Inverse Monte Carlo Simulation of Biomolecular Conformation and Coarse-grained Molecular Modeling of Chondroitin Sulfate Conformation, Titration, and Osmotic Pressure

by

Mark Bathe

Bachelor of Science, Mechanical Engineering, 1998
Master of Science, Mechanical Engineering, 2000
Massachusetts Institute of Technology
Cambridge, MA

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

Massachusetts Institute of Technology

June 2004

Signature of Author

Department of Mechanical Engineering
May 7, 2004

Certified by

Alan J. Grodzinsky
Professor of Electrical, Mechanical, and Biological Engineering
Thesis Committee Chairman

Accepted by

Ain A. Sonin
Professor of Mechanical Engineering
Chairman, Committee for Graduate Studies

Inverse Monte Carlo Simulation of Biomolecular Conformation and Coarse-grained Molecular Modeling of Chondroitin Sulfate Conformation, Titration, and Osmotic Pressure

by

Mark Bathe

Submitted to the Department of Mechanical Engineering on May 7th, 2004,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

The first part of this thesis is concerned with the solution structure determination problem. Whereas many biomacromolecules, such as proteins, can be adequately characterized by a single conformation in solution, numerous other important molecules (e.g., nucleic acids, carbohydrates, and polypeptides) exhibit conformational isomerism and disorder. For these molecules, the term "structure" does not correspond to a single conformation but rather to an ensemble of conformations. Given a molecular model and experimental data, the goal of the structure determination problem is to solve for an ensemble of conformations that is consistent with the data. Traditional computational procedures such as simulated annealing, however, are not guaranteed to generate a unique ensemble. The computed ensemble is often simply dependent on the user-specific protocol employed to generate it.

As an alternative, a numerical method for determining the conformational structure of macromolecules is developed and applied to idealized biomacromolecules in solution. The procedure generates unique, maximum entropy conformational ensembles that reproduce thermodynamic properties of the macromolecule (mean energy and heat capacity) in addition to the target experimental data. As an evaluation of its utility in structure determination, the method is applied to a homopolymer and a heteropolymer model of a three-helix bundle protein. It is demonstrated that the procedure performs successfully at various thermodynamic state points, including the ordered globule, disordered globule, and random coil states.

In the second part of this thesis, a molecular model is developed and used to investigate the properties of anionic glycosaminoglycan (GAG) molecules. GAGs are critically important to the structure and biomechanical properties of articular cartilage, an avascular tissue that provides a low-friction, protective lining to the ends of contacting bones during join locomotion. The tissue consists predominantly of two types of macromolecules, collagen and aggrecan. Aggrecan consists of a linear protein backbone with a high mass fraction of covalently attached chondroitin sulfate (CS) GAGs, which endow cartilage with its high compressive modulus via osmotic action. During the onset and progression of osteoarthritis, a debilitating joint disease that affects millions in the US alone, the chemical composition of CS (sulfate type, sulfate pattern, and molecular weight) changes, concomitantly with alterations in the biomechanical properties of cartilage.

For this reason, it is of primary biological interest to understand the effects of CS chemical composition on its conformation, titration behavior, and osmotic pressure. To enable the investigation of these properties, a coarse-grained model of CS is developed. Systematically derived from an all-atom description, the model enables the atomistic-based simulation of large-scale macromolecular assemblies relevant to cartilage biomechanics. Extensive comparison with experimental data demonstrates that this computationally efficient model is also quantitatively predictive, despite the absence of any adjustable parameters. 4-sulfation of CS is found to significantly increase the intrinsic stiffness of CS, as measured by the characteristic ratio and persistence length in the limit of high ionic strength. Average sulfate density is found to increase CS stiffness at finite ionic strength due to electrostatic interactions that tend to stiffen the chain backbone. Sulfation type and pattern (the statistical distribution of sulfates along a CS chain) are not found to influence the osmotic pressure, which is found to be sensitive primarily to the mean volumetric fixed charge density.

Thesis Committee Chairman: Alan J. Grodzinsky
Title: Professor of Electrical, Mechanical, and Biological Engineering

Acknowledgements

First and foremost I would like to thank my three advisors, Professors Greg Rutledge, Bruce Tidor, and Alan Grodzinsky, for their support in this work. I would also like to thank the Department of Defense for their full, three-year financial support in the form of an NDSEG graduate fellowship, as well as the professors that supported me in applying for that fellowship, Professors Roger D. Kamm, Mary C. Boyce, Nicolas Hadjiconstantinou, and Bora Mikič.

From the very beginning until the very end, Greg has taken the time to meet with me regularly to discuss my progress, answer my questions, or just chat, no matter how busy his schedule. He was tremendously patient with me during my times of floundering, and they were many, and wholly supported my pursuit of whichever research direction I chose, teaching me innumerable valuable lessons in research along the way. He has my deepest appreciation.

I thank Bruce for bringing my attention to the joint research project with Alan on cartilage biomechanics. Bruce has a unique ability to see the "big picture" in research and focus on maximizing one's impact on science. I have benefited enormously from his perspective and philosophy on research as well as from my contact with the ideas on protein design and optimization utilized in his lab and I am greatly indebted to him.

My third advisor on this project was Alan, whom I had less extensive contact with, but when I did it was memorable. Aside from leading me by the example of his obvious traits as a great researcher and scientist, Alan taught me something very important in life: it is possible to perform at the highest level of intensity and quality in one's daily work life and at the same time be as congenial, relaxed, and down to Earth as can be. I will strive to emulate Alan's attitude towards research and life during the remainder of my career.

I am also deeply indebted to Roger Kamm, who has been much more than a committee member to me during my time at MIT. My first, formative years in research were spent with Roger during my Bachelor's and Master's theses. Roger has a uniquely dynamic and open attitude towards science and research that has benefited me enormously. I could not have asked for a more supportive advisor to start my research career with and I thank him for motivating me to follow my interest in biophysics.

Davide Marini has been a constant source of support, energy, and friendship for me during, and prior to, my Ph.D. From our qualifiers preparation to lunches at the Dome, Davide has become a very close friend and confidant, in addition to being a great excuse to visit Italy from time to time!

I am extremely grateful to all of the members of the Rutledge and Tidor labs, past and present, for the countless discussions we have had and the help and support that they have provided me. Michael Altman and Pieter In't Veld, in particular, made substantial contributions to my research and always had time for me. Thank you Michael and Pieter. I also thank Wonmuk Huang for his help in analytically solving several integrals contained in Chapter 3 of this thesis that Mathematica, Maple, and Mathcad could not touch.

Finally, without the unconditional love and support of my parents, Klaus-Jürgen and Zorka, my sister Ingrid, my soul-mate Olga, and all of my close friends, you know who you are, I would not be where or who I am today.

Table of Contents

Chapter 1

An inverse Monte Carlo procedure for conformation determination of macromolecules

(published in the *Journal of Computational Chemistry*, 2003, 24 (7): 876-890)

1.1 Introduction

The study of macromolecular structure in solution is of central importance to structural biology. Solution structure studies may be carried out at physiological levels of salt concentration, pH, and temperature, enabling the study of macromolecular structure under native conditions. Whereas many proteins can be adequately characterized by a single conformation in solution, numerous other important biomacromolecules (e.g., carbohydrates, nucleic acids, and polypeptides) exhibit conformational isomerism and disorder. Additionally, their equilibrium conformational ensembles may depend significantly upon specific environmental conditions such as pH and salt concentration. Examples include the pH-dependent helix-coil transition of some polypeptides (Poland and Scheraga 1970) and the salt-concentration conformational dependence of charged biopolymers such as DNA and anionic glycosaminoglycans (Cleland 1977; Skibinska and others 1999). Solution structure studies are therefore particularly important for flexible macromolecules, because x-ray crystallography typically results in a single conformation that may not be relevant to the biologically active confomer(s) (Groth and others 2001; Nikiforovich and Marshall 2001).

NMR and x-ray or neutron scattering are the predominant experimental approaches to structure determination in solution. Each method yields ensemble-averaged observables (distance and dihedral angle restraints in NMR and the molecular structure factor in scattering) that are used as input to computational structure determination methods. Given the experimental data and a molecular model, the goal of the computational methods is to find the ensemble of model conformations that is consistent with the experimental observables. A primary concern in this process regards

the completeness and uniqueness of the generated ensemble (Bonvin and Brunger 1996; Groth and others 2001).

Computational methods for structure determination of conformationally flexible molecules from NMR data fall into one of two classes. In the first, the potential energy function of the molecular model is modified to include a penalty term that accounts for differences between model and experimental observables. Model observables are computed either as time-averages over an adjustable window size during a single simulation or as ensemble averages over multiple simulations (Bonvin and Brunger 1995; Torda and others 1993). In the second approach, an initial "basis set" of conformations is first generated using molecular mechanics (typically using systematic search or simulated annealing followed by energy minimization). The weights of the various conformations are then adjusted so as to maximize agreement between the model and experimental observables (Groth and others 1999; Nikiforovich and others 1987; Shenderovich and others 1988). Limitations of the abovementioned methods include their dependence on molecular mechanical force fields (which may result in an incomplete conformational basis set) and the use of adjustable parameters in the time averaging and weight determination methods (Groth and others 2001).

Computational methods for structure determination from scattering data typically define an objective function (analogous to the penalty term referred to above) that represents the difference between the experimental and model structure factors. The objective function is then employed in a conventional molecular simulation in one of two ways. In the first, conformational search and minimize techniques, such as simulated annealing followed by quenching, are used to minimize the objective function, thereby

maximizing agreement between the experimental and model structure factors. Multiple simulations of this type may be performed to generate a set of conformers to the data (Svergun 1999). In the second approach, called Reverse Monte Carlo, a single Monte Carlo simulation is performed using the objective function to sample statistically an ensemble of conformations (Mcgreevy and Pusztai 1988; RosiSchwartz and Mitchell 1996). Both of the above procedures effectively replace a unique, thermodynamically consistent ensemble of conformations with a non-unique ensemble that depends upon the choice of adjustable parameters and the simulation protocol employed (Groth and others 1999). While this may not be a serious limitation for identifying a single best-fit conformer, such as is often the case for folded proteins, it is unsatisfactory for describing the correct distribution of conformers for flexible macromolecules.

In a recent communication we introduced a parameter-free Monte Carlo procedure for determining molecular structure that is formulated in the semi-grand canonical statistical thermodynamic ensemble (referred to hereafter simply as SGMC, for semi-grand canonical Monte Carlo simulation) (Rutledge 2001). The inputs to the method are the inter-molecular radial distribution function (RDF), which is directly related to the spherically averaged molecular structure factor by Fourier transformation, and a molecular model. The output is the (unique) structural ensemble that maximizes the conformational entropy of the system subject to the RDF provided and thermodynamic constraints (number of molecules N, volume V, and temperature T). We demonstrated its utility by applying it to a model Lennard-Jones fluid at various thermodynamic state points. It was shown that the procedure generates an ensemble of configurations that correctly reproduces the mean energy and heat capacity of the system

in addition to the target RDF. We have also demonstrated how the method can be used to analyze chain stretching during flow, as measured by orientation distributions deduced from DECODER NMR experiments for a polystyrene melt (Colhoun and others 2002).

The goal of the present communication is to evaluate the utility of the SGMC procedure for structure determination of macromolecules in solution. For this purpose, it is desirable to study model biopolymers for which the exact conformational ensembles and thermodynamic properties are known a priori. Following the approaches of Taylor and Lipson (1996), Zhou and Karplus (1997), Zhou and Karplus (1999), and Zhou and others (1997), we choose to model a protein as a series of N freely jointed beads, each representing an amino acid centered at its C^{α} position. The beads interact via non-bonded potentials that implicitly include the effects of solvent. We employ the Metropolis Monte Carlo (MMC) method to generate test sets of "experimental data" (residue-specific inter-residue RDFs, hereafter referred to simply as residue-specific RDFs, the input to the SGMC procedure) for a homopolymer and heteropolymer protein model at two state points each, so as to test the method for both ordered and conformationally flexible molecules (Figure 1). The utility of the SGMC procedure is then evaluated by employing solely the residue-specific RDFs to solve the inverse problem, namely to compute iteratively effective inter-residue interaction potentials that correctly reproduce the RDFs in the semi-grand canonical ensemble. The accuracy of the resulting conformational ensembles is evaluated by comparing not only the radial distribution functions but also the mean energy and heat capacity of the "experimental" and SGMC ensembles.

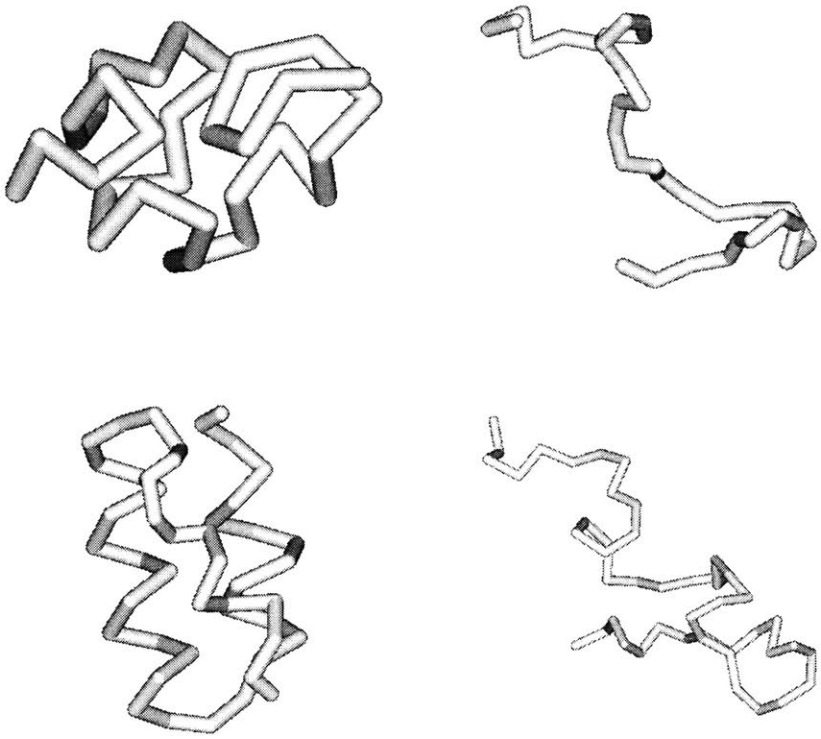


Figure 1 Instantaneous depictions of the homopolymer in its disordered globule (top left, $T^* = 1.0$) and random coil (top right, $T^* = 4.0$) states and of the three-helix bundle protein model in its native, ordered globule (RMS structure) (bottom left, $T^* = 0.65$) and random coil (bottom right, $T^* = 1.5$) states.

We note that although the SGMC procedure is tested on model proteins in this study, it is by no means limited to poly-amino acids. It is generally applicable to homopolymer and heteropolymer chain molecules. Its application to real macromolecules, however, is currently limited by the fact that residue-specific radial distribution functions are not readily resolved for heteropolymers in most solution scattering measurements. Such measurements may become feasible in the future, however, and we suggest at least one avenue for obtaining the required residue-specific RDFs in the Concluding Discussion. Moreover, because residue-unspecific RDFs are readily available experimentally for heteropolymers, we also test the capability of the

SGMC procedure to predict conformational ensembles for the heteropolymer using a single, residue-unspecific RDF in the Results section. We find promising preliminary results for the heteropolymer in its conformationally flexible state.

1.2 Theory and numerical methods

1.2.1 Theoretical basis of the SGMC procedure

We outline the theoretical basis of the SGMC procedure here and refer the reader to Rutledge (2001) and Kofke (1999) for details. Consider a system of N particles in a fixed volume V and at a fixed temperature T. Particles interact through a pair-wise additive potential $\phi_{ij} = \phi\left(\mathbf{r}_i, \mathbf{r}_j, I_i, I_j\right)$ where $\mathbf{r}_i$ is the position of the i^{th} particle with respect to an inertial frame of reference and I_i is its speciation index, a component designating random variable distributed according to $p(I)$ such that $p(I)$ is normalized and everywhere non-negative. In this framework, the isomolar semi-grand canonical partition function of the N-particle system is

$$Y_N\left(N,V,T,a^*(I)/a^*_{ref}\right) = \frac{Z_N}{N!}\left[\frac{q_{int}}{\Lambda^3}\right]^N \int_{I_1}\cdots\int_{I_N}\prod_{i=1}^{N}\left[\frac{a^*(I_i)p(I_i)}{a^*_{ref}p_{ref}}\right]\prod_{i=1}^{N} dI_i \qquad (1)$$

where q_{int} is the internal partition function and Λ is the de Broglie wavelength, each of which is independent of I in the applications that follow. We also assume that the interaction potential, ϕ_{ij}, is independent of I and set it equal to a predetermined "base" interaction potential, $\phi(\mathbf{r}_i, \mathbf{r}_j) = \phi_0(\mathbf{r}_i, \mathbf{r}_j)$. The configurational partition function is then, $Z_N = \int_{\mathbf{r}_1}\cdots\int_{\mathbf{r}_N}\exp\left[-\beta U(\mathbf{r}_1,\ldots,\mathbf{r}_N)\right]\prod_{i=1}^{N} d\mathbf{r}_i$, where $U(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \sum_{i<j}\phi_0(\mathbf{r}_i,\mathbf{r}_j)$ is the potential energy of the system. The residual activity, $a^*(I_i)$, is defined to be $a(I)/p(I)$, where $a(I) = \exp\left[\beta\mu(I)\right]$, μ is the chemical potential, and $\beta = (k_B T)^{-1}$, where k_B is the

Boltzmann constant and T is absolute temperature. A reference activity, $a_{ref} = a^*_{ref} p_{ref}$, has been introduced in the partition function of the N-particle semi-grand canonical ensemble because the total number of particles is fixed. Although the speciation index I in Eq. (1) is a scalar variable that is typically used to label chemically distinct species, it is not limited as such. For systems in which more than one parameter is required to define a species (Rutledge 2001), the speciation index may equally well be written as a vector, **I**. The residual activity and distribution functions then take the forms, $a^*(\mathbf{I})$ and $p(\mathbf{I})$, respectively. As an example, anisotropic fluid particles in a nematic liquid crystal could be speciated according to their orientation, which would require two indices, e.g. the polar and azimuthal angles, to characterize the species of each particle.

With reference to Eq. (1), the Monte Carlo acceptance criterion in a simulation of N particles in a semi-grand canonical ensemble is

$$rn \le \min\left\{1, \frac{\left[e^{-\beta U}\prod_{i=1}^{N} a^*(\mathbf{I}_i)\, p_{tar}(\mathbf{I}_i)\right]_{new}}{\left[e^{-\beta U}\prod_{i=1}^{N} a^*(\mathbf{I}_i)\, p_{tar}(\mathbf{I}_i)\right]_{old}}\right\} \tag{2}$$

where rn is a pseudorandom deviate between 0 and 1 and *tar* is used to differentiate between the target (experimental) and SGMC-simulated probability distributions. To incorporate the radial distribution function into the semi-grand canonical framework, we first assign each particle to a unique species defined by its position relative to the $(N-1)$ remaining particles. We thereby associate $p(\mathbf{I}_i)$ with the N-particle distribution function

$g^{(N)}(r_{ij,j\neq i})$. Next, we invoke a Kirkwood-like superposition approximation, $g^{(N)}(r_{ij,j\neq i}) \approx \prod_{j,j\neq i} g^{(2)}(r_{ij})$, where the two-particle distribution function $g^{(2)}(r_{ij})$ is simply the radial distribution function $g(r)$. The product over all particles in Eq. (2) can be written equivalently as a product over all pairs of particles. Grouping contributions of similar r, the number of pair-wise interactions between r and $r + \delta r$ is $n(r) = \rho N g(r) 2\pi r^2 \delta r$, where ρ is the average number density, N/V. The acceptance criterion in Eq. (2) then becomes

$$rn \leq \min\left\{1, e^{-\beta\Delta U} \prod_{r=0}^{r_{cut}} [a^*(r) g_{tar}(r)]^{\Delta n(r)}\right\} \tag{3}$$

where $a^*(r)$ is the residual activity function, $g_{tar}(r)$ is the target experimental radial distribution function, and ΔU and $\Delta n(r)$ are differences in the configurational energy and the number of unique pair-wise interactions at distance r, respectively, between successive configurations.

1.2.2 Extension and application of the SGMC procedure to isolated chain molecules

The preceding theory, which was applied in our previous study to a Lennard-Jones fluid of identical molecules with periodic boundary conditions, can be extended to the case of an isolated homopolymer chain molecule by removing the periodic boundary conditions and treating the individual residues as fluid molecules (Taylor and Lipson

1996). In this case, the non-bonded intra-molecular radial distribution function is given by

$$g(r) = \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} w_{ij}(r) \tag{4}$$

where $w_{ij}(r)$ denotes the partial intramolecular radial distribution function between monomers i and j. Extension of the SGMC theory to heteropolymers requires that each pair of particles is speciated not only according to its relative separation distance, but also by its type of chemical (pair-wise) interaction. The speciation vector **I** is then composed of the sets $\{\eta, r\}$, where η ($\eta = 1,...,M$) designates the pair-wise interaction type and r designates the pair-wise distance index. Although in general both r and η may be continuous parameters, in the present application η is discrete. As will be seen in the next section, $M = 2$ for the heteropolymer model studied in this work because it has two distinct types of pair-wise interactions called "contacts" and "non-contacts". Moreover, composition in η-space is fixed in the MMC and SGMC simulations because for each pair of particles i and j, the value of η_{ij} is known and held constant during the simulations (i.e., "contacts" remain "contacts" and "non-contacts" remain "non-contacts").

In order to completely solve for the structure of a heteropolymer, the SGMC procedure requires that each distinct interaction type has a separate target radial distribution function, $g_{tar}(\eta, r)$, associated residual activity, $a^*(\eta, r)$, and $n(\eta, r)$ in Eq. (3). The modified sampling criterion for multiple types of pair-wise interactions is

$$rn \le \min\left\{1, e^{-\beta\Delta U}\prod_{\eta=1}^{M}\prod_{r=0}^{r_{cut}}[a^*(\eta,r)g_{tar}(\eta,r)]^{\Delta n(\eta,r)}\right\}. \quad (5)$$

where we have allowed for distinct "base" interaction potentials, $\phi_0(\eta,r)$, specific to pair-wise interaction type η, so that the potential energy is now, $U = \sum_{i<j}\phi_0(\eta_{ij},r_{ij})$. Note that unlike the case of a continuous liquid in which $g(r)$ approaches one in the limit of large r, $g(r)$ for an isolated chain molecule approaches zero for values of r that are large relative to the chain's root-mean-squared radius of gyration, $\left\langle R_g^2\right\rangle^{1/2}$. Moreover, the average number density, ρ, may be arbitrarily small for an isolated (infinitely dilute) chain, since V can be taken as large as desired.

As shown by Rutledge (2001), the residual activities, $a^*(\eta,r)$, that result in the target probability distributions, $g_{tar}(\eta,r)$, are in general not known a priori, and therefore the SGMC procedure requires iteration. Successive updates on the estimates of $a^*(\eta,r)$ converge to estimates of effective pair-wise interaction potentials

$$\phi_{eff}(\eta,r) = \lim_{k\to\infty}\left\{\phi_0(\eta,r) - k_BT\ln[g_{tar}(\eta,r)\frac{a^*_{(k)}(\eta,r)}{a^*_{ref}(\eta)}]\right\} \quad (6)$$

where k is the iteration number and the $\phi_0(\eta,r)$ were taken to be the hard-core portions of the original MMC potentials in the present study. In alternative applications of the SGMC procedure to problems of macromolecular structure determination it may be desirable to include additional known potential energy terms in ϕ_0.

1.2.3 Residual activity iteration procedure

To obtain successive updates in the unknown residual activity, $a^*(\eta,r)$, we developed and implemented a multi-dimensional Newton-Raphson iteration scheme similar to that of Lyubartsev and Laaksonen (1997). The goal of the SGMC simulation is to find $a^*(\eta,r)$ such that $\langle n(\eta,r)\rangle = n_{tar}(\eta,r)$ to within statistical uncertainty for all η and r. For purposes of updating $a^*(\eta,r)$, $\langle n(\eta,r)\rangle$ is an ensemble average quantity that is obtained during a single iteration of the SGMC simulation at the current $a^*(\eta,r)$. $n_{tar}(\eta,r)$ is known a priori and directly related to $g_{tar}(\eta,r)$. We begin by discretizing the functions $a^*(\eta,r)$, $\langle n(\eta,r)\rangle$, and $n_{tar}(\eta,r)$ in η and r-space to form three $M \times (r_{cut}/\delta r)$ matrices denoted **A**, $\langle \mathbf{N} \rangle$, and $\mathbf{N}_{tar}$, respectively. Subscripts will be used to denote components of these matrices when necessary and a parenthetical subscript (k) will denote its k^{th} iterative estimate. We proceed by expanding $\left[\mathbf{N}_{tar} - \langle \mathbf{N} \rangle\right]$ in a Taylor series in **A** about the initial guess $\mathbf{A} = \mathbf{A}_{(0)}$, where $A_{\eta r(0)} = a^*_{ref,\eta} / g_{tar,\eta r}$ and $a^*_{ref,\eta} = 1$. Note that each pair-wise interaction type η has an associated reference residual activity $a^*_{ref,\eta}$. Dropping nonlinear terms and setting the expansion equal to zero, we obtain the resulting set of linear equations that is to be solved at every iteration k for the increment in the residual activity, $\Delta\mathbf{A}_{(k)}$,

$$\left.\frac{\partial \langle \mathbf{N} \rangle}{\partial \mathbf{A}}\right|_{(k)} \Delta\mathbf{A}_{(k)} = \left[\mathbf{N}_{tar} - \langle \mathbf{N} \rangle_{(k)}\right] \tag{7}$$

where $\partial\langle N_{\eta r}\rangle/\partial A_{\mu\tau} = (A_{\mu\tau})^{-1}\left(\langle N_{\eta r}N_{\mu\tau}\rangle - \langle N_{\eta r}\rangle\langle N_{\mu\tau}\rangle\right)$ (see Appendix A). The updated residual activity, $\mathbf{A}_{(k+1)}$, is computed from $\mathbf{A}_{(k+1)} = \mathbf{A}_{(k)} + \alpha\Delta\mathbf{A}_{(k)}$, where α is a damping factor ($\alpha = 0.5$ was found to be optimal in the present study). An overview of the SGMC structure determination procedure is provided in Figure 2.

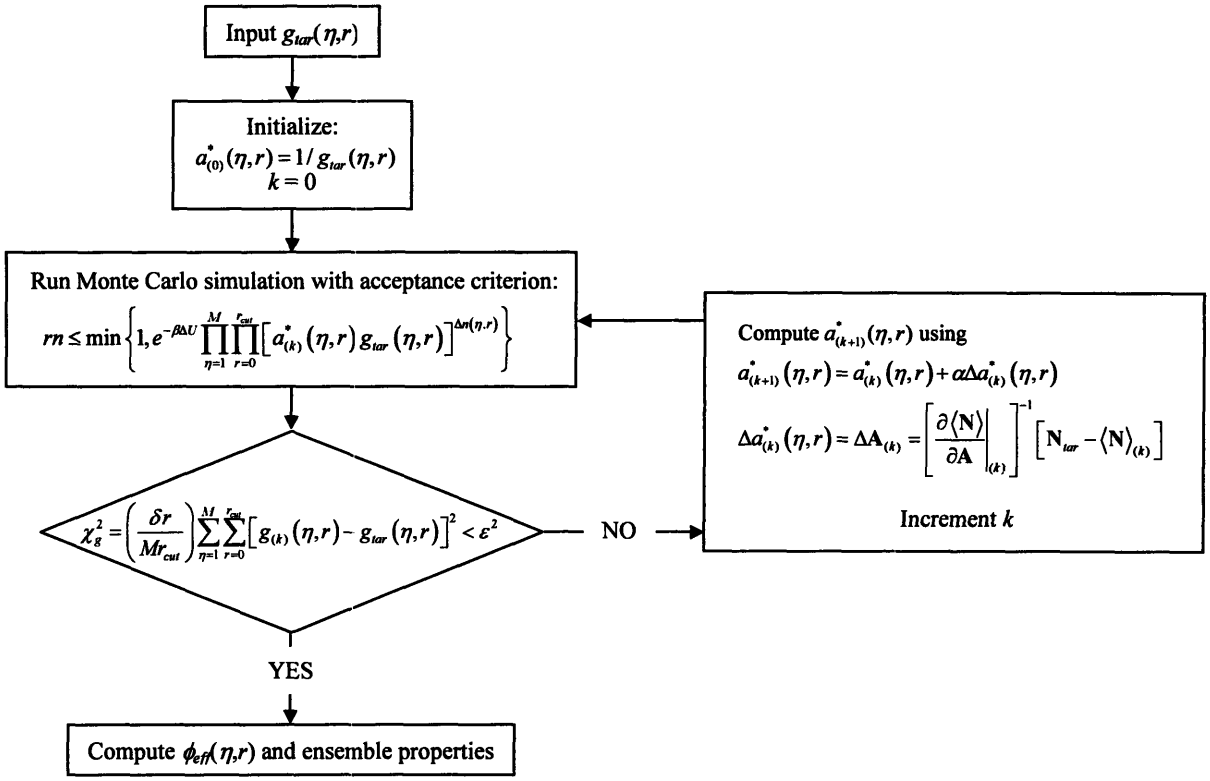


Figure 2 Flow chart of the SGMC structure determination procedure utilizing inter-residue radial distribution functions, $g_{tar}(\eta,r)$, as experimental input.

1.2.4 Properties of the SGMC procedure

Formulating the macromolecular structure determination problem in a statistical mechanical framework (namely, the semi-grand canonical ensemble) results in the following important property of the SGMC procedure: given the fixed values

$(N(\eta),V,T,a^*(\eta,r)/a^*_{ref}(\eta))$, the distribution (or ensemble) of conformations produced by the method maximizes the conformational entropy of the macromolecule subject to the constraints $\sum_q P_q U_q = \langle E \rangle$ and $\sum_q P_q n_q(\eta,r) = \langle n(\eta,r) \rangle$, where $N(\eta)$ is the number of pairs of type η, P_q is the probability of the q^{th} conformation, and U_q and $n_q(\eta,r)$ are the corresponding energy and number of particles of type (η,r) in conformation q (Kofke 1999; McQuarrie 1975). Since iteration is performed to find the appropriate $a^*(\eta,r)$ that reproduce the target $p_{tar}(\eta,r)$ (which in the present application are associated with the non-bonded radial distribution function), this is equivalent to maximizing the conformational entropy subject to the above constraints and the fixed distribution $p_{tar}(\eta,r)$ (assuming the existence of $a^*(\eta,r)$ and convergence of the iterative procedure). Thus, in the current application the method generates the unique, maximum entropy ensemble consistent with the structural data, $g_{tar}(\eta,r)$, and statistical mechanics.

In the current application, the SGMC procedure amounts to computing a pair-wise additive Hamiltonian [$H_{eff} = \sum_{i<j} \phi_{eff}(\eta_{ij}, r_{ij})$] (Eq. (6)) for the macromolecule that reproduces the target RDFs, $g_{tar}(\eta,r)$. Although in the present study H_{eff} coincides with the original, full Hamiltonian of the system, *H*, this will not be the case in applications of the method to real macromolecules. The full Hamiltonian, *H*, of the macromolecular system (protein or other solute + solvent) under study will typically depend upon *n* "interesting" degrees of freedom $(\mathbf{x}_i;\ i=1,...,n)$ (taken to be the positions of the C^α atoms in the current application) and *m* "uninteresting" degrees of freedom $(\mathbf{y}_i;\ i=n+1,...,n+m)$ (taken to be the positions of the remaining protein atoms and all

solvent atoms). Averaging H over the "uninteresting" degrees of freedom results in a new, averaged Hamiltonian, H^*, that explicitly depends upon only the "interesting" degrees of freedom (Roux and Simonson 1999). In general, the new Hamiltonian (a potential of mean force) will not be a pair-wise additive function of the interesting degrees of freedom, but will contain multi-body terms. The superposition approximation employed in the SGMC theory (sec. 2.1), however, assumes that this new Hamiltonian is well approximated using solely pair-wise interaction terms $H_{eff} \approx H^*$. Although it is difficult at present to state how good an approximation this will be in the general case, two important attributes of H_{eff} have been proven. First, of the set of Hamiltonians that reproduce the target RDFs, there is only one that is pair-wise additive (up to an additive constant) (Henderson 1974). Second, of the set of Hamiltonians that reproduce the target RDFs, the pair-wise additive one gives maximum conformational entropy to the system (Lyubartsev and Laaksonen 1997).

The distinguishing feature between applications of the SGMC method to real macromolecules and the simplified square-well models used here is the source of the radial distribution functions employed and the corresponding implicit (solvent + non-C^{α} solute atoms) information that they contain.

1.3 Protein models

Each protein model consists of N freely jointed spherical monomers with hard-core diameters σ_c, connected by bonds of fixed length. The monomers interact via spherically symmetric square-well non-bonded potentials with square-well diameter $\sigma_d = 1.5\sigma_c$. For the homopolymer studies, the interaction potential has a well depth of ε, and its thermodynamic state is therefore defined by the single dimensionless parameter $T^* = k_BT/\varepsilon$, the reduced temperature of the system. The hard-core diameter and all bond lengths are set to be equal and all distances are rendered dimensionless (Zhou and others 1997) upon scaling by σ_c.

The heteropolymer is an off-lattice model of the three-helix bundle fragment of *Staphylococcus aureus* protein A (PDB ID: 1BDD). The model consists of 46 spherical monomers that interact via two different non-bonded square-well potentials of the Gō-type (Taketomi and others 1975). In this model, a distinction is made between "contact" interactions and "non-contact" interactions; the difference in interaction is set by the "bias gap" $g \equiv (1 - B_O/B_N)$, where $B_N\varepsilon$ is the square-well depth of contact pair interactions and $B_O\varepsilon$ is the square-well depth of non-contact pair interactions ($B_N < 0$, $\varepsilon > 0$, and $B_O \geq B_N$). For values of g near or greater than 1 the model strongly favors its native (crystal) structure, whereas for $g \to 0$ the model degenerates to a homopolymer. As demonstrated by Zhou and Karplus (1997, 1999), the model is capable of reproducing not only the three-helix structural motif of the native protein, but also the thermodynamic transitions exhibited by real proteins. The two relevant thermodynamic parameters in the heteropolymer model are the reduced temperature, $T^* = k_BT/\varepsilon$, and the bias gap, g. As was the case for the homopolymer, T^* accounts for the effects of temperature, or

equivalently solvent quality, on the heteropolymer's state, and g accounts for the relative strength of interaction between contacts and non-contacts in the model. Although square-well potentials are not very representative of true interaction potentials, their discontinuous nature together with the multiple state points analyzed in this work provide for a challenging test of the SGMC method.

Following Zhou and Karplus (1997), several steps were required to determine the parameter values of the heteropolymer model. First, 46 spherical monomers were placed at the C^{α} positions of residues 10 through 55 of *Staphylococcus aureus* protein A. All bond lengths were fixed at their initial C^{α}-C^{α} lengths (average bond length 3.79 Å) and the hard-core monomer diameter, σ_c, was chosen to be equal to the minimum non-bonded distance between any two C^{α}'s in the model (σ_c = 4.27 Å). This determined the square-well diameter, σ_d, to be 6.41 Å. Next, we define two non-bonded residues to be in "contact" in a given conformation if their inter-residue distance is within the square-well diameter σ_d. Using this definition, it was determined that there were 97 pairs in contact in the initial conformation. In order to obtain reasonable packing in the model in the absence of any side chains, the model was subsequently annealed in four successive MMC simulations from T^* = 1.0 to 0.001, while only imposing that the 97 initial contacts remain intact (i.e. within the distance σ_c to σ_d) and that the remaining non-bonded pairs satisfy the hard-core portion of the square-well potential. This was achieved by rejecting Monte Carlo moves that violated either of these conditions. After annealing, the total number of non-bonded pairs in contact had increased from 97 to 216*. The resulting 216 pairs were assigned to be "contacts", interacting with a square-well depth $B_N\varepsilon$, whereas

* 254 final contacts resulted from the annealing simulations of Zhou and Karplus (1997) because they were obliged to allow the bond lengths in their model to vary between $0.9\sigma_b$ and $1.1\sigma_b$ in performing Discrete Molecular Dynamics, thus enabling even closer packing.

the remaining non-bonded pairs were assigned to be "non-contacts", interacting with a square-well depth of $B_O\varepsilon$, thus completing the heteropolymer model definition.

In mathematical form, the non-bonded interaction potentials, $u(\eta,r)$, for the homopolymer and heteropolymer models are

Homopolymer:

$$u(r)=\begin{cases}\infty, & r<\sigma_c\\ -\varepsilon, & \sigma_c\le r<\sigma_d\\ 0, & \sigma_d\le r\end{cases} \tag{8}$$

Heteropolymer:

$$u(\eta,r)=\begin{cases}\infty; & \forall\eta, & r<\sigma_c\\ B_N\varepsilon; & \eta=\text{contact}, & \sigma_c\le r<\sigma_d\\ B_O\varepsilon; & \eta=\text{non-contact}, & \sigma_c\le r<\sigma_d\\ 0; & \forall\eta, & \sigma_d\le r\end{cases} \tag{9}$$

1.4 Simulation protocol and results

Two state points were selected for each model to evaluate the utility of the SGMC procedure for macromolecular solution structure determination. The two state points chosen for the homopolymer, $T^* = 1.0$ and 4.0, correspond to disordered globule and random coil states, respectively, and the two heteropolymer state points, $g = 0.70$, $T^* = 0.65$ and 1.5, correspond to the native, ordered globule and random coil states, respectively (Figure 1).

For each state point, a conventional MMC simulation was first performed in the canonical ensemble with the square-well interaction potentials turned on to generate target radial distribution functions for the respective models and state points. The SGMC procedure was then employed with the attractive portion of the interaction potentials turned off (as mentioned earlier, the "base" SGMC interaction potentials, $\phi_0(\eta,r)$, coincided with the respective hard-core portions of the MMC potentials, $u(\eta,r)$) to test the ability of the method to solve the inverse problem and regenerate the original interaction potentials employing only the target RDFs as input. The total non-bonded RDF was used as input for the homopolymer and the individual contact and non-contact RDFs (i.e., "residue-specific" RDFs) were used for the heteropolymer. In addition, a second simulation was performed for the heteropolymer using solely the *total* non-bonded RDF (i.e., "residue-unspecific" RDF). The latter simulation was performed in order to test the ability of the method to reproduce the target conformational ensemble in the absence of fully resolved, residue-specific experimental data. To evaluate the accuracy of the ensemble (or distribution) of conformations generated by the SGMC-generated effective interaction potentials, the mean energy and heat capacity per residue,

$E^* = \langle E \rangle / (\varepsilon N)$ and $C_v^* = C_v / (k_B N) = \beta^2 \left[\langle E^2 \rangle - \langle E \rangle^2 \right] / N$, were computed for each ensemble as measured by the *original* interaction potentials. This measure was chosen because it provides for an unambiguous comparison of the similarity between the MMC and SGMC structural ensembles up to the second moments of their distributions in energy.

For each Monte Carlo simulation (MMC and at every iteration of SGMC), initial molecular conformations were generated using a self-avoiding random walk. The models were then equilibrated using $0.25 - 3.0 \times 10^6$ moves, after which $10 - 25 \times 10^6$ moves were used to compute structural and thermodynamic averages, as well as the derivatives in Eq.(7) (SGMC only). Trial conformations were generated using the "translate-jiggle" algorithm of Dickman and Hall (1986). The move consists of selecting a bead at random along the chain, displacing the bead by a random displacement vector (the maximum magnitude of which is adjusted to maintain an acceptance/rejection ratio between 0.4 – 0.6), reconnecting the bead to the previous bead in the chain by normalizing the relevant bond vector to its original length, and rigidly translating the remaining portion of the chain to reconnect it to the displaced bead. Discretization in r-space was chosen for each simulation such that $\delta r = L/1200$, where L denotes the contour length of the relevant chain. $r_{cut} = L/2$ for $g(\eta, r)$ in each simulation whereas $r_{cut} = 4\sigma_c$ and $3\sigma_c$ for $a^*(\eta, r)$ and the acceptance criterion in the homopolymer and heteropolymer models, respectively. Prior to applying the SGMC procedure to the MMC-generated RDFs, we confirmed that the thermodynamic averages computed using the MMC algorithm were in excellent quantitative agreement with the results of Zhou and Karplus (1997) and Zhou and others (1997) (data not shown).

1.4.1 Homopolymer

Figure 3 shows the agreement between the homopolymer target (MMC) and converged SGMC RDFs for the two simulated state points, $T^* = 1$ and $T^* = 4$ after seven iterations.

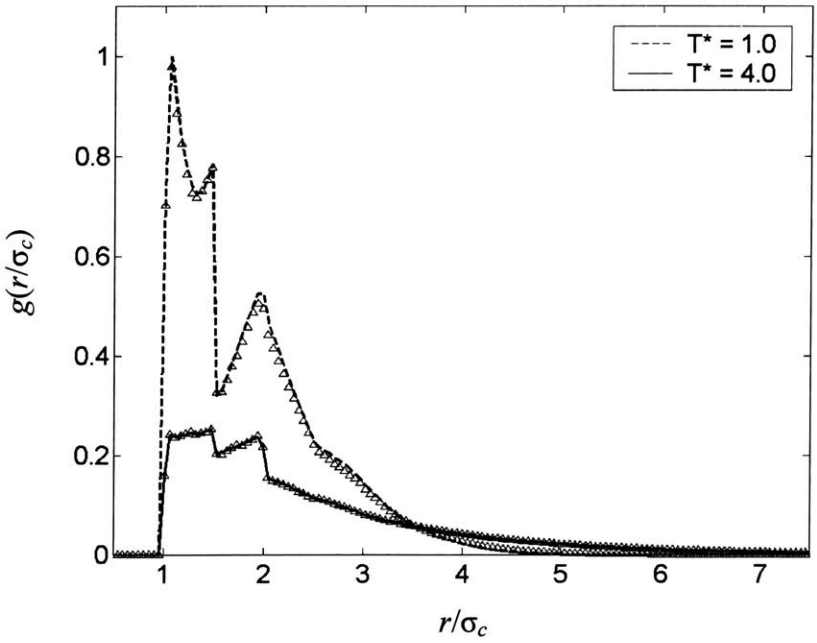


Figure 3 Target MMC (lines) and reconstructed SGMC (symbols) homopolymer non-bonded radial distribution functions for $T^* = 1.0$ (– –) and 4.0 (–), corresponding to the disordered globule and random coil states, respectively. SGMC results required 7 iterations ($\alpha = 0.50$) for each state point.

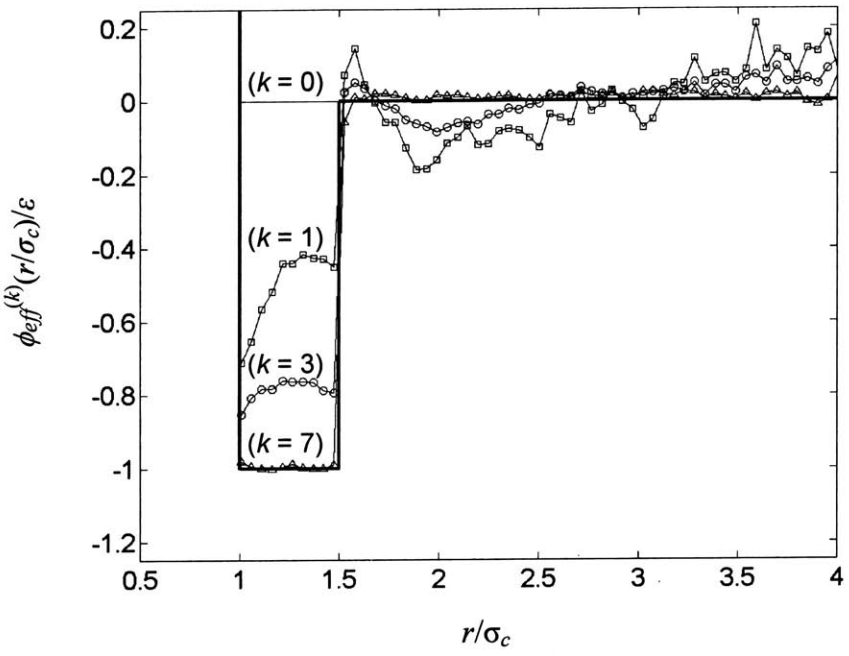
Evidence of the random coil nature of the homopolymer at $T^* = 4$ is provided by its RDF which exhibits significantly reduced first and second neighbor peaks as compared with its disordered globule ($T^* = 1$) RDF, as well as a slowly decaying tail that remains non-zero for distances as large as $r = 7\sigma_c$. The convergence histories of the effective interaction potentials (iteration numbers $k = 0$, 1, 3, and 7) are illustrated in Figs 4a and 4b for $T^* = 1$ and 4, respectively. The relative agreement between the MMC and SGMC RDFs is seen to be better than the agreement between the effective interaction potentials, particularly

for the high temperature, random coil case. These features illustrate the insensitivity of the RDF to certain details of the interaction potential, particularly for higher temperatures (or equivalently, better solvents) as entropic effects begin to dominate over the interaction energy. The MMC and SGMC ensemble average mean energies and heat capacities as measured by the original MMC interaction potentials are provided in Table 1.

Table 1 MMC and converged SGMC thermodynamic averages for the homopolymer. Statistical uncertainties evaluated using the blocking method (Flyvbjerg and Petersen 1989).

	E^*		C_v^*	
T*	**MMC**	**SGMC**	**MMC**	**SGMC**
1.00	$-2.41 \pm 1.3 \times 10^{-2}$	$-2.39 \pm 9.7 \times 10^{-3}$	1.27	1.30
4.00	$-0.73 \pm 9.1 \times 10^{-4}$	$-0.73 \pm 8.0 \times 10^{-4}$	8.43×10^{-2}	8.50×10^{-2}

(a)



(b)

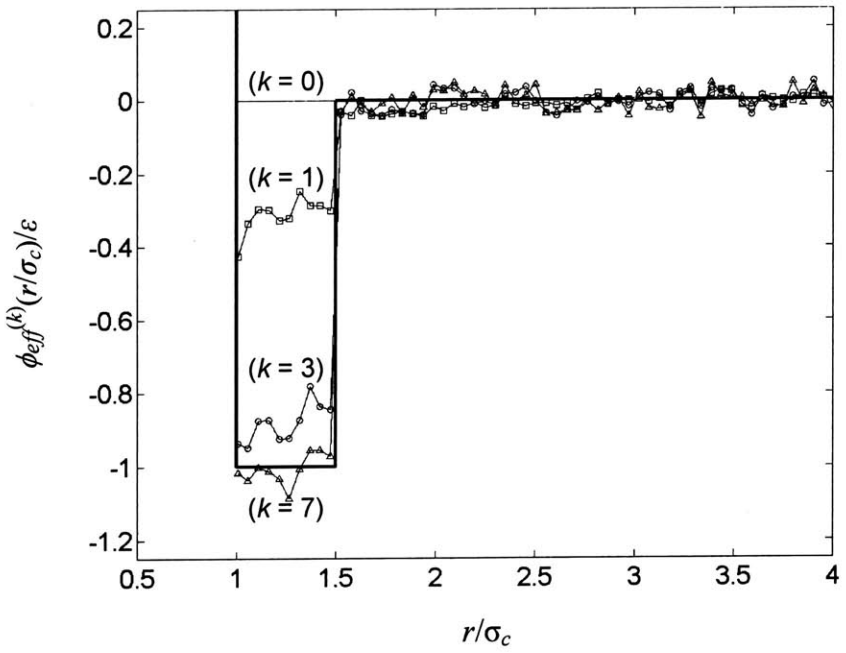


Figure 4 Convergence history (k = iteration #) of the SGMC-reconstructed effective potential and the square-well target potential (thick solid line) for $T^* = 1.0$ (a) and $T^* = 4.0$ (b), corresponding to the disordered globule and random coil states, respectively.

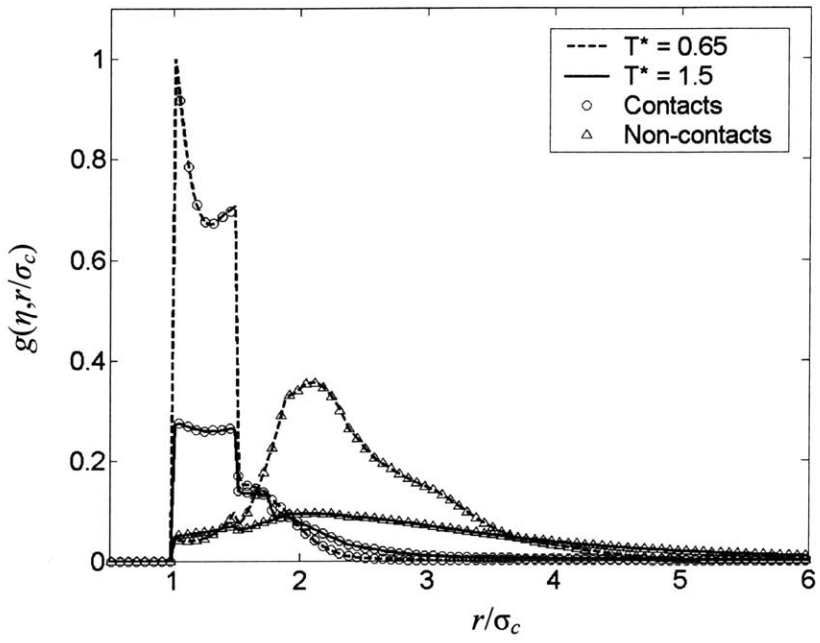
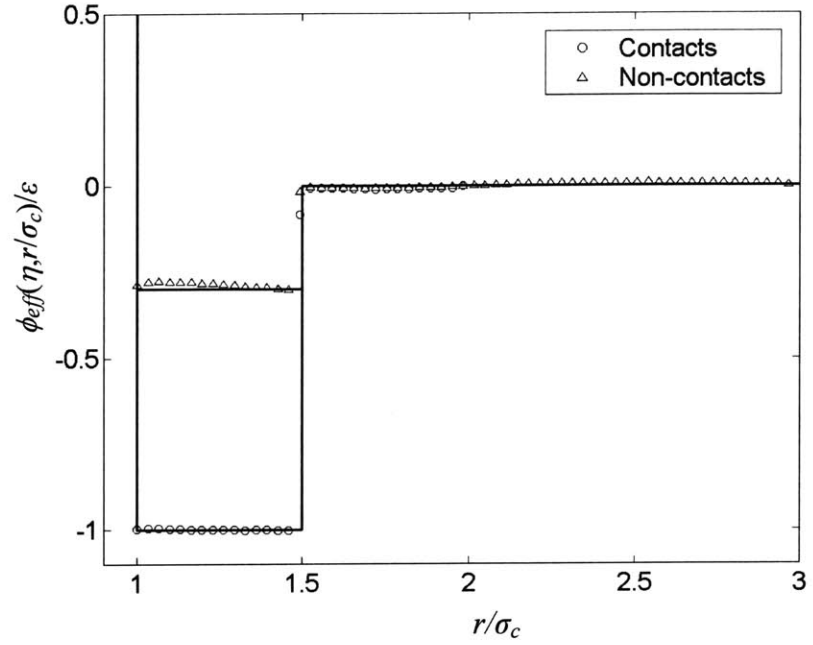


Figure 5 Target (MMC: lines) and reconstructed (SGMC: symbols) heteropolymer (g = 0.70) radial distribution functions for contact (○) and non-contact (Δ) pair interaction types at T^* = 0.65 (– –) and 1.5 (–), corresponding to the native, ordered globule and random coil states, respectively. SGMC results required 13 and 8 iterations (α = 0.50) for the low and high temperature cases, respectively.

1.4.2 Heteropolymer with residue-specific RDFs

Figure 5 shows the agreement between the heteropolymer target (MMC) and converged SGMC RDFs for contact and non-contact pairs types at the two simulated state points, T^* = 0.65 and 1.50. Contact pair residues are seen to reside within about $3\sigma_c$ of each other at each state point, with a strong peak corresponding to the square-well region of r = 1-1.5σ_c. In contrast, non-contact pair residues exhibit considerably less close-range order due to their relatively weaker interaction energy, particularly in the high temperature case. The SGMC-reconstructed contact and non-contact effective interaction potentials are compared with their target potentials in Figure 6.

(a)



(b)

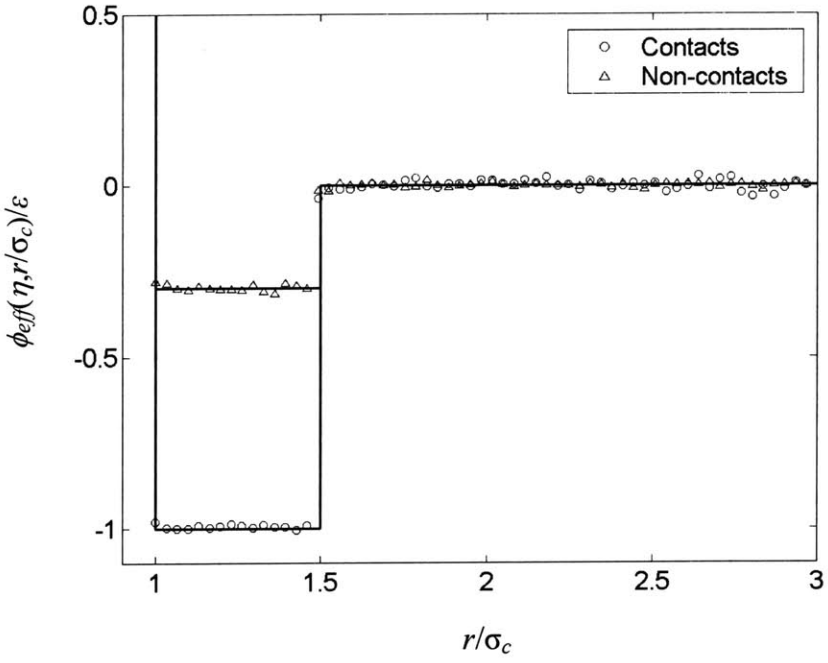


Figure 6 Target (MMC: lines) and reconstructed (SGMC: symbols) heteropolymer (g = 0.70) contact and non-contact effective interaction potentials for the ordered globule, T^* = 0.65 (a) and random coil, T^* = 1.5 (b) states. SGMC convergence required 13 and 8 iterations (α = 0.50) for the low and high temperature states, respectively.

The regenerated contact interaction potentials are in better agreement with their targets than the non-contacts, which we attribute to the lower relative sampling of the latter (non-contacts pairs are within the square-well distance of each other less frequently than contact pairs, resulting in a poorer estimation of the relevant derivatives in Eq. (7)). The effect of temperature on the accuracy of the generated potentials is also illustrated by the relatively low degree of noise in the $T^* = 1.5$ case, with respect to the $T^* = 4.0$ case of the homopolymer (Figure 6b vs. 4b), which is nearer to the transition temperature $T^* = 1.0$. Finally, Table 2a demonstrates the ability of the SGMC effective interaction potentials to reproduce the target distribution of conformations for the heteropolymer.

Table 2a MMC and converged SGMC thermodynamic averages for the heteropolymer employing each partial radial distribution function (contact and non-contact) to solve for distinct residual activities for those pairs.

	E^*		C_v^*	
T*	**MMC**	**SGMC**	**MMC**	**SGMC**
0.650	$-3.33 \pm 2.2 \times 10^{-2}$	$-3.28 \pm 2.4 \times 10^{-2}$	2.99	3.15
1.50	$-1.26 \pm 1.6 \times 10^{-3}$	$-1.26 \pm 2.6 \times 10^{-3}$	0.56	0.56

The mean conformational energies (E^*) correctly reproduced by the SGMC procedure in the homopolymer and heteropolymer cases is a simple consequence of the identity between the MMC and SGMC radial distribution functions

$$< E >= \sum_{\eta=1}^{M} \frac{N(\eta)(N(\eta)-1)}{V} \int_0^\infty u(\eta,r) g(\eta,r) 4\pi r^2 dr \,, \qquad (10)$$

where $u(\eta,r)$ denotes the original MMC interaction potential for pair-type η, which is satisfied by construction. More importantly, the correctly reproduced heat capacities (C_v^*) are indicative of the fluctuations in energy due to sampling of the structural ensemble. The fact that these too are correctly reproduced by the SGMC procedure is a strong indication that the *distribution* of conformations at a particular thermodynamic state point is correctly sampled (at least up to its second moment). This feature distinguishes the SGMC method from other structure determination procedures.

1.4.3 Heteropolymer without residue-specific RDFs

As noted earlier, current experimental limitations make it difficult to obtain residue-specific radial distribution functions for macromolecules in solution. Typically, only the total radial distribution function is readily accessible from wide-angle x-ray scattering, for example. To test the ability of the SGMC procedure to accurately compute structural ensembles for heteropolymers employing only the total, residue-unspecific RDF (i.e., ignoring the distinction between different types of pair-wise interactions) we next employ the total non-bonded RDF of the heteropolymer model to compute a single non-bonded effective interaction potential for each state, $T^* = 0.65$ and $T^* = 1.5$, and evaluate the resulting ensembles.

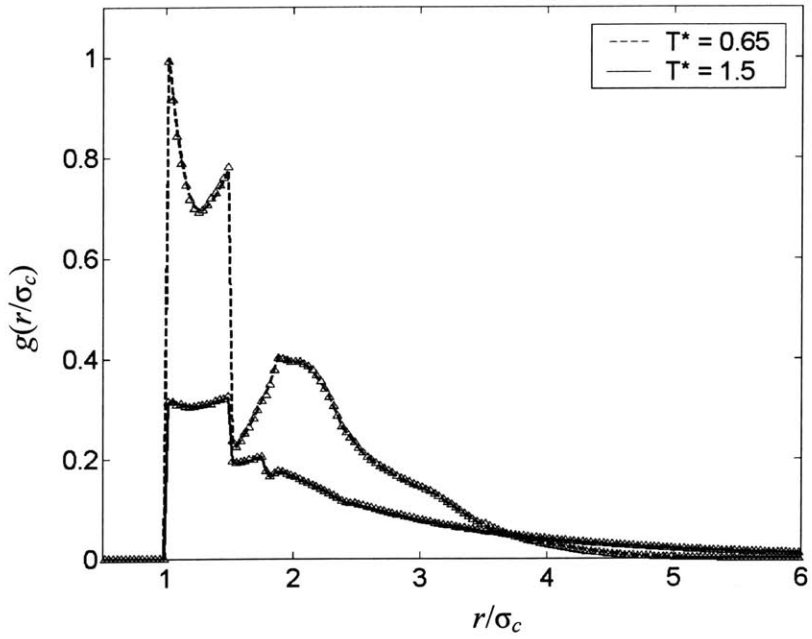


Figure 7 Target MMC (lines) and reconstructed SGMC (symbols) heteropolymer (g = 0.70) total non-bonded radial distribution function at T^* = 0.65 (– –) and T^* = 1.5 (–). SGMC results required 10 and 9 iterations (α = 0.50) for the low and high temperature cases, respectively.

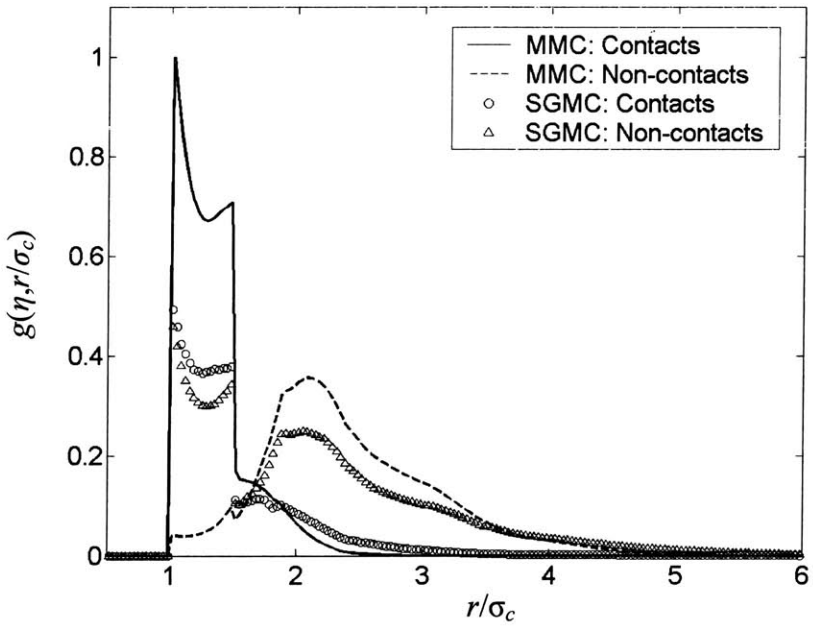
Figure 7 demonstrates that the SGMC-computed effective interaction potential is capable of reproducing the target, total non-bonded heteropolymer RDF at each state point. In order to evaluate the accuracy of the corresponding structural ensembles, however, it is necessary to compare the SGMC contact and non-contact RDFs with their counterparts in the target MMC case (Figure 8), as well as their respective thermodynamic averages, shown in Table 2b. While considerable disagreement is seen to exist in Figure 8 between the contact and non-contact RDFs as well as in the thermodynamic averages in the low temperature, ordered globule state, the SGMC procedure performs reasonably well at the higher temperature state. At T^* = 1.5, the mean energies per residue of the MMC and SGMC systems are –1.26 and –1.16, respectively, and the heat capacities per residue are 0.56 vs. 0.55. Apparently the

disagreements present in the non-bonded RDFs at close range (within the square-well region) do not incur a serious penalty in the computed thermodynamic averages.

Table 2b MMC and converged SGMC thermodynamic averages for the heteropolymer employing the total non-bonded radial distribution function, assuming a single pair-type interaction.

	E^*		C_v^*	
T*	**MMC**	**SGMC**	**MMC**	**SGMC**
0.650	$-3.33 \pm 2.2 \times 10^{-2}$	$-2.17 \pm 6.8 \times 10^{-2}$	2.99	2.09
1.50	$-1.26 \pm 1.6 \times 10^{-3}$	$-1.16 \pm 3.1 \times 10^{-3}$	0.56	0.55

(a)



(b)

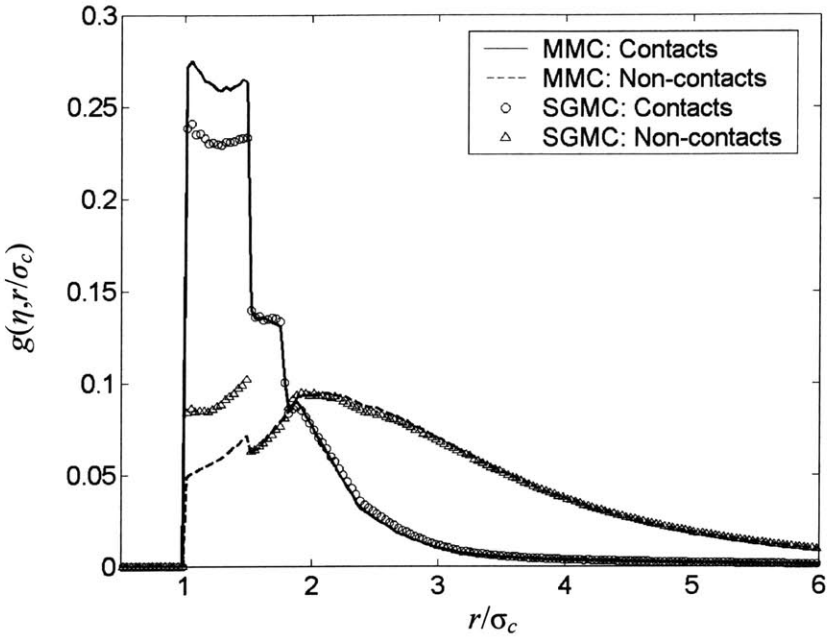
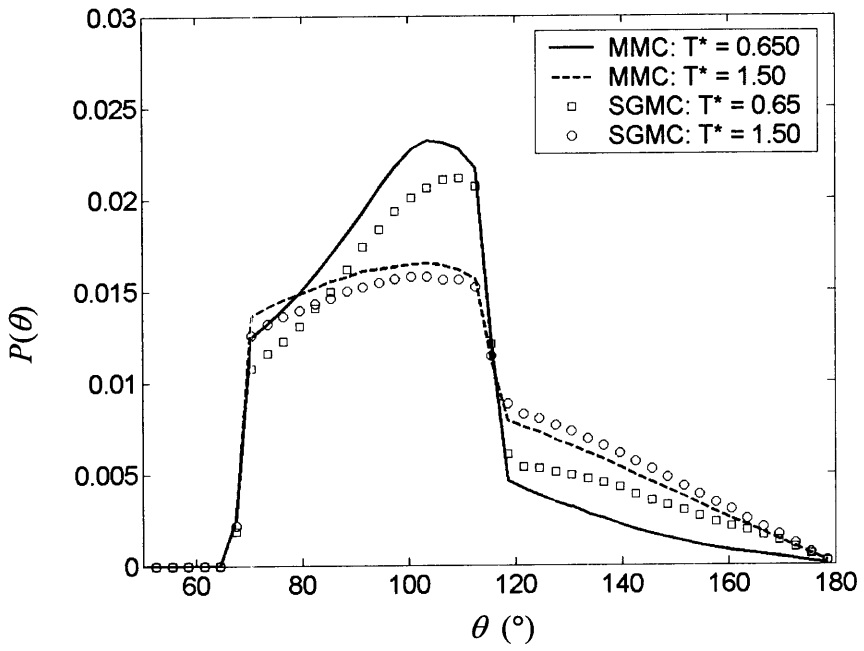


Figure 8 Target (MMC) and reconstructed (SGMC) heteropolymer ($g = 0.70$) radial distribution functions for contact (○) and non-contact (Δ) pair interaction types at $T^* = 0.65$ (a) and 1.5 (b), corresponding to the model protein's native, ordered globule and random coil states, respectively. SGMC results were obtained employing a single residual activity applied to all pair types.

To further investigate the differences between the MMC and SGMC structural ensembles, we compare the bond and dihedral angle distribution functions, $P(\theta)$ and $P(\varphi)$, in Figs. 9a and b, respectively. While the bond angle distribution is in good overall agreement at each state point (Figure 9a), in part due to the hard-core repulsion constraint between residues, the distribution in dihedral angles in the low temperature SGMC case is in considerable disagreement with its target, MMC counterpart (Figure 9b). Dihedral angles are in good agreement, however, in the denatured state, $T^* = 1.5$, further demonstrating that the SGMC procedure is capable of producing accurate structural ensembles for conformationally flexible heteropolymers employing solely the total non-bonded radial distribution function.

(a)



(b)

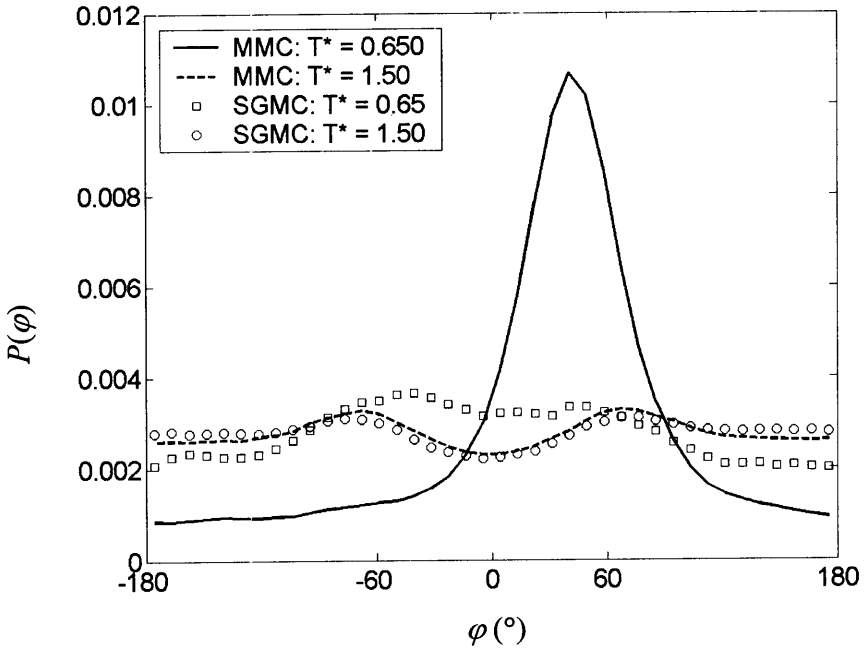


Figure 9 Target (MMC) and reconstructed (SGMC) heteropolymer ($g = 0.70$) bond (a) and dihedral (b) angle distributions at $T^* = 0.65$ and 1.5, corresponding to the model protein's native, ordered globule and random coil states, respectively. SGMC results were obtained employing a single residual activity applied to all pair types.

In modeling the heteropolymer as a homopolymer, we impose a weaker constraint on the system than in the previous application in which residue-specific RDFs are used. In effect, we require that the total non-bonded RDF equal the sum of the individual contact and non-contact RDFs, as opposed to imposing that the contact and non-contact pairs individually satisfy their respective RDFs. Since the SGMC method maximizes conformational entropy subject to these constraints, the resulting structural ensembles as represented by the distribution of bond and dihedral angles in Figure 9 are predicted to be somewhat less ordered at each state point and the mean energies are correspondingly lower in magnitude. The fact that the residue-specific RDFs are better reproduced at the higher temperature state point may be attributed to the diminished distinction between contacts and non-contacts at this state point.

1.5 Concluding discussion

A novel numerical method for the determination of macromolecular structure in solution has been presented and evaluated. The procedure generates effective, residue-specific interaction potentials from the corresponding radial distribution functions by simulating a model macromolecule in a semi-grand canonical ensemble. When provided with each residue-specific radial distribution function corresponding to distinct pair-wise interactions, the procedure generates a structural ensemble that reproduces the mean energy and heat capacity of the macromolecule in addition to the input RDFs, indicating the accuracy of the generated distribution of conformations for random coil, disordered globule, and ordered globule states. When only provided the total, residue-unspecific inter-residue radial distribution function for a heteropolymer, the procedure generates molecular ensembles that can generally be considered as less ordered than their previously mentioned counterparts due to the lack of sufficient experimental data and because it is a maximum entropy approach.

In principle, the SGMC procedure can reconstruct the correct structural ensemble and effective potentials if given sufficient input information, e.g., in the form of residue-specific RDFs. However, the ability of the SGMC procedure to recreate accurate target interaction potentials (at all values of r for each pair-type η) rests on its ability to compute accurately the derivative matrix in Eq. (7). In certain situations, this step may encounter practical limitations, making very long simulations necessary. One such situation occurs when the system structure is determined primarily by entropy. In this case, the RDFs input to the simulation are insufficiently sensitive to the underlying interaction potentials, making their accurate determination a laborious task of long

simulations and low allowable errors in χ^2. Another situation, which turns out to be the case in the low temperature, surface molten solid state of Zhou and Karplus,[1] occurs when regions of the RDFs are essentially unpopulated due to a combination of unfavorable interaction energies associated with those regions and weak entropic effects. As in the entropy-dominated case, very long simulations are required here to sample these regions of *r*-space with sufficient statistical accuracy to ensure convergence of the effective potentials to their true values.

Two future avenues for application of the current methodology to real macromolecules are envisioned. In the first, homopolymers exhibiting helix-coil transitions depending upon pH level [for example poly-L-lysine and poly-L-glutamic acid (Muroga 2001)] may be studied using the current methods applied to dilute solution wide-angle x-ray scattering data of Grigoryev and others (1971). Similar studies may be performed on carbohydrates or DNA utilizing scattering data and treating sugar or nucleic acid monomers or larger chemical subunits as "residues". Following the approaches of Diaz and others (1993) and Svergun and others (2001), the macromolecular structure factor may be obtained by modeling each "residue" as a monomer with an effective scattering intensity equal to the spherically averaged scattering intensity of the entire chemical sub-unit. The inverse Fourier transform of the experimental structure factor (the structure factor is directly related to the intensity of x-rays scattered by the macromolecule) yields the inter-residue radial distribution function. Taken together with an adequate molecular model, the experimental RDF may be used as input to the SGMC procedure to compute a thermodynamically consistent ensemble of conformations without any adjustable parameters, as demonstrated in this study.

Moreover, application of the method is not limited to equilibrium state points, but may include non-equilibrium state points as well. In the second application envisioned, selective C^{α} labeling (heavy atoms or deuterons) may be used in conjunction with scattering (x-rays or neutrons) to obtain residue-specific radial distribution functions that would enable heteropolymeric systems to be studied, either in their entirety or only in specific regions of interest (e.g. conformationally flexible loop regions in proteins). Though theoretically feasible, current experimental limitations to this approach will determine the degree of its future success.

Appendix A Evaluation of $\partial\langle \mathbf{N} \rangle / \partial \mathbf{A}$

Referring to Eq. (5) and the notation of Section 2 we begin by defining

$$\Omega(q) \equiv \left\{ e^{-\beta U} \prod_{\eta=1}^{M} \prod_{r=0}^{r_{cut}} \left(a^{*}(\eta, r) g_{tar}(\eta, r) \right)^{n(\eta,r)} \right\}$$

where U and $n(\eta,r)$ are each implicitly functions of the conformation variable q. $\langle \mathbf{N} \rangle$ is then

$$\langle \mathbf{N} \rangle = \frac{\sum_{q} \mathbf{N}(q)\Omega(q)}{\sum_{q} \Omega(q)}$$

where the dependence of $\mathbf{N}$ on q has been made explicit. The derivative

$$\frac{\partial \langle \mathbf{N} \rangle}{\partial \mathbf{A}} = \frac{\partial}{\partial \mathbf{A}} \left[\frac{\sum_{q} \mathbf{N}(q)\Omega(q)}{\sum_{q} \Omega(q)} \right]$$

can be evaluated by first separately computing the derivatives of the numerator and denominator as

$$\frac{\partial}{\partial A_{\mu\tau}} \left[\sum_{q} N_{\eta r}(q)\Omega(q) \right] = \frac{1}{A_{\mu\tau}} \sum_{q} N_{\mu\tau}(q) N_{\eta r}(q)\Omega(q)$$

$$\frac{\partial}{\partial A_{\mu\tau}} \left[\sum_{q} \Omega(q) \right] = \frac{1}{A_{\mu\tau}} \sum_{q} N_{\mu\tau}(q)\Omega(q)$$

and then using the product rule to obtain the final result

$$\frac{\partial \langle N_{\eta r} \rangle}{\partial A_{\mu\tau}} = \frac{1}{A_{\mu\tau}} \left[\langle N_{\eta r} N_{\mu\tau} \rangle - \langle N_{\eta r} \rangle \langle N_{\mu\tau} \rangle \right].$$

References for Chapter 1

Bonvin A, Brunger AT. 1995. Conformational variability of solution Nuclear-Magnetic-Resonance structures. Journal of Molecular Biology 250(1):80-93.

Bonvin A, Brunger AT. 1996. Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? Journal of Biomolecular NMR 7(1):72-76.

Cleland RL. 1977. Persistence length of hyaluronic-acid - Estimate from small-angle x-ray scattering and intrinsic viscosity. Archives of Biochemistry and Biophysics 180(1):57-68.

Colhoun FL, Armstrong RC, Rutledge GC. 2002. Analysis of experimental data for polystyrene orientation during stress relaxation using semigrand canonical Monte Carlo simulation. Macromolecules 35(15):6032-6042.

Diaz FG, Cascales JJL, Delatorre JG. 1993. Bead-model calculation of scattering diagrams - Brownian dynamics study of flexibility in immunoglobulin IGG1. Journal of Biochemical and Biophysical Methods 26(4):261-271.

Dickman R, Hall CK. 1986. Equation of state for chain molecules - Continuous-space analog of Flory Theory. Journal of Chemical Physics 85(7):4108-4115.

Flyvbjerg H, Petersen HG. 1989. Error-estimates on averages of correlated data. Journal of Chemical Physics 91(1):461-466.

Grigoryev AL, Volkova LA, Ptitsyn OB. 1971. FEBS Letters 14:189-191.

Groth M, Malicka J, Czaplewski C, Oldziej S, Lankiewicz L, Wiczk W, Liwo A. 1999. Maximum entropy approach to the determination of solution conformation of flexible polypeptides by global conformational analysis and NMR spectroscopy - Application to DNS1-c- D-A(2)bu(2), Trp(4),Leu(5) -enkephalin and DNS1-c- D-A(2)bu(2), Trp(4), d-Leu(5) enkephalin. Journal of Biomolecular NMR 15(4):315-330.

Groth M, Malicka J, Rodziewicz-Motowidlo S, Czaplewski C, Klaudel L, Wiczk W, Liwo A. 2001. Determination of conformational equilibrium of peptides in solution by NMR spectroscopy and theoretical conformational analysis: Application to the calibration of mean-field solvation models. Biopolymers 60(2):79-95.

Henderson RL. 1974. Uniqueness theorem for fluid pair correlation-functions. Physics Letters A A 49(3):197-198.

Kofke DA. 1999. Semigrand canonical Monte Carlo simulation; Integration along coexistence lines. Monte Carlo Methods in Chemical Physics. p 405-441.

Lyubartsev AP, Laaksonen A. 1997. Osmotic and activity coefficients from effective potentials for hydrated ions. Physical Review E 55(5):5689-5696.

Mcgreevy RL, Pusztai L. 1988. Reverse Monte Carlo simulation: A new technique for the determination of disordered structures. Molecular Simulation 1:359-367.

McQuarrie DA. 1975. Statistical Mechanics. New York: Harper & Row. 641 p.

Muroga Y. 2001. Derivation of the small-angle x-ray scattering functions for local conformations of polypeptide chains in solution. Biopolymers 59(5):320-329.

Nikiforovich GV, Marshall GR. 2001. Editorial: Current developments in computational studies of peptides. Biopolymers 60(2):77-78.

Nikiforovich GV, Vesterman B, Betins J, Podins L. 1987. The space structure of a conformationally labile oligopeptide in solution - Angiotensin. Journal of Biomolecular Structure & Dynamics 4(6):1119-1135.

Poland D, Scheraga HA. 1970. Theory of helix-coil transitions in biopolymers; statistical mechanical theory of order-disorder transitions in biological macromolecules. New York: Academic Press.

RosiSchwartz B, Mitchell GR. 1996. Extracting force fields for disordered polymeric materials from neutron scattering data. Polymer 37(10):1857-1870.

Roux B, Simonson T. 1999. Implicit solvent models. Biophysical Chemistry 78(1-2):1-20.

Rutledge GC. 2001. Modeling experimental data in a Monte Carlo simulation. Physical Review E 6302(2):art. no.-021111.

Shenderovich MD, Nikiforovich GV, Saulitis JB, Chipens GI. 1988. Determination of cyclopeptide conformations in solution using NMR data and conformational energy calculations. Biophysical Chemistry 31(1-2):163-173.

Skibinska L, Gapinski J, Liu H, Patkowski A, Fischer EW, Pecora R. 1999. Effect of electrostatic interactions on the structure and dynamics of a model polyelectrolyte. II. Intermolecular correlations. Journal of Chemical Physics 110(3):1794-1800.

Svergun DI. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). Biophysical Journal 77(5):2896-2896.

Svergun DI, Petoukhov MV, Koch MHJ. 2001. Determination of domain structure of proteins from X-ray solution scattering. Biophysical Journal 80(6):2946-2953.

Taketomi H, Ueda Y, Go N. 1975. Studies on protein folding, unfolding and fluctuations by computer-simulation .1. Effect of specific amino-acid sequence represented by specific inter-unit interactions. International Journal of Peptide and Protein Research 7(6):445-459.

Taylor MP, Lipson JEG. 1996. Collapse of a polymer chain: A Born-Green-Yvon integral equation study. Journal of Chemical Physics 104(12):4835-4841.

Torda AE, Brunne RM, Huber T, Kessler H, Vangunsteren WF. 1993. Structure refinement using time-averaged J-coupling constant restraints. Journal of Biomolecular NMR 3(1):55-66.

Zhou YQ, Karplus M. 1997. Folding thermodynamics of a model three-helix-bundle protein. Proceedings of the National Academy of Sciences of the United States of America 94(26):14429-14432.

Zhou YQ, Karplus M. 1999. Interpreting the folding kinetics of helical proteins. Nature 401(6751):400-403.

Zhou YQ, Karplus M, Wichert JM, Hall CK. 1997. Equilibrium thermodynamics of homopolymers and clusters: Molecular dynamics and Monte Carlo simulations of systems with square-well interactions. Journal of Chemical Physics 107(24):10691-10708.

Chapter 2

A coarse-grained molecular model for anionic glycosaminoglycans: Application to chondroitin sulfate and hyaluronan

2.1 Introduction

Proteoglycans are high molecular weight comb biopolymers that play a fundamental role in the biomechanical properties of the extra-cellular matrices (ECM) of various tissues, including articular cartilage, skin, and corneal stroma. They consist of a linear protein backbone with covalently attached linear glycosaminoglycan (GAG) chains. Of particular interest is aggrecan, the predominant proteoglycan in articular cartilage. A single aggrecan molecule consists of ~ 100 anionic chondroitin sulfate (CS) GAGs (each 10-30 kDa, 20-60 disaccharides, or 200-600 Å contour length) covalently linked to serine-glycine residue pairs on its ~ 300 kDa protein backbone (~ 4000 Å contour length), as well as ~ 60 keratan-sulfate (KS) oligosaccharides of considerably lower molecular mass. With the aid of link protein, aggrecan associates non-covalently with the high molecular mass anionic GAG hyaluronan, which serves to retain aggrecan in the ECM. The abundance of charged CS present on aggrecan generates an osmotic swelling pressure that is opposed by tensile stresses in the elastic collagen network. The combination of these effects provides cartilage with its unique biomechanical properties, namely a very low coefficient of sliding friction even under substantial compressive loads, which serves to protect the ends of bones from wear during joint locomotion.

Interestingly, the sulfation type (4- vs. 6-sulfation of chondroitin disaccharides), pattern, and molecular weight of CS chains present on aggrecan varies significantly with disease (osteoarthritis or rheumatoid arthritis), age, depth within the cartilage layer, and anatomical site (Bayliss and others 1999; Hardingham 1998; Lewis and others 1999; Plaas and others 1998; Plaas and others 1997; Platt and others 1998; Roughley and Lee 1994) concomitantly with the biomechanical and biochemical properties of articular

cartilage. For these reasons of biomechanical and biological relevance, it is of interest to characterize the solution conformation and osmotic pressure of chondroitin sulfate under physiological conditions, as well as to investigate their dependencies on sulfation type, MW, pH, and ionic strength. Although ultimately our aim is to simulate multiple aggrecan molecules interacting in solution at physiological concentration, in the present study we focus on developing a systematically coarse-grained GAG model that is a stepping-stone towards that goal. We develop and apply the model to single GAGs (i.e., in infinitely dilute solution) in the present study, and examine the effects of finite, physiological GAG concentration, in addition to sulfation pattern, on CS conformation and osmotic pressure in a subsequent work.

Unlike chondroitin sulfate (CS), the solution conformation and titration behavior of HA has been extensively characterized by a variety of experimental techniques including light scattering, neutron scattering, viscometry, and potentiometric titration (Buhler and Boue 2003; Buhler and Boue 2004; Cleland 1977; Cleland 1991; Cleland and others 1982; Fouissac and others 1992; Ghosh and others 1990; Gribbon and others 2000; Hayashi and others 1995; Hayashi and others 1996; Mathews and Decker 1977; Reed and others 1989; Scott and Heatley 1999; Tsutsumi and Norisuye 1998). The intrinsic persistence length (the persistence length in the limit of infinite ionic strength) has been measured by various groups to be between 40-90 Å (Buhler and Boue 2004; Cleland 1977; Fouissac and others 1992; Ghosh and others 1990; Hayashi and others 1995), and the *apparent* persistence length (the persistence length in the presence of long-ranged, excluded volume interactions) has been measured to be proportional to κ^{-1}, where κ is the inverse Debye length (Ghosh and others 1990). On the theoretical side,

Reed and Reed (1991) used an off-lattice Monte Carlo simulation algorithm to simulate HA in solution and found agreement with the κ^{-1} persistence length scaling. More recently, Almond and coworkers performed a series of meticulous studies using fully atomistic models of solvated HA di- (Almond and others 1997), tetra- (Almond and others 1998), and decasaccharides (Almond and others 2000) to illuminate the trans-glycosidic linkage hydrogen-bonding structure and its effect on glycosidic linkage flexibility. Conformational results from the decasaccharide simulation were extrapolated to estimate a persistence length for HA of 139 ± 22 Å in the absence of added salt or cations. The effects of physiological ionic strength on conformation, however, were not examined. At the other extreme, Haxaire and others (2000) studied the infinite ionic strength conformation of HA, finding an intrinsic persistence length of 75 Å. They also performed a detailed study of the stability of various helical forms of HA in the solid-state by comparing with fiber diffraction data. Almond and Sheehan (2000) also studied the conformation of chondroitin tetrasaccharides using an explicit solvent model without added salt and contrasted its behavior with that of HA.

Several simplified, coarse-grained models of HA in solution have been developed to study of its titration behavior and conformation. One of the earliest models employed RIS theory to simulate its unperturbed state (Cleland 1971). The model was later extended to simulate the titration behavior of HA using an extended Bragg-Williams approximation for the discrete charge site Debye-Hückel interaction energy together with either a rigid worm-like chain approximation or the Metropolis Monte Carlo method (Cleland 1984). The aforementioned work was limited to studying the titration behavior of HA in finite ionic strength solutions and did not study HA conformation. As

mentioned previously, a Metropolis Monte Carlo algorithm with a three-state RIS model and Debye-Hückel theory was used to simulate HA titration and conformational properties simultaneously (Reed and Reed 1991; Reed and Reed 1992). That model was not based on an underlying atomic model of HA, however, and was therefore incapable of predicting the intrinsic persistence length of HA ab initio, instead treating it as an adjustable fitting parameter. With regards to chondroitin sulfate, there is only one study we are aware of that examined its solution conformation (Rodriguez-Carvajal and others 2003). In the limit of infinite ionic strength, C4S was determined to be stiffer than CH, which in turn was found to be stiffer than C6S. The finding that C4S is intrinsically stiffer than C6S is in contradiction to experimental results of Tanaka (1978a) who found the reverse to be true. The present study may be viewed as an extension of that of Rodriguez-Carvajal and others (2003) in that we examine the effects of finite ionic strength and pH on the conformation and titration behavior of CH, C4S, and C6S, in addition to presenting results for the high ionic strength limit.

The goal of the present study is to develop a molecular-level model for CS that simultaneously predicts its titration behavior, solution conformation, and osmotic pressure. A conventional, all-atom simulation of the polysaccharide and its suspending solvent (water plus counter ions and added salt ions) could, in principle, provide the conformational and thermodynamic properties of interest, but that approach is precluded by the large system sizes of interest. For that reason, we developed a coarse-grained model that is similar in spirit to the aforementioned models of HA as well as models for other aqueous polysaccharides such as pullulan, xythan (Perico and others 1999), cellulose, and laminaran (Cleland 1971). We begin with explicit, all-atom models of the

constituent GAG disaccharides. Through systematic coarse-graining we arrive at a molecular model that is free of adjustable parameters and capable of predicting quantitatively the solution titration behavior, conformation, and osmotic pressure of CS solutions of physiologically relevant molecular weight and concentration, while retaining the molecular-level identity of the individual monosaccharides (e.g., the sulfate type and pattern). The linearized Poisson Boltzmann equation is used to calculate the electrostatic energy, where it is assumed that each GAG is suspended in an aqueous 1:1 salt solution that is in osmotic equilibrium with a reservoir of fixed ionic strength. The model is applied to HA in addition to CS because of the relevance of HA to cartilage and synovial fluid biomechanics, as well as to provide an experimentally well-characterized benchmark molecule with which to validate our GAG modeling approach.

2.2 Modeling

Chondroitin is a linear (unbranched) polysaccharide consisting of repeating disaccharide units of D-glucuronic acid (GlcUA) and N-acetyl-galactosamine (GalNAc), alternately linked in $\beta 1,3$ and $\beta 1,4$ glycosidic linkages, respectively (Figure 1). C4S and C6S consist of repeating chondroitin disaccharides that are sulfated at the 4- or 6-carbon of N-acetyl-D-galactosamine, respectively (Figure 1). Hyaluronan is similar in chemical structure to chondroitin, consisting of repeating disaccharide units of D-glucuronic-acid and N-acetyl-D-glucosamine (GlcNAc) (Figure 1), but is typically of very high molecular mass, ranging from 10^5 to 10^6 Da, equivalent to about 250 to 2,500 disaccharide repeat units. For ease of notation, A, B, C, D, and E, will be used to refer to the monosaccharide units, GlcUA, GalNAc, GlcNAc, GlcNAc4S, and GlcNAc6S, respectively.

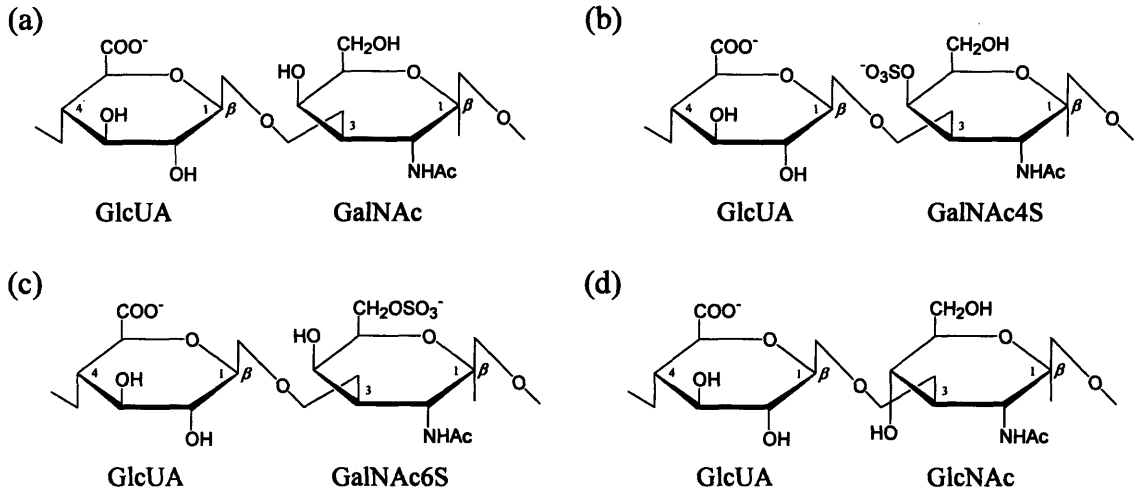


Figure 1 Disaccharide repeat units of chondroitin (a), chondroitin 4-sulfate (b), chondroitin 6-sulfate (c), and hyaluronan (d).

2.2.1 Topology

Following previous polysaccharide modeling studies (Brant and Goebel 1975; Perico and others 1999), the coarse-grained model backbone is defined by the sequence of chemical and virtual bonds depicted in Figure 2.

(a)

(b)

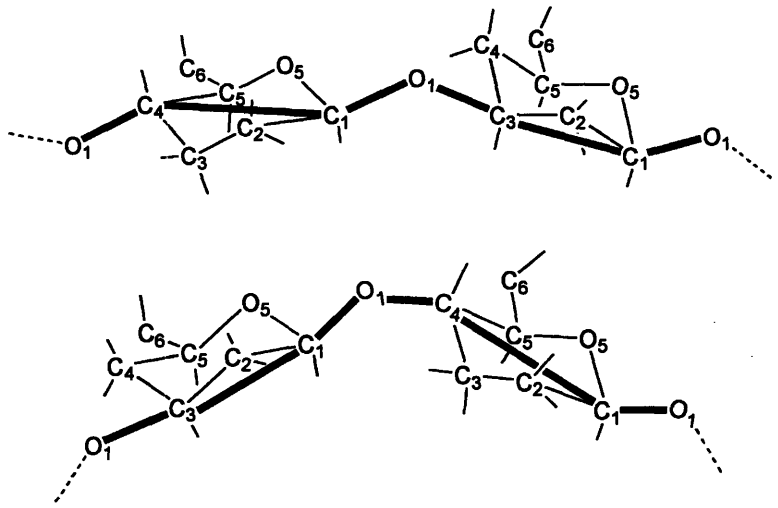


Figure 2 Definition of the coarse-grained model bonded backbone structure (thick solid lines) based on the all-atom disaccharide representation for $\beta 1,3$ (a) and $\beta 1,4$ (b) linkages.

Each monosaccharide unit (A-E) consists of three backbone sites that coincide with atoms CJ, C1, O1, (J = 3 for units B-E and J = 4 for A) and define the polysaccharide topology (Figure 2). Additional center of charge and repulsive Lennard-Jones sites (not shown in Figure 2) are used to model non-bonded electrostatic interactions and steric interactions, respectively, and are described in the subsequent modeling section. All bond lengths, valence angles, and dihedral torsion angles in the coarse-grained model are treated as rigid, except for the glycosidic torsion angles, which provide the majority of the conformational flexibility in polysaccharides (Brant and Goebel 1975; Cleland 1971; Perico and others 1999; Rodriguez-Carvajal and others 2003).

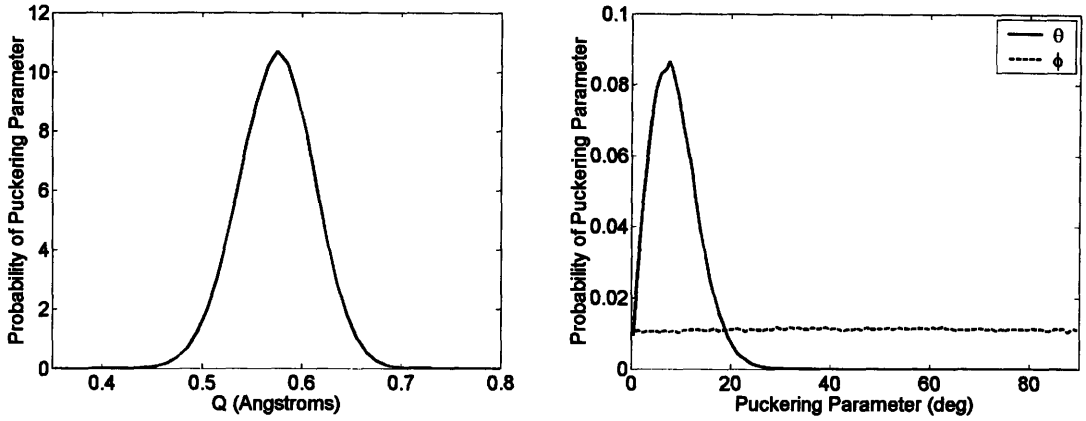


Figure 3 Cremer-Pople ring puckering parameter probability distributions for GlcUA at the non-reducing end of a CH β1,3 disaccharide.

Cremer-Pople ring puckering parameters may be used to measure the equilibrium conformation of pyranose rings (Cremer and Pople 1975). Six membered pyranose rings require the three puckering parameters, (Q,θ,ϕ), where a value of $\theta = 0°$ indicates a stable $^4C_1$ chair conformation. Analyses of the (Q,θ,ϕ) probability distributions computed from all-atom solvated disaccharide simulations demonstrate that each type of ring in the GAGs studied is stable in a slightly distorted $^4C_1$ chair conformation (e.g., Figure 3), independent of whether it is located at the reducing or non-reducing terminus (Table 1), thus lending support to the rigid-monosaccharide assumption.

Table 1 Mean values of the Cremer-Pople puckering parameters for each pyranose ring at the reducing and non-reducing ends of each disaccharide (HA, CH, C4S, C6S). The mean value of the third puckering parameter, ϕ, is not included because it was uniformly distributed between 0 and 90° in every case (e.g., Figure 3). $\theta = 0°$ indicates an exact chair conformation (Cremer and Pople 1975) and the absence of error estimates indicates that the estimated error was less than the smallest significant figure shown.

		At Reducing Terminus		At Non-reducing Terminus	
		Q (Å)	θ (°)	Q (Å)	θ (°)
HA	**GlcUA**	0.58	8.7 ± 0.4	0.574 ± 0.002	8.6 ± 0.3
	GlcNAc	0.58	8.4 ± 0.3	0.574 ± 0.002	8.7 ± 0.3
CH	**GlcUA**	0.58	8.6 ± 0.3	0.574 ± 0.002	8.8 ± 0.2
	GalNAc	0.57	8.7 ± 0.4	0.580 ± 0.001	8.2 ± 0.4
C4S	**GlcUA**	0.58	8.6 ± 0.2	0.573 ± 0.00	8.8 ± 0.3
	GalNac4S	0.58	8.1 ± 0.6	0.576 ± 0.003	8.0 ± 0.2
C6S	**GlcUA**	0.58	8.6 ± 0.4	0.574 ± 0.001	8.7 ± 0.3
	GalNac6S	0.57	9.1 ± 0.5	0.581 ± 0.002	8.2 ± 0.5

Values of the internal degrees of freedom of the coarse-grained model were assigned from the corresponding mean values obtained from equilibrium room temperature all-atom simulations of isolated, solvated disaccharides (Table 2). In the disaccharide simulations water was treated implicitly as a uniform dielectric medium with dielectric permittivity, $\varepsilon = 78.5$. In this somewhat crude approximation for the solvent, the effects on conformation induced by the low dielectric of the sugar region, the hydrophobic effect, as well as the counter-ions associated with the carboxylate and sulfate groups were neglected. We opted not to use one of the more recent implicit solvent models (Roux and Simonson 1999) such as the Generalized Born treatment of electrostatics with the Solvent-Accessible-Surace-Area (SASA) model for the hydrophobic effect due to the lack of well-validated parameter values for polysaccharides.

Table 2 Coarse-grained model bond lengths, b, valence angles, θ, and dihedral torsion angles, ν. A prime denotes that the atom belongs to the reducing sugar terminus. ϕ and ψ denote the glycosidic torsion angles, the only flexible internal degrees of freedom present in the model. Q denotes the center of charge site used to model electrostatic interactions between charged monosaccharide units and V denotes the repulsive Lennard-Jones site used to model steric interactions between monosaccharides.

Internal Coordinate (Å, degrees)	**GlcUA** $J=4, I=3$	**GlcNAc** $J=3, I=4$	**GalNAc** $J=3, I=4$	**GalNAc4S** $J=3, I=4$	**GalNAc6S** $J=3, I=4$
b(O1'–CJ')	1.42	1.42	1.42	1.42	1.42
b(CJ'–C1')	2.89	2.52	2.52	2.52	2.53
b(C1'–O1')	1.41	1.41	1.41	1.41	1.41
b(C1'–V')	1.69	1.51	1.37	1.72	1.48
b(C1'–Q')	3.31	*na*	*na*	5.78	5.90
θ(C1–O1–CJ')	117.5	117.7	117.6	117.2	117.7
θ(O1–CJ'–C1')	147.8	144.0	145.1	144.4	145.0
θ(CJ'–C1'–O1')	148.4	144.6	144.4	144.4	144.2
θ(CJ'–C1'–V')	6.1	7.2	10.0	19.4	26.3
θ(CJ'–C1'–Q')	42.8	*na*	*na*	53.0	86.3
ϕ(CI–C1–O1–CJ')	-173.2	-110.9	-111.1	-112.2	-111.8
ψ(C1–O1–CJ'–C1')	-179.6	-123.2	-142.6	-137.0	-145.9
ν(O1–CJ'–C1'–O1')	171.4	2.0	-1.7	-0.9	-1.6
ν(O1–CJ'–C1'–V')	-76.7	-84.9	-31.8	-74.7	-113.4
ν(O1–CJ'–C1'–Q')	93.5	*na*	*na*	-73.9	-116.4

2.2.2 Bonded energy

In the limit of infinite ionic strength, the conformational energy and statistics of successive glycosidic linkage torsion angles are assumed to be independent. This modeling assumption implies that specific chemical interactions (e.g., hydrogen bonding) occur only across single glycosidic linkages and is justified by the presence of the nearly rigid intervening sugar monomers that provide spatial separation between successive linkages (Brant and Christ 1990). Individual torsion angles, ϕ and ψ, within a given

linkage, however, are typically highly interdependent and were treated as such. Simple Boltzmann inversion was used to compute and tabulate potentials of mean force (PMFs) for the glycosidic torsion angles from equilibrium probability distributions obtained from the isolated all-atom disaccharide simulations, $F_{kl}^{B}(\phi,\psi) = -k_B T \ln P_{kl}(\phi,\psi)$, where k and l refer to the monosaccharide types A, B, C, D, or E, and there are eight different PMFs for the GAGs modeled (AB, AC, AD, and AE for the $\beta 1,3$ linkage; BA, CA, DA, and EA for the $\beta 1,4$ linkage). The tabulated free energies were subsequently employed in the coarse-grained simulations and implicitly include the effects of covalent, van der Waals, solvent-mediated electrostatic interactions, and hydrogen-bonding occurring across a single glycosidic linkage, as well as the configurational entropy of the internal disaccharide degrees of freedom. The total bonded free energy of a GAG consisting of N monosaccharide units in a conformation specified by the set of glycosidic torsions $\{\phi,\psi\}$ is given by,

$$F^{B}(\{\phi,\psi\}) = \sum_{\substack{i=1 \\ j=i+1}}^{N-1} F_{i(k)j(l)}^{B}(\phi,\psi) = -k_B T \sum_{\substack{i=1 \\ j=i+1}}^{N-1} \ln P_{i(k)j(l)}(\phi,\psi) \qquad (1)$$

where a single summation is performed over all glycosidic linkages, with the constraint that for each monomer $i, j = i+1$, and $i(k)$ and $j(l)$ denote the types of monomers (k,l) at linkage (i,j).

Two alternative approaches for computing $F_{kl}^{B}(\phi,\psi)$ that are frequently employed in polysaccharide simulations were also investigated. These were: (1) treating each monosaccharide as rigid in its minimum energy configuration and stepping incrementally through (ϕ,ψ)-space to compute the bonded energy at every grid point and (2)

performing the same procedure as (1) followed by a constrained energy minimization at every grid point to obtain a local energy minimum. In each of these approaches, the configurational entropy of the non-(ϕ,ψ) internal degrees of freedom is neglected. The first approach immediately proved inadequate due to steric clashing of the bulky side groups present in neighboring monosaccharides, which artificially precludes normally accessible regions of (ϕ,ψ)-space. Energy minimization employed in the second approach relieved the problem of steric clashing encountered in (1). However appreciably different energy surfaces were obtained with respect to the finite temperature simulations, due to the effects of configurational entropy of the non-(ϕ,ψ) degrees of freedom included in the latter. For this reason, we opted to employ the finite temperature simulation approach to compute $F_{kl}^{B}(\phi,\psi)$.

In general, $F_{kl}^{B}(\phi,\psi)$ will depend on environmental conditions such as temperature, ionic strength, and external pH, the latter two of which may affect the protonation state of the carboxylate group on GlcUA and hence specific inter-glycosidic linkage interactions such as hydrogen bonding between the carboxylate group and the adjacent monosaccharides. Since our interest is limited to room temperature conditions in this study, we only tested the effects of ionic strength and pH on $F_{kl}^{B}(\phi,\psi)$ by repeating each disaccharide simulation (1) with a protonated carboxylate group and (2) at 1 M ionic strength using an all-atom Debye-Hückel interaction energy implemented in CHARMM. Interestingly, the various $F_{kl}^{B}(\phi,\psi)$ obtained were insensitive to both the protonation state of the carboxylate group and ionic strength. Accordingly, the generally

solvent-dependent $F_{kl}^{B}(\phi,\psi)$ were approximated to be independent of pH and ionic strength, for the purposes of this study.

The insensitivity of $F_{kl}^{B}(\phi,\psi)$ to ionic strength is attributed within the context of the present, simple implicit solvent model to the energetic dominance over electrostatic interactions of (1) covalent interactions occurring across glycosidic linkages, including the exo-anomeric effect, (2) steric interactions due to trans-glycosidic linkage van der Waals interactions, and (3) trans-glycosidic linkage hydrogen bonding, which is modeled in CHARMM by a combination of van der Waals and Coulombic terms, both of which are present in our implicit solvent model at 1 M ionic strength because the Debye length is 3 Å and hydrogen bond interactions take place over length scales that are < 3 Å. The insensitivity of $F_{kl}^{B}(\phi,\psi)$ to the protonation state of the carboxylate group on GlcUA is to be reasonably expected in our model because protonation did not significantly alter the partial charge distribution on the carboxylate group, so that hydrogen bonds between it and groups belonging to adjacent monosaccharides were not disrupted. Further analysis of the specific energetic contributions to $F_{kl}^{B}(\phi,\psi)$, as well as their sensitivity to the particular solvent model employed, is certainly warranted but beyond the scope of the present study.

At this point we make a necessary digression to point out that two definitions for the glycosidic torsion angles are required in this study: (1) the conventional hydrogen bond definition: $\phi^{H}(\mathrm{H1}-\mathrm{C1}-\mathrm{O}J'-\mathrm{C}J')$ and $\psi^{H}(\mathrm{C1}-\mathrm{O}J'-\mathrm{C}J'-\mathrm{H}J')$, advocated by the (Nomenclature 1971), where a prime denotes an atom belonging to the reducing sugar and $J'=3'$ or $4'$ depending on the linkage type and (2) a definition based on the virtual

bonds in the coarse grained model, namely $\phi^V(CI-C1-OJ'-CJ')$ and $\psi^V(C1-OJ'-CJ'-C1')$, where $I=3$ when $J'=4'$ and $I=4$ when $J'=3'$. To be consistent with convention we employ the first definition for the presentation of results from this point forward, and therefore drop the use of superscripts. To afford the most accurate mapping from the all-atom model to the coarse grained model, however, we employed the virtual bond definition to define the coarse-grained model topology and the glycosidic torsion angle potentials of mean force. Glycosidic torsion angle results computed using the coarse-grained model may be subsequently transformed to the hydrogen bond definition for purposes of presentation using the following torsion angle offsets that were computed as ensemble averages from the all-atom disaccharide simulations, $\Delta\phi \equiv \langle\phi^H\rangle - \langle\phi^V\rangle$ and $\Delta\psi \equiv \langle\psi^H\rangle - \langle\psi^V\rangle$ (Table 3).

Table 3 Offsets in hydrogen bond and virtual bond glycosidic torsion angle definitions.

Monosaccharide	$\Delta\phi$ (deg)	$\Delta\psi$ (deg)
GlcUA	234.9	174.9
GlcNAc	175.5	126.9
GalNAc	174.4	126.1
GalNAc4S	174.7	125.3
GalNAc6S	174.5	126.3

2.2.3 Non-bonded energy

A charge site (Q) was defined for each charged monosaccharide (GlcUA, GalNAc4S, GalNAc6S) by computing its mean center of charge from the explicit atom disaccharide simulations and defining its mean virtual bond length, valence angle, and

dihedral torsion angle (Table 2). Electrostatic interactions between charge sites were computed using a Debye-Hückel interaction potential assuming zero ionic radii,

$$u^{DH}(r_{ij}) = k_B T \lambda_B \frac{e^{-\kappa r_{ij}}}{r_{ij}} \tag{2}$$

where k_B is the Boltzmann constant, T is temperature, κ is the inverse Debye length $\kappa^2 = 8\pi\lambda_B N_A c_s$, where N_A is Avogadro's number, c_s is molar ionic strength of the fully dissociated 1:1 electrolyte, $\lambda_B = e^2 / \varepsilon k_B T$, is the Bjerrum length, defined to be the distance at which the Coulombic interaction energy of two monovalent charges embedded in a uniform dielectric medium is equal to $k_B T$, where ε is dielectric permittivity, e is the electronic charge ($\lambda_B = 7.14$ Å in water at 298 K), r_{ij} is the distance of separation between center of charge sites *i* and *j*, and all charge sites have been assumed to have charge $-e$ (monovalent sulfate and carboxylate groups). Higher order intra-molecular non-bonded electrostatic interactions such as dipole-monopole, dipole-dipole, etc. were neglected in this approximation. Non-bonded van der Waals and hydrophobic interactions were also ignored, which implicitly assumes that bonded and non-bonded electrostatic interactions dominate in determining the conformation of anionic GAGs. The neglect of van der Waals, hydrophobic, and higher order electrostatic interactions will be most valid at Debye lengths that are larger than the monosaccharide length scale of 5-10 Å.

The Debye-Hückel interaction potential in Eq. (2) is derived from the electrostatic potential solution to the linearized Poisson Boltzmann equation, also called the Debye-

Hückel equation (Debye and Hückel 1923), assuming spherical symmetry and linearization about a zero electrostatic potential (McQuarrie 1975). The Poisson-Boltzmann equation, and therefore the Debye-Hückel interaction potential as well, implicitly assumes uniform electrochemical potentials of the salt species. In modeling a single GAG using the Debye-Hückel potential, we are therefore implicitly assuming that the polyelectrolyte is suspended in a salt solution in which the salt species are in electrochemical equilibrium with a salt reservoir of fixed molarity, where the reference electrostatic potential is assumed to be zero. Moreover, the Poisson-Boltzmann equation assumes that the chemical (non-electrostatic) contribution to the activity coefficients of the salt species is unity (ideality). In using the Debye-Hückel interaction potential with zero ionic radius we have neglected several potentially important effects. These include the low dielectric and salt exclusion of the GAGs, as well as the distributed, asymmetric charge distributions of the various monosaccharides, as noted above. We have nonetheless employed Eq. (2) to model electrostatic interactions, as is common in the literature, with the zero ionic radii assumption as a first modeling attempt to keep the physical model as simple as possible with no adjustable parameters. The current model may be extended in a straightforward manner to account for the aforementioned higher order effects without altering the basic modeling formalism.

According to the Manning criterion (Manning 1969), if the equivalent line charge density of a polyelectrolyte, $\xi \equiv \lambda_B / d$, is greater than unity, where d is the inter-charge spacing, counterions will "condense" on the polyelectrolyte until its effective charge density is reduced to unity. Whereas chondroitin and hyaluronan have equivalent line charge densities of 0.7 (assuming an inter-charge spacing equal to the approximate

disaccharide repeat unit length of 10 Å) and therefore meet the criterion for full ionization, chondroitin sulfate has an equivalent line charge density of 1.4, and would therefore be partially deionized to reduce its effective line charge density to unity, according to Manning's theory. In utilizing the Debye-Hückel equation to model the electrostatic free energy of chondroitin sulfate, one would therefore be required to scale down the carboxylate and sulfate charges to $0.7e$ in order to strictly abide by Manning's criterion. (The alternative of solving the non-linear Poisson-Boltzmann equation for every configuration of chondroitin sulfate sampled is intractable given current computational resources and the molecular weight range of interest in this study.) Manning's criterion, however, is strictly valid only for idealized systems such as line charges and cylindrical cell models, neither of which takes into account the detailed molecular nature of chondroitin sulfate, including the distribution of atomic partial charges and its molecular shape. It is our view that the linear charge density of chondroitin sulfate is close enough to the borderline case of unity that in order to accurately determine the validity of employing the Debye-Hückel equation with fully ionized charge groups one would need to perform a detailed analysis employing the non-linear Poisson-Boltzmann equation and an atomic-level model that accounts for the distribution of atomic partial charges in chondroitin sulfate as well as the shape of its low dielectric and salt-excluding domain. Such a level of analysis is beyond the scope of the present study; we retain the fully ionized charge values of the carboxylate and sulfate groups in analyzing the conformation of chondroitin sulfates using Eq. (2).

A van der Waals site (V) was defined for each monosaccharide by computing its mean center of geometry (defined as the center of mass with a unit mass weighting for

each atom type) from the explicit atom disaccharide simulations and defining its mean virtual bond length, valence angle, and dihedral torsion angle (Table 2). Steric interactions between monosaccharides are modeled using the standard repulsive Lennard-Jones potential,

$$u^{RLJ}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4}\right]; & r < r_c \\ 0; & r \geq r_c \end{cases} \quad (3)$$

where $r_c = 2^{1/6}\sigma$ is the cut-off radius, and ε and σ are the van der Waals energy and radius parameters. σ was equated with the sphere radius obtained from equating the volume of a sphere with that obtained from the molecular volume of a monosaccharide (σ = 3.29, 3.56, 3.56, 3.66, and 3.66 Å for GlcUA, GalNAc, GlcNAc, GlcNAc4S, and GlcNAc6S), calculated using a 1.4 Å spherical probe. ε was taken to be the value for carbon-carbon interactions (0.15 *kcal/mol* = 0.253 k_BT).

Accounting for bonded and non-bonded interactions, the free energy of the GAG model in a conformation defined by the set of glycosidic torsion angles $\{\phi,\psi\}$ is,

$$F(\{\phi,\psi\}) = \sum_{i,j=i+1} F_{ij}^{B}(\varphi,\psi) + \sum_{i<j}[u^{DH}(r_{ij}) + u^{RLJ}(r_{ij})] \quad (4)$$

where the first summation is a single summation over all glycosidic linkages in the model and the second summation is a double summation over all pairs of monosaccharides, excluding nearest-neighbors, whose electrostatic interaction is included in the bonded

PMF. It is noted that the Debye-Hückel energy is zero if either monosaccharide has zero net charge and the repulsive Lennard-Jones potential need not be activated.

2.2.4 Titration

The theoretical study of the titration of macromolecules has been modeled and studied extensively using discrete site models for proteins and nucleic acids (Bashford and Karplus 1990; Beroza and others 1991; Havranek and Harbury 1999; Hill 1956; Tanford 1957; Tanford and Kirkwood 1957) as well as for polyelectrolytes (Jonsson and others 1996; Sassi and others 1992; Ullner and Jonsson 1996), such as hyaluronic acid (Cleland 1984; Reed and Reed 1992). Titratable sites in polyacids interact electrostatically due to the spatial proximity that is imposed by backbone connectivity, leading to markedly different titration behavior from isolated monoacids. Unfavorable electrostatic interaction involving titratable sites present in polyacids is relieved by protonation, which manifests in a variable apparent acidity that increases with increasing degree of ionization. Salt screening and conformational flexibility directly mediate the strength of electrostatic interactions, introducing additional complexity into the titration behavior of polyacids.

The apparent acidity of a polyacid is expressed by defining an apparent dissociation constant, pK^{app} that, unlike for monoacids, is a function of the average degree of ionization of the polyacid, $\alpha \equiv \left\langle \sum_i \alpha_i \right\rangle$, where α_i is the ionization state (0 or 1 for deionized or ionized, respectively) of the i^{th} titratable site,

$$pK^{app} \equiv \text{pH} - \log\frac{\alpha}{1-\alpha}. \tag{5}$$

To model the titration of GAGs, we follow (Reed and Reed 1992) and (Ullner and Woodward 2000) in employing a semi-grand canonical ensemble in which the titrating sites are the chemical 'species'. The titrating sites in CH, CS, and HA are the weak acid carboxylate groups present on glucuronic acid monomers. The sites are assumed to have a variable state of protonation, $x_i = \{0,1\}$, where x_i denotes the number of protons bound to the i^{th} site. The sulfate groups present in CS are strong acids and are thereby assumed to remain ionized at all pH. For a given conformation, $\{\phi,\psi\}$, of a GAG with a total of N titrating and non-titrating monosaccharide units, the free energy associated with protonation of a sub-set of P titrating sites may be written:

$$F^{prot}(\{\phi,\psi\};\{\mathbf{x}^P\}) = \sum_{i=1}^{P}\mu_i x_i + \sum_{i<j}^{N} F_{ij}^{elec}(q_i,q_j) \tag{6}$$

where $F_{ij}^{elec}(q_i,q_j)$ is now the electrostatic interaction energy between the charge sites on monosaccharide units i and j, including *all* charge group pairs (nearest neighbors, whether protonatable or not), P is the number of titratable sites, and $\mu_i = \varepsilon_i - \mu_{H^+} = k_B T[\ln 10](\text{pH} - \text{p}K_i^{int})$ is the total chemical potential associated with protonation. $\mu_{H^+} = -k_B T \ln(10)\text{p}H$ is the free energy change of protonation associated with removal of a hydrogen ion from solution and $\varepsilon_i = -k_B T \ln(10)\text{p}K_i^{int}$ is the intrinsic free energy change associated with protonating site i in an otherwise electrically neutral

molecule (Beroza and others 1991). As pointed out by (Beroza and others 1991), ε_i is a constant in rigid macromolecules that accounts for free energy contributions due to solvation, electrostatic interactions of the site with the fixed charge on the polyacid backbone, and standard-state quantum mechanical bonding effects. In conformationally flexible molecules such as those considered here, however, ε_i will generally be a function of conformation. Despite this complicating factor, we follow (Ullner and Jonsson 1996) and (Reed and Reed 1992) in treating ε_i as an ensemble averaged quantity that is measured experimentally and expressed in terms of the intrinsic dissociation constant, $\mathrm{p}K^{int}$.

The intrinsic dissocation constant of hyaluronic acid has been measured experimentally to be $\mathrm{p}K^{int} = 2.9 \pm 0.1$, (Cleland and others 1982). In the present study we assume that chondroitin has the same $\mathrm{p}K^{int}$ as hyaluronic acid. This assumption is justified because the molecules are chemically nearly identical, differing only in the stereochemistry of the hydroxyl group at the C_4 of GlcNAc. The preceding titration model may then be used directly to study the titration of both chondroitin and chondroitin sulfate by performing single site protonation/deprotonation moves, in which proton exchanges with a hydrogen ion reservoir of fixed pH are simulated, in addition to conformation-altering moves. The protonation/deprotonation moves are accepted or rejected according to the standard Metropolis criterion, $rn \leq \min\{1, e^{-\beta \Delta F^{prot}}\}$, where rn is a pseudo-random deviate between 0 and 1 and $\beta = (k_B T)^{-1}$. Configurational and thermodynamic averages, including the mean overall degree of ionization, α, are

computed as ensemble averages over the equilibrium distribution of states in protonation-configuration space.

2.3 Simulation protocol

2.3.1 All-atom disaccharide simulations

All-atom disaccharide simulations were performed using CHARMM (Brooks and others 1983) employing the all-atom polysaccharide force field provided with Quanta by MSI Inc. The Metropolis Monte Carlo method in the canonical ensemble was used to sample equilibrium configuration space. Dihedral torsion angle and Cartesian displacement moves were performed with equal frequency. 10^6 moves were used for equilibration followed by 3×10^7 moves to compute averages. Isolated disaccharides were simulated at 298 K using atom-based van der Waals and Coulombic interactions with dielectric permittivity equal to that of water ($\varepsilon = 78.5$) and no cut-offs.

2.3.2 Coarse-grained GAG model simulations

The Metropolis Monte Carlo (MMC) algorithm in the canonical ensemble was employed to simulate single isolated GAGs consisting of up to 1000 disaccharides. Trial configurations were generated using the pivot move, in which a glycosidic torsion angle was selected at random and changed by a random amount between $[-\Delta_{tormax}, \Delta_{tormax}]$, where Δ_{tormax} was adjusted on the fly to maintain an acceptance/rejection ratio between 40% to 60%. The pivot algorithm was first introduced by (Lal 1969) and later shown to be highly effective for the simulation of self-avoiding random walks by (Madras and Sokal 1988). 10^6 cycles were used to equilibrate the molecule followed by 10^6 cycles to compute equilibrium conformational and thermodynamic properties. Each cycle consisted of performing N pivot moves, where N is equal to the number of glycosidic linkages in the GAG. A non-bond electrostatic cut-off of three Debye lengths was used

and the Debye-Hückel energy shifted to equal zero at the cut-off. Configuration-altering moves were accepted according to the standard Metropolis criterion, $rn \leq \min\{1, e^{-\beta\Delta F}\}$, where $\Delta F \equiv F^{new} - F^{old}$ is the difference in configurational free energy between the new and old configurations.

For the titration simulations, single site protonation/deprotonation moves were performed in addition to configuration altering pivot moves. The protonation/deprotonation moves simulate the exchange of protons with a hydrogen reservoir of fixed pH and hence chemical potential, as described earlier. Protonation/deprotonation moves and pivot moves were selected at random to satisfy detailed balance and were performed with a relative frequency of P:1, where P is the number of titratable sites in the molecule. 25×10^6 cycles were used to equilibrate the molecule followed by 50×10^6 cycles to compute equilibrium averages, where a cycle now consists of N pivot moves + NP protonation/deprotonation moves. As opposed to previous Monte Carlo titration studies of proteins, in which strong coupling between titrating sites (Beroza and others 1991) required intramolecular proton exchange moves, efficient sampling was achieved in the present study using single site bulk-exchange moves.

2.4 Results

2.4.1 Glycosidic torsion angle potentials of mean force

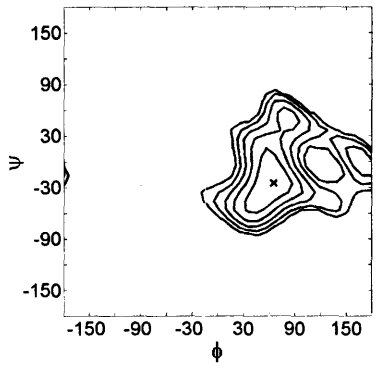
The glycosidic dihedral angle potentials of mean force computed from all-atom disaccharide simulations are shown in Figure 4 for each disaccharide (CH, C4S, C6S, and HA) and linkage type (β1,3 and β1,4). Contour lines of constant energy are drawn in increments of k_BT above the minimum energy conformation which is denoted by an 'X' in the plot. We reiterate that trans-glycosidic Coulombic interactions between sulfate and carboxylate groups have a negligible influence on the C4S and C6S PMFs. This implies that trans-glycosidic covalent bonded, van der Waals, and hydrogen bond interactions were the primary determinants of the PMFs.

The PMFs for the β1,4 linkages are highly similar across all disaccharide types (Figure 4, b, d, f, h). The reason for the similarities is that the chemical perturbations at the C_4 and C_6 of GlcNAc and GalNAc are located distal to the β1,4 linkage region and thus do not influence the specific trans-glycosidic interactions. Even the influence of Coulombic repulsion between the sulfate and carboxylate groups in C4S and C6S is negligible on the linkage flexibility.

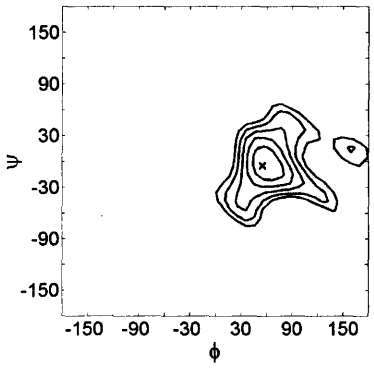
Turning to the β1,3 linkages, the CH and C6S PMFs are similar because the sulfate group in C6S is located distal from the linkage region (Figure 4 a and e). The PMF for CH is notably different from that of HA, however, which has a smaller minimum energy well (Figure 4 a and g). The observed difference is a consequence of the difference in stereochemistry of the hydroxyl group at the C_4 position in GlcNAc and GalNAc, which interacts with the ring oxygen on GlcUA via hydrogen-bonding (Almond and others 1997). The most significant difference in the β1,3 PMFs, however, is seen

for C4S, in which the sulfate group is located proximal to the $\beta 1,3$ linkage (Figure 4, c). The sulfate group is capable of making direct van der Waals contact with the neighboring GlcUA residue, thereby significantly restricting the flexibility of the C4S $\beta 1,3$ linkage relative to CH and C6S. The reduction in flexibility is illustrated by the considerable narrowing of the contour lines around the minimum energy point in Figure 4c with respect to Figure 4a and e.

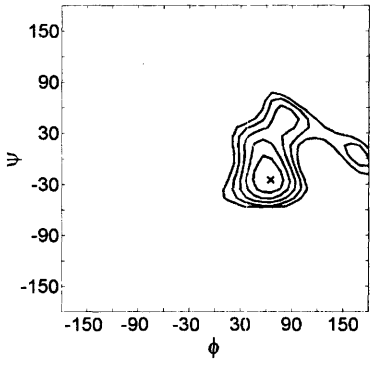
(a)


150
90
30
-30
-90
-150
ψ
-150 -90 -30 30 90 150
ϕ

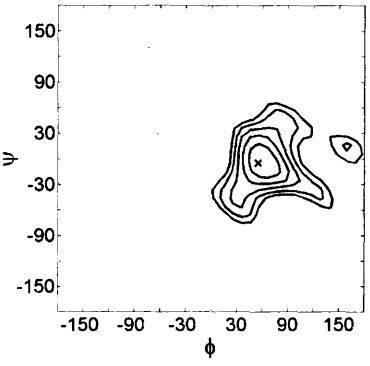
(b)


150
90
30
-30
-90
-150
ψ
-150 -90 -30 30 90 150
ϕ

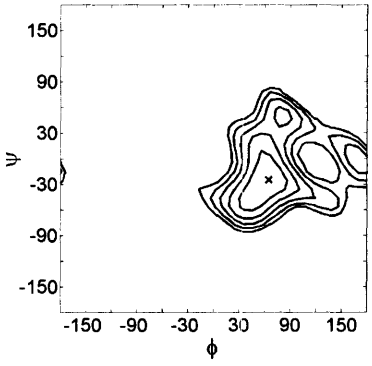
(c)


150
90
30
-30
-90
-150
ψ
-150 -90 -30 30 90 150
ϕ

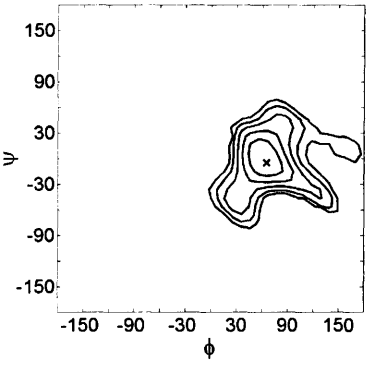
(d)


150
90
30
-30
-90
-150
ψ
-150 -90 -30 30 90 150
ϕ

(e)


150
90
30
-30
-90
-150
ψ
-150 -90 -30 30 90 150
ϕ

(f)


150
90
30
-30
-90
-150
ψ
-150 -90 -30 30 90 150
ϕ

(g) (h)

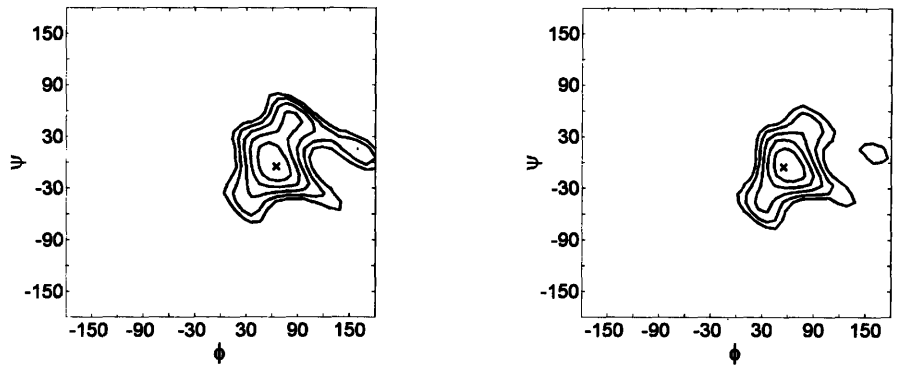


Figure 4 Glycosidic torsion angle potentials of mean force for the various disaccharides and linkage types (a) CH $\beta 1,3$ (b) CH $\beta 1,4$ (c) C4S $\beta 1,3$ (d) C4S $\beta 1,4$ (e) C6S $\beta 1,3$ (f) C6S $\beta 1,4$ and (g) HA $\beta 1,3$ (h) HA $\beta 1,4$. An 'X' is used to denote the minimum energy conformation and contours are drawn in increments of $k_B T$ above the minimum. H-bond definitions are used for ϕ and ψ.

2.4.2 Fully ionized conformation

Characteristic ratio, C_N

The characteristic ratio, C_N, is a standard measure of the effects of local chemical structure and bonded interactions on the overall dimensions of a polymer in its θ-state. It is defined as the dimensionless ratio of the mean square unperturbed end-to-end distance of the real chain to the hypothetical mean squared end-to-end distance of an equivalent Gaussian chain,

$$C_N \equiv \frac{\langle r^2 \rangle_0}{N l_v^2} \tag{1}$$

where $\langle r^2 \rangle_0$ is the mean square unperturbed end-to-end distance of the real chain, N is the number of bonds in the chain and l_v is the root-mean-square bond length (l_v = 5.2 Å for HA, CH, C4S, and C6S), where the subscript v denotes "virtual" because we utilize the conventional glycosidic oxygen-to-glycosidic oxygen virtual bond definition (Cleland 1971); (Brant and Goebel 1975) in analyzing the conformation of GAGs in this study. Utilizing this definition, each virtual bond is rigid in the current model and spans a single monosaccharide. The room temperature unperturbed state is achieved in the context of the current model by switching off all non-bond interactions. This is equivalent to taking the limit of infinite concentration of added salt, since the bonded PMFs are assumed to be independent of ionic strength.

The dependence of C_N on chain length is shown in Figure 5 for up to 512 monosaccharides for CH, C4S, C6S, and HA. C_N necessarily attains a limiting value at long chain lengths, $C_\infty = \lim_{N\to\infty} C_N$, because it does not contain the effects of excluded volume (Table 4). The model predicts that C_N is considerably larger for C4S than for the other GAGs, with a limiting value of C_∞ that is ~ 40% larger. There are two contributions to C_N in the current model: the bonded backbone chemical structure and the glycosidic torsion potentials of mean force. Noting that the bonded backbone chemical structures of the GAGs are similar (Table 2) leads one to conclude that it is the lack of flexibility present in the $\beta 1,3$ PMF of C4S that leads to the observed differences. The effect on C_∞ of the sulfate group at the C_6 position of GalNAc is minor because of its distal location from the $\beta 1,3$ and $\beta 1,4$ linkage regions. Interestingly, the C_N of HA is similar to those of CH and C6S despite the notable differences in the $\beta 1,3$ PMF of

HA. This result is attributable to the fact that the *flexibility* of the $\beta 1,3$ linkage of HA is similar to CH, unlike C4S. To ensure that the foregoing conclusions regarding C_N are insensitive to the choice of virtual bonds, the preceding calculations were repeated using the n (= $3N$) backbone virtual bonds of the coarse-grained model (Figure 2). The results obtained were qualitatively similar, although clearly the numerical values of C_n were not equivalent to those of C_N.

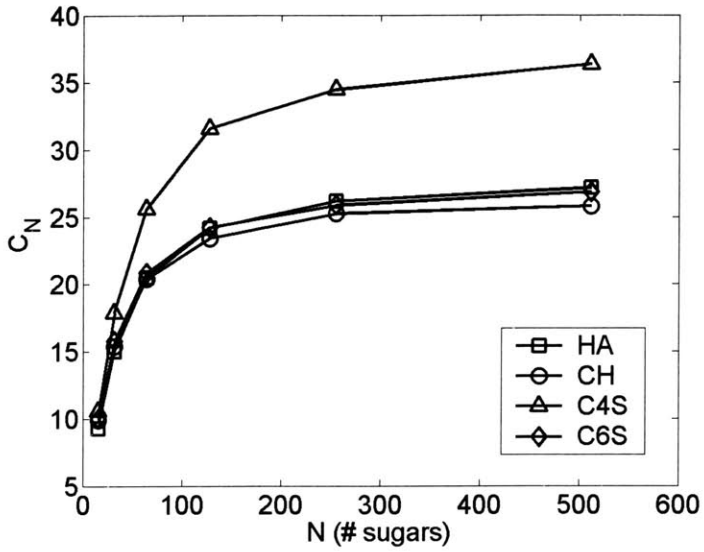


Figure 5 Unperturbed characteristic ratios, C_N, as a function of molecular weight (number of monosaccharides) for the four GAGs studied.

Table 4 Limiting GAG characteristic ratios, C_∞, and intrinsic persistence lengths, a_0.

GAG	C_∞	a_0 (Å)
HA	28	71 ± 3
CH	27	70 ± 4
C4S	38	96 ± 6
C6S	28	71 ± 4

Persistence length

The persistence length, a, of a polyelectrolyte may be decomposed into two contributions: intrinsic, a_0, and electrostatic, a_e. a_0 is strictly defined to be equal to the persistence length of the polyelectrolyte in the limit of infinite ionic strength, $a_0 \equiv \lim_{c_s \to \infty} a$. Like C_N, a_0 is attributable to the bonded chain structure and short-range non-electrostatic bonded interactions. As its name implies, a_e is due to electrostatic interactions occurring along the chain backbone, which lead to additional stiffening of the chain backbone. The total persistence length may be computed from the chain structure using (Flory 1988),

$$a = \frac{\langle \mathbf{l}_1 \cdot \mathbf{R}_{ee} \rangle}{l_1} \tag{2}$$

where $\mathbf{l}_1$ is the first bond vector in the chain, l_1 is its root-mean square value, $\mathbf{R}_{ee}$ is the end-to-end vector, and brackets are used to denote ensemble average. For polymers in the θ-state, a and C_N both necessarily attain limiting values with increasing molecular weight due to the absence of excluded volume effects and are related by $C_\infty = 2a/l - 1$. For polymers and polyelectrolytes that are not in the θ-state, however, long-range interactions will eventually lead to excluded volume effects that cause a to increase with increasing molecular weight. This artifact will lead to the observation of an *apparent* persistence length that is greater than the *true* persistence length, the latter of which is due solely to local interactions leading to back-bone stiffness. For this reason, it is

important to exercise care in computing the persistence length of polymers that are not in the θ-state, such as the polyelectrolytes considered here at finite ionic strength.

For each GAG and salt concentration studied, a attains a limiting value for $L \approx 4$-$6a$, where L denotes the contour length (Figure 6). This result is intuitive because for contour lengths on the order of a persistence length, the chain will remain relatively extended, ensuring that long-range intramolecular interactions are absent.

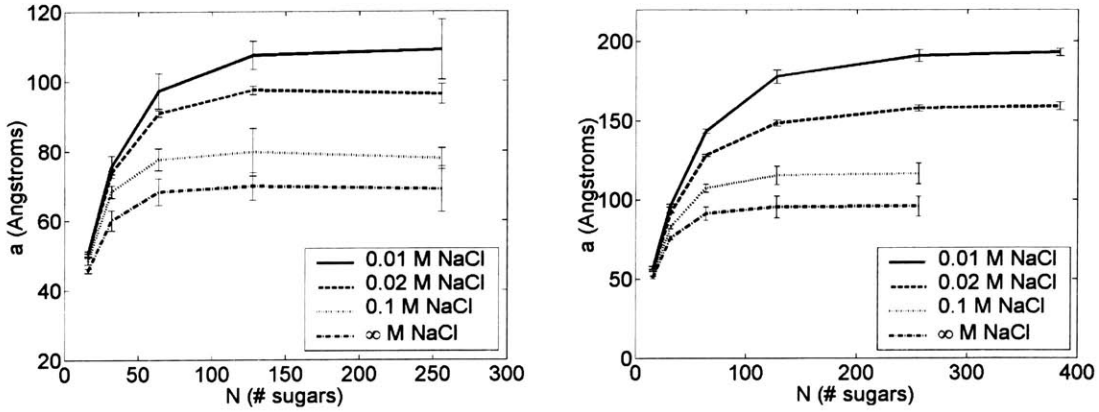


Figure 6 Ionic strength and molecular weight dependence of the total persistence length, a, of chondroitin (left) and chondroitin 4-sulfate (right).

The following expressions were obtained by a least-squares fit of the data for persistence length versus salt concentration,

$$\begin{aligned} a_{HA} &= 71 + 145 c_s^{-0.56} \\ a_{CH} &= 70 + 153 c_s^{-0.59} \\ a_{C4S} &= 96 + 447 c_s^{-0.67} \\ a_{C6S} &= 71 + 397 c_s^{-0.64} \end{aligned} \tag{3}$$

where a is in units of Å, c_s is in mM, $R^2 > 0.99$ for each fit, and persistence length data for $N = 128$ and 256 sugars were used for HA/CH and C4S/C6S, respectively. As an alternative to using the 'microscopic' definition in Eq. (2), the persistence length may be inferred from the mean square radius of gyration assuming a worm-like chain model (Ullner and others 1997),

$$s^2 = \frac{La}{3} - a^2 + 2\frac{a^3}{L} - 2\frac{a^4}{L^2}\left(1 - e^{-L/a}\right) \tag{4}$$

where L is the contour length, assumed to be 10 Å per disaccharide for each of the GAGs studied here. Indeed this is the approach typically taken experimentally because the radius of gyration is the usually the relevant observable (e.g., scattering experiments). The solution to Eq. (4) for HA and CH (128-sugars) and C4S and C6S (256-sugars) yields ($R^2 > 0.99$),

$$\begin{aligned} a_{HA} &= 72 + 172 c_s^{-0.56} \\ a_{CH} &= 71 + 191 c_s^{-0.58} \\ a_{C4S} &= 100 + 480 c_s^{-0.63} \\ a_{C6S} &= 74 + 463 c_s^{-0.60} \end{aligned} \tag{5}$$

While the intrinsic persistence lengths and salt concentration exponents are similar in Eqs. (3) and (5), significant differences exist between the numerical pre-factors multiplying the salt concentration dependent terms. Thus, quantitative differences occur in the predicted persistence length depending on whether the 'microscopic' (bond vectors) or 'macroscopic' (radius of gyration) definition is employed.

Although there is no clear consensus on the conformational behavior of polyelectrolytes in the presence of salt, the electrostatic persistence length is typically reported to be $\propto \kappa^{-\beta}$, where $\beta \approx 1$ for flexible chains and $\beta \approx 2$ for stiff chains such as DNA (Ullner 2003; Ullner and Woodward 2002). Molecular simulations of flexible Debye-Hückel chains (Reed and Reed 1991; Ullner and others 1997; Ullner and Woodward 2002) and experiments on HA and other flexible polyelectrolytes have reported $\beta \approx 1$ (Ghosh and others 1990; Reed and others 1991) as well as a linear dependence of the electrostatic persistence length on the linear charge density, ξ, below the threshold value of $\xi = 1$. A more precise value of $\alpha = 1.2$ has also been reported for simulations of Debye-Hückel freely jointed chains, where the slightly higher exponent was attributed to excluded volume effects (Ullner and Woodward 2002). The reported results of $a_e \propto \kappa^{-1}, \kappa^{-1.2}$, agree very will with the observed, $a_e \propto c_s^{-0.6}$ found in the present study, where it is noted that $\kappa \propto c_s^{0.5}$. We do find, however, a significantly stronger than linear dependence of a_e on ξ. In going from HA and CH to C4S and C6S the linear charge density doubles, whereas the electrostatic persistence length pre-factor changes by between (2.4 – 3.1).

Comparison with experiment

The persistence length of HA in high ionic strength aqueous solution has been measured experimentally using a variety of methods including light scattering, neutron scattering, x-ray scattering, and viscometry (Table 5). Results from different studies exhibit considerable deviation ranging from 40 to 90 Å between 0.1–1.0 M NaCl, with no

consensus on the correct value of a_0. The spread in the experimental data brackets the results predicted by Eq. (3), namely 74-83 Å at 0.1–1.0 M NaCl, although clearly the comparison does not represent a stringent test of the predictive capabilities of the model. Ghosh and others (1990) published the ionic strength dependence of the *apparent* persistence length of 1 MDa HA, but as its name suggests, their persistence length measurements included excluded volume effects not believed to be present in the persistence length results presented in Eqs. (3) and (5). Their results are therefore not of utility in validating our model predictions.

Table 5 HA experimental persistence length at high ionic strength.

a (Å)	[NaCl] (M)	Experimental Method	Reference
40	0.2	SAXS	(Cleland 1977)
40	0.5	viscometry	
90	1.0	light scattering	(Ghosh and others 1990)
80	0.3	light scattering	(Fouissac and others 1992)
42	0.2	viscometry	(Hayashi and others 1995)
41	0.5	viscometry, light scattering	
90	0.1	neutron and light scattering	(Buhler and Boue 2004)

To further test the model, we probe its ability to predict the ionic strength dependence of the radius of gyration of HA in the range 0.01 to 1.0 M NaCl (Table 6). We also compare the theoretical root-mean-square end-to-end distance, r, with the only solution conformation experimental data known to the authors for C4S and C6S (Table 7). Experimental methods include light scattering in the cases of Esquenet and Buhler (2002), Fouissac and others (1992), and Hayashi and others (1995) as well as viscometry

in the study of Hayashi and others (1995). Although disagreement between the theoretical and experimental results for HA is as large as 30% in some cases, disagreement between the various published experimental studies is also within that range. While these circumstances make a quantitative comparison difficult, it is the best that can be done given the existing experimental data.

Table 6 Experimental and theoretical root-mean-squared radius of gyration, *s*, for HA.

MW (kDa)	[NaCl] (M)	*s* (Å)		Difference (%)	Reference
		Experimental	Theoretical		
130	0.01	557	388 ± 12	-30	(Fouissac and others 1992)
130	0.06	444	331 ± 12	-26	(Fouissac and others 1992)
130	0.3	424	302 ± 15	-29	(Fouissac and others 1992)
104	0.5	228	234 ± 10	3	(Hayashi and others 1995)
69	0.5	173	207 ± 10	20	(Hayashi and others 1995)
85	0.1	270 ± 20	250 ± 9	-7	(Esquenet and Buhler 2002)

Table 7 Root-mean-squared end to end distance, *r*, for C4S and C6S.

GAG	MW (kDa)	[NaCl]	*r* (Å)		Difference (%)	Reference
			Experimental	Theoretical		
C4S	27	0.2	273	324 ± 25	19	(Tanaka 1978b)
C6S	44	0.2	408	396 ± 21	-3	(Tanaka 1978b)

2.4.3 Titration

Theoretical titration curves at physiological ionic strength (0.15 M NaCl) demonstrate the effect of increasing pH on dissociation of the carboxylate group on GlcUA in CH, C4S, and C6S (40-disaccharides each) (Figure 7, left). As α increases, further dissociation becomes increasingly difficult due to unfavorable anionic-anionic electrostatic interactions, as demonstrated by the increasing apparent acidity, pK^{app} (Figure 7, right). The pK^{app} is significantly larger in C4S and C6S than in CH due to the presence of the sulfate groups, which interact unfavorably with ionized carboxylate groups and drive the equilibrium towards the more associated, or protonated, state. Interestingly, the titration behavior observed for the 4- and 6- sulfated forms of chondroitin are highly similar, except for a very small shift of about 0.04 pK units of C4S over C6S (Figure 7, right).

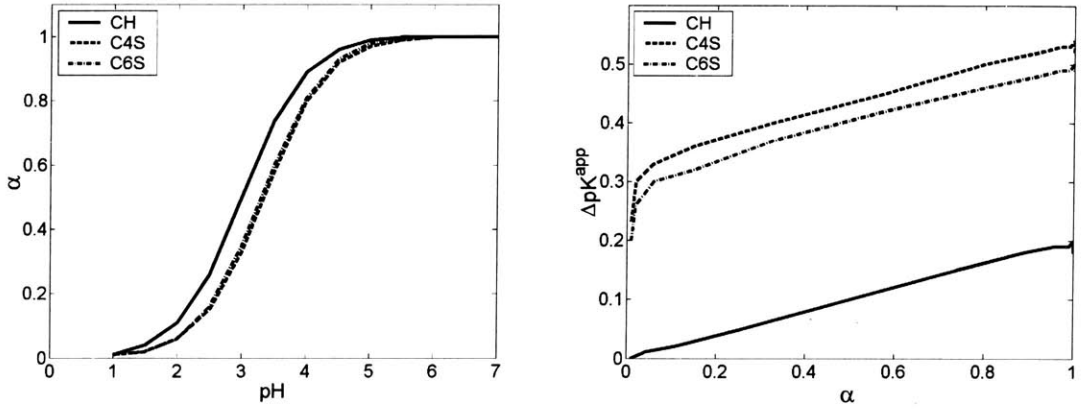


Figure 7 Theoretical titration curves of 40-disaccharide GAGs (CH, C4S, C6S) at physiological ionic strength (0.15 M).

These findings may be rationalized by examining mean carboxylate-carboxylate and carboxylate-sulfate group separations (Figure 8). Figure 8 demonstrates that the mean carboxylate-carboxylate group spacing is ~ 10 Å in both C4S and C6S, whereas the mean

carboxylate-sulfate spacing in C4S is smaller than in C6S by a bit more than an Ångstrom. The resulting increased electrostatic interactions lead to the larger observed shift in the pK of the carboxylate group in C4S. Finally, ΔpK^{app} depends linearly on α for each GAG, except for very small α in C4S and C6S (Figure 7). When $\alpha \approx 0$ there is a large change in electrostatic energy due to deprotonation, $\Delta pK = \frac{1}{Nk_BT \ln 10} \frac{\partial F}{\partial \alpha}$, because of the reference state employed.

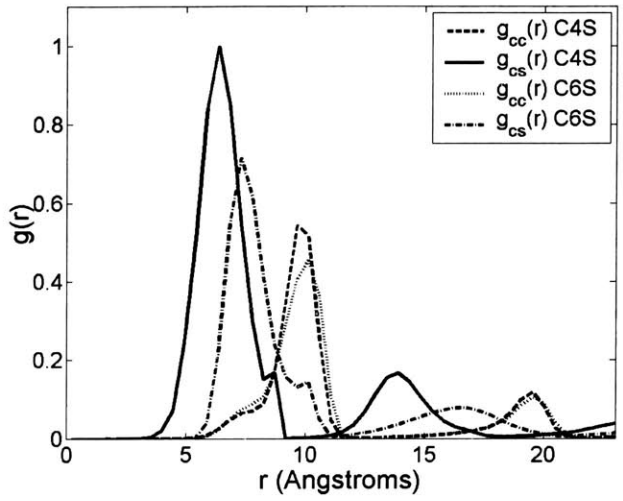


Figure 8 Intramolecular carboxylate-carboxylate, $g_{cc}(r)$, and carboxylate-sulfate, $g_{cs}(r)$, charge group radial distribution functions at 0.15 M NaCl and pH 4 for 40-disaccharide GAGs of CH, C4S, and C6S.

For any given GAG (e.g., C4S), the strength and range of electrostatic interactions between charged groups increase with decreasing ionic strength, shifting the titration curve towards lower α at all pH (Figure 9). In contrast, as the ionic strength approaches 1.0 M the Debye length is reduced to ~ 3 Å, effectively screening almost all electrostatic interactions and neighbor site coupling so that the Langmuir isotherm of the acid is nearly reproduced with $pK^{int} = 2.9$.

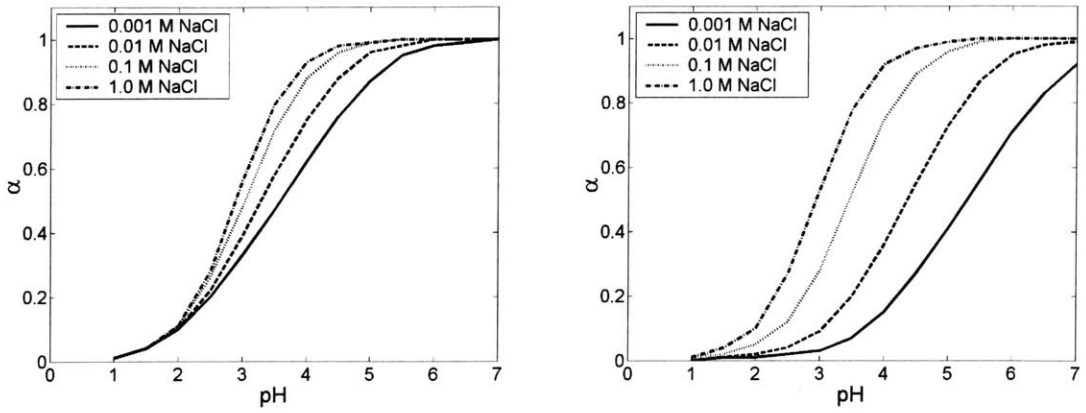


Figure 9 Ionic strength dependence of titration curves for CH (left) and C4S (right).

In analyzing the effects of pH on GAG conformation, we first consider the infinite ionic strength limit. At infinite ionic strength, non-bonded electrostatic interactions between charged monomers are absent, so that GAG conformation is determined exclusively by the bonded structure and glycosidic PMFs. As noted in the modeling section, however, the bonded PMFs were assumed to be independent of pH and ionic strength. For this reason, pH will not affect GAG conformation in the limit of infinite ionic strength.

For finite ionic strengths, non-bonded electrostatic interactions are present. Decreasing the pH from the physiological value of 7 will cause carboxylate groups to tend towards their protonated, deionized state. This will directly affect GAG conformation by decreasing intra-molecular electrostatic repulsion, which will shift the equilibrium distribution of states to more compact conformations due to the disordering effects of entropy. The effects of pH on conformation will be weaker, however, for higher ionic strengths since electrostatic interactions will already be screened in that case.

The perturbed characteristic ratio, $C_{80} \equiv \langle r_{ee}^2 \rangle / N l_v^2$, for CH (Figure 10, left) and C4S (Figure 10, right) demonstrate the aforementioned effects. At 1.0 M ionic strength the pH has no noticeable effect on the conformation of chondroitin due to the screening of electrostatic interactions (Debye length ~ 3 Å). Indeed, short-range electrostatic interactions along the chain backbone are almost entirely absent because 10 Å (~ 3 Debye lengths) is the minimum carboxylate-carboxylate spacing in CH. Long-range electrostatic interactions are also minimal because the contour length, $L \sim 400$ Å, is only about $6a_0$, implying that the chain adopts a more or less extended state. In contrast, the pH has a noticeable, albeit small, effect on C4S conformation at 1.0 M ionic strength, due to the higher linear charge density present on that molecule. Interestingly, at 0.001 M, the lowest ionic strength considered, C_{80} reaches a plateau in the case of CH but not C4S, because α does not reach its full value of 1 at pH 7 in the latter case (Figure 9).

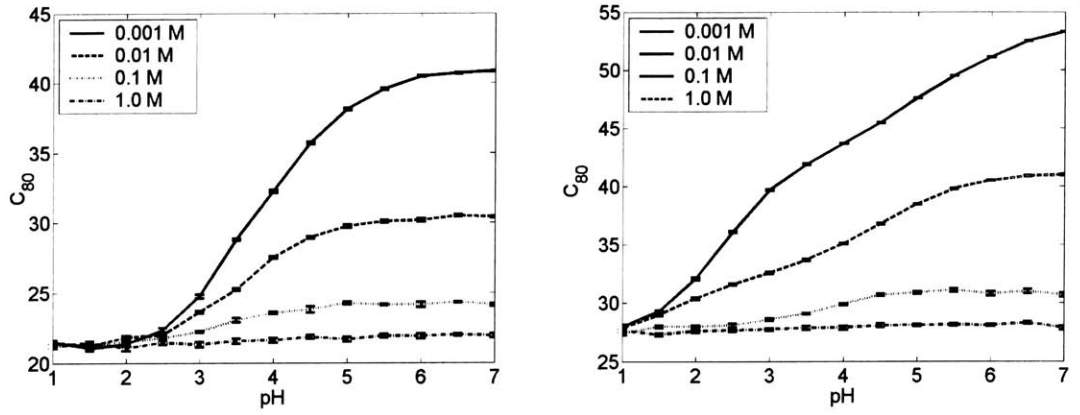


Figure 10 Dependence of conformation on pH at different ionic strengths for CH (left) and C4S (right).

Comparison with experiment

(Cleland and others 1982) employed potentiometric titration [as well as various theoretical approaches including a discrete site Monte Carlo model similar to that developed in this work (Cleland 1984)] to study the titration properties of HA. They found that pK^{app} and ΔpK^{app} depend linearly on α for ionic strengths greater than 0.001 M, a result that was also found in the present study (Figure 11, left). In Figure 11 (right) we compare the theoretical slope of ΔpK^{app} vs. α as a function of ionic strength as computed using the present model for HA and from experimental results of Cleland and others (1982). The model tends to over-predict the slope with respect to experiment at low ionic strengths and under-predict it at high ionic strength. Disagreement at low ionic strength may be attributed to the overestimation of electrostatic interactions by the Debye-Hückel interaction potential. At high ionic strengths, the Debye length becomes small so that only short-range electrostatic interactions contribute to ΔpK^{app}. When short length scale interactions dominate the electrostatic energy, the model's failure to account for molecular detail such as the low dielectric and salt exclusion properties of the GAG domain gain importance, offering a source of discrepancy between the model and experiment.

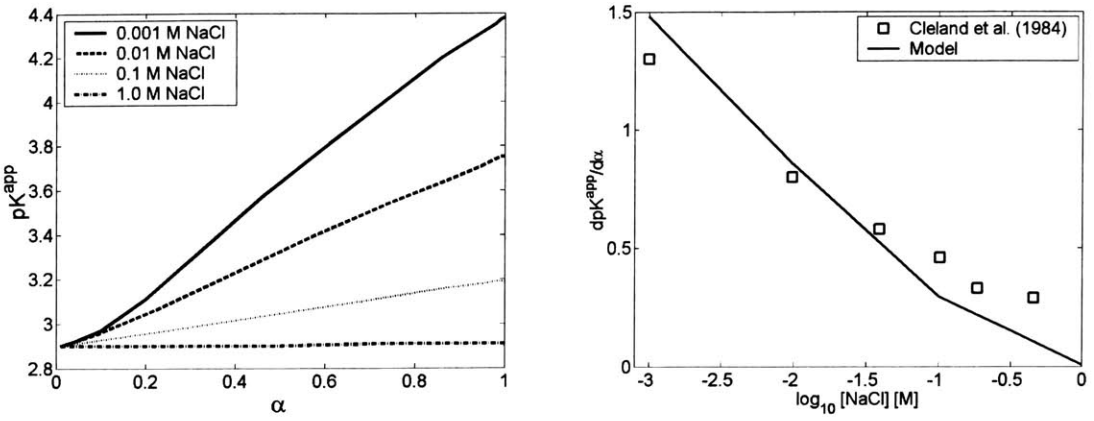


Figure 11 Theoretical titration curves of hyaluronic acid at four different ionic strengths 0.001 – 1.0 M NaCl (left). Slopes of experimental (Cleland and others 1982) and theoretical titration curves of hyaluronic acid as a function of ionic strength (right).

2.5 Concluding discussion

A computationally efficient, coarse-grained molecular model is developed for the investigation of the conformational and titration properties of chondroitin sulfate and hyaluronan, two anionic glycosaminoglycans that play a central role in articular cartilage biomechanics. Predictions for the intrinsic persistence length and the ionic strength dependence of the radius of gyration and $\partial \Delta \mathrm{p}K^{app} / \partial \alpha$ are shown to be in quantitative agreement with experiment, despite the absence of any adjustable parameters. The predictive capabilities of the model are attributed in part to its being based on an underlying all-atom model that uniquely defines the topology and energetic interactions. Short-ranged, chemically specific trans-glycosidic interactions between monosaccharides are included using a bonded potential of mean force that is obtained from all-atom simulations of isolated disaccharides. Long-ranged electrostatic interactions between monosaccharides are modeled using a Debye-Hückel interaction potential applied to the monosaccharide centers of charge.

Interestingly, 6-sulfation is not found to affect the inherent stiffness of chondroitin, as measured by C_∞ and a_0, whereas 4-sulfation has a considerable stiffening effect. C_∞ increases from 27 and 28 to 38 and a_0 from 68 and 71 to 96 Å in going from CH and C6S to C4S, respectively. This finding is attributed to the proximity of the sulfate group in C4S to the $\beta 1,3$ linkage, where it sterically restricts glycosidic flexibility by interacting with GlcUA. In contrast, the sulfate group in C6S is distal from both the $\beta 1,3$ and $\beta 1,4$ linkages so that it does not significantly alter glycosidic linkage flexibility with respect to CH. Our finding for C4S is consistent with a recent study by (Rodriguez-Carvajal and others 2003), but our result for C6S is in contradiction with that

work, in which C6S was found to be more flexible than CH. The present study goes beyond the work of Rodriguez-Carvajal and others (2003) by accounting for long-range electrostatic interactions and pH on conformation and titration properties. 6- and 4-sulfation are equally found to significantly increase the electrostatic persistence length at all ionic strengths due to increase in CH charge density associated with sulfation, and the electrostatic persistence length of HA is found to be similar to that of CH.

Despite the advances made in the present study towards our ability to quantitatively model GAG conformation and titration, several modeling deficiencies remain. Principal among them are the neglect of the low dielectric and salt exclusion of the GAG domain and the effects of Ca^{2+} cations on electrostatic interactions. Moreover, internal monosaccharide degrees of freedom as well as glycosidic bond lengths and valence angles were fixed at their mean values obtained from isolated disaccharide simulations. In future studies it would be prudent to further investigate the effects of these assumptions on the predicted properties. Unfortunately, well-established experimental conformational and titration data are scarce for anionic GAGs, in particular for CS, and detailed atomistic simulations are too computationally expensive to be performed for the molecular weight range of interest, posing inherent difficulties in discriminating between various coarse-grained modeling assumptions.

References for Chapter 2

Almond A, Brass A, Sheehan JK. 1998. Dynamic exchange between stabilized conformations predicted for hyaluronan tetrasaccharides: comparison of molecular dynamics simulations with available NMR data. Glycobiology 8(10):973-980.

Almond A, Brass A, Sheehan JK. 2000. Oligosaccharides as model systems for understanding water-biopolymer interaction: Hydrated dynamics of a hyaluronan decamer. Journal of Physical Chemistry B 104(23):5634-5640.

Almond A, Sheehan JK. 2000. Glycosaminoglycan conformation: Do aqueous molecular dynamics simulations agree with x-ray fiber diffraction? Glycobiology 10(3):329-338.

Almond A, Sheehan JK, Brass A. 1997. Molecular dynamics simulations of the two disaccharides of hyaluronan in aqueous solution. Glycobiology 7(5):597-604.

Bashford D, Karplus M. 1990. pKa's of ionizable groups in proteins - Atomic detail from a continuum electrostatic model. Biochemistry 29(44):10219-10225.

Bayliss MT, Osborne D, Woodhouse S, Davidson C. 1999. Sulfation of chondroitin sulfate in human articular cartilage - The effect of age, topographical position, and

zone of cartilage on tissue composition. Journal of Biological Chemistry 274(22):15892-15900.

Beroza P, Fredkin DR, Okamura MY, Feher G. 1991. Protonation of interacting residues in a protein by a Monte Carlo method - Application to lysozyme and the photosynthetic reaction center of Rhodobacter-Sphaeroides. Proceedings of the National Academy of Sciences of the United States of America 88(13):5804-5808.

Brant DA, Christ MD. 1990. Realistic conformational modeling of carbohydrates. In: French AD, Brady JW, editors. Computer Modeling of Carbohydrate Molecules. Washington, DC. p 42-68.

Brant DA, Goebel KD. 1975. General treatment of configurational statistics of polysaccharides. Macromolecules 8(4):522-530.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry 4(2):187-217.

Buhler E, Boue F. 2003. Persistence length for a model semirigid polyelectrolyte as seen by small angle neutron scattering: A relevant variation of the lower bound with ionic strength. European Physical Journal E 10(2):89-92.

Buhler E, Boue F. 2004. Chain persistence length and structure in hyaluronan solutions: Ionic strength dependence for a model semirigid polyelectrolyte. Macromolecules 37(4):1600-1610.

Cleland RL. 1971. Ionic polysaccharides. V. Conformational studies of hyaluronic acid, cellulose, and laminaran. Biopolymers 10:1925-1948.

Cleland RL. 1977. Persistence length of hyaluronic acid - Estimate from small-angle x-ray scattering and intrinsic viscosity. Archives of Biochemistry and Biophysics 180(1):57-68.

Cleland RL. 1984. Theory of potentiometric titration of polyelectrolytes - A discrete-site model for hyaluronic acid. Macromolecules 17(4):634-645.

Cleland RL. 1991. Electrophoretic mobility of wormlike chains. 1. Experiment - hyaluronate and chondroitin 4-sulfate. Macromolecules 24(15):4386-4390.

Cleland RL, Wang JL, Detweiler DM. 1982. Polyelectrolyte properties of sodium hyaluronate. 2. Potentiometric titration of hyaluronic acid. Macromolecules 15(2):386-395.

Cremer D, Pople JA. 1975. General definition of ring puckering coordinates. Journal of the American Chemical Society 97(6):1354-1358.

Debye PW, Hückel E. 1923. Zur Theorie der Elektrolyte I. Zeitschrift für Physik 24:305-324.

Esquenet C, Buhler E. 2002. Aggregation behavior in semidilute rigid and semirigid polysaccharide solutions. Macromolecules 35(9):3708-3716.

Flory PJ. 1988. Statistical Mechanics of Chain Molecules. New York: Oxford University Press.

Fouissac E, Milas M, Rinaudo M, Borsali R. 1992. Influence of the ionic strength on the dimensions of sodium hyaluronate. Macromolecules 25(21):5613-5617.

Ghosh S, Xiao L, Reed CE, Reed WF. 1990. Apparent persistence lengths and diffusion behavior of high molecular weight hyaluronate. Biopolymers 30(11-12):1101-1112.

Gribbon P, Heng BC, Hardingham TE. 2000. The analysis of intermolecular interactions in concentrated hyaluronan solutions suggest no evidence for chain-chain association. Biochemical Journal 350:329-335.

Hardingham T. 1998. Chondroitin sulfate and joint disease. Osteoarthritis and Cartilage 6:3-5.

Havranek JJ, Harbury PB. 1999. Tanford-Kirkwood electrostatics for protein modeling. Proceedings of the National Academy of Sciences of the United States of America 96(20):11145-11150.

Haxaire K, Braccini I, Milas M, Rinaudo M, Perez S. 2000. Conformational behavior of hyaluronan in relation to its physical properties as probed by molecular modeling. Glycobiology 10(6):587-594.

Hayashi K, Tsutsumi K, Nakajima F, Norisuye T, Teramoto A. 1995. Chain stiffness and excluded volume effects in solutions of sodium hyaluronate at high ionic strength. Macromolecules 28(11):3824-3830.

Hayashi K, Tsutsumi K, Norisuye T, Teramoto A. 1996. Electrostatic contributions to chain stiffness and excluded-volume effects in sodium hyaluronate solutions. Polymer Journal 28(10):922-928.

Hill TL. 1956. Titration curves and ion binding on proteins, nucleic acids, and other macromolecules with a random distribution of binding sites of several types. Journal of the American Chemical Society 78:5527-.

Jonsson B, Ullner M, Peterson C, Sommelius O, Soderberg B. 1996. Titrating polyelectrolytes - Variational calculations and Monte Carlo simulations. Journal of Physical Chemistry 100(1):409-417.

Lal M. 1969. 'Monte Carlo' computer simulation of chain molecules, I. Molecular Physics 17:57-64.

Lewis S, Crossman M, Flannelly J, Belcher C, Doherty M, Bayliss MT, Mason RM. 1999. Chondroitin sulphation patterns in synovial fluid in osteoarthritis subsets. Annals of the Rheumatic Diseases 58(7):441-445.

Madras N, Sokal AD. 1988. The pivot algorithm - A highly efficient Monte Carlo method for the self-avoiding walk. Journal of Statistical Physics 50(1-2):109-186.

Manning GS. 1969. Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. Journal of Chemical Physics 51:924.

Mathews MB, Decker L. 1977. Conformation of hyaluronate in neutral and alkaline solutions. Biochimica Et Biophysica Acta 498(1):259-263.

McQuarrie DA. 1975. Statistical Mechanics. New York: Harper & Row.

Nomenclature I-ICoB. 1971. Archives of Biochemistry and Biophysics 145:405-436.

Perico A, Mormino M, Urbani R, Cesaro A, Tylianakis E, Dais P, Brant DA. 1999. Local dynamics of carbohydrates. 1. Dynamics of simple glycans with different chain linkages. Journal of Physical Chemistry B 103(38):8162-8171.

Plaas AHK, West LA, Wong-Palms S, Nelson FRT. 1998. Glycosaminoglycan sulfation in human osteoarthritis - Disease-related alterations at the non-reducing termini of chondroitin and dermatan sulfate. Journal of Biological Chemistry 273(20):12642-12649.

Plaas AHK, Wong-Palms S, Roughley PJ, Midura RJ, Hascall VC. 1997. Chemical and immunological assay of the nonreducing terminal residues of chondroitin sulfate from human aggrecan. Journal of Biological Chemistry 272(33):20603-20610.

Platt D, Bird JLE, Bayliss MT. 1998. Ageing of equine articular cartilage: structure and composition of aggrecan and decorin. Equine Veterinary Journal 30(1):43-52.

Reed CE, Li X, Reed WF. 1989. The effects of pH on hyaluronate as observed by light scattering. Biopolymers 28(11):1981-2000.

Reed CE, Reed WF. 1991. Monte Carlo electrostatic persistence lengths compared with experiment and theory. Journal of Chemical Physics 94(12):8479-8486.

Reed CE, Reed WF. 1992. Monte Carlo study of titration of linear polyelectrolytes. Journal of Chemical Physics 96(2):1609-1620.

Reed WF, Ghosh S, Medjahdi G, Francois J. 1991. Dependence of polyelectrolyte apparent persistence lengths, viscosity, and diffusion on ionic strength and linear charge density. Macromolecules 24(23):6189-6198.

Rodriguez-Carvajal MA, Imberty A, Perez S. 2003. Conformational behavior of chondroitin and chondroitin sulfate in relation to their physical properties as inferred by molecular modeling. Biopolymers 69(1):15-28.

Roughley PJ, Lee ER. 1994. Cartilage proteoglycans - Structure and potential functions. Microscopy Research and Technique 28(5):385-397.

Roux B, Simonson T. 1999. Implicit solvent models. Biophysical Chemistry 78(1-2):1-20.

Sassi AP, Beltran S, Hooper HH, Blanch HW, Prausnitz J, Siegel RA. 1992. Monte Carlo simulations of hydrophobic weak polyelectrolytes - Titration properties and pH-induced structural transitions for polymers containing weak electrolytes. Journal of Chemical Physics 97(11):8767-8774.

Scott JE, Heatley F. 1999. Hyaluronan forms specific stable tertiary structures in aqueous solution: A C-13 NMR study. Proceedings of the National Academy of Sciences of the United States of America 96(9):4850-4855.

Tanaka K. 1978a. Physicochemical properties of chondroitin sulfate. 1. Ion binding and secondary structure. Journal of Biochemistry 83(3):647-653.

Tanaka K. 1978b. Physicochemical properties of chondroitin sulfate. 2. Molecular weight dependence of intrinsic viscosity and sedimentation coefficient, and analysis as a semi-flexible coil. Journal of Biochemistry 83(3):655-659.

Tanford C. 1957. Theory of protein titration curves. II. Calculations for simple models at low ionic strength. Journal of the American Chemical Society 79(20):5340-5347.

Tanford C, Kirkwood JG. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. Journal of the American Chemical Society 79(20):5333-5339.

Tsutsumi K, Norisuye T. 1998. Excluded-volume effects in sodium hyaluronate solutions revisited. Polymer Journal 30(4):345-349.

Ullner M. 2003. Comments on the scaling behavior of flexible polyelectrolytes within the Debye-Hückel approximation. Journal of Physical Chemistry B 107(32):8097-8110.

Ullner M, Jonsson B. 1996. A Monte Carlo study of titrating polyelectrolytes in the presence of salt. Macromolecules 29(20):6645-6655.

Ullner M, Jonsson B, Peterson C, Sommelius O, Soderberg B. 1997. The electrostatic persistence length calculated from Monte Carlo, variational and perturbation methods. Journal of Chemical Physics 107(4):1279-1287.

Ullner M, Woodward CE. 2000. Simulations of the titration of linear polyelectrolytes with explicit simple ions: Comparisons with screened Coulomb models and experiments. Macromolecules 33(19):7144-7156.

Ullner M, Woodward CE. 2002. Orientational correlation function and persistence lengths of flexible polyelectrolytes. Macromolecules 35(4):1437-1445.

Chapter 3

Osmotic pressure of chondroitin sulfate glycosaminoglycans: A molecular modeling investigation

3.1 Introduction

Articular cartilage is an avascular tissue that provides a low-friction, protective lining to the ends of contacting bones during joint locomotion. The tissue consists of a dense extra-cellular matrix of aggrecan and type II collagen that is maintained by a sparse volume fraction (~ 2%) of cells. Aggrecan is a high molecular weight proteoglycan (1-3.5 MDa) that consists of a linear protein backbone (~ 300 Å) with roughly 100 covalently bound anionic chondroitin-sulfate (CS) glycosaminoglycans (GAGs), as well as a smaller molecular weight fraction of keratan-sulfate GAGs and other oligosaccharides (Doege and others 1991; Hascall and Hascall 1981). The high negative charge density presented by the CS-GAGs on aggrecan generates an osmotic swelling pressure that maintains articular cartilage in a hydrated state even under substantial compressive loads, playing an important role in determining its unique biomechanical properties.

The CS-GAGs present on aggrecan vary in chemical composition depending on the state of health or disease of articular cartilage [Osteoarthritis (OA) or Rheumatoid arthritis (RA)], anatomical site, depth within the cartilage layer, and age of the organism (Bayliss and others 1999; Hardingham 1998; Lewis and others 1999; Plaas and others 1998; Plaas and others 1997; Platt and others 1998; Roughley and Lee 1994). For example, the fraction of 6-sulfated CS disaccharides in human femoral condyle cartilage increases with age from about 0.5 to 0.8 from birth to the age of 20 years with a concomitant decrease in 4-sulfation, after which it plateaus. Additionally, the concentration of 6-sulfated CS disaccharides in knee synovial fluid was observed to be significantly lower in RA and OA than in healthy tissue (Lewis and others 1999) and the

concentration of 4-sulfated CS disaccharides higher in osteoarthritic hip cartilage, with only slight changes in overall GAG content (Mankin and Lippiello 1971). Considering the important role that CS plays in determining the mechanical properties of articular cartilage and the significant differences observed in CS chemical composition, it is of significant biological interest to understand the connection between CS chemical composition and osmotic pressure in detail.

The objective of the present study is to investigate the effects of CS chemical composition, namely sulfation type and pattern and molecular weight, on CS osmotic pressure at physiological ionic strength (0.15 M NaCl). To this end, a recently developed coarse-grained molecular model of CS is employed that enables the efficient computation of conformational and thermodynamic properties of anionic GAGs in solution. In a previous work the model was applied to isolated GAGs in infinitely dilute solution and the effects of sulfation type, ionic strength, and pH on their conformation and titration behavior was studied. In the current work it is demonstrated that the model can be used to compute directly CS osmotic pressure and thereby investigate its dependence on CS sulfate composition, molecular weight, and concentration. Although the ultimate aim of this ongoing research project is to study the mechanical properties of aggrecan, the current coarse-grained modeling investigation into the behavior of CS solutions is viewed as a valuable and necessary stepping-stone towards that end, in part because the modeling of proteoglycans has no precedence in the molecular simulation community.

3.2 Modeling

Chondroitin (CH) is a linear (unbranched) polysaccharide consisting of repeating disaccharide units of glucuronic acid (GlcUA) and N-acetyl-galactosamine (GalNAc), alternately linked in β1,3 and β1,4 glycosidic linkages, respectively (Figure 1). When the 4- or 6-carbon of N-acetyl-D-galactosamine is sulfated, chondroitin is termed chondroitin 4-sulfate (C4S) or chondroitin 6-sulfate (C6S), respectively (Figure 1). The chondroitin-sulfates that are covalently bound to aggrecan typically range in molecular mass from 10 to 30 kDa, equivalent to about 20 to 60 disaccharide units.

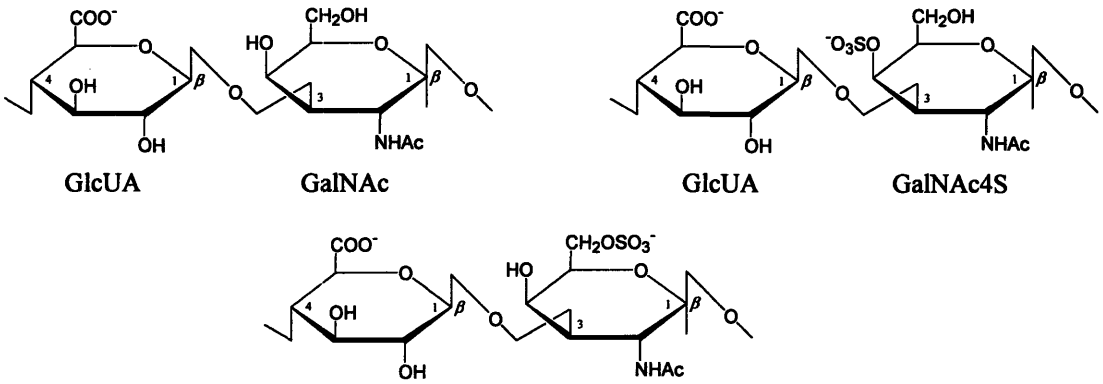


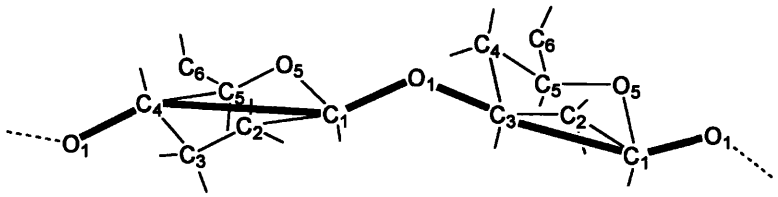
Figure 1 Disaccharide repeat units of chondroitin (top left), chondroitin 4-sulfate (top right), chondroitin 6-sulfate (bottom).

3.2.1 GAG model

The polysaccharide model is topologically defined by the sequence of chemical and virtual bonds depicted in (Figure 2). All internal degrees of freedom, including bond lengths, valence angles, and dihedral torsion angles are treated as rigid except for the glycosidic torsion angles, which provide the majority of conformational flexibility to polysaccharides. Values of the rigid internal degrees of freedom are computed from

equilibrium all-atom simulations of solvated disaccharides. The isolated disaccharide simulations are also used to compute bonded, glycosidic torsion angle free energies for each linkage and type of GAG, $F_{ij}^{B}(\phi,\psi)$, that include the effects of specific chemical interactions (e.g., hydrogen bonding) occurring across each glycosidic linkage, as well as the effects of the configurational entropy of all degrees of freedom other than the glycosidic torsions, (ϕ,ψ). Electrostatic non-bonded interactions are modeled using a simple Debye-Hückel interaction potential between monosaccharide center-of-charge sites and the effects of non-electrostatic steric interactions are included using a purely repulsive Lennard-Jones potential between the centers of geometry (= center of mass with all atoms weighted by unity) of monosaccharides. In the following section we describe how the foregoing model may be applied to compute the osmotic pressure of concentrated solutions of anionic GAGs.

(a)



(b)

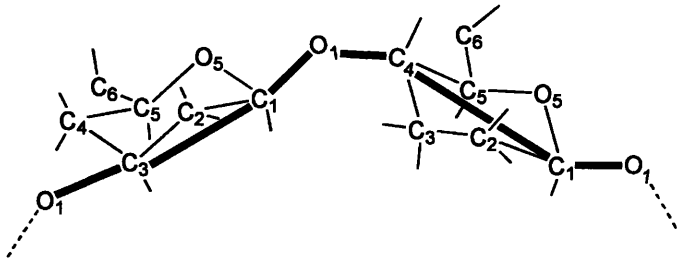


Figure 2 Definition of the coarse-grained model bonded backbone structure (thick solid lines) based on the all-atom disaccharide representation for (a) β1,3 and (b) β1,4 linkages.

3.2.2 Osmotic pressure

To derive the appropriate expression that is used to calculate the osmotic pressure of a solution of flexible polyelectrolytes, we begin by considering N identical negatively charged colloids in a volume V at temperature T, suspended in a solution of fully dissociated 1:1 electrolyte in water. The colloid result is then directly extended to treat polyelectrolytes. Consistent with the framework of Poisson-Boltzmann (PB) theory, water is treated as a uniform dielectric with dielectric permittivity ε, and the mobile ion species as spatially varying fields with number densities, $n_\alpha(\mathbf{r})$ $(\alpha = \pm)$. Each mobile ion species is in osmotic (electrochemical) equilibrium with an electrolyte-water reservoir of fixed ionic strength and electrolyte electrochemical potentials, $\mu_r = k_B T \ln n_r$, where $n_\alpha = n_r$ and $\mu_\alpha = \mu_r$ $(\alpha = \pm)$ are the mean number densities and electrochemical potentials of the electrolyte species in the reservoir. Within the PB framework, the free energy of the colloidal solution is given by (von Grünberg and others 2001),

$$F(N,V,T;\mu_r) = -k_B T \ln \frac{1}{N!\Lambda^{3N}} \int d\mathbf{R}^N e^{-\beta\Omega(\{\mathbf{R}^N\})} \tag{1}$$

where $\Omega = U - TS - \sum_{\alpha=\pm} \mu_\alpha N_\alpha$ is the grand potential, or equivalently the effective Hamiltonian or potential of mean force, of the system when the colloids are in configuration $\{\mathbf{R}^N\}$, k_B is the Boltzmann constant, T is absolute temperature, $\beta = (k_B T)^{-1}$, and Λ is the de Broglie wavelength. U is the total internal electrostatic energy, S is the mobile ion entropy, and the electrochemical potential terms ensure osmotic equilibrium with the electrolyte reservoir. Ω is a *free* energy that depends both

on the colloid configuration $\{\mathbf{R}^N\}$ and the thermodynamic state (N,V,T,μ_r) and includes the averaged effects of the 'fast' mobile ion degrees of freedom on colloid-colloid interactions. The electrostatic internal energy, U, is (Belloni 2000; Deserno and von Grünberg 2002; Trizac and Hansen 1996),

$$U = \frac{1}{2}\int_V [\rho_{fc}(\mathbf{r}) + \rho_m(\mathbf{r})]\psi(\mathbf{r})d\mathbf{r} \tag{2}$$

where $\rho_{fc}(\mathbf{r})$ is the colloidal (fixed charge) density $\rho_m(\mathbf{r})$ is the mobile ion charge density, $\psi(\mathbf{r})$ is the electrostatic potential, arbitrarily set equal to zero in the reservoir. The dependence of ρ_{fc}, ρ_m, and ψ on $\{\mathbf{R}^N\}$ is implicit, but has been dropped for notational clarity. The volume integral in Eq. (2) is carried out over all space. The mobile ion entropy, S, corresponding to non-linear PB theory is,

$$S = -k_B \sum_{\alpha=\pm} \int_V n_\alpha(\mathbf{r})[\ln n_\alpha(\mathbf{r}) - 1]d\mathbf{r} \tag{3}$$

where the mobile ions are treated as an inhomogeneous ideal gas. To be consistent with linearized PB theory, $S[n_\alpha(\mathbf{r})]$ must be expanded to quadratic order about a suitable expansion point, $\hat{n}_\alpha$. Only then will the functional minimization of Ω, subject to the constraints of Poisson's equation and electroneutrality, yield the linearized PB equation (Deserno and von Grünberg 2002). In the present study we choose, $\hat{n}_\alpha = n_r$, corresponding to linearization of the PB equation about the reference potential in the

reservoir. Although this is by far the most common choice employed in the literature, it has been shown that the optimal expansion point corresponds to the Donnan potential and its corresponding number densities (Deserno and von Grünberg 2002; Tamashiro and Schiessel 2003; von Grünberg and others 2001). Functional expansion of S to quadratic order yields (Appendix A),

$$S = \sum_{\alpha=\pm}\left(-Vk_B \langle n_\alpha \rangle \ln n_r + Vk_B n_r - \frac{k_B}{2n_r}\int_V [n_\alpha(r) - n_r]^2 \, d\mathbf{r} \right) \tag{4}$$

where $\overline{n}_\alpha \equiv V^{-1}\int_V n_\alpha(\mathbf{r})d\mathbf{r}$ is the mean mobile ion number density and all volume integrals are carried out over the space accessible to the solvent. Finally, the electrochemical potential terms are,

$$\sum_\alpha \mu_\alpha N_\alpha = \sum_\alpha k_B T \ln n_r \int_V n_\alpha(\mathbf{r})d\mathbf{r} \tag{5}$$

where equality of the electrochemical potentials in the solution and reservoir has been employed.

To obtain a closed-form analytical expression for Ω that can be employed in molecular simulations requires analytical expressions for $\psi(\mathbf{r})$, $\rho_{fc}(\mathbf{r})$, $n_\alpha(\mathbf{r})$, and $\rho_m(\mathbf{r})$. The mean-field Boltzmann relation in linearized form relates $n_\alpha(\mathbf{r})$ to $\psi(\mathbf{r})$,

$$n_\alpha(\mathbf{r}) = n_r\left[1 - \beta\nu_\alpha e\psi(\mathbf{r})\right] \tag{6}$$

where ν_α is the valence of mobile ion species α. The fixed and mobile ion charge densities are,

$$\begin{aligned}\rho_{fc}(\mathbf{r}) &= -e\sum_{i=1}^{N}\delta^3(\mathbf{r}-\mathbf{R}_i)\\ \rho_m(\mathbf{r}) &= e[n_+(\mathbf{r})-n_-(\mathbf{r})]\end{aligned} \tag{7}$$

where the colloids have been assumed to have zero radius and a 1:1 electrolyte has been assumed (e.g., NaCl). Poisson's equation relates the electrostatic potential to the charge density for any configuration of colloids $\{\mathbf{R}^N\}$,

$$\bar{\nabla}^2\psi(\mathbf{r}) = -\frac{1}{\varepsilon}\left[\rho_{fc}(\mathbf{r})+\rho_m(\mathbf{r})\right] \tag{8}$$

The solution to Eq. (8), subject to the charge source terms in Eq. (7) and the far field boundary condition, $\psi(|\mathbf{r}|\to\infty)=0$, is given by a superposition of Debye-Hückel potentials,

$$\psi(\mathbf{r};\{\mathbf{R}^N\}) = -e\sum_{i=1}^{N}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{\varepsilon|\mathbf{r}-\mathbf{R}_i|} \tag{9}$$

where, $\kappa=(8\pi\lambda_B n_r)^{1/2}$ is the inverse Debye length (equal in the colloid compartment and reservoir due to our choice of expansion point for S), $\lambda_B \equiv e^2/\varepsilon k_B T$ is the Bjerrum length, and the dependence of ψ on the colloid configuration has been made explicit.

The previous expressions may be used to evaluate the separate contributions to the dimensionless grand potential, $\beta\Omega$, (Appendix A),

$$\begin{aligned}
\beta U &= -\frac{3}{4}N\kappa\lambda_B + \lambda_B \sum_{i<j}^{N} \frac{e^{-\kappa R_{ij}}}{R_{ij}}\left(1-\frac{\kappa R_{ij}}{2}\right) \\
\beta TS &= 2Vn_r \ln n_r - 2n_r V + \frac{1}{4}N\kappa\lambda_B + \frac{1}{2}\kappa\lambda_B \sum_{i<j}^{N} e^{-\kappa R_{ij}} \\
\sum_{\alpha=\pm} \beta\mu_\alpha N_\alpha &= -2Vn_r \ln n_r
\end{aligned} \tag{10}$$

yielding the final result for $\beta\Omega = \beta U - \beta TS - \beta\sum_\alpha \mu_\alpha N_\alpha$,

$$\beta\Omega = -\frac{1}{2}N\kappa\lambda_B + \lambda_B \sum_{i<j}^{N} \frac{e^{-\kappa R_{ij}}}{R_{ij}} - 2n_r V\,. \tag{11}$$

The first term in the grand potential is the favorable electrostatic internal self-energy of the point charge colloids that is due to the local electrostatic energy well that each colloid resides in due to its surrounding oppositely charged mobile ion cloud. The second term is the standard Debye-Hückel interaction energy between colloids *i* and *j*, and the third term is the pressure contribution due to the mobile ion translational entropy. Interestingly, the electrostatic internal self-energy of the mobile ion charge clouds ($-N\kappa\lambda_B/4$) as well as the interaction of charge cloud *j* with the electrostatic field due to colloid *i* in Eq. (10) for *U* have cancelled with equal contributions to the entropy of the opposite sign.

The expression in Eq. (11) can be derived in a more straightforward manner yet with less physical insight by employing the generalized grand potential of linearized PB theory presented by (Deserno and von Grünberg 2002) (their Eq. 20),

$$\beta\Omega_{lin,eq} = \frac{1}{2}\int_V \rho_{fc}(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} - 2n_r V \tag{12}$$

where the expansion point, $\bar{\psi} = 0$, $\bar{n}_\alpha = n_r$, has been assumed (an overbar is used in their notation to denote the expansion point, not a volume average, and their n_b is equal to our n_r). Evaluation of the integral in Eq. (12) leads directly to Eq. (11) (Appendix A).

The foregoing theory may be directly applied to polyelectrolytes by noting that each charge site on the chain is bonded to the chain backbone and interacts with other charge sites via conventional Debye-Hückel interactions. The effective Hamiltonian for a specific configuration of polyelectrolytes is assumed to consist of independent electrostatic, non-electrostatic, and kinetic energy contributions,

$$H^{eff} = \Omega^{elec} + F^{non-elec} + K \tag{13}$$

where $\Omega^{elec} = \Omega$ from the colloid result and $F^{non-elec}$ contains the non-electrostatic bonded contributions as well as the repulsive Lennard-Jones non-bond potential if included in the model. It should be noted that the latter assumption neglects the effect of excluded volume on the grand potential. $F^{non-elec}$ is expressed as a free energy because it

contains the glycosidic torsion angle potential of mean force, $F^B(\phi,\psi)$, in the GAG model,

$$F^{non-elec} = \sum_{i,j=i+1} F^B_{i(k)j(l)}(\varphi,\psi) + \sum_{i<j} u_{RLJ}(r_{ij}) \tag{14}$$

where the first summation is over all glycosidic torsion angle pairs (i, j), where $i(k)$ denotes monomer type k at position i and $j(l)$ denotes monomer type l at position j, and the second summation is over all non-bonded monosaccharide pairs and models the effects of steric interactions between monosaccharides utilizing a repulsive Lennard-Jones (RLJ) potential.

The osmotic pressure, Π, is defined as the difference in pressure between the colloid and reservoir compartments,

$$\Pi \equiv -\left.\frac{\partial F^I}{\partial V}\right|_{N,T,\mu_r} + \left.\frac{\partial F^{II}}{\partial V}\right|_{T,\mu_r} \tag{15}$$

where $-\partial F^{II}/\partial V\big|_{T,\mu_r} = 2n_r k_B T$ is the pressure in the reservoir in the van't Hoff approximation, consistent with PB theory. The pressure in the polyelectrolyte solution is (Appendix B),

$$-\left.\frac{\partial F^I}{\partial V}\right|_{N,T,\mu_r} = \frac{N}{V}k_B T - \left\langle \frac{\partial H^{eff}}{\partial V} \right\rangle \tag{16}$$

where N/V is the mean number density of polyelectrolyte interaction sites. The evaluation of Eq. (16) using the standard atomic virial theorem is precluded in the current GAG model due to the presence of rigid internal degrees of freedom, including bond lengths, angles, and torsions. For this reason, the molecular virial theorem is employed for which only intermolecular interactions are required (Theodorou and others 1993). Using the molecular virial theorem for the pressure in Eq. (16) the osmotic pressure is,

$$\Pi = n_p k_B T - \frac{1}{3V}\left\langle \sum_{i<j}^{N_{nb}} R_{c_i c_j} \frac{\partial u^{DH}(r_{ij})}{\partial r_{ij}} \right\rangle \tag{17}$$

where n_p is the average number density of polyelectrolytes, $u^{DH}(r)$ is the standard Debye-Hückel interaction energy, $R_{c_i c_j}$ is the distance between the centers of mass of the polyelectrolyte to which non-bond sites i and j belong, and the summation is over only intermolecular non-bonded interactions. When the minimum image convention is employed to simulate periodic boundary conditions care must be taken to ensure that the summation in Eq. (17) is over *all* possible non-bond sites present in the system, including interactions between sites that are on different images of the same chain. In this case, $R_{c_i c_j}$ is the still the distance between the centers of mass of the chain *images* to which non-bond sites i and j belong. It is worthy of note that the mobile ion ideal gas contribution due to translational entropy present in Eq. (11) has cancelled with the reservoir pressure and that the self-energy in Eq. (11) does not contribute to the osmotic pressure because it does not depend on volume (κ is fixed at its reservoir value).

Finally, when the repulsive Lennard-Jones interaction potential is utilized in the model the appropriate interaction force is added to the virial equation in Eq. (17).

Model validation

To demonstrate the validity of Eq. (17), we compute the excess osmotic pressure of a system of point-charges on an FCC lattice. The spherical cell model solution to the linearized PB equation [sub-optimal linearized theory to be consistent with Eq. (17)] is used as a benchmark. The spherical cell model is commonly used to model the Wigner-Seitz cell of an FCC lattice (von Grünberg and others 2001) and an analytical solution to the linearized PB equation exists for that geometry (Deserno and von Grünberg 2002). An FCC configuration of spatially fixed point-charges is considered, wherein the translational entropy, or ideal gas, contribution to the osmotic pressure in Eq. (17) is zero. The finite radius of the colloid in the cell model is chosen to be 0.1 Å, which is much less than the two relevant length scales in the problem: R, the outer cell boundary and r_D, the Debye length. The cell model therefore approximates the point charge model, in this case. Results for the two models are compared in Figure 3 for reservoir ionic strengths of 0.01, 0.1, and 1.0 M NaCl. As shown in the figure, agreement between the two models is perfect up to very high fixed charge densities, demonstrating the validity of Eq. (17).

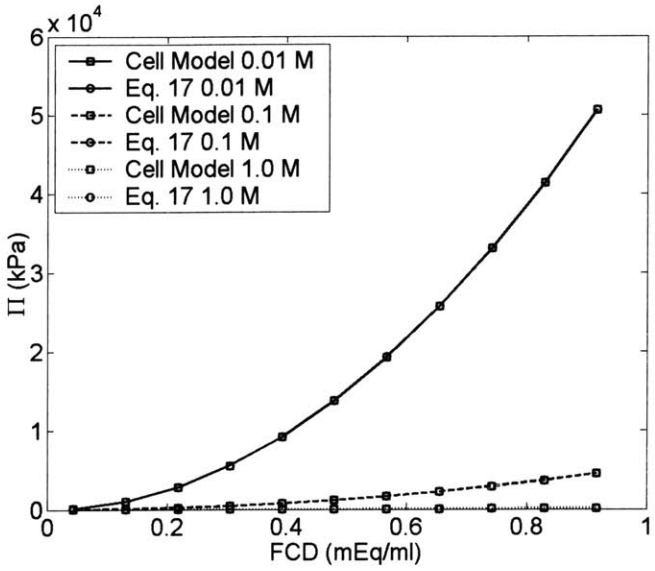


Figure 3 Osmotic pressure predicted by spherical cell model with radius 0.1 Å and Eq. (17) for point charges in an FCC lattice configuration (100 mg/ml fully sulfated CS-GAG has an equivalent fixed charge density of 0.44 mEq/ml).

3.3 Simulation protocol

GAG concentrations up to 100 mg/ml are simulated at room temperature (298 K) in a cubic cell employing periodic boundary conditions. The equilibrium distribution of states in the constant (N,V,T)-ensemble is generated using the conventional Metropolis Monte Carlo algorithm and the standard acceptance criterion, $rn \leq \min\left\{1, e^{-\beta\Delta H^{eff}}\right\}$, where rn is a pseudo-random deviate $\in]0,1]$ and ΔH^{eff} is the change in the effective Hamiltonian between old and new configurations. Although the simulation appears to be carried out in a standard *canonical* (N,V,T) ensemble, a *semi-grand canonical* ensemble (N,V,T,μ_r) is actually simulated by use of H^{eff}, because it implicitly includes the averaged, thermally equilibrated effects of the mobile ion degrees of freedom on GAG-GAG interactions and thermodynamic properties.

Configuration space is sampled using the pivot move applied to glycosidic torsion angles (Madras and Sokal 1988; Sokal 1996), rigid chain translation moves, and rigid chain rotation moves with relative frequencies of 0.4, 0.4, and 0.2, respectively, and rigid rotations were performed about the chain center of mass. Detailed balance is satisfied by selecting move types at random and by applying the pivot move to random glycosidic torsion angles. Maximum move sizes are adjusted on the fly to maintain an acceptance ratio between 40 – 60% for each move type. Finite size effects on GAG conformation and osmotic pressure were investigated, and it was found that at least 100 chains must be simulated for the highest GAG molecular weight (40 disaccharides) considered in this study. The Debye-Hückel potential was truncated at 3 Debye lengths and shifted by a constant factor to equal zero at the cut-off.

Initial configurations of 100 helical GAGs were generated with random orientations and locations in the simulation cell. 10^4 cycles were subsequently used to equilibrate the system and 10^4 cycles were used to compute averages, where a cycle consisted of N configuration-altering moves, where N is equal to the number of monosaccharides in the system. Although coarse-graining the atomistic description of GAGs is in part responsible for enabling the simulation of the large system sizes considered in this study, the computational efficiency of the model is largely due to the use of mean-field PB theory for the solvent and microion degrees of freedom. Use of the effective Hamiltonian eliminates the need to treat long-range Coulombic interactions using computationally expensive algorithms such as (Particle Mesh or conventional) Ewald summation and it eliminates the "fast" moving degrees of freedom that would otherwise restrict efficient sampling of the "slower" moving GAG degrees of freedom that are of principal interest in this study. The speed-up obtained by coarse-graining the solvent is better of course at dilute GAG concentrations; as one moves to higher GAG concentrations the speed-up achieved by eliminating solvent proportionately decreases. The GAG concentrations and system sizes considered in the present study are in a regime where a coarse-grained treatment of both the solvent *and* polysaccharide degrees of freedom is required to obtain a computationally tractable model in terms of both memory and CPU-time requirements.

3.4 Results

The osmotic pressure of chondroitin 4-sulfate solution in contact with a 1:1 salt reservoir at physiological ionic strength (0.15 M) demonstrates a strong dependence on molecular weight for chain lengths less than 16 disaccharides, after which it attains a limiting dependence on c (Figure 4).

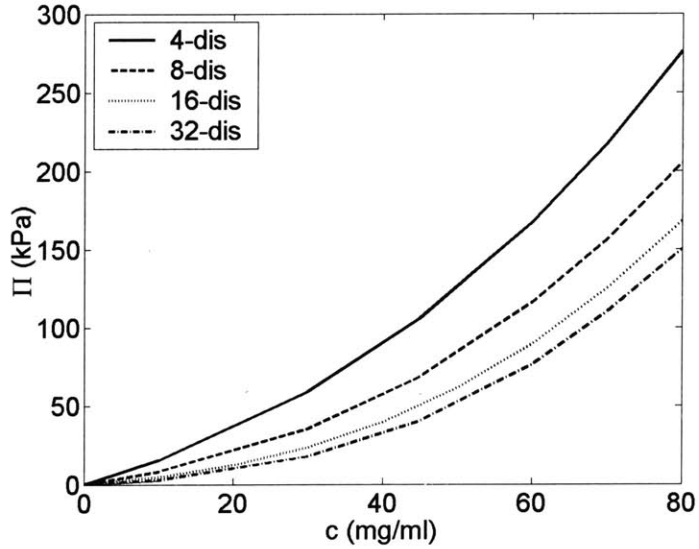


Figure 4 Dependence of C4S osmotic pressure on concentration and molecular weight (4 – 32 disaccharides or 1.8 – 14.6 kDa) at physiological ionic strength (0.15 M). Statistical error is < 1% for all data.

The osmotic pressure exhibits clearly non-ideal behavior for each molecular weight considered. A virial expansion may be used to quantify the extent of non-ideality,

$$\frac{\Pi}{cRT} = A_1 + A_2 c + A_3 c^2 \tag{18}$$

where $A_1 = 1/\bar{M}_n$, $\bar{M}_n$ is the number average molecular weight, $R = 8.3145 \times 10^3$ Pa·l/K·mol is the molar ideal gas constant, c is concentration (g/l or mg/ml) and the virial equation has been truncated after the third virial coefficient, A_3. A_1 is due solely to the ideal gas or van't Hoff contribution to the osmotic pressure, which is physically due to the translational entropy of the chains. The data in Figure 4 was fit using a least-squares regression to obtain the virial coefficients in Table 1. As expected, $1/A_1$ is equal to $\bar{M}_n$. The second virial coefficient, A_2, represents the contribution of two-body GAG-GAG interactions to the osmotic pressure. The ratio A_2/A_1 increases significantly with molecular weight, indicating that the relative contribution of intermolecular GAG interactions to the osmotic pressure increases with increasing molecular weight (Table 1). This result is to be expected because at any given mass concentration, increasing the molecular weight decreases the GAG number density and thus the total translational entropy in the system.

Table 1 Osmotic pressure virial coefficients for C4S solution in contact with a 0.15 M NaCl reservoir.

$\bar{M}_n$		$A_1 \times 10^{-4}$ (mol/g)	$A_2 \times 10^{-6}$ (mol·l/g^2)	$A_3 \times 10^{-8}$ (mol·l^2/g^3)
# of disaccharides	kDa			
4	1.83	5.45	6.92	4.56
8	3.66	2.74	5.16	5.43
16	7.32	1.37	4.30	5.74
32	14.6	0.67	3.98	5.83

Increasing the ionic strength of the salt reservoir from 0.15 to 1.0 M NaCl significantly reduces the osmotic pressure due to the increased screening of

intermolecular GAG electrostatic interactions (Figure 5, left). Further increasing the ionic strength to 5.0 M results in a nearly complete elimination of electrostatic effects on Π. In the infinite ionic strength (or θ-state) limit, the van't Hoff equation applies. In that limit, Π is due solely to the translational entropy of the homogeneous 'ideal gas' of GAG molecules, $\Pi = n_{GAG} k_B T$, where n_{GAG} is the average number density of GAG molecules.

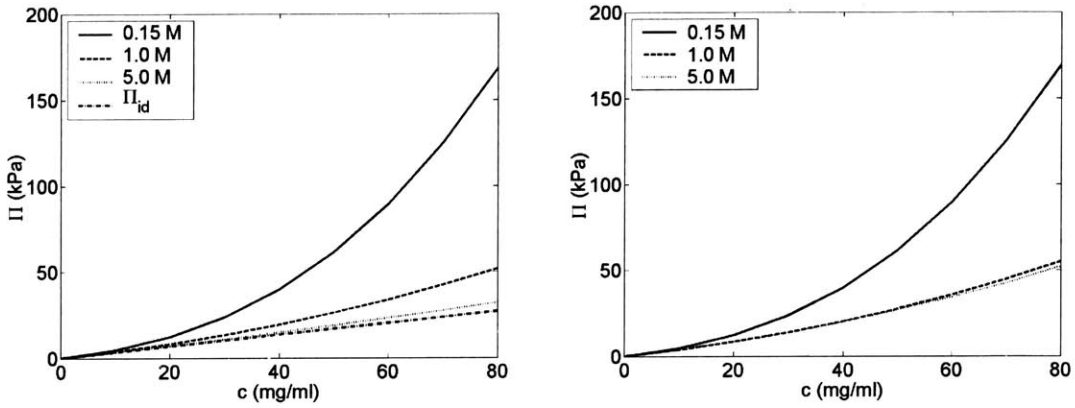


Figure 5 Dependence of C4S (16-disaccharides) osmotic pressure on ionic strength. Debye-Hückel non-bonded interactions only (left) and Debye-Hückel only (DH) as well Debye-Hückel plus steric non-bond interactions (DH-RLJ) (right).

Intermolecular center-of-mass radial distribution functions (RDFs) for 40-disaccharide C4S-GAGs demonstrate that at 0.15 and 1.0 M NaCl chains are maximally spaced due to electrostatic repulsion (Figure 6, left and right, respectively). Decreased electrostatic interactions at 1.0 M with respect to 0.15 M result in a higher population of small-r states in the former, because thermal energy is able to bring chains into closer proximity. At both salt concentrations, however, as the GAG concentration is increased the population of chains at close distances ($r < 25$ Å) increases, resulting in the observed increases in Π.

To quantify the effects of non-electrostatic, steric intermolecular interactions on Π, Π was recomputed with the repulsive Lennard-Jones interaction potential included (Figure 5, right). Interestingly, at 0.15 M NaCl steric interactions do not have an observable effect on Π. At 0.15 M NaCl intermolecular electrostatic interactions dominate over the disordering effects of thermal energy, causing GAG molecules to be maximally spaced from one another and thereby avoid the close contact required for steric interactions. The intermolecular center-of-mass RDFs discussed previously lend support to this interpretation. At higher ionic strengths, however, repulsive electrostatic interactions are diminished so that the disordering effects of entropy allow GAGs to come into close proximity. This leads to a noticeable contribution of steric interactions to Π with respect to the Debye-Hückel only model at high ionic strength (Figure 5, compare 5 M NaCl left and right).

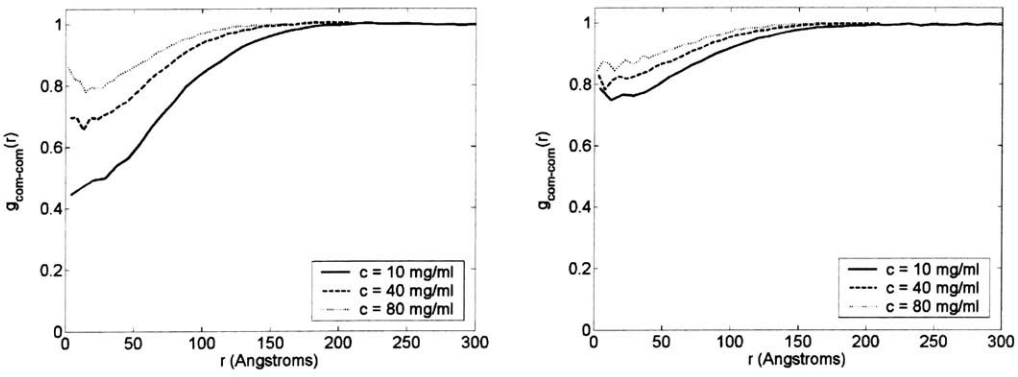


Figure 6 Intermolecular center-of-mass radial distribution functions for 40-disaccharide C4S-GAGs as a function of GAG concentration and ionic strength: 0.15 M NaCl (left) and 1.0 M NaCl (right).

Interestingly, the root-mean-squared radius of gyration, *s*, of 40-disaccharide C4S-GAGs does not exhibit a dependence on GAG concentration between the physiological concentration range of 20-80 mg/ml (Figure 7). While the conformation of

semi-flexible polyelectrolytes in the semi-dilute regime is not well understood (Barrat and Joanny 1996), it is postulated that the insensitivity of s to GAG concentration is due to the fact that at ionic strengths greater than or equal to 0.15 M, electrostatic interactions are already highly screened (Debye length < 10 Å). CS-GAGs therefore adopt their nearly unperturbed conformation even in dilute solution, and additional screening effects due to increased GAG concentration are insignificant.

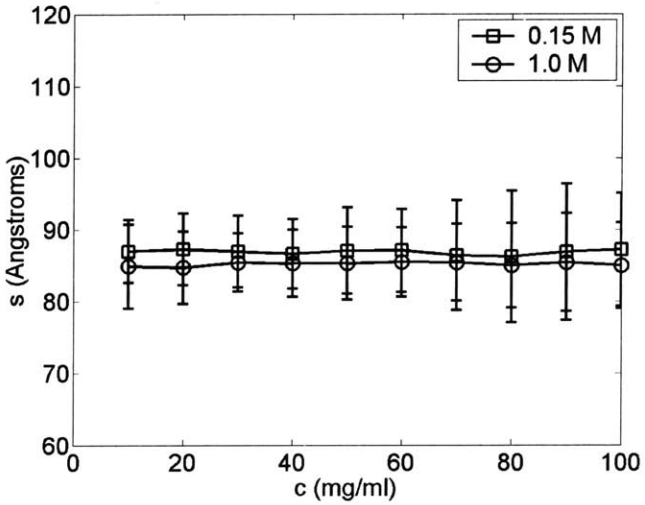


Figure 7 Root-mean-squared radius of gyration, $s \equiv \langle s^2 \rangle^{1/2}$, vs. GAG concentration for 40-disaccharide C4S chains at 0.15 and 1.0 M NaCl.

As discussed in the Introduction, the chondroitin-sulfate chains in aggrecan are typically sulfated at either the C_4 or C_6 position of the GalNAc monosaccharide. To gauge the effects of 4- vs. 6-sulfation on Π, pure solutions of 16-disaccharide C4S-GAGs and pure C6S-GAGs were studied (with steric interactions turned off). The effects of sulfate type on Π were found to be insignificant (Figure 8, left). This result is rather intuitive if one considers that the two relevant length scales in the system, the Debye

length (~ 10 Å) and the intermolecular spacing (~ 20 – 40 Å) are much greater than the difference between the C_4 and C_6 sulfate positions (~ 2 – 3 Å).

In addition to the sulfation *type*, the sulfation *pattern* may also vary in CS-GAGs, as well as in other sulfated GAGs that are relevant to tissue biomechanics, such as keratan-sulfate (KS). The statistical description of sulfate patterns in CS-GAGs is described using a traditional first-order Markov model of random copolymerization (Appendix C). GAGs consist of two monomer types, sulfated and unsulfated disaccharides, represented by the letters A and B, respectively. The two parameters (f, λ) fully describe the statistical distribution of copolymers in the system, where $f \equiv p(A)$ denotes the overall fraction of sulfated disaccharides and $\lambda \equiv p(A|A) + p(B|B) - 1$ $(-1 \le \lambda \le 1)$ is a coupling parameter that describes the copolymer sequence. As described in Appendix C, $\lambda = -1$ denotes ideal alternating AB copolymers, $\lambda = 0$ denotes random AB copolymers, and $\lambda = 1$ denotes a mixture of A and B homopolymers.

The effect of sulfate pattern on Π is investigated for 24-disaccharide C4S-GAGs by setting $f = 0.5$ and varying λ between –1 and 1 (Figure 8, right), again in the absence of steric interactions (no repulsive Lennard-Jones potential). Somewhat surprisingly, Π is found to be relatively insensitive to λ, varying by less than 10% at even the highest computed concentration of 80 mg/ml. If one considers that the fraction of sulfated chondroitin disaccharides in aggrecan is $\ge 90\%$ $(f \ge 0.9)$ in human articular cartilage (Bayliss and others 1999), it is clear that the CS sulfate pattern will have an even smaller effect in articular cartilage. Nonetheless, the effect of λ on Π for $f = 0.5$ is still of interest for understanding the biomechanical properties of other tissues where GAG

sulfation may be less prominent. Results of this study clearly indicate, however, that the sulfate pattern does not significantly affect the osmotic pressure of sulfated GAGs. In other words, only the fixed charge density matters, an observation that has been made previously by several experimental groups (Ehrlich and others 1998; Maroudas and Bannon 1981; Urban and others 1979). Insight into this somewhat surprising result may be gained by examining intermolecular charge-group (carboxylate-carboxylate, carboxylate-sulfate, and sulfate-sulfate) RDFs, as explained next.

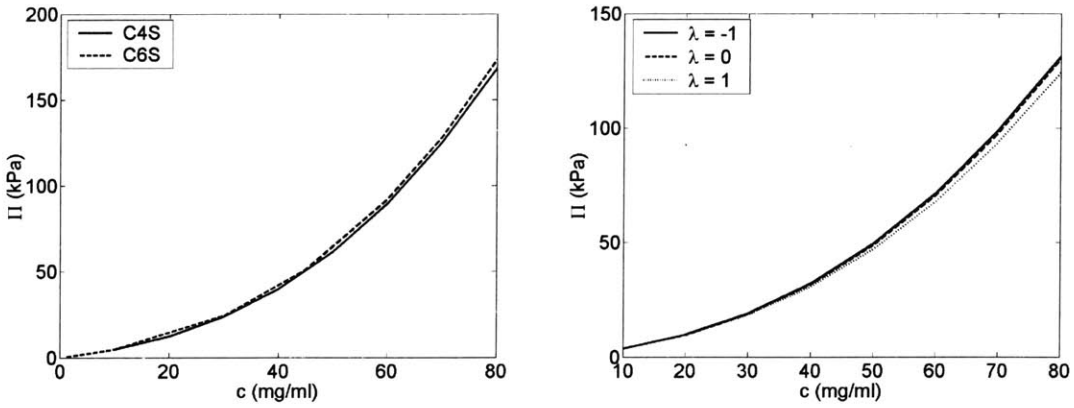


Figure 8 Effect of sulfate type (4 vs. 6) (left, 16-disaccharide MW) and 4-sulfation pattern (right, 24-disaccharide MW) on CS osmotic pressure at physiological ionic strength (0.15 M NaCl) (CS-GAGs are 100% sulfated).

As shown by Eq. (17), in the absence of steric interactions Π is fully determined by GAG translational entropy (the 'ideal gas' term) and intermolecular Debye-Hückel interactions. Since monodisperse samples of 24-disaccharide CS-GAGs are considered here, the ideal gas contribution to Π is invariant across different samples. Differences in intermolecular Debye-Hückel interactions are therefore solely responsible for observed differences in Π caused by changes in λ. For this reason, intermolecular charge group

radial distribution functions (RDFs) may be examined to explain the insensitivity of Π to λ.

Each intermolecular charge group RDF exhibits the same sigmoidal-like curve as was seen for the GAG center of mass RDFs, in which it is clear that electrostatic interactions maintain charge groups maximally separated. If one tentatively assumes that the equilibrium distribution of GAG conformations is unaffected by λ, then it follows that $g_{cc}(r)$ is insensitive to λ because the distribution of carboxylate groups is unaffected by λ. This observation is borne out in Figure 9, where it is seen that $g_{cc}(r)$ is insensitive to λ.

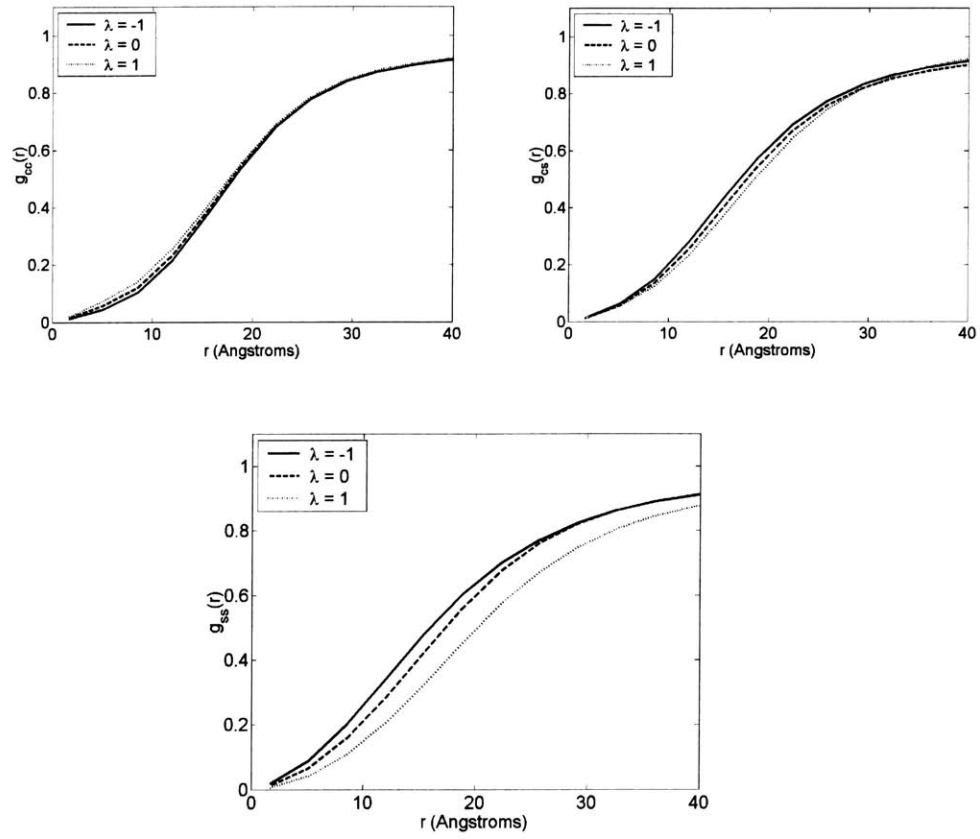


Figure 9 Intermolecular charge group radial distribution functions at 80 mg/ml solution of 24-disaccharide C4S-GAGs at 0.15 M NaCl. $g_{cc}(r)$, $g_{cs}(r)$, and $g_{ss}(r)$ denote the carboxylate-carboxylate, carboxylate-sulfate, and sulfate-sulfate RDFs, respectively.

The intermolecular carboxylate-sulfate group RDF, $g_{cs}(r)$, is also relatively insensitive to λ, despite the variation in sulfate patterns affected by λ. Only the sulfate-sulfate group RDF, $g_{ss}(r)$, is noticeably affected by λ, exhibiting a definitive decrease at small r ($r \leq 40$ Å) in the limit that $\lambda = 1$. As $\lambda \to 1$, sulfate patterns become increasingly 'blocky' as higher and higher proportions of sulfate groups are located on part of the same chain. This results in an increase in the proportion of sulfate groups making *intra*molecular interactions in the system and proportionately decreases the fraction involved in *inter*molecular interactions, only the latter of which contribute to Π (ignoring interactions between sulfate group on different images of the same chain, an effect that goes to zero as the box size becomes large compared to the chain length). This effect leads directly to the observed decrease in the *inter*molecular $g_{ss}(r)$ for $\lambda = 1$ and the concomitant reduction in Π affected by these 'deactivated' sulfate groups.

As with our earlier line of reasoning for $g_{cc}(r)$, the foregoing explanation of the observed changes in $g_{ss}(r)$ hinges on the assumption that the equilibrium distribution of GAG conformations is unaffected by λ. Although this assumption is reasonable because we consider fully mobile, untethered chains, one could argue that chains with higher net charge will on average be spaced further from one another than from chains of lower net charge. This effect, however, would tend to further decrease $g_{ss}(r)$ as $\lambda \to 1$, however, lending further credence to the foregoing explanation.

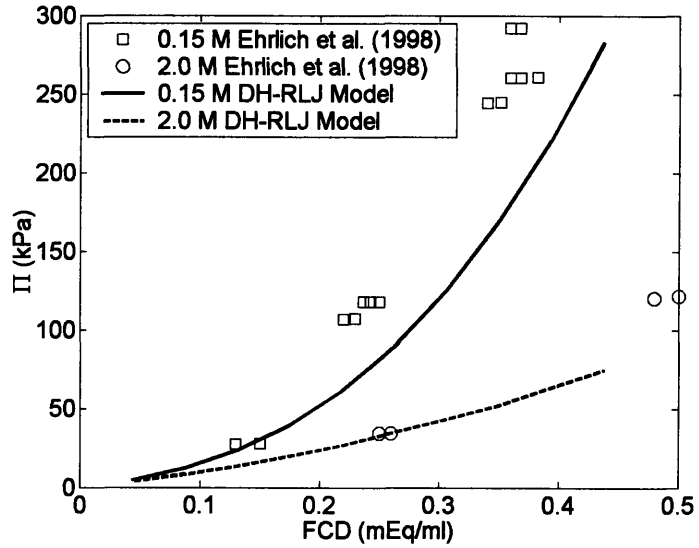


Figure 10 Comparison of theoretical C4S-GAG (16-disaccharide MW) osmotic pressure with the RLJ potential activated (DH-RLJ) and experiment (unknown MW) (Ehrlich and others 1998). Note: 0.44 mEq/ml CS fixed charge density is equivalent to 100 mg/ml of fully sulfated C4S.

Comparison to Experiment

The osmotic pressure of CS solutions has been measured experimentally as a function of NaCl concentration using equilibrium dialysis (Ehrlich and others 1998). Although the overall CS sulfation fraction was measured to be 80% in that study, CS molecular weight was not measured. Considering that Π was found in the present study to have a significant dependence on molecular weight, use of the experimental results as a basis for model validation may reasonably be questioned. Notwithstanding, the theoretical model (16-disaccharide C4S-GAGs) is seen to somewhat underpredict Π at 0.15 and 2.0 M NaCl (Figure 10). Because 16-disaccharides represents a lower bound on the chain length of CS-GAGs present on aggrecan in cartilage (from which the experimental CS was obtained), the model results in Figure 10 represent the *best* possible agreement with experiment, since it was shown that Π decreases with increasing

molecular weight (see Figure 4). The origin of the discrepancy between the model and experimental data in Figure 13 are attributed to a neglect of the low dielectric polysaccharide molecular domain and the use of sub-optimal Debye-Hückel theory (Appendix D).

3.5 Concluding discussion

To the best of my knowledge, the present study represents the first realistic molecular modeling attempt to predict the conformation and osmotic pressure of CS-GAGs in physiological ionic strength solution. Although the CS-GAG and solvent models are coarse-grained in order to achieve computational efficiency, each is based on an underlying atomistic model and therefore has no adjustable parameters. The GAG model accounts for the conformational flexibility of the glycosidic linkages, retains the chemical identity of the carboxylate and sulfate chemical groups, and models both electrostatic and steric non-bond interactions. The suspending 1:1 aqueous salt solution is in osmotic equilibrium with a reservoir of fixed (0.15 M) ionic strength and is modeled using linear PB theory, which is based on a mean-field approximation to the primitive model of electrolytes.

As discussed in the Introduction, the chemical composition of CS-GAGs present on aggrecan varies considerably with the anatomical site and depth of articular cartilage, as well as with the age of the organism and disease state of the tissue. Considering that CS-GAGs play a fundamental role in determining the compressive mechanical properties of the tissue and that chondrocytes, the cells that maintain the matrix, are continuously subject to compressive mechanical loading in their biomechanical environment. It is of primary biological interest to understand how changes in CS chemical composition affect its mechanical properties. Interestingly, the findings of this study suggest that while the average fixed charge density of CS-GAGs affects their osmotic pressure, the disaccharide sulfation type (4 vs. 6) and sulfation pattern do not.

In calculating the osmotic pressure of CS-GAGs at physiological ionic strength (0.15 M NaCl), previous authors have alternatively assumed either that steric intermolecular interactions make no contribution (*i.e.* the osmotic pressure is fully determined by electrostatic effects) (Buschmann and Grodzinsky 1995) or that the steric contribution is independent of ionic strength and is therefore equal to its contribution in the high ionic strength limit (Ehrlich and others 1998). To test these hypotheses the osmotic pressure was computed at 0.15 M NaCl with steric interactions alternatively turned on and off in the model. It was found that the osmotic pressure curves coincided, from which it is concluded that steric interactions do not contribute to the osmotic pressure at physiological ionic strength, lending credence to the assumption of Buschmann and Grodzinsky (1995).

Appendix A Analytical evaluation of the grand potential, Ω

Three integral results are required to evaluate Ω. The results are provided here and the details of their evaluation are provided at the end of this appendix.

Integral result 1

$$\int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} d\mathbf{r} = \frac{4\pi N}{\kappa^2} \tag{19}$$

Integral result 2

$$\int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \sum_{j=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|} d\mathbf{r} = \frac{2\pi}{\kappa}\left(N + \sum_{\substack{i=1 \\ j\neq i}}^{N} e^{-\kappa R_{ij}} \right) \tag{20}$$

Electrostatic internal energy, U

The electrostatic internal energy, U, may be divided into two contributions, that due to the colloid fixed charge density, $U_{fc} = (1/2)\int_V \rho_{fc}(\mathbf{r})\psi(\mathbf{r})d\mathbf{r}$, and that due to the mobile ion charge density, $U_m = (1/2)\int_V \rho_m(\mathbf{r})\psi(\mathbf{r})d\mathbf{r}$, where,

$$\beta e\psi(\mathbf{r};\{\mathbf{R}^N\}) = -\lambda_B \sum_{i=1}^{N_c} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \tag{21}$$

$$\rho_{fc}(\mathbf{r}) = -e\sum_{i=1}^{N} \delta^3(\mathbf{r}-\mathbf{R}_i) \tag{22}$$

$$\rho_m(\mathbf{r}) = -2\beta n_r e^2 \psi(\mathbf{r}) = \frac{e\kappa^2}{4\pi} \sum_{i=1}^{N_c} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \tag{23}$$

where the relations $\rho_m(\mathbf{r}) = e[n_+(\mathbf{r}) - n_-(\mathbf{r})]$ and $n_\alpha(\mathbf{r}) = n_r\left[1 - \beta\nu_\alpha e\psi(\mathbf{r})\right]$ have been used to obtain Eq. (23). Substitution of the foregoing relations results in the mobile ion contribution to the electrostatic internal energy yields (in dimensionless form),

$$\beta U_m = -\frac{1}{8\pi}\kappa^2\lambda_B \int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \sum_{j=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|} d\mathbf{r} \tag{24}$$

which becomes,

$$\beta U_m = -\frac{1}{4}N\kappa\lambda_B - \frac{1}{2}\kappa\lambda_B \sum_{i<j}^{N} e^{-\kappa R_{ij}} \tag{25}$$

upon use of Integral Result 2, and the fixed charge density contribution,

$$\beta U_{fc} = -\frac{1}{2}N\kappa\lambda_B + \lambda_B \sum_{i<j}^{N} \frac{e^{-\kappa R_{ij}}}{R_{ij}} \tag{26}$$

where the result, $\psi_0 = -e\kappa/\varepsilon$, has been employed to evaluate the colloid self-energy, where ψ_0 is the electrostatic potential at the center of a negatively charged (unit valence) point ion due to its surrounding cloud of positively charged mobile ions (Eq. 15-42 in McQuarrie, Statistical Mechanics, 2000). The self-energy and ψ_0 do not include the

screened potentials due to the other $(j \neq i)$ colloids, that contribution is included in the second, Debye-Hückel interaction term in Eq. (26).

Combining the mobile ion and fixed charge contributions yields the final result for the dimensionless electrostatic internal energy,

$$\beta U = -\frac{3}{4} N \kappa \lambda_B + \lambda_B \sum_{i<j}^{N} \frac{e^{-\kappa R_{ij}}}{R_{ij}} \left(1 - \frac{\kappa R_{ij}}{2} \right) \tag{27}$$

Mobile ion entropy, S

In nonlinear PB theory the mobile ion entropy is given by that of an inhomogeous, non-interacting ideal gas,

$$S = -k_B \sum_{\alpha=\pm} \int_V n_\alpha(\mathbf{r}) [\ln n_\alpha(\mathbf{r}) - 1] d\mathbf{r} \tag{28}$$

where the volume integral extends over the space accessible to the mobile ions. To be consistent with linearized PB theory Eq. (28) is expanded to quadratic order about the reservoir densities, n_r,

$$S = \sum_{\alpha=\pm} \left\{ -V k_B \bar{n}_\alpha \ln n_r + V k_B n_r - \frac{k_B}{2 n_r} \int_V \left[n_\alpha(\mathbf{r}) - n_r \right]^2 d\mathbf{r} \right\} \tag{29}$$

where $\overline{n}_\alpha = V^{-1}\int_V n_\alpha(\mathbf{r})d\mathbf{r}$ is the volume-averaged mobile ion density in the colloid/polyelectrolyte solution. Expanding the quadratic term in the integral in Eq. (29) and using the result, $\sum_\alpha \overline{n}_\alpha = 2n_r$ [Eq. (39)], yields,

$$S = 2S_{id,r} - \frac{k_B}{2n_r}\sum_{\alpha=\pm}\left[\int_V n_\alpha(\mathbf{r})^2 d\mathbf{r} - 2n_r\overline{n}_\alpha V + n_r^2 V\right] \tag{30}$$

where $S_{id,r} \equiv Vk_B n_r\left(\ln n_r - 1\right)$ has been defined. The quadratic term in Eq. (30) may be evaluated by noting that the mobile ion densities are given by,

$$n_\alpha(\mathbf{r}) = n_r\left[1 - \beta\nu_\alpha e\psi(\mathbf{r})\right] \tag{31}$$

which, upon substitution of the electrostatic potential in Eq. (21), result in,

$$n_\alpha(\mathbf{r}) = n_r\left(1 + \nu_\alpha\lambda_B\sum_{i=1}^{N}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|}\right) \tag{32}$$

so that,

$$\int_V n_\alpha(\mathbf{r})^2 d\mathbf{r} = \int_V\left(n_r^2 + 2\nu_\alpha\lambda_B n_r^2\sum_{i=1}^{N}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} + \nu_\alpha^2\lambda_B^2 n_r^2\sum_{i=1}^{N}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|}\sum_{j=1}^{N}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|}\right)d\mathbf{r}. \tag{33}$$

Eq. (33) may be simplified to,

$$\int_V n_\alpha(\mathbf{r})^2 d\mathbf{r} = V n_r^2 + \lambda_B^2 n_r^2 \int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \sum_{j=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|} d\mathbf{r} \tag{34}$$

upon anticipation of the results, $\sum_\alpha \nu_\alpha = 0$ and $\nu_\alpha^2 = 1$. Use of Integral Result 2 and $\kappa = (8\pi\lambda_B n_r)^{1/2}$ finally yields,

$$\int_V n_\alpha(\mathbf{r})^2 d\mathbf{r} = V n_r^2 + \frac{1}{4}\lambda_B \kappa n_r \left(N + \sum_{\substack{i=1 \\ j\neq i}}^{N} e^{-\kappa R_{ij}} \right). \tag{35}$$

Substitution of this result back into Eq. (30) results in the final expression for the mobile ion entropy (in dimensionless form),

$$\boxed{\beta TS = 2\beta TS_{id,r} - \frac{1}{4} N \lambda_B \kappa - \frac{1}{4} \lambda_B \kappa \sum_{\substack{i=1 \\ j\neq i}}^{N} e^{-\kappa R_{ij}}}. \tag{36}$$

Electrochemical potential terms, $\mu_\alpha N_\alpha$

The mobile ion electrochemical potential terms may be written,

$$\beta \sum_{\alpha=\pm} \mu_\alpha N_\alpha = \beta \sum_{\alpha=\pm} \overline{n}_\alpha V \ln n_r \tag{37}$$

where equality of the electrochemical potentials in the solution and reservoir has been used, $\mu_\alpha = \mu_r$, and the electrochemical potential in the reservoir is given by, $\mu_r = k_B T \ln n_r$, in the van't Hoff approximation, consistent with PB theory. Volume-

averaging the linearized Boltzmann relation using Integral Result 1 for the mobile ion densities,

$$\bar{n}_\alpha V = n_r V\left(1-\beta\nu_\alpha e\bar{\psi}\right) \tag{38}$$

results in,

$$\bar{n}_\alpha V = n_r V + \frac{1}{2}\nu_\alpha N \tag{39}$$

which upon substitution into Eq. (37) yields the electrochemical potential terms in dimensionless form,

$$\boxed{\beta\sum_{\alpha=\pm}\mu_\alpha N_\alpha = 2Vn_r \ln n_r} \tag{40}$$

Putting together Eqs. (27), (36), and (40) yields the dimensionless grand potential of linearized PB theory,

$$\boxed{\beta\Omega = -\frac{1}{2}N\kappa\lambda_B + \lambda_B\sum_{i<j}^{N}\frac{e^{-\kappa R_{ij}}}{R_{ij}} - 2n_r V} \tag{41}$$

Evaluation of integral results 1, 2, and 3*

Integral result 1

Utilizing spherical polar coordinates and choosing the origin of the basis to coincide with $\mathbf{R}_i$ we have,

$$\int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} d\mathbf{r} = 4\pi N \int_0^\infty r e^{-\kappa r} dr \tag{42}$$

where it is assumed that the volume integration is performed over all of space $(V^{1/3} \gg \kappa^{-1})$. Integration by parts yields,

$$4\pi N \int_0^\infty r e^{-\kappa r} dr = \frac{4\pi N}{\kappa^2}. \tag{43}$$

Thus Integral Result 1 is,

$$\boxed{\int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} d\mathbf{r} = \frac{4\pi N}{\kappa^2}} \tag{44}$$

Integral result 2

We begin by splitting the summation into two integrals involving only self- and cross terms

$$\int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \sum_{j=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|} d\mathbf{r} = \sum_{i=1}^{N} \int_V \frac{e^{-2\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|^2} d\mathbf{r} + \sum_{\substack{i=1 \\ j\neq i}}^{N} \int_V \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|} d\mathbf{r}. \tag{45}$$

* The aid of Dr. Wonmuk Huang in evaluating the following integrals is gratefully acknowledged.

The self-term integral is

$$\sum_{i=1}^{N}\int_{V}\frac{e^{-2\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|^2}d\mathbf{r}=4\pi N\int_0^\infty e^{-2\kappa r}dr \tag{46}$$

which is evaluated directly as,

$$4\pi N\int_0^\infty e^{-2\kappa r}dr=\frac{2\pi N}{\kappa}. \tag{47}$$

where the approach used to evaluate Integral Result 1 has been employed.

To evaluate the cross-terms $(i\neq j)$, we begin by defining the vectors, $\mathbf{r}'\equiv\mathbf{r}-\mathbf{R}_i$, and, $\mathbf{R}_{ij}\equiv\mathbf{R}_j-\mathbf{R}_i$, so that,

$$I_{ij}\equiv\sum_{\substack{i=1\\j\neq i}}^{N}\int_{V}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|}\frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|}d\mathbf{r}=\sum_{\substack{i=1\\j\neq i}}^{N}\int_{V}\frac{e^{-\kappa|\mathbf{r}'|}}{|\mathbf{r}'|}\frac{e^{-\kappa|\mathbf{r}'-\mathbf{R}_{ij}|}}{|\mathbf{r}'-\mathbf{R}_{ij}|}d\mathbf{r}'. \tag{48}$$

Next, we employ spherical polar coordinates, $|\mathbf{r}'-\mathbf{R}_{ij}|=\sqrt{r'^2+R_{ij}^2-2r'R_{ij}\cos\theta}$, to obtain,

$$I_{ij}=\sum_{\substack{i=1\\j\neq i}}^{N}2\pi\int_0^\infty dr\int_0^\pi d\theta\left(re^{-\kappa r}\frac{e^{-\kappa\sqrt{r^2+R_{ij}^2-2rR_{ij}\cos\theta}}}{\sqrt{r^2+R_{ij}^2-2rR_{ij}\cos\theta}}\sin\theta\right) \tag{49}$$

where we have dropped the primes on r' for convenience, and we were able to integrate over, φ, because we aligned r' with the z-axis. Next we use the substitutions, $a \equiv \frac{r^2 + R_{ij}^2}{2rR_{ij}}$ and $b \equiv \sqrt{2rR_{ij}}$, to obtain,

$$I_{ij} = \sum_{\substack{i=1 \\ j \neq i}}^{N} 2\pi \int_0^{\infty} dr \int_0^{\pi} d\theta \left(re^{-\kappa r} \frac{e^{-\kappa b\sqrt{a-\cos\theta}}}{b\sqrt{a-\cos\theta}} \sin\theta \right). \tag{50}$$

Next we use, $x \equiv \cos\theta$ and $dx = -\sin\theta d\theta$, to obtain

$$I_{ij} = \sum_{\substack{i=1 \\ j \neq i}}^{N} -2\pi \int_0^{\infty} dr \int_1^{-1} dx \left(re^{-\kappa r} \frac{e^{-\kappa b\sqrt{a-x}}}{b\sqrt{a-x}} \right) \tag{51}$$

and then we finally use, $y \equiv \sqrt{a-x}$ and $dy = -\frac{dx}{2\sqrt{a-x}}$, to yield,

$$I_{ij} = 4\pi \int_0^{\infty} dr \frac{re^{-\kappa r}}{b} \int_{\sqrt{a-1}}^{\sqrt{a+1}} e^{-\kappa by} dy \tag{52}$$

or

$$I_{ij} = -4\pi \int_0^{\infty} dr \frac{re^{-\kappa r}}{\kappa b^2} \left(e^{-\kappa b\sqrt{a+1}} - e^{-\kappa b\sqrt{a-1}} \right). \tag{53}$$

Substituting back a and b yields,

$$I_{ij} = \sum_{\substack{i=1 \\ j\neq i}}^{N} -\frac{2\pi}{\kappa R_{ij}} \int_0^\infty dr e^{-\kappa r} \left(e^{-\kappa\sqrt{r^2+R_{ij}^2+2rR_{ij}}} - e^{-\kappa\sqrt{r^2+R_{ij}^2-2rR_{ij}}} \right) \tag{54}$$

or

$$I_{ij} = \sum_{\substack{i=1 \\ j\neq i}}^{N} \frac{2\pi}{\kappa R_{ij}} \int_0^\infty dr e^{-\kappa r} \left(e^{-\kappa|r-R_{ij}|} - e^{-\kappa|r+R_{ij}|} \right) \tag{55}$$

or equivalently,

$$I_{ij} = \sum_{\substack{i=1 \\ j\neq i}}^{N} \frac{2\pi}{\kappa R_{ij}} \int_0^\infty dr \left(e^{-\kappa r-\kappa|r-R_{ij}|} - e^{-\kappa r-\kappa|r+R_{ij}|} \right). \tag{56}$$

Next we split the first integral into two parts to account for the absolute value sign,

$$I_{ij} = \sum_{\substack{i=1 \\ j\neq i}}^{N} \frac{2\pi}{\kappa R_{ij}} \left(\int_0^{R_{ij}} e^{-\kappa R_{ij}} dr + \int_{R_{ij}}^\infty e^{-2\kappa r+\kappa R_{ij}} dr - \int_0^\infty e^{-2\kappa r-\kappa R_{ij}} dr \right) \tag{57}$$

which yields upon integration,

$$I_{ij} = \sum_{\substack{i=1 \\ j\neq i}}^{N} \frac{2\pi}{\kappa} e^{-\kappa R_{ij}} . \tag{58}$$

The final result for **Integral Result 2** is then,

$$\boxed{\int_V \sum_{i=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_i|}}{|\mathbf{r}-\mathbf{R}_i|} \sum_{j=1}^{N} \frac{e^{-\kappa|\mathbf{r}-\mathbf{R}_j|}}{|\mathbf{r}-\mathbf{R}_j|} d\mathbf{r} = \frac{2\pi}{\kappa} N + \frac{2\pi}{\kappa} \sum_{\substack{i=1 \\ j \neq i}}^{N} e^{-\kappa R_{ij}}} \tag{59}$$

Evaluation of Ω from generalized grand potential of linear PB theory

Eq. 20 from (Deserno and von Grünberg 2002) is (using our notation),

$$\beta\Omega_{lin,eq} = \frac{1}{2} \int_V \rho_{fc}(\mathbf{r}) \psi(\mathbf{r}) d\mathbf{r} - 2n_r V \tag{60}$$

where the integral term is simply given by Eq. (26),

$$\frac{1}{2} \int_V \rho_{fc}(\mathbf{r}) \psi(\mathbf{r}) d\mathbf{r} = -\frac{1}{2} N\kappa\lambda_B + \lambda_B \sum_{i<j}^{N} \frac{e^{-\kappa R_{ij}}}{R_{ij}} \tag{61}$$

so that, upon substitution, Eq. (60) becomes,

$$\beta\Omega_{lin,eq} = -\frac{1}{2} N\kappa\lambda_B + \lambda_B \sum_{i<j}^{N} \frac{e^{-\kappa R_{ij}}}{R_{ij}} - 2n_r V \tag{62}$$

which is the same result as in Eq. (41).

Appendix B Evaluation of the polyelectrolyte pressure

Consider a system of N particles in volume V at temperature T interacting with the pair-wise additive and volume-dependent effective Hamiltonian, $H^{eff} = u_0(V) + \sum_{i<j} u(r_{ij})$. Using the statistical mechanical result, $F = -k_B T \ln \Upsilon$, where Υ is the partition function,

$$\Upsilon = \frac{1}{N!\Lambda^{3N}} \int d\mathbf{r}^N \exp\left[-\beta H^{eff}(\{\mathbf{r}^N\})\right] \tag{63}$$

and Λ is the de Broglie wavelength, the thermodynamic pressure may be written,

$$P = -\frac{\partial F}{\partial V} = k_B T \frac{\partial \ln \Upsilon}{\partial V}. \tag{64}$$

Following Smith (1987), Eq. (64) may be evaluated using the scaled coordinates, $\mathbf{r} = V^{1/3}\mathbf{s}$, so that the volume dependence of Υ becomes explicit viz., $\Upsilon = \frac{V^N}{N!\Lambda^{3N}} \int d\mathbf{s}^N \exp\left[-\beta H^{eff}(\{V^{1/3}\mathbf{s}^N\})\right]$, as,

$$\frac{\partial \ln \Upsilon}{\partial V} = \frac{1}{\Upsilon}\frac{\partial \Upsilon}{\partial V} = \frac{N}{V} - \beta\left\langle \frac{\partial H^{eff}}{\partial V} \right\rangle. \tag{65}$$

Finally, using the transformation, $dV = 3V dr_{ij} / r_{ij}$ (at fixed **s**, perform a homeothetic rescaling of the volume so that, $\mathbf{r} = V^{1/3}\mathbf{s}$, becomes, $d\mathbf{r} = \mathbf{s}V^{-2/3}dV/3$), the volume derivative of the effective Hamiltonian becomes,

$$\frac{\partial H^{eff}}{\partial V} = \frac{\partial u_0}{\partial V} + \frac{1}{3V}\sum_{i<j}^{N} r_{ij}\frac{\partial u(r_{ij})}{\partial r_{ij}}. \tag{66}$$

so that the pressure may be written,

$$P = nk_BT - \frac{\partial u_0}{\partial V} - \frac{1}{3V}\left\langle \sum_{i<j}^{N} r_{ij}\frac{\partial u(r_{ij})}{\partial r_{ij}} \right\rangle \tag{67}$$

where $n \equiv N/V$ is the average number density of interacting particles in the system and nk_BT is their ideal gas contribution to the pressure.

Turning to the polyelectrolyte model considered in the present study, in which $u_0 = -2n_rVk_BT$, and employing the molecular virial theorem (Theodorou and others 1993), the pressure in the polyelectrolyte compartment is,

$$P^I = n_pk_BT + 2n_rk_BT + \frac{1}{3V}\left\langle \sum_{m}\sum_{i<j}^{N} R_{c_ic_j}\frac{\partial u^{(m)}(R_{ij})}{\partial R_{ij}} \right\rangle \tag{68}$$

where n_p is the mean number density of polyelectrolytes, n_r is the reservoir mobile ion density, the outer summation (m) is over the different types of non-bonded interaction

potentials present in the model (the Debye-Hückel and the repulsive Lennard-Jones potential when activated), the inner summation (i, j) is over all intermolecular non-bond sites in the system (including non-bond interactions between sites on different images of the same chain), and $R_{c_i c_j}$ is the distance between the centers of mass of the chain images to which non-bond sites i and j belong.

Appendix C Statistical description of CS-GAG copolymers

A traditional first-order Markov model is used for the statistical description of CS-GAG copolymers (Odian 1981). Following Fredrickson and Milner (1991) and Fredrickson and others (1992), copolymers consist of two types of monomers, A and B. In the present context A and B represent sulfated and unsulfated disaccharide 'monomers', respectively, where each chain has the same degree of polymerization, N. The copolymer sequences are statistically determined by the overall fraction of A, $f \equiv p(\mathrm{A})$, and the transition probability matrix,

$$\mathbf{p} = \begin{bmatrix} p(\mathrm{A}|\mathrm{A}) & p(\mathrm{A}|\mathrm{B}) \\ p(\mathrm{B}|\mathrm{A}) & p(\mathrm{B}|\mathrm{B}) \end{bmatrix} \tag{69}$$

where $p(K|L)$ is the conditional probability that a monomer of type K immediately follows a monomer of type L on the chain. Conservation of probability requires that, $p(\mathrm{A}|\mathrm{A}) + p(\mathrm{B}|\mathrm{A}) = 1$ and $p(\mathrm{B}|\mathrm{B}) + p(\mathrm{A}|\mathrm{B}) = 1$, and stationarity of the Markov process requires that, $f = fp(\mathrm{A}|\mathrm{A}) + (1-f)p(\mathrm{A}|\mathrm{B})$, be satisfied. These three constraints leave only one degree of freedom in **p**, which is chosen to be the linear combination, $\lambda \equiv p(\mathrm{A}|\mathrm{A}) + p(\mathrm{B}|\mathrm{B}) - 1$. The two parameters (f, λ) then fully specify the statistical description of AB copolymers in the system, where $(0 \le f \le 1)$ and $(-1 \le \lambda \le 1)$.

As demonstrated by (Fredrickson and others 1992), λ denotes the strength of the chemical correlations along the chain. $\lambda = -1$ requires that $p(\mathrm{A}|\mathrm{A}) = p(\mathrm{B}|\mathrm{B}) = 0$,

resulting in an alternating AB copolymer where there is only one molecular species present. When $\lambda = 0$ there are no chemical correlations between A and B monomers, resulting in a mixture of ideal random AB copolymers. Finally, $\lambda = 1$ results in a mixture of A and B homopolymers with population fractions f and $(1-f)$, respectively.

Appendix D Evaluation of model shortcomings in predicting Π

There are two primary shortcomings of the current GAG model as pertains to the computation of osmotic pressure, (I) use of the Debye-Hückel equation instead of the nonlinear PB equation and (II) neglect of the effects of GAG low dielectric and salt exclusion on the electrostatic internal energy.

I. Linearized Poisson-Boltzmann theory

Linearized PB theory is employed to compute the osmotic pressure of GAG solutions in osmotic equilibrium with a 1:1 salt reservoir of fixed ionic strength. In deriving Eq. (17), we employed the conventional, Debye-Hückel linearization of the PB equation, where linearization is performed about the reference (zero) electrostatic potential in the reservoir and the corresponding reservoir mobile ion densities, n_r. This choice requires that, $e\psi(\mathbf{r}) \ll k_B T$, be satisfied everywhere in the spatial domain, $\mathbf{r}$. In order to gauge the error in Π associated with linearization of the PB equation we compare the osmotic pressure predicted by the linear and nonlinear PB equation in the context of the cylindrical cell model, for which a numerical solution to the nonlinear PB equation may be obtained.

The cylindrical cell model consists of an inner cylinder of radius, a, with uniform charge density, σ, and a concentric outer cylinder of radius R where the component of the electric field normal to the surface is zero by Gauss' law (electroneutrality in the cell). The analytical solution for the osmotic pressure for the ($\bar{\psi} = 0$) linearized PB equation in

the context of the cylindrical cell model and a 1:1 electrolyte is given by (Deserno and von Grünberg 2002),

$$\Pi = n_r \theta^2 \gamma^2 k_B T - 2 n_r k_B T \tag{70}$$

where $\theta \equiv \bar{n}_c / 2n_r$, $\bar{n}_c$ is the average counterion density in the cell and n_r is the mobile ion reservoir density. The parameter, $\gamma \equiv (1-\phi)/(2D\phi^{1/2})$, where $\phi = (a/R)^2$ is the volume fraction and $D = K_1(x_0)I_1(X) - K_1(X)I_2(x_0)$, where $x_0 \equiv \kappa a$ and $X \equiv \kappa R$ are the dimensionless inner and outer radii, respectively. Finally, I_1 and K_1 are first order modified Bessel functions of the first and second kind, respectively.

An analytical solution is not available for the nonlinear Poisson-Boltzmann equation with added salt for the cylindrical cell model, so we utilize a simple finite difference scheme with Newton-Raphson iteration to obtain its solution. The nonlinear PB equation with its appropriate boundary conditions on the inner (S_{in}) and outer (S_{out}) cylindrical surfaces is (in dimensionless form),

$$\begin{aligned} \nabla^2 \psi &= \kappa^2 \sinh \psi && \text{in } V_{cell} \\ \hat{\mathbf{n}} \cdot \vec{\nabla} \psi &= 4\pi \lambda_B \sigma && \text{on } S_{inner} \\ \hat{\mathbf{n}} \cdot \vec{\nabla} \psi &= 0 && \text{on } S_{outer} \end{aligned} \tag{71}$$

where $\hat{\mathbf{n}}$ is the surface unit normal. A standard finite difference representation of the Laplacian in (71) in cylindrical coordinates assuming cylindrical symmetry ($\partial / \partial \theta = 0$) and spatial invariance in the axial direction ($\partial / \partial z = 0$) yields,

$$\frac{\psi_{i-1}-2\psi_i+\psi_{i+1}}{2(\Delta r)^2}+\frac{\psi_{i+1}-\psi_{i-1}}{r_i\Delta r}=\kappa^2\sinh\psi_i \tag{72}$$

which may be written in matrix form as $\mathbf{K}\boldsymbol{\psi}=\mathbf{R}(\boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is the discretized vector of ψ values on the domain $[a,R]$ and $\mathbf{R}$ denotes the nonlinear forcing term, $\sinh(\psi_i)$. The boundary conditions are applied on $\boldsymbol{\psi}$ by imposing the first-order spatial derivatives provided in Eq. (71). Newton-Raphson iteration may be used to solve for $\boldsymbol{\psi}_{(k+1)}$ given $\boldsymbol{\psi}_{(k)}$ using,

$$\left[\mathbf{K}-\frac{\partial\mathbf{R}}{\partial\boldsymbol{\psi}}\right]_{(k)}\Delta\boldsymbol{\psi}=-\left[\mathbf{K}\boldsymbol{\psi}-\mathbf{R}\right]_{(k)} \tag{73}$$

where $\Delta\boldsymbol{\psi}\equiv\boldsymbol{\psi}_{(k+1)}-\boldsymbol{\psi}_{(k)}$. Once the converged solution to the electrostatic potential is obtained, the osmotic pressure is calculated from the boundary density rule (Deserno and von Grünberg 2002),

$$\beta\Pi=\sum_{i=\pm}n_i(R)-2n_r \tag{74}$$

where $n_i(r)=n_r e^{-\beta\nu_i e\psi(r)}$.

We choose the following cell model parameters for chondroitin-sulfate: a = 6 Å (Ogston and others 1973) and b = 5 Å, where $2b$ is the average length of a CS-GAG disaccharide, found to be 9.5 Å in this study (the disaccharide length was computed to be 9.5 Å from the rod-like configuration end-to-end distance of C4S- and C6S-GAGs). The surface charge density is related to a and b by, $\sigma=e/(2\pi ab)$. The outer cell radius is

related to the CS concentration by, $R = (2bN_A \pi c_{CS} / M_{CS})^{-1/2}$, where $N_A = 6.022 \times 10^{23}$ is Avogadro's number, $M_{CS} = 457$ g/mol is the molecular weight of a CS disaccharide, and c_{CS} is the CS concentration.

Figure 11 demonstrates that the linear PB solution considerably over-predicts the osmotic pressure with respect to the nonlinear solution, particularly for higher concentrations, where the relative error, $\varepsilon \equiv (\Pi_{lin} - \Pi_{nlin}) / \Pi_{lin}$ is nearly 0.6! Although the cylindrical cell model is clearly only qualitatively similar to the discrete-site, conformationally flexible CS-GAG model employed in the present study, it does provide a qualitative understanding of the magnitude of the error that is involved in utilizing the Debye-Hückel potentials to compute Π.

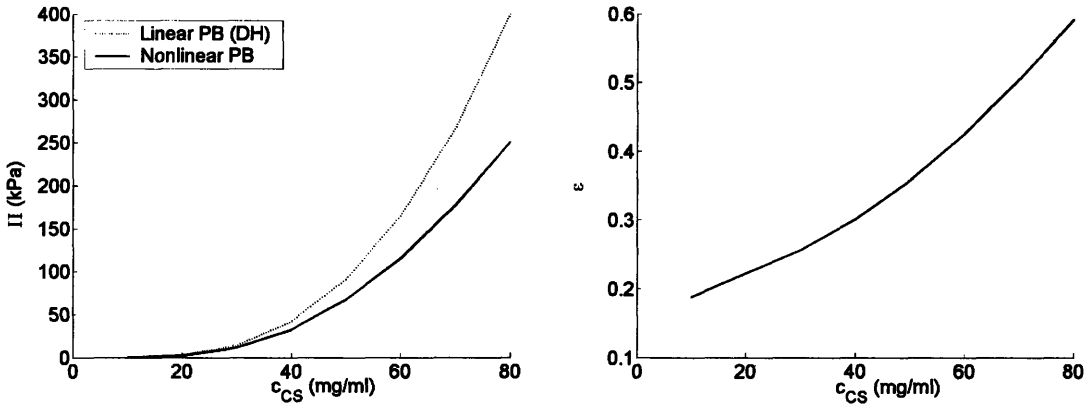


Figure 11 Osmotic pressure (left) as computed from the cylindrical cell model employing the linear and nonlinear PB equation and cell model parameters corresponding to CS-GAGs (a = 5.5 Å and b = 5 Å) and relative error, ε, (right) between the linear and nonlinear PB-based solutions.

Interestingly, if linearization of the PB equation is performed about the optimal, *Donnan* potential, $\bar{\psi}$, then the criterion, $e(\psi(\mathbf{r}) - \bar{\psi}) \ll k_B T$, must be satisfied throughout the domain, $\mathbf{r}$ (Deserno and von Grünberg 2002). This criterion is less strict than the

foregoing Debye-Hückel criterion, $e\psi \ll k_B T$, and may certainly be satisfied in cases when the latter is violated. Optimal linearization will perform particularly well with respect to Debye-Hückel linearization when the Debye length is on the order of or less than the inter-molecular spacing. Application of optimal linear PB theory to the polyelectrolyte model contained herein is in progress.

II. Effect of neglecting the low dielectric and salt-exclusion of the GAG molecular domain

In the current study GAGs are treated as a series of bonded point charges, which ignores the fact that the GAG molecular domain (1) excludes mobile salt ions and (2) has a dielectric permittivity between 1-4, significantly lower than that of water (78.5).

The ramifications of this assumption are best illustrated by the completely general result for Ω of linearized PB theory ($\bar{\psi} = 0$, $\bar{n}_i = n_r$) presented by (Deserno and von Grünberg 2002) (their Eq. 20),

$$\beta\Omega_{lin,eq} = \frac{1}{2}\int_V \rho_{fc}(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} - 2Vn_r. \tag{75}$$

Thus only the fixed charge density contribution to the electrostatic internal energy needs to be considered. Moreover, as shown in the Modeling section the self-energy contribution to the GAG osmotic pressure is independent of volume, and therefore only the GAG-GAG *interaction* contribution to Eq. (75) needs to be considered. This

fortuitous result is a consequence of the fact that when linearized PB theory with ($\bar{\psi} = 0$, $\bar{n}_i = n_r$) is used, the mobile ion entropy cancels with the mobile ion electrostatic internal energy and some terms of the electrochemical potential (see Appendix A), leaving only those terms shown in Eq. (75).

To approximate the effects of the GAG molecular domain on the electrostatic interaction (internal) energy, the linear PB equation was solved in 3D using an all-atom representation of GAGs. An additional complication arises in the evaluation of, $\frac{1}{2}\int_V \rho_{fc}(\mathbf{r})\psi(\mathbf{r})d\mathbf{r}$, because of the non-linearity introduced by the dielectric and ionic strength discontinuity between the solvent and GAG domains. As a first approximation to obtaining a pair-wise additive interaction energy for use in the Monte Carlo simulations the molecular configuration corresponding to two parallel helices was chosen. The electrostatic interaction energy of the helices was computed as a function of the distance separating their central axes using (A) the coarse-grained GAG model as presented in the Modeling section and (B) the continuum linear PB equation with explicit partial atomic charges, low dielectric, and salt exclusion accounted for using the finite difference program DELPHI (Nicholls and Honig 1991).

The electrostatic interaction energy per sugar computed using the coarse-grained GAG model is shown to significantly underestimate the all-atom PB solution (Figure 12, left). This is most likely due to the electric field-line focusing that occurs between the two interacting helices when the low GAG dielectric is accounted for. It was verified that the disagreement is not due to the monopole approximation utilized in the coarse-grained model. Verification was performed by comparing the coarse-grained model interaction

energy to that of the all-atom GAG model using a pair-wise atomic Debye-Hückel interaction potential implemented in CHARMM (Brooks and others 1983).

In order to correct for the dielectric/salt exclusion effect in the coarse-grained model, the electrostatic non-bond potential was empirically adjusted in order to match the all-atom PB result. As it turns out, a Debye-Hückel interaction potential with a dielectric permittivity of $\varepsilon = 37.4$ fits the all-atom model results very well for the three helix orientations considered. It is emphasized that there is no (to the best of the author's knowledge) physical reason why a Debye-Hückel potential with a dielectric permittivity of $\varepsilon = 37.4$ matches the all-atom results. Additionally, further work is required to test whether the effective ε obtained is predictive at other salt concentrations than that which it was parameterized at (0.1 M).

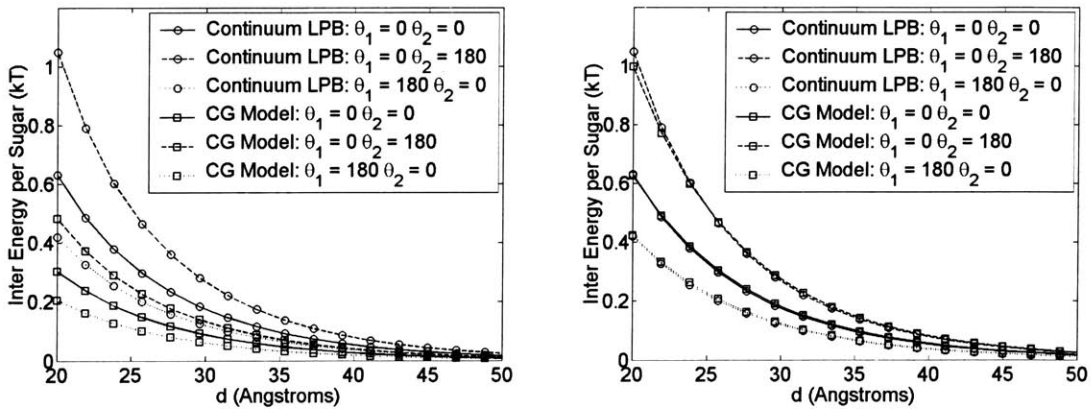


Figure 12 Electrostatic interaction energy for two identical parallel C4S helices (16-disaccharide MW) at different relative orientations specified by θ_1 and θ_2 computed using the continuum 3D linear PB (LPB) equation and the simple Coarse-grained model Debye-Hückel potentials assuming a uniform dielectric of 78.5 (left) and 38.5 (right). The ionic strength is 0.1 M.

The osmotic pressure was recomputed for 16-disaccharide C4S-GAGs at 0.15 M using the parameterized Debye-Hückel non-bond potential with $\varepsilon = 37.4$ (Figure 13) and

compares better with experiment, as expected since the interaction potential is considerably stronger now. There are two points to make. First, the comparison with experiment is not rigorous nor strictly valid because we do not know the molecular weight of the sample used in the experiment. Second, the non-bond Debye-Hückel potential was parameterized to the all-atom PB solution at 0.1 M NaCl and subsequently employed at 0.15 M NaCl. The inverse Debye length, κ, was, however, not adjusted to match the PB solution and therefore it is hoped that the empirical potential is transferable to other salt concentrations. Further work is required to test this hypothesis.

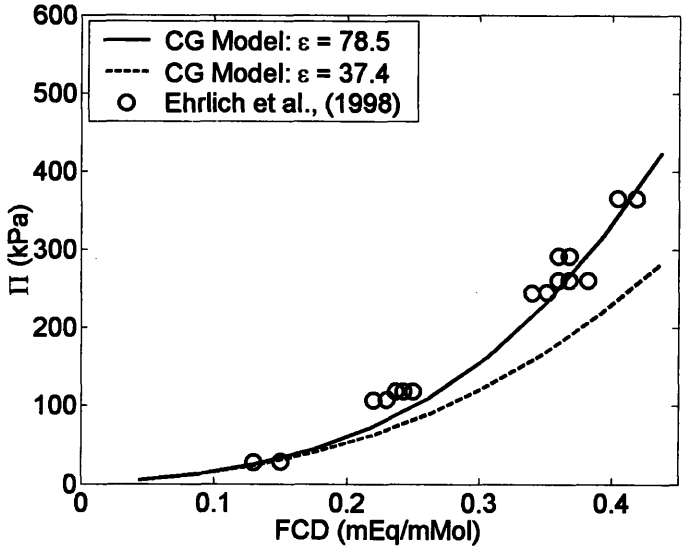


Figure 13 Comparison between experimental (Ehrlich and others 1998) and theoretical osmotic pressure for the original, simple Debye-Hückel coarse-grained model ($\varepsilon = 78.5$) and the 'continuum Poisson-Boltzmann-corrected model' ($\varepsilon = 37.4$).

References for Chapter 3

Barrat JL, Joanny JF. 1996. Theory of polyelectrolyte solutions. Advances in Chemical Physics, Vol Xciv. New York: John Wiley & Sons, Inc. p 1-66.

Bayliss MT, Osborne D, Woodhouse S, Davidson C. 1999. Sulfation of chondroitin sulfate in human articular cartilage - The effect of age, topographical position, and zone of cartilage on tissue composition. Journal of Biological Chemistry 274(22):15892-15900.

Belloni L. 2000. Colloidal interactions. Journal of Physics-Condensed Matter 12(46):R549-R587.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry 4(2):187-217.

Buschmann MD, Grodzinsky AJ. 1995. A molecular-model of proteoglycan-associated electrostatic forces in cartilage mechanics. Journal of Biomechanical Engineering-Transactions of the Asme 117(2):179-192.

Deserno M, von Grünberg HH. 2002. Osmotic pressure of charged colloidal suspensions: A unified approach to linearized Poisson-Boltzmann theory. Physical Review E 66(1):art. no. 011401.

Doege KJ, Sasaki M, Kimura T, Yamada Y. 1991. Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan - Human-specific repeats, and additional alternatively spliced forms. Journal of Biological Chemistry 266(2):894-902.

Ehrlich S, Wolff N, Schneiderman R, Maroudas A, Parker KH, Winlove CP. 1998. The osmotic pressure of chondroitin sulphate solutions: Experimental measurements and theoretical analysis. Biorheology 35(6):383-397.

Fredrickson GH, Milner ST. 1991. Thermodynamics of random copolymer melts. Physical Review Letters 67(7):835-838.

Fredrickson GH, Milner ST, Leibler L. 1992. Multicritical phenomena and microphase ordering in random block copolymer melts. Macromolecules 25(23):6341-6354.

Hardingham T. 1998. Chondroitin sulfate and joint disease. Osteoarthritis and Cartilage 6:3-5.

Hascall VC, Hascall GK. 1981. Proteoglycans. Cell biology of extracellular matrix:39-63.

Lewis S, Crossman M, Flannelly J, Belcher C, Doherty M, Bayliss MT, Mason RM. 1999. Chondroitin sulphation patterns in synovial fluid in osteoarthritis subsets. Annals of the Rheumatic Diseases 58(7):441-445.

Madras N, Sokal AD. 1988. The pivot algorithm - A highly efficient Monte Carlo method for the self-avoiding walk. Journal of Statistical Physics 50(1-2):109-186.

Mankin HJ, Lippiello L. 1971. The glycosaminoglycans of normal and arthritic cartilage. The Journal of Clinical Investigation 50:1712-1719.

Maroudas A, Bannon C. 1981. Measurement of swelling pressure in cartilage and comparison with the osmotic pressure of constituent proteoglycans. Biorheology 18(3-6):619-632.

Nicholls A, Honig B. 1991. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. Journal of Computational Chemistry 12(4):435-445.

Odian G. 1981. Principles of Polymerization. New York: Wiley-Interscience

Ogston AG, Preston BN, Wells JD, Snowden JM. 1973. Transport of compact particles through solutions of chain polymers. Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences 333(1594):297-316.

Plaas AHK, West LA, Wong-Palms S, Nelson FRT. 1998. Glycosaminoglycan sulfation in human osteoarthritis - Disease-related alterations at the non-reducing termini of chondroitin and dermatan sulfate. Journal of Biological Chemistry 273(20):12642-12649.

Plaas AHK, Wong-Palms S, Roughley PJ, Midura RJ, Hascall VC. 1997. Chemical and immunological assay of the nonreducing terminal residues of chondroitin sulfate from human aggrecan. Journal of Biological Chemistry 272(33):20603-20610.

Platt D, Bird JLE, Bayliss MT. 1998. Ageing of equine articular cartilage: structure and composition of aggrecan and decorin. Equine Veterinary Journal 30(1):43-52.

Roughley PJ, Lee ER. 1994. Cartilage proteoglycans - Structure and potential functions. Microscopy Research and Technique 28(5):385-397.

Smith W. 1987. Coping with the pressure! How to calculate the virial. CCP Newsletter 26(43).

Sokal AD. 1996. Monte Carlo methods for the self-avoiding walk. Nuclear Physics B:172-179.

Tamashiro MN, Schiessel H. 2003. Where the linearized Poisson-Boltzmann cell model fails: Spurious phase separation in charged colloidal suspensions. Journal of Chemical Physics 119(3):1855-1865.

Theodorou DN, Boone TD, Dodd LR, Mansfield KF. 1993. Stress tensor in model polymer systems with periodic boundaries. Makromolekulare Chemie-Theory and Simulations 2(2):191-238.

Trizac E, Hansen JP. 1996. Free energy of electric double layers around finite particles. Journal of Physics-Condensed Matter 8(47):9191-9199.

Urban JPG, Maroudas A, Bayliss MT, Dillon J. 1979. Swelling pressures of proteoglycans at the concentrations found in cartilaginous tissues. Biorheology 16(6):447-464.

von Grünberg HH, van Roij R, Klein G. 2001. Gas-liquid phase coexistence in colloidal suspensions? Europhysics Letters 55(4):580-586.