Electrostatic Interactions In and Between Biomolecules

Lund, Mikael

2006

Electrostatic Interactions In and Between Biomolecules

Mikael Lund
Department of Theoretical Chemistry
Lund University – Sweden

The defense is open to the public and will take place Monday 11th of December 2006, 13:15 in Auditorium B, Chemical Center, University of Lund, Sweden.
The faculty opponent is Prof. Dusan Bratko - Virginia Commonwealth University.

Organization LUND UNIVERSITY	Document name DOCTORAL DISSERTATION
	Date of issue Dec. 11, 2006
Author(s) Mikael Lund	Sponsoring organization The Research School in Pharmaceutical Sciences (FLÄK).

Title and subtitle

Electrostatic Interactions In and Between Biomolecules

Abstract

Electrostatic interactions in bio-molecular systems are important not only in the living cell but also in more technical applications. Using molecular simulation as well as approximate theories the properties of a number of aqueous protein solutions have been studied. This include interactions with other proteins, protons, charged membranes as well as flexible polyelectrolytes. The focus is on electrostatic interactions and special attention is put on charge regulation. I.e. how the protonation state of a biomolecule is influenced by nearby charged species. We show that this gives an important contribution to the free energy and that the mechanism can be accounted for by a simple statistical mechanical model. In particular we introduce the concept of "protein capacitance" that is the key intrinsic property for quantifying the charge regulation.

Key words:

Molecular modelling, charge regulation, protein interactions

Classification system and/or index termes (if any):

Supplementary bibliographical information:	Language English
ISSN and key title:	ISBN 91-7422-130-2

Recipient's notes	Number of pages	Price
	Security classification	

DOKUMENTDATABLAD enl SIS 61 41 21

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 Mikael Lund Date 19 Oct. 2006

Printed by Media Tryck in Lund

ISBN 91-7422-130-2

Contents

Preface

This text represents the formal work of my doctoral studies at the Department of Theoretical Chemistry at Lund University in the period 2002–2006. The structure of this document more or less follows the "standard" procedure with an introduction followed by some scientific papers, describing the research results in a fairly concise manner. Since the important findings are described in these latter articles the initial sections shall be used to elaborate mostly on general concepts and themes not covered explicitly in the papers. Occasionally I will – briefly – relate to the papers but these will not be the main center of attention. Reading through this thesis as well as the attached papers, it will become apparent that a major theme revolves around molecular coarse graining. That is, to transform a complex system into something more simple, yet realistic, that can be investigated using clear physical methods – preferably with a minimal set of operational parameters. This approach is in sharp contrast to an increasing amount of biophysical research with the mantra "more detail". While such undertakings are valiant one should always weigh input versus output and the paramount measure of scientific quality must be the *gain in physical insight*.

Interactions between molecules of biological origin are important not only for processes taking place in living cells, but can also be utilized in a number of more technical applications. Proteins are complex molecules and their mutual interactions can be predicted only if a solid foundation of the basic, physical mechanisms is established. For example, why does a protein solution precipitate upon addition of multivalent salt? And why is this precipitate sometimes crystalline, other times amorphous? The answers to these questions are important not only to crystallographers, attempting to produce protein crystals, but also many diseases are connected with malicious protein aggregation. For example, the formation of amyloid fibrils, as evident in Alzheimer's disease, is known to be enhanced by high valency metal ions and attempting to intelligently design a suitable cure, knowledge of the underlying mechanisms is indispensable. Designing new drugs typically involves a large number of candidate compounds targeting bio-molecules such as enzymes or membranes. Probing the binding affinity via theoretical means requires a description of the intermolecular interactions that, in many cases, are profoundly influenced by electrostatics. For example, the protein net charge is sensitive to the solution pH and if the targeting molecule is also charged, long ranged electrostatic interactions can significantly affect binding properties. A related *in vivo* example is hisactophilin binding to a lipid membrane. Here, the stability of the protein-membrane complex is governed mainly by electrostatic interactions and even small intracellular pH changes can influence this equilibrium.

Another area where bio-molecular interactions enjoy much attention is in food sciences. Proteins are essential constituents of our diet and properties such as solution stability, viscosity, emulsions etc. are important when designing food products. In contrast to biological systems, there is a reasonably large degree of freedom in which the solution conditions can be varied – *i.e.* pH, salt concentration and valency, temperature, and addition of other macro-molecules such as polymers. While the physical properties can be measured via thermodynamic techniques, it is from these results difficult to decipher what is taking place at the molecular level. This gap between macroscopic and microscopic properties, is accounted for by statistical mechanics and as shall be shown in this thesis, this is a very useful tool for studying *molecular mechanisms.*

Now, during the past four to five years a number of fine people have – in one way or

another – contributed to this outcome and I will now spend a few (Danish) words acknowledging these. **Med** udsigt til mange års ekstra studier, kurser og svensk valuta, var det med tvivlende skridt, jeg indtog skrivebordet i Svinestien for 4 1/2 år siden. Trods den mærkelige dialekt man talte her, blev det dog hurtigt klart, at de svenskere nu ikke var helt så slemme endda. Faktisk, har det været en udsøgt fornøjelse at færdes på både Teoretisk- såvel som Biofysisk kemi, hvilket her fjernet enhver tvivl om mit valg, og endda opvejet de ca. 2000 timer ens luksuslegeme er blevet udsat for Øresundstogenes tvivlsomme komfort. Tak for det! Af mere videnskabelig karakter, har både vejleder Bo, samt Torbjörn, Cliff og Hr. Forsman(n) været uvurderlige kapaciteter, og altid udvist stor velvilje samt pågående kritik. Herunder skal naturligvis også nævnes alle andre fra statmek gruppen: Martin T & Martin T, Gunnar, Magnus U, Fernando, Asbjørn, Daniel, Monsieur Labbez og laboratorievennerne Sara, Tonu og Ingemar. Jeg takker for alle kaffepointene, ophold i "Tåget" og på "Filippa In", indføring i tykmælkens lyksaligheder, "kemisk potatis", gode råd om træhusets finurligheder og alt muligt andet stort, som småt. Også en stor tak til min tidligere vejleder og køkkenkemiker, Thorvald Pedersen, som er den egentlige årsag til at jeg havnede i Lund. Udenfor universitetets mure har familie og trofaste venner altid ventet med behageligt og til tider lummert selskab. Tak er kun et fattigt ord. **Last**, but definitely not least I wish to thank Miss Forecast, Wonkeun, Chinmayee, Bianka, Juliette, Dan and of course Cliff for making my six months in Canberra feel more like six weeks.

M.L.
November 2006

Dansk sammenfatning

Elektrostatiske vekselvirkninger mellem biomolekyler er vigtige, ikke bare i biologiske systemer, men også i flere tekniske sammenhænge. I denne afhandling studereres en række forskellige egenskaber, relevant for særligt proteiner. Således har vi ved hjælp af molekylær simulering undersøgt hvorledes proteiner vekselvirker med protoner, salte, andre peptider og proteiner, ladede membraner samt polymerer. Som det fremgår af titlen er der lagt vægt på elektrostatik, dvs. vekselvirkninger mellem ladede partikler, men andre afledede former – van der Waals vekselvirkninger mfl. – er også behandlet.

En særlig egenskab ved proteiner og andre bio-molekyler, er at de indeholder titrerbare grupper, hvis protoniseringstilstande afhænger ikke bare af opløsningens pH værdi, men også af det elektriske potential genereret af omkringliggende, ladede makromolekyler. Dette indebærer, at når et protein for eksempel nærmer sig at andet protein, et DNA molekyler, eller en ladet membran, vil dets elektriske ladning ændres. Denne mekanisme kaldes *ladningsregulering*, og bidrager til en sænkning af systemets frie energi. Denne mekanisme er studeret vha. Monte Carlo simuleringer af proteinopløsninger, og vi har endvidere udarbejdet en tilnærmet model, baseret på statistisk termodynamik. Ud fra disse teoretiske betragtninger, har vi udledt en ny protein egenskab – *ladningskapacitansen* – som er helt på linje med mere velkendte egenskaber som den totale ladning og dipolmomentet. Kapacitancen kan findes ved at differenciere den målte, eller beregnede protein titreringskurve, og kan anvendes til at estimere hvordan proteinet påvirkes af andre ladninger i systemet.

Derudover har vi undersøgt hvorledes protein-protein vekselvirkninger påvirkes af pH, salt koncentration og ikke mindst salttype. Ved at beregne den frie energi som funktion af den indbyrdes afstand mellem proteinmolekylerne, kan vi opdrive den *anden virial koefficient*, B_2, som er en velkendt indikator for den overordnede vekselvirkning. B_2 kan måles eksperimentelt vha. lysspredning, og sådanne målinger (udført af andre forskningsgrupper) er derfor anvendt til at verificere vores modeller; oftest foreligger glimrende overensstemmelse. Ved tilsætning af trivalente ioner, kan makromolekyler af samme (høje) ladning bringes til at aggregere. Ifølge vores studier er dette pga. ion-ion korrelationer, der ikke lader sig beskrive med gængse Poisson-Boltzmann metoder, men kræver istedet en eksplicit saltbeskrivelse.

I udarbejdelsen af de teoretiske modeller, er der lagt vægt på at disse skal være så simple som mulige, men samtidig beskrive systemet i en sådan grad af detalje, så alle væsentlige egenskaber bibeholdes. Sådanne molekylære forenklinger – *coarse graining* på engelsk – reducerer ikke bare antallet af justerbare parametre, men kræver også væsentlig mindre regnekraft og bidrager med et klart billede af de basale fysiske mekanismer.

Symbols and Abbreviations

A, w	Free energy or effective potential
β	$1/kT$
B	Virial coefficient
c_x	Concentration of the component x
C	Charge capacitance
e	Electron unit charge
ϵ_0	Permittivity of vacuum
ϵ_r	Relative dielectric constant
ϕ	Electric potential
γ	Activity coefficient
$g(R)$	Radial distribution function
k	Boltzmann's constant
K^*	Stoichiometric equilibrium constant
κ	Inverse Debye screening length
l_B	Bjerrum length
μ	Electric dipole moment
μ^{ex}	Excess chemical potential
M_w	Molecular weight
Q, q	Charge
R, r	Distance
$\mathcal{R}^N$	Coordinate space *i.e.* $[\mathbf{r}_1 \dots \mathbf{r}_N]$
σ	Particle diameter
S	Entropy
T	Temperature
U, u	Energy
Z, z	Charge number or valence
$\langle ... \rangle_x$	Statistical mechanical average over x

DH	Debye-Hückel
DLVO	Derjaguin, Landau, Vervey and Overbeek
MC	Monte Carlo
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance

List of Papers

1. **A Mesoscopic Model for Protein-Protein Interactions in Solution**
   Mikael Lund and Bo Jönsson
   Biophysical Journal 85:2940–2947 (2003)

2. **On the Charge Regulation of Proteins**
   Mikael Lund and Bo Jönsson
   Biochemistry 44:5722–5727 (2005)

3. **Enhanced Protein Adsorption Due to Charge Regulation**
   Mikael Lund, Torbjörn Åkesson, and Bo Jönsson
   Langmuir 21:8385–8388 (2005)

4. **On the Complexation of Proteins and Polyelectrolytes**
   Fernando L. B. da Silva, Mikael Lund, Bo Jönsson, and Torbjörn Åkesson
   Journal of Physical Chemistry B 110:4459–4464 (2006)

5. **Implications of a High Dielectric Constant in Proteins**
   Mikael Lund, Bo Jönsson, and Clifford E. Woodward
   Submitted (2006).

6. **Driving Forces Behind Ion-Ion Correlations**
   Mikael Lund and Bo Jönsson
   Journal of Chemical Physics, Accepted for publication (2006).

7. **Electrostatic Interactions Between Proteins in a Salt Solution - A Monte Carlo Simulation Study**
   Mikael Lund and Bo Jönsson
   Manuscript.

8. **Electrostatics in Macromolecular Solutions**
   Bo Jönsson, Mikael Lund, and Fernando L. B. da Silva
   Food Colloids: Self-Assembly and Material Science
   Eds: E. Dickinson and M. E. Leser, Royal Society of Chemistry, Cambridge. In Press.

9. **Monte Carlo simulering – fra proteiner til cement**
   Mikael Lund and Bo Jönsson
   Dansk Kemi 6/7 (2006).

Publications Not Included in this Thesis

10. **Coarse Graining Biomolecular Systems**
    Mikael Lund
    Springer Lecture Notes in Computer Science, Submitted (2006).

11. **Activity Coefficients in Sea Water Using Monte Carlo Simulations**
    Mikael Lund, Bo Jönsson, and Thorvald Pedersen
    Marine Chemistry 80:95-101 (2003).

12. **Fluxes of Ammonia in the Coastal Marine Boundary Layer**
LISE-LOTTE SØRENSEN, OLE HERTEL, MIKAEL LUND, AND BRITTA PEDERSEN
Atmospheric Environment 37S1:167-177 (2003).

13. **Improved SO_2 Control with New Method for At-line Measurements**
BIRTHE SKANDS, KJELD PEDERSEN, MIKAEL LUND, AND THORVALD PEDERSEN
Proceedings from the 29th Congress of the European Brewery Convention (2003).

14. **Ammonia Deposition in a Coastal Area**
LISE-LOTTE SØRENSEN, BRITTA PEDERSEN, AND MIKAEL LUND
Coastal Air Pollution Meteorology and Air-sea Nutrient Exchange. Annual report 1999. Geernaert, G.L. (Ed.). Technical report, GSF-Forschungszentrum für Umvelt und Gesundheit Gmbh. EUROTRAC-2 International Scientific Secretariat, München. (2000).

1 Concerning Proteins

1.1 Various Views

This first part shall be used to describe proteins as seen from different perspectives and even through somewhat blurry glasses. I will point out and identify essential features that can be quantified using a well tested toolbox of physical methods. So, what is a protein? A chemist would merely call it a big molecule as it consists of one or more sequences of amino acid residues, linked through covalent peptide bonds[1]. This backbone has side chains and due to different chemical properties of the individual amino acids, the chain(s) can coil to form a well defined macromolecular structure. In hydrophilic, globular proteins the structure is constructed so that non-polar parts of the chains are well hidden in the protein interior, while more polar groups are solvated – *i.e.* exposed to the (aqueous) solvent at the protein surface. Polar or hydrophilic groups can be small parts of the macromolecule where electronic displacements create local dipole moments or actual charges on (de)protonated acidic or alkaline amino acid residues. Furthermore, various multivalent metal ions can be incorporated. By now it should be clear that a protein is full of (partial) charges and it seems only reasonable to assume that electrostatics may play an important role in biological systems[2,3].

Of similar importance is the surrounding solvent and other solutes that may significantly influence the electrostatic behavior of the protein. In the living cell, proteins are solvated in an aqueous salt solution containing roughly 0.15 M salt as well as an appreciable amount of other macromolecules. The proton concentration or pH is an essential property of the solution as it can be used to control the protonation status – and hence the charge – of acidic and alkaline (*titratable*) groups in the protein. In fact, for certain proteins even small intracellular pH variations can change the sign of the overall charge (Paper 3). How the protein charge varies with pH is illustrated by the titration curve as normally measured electrochemically in a titration experiment or constructed from individual dissociation constants (pK_a) of the titratable sites. The latter approach is more elaborate as the determination of individual pK_a values require extensive NMR studies.

Now that we have an idea of what a protein is, we should be able to draw one. And this is where the different points of view become evident. Scientists from different fields tend to come up with their own interpretations and Figure 1 illustrates how the same protein may look in various disciplines. What is shown is in fact a kind of *coarse graining*[4] where the most detailed model (quantum level) is transformed into more simple alternatives, tailored to capture only properties of interest. This is a powerful technique as the mathematical complexity can be drastically reduced (or even removed) yet it still provides useful insight. It can also be a pitfall, though. It is not uncommon to see coarse graining applied to problems where it is not suitable. For example, it would be silly to measure protein aggregation using the "cartoon model" as it provides no controlled way of quantifying intermolecular interactions. Likewise

[1] T. E. Creighton. *Proteins - Structures and Molecular Properties*. W.H. Freeman, New York, 2nd edition, 1993.

[2] Jonathan P K Doye, Ard A Louis, and Michele Vendruscolo. Inhibition of protein crystallization by evolutionary negative design. *Physical Biology*, 1(1):P9–P13, 2004.

[3] I. Gitlin, J. D. Carbeck, and G. M. Whitesides. Why are proteins charged? networks of charge–charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew. Chem. Int. Ed.*, 45:3022–3060, 2006.

[4] W. C. K. Poon. *Soft Condensed Matter Physics In Molecular And Cell Biology*, chapter 1. Taylor and Francis, New York, first edition, 2006.

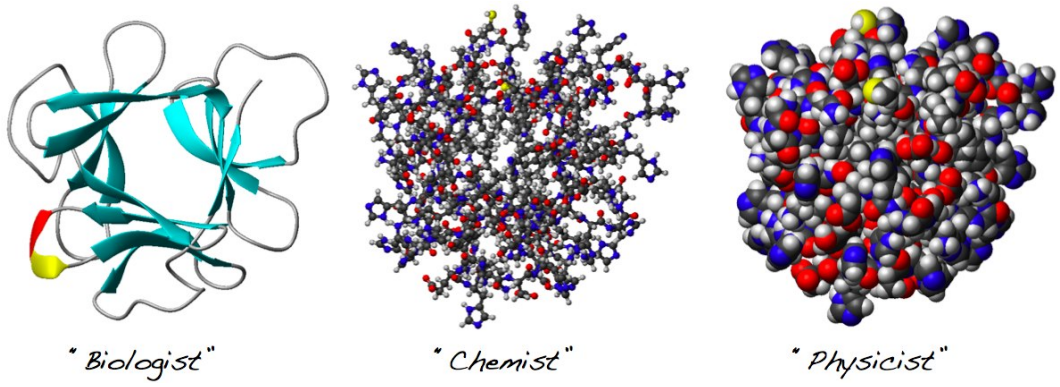


Figure 1: A protein as seen from three different scientific view points: The "cartoon" model (left) showing secondary structure, the ball-and-stick model (middle) illustrating individual atoms and chemical bonds, and finally a space filling model (right), where each atom is presented by a sphere.

it would seem excessive to use a full quantum mechanical model as we deal with long range interactions and a detailed description of internal electron densities is not really needed (nor tractable for that matter). The trick is to find the right level of detail for the problem at hand, and at the same time ensure that the results are not artificially modulated by an excessive number of input parameters.

1.2 More Coarse Graining

The "biologist", "chemist", and "physicist" models shown in Figure 1 all illustrate – some better than others – that proteins have a detailed surface topology and indeed occupy a volume from which other molecules are excluded. This gives rise to significant intermolecular interactions and any quantifying model should capture this effect. From a computational point of view it is appealing to coarse grain as much as possible yet preserve dominant physical properties – in this case the excluded volume. The "colloid scientist" would (possibly) replace the protein with a sphere matching the protein volume so as to reduce the number of particles from thousands to merely one[5,6]. For certain properties this might be a reasonable approximation but obviously it is a poor description for more elongated molecules. Another important disadvantage of the spherical model is that the detailed charge distribution is usually replaced by a single point charge, thus neglecting possible electric multipole moments. Figure 2 shows other alternatives; in particular we note the amino acid model where all amino acid residues in the protein are replaced by a sphere. This method was scrutinized in Paper 1 and – while seemingly crude – it captures many details of the protein and is advantageous for several reasons: 1) the surface topology is well preserved (see Figure 3), 2) the detailed charge distribution is maintained, 3) it allows for angularly dependent short range interactions (van der Waals, for example), and 4) the number of particles is reduced by one order of magnitude compared to the atomistic representation, allowing even large proteins to be handled, 5) and finally, the calculated potential of mean force between proteins seems to be identical to that of the all-atom model – see page 44 (Paper 1).

[5] B. V. Derjaguin and L. Landau. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Acta Phys. Chim. URSS*, 14:633–662, 1941.

[6] E. J. W. Verwey and J. Th. G. Overbeek. *Theory of the Stability of Lyophobic Colloids*. Elsevier Publishing Company Inc., Amsterdam, 1948.

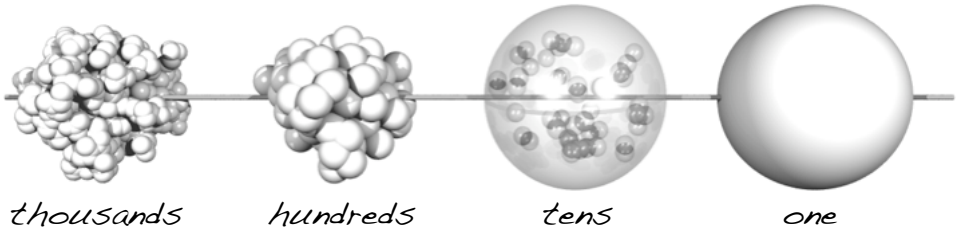


Figure 2: Various protein models and the number of particles involved. From left to right: 1) Atomistic model where all atoms are represented by spheres. 2) Amino acid model where entire amino acid residues are approximated by spheres. 3) Point charges in an encapsulating sphere. 4) A sphere with a point charge in the middle (DLVO type).

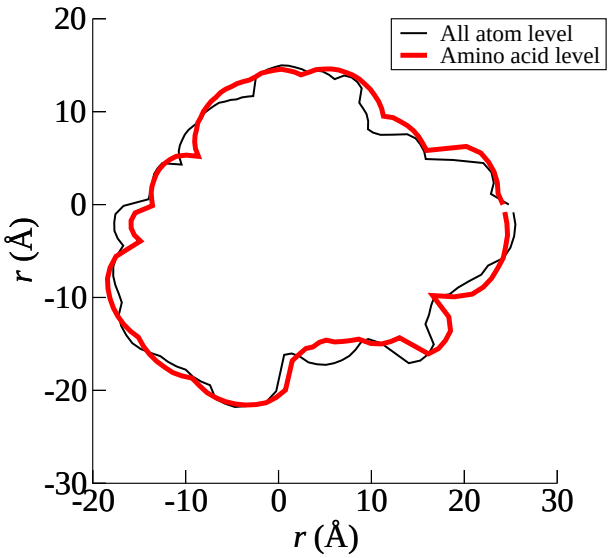


Figure 3: Cross section of the protein ribonuclease using both an atomistic- and an amino acid representation.

1.3 Rigidity

The models proposed so far all neglect structural degrees of freedom within the protein molecule. Clearly this is an approximation as side chains do have some flexibility and certain proteins are known to function via structural perturbations. To incorporate this we need a set of energy functions for the stretching, twisting and bending of intra-molecular bonds in the protein. Such *force fields*[7,8] are typically constructed in a semi-empirical manner and may contain hundreds of parameters where some are suitable for certain molecules, some are not. This is an important matter and one should thoroughly make sure that a given set of parameters do not artificially influence the final results.

In cases where structural fluctuations are minor or irrelevant for the properties of interest the molecule can be kept rigid. For a number of proteins NMR studies indicate that the structure is invariant to even large changes in solution conditions (pH, salt).

[7] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2):187–217, 2004.

[8] D.A. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, Jr. K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods. The amber biomolecular simulation programs. *J. Computat. Chem.*, 26:1668–1688, 2005.

In theoretical calculations the (fixed) structural coordinates are acquired from either NMR or X-ray crystallography, but these are only partly experimental: After the data collection the experimental constraints are – together with a force field – used in a simulation, so as to obtain the most probable structure. Therefore the "experimental" structure does hold a reminiscence of theory and should be regarded as an average structure. This is especially true for solution NMR structures where the flexible side chains can appear in a multitude of conformations and are often averaged to merely one. These facts support the usage of a rigid but *likely* structure. Nevertheless it is important to recall that a rigid model may becomes less appropriate at denaturating conditions – pH extremes etc.

1.4 Protonation State

A number of amino acid side chains contain titratable groups such as -COOH, $-NH_2$, and –SH with proton affinities covering most of the pH-range (Table 1). Since a protein is just a sequence of amino acid residues, with perhaps hundreds of titratable groups, it can be viewed as a large polyampholyte where the degree of protonation for each site is given by

$$\alpha = \frac{1}{10^{pH-pK_a}+1} \tag{1}$$

The protonation state of one group will influence the state of a neighboring group, which will effect another and so on. Nearby solutes such as salt and charged macromolecules will have a similar effect and the real, *average* protonation state is an asymmetric function as shown in Figure 4 which is also found experimentally[9]. Using a stoichiometric dissociation constant in eq 1 will only cause the titration curve to shift along the pH-axis. A better fit may be obtained by introducing a factor in the exponential so as to flatten the curve. Still, such a "Hill-parameter" of course fails to describe the asymmetry.

Table 1: Acid dissociation constants for "isolated" amino acid residues - *i.e.* their intrinsic values, unaffected by nearby charges and other solutes.

Residue	Abbreviation	pK_a[1]
C-terminal	Ctr	3.8
Aspartic acid	Asp	4.0
Glutamic acid	Glu	4.4
Histidine	His	6.3
N-terminal	Ntr	7.5
Tyrosine	Tyr	9.6
Lysine	Lys	10.4
Cysteine	Cys	10.8
Arginine	Arg	12.0

In a statistical mechanical framework the problem can be formulated as an average protonation state, π for a site, i by an average over all particle positions and protonation

[9] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse. Ionization behaviour of acidic residues in calbindin D_{9k}. *Proteins*, 37:106–115, 1999.

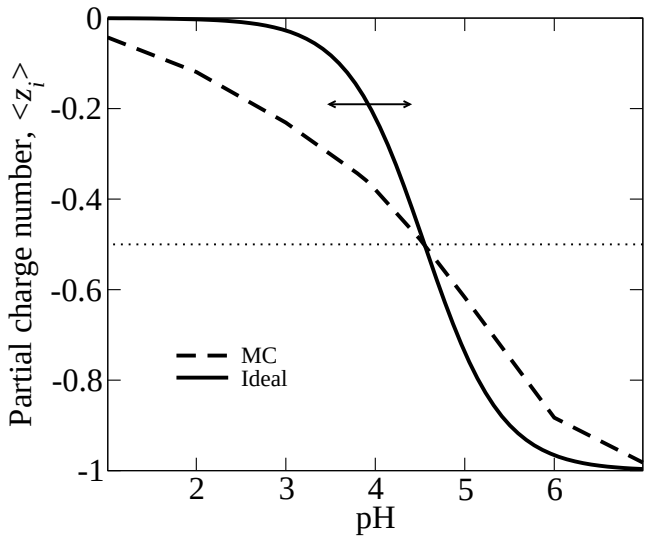


Figure 4: Average protonation state of one of the glutamic acid residues in calbindin, calculated using Monte Carlo simulations. The "ideal" curve is the best fit to Equation 1 – that is pK_a^*=pH for $\langle z_i \rangle = 0.5$.

states,

$$\langle \pi_i \rangle_{\boldsymbol{\pi},\mathcal{R}} = \frac{\int \pi_i e^{-\beta U(\boldsymbol{\pi}^M,\mathcal{R}^N)} d\boldsymbol{\pi}^M d\mathcal{R}^N}{\int e^{-\beta U(\boldsymbol{\pi}^M,\mathcal{R}^N)} d\boldsymbol{\pi}^M d\mathcal{R}^N} \tag{2}$$

Multidimensional integrals like this must be solved numerically, but as a prerequisite we need an energy function for (de)protonating a given site. Starting out with a single site, we shall now use thermodynamic arguments to derive equations for just that. For any given monoprotic acid one can write the dissociation process

$$\mathrm{HA} \rightleftharpoons \mathrm{H^+} + \mathrm{A^-}$$

where the thermodynamic equilibrium constant is defined as

$$K_a = \overbrace{\frac{\gamma_{\mathrm{H^+}}\gamma_{\mathrm{A^-}}}{\gamma_{\mathrm{HA}}}}^{\Gamma} \cdot \overbrace{\frac{c_{\mathrm{H^+}}c_{\mathrm{A^-}}}{c_{\mathrm{HA}}}}^{K_a^*} \Leftrightarrow pK_a = -\log\Gamma - \log\frac{c_{\mathrm{H^+}}c_{\mathrm{A^-}}}{c_{\mathrm{HA}}} \tag{3}$$

If we now define pH as the negative logarithm of the proton *concentration*[1] we get that

$$-\log\frac{c_{\mathrm{A^-}}}{c_{\mathrm{HA}}} = \log\Gamma - (pH - pK_a) \tag{4}$$

The fraction $c_{\mathrm{A^-}}/c_{\mathrm{HA}}$ is nothing but the probability of deprotonating the acid and can immediately be transformed to a free energy change,

$$\beta\Delta A_{HA\to A^-} = -\ln\frac{c_{\mathrm{A^-}}}{c_{\mathrm{HA}}} = \ln\Gamma - (pH - pK_a)\cdot\ln 10 \tag{5}$$

The first term on the right hand side is identified as the sum of excess chemical potentials ($\mu^{ex} = \ln\gamma$) for the three species HA, $\mathrm{A^-}$, and $\mathrm{H^+}$ and thus accounts for all effects not captured in the reference state - *i.e.* the intrinsic or thermodynamic dissociation constant, K_a. This is an important observation and implies that:

[1]Note the deviation from the formal definition, $-\log a_{H^+}$, done solely to conform with later Monte Carlo simulations – see section 3.2.

- The excess chemical potentials should contain only the *difference* between the model compound[2] and the residue in the protein (solution).

- Changes in charge solvation can most likely be neglected as these are approximately the same in the protein and in the model compound. Exceptions are deeply buried groups.

- Γ is a function of pH and salt.

The last point – that Γ and hence the stoichiometric dissociation constant is a function of pH – is the origin of the asymmetric nature of the single-site titration curve. Single-site titration behavior will be picked up again in Section 3.2 where we will use the above findings to construct a Monte Carlo simulation method for evaluating eq 2.

1.4.1 Total Charge

Now, the *total* protein charge is just the sum of all single-site charges which are – as we have just seen – governed by factors such as pH, other solutes as well as protein structure and sequence. This is of great importance for the solution stability – it is for example well known that proteins aggregate when pH approaches the iso-electric point[3] or when mixed with multivalent ions. The overall protonation state of a protein is usually measured electrochemically[10], while individual sites can be observed using NMR.

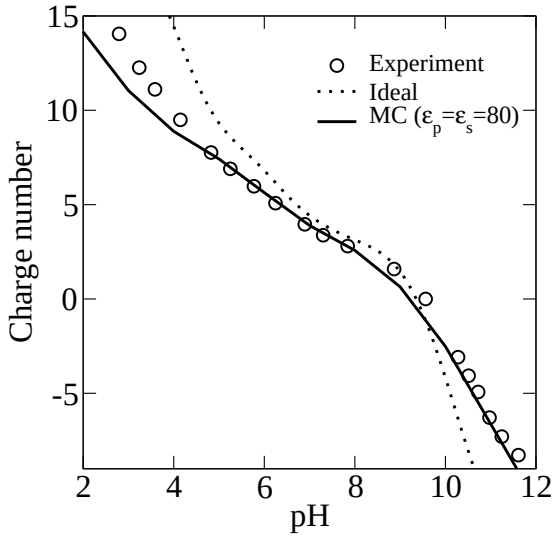


Figure 5: The average net-charge number of ribonuclease as a function of pH as measured using electrochemical methods[10] and calculated using Monte Carlo simulations and a non-interacting model ("ideal").

[2]Intrinsic dissociation constants for amino acid side chains are usually measured for small peptides – model compounds – in a dilute aqueous solution (the reference state). Neighboring groups are chosen so as to disturb as little as possible.

[3]The iso-electric point, pI is defined as the pH value where the overall protein charge Z is zero – that is when all negative and positive charges balance.

[10] C. Tanford and J. D. Hauenstein. Hydrogen ion equilibria of ribonuclease. *J. Am. Chem. Soc.*, 78:5287–5291, 1956.

1.4.2 Dipole Moment

The protonation state of the individual sites not only control the overall protein charge, but also the *electric dipole moment* defined as

$$\bar{\mu} = \sum_i \bar{r}_i z_i. \tag{6}$$

Due to the large size of proteins, dipole moments of several hundred Debye[4] is not un-common and as shown in Figure 6 the pH variation can be substantial. For the elongated "Fab" fragment (PDB entry 3HFL) going from pH 3 to 7 μ increases by a factor of three. In turn this will affect interactions with other multipoles (proteins etc.) and in the case of "fab" it is noteworthy that this happens at physiological pH.

However, there is an issue with eq 6 since if $\sum q_i \neq 0$ the dipole moment will rely on the choice of origin. While this is unfortunate, it rarely poses a problem when estimating protein-protein interactions using a multipole expansion. In this approximation (see section 2.4.1), the distance between the two interacting charge distributions ideally should be much greater than the individual charge vectors ($R >> r_i \wedge r_j$ in Figure 8). This condition is favored by placing origin at the center-of-charge which – for globular proteins – roughly coincides with the center-of-mass.

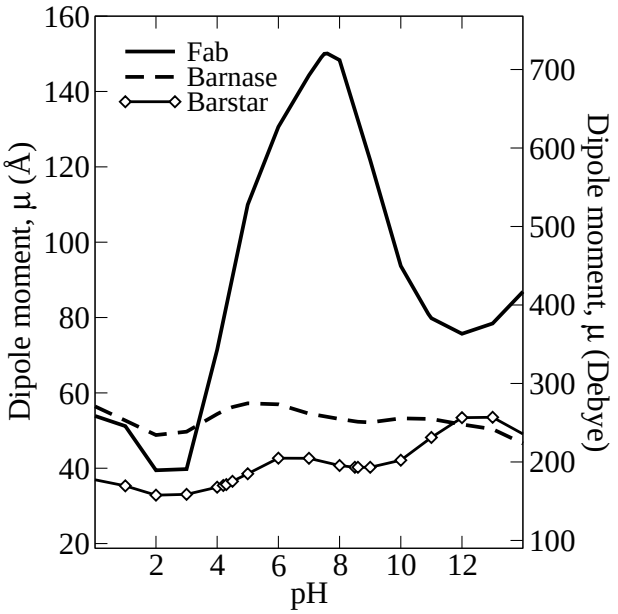


Figure 6: Calculated electric dipole moments for three different proteins at various pH values.

1.4.3 Charge Capacitance

So far we have studied two average protein properties, the total charge and the dipole moment. Since a titratable site can be found in two states these averages will fluctuate around their mean values, leading to interesting physical mechanisms. As for the protein charge the fluctuation is

$$\left\langle Z^2 \right\rangle - \left\langle Z \right\rangle^2 = C \tag{7}$$

[4] 1 D(ebye) = $3.336{\cdot}10^{-30}$ C·m = 0.208 eÅ.

Assuming that there is no salt and that the protein is rigid we can write the net charge, Z as an ensemble average over all protonation states,

$$\langle Z \rangle = \frac{\int Z e^{-\beta U(pH,\boldsymbol{\pi}^M)} d\boldsymbol{\pi}^M}{\int e^{-\beta U(pH,\boldsymbol{\pi}^M)} d\boldsymbol{\pi}^M} \tag{8}$$

If we disturb the system by some small perturbation quantified by α, the response in $\langle Z \rangle$ is given by,

$$\frac{\partial \langle Z \rangle}{\partial \alpha} = \langle Z \rangle \left\langle \frac{\partial \beta U}{\partial \alpha} \right\rangle - \left\langle Z \frac{\partial \beta U}{\partial \alpha} \right\rangle \tag{9}$$

For example, exposing the protein to an external potential, ϕ_{ext} the energy function for a given configuration is

$$\beta U = \sum_i \beta e z_i [\phi_i + \phi_{ext}] \tag{10}$$

Taking the derivative with respect to the external potential: $\partial \beta U / \partial \beta e \phi_{ext} = \sum z_i = Z$ and inserting into eq 9 we arrive at the following relationship which shows that C in fact is a *capacitance*,

$$-\frac{\partial \langle Z \rangle}{\partial \beta e \phi_{ext}} = \langle Z^2 \rangle - \langle Z \rangle^2 = C. \tag{11}$$

This is a general result from linear response theory[11] and states that the response function for a small perturbation is related to spontaneous equilibrium fluctuations. This interesting outcome provides a theoretical basis for the charge regulation mechanism as studied in several of the included papers. The relation can be further expanded by realizing that each site in the protein is affected not only by the local electrostatic environment, but also there is a contribution from the intrinsic pK_0-value as discussed earlier. Viewing this as an external potential, Eq. 10 can be rewritten into,

$$\beta U = \sum_i \beta e z_i \left[\phi_i + \ln 10 (pH - pK_0)\right] \tag{12}$$

and performing the above analysis once again we get that,

$$C = -\ln 10 \frac{\partial Z}{\partial pH}. \tag{13}$$

There are a number of observations related to charge fluctuations:

- C is an intrinsic protein property in league with the total charge and the dipole moment.
- C is a function of the solution conditions as well as protein structure and sequence.
- C can be used to estimate intermolecular interactions. More on this in Section 2.4.1.
- Knowing C we can estimate how an external potential (another protein, DNA etc.) influences the protonation state of the protein. *i.e.* we can calculate the induced charge.
- C can be measured experimentally as the derivative of the protein titration curve.

[11] D. A. McQuarrie. *Statistical Mechanics*. Harper Collins, New York, 1976.

The fluctuation ability for a single titratable site peaks when pH=pK_a and hence the overall capacitance increases when many of such sites are present in the protein. For example, in hisactophilin 26% of the residues are histidines (p$K_{his} \sim 6.3$) resulting in a capacitance maximum close to pH 7 as seen in the top-right figure on page 56. This particular example is unusual as the capacitance for most proteins peaks at high and low pH, corresponding to a majority of basic and acidic groups.

2 Interactions in the condensed phase

2.1 Effective Potentials

As we have just seen, proteins are complex molecules with a multitude of charge dislocations and they can even exist in many protonation states depending on the solution conditions. Still, interactions with and within proteins can be quantified using classical (as in non-quantum) methods for small molecules and colloids. One obvious but important feature of an aqueous solution is that the particle density is *high*; the concentration of water in water is roughly 55 mol/l and we can expect this to significantly influence any interaction taking place. From a computational point of view this is challenging as one must consider thousands of solvent molecules and their mutual interactions throughout coordinate space. This is manageable by "averaging out" some variables, leaving us with *effective potentials* that have the characteristics of free energies and can be derived in a step wise manner,

$$\begin{aligned} \beta A(\mathbf{x}^{n-1}) &= -\ln \int e^{-\beta U(\mathbf{x}^n)} dx_n \\ \beta A(\mathbf{x}^{n-2}) &= -\ln \int e^{-\beta U(\mathbf{x}^n)} dx_n dx_{n-1} \\ &= -\ln \int e^{-\beta A(\mathbf{x}^{n-1})} dx_{n-1} \end{aligned}$$

For example, for polar fluids structural degrees of freedom can be integrated out to reduce the effect of the enormous number of solvent molecules to a single, macroscopic number, namely the relative *dielectric constant*, ϵ_r. Very appropriately, the solvent is now referred to as a dielectric continuum, indicating that structural detail is nonexistent. The (effective) interaction energy between two charges, i and j in the solution is now given by Coulombs law, reduced by a factor of $1/\epsilon_r$:

$$\langle \beta w_{ij} \rangle_{solv} \approx \frac{1}{\epsilon_r} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij} kT} = \frac{l_B z_i z_j}{r_{ij}} \tag{14}$$

In a similar fashion we can average over salt particles (Figure 7) so as to arrive at the classic Debye-Hückel result[12],

$$\langle \beta w_{ij} \rangle_{solv,salt} \approx \frac{l_B z_i z_j}{r_{ij}} e^{-\kappa r_{ij}} \tag{15}$$

where the inverse Debye screening length, κ is proportional to the square root of the ionic strength. Equations 14 and 15 are valid for spherical symmetric charges, but for

[12] P. Debye and E. Huckel. *Z. Physik*, 24:185, 1923.

dipolar molecules – such as proteins – angular averaging produces additional terms – dipole-dipole interactions for example:

$$\left\langle \beta w_{ij}^{dip-dip} \right\rangle_{solv,angles} \approx -\frac{\left(l_B \mu_i \mu_j\right)^2}{3r_{ij}^6} \tag{16}$$

Deriving such expressions, assumptions and mathematical approximations are almost always applied which of course has consequences. For example, in the Debye-Hückel theory correlations between ions are partly neglected and unphysical results usually appear for strongly coupled systems. At a later stage I will go into further detail about multipole expansions, but at this point it suffices to say that effective potentials reduce computational costs significantly – usually at the expense of molecular detail.

A somewhat camouflaged feature of effective potentials is that they are temperature dependent. For example, in the expression for the solvent screened potential ϵ_r really should be $\epsilon_r(T)$ but is seldom written so. This has some counter-intuitive consequences and for example, in water the interaction energy between two like charged ions is *attractive* according to the dielectric continuum model[13]. Within the primitive model it is possible to perform a simple partition of the potential of mean force into contributions from energy and entropy (see bottom-left figure on page 46, Paper 1) but to obtain the *true* system energy, including the solvent, a more subtle approach must be taken. This is given in paper 6 where we show that the driving force behind ion-ion correlations in fact stems from entropy and not energy as has been argued for more than twenty years.

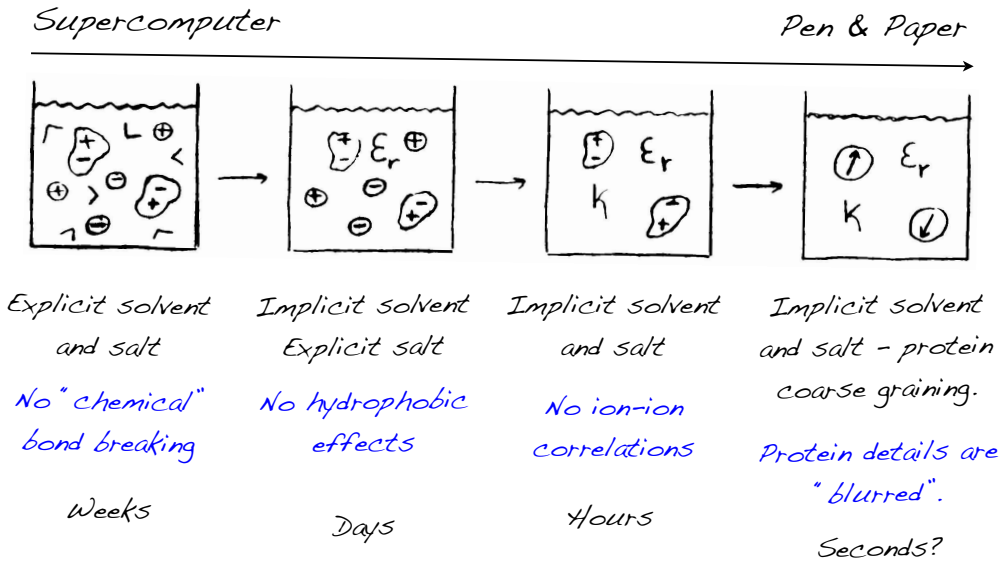


Figure 7: Simplifying an aqueous protein solution using statistical mechanical averaging results in faster and faster calculations.

2.2 The Simple Model of Electrolytes

We have now seen how molecular detail can be implicitly encompassed in effective potentials. Many classical theories for electrolyte solutions are based on the dielectric assumption – *i.e.* that charge-charge interactions can be quantified by a solvent screened

[13] D. F. Evans and H. Wennerström. *The Colloidal Domain - Where Physics, Chemistry, Biology and Technology Meet.* VCH Publishers, New York, 1994.

Coulomb potential (eq 14). In this description solutes such as ions and molecules must be explicitly accounted for, while the solvent is an "invisible" continuum. It is clear that any solute will occupy a volume inaccessible to other solutes and for ions and other small molecules this volume can be fitted well using non-overlapping (charged) spheres of diameter σ. This is achieved by the following conditional pair-potential,

$$\beta w_{ij} = \begin{cases} \frac{l_B z_i z_j}{r_{ij}} & r_{ij} \geq (\sigma_i + \sigma_j)/2 \\ +\infty & r_{ij} < (\sigma_i + \sigma_j)/2 \end{cases} , \tag{17}$$

known as the "simple model of electrolytes". Larger, more complex molecules can be constructed by collections of spheres using various levels of coarse graining (Figure 2, page 13). The predictive success of this simple model has been proven in numerous applications, including the well-known Debye-Hückel theory for determining excess chemical potentials[5]. For concentrated systems or when multivalent ions are present the DH theory becomes less applicable but this is not due to the dielectric assumption but rather connected with the underlying approximations. Using explicit ion Monte Carlo simulations combined with a particle insertion method[14,15], ion activity coefficients can be estimated up to molar concentrations. This is illustrated in Table 2 where we have calculated mean activity coefficients for salts in sea water solutions[16,17,18]. Note that this is a concentrated ($I \approx 0.8$ mol/l), multicomponent system consisting of several ions: Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^- and SO_4^{2-}. Despite this complexity, excellent agreement is found with experimental data.

salt	T (K)	S (‰)	$\gamma^{\text{Exp.}}$	γ^{MC}	γ^{Pitzer}
Na_2SO_4	298	35	0.374±0.016	0.374	0.364
		25	0.405±0.016	0.399	0.394
		15	0.435±0.016	0.445	0.444
		5	0.620±0.016	0.560	0.564
	288	35	0.385±0.016	0.381	0.369
K_2SO_4	298	35	0.352±0.018	0.360	0.344
NaCl	298	35	0.672±0.007	0.664	0.667
		15	0.730±0.001	0.699	0.709
KCl	298	35	0.645±0.008	0.645	0.639
$CaSO_4$	298	35	0.136	0.152	0.150
s				0.023	0.021

Table 2: Mean activity coefficients in sea water measured and calculated using MC simulations and the semi-empirical Pitzer model[17,18]. S is the salinity.

This success is in part due to the fact that we have fine tuned ionic sizes to experimental data for single electrolyte systems – NaCl, KCl etc. In principle the ion radius

[5] Extended DH: $-\ln\gamma = \frac{l_B \kappa z^2}{2(1+\kappa\sigma)}$

[14] B. Widom. Some topics in the theory of fluids. *J. Chem. Phys.*, 39:2808–2812, 1963.

[15] B. R. Svensson and C. E. Woodward. *Molec. Phys.*, 64:247, 1988.

[16] M. Lund, B. Jönsson, and T. Pedersen. Activity coefficients in sea water using monte carlo simulations. *Marine Chemistry*, 80:95–101, 2003.

[17] K.S. Pitzer. Thermodynamics of electrolytes. 1. theoretical basis and general eq uations. *J. Phys. Chem.*, 77:268–277, 1973.

[18] K.S. Pitzer, editor. *Activity coefficients in electrolyte solutions*, chapter 6. CRC-Press, 2 edition, 1991.

should reflect the size of the hydrated ion and this is also the trend found when fitting to experimental data (Table 3) but most likely it will also cover model deficiencies such as a possible breakdown of the continuum assumption at short separations. In addition, ions are not *hard* and the Pauli repulsion could be better described using a smooth function. Finally, we have neglected dispersion interactions that – for certain ions – can be significant; the Hofmeister series is a good example[19,20]. Still, Table 2 indeed is encouraging and illustrates well a practical application of the continuum model, providing clear physical insight with a minimal set of operational parameters.

	r_{mc}	r_{stokes}	r_{cryst}
Na^+	1.8	1.8	1.2
K^+	1.5	1.3	1.5
Mg^{2+}	3.0	3.5	0.7
Ca^{2+}	2.8	3.1	1.1
Cl^-	1.7	1.2	1.7
SO_4^{2-}	1.6	2.3	2.4
NH_4^+	1.5		
H^+	2.5		

Table 3: Hard-sphere radii of individual ions fitted to experimental data for a single simple salt.

2.3 Dielectric Boundaries

Averaging out solvent degrees of freedom can drastically reduce the complexity when simulating liquids. However, this continuum approach brings about another concern: The deep interior of proteins consists mainly of non-polar matter, while the surrounding solvent is highly polar. The effect of this can be captured by solving the Poisson(-Boltzmann) equation taking into account the dielectric boundary between the low and high dielectric regions. However, this boundary is not well defined nor sharp; surface groups of water soluble proteins are partially charged, polar and polarizable, stemming from proton fluctuations, structural flexibility of amino acid side chains and backbone dipole moments. Further, since the dielectric constant is a macroscopic property influenced by distant molecules, setting up boundaries within a few nanometers may push the model too far. Note that it is not uncommon (in fact it is normal) to include charges in the low dielectric protein interior even though the Born energy suggests costs of $\sim$10-100 kT per charge relative to positions just a few Ångströms away. This is typically remedied by partly solvating charges using surface accessible areas and by doing so good agreement can be obtained with experimental data for properties such as side-chain pK_a-values and overall titration behavior.

The above complications suggest a different and simpler approach. If we can prove that surface charges are established in regions with a relatively high dielectric response, electrostatic interactions between them can be conveniently estimated using a uniform

[19] F. W. Tavares, D. Bratko, and J. M. Prausnitz. The role of salt-macroion van der waals interactions in the colloid-colloid potential of mean force. *Current Opinion in Colloid & Interface Science*, 9(1-2):81–86, 2004.

[20] M. Bostrom, F.W. Tavares, D. Bratko, and B.W. Ninham. Specific ion effects in solutions of globular proteins: Comparison between analytical models and simulation. *Journal of Physical Chemistry B*, 109(51):24489–24494, 2005.

dielectric constant equal to that of water. Mathematically, this is trivial as the pair interaction between any two charges will be of the type shown in eq 14 and hence, many-particle systems can be studied relatively easily. Using this approach a number of researchers[21,22,23,24,25,9] have obtained good agreement with experimental data – even though effects from the low dielectric interior have been ignored. This indicates that the uniform dielectric approximation is suitable for the overall electrostatic environment but may become less applicable for deeply buried charges. In Paper 5 we embedded a low dielectric sphere in a protein (illustrated on page 75) and studied how this affected the protonation behavior of several proteins. Electrostatic pair potentials must now include a truncated Legendre polynomial as derived in Appendix A and as such is not as easily handled as in the uniform dielectric model.

2.4 Two-body Interactions

For molecules in solution it is convenient to create an effective pair-potential averaged over all degrees of freedom save the distance between them. For two proteins this is an involved computation as the average is over all salt and solvent positions and orientations, protonation states and inter-protein orientations. This *potential of mean force*, $w(R)$ can be obtained theoretically and experimentally using scattering techniques. The interaction free energy is directly related to the probability of finding the two molecules at a certain separation,

$$\beta w(R) = -\ln g(R) = -\ln \frac{\rho(R)}{\rho_{bulk}} \tag{18}$$

where ρ is the particle density, $g(R)$ is the radial distribution function. Although possible, the latter microscopic property is difficult to obtain experimentally and further averaging is often required for proteins. Since $w(R)$ implicitly incorporates the effects from the liquid surroundings we can – assuming a low protein concentration – express the interaction using the virial expansion for an imperfect gas[26]. The effective second virial coefficient is thus given by,

$$B_2 = -2\pi \int_0^\infty \left(e^{-\beta w(R)} - 1\right) R^2 dR \tag{19}$$

[21] A. Warshel, S. T. Russel, and A. K. Churg. Macroscopic models for studies of electrostatic interactions in proteins: Limitations and applicability. *Proc. Natl. Acad. Sci. USA*, 81:4785–4789, 1984.

[22] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse. Binding of Ca^{2+} to Calbindin D_{9k}: Structural Stability and Function at High Salt Concentration. *Biochemistry*, 33:14170–14176, 1994.

[23] E. Demchuk and C. Wade. Improving the continuum dielectric approach to calculating pK_a's of ionizable groups in proteins. *J. Phys. Chem.*, 100:17373–17387, 1996.

[24] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse. Measurement and modelling of sequence-specific pK_a values of calbindin D_{9k}. *J. Mol. Biol.*, 259:828, 1996.

[25] R. Penfold, J. Warwicker, and B. Jönsson. Electrostatic models for calcium binding proteins. *J. Phys. Chem. B*, 102:8599–8610, 1998.

[9] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse. Ionization behaviour of acidic residues in calbindin D_{9k}. *Proteins*, 37:106–115, 1999.

[26] T. L. Hill. *An Introduction to Statistical Thermodynamics*. Dover Publications Inc., New York, 1986.

which is a macroscopic entity[6] readily obtainable via various techniques[27,28]. If positive, B_2 indicates that repulsive interactions dominate and if negative the net-effect is an attraction. In fact, it has been observed[29] that if a protein's B_2 is within a certain range – the so-called crystallization window – the protein is likely to form crystals. When even more negative, amorphous aggregates are built.

At this stage it is pertinent to point out that B_2 does *not* depend on the protein concentration as it accounts for the interaction between two molecules only. Many body effects in more dense protein solutions are captured by higher order virials even though these are sometimes incorporated into the second virial[30] in an attempt to better conform with experimental conditions.

2.4.1 Multipole Expansion

A protein can be viewed as a set of spatially distributed charges defined by $[\mathbf{r}_i, z_i]$ where $\mathbf{r}_i$ defines the vector from the center-of-mass to the charge, z_i. The potential of mean force between two such charge distributions (Figure 8) can be evaluated in a multipole expansion where one averages over all possible angular orientations and protonation states. This is shown for the salt-free case on page 49 (Paper 2) and the resulting interaction free energy becomes,

$$\beta w(R) \approx \overbrace{\frac{l_B Z_a Z_b}{R}}^{Ion-ion} - \overbrace{\frac{l_B^2}{2R^2}\left(Z_a^2 C_b + Z_b^2 C_a\right)}^{Ion-induced} - \overbrace{\frac{l_B^2 C_a C_b}{2R^2}}^{Induced} - \underbrace{\frac{l_B^2}{6R^4}\left(Z_a^2\mu_b^2 + Z_b^2\mu_a^2\right)}_{Ion-dipole} - \underbrace{\frac{\left(l_B\mu_a\mu_b\right)^2}{3R^6}}_{Dipole-dipole} + \ldots \tag{20}$$

where we identify the familiar intrinsic protein properties, $Z = \langle \sum z_i \rangle$, $\mu = \langle | \sum \mathbf{r}_i z_i | \rangle$ and $C = \langle Z^2 \rangle - \langle Z \rangle^2$. The induced terms are often left out under the assumption that the molecular charge state is constant – *i.e.* that $\langle Z^2 \rangle = \langle Z \rangle^2$. For proteins this is of course not the case, which was recognized by Kirkwood and Shumaker[31] and also by Phillies[32]. However, they did not interrelate charge fluctuations and the protein titration curve (eq 11) and so had little means of quantifying the fluctuation term.

[6] Note the dimension of *volume* which is commonly converted into ml·mol/g^2 by the factor $N_{Av}M_w^{-2}$ *or* to a unit-less entity by division with the hard-core term, $B_2^{hc} = \frac{2\pi\sigma^3}{3}$ where σ is the distance of closest approach.

[27] O. D. Velev, E. W. Kaler, and A. M. Lenhoff. Protein interactions in solution characterized by light and neutron scattering: Comparison of lysozyme and chymotrypsinogen. *Biophys. J.*, 75:2682–2697, 1998.

[28] J. Bloustine, V. Berejnov, and S. Fraden. Measurements of Protein-Protein Interactions by Size Exclusion Chromatography. *Biophys. J.*, 85(4):2619–2623, 2003.

[29] A. George and W. W. Wilson. Predicting protein crystallization from a dilute solution theory. *Acta Cryst.*, D50:361–365, 1994.

[30] S.R. McGuffee and A.H. Elcock. Atomically Detailed Simulations of Concentrated Protein Solutions: The Effects of Salt, pH, Point Mutations, and Protein Concentration in Simulations of 1000-Molecule Systems. *Journal of the American Chemical Society*, 128(37):12098–12110, 2006.

[31] J. G. Kirkwood and J. B. Shumaker. Forces between protein molecules in solution arising from fluctuations in proton charge and configuration. *Proceedings of the National Academy of Sciences*, 38:863–871, 1952.

[32] G. D. Phillies. Excess chemical potential of dilute solutions of spherical polyelectrolytes. *J. Chem. Phys.*, 60:2721–2731, 1974.

Ståhlberg and Jönsson[33,34], however, showed that the charge regulation interaction between a protein and a charged surface is controlled by the slope of the protein titration curve, *i.e* dZ/dpH. The range of interaction for the multipole terms in eq 20 becomes increasingly short ranged for higher order terms and we note that the ion-induced and induced-induced interactions are *more* long ranged than both ion-dipole and dipole-dipole interactions. This is important for the solution stability since the contribution to virial coefficients and binding constants can be significant as long ranged, attractive interactions are favored c.f. eq 19.

The multipole expansion shown is valid in the limit of no or very little salt. Rather often we need to go beyond this limitation and can proceed in at least two ways: the correct way or the not-completely-wrong way! Phillies[32] and Bratko *et al.*[35] made a rigorous analysis of the interaction between two charged, low dielectric charge distributions in a high dielectric salt solution at the Debye-Hückel level. This takes into account the salt-excluded volume taken up by the macromolecules and is similar to the Tanford-Kirkwood model[36], but for two spheres instead of one. The superficial alternative is to argue in terms of salt screened potentials which will immediately yield an approximate solution. For example, regarding protein a as a fluctuating monopole interacting with some potential, ϕ_b generated by another protein, b, the energy is $eZ_a\phi_b$ and we get that,

$$\begin{aligned}\beta w &= -\ln\left\langle e^{-\beta U}\right\rangle \\ &\approx \beta\left\langle U\right\rangle - \frac{\beta^2}{2}\left[\left\langle U^2\right\rangle - \left\langle U\right\rangle^2\right] \\ &\approx \beta e\phi_b\left\langle Z_a\right\rangle - \frac{\left(\beta e\phi_b\right)^2}{2}\left[\left\langle Z_a^2\right\rangle - \left\langle Z_a\right\rangle^2\right]\end{aligned}$$

Thus applying a Debye-Hückel potential, $\beta e\phi_b = \frac{l_B Z_b}{R(1+\kappa a)}e^{-\kappa(R-a)}$ we get the following,

$$\beta w(R) = \overbrace{\frac{l_B Z_a Z_b}{R}\cdot\frac{e^{-\kappa(R-a)}}{(1+\kappa a)}}^{Ion-ion} - \overbrace{\frac{l_B^2 Z_b^2 C_a}{2R^2}\cdot\frac{e^{-2\kappa(R-a)}}{(1+\kappa a)^2}}^{Ion-induced}. \tag{21}$$

In this context the size parameter, a defines the closest distance between the "central-ion" (protein b) and surrounding ions which is approximately the protein radius. However, salt-exclusion is not considered for the fluctuating monopole and as such a is arguably best treated as an adjustable parameter. This approximate procedure is equivalent to the analysis made in Paper 3 (see page 57) for the interaction between a protein and a charged surface. Qualitatively, no difference was seen between these estimates and explicit MC simulations. Equation 21 can be more thoroughly tested by comparison with a potential of mean force as simulated using explicit mobile ions. This is shown for the ion-ion term in Figure 12 on page 31 where it is found that the best agreement between theory and exact numerical simulations is obtained when

[33] J. Ståhlberg, U. Appelgren, and B. Jönsson. *J. Coll. Inteface Sci.*, 176:397–407, 1995.

[34] J. Ståhlberg and B. Jönsson. *Anal. Chem.*, 68:1536–1544, 1996.

[32] G. D. Phillies. Excess chemical potential of dilute solutions of spherical polyelectrolytes. *J. Chem. Phys.*, 60:2721–2731, 1974.

[35] D. Bratko, A. Striolo, J. Z. Wu, H. W. Blanch, and J. M. Prausnitz. Orientation-averaged pair potentials between dipolar proteins or colloids. *J. Phys. Chem.*, 106:2714–2720, 2002.

[36] C. Tanford and J. G. Kirkwood. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.*, 79:5333, 1957.

$a \simeq \sigma_{macroion}/2$. Similarly in Paper 2 (top right figure on page 53) we compared the ion-induced term with results obtained from detailed MC simulations of the lysozyme-calbindin interaction. Again, the multipole expansion seems to capture the electrostatic interaction remarkably well.

This success is in a way surprising, since that in the multipole expansion we assume that $-1 < \beta U \leqq 1$ and thus, for highly coupled systems this approach is expected to become less applicable. To get around this problem one needs to perform the average without the above mentioned assumption[7] so as to account for specific coupling between the multipole terms. This has been examined by Bratko *et al.*[35] who showed that this coupling reduces the dipolar terms, compared with independent averaging as outlined here. However, in this work they exemplified the theory using a dipole moment of 400 Debye and a radius of 10 or 20 Å which for real proteins may be a bit drastic; for more moderate conditions the coupling effect is less dramatic.

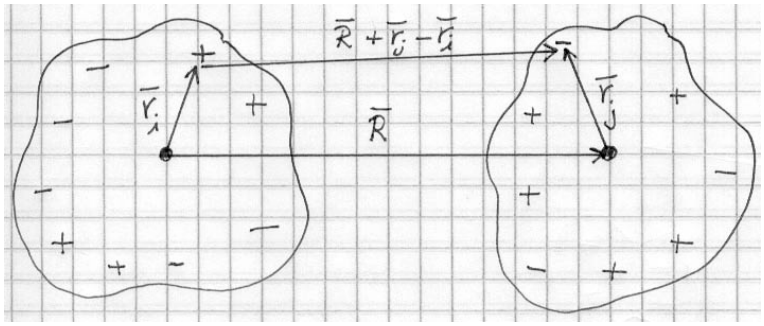


Figure 8: Two interacting, non-overlapping charge distributions.

2.4.2 Protein-Protein Interactions

We have now seen how electrostatic two-body interactions can be approximated by a multipole expansion. However, proteins are not idealized spheres and a number of other mechanisms may come into play,

- Hydrophobic effects.
- Molecular crowding.
- Non-spherical surface topology.
- van der Waals interactions.

Hydrophobicity. The concept "hydrophobic effect" is loosely formulated but shall in this context be used to cover attractive interactions related to correlations between non-hydrophilic surface groups on two interacting proteins. The origin is intimately connected with solvent-solvent and solvent-solute interactions[13,37] and as such difficult to capture within a continuum model. One possible way to incorporate hydrophobic

[7] That is eq 1 → 2 in Paper 2.

[13] D. F. Evans and H. Wennerström. *The Colloidal Domain - Where Physics, Chemistry, Biology and Technology Meet.* VCH Publishers, New York, 1994.

[37] J. Forsman, B. Jönsson, and C. E. Woodward. Computer simulations of water between hydrophobic surfaces: the hydrophobic force. *J. Phys. Chem.*, 100:15005, 1996.

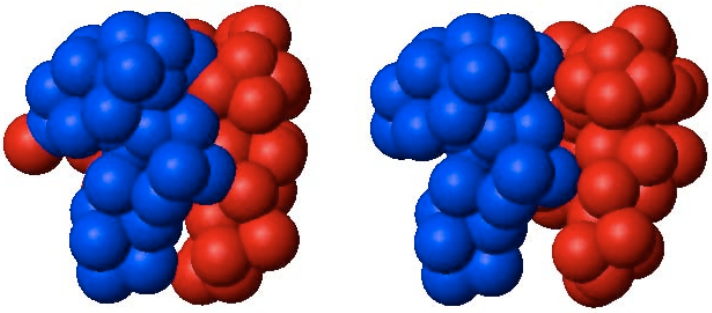


Figure 9: Calbindin represented using the amino acid model with its two strands marked as red (1-45) and blue (46-75), respectively. Left: the X-ray structure of intact calbindin. Right: dominant close-contact configuration obtained from a Monte Carlo simulation of the two fragments (pH 7 and low salt).

interactions will be to specify an effective, short ranged pair-potential between supposedly non-polar residues in the proteins. The choice of such a potential is not obvious and ideally it should capture the effect of salt which is known to enhance hydrophobic interactions[38,39]. Throughout this work we have assumed that hydrophobic interactions are of minor importance for hydrophilic, charged proteins[40]. However, we have made a few preliminary investigations of the association of cleaved calbindin, where hydrophobic residues are exposed to the solvent. To capture the hydrophobic attraction we have invoked a $-1/r^9$ potential between non-polar sites (approx. 1 kT at contact). As shown in Figure 9 the two fragments come together in a configuration very close to the original structure.

Another interesting and possibly related feature is that interactions between certain proteins, often with hydrophobic surface "patches", are known to be influenced by the type of salt, following a Hofmeister series[41].

Crowding. Proteins in the living cell are under the influence of other intra-cellular solutes which may take up as much as 40% of the available volume[42,43] and will inevitably have an impact upon inter-molecular interactions. To a first approximation we will assume that the extra solutes, or crowding agents, are neutral and that their contribution will be a volume excluding effect only. As seen in Figure 10, by increasing the volume fraction of the crowding agent, the attraction between oppositely charged macroions is increased and at dense conditions structural oscillations appear. Since the crowding causes further population of close-contact configurations this may impact dipolar correlations as well as interactions related to surface complimentarity. We could

[38] R. A. Curtis, C. Steinbrecher, M. Heinemann, H. W. Blanch, and J. M. Prausnitz. Hydrophobic forces between protein molecules in aqueous solutions of concentrated electrolyte. *Biophysical Chemistry*, 98(3):249–265, 2002.

[39] M. Jönsson, M. Skepö, and P. Linse. Monte carlo simulations of the hydrophobic effect in aqueous electrolyte solutions. *Journal of Physical Chemistry B*, 110(17):8782–8788, 2006.

[40] J. Dzubiella and J. P. Hansen. Reduction of the hydrophobic attraction between charged solutes in water. *J. Chem. Phys.*, 119:12049, 2003.

[41] M. Boström, D. R. M. Williams, and B. W. Ninham. Specific ion effects: Why the properties of lysozyme in salt solutions follow a hofmeister series. *Biophys. J.*, 85:686–694, 2003.

[42] A P Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*, 55(2):119–40, 1983.

[43] R J Ellis. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci*, 26(10):597–604, 2001.

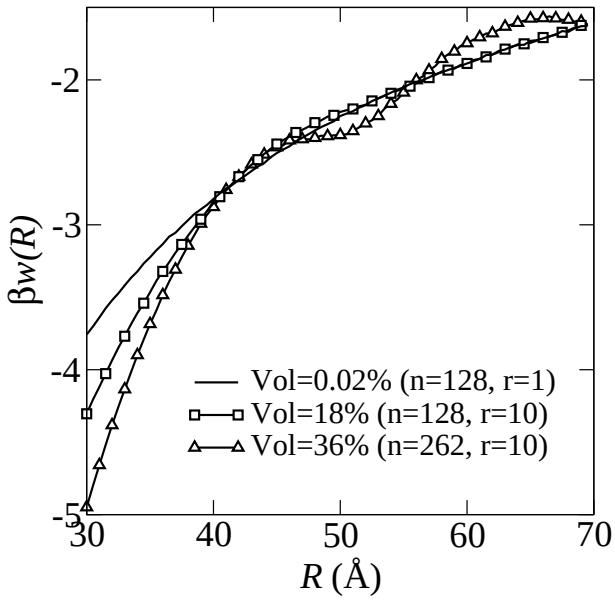


Figure 10: MC simulation of the interaction free energy between two charged spheres, a and b of opposite charge (Z_a=+5, Z_b = −5, σ=30 Å) in the presence of n spherical crowding agents of radius r. No salt added.

also imagine charged or dipolar crowding agents – the list can be made long – which could bring about further mechanisms. But this is for another research project and we will now leave molecular crowding altogether and consider protein-protein interactions under more idealized conditions.

Topology. Proteins come in many shapes and their space filling properties will influence interactions with the solvent as well as other solutes. In particular, the effective interaction between two proteins involves an angular average and the surface topology here contributes with a non-symmetric, repulsive interaction. For small nearly spherical molecules the interaction range is relatively short, whereas for elongated proteins it will decay more slowly. Figure 11 illustrates how overlapping configurations create a repulsion that fades out only after 6 nm. In this particular example we have used the very elongated protein, "fab".

van der Waals. In this context van der Waals interactions is the combined contribution from dispersion-, averaged electronic dipole- and induced dipole interactions. For atoms or small molecules this attractive interaction decays as $1/r^6$ but when added up for large spheres the decay varies approximately as $1/r$[44]. The latter is utilized in the DLVO theory traditionally used for colloids but has also been applied for proteins[45]. This assumes a spherical symmetric molecule and as just seen, this is not always a good approximation. The natural extension is to use the amino acid model (see Figure 2) and apply an attractive potential between the residues. As the individual amino acids are treated as relatively small spheres we assume a potential of the type $\beta u_{vdW} = -C_{vdW}/r^6$. The constant, C_{vdW}, depends on the solvent as well as

[44] J. N. Israelachvili. *Intermolecular and Surface Forces*. Academic Press, London, UK., 2nd edition, 1992.

[45] Martin Muschol and Franz Rosenberger. Interactions in undersaturated and supersaturated lysozyme solutions: Static and dynamic light scattering results. *The Journal of Chemical Physics*, 103(24):10424–10432, 1995.

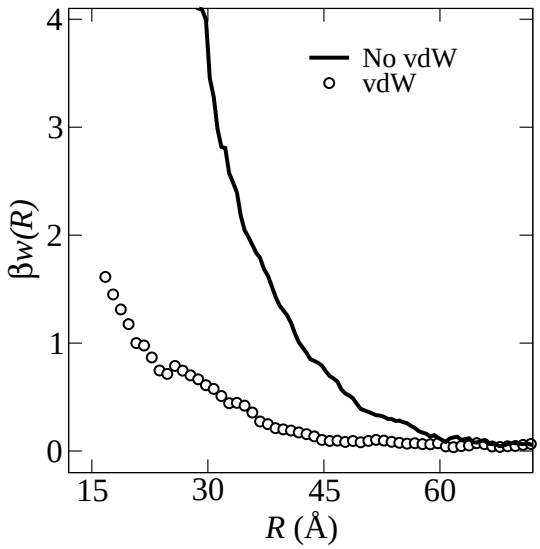


Figure 11: The interaction between two neutral fab molecules (pH=pI) with and without attractive van der Waals interactions.

other solutes[46] and its numeric value is naturally up for discussion. In this work we have generally chosen a value of C_{vdW}=25000 Å^6 which, for two normal sized residues, corresponds to an interaction strength of $\sim 0.2\ kT$ at contact – see Appendix C.

that for most proteins are in the range 3-10 kT. For example, Returning to the "fab" example in Figure 11 it is seen that van der Waals interactions can significantly reduce the overlap repulsion, mainly due to a large contact area as also indicated by the reduction in the closest protein-protein separation.

Dispersion interactions between ions and the macromolecules can also be important; it is well known that protein aggregation is influenced not only by ion valency and concentration, but also by the type of ion. In a crystallographic context this is exploited when producing protein crystals. This ion-specificity is believed to be connected to the polarizability of ions, the solvent and the macromolecules as discussed by a number of workers[19,20,47].

3 Monte Carlo Simulation

Throughout this text a not too small number of multidimensional integrals are listed and their solutions can be found only through numerical integration. This can be performed in various ways – Monte Carlo simulation[8] being one of them.

[8] All simulations in this thesis have been performed using a personal MC simulation suite written in C++. See Appendix B. Interested readers are welcome to contact the author.

[46] R.R. Netz. Static van der waals interactions in electrolytes. *Eur. Phys. J. E*, 5:189—205, 2001.

[19] F. W. Tavares, D. Bratko, and J. M. Prausnitz. The role of salt-macroion van der waals interactions in the colloid-colloid potential of mean force. *Current Opinion in Colloid & Interface Science*, 9(1-2):81–86, 2004.

[20] M. Bostrom, F.W. Tavares, D. Bratko, and B.W. Ninham. Specific ion effects in solutions of globular proteins: Comparison between analytical models and simulation. *Journal of Physical Chemistry B*, 109(51):24489–24494, 2005.

[47] L. Vrbka, P. Jungwirth, P. Bauduin, Touraud D., and Kunz W. Specific ion effects at protein surfaces: A molecular dynamics study of bovine pancreatic trypsin inhibitor and horseradish peroxidase in selected salt solutions. *J. Phys. Chem. B*, 110:7036–7043, 2006.

Taking a thermal average[9] of an observable, x,

$$\langle x \rangle = \frac{\int x e^{-U(\mathcal{R}^N/kT)} d\mathcal{R}^N}{\int e^{-U(\mathcal{R}^N/kT)} d\mathcal{R}^N} \tag{22}$$

we see that the Boltzmann factor favors low energy configurations, whereas high energy states contribute with $\simeq 0$. Therefore, if we were to integrate through all of coordinate space a large part of this effort would be spend by adding zeros. To perform the integration more efficiently Metropolis *et al.*[48] developed the Monte Carlo method where non-zero configurations are sampled more frequently. I will not describe this algorithm in great detail and the reader may instead consult standard text-books on computer simulations[49,50]. For the canonical ensemble the procedure is as follows:

1. Generate a new, random configuration.
2. Calculate the energy difference, ΔU between the new- and old configuration.
3. If the energy is lowered accept the new configuration and start over from (1).
4. If the energy increases accept the configuration with the probability $\exp(-\beta \Delta U)$.
5. If rejected, restore the old configuration.
6. Go back to (1).

In this *importance sampling*[10] scheme configurations of low energy are sought out by staying in the neighborhood of other low-energy states. Averages can now be sampled directly as $\langle x \rangle = \sum_i^n x_i/n$ since the probabilities of the n-generated configurations already follow a Boltzmann distribution according to the acceptance rule.

In a normal MC simulation (run within a reasonable amount of time) coordinate space will be truncated and the statistical basis of high energy configurations will be poorer than those of low energy. If sufficiently large energy differences divide regions in configurational space there is a risk that the simulation gets "stuck" in a local minimum. As a rule, differences of $\sim 10kT$ is manageable but hereafter it can be necessary to divide the simulation into two or more regions and combine the results afterwards. This is known as "umbrella sampling"[49] and one example of its usefulness is when simulating the interaction between two strongly interacting aggregates. The

[9] The average is really over all particle positions *and* momenta and the Hamiltonian is of the form $U = U_{pot}(\mathcal{R}^N) + U_{kin}(\mathbf{p}^N)$. The kinetic part can be solved analytically and appears only as a pre-factor and hence has no influence on static properties.

[10] Here is a rather silly example: Suppose I want to count all pigs in Denmark, I could visit each and every home and ask "How many pigs do you have?". Starting out from my home in Copenhagen, most of the answers would (conceivably) be "none". However, I eventually reach more rural regions and indeed find a pig. Proceeding to a random neighbor to the pig-owner I get "none" and therefore return to the previous location, since I had so good luck there. In this manner, the counting is intensified in regions with many pigs, while pig-less areas are visited more sporadically.

[48] N. A. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1097, 1953.

[49] D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, San Diego, 1996.

[50] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, Oxford, 1989.

[49] D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, San Diego, 1996.

majority of the generated configurations will be in the deep minima at close contact, neglecting the more long ranged interactions. We can circumvent this by running a number of simulations, each restricted to ranges of the inter-molecular separation. An advantageous feature of this approach is that the division into several simulations allows for efficient computational parallelization.

3.1 Boundaries

Due to computational limitations a simulation can include only a limited volume of a macroscopic system – typically a small box or spherical cell, hopefully representative for the system of interest. The boundary of this container is artificial in that the particle density here goes from a finite number to zero. In box simulations, this can be partly remedied by implementing periodic boundaries and the minimum image convention[50]. This, however, is not so easily done in a cell model[51] but staying at a "safe" distance from the boundary this rarely becomes a problem. Figure 12 shows a calculation of the potential of mean force between two charged aggregates in the presence of salt. For inter-aggregate separations, R, below the cell radius the agreement with the Debye-Hückel result (eq 21) is excellent but beyond, the results are doubtful. When the charged spheres approach the cell boundary they loose counter ion stabilization as no ions are allowed outside the hard, impenetrable cell wall. The shown example contains relatively little salt (10 mM) but further addition will screen electrostatic interactions and hence reduce effects from the boundary. In turn this means that the cell size can be reduced so as to decrease the computational load.

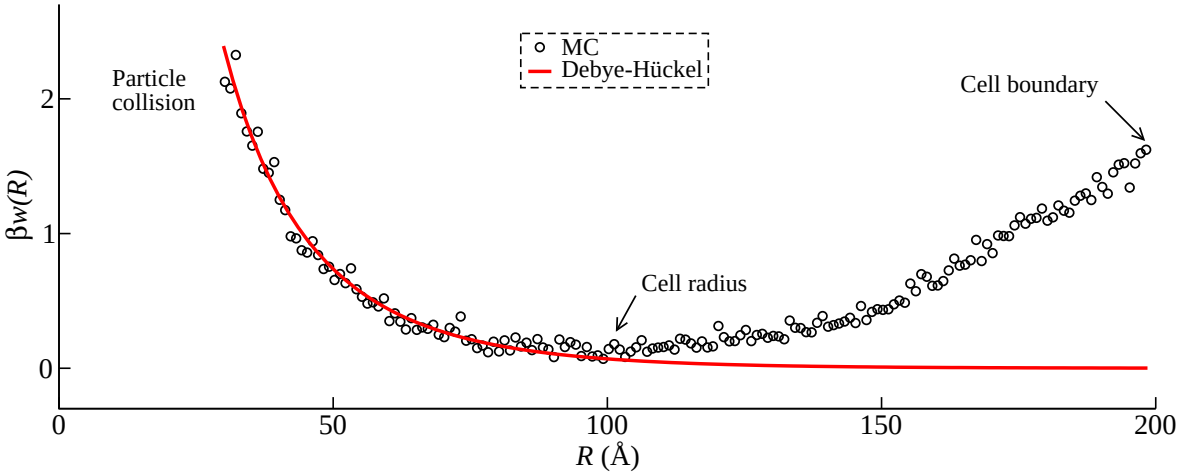


Figure 12: The potential of mean force between two like charged spheres (Z=5, σ=30 Å) in an aqueous solution with 10 mM 1:1 salt. The Debye-Hückel result is calculated using the ion-ion term from eq 21 with $1/\kappa$=30 Å and a set to 15 Å. The MC simulation is performed in a spherical cell with hard boundaries and a radius of 100 Å together with 26 explicit 1:1 ion pairs as well as counter ions.

[50] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 1989.

[51] R. A. Marcus. Titration of polyelectrolytes at higher ionic strengths. *J. Phys. Chem.*, 58:621–623, 1954. delta-pK from nearest neighbour interactions, Bragg-Williams and Ising.

3.2 pH Titration

In section 1.4 we used thermodynamic arguments to derive the free energy difference upon deprotonisation of a titratable group. This can be transformed into a MC move[52] where we will randomly pick a titratable group and try to move its proton to a random site in the bulk. According to eq 5 the energy difference for such a move has two contributions,

$$\Delta U = \Delta U^{el}_{HA\rightarrow A^-} - \ln 10(pH - pK_a) \tag{23}$$

where ΔU^{el} is the electrostatic part and $pH - pK_a$ accounts for chemical effects captured implicitly by the experimental dissociation constant, determined under ideal conditions. For the protonation process we then have

$$\Delta U = \Delta U^{el}_{A^-\rightarrow AH} + \ln 10(pH - pK_a). \tag{24}$$

Of special note is that the calculated electrostatic energy will take in the interaction of the proton and hence pH should *not* contain γ_{H^+} and is therefore – formally incorrect – defined as $-\log C_{\mathrm{H}^+}$. Since γ_{H^+} is a function of the distance from the charged macromolecule, this implies that at constant "pH" the total chemical potential for protons *cannot* be constant throughout the simulation cell. While this non-equilibrium situation is problematic, the effect is diminished for 1) low protein concentrations, 2) high salt concentrations, and 3) low protein charge. Recently Labbez and Jönsson[53] proposed a modified scheme that remedies this and showed that this correction can be important for highly charged systems, such as mineral surfaces.

[52] M. Ullner, B. Jönsson, and P.-O. Widmark. Conformational properties and apparent dissociation constants of titrating polyelectrolytes: Monte Carlo simulation and scaling arguments. *J. Chem. Phys.*, 100:3365–3366, 1994.

[53] C. Labbez and B. Jönsson. A New Monte Carlo Method for the Titration of Molecules and Minerals. *Submitted*, 2006.

A Charges Outside a Low Dielectric Sphere

We consider a neutral sphere with the dielectric constant ϵ_i and radius a in a dielectric solvent, ϵ_o and stride to calculate the potential $\phi(\mathbf{r})$ due to a charge, q located at $\mathbf{r_0}$ in the solvent. Origo is chosen to coincide with the center of the sphere.

The potential inside and outside the sphere must satisfy the Laplace and the Poisson equation, respectively.

$$\nabla^2\phi_i = 0 \tag{25}$$
$$\nabla^2\phi_o = \rho\delta\left(\mathbf{r}_0 - \mathbf{r}\right)/\epsilon_o \tag{26}$$

In our case of spherical symmetry, the general solution to the Laplace equation can be written as,

$$\phi_L = \sum_{n=0}^{\infty}\left(A_n r^n + \frac{B_n}{r^{n+1}}\right)P_n\left(\cos\theta\right) \tag{27}$$

where P_n is the n'th order Legendre polynomial and θ is the angle between $\mathbf{r}_0$ and $\mathbf{r}$. At the surface ($r = a$) the potential must be continuous while the field will experience a discontinuity. At large separation the potential must approach zero, which leads to the following boundary conditions:

$$\phi_i(r=0) \neq \infty \tag{28}$$
$$\phi_o(r\to\infty) = 0 \tag{29}$$
$$\phi_i(r=a) = \phi_o(r=a) \tag{30}$$
$$\epsilon_i\frac{\partial\phi_i(r=a)}{\partial r} = \epsilon_o\frac{\partial\phi_o(r=a)}{\partial r} \tag{31}$$

One particular solution that satisfies the Poisson equation - but not the boundary conditions - is the Coulomb potential,

$$\phi_c = \frac{q}{\epsilon_o|\mathbf{r}_0 - \mathbf{r}|} = \sum_{n=0}^{\infty}\left(\frac{r}{r_0}\right)^n\cdot\frac{q}{\epsilon_o r_0}P_n(\cos\theta) \tag{32}$$

here re-written using a binomial expansion[11] that completely converges for $r_0 > r$. To match the boundary conditions for ϕ_o we add to the Coulomb potential a solution to the Laplace equation (which is perfectly OK, since $\nabla^2\phi_L = 0$) and we can now write the potentials as,

$$\phi_i = \sum_{n=0}^{\infty}A_n r^n P_n(\cos\theta) \tag{33}$$
$$\phi_o = \sum_{n=0}^{\infty}\left(\frac{qr^n}{\epsilon_o r_0^{n+1}} + \frac{B_n}{r^{n+1}}\right)P_n(\cos\theta) \tag{34}$$

where we have exploited the first two boundary conditions. Next, the potentials and fields are equated at $r = a$, yielding

$$A_n = \frac{q}{\epsilon_o r_0^{n+1}}\left(1 + \frac{\epsilon_o - \epsilon_i}{\epsilon_i + \epsilon_o(1+1/n)}\right) \tag{35}$$
$$B_n = \frac{qa^{2n+1}}{\epsilon_o r_0^{n+1}}\left(\frac{\epsilon_o - \epsilon_i}{\epsilon_i + \epsilon_o\left(1+1/n\right)}\right) \tag{36}$$

[11] The Legendre form is convenient when matching the boundary conditions. See Böttcher p25.[54]

valid for $r_0 > a$ and $n \neq 0$. The $n = 0$ term has to be dealt with separately and one easily finds that $A_0 = q/(\epsilon_o r_0)$ and $B_0 = 0$.

Remembering that the expanded form of the Coulomb potential is no longer needed and that - until now - all $4\pi\epsilon$ factors have been omitted, we arrive at

$$\begin{aligned}
\phi_i &= \frac{q}{4\pi\epsilon\epsilon_o}\left(\frac{1}{r_0} + \sum_{n=1}^{\infty}\frac{r^n P_n(\cos\theta)}{r_0^{n+1}}\,(1+\zeta)\right) \\
&= \frac{q}{4\pi\epsilon\epsilon_o}\left(\frac{1}{|\mathbf{r}_0 - \mathbf{r}|} + \sum_{n=1}^{\infty}\frac{r^n P_n(\cos\theta)}{r_0^{n+1}}\zeta\right) \\
\phi_o &= \frac{q}{4\pi\epsilon\epsilon_o}\left(\frac{1}{|\mathbf{r}_0 - \mathbf{r}|} + \sum_{n=1}^{\infty}\frac{a^{2n+1} P_n(\cos\theta)}{(r_0 r)^{n+1}}\zeta\right)
\end{aligned}$$

where $\zeta = \frac{\epsilon_o - \epsilon_i}{\epsilon_i + \epsilon_o(1+1/n)}$. The scribbles in Figure 13 outline a similar analysis, but with the charge placed inside the sphere.

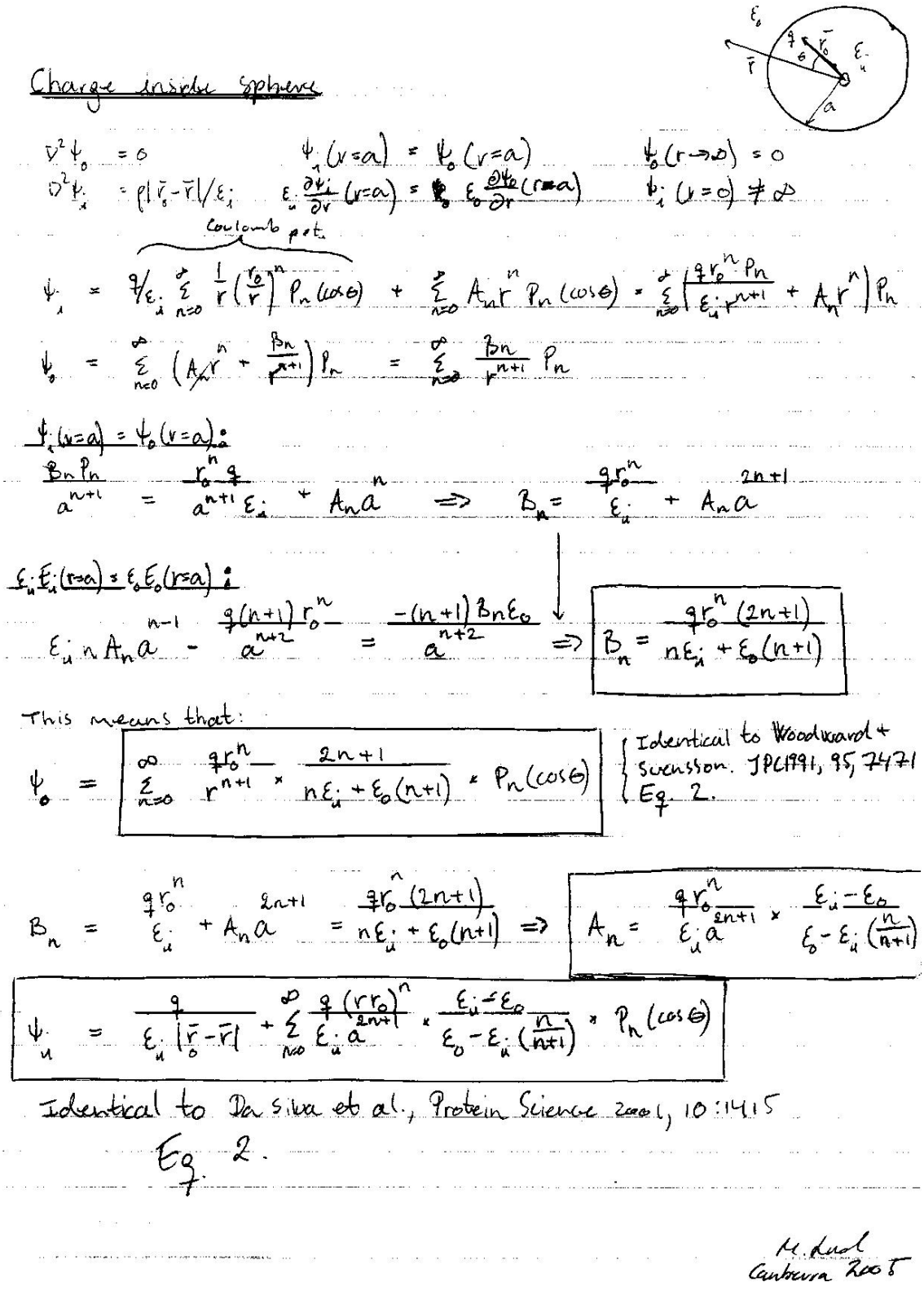


Figure 13: Derivation of the case of a charge located inside the sphere.

B Object Oriented Programming

The purpose of this appendix is to illustrate some of the advantages offered by an object oriented programming language, here exemplified using C++. Writing a new computer simulation program is routinely trivial in the sense that most of the (well tested) code is available in existing programs. Access to this older code base can be more or less convenient, but is often simply copied into the new program, *i.e.* in a "quick-and-dirty" manner. In the following example for handling particles I will demonstrate a different approach where data and functions (methods) are abstracted from the main program and can be directly re-used in other applications.

```
#include <iostream>
#include <cmath>
using namespace std;

class point {
 public:
  double x,y,z;
  double sqdist(point &a) {
    double dx,dy,dz;
    dx=x-a.x;
    dy=y-a.y;
    dz=z-a.z;
    return dx*dx + dy*dy + dz*dz;
  }
  double dist(point &a) {
    return sqrt( sqdist(a) );
  }
  point() { x=y=z=0; };
};

class particle : public point {
 public:
  double radius, charge;
  bool overlap(particle &a) {
    double d=radius+a.radius;
    if ( sqdist(a)<d*d )
      return true;
    return false;
  }
  particle() { radius=charge=0; }
};

// Main program
int main() {
  particle a,b;
  a.x=5.5;
  b.z=-2;
  a.radius=2.5;
  b.radius=3;
  cout << "Distance between a and b = " << a.dist(b) << endl
       << "Closest allowed distance = " << a.radius + b.radius << endl
       << "Particle overlap? ";
  if ( a.overlap(b)==true )
       cout << "yes.\n";
  else
    cout << "no.\n";
}
```

The output will look something like this:

```
Distance between a and b = 5.85235
Closest allowed distance = 5.5
Particle overlap? no.
```

The first class, `point` has three data members `x, y, z` describing a three-dimensional vector. It has methods – or member functions – like `sqdist()` and `dist()` used to calculate the (squared) distance between points. Finally, it has a constructor, `point()` called whenever a new instance of the class is created – in this example it zeros the

coordinates. A "particle" is very similar to a point in that it also has coordinates but also a radius, a charge etc. Object oriented languages allow inheritance from one class to another to let the programmer re-use existing code. Thus, we will run off with everything from `point` and build `particle` on top of that. Besides two new data members, `radius` and `charge` it will also contain a new function, `overlap()` that tests for hard core overlap between any two particles. The two classes need be written only once and will normally be kept in a separate file, abstracting as much as possible from the central program.

Due to the convenient access to data and methods, the main program should be pretty self-explanatory and can be kept short. Most likely a simulation program will involve numerous particles and here the C++ STL `vector` library comes in handy. The `vector` class is approximately as fast as hard coded arrays but has a number of convenient member functions to avoid out-of-bound type errors – `size()` and `push_back()` to mention a few.

```
#include<vector>
...
int main() {
  vector<particle> p(2);
  particle c;

  p[0].x=5.5;
  p[1].z=-2;
  c.y=0.5;

  cout << p.size(); // -> 2
  p.push_back(c);   // Dynamic particle addition
  cout << p.size(); // -> 3
  cout << p[2].y ;  // -> 0.5
  ...
```

Generating a vector of a class does not mean that n instances of the binary member functions are allocated in memory. Generally, member functions exist in only one place but can be *inlined* – manually or by the compiler – so as to increase the performance.

C van der Waals Parameter

The van der Waals interaction between two proteins can be approximated by integrating all atomic C_{vdW}/r^6 terms on the two surfaces. The DLVO theory captures this for two spheres, describing the strength of interaction via the Hamaker constant,

$$A = \pi^2 \rho_1 \rho_2 C_{vdW} \tag{37}$$

where ρ are the atom densities for the two proteins. Since densities of different proteins are subject to only small variations, the Hamaker constant is fairly constant; usually in the range 3-10 kT. In the amino acid model each amino acid is treated as a sphere and hence corresponds to an "atom", albeit larger. To calculate C_{vdW} for two such residues we therefore use the amino acid number density, which can be easily calculated from the number of residues and the estimated protein radius. For calbindin with 75 residues and a radius of $\sim$17 Å a Hamaker constant of 3 kT gives $C_{vdW} \sim 23000$ Å^6.

2940 Biophysical Journal Volume 85 November 2003 2940–2947

A Mesoscopic Model for Protein-Protein Interactions in Solution

Mikael Lund and Bo Jönsson
Department of Theoretical Chemistry, Lund University, Lund, Sweden

ABSTRACT Protein self-association may be detrimental in biological systems, but can be utilized in a controlled fashion for protein crystallization. It is hence of considerable interest to understand how factors like solution conditions prevent or promote aggregation. Here we present a computational model describing interactions between protein molecules in solution. The calculations are based on a molecular description capturing the detailed structure of the protein molecule using x-ray or nuclear magnetic resonance structural data. Both electrostatic and van der Waals interactions are included and the salt particles are explicitly treated allowing investigations of systems containing mono-, di-, and trivalent ions. For three different proteins—lysozyme, α-chymotrypsinogen, and calbindin D_{9k}—we have investigated under which conditions (salt concentration, ion valency, pH, and/or solvent) the proteins are expected to aggregate via evaluation of the second virial coefficient. Good agreement is found with experimental data where available. Calbindin is investigated in more detail, and it is demonstrated how changes in solvent and/or counterion valency lead to attractive ion-ion correlation effects. For high valency counterions we have found abnormal trends in the second virial coefficient. With trivalent counterions, attraction of two negatively charged protein molecules can be favored because the repulsive term is decreased for entropic reasons due to the low number of particles present.

INTRODUCTION

Protein-protein interactions in aqueous solution are of fundamental biological interest and a complete description of the forces acting between proteins and other biomolecules is necessary in an attempt to understand the processes taking place in the living cell. Many of these processes involve weak noncovalent interactions causing the formation of both temporary and more permanent supramolecular structures. The list of examples can be made long: the binding of atomic ions or small molecule cofactors, signal peptides binding to receptor proteins, formation of biological membranes, protein-protein aggregation, etc. In a crystallographic context it is of particular interest to investigate under which conditions the proteins associate to form crystals suitable for diffraction experiments. It is well-known that aggregation can be induced by changes in pH, the salt concentration, valency of ions, or the polarity of the solvent. Presently these matters largely rely on experimental findings but as shall be shown here, valuable information can be derived using computational methods.

The second virial coefficient, B_2, is a useful indicator of the overall interaction between two molecules and its importance in describing protein aggregation has been stressed by several workers (Neal et al., 1999; George and Wilson, 1994). George and Wilson have shown that B_2, to form crystals suitable for diffraction studies, must lie within a narrow interval—the so-called crystallization slot. Second virial coefficients for large molecules can be measured using light- and/or neutron scattering, but is also readily obtained from Monte Carlo or molecular dynamics simulations (Allahyarov et al., 2002). During the years more and more computing power has become available and simulations of a single protein in solution has become a standard approach in theoretical biochemistry. However, to perform a Monte Carlo or molecular dynamics simulation based on an atomistic representation of, say, two proteins in a salt solution, in an attempt to calculate the second virial coefficient, is still beyond reach. The problem comes from the fact that B_2 requires a sampling of all protein separations and orientations, which is a time-consuming process compared to the simulation of one single protein molecule. Thus, to make any progress one has to resort to more coarse-grained models.

A significant simplification is obtained if the water molecules are replaced by a structureless dielectric continuum. This means that the only remaining molecules of the solvent are salt particles. However, systems with high salt concentration still require substantial computation times and to overcome this, screened Coulomb potentials are often utilized (Carlsson et al., 2001). The screened Coulomb approximation usually works well for weakly charged macromolecules in monovalent salt solution, but in solutions with multivalent salt and/or low dielectric permittivity it becomes less applicable. In such cases, salt particles must explicitly be taken into account to correctly reflect the electrostatic interactions in the system.

As for the protein description, it is crucial to capture the discrete charge distribution originating from (de-)protonated amino acids. Specific angular orientations are not without importance for the electrostatic interactions and treating the protein as an object with a central net charge only has been shown to fail (Allahyarov et al., 2002). Several workers have incorporated discreteness by placing point charges within or on the surface of a large sphere (Carlsson et al., 2001; Allahyarov et al., 2002). This approach assumes that the

Submitted April 9, 2003, and accepted for publication July 23, 2003.

Address reprint requests to Mikael Lund, Theoretical Chemistry, Chemical Center, P.O.B. 124, S-221-00 Lund, Sweden. Tel.: 46-46-222-0381; Fax: 46-46-222-4543; E-mail: mikael.lund@teokem.lu.se.

0006-3495/03/11/2940/08 $2.00

excluded volume of the protein possesses spherical symmetry which, even for globular proteins, may be a too-crude approximation considering that, at short protein-protein separations, attractive van der Waals interactions can be highly angular-dependent (Asthagiri et al., 1999). With *van der Waals interactions* we mean the sum of quantum mechanical dispersion forces, thermally averaged dipole-dipole, and dipole-induced dipole terms (Israelachvili, 1991). Between atoms or small molecules, the van der Waals term is relatively short-ranged, with a potential proportional to $1/r^6$. However, when integrated for large (spherical) molecules it becomes appreciably more long-ranged, scaling by $\sim 1/r$ at short and intermediate separations (Israelachvili, 1991). Thus, when incorporating van der Waals interactions it is necessary to either include the integrated term assuming spherical symmetry or explicitly evaluate the $1/r^6$ term for all atomic components in the two proteins.

In this work, we have chosen the latter approach, which captures the detailed structural properties of the protein but at the same time is more computationally demanding. The electrostatic interactions have been evaluated with discrete charges on all titratable amino acids and explicit salt particles and counterions in the solution. Three proteins have been studied: lysozyme, chymotrypsinogen, and calbindin. All three are structurally well-defined, and for lysozyme and chymotrypsinogen, experimental data is readily available in the literature. Calbindin is included as a good representative of a highly charged protein, and in addition, there exists a wealth of experimental data for its calcium binding properties (Linse et al., 1988, 1991; Svensson et al., 1991) and its ionization behavior is also well-documented (Kesvatera et al., 1999).

METHODOLOGY

Interactions

We use a dielectric continuum model for the solution assuming that all charges are uniformly screened by a constant relative permittivity with a value equal to that of pure water. The protein is modeled as a collection of hard spheres representing either single atoms or whole amino acids. A sphere in this model will carry an average charge determined by the pH and pK_a of the particular amino acid. The average charges on the titratable sites have been determined in separate Monte Carlo (MC) simulations of a single protein, in which the amino acids are allowed to titrate. Furthermore, attractive van der Waals interactions between all amino acids are explicitly taken into account.

The choice of a uniform relative permittivity has been and still is very much debated (Antosiewicz et al., 1994, 1996), the main argument being that the charges, at least within the same protein, should be scaled with a much lower value than that of water. This is in principle true but since most charges are located in the outer polar regions of the protein, the effect might be smaller than at first anticipated. Recent experimental and theoretical studies of the ionization behavior of calbindin does not seem to support the idea of a low dielectric permittivity of the protein interior (Spassov and Bashford, 1998; Kesvatera et al., 2001), and the best agreement between experiment and theory is obtained with a high uniform dielectric permittivity.

The objective of the MC simulations is to calculate the free energy change associated with bringing two protein molecules together in an aqueous salt solution. This free energy of interaction or potential of mean force, $w(r)$, must take into account changes in energy and entropy originating from the solvent, the ions, and the proteins. The energy of interaction can be split into contributions from short-range repulsion (*hs*), electrostatics (*el*), and van der Waals (*vdW*) terms, and the interaction between any two sites can be written as

$$u_{ij} = u_{ij}(hs) + u_{ij}(el) + u_{ij}(vdW). \quad (1)$$

The hard-sphere (*hs*) term accounts for the repulsion arising when the electron clouds from two atoms or molecules come into contact. An exact description of this contribution requires a complex quantum mechanical treatment, and hence the simpler hard-sphere term is usually applied as

$$u_{hs}(r) = \infty \quad r_{ij} < \frac{\sigma_i + \sigma_j}{2}, \quad (2)$$

where σ_i is the diameter of site i. The electrostatic term includes Coulombic interactions between charged sites, and in the dielectric continuum approximation it can be written as

$$u_{el}(r) = \frac{z_i z_j e^2}{4\pi\varepsilon_0\varepsilon_r r_{ij}}, \quad (3)$$

where ε_r is the relative dielectric permittivity, z_i the valency of site i, r_{ij} the site-site distance, e the electron charge, and ε_0 the permittivity of vacuum.

To describe the short-range interaction between two protein molecules we invoke a van der Waals-type interaction,

$$u_{vdW}(r) = -\frac{C}{r_{ij}^6}. \quad (4)$$

Here C determines the magnitude of the attraction and is related to the Hamaker constant, A (Israelachvili, 1991),

$$A = \pi^2\rho_1\rho_2 C, \quad (5)$$

where the ρ_i values are particle densities. As a first approximation, we have decided to use the same C for all amino-acid-to-amino-acid interactions. A straightforward improvement would be to let the amino-acid size affect the interaction, hence one would have a different C_{ij} for each pair of amino acids. Calculation of Hamaker constants can be done using the Lifshitz theory (Israelachvili, 1991), but detailed knowledge of the electronic properties is required and as a consequence, A is often treated as an adjustable parameter. Fortunately, Hamaker constants are not subject to large variations and for proteins in water, A is $\sim$3–10 kT (Farnum and Zukoski, 1999) ($1\ kT = 4.11 \times 10^{-27}$ J at 298.15 K).

The effects of the described interactions are shown in Fig. 1 and it is to be noted that the effective hard-sphere contribution, reflecting the nonspherical shape of the protein, is surprisingly long-ranged. Including the electrostatic interaction makes the free energy of interaction even more repulsive, whereas the van der Waals term decreases the repulsion at short separation. Unless stated otherwise, the calculations presented in this work include all of the three terms discussed here.

One effect not taken into account here is the hydrophobic attraction arising at very short separation. As the two protein molecules approach, the water interactions in between become unfavorable and eventually the solvent seeks to the bulk. This results in a free energy gain and effectively creates an attraction. One can view the hydrophobic effect as a correction to the van der Waals term (Forsman et al., 1997) and the effect can be partially incorporated into the van der Waals term by adjusting the C coefficient. A more stringent alternative would be to include another short-range attractive term into Eq. 1. The argument against these refinements is that the proteins

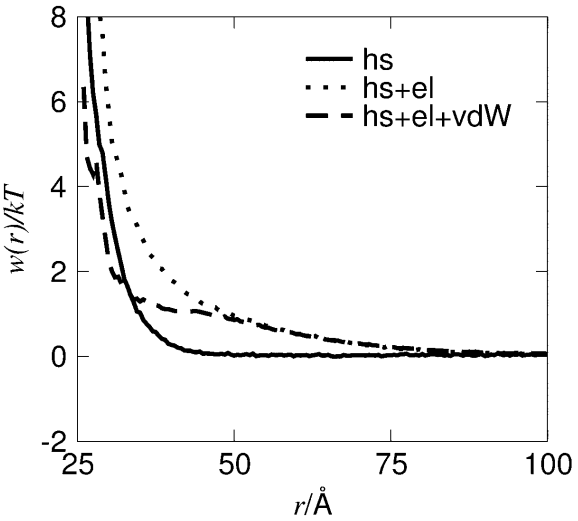


FIGURE 1 Contributions to the interaction free energy, $w(r)$ from hard-sphere (*hs*), electrostatic (*el*), and van der Waals (*vdW*) interactions for lysozyme at pH 9.0.

treated here are hydrophilic, and the hydrophobic effect is thus less important.

Model

The simulated model system consists of two protein molecules built from spheres immersed in a spherical cell (Fig. 2). To maintain electroneutrality and the desired salt concentration, mobile salt particles with hard-sphere diameters, $\sigma = 4$ Å, are added. As for the protein shape, two models have been utilized—both based on structural data obtained from the Brookhaven Protein Databank (PDB). In the first (atomic) model, the protein molecules are mimicked by replacing each nonhydrogen atom in the protein by a hard sphere with diameter $\sigma = 4$ Å, which, for chymotrypsinogen, results in >1800 particles per protein molecule. In addition, one has to include the salt particles and in a dilute protein solution with high salt concentration, the total number of interacting particles can add up to many thousands, which leads to lengthy simulations. To improve the simulation efficiency, a slightly simplified mesoscopic model has been developed. Here the atoms in each amino acid are replaced by a single sphere located at the amino-acid center of mass. The size of these spheres are set equal for all residues and adjusted so that the total excluded volume of the protein is equal to that of the atomic model. This amounts to a diameter for the amino-acid spheres of $\sigma = 6.8$ Å. Average charges are assigned to the center of each sphere according to the actual pH. Despite this seemingly coarse description, the geometry of the surface is found to be remarkably similar to that of the atomic model (Fig. 3). More important, the simulated potentials of mean force for mono-, di-, and trivalent counterions are virtually identical to the more detailed model as shown in Fig. 4. In the mesoscopic model, the van der Waals term, $-C/r_{ij}^6$, is evaluated for amino-acid pairs and the parameter C must reflect this. With a Hamaker constant of 9 kT, C/kT can be estimated to 25,000 Å^{-6} for amino-acid pairs.

The actual charge on an amino-acid residue is pH-dependent, since acidic and basic amino acids can titrate. To specify the average charge on an amino acid at a particular pH, the relevant pK_a values must be known—either from experiment or simulations. The theoretical approach has been shown to be in good agreement with nuclear magnetic resonance studies (Kesvatera et al., 1999, 2001). Hence, we

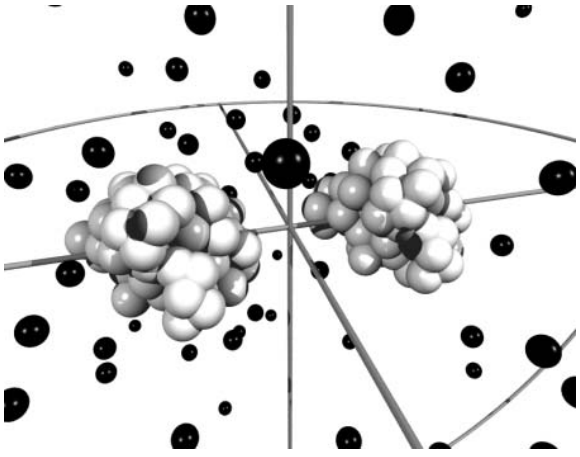


FIGURE 2 Snapshot from a Monte Carlo simulation of calbindin using the amino-acid model. The black spheres illustrate ions whereas amino acids are depicted by white spheres, clustered to form the two proteins. In the simulations, the proteins are displaced along the z-axis and rotated independently. Ions are displaced in all three directions.

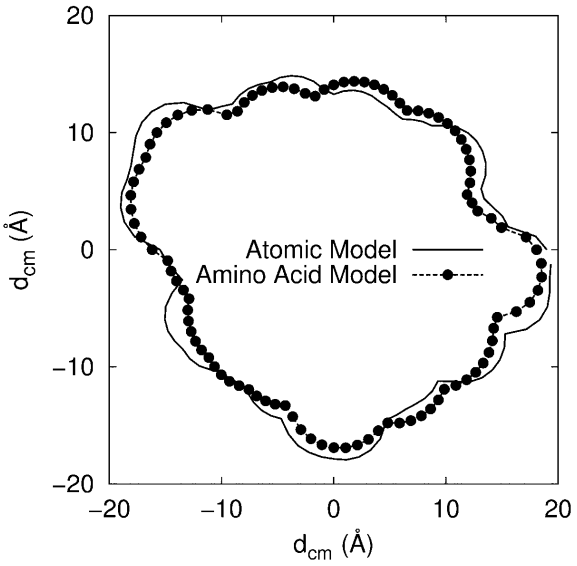


FIGURE 3 Cross-sections of calbindin using the atomic and amino-acid models. A small ion (diameter 4 Å) is rolled on the surface so as to define the minimum distance to the center of mass.

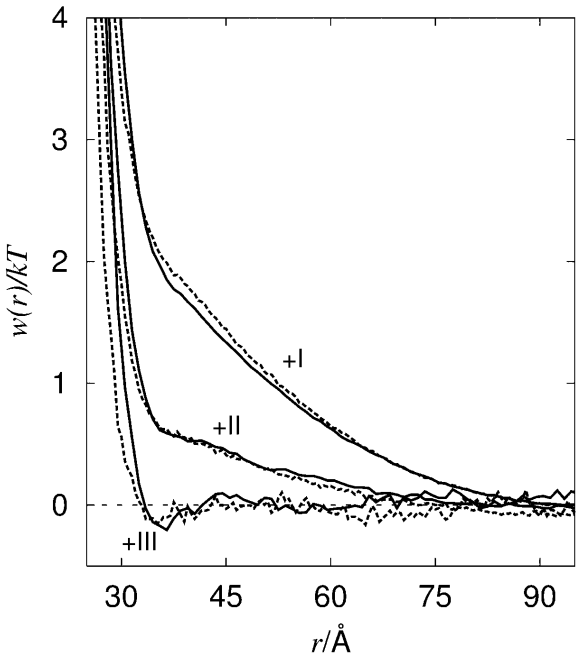


FIGURE 4 Free energy of interaction for calbindin ($c_p = 0.79$ mM) with different counterion valencies (+I, +III, and +III) simulated using the atomic (*solid lines*) and amino-acid (*dashed lines*) models. The van der Waals interactions are not included.

shall use this method to obtain pH-dependent average charges on titratable residues in the proteins. The average net charge of lysozyme, chymotrypsinogen, and calbindin as a function of pH is shown in Table 1.

Monte Carlo simulations

Most simulations were performed in the canonical ensemble using the traditional Metropolis Monte Carlo algorithm (Allen and Tildesley, 1989) supplemented with a few semicanonical simulations of a single protein allowed to titrate (Kesvatera et al., 1999). The energy evaluation for each configuration includes all pair interactions,

TABLE 1 Average net charges, *Z* on calbindin, lysozyme, and α-chymotrypsinogen at different pH values obtained from Monte Carlo simulation

Calbindin		Lysozyme		Chymotrypsinogen	
pH	Z	pH	Z	pH	Z
3.5	3.7	4.5	9.0	3.0	12.9
4.0	1.6	6.0	8.0	4.0	9.2
5.0	−2.2	7.5	6.9	5.3	6.7
6.0	−5.0	9.0	4.9	6.8	5.2
7.0	−6.6	10.0	2.8		
8.0	−7.4	10.5	1.5		
9.0	−8.0				
10.0	−8.8				
11.0	−10.6				

$$U = \sum_{i,j\in p,s}\left(u_{ij}^{el} + u_{ij}^{hs}\right) + \sum_{i,j\in p} u_{ij}^{vdW} \quad i \neq j, \tag{6}$$

where *s* and *p* mean salt and protein particles, respectively. During the MC simulation the proteins are allowed to translate symmetrically along the *z*-axis and individually rotate around vectors going through their center-of-mass. Mobile ions may translate in any direction. By these random displacements and rotations all possible configurations are explored; if a move leads to an energy decrease, the new state is accepted. If the energy increases, the state is accepted with the probability exp $(-\Delta U/kT)$. Proceeding this way, the system eventually reaches equilibrium and its properties can be sampled. The distribution function, $\rho(r)$, is readily obtained by sampling the probability of finding the two proteins at a certain separation. This is directly related to the change in free energy of interaction,

$$w(r)/kT = -\ln\frac{\rho(r)}{\rho(\infty)} + \text{const}, \tag{7}$$

where the $\rho(\infty)$ is determined from the asymptote of $\rho(r)$ and the constant is set to *0*. To represent this in a more convenient manner, $w(r)$ can be integrated to yield the second virial coefficient,

$$B_2 = -2\pi\int_0^\infty (e^{-w(r)/kT} - 1)r^2dr = -2\pi\int_0^\infty \left(\frac{\rho(r)}{\rho(\infty)} - 1\right)r^2dr. \tag{8}$$

The second virial coefficient comprises the predominant effect of the interaction between two proteins molecules; if B_2 is positive, then there is a net repulsion; and if negative, the net interaction is attractive. B_2 has some interesting properties; it is, in general, rather easy to fit experimental data with a variety of $w(r)$ values with adjustable parameters.

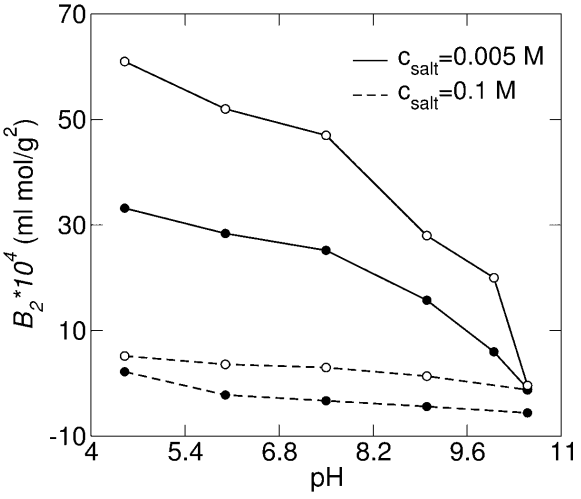


FIGURE 5 Measured (*solid symbols*) and simulated (*open symbols*) second virial coefficients, B_2, for lysozyme as a function of pH at two different NaCl concentrations—monovalent counterions.

At the same time, the virial coefficient will be very sensitive to these parameters. This means that if one is able to produce a potential of mean force without adjusting parameters and if this $w(r)$ reproduces experimental B_2 values, then the underlying physics is probably correct.

RESULTS AND DISCUSSION

Lysozyme

To verify our model we have simulated virial coefficients of lysozyme and compared with experimental data from Velev et al. (1998), who measured B_2 for lysozyme as a function of pH at 5 mM NaCl concentration. We have used an x-ray structure of egg-white lysozyme (Ramanadham et al., 1990) to determine the positions of the amino-acid spheres as explained above. Monovalent counterions are included to maintain electroneutrality and an appropriate amount of 1:1 salt to yield 5 mM is added. The amino-acid charges are assigned for each pH according to simulated pK_a values (Table 1). Each calculation includes 430 million configurations corresponding to a simulation time of ~20 h on a standard PC. As seen in Fig. 5, reasonable agreement is found considering that we have made no efforts to adjust our parameters to fit the experimental data. At low pH the calculated B_2 values are found to be somewhat higher than those obtained from experiments. One possible explanation is that the ionic strength in the experiment is >5 mM due to residual salt from the protein preparation or from pH-adjusting agents. A higher salt content gives rise to a higher screening of the repulsive protein interactions, which in turn diminishes B_2. This agrees with the observed trend that the simulated data fits better at high pH values. At pH 10.5 the electrostatic repulsion is of minor importance and the van der Waals attraction dominates. Since at this point we perfectly match the experimental data, it can be concluded that the chosen C parameter is indeed reasonable. In fact, the virial coefficient is very sensitive to the C parameter at high pH whereas at low pH it is essentially negligible. For example, doubling the C parameter at pH 4.5 changes B_2 from 61 to 60 ml $\times$ mol/g^2, whereas at pH 10.5, B_2 is decreased from -0.4 to -149 ml $\times$ mol/g^2. This behavior is also evident from Eq. 8, where long-range interactions in general have a larger impact on B_2 than those of short-range.

Another possible explanation for the overestimated virial coefficient at low pH is that the charge distribution on the two protein molecules is fixed independently of their separation. It seems reasonable to assume that two positively charged proteins coming in close contact will release protons to reduce their net charge. This is, of course, not possible at all pH values, but if pH is close to the pK_a of some amino acids in the protein, it is certainly a mechanism that will lower the repulsive interaction leading to a reduction of the B_2 (André et al., 2003, unpublished results).

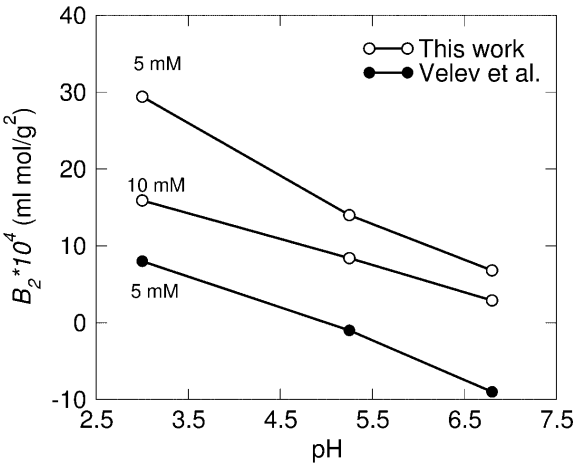


FIGURE 6 Measured (*solid symbols*) and simulated (*open symbols*) second virial coefficients, B_2, for chymotrypsinogen as a function of pH at 5 mM/10 mM NaCl concentration.

Chymotrypsinogen

Also measured by Velev et al. (1998) are second virial coefficients for α-chymotrypsinogen. This protein is—in a computational context—fairly large, containing 245 residues. The three-dimensional structure (PDB Id: 1CHG) used in the calculations is obtained by x-ray diffraction by Freer et al. (1970). Comparing with measured virial coefficients at 5 mM salt concentration (Fig. 6) shows that our calculations follow the experimental trend, but are more repulsive. However, the system is very sensitive to the ionic strength as illustrated by doubling the salt concentration from 5 mM to 10 mM, effectively decreasing the repulsion. This indicates that at low salt concentrations even trace amounts of residual ions may lower the experimental B_2, making direct comparison with theoretical calculations less accurate. Another explanation of the difference between theory and experiment could be that the protein average charge is set too high. When the two macro molecules come into close contact the titratable sites are perturbed, effectively lowering the net charge. This can be remedied by allowing both proteins to titrate during simulation so as to adjust charges as a function of protein-protein separation.

It is to be noted that the chymotrypsinogen interactions are, in general, less repulsive than those between lysozyme, as also observed by Velev and co-workers. One explanation could be the high dipole moment of chymotrypsinogen leading to angular correlations (Velev et al., 1998), but the importance of such dipole-dipole interactions remain left for further study.

Calbindin

Currently there is no experimental data available for calbindin, but we hope to obtain second virial coefficients

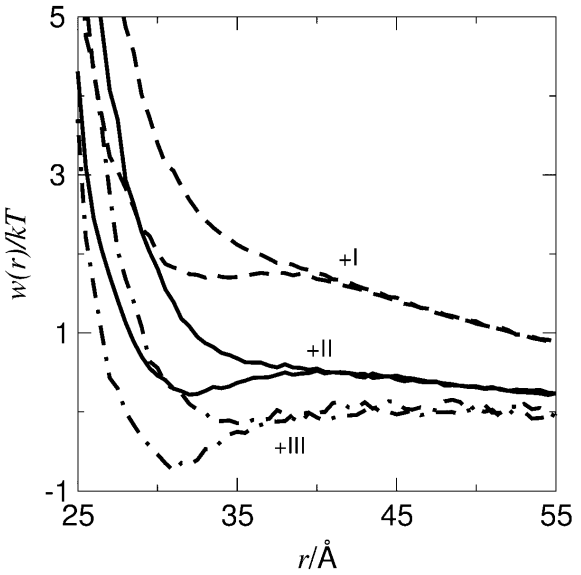


FIGURE 7 Free energy of interaction for two calbindin molecules, c_p = 0.79 mM, with various counterions. Graphs both with and without short-range van der Waals interactions are shown.

for calbindin under a variety of conditions in the near future. Calbindin is considerably smaller than lysozyme and chymotrypsinogen and is thus suitable for more exploratory simulations. The structure of calbindin used in the simulations derives from an x-ray determination of the calcium-free protein (Szebenyi and Moffat, 1986) and the overall charge is −7 at neutral pH; compare to Table 1. In a salt-free environment with only counterions present, we note that B_2 is strongly dependent on the counterion valency. Fig. 7 demonstrates that an increasing valency significantly decreases the repulsion between the proteins. This is qualitatively in accordance with mean field theories (Derjaguin and Landau, 1941; Verwey and Overbeek, 1948), and is a consequence of an increased electrostatic screening from multivalent ions. Note that in the case of trivalent counterions the electrostatics are completely screened and the van der Waals term hence leads to an attractive minimum. This qualitatively explains why proteins are found to precipitate upon addition of even small amounts of multivalent salts.

It is illustrative to split the free energy, $A(r) = U(r) - TS(r)$, into the individual contributions from energy and entropy as shown in Fig. 8. The derivatives of these terms give the force acting between the two proteins, that is

$$-\frac{\partial A}{\partial r} = -\frac{\partial U}{\partial r} + T\frac{\partial S}{\partial r}; \quad \text{or} \quad F_{\text{tot}} = F_{\text{u}} + F_{\text{s}}. \qquad (9)$$

The energetic force component, F_{u}, is attractive between the two negatively charged proteins, which at first sight might seem counterintuitive. However, this is a general result; for an overall neutral system of charges, the energy will always favor a compaction. Hence, the origin of repulsion is the entropy, and not the energy. This means that the reduced repulsion seen with trivalent compared to monovalent counterions (see Fig. 7) is a direct consequence of the fact that the number of particles has been reduced by a factor of three. Thus, one can favor attractive interactions by decreasing the entropic term. Similarly, one can also strengthen the attraction by favoring the energy term. One simple way to do this is by lowering the solvent polarity and hence the dielectric constant, which substantially enhances the energy term. From an experimental point of view,

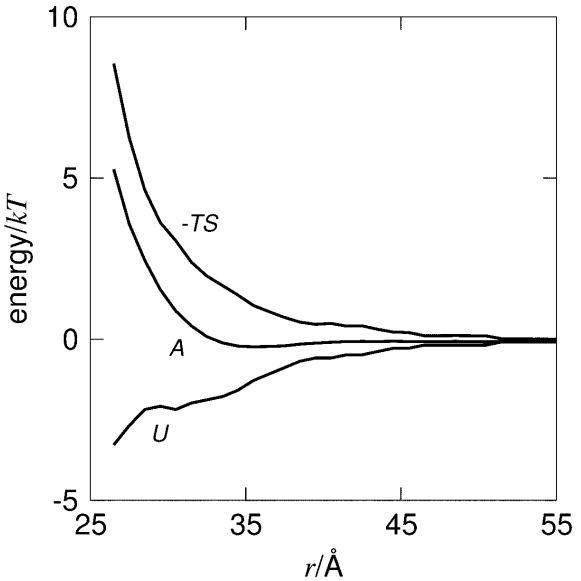


FIGURE 8 The interaction free energy (A), energy (U), and entropy ($-TS$) for a solution of calbindin with trivalent counterions. The van der Waals interactions are not included.

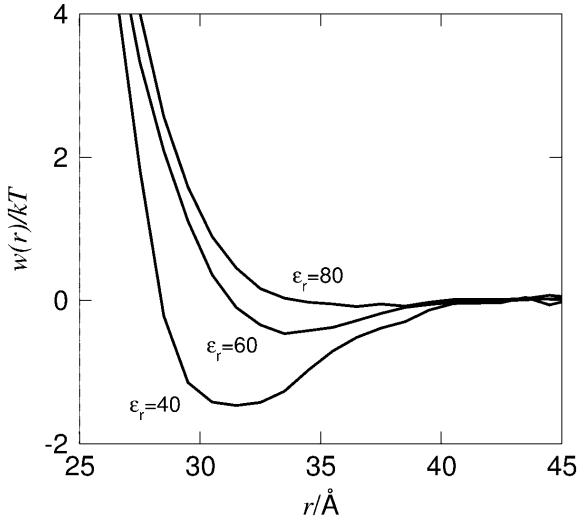


FIGURE 9 Interaction free energy for two calbindin molecules, 0.79 mM, with trivalent counterions at different dielectric constants, ε_r. The van der Waals interactions are not included.

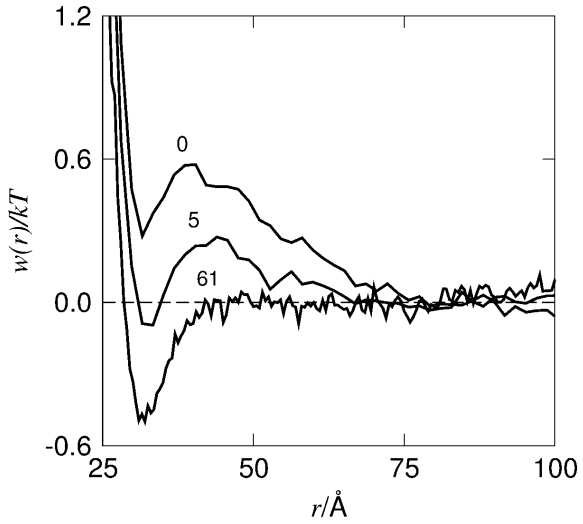


FIGURE 10 Free energy of interaction for two calbindin molecules, 0.79 mM with mono- (*top*) and divalent (*bottom*) counterions at various 1:1 salt concentrations (mM).

a gradual change of the dielectric constant can be achieved by addition of methanol which is miscible with water in all proportions. As a matter of fact, this is a well-known method for precipitating proteins, although the underlying mechanism has not been fully understood. Fig. 9 shows how the potential of mean force for a system containing trivalent counterions can display a net attraction, even though no van der Waals forces are included. Thus, a moderate reduction of the solvent polarity leads to a free energy minimum and eventually to a precipitation of the protein. This is a well-known phenomenon in colloid chemistry and is usually discussed in terms of ion-ion correlations (Guldbrand et al., 1984; Jönsson and Wennerström, 2001). When the electrostatic coupling strength is increased by increasing the valency or lowering the dielectric permittivity, it is accompanied by a strong accumulation of counterions close to the protein. In particular, trivalent counterions will be found very near the protein surface; however, they will not be bound in a chemical sense.

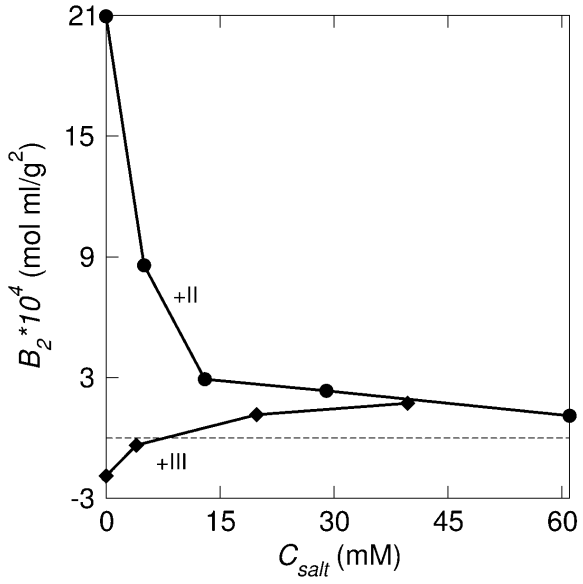


FIGURE 11 Second virial coefficient, B_2 for calbindin, $c_p = 0.79$ mM, as a function of counterion valency (+II/+III) and 1:1 salt concentration.

Increasing the salt concentration is also known to induce protein aggregation and in the case of monovalent ions a relatively high amount is required to effectively reduce the repulsion. As expected, divalent counterions more readily support salt-induced aggregation as is illustrated in Fig. 10. In the monovalent case, a minimum in $w(r)$ occurs at a salt concentration of ~0.1 M, whereas for divalent counterions, this minimum is already found at ~0.005 M salt. From the second virial coefficient, Fig. 11, one can note that even though an attractive minimum occurs in $w(r)$, B_2 still remains positive.

Proteins with trivalent counterions, however, show a different behavior. Here the second virial coefficient is increased when 1:1 salt is added, which is in direct conflict with predictions from the DLVO theory (Derjaguin and Landau, 1941; Verwey and Overbeek, 1948). The origin of this effect stems from a competition of mono- and trivalent counterions. The latter accumulate close to the charged protein and give rise to an efficient screening. However, when more salt is gradually added, the monovalent counterions will replace the trivalent ones with a concomitant reduction of the protein screening. A similar behavior has been seen in DNA solutions (Khan et al., 1999).

It is to be noted that in these calbindin simulations a cell radius of 100 Å is used, which is adequate for systems where the protein-protein interactions are screened. However, in cases with no or very low 1:1 salt concentration, an artificial boundary effect may interfere and change the numerical results. This can be remedied by increasing the cell radius, but unfortunately this drastically increases the computation time. However, the general trends are preserved, and in cases with di- and trivalent ions, this effect is of no importance.

CONCLUSION

The protein model presented in this article includes electrostatic and van der Waals interactions, while at the same time it takes the specific protein structure into account. Considering the experimental uncertainties, calculated second virial coefficients for lysozyme and chymotrypsinogen are in good agreement with measurements. In the case of low protein net charge the agreement is actually very good, indicating that the magnitude of the attractive van der Waals term is reasonably chosen. The agreement with experiment can be improved by, for example, including an additional short-range attraction mimicking the hydrophobic interaction. Also, letting the proteins titrate during simulation

allows the net-charge to vary as a function of protein-protein separation, effectively decreasing the electrostatic repulsion.

Calbindin has been simulated more intensively and the effect of salt concentration and salt valency have been investigated. Addition of multivalent counterions causes a dramatic reduction of the electrostatic repulsion between two proteins. Preliminary scattering experiments indicate that calbindin self-associates upon addition of small amounts of lanthane (III) ions. The present simulations predict this effect to be due to ion-ion correlation and it can be depleted by addition of monovalent salt. Correlation effects can be invoked by lowering the solvent polarity, effectively increasing the electrostatic interactions.

We thank Dr. Sara Linse, Biophysical Chemistry, University of Lund, Sweden for informative discussions and for comments on the manuscript. We also thank Dr. Tonu Kesvatera, National Institute of Chemical Physics and Biophysics, Tallinn, Estonia, for providing preliminary experimental data on calbindin.

REFERENCES

Allahyarov, E., H. Löwen, J. P. Hansen, and A. A. Louis. 2002. Discrete charge patterns, Coulomb correlations and interactions in protein solutions. *Euro. Phys. Lett.* 57:731–737.

Allen, M. P., and D. J. Tildesley. 1989. Computer Simulations of Liquids. Oxford University Press, Oxford, UK.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pK_as in proteins. *Biochemistry.* 35:7819–7833.

Asthagiri, D., B. L. Neal, and A. Lenhoff. 1999. Calculation of short-range interactions between proteins. *Biophys. Chem.* 78:219–231.

Carlsson, F., M. Malmsten, and P. Linse. 2001. Monte Carlo simulations of lysozyme self-association in aqueous solution. *J. Phys. Chem.* 105:12189–12195.

Derjaguin, B. V., and L. Landau. 1941. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Acta Phys. Chim. URSS.* 14:633–662.

Farnum, M., and C. Zukoski. 1999. Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophys. J.* 76: 2716–2726.

Forsman, J., B. Jönsson, C. E. Woodward, and H. Wennerström. 1997. Attractive surface forces due to liquid density depression. *J. Phys. Chem. B.* 101:4253.

Freer, S. T., J. Kraut, J. D. Robertus, H. T. Wright, and N. H. Xuong. 1970. Chymotrypsinogen: 2.5-Å crystal structure, comparison with α-chymotrypsin, and implications for zymogen activation. *Biochemistry.* 9:1997–2009.

George, A., and W. W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr.* D50:361–365.

Guldbrand, L., B. Jönsson, H. Wennerström, and P. Linse. 1984. Electrical double layer forces. A Monte Carlo study. *J. Chem. Phys.* 80:2221.

Israelachvili, J. 1991. Intermolecular and Surface Forces, 2nd Ed. Academic Press, London, UK.

Jönsson, B., and H. Wennerström. 2001. When ion-ion correlations are important in charged colloidal systems. *In* Electrostatic Effects in Soft Matter and Biophysics. C. Holm, P. Kekicheff, and R. Podgornik, editors. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Kesvatera, T., B. Jönsson, E. Thulin, and S. Linse. 1999. Ionization behavior of acidic residues in calbindin d_{9k}. *Proteins.* 37:106–115.

Kesvatera, T., B. Jönsson, E. Thulin, and S. Linse. 2001. Focusing of the electrostatic potential at EF-hands of calbindin D_{9k} titration of acidic residues. *Proteins.* 45:129–135.

Khan, M. O., S. M. Mel'nikov, and B. Jonsson. 1999. Anomalous salt effects on DNA conformation: experiment and theory. *Macromolecules.* 32:8836–8840.

Linse, S., P. Brodin, C. Johansson, E. Thulin, T. Grundström, and S. Forsen. 1988. The role of protein surface charges in ion binding. *Nature.* 335:651–652.

Linse, S., C. Johansson, P. Brodin, T. Grundström, T. Drakenberg, and S. Forsen. 1991. Electrostatic contributions to the binding of Ca^{2+} in calbindin D_{9k}. *Biochemistry.* 30:154–162.

Neal, B. L., D. Asthagiri, O. D. Velev, A. M. Lenhoff, and E. W. Kaler. 1999. Why is the osmotic second virial coefficient related to protein crystallization? *J. Cryst. Growth.* 196:377–387.

Ramanadham, M., L. C. Sieker, and L. H. Jensen. 1990. Refinement of triclinic lysozyme. II. The method of stereochemically restrained least squares. *Acta Crystallogr.* B46:63–69.

Spassov, V., and D. Bashford. 1998. Electrostatic coupling to pH-titrating sites as a source of cooperativity in protein-ligand binding. *Protein Sci.* 7:2012–2025.

Svensson, B., B. Jönsson, C. E. Woodward, and S. Linse. 1991. Ion binding properties of calbindin D_{9k}—a Monte Carlo simulation study. *Biochemistry.* 30:5209–5217.

Szebenyi, D. M. E., and K. Moffat. 1986. The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. molecular details, ion binding, and implications for the structure of other calcium-binding proteins. *J. Biol. Chem.* 261:8761–8777.

Velev, O. D., E. W. Kaler, and A. M. Lenhoff. 1998. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophys. J.* 75:2682–2697.

Verwey, E. J. W., and J. T. G. Overbeek. 1948. Theory of Stability of Lyophobic Colloids. Elsevier Publishing Company, Amsterdam, The Netherlands.

5722 *Biochemistry* **2005,** *44,* 5722−5727

On the Charge Regulation of Proteins

Mikael Lund* and Bo Jönsson

Theoretical Chemistry, Chemical Center, Post Office Box 124, S-221 00 Lund, Sweden

Received November 8, 2004; Revised Manuscript Received January 5, 2005

ABSTRACT: It is known that the overall charge of a protein can change as the molecule approaches a charged object like another protein or a cell membrane. We have formalized this mechanism using a statistical mechanical framework and show how this rather overlooked interaction increases the attraction between protein molecules. From the theory, we can identify a unique property, the *protein charge capacitance*, that contains all information needed to describe the charge regulation mechanism. The capacitance can be obtained from experiment or theory and is a function of pH, salt concentration, and the number of titrating residues. For a range of different protein molecules, we calculate the capacitance and demonstrate how it can be used to quantify the charge regulation interaction. With minimal effort, the derived formulas can be used to improve existing models by including a charge regulation term. Good agreement is found between theory, simulations, and experimental data.

Knowledge of the basic intermolecular interactions between biomolecules is of vital importance for our understanding of processes in the living cell. This includes, for example, the interaction of a protein with small ligands, with DNA, or with a membrane surface, as well as the interaction between two or more protein molecules. The interaction of two proteins, each with a nonzero net charge, is at long distances dominated by a direct Coulomb interaction. The protein charges come from ionized amino acid residues and will vary with pH and other solution conditions. The majority of models presented in the literature describing protein−protein interactions, however, implement the protein charge distribution as a *fixed* set of charges (*1*−*4*) or even by a single-point charge. This is probably still a valid approach for proteins carrying a significant net charge, but when an approximately neutral protein, pH ≈ pI, is approaching a charged surface or another highly charged protein, its charge distribution will change. That is, the protonation state of the protein not only depends on pH but also on nearby molecules; the electrostatic potential from a neighboring molecule will perturb the titrating groups in the protein. This is of course the same effect as is seen internally in a protein, where any charged amino acid may affect the apparent pK_a values of all other groups. Our aim is to formally describe this interaction in mathematical terms and then to numerically calculate the appropriate response function. For the latter, we will use Monte Carlo (MC) simulations describing the protein in atomic detail including all ion−ion interactions within the protein and with the surrounding salt solution.

Charge regulation for colloidal particles has been discussed by Kirkwood and Shumaker (*5*) and later by Carnie et al. (*6*, *7*) within the Debye−Hückel approximation. Zydney et al. have used a similar approach and found the regulation mechanism important for proteins in porous media (*8*), in membranes (*9*), and in capillary electrophoresis experiments (*10*). Bowen and Williams (*11*) solved the Poisson−Boltzmann equation for a sphere in a Wigner−Seitz cell so as to mimic bovine serum albumin. They found that inclusion of a charge regulation term significantly improved the agreement with measured osmotic coefficients. Ståhlberg et al. (*12*−*14*) have studied the net charge of lysozyme in the context of ion-exchange chromatography, where they show how the charge regulation, i.e., the capacitance of a protein, can be derived from the experimental titration curve.

* To whom correspondence should be addressed. Telephone: +46-46-222 0381. Fax: +46-46-222 4543. E-mail: mikael.lund@teokem.lu.se.

THEORY AND METHODS

Protein Charge Capacitance: Statistical Mechanical Derivation. In this section, we will derive a formal expression for the capacitance in terms of charge fluctations. The essence is captured in eqs 5 and 6, and the reader may want to proceed directly to this part. Let us start by considering two proteins in a salt solution, each described by the charge distributions [$\mathbf{r}_i$, q_i] and [$\mathbf{r}_j$, q_j], respectively. The mass centra of the distributions are separated by $\mathbf{R}$, which means that the distance between two charges i and j is given by $r_{ij} = |\mathbf{R} + \mathbf{r}_j - \mathbf{r}_i|$. The average net charge number of the distributions does not need to be zero, that is, $\langle Q_A \rangle \neq 0$, where $\langle Q_A \rangle = \langle \sum q_i \rangle$. The free energy of interaction can be written as,

$$\beta A(R) = -\ln\langle e^{-\beta U(R)} \rangle \approx -\ln\left[1 - \langle \beta U(R) \rangle + \frac{1}{2}\langle (\beta U(R))^2 \rangle\right] \quad (1)$$

where $U(R)$ is the interaction between the two charge distributions, $\beta = 1/kT$, with k being the Boltzmann constant and T being the temperature, and $\langle ... \rangle$ denotes an average over the unperturbed system, which in the present case is the single isolated protein in salt solution. The average runs over all orientations and ionization states of the protein as well as over the positions of all salt particles. This means that the calculated averages will depend on both the salt and protein concentrations and pH. Equation 1 can be expanded to a second order as $\ln(1 - x) \approx -x - x^2/2$ for small x,

10.1021/bi047630o CCC: $30.25
Published on Web 03/19/2005

$$\beta A(R) \approx \langle \beta U(R) \rangle - \frac{1}{2}[\langle (\beta U(R))^2 \rangle + \langle \beta U(R) \rangle^2] \quad (2)$$

The interaction energy is simply the Coulomb interaction between the two charge distributions,

$$\beta U(R) = \sum_i \sum_j \frac{l_B q_i q_j}{r_{ij}} \quad (3)$$

where the Bjerrum length, $l_B = e^2/4\pi\epsilon_0\epsilon_r kT$ has been introduced. We can make a multipole expansion of the energy, assuming that $R \gg r_i, r_j$. This expansion will include an ion−ion interaction, an ion−dipole interaction, a dipole−dipole interaction, etc. It will also include charge-induced charge and induced charge-induced charge interactions. Thus, we can write an approximation to the free energy including all terms of order up to $1/R^2$. Note that the ion−dipole interaction disappears in the first order and that the first nonvanishing dipole term will be of the order $1/R^4$.

$$\beta A(R) \approx \frac{l_B \langle Q_A \rangle \langle Q_B \rangle}{R} - \frac{l_B^2}{2R^2}[\langle Q_A^2 \rangle \langle Q_B^2 \rangle - \langle Q_A \rangle^2 \langle Q_B \rangle^2] \quad (4)$$

Note also that $\langle Q^2 \rangle \neq \langle Q \rangle^2$. We now define a "charge polarizability" or capacitance, C, that quantifies the charge fluctuations of the protein,

$$C \equiv \langle Q^2 \rangle - \langle Q \rangle^2 = -\frac{\partial Q}{\partial \beta e \Phi} \quad (5)$$

where linear response theory gives the relation to the electrical potential, Φ. With this definition of the capacitance, eq 4 can be rewritten in a more useful form,

$$\beta A(R) \approx \frac{l_B \langle Q_A \rangle \langle Q_B \rangle}{R} - \frac{l_B^2}{2R^2}(C_A C_B + C_A \langle Q_B \rangle^2 + C_B \langle Q_A \rangle^2) \quad (6)$$

The first term is the direct Coulomb term, and the following terms are the *induced charge-induced charge* and *charge-induced charge* interactions. If the protein molecules are identical, that is, $\langle Q_A \rangle = \langle Q_B \rangle = \langle Q \rangle$, the expression then simplifies to,

$$\beta A(R) \approx \frac{l_B \langle Q \rangle^2}{R} - \frac{l_B^2}{2R^2}(C^2 + 2C\langle Q \rangle^2) \quad (7)$$

and if the proteins happen to be at pH = pI, then $\langle Q \rangle = 0$ and the leading term is the induced charge-induced charge interaction,

$$\beta A(R) \approx -\frac{l_B^2 C^2}{2R^2} \quad (8)$$

which is equivalent to the findings of Bratko et al. (*15*) for the intermolecular interaction between micelles subject to charge fluctuations of the number of counterions.

The above equations show that the fluctuating charge of a protein may under certain circumstances contribute significantly to the net interaction between two proteins. From eq 5, we can also write the induced charge as $Q_{ind} = -C\beta e\Delta\Phi$, valid for small potentials. Note that in the derived equations we have used an unscreened Coulomb potential (eq 3) valid for no or very low salt concentrations, only. To include the effect of salt, it is necessary to use a screened potential; an example of this is shown in eq 14.

Macroscopic Picture. The capacitance, C, can be derived from the protein titration curve. For a single titrating acid, the degree of ionization, α, can be found in any elementary physical chemistry textbook,

$$\mathrm{pK} = \mathrm{pH} - \log\frac{\alpha}{1-\alpha} \quad (9)$$

Taking the derivative of α with respect to pH gives

$$\frac{\partial \alpha}{\partial \mathrm{pH}} = \alpha(1-\alpha)\ln 10 = C \ln 10 \quad (10)$$

where in the second step we have identified the capacitance defined in eq 5. We can obtain an approximate value for the capacitance in a protein assuming that there is no interaction between the titrating sites: A protein contains several titrating groups such as aspartic and glutamic acid, histidine, etc., each with an ideal pK_0 value. When different titrating groups are denoted with γ and their number with n_γ, the total capacitance can then be approximated with,

$$C^{\mathrm{ideal}} = \sum_\gamma n_\gamma \frac{10^{\mathrm{pH}-\mathrm{pK}_{0,\gamma}}}{(1+10^{\mathrm{pH}-\mathrm{pK}_{0,\gamma}})^2} \quad (11)$$

More realistic capacitances, where intramolecular interactions are included, can be obtained from experimental protein titration curves, readily available in the literature. As evident from eq 10, the capacitance can be extracted from the slope,

$$C \ln 10 = -\frac{\partial Q}{\partial \mathrm{pH}} \quad (12)$$

Thus, we can estimate the magnitude of the regulation interaction using the $1/R^2$ terms derived in the previous section. Where experimental titration data are not available, one may resort to theoretical models; the Tanford−Kirkwood theory, MC simulation, etc. This shall be the topic of the next section.

Simulation Model. In some recent studies, we have demonstrated how MC simulations can be used to predict protein−protein interactions in solution (*3*) as well as the titration behavior of an isolated protein in salt solution (*21*). Here, we shall give a brief introduction to these methods; for a more thorough survey, the reader may consult the references mentioned above.

The simulations are based on a simple dielectric continuum model and exploit the Protein Data Bank to provide a detailed structural description of the protein. The protein is treated either in full atomic detail, "atomistic model", or in a simplified version, where we represent each amino acid as a sphere, "the amino acid model". The aqueous solvent is treated as a structureless continuum described by the dielectric constant ϵ_r equal to that of pure water. The ions in the surrounding salt solution are described by charged, hard spheres allowed to move in any direction within a spherical cell enclosing the system of interest, see Figure 1.

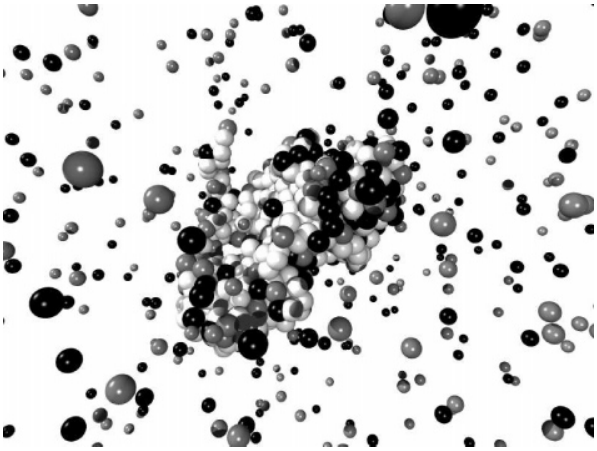


FIGURE 1: Schematic picture of the model system with the protein and salt solution enclosed in a cell.

Some of the simulations are performed for *two* proteins in a spherical cell. The same type of interactions are included, but now the two proteins are allowed to translate and rotate as well. From these simulations, we can calculate the free energy of interaction between the proteins.

In the amino acid model, each residue is replaced with a single sphere located at a center-of-mass of the residue according to the crystal structure and the radius is adjusted, so that the total protein volume is equal to that of the protein in full atomic detail. This amounts to an average amino acid radius of around 3.5 Å. The degree of protonation for each amino acid is affected by pH and by the potential produced by all other charges in the system. This has been incorporated into the simulations via a titration scheme, where protons are allowed to exchange between sites in the protein and the surrounding solution. The trial energy for such an exchange is calculated according to

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

where ΔU_{el} is the change in electrostatic energy, pK_0 is the dissociation constant of the isolated amino acid, (+) applies when protonating an amino acid, and (−) applies when deprotonating an amino acid. Sites on the protein are selected randomly as is the proton (a positive charge) in the salt solution, when deprotonating the proton is inserted at a random position. Thus, during simulation, the protein charge will respond to changes in the electrostatic surroundings, which of course will be more important when pH ≈ pK_0. Note that, because the ionized residues are located near the surface, we have assumed a relatively high dielectric response from the surroundings. This allows us to use a uniform dielectric permittivity for the whole system equal to the value of pure water. Naturally, this assumption becomes less applicable for charges buried in the protein interior where the dielectric response may be smaller; in such cases, more sophisticated models should be utilized.

The total interaction energy for the system is written as a sum of contributions from electrostatics and hard-core repulsions

$$\beta U_{\mathrm{tot}} = \sum_{i,j\in p,s}\left(l_{\mathrm{B}}\frac{q_i q_j}{r_{ij}} + u_{ij}^{\mathrm{hs}}\right) \qquad i \neq j \qquad (13)$$

Table 1: Titrating Residues in the Investigated Proteins[a]

protein	residues	Asp	Glu	His	Tyr	Lys	Cys[b]	Arg
pK_0		4.0	4.4	6.3	9.6	10.4	10.8	12.0
smMLCK (*16*)	19	0	0	1	0	2	0	3
calbindin D_{9k} (*17*)	75	4	13	0	1	10	0	0
calmodulin (*16*)	142	16	19	1	2	5	0	6
hisactophilin (*18*)	118	6	7	31	3	9	1	1
M-m-CoA mutase (*19*)	726	47	51	12	22	36	2	43
lysozyme (*20*)	129	7	2	1	3	6	0	11

[a] The dissociation constants for the isolated amino acids are given in the second line and the corresponding pK_0 for C and N termini are 3.8 and 7.5, respectively. [b] Only cysteines not engaged in sulfide bridges can titrate.

where the indexes i and j refer to salt (s) and protein (p) particles, separated by the distance r_{ij}. The hard-core term, u_{ij}^{hs} is ∞ for $r_{ij} < \sigma_i + \sigma_j$ and zero otherwise.

The model system is solved using the traditional Metropolis MC method (*22*), performed in a semicanonical ensemble. This means that the salt particles, positive or negative hard spheres, are subject to random displacement in the surrounding solution, while "protons", i.e., positive ions, can exchange between the solution and titrating groups on the protein. The whole system, including protein(s) and all ions, is electroneutral. The results from the simulation are the average charge on each titrating group and the distribution of co- and counterions around the protein. To calculate the response function, i.e., the protein capacitance, we will also calculate the protein average net charge, $\langle Q\rangle$, and the averaged squared net charge, $\langle Q^2\rangle$.

Again, we stress that MC simulation is just one of many methods that can be used to estimate capacitances; Equation 6 holds for any model.

RESULTS

Using MC simulation, we have calculated the capacitance for a number of proteins with different characteristics in terms of number and type of residues (see Table 1). Unless otherwise stated, we have used a salt concentration of 70 mM and a protein concentration of 0.7 mM. This choice of conditions allows us to treat very large protein complexes with more than 1000 residues. Figure 2a shows the capacitance for calbindin D_{9k} derived from the experimental titration curve (*21*, *23*) as well as from the atomistic and the amino acid model. The two models give virtually identical results, and below pH 9, the agreement with experimental data is very good. The discrepancy at high pH could be due to a minor unfolding of the protein; this would decrease the internal electrostatic perturbation and, as seen, shift capacitances toward the ideal curve. The main difference from the ideal capacitance curve is a strong broadening of two peaks corresponding to the response from acidic and basic residues, respectively.

Calmodulin is another calcium-binding protein, and its capacitance has the same qualitative appearance as calbindin. Both proteins have a large number of Asp/Glu residues giving rise to a large capacitance at pH 4. They also have a large portion of Lys/Arg resulting in a second peak at pH 11−12. Figure 2b shows the capacitance for calmodulin and for a small positively charged peptide from smooth muscle myosine light-chain kinase (smMCLK). The capacitance for the peptide is essentially zero below pH 10, while it peaks

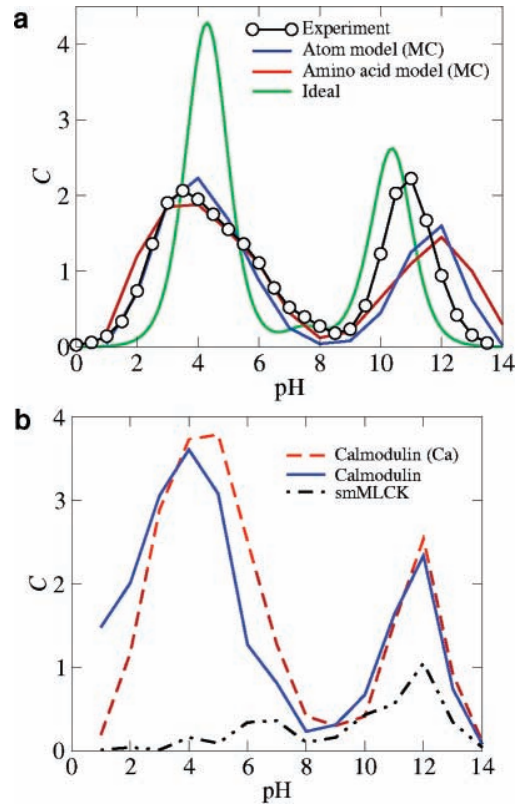


FIGURE 2: (a) Simulated, ideal, and measured [Kesvatera et al. (*21*, *23*)] capacitances for calbindin D_{9k} as a function of pH. pI for calbindin is approximately 4.2. (b) Simulated (atomic model) capacitances for smMLCK, calmodulin (calcium free), and calmodulin with four calcium ions bound. pI for calmodulin is approximately 3.9.

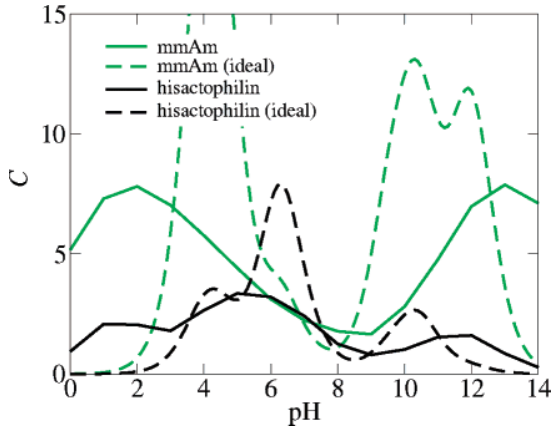


FIGURE 3: Simulated (amino acid model) and ideal capacitances for hisactophilin and methylmalonyl-CoA mutase (mmAm) as a function of pH. pI for hisactophilin is 7.3, and pI for methylmalonyl-CoA mutase is 5.2.

around pH 12 because of lysine and arginine residues. This maximum is much smaller than for calmodulin, but for the specific capacitance, i.e., $C_{sp} = C/N_{res}$, the situation is reversed.

The protein hisactophilin is of the same size as calbindin and calmodulin, but it has a slightly different capacitance curve, see Figure 3. Hisactophilin contains 31 histidine residues, which is reflected in a broad maximum for $C_{Hisacto}$ at pH 5−6. Because of the high positive charge of hisactophilin at acidic conditions (+28 at pH 3 and +23 at pH 4),

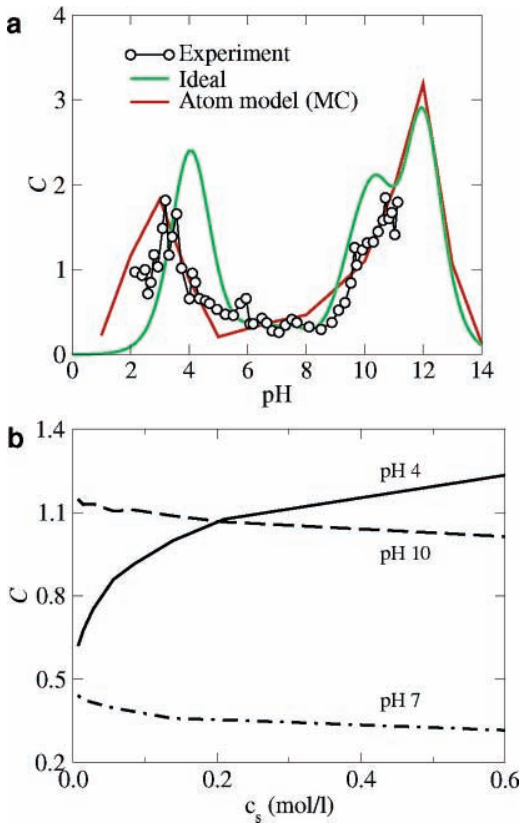


FIGURE 4: (a) Simulated, ideal, and measured capacitances for lysozyme. Experimental data are taken from Sakakibara and Hamaguchi (*26*). (b) Simulated capacitance for lysozyme (atomistic model) at different pH and salt concentrations.

the maximum is shifted downward. The isoelectric point found from the simulations is 7.3 in perfect agreement with experimental estimates (*24*). The relatively high capacitance of hisactophilin at physiological pH suggests that charge regulation may be of biological importance, for example, when coupling actin skeleton to negatively charged plasma membranes (*25*). In general, histidine-rich molecules are likely to have high capacitances around pH 7, thus making them good candidates for *in vivo* charge regulation.

The capacitance increases with protein size or more correctly with the number of titratable groups. Figure 3 also shows the capacitance for chain A in methylmalonyl-CoA mutase, which consists of more than 700 amino acids. The protein can achieve a very high net charge (+14 at pH 4 and −26 at pH 10) and a significant capacitance at extreme pH's. The isolelectric point is 5.2.

Lysozyme is another well-studied protein, and from the measured titration curve (*26*), we have extracted capacitances and, as shown in Figure 4a, agreement with the simulation is reasonable. Salt particles influence the capacitance (Figure 4b), but it is much less pronounced than the pH dependence. The salt effect is rather complex because it is governed by several mechanisms. In general, the effect of salt is to screen the electrostatic interactions, and hence, capacitances ought to approach their ideal values. However, the detailed charge distribution of the protein will modulate this effect. From this, it is difficult to derive any general statements because the balance depends not only on pH but also on the protein sequence and structure.

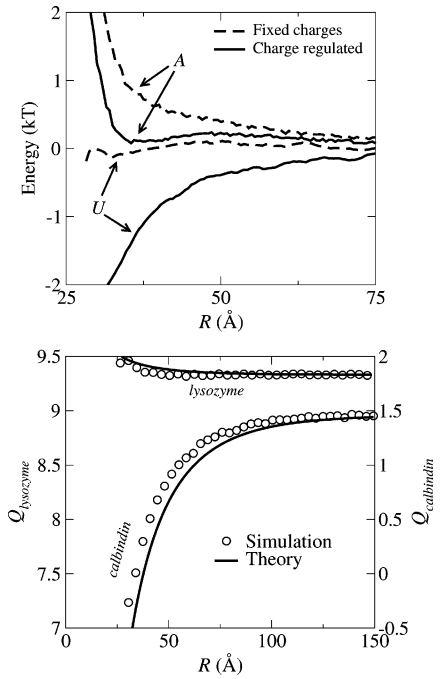


FIGURE 5: (a) Simulated energy (*U*) and free energy (*A*) of the interaction between calbindin and lysozyme at pH 4 for a protein model with fixed charges and one with charge regulation. The amino acid model is used, and the salt concentration is 6 mM. (b) Variation of the net charge of calbindin and lysozyme as a function of their separation. Circles represent simulated data based on the amino acid model, and lines are calculated from the induced charge cf. eq 5. The pH is 4, and the salt concentration is 6 mM.

The electrostatic interaction between two proteins will, at long distances, be dominated by the direct Coulomb interaction provided that the net charge, *Q*, is sufficiently different from zero. The induced interactions will play an important role only for protein−protein interactions at pH values close to the isoelectric point of one of the proteins; this can be seen from eq 6. We shall now illustrate this with an (artificial) example, where calbindin is interacting with lysozyme. Figure 5a shows the simulated free energy of interaction between the two proteins at pH 4, which is close to the isoelectric point for calbindin. At contact, there is a difference in the interaction energy of 1 kT between a model with fixed charges compared to a situation where the proteins are free to adjust their charges.

This difference between the two models is mainly due to the interaction between the induced charge in calbindin and the permanent charge in lysozyme. This is a typical result, and significant effects from charge regulation can be expected when one of the interacting proteins has a large net charge and the other has a large capacitance. Experimentally, this is confirmed by Zydney and Pujar (*9*), who were able to separate proteins by charge regulation at pH with high capacitances.

Following eq 6, we can approximate the difference between the fixed and regulated case as

$$\beta(A_{reg}(R) - A_{fix}(R)) = \beta\Delta A(R) = -\frac{l_B^2 e^{-2\kappa R}}{2R^2}(C_{calb}C_{lys} + C_{lys}Q_{calb}^2 + C_{calb}Q_{lys}^2) \quad (14)$$

where we have replaced the Coulomb interaction with a simple screened version to approximately account for the effect of salt. Figure 6 shows a perfect agreement between the simulated free-energy difference and the calculated one according to eq 14. Therefore, with minimal effort, eq 14 can be applied to existing models, for example, the DLVO approximation, so as to account for the induced charge interaction.

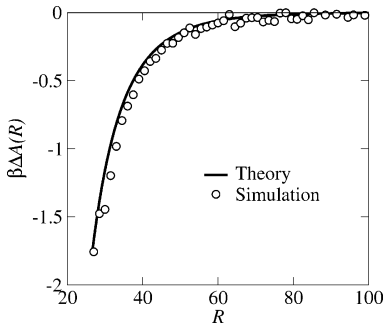


FIGURE 6: Difference in the free energy of interaction between calbindin and lysozyme at pH 4, with 6 mM salt for a protein model with charge regulation and one with fixed charges. Symbols denote the simulated difference (see Figure 5), and the solid line is obtained from eq 14 with $Q_{calb} = 1.46$, $C_{calb} = 2.23$, $Q_{lys} = 9.33$, $C_{lys} = 0.88$, and $1/\kappa = 39$ Å.

An interesting result is that, despite the fact that both calbindin and lysozyme are positively charged when isolated at pH 4, there is still an attractive electrostatic interaction between the two. Such an attraction could of course be due to charge−dipole and/or dipole−dipole interactions, but in the present case, the main contribution to the interaction free energy comes from the induced charges. This is further demonstrated in Figure 5b, where one can follow how the net charge of calbindin changes from +1.5 at infinite separation to −0.5 at contact between calbindin and lysozyme.

DISCUSSION

We have derived an expression for the protein charge capacitance and showed how it appears as a response to an externally applied potential. The same capacitance also enters the expression for the free energy of interaction between two proteins allowed to regulate their charges. The induced interaction coming from a charge regulation mechanism can be important if one of the proteins is close to its isoelectric point, while the other carries a net charge. Via the presented formalism, existing models for protein−protein or protein−membrane interactions can be easily improved to correctly describe the induction interaction. Predictions of interaction free energies and induced protein charges are in excellent agreement with MC simulations that represent the exact theoretical solution.

The protein charge capacitance, *C*, can be obtained from measured titration curves and from any theoretical model able to predict the protein charge; our simple continuum simulations for the protein protonation status reproduce the experiment very well.

The capacitance varies with pH, and its magnitude is related to the number of titratable groups in a protein. Typically, *C* will be large at pH values in the neighborhood of the pK_a values of titrating residues, although the capacitance maximum can be shifted one or two pH units from the ideal maximum. The capacitance curve will also be

broadened because of interactions within the protein. This means that, in a protein with many aspartates and/or glutamates, the capacitance will have a maxium around pH 4, while a large number of histidines will lead to a maximum at pH $\approx$ 6 and similar for the basic residues.

REFERENCES

1. Allahyarov, E., Löwen, H., Hansen, J. P., and Louis, A. A. (2002) Discrete charge patterns, coulomb correlations, and interactions in protein solutions, *Europhys. Lett. 57*, 731−737.
2. Carlsson, F., Malmsten, M., and Linse, P. (2001) Monte Carlo simulations of lysozyme self-association in aqueous solution, *J. Phys. Chem. 105*, 12189−12195.
3. Lund, M., and Jönsson, B. (2003) A mesoscopic model for protein−protein interactions in solution, *Biophys. J. 85*, 2940−2947.
4. Striolo, A., Bratko, D., Wu, J. Z., Elvassore, N., Blanch, H. W., and Prausnitz, J. M. (2002) Forces between aqueous nonuniformly charged colloids from molecular simulation, *J. Chem. Phys. 116*, 7733−7743.
5. Kirkwood, J. G., and Shumaker, J. B. (1952) Forces between protein molecules in solution arising from fluctuations in proton charge and configuration, *Proc. Natl. Acad. Sci. U.S.A. 38*, 863−871.
6. Carnie, S. L., and Chan, D. Y. C. (1993) Interaction free energy between plates with charge regulation: A linearized model, *J. Colloid Interface Sci. 161*, 260−264.
7. Carnie, S. L., Chan, D. Y. C., and Gunning, J. S. (1994) Electrical double layer interaction between dissimilar spherical colloidal particles and between a sphere and a plate: The linearized Poisson−Boltzmann theory, *Langmuir 10*, 2993−3009.
8. Pujar, N. S., and Zydney, A. L. (1997) Charge regulation and electrostatic interactions for a spherical particle in a cylindrical pore, *J. Colloid Interface Sci. 192*, 338−349.
9. Zydney, A. L., and Pujar, N. S. (1998) Protein transport through porous membranes: Effects of colloidal interactions, *Colloids Surf., A 138*, 133−143.
10. Menon, M. K., and Zydney, A. L. (2000) Determination of effective protein charge by capillary electrophoresis: Effects of charge regulation in the analysis of charge ladders, *Anal. Chem. 72*, 5714−5717.
11. Bowen, W. R., and Williams, P. M. (1996) The osmotic pressure of electrostatically stabilized colloidal dispersions, *J. Colloid Interface Sci. 184*, 241−250.
12. Ståhlberg, J., Appelgren, U., and Jönsson, B. (1995) Electrostatic interactions between a charged sphere and a charged planar surface in an electrolyte solution, *J. Colloid Interface Sci. 176*, 397−407.
13. Ståhlberg, J., and Jönsson, B. (1996) Influence of charge regulstion in electrostatic interaction chromatography of proteins, *Anal. Chem. 68*, 1536−1544.
14. Jönsson, B., and Ståhlberg, J. (1999) The electrostatic interaction between a charged sphere and an oppositely charged planar surface and its application to protein adsorption, *Colloids Surf., B 14*, 67−75.
15. Bratko, D., Woodward, C. E., and Luzar, A. (1991) Charge fluctuation in reverse micelles, *J. Chem. Phys. 95*, 5318.
16. Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin−peptide complex, *Science 257*, 1251−1255.
17. Szebenyi, D. M. E., and Moffat, K. (1986) The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. Molecular details, ion binding, and implications for the structure of other calcium-binding proteins, *J. Biol. Chem. 261*, 8761−8777.
18. Habazettl, J., Gondol, D., Wiltscheck, R., Otlewski, J., Schleicher, M., and Holak, T. A. (1992) Structure of hisactophilin is similar to interleukin-1b and fibroblast growth factor, *Nature 359*, 855−858.
19. Mancia, F., and Evans, P. R. (1998) Conformational changes on substrate to methylmalonyl coa mutase and new insights into the free radical mechanism, *Structure 6*, 711−720.
20. Ramanadham, M., Sieker, L. C., and Jensen, L. H. (1990) Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares, *Acta Crystallogr., Sect. B 46*, 63−69.
21. Kesvatera, T., Jönsson, B., Thulin, E., and Linse, S. (1999) Ionization behaviour of acidic residues in calbindin D_{9k}, *Proteins 37*, 106−115.
22. Metropolis, N. A., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A., and Teller, E. (1953) Equation of state calculations by fast computing machines, *J. Chem. Phys. 21*, 1087−1097.
23. Kesvatera, T., Jönsson, B., Thulin, E., and Linse, S. (2001) Focusing of the electrostatic potential at ef-hands of calbindin D_{9k}. Titration of acidic residues, *Proteins 45*, 129−135.
24. Hanakam, F., Eckerskorn, C., Lottspeich, F., Müller-Taubenberger, A., Schäfer, W., and Gerisch, G. (1995) The pH-sensitive actin-binding protein hisactophilin of dictyostelium exists in two isoforms which both are myristoylated and distributed between plasma membrane and cytoplasm, *J. Biol. Chem. 270*, 596−602.
25. Hanakam, F., Gerisch, G., Lotz, S., Alt, T., and Seelig, A. (1996) Binding of hisactophilin I and II to lipid membranes is controlled by a pH-dependent myristoyl-histidine switch, *Biochemistry 35*, 11036−11044.
26. Sakakibara, R., and Hamaguchi, K. (1968) Structure of lysozyme, *J. Biochem. 64*, 613−618.

BI047630O

Enhanced Protein Adsorption Due to Charge Regulation

Mikael Lund,* Torbjörn Åkesson, and Bo Jönsson

Theoretical Chemistry, Chemical Center, POB 124, S-221 00 Lund, Sweden

Received March 7, 2005. In Final Form: May 27, 2005

When a protein molecule approaches a charged surface, its protonation state can undergo dramatic changes due to the imposed electric potential. This has a large impact on adsorption strengths that may be enhanced by several kT. Using mesoscopic simulation techniques as well as analytical theories, we have investigated this regulation mechanism and demonstrate how it is influenced by salt concentration and solution pH. Using hisactophilin as a test case, we show how the binding to a lipid membrane is governed by small changes in pH and that this is intimately coupled to the charge regulation mechanism.

Introduction

More than 50 years ago, Kirkwood and Shumaker[1] used statistical mechanical perturbation theory to discuss charge regulation in bio-molecules, i.e., that the total protein charge is *not constant* but affected by nearby charged objects (other proteins, membranes, DNA, etc.). Although Hill[2] covered this work, today charge regulation is still considered somewhat exotic and is absent from most text books. This is in spite of the fact that the regulation interaction is on line with ion−dipole and dipole−dipole interactions and can be more long ranged. Several workers have presented applications of charge regulation, including its importance for protein−protein interactions,[3−5] protein adsorption to surfaces,[6−8] and pores.[9] In particular, Ståhlberg and Jönsson[6] noted that a proteins ability for charge induction is proportional to the slope of the pH titration curve. In a recent study,[5] we showed that this slope, $\partial Q/\partial$pH, is directly related to charge fluctuations and that it enters the original expression for the free energy presented by Kirkwood and Shumaker back in 1952. In the present work, we shall advance on this concept and derive expressions valid for a protein near a charged surface.

Theoretically, a charged planar surface in a salt solution can be described using the well-known Gouy−Chapman (GC) theory[10] and in the most simple case the protein is described as a charged sphere. Although this very simple approach is able to qualitatively account for many phenomena,[11] it neglects several features, namely that (1) the protein charge distribution is not spherical symmetric. and (2) the protein net charge is not constant. Furthermore, the GC theory fails to describe ion correlation effects, important for multivalent ions. These issues can be resolved using Monte Carlo simulations that represent the exact numerical solution to the given statistical mechanical model. Using a mesoscopic protein description, the essential physics of protein−protein interactions can be captured,[12,5] and we shall here expand the model to include a charged wall.

Hisactophilin, a small 13 kDa globular protein rich on histidines (26%) is used as a test case to demonstrate charge regulation under different solution conditions. It binds in vivo to negatively charged membranes, and the process is governed by small changes in intracellular pH.[13] At pH $\approx$ 7.5, no binding is observed, but as the pH is lowered to around 6.5, the protein is attached to the membrane. This conforms with the isoelectric point of 7.1 so that lowering pH will render the protein positively charged, thus attracting it to the membrane. In addition to this purely electrostatic interaction, there is evidence of an anchoring of a myristoyl chain attached to the N-terminus. We will not penetrate this mechanism, but concentrate solely on the electrostatic contribution.

Model and Theory

Monte Carlo Simulation. The protein-charged wall system is investigated using the traditional Metropolis Monte Carlo simulation method.[14] Although atomistic protein models are tractable within this scheme, coarse graining the atomic NMR or X-ray structure can reduce the computational cost significantly. An efficient and accurate method is to present each amino acid by a single hard sphere located in its mass center. This simplification

* Corresponding author. Phone: +46-46-222-0381. E-mail: mikael.lund@teokem.lu.se.

(1) Kirkwood, J. G.; Shumaker, J. B. Forces between protein molecules in solution arising from fluctuations in proton charge and configuration. *Proc. Natl. Acad. Sci.* **1952**, *38*, 863−871.

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

(3) Phillies, G. D. J. Excess chemical potential of dilute solutions of spherical polyelectrolytes. *J. Chem. Phys.* **1973**, *60*, 2721−2731.

(4) Grant, M. L. Nonuniform charge effects in protein−protein interactions. *J. Phys. Chem. B* **2001**, *105*, 2858−2863.

(5) Lund, M.; Jönsson, B. On the charge regulation of proteins. *Biochemistry* **2005**, *44*, 5722−5727.

(6) Ståhlberg, J.; Jönsson, B. Influence of charge regulation in electrostatic interaction chromatography of proteins. *Anal. Chem.* **1996**, *68*, 1536−1544.

(7) Biesheuvel, P. M.; van der Veen, M.; Norde, W. A modified poisson-boltzmann model including charge regulation for the adsorption of ionizable polyelectrolytes to charged interfaces, applied to lysozyme adsorption on silica. *J. Phys. Chem. B* **2005**, *109*, 4171−4180.

(8) Biesheuvel, P. M.; Wittemann, A. A modified box model including charge regulation for protein adsorption in a spherical polyelectrolyte brush. *J. Phys. Chem B* **2005**, *109*, 4209−4214.

(9) Pujar, N. S.; Zydney, A. L. Charge regulation and electrostatic interactions for a spherical particle in a cylindrical pore. *J. Colloid Interface Sci.* **1997**, *192*, 338−349.

(10) Israelachvili, J. N. *Intermolecular and Surface Forces*, 2nd ed.; Academic Press: London, 1992.

(11) Seelig, A.; M., M. P. Binding of a neuropeptide, substance p, to neutral and negatively charged lipids. *Biochemistry* **1989**, *28*, 2490−2496.

(12) Lund, M.; Jönsson, B. A mesoscopic model for protein−protein interactions in solution. *Biophys. J.* **2003**, *85*, 2940−2947.

(13) Hanakam, F.; Gerisch, G.; Lotz, S.; Alt, T.; Seelig, A. Binding of hisactophilin i and ii to lipid membranes is controlled by a ph-dependent myristoyl-histidine switch. *Biochemistry* **1996**, *35*, 11036−11044.

(14) Metropolis, N. A.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087−1097.

10.1021/la050607z CCC: $30.25
Published on Web 00/00/0000 PAGE EST: 3.7

may seem drastic but in fact the charge distribution as well as the protein surface is well conserved.[12] In this study, we use the NMR structure of hisactophilin as determined by Habazettl et al.[15] and an amino acid diameter of 7 Å.

Mobile ions (salt and counterions) and the protein are immersed in a cubic box of side length 150 Å where one side is coated with charges so as to mimic a charged wall. Sampling is done in the NVT-ensemble using the minimum image convention as well as periodic boundary conditions in two dimensions. The protein molecule can freely rotate around its center of mass as well as translate perpendicular to the charged wall. Ions can move in any direction, except those representing the charged wall; these are restricted to movements on their box side. Unless otherwise stated, the wall charge is negative with a density of 300 Å^2 per charge, corresponding to -0.053 C/m^2. In a biological context this value is reasonable.

The solvent is treated as a structure-less medium, described solely by a temperature-dependent dielectric constant, ϵ_r. Ions and other particles are represented by hard spheres with diameter σ and charge number q. This comprises the primitive model of electrolytes and the energy of a particle, i, interacting with another, j, is evaluated according to

$$u_{ij} = \begin{cases} \dfrac{q_i q_j e^2}{4\pi\epsilon\epsilon_r r_{ij}} & r_{ij} \geq (\sigma_i + \sigma_j)/2 \\ +\infty & r_{ij} < (\sigma_i + \sigma_j)/2 \end{cases} \qquad (1)$$

where r_{ij} is the separation, e is the electronic charge, and ϵ_0 is the permittivity of vacuum. Here we have assumed a uniform dielectric response throughout the cell. Although this is a valid description for proteins with charges near their surfaces, it may be less applicable for proteins with buried, charged groups. In this study, we concentrate on electrostatic interactions only and thus ignore dispersion and hydrophobic interactions.

The ionization states of the titrateable residues are determined by continuously exchanging protons between the protein and the salt solution. This is done by randomly selecting residues and move their protons to the solution, and visa versa. The proper Boltzmann factor for this process is[16,5]

$$\exp[-\Delta U_{el}/kT \pm \ln 10(\mathrm{pH} - pK_a)] \qquad (2)$$

where ΔU_{el} is the change in electrostatic energy, pK_a is the dissociation constant of the isolated amino acid, (+) applies when protonating an amino acid, and (−) when deprotonating. In terms of thermodynamics, the left-hand side of the exponential, ΔU_{el}, corresponds to the sites excess chemical potential or activity coefficient. This semicanonical approach allows the protein to respond to changes in the electrostatic surroundings, i.e., the protonation state will depend on pH, the protein structure, salt concentration, and on the protein−wall separation.

To summarize, the simulation model describes the protein in mesoscopic detail, treats mobile ions explicitly, and finally lets the protein charge respond to electric perturbations.

(15) Habazettl, J.; Gondol, D.; Wiltscheck, R.; Otlewski, J.; Schleicher, M.; Holak, T. A. Structure of hisactophilin is similar to interleukin-1b and fibroblast growth factor. *Nature* **1992**, *359*, 855−858.

(16) Ullner, M.; Woodward, C. E.; Jönsson, B. A Debye−Hückel theory for electrostatic interactions in proteins. *J. Chem. Phys.* **1996**, *105*, 2056−2065.

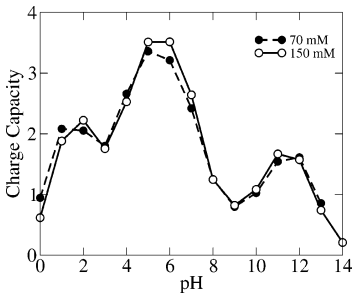


Figure 1. Charge capacitance for hisactophilin obtained from a simulation of a single protein molecule at two different salt concentrations.

Binding Free Energy. From the simulation, we obtain the protein-wall potential of mean force, $w(r)$ which can be represented in a more compact manner via a binding free energy

$$\langle w \rangle = \frac{\int_0^\tau w(r)\exp(-\beta w(r))\,\mathrm{d}r}{\int_0^\tau \exp(-\beta w(r))\,\mathrm{d}r} \qquad (3)$$

where r is the distance from the wall to the proteins mass center and $\beta = 1/kT$ where k is Boltzmann's constant. The range $[0;\tau]$, defining the protein−wall complex is subject to an operational definition and should be close to the range of interaction.[17] For salt solutions, this is approximately the Debye-length, $1/\kappa$ and thus we set τ to

$$\tau = 1/\kappa + contact$$

From an experimental point of view, τ depends on the method since different apparatuses have different criteria for discriminating the bound and nonbound state. Hence, slightly different binding constants can be expected from different experiments.

Protein Charge Capacitance. Although charge regulation may be investigated using molecular simulation, we will here present formal expressions to elucidate the observed behavior. A key concept in our analysis is the protein charge capacitance[5] that quantifies the proteins ability for charge fluctuations

$$C \equiv \langle Q^2 \rangle - \langle Q \rangle^2 = -\frac{1}{\ln 10}\frac{\partial Q}{\partial \mathrm{pH}} \qquad (4)$$

C is an intrinsic protein property and can be obtained experimentally or from theory as the derivative of the pH titration curve. When pH is close to pK_a of a titrateable site, the protonation status is easily disturbed by electrostatics, thus increasing the capacitance. For example, for a histidine rich protein (p$K_a^{his} \approx 6.5$) C is expected to peak around physiological pH. Of course this is slightly modulated by solution conditions as shown in Figure 1.

An important property of the capacitance is that it enters the expression for the free energy of interaction as shall now be derived. The interaction energy, u between an external electric potential, Φ and a protein molecule with partial charge numbers, q_i can be written as

$$\beta u = \frac{e\Phi}{kT}\sum q_i \qquad (5)$$

The origin of the external potential may be a charged surface or another molecule, assuming that the spatial

(17) Wennerström, H. *Organized solutions: surfactants in science and technology*; Marcel Dekker: New York, 1992; p 410.

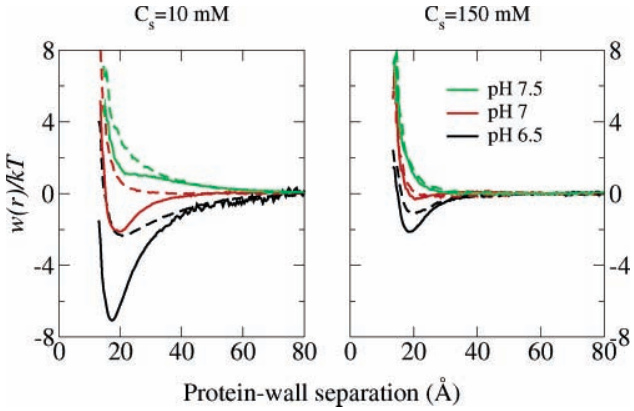


Figure 2. Free energy of interaction between hisactophilin and a negatively charged wall at 10 mM (left) and 150 mM (right) 1:1 salt. Dashed lines are for a model with fixed protein charges.

variation is small compared to the radius of the protein. From statistical mechanical perturbation theory,[18] the free energy can be approximated with

$$\beta A = -\ln \langle \exp(-\beta u) \rangle \approx \langle \beta u \rangle - \frac{1}{2}(\langle (\beta u)^2 \rangle - \langle \beta u \rangle^2) \qquad (6)$$

where $\langle ... \rangle$ denotes an average over all configurations of the unperturbed system. Now, combining eq 4−6, we arrive at the following expression:

$$\beta A = \frac{e\Phi}{kT} Q - \frac{1}{2}\left(\frac{e\Phi}{kT}\right)^2 C \qquad (7)$$

The first term is the direct Coulomb interaction between the protein and the potential and the second term is the interaction arising from charge fluctuations. To account for the proteins nonspherical charge distribution, one can easily add a dipole term to eq 7.

Analytical theories such as that of Gouy and Chapman, provide distance dependent expressions for Φ and take into account the effect of salt and counterions. Thus, using the protein charge and capacitance (as obtained from a titration curve), eq 7 can be used to estimate the adsorption free energy at different solution conditions.

Results and Discussion

Often, when estimating protein−wall interactions, the protein charge is kept constant; that is, no charge fluctuations are permitted. The free energy of interaction for such a system is shown in Figure 2 (dashed lines), and it is clear that the binding is strongly affected by pH and salt concentration. However, if charge regulation is included, the attractive interaction is drastically enlarged (Figure 2, solid lines), especially at low salt concentration. Comparing with the traditional fixed-charge case at pH 6.5 and 10 mM salt, we have lowered the free energy minimum from −2 to −7 kT or more than 12 kJ/mol. This is a difference not to be neglected, and it clearly demonstrates that charge fluctuations indeed can be crucial. A similar finding was done by Biesheuvel et al.,[7] who, by solving the Poisson−Boltzmann equation, showed that lysozyme binds to a negatively charged silica surface for pH > p*I* at low salt concentrations and attributed this to charge regulation.

(18) McQuarrie, D. A. *Statistical Mechanics*; Harper Collins: New York, 1976.

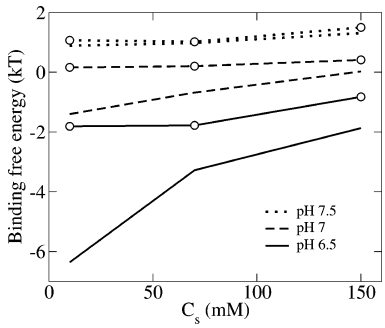


Figure 3. Electrostatic binding free energy at different pH and salt concentrations. Lines are for a model with charge regulation; lines with symbols for one with a fixed protein charge.

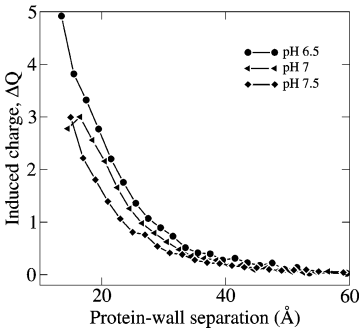


Figure 4. Induced charge as a function of the protein-wall separation at different pH. C_s = 70 mM.

Increasing the salt concentration rapidly screens electrostatic interactions, but even at 150 mM salt, a free energy minimum is found for pH 6.5. Again this minimum is enhanced by charge regulation, this time by around 1 kT/molecule. Figure 3 shows how the binding free energy varies with pH and salt concentration as well as the effect of charge regulation. At high pH, there is a net repulsion due to the negative protein charge and the effect of charge regulation seems insignificant. As pH is lowered, the protein goes through its isoelectric point, and at pH 6.5, the charge is around +5. Simultaneously, charge regulation becomes increasingly more important.

This is further demonstrated by following the induced charge as a function of protein−wall separation in Figure 4. At long distances, the protein is unaffected by the wall potential, but at shorter separations, it becomes increasingly more protonated. This strengthens the electrostatic attraction, especially at pH 6.5 where the protein charge has doubled at contact.

Another interesting mechanism when a protein adsorbs to a surface is the polarizability of charged headgroups in the lipid membrane. In Figure 5, we show the variation in average surface charge density, calculated around the perpendicular axis connecting the protein and the surface. In the case of a fixed protein charge, no change in surface charge density is observed, but when charge regulation is included, we see a narrow band where the surface charge density has been increased by more than six times. As shown in Figure 4 the induced charge can be substantial, leading to an increased potential exerted on the planar surface. Depending on the protein structure and charge distribution, the charged headgroups may then organize to minimize the free energy. However, since the band is very narrow, corresponding to roughly one charge, the contribution to the total free energy of interaction is small.

We will now account for the above, simulated findings in terms of the charge capacitance as shown in Figure 1.

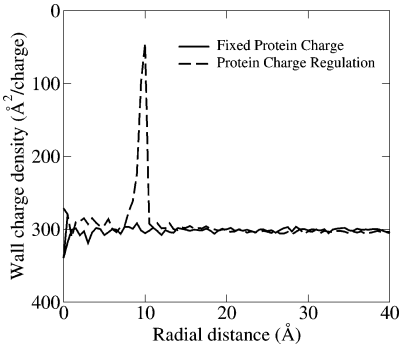


Figure 5. Variation in surface charge density in spherical shells around the perpendicular axis connecting the wall and the protein, averaged over configurational space. pH = 6.5 and C_s = 150 mM.

According to eq 7, the regulation free energy scales linearly with the capacitance and thus, for hisactophilin, we expect charge regulation to be most important at pH 5−7. In this interval it is interesting to note the marked capacitance gradient around pH 7 and that changing pH from 7.5 to 6 doubles the capacitance, rendering the protein much more susceptible to electrostatic perturbations. That this occurs precisely at physiological conditions is hardly a coincidence, but rather an efficient evolutionary design to control binding properties. Now, to quantify the interaction, we need to evaluate the magnitude of the potential from the charged surface, salt, and counterions. For this, we use the simple, linearized solution of the Gouy−Chapman potential

$$\Phi(r) = \frac{4\gamma kT}{e} \exp(-\kappa r) \qquad (8)$$

where $1/\kappa$ is the Debye length and $\gamma = \tanh(e\psi_0/4kT)$, ψ_0 being the surface potential. Figure 6 shows how the regulation interaction varies with pH and salt concentration and is very similar to that obtained by the more elaborate simulation method. Notable, this simple approach is able to predict the strong regulation dip below pH 7 and low salt concentration as is also found in Figure 3. As the salt concentration is increased, the potential from the surface is efficiently screened by $\exp(-2\kappa r)$ and the induced interactions play only a minor role. To some extent, this effect may be counter-acted by the fact that capacitance peaks are often enlarged when increasing the salt concentration. This stems from screening of internal

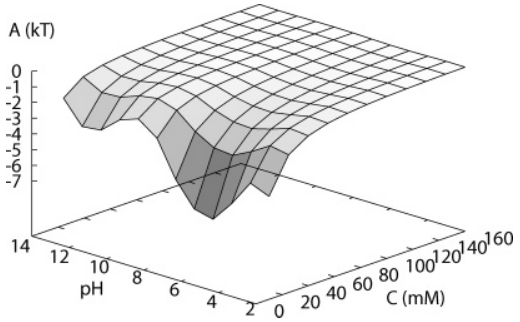


Figure 6. Charge regulation interaction at different pH and salt concentrations calculated from the Gouy−Chapman potential at protein-wall contact (15 Å), using a surface charge density of 300 Å^2.

electrostatic interactions in the biomolecule, causing the capacitance profile to resemble the ideal capacitance, which has a more distinct form.[5] The results presented in Figure 6 are calculated using capacitances for 70 mM salt and as such does not include this effect.

Conclusions

We have investigated the physiologically important interaction between hisactophilin and a charged, lipid membrane. As also found experimentally, the binding can be controlled by very small changes in pH and salt plays an important role for the adsorption. At pH 6.5 and 10 mM salt concentration, the charge regulation mechanism can contribute significantly (5 kT) to the adsorption free energy. This is explained by the unusually high capacitance of the histidine-rich protein at this pH. Increasing pH strongly reduces charge regulation, since hisactophilin's capacitance is considerably diminished by even small changes in pH. At physiological salt concentrations, the same behavior is observed, but the contribution from charge regulation is now reduced to around 1 kT, still creating a distinct minimum in the potential of mean force.

The derived expression for the regulation free energy between a protein and an electrical potential is general and can, in combination with the Gouy−Chapman theory, account for protein adsorption at different pH and salt conditions. Using this framework together with the slope of the titration curve, i.e., the capacitance, we arrive at the same conclusions as obtained by the more involved Monte Carlo simulation method.

LA050607Z

J. Phys. Chem. B **2006,** *110,* 4459−4464

On the Complexation of Proteins and Polyelectrolytes

Fernando Luís B. da Silva,*,† Mikael Lund,‡ Bo Jönsson,‡ and Torbjörn Åkesson‡

Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Av. do café, s/no., Universidade de São Paulo, 14040-903 Ribeirão Preto, SP, Brazil, and Department of Theoretical Chemistry, Lund University, POB 124, S-221 00 Lund, Sweden

Received: August 29, 2005; In Final Form: January 4, 2006

Both natural and synthetic polyelectrolytes form strong complexes with a variety of proteins. One peculiar phenomenon is that association can take place even when the protein and the polyelectrolyte carry the same charge. This has been interpreted as if the ion−dipole interaction can overcome the repulsive ion−ion interaction. On the basis of Monte Carlo simulations and perturbation theory, we propose a different explanation for the association, namely, charge regulation. We have investigated three different protein−polymer complexes and found that the induced ionization of amino acid residues due to the polyelectrolyte leads to a surprisingly strong attractive interaction between the protein and the polymer. The extra attraction from this charge-induced charge interaction can be several *kT* and is for the three cases studied here, lysozyme, α-lactalbumin, and *β*-lactoglobulin, of the same magnitude or stronger than the ion−dipole interaction. The magnitude of the induced charge is governed by a response function, the protein charge capacitance $\langle Z^2\rangle - \langle Z\rangle^2$. This fluctuation term can easily be calculated in a simulation or measured in a titration experiment.

I. Introduction

The complexation of polyelectrolytes and proteins is extensively used in pharmaceutics, foods, and cosmetics.[1−9] The subject has been addressed by a number of authors exploring it from experimental measurements[8−13] to theoretical modeling.[10,14,15] The strength of interaction is to a large extent regulated by electrostatic interactions, governed by key parameters such as pH and salt concentration.[8−10]

A particularly interesting issue is the apparently paradoxical formation of soluble complexes at conditions where the net charges of the protein and the polyelectrolyte have the same sign. Experimental studies of Dubin, Kruif, and co-workers[9−12,16] have demonstrated this special feature of the polymer/protein complexation. The term complexation "on the wrong side" has been used, meaning that a polyanion forms a complex with a protein at a pH above the isoelectric point of the protein.[4,9,14,16] The molecular interpretation of such studies has focused on the assumption of "charged patches" on the protein surface.[9,11,12,14,17] This mechanism has also been used in order to explain protein chromatography data.[18]

Following this reasoning, a polyanion monomer should bind in a positive protein region, and vice versa. The same kind of argument has been used when discussing the interaction between two protein molecules at the isoelectric point, pH ≈ pI. A more formal way to describe the interaction between oppositely charged patches on two protein molecules is in terms of a multipole expansion. That is, for two neutral protein molecules the leading terms would then be dipole−dipole, dipole−quadrupole, etc. Other electrostatic properties of the protein, however, may be more important, and Kirkwood and Shumaker[19] demonstrated theoretically in 1952 that fluctuations of residue charges in two proteins can result in an attractive force. Recently, Lund and Jönsson[20] have taken up this idea and used Monte Carlo simulations and a charge regulation theory in order to explain protein−protein association in a purely electrostatic model.

In a system were electrostatic interactions are known to be relevant, protein charge fluctuations could be an important component of models aimed to describe such systems. This potentially relevant contribution was neglected in previous simulations.[9,14−16] Conversely, the authors invoked an additional $1/R^6$ attractive potential in order to better describe the complexation "on the wrong side".[9,14,15]

The purpose of this work is 2-fold: The first is to demonstrate that a polyelectrolyte and a protein molecule at its isoelectric point do form complexes in a *purely* electrostatic model. We will also show that the driving force for the complexation can be due to charge fluctuations in the protein. The second issue is the relative importance of "charged patches", i.e. charge−dipole interactions, etc., versus the charge regulation term. Second-order perturbation theory offers an easy way to get a qualitative picture of the significance of these terms. We have chosen a set of globular proteins extensively investigated in the literature in order to demonstrate these interactions, namely, lysozyme (lys), α-lactalbumin (α-lac), *β*-lactoglobulin (*β*-lac), bovine serum albumin (bsa), insulin (ins), and calmodulin (CaM).

Lysozyme, α-lactalbumin, and *β*-lactoglobulin were chosen to be extensively studied due to both their biological relevance and particular physicochemical features. These milk proteins exhibit some interesting properties and are often used as models of protein folding, stability, complex formation, and other biophysical and biochemical studies.[21−28] Lysozyme is a small enzyme that hydrolyzes the glycosidic bond between *N*-acetylmuramic acid and *N*-acetylglucosamine. In the human body it acts as a barrier preventing infections.[21,22] α-Lactalbumin is a calcium metalloprotein that is responsible for the formation

* To whom correspondence may be addressed. Phone: +55 (16) 3602 42 19. Fax: +55 (16) 3633 29 60. E-mail: fernando@fcfrp.usp.br.
† Universidade de São Paulo.
‡ Lund University.

10.1021/jp054880l CCC: $33.50
Published on Web 02/09/2006

of lactose in the mammary gland.[23] Different other functional properties such as apoptosis and induction of cell growth inhibition have also been attributed to this whey protein.[23−25] Even though β-lactoglobulin, the primary component of whey, might be related to milk allergy, shows the ability to bind small hydrophobic molecules such as retinol and fatty acids, and has significant industrial interest,[26−30] its biological function remains unclear.[27,28,30]

II. Model and Simulation

The proteins were modeled as rigid bodies in full atomistic detail according to the X-ray structures provided by the Protein Data Bank[31,32] (PDB identities are 2LZT, 1HFY, 1BEB, 1AO6, 1APH, and 1CLL for lys, α-lac, β-lac, bsa, ins, and CaM, respectively). To account for the acid−base equilibrium, the initial atomic charges were allowed to change their charge state according to the solution pH.[33]

All protein atoms present in the X-ray structure are described by hard spheres of radius $R_a = 2$ Å. This is a reasonable size, and the results are not sensitive to this particular detail. The protein was kept fixed at the center of an electroneutral spherical cell, whose radius R_{cell} was determined by the protein concentration. This so-called *cell model*[34,35] has been used successfully in the past.[36,37] The electrolyte solution surrounding the macromolecule, including neutralizing counterions, is described by the restricted primitive model.[38] Each mobile ion k with charge q_k is treated explicitly as a hard sphere of radius $R_a = 2$ Å, while the solvent is treated as a structureless dielectric medium characterized by a relative dielectric permittivity ϵ_s.

The interaction between any two particles is given by

$$u(r_{ij}) = \begin{cases} \infty & r_{ij} \le 2R_a \\ \dfrac{q_i\, q_j}{4\pi\epsilon_0\epsilon_s r_{ij}} & \text{otherwise} \end{cases} \tag{1}$$

where ϵ_0 is the vacuum permittivity, q_i and q_j denote the charges on particles i and j, respectively, and r_{ij} represents their separation.

A single flexible polyelectrolyte is modeled as a chain of $N_{mon} = 21$ charged hard spheres of radii $R_{mon} = 2$ Å and charges of $q_{mon} = -e$ (e is the elementary charge) connected by harmonic springs. The bond interaction potential between neighboring monomers is calculated as

$$\beta u^{bond} = \frac{l_B}{2r_{min}{}^3} \sum_{i=1}^{N_{mon}-1} (r_{i,i+1})^2 \tag{2}$$

where $r_{i,i+1}$ is the distance between monomer i and $i+1$, r_{min} is the separation corresponding to the energy minimum for a dimer and $l_B = e^2/4\pi\epsilon_0\epsilon_s kT$ is the Bjerrum length. We have used a value of 4 Å for r_{min}, which results in an average monomer−monomer separation of approximately 7.4 Å. The polymer is not allowed to titrate. The total energy of the system for a given configuration is then

$$U = \sum_{i=1}^{N_{mob}} v^{ex}(r_i) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} u(r_{ij}) + u^{bond} \tag{3}$$

where $N_{mob} = N_c + N_s + N_{mon}$ is the total number of mobile particles comprising N_c counterions, N_s added salt ions, and N_{mon} number of polyanion beads and $N = N_{mob} + N_p$ is the total number of particles including the N_p protein atoms. The term $v^{ex}(r_i)$ is the imposed hard wall that defines the cell

$$v^{ex}(r_i) = \begin{cases} 0 & r_i \le R_C \\ \infty & \text{otherwise} \end{cases} \tag{4}$$

The dielectric constant, ϵ_s, was set to 77.8 at room temperature of 300 K.

Simulation Details. Single protein properties, average residue charges on lysozyme, α-lactalbumin, and β-lactoglobulin together with their dipole moments, and capacitances at different pH values were obtained initially in the simulations. The simulation cell, $R_C = 189$ Å, contained a single titrating protein fixed at the center plus 20 ion pairs as well as neutralizing counterions. The polyelectrolyte was absent in these simulations, and the corresponding pI values for the studied proteins were acquired from the average protein charge as a function of pH. Partial residual charges of the protein at pI were saved and used in simulations B and C.

The complexation between the protein and the polyelectrolyte was studied in the second set of simulations. The protein was kept fixed at the center of the cell, while the polyelectrolyte was free to move within the cell. The probability distribution, $P(R)$, for the separation between the center of mass of the two was sampled, and during a first production run an approximate potential of mean force, $w(R)$, was generated

$$\beta w(R) = -\ln P(R) + \text{constant} \tag{5}$$

The probability distribution, $P(R)$, was updated during a second production run, and a final well-converged potential of mean force was obtained. The cell boundary introduces an artificial repulsion in the potential of mean force at separations close to R_C. It is easy to correct for this depletion effect by subtracting off the potential of mean force for a polyelectrolyte in a cell without a protein. All simulations were done with a protein concentration of 0.06 M and salt concentration of approximately 1 mM. An appropriate number of counterions was always present in order to obtain an electroneutral system.

The simulations were performed in a semi-grand-canonical ensemble using the standard Metropolis Monte Carlo algorithm[39] with random displacements of mobile species (salt, counterions, and polyanion beads) within the cell. In addition, the simulation cell was coupled to a proton bath in order to establish a constant pH in the system. After every tenth attempted move of the mobile charges, an attempt was made to delete/insert protons on the titrating groups. In reality, protonation of an acidic group means that acid has been added to the solution. Hence in the simulation, a protonation was balanced by the insertion of a negative mobile charge in order to maintain electroneutrality. The acceptance/rejection of an attempt to change the ionization state of a residue was based on the trial energy

$$\Delta U_{titra} = \Delta U_c \pm kT \ln 10(\text{pH} - \text{p}K_0) \tag{6}$$

where ΔU_c is the corresponding change in Coulomb energy, and pK_0 is the dissociation constant of the model compound. These values were taken from ref 40 and are given in Table 1. This method is accurate for weakly charged systems, whereas for really highly charged titrating objects, it has to be corrected for the excess chemical potential of the proton.

Perturbation Theory. In this section we present a perturbation approach for the interaction between a protein molecule at pH = pI and a charged molecule, i.e., in our case a polyelectrolyte. If the charges of a neutral protein are fixed, that is, the amino acid residues are not allowed to titrate, then the leading term in the perturbation expansion is the ion−dipole interaction; the polyelectrolyte net charge interacts with the dipole moment

TABLE 1: Titrating Residues in the Investigated Proteins[a]

protein	residues	Asp	Glu	His	Tyr	Lys	Cys[b]	Arg
pK_0[40]		4.0	4.4	6.3	9.6	10.4	10.8	12.0
lysozyme (2LZT)	129	7	2	1	3	6	0	11
α-lactalbumin (1HFY)	123	14	4	3	4	13	0	1
β-lactoglobulin (1BEB)	320	20	32	2	8	30	2	6
bovine serum albumin (1A06)	585	35	62	15	18	58	35	24
insulin (1APH)	41	0	4	2	4	1	6	1
calmodulin (1CLL)	142	16	19	1	2	5	0	6

[a] The PDB identities are reported together with the corresponding protein name. The dissociation constants for the isolated amino acids are given in the second line and the corresponding pK_0 for C- and N-termini are 3.8 and 7.5, respectively.[40] [b] Only cysteins not engaged in sulfide bridges can titrate.

of the protein. If the amino acid residues can titrate and change their charges as a function of protein−polyelectrolyte separation, then an additional induced charge−charge interaction will appear (this term is sometimes referred to as the "regulation term"). Both terms are identically zero in first order, but contribute attractively to the free energy in second order. The thermally averaged ion−dipole term varies as R^{-4}, while the ion-induced charge interaction is more long ranged and decays like R^{-2}. The relative importance of these two terms is the focus of the present work. Below follows a formal derivation of the terms.

Consider a protein described by a charge distribution $[\mathbf{r}_i,q_i]$, assuming its center of mass is placed at the origin, and the polyelectrolyte simply modeled by a point charge Q_α at $[\mathbf{R}]$. The electrostatic energy can then be written as

$$\beta U(R) = \sum_i \frac{l_B z_i Z_\alpha}{|\mathbf{R} - \mathbf{r}_i|}$$

$$R = |\mathbf{R}|$$

where we have used $q_i = ez_i$. From statistical mechanical perturbation theory[41,42] we can, provided that $\beta U(R)$ is small, write the interaction free energy as

$$\beta A(R) = -\ln\langle e^{-\beta U(R)}\rangle_0 \approx -\ln\left[1 - \langle \beta U(R)\rangle_0 + \frac{1}{2}\langle(\beta U(R))^2\rangle_0\right]$$

where $\langle ...\rangle_0$ denotes an average over all configurations in the unperturbed system. Assuming that $R \gg r_i$, we now perform a multipole expansion and obtain the final expression for the ion−ion, ion-induced charge, and ion−dipole interactions

$$\beta A(R) \approx l_B Z_\alpha \left(\frac{\langle Z\rangle_0}{R}\right) - {l_B}^2 {Z_\alpha}^2 \left(\frac{C}{2R^2} + \frac{\langle \mu\rangle_0^2}{6R^4}\right) \qquad (7)$$

where $\langle Z\rangle_0$ is the protein average charge number, $\langle \mu\rangle_0 = |\sum_i z_i \mathbf{r}_i|$ the average dipole moment number,and finally C is the charge fluctuations or the *protein charge capacitance*[43]

$$C \equiv \langle Z^2\rangle_0 - \langle Z\rangle_0^2 = -\frac{1}{\ln 10}\frac{\partial \langle Z\rangle_0}{\partial \mathrm{pH}}$$

The capacitance is an intrinsic property of a protein defining its ability for charge regulation. It is easily obtained as the derivative of the titration curve and is thus strongly pH-dependent. Since charge fluctuations are largest when pH = pK_a of a certain residue, the capacitance of a protein rich on, say, glutamic acid will peak at pH 4−5 (p$K_a^{glu} \approx 4.4$).

The leading term in eq 7 is the ion−ion interaction, but this vanishes at the isoelectric point where $\langle Z\rangle_0 = 0$. Left is the regulation and the ion−dipole terms—which one is more important depends of course on C and $\langle \mu\rangle_0$ but it is interesting to note that the charge regulation term is more long ranged than the ion−dipole interaction.

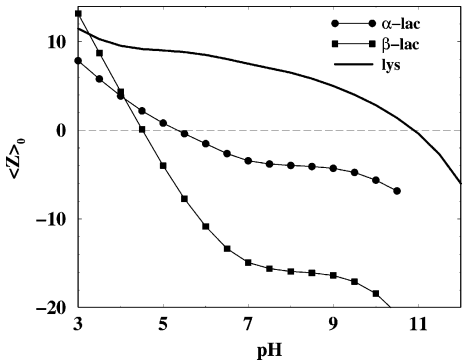


Figure 1. The simulated charge number of α-lactalbumin (spheres), *β*-lactoglobulin (squares), and lysozyme (no symbols). The salt concentration is 1 mM, and the protein concentration is 0.1 mM.

Equation 7 is valid in the limit of no salt. To approximately account for the effect of salt, the above expansion can be performed using a screened potential, where the ionic strength is expressed through the Debye length, $1/\kappa$. In the case of the ion−ion term, this brings about an additional factor of exp(−κR), while the ion-induced charge interaction is screened by exp(−2κR).

III. Results

The Isoelectric Point. As a starting point, we have simulated a single protein in salt solution varying the pH in order to determine the isoelectric point. Figure 1 shows the net charge of the three proteins as a function of pH. The corresponding pI values for α-lactalbumin, *β*-lactoglobulin, and lysozyme are 5.4, 4.5, and 10.9, respectively. These values have been obtained at low salt concentration. Addition of salt changes pI, and dimer or oligomer formation can also affect the isoelectric point. For example, the addition of 1 M salt to an α-lactalbumin solution decreases pI by approximately 0.4 units. Due to these facts, the experimental values for α-lactalbumin varies between 4.1 and 4.6, while for *β*-lactoglobulin the experimental data are slightly more scattered, between 4 and 5.5.

Figure 2 describes how the capacitance and the protein dipole moment vary with pH. The dipole moment is strictly well defined only at pI. For a nonneutral molecule the value of μ will depend on the coordinate origin—here we have used the center of mass. Both α-lactalbumin and *β*-lactoglobulin have large dipole moments over an extensive pH interval, while lysozyme has a comparatively small dipole moment. The capacitance for the three proteins varies significantly with pH, but at pI it is considerable for all three proteins. The smallest capacitance value is found for α-lactalbumin [$C_{\alpha-lac}$(pH = pI = 5.4) = 0.99], and it is related to the number of amino acid residues that titrate around pI. The relevant properties for the three proteins are collected in Table 2. Note, however, that the data for *β*-lactoglobulin is obtained for the dimer.[26,28]

Perturbation Calculations. Table 2 contains the basic physical data for the proteins used in this study. We now use this information to analytically calculate the ion-induced charge and ion−dipole contributions to the interaction free energy according to eq 7. The magnitude of the regulation and ion−dipole terms at contact are also given in Table 2. The results indicate that the regulation term is by far the most important term for lysozyme, while for α-lactalbumin and *β*-lactoglobulin

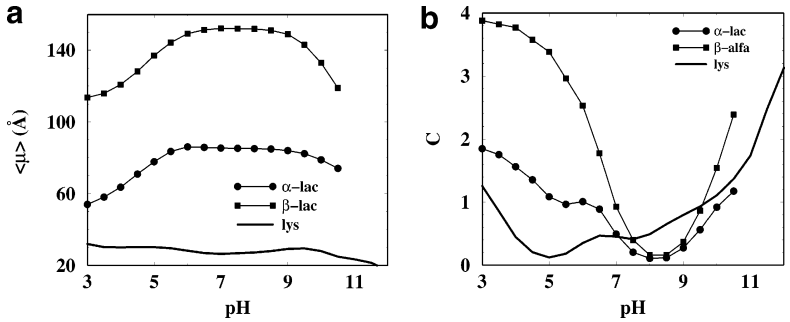


Figure 2. (a) The simulated dipole moment number, $\mu = \langle|\sum z_i \mathbf{r}_i|\rangle$, of α-lactalbumin (no symbols), β-lactoglobulin (spheres), and lysozyme (squares). The salt concentration is 1 mM, and the protein concentration is 0.1 mM. (b) Same as (a) but the protein capacitance, C.

TABLE 2: Charge Capacitance and Dipole Moment Number for the Investigated Proteins at Their Isoelectric Points[a]

	pI	C	μ	$R_p + R_{pe}$	$-\beta A_{reg}$	$-\beta A_{dip}$
lysozyme	10.9	1.7	24	58	5.7	0.2
α-lactalbumin	5.4	0.99	82	58	3.3	2.2
β-lactoglobulin	4.5	3.5	128	73	7.4	2.2
bovine serum albumin	5.5	3.2	297	81	5.5	7.7
insulin	5.4	0.36	49	51	1.6	1.3
calmodulin	4.0	3.7	51	58	12	0.9

[a] R_p is an estimate of the protein radius. The two last columns give the interaction between the protein and the polyelectrolyte at contact, that is $\beta A_{reg} = -l_B^2 Z_\alpha^2 C/2(R_p + R_{pe})^2$ and $\beta A_{dip} = -l_B^2 Z_\alpha^2 \mu^2/6(R_p + R_{pe})$,[4] where R_{pe} has been chosen as half the end-to-end separation of the corresponding neutral polymer (30 Å) and $Z_\alpha = -21$.

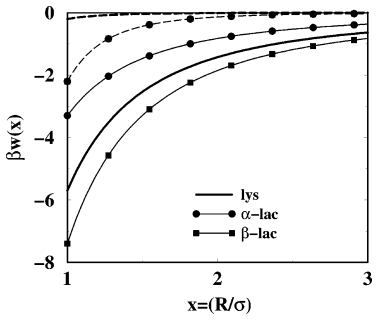


Figure 3. The contribution to the free energy of interaction from the charge−induced charge term (solid lines) and the ion−dipole term (dashed lines). Lines without symbols describe lysozyme, filled circles refer to α-lactalbumin, and filled squares refer to β-lactoglobulin, respectively. The free energies are calculated from eq 7 using simulated capacitances and dipole moments from Table *2* and $\sigma = R_p + R_{pe}$. Note that the ion−dipole terms for α-lactalbumin and β-lactoglobulin coincide.

the two terms are of comparable magnitude. The actual numbers in Table 2 should of course be regarded as qualitative and not quantitative. However, they still give, as will be seen below, a correct picture of the behavior of the three proteins. The contact separation has been defined as the protein radius plus the polyelectrolyte radius, $R_p + R_{pe}$. The latter has been chosen as half the end-to-end separation of the corresponding neutral ideal polymer. Both the protein and polyelectrolyte radii are approximate, but even with a rather generous variation of these values, the general picture of Table 2 will remain the same.

The regulation term decays slower than the ion−dipole term, which means that it will gain in relative importance at larger separation; see Figure 3. This means that even if the two terms are comparable at contact, the regulation term can still dominate the contribution to, for example, the second virial coefficient.

Monte Carlo Simulations. We have performed four different simulations for each protein: A, the "neutral" protein, that is all charges have been set to zero; B, the protein with fixed charges at each amino acid residue; C, the protein with an ideal dipole at its center of mass; D, the protein with titrating amino acid residues.

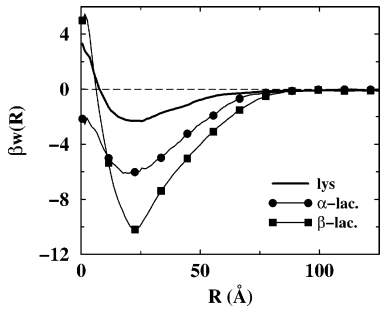


Figure 4. The potential of mean force between the centers of mass of the protein and the polyelectrolyte obtained from MC simulations with model D. The curves have been calculated at the respective isoelectric points for lysozyme (no symbols), α-lactalbumin (filled circles), and β-lactoglobulin (filled squares).

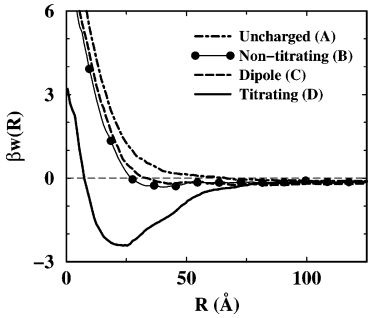


Figure 5. The potential of mean force between the centers of mass of lysozyme and the polyanion. The curves have been calculated at pI, and the four curves correspond to the different cases mentioned in the text.

The first set of simulations (A) describes only the shape of the protein, and the free energy of interaction is of course everywhere repulsive. These energy curves also give an indication of how difficult it is to deform the polyelectrolyte. The second set of simulations (B) uses fixed fractional charges on all residues, which has been determined in a separate simulation of the isolated protein at the appropriate pH. In the next set (C), the charge distribution of the protein is replaced by an ideal dipole; see Table 2. In the fourth and final set (D) the amino acids are allowed to titrate and this simulation contains all electrostatic contributions including the ion-induced charge term. The difference between set B and C describes the importance of higher order electrostatic moments, quadrupole, octupole, etc., in the protein, while a comparison of sets B and D will reveal the effect of the regulation mechanism.

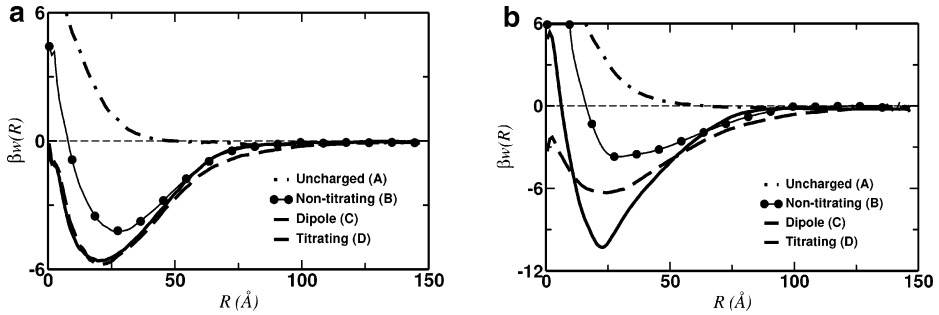


Figure 6. The potential of mean force between the centers of mass of (a) α-lactalbumin and (b) *β*-lactoglobulin and a polyanion, calculated at the respective pI. The four curves correspond to the different cases mentioned in the text.

The calculated potential of mean force, $w(R)$, for the three proteins at their respective pI all show a clear minimum; see Figure 4. The relative strength of the minima are in qualitative agreement with perturbation calculations, cf. Figure 3, while the actual numbers are approximately half the values predicted by second-order perturbation theory. A more quantitative comparison is difficult since the size of the polyelectrolyte is not unambiguously defined. The minima appear at roughly the same separation despite the fact that *β*-lactoglobulin is more than twice as big as the two others. This can be explained by the elongated form of the former, which also results in a more long ranged attraction. The separation R can approach zero, which corresponds to a situation where the polyelectrolyte wraps around the protein. Note, however, that $w(0)$ is repulsive indicating that the "wrapping" of the chain around the proteins is an entropically unfavorable structure.

β-Lactoglobulin forms a stronger complex with the polyanion than α-lactalbumin. Experimental results with whey proteins and gum arabicum by Weinbreck et al.[44] indicate a significant complex formation between the polyelectrolyte and *β*-lactoglobulin, which was the main component in the whey protein in their experiment. An experimental comparison of the relative strength of complex formation between α-lactalbumin and *β*-lactoglobulin at their respective pI is difficult, since the charge of gum arabicum varies in this pH interval. The attractive minimum in the protein-polyelectrolyte complex is reduced upon addition of salt,[14] and we can use the minima of $w(R)$ in Figure 4 in order to estimate the critical ionic strength. Assuming that the salt screening can be described by simple Debye−Hückel theory and that the complex can be defined as dissolved when the interaction is less than kT, we get the following relation

$$\exp(-2\kappa R_{\min})|\beta w(R_{\min})| \leq 1 \qquad (8)$$

The factor of 2 in the exponent comes from the fact that the second-order terms dominate the interaction. Following this recipe we find that approximately 10 and 20 mM salt is sufficient to dissociate the α-lactalbumin and *β*-lactoglobulin polymer complexes, respectively. This is in fair agreement with experiments,[9,14] but the actual numbers are of course dependent on the criterion in eq 8.

Thus, we have shown that a polyanion can form a complex with a neutral protein molecule. Next, we will make a numerically more rigorous partitioning of contributions to the potential of mean force shown in Figure 4. The minimum for lysozyme is solely due to charge regulation, Figure 5. If the charge distribution on lysozyme is considered fixed, then the polyanion−lysozyme interaction is essentially everywhere repulsive. Replacing the detailed charge distribution with an ideal dipole at the mass center has a small effect on the free energy. This means that the ion−dipole interaction gives a very small attractive contribution, while the effect from higher order moments is negligible.

As shown in Figure 6, the polyanion interacts more strongly with α-lactalbumin and *β*-lactoglobulin than with lysozyme. For α-lactalbumin the regulation term increases the depth of the minimum from approximately 4 to $6kT$. An interesting effect is that the dipolar protein shows a stronger interaction than the protein with a detailed but fixed charge distribution. This means that the ion−quadrupole interactions etc. add repulsive contributions to the interaction. The potential of mean force for *β*-lactoglobulin is even more attractive ($\approx 10kT$) in good agreement with the predictions based on the perturbation calculations.

IV. Conclusions

Strong protein−polyelectrolyte complexes can form due to purely electrostatic interactions even when the protein has a zero net charge at its isoelectric point. Two mechanisms contribute to the attractive interaction. One is the ion−dipole term and the other is a charge−induced charge interaction. The latter is often the dominating term and its strength is related to charge regulation of titrating amino acid residues in the protein. This ability can be quantified by the protein charge capacitance $C = \langle Z^2 \rangle - \langle Z \rangle^2$. The capacitance can be calculated from a simulation or measured in a titration experiment. Three different protein−polymer complexes have been investigated here by Monte Carlo simulations: lysozyme, α-lactalbumin, and *β*-lactoglobulin. The contribution from the charge−induced charge interaction to the free energy of interaction can be several kT s and is for the three cases studied here of the same magnitude or stronger than the ion−dipole interaction. We believe that this type of interaction is of importance also for the interaction between two proteins.

Acknowledgment. We thank Drs. Renko de Vries, Magnus Ullner, and Andrei Broukhno for stimulating discussions. FLBDS also acknowledges the CNPq and FAPESP/Brazil for the financial support during the development of this work.

Note Added after ASAP Publication. The following Note Added in Proof was submitted by the authors after this paper was published on the Web on 2/9/2006.

After publication of our paper on the Web, we were informed of the paper by P. M. Biesheuvel and M. A. Cohen Stuart, *Langmuir* **2004**, *20*, 2785, which, using mean-field theory, also suggests charge regulation as a relevant mechanism to describe polyelectrolyte−protein complexation at the wrong side of the isoelectric point.

This paper was reposted on 2/21/2006.

References and Notes

(1) Schmitt, C.; Sanchez, C.; Desobry-Banon, S.; Hardy, J. *Crit. Rev. Food Sci. Nutr.* **1998**, *38*, 689−753.
(2) Doublier, J. L.; Garnier, C.; Renard, D.; Sanchez, C. *Curr. Opin. Colloid Interface Sci.* **2000**, *5* (3, 4), 202−214.
(3) Zancong, S.; Mitragotri, S. *Pharm. Res.* **2002**, *19*, 391−395.
(4) Xia, J.; Dubin, P. In *Macromolecular Complexes in Chemistry and Biology*; Dubin, P., Bock, J., Davies, R. M., Schulz, D. N., Thies, C., Eds.; Springer-Verlag: Berlin, 1994.
(5) Jiang, G.; Woo, B. H.; Kangb, F.; Singhb, J.; DeLuca, P. P. *J. Controlled Release* **2002**, *79*, 137−145.
(6) Simon, M.; Wittmar, M.; Bakowsky, U.; Kissel, T. *Bioconjugate Chem.* **2004**, *15*, 841−849.
(7) Hubbell, J. A. *Science* **2003**, *300*, 595−596.
(8) Girard, M.; Turgeon, S. L.; Gauthier, S. F. *J. Agric. Food Chem.* **2003**, *51*, 6043−6049.
(9) de Kruif, C. G.; Weinbreck, F.; de Vries, R. *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 340−349.
(10) Grymonpré, K. R.; Staggemeier, B. A.; Dubin, P. L.; Mattison, K. W. *Biomacromolecules* **2001**, *2*, 422−429.
(11) Hattori, T.; Hallberg, R.; Dubin, P. L. *Langmuir* **2000**, *16*, 9738−9743.
(12) Seyrek, E.; Dubin, P. L.; Tribet, C.; Gamble, E. A. *Biomacromolecules* **2003**, *4*, 273−282.
(13) Hallberg, R.; Dubin, P. L. *J. Phys. Chem. B* **1998**, *102*, 8629−8633.
(14) de Vries, R. *J. Chem. Phys.* **2004**, *120* (7), 3475−3481.
(15) Carlsson, F.; Linse, P.; Malmsten, M. *J. Phys. Chem. B* **2001**, *105*, 9040−9049.
(16) de Vries, R.; Weinbreck, F.; de Kruif, C. G. *J. Chem. Phys.* **2003**, *118* (10), 4649−4659.
(17) Park, J. M.; Muhoberac, B. B.; Dubin, P. L.; Xia, J. *Macromolecules* **1992**, *25*, 290−295.
(18) Regnier, F. E. *Science* **1987**, *238*, 319−323.
(19) Kirkwood, J. G.; Shumaker, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **1952**, *38*, 863−871.
(20) Lund, M.; Jönsson, B. *Biophys. J.* **2003**, *85*, 2940−2947.
(21) Humphrey, B. D.; Huang, N.; Klasing, K. C. *J. Nutr.* **2002**, *132*, 1214−1218.
(22) Proctor, V. A.; Cunningham, F. E. *CRC Crit. Rev. Food Nutr.* **1988**, *26* (4), 359−3958.
(23) Fast, J.; Mossberg, A.-K.; Svanborg, C.; Linse, S. *Protein Sci.* **2005**, *14*, 329−340.
(24) Svensson, M.; Hakansson, A.; Mossberg, A.-K.; Linse, S.; Svanborg, C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (8), 4221−4226.
(25) Kuwajima, K. *FASEB J.* **1996**, *10*, 102−109.
(26) Gottschalk, M.; Nilsson, H.; Roos, H.; Halle, B. *Protein Sci.* **2003**, *12*, 2404−2411.
(27) Oliveira, K. M. G.; Valente-Mesquita, V. L.; Botelho, M. M.; Sawyer, M. L.; Ferreira, S. T.; Polikarpov, I. *Eur. J. Biochem.* **2001**, *268*, 477.
(28) Fogolari, F.; Ragona, L.; Licciardi, S.; Romagnoli, S.; Michelutti, R.; Ugnolini, R.; Molilari, H. *Proteins*: *Struct., Funct., Genet.* **2000**, *39*, 317−330.
(29) Motrich, R. D.; Gotteroa, C.; Rezzonico, C., Jr.; Rieraa, C. M.; Rivero, V. *Clin. Immunol.* **2003**, *109*, 203−211.
(30) Pellegrini, A.; Engels, M. *Curr. Med. Chem.*: *Anti-Infect. Agents* **2005**, *4*, 55−66.
(31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(32) Protein data bank. http://www.rcsb.org/pdb, 2005.
(33) Kesvatera, T.; Jönsson, B.; Thulin, E.; Linse, S. *Proteins: Struct., Func., Genet.* **1999**, *37*, 106−115.
(34) Hill, T. L. *Statistical Mechanics*; McGraw-Hill: New York, 1956.
(35) Marcus, R. A. *J. Chem. Phys.* **1955**, *23*, 1057.
(36) Jönsson, B. The Thermodynamics of Ionic Amphiphile−Water Systems−A Theoretical Analysis. Ph.D. Thesis, Lund University, Lund, Sweden, 1981.
(37) Svensson, B.; Jönsson, B.; Thulin, E.; Woodward, C. *Biochemistry* **1993**, *32*, 2828−2834.
(38) Levesque, D.; Weis, J. J.; Hansen, J. P. In *Monte Carlo Methods in Statistical Physics*; Binder, K., Ed.; Springer-Verlag: Berlin, 1986; Vol. 5, pp 47−119.
(39) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*: *From Algorithms to Applications*; Academic Press: San Diego, CA, 1996.
(40) Nozaki, Y.; Tanford, C. *Methods Enzymol.* **1967**, *11*, 715−734.
(41) McQuarrie, D. A. *Statistical Mechanics*; Harper Collins: New York, 1976.
(42) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: Oxford, 2001.
(43) Lund, M.; Jönsson, B. *Biochemistry* **2005**, *44* (15), 5722−5727.
(44) Weinbreck, F.; de Vries, R.; Schrooyen, P.; de Kruif, C. G. *Biomacromolecules* **2003**, *4*, 293−303.

Implications of a high dielectric constant in proteins

Mikael Lund* and Bo Jönsson
Theoretical Chemistry, Chemical Center, Lund University
POB 124, S-22100 Lund, SWEDEN.
Phone: +46-46-222 0381 / Fax: +46-46-222 4543
Email: mikael.lund@teokem.lu.se

Cliff E. Woodward
School of Chemistry, University College, University of New South Wales
Australian Defence Force Academy, Canberra ACT 2600, AUSTRALIA.

Abstract

Solvation of protein surface charges plays an important role for the protonation states of titratable surface groups and is routinely incorporated in low dielectric protein models using surface accessible areas. For many-body protein simulations, however, such dielectric boundary methods are rarely tractable and a higher level of coarse graining is desirable. We here propose the inverse scheme and scrutinize how charges on a high dielectric surface are affected by the non-polar protein interior. A simple dielectric model combined with explicit ion Monte Carlo simulations show that for small, hydrophilic proteins this effect is mostly negligible, suggesting that the protein (solution) can be approximated with a uniform high dielectric constant. This conception is verified by estimates for titration curves and acidity constants for four different proteins (BPTI, calbindin D_{9k}, ribonuclease A, and turkey ovomucoid third domain) that all correlate well with experimental titration data.

Keywords: Protein electrostatics, Monte Carlo Simulation, Coarse Graining, Dissociation Constant, Charge Capacitance, Uniform dielectric.

Introduction

In the past century protein ionization has caught the attention of many researchers and a wealth of experimental data is available, especially potentiometric titration curves and NMR determined pK_a values. Theoreticians have not neglected the subject either and numerous protein solution models have been proposed for the calculation of pK_a values. These range from simple, spherical models as that of Tanford and Kirkwood[1–3] (TK), numeric Poisson-Boltzmann (PB) and generalized Born calculations[4–9] as well as explicit solvent models.[10,11]

In the original TK model[1–3] titratable charged groups were assumed to be buried within a spherical protein with a dielectric constant much less than that of the surrounding (aqueous) solvent. Modern approaches utilizing electrostatic continuum models, can be viewed as generalizations of the original TK model. Two important questions arise: (i) *What is the dielectric constant of the protein?* and (ii) *Where is the dielectric interface between protein and solvent to be located?* The interior of a protein may have a higher than expected dielectric constant due to polarization contributions from the configurational freedom of polar side chains[12–14] as well

as proton fluctuations in titratable groups.[12,15] Many values for the interior dielectric response of proteins have been reported, ranging from 2 to 40, approximately. We will not review this whole branch of biophysics but merely note that there is little agreement about this quantity and, in macroscopic models it may ultimately be best treated as an adjustable parameter[16] matched to experimental results.

The majority of charges in globular proteins are located "close" to the protein solution interface (Fig. 1). These surface groups interact directly with the surrounding solvent, as evidenced by the fact that proteins dissolve in water through solvation of polar surface groups and that protons exchange between side chains and the solvent. This means that the position of the interface between a protein and the aqueous solvent is ambiguous. This fact led Tanford to concede some thirty years ago[17] *"...that the Kirkwood model is not appropriate for proteins because the ionic groups often extend into the surrounding solvent sufficiently far so that the medium between them is pure solvent and the effect of the large dielectric cavity several Ångstroms away is minimal."* In connection with this, it is pertinent to note, that a number of researchers[6,18–22] have obtained good agreement with experimental titration results using a *uniform* dielectric constant model, i.e., where the internal dielectric constant of the protein is set to that of the surrounding solvent. That such simple model can have predictive success may seem puzzling, but it is worthwhile to note that todays descendants of the original (low dielectric) TK approach are not entirely different from the uniform dielectric model: In recognition of surface charge solvation most models incorporate surface accessible areas, rendering them much more successful than what was previously reported by Tanford.

In this paper we will turn the problem upside down: Instead of consider solvent effects on charges in a low dielectric, we will study interior perturbations of charges in a high dielectric. For this purpose we propose a slight modification of the TK model by imbedding a low dielectric sphere within a static protein model (Fig. 2). This sphere will mimic the expected low dielectric deep interior of the protein. In the region between the sphere and the protein surface, we shall assume that the dielectric constant is equal to that of the solvent. This is in recognition of the ability of the solvent to permeate this region to solvate charged residues and, further, that the residues themselves will increase the local polarization of this region. We shall use this model to calculate theoretical protein titration curves and make a comparison with a range of experimental results. Though crude, this model should give us some insight into the role played by the low dielectric interior in determining titration behavior in proteins.

Model and theory

X-ray or nuclear magnetic resonance determined protein structures are represented by either:(a) a collection of spheres (diameter, σ=4 Å) each representing an atom or (b) spheres representing amino acid residues located at their center-of-mass and with a diameter, σ, determined from their molecular weight, M_w, according to the formula,

$$\sigma/2 = \left(\frac{3M_w}{4\pi\rho N_{Av}}\right)^{1/3} \qquad \rho = 1 \text{ g/cm}^3. \tag{1}$$

The latter (coarse grained) model reduces the number of particles from the order of thousands to hundreds and can be useful when studying several protein molecules.[23–25] The non-polar deep interior is described by a spherical cavity with radius a and dielectric constant ϵ_p, centered at the protein center of mass. The solvent is treated as a dielectric continuum with dielectric constant, ϵ_s, while counter ions and salt are explicitly treated as mobile, charged hard spheres (σ=4 Å). The canonical ensemble distribution is sampled using the Metropolis Monte Carlo

(MC) algorithm[26] where the total energy is calculated as a sum of pair interactions,

$$U = \sum_{i,j} \left(u^{el}(r_{ij}) + u^{hs}(r_{ij}) \right) + \sum_i u^{rf}(r_i) \qquad i \neq j \tag{2}$$

$$u^{hs}(r_{ij}) = \begin{cases} 0 & r_{ij} > (\sigma_i + \sigma_j)/2 \\ \infty & \text{otherwise} \end{cases} \tag{3}$$

The electrostatic part of the pair potential, u^{el}, is found by identifying the appropriate boundary conditions and solving the Poisson and Laplace equations for a spherical discontinuity, yielding:[27,28]

$$u^{el}(r_{ij}) = \begin{cases} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_s} \left(\frac{1}{r_{ij}} + \sum_{n=1}^{\infty} \frac{r_j^n}{r_i^{n+1}} \Gamma_1 \right) & r_i > a > r_j \\ \frac{q_i q_j}{4\pi\epsilon_0\epsilon_s} \left(\frac{1}{r_{ij}} + \sum_{n=1}^{\infty} \frac{a^{2n+1}}{(r_i r_j)^{n+1}} \Gamma_1 \right) & r_i > a \wedge r_j > a \\ \frac{q_i q_j}{4\pi\epsilon_0\epsilon_p} \left(\frac{1}{r_{ij}} + \sum_{n=0}^{\infty} \frac{(r_i r_j)^n}{a^{2n+1}} \Gamma_2 \right) & r_i < a \wedge r_j < a \end{cases} \tag{4}$$

with $\Gamma_1 = \frac{(\epsilon_s - \epsilon_p) P_n(\cos\theta)}{\epsilon_p + \epsilon_s (n+1)/n}$ and $\Gamma_2 = \frac{(\epsilon_p - \epsilon_s) P_n(\cos\theta)}{\epsilon_s - \epsilon_p n/(n+1)}$. P_n is the Legendre polynomial evaluated for the angle, θ between $\mathbf{r}_i$ and $\mathbf{r}_j$. In the case of a homogeneous dielectric medium ($\epsilon_p = \epsilon_s$) Eq. 4 reduces to the plain Coulomb potential and for $i = j$ the second term contributes to the self energy,

$$u^{rf}(r) = \frac{1}{2} q \phi^{rf}(r) = \frac{q^2}{8\pi\epsilon_0\epsilon_s} \sum_{n=1}^{\infty} \frac{\Gamma_1}{r} \left(\frac{a}{r} \right)^{2n+1} \tag{5}$$

here shown for a charge outside the spherical cavity.

Titration is accomplished by exchanging protons (+1 charges) between random sites in the protein and the solvent which brings about an additional energy term:[15]

$$\Delta U = \Delta U^{el} \pm (\mathrm{pH} - \mathrm{pK}_a) kT \ln 10 \tag{6}$$

where kT is the thermal energy, ΔU^{el} is the change in electrostatic energy, pK_a is the site's intrinsic acidity constant, (+) applies for protonation and (-) for deprotonation. The first term in Eq. 6 is purely electrostatic while $(\mathrm{pH} - \mathrm{pK}_a) kT \ln 10$ accounts for "chemical" interactions, as measured for small model compounds (pK_a: Ctr 3.8, Asp 4.0, Glu 4.4, His 6.3, Ntr 7.5, Tyr 9.6, Lys 10.4, Cys 10.8, Arg 12.0) – more on this in the next section.

We have also investigated excess chemical potentials, μ_{ex} for ion pairs in the solution using the Widom particle insertion method.[29] Here a cat- and an anion ($\sigma = 4$ Å) are simultaneously inserted at random positions in the cell and μ_{ex} can be sampled according to

$$\mu_{\mathrm{ex}} = -kT \ln \left\langle e^{-\Delta U / kT} \right\rangle_0 , \tag{7}$$

where ΔU is the total energy change for the process evaluated according to Eq. 2

Results and discussion

Eq. 6 embodies an approach often employed for protein titration calculations. The term, $pH - pK_a$, includes bonding and (de)solvation components and obtaining it from first principles would require quantum calculations and more complex simulations. Instead, we will adopt the approach of obtaining the $\mathrm{p}K_a$ values from titration data on small amino acid *model* compounds. Thus, ΔU^{el}, accounts for the energy change that occurs when a proton is transferred from the model compound to the corresponding site in the protein. As in other work, we will assume that ΔU^{el} contains only electrostatic contributions, i.e., we assume that non-electrostatic contributions to proton binding are the same in the protein and the model compound. In order

to calculate, ΔU^{el}, one needs to properly model the electrostatic environment of protons in the model compound as well as that of the protein. In this work, we shall assume that the low dielectric region of the model compounds are small enough to be neglected. Furthermore, we shall assume that the titratable sites on the protein lie *outside* the low dielectric sphere, modelling the interior of the protein. The validity of these assumptions will eventually be verified upon comparison with experiment.

The quantity ΔU^{el} can itself be separated into one- and two-body components. The one-body contribution is the so-called *generalized* Born energy, ΔU^{Born}, and is due to the interaction of a proton with its own reaction potential arising from the polarization of the surrounding dielectric continuum. If the proton were immersed in a uniform dielectric continuum with dielectric constant ϵ_s, one would obtain the usual Born expression,

$$U_{\infty}^{Born} \propto -(1 - 1/\epsilon_s)/\sigma_B \tag{8}$$

where the proton is assigned a Born diameter, σ_B. The generalized Born energy is a sum of this term and a contribution due to the polarization of dielectric boundaries. In our model, the proton is assumed to be located in the solvent region of the dielectric medium in both the protein and the model compound. This leads to a cancellation of the terms of the type in Eq.(8). Thus, for a proton removed from the model compound and placed in the protein at a distance r_0 from the center of the low dielectric sphere ($r_0 > a$),

$$\Delta U^{Born} = u^{rf}(r_0) \tag{9}$$

where u^{rf} is the reaction potential defined by Eq.(5). We note that ΔU^{Born}, is generally positive, i.e., it disfavours the transfer of charge to the protein, due to the presence of the low dielectric region, which is assumed absent in the model compound. A charge placed near the dielectric interface will be subject to an unfavourable reaction potential. The magnitude of this effect is illustrated in Fig. 4 where we have used MC simulations to calculate the excess chemical potential, $\mu_{\rm ex}$, for an ion outside a low dielectric sphere. We have investigated the dependence on the radius of the sphere and the external salt concentration. Firstly, we note that the larger is the radius of the low dielectric region, the more unfavourable the reaction potential becomes. This is due to a greater amount of lower dielectric material being closer to the ion at a given distance from the surface. Since the reaction field is relatively short ranged we observe essentially no screening at physiological salt concentrations. The effect of the low dielectric region on the Born energy is not large. With reasonable choices for the low dielectric sphere and salt concentrations, $\mu_{\rm ex}$ (stemming from the Born energy) is only of the order of a few tenths of a kT. Furthermore, we expect that our approximations will overestimate the penalty of charging the protein, as we have assumed that the dielectric environment of the model compounds can be replaced by a uniform dielectric continuum modelling water.

We also note, that the same approximations, as embodied in Eq.(9), but applied to a model whereby the proton is transferred into the low dielectric sphere of the protein instead ($r_0 < a$), would give rise to a negative Born energy, and thus *favours* the creation of charge in the protein. This apparent paradox arises, in this case, as the model compound would now be assumed to be equivalent to an infinite low dielectric continuum. This is obviously unphysical.

The two-body contribution to ΔU is due to the electrostatic interaction between the titrating proton and all the other charges present on the protein. This interaction is mediated by the intervening dielectric material and any dielectric boundaries present. The presence of the boundaries can have some unexpected consequences. To illustrate this, we present in Fig 3, the interaction between two like charges close to the spherical dielectric boundary in the protein. One may expect that the presence of the low dielectric region would have a tendency to enhance the electrostatic interaction between charges. However, our results show that, even when the

charges are only a few Ångstroms from the surface of the low dielectric sphere, there is little change in their interaction, whether the sphere is present or not. The electric field lines tend to avoid the low dielectric region and the interaction is well approximated by that in a uniform solvent. By looking at the *difference* between the ion pair interaction in the uniform and non-uniform dielectric model, we are better able to see the effect of the dielectric boundary, Fig.3 (b). For example, for charges just outside the spherical surface, the pair interaction is more repulsive than the direct Coulomb potential at small angles, but then becomes slightly more attractive at large angles. This can be understood by realising that the effect of the boundary on the potential set up by a given charge, can be represented by induced multi-poles, which are unfavourably aligned with the charge. Considering just the induced dipole component, when a second charge is brought into the vicinity of the first charge and the induced dipole, it is repelled by both at small angles and repelled by the charge and attracted by the dipole at large angles (greater than 90). Though truncation at the dipole term is a poor approximation, addition of the higher order induced multipoles does not alter the qualitative picture.

Changing the radius, a, of the low dielectric region does have a significant influence on the pair interaction. If a is decreased the effect of the induced multipoles is diminished. On the other hand, the discontinuity has a much larger effect when the interacting charges become immersed in the low dielectric region. When the charges are just below the surface ($a = r_i + 0.1$) the small angle interactions are *more screened* than in the case of the uniform, high-dielectric description. This counter-intuitive result is caused by a strong polarization of the high dielectric region. The multipoles induced by a charge are attractive to a second charge at small angles and repulsive at large angles. When the charges are even further below the surface of the low dielectric sphere the interaction is much more repulsive than that without the sphere. These results demonstrate how sensitive the interaction between charges are to the choice of the position of the dielectric boundary.[30]

Using our model, we have investigated the titration behaviour of four different proteins: calbindin, bovine pancreas trypsin inhibitor (BPTI), ribonuclease A, and turkey ovomucoid third domain. We used MC simulations to calculate the titration curve for ribonuclease A. In this case we have assumed that the dielectric constant was uniform throughout the solution and the protein, with a value, $\epsilon_p = \epsilon_s$=80. The results are given in Fig. 5. Good agreement is found between the computed and measured titration curve, despite the absence of a low dielectric protein interior. We found that the calculated results remained equally good, even in the presence of a low dielectric spherical region, provided the it did not include ionic groups. The calculated results differ significantly from the ideal or "null" curve, which does not include electrostatic interactions. This is an important observation and hints that the underlying physical mechanisms captured by our model are probably correct.

For turkey ovomucoid third domain and BPTI we have calculated individual pK_a values (Table 1 and 2) and compared with measurements and theoretical predictions based on the PBE[5] and a modified Tanford-Kirkwood (MTK) approach.[31] Again we have assumed a uniform high dielectric constant ($\epsilon_p = \epsilon_s = 80$). In both cases the model agrees well with experimental data and is very competitive with the alternative models (see also Khandogin and Brooks[32] work on ovomucoid), which both assume low dielectric interiors for the protein.

It is worth noticing that the MTK calculations are significantly improved by introducing either a higher dielectric constant (reported up to 20), or letting the flexible side chains relax in the electric field. As is discussed by Havranek and Harbury[31] and others,[16,33] buried polar groups may require a more subtle description.

As a final example, Fig. 6 shows measured[22,34] and our calculated titration curves for calbindin D_{9k}. In this case, we have incorporated a low dielectric spherical interior in the protein, The radius of the interior was made as large as possible, while still keeping the surface protein charges external to it. We observed no notable deviations from results obtained using

a uniform dielectric constant (not shown in graph). The agreement with the experimental data is good, with the electrostatic interactions tending to flatten the titration curve, compared to the ideal model. This effect is slightly overestimated by the simulations.

The decomposition of intrinsic pK_a-values is not trivial and it is of interest to analyze how minor shifts in these values will influence the titration curve. This can be conveniently estimated using the *protein charge capacitance*[15] which is a molecular measure of how the protonation state responds to an external electric potential, ϕ_{ext}:

$$C = \left\langle Z^2 \right\rangle - \langle Z \rangle^2 = -\frac{\partial Z}{\beta e \partial \phi_{ext}} = -\frac{1}{\ln 10}\frac{\partial Z}{\partial \mathrm{pH}} \tag{10}$$

where Z is the total protein charge number. As evident from Eq. 6 shifting the pK_a values is equivalent to applying an external potential. At pH 4 the capacitance for calbindin is approximately 2 (from the derivative of the titration curve) and shifting the intrinsic pK_a values by 0.4 units will cause the protein to bind two more protons ($\Delta Z \approx -C \ln 10 \Delta pK_a$). If at a pH where C is small, much larger shifts are required to significantly influence the titration curve. Generally, electrostatic interactions tend to lower and broaden the capacitance peaks, thus making the model calculations less sensitive to the choice of intrinsic parameters. For example, the ideal capacitance for calbindin at pH 4 is more than twice as large as in the real protein. As for BPTI and ovomucoid third domain, the root-mean-square value (rms) was found to be 0.4-0.5 pK units (Table 1 and 2) indicating that better agreement with experiment may be obtained by slightly changing the intrinsic pK_a values. At the moment we do not offer any controlled way of effecting this, we merely note this as an operational possibility.

Conclusions

We have investigated the effect of incorporating a low dielectric sphere to model protein interiors in simulations. The model highlights some counterintuitive behaviours, for example, the interaction between a pair of charges is not always enhanced by placing them in the vicinity of the low dielectric core. This is due to the orientation of the induced multi-poles, relative to the charges. This suggests that the placement of the dielectric boundaries in a protein/solvent system is an important consideration, when attempting to model phenomena such as protein titration. Buried ionic groups may thus prove particularly problematic in both uniform dielectric models as well as in those containing a dielectric boundary.

This notwithstanding, we have found that an electrostatic model, which assumes a uniform value for the dielectric constant (equal to that of water) is able to accurately predict both site and overall titration behaviour for a range of small proteins. This is invaluable for many-body protein simulations (protein aggregation etc.) where dielectric boundaries pose a difficult computational problem. Further, since salt particles are treated explicitly the model is capable of describing multivalent ions and/or high salt concentrations.

Incorporating a low dielectric sphere to model the interior of a protein has little effect on the results, provided the sphere does not include surface charges. In the light of these findings and from the fact that molecular continuum models are idealized approximations in the first place we suggest that the straightforward strategy of assuming a uniform dielectric model seems to provide an accurate an simple model for protein electrostatics.

Acknowledgments

We wish to thank Professor Bengt Jönsson, Department of Biophysical Chemistry, Lund University for useful discussions and, for financial support: The Solander Program, The Swedish

Foundation for Strategic Research and The Research School in Pharmaceutical Sciences, Sweden.

References

1. Kirkwood, J. G. (1934) Theory of solutions of molecules containing widely separated charges with special application to zwitterions *J. Chem. Phys. 2,* 351–361.

2. Tanford, C., Kirkwood, J. G. (1957) Theory of protein titration curves. i. general equations for impenetrable spheres *J. Am. Chem. Soc. 79,* 5333.

3. da Silva, F. L. B., Jönsson, B., Penfold, R. (2001) A critial investigation of the tanford-kirkwood scheme by means of monte carlo simulations *Protein Science 10,* 1415–1425.

4. Warwicker, J., Watson, H. C. (1982) Calculation of the electric potential in the active site cleft due to alpha-helix dipoles *J. Mol. Biol. 157,* 671–679.

5. Antosiewicz, J., Briggs, J. M., Elcock, A. H., Gibson, M. K., McGammon, J. A. (1996) Computing ionization states of proteins with a detailed charge model *J. Comp. Chem. 17,* 1633-1644.

6. Demchuk, E., Wade, C. (1996) Improving the continuum dielectric approach to calculating pk_a's of ionizable groups in proteins *J. Phys. Chem. 100,* 17373–17387.

7. Georgescu, E. R., Alexov, E. G., Gunner, M. R. (2002) Combining conformational flexibility and continuum electrostatics for calculating pka's in proteins *Biophys J. 83,* 1731–1748.

8. Warwicker, J. (2004) Improved pk_a calculations through flexibility based sampling of a water-dominated interaction scheme *Protein Science 13,* 2793–2805.

9. Feig, M., III, C. L. B. (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations *Current Opinion in Structural Biology 14,* 217–224.

10. Simonson, T., Carlsson, J., Case, D. A. (2004) Proton binding to proteins: pk_a calculations with explicit and implicit solvent models *J. Am. Chem. Soc. 126,* 4167–4180.

11. Tan, C., Yang, L., Luo, R. (2006) How well does poisson-boltzmann implicit solvent agree with explicit solvent? a quantitative analysis *J. Phys. Chem. 110,* 18680–18687.

12. Kirkwood, J. G., Shumaker, J. B. (1952) Forces between protein molecules in solution arising from fluctuations in proton charge and configuration *Proceedings of the National Academy of Sciences 38,* 863–871.

13. Smith, P. E., Brunne, R. M., Mark, A. E., van Gunsteren, W. F. (1993) Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations *J. Phys. Chem. 97,* 2009–2014.

14. Alexov, E. G., Gunner, M. R. (1997) Incorporating protein conformational flexibility into the calculation of ph-dependent protein properties *Biophys J. 5,* 2075–2093.

15. Lund, M., Jönsson, B. (2005) On the charge regulation of proteins *Biochemistry 44,* 5722–5727.

16. Schutz, C. N., Warshel, A. (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models *Proteins: Structure, Function, and Genetics 44,* 400-417.

17. Tanford, C., Roxby, R. (1972) Interpretation of protein titration curves. application to lysozyme. *Biochemistry 11,* 2192–2198.

18. Warshel, A., Russel, S. T., Churg, A. K. (1984) Macroscopic models for studies of electrostatic interactions in proteins: Limitations and applicability *Proc. Natl. Acad. Sci. USA 81,* 4785–4789.

19. Kesvatera, T., Jönsson, B., Thulin, E., Linse, S. (1994) Binding of ca2+ to calbindin d9k: Structural stability and function at high salt concentration *Biochemistry 33,* 14170–14176.

20. Kesvatera, T., Jönsson, B., Thulin, E., Linse, S. (1996) Measurement and modelling of sequence-specific pK_a values of calbindin D_{9k} *J. Mol. Biol. 259,* 828.

21. Penfold, R., Warwicker, J., Jönsson, B. (1998) Electrostatic models for calcium binding proteins *J. Phys. Chem. B 102,* 8599–8610.

22. Kesvatera, T., Jönsson, B., Thulin, E., Linse, S. (1999) Ionization behaviour of acidic residues in calbindin d_{9k} *Proteins 37,* 106–115.

23. Carlsson, F., Malmsten, M., Linse, P. (2001) Monte carlo simulations of lysozyme self-association in aqueous solution *J. Phys. Chem. 105,* 12189-12195.

24. Lund, M., Jönsson, B. (2003) A mesoscopic model for protein-protein interactions in solution *Biophys. J. 85,* 2940–2947.

25. McGuffee, S., Elcock, A. (2006) Atomically detailed simulations of concentrated protein solutions: The effects of salt, ph, point mutations, and protein concentration in simulations of 1000-molecule systems *Journal of the American Chemical Society 128,* 12098-12110.

26. Metropolis, N. A., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A., Teller, E. (1953) Equation of state calculations by fast computing machines *J. Chem. Phys. 21,* 1087–1097.

27. Böttcher, C. J. F. *Theory of Electric Polarization;* Elsevier: Amsterdam, 1973.

28. Woodward, C. E., Svensson, B. R. (1991) Potentials of mean force in charged systems: Application to superoxide dismutase *J. Phys. Chem. 95,* 7471–7477.

29. Widom, B. (1963) Some topics in the theory of fluids *J. Chem. Phys. 39,* 2808–2812.

30. Scarsi, M., Apostolakis, J., Caflisch, A. (1997) Continuum electrostatic energies of macromolecules in aqueous solutions *J. Phys. Chem. 101,* 8098–8106.

31. Havranek, J. J., Harbury, P. B. (1999) Tanford-kirkwood electrostatics for protein modeling *Proc. Natl. Acad. Sci. 96,* 11145-11150.

32. Khandogin, J., Brooks, C. L. r. (2005) Constant ph molecular dynamics with proton tautomerism. *Biophys J 89,* 141-57.

33. Fitch, C. A., Karp, D. A., Lee, K. K., Stites, W. E., Lattman, E. E., E., B. G.-M. (2002) Experimental pk_a values of buried residues: Analysis with continuum methods and role of water penetration *Biophys J. 82,* 3289–3304.

34. Kesvatera, T., Jönsson, B., Thulin, E., Linse, S. (2001) Focusing of the electrostatic potential at ef-hands of calbindin d_{9k}. titration of acidic residues. *Proteins 45,* 129–135.

35. Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R. (1983) The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors *Acta Crystallogr., Sect. B 39,* 480.

36. Richarz, R., Wüthrich, K. (1978) High-field ^{13}c nuclear magnetic resonance studies at 90.5 mhz of the basic pancreatic trypsin inhibitor *Biochemistry 17,* 2263–2269.

37. Wuthrich, K., Wagner, G. (1979) Nuclear magnetic resonance of labile protons in the basic pancreatic trypsin inhibitor *J. Mol. Biol. 130,* 1–18.

38. Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J., Neumann, S. (1986) X-ray crystal structure of the complex of human leukocyte elastase (pmn elestase) and the third domain of the turkey ovomucoid inhibitor *EMBO 5,* 2453.

39. Shaller, W., Robertson, A. D. (1995) ph, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid thirs domain *Biochemistry 34,* 4714–4723.

40. Tanford, C., Hauenstein, J. D. (1956) Hydrogen ion equilibria of ribonuclease *J. Am. Chem. Soc. 78,* 5287–5291.

41. Leonidas, D. D., Shapiro, R., Irons, L., Russo, N., Acharya, K. R. (1997) Crystal structures of ribonuclease a complexes with 5'-diphosphoadenisine 3'-phosphate and 5'diphosphoadenosine 2'-phosphate at 1.7 a resolution *Biochemistry 36,* 5578.

42. Szebenyi, D. M. E., Moffat, K. (1986) The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. molecular details, ion binding, and implications for the structure of other calcium-binding proteins *J. Biol. Chem. 261,* 8761–8777.

Table 1: Measured (∼25-50 mM salt) and calculated pK_a values for BPTI. The MC (35 mM salt) and PB (150 mM salt) results are based on the crystal structure[35] (4PTI). The MC calculation was performed using an all atom description and a uniform dielectric response ($\epsilon_p = \epsilon_s = 80$).

	Ideal	PB[5]	MC	Exp.[36,37]
Asp3	4.0	3.3	2.8	3.0
Asp50	4.0	2.5	2.5	3.4
Glu7	4.4	4.7	2.9	3.7
Glu49	4.4	3.5	3.8	3.8
Lys15	10.4	10.4	10.3	10.6
Lys26	10.4	10.4	10.7	10.6
Lys41	10.4	10.3	11.1	10.8
Lys46	10.4	10.0	10.4	10.6
Ctr	3.8	3.8	2.7	2.9
Ntr	7.5	7.2	7.3	8.1
rms	0.6	0.7	0.5	

Figure 1. Distances (Å) to the center-of-charge of: titrateable sites (circles) and the protein surface closest to that site (line). Plotted for calbindin (left) and ribonuclease (right) using an all atom description based on the crystal structures.

Table 2: Measured and calculated pK_a values for turkey ovomucoid third domain at 10 mM 1:1 salt concentrations. The MC and MTK calculations are both based on the crystal structure[38] (1PPF). In the simulations, the amino acid model was applied.

	Ideal	MTK[31]		MC	Exp.[39]
		ϵ_p=4, relaxed	ϵ_p=20	ϵ_p=80	
Asp7	4.0	2.1	3.0	3.2	< 2.6
Glu10	4.4	4.0	3.5	3.9	4.1
Glu19	4.4	3.1	2.7	3.4	3.2
Asp27	4.0	2.9	3.7	2.7	< 2.3
Glu43	4.4	5.6	4.7	4.1	4.7
Ctr56	3.8	2.6	3.2	2.5	< 2.5
rms	1.2	0.5	0.7	0.4	

Figure 2. Illustration of a protein model, where all charges (salt and protein) are located in a high-dielectric region (ϵ_s) and the non-polar interior is described by a spherical, charge-less low dielectric cavity (ϵ_p) of radius a.

Figure 3. Top: Effect of the non-polar protein interior (ϵ_p) on the electrostatic pair interaction between two positive charges situated in a polar region (ϵ_s) a few Ångstroms away from the dielectric interface. Calculated using Eq. 4 with $r_i = r_j = 15$ Å and $a = 13$ Å. Bottom: Plotted as the difference with various values of a.

Figure 4. MC simulation of the decay of the excess chemical potential of a univalent ion outside a low dielectric, neutral sphere (ϵ_p=2, ϵ_s=80) of different radii, a. The energy offsets are arbitrary.

Figure 5. Measured[40] and calculated titration curve for ribonuclease A at 10 mM salt. The MC results are based on the crystal structure 1AFU.[41]

Figure 6. Titration curve for calbindin obtained from intrinsic pK_a-values (ideal), measurements,[22,34] and MC simulations assuming the protein core to be a low dielectric spherical cavity of radius a and dielectric constant, ϵ_p=2. The simulations were performed using the full atomic structure (PDB entry 3ICB[42]) and with a ionic strength of 5 mM matching the experimental conditions.

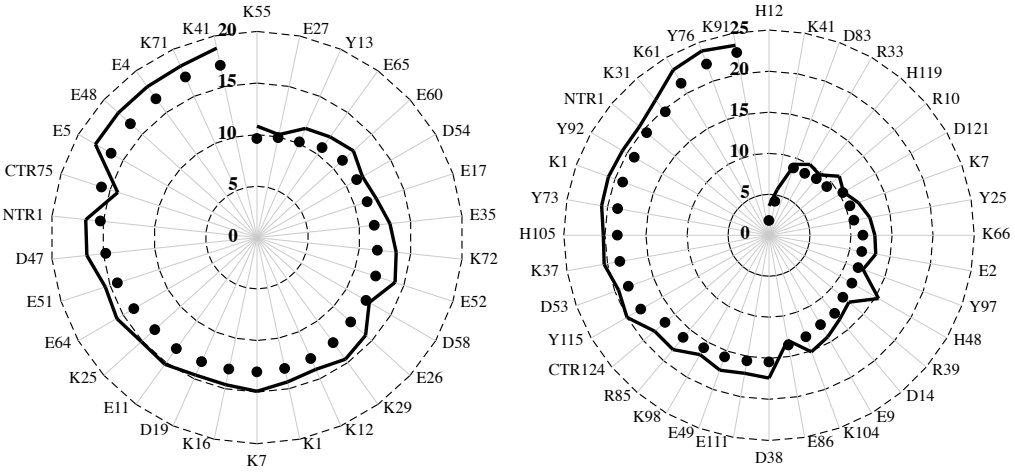


Figure 1:

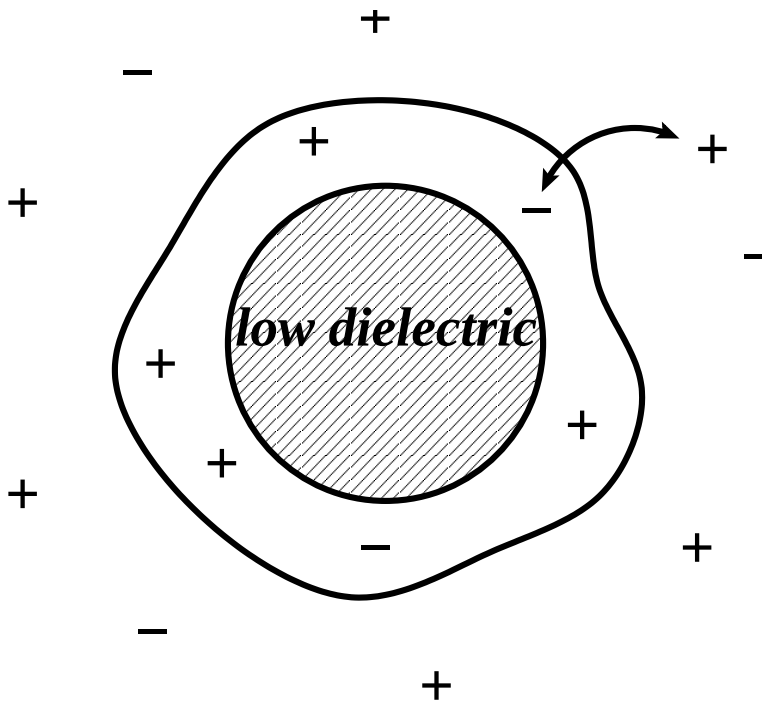


Figure 2:

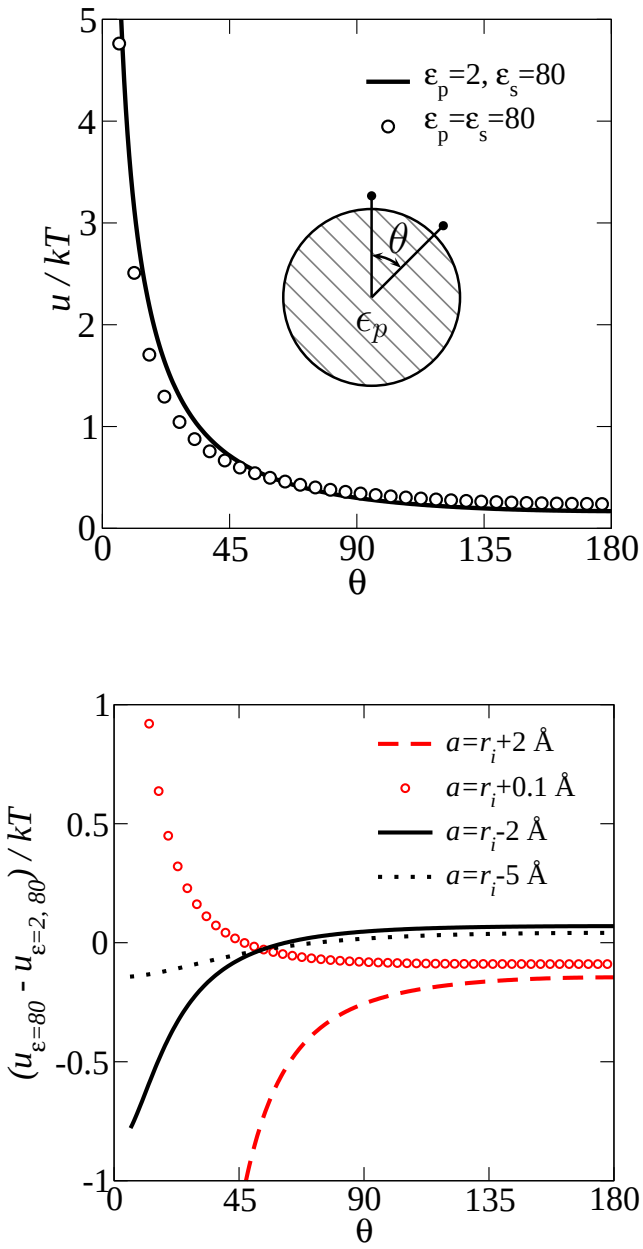

5
4
3
2
1
0
0
45
90
135
180
θ
u / kT
ε_p=2, ε_s=80
ε_p=ε_s=80
θ
ε_p
1
0.5
0
-0.5
-1
(u_ε=80 - u_ε=2, 80) / kT
a=r_i+2 Å
a=r_i+0.1 Å
a=r_i-2 Å
a=r_i-5 Å

Figure 3:

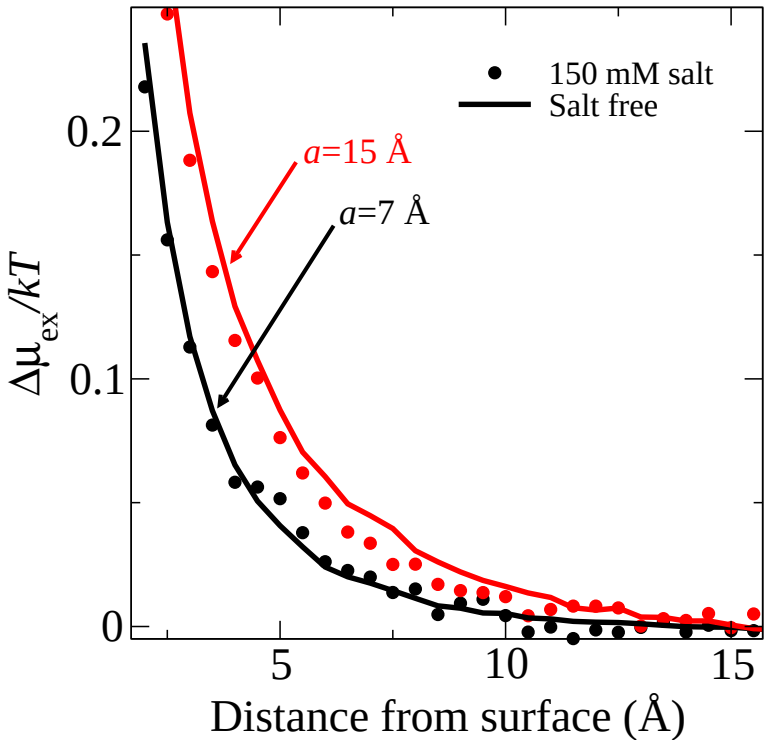


Figure 4:

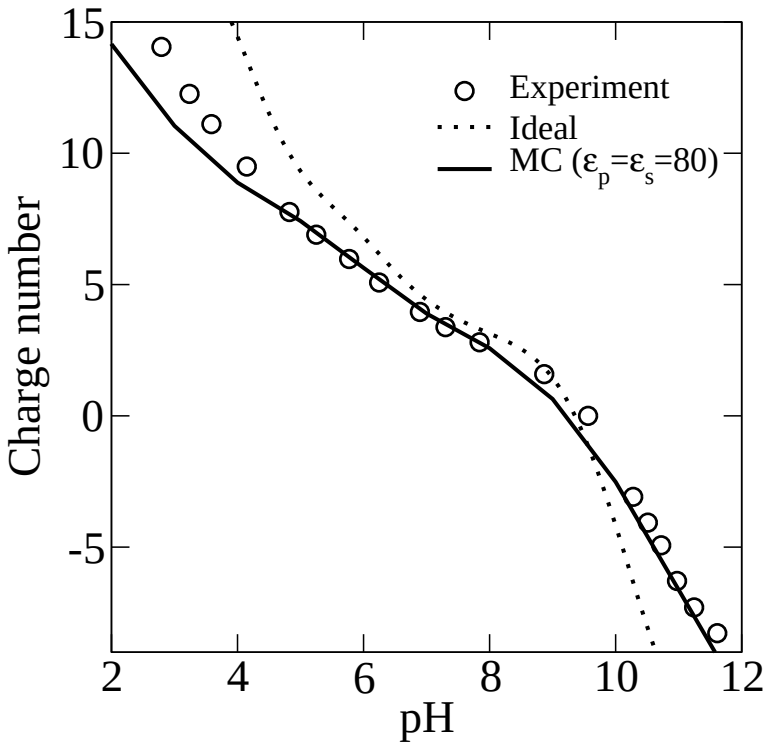


Figure 5:

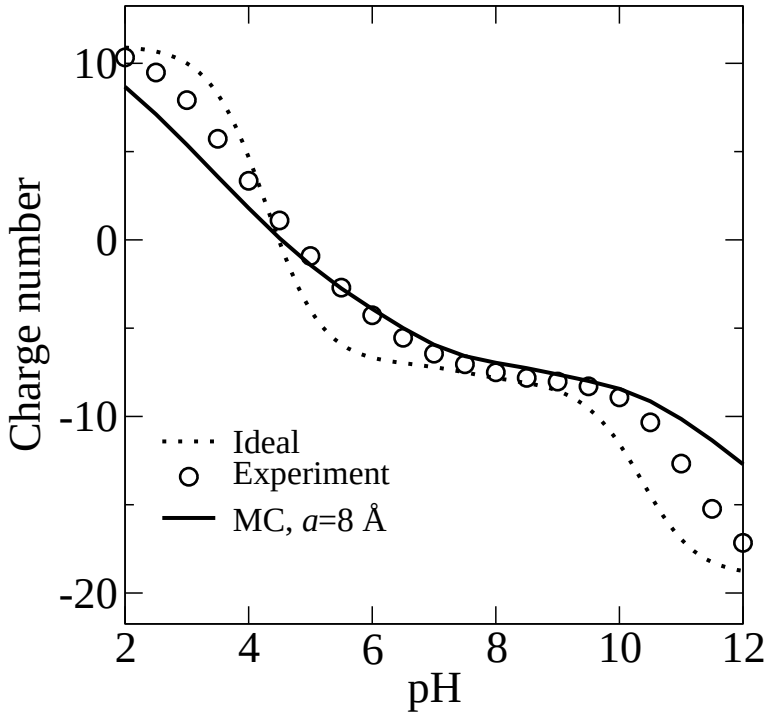


Figure 6:

Driving Forces Behind Ion-Ion Correlations

Mikael Lund and Bo Jönsson
Theoretical Chemistry
Chemical Center, POB 124, S-221 00 Lund, SWEDEN
Phone: +46-46-222 0381 / Fax: +46-46-222 4543
Email: mikael.lund@teokem.lu.se
(October 28, 2006)

It is well known that attractive electrostatic interactions can occur between strongly coupled species of like charge [1, 2, 3, 4, 5, 6]. This is known as ion-ion correlations and in electrostatic continuum models it is often explained via a partitioning of free energies into individual contributions from energy and entropy. For example, Guldbrand *et al.* [1] found that the attraction between like charged surfaces in the presence of divalent counter ions were due mainly to a strong energetic component that could overcome the repulsive entropic contribution from the counter ions. This is true within the primitive model where the solvent screened Coulomb potential is treated as an energy. However, the dielectric constant – and especially that for water – is temperature dependent since it contains solvent degrees of freedom. In the following we will scrutinize the implications of this basic fact and see that the physics behind ion-ion correlations can in fact be explained using alternative arguments.

Consider a N-particle system, $\mathcal{R}^N$, interacting through an effective, temperature dependent potential. The system free energy is then given by

$$A = -kT \ln \int e^{-U(T,\mathcal{R}^N)/kT} d\mathcal{R}^N \tag{1}$$

Even though the configurational "energy" $U(T,\mathcal{R}^N)$ has the character of a free energy, Eq. 1 is still valid and A is the true system free energy. The average energy can always be obtained through

$$\begin{aligned} U &= A - T\frac{\partial A}{\partial T} \\ &= \frac{\int \left(U(T,\mathcal{R}^N) - T\frac{\partial U(T,\mathcal{R}^N)}{\partial T}\right) e^{-U(T,\mathcal{R}^N)/kT} d\mathcal{R}^N}{\int e^{-U(T,\mathcal{R}^N)/kT} d\mathcal{R}^N} \end{aligned}$$

We now make the assumption that the configurational energy function, $U(T,\mathcal{R}^N)$, is a sum of pairwise additive, solvent screened Coulomb potentials. This corresponds to the primitive model and while Eq. 1 now becomes approximate, the agreement with experimental data – both for free energies and derivatives – is often excellent, even at different temperatures [7, 8, 9]. The total energy can now be written as:

$$\begin{aligned} U &= \frac{\left(\frac{T\partial\epsilon}{\epsilon\partial T} + 1\right) \int U(T,\mathcal{R}^N) e^{-U(T,\mathcal{R}^N)/kT} d\mathcal{R}^N}{\int e^{-U(T,\mathcal{R}^N)/kT} d\mathcal{R}^N} \\ &= \left(\frac{T\partial\epsilon}{\epsilon\partial T} + 1\right) \left\langle U(T,\mathcal{R}^N) \right\rangle_{\mathcal{R}} \end{aligned}$$

At room temperature $\frac{T\partial\epsilon}{\epsilon\partial T} \simeq -1.37$ (Ref. [10]) and we arrive at the following

$$U \simeq -0.37 \left\langle U(T,\mathcal{R}^N) \right\rangle_{\mathcal{R}} \tag{2}$$

U in Eq. 2 accounts for both implicit and explicit energy components and should be regarded as the total system energy. This has the interesting implications that interactions in a polar solvent are mostly governed by *entropy*; either from ions or dipoles.

Figure 1 shows a Monte Carlo simulation study (see ref. [6] for technical details) of the interaction between two like charged colloidal particles in the presence of different counter ions. In the case of monovalent counter ions a strong repulsion prevails and in the traditional picture, i.e. where $\partial\epsilon/\partial T = 0$ this is explained by a decreased entropy arising when bringing the macromolecules (and ions) into contact. Also, note that the energy is attractive. In the case of trivalent counter ions this entropic cost is reduced since there is three times less ions and, further, the attractive energetic component is enhanced.

However, if the temperature variation of the dielectric constant of water is included (Fig. 1, right) a different picture emerges. In the first case (monovalent counter ions) the repulsion stems

mainly from entropy, but the energy has changed sign and is no longer attractive. For trivalent ions the inter-colloid, ion-ion correlation attraction is caused by entropy and *not* energy as previously argued in the continuum model. These findings are derived from empirical knowledge of the behavior of the dielectric constant and an attempt to explain them will inevitably involve water dipoles on a microscopic level. Thus, we are beyond the continuum description and the following argumentation is of course open for discussion. If we start with trivalent counter ions, bringing the two spheres into close contact the "aggregate" appears neutral, thus diminishing the electric perturbations of the surrounding medium. For the water dipoles this is entropically favorable as they become less oriented at the cost of interaction energy with the aggregate. In the case of monovalent counterions this effect is still present but is opposed by a strong entropic repulsion brought about by constraining the much larger counterion cloud. This is also reflected in the excess chemical potential for a 1:1 salt, averaged over the entire simulation volume using the Widom particle insertion method [11]; for monovalent counter-ions μ^{ex}=-1.3 kT but only -0.16 kT in the trivalent case. The above considerations can be condensated to the following:

- Weak coupling → Entropic repulsion (Ionic)
- Strong coupling → Entropic attraction (Dipolar)

This is contrary to the correlation argument that in an electro-neutral system within the dielectric continuum model, the energy will always contribute with an attractive interaction. While $\left\langle U(T,\mathcal{R}^N)\right\rangle_{\mathcal{R}}$ can be useful for deciphering simulation results it is important to keep in mind that the real energy – given by Equation 2 – dictate a *change in sign* turning the entropy/energy balance upside down. This is of course critical when comparing against calorimetric data and since $T/\epsilon\partial\epsilon/\partial T$ decrease with increasing T larger deviations can be expected at elevated temperatures.

Acknowledgments

For financial support we thank The Swedish Foundation for Strategic Research and The Research School in Pharmaceutical Sciences, Sweden.

References

[1] L. Guldbrand, B. Jönsson, H. Wennerström, and P. Linse. Electric double layer forces. a monte carlo study. *J. Chem. Phys.*, 80:2221, 1984.

[2] R. Kjellander and S. Marčelja. Correlation and image charge effects in electric double layers. *Chem. Phys. Letters*, 112:49–43, 1984.

[3] R. M. Pashley and J. N. Israelachvili. Molecular layering of water in thin films between mica surfaces and its relation to hydration forces. *J. Coll. Interface Sci.*, 101:511–523, 1984.

[4] A. Khan, K. Fontell, and B. Lindman. Liquid crystallinity in systems of magnesium and calcium surfactants : Phase diagrams and phase structures in binary aqueous systems of magnesium and calcium di-2-ethylhexylsulfosuccinate. *J. Colloid Interface Sci.*, 101:193–200, 1984.

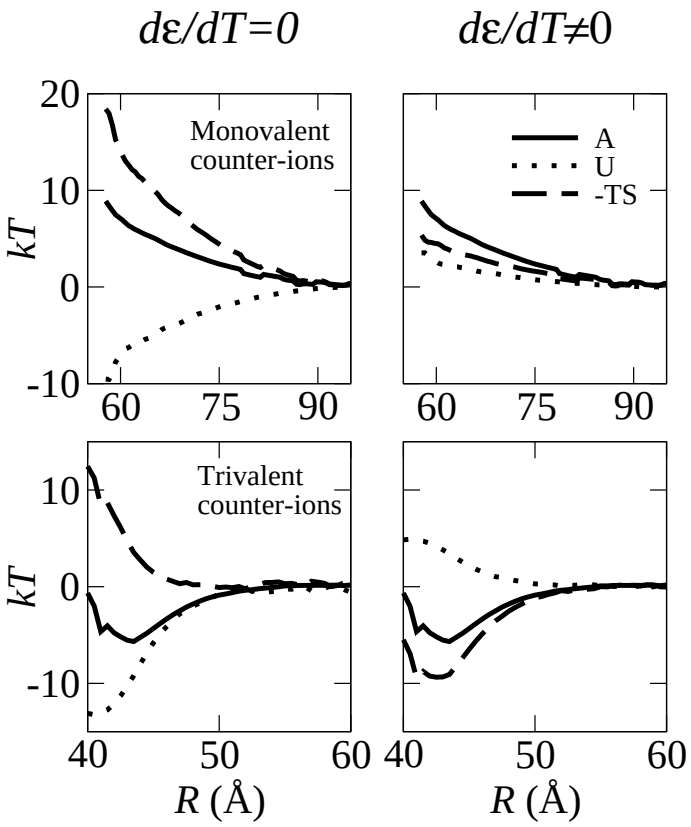


Figure 1: Energy decomposition of the interaction between two like charged spheres ($Z = 60$, radius 2 nm) in an aqueous solution with different counter-ions. Left: Temperature dependence of the dielectric constant is neglected. Right: Temperature dependence included according to Eq. 2. (Full drawn lines = A, dashed = $-TS$, dotted = U)

[5] B. Jönsson and H. Wennerström. *When ion-ion corelations are important in charged colloidal systems. In Electrostatic Effects in Soft Matter and Biophysics.* Kluwer Academic Publishers, Eds. C. Holm and P. Kekicheff and R. Podgornik, 2001.

[6] M. Lund and B. Jönsson. A mesoscopic model for protein-protein interactions in solution. *Biophys. J.*, 85:2940–2947, 2003.

[7] K. S. Pitzer. *Activity coefficients in electrolyte solutions.* CRC-Press, London, 2 edition, 1991.

[8] I. Carlsson, A. Fogden, and H. Wennerstrom. Electrostatic interactions in ionic microemulsions. *Langmuir*, 15(19):6150–6155, 1999.

[9] M. Lund, B. Jönsson, and T. Pedersen. Activity coefficients in sea water using monte carlo simulations. *Marine Chemistry*, 80:95–101, 2003.

[10] R. C. Weast, editor. *CRC Handbook of Chemistry and Physics.* CRC Press, Boca Raton, Fl, 1974.

[11] B. Widom. Some topics in the theory of fluids. *J. Chem. Phys.*, 39:2808–2812, 1963.

Electrostatic Interactions Between Proteins in a Salt Solution - A Monte Carlo Simulation Study

M. Lund and Bo Jönsson

Theoretical Chemistry, Chemical Center, POB 124, S-221 00 Lund, SWEDEN

October 31, 2006

Introduction

Due to their complex structure and mixed composition biomolecules interact via a wide range of physical mechanisms. This includes hydrophobic-, steric-, dispersion-, and electrostatic forces and in order to perform predictive studies an understanding of these interactions on a microscopic level is essential. It is generally assumed that electrostatic interactions play key roles for both structure and function of these molecules. Hence, electrostatic interactions in biomolecules have been intensively studied over the years. Many studies deal with the properties of a single biomolecule [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], while the electrostatic interaction between, for example, two proteins have not enjoyed the same intense interest [12, 13, 14, 15]. In this communication we will treat both the electrostatic and van der Waals interaction between two proteins/peptides in a solution with varying salt concentration and pH. The focus will, however, be on the former type of interaction. For two equally charged proteins the free energy of interaction is dominated by a strong repulsion, which can be significantly modulated by addition of salt. The energetic and entropic components of this interaction can sometimes vary in a counterintuitive way and we will try to emphasize this conceptually important fact. Additional van der Waals and/or hydrophobic interactions can sometimes overcome the Coulombic repulsion and we will discuss the relative importance of these terms, which can be conveniently done via the second virial coefficient. At short separation, close to contact, the interaction between two proteins will not only depend on their net charge but on the detailed charge distribution in the protein. The non-uniform charge distribution depends on the protein structure but also on solution conditions such as pH and salt concentration. Often, when transcribing protein structure to functionality the charge distribution is characterized by loosely formulated terms like "charge patchiness" and "charge compartments". While these terms may be intuitively appealing, they fail to quantify any kind of binding properties. We will investigate the importance of the actual charge distribution in the protein and try to describe it in terms of a multipole expansion. That is, we will discuss the free energy of interaction in terms of ion-ion, ion-dipole, ion-quadrupole, dipole-dipole etc. interactions and demonstrate its applicability for a number of protein complexes in aqueous salt solutions.

The non-uniform charge distribution, as well as a non-uniform distribution of hydrophobic amino acids, should in principle induce orientational correlations between two protein molecules. We will try to describe this and to quantify the importance of the different terms. Note, however, that the ion-dipole interaction is subject to an orientational average, which tend to reduce its importance - the free energy of interaction between an ion and a dipole decays asymptotically as R^{-4}, where R is their separation. This term can ce compared with the ion-induced ion interaction, which is not orientationally averaged and decays as R^{-2}. The ion-induced ion term comes from the fact that the protonation state of an amino acid residue is perturbed by nearby charged molecules (ions, other proteins, DNA etc.). That is, when two proteins approach, their

ionization states become correlated lowering the free energy. This *charge regulation* interactions can be included in the multipole expansion and can be quantified in terms of the *capacitance*, C, which is the key property for describing the ability for charge regulation.

Methodology and Theory

Interaction Model

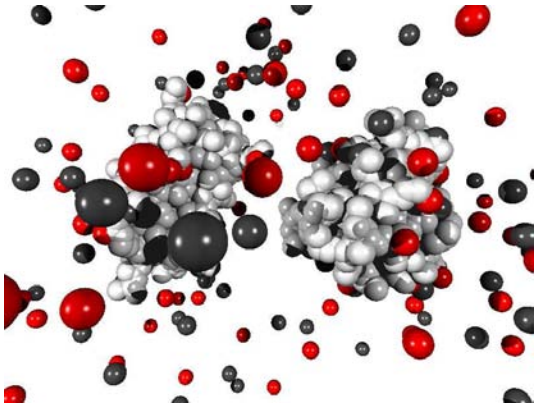


Figure 1: Snapshot from a MC simulation of two proteins. The red and grey spheres illustrate mobile cations and anions, while amino acids are depicted as a white spheres, clustered to form the two proteins. In a simulations, the proteins are displaced along a line and rotated independently. Ions are displaced in all three directions and the whole system is enclosed in sphere of appropriate radius.

The simulated model system consists of two protein molecules built from spheres immersed in a spherical cell see Fig.. To maintain electroneutrality and the desired salt concentration, mobile salt particles with hard sphere diameters, $\sigma = 4$ Å, are added. As for the protein shape, two models have been utilized both based on structural data obtained from the Brookhaven Protein Databank (PDB). In the first *atomic model* the protein molecules are mimicked by replacing each nonhydrogen atom in the protein by a hard sphere with diameter $\sigma = 4$ Å, which, for the largest system results in ??? particles/protein. In addition, one has to include salt particles and in a dilute protein solution with high salt concentration, the total number of interacting particles can add up to many thousands. To improve the simulation efficiency, a slightly simplified *mesoscopic model* has been developed. Here the atoms in each amino acid are replaced by a single sphere located at the amino-acid center of mass. The size of these spheres are set equal for all residues and adjusted so that the total excluded volume of the protein is equal to that of the atomic model. This amounts to a diameter for the amino acid spheres of $\sigma = 6.8$ Å. Charges are assigned to the center of each sphere according to the actual pH. The atomic and mesoscopic models produce virtually identical results for the charge distribution as well as the free energy of interaction of two proteins.

The actual charge on an amino-acid residue is pH dependent, since acidic and basic amino acids can titrate. This can be taken into account in the simulations, and the charge on a residue will fluctuate and its average value will be determined by the difference $pK_a - pH$, where pK_a is the acid constant of the hypothetical free amino acid [16]. This theoretical approach has been applied to single proteins and shown to be in good agreement with experimental data from nuclear magnetic resonance studies [17, 18].

We use a dielectric continuum model for the solution assuming that all charges are uniformly screened by a constant relative permittivity with a value equal to that of pure water. Thus, the simulation cell will contain a collection of spheres, which are fixed in space and can be

either charged or neutral. In addition the mobile salt partcles are also described as charged hard spheres - see Fig.. The total interaction energy can be split into short-range repulsive (hs), electrostatic (el), and van der Waals (vdW) terms,

$$U = \sum_{i>j} [u_{hs}(r_{ij}) + u_{el}(r_{ij}) + u_{vdW}(r_{ij})] \tag{1}$$

where r_{ij} is the distance between two sites. The repulsive u_{hs} term accounts for the repulsion arising when the electron clouds from two atoms or molecules come into contact. An exact description of this contribution requires a complex quantum mechanical treatment, and hence the simpler hard-sphere term is usually applied as,

$$u_{hs}(r_{ij}) = \infty \qquad r_{ij} \leq \frac{\sigma_i + \sigma_j}{2} \tag{2}$$

where σ_i is the diameter of site i. The electrostatic term includes Coulombic interactions between charged sites, and in the dielectric continuum approximation it can be written as,

$$u_{el}(r_{ij}) = \frac{z_i z_j e^2}{4\pi\epsilon_0\epsilon_r r_{ij}} \tag{3}$$

where ϵ_r is the relative dielectric permittivity, z_i the valency of site i, e the electron charge, and ϵ_0 the permittivity of vacuum. To describe the short-range attraction between two protein sites we invoke a van der Waals-type interaction,

$$u_{vdW}(r_{ij}) = -\frac{D_{ij}}{r_{ij}^6} \tag{4}$$

Here D_{ij} determines the magnitude of the attraction and is related to the Hamaker constant, A [19],

$$A_{ij} = \pi^2 \rho_i \rho_j D_{ij} \tag{5}$$

where the ρ values are particle densities. As a first approximation, we have decided to use the same D_{ij} for all amino acids. A straightforward improvement would be to let the amino-acid size affect the interaction, hence one would have a different D_{ij} for each pair of amino acids. Calculation of Hamaker constants can be done using the Lifshitz theory [19], but detailed knowledge of the electronic properties is required and as a consequence, A is often treated as an adjustable parameter. Fortunately, Hamaker constants are not subject to large variations and for proteins in water, $A \sim 3 - 10$ kT [20] ($1\ kT = 4.11 \cdot 10^{-27}$ J at 298.15 K). In addition, there will also be hydrophobic interactions between some amino acids. One way to treat this attractive term would be to augment the van der Waals term with an extra attraction at short separations. We have neglected this interaction as it is probably only important when there is a structural rearrangement in the aggregation process and this is not taken into account in the present model.

Monte Carlo simulations

Most simulations were performed in the canonical ensemble using the traditional Metropolis Monte Carlo (MC) algorithm [21] supplemented with a semicanonical procedure allowing the protein to titrate[17]. The energy evaluation for each configuration includes all pair interactions as in eq.(1). During the MC simulation the proteins are allowed to translate symmetrically along the z-axis and individually rotate around vectors going through their center-of-mass. Mobile ions may translate in any direction. By these random displacements and rotations all possible configurations are explored. If a move leads to an energy decrease, the new state is accepted.

If the energy increases, the state is accepted with the probability $\exp(-\Delta U_{el}/kT)$. In addition to these particles moves, the protein is also allowed to titrate. This is done via a random change of the ionization status of titrating residues. The acceptance for the titration process is controlled by a change in electrostatic interactions, ΔU_{el}, plus the cost for ionizing/neutralizing the randomly chosen amino acid. The appropriate Boltzmann factor reads,

$$\exp(-\Delta U_{tit}/kT) = \exp[-\Delta U_{el}/kT \pm \ln 10(pK_a - pH)]) \tag{6}$$

where pH is the chosen pH and pK_a is the acid constant for the particular amino acid. The second term in the exponential can be either positive or negative, depending on whether the group is ionized or neutralized. After completion of this semicanonical MC scheme, one obtains the average charge on each titrating residue and hence the proper net charge of the protein. Note that this procedure mimics the experimental situation, in which a proton released from the protein is absorbed by buffer maintaining a constant pH. In a few simulations, we have suppressed the titration and instead used fixed charges appropriate for that particular pH. The results from these simulations give an indication of the importance of charge regulation.

Proceeding this way, the system eventually reaches equilibrium and its properties can be sampled. The probability distribution, $P(R)$, for the protein-protein separation is readily sampled and the free energy of interaction is,

$$\beta A(R) = -\ln P(R) + \text{const} \tag{7}$$

where $\beta = 1/kT$ and R is the distance between the center-of-masses of the two proteins and the constant can be found from the asymptote of $P(R)$.

The MC simulations give the exact free energy, but it is sometimes more informative and gives a better insight to make an approximate partitioning of the free energy. This will allow us to identify different physical components of the interaction free energy. For this purpose, we will use statistical mechanical perturbation theory [22] to derive expressions for the free energy of interaction, A between two charged molecules. Performing a thermal averaging of the intermolecular interaction energy, $U(R, \Omega, X)$ we obtain the *exact* solution,

$$\beta A(R) = -\ln \left\langle e^{-\beta U(R,\Omega,X)} \right\rangle \tag{8}$$

where the brackets denote an unbiased average over all possible configurations, including orientations (Ω) and protonation states (X). If the interaction is small compared to the thermal energy, kT, then Eq. 8 can be expanded to second order to give,

$$\beta A(R) \approx \beta \left\langle U(R) \right\rangle - \frac{\beta^2}{2} \left[\left\langle U(R)^2 \right\rangle - \left\langle U(R) \right\rangle^2 \right] \tag{9}$$

The electrostatic energy, $U(R, \Omega, X)$, is simply the Coulomb potential, summed over all charged groups, i and j in the two proteins. If salt is present, a screened Debye-Huckel potential is more appropriate,

$$\beta U(R, \Omega, X) = \begin{cases} \dfrac{l_B \sum\sum z_i z_j}{r_{ij}} & \text{(no salt)} \\ \dfrac{l_B \sum\sum z_i z_j e^{-\kappa r_{ij}}}{r_{ij}} & \text{(salt)} \end{cases} \tag{10}$$

Here, $l_B = e^2/4\pi\epsilon_0\epsilon_r kT$ is the Bjerrum length, r_{ij} is the distance between the charges and κ is the inverse Debye length. Combining eq.(9) and (10) and performiing a multipole expansion of the charge distribution on each protein, we obtain the free energy expressions as a function of the protein center-center separation, R (See Table 1). Now we can identify different terms like the direct ion-ion interaction, ion-induced ion, ion-dipole etc. These contributions are

expressed in terms of the charge number, Z, the capacitance, $C = \langle(\sum_i z_i)^2\rangle - \langle\sum_i z_i\rangle^2$ and the average dipole number, μ. (The dipole number has the dimension of length and by multiplying with the elementary charge one obtains the proper dipole moment, μe.) Note that the direct ion-dipole interaction varies like $1/R^2$, but that the orientational averaging changes it into $1/R^4$ and similarly for the dipole-dipole interaction. The Debye-Hückel theory can, at least, approximately take into account the salt-excluded volume taken up by the proteins, by including a size parameter, a. We have found that an accurate choice of a is to set it equal to the average protein radius.

Table 1: Expressions for the free energy of interaction between two charged molecules A and B. The third column shows the Debye-Hückel screening factor for the various types of interactions. R is the center-to-center distance and κ is the inverse Debye length. The size parameter, a is the average protein radius. The subscripts A and B refers to the two proteins.

Interaction	$\beta A(R)$	Screening factor
Ion-ion	$\frac{l_B Z_A Z_B}{R}$	$\frac{e^{-\kappa(R-a)}}{1+\kappa a}$
Ion-induced	$-\frac{l_B^2 Z_A^2 C_B}{2R^2}$	$\left(\frac{e^{-\kappa(R-a)}}{1+\kappa a}\right)^2$
Induced-induced	$-\frac{l_B^2 C_A C_B}{2R^2}$	$e^{-2\kappa R}$
Ion-dipole	$-\frac{(l_B Z_A \mu_B)^2}{6R^4}$	$\left(\frac{e^{-\kappa(R-a)}}{1+\kappa a}\right)^2 (1+\kappa R)^2$
Dipole-dipole	$-\frac{(l_B \mu_A \mu_B)^2}{3R^6}$	

Results

Basic protein properties

In order to demonstrate the presented methods we shall analyze two protein complexes, namely fab-lysozyme (antibody-antigene) and barnase-barstar (substrate-inhibitor). The structure of both complexes are known (PDB entries 3HFL and 1BRS) and the individual protein structures can be directly employed in the MC simulations - Table 2 shows their amino acid composition.

Table 2: Titrating residues in the investigated proteins. The standard dissociation constants for the isolated amino acids [16] are given in the second line and the corresponding pK_0 for C- and N-termini are 3.8 and 7.5, respectively. †Only cysteins not engaged in sulfide bridges can titrate.

Protein	residues	Asp	Glu	His	Tyr	Lys	Cys†	Arg
pK_0	-	4.0	4.4	6.3	9.6	10.4	10.8	12.0
Lysozyme	129	7	2	1	3	6	0	11
Fab (L-chain)	214	10	10	2	11	13	1	7
Barnase (A-chain)	110	9	3	2	7	8	0	6
Barstar (D-chain)	89	4	10	1	3	6	0	3

Valuable information is obtained by simulating the protonation state of a *single* protein in a salt solution. Computationally, this is a fast procedure and the resulting data immediately provides insight to the intermolecular interactions. The essential properties are: the total charge number (Z), the charge capacitance (C), and the dipole moment (μ). Values are shown in Fig.3 and 2 and we shall use data from these graphs as input for the perturbation expressions in Table 1. Before doing so, we give a description of the obtained properties:

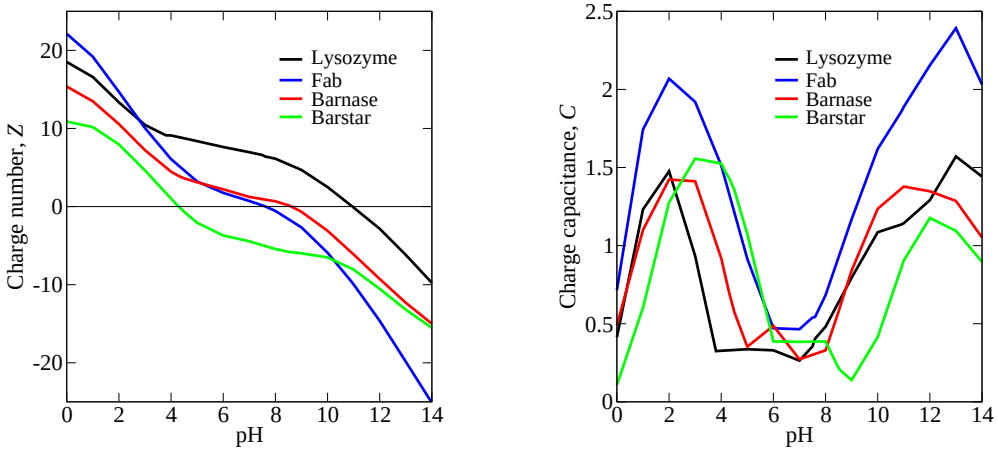


Figure 2: Charge numbers and capacitances for some of the investigated proteins. All data are calculated using amino acid level Monte Carlo simulation of a single, titrating protein in a dilute salt solution.

At low pH all proteins are fully protonated, but as pH is increased they pass their iso-electric points which, for the investigated proteins, span a large interval, from barstar with pI=4.3 to lysozyme with pI=10.9. At high pH, eventually all protons are released giving rise to substantial net negative charge and possibly denaturation. The presented calculations are based on the x-ray structure of the proteins and any structural rearrangement is not taken into account in the simulation. If the protein net charge is high, far from pI, it will in most cases dominate

the interaction via the direct ion-ion term, $l_B Z_A Z_B/R$. The ion-induced and ion-dipole terms are more short range and will under these conditions usually only make a small contributions. However, the ion-induced and ion-dipole interactions both scale as *charge squared* and thus, these interactions are more sensitive to the proteins protonation state. Close to pI these terms will dominate the electrostatic interaction.

The iso-electric point of barnase is approximately 8.5, which agrees with the experimental stability measurements by Schreiber and Fersht [23]. At pH=8 these authors found that the stability of the barnase-barstar is reduced upon mutation of positively charged lysine and arginine into alanine. According to Fig.2a barnase is positively and barstar negatively charged at this pH. The mutation of acidic residues in barnase should then based on electrostatic arguments increase the stability. This is seen in two out of three cases in the study and in all three cases the changes are smaller than for the basic residues.

The charge capacitance is a measure of how the total protonation state is affected by an external potential, stemming from another charged object. As can be seen in Table 1, the induced interactions scale linearly with the capacitance and the regulation mechanism can therefore be important at pH where C peaks. All four capacitance curves in Fig 2 have similar forms with peaks at low and high pH, caused by the large amount of acidic and basic residues (see Table 2). Notable, in all cases at neutral pH the capacitance is relatively small. This is a typical result and a result of the lack of histidine residues. Histidine-rich proteins are not abundant, but some examples are known, namely hisactophilin and histidine-rich glycoprotein (HGR). It has been proposed [24] that charge regulation is an important driving force when hisactophilin binds to lipid membranes *in vivo*.

The dipole moment is the first order approximation for describing the uneven charge distribution of a molecule. Note, however, that the dipole moment is only strictly defined for a neutral molecule. Here we have evaluated it around the protein mass centra, which are also used for defining the protein protein separation. Fig 3 shows that the dipole moment varies rather modestly for lysozyme, barnase and barstar, but changes dramatically for fab when pH is varied. The fab dipole moment is almost 700 D at pH=7. This is due to an uneven distribution of the Asp, Glu and Tyr residues. The two former start to titrate at around pH=4, creating a large dipole moment, while when the tyrosines begin to titrate we see a reduction of the dipole moment. The large dipole moment of fab will have consequences for its orientational beahviour in the fab-lysozyme complex.

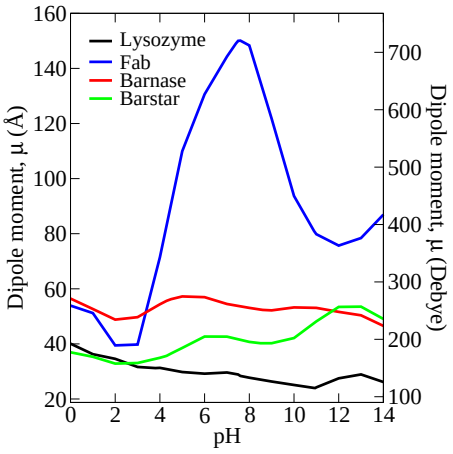


Figure 3: Dipole moments for some of the investigated proteins. All data are calculated using amino acid level Monte Carlo simulation of a single, titrating protein in a dilute salt solution.

The second virial coefficient

Table 3: The interaction of two different proteins, A and B, different pH. The isoelectric points of the proteins are denoited by pI_A and pI_B, respectively. The interactions can be divided into: ion-ion, ion-induced charge (ion-ind), ion-dipole (ion-dip), dipole-dipole (dip-dip), induced charge-induced charge (ind-ind), and van der Waals (vdW) interactions.

pH range	Interactions
$pI_A < pH > pI_B$	Repulsive (ion-ion)
$pI_A > pH < pI_B$	Repulsive (ion-ion)
$pI_A < pH < pI_B$	Attractive (ion-ion, ion-ind, ion-dip, vdW)
$pI_A \approx pH \approx pI_B$	Attractive (ind-ind, dip-dip, vdW)
pI_A or $pI_B \approx pH$	Attractive (ion-ind, ion-dip, vdW)

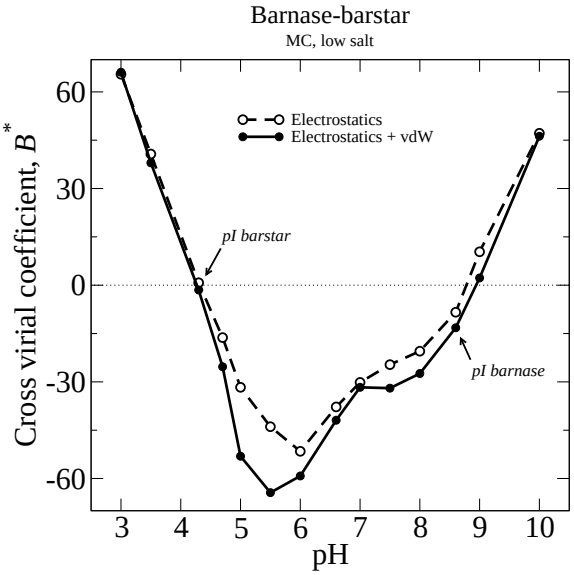


Figure 4: Cross second virial coefficient for barnase-barstar as a function of pH - both with and without van der Waals interactions. Calculated from MC simulations at low salt concentration.

The potential of mean force, $A(r)$ is a microscopic quantity, but can be integrated to yield measurable thermodynamic entities. In particular, the second virial coefficient is commonly used to weigh protein-protein interactions and is defined by,

$$B = -2\pi \int_0^\infty \left(e^{-\beta A(r)} - 1\right) r^2 dr \tag{11}$$

When B is positive, the net interaction is repulsive, while if negative the proteins attract each other. Virial coefficients can be defined for both homo- and heterodimers, the latter being the cross virial coefficient. Closely related to B is the binding constant [25],

$$K \approx 4\pi \int_\sigma^\infty \left(e^{-\beta A(r)} - 1\right) r^2 dr = 2\left(B_{hc} - B\right) \tag{12}$$

valid for relatively short range interactions. B_{hc} is the hard core contribution equal to $2\pi\sigma^3/3$, where σ is the average protein diameter.

The protein protonation state is influenced by solution pH and it has a large impact on protein-protein interactions - Table 3 gives an overview of the dominant interactions at various

pH. Figure 4 shows the cross virial coefficient for the barnase-barstar complex and at extreme pH when the proteins are like charged, strong repulsive interactions prevail. The opposite is observed at pH values spanning the two iso-electric points, where the net charges are of opposite signs, causing the proteins to attract. The virial coefficient increases steadily when pH is raised from 5.5 to 10. This reduction in complex stability is also found in the experimental study of Schreiber and Fersht [23], where they measure decreasing equilibrium constant for the barnase-barstar complex in the pH range 4.5 to 9.

For attractive interactions, close-contact configurations are more populated and short range interactions become increasingly important. This is demonstrated at $pH \approx 5-6$ where the cross virial coefficient is noticeably lowered by van der Waals interactions, whereas at high and low pH van der Waals interactions have only little or no effect. Due to the short range nature of van der Waals interactions, surface complimentarity is an important aspect [12]. Ion-dipole and dipole-dipole interactions tend to orient the protein molecules, thus constraining the number of possible contact configurations. This has consequences for the van der Waal's interactions that may be modulated in non-trivial manners as evident in Figure 4.

At the iso-electric points the direct ion-ion interaction has vanished, leaving only van der Waals and higher order electrostatic terms (ion-dipole, regulation etc.). Importantly, we note a significant attraction at pH 8.6 (pI for barnase) caused mainly by ion-dipole and regulation interactions.

Electrostatic interactions near the isoelectric point

When pH=pI for one of the proteins the strong ion-ion term vanishes and the effective terms are ion-induced charge and ion-dipole interactions. To investigate the relative importance of these, we have used Eqs. 11 and 12 and the expressions from Table 1 to calculate the cross second virial coefficients and binding constants at the iso-electric points and at different salt concentrations. This is shown for the fab-lysozyme system in Figure 5a. The interaction is strongly

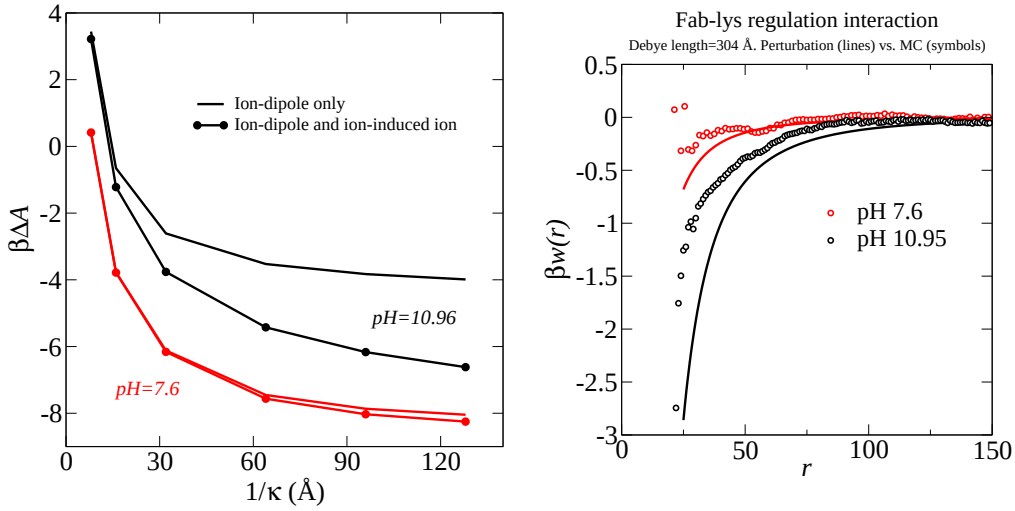


Figure 5: a) CHANGE! Lysozyme-fab association free energies ($\beta\Delta A = -\ln K$) at different salt concentrations and pH, calculated at the iso-electric points using perturbation formulas from Table 1. Lines represents calculations with ion-dipole interactions only, while for lines with symbols the regulation interaction has been included. b) The charge regulation interaction for the lysozyme-fab system calculated using Monte Carlo simulation (circles) and perturbation theory (lines). Debye length=300 Å.

salt dependent and at physiological concentrations there remains only small contribution. The

strong salt dependence comes from the fact that these term only contribute in second order and the Debye-Hückel screening is $\exp(-2\kappa r)$. At pH 7.6 (pI for fab), including the ion-induced charge term has little or no effect on the binding constant. Thus the ion-dipole term is completely dominant due to the unusually high dipole moment of fab at this pH ($\mu_{fab} \approx 700$ D). and its relatively low capacitance ($C_{fab} = 0.5$). At pH 10.9 (pI for lysozyme) the induced interaction becomes increasingly important at low salt concentrations and at 1 mM salt ($1/\kappa = 96$ Å) the association constant has been enlarged by one order of magnitude, compared to the pure ion-dipole description. Again, this is explained using data from Fig 2 and 3: At pH 10.9 the dipole moment of lysozyme is very small ($\mu_{lys} = 24$ Å) while the capacitance is around 1. The above findings are further supported by numeric simulations of the lysozyme-fab system as illustrated in Figure 5b. The shown data are obtained by subtracting the potential of mean force obtained from two simulations - one with charge titration and another without. Although the qualitative picture is retained, the highly non-spherical shape of the fab molecule causes some discrepancies between the simulation and the perturbation results.

Orientational dependency

An interesting feature of the ion-dipole interaction, is that it strides to orient a dipolar molecule and thus can be of importance for recognition processes in the cell. In a MC simulation the orientational ordering can be probed by averaging the z-components of the dipole moment vectors. That is, we can calculate the angle ,Θ, between the vector connecting the two proteins and the dipole moment vector. This average will be zero for a completely random orientation and unity for a perfectly aligned dipole. Figure 6 shows the average angle, $\langle \cos \Theta \rangle$, as a function of the fab-lysozyme separation at different pH. When the proteins are far apart, $\langle \cos \Theta \rangle$ approaches zero, but at shorter separations an alignment occurs. At around neutral pH, fab is electroneutral and has a large dipole moment. At the same condition, $Z_{lysozyme} \approx 7$ and we therefore expect the ion-dipole interaction to be significant. This is indeed the case and manifested by a pronounced alignment of the fab molecule, $\langle \cos \Theta \rangle \approx 0.65$. We also note that the alignment occurs at intermediate protein-protein separations and is a direct consequence of the elongated shape of fab. As pH is raised, Z_{fab} becomes negative while $Z_{lysozyme}$ and μ_{fab} decreases. Eventually the picture is reversed and lysozyme is aligned, while fab is more or less freely rotating. However, since $\mu_{lysozyme}$ is quite small, the alignment effect is not as large as in the fab case. The maximum alignment happens at closest possible contact, because the shape of lysozyme is approximately spherical.

References

[1] J. Warwicker and H. C. Watson, J. Mol. Biol. **157**, 671 (1982).

[2] A. Warshel, S. T. Russel, and A. K. Churg, Proc. Natl. Acad. Sci. USA **81**, 4785 (1984).

[3] M. K. Gilson, A. Rashin, R. Fine, and B. Honig, J. Mol. Biol. **183**, 503 (1985).

[4] K. Sharp, R. Fine, and B. Honig, Science **236**, 1460 (1987).

[5] D. Bashford, M. Karplus, and G. W. Canters, J. Mol. Biol. **203**, 507 (1988).

[6] B. Svensson, B. Jönsson, and C. E. Woodward, Biophys. Chem. **38**, 179 (1990).

[7] B. Svensson, B. Jönsson, C. E. Woodward, and S. Linse, Biochemistry **30**, 5209 (1991).

[8] J. Antosiewicz, J. A. McCammon, and M. K. Gilson, J. Mol. Biol. **238**, 415 (1994).

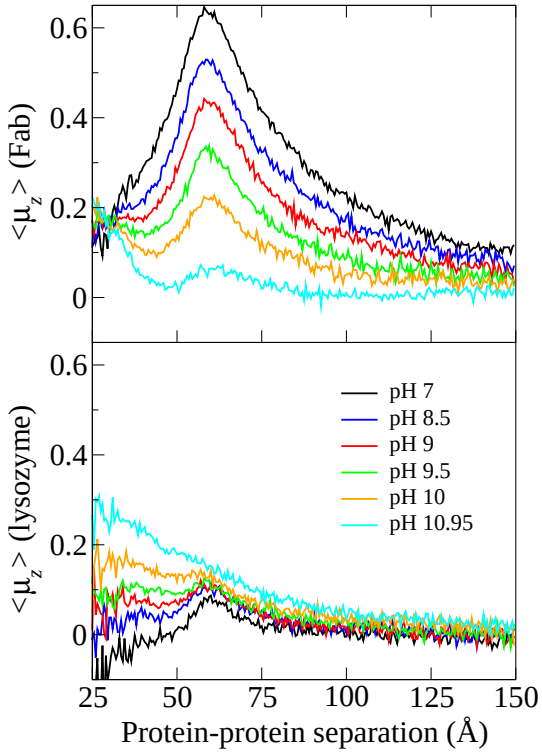


Figure 6: Average orientation ,$\langle \cos\Theta \rangle$, for the dipole moment vector in the fab-lysozyme complex at different pH and protein-protein separations. Obtained from MC simulation at low salt conditions. $\langle \cos\Theta \rangle = 0$ means that the protein is completely randomly oriented and if $\langle \cos\Theta \rangle = 1$ it is perfectly aligned.

[9] J. Antosiewicz, J. A. McCammon, and M. K. Gilson, Biochemistry **35**, 7819 (1996).

[10] B. Honig and A. Nicholls, Science **268**, 1144 (1995).

[11] V. Spassov and D. Bashford, Protein Sci. **7**, 2012 (1998).

[12] D. Asthagiri, B. L. Neal, and A.M. Lenhoff, Biophys. Chem. **78**, 219 (1999).

[13] E. Allahyarov, H. Löwen, J. P. Hansen, and A. A. Louis, Europhys. Lett. **57**, 731 (2002).

[14] A. H. Elcock and J. A. McCammonm, Biophys. J. **80**, 613 (2001).

[15] M. Lund and B. Jönsson, Biophys. J. **85**, 2940 (2003).

[16] Y. Nozaki and C. Tanford, Methods Enzymol. **11**, 715 (1967).

[17] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse, J. Mol. Biol. **259**, 828 (1996).

[18] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse, Proteins **45**, 129 (2001).

[19] J. Israelachvili, *Intermolecular and Surface Forces*, 2nd ed. (Academic Press, London, 1991).

[20] M. Farnum and C. Zukoski, Biophys. J. **76**, 2716 (1999).

[21] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1989).

[22] D. A. McQuarrie, *Statistical Mechanics* (Harper Collins, New York, 1976).

[23] G. Schreiber and A. Fersht, Biochemistry **32**, 5145 (1993).

[24] M. Lund, T. Åkesson, and B. Jönsson, Langmuir **21**, 8385 (2005).

[25] D. F. Evans and H. Wennerström, *The Colloidal Domain - Where Physics, Chemistry, Biology and Technology Meet* (VCH Publishers, New York, 1994).

October 31, 2006

Electrostatics in Macromolecular Solution

Bo Jönsson, Mikael Lund and Fernando L. Barroso daSilva

Theoretical Chemistry, Chemical Center, POB 124, S-221 00 Lund, SWEDEN

Abstract

An overview of the interaction between charged macromolecules in aqueous solution is presented. The starting point is the dielectric continuum model and the Debye-Hückel equation. The usefulness of the simple theory is emphasized in particular for biological macromolecules, whose net charge or surface charge density often is low. With more highly charged macromolecules or aggregates it may be necessary to go beyond the simple Debye-Hückel theory and invoke the non-linear Poisson-Boltzmann equation or even to approach an exact solution using Monte Carlo simulations or similar techniques. The latter approach becomes indispensable when studying systems with divalent or multivalent (counter)-ions. The long range character of the electrostatic interactions means that charged systems of varying geometry - spheres, planes, cylinders... - often have many properties in common. Another consequence is that the detailed charge distribution on a macromolecule is less important. Many biological macromolecules contain titratable groups, which means that the net charge will vary as a consequence of solution conditions. This gives an extra attractive contribution to the interaction between two macromolecules, which might be particularly important close to their respective isoelectric points. The treatment of flexible polyelectrolytes/polyampholytes requires some extra efforts in order to handle the increasingly complex geometry. A theoretical consequence is that the number of parameters - chain length, charge density, polydispersity etc - prohibits the presentation of a simple unified picture. An additional experimental, and theoretical, difficulty in this context is the slow approach towards equilibrium, in particular with high molecular weight polymers. A few generic situations where polyelectrolytes can act both as stabilizers and coagulants can, however, be demonstrated using simulation techniques.

Introduction - The Dielectric Continuum Model

An aqueous solution containing biological molecules can in a general sense be described as an electrolyte solution. That is, it contains simple ions such as Na^+, K^+, Cl^- etc., but it can also include macromolecules with a net charge significantly different from unity. DNA, proteins and polysaccharides are important examples of natural origin but different synthetic additives can also be described as charged macromolecules, sometimes collectively refered to as *polyelectrolytes*. It is our intention to discuss the interaction/stability of biological polyelectrolytes in a few generic situations, some of which hopefully are of interest for a food chemist.

Despite the progress in computer technology and numerical algorithms during the last decades, it is still not feasible to treat a general solution of charged macromolecules in an atomistic model. This becomes especially clear when we are trying to calculate the interaction between macromolecules and how the interaction can be modulated by other charged species.

The alternative at hand is to use the dielectric continuum model, crudely refered to as *the Primitive Model*. The solvent is then described as a structureless medium solely characterized by its relative dielectric permittivity, ϵ_r. This simplification facilitates both the theoretical treatment and the conceptual understanding of electrostatic interactions in solution. In contrast to its name, it is a very sophisticated approximation, which allows an almost quantitative description of widely different phenomena such as sea water and cement paste! In the Primitive Model we treat all charged species as charged hard spheres and the interaction, between two charges i and j separated a distance r, can be formally described as,

$$u(r) = \frac{Z_i Z_j e^2}{4\pi\epsilon_0\epsilon_r r} \qquad r > d_{hc} \tag{1}$$

$$u(r) = \infty \qquad r < d_{hc} \tag{2}$$

where Z_i is the ion valency, e the elementary charge, ϵ_0 the dielectric permittivity of vacuum and d_{hc} is the hard sphere diameter of the ion. For simplicity, we will in this communication mostly assume it to be the same for all ionic species and equal to 4 Å.

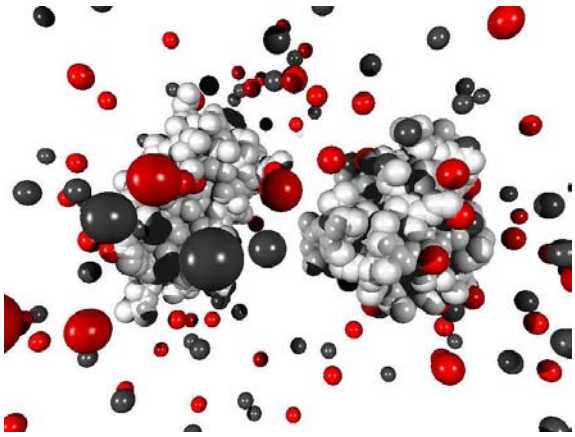


Figure 1: Snapshot from a MC simulation of two proteins. The black and grey spheres illustrate mobile cations and anions, while amino acids are depicted as a white spheres, clustered to form the two proteins. In a simulations, the proteins are displaced along a line and rotated independently. Ions are displaced in all three directions and the whole system is enclosed in sphere of appropriate radius.

These charges can be the small mobile ions in a salt solution, but they can also be the charged groups on a protein or some other macromolecule. The model is schematically depicted in Figure 1 with two macromolecules in a salt solution. We will solve this model exactly using Monte Carlo (MC) simulations or in an approximate way with either the Poisson-Boltzmann (PB) equation or its linearized version, the Debye-Hückel (DH) equation. For an introduction to the DH theory, the reader is recommended to consult the excellent textbook of Hill [1]. Engström and Wennerström [2] has solved the PB equation for a charged surface with neutralising counterions and their paper is a good starting point on this subject. Monte Carlo and other simulations are well described in the textbooks by Allen and Tildesley [3] and by Frenkel and Smit [4]. MC simulations allow us to emphasize where the simple theory is applicable and where a more accurate treatment is needed. The simulations also give an oppurtunity to clarify certain physical mechanisms, providing a deeper understanding of the system at hand.

The paper is arranged as follows:

- A simple electrolyte solution.
- A charged macromolecule in a salt solution.

- The interaction between two charged macromoleules.
- The addition of polyelectrolytes/polyampholytes.
- Attraction due to charge regulation.
- Protein polyelectrolyte complexes.

A Simple Electrolyte Solution

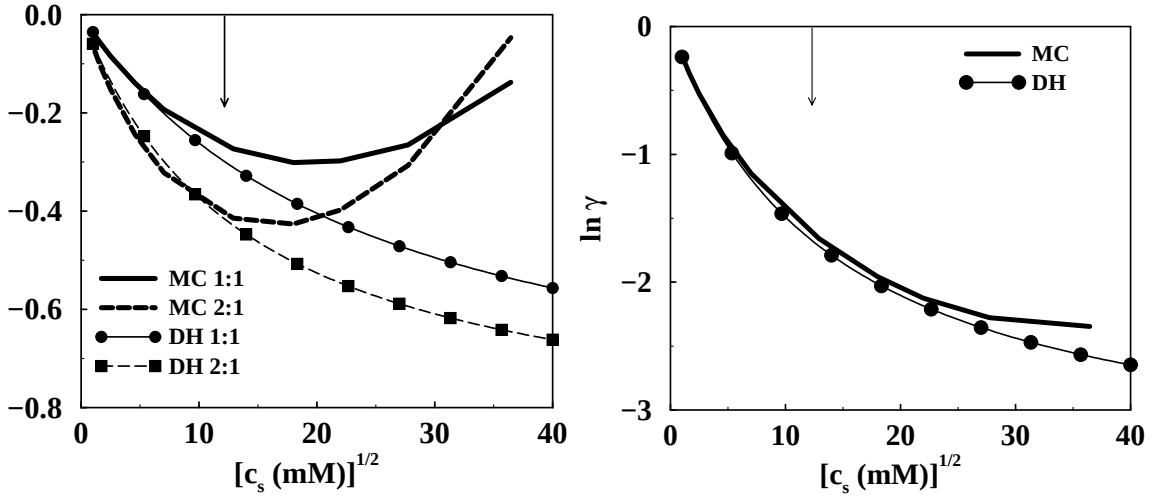


Figure 2: Individual activity factors from MC simulations and from the DH theory with the hard core diameter equal to 4 Å. a) Monovalent ion in a 1:1 and 2:1 salt and b) Divalent ion in a 2:1 salt. The arrows indicate physiological salt condition.

An important property in an electrolyte solution is the activity factor, γ, or excess chemical potential, μ_{ex}, which is a part of the total chemical potential, μ,

$$\mu = \mu_0 + kT \ln c + kT \ln \gamma = \mu_0 + \mu_{id} + \mu_{ex} \tag{3}$$

μ_0 is an uninteresting reference chemical potential and c is the concentration. It is straightforward to calculate γ in a Monte Carlo simulation, but we can also obtain it from the Debye-Hückel approximation,

$$kT \ln \gamma^{DH} = -\frac{Z^2 e^2 \kappa}{8\pi\epsilon_0\epsilon_r(1+\kappa d_{hc})} \tag{4}$$

The important quantity in eq.(4) is the inverse screening length, κ,

$$\kappa^2 = \frac{e^2}{\epsilon_0\epsilon_r kT} \sum_i c_i z_i^2 \tag{5}$$

which is proportinal to the ionic strength. Figure 2 shows how γ varies as a function of salt concentration for two different salts. The accuracy of the simple DH theory is surprisingly good and the main discrepancy comes from the too approximate treatment of the excluded volume effect, *i.e.* the hard core interaction. A knowledge of γ allows us to calculate a number of interesting quantities. For example, we can calculate the dissolution of carbon dioxide in the ocean. The high salt content of the oceans increases the solubility of CO_2, which is apparent from the equilibrium relations,

$$H_2CO_3 \rightarrow HCO_3^- + H^+ \qquad K_1 = \frac{c_{HCO3}c_H}{c_{H2CO3}}\frac{\gamma_{HCO3}\gamma_H}{\gamma_{H2CO3}} = K_1^S \frac{\gamma_{HCO3}\gamma_H}{\gamma_{H2CO3}} \tag{6}$$

$$HCO_3^- \rightarrow CO_3^{2-} + H^+ \qquad K_2 = \frac{c_{CO3}c_H}{c_{HCO3}}\frac{\gamma_{CO3}\gamma_H}{\gamma_{HCO3}} = K_2^S\frac{\gamma_{CO3}\gamma_H}{\gamma_{HCO3}} \tag{7}$$

Note that *thermodynamic* equilibrium constants, K_1 and K_2, are true constants in contrasts to the *stoichiometric* ones, K_1^S and K_2^S. Table 1 presents experimental and simulated activity factors for some salts relevant for sea water. The departure from ideality ($\gamma = 1$) is non-negligible and as a consequence the dissolution of CO_2 in sea water is significantly larger than in fresh water. The excellent agreement between measured and simulated activity factors in Table 1 gives a strong support for the Primitive Model.

Salt	γ_{Exp}	γ_{Sim}
Na_2SO_4	0.37	0.37
K_2SO_4	0.35	0.36
NaCl	0.67	0.67
KCl	0.66	0.66
$CaSO_4$	0.14	0.15

Table 1: Experimental [5, 6] and simulated [7] mean activity factors in sea water at 298 K. The salinity is 3.5 %.

A Charged Macromolecule in a Salt Solution

We can use the activity factors in order to study how the binding of a charged ligand to a charged macromolecule is affected by addition of salt or changes in pH - a change in pH means that the net charge of both ligand and macromolecule can vary. The changes will affect the electrostatic interactions and are almost quantitatively captured by the activity factors. The simplest approach would then be to treat the macromolecule as a charged spherical object and directly apply eq.(4). Let us take the calcium binding to the small chelator 5,5'-Br_2BAPTA as an example [8],

$$Ch + Ca^{2+} \rightarrow ChCa \qquad K = \frac{c_{ChCa}}{c_{Ch}c_{Ca}}\frac{\gamma_{ChCa}}{\gamma_{Ch}\gamma_{Ca}} = K_s\frac{\gamma_{ChCa}}{\gamma_{Ch}\gamma_{Ca}} \tag{8}$$

Since K is a true constant we can write a relation between the stoichiometric binding constants at two different salt concentrations as,

$$K_s^I\frac{\gamma_{ChCa}^I}{\gamma_{Ch}^I\gamma_{Ca}^I} = K_s^{II}\frac{\gamma_{ChCa}^{II}}{\gamma_{Ch}^{II}\gamma_{Ca}^{II}} \tag{9}$$

The charge of the chelator is $-4e$ at neutral pH and it is assumed to have a radius of 7 Å. When calcium is bound to the chelator it is simply modeled by a reduction of the chelator charge from $-4e$ to $-2e$. This simple model captures the salt dependence from 1 mM to 1 M salt. Table 2 shows how the stoichiometric binding constant, K^s, varies with salt concentration. Both simulated and DH results are in excellent agreement with experiment.

A quantitatively more correct alternative is to use the so-called Tanford-Kirkwood (TK) model [9]. The TK model takes the detailed charge distribution into account and solves the electrostatic problem using a variant of the DH approximation. The final result is the free energy for the macromolecule in a salt solution. For not too highly charged macromolecule this is usually a very efficient and reliable approach and the relevant equations are easily evaluated numerically. Figure 3 shows how the calcium binding constant to the small protein calbindin D_{9k} varies with salt concentration [10]. Both simulated and TK results are based on the detailed charge distribution of the protein with the calbindin structure obtained from an x-ray study [11].

c_s (mM)	ΔpK_s^{Exp}	ΔpK_s^{Sim}	ΔpK_s^{DH}
2	0.00	0.00	0.00
10	0.26	0.32	0.32
25	0.64	0.60	0.59
50	0.89	0.85	0.84
100	1.20	1.12	1.11
300	1.58	1.58	1.59
500	1.77	1.79	1.81
1000	1.97	2.05	2.09

Table 2: Shift in the stoichiometric calcium binding constant for the chelator BAPTA. 2 mM salt has been taken as a reference point and the shifts are calculated relative this value.

The agreement between the two theoretical approaches is excellent and so is the comparison with experimental results.

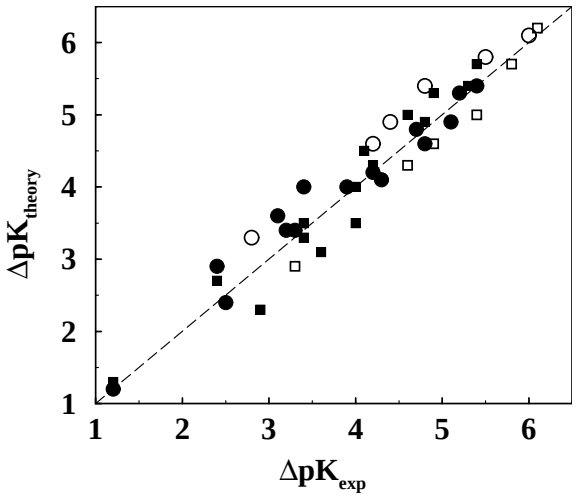


Figure 3: A comparison of experimental and theoretical binding constant shifts for the calcium binding protein calbindin D_{9k}. The electrostatic interactions have been modified by adding salt in the range 2-150 mM and by mutating (neutralizing) charge residues in the protein [10]. The symbols represent different mutations (charge neutralization of acidic residues) and different salt concentrations. Spheres are simulated data and squares are calculated using the TK approach. Filled symbols describe the addition of KCl and open symbols the addition of K_2SO_4. The dashed line corresponds to perfect agreement. The shifts are calculated relative to the native protein at 2 mM salt concentration.

It is interesting to investigate the limitations of the TK approach and one should expect deviations from the simulated values for a really highly charged protein. This is indeed the case and Figure 4 reveals a typical behaviour for the binding of a charged ligand to an oppositely charged macromolecule or particle. That is, when the charge reaches a certain niveau, then the electrostatic response is no longer linear but it approaches an asymptotic value. This means that the binding becomes "saturated" and, for example, a further increase of negatively charged residues in a protein does not lead to an increased binding of calcium.

The electrostatic model in colloid chemistry has always been one with a uniform dielectric permittivity for the whole system, typically chosen to be equal to that of water. In the calculations reported above we have followed this tradition. Obviously, the dielectric permittivity of a protein is different from that of bulk water, but we do not know its exact value and to be more formal, it is not a well-defined quantity. We also note that charged species prefer the high dielectric region - ions dissolve in water and not in oil! Another way to express it is to say that the electric field lines remain in the aqueous phase, hence a small body of low dielectric material

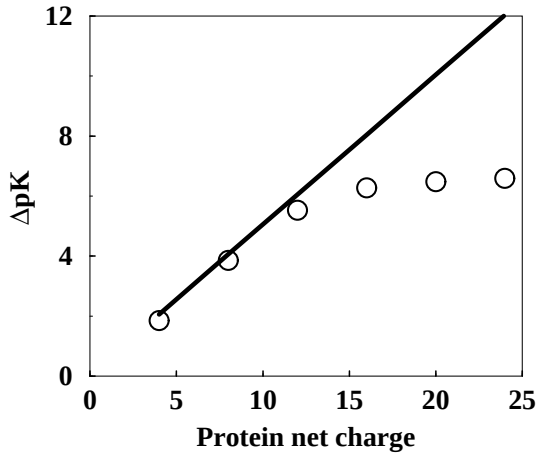


Figure 4: Binding constant shifts as a function of protein net charge - comparison of DH (line) and MC simulations (symbols). The protein is modelled as a sphere of radius is 14 Å with two binding sites close to the surface. The shift refers to a change in salt concentration from 1 to 500 mM. The protein concentration is $20\mu M$ and the binding process involves two divalent ions.

has only a marginal effect on the electrostatic interactions. These conclusions are supported by a wealth of experimental results on colloidal systems.

In biophysics, the opposite paradigm prevails and the low dielectric interior of a protein is usually assumed to be the clue to many properties of biochemical interest. The electrostatic approach is based on the PB or DH equation. A technical feature with the "low dielectric" assumption is that the calculations contain a divergence, which can cause numerical problems. Or, it can be used as a "fitting parameter". The divergence in electrostatic calculations invoking a low dielectric region is apparent in many applications. One very clear such example is the determination of apparent pK_a's in the protein calbindin - see Table 3, which have been determined experimentally by Kesvatera *et al.* [12] and theoretically by Spassov and Bashford [13] using a low dielectric response for the protein. Juffer and Vogel [14] have extended the Debye-Hückel calculations of Spassov and Bashford and allowed for a high dielectric response from the protein. The paper by Kesvatera also contains results from MC simulations using a uniform dielectric response equal to that of water. Obviously the calculations using a low dielectric interior containing charged groups are unable to describe the electrostatic interactions in calbindin and the results are unphysical.

Amino Acid	Exp.	Theory-Spassov	Theory-Juffer	Theory-Kesvatera
Glu-27	6.5	**21.8**	5.2	4.7
Asp-54	3.6	**16.9**	4.8	4.4
Asp-58	4.4	**9.1**	4.8	4.8
Glu-60	6.2	**13.2**	5.6	6.0
Glu-65	5.4	**12.7**	4.6	5.0
Rms.	-	**10.6**	0.93	0.92

Table 3: The apparent pK_a of titrating acidic groups in calbindin D_{9k}. Experimental and various theoretical results. The "low dielectric" results of Spassov and Bashford have been highlighted. Both Spassov-Bashford and Juffer-Vogel have used the DH approximation, but the latter authors have assumed a uniformly high dielectric permittivity in the same way as Kesvatera *et al.*. The rms deviations are given in units of pK_a.

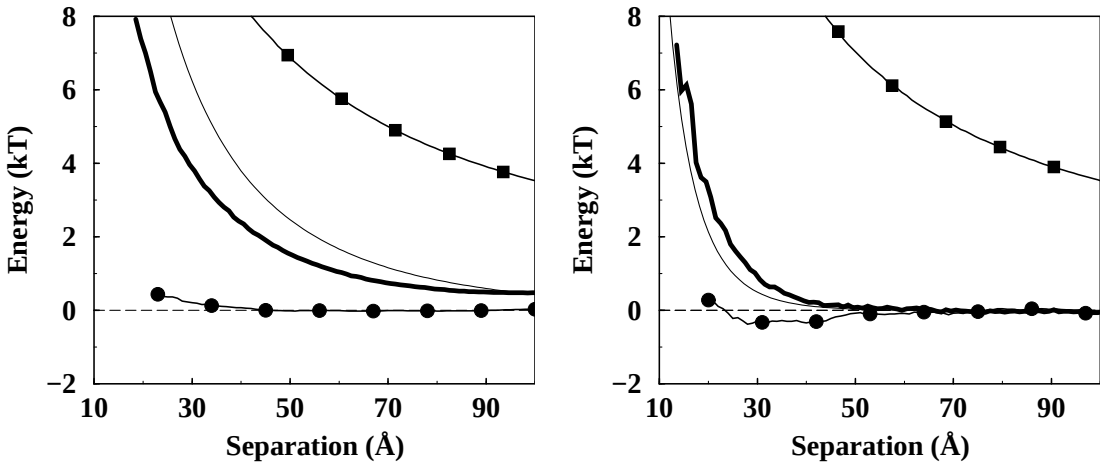


Figure 5: The interaction between two smMLCK peptides at two different salt concentrations; left=4 mM and right=100 mM of a monovalent salt. The smMLCK peptide consists of 15 amino acids and its net charge is +7e. Solid fat lines show the simulated free energy of interaction, while thin solid line is from the screened Coulomb interaction, eq.(10). The thin line with filled circles is the simulated total energy of interaction and the line marked with filled squares is the electrostatic interaction between the charges on the two peptides only.

The Electrostatic Interaction Between Two Proteins

The interaction of two peptides

Calmodulin binds to myosin light chain kinase (MLCK) via a small peptide rich in basic residues. Calmodulin and the peptide forms a complex, which has been isolated and crystallized. We have taken the peptide, smooth muscle MLCK (= smMLCK), from this complex and studied the interaction between a pair. The net charge of smMLCK at neutral pH is close to $+7e$ and the two peptides repel each other, see Figure 5a, that is the free energy of interaction is positive. The unscreened direct electrostatic interaction between the peptides is of course strongly repulsive, but the total electrostatic energy, including the background electrolyte, is essentially zero or sligthly attractive for all separations. Thus, the repulsion between the equally charged peptides is totally dominated by the entropy - the entropy of salt and counterions. An increase in salt concentration from 4 to 100 mM does not change this picture - Figure 5b.

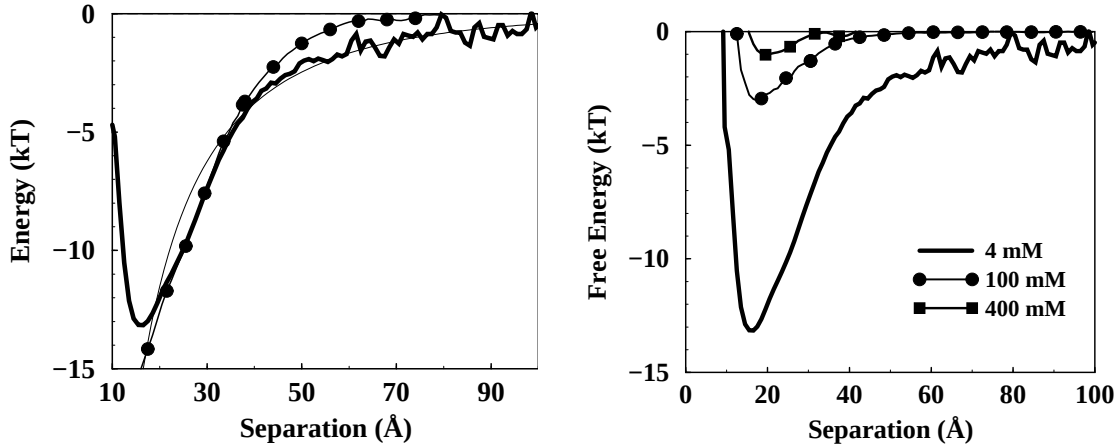


Figure 6: The interaction between an smMLCK peptide and a fragment of calmodulin. The net charge of smMLCK is $+7e$ and the calmodulin fragment has a charge of $-8e$. Left: The salt concentration is 4 mM. The solid fat line is the simulated free energy of interaction, while the thin line is the corresponding screened Coulomb interaction. The thin line with filled circles is the simulated total energy of interaction. Right: The effect of added salt on the free energy of interaction.

A different picture emerges for the interaction of two oppositely charged peptides. Figure 6

shows the free energy of interaction between smMLCK and a peptide section from calmodulin, comprising Glu45-Glu67 with a net charge of $-8e$. The interaction free energy is strongly attractive and so is the total energy. Thus, the attraction is energy driven and the entropy change is in this case only marginal.

The interaction between two charged macromolecules in a salt solution is screened by salt particles and one can derive an expression for their free energy of interaction, $A(r)$, based on the DH approximation,

$$A(r)/kT = l_B Z_1 Z_2 \frac{\exp(-\kappa r)}{r} \tag{10}$$

where we for convenience have introduced the Bjerrum length, $l_B = e^2/4\pi\epsilon_0\epsilon_r kT$. Note that $A(r)$ is a free energy. Figure 5 shows that the screened Coulomb potential is a good approximation and it is semi-quantitatively correct at both salt concentrations.

The results presented here for these peptides is generic and is found in many cases with charged macromolecules or particles. The geometry is not crucial and the same qualitative behaviour is found for both interacting planes and interacting spheres. The screened Coulomb potential captures the change in free energy when the two macromolecules approach each other. It is, however, questionable to partition the screened Coulomb interaction into energy and entropy terms. More elaborate forms of the screened Coulomb potential can be derived [15], where the macromolecular size is taken into account. The comparison in this section has been limited to a uni-uni valent electrolyte and to situations where κ^{-1} is of the same order or larger than the macromolecular dimension. To extend the use of the screened Coulomb potential to multivalent electrolytes usually leads to qualitatively incorrect results - see next section.

The effect of multivalent ions

Above we have shown how the simple theory, the screened Coulomb potential, is capable of an almost quantitative description of the interaction between two charged proteins. This good agreement is limited to systems containing only monovalent counterions. There is a qualitative difference between the interaction of two charged macromolecules in the presence of monovalent and in the presence of multivalent counterions. In the latter case the mean field approximation behind the DH equation breaks down and one has to rely on simulations or more accurate theories like the hypernetted chain equation [16, 17]. The deviation from the mean field description due to *ion-ion correlations* has such a physical origin that the effect should be independent of the particular geometry of the charged aggregates. Clearly there are quantitative differences between cylindrical, spherical or irregularly shaped or flexible charged colloidal species, but the basic mechanism operates in the same way. The importance of ion-ion correlations can be seen from Fig.7, where the free energy of interaction for two charged spherical aggregates has been calculated from an MC simulation. For monovalent counterions there is a monotonic repulsion in accordance with the screened Coulomb equation, eq.(10), but with multivalent counterions or a solvent with a low dielectric permittivity, the entropic double layer repulsion decreases and eventually the correlation term starts to dominate. This phenomenon can be seen as a balance between entropy and energy. For two weakly or moderately charged macromolecules with monovalent counterions, the dominant contribution to the free energy of interaction comes, as we have seen in Figure 5, from a reduction in entropy when the two counterion clouds start to overlap. The energy of interaction is always attractive and is only weakly dependent on the counterion valency. The important difference between a system with monovalent or divalent counterions, is the reduced entropy of the latter due to a lower number density of counterions. Thus, any change that reduces the entropy and/or increases the electrostatic interactions will eventually lead to a net attractive interaction.

This is for a model system with spheres with net charges, but the same mechanism is

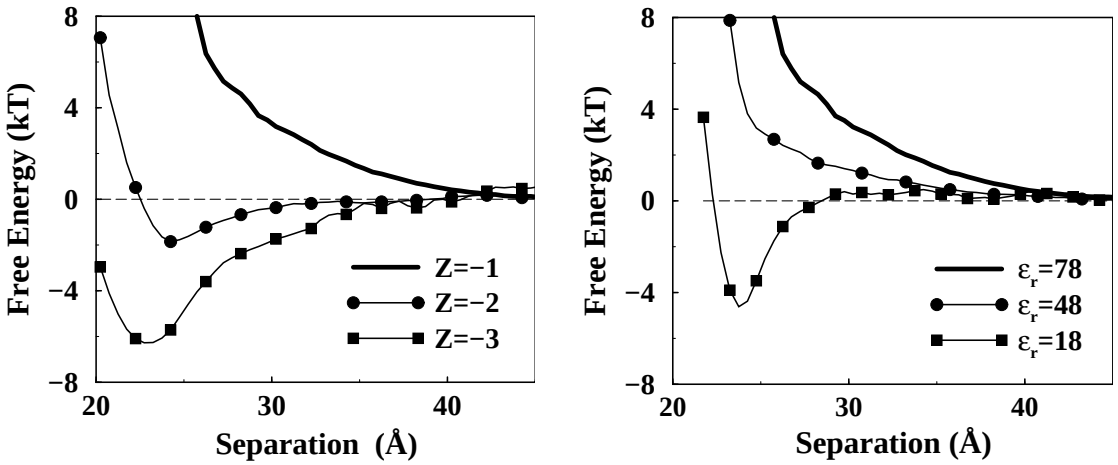


Figure 7: a) The free energy of interaction between two spherical aggregates of radius 10 Å and net charge 24. The system contains no salt but only counterions of different valency. The dielectric permittivity is 78 and the temperature 298 K. b)The same as in a) with monovalent counterions and variation of the relative dielectric permittivity.

also operating between two protein molecules with discrete charge distributions and irregular form[18] and between two DNA molecules [19].

The effect of titrating groups

All proteins and many other macromolecules contain ionizable residues whose ionization status depends on the interaction with other molecules. This means that the electrostatic interaction between two proteins, besides the interaction between their average charges, also will contain terms originating from induced charges. These interactions can be formalized in a statistical mechanical perturbation approach [20, 21] and a protein is characterized not only by its average net charge, but also by its *capacitance*. The induction interaction is important for the interaction of an approximately neutral protein with another charged macromolecule. The protein capacitance is a function of the number of titrating residues and will display maxima close to the pK_a's of the titrating amino acids. In this section we will derive a formal expression for the capacitance. Consider the macromolecules A and B, described by two set of charges $[\mathbf{r}_i, z_i]$ and $[\mathbf{r}_j, z_j]$, respectively. Their mass centra are separated by $\mathbf{R}$, which means that the distance between two charges i and j is given by $r_{ij} = |\mathbf{R} + \mathbf{r}_j - \mathbf{r}_i|$. The average net charge of the distributions need not be zero, that is $\langle Z_A \rangle \neq 0$, where $\langle Z_A \rangle = \langle \sum z_i \rangle$. The free energy of interaction can be written as,

$$\begin{aligned} A(R)/kT &= -\ln \langle \exp(-U(R)/kT) \rangle_0 \approx \langle U(R)/kT \rangle_0 \\ &-\frac{1}{2}\left\langle (U(R)/kT)^2 \right\rangle_0 + \frac{1}{2}[\langle U(R)/kT \rangle_0 + \frac{1}{2}\left\langle (U(R)/kT)^2 \right\rangle_0]^2 \end{aligned} \tag{11}$$

where $U(R)$ is the interaction between the two charge distributions and $\langle ... \rangle_0$ denotes an average over the unperturbed system, which in the present case is the single isolated protein in solution. The interaction energy is simply the direct Coulomb interaction between the two charge distributions,

$$U(R)/kT = \sum_i \sum_j \frac{l_B z_i z_j}{r_{ij}} \tag{12}$$

We can make a Taylor series expansion of U, assuming that $R >> r_i$. This expansion will include ion-ion interaction, ion-dipole interaction, dipole-dipole interaction etc. It will also include charge-induced charge and induced charge-induced charge interactions. Thus, we can

write an approximation to the free energy including all terms of order up to $1/R^2$. Note that the ion-dipole interaction disappears in first order and that the first non-vanishing dipole term, $-l_B^2 Z^2 \mu^2 / 6R^4$ is of order $1/R^4$.

$$A(R)/kT \approx \frac{l_B \langle Z_A \rangle \langle Z_B \rangle}{R} - \frac{l_B^2}{2R^2}(\left\langle Z_A^2 \right\rangle - \langle Z_A \rangle^2)(\left\langle Z_B^2 \right\rangle - \langle Z_B \rangle^2)$$
$$-\frac{l_B^2}{2R^2}((\left\langle Z_A^2 \right\rangle - \langle Z_A \rangle^2) \langle Z_B \rangle^2 + (\left\langle Z_B^2 \right\rangle - \langle Z_B \rangle^2) \langle Z_A \rangle^2) \tag{13}$$

The first term is the direct Coulomb term and the following term is the *induced charge-induced charge* and the last terms are the *charge-induced charge* interactions. Note also that $\langle Z^2 \rangle \neq \langle Z \rangle^2$. If the molecules are identical, that is $\langle Z_A \rangle = \langle Z_B \rangle = \langle Z \rangle$, then the expression simplifies to,

$$A(R)/kT \approx -\frac{l_B \langle Z \rangle^2}{R} - \frac{l_B^2}{2R^2}(\left\langle Z^2 \right\rangle - \langle Z \rangle^2)^2 - \frac{l_B^2}{R^2}(\left\langle Z^2 \right\rangle - \langle Z \rangle^2) \langle Z \rangle^2 \tag{14}$$

and if $pH = pI$, then $< Z >= 0$ and the induced charge-induced charge interaction becomes the leading term,

$$A(R) \approx -\frac{l_B^2 \langle Z^2 \rangle^2}{2R^2} \tag{15}$$

The above equations show that the fluctuating charge of a protein or macromolecule may under certain circumstances contribute significantly to the net interaction. We can define a "charge polarizability" or a capacitance, C, as

$$C = \left\langle Z^2 \right\rangle - \langle Z \rangle^2 \tag{16}$$

With this definition of the capacitance, Eq.(13) can be rewritten in a more compact form,

$$A(R)/kT \approx \frac{l_B \langle Z_A \rangle \langle Z_B \rangle}{R} - \frac{l_B^2}{2R^2}(C_A C_B + C_A \langle Z_B \rangle^2 + C_B \langle Z_A \rangle^2) \tag{17}$$

We can use general electrostatic equations and relate the capacitance to the charge induced by a potential $\Delta\Phi$,

$$Z_{ind} = \frac{C \Delta\Phi}{kT} \tag{18}$$

The capacitance, C, can also be derived from the experimental titration curve. For a single titrating acid the ionization degree, α, can be found in any elementary physical chemistry textbook,

$$\log K = -pH + \log \frac{\alpha}{1-\alpha} \tag{19}$$

Taking the derivative of α wrt to pH gives,

$$\frac{d\alpha}{dpH} = \alpha(1-\alpha) = C \ln 10 \tag{20}$$

where in the second step we have identified the capacitance defined in Eq.(16). We can obtain an approximate value for the capacitance in a protein assuming that there is no interaction between the titrating sites. A protein contains several titrating groups like aspartic and glutamic acid, histidine etc., each with an ideal pK value. Denoting different titrating groups with γ and their number with n_γ, then the total capacitance can be approximated with,

$$C^{ideal} = \frac{1}{\ln 10} \sum_\gamma n_\gamma \frac{10^{pH-pK_\gamma}}{(1+10^{pH-pK_\gamma})^2} \tag{21}$$

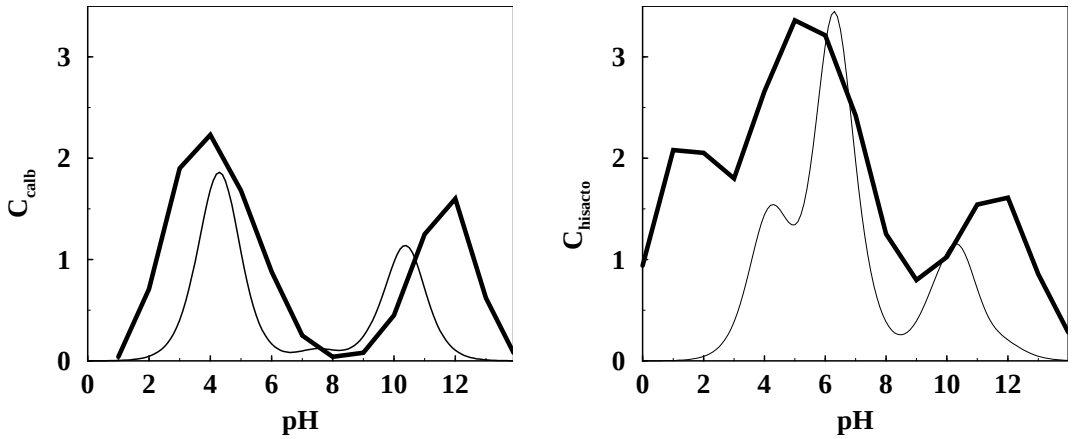


Figure 8: Left: The capacitance for calbindin D_{9k} as a function of pH. The thick solid curve is from a MC simulation of the atomistic model , while the thin solid line is the ideal capacitance calculated from Eq.(21). pI for calbindin is approximately 4.2. Right: The capacitance for hisactophilin as a function of pH with symbols as before. pI for hisactophilin is 7.3

We have calculated the capacitance for a number of proteins with different characteristics in terms of number and type of residues. A MC simulation has to be performed at each pH at given salt and protein concentrations. Unless otherwise stated we have used a salt concentration of 70 mM and a protein concentration of 0.7 mM. Figure 8a shows the capacitance for calbindin. The main difference from the ideal capacitance curve is a strong broadening of two peaks corresponding to the response from acidic and basic residues, respectively. If the protein has a significant net charge, the true curve will also shift away from the ideal one, as is seen for calbindin at high pH.

The protein hisactophilin is of the same size as calbindin, but it has a slightly different capacitance curve, see Figure 8b. The protein contains 31 histidine residues, which is reflected in a large maximum for $C_{hisacto}$ at $pH \approx 5$. The downward shift of the maximum is due to the high positive charge of hisactophilin at low pH. The net charge is +28 at $pH = 3$ and +23 at $pH = 4$. The isoelectric point found from the simulations is $pI = 7.3$, which is in good agreement with experimental estimates.

The electrostatic interaction between two proteins will be dominated by the direct Coulomb interaction provided that the net charge, Z, is sufficiently different from zero. The induced interactions will only play an important role at pH values close to the isoelectric point of one of the proteins - this can be seen from Eq.(17). Figure 9a shows the free energy of interaction between the two proteins calbindin and lysozyme at $pH = 4$, which is close to the isoelectric point for calbindin. At contact there is a significant difference in interaction energy between a model with fixed charges compared to a situation where the proteins are free to adjust their charges.

The difference in free energy between the two models is mainly due to the interaction between the induced charge in calbindin and the permanent charge in lysozyme. This is a typical result and significant effects from charge regulation can be expected when one of the interacting proteins has a large net charge and the other a large capacitance. Following Eq.(17) we can approximate the difference as,

$$(A_{reg}(R) - A_{fix}(R))/kT = \Delta A(R)/kT = -\frac{l_B^2}{2R^2}(C_{calb}C_{lys} + C_{lys}Z_{calb}^2 + C_{calb}Z_{lys}^2) \qquad (22)$$

and Figure 10 shows an almost perfect agreement between the simulated free energy difference and the calculated one according to Eq.(22).

An interesting result is that despite that both calbindin and lysozyme are positively charged at $pH = 4$, there is still an attractive electrostatic interaction between the two. Such an

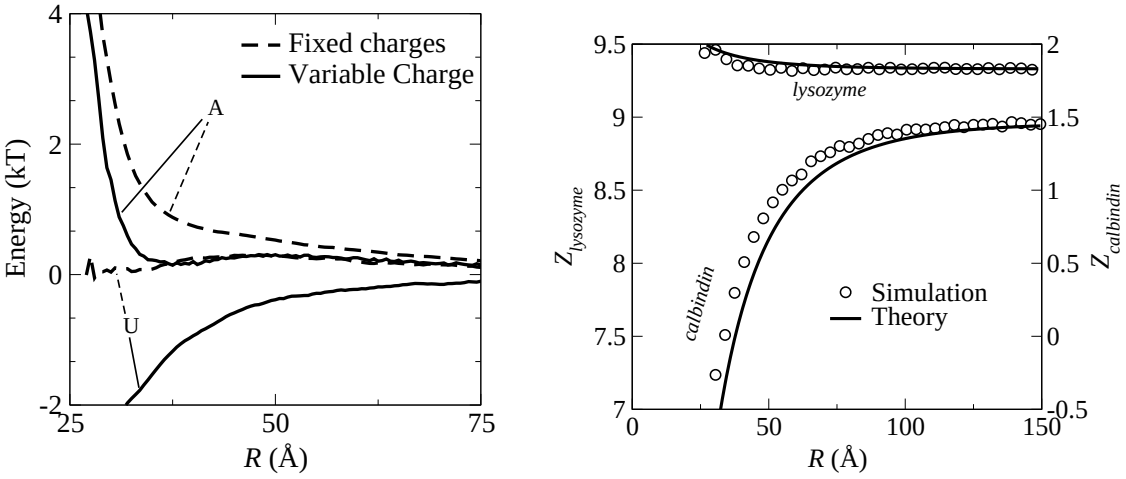


Figure 9: a) The energy amd free energy of interaction between calbindin and lysozyme at $pH = 4$ for a protein model with fixed charges (dashed lines) and one with charge regulation (solid lines). The amino acid model is used and the salt concentration is b) The variation of net charge of calbindin (solid line) and lysozyme (dashed line) as a function of their separation. The simulations are based on the amino acid model. $pH = 4$ and salt concentration is 5 mM.

attraction could of course be due to charge-dipole and/or dipole-dipole interactions, but they do not seem to be important in the present case: the main contribution to the interaction free energy comes from the induced charges. This is further demonstrated in Figure 9b, where one can follow how the net charge of calbindin goes from ≈ 1.4 at infinite separation to ≈ -0.5 at contact between calbindin and lysozyme. We will come back to this issue when discussing protein polyelectrolyte complexation.

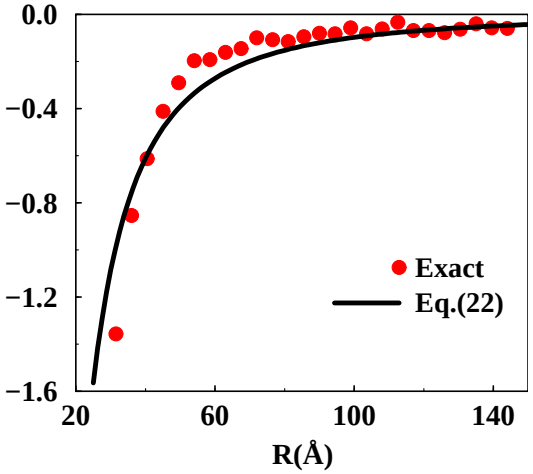


Figure 10: The difference in free energy of interaction between calbindin and lysozyme at $pH = 4$ for a protein model w regulation and one with fixed charges. R is the separation between the mass centra of the two proteins. Symbols denote the simulated difference (see Figure 9) and the solid line is obtained from Eq.(22) with $Z_{calb} = 1.16$, $C_{calb} = 2.23$, $Z_{lys} = 10.2$ and $C_{lys} = 0.88$.

Bridging attraction with polyelectrolytes

Adsorption of a polyelectrolyte to an aggregate is a necessary, but not sufficient condition, in order to attain a modulation of the free energy. It actually has to adsorb to both aggregates in order to form *bridges*, see Figure 11, that can lead to attractive interactions. For highly charged polyelectrolytes and oppositely charged macromolecules, bridge formation is usually a very effective way of destabilization. From simulations and mean field theories, we know that

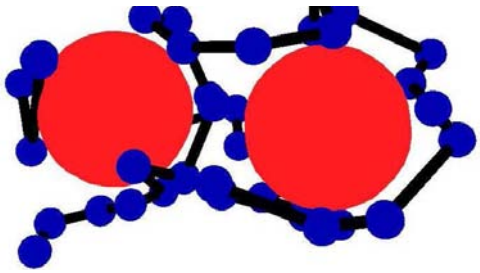


Figure 11: Snapshot from a MC simulation of system containing two charged macromolecules and an oppositely charged polyelectrolyte.

the attraction is rather short ranged and that it typically only extends over distances of the order of the monomer-monomer separation [22, 23, 24, 25]. Figure 12a shows what happens if a polyelectrolyte salt is added to a solution of two charged macromolecules. The double layer repulsion is replaced with a short range attraction with a minimum at a surface-to-surface separation of approximately a monomer-monomer distance.

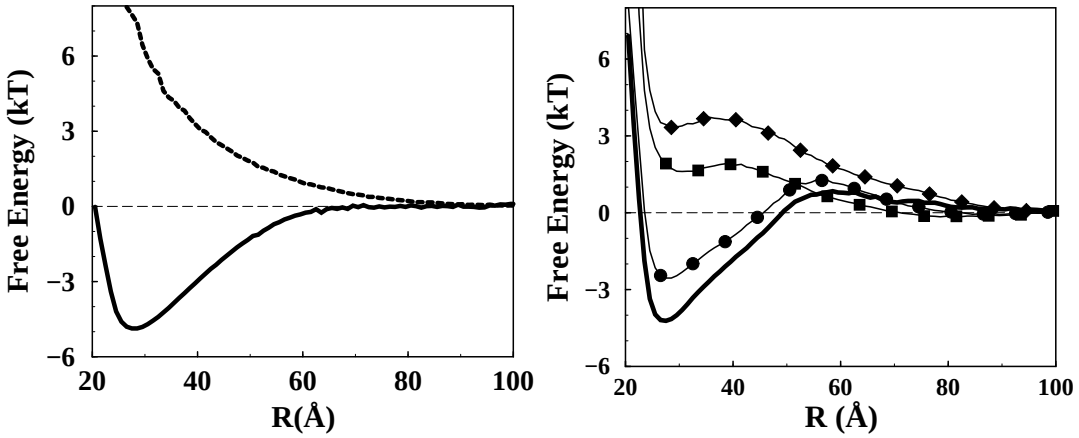


Figure 12: a) The free energy of interaction between two charged spheres as a function of separation, in the presence of a polyelectrolyte salt (solid line) and in the presence of a 1:1 salt (dashed line). The charge of the aggregates is 10e and the radius is 10 Å. The freely jointed polyelectrolyte chain contains 10 charged monomers separated a distance of 6 Å. b) The free energy of interaction between two negatively charged spheres in the presence of a single neutral polyampholyte chain with 40 monomers. The charge topology has been varied and the following notation is used: di-block (solid line with no symbols), tri-block (+10,-20,+10) (circles), tetra-block (+10,-10,+10,-10) (squares) and "reversed" tri-block (-10,+20,-10) (diamonds). Each macromolecule has a charge of +20e and the radius is 10 Å.

A polyelectrolyte adsorbs readily to an oppositely charged macromolecule and in the presence of several charged spheres it becomes of course entropically favourable for the chain to adsorb to more than one sphere. This can only be accomplished at short separations, since the chain tries to avoid placing charges far from the charged aggregates, where the potential is high. Thus, a weakly charged chain, *i.e.* a chain with large separation between the charged monomers, will lead to a more long ranged but weaker attraction. In general, one finds that highly charged systems give rise to fewer, but stronger "bridges", and there will be an optimal choice of polyelectrolyte structure for the attraction between the colloids.

The interaction between charged macromolecules is, from an electrostatic point of view, rather insensitive to the addition of neutral *random* polyampholytes. It is only with block-polyampholytes that the normal double layer repulsion can be decreased in the same way as with oppositely charged polyelectrolytes. The oppositely charged block acts in the same way as an oppositely charged polyelectrolyte. The only complication or constraint is that the equally charged blocks should avoid the aggregates. If the polyampholyte has a net charge, then it behaves qualitatively as a weakly charged polyelectrolyte. A mixing of positively and negatively

charged monomers allows a tailoring of the range and magnitude of the attraction. Figure 12b shows the free energy of interaction between two charged macromolecules with different types of polyampholytes. A naive picture of a tri-block between two adsorbing macromolecules, which seems to be true for neutral block-copolymers, is one where the two ends of the PA chain adsorb to one aggregate each and "pull" them together. Such a structure is quite common in a simulation, but it does not lead to a significant "pulling" force due to the weak force constant of a long segment of negatively charged monomers. Another way to express this is that the free energy gain of adsorbing a PA chain is approximately distance independent for a tri-block of the type (-10,+20,-10). Figure 13 is a snapshot from the simulation and demonstrates this conformation.

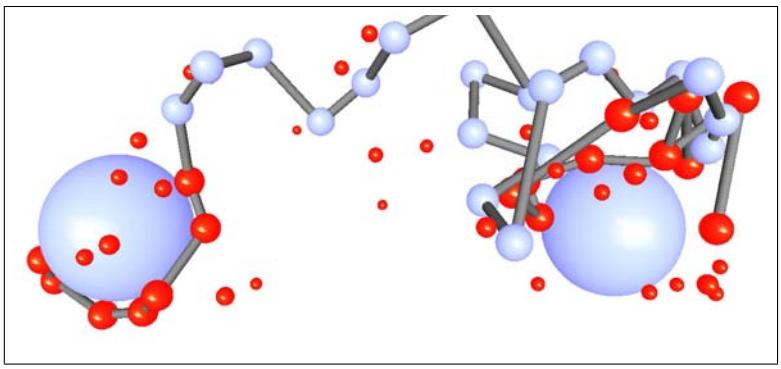


Figure 13: A tri-block, (+10,-20,+10), adsorbing to two negatively charged, $Z = -20$, macroions. Counterions and positively charged monomers are shown in grey and negatively charged monomers in black.

Protein polyelectrolyte complexation

The complexation of polyelectrolytes and proteins is extensively used in pharmaceutics, foods and cosmetics. [26, 27, 28, 29, 30, 31, 32, 33] The subject has been addressed by a number of authors exploring it from experimental measurements [32, 33, 34, 35] to theoretical modeling [36, 37, 38]. The strength of interaction is to a large extent regulated by electrostatic interactions, governed by key parameters such as pH and salt concentration.

A particularly interesting observation [33, 36, 39] is the apparently paradoxical formation of soluble complexes at conditions where the net charges of the protein and the polyelectrolyte have the same sign. Experimental studies of Dubin, Kruif and co-workers [33, 36, 39] have demonstrated this special feature of the polymer/protein complexation. The term complexation "on the wrong side" has been used, meaning that a polyanion forms a complex with a protein at a pH above the isoelectric point of the protein. The molecular interpretation of such studies has focused on the assumption of "charged patches" on the protein surface [33, 40, 34, 37].

A formal way to describe the interaction between oppositely charged patches on two macromolecules is in terms of a multipole expansion. That is, for two neutral protein molecules the leading terms would then be dipole-dipole, dipole-quadrupole, etc. Other electrostatic properties of the protein, however, may be more important and Kirkwood and Shumaker [20] demonstrated theoretically already in 1952 that fluctuations of residue charges in two proteins can result in an attractive force. Recently, we have taken up this idea and used MC simulations and a charge regulation theory in order to explain protein-protein and protein-polyelectrolyte association in a purely electrostatic model [21, 41]. A charge regulation mechanism has also been suggested by Biesheuvel and Cohen-Stuart [42].

We can use simulated capacitances and dipole moments in order to analytically calculate the ion-induced charge and ion-dipole contributions to the interaction free energy according

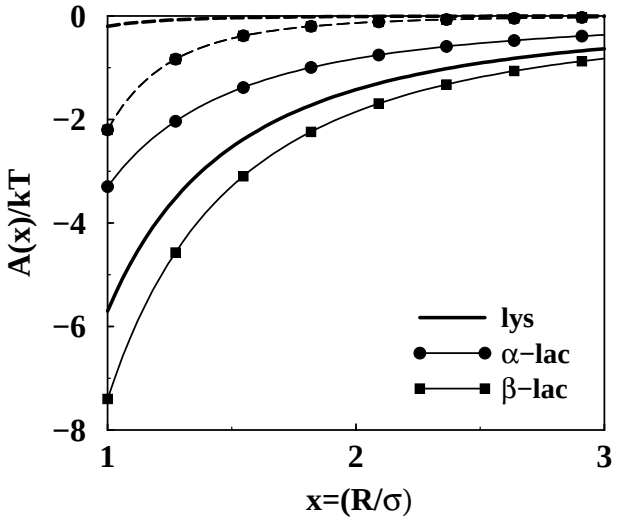


Figure 14: The contribution to the free energy of interaction from the charge - induced charge term (solid lines) and the ion-dipole term (dashed lines). Lines without symbols describe lysozyme, filled circles refer to α-lactalbumin and filled squares refer to β-lactoglobulin, respectively. The free energies are calculated from eq.(17) using simulated capacitances and dipole moments. Note that the ion-dipole terms for α-lactalbumin and β-lactoglobulin coincide.

to eq.(17). The results indicate that the regulation term is by far the most important term for lysozyme, while for α-lactalbumin and β-lactoglobulin the two terms are of comparable magnitude. The curves in Figure 14 should of course be regarded as qualitative and not quantitative. However, they still give, as will be seen below, a correct picture of the behaviour of the three proteins. The contact separation has been defined as the protein radius plus the polyelectrolyte radius, $R_p + R_{pe}$. The latter has been chosen as half the end-to-end separation of the corresponding neutral ideal polymer. Both the protein and polyelectrolyte radii are approximate, but even with a rather generous variation of these values the general picture of Figure 14 will remain the same. The regulation term decays slower than the ion-dipole term, which means that it will gain in relative importance at larger separation, see Figure 14. This means that even if the two terms are comparable at contact, the regulation term can still dominate the contribution to, for example, the second virial coefficient.

We have performed four different simulations for each protein-polyelectrolyte complex:

- A: the "neutral" protein, that is all charges have been set to zero.
- B: the protein with fixed charges at each amino acid residue.
- C: the protein with an ideal dipole at its center of mass.
- D: the protein with titrating amino acid residues.

The first set of simulations (A) describes only the shape of the protein and the free energy of interaction is of course everywhere repulsive. The second set of simulations (B) uses fixed fractional charges on all residues, which has been determined in a separate simulation of the isolated protein at the appropriate pH. In the next set (C), the charge distribution of the protein is replaced by an ideal dipole. In the fourth and final set (D) the amino acids are allowed to titrate and this simulation contains all electrostatic contributions including the ion-induced charge term. The difference between set B and C describes the importance of higher order electrostatic moments, quadrupole, octupole etc. in the protein, while a comparison of sets C and D reveals the effect of the regulation mechanism.

The calculated free energy of interaction, $A(R)$, for the three proteins at their respective pI all show a clear minimum, see Figure 15. The relative strength of the minima are in qualitative

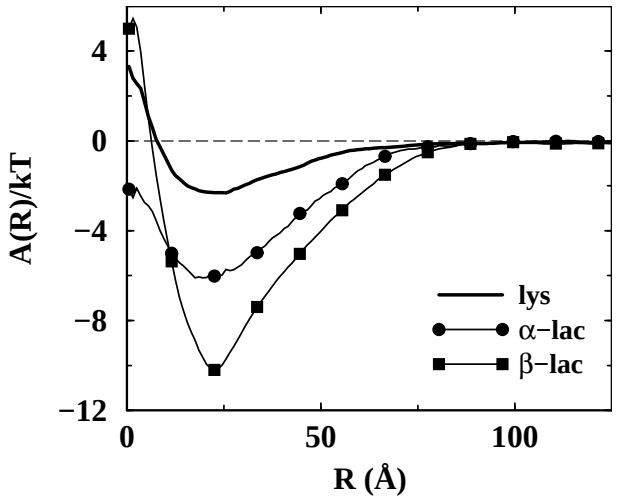


Figure 15: The free energy of interaction between the centers of mass of the protein and the polyelectrolyte at low salt concentration obtained from MC simulations with Model D. The curves have been calculated at the respective isoelectric points for lysozyme (no symbols), α-lactalbumin (filled circles) and β-lactoglobulin (filled squares).

agreement with perturbation calculations, *cf.* Figure 14, while the actual numbers are approximately half the values predicted by second order perturbation theory. The minima appear at roughly the same separation despite the fact that β-lactoglobulin is more than twice as big as the two others. This can be explained by the elongated form of the former, which also results in a more long ranged attraction. The separation R can approach zero, which corresponds to a situation where the polyelectrolyte wraps around the protein. Note, however, that $A(0)$ is repulsive indicating that the "wrapping" of the chain around the proteins is an entropically unfavourable structure.

The attractive minimum in the protein-polyelectrolyte complex is reduced upon addition of salt [37] and we can use the minima of $A(R)$ in Fig.15 in order to estimate the critical ionic strength. Assuming that the salt screening can be described by simple Debye-Hückel theory and that the complex can be defined as dissolved when the interaction is less than kT, we get the following relation,

$$\exp(-2\kappa R_{min})|A(R_{min})| \leq kT \tag{23}$$

The factor of two in the exponent comes from the fact that the second order terms dominate the interaction. Following this recipe we find that approximately 10 and 20 mM salt is sufficient to dissociate the α-lactalbumin and β-lactoglobulin polymer complexes, respectively.

Thus, we have shown that a polyanion can form a complex with a neutral protein molecule. Next, we will make a numerically more rigorous partitioning of contributions to the free energy of interaction shown in Fig. 15. The minimum for lysozyme is solely due to charge regulation, Fig. 16a. If the charge distribution on lysozyme is considered fixed, then the polyanion-lysozyme interaction is essentially everywhere repulsive. Replacing the detailed charge distribution with an ideal dipole at the mass center has a small effect on the free energy. This means that the ion-dipole interaction gives a very small attractive contribution, while the effect from higher order moments is negligible.

As shown in Fig. 16b, the polyanion interacts more strongly with α-lactalbumin than with lysozyme. For α-lactalbumin the regulation term increases the depth of the minimum from approximately 4 to 6 kT. An interesting effect is that the dipolar protein shows a stronger interaction than the protein with a detailed but fixed charge distribution. This means that the ion-quadrupole interactions etc. add repulsive contributions to the interaction.

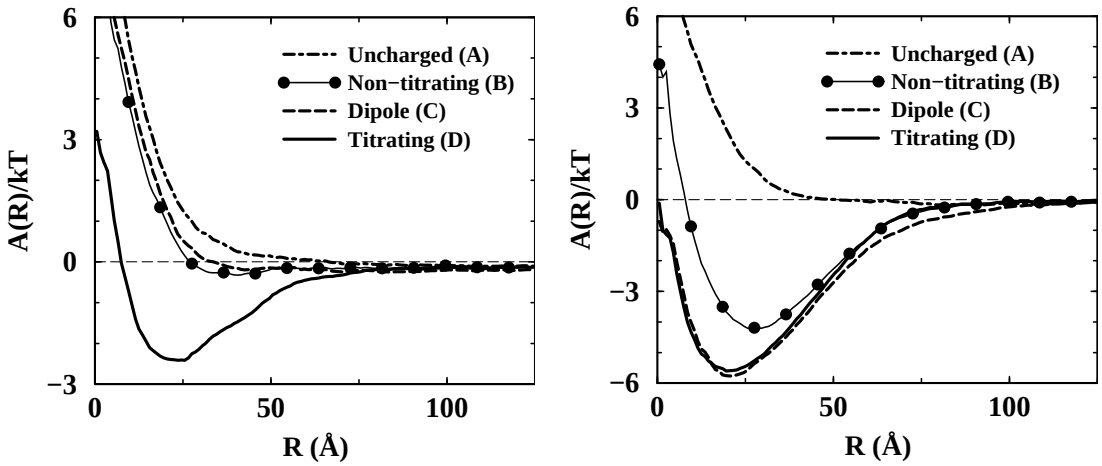


Figure 16: The free energy of interaction between the centers of mass of lysozyme and the polyanion. The free energies have been calculated at pI and the four curves correspond to the different cases mentioned in the text. a) Lysozyme and the polyanion and b) α-lactalbumin and the polyanion.

Conclusion

We have demonstrated a few generic situations where electrostatic interactions between charged macromolecules seem to play an important role. With Monte Carlo simulations we can obtain the exact answer within the given interaction model, which allows us to test the validity of approximate theories. Many biochemical systems are comparatively weakly charged, in contrast to many inorganic systems, and simple theories based on the Debye-Hückel approximation give accurate answers. The long range character of the Coulomb interaction usually means that the geometry and detailed distribution of the charged groups are less important for the interaction of two charged macromolecules.

References

[1] T. L. Hill, *An Introduction to Statistical Thermododynamics* (Dover Publications Inc., New York, 1986).

[2] S. Engström and H. Wennerström, J. Phys. Chem. **82**, 2711 (1978).

[3] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1989).

[4] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, San Diego, 1996).

[5] R. F. Platford and T. Dafoe, J. Mar. Res. **23**, 63 (1965).

[6] R. F. Platford, J. Mar. Res. **23**, 55 (1965).

[7] M. Lund, B. Jönsson, and T. Pedersen, Mar. Chem. **80**, 95 (2003).

[8] L. A. Svensson, E. Thulin, and S. Forsén, J. Mol. Biol. **223**, 601 (1992).

[9] C. Tanford and J. G. Kirkwood, J. Am. Chem. Soc. **79**, 5333 (1957).

[10] B. Svensson, B. Jönsson, C. E. Woodward, and S. Linse, Biochemistry **30**, 5209 (1991).

[11] D. M. E. Szebenyi and K. Moffat, J. Biol. Chem. **261**, 8761 (1986).

[12] T. Kesvatera, B. Jönsson, E. Thulin, and S. Linse, Proteins **45**, 129 (2001).

[13] V. Spassov and D. Bashford, Protein Sci. **7**, 2012 (1998).

[14] A. H. Juffer and H. J. Vogel, Proteins **41**, 554 (2000).

[15] B. Beresford-Smith. *Some aspects of strongly interacting colloidal interactions*. PhD thesis, Australian National University, Canberra, (1985).

[16] R. Kjellander and S. Marčelja, Chem. Phys. Letters **112**, 49 (1984).

[17] R. Kjellander and S. Marčelja, J. Chem. Phys. **82**, 2122 (1985).

[18] M. Lund and B. Jönsson, Biophys. J. **85**, 2940 (2003).

[19] S. M. Mel'nikov, M. O. Khan, B. Lindman, and B. Jönsson, J. Am. Chem. Soc. **121**, 1130 (1999).

[20] J. G. Kirkwood and J. B. Shumaker, Chemistry **38**, 863 (1952).

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

[22] C. E. Woodward, B. Jönsson, and T. Åkesson, J. Chem. Phys. **89**, 5145 (1988).

[23] R. Podgornik, J. Phys. Chem. **95**, 5249 (1991).

[24] J. Ennis, L. Sjöström, T. Åkesson, and B. Jönsson, Langmuir **16**, 7116 (2000).

[25] L. Sjöström and T. Åkesson, J. Coll. Interface Sci. **181**, 645 (1996).

[26] C. Schmitt, C. Sanchez, S. Desobry-Banon, and J. Hardy, Crit. Rev. Food Sci. Nutr. **38**, 689 (1998).

[27] J. L. Doublier, C. Garnier, D. Renard, and C. Sanchez, Curr. Opin. Colloid. Interface Sci. **5**, 202 (2000).

[28] S. Zancong and S. Mitragotri, Pharm. Res. **19**, 391 (2002).

[29] G. Jiang, B. H. Woo, F. Kangb, Jagdish Singhb, and Patrick P. DeLuca, J. Controlled Release **79**, 137 (2002).

[30] M. Simon, M. Wittmar, U. Bakowsky, and T. Kissel, Bioconjugate Chem. **15**, 841 (2004).

[31] J. A. Hubbell, Science **300**, 595 (2003).

[32] M. Girard, S. L. Turgeon, and S. F. Gauthier, J. Agric. Food Chem. **51**, 6043 (2003).

[33] C. G. de Kruif, F. Weinbreck, and R. de Vries, Curr. Opin. Colloid Interface Sci. **9**, 340 (2004).

[34] E. Seyrek, P. L. Dubin, C. Tribet, and E. A. Gamble, Biomacromolecules **4**, 273 (2003).

[35] R. Hallberg and P. L. Dubin, J. Phys. Chem. B **102**, 8629 (1998).

[36] K. R. Grymonpré, B. A. Staggemeier, P. L. Dubin, and K. W. Mattison, Biomacromolecules **2**, 422 (2001).

[37] R. de Vries, J. Chem. Phys. **120**, 3475 (2004).

[38] F. Carlsson, P. Linse, and M. Malmsten, J. Phys. Chem. B **105**, 9040 (2001).

[39] R. de Vries, F. Weinbreck, and C. G. deKruif, J. Chem. Phys. **118**, 4649 (2003).

[40] T. Hattori, R. Hallberg, and P. L. Dubin, Langmuir **16**, 9738 (2000).

[41] F. L. B. da Silva, M. Lund, B. Jönsson, and T. Åkesson, J. Phys. Chem. B **110**, 4459 (2006).

[42] M.A. Cohen Stuart P.M. Biesheuvel, Langmuir **20**, 2785 (2004).

MOLEKYLÆR MODELLERING

Monte Carlo simulering – fra proteiner til cement

Vi har kastet kittel og murerarbejdstøj i ringen og vil i stor magelighed studere proteiner, cement, DNA og havvand. Kan biomolekyler virkelig beskrives med simple, hårde kugler? Og hvad siger vores kollegaer til sådanne fantasifulde tilnærmelser? Læs mere her, hvor der gives en introduktion til molekylær modellering

Af Mikael Lund og Bo Jönsson - Teoretisk Kemi, Lunds Universitet

Hvad har proteiner og DNA tilfælles med havvand og cement? Ikke meget, vil nogen måske mene, og der er da ved første øjekast heller ikke mange ligheder. Men ser man lidt nærmere efter, viser det sig, at de deler den samme egenskab, nemlig at de er stærkt ioniske. Og netop ladninger i form af frie ioner og molekyler er uhyre vigtige for de fysiske og kemiske egenskaber. Et synligt eksempel er, når mælk og eddike blandes: pH-sænkningen foranlediger protonering af proteinet kasein, så det går fra at være negativt ladet til omtrent neutralt. Der er nu ikke længere nogen frastødende kraft mellem kaseinmolekylerne, og de udfælder derfor.

Den klassiske teori for ioniske opløsninger blev grundlagt i starten af 1900-tallet. Den nok mest berømte er Debye og Hückels vellykkede teori til beregning af aktivitetskoefficienter. Modellen de anvendte, behandler solventet som strukturløst (et dielektrisk kontinuum), mens ioner antages at være ladede, hårde kugler. Parvekselvirkningsenergien mellem to kugler kan derfor beskrives med Coulombs lov:

$$E_{12} \propto z_1 z_2 / \varepsilon r_{12}$$

hvor z er ladningstallene (valens), r den indbyrdes afstand og ε er solventets dielektricitetskonstant. For at beskrive systemets termodynamiske egenskaber skal man summere energierne for *alle* tænkbare molekylorienteringer og -positioner (mikrotilstande), hvorved den eftertragtede tilstandssum kan skrives som:

$$Z = \sum e^{-E_{tot}/kT}$$

Det bliver hurtigt et ubehageligt regnestykke, og for at løse problemet måtte Debye og Hückel ty til både linearisering og middelfeltsteori (Poisson-Boltzmann). Disse tilnærmelser har betydning for opløsninger med multivalente ioner og/eller høj ionstyrke, og Debye-Hückel-teorien er da også kendt for at fejle for sådanne systemer.

I dag er sagen en ganske anden. Tilstandssummen - som er alfa og omega i statistisk mekanik - kan opnås ved at lade en eller mange computere udregne energierne for de utallige partikelkombinationer. For at beskrive en 0.1 M NaCl-opløsning kan man f.eks. tage 200 kugler - halvdelen positivt ladede, resten negative - og lægge dem i en virtuel, kubisk kasse med en sidelængde på 119 Ångstrøm (regn selv efter!) og dielektricitetskonstanten 80, svarende til vand. I en Monte Carlo (MC)-simulering kan mikrotilstandene herefter udforskes ved at udføre *tilfældige* kugleflytninger. Systemets energi udregnes før og efter hver flytning, og forskellen afgør om gerningen skal godkendes eller forkastes. Udvælgelsesmetoden er snedigt opbygget, så de konfigurationer, der bidrager mest til tilstandssummen (lavest energi), besøges oftere end dem, der giver et ubetydeligt bidrag (høj energi). Efter en stund med sådanne rokeringer opnår systemet termodynamisk ligevægt, og man kan nu - under stadig flytning - påbegynde undersøgelser af både mikro- og makroskopiske egenskaber. Slutresultatet er *eksakt* for den opsatte model, og simuleringen betegnes populært som et »computereksperiment«.

Havvand og drivhuseffekten

Lad os tage opløsningen af kuldioxid i havvand som et eksempel,

$$CO_2(g) + H_2O(l) \leftrightarrow CO_2(aq) \leftrightarrow H_2CO_3(aq)$$

$$H_2CO_3(aq) \leftrightarrow H^+(aq) + HCO_3^-(aq)$$

$$HCO_3^-(aq) \leftrightarrow H^+(aq) + CO_3^{2-}(aq)$$

Med tilhørende dissociationskonstanter,

$$K_1 = \frac{c_{H^+} c_{HCO_3^-}}{c_{H_2CO_3}} \cdot \frac{\gamma_{H^+}\gamma_{HCO_3^-}}{\gamma_{H_2CO_3}}$$

$$K_2 = \frac{c_{H^+} c_{CO_3^{2-}}}{c_{HCO_3^-}} \cdot \frac{\gamma_{H^+}\gamma_{CO_3^{2-}}}{\gamma_{HCO_3^-}}$$

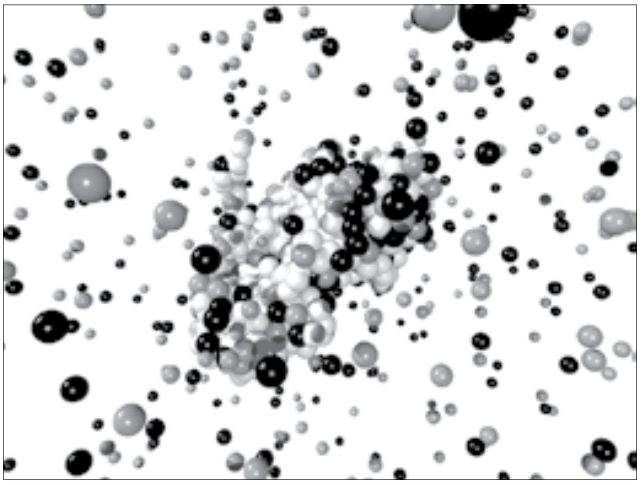
I fortyndede opløsninger, kan alle aktivitetskoefficienter, γ sættes til 1, men i havvand er ionstyrken tæt på 1 M og de ovenstående γ-brøker er hhv. 0,3 og 0,04. Disse afvigelser har naturligvis betydning for kuldioxids opløselighed i verdenshavene, og det er derfor vigtigt, at klimamodeller inkluderer præcise aktivitetskoefficienter. Sådanne målinger er dog ofte problematiske (T. Pedersen, Dansk Kemi 2000, nr. 9 og 10), og MC-simuleringer kan her hjælpe til [1]. Tabel 1 viser hhv. målte og beregnede middelaktivitetskoefficienter i kunstigt havvand[1].

	$\gamma_{\pm}$(Målt)	$\gamma_{\pm}$(Simuleret)
Na_2SO_4	0,374±0,016	0,374
K_2SO_4	0,352±0,018	0,360
KCl	0,645±0,008	0,645
$CaSO_4$	0,136	0,152

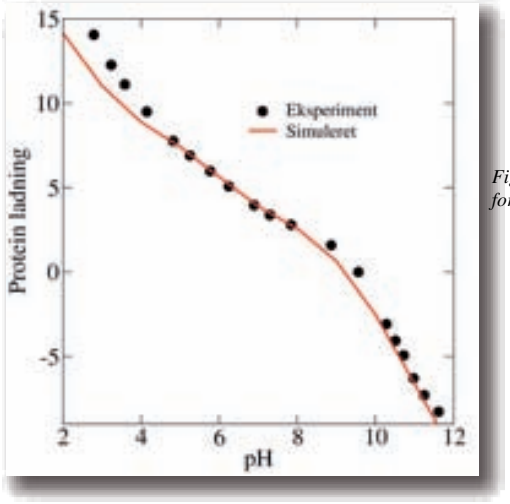
Tabel 1. Middelaktivitetskoefficienter for en række salte i havvand. Salinitet: 35 promille.

Proteiner som kugler:

Det er nu skitseret, hvordan ioniske opløsninger kan beskrives med kugler i en kasse, men der er ingen grund til at stoppe der. Kuglesystemet kan let bruges til at opbygge større molekyler som proteiner, DNA, miceller og polymerer. Figur 1 viser et billede fra en simulering af et protein i en vandig saltopløsning, hvor målet var at bestemme støkiometriske pK_a-værdier for titrerbare residier [2]. Som det ses i figur 2, er titrerkurverne ofte i særdeles god overensstemmelse med det eksperimentelt bestemte resultat.



Figur 1. Monte Carlo-simulering af et protein i en saltopløsning, hvor alle atomer og ioner er beskrevet med kugler.

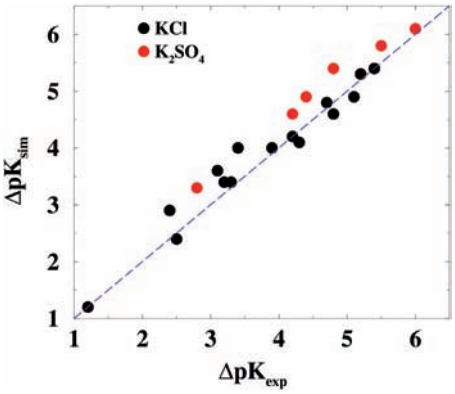


Figur 2. Titreringskurve for ribonuclease A.

Der findes mange calciumbindende proteiner i biologiske systemer - et af dem er calbindin D_{9k}, der binder to calciumioner. Bindingsprocessen er grundigt studeret med forskellige spektroskopiske metoder og ved at variere saltkoncentration, salttype og endda ved at mutere aminosyrer i proteinet. Man kan - næsten uden sved på panden - foretage tilsvarende eksperimenter på en computer, og som det fremgår af figur 3 er overensstemmelsen mellem teori og praksis særdeles god. Bemærk at systemet udelukkende er opbygget af små kugler.

Ion-ionkorrelation

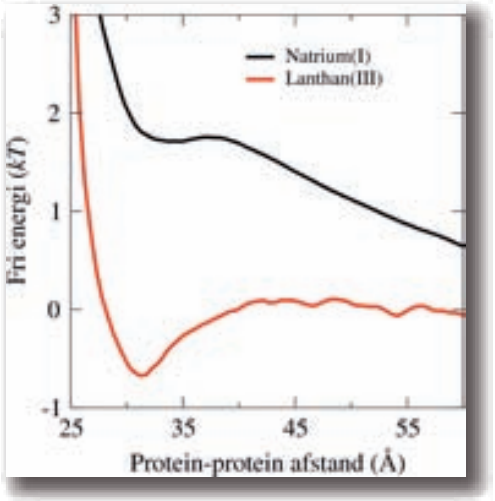
Det er også muligt at studere to eller flere proteinmolekyler, hvorved opløsningens stabilitet kan følges under forskellige forhold som pH, saltkoncentration og valens [3]. Det har betydning



Figur 3. Målte og simulerede skift i Ca^{2+}-bindingskonstanter til calbindin D_{9k} ved forskellige pH, saltkoncentrationer og proteinmuteringer.

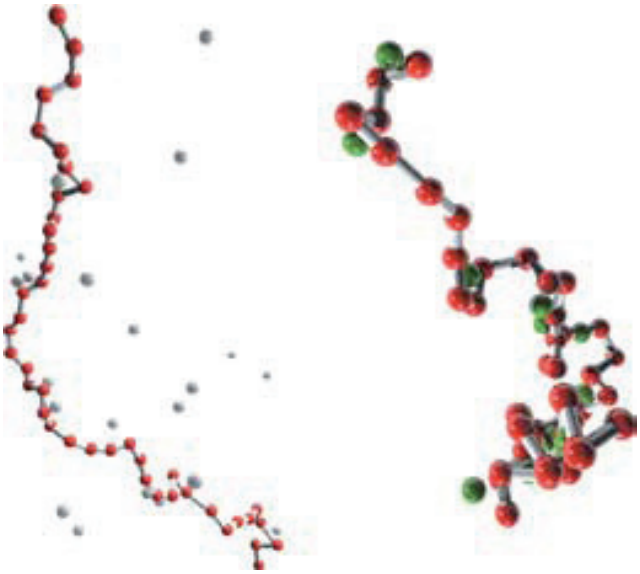
for protein-krystallografer, som ofte prøver sig frem til de forhold, der giver fine krystaller. Følgende eksempel illustrerer en interessant mekanisme, der kan have stor indflydelse på protein-protein-vekselvirkninger.

Proteinet calbindin har ved pH 7 en ladning på -7, og den vandige opløsning er derfor ganske stabil, da molekylerne frastøder hinanden. Tilsættes NaCl, vil frastødningen gradvist mindskes, men man får *ikke* et fri energi minimum, der forårsager udfældning. Tilsættes derimod meget små mængder $LaCl_3$, udfælder proteinet straks. Det skyldes *ionkorrelationer*, som ikke kan studeres med Poisson-Boltzmann-metoder, men kræver en eksplicit beskrivelse af frie ioner. Meget fundamentalt, mindskes energien i et elektroneutralt system (proteiner og modioner) *altid*, når ladningerne flyttes tættere sammen; elektrostatikken bidrager med en tiltrækkende kraft mellem molekylerne - til trods for at de har samme ladning. Det strider måske mod vanlig intuition, men er helt analogt med en saltkrystal, hvor de skiftende kat- og anioner sørger for gitterenergien. For at neutralisere to calbindinmolekyler kræves enten 14 natrium(I)-ioner eller ca. 5 lanthan(III)-ioner. Disse skal inddrages fra opløsningen, hvilket - da antallet af frihedsgrader mindskes - koster *entropi*, og de giver derfor et frastødende bidrag til protein-protein-vekselvirkningen. For Na^+ er dette bidrag større end den elektriske tiltrækning, og proteinerne frastøder hinanden. For La^{3+} kræves der tre gange færre partikler, og prisen (entropi) er derfor mindre. Da trivalente ioner samtidig giver anledning til stærke parvekselvirkninger, dominerer den elektrostatiske energi, og proteinerne udfælder (figur 4).



Figur 4. Ændring i fri energi som funktion af afstand mellem to calbindinmolekyler ved pH 7 og med hhv. mono- og trivalente modioner.

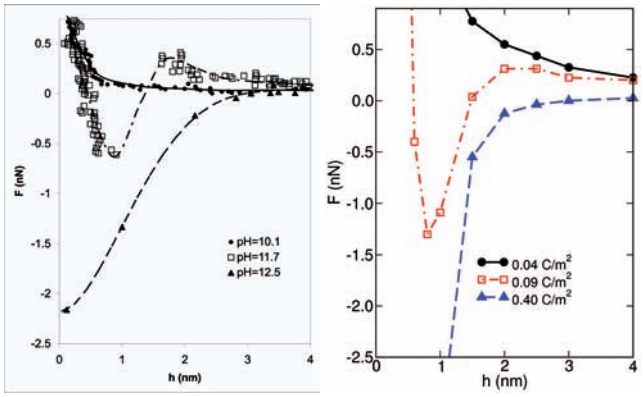
MOLEKYLÆR MODELLERING



Figur 5. Simuleret DNA-streng med hhv. monovalente (venstre) og trivalente (højre) modioner.

Cement og DNA:
Ionkorrelationer er vigtige for systemer med høj ladning og di- eller trivalente ioner. Til trods for deres umiddelbare forskelle er både DNA og cement gode eksempler på sådanne systemer. DNA er en negativ ladet polymer, og ifølge gængse teorier burde det derfor være stift som en pind. Og det er da også, hvad man observerer, hvis der kun er Na^+ til stede. I en biologisk celle findes der dog både tri- og tetravalente kationer (spermidin og spermin), der medfører, at DNA-strengen krøller sig sammen som vist i figur 5. Det er således ionkorrelationer, der gør det muligt at pakke de meget lange DNA-strenge i cellen [4].

Cement har været kendt siden antikken, og bygninger som Pantheon og Pont du Gare fryder den dag i dag vores øjne. Den antikke cement adskiller sig ikke nævneværdigt fra moderne Portlandcement, der har været i brug siden 1800-tallet. Hovedbestanddelen er tricalciumsilicat, som kan opløses i vand under dannelse af en stærk basisk opløsning. Efter en stund er ionkoncentrationen øget tilstrækkelig til at udfælde et andet mineral, calciumsilicathydrat (C-S-H). Udfældningen sker i form af nanostore lameller, der under basiske forhold har en meget høj ladning pga. titrerende silanolgrupper. Da der samtidig forefindes calciumioner, leder ionkorrelationer til en tiltrækkende vekselvirkning mellem de højt ladede C-S-H-lameller [5]. At simuleringerne virkelig formår at beskrive cementkohesionen, bevises i figur 6.



Figur 6. Målt (venstre) og simuleret (højre) kraft mellem cementpartikler ved forskellige pH.

Tak til Thorvald Pedersen for venlig opfordring til dette indlæg - og for i første omgang at igangsætte samarbejdet på tværs af sundet.

E-mail-adresser
Mikael Lund: mlund@mac.com
Bo Jönsson: bo.jonsson@teokem.lu.se

Fodnote
[1] En vandig opløsning indeholdende de vigtigste komponenter af havvand (Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^- og SO_4^{2-}). Aktivitetskoefficienter for andre species antages udelukkende at afhænge af disse »hovedingredienser«.

Referencer:
1. Lund, M., Jönsson B. og Pedersen, T., *Marine Chemistry* 80, 95-101 (2003).
2. Lund, M. og Jönsson, B. *Biochemistry* 44, 5722-5727 (2005).
3. Lund, M. og Jönsson, B. *Biophysical Journal* 85, 2940-2947 (2003).
4. Khan, M. O., Melkinov, S. M., Jönsson, B. *Macromolecules* 32, 8836-8840 (1999).
5. Jönsson, B., Wennerström, H., Nonat, A., Cabane, B. *Langmuir* 20, 6702-6709 (2004)

Nyt om...

...uran

Margaret-Jane Crawford og medarbejdere ved universitetet i München har fremstillet et ammoniumsalt af uranpolyazid

$$U(N_3)_7^{3-}$$

Crawford brugte propannitril

$$CH_3CH_2CN$$

som opløsningsmiddel

Gruppen har senere fremstillet analoge heptaazidforbindelser af molybden og wolfram.

Fremstilling af disse tungmetalpolyazider bliver aldrig brugt til kemiske småforsøg, det er alt for farligt: Opvarmes et af disse stoffer til stuetemperatur, dekomponerer det under eksplosion. Henlægges det i atmosfærisk luft i laboratoriet, bryder det i brand.

Bos

Litteratur:
S.K. Ritter 2005: New Wrinkles in Uranium Chemistry. Synthesis of novel uranium-nitrogen compounds. *Chemical & Engineering News.* November 28: 31