beta\Delta U+(n_A\beta\mu_A+n_B\beta\mu_B)\right)\right]. \quad (8)$$

V is the volume of the simulation container and n_i is the stoichiometric coefficient of the electroneutral pair. N_i is the number of ions i in any given configuration, μ_i is the chemical potential of the ion and ΔU is the interaction energy difference of the new and old configuration. The calculations are performed at constant pH, or constant chemical potential of hydronium ions. Explicit presence of protons leads to numerical difficulties at intermediate and high pH-values. This is avoided with a routine that exploits the possibility to exchange protons with monovalent cations while retaining an accurate description of the chemical equilibrium. Charge states may be sampled with the following acceptance criterion

$$P = \min\left[1, \frac{N_A!}{(N_A-n)!V^n}\exp\left(-\beta\Delta U-\beta\mu_A-\ln(10)(n\cdot pH-pK_{a,i})\right)\right]$$
$$P = \min\left[1, \frac{N_A!V^n}{(N_A+n)!}\exp\left(-\beta\Delta U+\beta\mu_A+\ln(10)(n\cdot pH-pK_{a,i})\right)\right], \quad (9)$$

where the first expression describes protonation and second deprotonation.[13] n is the number of protons exchanged and N_A is the number of exchangeable ions in any given configuration. Since the chemical potentials for the ions are determined from bulk solutions, there is a risk that the dendrimer will make a significant contribution to the excess chemical potential and thereby changing the salt concentration as function of pH. We have investigated this by controlling the magnesium concentration in the cell at high pH. This should be the case which is most sensitive to such effects since magnesium is divalent and the charge of the dendrimer is maximum at these conditions. Chemical potentials corresponding to a solute free solution of 8 mM $MgCl_2$ and 2.2 mM KCl gave rise 8.3 mM Mg^{2+} and 2.2 mM K^+. We may therefore conclude that this effect is small under the present circumstances. Also, we have investigated the sensitivity to the radii of cations. Neither increasing the radii of K or decreasing the same for Mg had any noticable effect. The radius of

the bulk simulation cell was 583, 270.6 and 125.6 Å for solute concentrations of 2, 20 and 200 μM. The simulation of bulk solution was performed with an equilibration and production run of $3 \cdot 10^5$ cycles respectively 10^6 cycles. Each cycle composed of one core titration, 32 glutamate titrations with randomly chosen glutamates, random displacements of one third of the ions picked at random, one random change of the number of salt particles and random update of the harmonic bonds. In the simulations with phospholipid walls each cycle was augmented with an update of the dendrimers location in z-direction and the location of phospholipid head group beads. The equilibration was extended to $5 \cdot 10^5$ cycles and production to $8 \cdot 10^6$. To improve convergence umbrella sampling was preformed in two windows in the z-direction, 0-100 Å and 70 to 150 Å . The final result was obtained by averaging three independent productions for each window. The size and charges of the particles in the model are summarized in Table 4.

Table 4: Particle properties.

Type	Charge (e)	$\sigma/2$ (Å)
Core	0	2.0
Dendrite	0	2.0
K^+	+1	1.2
Cl^-	-1	2.2
Mg^{2+}	+2	2.0
POPC	-1/0/+1	2.5

References

(1) Borisov, S. M.; Klimant, I. *Analyst* **2008**, *133*, 1302–1307.

(2) Han, J.; Burgess, K. *Chemical Reviews* **2010**, *110*, 2709–2728.

(3) Vinogradov, S. A.; Wilson, D. E. *Chemistry: A European Journal* **2000**, *6*, 2456–2461.

(4) Finikova, O.; Galkin, A.; Rozhkov, V.; Cordero, M.; Hägerhäll, C.; Vinogradov, S. *J. Am. Chem. Soc.* **2003**, *125*, 4882–4893.

(5) Leiding, T.; Górecki, K.; Kjellman, T.; Vinogradov, S. A.; Hägerhäll, C.; Årsköld, S. P. *Analytical Biochemistry* **2009**, *388*, 296–305.

(6) Leiding, T.; Wang, J.; Martinsson, J.; DeGrado, W. F.; Årsköld, S. P. *PNAS* **2010**, *107*, 15409.

(7) Thyagarajan, S.; Leiding, T.; Årsköld, S. P.; Cheprakov, A. V.; Vinogradov, S. A. *Inorganic Chemistry* **2010**, *49*, 9909–9920.

(8) Hill, A. V. *Proceedings of the Physiological Society* **1910**, iv.

(9) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications Inc.: New York, 1986.

(10) `http://avogadro.openmolecules.net`.

(11) Lund, M.; Trulsson, M.; Persson, B. *Source Code for Biology and Medicine* **2008**, *3*, 1.

(12) Metropolis, N. A.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1097.

(13) Labbez, C.; Jönsson, B. *Applied Parallel Computing. State of the Art in Scientific Computing*; Springer Berlin / Heidelberg, 2007; Vol. Volume 4699/2010; Chapter A New Monte Carlo Method for the Titration of Molecules and Minerals.

(14) Granfeldt, M. K.; Miklavic, S. J. *J. Phys. Chem.* **1991**, *95*, 6351–6360.

(15) Torrie, G. M.; Valleau, J. P. *J. Chem. Phys.* **1980**, *73*, 5807–5816.

(16) Lund, M.; Jönsson, B. *Biochemistry* **2005**, *44*, 5722–7.

(17) Svensson, B.; Jönsson, B.; Woodward, C. E.; Linse, S. *Biochemistry* **1991**, *30*, 5209–5217.

(18) Svensson, B.; Jönsson, B.; Fushiki, M.; Linse, S. *J. Phys. Chem.* **1992**, *96*, 3135–3138.

(19) Linse, S.; Jönsson, B.; Chazin, W. J. *Proceedings of the National Academy of Sciences* **1995**, *92*, 4748–4752.

(20) Smith, M. *Critical Stability Constants*; Plenum Press, 1977; Vol. 3.

(21) Frenkel, D.; Smith, B. *Understanding Molecular Simulations*; Academic Press, 2002.

Paper IX

Persson, Lund, Forsman and Åkesson
Manuscript

Multibody Protein Interactions and Genetic Variants

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

Department of Theoretical Chemistry, Lund University, POB 124, 221 00 Lund, Sweden

E-mail: bjorn.persson@teokem.lu.se

Abstract

A number of experimental studies suggest that lactoferrin – a biologically versatile whey protein – self assembles into dimers as well as higher order oligomers. Using multiscale Metropolis Monte Carlo simulations we investigate the association of lactoferrin in bulk protein solutions and show (1) that stereospecific dimer formation is driven by a symbiotic coupling between electrostatic and van der Waals interactions and (2) that this coupling is qualitatively maintained for lactoferrin from four different genetic sources – bovine, buffalo, horse and human. Salt concentration and pH strongly influence the aggregation and for the conditions studied here, human and horse lactoferrin are found to form larger aggregates than do bovine and buffalo lactoferrin.

Introduction

Lactoferrin, a whey protein first discovered in bovine milk,[1] is excreted in mammals with significant concentration variations between species and biological fluids. Lactoferrin has a prolate geometry that resembles a dumbbell with two major lobes, each able to bind a ferric ion.[2] The molecular weight is about 80 kDa, stemming from a single polypeptide of roughly 700 amino acids that can be approximately described as a biaxial ellipsoid with half-axes 26 Å and 47 Å, respectively.[3]

Besides being secreted in the mammary glands, lactoferrin is also found in many other exocrine fluids like saliva, pancreatic fluid, bile and tears.[4] Several biological functions have been ascribed to lactoferrin – for example as an enhancer of iron adsorption, bactericidal action and growth factor. The protein has been isolated in two different forms: the iron-bound (*holo*) and iron-free (*apo*) form, where *apo* is slightly more open than *holo*. Some biological activities do not depend on the presence of iron, but the apo and holo forms have slightly different concentration dependencies on sedimentation coefficients.[5]

Experimentally, there are several observations suggesting that lactoferrin self assembles. Early studies of the physical properties of the protein revealed anomalous electrophoretic behavior with the appearance of a faster moving component at pH 7-9.[6] It has also been noted that dimers of bovine lactoferrin can be mistaken for immunoglobulin[7] which has a molecular weight twice that of lactoferrin. Analysis of lactoferrin in bovine milk, obtained during infection, suggests the existence of monomers, dimers as well as trimers.[8,9] Further, a recent scattering study has shown that human lactoferrin forms higher order aggregates and that these are diminished at highly elevated salt concentrations.[3] It has also been found that human and bovine lactoferrin form large aggregates at intermediate salt concentrations[10] and the osmotic pressure of human lactoferrin is found to surprisingly low, as function of the volume fraction.[11]

The above results all indicate that lactoferrin is able to self assemble. We have previously developed a molecular model of bovine lactoferrin that predicts stereospecific dimers, in accordance with experimental results.[12] In this article we present an extended theoretical study of the physical interactions in aqueous lactoferrin solutions and establish whether or not such self assembly can be modeled

*To whom correspondence should be addressed

and understood on the basis of a coarse grained description of molecular entities.

Method

Metropolis Monte Carlo method is used to study the interaction and aggregation of genetic variants of lactoferrin in the canonical ensemble.[13,14] We use a rigid, coarse grained protein model – derived from experimental protein structures – where amino acids are represented by soft spheres. This description has previously been applied to several different systems,[12,15] including studies of large proteins and thermodynamic properties of protein solutions.[16] Protein structures are taken from the PDB repository of bovine (1BLF), buffalo (1CE2), horse (1B1X), and human (1LFG) lactoferrin.[17–20] pH dependent charges of titratable residues are placed in the center of the amino acid spheres, and determined from a separate titration simulation described elsewhere.[12,21] Excluded volume and van der Waals interactions are described by a Lennard-Jones potential and the total energy is,

$$U = \sum_{i<j} \frac{q_i q_j e^{-\kappa r_{ij}}}{4\pi\varepsilon_0\varepsilon_r r_{ij}} + 4\varepsilon_{lj}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (1)$$

Here, r_{ij} is the inter-particle separation, σ_{ij} is the arithmetic mean of the particle diameters, $\varepsilon_{lj} = 0.05\,kT$, $\varepsilon_0\varepsilon_r$ is the dielectric permittivity of water, $\varepsilon_r = 80$, q are the particle charges, and κ is the inverse Debye screening length. All calculations are performed at room temperature, $T = 298$ K and k is Boltzmann's constant,

At close protein-protein separations, the above sum runs over all amino acids and thus explicitly captures anisotropy in near-atomic detail – see Figure 1. When two proteins are separated more than a threshold mass-center distance, R_{cut}, we further coarse grain by reducing each protein to a dipolar molecule with a cationic (+) and anionic (−) center, defined as follows:

$$\vec{r}_{\pm} = \sum_{i}^{N_{\pm}} q_i \vec{r}_i / q_{\pm}, \quad q_{\pm} = \sum_{i}^{N_{\pm}} q_i.$$

Since each charge center mimics half of the protein, we spread the charge over a sphere of radius $R_{\pm}$ which in the Debye-Hückel theory corresponds to a simple charge scaling,[22]

$$q'_{\pm} = q_{\pm}\frac{\sinh(\kappa R_{\pm})}{\kappa R_{\pm}}. \quad (2)$$

Per protein pair, this multiscale coarse graining approach reduces the number of site interactions from about half a million to merely *four*, making bulk simulations with many proteins tractable. For lactoferrin, suitable values for R_{cut} and $R_{\pm}$ are 150 Å and 25 Å, respectively, and while heuristically chosen, the latter value incidentally agrees with the lob size of lactoferrin.

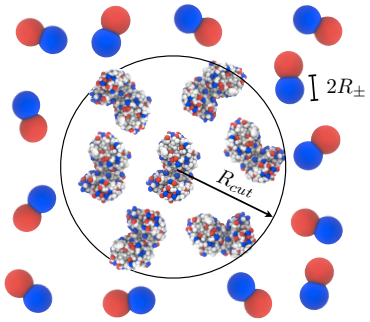


Figure 1: Multiscale interaction model of proteins coarse grained to the amino acid level (inside circle) and, when the inter-protein distance is larger than R_{cut}, to dipolar molecules.

During the MC simulations – performed using the Faunus framework[23] – the rigid proteins are translated, rotated and subjected to non-rejective cluster translation.[24] Two simulation geometries are used: (1) the cell model, where the system is enclosed in a spherical simulation cell with hard boundaries, and (2) a cubic cell with periodic boundaries and minimum image distances. Geometry (1) is used for studies of two proteins, only, while (2) is used for bulk solutions consisting of 40 proteins at a concentration of 64 μM. The angularly averaged potential of mean force, $w(R)$, between proteins is calculated from the radial distribution function, $g(R) = \exp(-w(R)/kT)$, where R is the protein-protein mass center separation. Due to strong inter-protein attraction, $g(R)$, is sampled in divided windows. The probability, P, of finding a protein in an aggregate of size N is evalu-

ated from bulk simulations: If the separation of two amino acids of two different proteins is shorter than $\sigma_{ij}+2$ Å, the proteins are considered to be part of the same aggregate.

Results and Discussion

Twobody Interactions

In a previous study we have shown that bovine lactoferrin aggregates to form dimers at pH 7 and 5 mM 1:1 salt.[12] We now revisit this result and elaborate further on the interaction mechanism leading to dimer formation.

Figure 2 shows a decomposition of the angularly averaged interaction energy, as well as the free energy (PMF) as a function of the protein-protein separation. The full PMF – where both electrostatic and van der Waals (vdW) interactions are included – shows a distinct and narrow minimum, corresponding to a tightly bound and highly stereo-specific dimer configuration.[12] Artificially turning off either electrostatic or vdW interactions completely removes this minimum, showing that both interactions are needed. Individually, neither vdW nor electrostatics lead to specific dimer formation but when combined, the two terms *couple* due to non-linearity of the Boltzmann weight. From Figure 2 it is also anticipated that the proteins have a large contact area at the free energy minimum and that few configurations dominate the interaction free energy.

To investigate if this is a general result for lactoferrin and if the vdW-electrostatic cooperativity exists for other variants of lactoferrin we calculate the PMF also for the buffalo, horse and human forms – see Figure 3. All four proteins display narrow PMF minima at around 45 Å and, compared to the case where electrostatics are turned off, these minima are deep and shifted to shorter separations. While the trends are similar for all four variants, the location of the barrier differs with bovine lactoferrin having the largest barrier and horse lactoferrin the lowest. This correlates well with the protein net charges listed in Table 1.

The *depths* of the minima, however, do not show a similar correlation with the net charge since the highest charged protein (bovine) has a similar depth as the lowest (horse). This is because

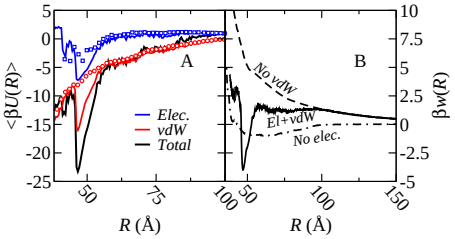


Figure 2: *Left:* Contributions to the average energy, $\langle U \rangle$, as a function of mass-center separation, R. Blue symbols indicate the electrostatic contribution in the absence of vdW interactions, while red symbols indicate vdW interactions in the absence of electrostatics. Simulated at 5 mM salt and pH 7. *Right:* Potential of mean force with and without electrostatics and vdW interactions.

Table 1: Protein net charge numbers at pH 7 for genetic lactoferrin variants at two different 1:1 salt concentrations. Calculations performed using MC titration simulations with 200 salt pairs.[21]

	5 mM	25 mM
Bovine	13.8	14.3
Buffalo	9.4	9.7
Horse	7.9	8.3
Human	9.8	10.2

the source of the attraction is in all four cases the mutual amplification of the electrostatic and vdW energy components, stemming from high surface complementarity of the dimer. For all proteins, increasing the salt concentration to 25 mM completely suppresses the barrier. At the same time the minimum decreases for bovine while it increases for the three other variants.

Many-body Interactions

We now proceed to study the simultaneous interaction of *many* proteins in a bulk solution. Despite that each lactoferrin has around 700 amino acid residues we simulate up to 40 molecules at a concentration of 64 μM, corresponding to a volume fraction of 0.004 by assuming a molec-

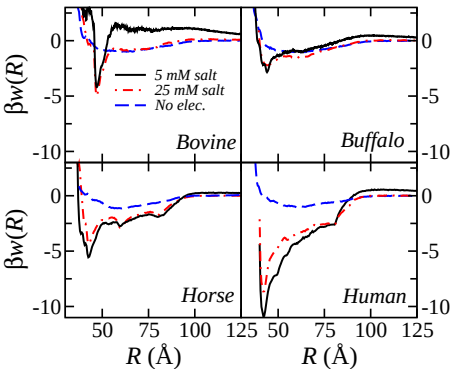


Figure 3: Potential of mean force for four generic variants of lactoferrin at pH 7 at different 1:1 salt concentrations as well as with electrostatic interactions turned off.

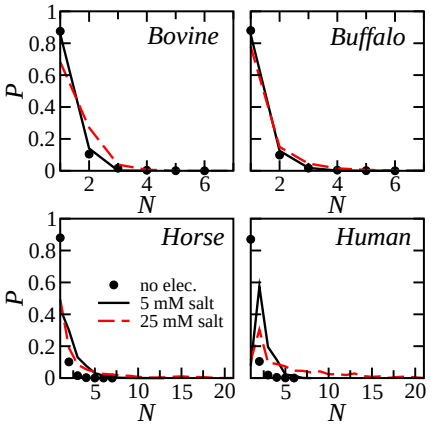


Figure 4: Probability distribution of protein aggregation number, N, for four genetic variants of lactoferrin (pH 7) and at different salt concentrations as well as when electrostatics are turned off.

ular weight of 78 kDa and a specific volume of 0.7235 ml/g.[11] These 30.000 particle simulations are possible with the aid of the previously described multiscale interaction model (Figure 1) that greatly reduces the simulation time, while giving identical results as a full amino acid particle description.

Figure 4 shows aggregate distributions at two different salt concentrations and for all four variants. The aggregation of human and horse lactoferrin is remarkable and in both cases the probability of finding oligomers is substantially higher at low salt than for bovine and buffalo. When the salt concentration is increased to 25 mM, the electrostatic barrier is reduced. This leads to an expansion of the aggregation profile and aggregates as large as 20 proteins are recorded. In the case of human lactoferrin, dimers are more common than monomers at any condition. This is noteworthy, considering the rather dilute protein concentration. The same trend is found for bovine and buffalo lactoferrin, though less pronounced.

Figure 5 shows the potential of mean force in human and bovine lactoferrin solutions at 64 μM. In contrast to $w(R)$ shown in Figure 3 the effective interaction is diminished due to many-body protein interactions. Human lactoferrin – which forms large aggregates – still interact via relatively long ranged interactions and several distinct minima are observed, corresponding to first, second and third neighbors. Bovine lactoferrin interacts much less and is dominated by monomers, although the $w(R)$ does have two minor minima, corresponding to second and third interaction shells. Turning off electrostatics, human and bovine lactoferrin have very similar interactions with no specific binding modes and only weak attractive interactions. While for human lactoferrin, the attraction *decreases* significantly, it does not change for bovine. Note though that removal of electrostatics in both cases retards all specific binding modes.

Experimentally, it has been shown that human and bovine lactoferrin form 100 nm sized aggregates as the salt concentration is increased from 1 to 10 mM.[10] Aggregates of this size are of course not seen in the present simulations due to the limited number of particles. At present it is therefore not possible to infer if the model will give rise to such large aggregates. In a previous study of the osmotic pressure of lactoferrin solutions[11] it was necessary to introduce an extra attractive term of unknown origin to explain the experimental results, despite already accounting for excluded volume-, electrostatic- and vdW interactions. However, this was for an isotropic

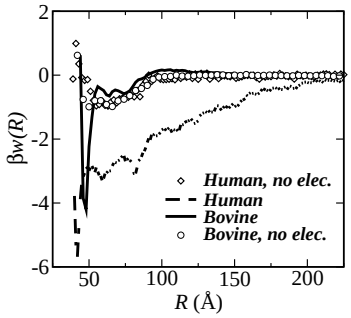


Figure 5: Inter-protein potential of mean force between lactoferrin molecules in 64 μM protein solutions of human lactoferrin and bovine lactoferrin, respectively. The salt concentration is 25 mM and pH is 7.

interaction model. We suggest that this extra attraction may very well stem from the vdW-electrostatic coupling mechanism described in this work. The presented decomposition of the protein-protein interaction shows that a DLVO-approach with an overlap approximation of the electrostatic and vdW interactions will be qualitatively wrong; the prolate shape and anisotropic charge distribution of lactoferrin simply do not allow for a centro-symmetric description. In the model presented here, correlations between the electrostatic and vdW interactions arise naturally and the aggregation is explained with no need for extra attractive terms.

The fact that anisotropic interactions must be included even at large separations is due to ion-dipole correlations of lactoferrin which has a significant dipole moment as shown in Figure 6. Interestingly, the molecular dipole moments of all lactoferrin variants are at maximum exactly at physiological pH. This may not be entirely coincidental, but rather an example of an evolutionary design, tailored to tune biologically important interactions.

The simulated isoelectric points, pI, of the lactoferrin variants – see Figure 6 – are all around nine which is in good agreement with experimental findings (8.0-8.5).[10] In the experimental data, pI varies significantly with salt concentration as well as salt type,[10] which is likely due to specific ion binding to the protein surface. This effect is missing in our implicit salt model.

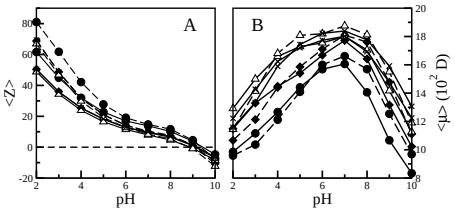


Figure 6: Titration curves of different sources of lactoferrin. Dots: Bovine, Diamonds: Buffalo, Triangles: Horse and Crosses: Human. Full lines are for 5 mM salt and staggered 100 mM salt. A: Average net valency, B: Average dipole moment, relative center of mass, in Debye.

Conclusion

The above simulation results strongly suggest that despite a substantial protein net charge, lactoferrin – obtained from four different genetic sources – forms dimers as well as higher order aggregates. The protein-protein attraction is dominated by specific orientations where van der Waals and electrostatic interactions reinforce each other due to non-linear coupling. This is manifested in deep and narrow free energy minima at short protein separations. The attraction leads to a stabilization of aggregates and, as the salt concentration is increased, the electrostatic barrier stemming from the net charge diminishes.

Protein aggregation is controlled by an intricate balance between a range of interaction mechanisms and we have here demonstrated that a detailed knowledge of protein structure and charge distribution is a prerequisite for studying their interactions. Although this study underpins that subtle amino acid interactions at short separations are crucial for the formed protein-protein complex, all four genetic variants of lactoferrin show similar interaction patterns. This suggests that the model and coarse graining level is robust. To study systems of experimental relevance, simplified models for proteins and their mutual interactions are

needed; the present work describes one possible route.

Acknowledgement The authors wish to thank the Linneaus center Organizing Molecular Matter in Lund, Sweden as well as the LUNARC computational facility.

References

(1) Sørensen, M.; Sørensen, S. P. L. *C. R. Trav. Lab. Carlsberg* **1939**, *23*, 55–99.

(2) Brock, J. H. *Biochemistry and Cell biology* **2002**, *80*, 1–6.

(3) Babina, S. E.; Tuzikov, F. V.; Tuzikova, N. A.; Buneva, V. N.; Nevinskii, G. A. *Molecular Biology* **2006**, *40*, 121–131.

(4) Fox, P. F.; McSweeney, P. L. H. In *Advanced Dairy Chemistry*, 3rd ed.; Fox, P. F., McSweeney, P. L. H., Eds.; Kluwer Academic/Plenum Publishers, 2003; Vol. 1: Proteins.

(5) Querinjean, P.; Masson, P. L.; Heremans, J. F. *Eur J Biochem* **1971**, *20*, 420–425.

(6) Szuchet-Derechin, S.; Johnson, P. *European Polymer Journal* **1965**, *1*, 283–291.

(7) Schanbacher, F. L.; Smith, K. L.; Ferguson, L. C. *Fed. Proc.* **1971**, *30*, 532.

(8) Harmon, R. J.; Schanbacher, F. L.; Fergusson, L. C.; Larry, K. *Infection and Immunity* **1976**, *13*, 533–542.

(9) Wang, H.; Hurley, W. L. *J Dairy Sci* **1998**, *81*, 1896–1903.

(10) Mela, I.; Aumaitre, E.; Williamson, A.-M.; Yakubov, G. E. *Coll. Surf. B: Bioint.* **2010**, *78*, 53.

(11) Bowen, W. R.; Cao, X.; Williams, P. M. *Proc. R. Soc. Lond. A* **1999**, *455*, 2933.

(12) Persson, B. A.; Lund, M.; Forsman, J.; Chatterton, D. E.; Åkesson, T. *Biophys. Chem.* **2010**, *151*, 187.

(13) Metropolis, N. A.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1097.

(14) Frenkel, D.; Smith, B. *Understanding Molecular Simulations*; Academic Press, 2002.

(15) Persson, B. A.; Lund, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8879.

(16) *Paper VII*

(17) Moore, S. A.; R, B. F. A. C.; Groom,; Haridas, M.; Baker, E. N. *J. Mol. Biol.* **1997**, *274*, 222.

(18) Karthikeyan, S.; Paramasivam, M.; Yadav, S.; Srinivasan, A.; Singh, T. P. *Acta Crystallogr.,Sect.D* **1999**, *55*, 1805.

(19) Sharma, A. K.; Paramasivam, M.; Srinivasan, A.; Yadav, M. P.; Singh, T. P. *J. Mol. Biol.* **1999**, *289*, 303.

(20) Norris, G. E.; Anderson, B. F.; Baker, E. N. *Acta Crystallogr.,Sect.B* **1991**, *47*, 998.

(21) Lund, M.; Jönsson, B. *Biochemistry* **2005**, *44*, 5722–7.

(22) Trulsson, M.; Forsman, J.; Åkesson, T.; Jönsson, B. *Langmuir* **2009**, *25*, 6106.

(23) Lund, M.; Trulsson, M.; Persson, B. *Source Code for Biology and Medicine* **2008**, *3*, 1.

(24) Liu, J.; Luijten, E. *Phys. Rev. Lett.* **2004**, *92*, 035504.

(25) Liang, Y.; Bowen, W. R. *J. Coll. Interface Sci.* **2005**, *284*, 157.