Protein Science (1995), *4*:2019–2031. Cambridge University Press. Printed in the USA.

Prediction of polyelectrolyte polypeptide structures using Monte Carlo conformational search methods with implicit solvation modeling

JOHN SPENCER EVANS,[1,3] SUNNEY I. CHAN,[1] AND WILLIAM A. GODDARD III[2]

[1] Arthur Amos Noyes Laboratory for Chemical Physics, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

[2] Materials and Molecular Simulation Center, Beckman Institute, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

(RECEIVED April 17, 1995; ACCEPTED July 11, 1995)

Abstract

Many interesting proteins possess defined sequence stretches containing negatively charged amino acids. At present, experimental methods (X-ray crystallography, NMR) have failed to provide structural data for many of these sequence domains. We have applied the dihedral probability grid–Monte Carlo (DPG-MC) conformational search algorithm to a series of N- and C-capped polyelectrolyte peptides, $(Glu)_{20}$, $(Asp)_{20}$, $(PSer)_{20}$, and $(PSer\text{-}Asp)_{10}$, that represent polyanionic regions in a number of important proteins, such as parathymosin, calsequestrin, the sodium channel protein, and the acidic biomineralization proteins. The atomic charges were estimated from charge equilibration and the valence and van der Waals parameters are from DREIDING. Solvation of the carboxylate and phosphate groups was treated using sodium counterions for each charged side chain (one Na^+ for COO^-; two Na for $CO(PO_3)^{-2}$) plus a distance-dependent (shielded) dielectric constant, $\epsilon = \epsilon_0 R$, to simulate solvent water. The structures of these polyelectrolyte polypeptides were obtained by the DPG-MC conformational search with $\epsilon_0 = 10$, followed by calculation of solvation energies for the lowest energy conformers using the protein dipole-Langevin dipole method of Warshel.

These calculations predict a correlation between amino acid sequence and global folded conformational minima:

1. Poly-L-Glu_{20}, our structural benchmark, exhibited a preference for right-handed α-helix (47% helicity), which approximates experimental observations of 55–60% helicity in solution.

2. For Asp- and PSer-containing sequences, all conformers exhibited a low preference for right-handed α-helix formation ($\leq$10%), but a significant percentage ($\approx$20% or greater) of β-strand and β-turn dihedrals were found in all three sequence cases: (1) Asp_n forms supercoil conformers, with a 2:1:1 ratio of β-turn:β-strand:α-helix dihedral angles; (2) $PSer_{20}$ features a nearly 1:1 ratio of β-turn:β-sheet dihedral preferences, with very little preference for α-helical structure, and possesses short regions of strand and turn combinations that give rise to a collapsed bend or hairpin structure; (3) $(PSer\text{-}Asp)_{10}$ features a 3:2:1 ratio of β-sheet:β-turn:α-helix and gives rise to a superturn or C-shaped structure.

Keywords: computational chemistry; counterion condensation; dihedral angles; internal coordinate method; Monte Carlo; peptide conformation; polyanionic sequences; protein conformation; protein folding

Reprint requests to: William A. Goddard III, Materials and Molecular Simulation Center, Beckman Institute (139-74), California Institute of Technology, Pasadena, California 91125; e-mail: wag@wag.caltech.edu.

[3] Present address: Department of Chemistry, New York University, New York, New York 10010.

Abbreviations: DPG-MC, dihedral probability grid–Monte Carlo method; PSer, *o*-phospho-L-serine; Q, net partial atomic charge; Q_{Na}, sodium partial atomic charge; vdW, van der Waals; PDB, Brookhaven Protein Data Bank; PDLD, protein dipole-Langevin dipole solvent electrostatics; $\Delta\Delta G_p^{sol}$, free energy of solvation; ΔG_{Born}, residual bulk energy of the solvent; ΔG_{Lang}, free energy of water dipole interactions with solute; $-T\Delta S$, entropy contribution to solvation free energy; QEq, charge equilibration algorithm; E_{total}, total potential energy; μ_k^L, kth effective Langevin dipole; ξ_k^0, local field on the kth dipole; $F(\xi)$, function of electric field on the ith Langevin dipole at the solvation surface; ϵ_0, static dielectric constant; H64, high-resolution protein structure database containing 64 proteins; GluCOONa, poly-L-Glu, in the sodium ion complexed form; HF, Hartree-Fock wavefunction.

Contiguous sequence repeats of negatively charged amino acids (e.g., Asp, Glu, or posttranslationally modified Glu, Ser, Thr, and Tyr) occur occasionally in proteins (Gorski, 1992). Some examples of polyanionic sequences can be found in parathymosin (Frangou-Lazaridis et al., 1988), calsequestrin (Scott et al., 1988; Zarain-Herzberg et al., 1988), aspartactin (Clegg et al., 1988), and the sodium channel protein (Noda et al., 1984). A preponderance of polyanionic sequences are also found in the acidic template protein superfamily (Table 1). This superfamily of proteins plays an important role in controlling crystal and amorphous mineral growth in the process of biomineralization (Mann, 1988; Lowenstam & Weiner, 1989). Under solution conditions, polyanionic sequence stretches would be expected to exhibit polyelectrolyte behavior; that is, they would exist as charged species with counterions condensed along the volume of the repeat sequence (Manning, 1969a, 1969b, 1978; Record et al., 1978). Polyelectrolyte conformation and dynamics are significantly affected by factors such as pH, solution ionic strength, and the charge/radius ratio of the counterion(s) (Krimm & Mark, 1968; Manning, 1969a, 1969b, 1978; Manning & Holtzer, 1973; Record et al., 1978; Seidel, 1994). Hence, polyelectrolyte anionic sequences in proteins may play an important role in protein–ion and protein–protein interactions, solvation, and protein folding. However, due to conformational lability and other factors, experimental determination of polyanionic polypeptide conformational states (e.g., X-ray crystallography, NMR spectroscopy) has proved difficult (Evans & Chan, 1994; Evans et al., 1994).

As an adjunct to experimental methods of protein structure determination, theoretical methodologies have been developed and applied toward determining the native conformation (i.e., the global minima) of globular proteins (Anfinsen, 1973). Methods developed to search conformational space include Monte Carlo conformational search (Holm & Sander, 1992; Nayeem et al., 1992; Carlacci & Englander, 1993; Evans, 1993; Evans et al., 1995), on-lattice search methods (Lau & Dill, 1989; Kolinski & Skolnick, 1994), off-lattice search methods (Gregoret & Cohen, 1991; Gunn et al., 1994), and genetic algorithms (Goldberg, 1989; Dandekar & Argos, 1994). These methodologies have not been applied to the problem of determining polyanionic polypeptide structure. In part, this may be due to the problem of accurately modeling the considerable electrostatic interactions existing in these systems, as well as the problem of how to best represent solvent effects and counterions (either implicitly or explicitly) during the simulation process.

Our aims are: (1) to develop computational methods for rapidly exploring the preferred solution conformation(s) of polyelectrolyte polypeptide sequences; and (2) to determine the stabilizing interactions (hydrogen bonding, electrostatic, etc.) that define the conformational ensemble.

Thus, we begin with a simple level of theory and validate the approach against a structural benchmark. We assume that the conformation of polyanionic polypeptides is influenced by amino acid-specific secondary structure and side-chain interactions. Experimental evidence for this hypothesis has been established with structural studies of poly-L-Glu, poly-L-Asp, and copolypeptides of L-Glu and L-Leu. It was shown that poly-L-Glu and poly-L-Asp exhibit abrupt changes in Na^+ activity coefficients, compared to simple synthetic polyelectrolytes (Berger & Katchalski, 1951; Ascoli & Botre, 1963). This evidence inferred the presence of structural stabilization in polyanionic peptides. Furthermore, Fasman and others (Fasman et al., 1962, 1964), using L-Glu and L-Leu copolymers, demonstrated that addition of L-Leu in the sequence increases the helical stability of the Na^+ salt form of poly-L-Glu. This was attributed to hydrophobic interactions arising from the presence of Leu in the peptide sequence (in addition to backbone and/or side-chain hydrogen bonding). Protein database searches show that the 20 amino acids exhibit a range of dihedral angle preferences (ϕ, ψ, χ) in protein crystal structures (Rooman et al., 1991; Abagyan & Totrov, 1994; Evans et al., 1995) that reflect steric information encoded in the protein structures themselves (Mathiowetz, 1993; Evans et al., 1995). That is, specific combinations of amino acids lead to the creation of specific folds (Bowie et al., 1991). Thus, experimental studies of polyanionic polypeptides and database analyses both indicate that polyanionic sequences adopt folds that are influenced by amino acid side-chain type.

The dihedral probability grid–Monte Carlo method is a database-driven conformational search algorithm developed to rapidly search polypeptide conformational space (Mathiowetz,

Table 1. *Examples of polyanionic sequences in proteins*

Protein	Sequence	Reference
Osteopontin (bovine)	-DDLDDDDD-	Kerr et al., 1991
Phosphophoryn (bovine)[a]	-$(D)_n$-, -$(DPSer)_n$-, -$(PSer)_n$-	Sabsay et al., 1991; Evans and Chan, 1994
Parathymosin	-EDGEDDDEGDEEDEEEEEEEEDE-	Frangou-Lazaridis et al., 1988
Calsequestrin (skeletal)	-EDDDDEDDDDDDDD-	Zarain-Herzberg et al., 1988
Potassium ion transporter protein	-DMDDDDDDDDNDGDNNEE-	Noda et al., 1984
Calspermin	-SEEVEEEEGVKEEEEEEEEEET-	Ono et al., 1989
Prothymosin	-DEEEEEGGEEEEEEEEGDGEEEDGDEDEE-	Scott et al., 1988
Aspartactin	-DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD-	Clegg et al., 1988
Nucleolar nucleolin	-KEESEEDEDDEDDEEDEDEEEESDEEEEP-	Maridor et al., 1990
Nucleolar NO38	-SEDDEDDDEDEDDDEDDEDDLDDDEEEI-	Maridor et al., 1990
Ubiquinol-cytochrome *c* reductase	-GDEDEDEDEDEDDDDDDDEDEEEEEEV-	van Loon et al., 1984
DMP-1 (tooth dentine matrix)[b]	-DKEEDEDDSGDDTFGDED-	George et al., 1993
Mollusc shell acidic nucleation proteins	-$(DX)_n$-; -$(DPTD)_n$-; X = Ser or Gly	Weiner and Hood, 1975; Weiner, 1983

[a] PSer, *o*-phosphoserine; *n* denotes repeat blocks.
[b] This protein features several different Asp- and Glu-containing repeat regions; one of these is featured here.

1993; Evans et al., 1995; Mathiowetz & Goddard, 1995). In this method, the choices of internal coordinates are guided by side-chain-specific dihedral angle probabilities that are based on a collection of known protein crystal structures (Mathiowetz, 1993; Evans et al., 1995; Mathiowetz & Goddard, 1995). The DPG-MC algorithm successfully predicted the secondary structures of two polypeptides: (1) S3, a 20-residue polypeptide that adopts a right-handed α-helical conformation (contiguous from residue 2 through residue 15) (Lyu et al., 1993; Evans et al., 1995); and (2) the integrin receptor peptide, a 7-residue polypeptide that assumes a Type II β-turn in solution (Johnson et al., 1993; Evans et al., 1995). The DPG-MC method was shown to be free of starting structure bias (Evans et al., 1995).

Herein we apply the DPG-MC algorithm to the problem of determining polyanionic polypeptide conformational states. As a benchmark, we chose poly-L-$(Glu)_n$, a polyelectrolyte polypeptide extensively studied by UV absorption and CD spectroscopy. For ionic strengths at or exceeding the physiologic range (i.e., $\geq$0.1 M NaCl), the preferred peptide conformation for the Na^+ counterion form is a right-handed α-helix ($\leq$60% helical) (Ascoli & Botre, 1963; Fasman et al., 1964; Krimm & Mark, 1968; Sato et al., 1993). Using N^{α}-acetyl-L-$(Glu)_{20}$-C^{α}-amide as the model, we constructed a COO^- form in which each δ-carboxylate group is assigned a single Na^+ in a nonbonding geometry (Fig. 1). Using a branching algorithm (Abe et al., 1984), the Na atom is assigned to each side chain as a branch atom off the polypeptide backbone and moves with the side chain during the dihedral motion generated at each DPG-MC step. This approach is consistent with the counterion condensation model, in which the deprotonated state of the polypeptide is neutralized electrostatically by counterions that cannot be diluted away from the polymer surface (Schmitz & Yu, 1988; Manning, 1969a, 1969b, 1978). We find that poly-L-$(Glu)_{20}$ Na^+ prefers right-handed α-helical conformations. In addition, this approach was used to determine the low-energy conformational states for a number of representative capped polyanionic polypeptide sequences: L-Asp_{20}, L-$(PSer)_{20}$, and L-$(PSer\text{-}Asp)_{10}$ (Weiner, 1983; Oldberg et al., 1986; Sabsay et al., 1991; Evans & Chan, 1994). We find sequence-dependent conformational preferences for these polypeptides.

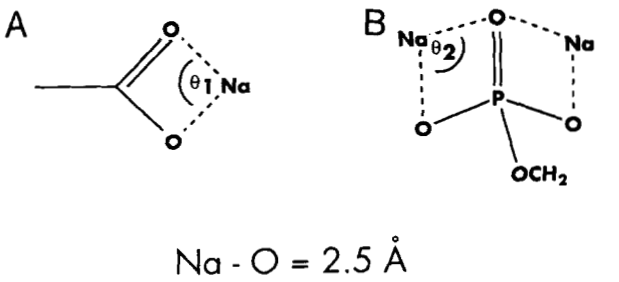

Fig. 1. Na-O geometries used to start the search for polyelectrolyte peptides. **A:** The 1 Na:1 Asp carboxylate. The Na-O distance is 2.5 Å, the O-C-O is 117.05°, resulting in $\theta_1 = 50.8°$. **B:** The 2 Na:1 PSer monophosphate ester. The Na-O distances for the bridging position are 2.5 Å, the O-P-O is 109.5°, resulting in $\theta_2 = 60°$.

Results

Charge distribution on Na-peptide complexes

Charges were obtained using the charge equilibration method (Rappè & Goddard, 1991), which leads to a consistent treatment of charges for counterions and/or side-chain atoms ion paired with the metal atoms. We calculated the charges for Na-counterion peptides in the extended secondary structure. These side-chain atom charges will influence the folding of the peptide into an energy-stabilized conformation. In addition, the peptide backbone charges influence the intrapeptide interactions along the chain (via dipole-dipole interactions) (Bellido & Rullman, 1989). Ab initio quantum chemistry calculations have shown that the partial charge on C_α depends strongly on the residue type and, to some extent, on the side-chain conformation (Bellido & Rullman, 1989). As shown in Figure 2, QEq generates consistent peptide backbone charges. With the exception of Na atoms (see below), the average standard deviation obtained for each heteroatom was found to be <0.01 charge units, whereas the deviation for hydrogen atoms was <0.001 charge units. This holds true for different sequence arrangements of Asp and PSer.

A comparison of the mean partial atomic charges obtained for the side-chain and Na atoms (Fig. 2) reveals that QEq assigns consistent values (within statistical error) for carboxylate and monophosphate ester atoms in each peptide, regardless of the sequence or the residue length. The QEq charge values for Na range from $Q_{Na} = 0.7$ to 0.9 (Fig. 2), which are plausible values for this ion. Indeed, variations in Q_{Na} values are of interest. For Na atoms ion paired with a carboxylate group, the Na partial atomic charge values ($Q_{Na} = +0.81$ for Asp_{20}, +0.86 for COO^- of $(PSer\text{-}Asp)_{10}$, +0.78 for Glu_{20}) were found to be very similar regardless of the sequence. However, the Na atom pairs assigned to the monophosphate ester, while similar in average charge ($Q_{Na} = +0.68$ to +0.79), exhibited a larger degree of variation (SD: $\pm$0.08–0.19) within a particular peptide-Na cluster (e.g., in $(PSer\text{-}Asp)_{10}$, Fig. 2). Because the geometries of the side chains and Na atoms are fixed identically in the starting structures, the variation in phosphate-specific Na charge assignments must result from nearest-neighbor Coulombic interactions (i.e., the influence from nearby monophosphate groups).

Analysis of poly-L-Glu potential energies

To determine the effects of the dielectric constant ϵ_0 on conformer selection during the DPG-MC runs, we examined the potential energies obtained for the benchmark peptide, N^{α}-acetyl-poly-L-$(Glu)_{20}$-C^{α}-amide, in the Na counterion form (abbreviated as GluCOONa) (Fig. 3). The following trends were observed in the potential energies for GluCOONa as ϵ_0 is increased from 1 to 80:

1. The electrostatic potentials decrease toward zero, becoming relatively constant in the range of $10 \leq \epsilon_0 \leq 80$ (Fig. 3).
2. The valence and van der Waals potentials decrease by 50–100%.
3. The hydrogen bonding interactions increase rapidly from $1 \leq \epsilon_0 \leq 10$, then remain constant.
4. The free energy of solvation, ΔG_p^{sol}, and the Langevin dipole free energy, ΔG_{Lang}, both decrease as ϵ_0 increases (Fig. 3).

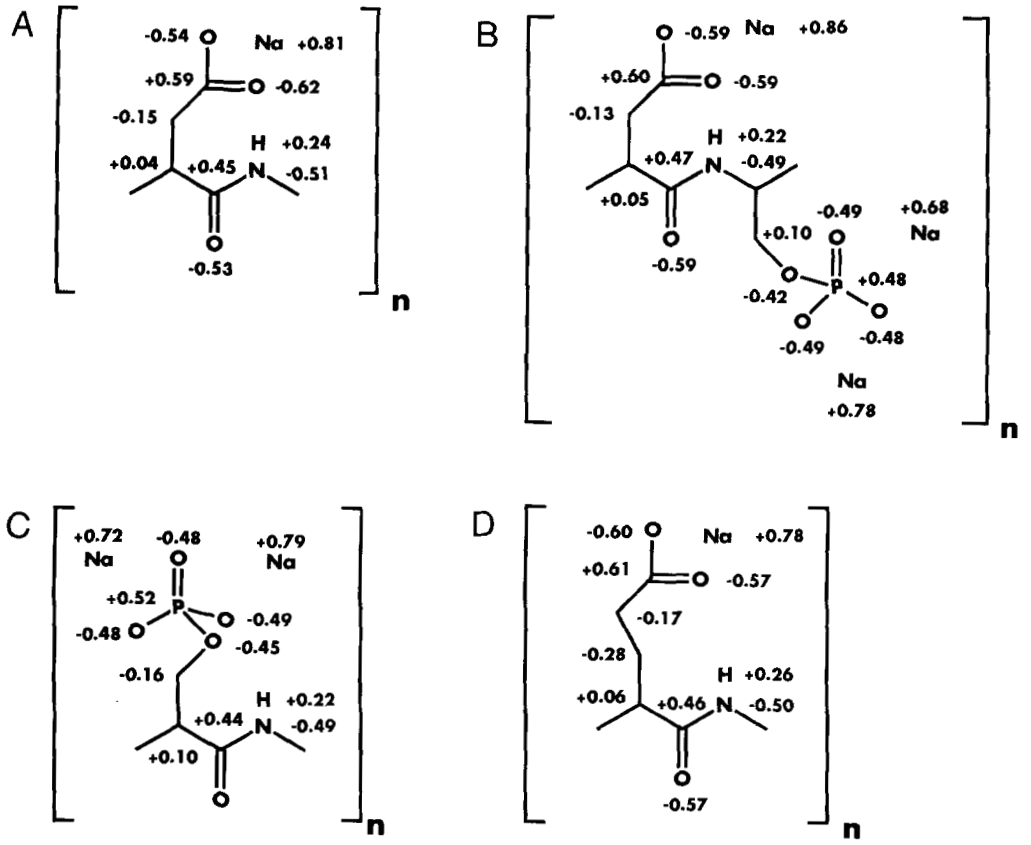

Fig. 2. Mean partial atomic charges obtained for backbone and side-chain atoms in the initial structures of Na^+-complexed peptides. Partial atomic charges were determined using the *Q*Eq method of Rappè and Goddard (1991) for each atom type in Na:peptides. All polypeptides were assigned an extended secondary structure (ϕ, $\psi = 180°$; χ^1, $\chi^2 = 60°$) prior to charge equilibration. The monomer atomic charges (in electron units (e.u.), ±SD, for $n = 20$) are given alongside the corresponding atom types. **A:** Asp. **B:** Asp-PSer. **C:** PSer. **D:** Glu.

The other POLARIS energy terms, ΔG_{Born} and $T\Delta S$, were unaffected by ϵ_0 (Evans, 1993). This is not surprising, because we are not introducing any variations in hydrophilicity or hydropathy. Interestingly, for $\epsilon_0 \geq 10$, both potentials experience a small linear decrease for GluCOONa. We find that the most favorable ΔG_p^{sol} and ΔG_{Lang} energies are observed for $\epsilon_0 \geq 10$.

Based on these analyses, we believe that the optimal dielectric for screened distance-dependent Monte Carlo simulations of polyanionic polypeptides is $\epsilon_0 = 10$. With $\epsilon_0 < 10$, the Coulombic energies seem to be unrealistically negative (leading to larger, more unfavorable valence and van der Waals interactions), whereas there are no significant changes in the energies for $\epsilon_0 \geq 10$. In addition, the PDLD grid-peptide dipole interaction energies are optimal for $\epsilon_0 \geq 10$.

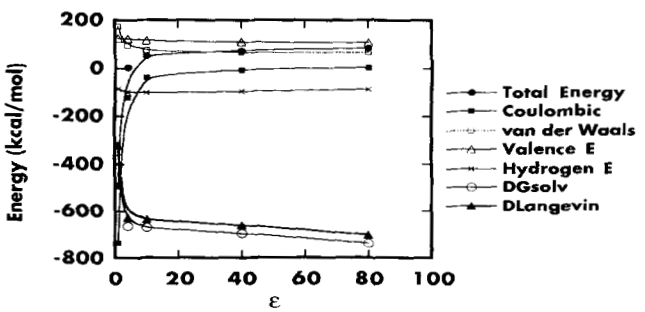

Fig. 3. Energetics for DPG-MC-accepted conformers of N^{α}-acetyl poly-L-$(Glu)_{20}$-C^{α}-amide polypeptide as a function of the dielectric constant, ϵ_0. In each case, the mean energy values are based on the DREIDING and POLARIS force fields and reflect accepted conformers ($n = 4$) determined for each ϵ_0. The SD is less than or equal to the symbol size. vdW denotes van der Waals. The valence energy is the sum of bonding interactions: $E_{Val} = E_{Bond} + E_{Angle} + E_{Torsion} + E_{Inversion}$.

Dihedral preferences for poly-L-Glu_{20} as a function of ϵ_0

We next examine the ϕ, ψ, and $\chi^{1,2,3}$ preferences and structures for the four lowest energy GluCOONa peptides as a function of ϵ_0 (Figs. 4, 5). Here, the energy is considered as the sum of the internal energy (using DREIDING) plus the external energy from POLARIS. (The DREIDING energy uses $\epsilon_0 = 1$ because the POLARIS calculations are performed at the same dielectric value.) The ϕ, ψ Ramachandran plots lead to two major populated regions (Fig. 4; Table 2): right-handed α-helix ($\phi = -55 \pm 20°$; $\psi = -55 \pm 20°$) and β-strand ($\phi = -120 \pm 45°$; $\psi = 130 \pm 30°$). There are subtle differences in the ϕ, ψ dihedral scatter for $\epsilon_0 < 10$ compared to values obtained for $\epsilon_0 > 10$. In general,

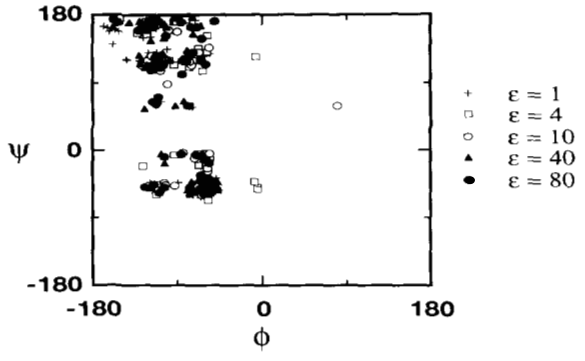

Fig. 4. ϕ/ψ Dihedral scatter plots for DPG-MC-accepted N$^{\alpha}$-acetyl poly-L-(Glu)$_{20}$-C$^{\alpha}$-amide polypeptide conformers. Dihedral pairs were obtained for all accepted structures ($n = 4$) for each dielectric constant studied. ϵ_0 Values are given at the top of each plot. Major ϕ, ψ regions are defined as shown in Table 2 (Garnier & Robson, 1989).

$\epsilon_0 \geq 10$, leads to smaller deviations. Hence, ϵ_0 has a minor effect on the selection of poly-L-Glu backbone conformations during the DPG-MC simulation. In contrast, there is a significant ϵ_0 dependence for the χ^3 torsions (Fig. 5). Although several poly-L-Glu *best* conformers possess similar χ^1, χ^2 distributions as a function of ϵ_0 (Fig. 5), we find very little agreement for χ^2, χ^3 distributions. The greatest variation is in the χ^3 dihedral, which involves the C_δ carboxylate group. This is plausible because the positioning of individual carboxylate side-chain groups on the peptide surface is influenced by electrostatic and van der Waals interactions with other neighboring side chains and with the counterions. Thus, DPG-MC selects poly-L-Glu low energy conformations primarily on the basis of (1) residue-dependent secondary structure, and (2) side-chain torsions that optimize nonbonding interactions.

The choice of ϵ_0 affects the selection of side-chain conformers and (to a lesser degree) the conformer secondary structure.

Comparison of GluCOONa structures

The ϕ, ψ distributions for the four lowest energy GluCOONa conformers ($\epsilon_0 = 10$; Fig. 6A) indicate a preference for right-handed α-helix conformations. This is also evident from the C_α traces and the all-atom representations for the lowest energy ensemble (Fig. 6B). As shown in Table 3, this ensemble has 47% average helicity, with β-strand and β-turn structures each comprising less than 10%. This value of 47% is close to the experimentally determined values (55–60%) for the sodium salt form of poly-L-Glu in solution (Ascoli & Botre, 1963; Fasman et al., 1964; Krimm & Mark, 1968; Sato et al., 1993). The backbone RMS Cartesian coordinate deviation (RMSD) for conformers in our ensemble is 5.1 Å, indicating significant structural variations within this ensemble. Some of the peptides feature nonhelical regions at the N- and C-termini, consistent with the end unraveling common to some α-helical peptide structures in solution (Lyu et al., 1992; Scholtz et al., 1993; Zhou et al., 1993). These nonhelical end regions (2–3 residues in length) feature backbone ϕ, ψ torsions that are β-strand, β-turn, or other secondary structures (Fig. 6A,B). Nonhelical ψ, ψ dihedrals are also found in 2–3-residue regions within the internal portion of the GluCOONa polypeptide (Fig. 6B).

Lowest energy conformational ensembles for Asp- and PSer- containing polyanionic polypeptides

We now examine the structural preferences for other Na^+ polyacidic sequences containing Asp and PSer residues. Using $\epsilon_0 = 10$, we generated a low-energy conformational ensemble for each model peptide series (Fig. 7). The dihedral populations (i.e., % helix, % β-sheet, % β-turn) are tabulated in Table 3 for each low-energy ensemble. In addition, a fit of the C_α backbone conformations was performed for each ensemble (Fig. 8A,B,C). The mean RMSDs of backbone C_α positions for Asp_{20}, $(PSer\text{-}Asp)_{10}$, and $PSer_{20}$ low-energy ensemble ($n = 4$) are 7.3 Å, 5.9 Å, and 8.0 Å, respectively. We observed the following trends:

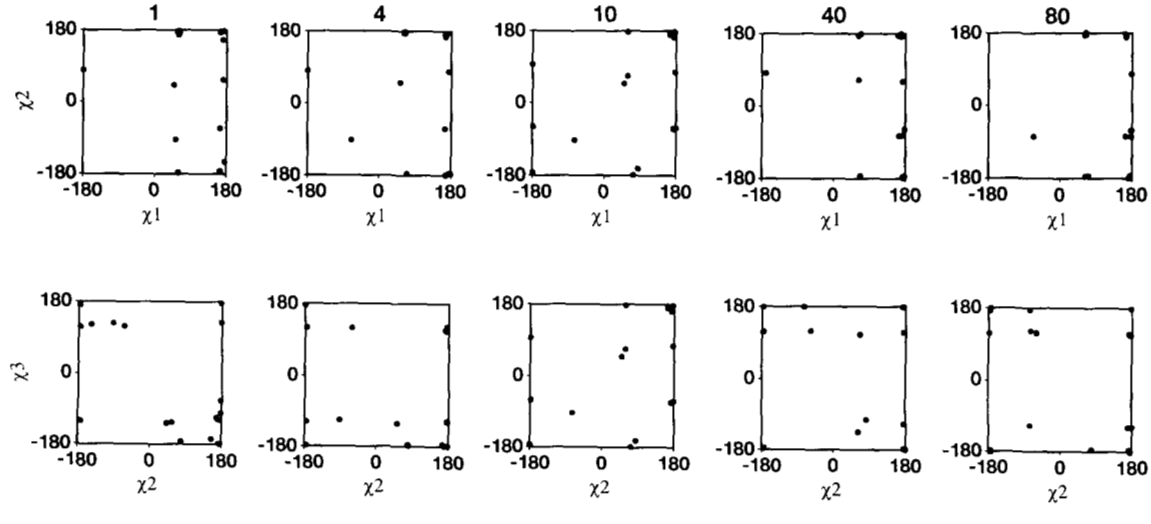

Fig. 5. χ Dihedral scatter plots for DPG-MC-accepted N$^{\alpha}$-acetyl poly-L-(Glu)$_{20}$-C$^{\alpha}$-amide polypeptide conformers. χ Dihedral pairs were obtained for the best or lowest-energy structures for each dielectric constant studied. ϵ_0 Values are given at the top of each plot.

Table 2. *Major ϕ, ψ regions*

Type	ϕ_i	ϕ_{i+1}	ψ_i	ψ_{i+1}
β-Strand	$-120° \pm 45°$	N/A	$130° \pm 30°$	N/A
Right-handed α-helix	$-55° \pm 20°$	N/A	$-55° \pm 20°$	N/A
β-Turn Type I	$-75° \pm 65°$	$-90° \pm 40°$	$-30° \pm 40°$	$-15°$ to $+40°$
β-Turn Type III	$-75° \pm 65°$	$-60° \pm 40°$	$-30° \pm 40°$	$-15°$ to $-70°$

1. For all sequences, the right-handed α-helix backbone percentage is small (<20%). There is a definite preference for β-turn and β-strand structures (Figs. 7, 8A,B,C; Table 3).
2. A significant percentage (>30%) of conformer backbone dihedrals was observed to lie outside the range for the three types of secondary structures examined. Of the three polypeptides, the PSer-Asp_{10} sequence exhibited the smallest content of other secondary structures (Table 3; Figs. 7, 8A,B,C).
3. The helix, turn, and strand preference was observed to be sequence dependent (Table 3).

As noted for the GluCOONa conformational ensemble, the RMSD for C_α positions indicates significant structural differences within each ensemble. The RMSD values obtained for $PSer_{20}$, Asp_{20}, and PSer-Asp_{10} are larger than the RMSD value for GluCOONa. An examination of the C_α traces reveals that there are qualitative structural similarities among the conformers within each PSer and Asp sequence family:

Asp_{20}

Secondary structures for this peptide reveal a 2:1:1 ratio for β-turn:β-sheet:right-handed α-helix (Fig. 8A; Table 3). The pep-

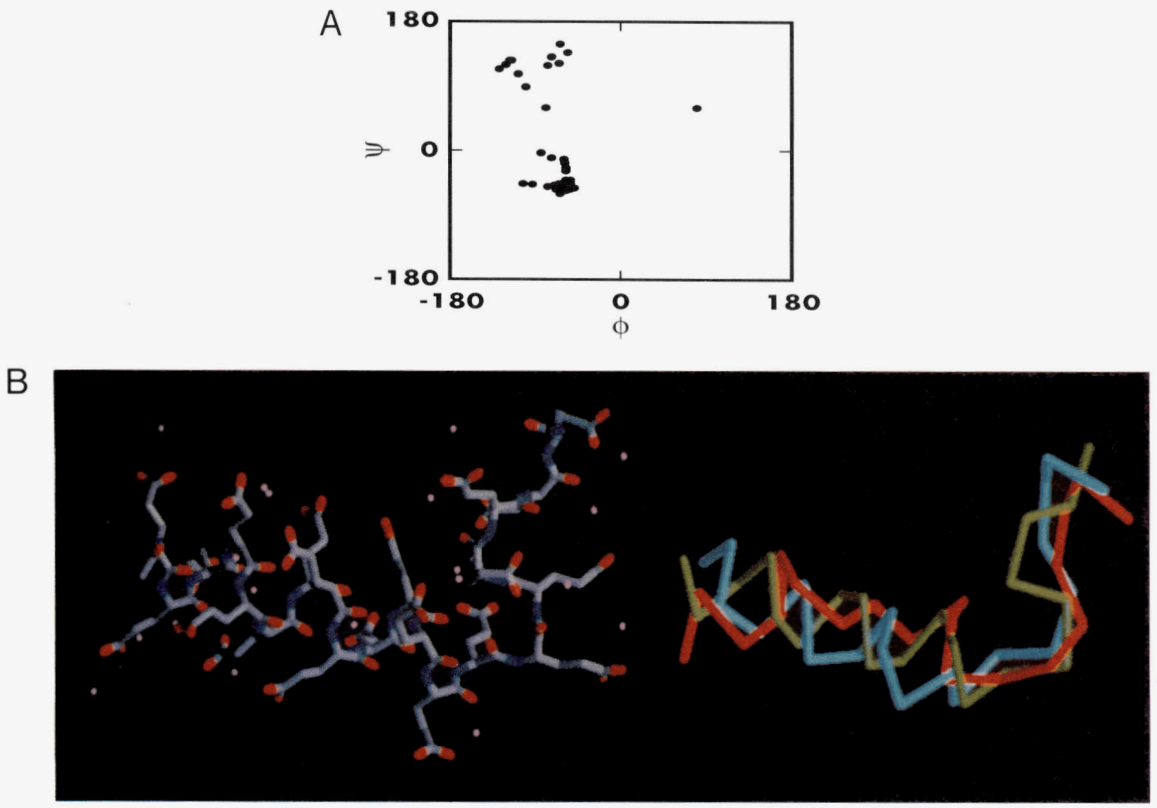

Fig. 6. ϕ, ψ Ramachandran plot (**A**) and C_α trace and ball-and-stick representations (**B**) of N^α-acetyl poly-L-$(Glu)_{20}$-C^α-amide, $\epsilon_0 = 10$. From the four-member low-energy ensemble, the global minimum structure was selected on the basis of the total potential energy (the internal energy from DREIDING [calculated at $\epsilon_0 = 1$] plus the external energy from POLARIS). In A, the dihedrals are presented for the global minimum structure; in B, the four-member low-energy ensemble (including the global minimum structure) is presented in terms of the C_α virtual bond representations. For the representations in B, a fit of atomic coordinates was performed for each ensemble structure. Alongside the C_α traces are a ball-and-stick representation of the global minima structure. Note that two of the four low-energy conformers in this ensemble in B are nearly identical. Hydrogen atoms are removed for clarity. DREIDING atom-type color legend: blue, nitrogen; red, oxygen; lavender, Na; gray or light blue, carbon.

Table 3. *Distribution of secondary structures in DPG-MC accepted conformers (for $\epsilon_0 = 10$)*[a]

Structure	% β-strand	% α-helix (rh)[b]	% β-turn[c]	Other[d]
Glu_{20}	8	47	12	32
Asp_{20}	18	13	26	43
$(PSer\text{-}Asp)_{10}$	36	10	19	35
$PSer_{20}$	25	4	29	42

[a] Percentages were determined from the total number of dihedral ϕ, ψ pairs ($n = 76$) taken from the four best accepted structures for each sequence. Classification of torsion angle ranges for defined secondary structures is given in Figure 4 (Garnier & Robson, 1989).
[b] rh, Right-hand.
[c] Type I and Type III β-turns are grouped under one classification.
[d] Percentage of dihedral pairs that were out of range for these classifications.

tide ensemble clearly adopts a collapsed coiled tertiary structure featuring a number of turns or bends (Fig. 8A).

$(PSer\text{-}Asp)_{10}$

This heteropolymer features a 3:2:1 ratio for β-sheet:β-turn:α-helix, differing significantly from the structures of either Asp or PSer homopolymer (Fig. 8B; Table 3). This peptide ensemble also features the lowest amount of other secondary structures. The resulting combination of sheet, turn, and helix gives rise to a superturn or C-shaped structure (Fig. 8B).

$PSer_{20}$

The PSer homopolymer features a nearly 1:1 ratio of β-turn: β-sheet dihedral preferences, with very little preference for α-helical structure (Fig. 8C; Table 3). Compared to Asp_{20}, the $PSer_{20}$ peptide ensemble is less coiled and possesses short regions of strand and turn combinations that give rise to a collapsed bend or hairpin structure (Fig. 8C).

Thus, the preference for dihedral pairs that each sequence exhibits (Table 3) translates into a unique secondary/tertiary structure for that particular sequence (Figs. 7, 8A,B,C).

Structural preferences and polyelectrolyte theory

Examination of the lowest energy conformers suggests that sequence-specific backbone and side-chain arrangements optimize the electrostatic interactions. It has been observed experimentally that delocalization of cations can occur along the length of the polymer if the nearest-neighbor charge spacing is sufficiently close such that two or more groups can cooperate in holding the cation (i.e., effective multivalency) (Manning, 1978). Experiments have also shown that in the presence of counterions, polyelectrolyte chains behave as flexible coils in solution (Krimm & Mark, 1968; Schmitz & Yu, 1988; Heath & Schurr, 1992). The polyanionic sequences in our study feature either

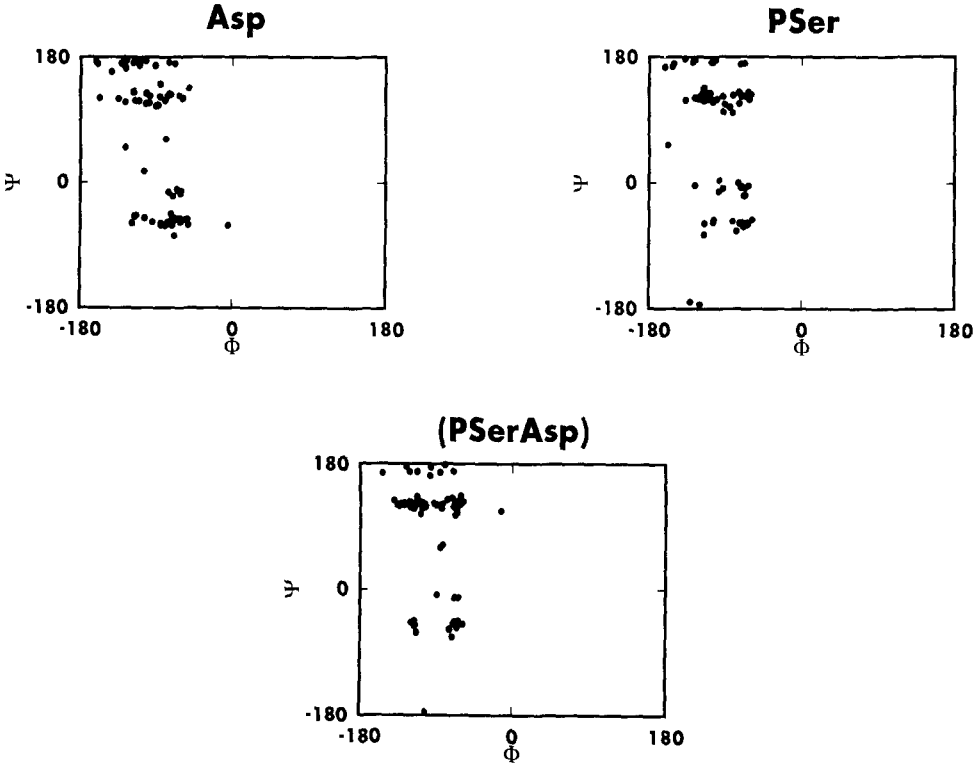

Fig. 7. ϕ, ψ Dihedral scatter plots for DPG-MC-accepted Asp- and PSer-containing polyanionic polypeptide conformers. Dihedral pairs were obtained for all accepted structures ($n = 4$) using $\epsilon_0 = 10$. Major ϕ, ψ regions are defined as shown in Table 2 (Garnier & Robson, 1989).

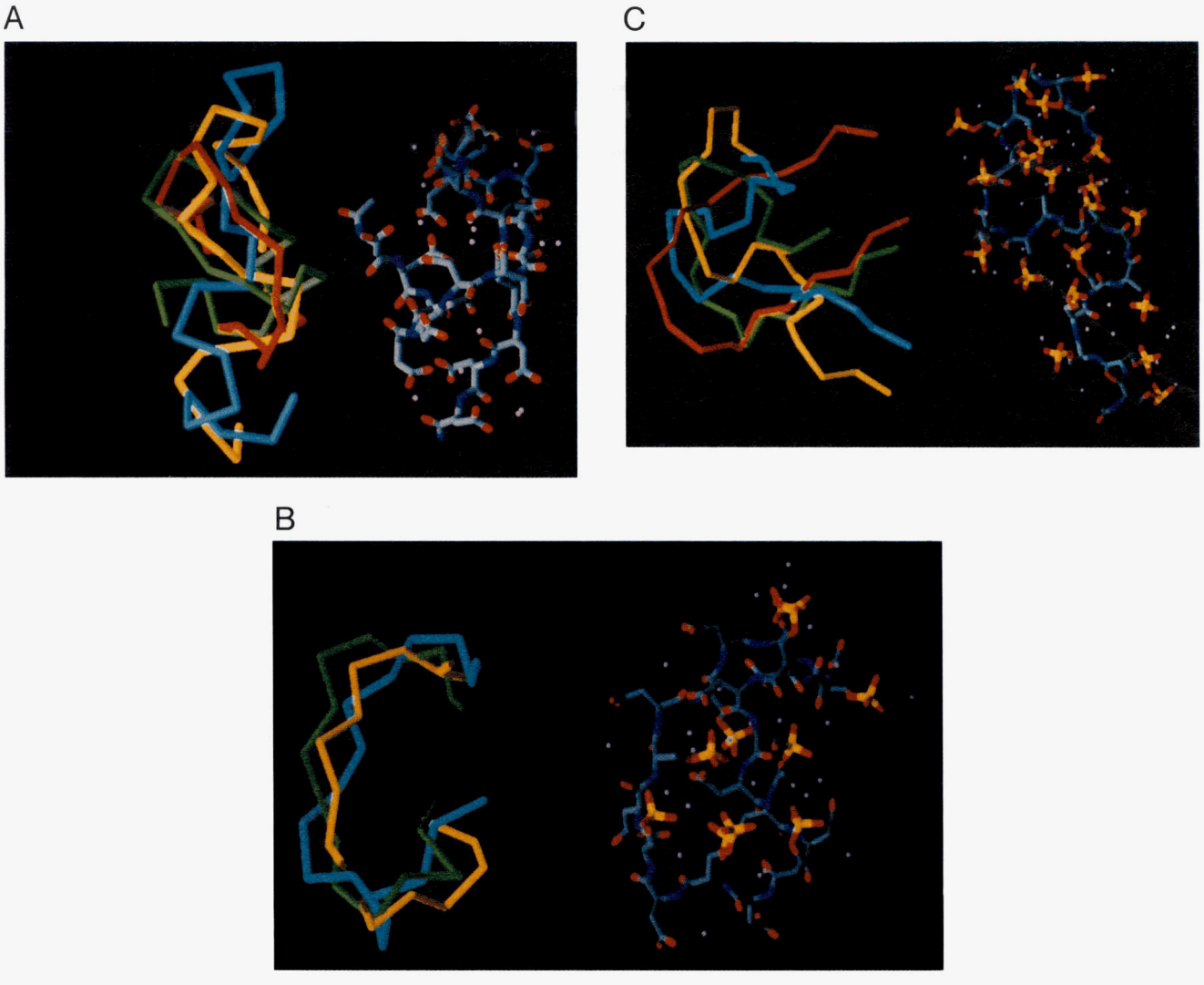

Fig. 8. C_α trace and ball-and-stick representations of accepted and *best* Asp- and PSer-containing polyanionic polypeptide conformers, $\epsilon_0 = 10$. For the four-member low-energy ensemble, the global minimal structure was selected on the basis of DREIDING (calculated at $\epsilon_0 = 1$) + POLARIS potentials, and a C_α fit of atomic coordinates was performed for each ensemble structure. Alongside each set of traces are ball-and-stick representations of the global minima structure. **A:** Asp_{20}. **B:** $(PSer\text{-}Asp)_{10}$. **C:** $PSer_{20}$. For $PSer\text{-}Asp_{10}$, two of the four low-energy conformers in this ensemble are nearly identical and thus overlap. Hydrogen atoms are removed for clarity. DREIDING atom-type color legend: yellow, P; blue, nitrogen; red, oxygen; lavender, Na; gray or light blue, carbon.

side-chain clustering and/or coiling conformations, which would promote Na^+ clustering effects (Figs. 6B, 8A,B,C). Counterion delocalization is also evidenced by the mean Na–O distances calculated for the phosphate and carboxylate complexes (Table 4); the mean phosphate and carboxylate oxygen–sodium distances are found to be within ±0.1 Å, indicative of uniform positioning of counterions between closely spaced anionic side chains.

The Glu_{20}, Asp_{20}, and $PSer_{20}$ homopolymers possess a uniform distribution of charged groups along the length of the polymer chain. Hence, chain collapse enhances counterion delocalization. In contrast, the $(PSer\text{-}Asp)_{10}$ sequence consists of polyion groups (PSer) alternating with monovalent ion groups (Asp), leading to an axial charge density of the polymer that fluctuates over its length. Thus, the $(PSer\text{-}Asp)_{10}$ conformational ensemble features partial segregation of the PSer and Asp groups on opposite sides of the peptide bond along the chain (Fig. 8B). Apparently, $(PSer\text{-}Asp)_{10}$ maximizes PSer Na-O clustering by backbone twisting, arranging the PSer groups so that Na atoms can be shared by adjacent PSer groups. The PSer-PSer clustering is favored electrostatically over the weaker PSer-Asp clustering where Na diffusion would be restricted. Thus, chain collapse in $(PSer\text{-}Asp)_{10}$ is more oriented than for the homopolymer cases. Consequently, the $(PSer\text{-}Asp)_{10}$ peptides adopt a compact C-shaped structure that features segregation (<70%) of the Asp side chains from the PSer side chains. Thus, electrostatic optimization induces clear conformational differences between polyanionic sequences.

Collectively, these simulations indicate that polyanionic sequences prefer coiled or collapsed conformations that provide a dense electrostatic molecular environment that permits cation delocalization and minimizes electrostatic repulsion within the

Table 4. *Mean Na-oxygen distances (Å) in DPG-MC-generated peptide conformers*

Structure	Na-ligand distance[a]	
	Carboxylate	Phosphate
Glu_{20}	3.40	–
Asp_{20}	3.37	–
$PSer_{20}$	–	3.39
PSer-Asp_{10}	3.46	3.35

[a] Mean distances computed for each assigned side-chain–Na cluster ($n = 100$ total Na-O pairs) within the low-energy ensemble of each peptide sequence.

complex. These theoretical results are consistent with polyelectrolyte theory and are also consistent with experimental evidence regarding polyelectrolyte chain conformations and ion distribution in solution (Manning, 1969a, 1969b, 1978; Katchalsky, 1971; Le Bret, 1982; Schmitz & Yu, 1988; Heath & Schurr, 1992; Sato et al., 1993; Seidel, 1994).

Discussion

We have approached the problem of defining the solution conformation of a conformationally elusive family of protein-polymer structures (namely, polyelectrolytes) by the use of a Metropolis Monte Carlo sampling algorithm (DPG-MC) and a generic microscopic force-field description of the electrostatic interactions. We describe a simple strategy to treat the monovalent counterions in nonbonding interactions with mono- and polyion groups along the length of the peptide chain by (1) allowing each counterion to remain with its side chain during random walk variations in the dihedral angles along the polypeptide chain, followed by (2) reoptimizing the positions of the counterions during a final minimization to relax the structure.

This provides a simple but effective way of modeling both solvation of the charged groups of polyelectrolyte and counterion clustering.

Using poly-L-Glu as a benchmark, we find that $\epsilon = \epsilon_0 R$ with $\epsilon_0 = 10$ seems to best simulate the solvation effects of the water electrostatic forces between the cations and the anionic sites and between anionic side chains. This gives reasonable results in terms of energetics (DREIDING plus POLARIS) and structures (Figs. 4, 5, 6).

Agreement of DPG-MC-predicted polyelectrolyte structures with polyelectrolyte theory and experimental data

Of major importance is the feasibility of DPG-MC and the counterion-pair approach for predicting polyelectrolyte conformations in solution. Using poly-L-Glu as a structural benchmark, we found a strong conformational preference toward right-handed α-helical structures (Fig. 6A,B). This is consistent with experimental observations (Ascoli & Botre, 1963; Fasman et al., 1964; Krimm & Mark, 1968; Sato et al., 1993). For Asp_{20} in the presence of Na^+, we found only a small percentage (<10%) of dihedral pairs in the ϕ, ψ region of α-helix (Figs. 7, 8A; Table 3). This is consistent with experimental CD studies on poly-L-(Asp) polypeptides, which found only a small degree of ordering (partial helicity) at low pH or in the presence of counterions (apparently due to steric repulsion effects) (Berger & Katchalski, 1951; Ascoli & Botre, 1963). Instead, we found β-turn and β-strand structures to be preferred (>10%). Recent experiments show that $(Asp)_n$, $(PSer\text{-}Asp)_n$, and $(PSer)_n$ repeat sequence regions in the biomineralization template protein, bovine dentine phosphophoryn, exhibit pH, thermal, and solvent denaturability consistent with tertiary folding of these sequences induced by counterion condensation at low ionic strength (Evans, 1993; Evans & Chan, 1994; J.S. Evans & S.I. Chan, manuscript in prep.). As the ionic strength increases, theory and experiment suggest that charge shielding decreases the electrostatic contributions to the persistence length or bending flexibility of a polyelectrolyte (Le Bret, 1982). Full stretching of flexible polyelectrolytes will not normally occur under most solution conditions (Seidel, 1994). In other words, the polyelectrolyte in the presence of counterions should adopt a more condensed solution conformation. Indeed, we observed this for the conformational ensemble of Asp- and PSer-containing peptides (Figs. 7, 8A,B,C). This suggests that such microscopic treatments of polyelectrolyte counterion-pairing can provide a feasible approach for studying the structure of polyanionic sequence repeats. This method for determining polyelectrolyte sequence conformation may also serve as a useful adjunct for NMR and X-ray structure determination methods of proteins containing these sequences.

Conformational snapshots of polyelectrolyte lowest energy conformers

One goal of this investigation into the low-energy conformers was to develop ideas about the conformational transitions for polyelectrolyte sequences in solution under conditions of counterion condensation. The distribution of ϕ, ψ dihedrals and the accompanying ball-and-stick figure representations (Figs. 6, 7, 8A,B,C) indicate that various peptide sequences utilize turn, helix, and β-strand or other secondary structures in different combinations to achieve a balance of electrostatic forces. For example, Asp_{20} and $PSer_{20}$ sequences both feature similar proportions of α-helical structure but slightly different percentages of β-strand and β-turn dihedral pairs (Figs. 7, 8A,C; Table 3). No evidence for Asp_{20} hairpins or $PSer_{20}$ supercoils was found. In comparison, the $(PSer\text{-}Asp)_{10}$ sequence forms a superturn or C-shaped structure (Fig. 8B) featuring a larger percentage of β-turn structure than either the Asp or PSer sequence and a lower percentage of α-helical structure (nearly 50%). This raises the possibility that the conformational ensembles for various polyanionic sequences in solution may be conservative. For example, in solution $(Asp)_n$ might fluctuate only between different types of large-turn coil structures but would avoid the formation of a C-shaped hairpin conformation. The reverse would be true for $(PSer\text{-}Asp)_n$. This restriction of conformational space may determine the electrostatic properties for each polyelectrolyte sequence.

Significance of polyanionic sequence clusters

For most proteins described in Table 1, it has been demonstrated or suggested that these polyanionic sequences are involved in metal ion interactions (Weiner & Hood, 1975; Weiner, 1983;

Noda et al., 1984; Gaber et al., 1988; Scott et al., 1988; Zarain-Herzberg et al., 1988; Kerr et al., 1991; Sabsay et al., 1991; Gorski, 1992; George et al., 1993). Our observation that the conformations depend upon polyanionic amino acid sequence may have some bearing on the solution behavior and cation binding properties of such polyanionic sequence domains, perhaps by modulating the charge density in specific regions of the sequence. One example can be cited in support of this hypothesis: the alkaline pK_a shifts observed for certain Asp clusters in phosphophoryn (Evans, 1993; Evans & Chan, 1994; Evans et al., 1994). This shift arises as a direct result of neighboring PSer residues exerting short-range electrostatic stabilization of protonation-deprotonation events at adjacent Asp carboxylate groups. Furthermore, each sequence cluster in phosphophoryn ($PSer_n$, $(PSer\text{-}Asp)_n$, Asp_n) exhibits different affinities of divalent cations and a sequential cation loading phenomenon (Evans et al., 1994). Thus, by subtle variations in the anionic sequence composition of a polyanionic domain, the tertiary folding of the domain may be changed in such a way as to bring about variation in oxoanion density. Consequently, variations in metal ion coordination, hydration, protein-protein interactions, or protein-nucleic acid interactions can occur. These studies indicate that molecular modeling can assist in predicting which specific anionic amino acid sequences give rise to particular conformational folds.

The counterion model utilized in this study could be modified to mimic conditions of low ionic strength where side-chain charge repulsion occurs. This may be accomplished in the simulations either by removing selective counterions, or by modifying partial atomic charge assignments on either the metal atom or the side-chain atom types. In this way, polyelectrolyte conformation and the resulting effects on pK_a's (Manning & Holtzer, 1973) could be examined in greater detail. Similarly, the effects of different counterions on charge shielding and polyelectrolyte conformation could be studied by these methods. The DPG-MC algorithm, along with the counterion ion pair model, could also be applied to conformational analysis of nonbiological polyelectrolyte polymers in solution. Further studies into these possibilities are currently in progress.

Materials and methods

All calculations presented here were performed using a modified version of BIOGRAF and version 3.22P of POLARIS (Molecular Simulations, Inc., Burlington, Massachusetts) running on Silicon Graphics workstations (R4400 processor). Most simulations used the DREIDING force field (Mayo et al., 1990), but the AMBER force field (Weiner et al., 1986) was used for some comparisons. The protein data set H64 was utilized for generating dihedral probabilities in our simulations (Mathiowetz, 1993; Evans et al., 1995). A full description of the implementation of the DPG-MC program is given elsewhere (Evans et al., 1995). Torques are not used in the DPG-MC method but are necessary for the conjugate-gradients internal-coordinate energy minimizer used to optimize the structures produced by the Monte Carlo simulations.

Description of polypeptides

The polyelectrolyte sequence peptides poly-L-$(Glu)_{20}$, poly-L-$(Asp)_{20}$, poly-L-$(PSer)_{20}$, and poly-L-$(PSer\text{-}Asp)_{10}$ were constructed in the extended backbone conformation (ψ, $\phi = 180°$; $\chi = 60°$), such that consecutive residues of the peptide were anticonformers relative to one another. We established from previous Monte Carlo studies that recognizable secondary structures will form in peptides with ≤ 20 residues (Evans et al., 1995). The Ser residues were modified by the base fragment addition of H_3PO_4 to form PSer, with subsequent removal of superfluous O and H atoms to create the deprotonated *o*-monophosphate ester. The peptides were capped by the addition of an acetyl group to the N-terminal α-NH_2 group and conversion of the C-terminal α-carboxylate group to the amide form. Explicit hydrogen atoms were included for the side-chain methylene and the peptide backbone, acetyl, and C-terminal amide for all peptides. The COO^- form was generated by deletion of the acid hydrogen from the corresponding side-chain carboxylate group.

Na^+ solvation

Anion sites such as Asp and Glu carboxylate and PSer monophosphate ester would cause reordering of solvent molecules around each peptide. This solvation can be simulated by using a large number of explicit water molecules and counterions (Smith & Pettitt, 1994). However, explicit water greatly slows the conformational search. We use instead a Na^+ near each anionic oxygen atom to represent aqueous solvation. This Na^+ partially represents real counterions and partially the solvation of these anionic sites. For the purposes of generating conformations, each Na^+ is considered a site-binding counterion species associated with the peptide. This situation mimics the condition of saturation (i.e., it accounts for charge screening due to the omnipresence of counterions at each charged peptide site) (Manning, 1969a, 1969b, 1978; Manning & Holtzer, 1973). However, the Na ion is free to reposition itself on the peptide structure during each phase of energy minimization that occurs after each Monte Carlo dihedral motion and during the quenching for each best accepted structure. Thus, in the fully deprotonated form, we have one Na^+ per COO^- and two Na^+ per PO_4^{2-}. That is, the system is minimally saturated with Na, and the net molecular charge of the Na paired peptide complex is neutral. (Note that we have not tried to model condensation, i.e., ions occupying a common free volume extending over the entire length of the polyion [Manning, 1969a, 1969b, 1978; Heath & Schurr, 1992].)

For Asp and Glu, we started with a single Na ion in the O-C-O bond plane of the carboxylate group, and for PSer, we started with a Na ion in each of the O-P-O bond planes as shown in Figure 1. Use of the Na-O bridging geometry is supported by structural evidence from databases compiled for carboxylates and phosphates (Baur, 1974; Carrell et al., 1988); this geometry has also been determined to be optimal for phosphate groups as determined from HF ab initio calculations (Pullman et al., 1978a, 1978b; Liebmann et al., 1982; Chen & Rossky, 1993). The atomic charges for each Na:peptide complex were assigned using the *Q*Eq method (Rappè & Goddard 1991) with a total charge of zero. Mean charges for the various Na:peptide complexes are shown in Figure 2. Partial charges for the peptide backbone atoms are also given in Figure 2. The extended structures of the Na:peptide complexes were then minimized to convergence prior to the DPG-MC runs (see below).

DPG-MC simulations of polyelectrolyte peptides

For the purposes of determining torsional motion, the DPG-MC procedure defines each dihedral ϕ, ψ pair side chain as a branching structure off of the peptide backbone (Abe et al., 1984). This branching feature permits the assignment of a nonbonded species, such as a cation or water molecule, to a specific side-chain residue. In this instance, each side-chain Na^+ ion was designated the last atom(s) of that particular branch structure off of the backbone. This side-chain–metal pair then remains associated via a minimized nonbonding distance throughout the dihedral motion and over the course of the conformational search. Although this approach is simplistic and qualitative in nature, it does provide a unique approach to maintaining nonbonding and electrostatic interactions with metal ions and side-chain ligand atoms. It also allows these interactions to influence the outcome of the global conformational minima search.

Using the minimized COO^- Na^+-peptide extended structure as the initial conformation, we carried out DPG-MC simulations at Boltzmann temperatures of 0, 300, 500, 1,000, and 2,000 K. Dihedral backbone pairs (ϕ, ψ) or side-chain χ were selected at each step of the simulation as described in our earlier investigations (Evans et al., 1995). For these peptide structures, we utilized 2×10^5 steps in each temperature run, for a total of 1×10^6 conformers generated for each peptide. Probability grids based on 5° increments were utilized. The dihedrals were incremented one pair at a time for each step, with the pairs selected randomly at each step of the Monte Carlo. The energy was calculated without nonbond cutoffs at each Monte Carlo step prior to the Metropolis evaluation. Accepted structures represent the lowest energy minima generated during a particular temperature run. From this set, the DPG-MC program determines the overall best conformer, which was then subjected to Cartesian minimization to convergence as described below.

Representing electrostatic forces and simulations in the presence of solvent

In the DREIDING force field (Mayo et al., 1990), the total potential energy function consists of the valence (E_{Val}) and nonbonding potential energy function (E_{NB}):

$$E_{Val} = E_{Bond} + E_{Angle} + E_{Torsion} + E_{Inversion} \tag{1}$$

$$E_{NB} = E_Q + E_{vdW} + E_{HB} \tag{2}$$

where E_Q, E_{vdW}, and E_{HB} denote the electrostatic, van der Waals, and hydrogen bonding potential energies between nonbonded atoms, and

$$E_Q = \frac{(322.0637)Q_iQ_j}{\epsilon R_{ij}}. \tag{3}$$

Here, Q_i is the charge of atom i in electron units, R_{ij} is the distance in angstroms, ϵ is the dielectric constant, and the constant 322.0637 converts units so that the energy is in kcal/mol. To include the effects of screening by solvent water, the ϵ term is expressed as (Mayo et al., 1990):

$$\epsilon = \epsilon_0 R_{ij}. \tag{4}$$

This should yield a good first approximation for short-range metal–side-chain and side-chain–side-chain electrostatic interactions (i.e., ≤9.0 Å) (Bellido & Rullman, 1989), while permitting long-range electrostatic interactions (Bellido & Rullman, 1989).

In order to establish the appropriate value of ϵ_0 in Equation 4, we considered values of 1, 4, 10, 40, and 80 (such values have been used for implicit modeling of bulk solvent effects on mechanistic and dynamic simulations [Warshel & Russell, 1984, 1985; Matthew, 1985; Harvey, 1989; Solmajer & Mehler, 1991; Lee et al., 1993]). The use of Equation 4 with $\epsilon_0 = 1$ will overestimate electrostatic forces (Smith & Pettitt, 1992, 1994), whereas $\epsilon_0 = 80$ should underestimate these forces (Matthew, 1985).

Because the calculated electrostatic energy of the system depends on ϵ_0, we need some criteria for selecting among the accepted structures for various values of ϵ_0 to obtain a best conformer. To do this, we used the PDLD of Warshel and colleagues (Warshel & Russell, 1984, 1985; Lee et al., 1993) to evaluate the free energy of solvation ($\Delta\Delta G_p^{sol}$) for each accepted DPG-MC structure. These calculations were carried out using the POLARIS force field (Lee et al., 1993) (version 3.22P, Molecular Simulations, Inc.). This model simulates solvation free energy in polar solvents via the Langevin function for average orientations of the solvent dipoles around charged atoms together with two additional solvation terms. Specifically,

$$\Delta\Delta G_p^{sol} = \Delta G_{Lang} + \Delta G_{Born} + (-T\Delta S) \tag{5}$$

$$\Delta G_{Lang} = -\frac{332}{2}\sum_k \mu_k^L \xi_k^0 \tag{6}$$

$$\Delta G_{Born} = -166\left(\frac{Q_2}{b} + \frac{(RQ)^2}{b^3}\right) \quad \text{for} \quad R \ll b \tag{7}$$

$$\Delta S = A + B\sum_{i \in s} CF(\xi_i^0). \tag{8}$$

$\Delta G_{Langevin}$ is the enthalpy energy arising from the interaction of the water dipoles with the solute (Lee et al., 1993), μ_k^L is the kth effective Langevin dipole, and ξ_k^0 is the local field on the kth dipole. The ΔG_{Born} term is the residual bulk energy of the solvent, b is the radius of the solvent sphere, and R is the distance of the ionized group from the center of the sphere. Finally, $(-T\Delta S)$ is the entropy contribution to the solvation free energy. Here, $F(\xi)$ is a function of the electric field on the ith Langevin dipole at the solvation surface, and A, B, and C are scaling constants obtained from studies of the free energies of transfer of different molecules from hydrocarbons to water (Lee et al., 1993). For each Na^+-peptide, the QEq charge assignments were utilized for dipole determinations, and a two-region model was defined (Lee et al., 1993): Region I contains the peptide, Region II is nonexistent, and Region III contains the water dipole grid in which the peptide is centered (Warshel & Russell, 1985). The Region III solvent sphere was defined by an outer radius of 30 Å, an inner radius of 12 Å, and an inner and outer polarizable grid spacing of 1 Å and 3 Å, respectively. No cutoffs were used. The value of 30 Å accommodates all of the DPG-MC-generated structures in our study. The Boltzmann-averaged solvation properties were calculated for each peptide using small random displacements of the Langevin grid center and averaging over five determinations.

Minimization methods

The following minimization protocol was utilized for both the extended starting structures and the best accepted structures: 10 steps of steepest descents minimization (Press et al., 1986) were initially used to relieve improper contacts, followed by conjugate-gradient minimization (Press et al., 1986) to convergence. No constraints or nonbonding cutoffs were used. In all instances, a hydrogen bonding cutoff value of 5.0 Å was used.

Acknowledgments

We thank Dr. Siddharth Dasgupta for his helpful advice during the course of this study. J.S.E. acknowledges a Postdoctoral National Research Service Award from the NIH (NIDR 1F32-DE-05445) and a fellowship award from AMGEN Pharmaceuticals. These studies were supported by a grant from DOE-AICD, using facilities supported also by NSF-CHE, NSF-GCAG, Aramco, Asahi Glass, Asahi Chemical, BP Chemical, Chevron Petroleum Technology Co., Oronite, Vestar, Hughes Research Labs, Xerox, and Beckman Institute.

References

Abagyan R, Totrov M. 1994. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol 235*:983–1002.

Abe H, Braun W, Noguti T, Go N. 1984. Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. General recurrent equations. *Comput Chem 8*:239–247.

Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science 181*:223–230.

Ascoli F, Botre C. 1963. Polyelectrolyte behavior of charged polyamino acids in aqueous solution. *Biopolymers 1*:353–357.

Baur WH. 1974. The geometry of polyhedral distortions. Predictive relationships for the phosphate group. *Acta Crystallogr B 30*:1195–1215.

Bellido MN, Rullman JAC. 1989. Atomic charge models for polypeptides derived from ab initio calculations. *J Comput Chem 10*:479–487.

Berger A, Katchalski E. 1951. Poly-L-aspartic acid. *J Am Chem Soc 73*: 4084–4088.

Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*: 164–170.

Carlacci L, Englander SW. 1993. The loop problem in proteins: A Monte Carlo simulated annealing approach. *Biopolymers 33*:1271–1286.

Carrell CJ, Carrell HL, Erlebacher J, Glusker JP. 1988. Structural aspects of metal-ion carboxylate interactions. *J Am Chem Soc 110*:8651–8656.

Chen SWW, Rossky PJ. 1993. Potential of mean force for a sodium dimethyl phosphate ion pair in aqueous solution: A further test of the extended RISM theory. *J Phys Chem 97*:6078–6082.

Clegg DO, Helder JC, Hann BC, Hall DE, Reichardt LF. 1988. Amino acid sequence and distribution of mRNA encoding major skeletal muscle laminin binding protein: Extracellular matrix-associated protein with unusual COOH-terminal polyaspartate domain. *J Cell Biol 107*:699–705.

Dandekar T, Argos P. 1994. Folding the main chain of small proteins with the genetic algorithm. *J Mol Biol 236*:844–861.

Evans JS. 1993. NMR and computational studies of the biomineralization template protein, bovine dentine phosphophoryn [thesis]. Pasadena, California: California Institute of Technology.

Evans JS, Chan SI. 1994. Phosphophoryn, a biomineralization template protein. pH-dependent protein folding experiments. *Biopolymers 34*:507–527.

Evans JS, Chiu T, Chan SI. 1994. Phosphophoryn, a biomineralization regulatory protein. Conformational folding in the presence of Cd (II). *Biopolymers 34*:1359–1375.

Evans JS, Mathiowetz AM, Chan SI, Goddard WA. 1995. De novo prediction of polypeptide conformations using dihedral probability grid Monte Carlo methodology. *Protein Sci 4*:1203–1216.

Fasman GD, Lindblow C, Bodenheimer E. 1962. The stabilization toward temperature of the helical conformation of copolypeptides of L-glutamic acid and L-leucine: An inverse temperature effect. *J Am Chem Soc 84*: 4977–4979.

Fasman GD, Lindblow C, Bodenheimer E. 1964. Conformational studies on synthetic poly-α-amino acids: Factors influencing the stability of the helical conformation of poly-L-glutamic acid and copolymers of L-glutamic acid and L-leucine. *Biochemistry 3*:155–166.

Frangou-Lazaridis M, Clinton M, Goodall GJ, Horecker BL. 1988. Prothymosin a and parathymosin: Amino acid sequences deduced from the cloned rat spleen cDNAs. *Arch Biochem Biophys 263*:305–310.

Gaber RF, Styles CA, Fink GR. 1988. TRK1 encodes a plasma membrane protein required for high-affinity potassium transport in *Saccharomyces cerevisiae*. *Mol Cell Biol 8*:2848–2859.

Garnier J, Robson B. 1989. Secondary structure prediction. In: Fasman GD, ed. *Prediction of protein structure and the principles of protein conformation*. New York: Plenum Press. pp 418–425.

George A, Sabsay B, Simonian PAL, Veis A. 1993. Characterization of a novel dentin matrix acidic phosphoprotein. Implications for induction of bio-mineralization. *J Biol Chem 268*:12624–12630.

Goldberg DE. 1989. *Genetic algorithms in search, optimization and machine learning*. Reading, Massachusetts: Addison-Wesley Publishing.

Gorski J. 1992. Acidic phosphoproteins from bone matrix: A structural rationalization of their role in biomineralization. *Calcif Tiss Int 50*:391–396.

Gregoret LM, Cohen FE. 1991. Protein folding. Effect of packing density on chain conformation. *J Mol Biol 219*:109–122.

Gunn JR, Monge A, Friesner RA, Marshall CH. 1994. Hierarchical algorithm for computer modeling of protein tertiary structure: Folding of myoglobin to 6.2 Å resolution. *J Phys Chem 98*:702–711.

Harvey SC. 1989. Treatment of electrostatic effects in macromolecular modeling. *Proteins Struct Funct Genet 5*:78–92.

Heath PJ, Schurr JM. 1992. Counterion condensation. Effects of site binding, fluctuations in nearest-neighbor interactions, and bending. *Macromolecules 25*:4149–4159.

Holm L, Sander C. 1992. Fast and simple Monte Carlo algorithm for side chain optimization in proteins: Application to model building by homology. *Proteins Struct Funct Genet 14*:213–223.

Johnson WC, Pagano TG, Basson CT, Madri JA, Gooley P, Armitage IA. 1993. Biologically active Arg-Gly-Asp oligopeptides assume a type II β-turn in solution. *Biochemistry 32*:268–273.

Katchalsky A. 1971. Polyelectrolytes. *Pure Appl Chem 26*:327–372.

Kerr JM, Fisher LW, Termine JD, Young MF. 1991. The cDNA cloning and RNA distribution of bovine osteopontin. *Gene 108*:237–243.

Kolinski A, Skolnick J. 1994. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins Struct Funct Genet 18*:338–352.

Krimm S, Mark JE. 1968. Conformations of polypeptides with ionized side chains of equal length. *Proc Natl Acad Sci USA 60*:1122–1129.

Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules 22*:3986–3997.

Le Bret M. 1982. Electrostatic contribution to the persistence length of a polyelectrolyte. *J Chem Phys 76*:6243–6255.

Lee FS, Chu ZT, Warshel A. 1993. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. *J Comput Chem 14*(2):161–185.

Liebmann P, Loew G, McLean AD, Pack GR. 1982. Ab initio SCF studies of Li, Na, Be, and Mg with H_2PO_4: Model for cation binding to nucleic acids. *J Am Chem Soc 104*:691–697.

Lowenstam HA, Weiner S. 1989. *On biomineralization*. New York: Oxford University Press. pp 7–50.

Lyu P, Wemmer DE, Zhou HT, Pinter RO, Kallenbach NR. 1993. Capping interactions in isolated α-helices: Position-dependent substitution effects and structure of a serine-copper peptide helix. *Biochemistry 32*:421–425.

Mann S. 1988. Regulation of biomineral formation. In: Mann S, Webb J, Williams RJP, eds. *Biomineralization: Chemical and biochemical perspectives*. New York: VCH Publishers. pp 35–61.

Manning GS. 1969a. Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties. *J Chem Phys 31*:924–933.

Manning GS. 1969b. Limiting laws and counterion condensation in polyelectrolyte solutions. II. Self-diffusion of the small ions. *J Chem Phys 31*:934–938.

Manning GS. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys 11*:179–246.

Manning GS, Holtzer A. 1973. Application of polyelectrolyte limiting laws to potentiometric titration. *J Phys Chem 77*:2206–2212.

Maridor G, Krek W, Nigg EA. 1990. Structure and developmental expression of chicken nucleolin and NO38: Coordinate expression of two abundant nonribosomal nucleolar proteins. *Biochim Biophys Acta 1049*: 126–133.

Mathiowetz AM. 1993. Dynamic and stochastic protein simulations: From

peptides to viruses [thesis]. Pasadena, California: California Institute of Technology.
Mathiowetz AM, Goddard WA. 1995. Building proteins from C-alpha coordinates using dihedral probability grid Monte Carlo. *Protein Sci 4*: 1217–1232.
Matthew JB. 1985. Electrostatic effects in proteins. *Annu Rev Biophys Chem 14*:387–417.
Mayo SL, Olafson BD, Goddard WA. 1990. DREIDING: A generic force field for molecular simulations. *J Phys Chem 94*:8897–8909.
Nayeem A, Vila J, Scheraga HA. 1992. A comparative study of the simulated annealing and Monte Carlo-with-minimization approaches to the minimum-energy structures of polypeptides: (Met)-enkephalin. *J Comput Chem 12*:594–605.
Noda M, Shimizu S, Tanabe T, Takai T, Kayano T, Ikeda T, Takahashi H, Nakayama H, Kanaoka Y, Minamino N, Kangawa K, Matsuo H, Raftery M, Hirose T, Inayama S, Hayashida H, Miyata T, Numa S. 1984. Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature 312*:121–127.
Ono T, Slaughter GR, Cook RG, Means AR. 1989. Molecular cloning sequence and distribution of rat calspermin, a high affinity calmodulin-binding protein. *J Biol Chem 264*:2081–2087.
Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1986. *Numerical recipes: The art of scientific computing*. Cambridge, UK: Cambridge University Press. pp 274–328.
Pullman A, Pullman B, Berthod H. 1978a. An SCF ab initio investigation of the *through-water* interaction of the phosphate anion with the Na^+ cation. *Theor Chim Acta 47*:175–192.
Pullman B, Pullman A, Berthod H. 1978b. SCF ab initio study of the *through-water* versus *direct* binding of the Na^+ and Mg^{2+} cations to the phosphate anion. *Int J Quantum Chem Quant Biol Symp 5*:79–90.
Rappè AK, Goddard WA. 1991. Charge equilibration for molecular dynamics simulations. *J Phys Chem 95*:3358–3363.
Record MT, Anderson CF, Lohman TM. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys 11*:103–178.
Rooman MJ, Kocher JPA, Wodak SJ. 1991. Prediction of protein backbone conformation based on 7 structure assignments: I. Influence of local interactions. *J Mol Biol 221*:961–979.
Sabsay B, Stetler-Stevenson WG, Lechner JH, Veis A. 1991. Domain structure and sequence distribution in dentin phosphophoryn. *Biochem J 276*:699–707.
Sato M, Kato F, Komiyama J. 1993. Specific helix stabilization of poly(L-glutamic acid) in mixed counterion systems. *Biopolymers 33*:985–993.
Schmitz KS, Yu JW. 1988. On the electrostatic contribution to the persistence length of flexible polyelectrolytes. *Macromolecules 21*:484–493.
Scholtz JM, Qian H, Robbins VH, Baldwin RL. 1993. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry 32*:9668–9676.
Scott BT, Simmerman HKB, Collins JH, Nadal-Ginard B, Jones LR. 1988. Complete amino acid sequence of canine cardiac calsequestrin deduced by cDNA cloning. *J Biol Chem 263*:8959–8964.
Seidel C. 1994. The polyelectrolyte brush: Implications from simulation results on polyelectrolytes in solution. *Macromolecules 27*:2227–2231.
Smith PE, Pettitt BM. 1992. Amino acid side chain populations in aqueous and saline solution: Bis-penicillamine enkephalin. *Biopolymers 32*:1623–1629.
Smith PE, Pettitt BM. 1994. Modeling solvent in biomolecular systems. *J Phys Chem 98*:9701–9711.
Solmajer T, Mehler EL. 1991. Electrostatic screening in molecular dynamics simulations. *Protein Eng 4*:911–917.
van Loon APGM, DeGroot RJ, DeHann M, Kekker A, Givell LA. 1984. The DNA sequence of the nuclear gene coding of the 17-kd subunit VI of the yeast ubiquinol-cytochrome *c* reductase: A protein with an extremely high content of acidic amino acids. *EMBO J 3*:1039–1043.
Warshel A, Russell ST. 1984. Calculations of electrostatic interactions in biological systems and in solution. *Q Rev Biophys 17*:283–422.
Warshel A, Russell ST. 1985. Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor. *J Mol Biol 185*:389–404.
Weiner S. 1983. Mollusk shell formation: Isolation of two organic matrix proteins associated with calcite deposition in the bivalve *Mytilus californianus*. *Biochemistry 22*:4139–4145.
Weiner S, Hood L. 1975. Soluble protein of the organic matrix of mollusk shells: A potential template for shell formation. *Science 190*:987–989.
Weiner SJ, Kollman PA, Nguyen DT, Case DA. 1986. An all-atom force field for simulations of proteins and nucleic acids. *J Comput Chem 7*:230–252.
Zarain-Herzberg A, Fliegel L, MacLennan DH. 1988. Structure of the rabbit fast-twitch skeletal muscle calsequestrin gene. *J Biol Chem 263*: 4807–4812.
Zhou NE, Kay CM, Sykes BD, Hodges RS. 1993. A single-stranded amphipathic helix in aqueous solution: Design, structural characterization, and its application for determining α-helical propensities of amino acids. *Biochemistry 32*:6190–6197.