Monte Carlo study of titration of linear polyelectrolytes

Christopher E. Reed and Wayne F. Reed

Citation: J. Chem. Phys. **96**, 1609 (1992); doi: 10.1063/1.462145
View online: http://dx.doi.org/10.1063/1.462145
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v96/i2
Published by the American Institute of Physics.

Monte Carlo study of titration of linear polyelectrolytes

Christopher E. Reed and Wayne F. Reed
Tulane University, Department of Physics, New Orleans, Louisiana 70018

(Received 1 May 1991; accepted 7 October 1991)

An off-lattice Metropolis Monte Carlo algorithm with reptation is used to find the average fractional ionization $\overline{\alpha}$ as a function of pH for a generic ionizable linear polyelectrolyte in a salt solution. The polyelectrolyte is treated as a threefold rotational isomeric state model polymer; each unit can bear a negative charge or not with intrinsic ionization constant pK_a. Debye–Hückel screening is assumed between the charges. For computational convenience, the dielectric constant of the polymer is taken to be that of the solvent. The number of units N was either 50 or 100. Monte Carlo results were collected for various Debye screening lengths at six combinations of number of chain units N, bond angle θ, and Manning parameter when fully charged, ξ_0. For four of the combinations, ξ_0 was 1 to take partial account of counterion condensation. These runs had N and θ of 50 and 1°, 50 and 70°, 100 and 1°, and 100 and 70°. The fifth combination had $N = 50$, $\theta = 70°$, and $\xi_0 = 2.85$. The sixth had $N = 50$, $\theta = 27.34°$, and $\xi_0 = 0.72$, for comparison with data for hyaluronate. The Monte Carlo results are compared to third nearest-neighbor linear Ising type calculations and to simple mean field theories in α. Mean field theory in α worked very well in the (nearly rodlike) $\theta = 1°$ cases using the known distance between units. Mean field theory in α using an estimate for the distance between units based on the ideas of electrostatic persistence length and excluded volume worked equally well for the $\theta = 1°$ cases and moderately well for the $\theta = 70°$ cases. The free energy and entropy per simulated chain were calculated by thermodynamic integration of the Monte Carlo results for $\overline{\alpha}$ as a function of pH.

I. INTRODUCTION AND BACKGROUND

A. Goals

Using a Metropolis Monte Carlo algorithm to simulate variably ionizable polyelectrolytes made of units having an intrinsic pK_a, we obtain $\overline{\alpha}$, the average fraction of the units which are ionized. (In what follows, α will mean the fractional ionization of some particular polyelectrolyte molecule; $\overline{\alpha}$ will mean the average of α over many molecules.) These results for $\overline{\alpha}$ are compared with the results of linear Ising model calculations to third nearest neighbors and with the results of simple mean field theories for the distribution of charges. Results from a scan of $\overline{\alpha}$ as a function of $pH - pK_a$ are used to calculate the free energy and entropy of a population of chains. Finally, comparison is made to experimental data of Cleland for the titration of hyaluronate.[1] The same fairly artificial assumptions were made about the polymer and the solvent in both this initial Monte Carlo work and in the calculations. These reduce the realism of the simulation, but allow testing of the additional simplifying approximations in the calculations.

B. Assumptions

The first assumption is of a very dilute solution; only one polymer molecule is simulated. The polymer units are assumed to bear either one charge or no charge, i.e., ionization of monomers is assumed to be complete. For definiteness, the charges are taken to be negative; the polyelectrolyte is thus a polyanion and polyacid rather than a polycation and polybase. Within our assumptions, what follows can be applied to polybases by changing the sign of $pH - pK_a$.

The $\overline{\alpha}$ of a polymer molecule is assumed to differ from that of isolated units only because of electrostatic repulsion between those units which are charged. To calculate this, Debye–Hückel screening is assumed with no corrections for the finite size of ions or polymer units. This approximation could be addressed by Monte Carlo simulations using explicit counterions.

It is assumed that the polymer has the same dielectric constant ϵ as the solvent, although polymers generally have dielectric constants less than those of common solvents and far less than that of water. This major, albeit common, approximation is made in all the Monte Carlo simulations of flexible polyelectrolytes of which we are aware because it simplifies greatly the task of calculating electrostatic energies; it is made in many theories for the same reason. It may be least inaccurate under conditions of low charge density along the chain and of highly extended chains, circumstances in which most of the material between any two charges would be solvent. Experimentally, this assumption might be realistic if a charged polymer could be dissolved in a solvent resembling its monomers, reducing the discrepancy in ϵ.

A threefold rotational isomeric state model is assumed with the charges located on the polymer backbone. This is probably the simplest rotational isomeric model. Our implementation of it has been described.[2] Flory's[3] book ably expounds rotational isomeric state models, more complex versions of which can yield "ball and stick" models of polymers accurate at the classical mechanics level, e.g., hyaluronate was so modeled in Ref. 4.

Briefly, polymer units are separated by a constant bond length D and the bond angle θ is also constant. The zero of

 0021-9606/92/021609-12$06.00

angle is defined so that $\theta = 0°$ for a rodlike polymer. Once the positions of three consecutive units are fixed, the next unit has three possible positions equally spaced in azimuthal angle corresponding to the three allowed bond types—*trans*, *gauche*$^+$, and *gauche*$^-$. Thus, if there are N polymer units, there are 3^{N-3} possible spatial configurations before considering possible steric hindrances. Specifically, Cartesian coordinates can be chosen so that any three consecutive units are at (0,0,0), D(0,0,1), and $D(\sin\theta,0,1+\cos\theta)$. Then the next unit is at

$$r = D(\sin\theta, 0, 2+\cos\theta) \tag{1a}$$

if these units are in the *trans* configuration, or

$$r = D\left(\sin\theta + \frac{3}{2}\sin\theta\cos\theta,\ \pm\frac{3^{1/2}}{2}\sin\theta,\ \frac{3}{2}+\cos\theta+\frac{3}{2}\cos^2\theta\right), \tag{1b}$$

if they are in the *gauche*$^\pm$ positions.

Counterion condensation is not considered except by remaining aware of the Manning limit[5]; if the valence of the counterions is V, the linear charge density along the polymer should not exceed about $1/V$ elementary charges per "Bjerrum length" $l_B \leq$,

$$l_B \equiv \frac{e^2}{4\pi\epsilon k_B T}, \tag{2}$$

in meter-kilogram-second-ampere units.

For a long thin straight uniformly charged rod in the Debye–Hückel approximation with infinitely small counterions, counterions will condense onto the polyelectrolyte until the net charge density is reduced to this critical value; this may be approximately true in more realistic models.[5]

We define ξ_0 to be the number of elementary charges per l_B along the polymer when it is fully charged and the Manning parameter ξ of a partially charged chain to be its average number of charges per l_B. Therefore $\xi = \alpha\xi_0$. If the counterions are of valence 1, ξ should not exceed 1[5]; ξ_0 can perhaps exceed 1 if $\bar{\alpha}$ is small enough. We consider the length L of a threefold rotational isomeric state model polymer to be its maximum end-to-end length, which occurs when all bonds are in the *trans* configuration. So

$$L \cong D(N-1)\cos(\theta/2), \tag{3}$$

where N is the number of units. Equation (3) is exact if N is odd. We consider ξ_0 to be

$$\xi_0 \equiv N(q/e)l_B/L, \tag{4}$$

where q/e is the number of electronic charges born by a charged unit.

It is shown in Appendix A that under our assumptions, the average Gibbs free energy per individual polymer chain is[6]

$$F = \bar{U} - TS - \ln(10)k_B TN\bar{\alpha}(pH - pK_a), \tag{5}$$

where $\bar{U}$ and S are the average electrostatic energy and configurational entropy per polymer molecule. The chain is considered to be one of a great many chains in dilute solution at a given T, $pH - pK_a$, κ^{-1}, and pressure P. Defining

$$X \equiv \ln(10) \times k_B TN(pH - pK_a), \tag{6}$$

$$F = \bar{U} - TS - X\bar{\alpha}. \tag{7}$$

This implicitly neglects, e.g., chemical effects in which pK_a of an ionizable group depends on its environment.[7]

In the Debye–Hückel model used here, the electrostatic energy in a particular spatial and charge configuration is

$$U = \frac{q^2}{4\pi\epsilon}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\alpha_i\alpha_j\frac{\exp(-\kappa r_{ij})}{r_{ij}}, \tag{8}$$

where the polymer units are numbered consecutively from 1 through N, α_i is 1 when unit i is charged and 0 if it is not, r_{ij} is the distance between units i and j, and κ^{-1} is the Debye screening length.

This model is specified by the dimensionless parameters N, θ, $pH - pK_a$, ξ_0, κ^{-1}/D, and D/l_B. This last is fixed at 0.838, approximately right for polysaccharides.[4]

We note that a given pH implies a minimum effective salt concentration. However, for a given $pH - pK_a$, pK_a can be chosen so that the $pH = 7$ and the minimum effective salt concentration $\cong 10^{-7}$ M in H_2O.

C. Previous Monte Carlo work on partially ionized polyelectrolytes

There are quite a few recent Metropolis Monte Carlo investigations of linear polyelectrolytes, some of which we have already mentioned.[2,8] Several have investigated partially ionized polyelectrolytes by using charged units regularly spaced among uncharged units.[9–12] It is not clear how to do this for values of $\bar{\alpha}$ that are not simple fractions, or indeed when $1/\bar{\alpha}$ is not an integer. As pointed out by Cleland,[1] there is some doubt whether this is a generally appropriate method for studying $pH - pK_a$ controlled partial ionization, although of course some polymers do have regularly alternating charge groups.

Also, Brender[13] has scanned the Manning parameter ξ_0 in fully charged chains by varying T.

II. METHODS

A. Metropolis Monte Carlo with ionizable units

The Metropolis Monte Carlo program used here has been discussed.[2] The only changes in it are that the units are ionizable and the incorporation of $pH - pK_a$ into the free energy for a polyanion [Eq. (5)]. This also was explained briefly[8]; as it is central to the present work, it seems appropriate to justify it in more detail.

With ionizable units, there are now six possible states for each unit—two charge states times three rotational states. The first three rotational states (counting from either end of the polymer) are irrelevant, although including them in the program does no harm. From Eq. (7), the probability p of a particular chain configuration should be

$$p \propto \exp[-(U - X\alpha)/k_B T], \tag{9}$$

where U and α are the energy and fractional ionization of that particular configuration. This is shown in Appendix A.

Simply generating configurations and then weighting them by this factor would waste computer time for large N, as most of the configurations generated would be of such low probability as to be irrelevant. The well-known Metropolis

Monte Carlo[14,15] method is a convenient means to avoid this problem. It can be used to generate samples with any desired probability distribution, as long as one knows the desired relative probabilities of any two states.[14,15]

To be explicit, an end of the polymer is chosen at random (probability 1/2). Then a new trial polymer is constructed by adding a new unit onto that end in one of six possible states (charged or uncharged and in a *trans*, *gauche*$^+$, or *gauche*$^-$ position) each with probability 1/6, and lopping one unit off the other end of the polymer. The program then finds ΔA, the change in $U - X\alpha$ in going to the trial state. A hard sphere repulsion can be incorporated by adding a large multiple of k_BT to U when two units are closer than the hard sphere diameter. The rule used was that if $\Delta A \leqslant 0$, an attempted transition always succeeds, while if $\Delta A \geqslant 0$, the transition succeeds with probability $\exp(-\Delta A/k_BT)$.[14] If the transition is successful, the trial configuration becomes the new configuration.

The proof that this procedure leads to the desired probability distribution is not given here. It uses the fact that if a set of states are connected one to the other by transition probabilities which are nonzero and less than one, their relative probabilities eventually approach some unique set of "equilibrium values."[16]

This cycle is repeated many times. Statistics on the model polymer are collected every N^2 reptation attempts, so-called "stroboscopic sampling."[9] This is because, in reptation, the polymer slides back and forth. If reptation attempts in each direction are successful with equal probability, it takes on the order of $N^2/4$ successful reptation attempts to generate an entirely new polymer; not all reptation attempts will be successful. Each Monte Carlo run reported consists of 500 samples. Since on the order of N calculations are required to calculate a difference in electrostatic energy, the computer time per sample required by the present algorithm is proportional to N^3.

Early attempts to use this algorithm for $pH - pK_a$ controlled ionization were unsuccessful because, at the beginning of a run, the model polymer often stuck for many samples in metastable states in which both ends of the polymer were highly charged, while the middle was uncharged, before finally making a transition to a more probable state. This problem was avoided by using a "quasithermal" method to generate the starting configuration. That is, the starting configuration was made by adding one unit at a time, starting with no units; the probability of adding a unit in a particular state being proportional to $\exp(-\Delta A/k_BT)$ incurred when adding it, ΔA having been defined above. While this does not give a thermalized initial sample, it comes much closer to it than, say, letting each unit in the starting configuration have a probability of 1/6 of being in each of the six possible states. There are doubtless many other ways to choose the starting configuration so as not to get stuck in a metastable configuration.

Theoretically, reptation has a drawback in that configurations, in which both ends are trapped in cul de sacs formed by other units, are not connected to the other states and so the probabilities of being in them do not equilibrate.[17] This nonergodicity should not be a serious problem in the present work. For a random coil, the probability of both ends being trapped is quite small[17] and should be further reduced when the coil is expanded by electrostatic repulsion, reducing the average number of units per unit volume in the vicinity of the ends. Also, if the hard sphere repulsion is turned off, as in the present work, there are no absolute culs de sac.

The computer used was an IBM 3081KX; the random number generator was IMSL's routine "GGUBS" with starting seed 123457. Some very small deviations from randomness were found for this random number generator by trying various starting seeds; it is believed that the errors caused are comparable to, or smaller than, those caused by the finite sample size of 500.

B. Free energy and entropy of a population of chains

The free energy F and entropy S per chain of a population of chains can be found from the Monte Carlo results for $\overline{U}$ and $\overline{\alpha}$ as a function of $pH - pK_a$ by thermodynamic integration of $\overline{\alpha}$ with respect to $pH - pK_a$. This is analogous to Ref. 18's procedure for finding F in Ising spin systems by thermodynamic integration of magnetization as a function of magnetic field.

For the present model of a polyelectrolyte, S and F are known exactly when $pH \ll pK_a$ for then $\overline{\alpha} \to 0$, and all uncharged states are equally probable. So in this threefold rotational isomeric state model, S is

$$S = S_0 \equiv k_B(N-3)\ln(3) \quad \text{when } pH \ll pK_a, \tag{10}$$

since the positions of the first three units can be specified, making the first three bond states irrelevant. When $\overline{\alpha} \to 0$, $\overline{U} \to 0$ also because so few units are charged. Also, as shown in Appendix A, when $\overline{\alpha} \to 0$, the relation between $\overline{\alpha}$ and $\overline{U}$ is the same as it would be for an isolated unit, so that

$$\overline{\alpha} \to 10^{pH-pK_a}/(1 + 10^{pH-pK_a}), \tag{11}$$

and clearly $\overline{\alpha}(pH - pK_a) \to 0$ as $pH - pK_a \to -\infty$. Putting this and $\overline{U} \to 0$ into Eq. (5) gives

$$F = F_0 \equiv -TS_0, \tag{12}$$

when $pH \ll pK_a$.

Then F can be found at values of pH, where $\overline{\alpha} \neq 0$ by numerical thermodynamic integration[18] over $pH - pK_a$ of the Monte Carlo results. To do this, we consider F to be a function of T, X, and P. Then, from Eq. (7),

$$\left.\frac{\partial F(T,X)}{\partial X}\right|_{T,P} = -\overline{\alpha}. \tag{13}$$

Integrating both sides of Eq. (13) with respect to X and using Eqs. (12) and (6) gives

$$F = -(N-3)k_BT\ln(3) - Nk_BT\ln(10) \times \int_{-\infty}^{pH-pK_a} \overline{\alpha}\, d(pH - pK_a). \tag{14}$$

S is now found from F via Eq. (7) using the known values of $\overline{\alpha}$ and $\overline{U}$.

III. RESULTS

All the results are reported in terms of their dependence on the dimensionless parameters N, θ, ξ_0, $pH - pK_a$, and

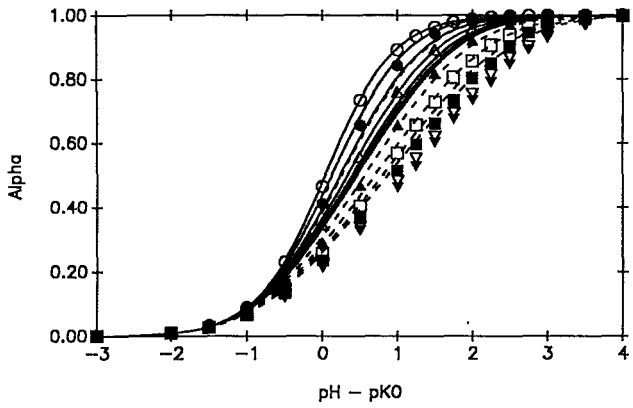

FIG. 1. Average fractional ionization $\overline{\alpha}$ vs $pH - pK_0$ for a 50 unit chain with Manning parameter 1, bond angle $\theta = 70°$, and various values of κ^{-1}/D (0.5, 1, 2, 4, 8, 16, 32, and 64 from top to bottom). The symbols are open circles, filled circles, open triangles (point up), filled triangles, open squares, filled squares, open inverse triangles, and filled inverse triangles for these eight κ^{-1}/D values, respectively. Also shown are two predicted sets of $\overline{\alpha}$ vs $pH - pK_0$ curves; a third-order linear Ising model type calculation (solid lines) and a mean field calculation in α [Eq. (24)] appropriate for a chain frozen in the all-*trans* position (dashed lines).

κ^{-1}/D, D/l_B being fixed. Hard-sphere diameter was zero for all the results reported here.

In terms of how the present Monte Carlo program works, it is natural to report results in terms of $\overline{\alpha}$ vs $pH - pK_a$ as in Figs. 1–5, although results are more commonly reported as pK or ΔpK vs $\overline{\alpha}$, where

$$pK \equiv pH - \log_{10}[\overline{\alpha}/(1-\overline{\alpha})] \tag{15a}$$

and

$$\Delta pK \equiv pK - pK_0, \tag{15b}$$

where pK_0 is defined as the limit of pK as $\overline{\alpha} \to 0$, i.e., at low pH. Equation (11) gives the form for $\overline{\alpha}$ which holds as $\overline{\alpha} \to 0$. Inserting Eq. (11) into Eq. (15a) yields

$$pK_0 = pK_a \tag{15c}$$

in the present model. In what follows, we will usually refer to pK_0 rather than pK_a.

Monte Carlo results were generated for several values of

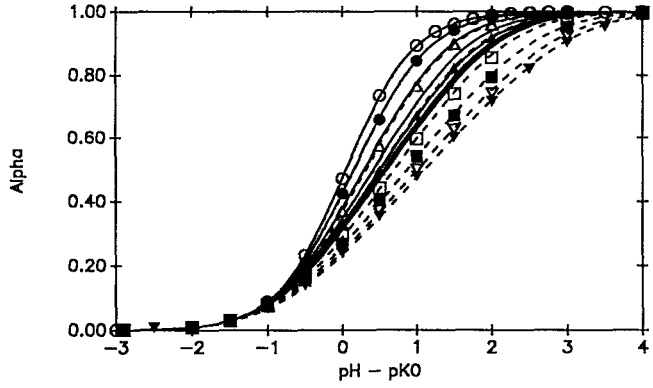

FIG. 2. The same as Fig. 1 except that $\theta = 1°$. The values of κ^{-1}/D are 0.5, 1, 2, 4, 8, 16, 32, and 64. Symbols for these, and line types, are the same as in Fig. 1.

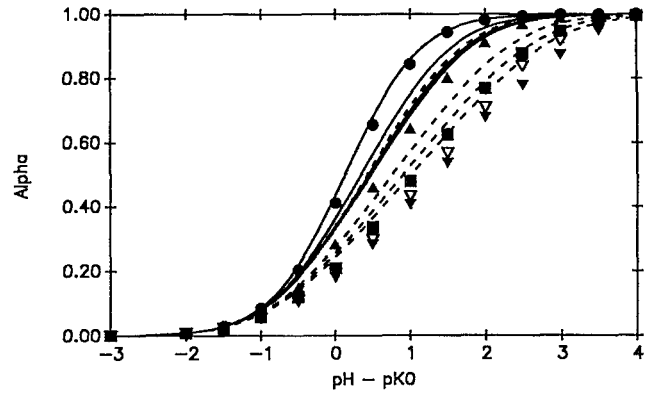

FIG. 3. The same as Fig. 1 except that $N = 100$. The values of κ^{-1}/D are 1, 4, 16, 32, and 64; the symbols and line types are the same as in Fig. 1.

Debye screening length at five combinations of maximum Manning parameter ξ_0, bond angle θ, and chain length N. These are shown in Figs. 1–5. Also shown in Figs. 1–5 are two calculations of $\overline{\alpha}$—a third-order linear Ising calculation and a mean field calculation in α which assumes the polymer to be in the all-*trans* configuration. For $\theta = 1°$, the polymers are quite rodlike, as is the *trans* configuration; thus for $\theta = 1°$, the distances between polymers are given correctly in this calculation. These calculations are discussed further below.

In Figs. 1–4, ξ_0 according to Eq. (4) was 1, so as to ensure that ξ remained below the Manning limit. Figures 1 and 2 show results for $\overline{\alpha}$ as a function of $pH - pK_0$ for $N = 50$ and $\theta = 70°$ and 1°, respectively, for $\kappa^{-1}/D = 0.5$, 1, 2, 4, 8, 16, 32, and 64. Note that at $\theta = 1°$, for $N \leqslant 100$, the simulated polymers are quite rodlike and there are essentially no electrostatic excluded volume effects. On the other hand, the polymers are very flexible at $\theta = 70°$, which is very close to $\theta = \arccos(1/3) \cong 70.529$, the bond angle at which the three-fold rotational isomeric state model becomes the tetrahedral lattice model. Figure 3 shows results for $N = 100$, $\theta = 70°$, and κ^{-1}/D values of 1, 4, 16, 32, and 64. Figure 4 shows results for $N = 100$, $\theta = 1°$, and κ^{-1}/D values of 0.5, 1, 2, 4, 16, and 64.

Figure 5 shows results for $\xi_0 = 2.85$, $\theta = 70°$, and

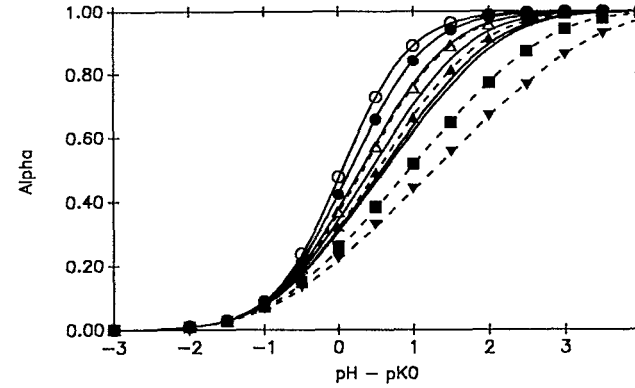

FIG. 4. The same as Fig. 1 except that $N = 100$ and $\theta = 1°$ for $\kappa^{-1}/D = 0.5$, 1, 2, 4, 16, and 64 with the same meanings for the symbols and line types.

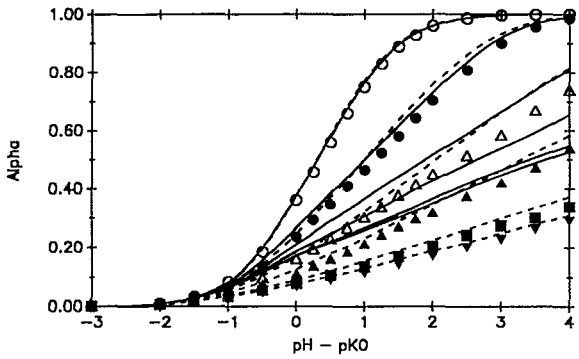

FIG. 5. The same as Fig. 1 except that $\xi_0 = 2.85$ for $\kappa^{-1}/D = 0.5$, 1, 2, 4, 16, and 64 with the same meanings for the symbols and line types.

$N = 50$ with κ^{-1}/D values of 0.5, 1, 2, 4, 16, and 64. The much higher ξ_0 value of 2.85 is the Manning parameter resulting if the backbone of the polymer is a single-bonded carbon chain and a charged group is attached to every other carbon; it was chosen for comparability to the model used in Refs. 11 and 12. As mentioned above, it is probably acceptable to consider values of $\xi_0 > 1$, if $\xi = \overline{\alpha}\xi_0$ remains less than 1. Thus $\overline{\alpha}$ values greater than 0.351 in Fig. 5 should be discounted.

For the $\theta = 1°$ cases, the mean field theory in α assuming an all-*trans* configuration works well; for $\theta = 70°$, it does not, but is better at the very high value of $\xi_0 = 2.85$ than for $\xi_0 = 1$. The third-order Ising prediction works well for small κ^{-1}/D and also works better at $\theta = 1°$ than at $\theta = 70°$.

Figure 6 shows entropy per chain calculated according to Eqs. (7) and (14) for the four lowest κ^{-1}/D values of Fig. 3.

Figures 7, 8, and 9 show the Monte Carlo results of Figs. 1, 3, and 5, respectively, plotted as ΔpK vs $\overline{\alpha}$. The $\log_{10}[\overline{\alpha}/(1-\overline{\alpha})]$ term in Eq. (15a) causes the large jitter in ΔpK near the limits $\overline{\alpha} = 0$ and $\overline{\alpha} = 1$, where the fractional fluctuation in $\overline{\alpha}$ or $(1-\overline{\alpha})$ is large, due to the very small number of charged or uncharged units, respectively. Also

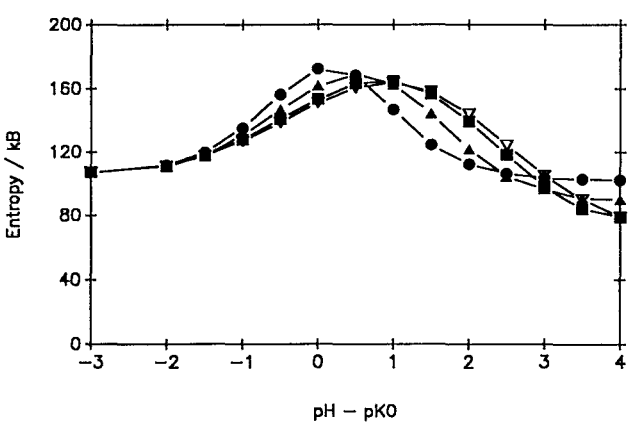

FIG. 6. S vs $pH - pK_0$ as found by thermodynamic integration of the Monte Carlo results of the top four κ^{-1}/D values of Fig. 3. The symbols are the same as in Fig. 1; the lines are guides to the eye.

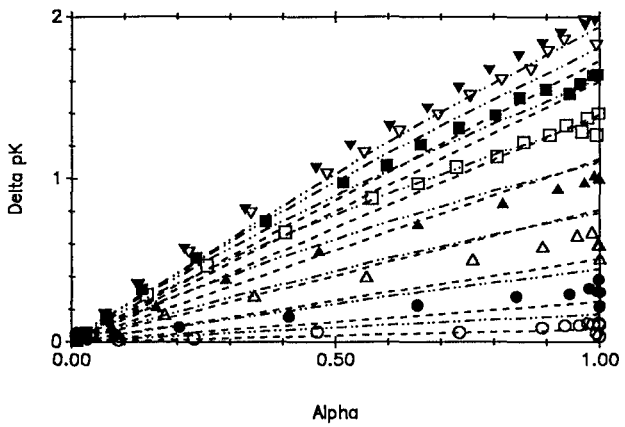

FIG. 7. ΔpK vs $\overline{\alpha}$ for the data of Fig. 1 compared with the predictions of Eqs. (24) (dashed lines) and (38) (dashed and dotted lines). The same symbols as in Fig. 1.

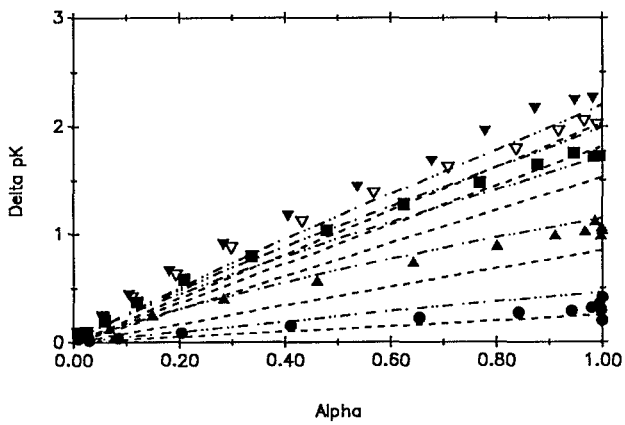

FIG. 8. ΔpK vs $\overline{\alpha}$ for the data of Fig. 3 compared with the predictions of Eqs. (24) (dashed lines) and (38) (dashed and dotted lines). The same symbols as in Fig. 1.

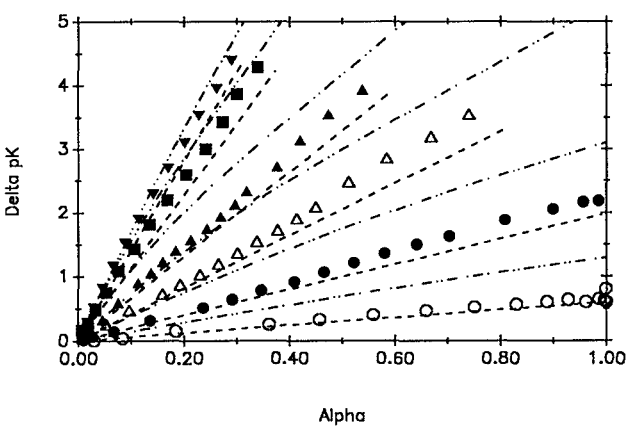

FIG. 9. ΔpK vs $\overline{\alpha}$ for the data of Fig. 5 compared with the predictions of Eqs. (24) (dashed lines) and (38) (dashed and dotted lines). The same symbols as in Fig. 1.

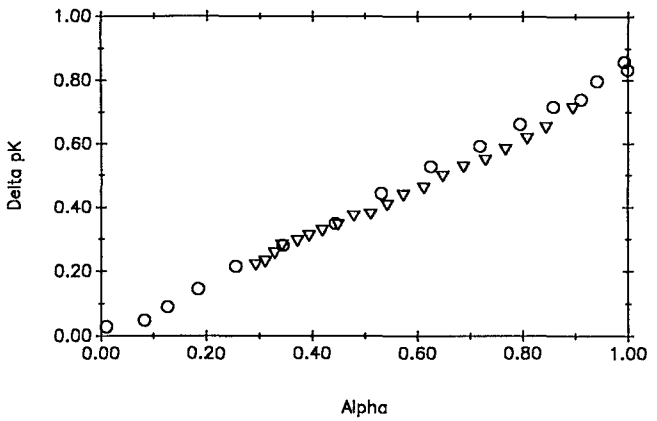

FIG. 10. ΔpK vs $\bar{\alpha}$ of Cleland's data compared to Monte Carlo results (see the text). Symbols are open circles for Monte Carlo and open inverse triangles for Cleland's data.

shown are the results of mean field calculations in α for two assumptions about the distances between polymer units. The first is the same all-*trans* assumption that was used in the dashed curves for Figs. 1–5. The second is an attempt to estimate the distances between units in a charged wormlike chain using the concepts of electrostatic persistence length[19,20] and excluded volume,[21,22] as described below. The second estimation worked better than the first for the $\xi_0 = 1$ cases (Figs. 7 and 8), and less well for $\xi_0 = 2.85$ (Fig. 9).

Figure 10 shows Monte Carlo results with $N = 50$, $\xi_0 = 0.716$, $\theta = 27.34°$, and $\kappa^{-1}/D = 3.07$; these results were chosen as discussed below for comparison to experimental results of Cleland[23] for the potentiometric titration of hyaluronate in dialysis equilibrium with a 10 mM salt solution. Considering that no attempt was made to extrapolate the Monte Carlo results to the very long chain limit, the agreement is surprising.

IV. DISCUSSION

A. Linear Ising model calculations

Linear Ising model type calculations were performed to third nearest neighbors. These methods are well known, so we merely recall the assumptions upon which they are based in order to discuss their strengths and limitations. Appendix B gives the bare formulas necessary to perform these calculations in first through third order; these have been explained by Cleland.[4] The methods behind the formulas were ably expounded by Flory.[3] The nearest-neighbor result for $\bar{\alpha}$ in the limit of infinitely long chains was given by Harris and Rice.[6]

These calculations consider only the electrostatic interactions between nearby units. Thus in an mth nearest-neighbor calculation, one uses the approximate form[4]

$$U \cong \frac{q^2}{4\pi\epsilon} \sum_{i=2}^{N} \sum_{j=i-m}^{i-1} \alpha_i \alpha_j \exp(-\kappa r_{ij})/r_{ij}. \tag{16}$$

Then one uses matrix methods to calculate the function Z,

$$Z \equiv \sum_{\substack{\text{all} \\ \text{states}}} \exp[-U/(k_B T) + N\alpha \ln(x)], \tag{17}$$

where U is the approximate form of Eq. (16), α is the fractional ionization in a particular state, and

$$x \equiv \exp[X/(Nk_B T)] = 10^{pH - pK_0}. \tag{18}$$

Then from Eq. (9), $\bar{\alpha}$ is

$$\bar{\alpha} = \sum_{\substack{\text{all} \\ \text{states}}} \alpha \exp[-U/(k_B T) + N\alpha \ln(x)] \div \sum_{\substack{\text{all} \\ \text{states}}} \exp[-U/(k_B T) + N\alpha \ln(x)] = \frac{x}{NZ} \frac{\partial Z}{\partial x}, \tag{19}$$

without making any nearest-neighbor assumptions. In the mth-nearest-neighbor approximation for U [Eq. (16)], Z and $\partial Z/\partial x$ can be calculated by matrix methods (Appendix B).

This method can explicitly take into account chains having a finite number of units, such as can be generated by a Monte Carlo program. In principle, these methods can be made exact by letting $m \to \infty$. However, this is impractical since the amount of computation required grows exponentially with m (Appendix B). In practice, one can include only short-range effects, i.e., effects with a range $\leqslant mD$ with m small. Also, these calculations cannot include excluded volume effects, i.e., strong interactions which occur between units far apart along the backbone of the chain when the chain bends so as to place them in close physical proximity.

The third-order linear Ising calculation is good for $\kappa^{-1}/D \leqslant 2$ in the $\xi_0 = 1$, $\theta = 1°$ cases (Figs. 2 and 4). For $\xi_0 = 1$ and $\theta = 70°$, it fails between $\kappa^{-1}/D = 1$ and 2 for $N = 50$ (Fig. 1) and between $\kappa^{-1}/D = 1$ and 4 for $N = 100$ (Fig. 3). In Fig. 5, for $\xi_0 = 2.85$, $\theta = 70°$, and $N = 50$, it fails between $\kappa^{-1}/D = 1$ and 2.

These trends make qualitative sense. As κ^{-1}/D increases, the electrostatic interactions become significant in relation to $k_B T$ for units separated by more than m bonds, and the mth order Ising calculation fails. For higher ξ_0, this happens sooner. For higher θ, $\langle \exp(-\kappa r_{ij})/r_{ij} \rangle$ increases for a given $|i-j|$ and the mth-order Ising calculation also fails sooner than at lower θ. Also, other parameters being the same, electrostatic excluded volume effects will be more severe at higher θ.

As κ^{-1}/D is increased, linear Ising type calculations for $\bar{\alpha}$ as a function of $pH - pK_0$ will converge to a limit as soon as $\kappa^{-1}/D \gg m$, while the Monte Carlo results will not converge til $\kappa^{-1}/D \gg \langle R^2 \rangle^{1/2}$, the rms end-to-end length.

In principle, linear Ising type models can be corrected for differing dielectric constants of the polymer and solvent by solving for the electrostatic interactions of charged units in short ($\leqslant m$ units) sections of polymers in all (3^m) configurations. Such corrections would be computationally expensive for large m.

The $\bar{\alpha}$ predicted by the third-order linear Ising Model is always equal to or greater than the Monte Carlo $\bar{\alpha}$. This is

also as expected, because neglecting electrostatic interactions between units separated by more than m bonds (here $m = 3$) leads to an underestimate of U. Hence the energy penalty for ionizing any particular unit in a particular ionization state and spatial configuration is always underestimated if at least one unit more than m bonds away is also charged. Finally, first-order linear Ising calculations (not shown) worked significantly less well than third-order calculations being accurate only for $\kappa^{-1}/D = 0.5$; second-order calculations, also not shown, were an improvement, but were, as expected, always less good than third order.

One notices that the third-order calculation still remarkably underestimates the electrostatic effect on $\overline{\alpha}$ for large κ^{-1}/D. Although one could of course go to still higher m, it is doubtful whether m greater than 6 or 7 would be practical on any computer now in existence, even with the most efficient algorithm.

B. Mean field calculations

1. Mean field calculation in α for all-trans configurations

Previously,[8] for equal N, κ^{-1}/D, θ, and ξ, our Monte Carlo program yielded very similar, although slightly smaller values of $\overline{U}/(Nk_bT)$ for a partially ionized $pH - pK_a$ controlled chain, such as we are discussing here, and for a chain with a smaller q, but $\alpha \equiv 1$. The radii of gyration were also very similar in the two cases. This, and the failure of the linear Ising calculations for even moderately large κ^{-1}/D, suggested trying a mean field theory (a.k.a. "Bragg–Williams approximation").

This kind of mean field calculation is also well known,[1] but because it worked quite well in some cases, is the basis of a more speculative theory described below, and has a short derivation, it will be described in moderate detail. Specifically, it is assumed that if $\overline{\alpha}$ is the average fractional ionization of a chain under given conditions, every unit has a probability $\overline{\alpha}$ to be charged, regardless of the charge states of other units or the spatial arrangement of the chain. Of course, this is an approximation, since if any given unit is charged, units near it will be less likely to be charged.

As each possible way in which $N\alpha$ out of N units are charged is assumed equally probable, $P(\alpha)$ the probability that $N\alpha$ of the units are charged is proportional to $\exp(-F/k_BT)$, with F given by Eq. (5), with α replacing $\overline{\alpha}$, and where S is k_B times the logarithm of the number of ways in which $N\alpha$ items can be chosen from N items. With Stirling's approximation,

$$S = -Nk_B[\alpha \ln(\alpha) + (1-\alpha)\ln(1-\alpha)]. \tag{20}$$

For large N, $P(\alpha)$ should be sharply peaked and $\overline{\alpha}$ will be the value of α that maximizes $P(\alpha)$ by minimizing F. So

$$\left.\frac{\partial F}{\partial \alpha}\right|_{\alpha=\overline{\alpha}} = 0. \tag{21}$$

To do this, one also needs an expression for $\overline{U}$, using the mean field assumption. By the mean field assumption,

$$\langle \alpha_i \alpha_j \rangle = \overline{\alpha}^2. \tag{22}$$

Then from Eq. (8),

$$\overline{U}(\alpha) \cong \frac{(q\alpha)^2}{4\pi\epsilon} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \langle \exp(-\kappa r_{ij})/r_{ij} \rangle \tag{23}$$

in the mean field approximation, where the distributions of the r_{ij} vary with α. Obviously, this is much easier to compute if r_{ij} are fixed, i.e., if the polymer is rigid. Thus if p_t the probability that a given bond is *trans* were 1, the polymer would be fixed in the *trans* position and the quantity in brackets could easily be computed numerically. Define $\overline{U}_t$ to be the resulting estimate of $\overline{U}$ in the *trans* state. Then $\overline{U}_t \leqslant \overline{U}$ because setting $p_t = 1$ means using the maximum value for every r_{ij} in Eq. (23).

Define g_t to be the value of $\overline{U}/Nk_BT$ in the *trans* state when $\alpha = 1$. Then assuming the polymer to be in the *trans* state, carrying out the differentiation (21), with Eqs. (20) and (23) for S and $\overline{U}$, and dividing by Nk_BT gives

$$0 = \ln(x) - 2g_t\overline{\alpha} + \ln[(1-\overline{\alpha})/\overline{\alpha}], \tag{24a}$$

or

$$\Delta pK = 2g_t\overline{\alpha}, x \ln(10) \tag{24b}$$

and

$$\overline{\alpha} = \frac{\exp(-2g_t\overline{\alpha})x}{1 + \exp(-2g_t\overline{\alpha})x}. \tag{25}$$

As $\theta \to 0$, assuming the chain to be in the all-*trans* configuration leads to the correct r_{ij}. So, if the mean field approximation is a good one, Eqs. (24) should work for small θ, while for larger θ, due to the too small value of $\overline{U}_t$, they should overestimate $\overline{\alpha}$ as a function of pH–pK_0 and underestimate ΔpK as a function of $\overline{\alpha}$.

In fact, this mean field calculation in α worked well for both the $\theta = 1°$ (rodlike) cases. For $\theta = 70°$, it worked much less well, always overestimating $\overline{\alpha}$, as it should. It is interesting that of the $\theta = 70°$ cases, the mean field theory in α with U_t as a reference worked best for $\xi_0 = 2.85$, although the interaction energy of nearest neighbors was $U_{nn} = (2.85 \cos\theta/2)^2 k_BT(l_B/D) = 6.50k_BT$. Naively, one might expect this mean field theory to work least well for the highest U_{nn}/k_BT. The reason for the greater success of mean field theory in α with U_t as a reference for $\xi_0 = 2.85$ may be that the polymer model was more highly stretched out by the greater electrostatic repulsion, so that the choice of the all-*trans* state as a reference was a better approximation.

2. Mean field with live worm approximation

To make a better mean field theory for a charged flexible chain, it is desirable to have some idea, implicit or explicit, of the distribution of the variably ionized segments around each other, so as to be able to estimate the electrostatic energy $\overline{U}(\alpha)$. We modified an idea of Cleland,[1] which he called the "frozen worm" approximation, and found it to be moderately successful.

Several approximations were made: (1) Equation (23) was assumed, i.e., that $\overline{U}(\alpha)$ could be estimated from the distribution of the r_{ij} in a uniformly charged chain with charges of size $q\overline{\alpha}$ (and the same N, θ, and κ^{-1}/D). As previously mentioned in Ref. 8, we found some evidence that this is not far from wrong. (2) Following Cleland,[1] it is assumed that the effective $\overline{r}_{ij}$, i.e., r_{ij} such that

$\exp(-\kappa\bar{r}_{ij})/\bar{r}_{ij}$ gives the correct value for $\langle\exp(-\kappa r_{ij})/r_{ij}\rangle$, is the root mean square value of r_{ij}, $\langle r_{ij}^2\rangle^{1/2}$. This will be an overestimate, especially for large $\kappa\langle r_{ij}^2\rangle^{1/2}$. (3) Following Cleland, it is assumed that in a given polymer, $\langle r_{ij}^2\rangle^{1/2}$ depends only on $|i-j|$. This also is generally false (Ref. 24, Chap. 3). Then it is assumed that $\langle r_{ij}^2\rangle^{1/2}$ can be calculated by (4) treating the polymer model as a wormlike chain with an intrinsic and an electrostatic persistence length and (5) multiplying the resulting $\langle r_{ij}^2\rangle^{1/2}$ by an expansion factor α_R estimated for the end-to-end length of wormlike chains with excluded volume due to short-range repulsive forces.[25] Of course, the screened electrostatic repulsion of Eq. (8) is not necessarily short range. Assumption (2) may be the weakest one.

Making assumptions (4) and (5) explicit requires several equations. Together they are very similar to our estimation[2,8] of the radius of gyration of polymer chains produced by the present Monte Carlo program. The major difference is that α_R replaces an expansion factor for α_S due to Gupta and Forsman.[26]

Specifically, for a wormlike chain with contour length L and persistence length L_T, the mean square end-to-end distance is (Ref. 24, Sec. 9c)

$$\langle R^2(L)\rangle = 2LL_T - 2L_T^2[1-\exp(-L/L_T)]. \quad (26)$$

If two points within the chain are a fraction x of the length of the chain apart, the mean-square separation between them is given by Eq. (26), with L replaced by xL. In making a correspondence between the threefold rotational isomeric state model and a wormlike chain,[2,8] we use Eq. (3) for L and $x \equiv |i-j|/(N-1)$ to find the contour length between units i and j.

According to the electrostatic persistence length theories of Odijk,[19] and of Skolnick and Fixman,[20]

$$L_T = L_0 + L_e, \quad (27)$$

where L_0 is an intrinsic persistence length (i.e., in the absence of electrostatic effects) and L_e is an electrostatic persistence length. We set

$$L_0 = \frac{1}{2\cos(\theta/2)}\frac{1+\cos\theta}{1-\cos\theta}D. \quad (28)$$

With Eq. (3) for L, this ensures that the mean-square end-to-end length of Eq. (26) has the correct limit as $L\to\infty$ when the chain is uncharged. This can be checked by noting that for the threefold rotational isomeric state model,[3] as $N\to\infty$,

$$\langle R^2\rangle \cong ND^2\frac{1+\cos\theta}{1-\cos\theta}, \quad (29)$$

while for Eq. (26), as $L\to\infty$,

$$\langle R^2\rangle \cong 2LL_T. \quad (30)$$

According to Odijk,[19] for finite $y \equiv \kappa L$,

$$L_e = \frac{\xi^2 y^2}{12\kappa^2 l_B}[3y^{-2} - 8y^{-3} + \exp(-y) \times (y^{-1} + 5y^{-2} + 8y^{-3})]. \quad (31)$$

Equations (26)–(31) implement assumption (4), with xL replacing L in Eq. (26). To implement assumption (5), following Odijk and Houwaart, consider the Kuhn segment length L_K of a wormlike chain to be $2L_T$. Then the expansion factor α_R is assumed to be given by[1,25]

$$\alpha_R^5 - \alpha_R^3 = zK(xL/L_K). \quad (32)$$

Here, z is the electrostatic excluded volume parameter and K is a complicated function of xL/L_K given by Eqs. (83)–(86) of Ref. 25. The electrostatic excluded volume parameter z is estimated according to[25]

$$z = (3/2\pi l_K^2)^{3/2}\beta n_K^{1/2}, \quad (33)$$

where $n_K \equiv L/L_K$ and β is an estimate[22] of the excluded volume between two segments;

$$\beta \cong (8L_T^2\kappa^{-1})R(w), \quad (34)$$

where in our notation, $w \equiv 2\pi\xi^2 l_B^{-1}\kappa^{-1}$ and $R(w)$ is

$$R(w) \equiv \int_0^{\pi/2} \sin^2\theta \int_0^{w/\sin\theta} (1/x)[1-\exp(-x)]\,dx\,d\theta. \quad (35)$$

The use of the Flory type expansion factor in Eq. (32) is an approximation even[25] for the end-to-end length of wormlike chains. We then use

$$\bar{r}_{ij} \cong \langle r_{ij}^2\rangle^{1/2} \cong \alpha_R\langle R^2(xL)\rangle^{1/2}, \quad (36)$$

with α_R given by Eqs. (32)–(36) with Eqs. (83)–(86) of Ref. 25.

Thus $\bar{U}(\bar{\alpha})$ is estimated from

$$\bar{U}(\bar{\alpha}) \cong \frac{(q\bar{\alpha})^2}{4\pi\epsilon}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\exp(-\kappa\bar{r}_{ij})/\bar{r}_{ij}, \quad (37)$$

using $\bar{r}_{ij}$ of Eq. (36). Finally, as in Eq. (24b), one uses

$$\Delta pK \cong 2x\ln(10)[\bar{U}(\bar{\alpha})/Nk_BT]/\bar{\alpha}. \quad (38)$$

Equation (38) differs from Cleland's[1] frozen worm approximation in that $\bar{U}$ is recalculated for each $\bar{\alpha}$ (hence the designation "live worm") and in the use of an electrostatic persistence length.

Given the rather long series of assumptions made in deriving it, Eq. (38) seems to work moderately well in Figs. 7–9 for the 70° cases. [For the $\theta = 1°$ cases, the predictions of Eqs. (38) and (24) do not differ significantly.] It is a definite improvement in the $\xi_0 = 1$ cases (Figs. 7 and 8), although it overestimates ΔpK for small κ^{-1}/D and underestimates it for large κ^{-1}/D. However, for the $\xi_0 = 2.85$ case in Fig. 9, Eq. (38) is inferior to the much simpler Eq. (24b). Possibly the main value of Eq. (38) is to show that even a long series of assumptions about the distances between units still gives a reasonable result for $\Delta pK(\alpha)$.

By contrast, attempts to construct a mean field theory in both α and p_t, the fraction of *trans* bonds, were quite unsuccessful.

C. Comparison to Cleland's titration of hyaluronate

Although the Monte Carlo program models a generic polyelectrolyte, it is of some interest to see how it compares with actual data for the potentiometric titration of a polyelectrolyte. Such a comparison is shown in Fig. 6. The data, for titration of hyaluronate by Cleland, Wang, and Detweiler,[23] by NaOH in 0.01 M $NaNO_3$, have been discussed and fitted at length by Cleland,[1] making them a natu-

ral choice for comparison. The parameters for the Monte Carlo run of $D = 10$ Å, $\kappa^{-1} = 30.7$ Å, and $\theta = 27.34°$ were chosen to simulate hyaluronate. The value of D is from Cleland.[4] The value of κ^{-1} was chosen to simulate the salt concentration and $\theta = 27.34°$ was chosen so that Eq. (28) with Eq. (3) would yield an intrinsic persistence length of 87 Å, our previous result.[27] Of course, q was one electronic charge.

Given the crudeness of the model and the fact that N was only 50, the agreement between its predictions and the data was better than expected.

Indeed, comparison of Monte Carlo results to data at higher salt concentrations (not shown) of Cleland, Wang, and Detweiler,[23] which Cleland[1] chose not to fit, shows the Monte Carlo results to underestimate ΔpK. This underestimate becomes increasingly severe as the salt concentration increases. We suspect that this is due to the neglect of local effects in the Monte Carlo program. At high salt concentrations, distant charges are screened effectively from each other. However, the polymer's dielectric constant is much less than that of water and salt obviously cannot penetrate the sterically excluded volume of the polymer monomers. This means that the electrostatic repulsion between nearby charges (which is the only remaining electrostatic repulsion at high salt) may be much greater than estimated by the program, leading to an underestimate of ΔpK.

D. Additional comments and speculations

There may be additional problems with mean field calculations on real polymers in water solution due to their low dielectric constants in relation to water.

Linear Ising type calculations are well adapted to including local effects, while mean field calculations are better at including large scale effects (although they are not necessarily good at including very large scale effects such as excluded volume). The two types of calculation can be combined[28] analogously to what, in solid state physics,[29] is called a Bethe–Peierls or quasichemical approximation. In this approximation, interactions between polymer units less than m bonds apart are treated in an mth-order linear Ising model type calculation like manner, while interactions between more distant units are treated in a mean field manner. This seems premature for the present model, for which the simple mean field theory worked well for the rigid ($\theta = 1°$) cases, while in the flexible ($\theta = 70°$) cases, it may be more urgent to find an improved treatment of chain flexibility and excluded volume effects.

However, in real polymers, local effects can be much more important than in the present model. These may be electrostatic effects due to a much lower polymer than solvent dielectric constant and to the inability of ions to penetrate the polymer. There may be other local effects, such as changes in the pK_0 of chemical groups, due to the nature of nearby chemical groups in copolymers.[7] Quasichemical type calculations could also perhaps be used to investigate the fact that, at the same $\bar{\alpha}$, $pH - pK_a$ calculations give[8] slightly lower energies than uniformly charged calculations.

One possible nonlocal effect would be a dependence of pK_a on κ^{-1}, e.g., due to electrostatic interactions with counterions.[30] Such a dependence could be inserted into the Monte Carlo algorithm and the theoretical predictions for $\bar{\alpha}$ "by hand."

Both the use of $pH - pK_a$ in a Monte Carlo algorithm and the numerical thermodynamic integration of $\bar{\alpha}$ with respect to pH–pK_a to find the free energy and entropy per chain of a population of chains could be applied straightforwardly to Monte Carlo simulations which use explicit counterions.

V. SUMMARY

Mean field theory in α worked amazingly well for the $\theta = 1°$ cases, even though the nearest-neighbor interaction energy was equal to $k_B T$ since ξ_0 was 1. It also worked moderately well when θ was 70° and ξ_0 was 1, using an expression for the typical distances between units based on the ideas of electrostatic excluded persistence and electrostatic excluded volume. When ξ_0 was 2.85 and θ was 70°, assuming that all the bonds were in the *trans* position worked better. When θ was 1°, the two expressions were indistinguishable. Even third-order Linear Ising type models did not work well for κ^{-1}/D larger than about 2, for $\theta = 1°$, or 1, for $\theta = 70°$.

ACKNOWLEDGMENTS

We acknowledge financial assistance from NSF Grant DMB #8803760. We also thank the Tulane Computing Center for its patience and support.

APPENDIX A

The purpose of this appendix is, for the present polymer model, to justify Eqs. (5), (9), and (11). These conclusions are not new; they are implicit in the work of Harris and Rice[6] and of Tanford and Kirkwood[30] for somewhat more complicated model polyelectrolytes.

The polymer has N ionizable sites, so it has 2^N possible charge states. It also has 3^{N-3} possible spatial configurations. A given configuration J is specified by giving both its spatial configuration and its charge states, so there are $8 \times 6^{N-3}$ possible states. In the Debye–Hückel approximation, the electrostatic energy of each such state J is fixed; call it U_J, or sometimes just U. Denote the fractional ionization in state J by $\alpha(J)$ (to distinguish it from α_i, the ionization state of a single site), or sometimes just α.

The polymer (polyacid) is considered to be in dilute solution at a given salt concentration, temperature, and pressure. Let n_J be the number of moles of the polymer that are in state J and let μ_J be the chemical potential of the molecules that are in state J. Then the total free energy of the polymer is

$$F = \sum_J n_J \mu_J, \tag{A1}$$

where the sum is over all configurations J. From the theory of solution thermodynamics,[31,32]

$$\mu_J = \mu_J^{(0)} + RT \ln(a_J) \cong \mu_J^{(0)} + RT \ln([J]), \tag{A2}$$

where a_J is the activity of polymer configuration, J, $[J]$ is its weight concentration, $\mu_J^{(0)}$ is its chemical potential at some reference set of concentrations, and we have used the fact

that when a substance becomes very dilute, its activity approaches its concentration.[31,32]

The idea is to approximately treat the polymer molecules as if they were the entire system, and the solvent, salt, and H^+ ions were "averaged out" leaving only effective interactions among and between the polymer molecules. Assuming the Debye–Hückel approximation already goes a long way in this direction. However, one must take into account the chemical reaction in which individual polymer units are ionized. Represent this reaction as

$$\mathscr{R} \rightarrow \mathscr{R}^- + H^+, \tag{A3}$$

$\mathscr{R}$ denoting uncharged polymer units.

Under reference conditions in which the n_J and the salt concentration are fixed, but the concentration of H^+ ions is not fixed, the change in chemical potential when one mole of $\mathscr{R}$ undergoes this reaction is $\mu^{(0)}_{\mathscr{R}^-} + \mu_{H^+} - \mu^{(0)}_{\mathscr{R}}$. This should be included in μ_J in order to fix Eqs. (A1)–(A2). It was assumed that the only interaction between polymer units is their electrostatic repulsion. Then the change in chemical potential in going between two polymer configurations under standard conditions is determined only by the number of ionizations that must take place and by the change in electrostatic energy. Specifically, denoting by O some particular polymer configuration in which no units are ionized, we make the approximation that

$$\mu^{(0)}_J - \mu^{(0)}_O \cong N_A \overline{U}(J) + N\alpha(J)\left[\mu^{(0)}_{\mathscr{R}^-} + \mu_{H^+} - \mu^{(0)}_{\mathscr{R}}\right], \tag{A4}$$

where N_A is Avogadro's number, since $N\alpha(J)$ units must be ionized in going from state O to state J. In Eq. (A1), the independent chemical species may be considered to be a polymer molecule in state J plus $N\alpha(J)$ hydrogen ions.

1. Derivation of Eq. (5)

Consider the quantity in square brackets in Eq. (A4). In it,

$$\mu_{H^+} = \mu^{(0)}_{H^+} + RT\ln(a_{H^+}) \tag{A5}$$

from solution thermodynamics. Also[31]

$$\ln(a_{H^+}) \cong -\ln(10)pH. \tag{A6}$$

Recall some more elementary solution thermodynamics.[32] Let K_a be the equilibrium association constant for reaction (A3). That is,

$$K_a \equiv \frac{a_{\mathscr{R}^-} a_{H^+}}{a_{\mathscr{R}}} \cong \frac{[\mathscr{R}^-]a_{H^+}}{[\mathscr{R}]}, \tag{A7}$$

where it is again assumed that the concentration of polymer units is very low. So $a_{\mathscr{R}} \cong [\mathscr{R}]$ and $a_{\mathscr{R}^-} \cong [\mathscr{R}]$. At equilibrium for reaction (A3), the chemical potential of the left-hand side $\mu_{\mathscr{R}}$ equals that of the right-hand side $\mu_{\mathscr{R}^-} + \mu_{H^+}$. Expressing these chemical potentials in terms of activities, as in Eq. (A2), and again assuming that the concentration of polymer units is small, as in Eq. (A7), yields

$$\ln(K_a) = (1/RT)(\mu^{(0)}_{\mathscr{R}} - \mu^{(0)}_{\mathscr{R}^-} - \mu_{H^+}). \tag{A8}$$

Combining Eqs. (A6)–(A8) and making the usual definition that $pK_a \equiv -\log_{10}(K_a)$,

$$(\mu^{(0)}_{\mathscr{R}^-} + \mu_{H^+} - \mu^{(0)}_{\mathscr{R}}) = \ln(10) \times RT(pK_a - pH). \tag{A9}$$

Combining Eqs. (A2), (A4), and (A9) gives

$$\mu_J = \mu^{(0)}_O + N_A U(J) + \ln(10)N\alpha(J)RT(pK_a - pH) + RT\ln([J]). \tag{A10}$$

Recall that the configurational entropy S of the polymer is given by, within a constant,

$$S = -R\sum_J n_J \ln([J]), \tag{A11}$$

as can be found from Stirling's approximation, assuming the total number of polymer molecules in the dilute solution is large. Inserting Eq. (A10) into Eq. (A1) and using Eq. (A11) gives, except for a constant that does not depend on the distribution of the polymer molecules among the states J,

$$F = \sum_J n_J [N_A U(J) + \ln(10)N\alpha(J)RT(pK_a - pH)] - TS. \tag{A12}$$

Dividing Eq. (A12) by the total number of polymer molecules, which is N_A times the sum over J of n_J, gives Eq. (5).

2. Derivation of Eq. (9)

At equilibrium, we must have $\mu_J = \mu_O$. Using Eq. (A10) for both μ_J and μ_O, recalling that by definition $\alpha(O) = U(O) = 0$, and solving for $[J]/[O]$ gives

$$[J]/[O] = \exp\{-[U(J)/k_BT] + \ln(10) \times N\alpha(J)(pH - pK_a)\}, \tag{A13}$$

which is Eq. (9) after using $X \equiv \ln(10)k_BTN(pH - pK_a)$.

3. Justification of Eq. (11)

We want to find the relationship between $\overline{\alpha}$ and pK_a when $\overline{\alpha}$ is very small. When $\overline{\alpha}$ is sufficiently small, it is unlikely that, if a given polymer unit is ionized, any other unit is ionized. That means that if $\overline{\alpha} \rightarrow 0$, so does $\overline{U}$, and thus the relative probabilities of any two states are given by Eq. (A13), or Eq. (9) with U set to zero.

When U is zero, the different ionizable sites on the model polymer do not interact and the relationship between between $\overline{\alpha}$ and pK_a must be the same as for isolated units. For an isolated unit, Eq. (A13) with $U = 0$ and $N = 1$ shows that the relative probabilities that it is or is not ionized are $10^{pH - pK_a}$ and 1, respectively. In this case, $\overline{\alpha}$ is the chance that this single unit is ionized, which is

$$\overline{\alpha} = 10^{pH - pK_a}/(1 + 10^{pH - pK_a}). \tag{A14}$$

This is enough to justify Eq. (11). It can of course also be derived directly from Eq. (A13) with $U = 0$ by defining

$$Z \equiv \sum_J \exp[\ln(10)N\alpha(J)(pH - pK_a)] = \sum_J x^{N\alpha(J)} = \sum_{(N\alpha)=0}^{N} x^{(N\alpha)}\binom{N}{N\alpha} = (1 + x)^N, \tag{A15}$$

where $x \equiv 10^{pH - pK_a}$ and then noticing that $\overline{\alpha} = (x/Nz)(\partial Z/\partial x)$.

APPENDIX B

We give the equations needed to use the linear Ising model in the first- through third-nearest-neighbor approximations to find $\bar{\alpha}$ as a function of $pH - pK_0$ for the threefold rotational isomeric state model defined in the text.

Equation (19) gives $\bar{\alpha}$ if one can find Z and $\partial Z/\partial x$. Recall that $\ln(x) = (pH - pK_0)\ln(10)$. In all orders of the linear Ising model, Z and $\partial Z/\partial x$ are given by equations having the forms

$$Z = \mathbf{J}_n^* \mathbf{U}^N \mathbf{J}_n \tag{B1}$$

and

$$\frac{\partial Z}{\partial x} = (\mathbf{J}_{2n}^*)\begin{pmatrix} \mathbf{U} & \partial \mathbf{U}/\partial x \\ \mathbf{0}_{n\times n} & \mathbf{U} \end{pmatrix}^N \begin{pmatrix} \mathbf{O}_{1\times n} \\ \mathbf{J}_n \end{pmatrix}. \tag{B2}$$

Here the statistical weight matrix[3] $\mathbf{U}$ of size $n\times n$ depends on the order m of the linear Ising approximation, as does n itself. In these equations, for $m = 1$, 2, and 3, $n(m) = 2$, 4, and 8 respectively. $\mathbf{0}_{a\times b}$ is a matrix of order $a\times b$, whose elements are all 0. $\mathbf{J}_n$ is a column vector of length n whose first (top) element is 1 and whose other elements are 0. $\mathbf{J}_n^*$ is a row vector with n elements, all of whose elements are 1.

It remains to specify the elements of $\mathbf{U}$ for $m = 1$, 2, and 3. First, we define some subsidiary quantities. The distance between nearest neighbors is D. Define y to be the Boltzmann factor for the Debye–Hückel electrostatic energy for a pair of nearest neighbors, both of which are charged. That is,

$$y \equiv \exp[-q^2\exp(-\kappa D)/(4\pi\epsilon k_B T D)]. \tag{B3}$$

Similarly, in the present model, the distance between next nearest neighbors is fixed. Call this distance $D_{\rm nn}$. Then

$$D_{\rm nn} = 2\cos(\theta/2)D, \tag{B4}$$

as found from the comments before Eq. (1). Analogously to Eq. (A3), define

$$w \equiv \exp[-q^2\exp(-\kappa D_{\rm nn})/(4\pi\epsilon k_B T D_{\rm nn})], \tag{B5}$$

so that w is the Boltzmann factor for a pair of charged next-nearest neighbors.

Next-next-nearest neighbors (third nearest neighbors) can, in the present model, be two distances apart, according to whether that section of the chain is in the *trans* configuration or in a *gauche* configuration. Call those two distances D_t and D_g, respectively. Then, from Eq. (1),

$$D_t = D(5 + 4\cos\theta)^{1/2} \tag{B6}$$

and

$$D_g = D(2 + 4\cos\theta + 3\cos^2\theta)^{1/2}. \tag{B7}$$

Again, let the Boltzmann factors for a pair of third-nearest neighbors in the *trans* and *gauche* configurations be, respectively,

$$t \equiv \exp[-q^2\exp(-\kappa D_t)/(4\pi\epsilon k_B T D_t)] \tag{B8}$$

and

$$g \equiv \exp[-q^2\exp(-\kappa D_g)/(4\pi\epsilon k_B T D_g)]. \tag{B9}$$

Define v to be a weighted average of t and g,

$$v \equiv (t + 2g)/3. \tag{B10}$$

Then the matrix $\mathbf{U}$ for the nearest neighbor ($m = 1$) approximation is

$$\mathbf{U} = \begin{pmatrix} 1 & x \\ 1 & xy \end{pmatrix}. \tag{B11}$$

For the next-next-nearest-neighbor ($m = 2$) approximation, $\mathbf{U}$ is

$$\mathbf{U} = \begin{pmatrix} 1 & x & 0 & 0 \\ 0 & 0 & 1 & xy \\ 1 & xw & 0 & 0 \\ 0 & 0 & 1 & xyw \end{pmatrix}. \tag{B12}$$

For the third-nearest-neighbor approximation, $\mathbf{U}$ is

$$\mathbf{U} = 3\begin{pmatrix} 1 & x & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & xy & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & xw & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & xyw \\ 1 & xv & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & xyv & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & xwv & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & xywv \end{pmatrix}. \tag{B13}$$

It may be appropriate to comment on computing the matrix equations (B1) and (B2). As pointed out by Flory,[3] this can be done in several ways. Repeatedly squaring the matrix $\mathbf{U}$ yields the second, fourth, ...2^p powers of $\mathbf{U}$, where 2^p is the largest power of $2 \leqslant N$. Writing N in binary notation so that it is a sum of powers of 2 and multiplying the corresponding powers of $\mathbf{U}$ yields $\mathbf{U}^N$, which can then be multiplied by the vectors $\mathbf{J}$ and $\mathbf{J}^*$ to obtain Z; $\partial Z/\partial x$ can be obtained by the same maneuver. This method can treat large values of N easily; the number of multiplications increases as $\ln(N)$. Its disadvantage is that the computer time required to perform each multiplication increases very rapidly with the order m of the calculation.

In Eqs. (B11)–(B13), the order $n(m)$ of the matrix $\mathbf{U}$ doubles as one increases m, the order of the linear Ising approximation, by 1. In general, however, in the absence of clever simplifications, $n(m)$ increases sixfold when m increases by 1. The time required to multiply two $n\times n$ matrices together is $\mathcal{O}(n^3)$, so the time required to multiply matrices of order $n(m)$ and thus the time required to evaluate Z and $\partial Z/\partial x$ increases $6^3 = 216$ times when m increases by 1 (for $m > \cong 3$). Thus the first method quickly becomes impractical, even on the fastest computers.

A second method (for finding Z, $\partial Z/\partial x$ is of course similar) is simply to multiply the vector $\mathbf{J}$ N times on the left by $\mathbf{U}$. This method is no longer suitable for very large values of N (CPU time is of course $\propto N$), but because the time required to multiply a vector of length n by an $n\times n$ matrix is only $\mathcal{O}(n^2)$, it quickly becomes more efficient as m increases. (The relative advantage increases 12-fold rather than sixfold when m increases by 1, because the proportion of nonzero elements in $\mathbf{U}$ drops in half when m increases by 1.)

Of course, if one is interested in the limit $\mathbf{N}\to\infty$, the appropriate approach is to find the largest eigenvalues of $\mathbf{U}$

and of the matrix constructed from it which plays the same role in finding $\partial Z/\partial x$ as **U** in finding Z.[3]

[1] R. L. Cleland, Macromolecules **17**, 634 (1984).
[2] C. E. Reed and W. F. Reed, J. Chem. Phys. **92**, 6916 (1990).
[3] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969).
[4] R. L. Cleland, Biopolymers **10**, 1925 (1971).
[5] G. S. Manning, J. Chem. Phys. **51**, 924 (1969); **51**, 3249 (1969); M. O. Fenley, G. S. Manning, and W. K. Olson, Biopolymers **30**, 1191 (1990).
[6] F. E. Harris and S. A. Rice, J. Phys. Chem. **58**, 725, 733 (1954).
[7] N. D. Truong, G. Medjhadi, D. Sarazin, and J. François, Polym. Bull. **24**, 101 (1990).
[8] C. E. Reed and W. F. Reed, J. Chem. Phys. **94**, 8479 (1991).
[9] G. A. Christos and S. L. Carnie, J. Chem. Phys. **91**, 439 (1989).
[10] G. A. Christos and S. L. Carnie, J. Chem. Phys. **92**, 7661 (1990).
[11] H. H. Hooper, H. W. Blanch, and J. M. Prausnitz, Macromolecules **23**, 4820 (1990).
[12] H. H. Hooper, S. Beltran, A. P. Sassi, H. W. Blanch, and J. M. Prausnitz, J. Chem. Phys. **93**, 2715 (1991).
[13] C. Brender, J. Chem. Phys. **92**, 4468; **93**, 2736 (1990); **94**, 3213 (1991).
[14] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Methuen, London, 1967).
[15] K. Binder and D. W. Heerman, *Monte Carlo Simulation in Statistical Physics* (Springer, Berlin, 1988).
[16] J. G. Kemeny and J. L. Snell, *Finite Markov Chains* (Van Nostrand, Princeton, 1960).
[17] F. T. Wall and F. Mandel, J. Chem. Phys. **63**, 4592 (1975).
[18] K. Binder and D. Stauffer, in *Applications of the Monte Carlo Method in Statistical Physics*, edited by K. Binder (Springer, New York, 1987), Chap. 1.
[19] T. Odijk, J. Polym. Sci. Polym. Phys. Ed. **15**, 477 (1977).
[20] J. Skolnick and M. Fixman, Macromolecules **10**, 944 (1977).
[21] T. Odijk and A. C. Houwaart, J. Polym. Sci. Polym. Phys. Ed. **16**, 627 (1978).
[22] M. Fixman and J. Skolnick, Macromolecules **11**, 863 (1978).
[23] R. L. Cleland, J. L. Wang, and D. M. Detweiler, Macromolecules **15**, 386 (1982).
[24] H. Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971).
[25] H. Yamakawa and W. H. Stockmayer, J. Chem. Phys. **57**, 2843 (1972).
[26] S. K. Gupta and W. C. Forsman, Macromolecules **5**, 779 (1972).
[27] S. Ghosh, X. Li, C. E. Reed, and W. F. Reed, Biopolymers **30**, 1101 (1990).
[28] A. Minakata, K. Matsuma, S. Sasaki, and H. Ohnuma, Macromolecules **13**, 1549 (1980).
[29] K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), Chap. 16.
[30] C. Tanford and J. G. Kirkwood, J. Am. Chem. Soc. **79**, 5333 (1957).
[31] C. Tanford, *Physical Chemistry of Macromolecules* (Wiley, New York, 1961), Chap. 4, and Chap. 9, Sec. 30.
[32] K. E. van Holde, *Physical Biochemistry*, 2nd ed. (Prentice-Hall, Englewood Cliffs, N.J., 1971), Chaps. 1–3.