University of New Orleans

ScholarWorks@UNO

Chemistry Faculty Publications

Department of Chemistry

The nonplanarity of the peptide group: Molecular dynamics simulations

Steven W. Rick
University of New Orleans, srick@uno.edu

Follow this and additional works at: https://scholarworks.uno.edu/chem_facpubs

Part of the Chemistry Commons

Recommended Citation

The nonplanarity of the peptide group: Molecular dynamics simulations with a polarizable two-state model for the peptide bond

Steven W. Rick[a)] and R. E. Cachau
Advanced Biomedical Computing Center, SAIC-Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 27102

The general properties of the peptide bond can be described from a linear combination of two states: a single bond neutral form and a double bond zwitterionic form. However, environmental effects can shift the balance of the linear combination. This would cause the rigidity of torsional rotations of the peptide bond to be environmentally dependent and, in fact, an analysis of protein structures in the protein data bank reveals a different degree of nonplanarity for different secondary structure elements. A potential is presented in which the peptide bond is treated as a linear combination of two states; the coefficients of the two states are updated as the simulation progresses using an extended Lagrangian formalism. The model is applied to the helix/coil transition of polyalanine. Fluctuations in the planarity of the peptide dihedral angle are found to increase the rate constant for the coil to helix transition by a factor of two. © *2000 American Institute of Physics.* [S0021-9606(00)50411-5]

I. INTRODUCTION

The properties of the peptide group are a result of the resonance between contributing structures, with a single bond state (A) and a zwitterionic double bond state (B) as the

$$\text{(1)}$$
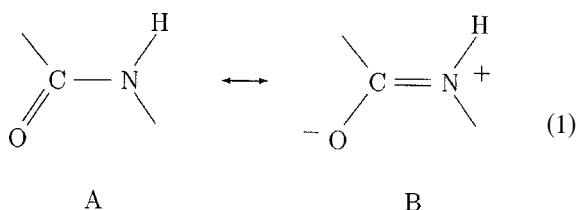
main contributors.[1] The rigidity of the peptide bond plane is due to the partial double bond character of the C–N bond, as well as the loss of resonance energy between the two states for nonplanar geometries. The energies of the two states and therefore the contribution each state makes to the superposition can be changed by interactions with the environment. For example, hydrogen bonds in the plane of the peptide bond, such as those formed in α-helices, will stabilize the zwitterionic form, while hydrogen bonds to the nitrogen atom perpendicular to the plane will stabilize the single bond neutral form. This type of hydrogen bonding, termed π-bond cooperativity because it involves the polarization of electrons with some π-bond character, is involved in the hydrogen bonding of other molecules, including nucleic acid base pairs.[2] As a consequence, not only the charge distributions on atomic sites, but also the flexibility of dihedral rotation about the peptide bond will be environmentally dependent. This polarization response would cause units of secondary structure with in-plane hydrogen bonds, such as α-helices

and β-sheets, to be more rigid while coils and turns would be less rigid. The two-state nature of the peptide bond would lead to nonadditive energies for hydrogen bond formation.

If the double bond character of the peptide bond can change considerably with its environment, then an analysis of the values of the $C_\alpha^i - C^i - N^{i+1} - C_\alpha^{i+1}$ dihedral angle (ω) from the protein data bank (PDB) should reveal a different distribution for the different secondary structure types. However, as pointed out by MacArthur and Thornton, the refinement process biases the ω-angle distribution to be artificially narrow relative to unbiased structures from the Cambridge structural database of small molecules.[3] Atomic resolution structures make available a small, but growing, database of unbiased protein structures. An analysis of 13 atomic-resolution structures shows a broad ω distribution.[4] In this article, we will present a detailed analysis of 22 atomic-resolution protein structures with an examination of the ω-angle distribution dependence on secondary structure. The increased rigidity of the peptide bond due to hydrogen bonding is also demonstrated by the shortening of the C–N bond and lengthening of the $C=0$ bond in hydrogen-bonded crystals of amide molecules compared to isolated electronic structure calculations.[2]

In an α-helix the dipoles of each peptide group will align. The electric field from the aligned dipoles will shift the equilibrium in Eq. (1) towards the more polar state, B. This cooperative effect will make peptide–peptide hydrogen bonds formed after the first hydrogen bond to be stronger than the first, which may be an important factor in the stability of the folded state, with peptide–peptide hydrogen bonds, relative to the unfolded state, with peptide–water hydrogen bonds. Cooperative effects on the energies and dipole moments have been observed both by experiment and by electronic structure methods for peptides and peptide-group

---

[a)]Author to whom correspondence should be addressed.

models, such as *N*-methylacetamide (NMA) and *N*-methylformamide (NMF). In 1962, Klotz and Franzen reported that their near-infrared experiments of NMA in water showed that there were cooperative effects in NMA aggregation, in their words ''once the dimer is formed, a trimer is formed even more readily.''[5] Dielectric measurements on polypeptide chains indicate that α-helix formation increases the dipole moment, μ, of the peptide backbone from about 3.5 debye, for isolated peptide groups, to 4.8–5.0 D.[6] An *ab initio* study with minimal basis sets on polyalanine and polyglycine found similar increases in the dipole moments from a fully extended chain to an α-helical conformation.[7] *Ab initio* calculations of NMA and NMF give about a 20% increase in the hydrogen bond strength of the trimer relative to the dimer, or about a 1 kcal/mol increase per hydrogen bond.[8,9]

The dependence of the flexibility of rotations around the ω-angle due to hydrogen bonding may affect the dynamics of proteins. Structural changes in the protein backbone involve rotations of the ϕ, and ψ, angles and the flexibility of the ω-angle may influence these rotations, leading to more flexible loop regions and more rigid helical and β-sheet regions. In this way, the two-state nature of the peptide bond may effect the protein folding time scales.

In order to study the influences of the polarizability of both nonbonded electrostatic and bonded interactions using molecular dynamics simulations, a new potential model has to be developed. A number of polarizable potentials have been used for molecular dynamics simulations. These models are based on point inducible dipoles[10–24] or fluctuating atomic charges.[25–32] Some of these models have been constructed for the peptide group.[22,27,32] These models treat the nonbonded electrostatic interactions (either Coulombic or dipole–dipole) with an electric field inducing a polarization response on an atomic site. They are insufficient to treat the polarization response implied by Eq. (1), in which an electric field on one atom can change the nature of the chemical bond of other atoms. A model to treat this would have to be constructed not of individual atomic but of collective molecular polarizabilities. Here we present a two-state model, with each state corresponding to state A and B from Eq. (1). The potential is then given by a linear combination of A and B, with the coefficient for each state (one for each peptide bond) being a variational parameter, determined by minimizing the energy. The coefficients are updated each time step using an extended Lagrangian formalism.[26,33–35] Even though the underlying chemical processes are much different, the two-state model is similar in some respects to empirical valence bond (EVB) models for studying reactions in solution.[36–39] In the EVB approaches, two or more diabatic states are coupled to each other and to the solvent. As for this two-state model, the different EVB states may have different charges and bond parameters.

The results of the database analysis of the ω-angle distributions for the PDB and atomic resolution structures will be given in the next section. The two-state model will be described in the Sec. III. Section IV describes our results using the two-state model on the helix/coil transition of polyalanine. Section V summarizes our conclusions.

TABLE I. Atomic resolution structures used for structural analysis.

PDB code	Protein	Resolution (Å)	R factor
1CEX	cutinase	1.00	0.0940
1IRO	rubredoxin	1.10	0.0903
1IXH	phosphate-binding pr.	0.98	0.1140
1LKS	lysozyme (hen)	1.10	0.0989
1JSF	lysozyme (human)	1.15	0.115
3LZT	lysozyme (gallus)	0.92	0.0903
1NLS	agglutinin	0.94	0.1270
2ERL	mating pheromone er-1	1.00	0.1290
2FDN	ferredoxin (c. a.-urici)	0.94	0.1003
2IGD	igg-binding protein	1.10	0.0930
3CHB	cholera toxin	1.25	0.1326
2KNT	kunits inhibitor	1.20	0.1489
1RGE	ribonuclease	1.15	0.109
1CTJ	cytochrome c6	1.10	0.1397
1A7S	serine protease homolog	1.12	0.159
1A6G	carbonmonoxy-myoglobin	1.15	0.1284
1B0Y	electron transfer protein	0.93	0.1545
1BRF	iron-sulfur protein	0.95	0.132
1BXO	hydrolase	0.95	0.1004
1MFM	oxidoreductase	1.02	0.118
2PVB	calcium binding protein	0.91	0.1098
	aldose reductase	0.65	0.100

II. DATABASE ANALYSIS OF THE ω ANGLE DISTRIBUTION

For an analysis of the distribution of dihedral angles, three different data sets of protein structures from the protein data bank (PDB)[40] were generated: atomic, 2.2 Å and 3.4 to 3.6 Å resolution structures. The data sets were generated as of May 1999.

A. Atomic resolution protein database

All structures in the PDB were checked and were excluded if they met any of the following criteria.

(a) Refinements done with restraints.
(b) Entries with ''0'' as their first number, indicating an incomplete refinement.
(c) Resolution above 1.4 Å.
(d) Deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and peptides with uncommon amino acids (for example, vancomycin).
(e) For multiple entries, including new versions, complexes, different space groups, few amino acid mutants, different species or strain (less than 2% of sequence modification), a single entry was chosen according to highest resolution or lowest R factor or youngest version, in that order.
(f) Number of amino acids less than 40.

This gave the 21 structures listed in Table I. An additional structure, of aldose reductase, was made available by A. Podjany.

B. 2.2 Å resolution structures

This data set included all protein structures in the PDB with a resolution equal to 2.2 Å and with an R factor less than 17. For multiple entries, the latest version was chosen. This set includes 72 structures.
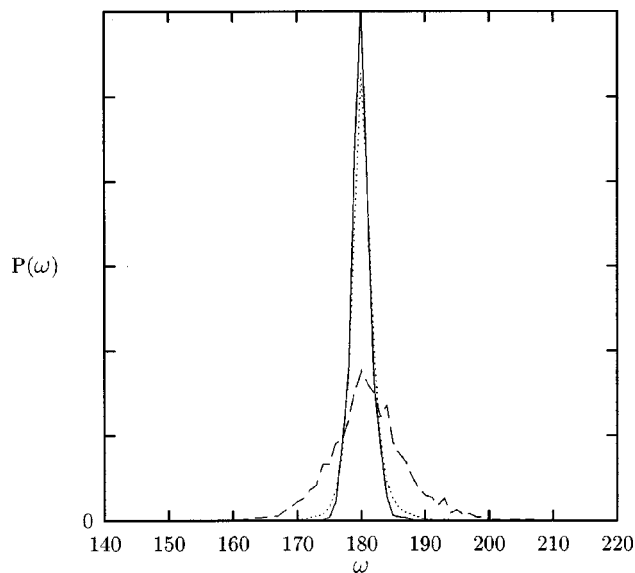
FIG. 1. Distribution of ω-angles for 3.6 Å (solid line), 2.2 Å (dotted line), and atomic (dashed line) resolution protein structures.

TABLE II. Statistics for the distributions of *trans* ω-angles from the data sets of different resolution for the various secondary structure motifs.

	Number of angles	Mean	Standard deviation
Atomic resolution			
Total	3546	181.0	6.10
α-helix	1102	180.1	4.12
Parallel β-sheet	333	180.3	6.01
Antiparallel β-sheet	505	184.8	6.69
Leftover	1606	180.6	6.61
2.2 Å resolution			
Total	13600	180.2	2.65
α-helix	3558	180.1	2.00
Parallel β-sheet	1778	179.7	2.08
Antiparallel β-sheet	2007	180.9	2.77
Leftover	6257	180.2	3.02
3.4 Å resolution			
Total	12610	180.0	1.81
α-helix	3944	179.9	1.34
Parallel β-sheet	1264	179.5	2.00
Antiparallel β-sheet	2186	180.2	1.95
Leftover	5216	180.0	1.97

C. 3.4 Å resolution structures

This data set included all protein structures in the PDB with a resolution between 3.4 and 3.6 Å and with an *R* factor less than 25. For multiple entries, the latest version was chosen. This set includes 16 structures.

The distribution of *trans* ω-angle, $P(\omega)$, for the three data sets are shown in Fig. 1. The $P(\omega)$ are all normalized. The distribution of ω angles is much broader for atomic resolution structures, which were refined without the use of restraints, than for both the 2.2 and 3.4 Å resolution structures, which were refined with restraints. The distribution from the atomic resolution data is similar to the distribution of ω-angles from peptides in the Cambridge structural database (CSD) and 11 atomic resolution PDB structures.[3,4] As has been pointed out, the distribution of ω angles from the lower resolution structures is strongly influenced by the refinement protocols.[3] Programs such as PROLSQ[41] and X-PLOR[42] place stiff restraints of 25 to 100 kcal/mol/deg to keep the peptide bond planar. Of the 3574 peptide bonds in the atomic resolution data set, 14, or 0.4%, are *cis*. Of the 2.2 and 3.4 Å resolution structures, only 0.03% are *cis*. A previous analysis has shown a trend for more *cis* groups the higher the resolution, particularly for proline residues.[43] This analysis has found 0.05% of nonproline ω angles were *cis*, less than the estimates of 0.1 to 1.5% from theoretical predictions and experimental structures of small peptide analogs.[43] These differences too may be due to refinement protocols.

There are now enough atomic resolution protein structures in the database to examine the ω-angle distributions for the different secondary structure units. For this analysis, a residue will be assigned a particular secondary structure if its ϕ and ψ angles are within 30° of the ideal value ($\phi,\psi=$ $-57,-47$ for the α-helix,$-139,135$ for an antiparallel β-sheet, and $-119,113$ for a parallel β-sheet). If a (ϕ,ψ) pair falls within $\pm30°$ of both types of β-sheet, then assignment is made according to which ideal ϕ,ψ values are closer. All

the residues that do not fall into one of these three categories are put in a leftover, or coil category. The mean and the standard deviation for the distributions are given in Table II. As illustrated in Fig. 1, the standard deviation of the atomic resolution structures is greater than the lower resolution structures which were refined using restraints and is close to value for the peptides in the Cambridge structural database of 5.9°.[3] However, the three data sets show some of the same trends. The standard deviations for the α-helical residues are all less than the overall standard deviations and the standard deviations for the antiparallel β-sheet and the leftover residues are all greater than the overall standard deviations. The mean ω-value for the antiparallel β-sheet residues is shifted towards values higher than 180°. The distributions from the atomic resolution data for the different secondary structure units are shown in Fig. 2 and 3. For the distribution of the
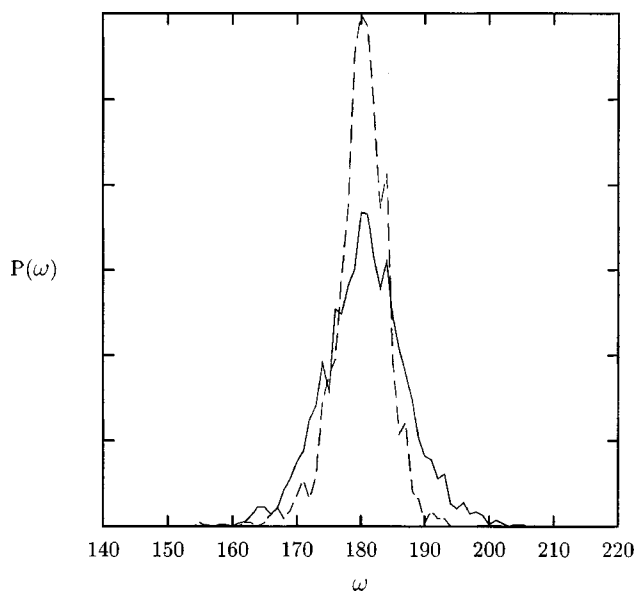

FIG. 2. Distribution of ω-angles for atomic resolution structures for α-helical (dashed line) and the leftover (solid line) residues.
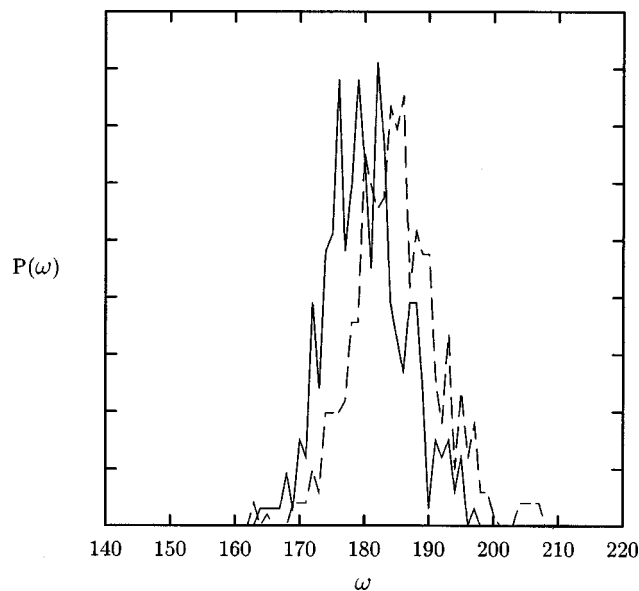
J. Chem. Phys., Vol. 112, No. 11, 15 March 2000

The nonplanarity of the peptide group   5233

FIG. 3. Distribution of ω-angles for atomic resolution structures for antiparallel β-sheet (dashed line) and parallel β-sheet (solid line) residues.

FIG. 4. Ramachandran diagram for the atomic resolution structures.

antiparallel β-sheet residues, there is a cluster of eight ω-values around 205°. Five of these values come from the structure of the cholera toxin (3*CHB*), a pentamer. All the rest of the ω-values greater than 20° from 180° in the leftover category. This group totals 11 and ranges from 136° to 223°. These 11 residues are all in loop regions and near the surface of the protein. There is no correlation between the deviations from planarity and the crystallographic B-factors in these structures.

It is significant that not only the width of $P(\omega)$ depends on secondary structure, but that ω-angles can be far from planar. Most of the angles are in the region near 180° but for the coil residues, there is a broad flat part of the distribution going out to $180° \pm 50°$. As more atomic resolution structures become available, we will see if the shapes of the ω-angle distributions remain the same. Based on the structures which are available now, it appears that an accurate description for loop regions and antiparallel β-sheets should include the possibility of largely nonplanar peptide bonds. The secondary structure should also induce in small changes in the C–N and C–O bond lengths, but the resolution for most currently available atomic resolution structures is not high enough to resolve this level of detail.

The Ramachandran plots for the 3560 ϕ,ψ pairs from the atomic resolution data are shown in Fig. 4. Figure 5 shows 3560 randomly chosen ϕ,ψ pairs from the 2.2 Å resolution data. The ϕ,ψ values from the atomic resolution data set are more closely clustered around the normally allowed regions of the Ramachandran diagram. Those from the 2.2 Å data are more widely distributed. This is the opposite trend than that of the distribution of ω-angles. The restraints used in the refinement process for the lower resolution structures forces the ω-angles to be more planar than those observed atomic resolution structures. The stress imposed in the structure by the artificially high restraint affects other parts of the structures. The refinement process overconstraining of the
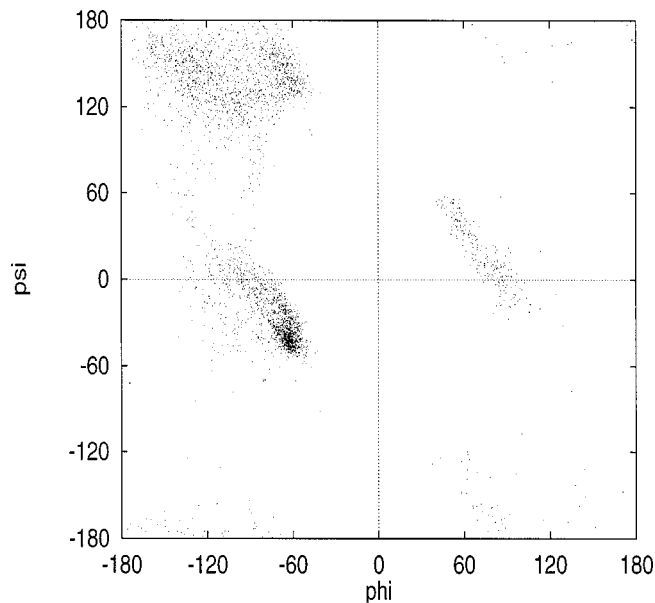
ω-angle is compensated by the broadening of the ϕ and ψ distributions.

The results of this section show that the properties of the peptide bond depend strongly on the local structure. In the next section, we will present a new potential in which the charge distribution and the dihedral flexibility respond to changes in the environment.

III. THE TWO-STATE MODEL

A. The potential

The potential model is taken to be a sum of two states, one corresponding to A and B from Eq. (1). Each state has associated with it a set of charges for the C_α, C, N, and H atoms and a dihedral force constant for the ω-angle. All other
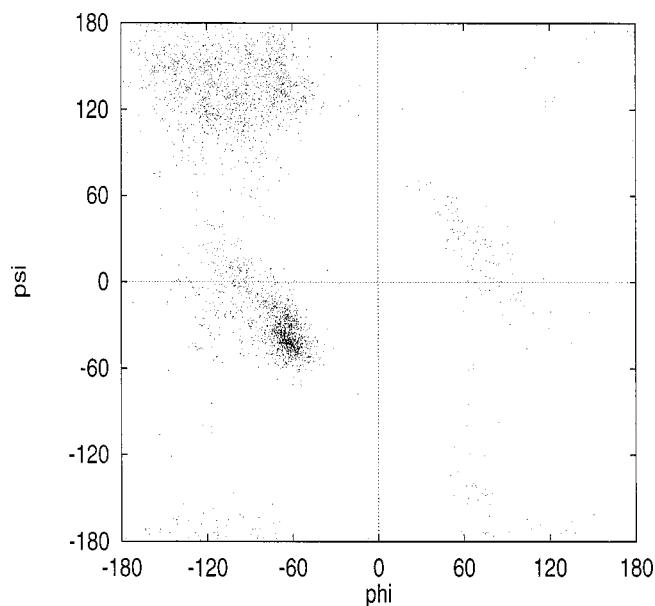

FIG. 5. Ramachandran diagram for the 2.2 Å resolution structures.

potential parameters, including the charges for the other atoms, are the same for both states. Each peptide bond group, i, has a coefficient for each state, C_{iA} and C_{iB}. Thus the dihedral force constants are given by

$$V_i = C_{iA}^2 V_A + C_{iB}^2 V_B, \tag{2}$$

and the charges are given by

$$q_{i,\alpha} = C_{iA}^2 q_{\alpha,A} + C_{iB}^2 q_{\alpha,B}, \tag{3}$$

with α indicating the atom type. The two states are coupled by an energy, $C_{iA} C_{iB} E_{AB}$, with the coupling E_{AB} assumed to be independent of geometry. The interaction model is based on standard potential functions.[44] The equations below are based directly on the versions used by AMBER4.1.[45] The interaction energy for a N_{res} residue peptide, each residue with N_i atoms, is given by

$$E = \sum_{i=1}^{N_{res}} \sum_{\alpha=1}^{N_i-1} \sum_{\beta=\alpha+1}^{N_i} U_{LJ}^{\alpha\beta}(r_{i\alpha i\beta}) + C_{iA}^2 q_{\alpha,A} q_{\beta,A}/\epsilon_0 r_{i\alpha i\beta} + C_{iB}^2 q_{\alpha,B} q_{\beta,B}/\epsilon_0 r_{i\alpha i\beta}$$

$$+ \sum_{i=1}^{N_{res}-1} \sum_{j=i+1}^{N_{res}} \sum_{\alpha=1}^{N_i} \sum_{\beta=1}^{N_j} U_{LJ}^{\alpha\beta}(r_{i\alpha j\beta}) + (C_{iA}^2 q_{\alpha,A} + C_{iB}^2 q_{\alpha,B})(C_{jA}^2 q_{\beta,A} + C_{jB}^2 q_{\beta,B})/\epsilon_0 r_{i\alpha j\beta}$$

$$+ \sum_{dihedrals} \frac{V_i}{2}(1 + \cos(n\phi - \gamma)) + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_i C_{iA}^2 \Delta E + C_{iA} C_{iB} E_{AB}, \tag{4}$$

where $r_{i\alpha j\beta} = |\mathbf{r}_{i\alpha} - \mathbf{r}_{j\beta}|$, ϵ_0 is the vacuum permittivity, and $U_{LJ}^{\alpha\beta}$ is the Lennard-Jones potential,

$$U_{LJ}^{\alpha\beta}(r_{i\alpha j\beta}) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{i\alpha j\beta}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{i\alpha j\beta}}\right)^6\right],$$

characterized by a well depth, $\epsilon_{\alpha\beta}$ and diameter $\sigma_{\alpha\beta}$. The Lennard-Jones and electrostatic interactions are only between atoms on the same molecule which are separated by at least three bonds. The interactions of atoms separated by three bonds (1–4 interactions) are reduced by a scale factor of 1/2 for the Lennard-Jones and 1/1.2 for the electrostatic interactions.[45] The first term in Eq. (4) describes the interactions between atoms in the same peptide bond group. The charge–charge interactions on a peptide group are only between charges of the same state, so that the only cross term mixing C_{iA} and C_{iB} involves E_{AB}. The second term in Eq. (4) describes the interactions between different residues.

The dihedral energy terms are given by the Fourier terms in the third line of Eq. (4). The terms for the ω-angle dihedral are given by $n=2$ and $\gamma=2\pi$. This describes four dihedral angles for rotations around the C–N bond, for each of four atom groups that can be made from one of the two atoms bonded to the C atom (H,C_α), the C atom, the N atom, and one of the two atoms bonded to the N atom (O,C_α). These dihedrals are polarizable, by Eq. (2). There is another term for the dihedral H–C–N–O, with $n=1$ and $\gamma=0$, which is adjusted to give a *cis/trans* energy difference of about 2 kcal/mol.[45] This torsional term is taken to be the same for both states. The bond angle force constants, K_θ, and equilibrium angles, Θ_{eq}, are taken to be state-independent. The term ΔE on the last line of Eq. (4) is a constant added to the energy of state A to bring the energies of the two states in balance. Bond length terms could also be included, however the calculations reported here are for rigid bonds, so these energies are not given in Eq. (4).

In summary, the differences between state A and B are defined in terms of four charges and one dihedral angle. The description of polarizability has introduced no new interactions, it just makes the interactions that are present in standard force fields responsive to changes in the environment. Therefore, the model should not be any more computationally expensive than conventional models. This is a property it shares with the fluctuating charge polarizable model.[26] It is straightforward to extend the model to include other properties. For instance, the bond angle constants around the nitrogen atom may change to make nitrogen pyramidalization easier for state A.[46] Equilibrium bond distance parameters may also change. Atomic polarizabilities could be added to the model, which may be different for the two states. The effects of the inclusion of some of these modifications will be the subject of future studies.

The coefficients, $C_{i\alpha}$, are found by minimizing the energy subject to the constraint that $C_{iA}^2 + C_{iB}^2 = 1$. If there are no interactions between the peptide groups, then the equations for the coefficients decouple and we have a simple expression for the energy of each peptide group,

$$E_i = \frac{C_{iA}^2 E_{iA} + C_{iB}^2 E_{iB} + C_{iA} C_{iB} E_{AB}}{C_{iA}^2 + C_{iB}^2}, \tag{5}$$

where E_{iA} is the Coulombic and dihedral energy of peptide group i in state A and E_{iB} is the energy for state B. E_i has two extrema, given by

$$E_{i\pm} = \tfrac{1}{2}\left(E_{iA} + E_{iB} \pm \sqrt{(E_{iA} - E_{iB})^2 + E_{AB}^2}\right). \tag{6}$$

The minimum energy (E_{i-}) coefficients are

$$C_{iA} = E_{AB}/\sqrt{4(E_{iA} - E_{i-})^2 + E_{AB}^2}, \tag{7}$$

and $C_{iB} = -\sqrt{1 - C_{iA}^2}$. ($E_{AB}$ is assumed to be ≥ 0, so the minimum energy coefficients have the opposite sign.) The coefficients for the interacting system can be found iteratively. Initial estimates of all the coefficients can be made

and the energies E_{iA} and E_{iB} can be calculated for all i. Then the values of C_{iA} can be found through Eq. (7). With these new coefficients, E_{iA} and E_{iB} can be recalculated and new coefficients can be found. This procedure can be iterated until converged. We found that convergence to five decimal places occurs in about five steps.

B. Extended Lagrangian dynamics

The extended Lagrangian method provides an efficient way to update the coefficients along with the molecular dynamics simulations. This method provides a basis for simulations in different ensembles,[33,34] Car–Parrinello *ab initio* molecular dynamics,[35] and simulations with polarizable potentials.[14,26,47] The coefficients are treated as dynamical variables by introducing fictitious kinetic energies into the Lagrangian. The extended Lagrangian is

$$L = \frac{1}{2}\sum_{i\alpha} m_{i\alpha}\dot{\mathbf{r}}_{i\alpha}^2 - E(\{\mathbf{r}\},\{C\}) + \frac{1}{2}\sum_i m_C \dot{\mathbf{C}}_i^2, \qquad (8)$$

where $m_{i\alpha}$ is the atomic mass and m_C is a fictitious mass for the coefficients, having units of energy time.[2] The coefficients evolve in time according Newton's equation

$$m_C \ddot{C}_{i\alpha} = -\frac{\partial U}{\partial C_{i\alpha}}, \qquad (9)$$

for $\alpha = A$ and B. The mass, m_C, is chosen to be small enough so that the coefficients respond quickly to changes in the nuclear degrees of freedom and so we remain on the ground state adiabatic potential surface. However, m_C should be large enough so that a small time step does not have to be used. The constraint $C_{iA}^2 + C_{iB}^2 = 1$ can be enforced with an algorithm like SHAKE[48] as is done for the set of coefficients in Car–Parrinello dynamics.[49] However for our model, it is easy to replace C_{iB} by $-\sqrt{1-C_{iA}^2}$ and just propagate C_{iA}. The coupling term, E_C becomes $-E_{AB}C_{iA}\sqrt{1-C_{iA}^2}$, leading to a force on the coefficients

$$\frac{\partial E_C}{\partial C_{iA}} = E_{AB}\frac{(2C_{iA}^2-1)}{\sqrt{1-C_{iA}^2}}, \qquad (10)$$

which diverges as C_{iA} approaches one. This divergence can be removed by replacing C_{iA} by $\cos(\phi_i)$ and C_{iB} by $-\sin(\phi_i)$ and treating ϕ_i as the dynamical variable. This gives a different Lagrangian,

$$L = \frac{1}{2}\sum_{i\alpha} m_{i\alpha}\dot{\mathbf{r}}_{i\alpha}^2 - E(\{\mathbf{r}\},\{\phi\}) + \frac{1}{2}\sum_i \sum_i m_C \dot{\phi}_i^2, \quad (11)$$

and ϕ_i is much easier to propagate than C_{iA}.

C. The polarizability tensor

The dipole moment of the peptide group is given by

$$\boldsymbol{\mu} = C_A^2 \boldsymbol{\mu}_A + C_B^2 \boldsymbol{\mu}_B, \qquad (12)$$

where $\boldsymbol{\mu}_\alpha$ is the dipole moment of state α. In the presence of an electric field, $\mathbf{E}$, the energy of each state is

$$E_{i\alpha} = E_{i\alpha}^0 - \boldsymbol{\mu}_\alpha \cdot \mathbf{E}, \qquad (13)$$

where $E_{i\alpha}^0$ is the energy of state α in the absence of the electric field. Putting Eq. (13) into Eq. (7) gives

$$C_A^2 = \tfrac{1}{2}E_{AB}^2 / (\Delta E_{AB}^2 + E_{AB}^2 - 2\Delta E_{AB}\Delta_{\boldsymbol{\mu}\cdot\mathbf{E}} + (\Delta\boldsymbol{\mu}\cdot\mathbf{E})^2$$
$$+ \Delta E_{AB}\Delta\boldsymbol{\mu}\cdot\mathbf{E}\sqrt{(\Delta E_{AB}-\Delta\boldsymbol{\mu}\cdot\mathbf{E})^2 + E_{AB}^2}), \qquad (14)$$

where $\Delta E_{AB} = E_A - E_B$ and $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_A - \boldsymbol{\mu}_B$. Using

$$\sqrt{(\Delta E_{AB}-\Delta\boldsymbol{\mu}\cdot\mathbf{E})^2 + E_{AB}^2}$$
$$= \sqrt{\Delta E_{AB}^2 + E_{AB}^2 - 2\Delta E_{AB}\Delta\boldsymbol{\mu}\cdot\mathbf{E} + (\Delta\boldsymbol{\mu}\cdot\mathbf{E})^2}$$
$$\approx \sqrt{\Delta E_{AB}^2 + E_{AB}^2}\Big(1 - \tfrac{1}{2}(2\Delta E_{AB}\Delta\boldsymbol{\mu}\cdot\mathbf{E} - (\Delta\boldsymbol{\mu}\cdot\mathbf{E})^2)/$$
$$(\Delta E_{AB}^2 + E_{AB}^2)\Big),$$

and keeping only the terms linear in the electric field leads to

$$C_A^2 = \frac{C_A^{0^2}}{1 - A\Delta\boldsymbol{\mu}\cdot\mathbf{E}} \approx C_A^{0^2}(1 + A\Delta\boldsymbol{\mu}\cdot\mathbf{E}), \qquad (15)$$

where $C_A^{0^2}$ is the value of C_A^2 in the absence of the field

$$C_A^{0^2} = \tfrac{1}{2}E_{AB}^2 / (\Delta E_{AB}^2 + E_{AB}^2 + \Delta E_{AB}\sqrt{\Delta E_{AB}^2 + E_{AB}^2}), \quad (16)$$

and

$$A = \frac{2C_A^{0^2}}{E_{AB}^2} \frac{(2\Delta E_{AB}^2 + E_{AB}^2 + 2\Delta E_{AB}\sqrt{\Delta E_{AB}^2 + E_{AB}^2})}{\sqrt{\Delta E_{AB}^2 + E_{AB}^2}}. \quad (17)$$

The value of C_A^0 from Eq. (16) is identical to that from Eq. (7). Inserting Eq. (15) into Eq. (12) gives

$$\boldsymbol{\mu} = \boldsymbol{\mu}^0 + (\boldsymbol{\mu}_A - \boldsymbol{\mu}_B)AC_A^{0^2}(\boldsymbol{\mu}_A - \boldsymbol{\mu}_B)\cdot\mathbf{E}, \qquad (18)$$

where $\boldsymbol{\mu}^0$ is the dipole moment with no field. The polarizability tensor, α, is then

$$\boldsymbol{\alpha} = (\boldsymbol{\mu}_A - \boldsymbol{\mu}_B)AC_A^{0^2}(\boldsymbol{\mu}_A - \boldsymbol{\mu}_B). \qquad (19)$$

For the planar peptide group, the polarizability is zero out of the plane, but there are nonzero polarizabilities in the plane and also there are nondiagonal components in the plane.

D. Potential parameters

The two-state model is added to an existing potential for proteins, the AMBER4.1 potential.[45] The two-state model requires ten additional parameters. For each state there are four additional charges for the C_α, C, N, and H atoms, but since these four charges must sum up to the same value as the four charges from the AMBER4.1 potential to preserve charge neutrality this only introduces three additional parameters for each state. There are the two parameters E_{AB} and ΔE. Another two parameters are the $n=2$ dihedral force constant, V_2^A and V_2^B [see Eq. (2)]. The $n=1$ dihedral force constant, V_1^α, is taken to be the same for both states and is adjusted to give a *cis/trans* energy difference equal to 2 kcal/mol. For the $n=2$ dihedral parameters, we will explore three different choices: one with V_2^A set equal to a value typical of a single bond and V_2^B to a value typical of a double bond (model 1), another with V_2^A equal to zero (model 2), and a third with both set equal to the AMBER4.1 value (model 3).

TABLE III. Potential parameters for the two-state models and the equivalent values in AMBER4.1 (Ref. 45). Charges, q_α, in units of e and dihedral force constants, V_i, in units of kcal/mol.

| | Model | Two-state | | AMBER4.1 |
		A	B	
q_C	1–3	0.5600	0.5850	0.5972
q_O	1–3	−0.3100	−0.7450	−0.5679
q_N	1–3	−0.5150	−0.3000	−0.4157
q_H	1–3	0.1505	0.3455	0.2719
V_1	1–3	1.0	1.0	2.0
V_2	1	5.0	20.0	10.0
V_2	2	0.0	25.0	10.0
V_2	3	10.0	10.0	10.0

TABLE IV. ΔE values for the three models with different dihedral force constants, for the Ace–(Ala)$_n$–Nme peptide, in kcal/mol.

| Model | Peptide bond position | | |
	N-terminal	Interior	C-terminal
1	−43.5	−39.3	−24.0
2	−43.1	−38.5	−23.5
3	−44.2	−40.3	−24.5

We picked the electrostatic polarization parameters to reproduce the data on the enhancement of the dipole moment upon helix formation,[6] the *ab initio* data on the nonadditivity on hydrogen bond formation of NMA,[8] and the experimental polarizability. Also, parameters were chosen so that the potential energy, on average, is close to the AMBER4.1 potential. This simplifies comparisons between the two models and will keep the two-state model compatible with conventional force fields, so that, for example, interactions with solvent molecules will be similar. Properties of both NMA and polyalanine were used to parameterize the model. However, since the focus of the present study is on polypeptides, the polyalanine properties were weighted more heavily. It should be noted that the nonadditivity, or cooperativity, of hydrogen bond formation is not due solely to polarizability. The cooperative energy is defined as[8,9]

$$\Delta E_{coop} = E_{I \cdot II \cdot III} - E_{I \cdot II} - E_{I \cdot III}, \qquad (20)$$

where the I, II, and III refer to three hydrogen bonded molecules, arranged in a line with I at the center (II$\cdots$I$\cdots$III). $E_{I \cdot II \cdot III}$ is the energy of the trimer and $E_{I \cdot II}$ and $E_{I \cdot III}$ are the energies of the dimers. All the energies in Eq. (20) are for the minimized geometry for each complex. A measure of nonadditivity through electronic polarization alone would be given by

$$\Delta E_{pol} = E_{I \cdot II \cdot III} - E^*_{I \cdot II} - E^*_{I \cdot III} - E^*_{II \cdot III}, \qquad (21)$$

where the star superscript denotes the dimer energies of the dimers in the same geometry as they are in the trimer. In other words, the addition of III to the I$\cdots$II dimer may induce a change in the I$\cdots$II orientation. For nonpolarizable potentials, Eq. (21) is zero, but Eq. (20) is not. In our calculation of the NMA trimer (with the two-state model) the difference between $E^*_{I \cdot II} + E^*_{I \cdot III}$ and $E_{I \cdot II} + E_{I \cdot III}$ is about 0.1 kcal/mol. The hydrogen bonds for the trimer are about 0.05 Å shorter. A bigger effect is the neglect of $E^*_{II \cdot III}$, the interaction energy of the molecules on the opposite side of the central molecule I. Our calculations show this energy to be about −0.4 kcal/mol (−6.0 kcal/mol using AMBER4.1) so some of ΔE_{coop} is due to energy gained by the II$\cdots$III interaction.

The parameters for the charges and dihedral force constants are listed in Table III. The coupling term, E_{AB}, is set equal to 16.0 kcal/mol. The parameter ΔE is chosen so that for the isolated *trans* NMA molecule or a peptide in an extended conformation (with no hydrogen bonds) the value of C_A^2 is 0.60, or the peptide bond has 60% single bond character in agreement with Pauling's estimate.[1] For NMA, ΔE equals −30.0 kcal/mol. The same set of parameters gives $C_A^2 = 0.45$ for *cis* NMA so the *cis* peptide bond has more double bond character. In the *cis* conformer, state B is stabilized by electrostatic interactions between the oxygen and amide hydrogen atom. For the polypeptide, three different changes in the torsional force constants are examined, in order to calculate the effect of torsional flexibility. For the best comparison of different V_2^A, V_2^B pairs, the electrostatic interactions should be equivalent. This requires using different values of ΔE for the three models. The narrower the torsional potential of state B relative to state A, the more this will shift the equilibrium towards state B so the values of ΔE need to be adjusted. In addition, for the polypeptide with C-terminal and N-terminal methyl blocking groups [Ace-(Ala)$_n$–Nme, Ace is acetyl and Nme is it N-methyl] different values of ΔE are needed for the first, the interior, and the last peptide bond, as listed in Table IV.

The properties of the two-state model, compared to the values from AMBER4.1, *ab initio* calculations, and experiment are shown in Table V. All the properties for polyalanine are for model 1 (see Table IV). The dipole moment for NMA, 3.42 D, is the value the two-state model gives with $C_A^2 = 0.6$. The trace of the polarizability tensor, α, is less than the experimental value because it lacks the component out of the plane and also because it lacks the polarization response of the two methyl groups. In this last respect, the polarizability of the two-state model is perhaps more comparable to the experimental value of formamide, 4.08 D.[52] The dimer and trimer energies for NMA are for antiparallel molecules, with a linear hydrogen bond (see Fig. 1 of Ref. 8). FEC stands for fully extended chain, the configuration of the polypeptide with ϕ and ψ near π and $E_{\alpha-helix} - E_{FEC}$ is the energy difference of the minimized geometries of the two configurations. The average dipole moment of the alanine residues for (Ala)$_{12}$ is given by $\mu_{residue}$. For AMBER4.1, $\mu_{residue}$ differs slightly between the α-helix and the FEC due to orientational changes. For the two-state model, this difference is greater due to polarization effects and is similar to the enhanced dipole moment reported in Refs. 6 and 7.

E. Simulation details

The simulations were performed with bond constraints enforced using the SHAKE algorithm,[48] a 1 fsec time step, and $m_c = 23.8$ fsec^2 kcal/mol. All simulations are done at constant temperature (at 300 K), using a Nosé–Hoover temperature

TABLE V. Properties of *trans* N-methylacetamide (NMA) and polyalanine [Ace–(Ala)$_N$–Nme] comparing the two-state model, AMBER4.1 (Ref. 45), *ab initio* calculations, and experimental values.

	Two-state model	AMBER4.1	*ab initio*	Experimental
NMA				
Dimer energy (kcal/mol)	−7.26	−8.06	−7.5[a]	
Trimer energy (kcal/mol)	−17.85	−16.75	−17.1[a]	
ΔE_{coop}	−3.31	−0.63	−2.1[a]	
μ_A (debye)	1.35			
μ_B (debye)	6.57			
μ (debye)	3.38	4.32	3.78[b]	3.73[c]
α (Å^3)	4.68	0.0		7.82[d]
polyalanine				
$E_{\alpha\text{helix}} - E_{\text{FEC}}$ (kcal/mol)				
$N=9$	−49.5	−50.7		
$N=12$	−75.9	−76.1		
$N=15$	−104.7	−102.2		
μ_{residue}, FEC (debye)	3.34	3.30	3.29[e]	3.5[f]
μ_{residue}, α-helix (debye)	4.68	3.69	5.0[e]	4.8–5.0[f]

[a]Reference 8.
[b]Reference 50.
[c]Reference 51.
[d]Reference 52.
[e]Reference 7.
[f]Reference 6.

bath with a mass for the Nosé variable equal to 0.994 kcal/mol psec^2.[34,53,54] At the beginning and during the simulation at 10 ps intervals, the exact set of coefficients are found using the iterative procedure described below Eq. (7). Restraints to keep the peptide in a FEC configuration are done by replacing the $n=2$ backbone torsional potential for the ϕ and ψ angles, $(V/2)[1+\cos(2X-180)]$, by the quadratic form with the same second derivative at $X=180°,2V(X-180)^2$. Restraints to keep the peptide in an α-helix are done by placing a one-sided harmonic restraint with a force constant of 6 kcal/mol/Å^2 on the 1–4 hydrogen bonds if they exceeded 2.7 Å.[55]

IV. APPLICATION OF THE TWO-STATE MODEL TO THE HELIX–COIL TRANSITION

Using the model described in Sec. II, we will examine the helix–coil transition for (Ala)$_N$ in the gas-phase. Simulations in the gas-phase, rather than in solution, allow us to focus on protein–protein interactions. More importantly, the folding processes occur on a time scale amenable to direct simulation at 300 K, so we can perform a number of simulations with different models for the peptide group to see how this affects the transition. There have been a number of simulations of conformational changes of polypeptides of 4 to 30 residues in the gas-phase.[56–65] These studies are generally of two types: free energy calculations using umbrella sampling[55,65] or thermodynamic integration,[61,62] and dynamical simulations which examine conformational space starting from an initial structure [either an α-helix (Refs. 56, 57, 59, 60, 63) or a fully extended chain (Ref. 64)]. These simulations and simulations including solvent, together with other theoretical approaches have increased our understanding of the process of protein folding (see Refs. 66 and 67, and references therein). In addition, recent experiments have begun to examine the folding of unsolvated proteins.[68–71]

A residue can be defined as α-helical either if the ϕ, ψ values are near the ideal α-helix values or if the 1–4 hydrogen bond is made. We choose to use the 1–4 hydrogen bond criteria since it is a stricter definition of helicity.[59] An order parameter for each residue is defined as

$$\sigma_i = \begin{cases} 1 & \text{if } r_{iO,i+3H} < r_{\text{cut}} \\ 0 & \text{otherwise} \end{cases}, \tag{22}$$

where $r_{iO,i+3H}$ is the distance between the oxygen atom on residue i and the hydrogen atom on residue $i+3$, and r_{cut} is 2.7 Å. For the entire peptide of N residues, the fraction helix is given by

$$\text{fraction helix} = \frac{1}{N-2} \sum_{i=0}^{N-3} \sigma_i, \tag{23}$$

and the sum starts at $i=0$ to include the oxygen atom on the acetyl blocking group. The numerous small polypeptide simulations, indicate that, unlike larger proteins, small polypeptides to not have a single conformational state corresponding to the free energy minimum. Circular dichroism experiments on small helix forming peptides also indicate that the peptide is about 75% helical.[72] In the analyses below, we will assign a structure as being in an α-helix structure if it has a fraction helix of 0.5 or greater. We will show results for two studies which start from different initial structures, an α-helix and a fully extended chain. It should be emphasized at this point that the helix–coil transition involves changes in the ϕ and ψ angles, which are not modified from the standard AMBER4.1 force field. Therefore, we are examining how the flexibility of the ω angle torsions, and also the polarizability of the charges, influences the ϕ and ψ torsional transitions.

Two sets of simulations are done, starting from different initial structures. One set, to examine equilibrium properties,

TABLE VI. Average value of the coefficient, C_A^2, for the simulations beginning in an α-helix for $(Ala)_N$ with $N=9,12$, and 15. Numbers in parenthesis indicate error estimates.

N	Model	All residues	$\langle C_A^2 \rangle$ helical	Nonhelical
9	1	0.51(1)	0.47(1)	0.54(1)
	2	0.50(2)	0.46(2)	0.52(3)
	3	0.51(1)	0.48(1)	0.55(1)
12	1	0.44(1)	0.42(1)	0.50(1)
	2	0.42(1)	0.40(1)	0.49(1)
	3	0.46(1)	0.44(1)	0.51(1)
15	1	0.38(1)	0.36(1)	0.46(1)
	2	0.35(3)	0.34(1)	0.44(1)
	3	0.39(1)	0.38(1)	0.47(1)

TABLE VII. Free energy difference, ΔG, energy difference, ΔE, and entropy difference, $T\Delta S$, between the α-helix and coil states for four different models of the peptide bond group, in kcal/mol and fraction helix.

N	Model	ΔG	ΔE	$T\Delta S$	Fraction helix
9	1	0.5(2)	−2.4(1)	−2.9(4)	0.37(6)
	2	0.6(1)	−1.8(1.0)	−2.4(1.1)	0.35(3)
	3	0.5(2)	−2.4(2)	−2.9(4)	0.37(6)
	4	0.3(3)	−2.5(2)	−2.8(4)	0.37(8)
12	1	−1.2(1)	−5.2(1)	−4.1(2)	0.67(1)
	2	−1.27(4)	−5.4(1)	−4.1(1)	0.68(1)
	3	−1.08(3)	−4.9(1)	−3.8(1)	0.66(1)
	4	−1.61(7)	−5.1(1)	−3.4(2)	0.71(1)
15	1	−2.8(1)	−9.0(3)	−6.2(4)	0.82(1)
	2	−3.0(1)	−9.5(6)	−6.4(6)	0.83(1)
	3	−2.56(7)	−8.5(2)	−5.9(2)	0.81(1)
	4	−2.6(2)	−6.7(1)	−4.1(3)	0.79(1)

starts in an ideal α-helix structure and another, to examine dynamical properties, starts in an unstable configuration.

A. Beginning from an α-helix

We begin these simulations starting with an ideal α-helix geometry. Initial velocities are assigned from a Maxwell–Boltzmann distribution with a temperature of 300 K. Equilibration consisted of 10 ps with α-helical restraints as described previously followed by an additional 10 ps of unrestrained dynamics. Data was collected over 20 nanoseconds. This procedure was repeated five times with a different set of initial velocities for 9, 12, and 15 residue polyalanine using the three parameters given Sec. II plus a nonpolarizable model.

Table VI gives the average values for the coefficient of state A squared, $\langle C_A^2 \rangle$. The parentheses in Table VI and the following tables give 95% confidence intervals. Averages are shown for all residues and also for residues in an α-helical geometry and those not in a helix. In general, those residues in an α-helix have about 10% more double bond character and the amount of double bond character (as given by $\langle C_B^2 \rangle = 1 - \langle (C_A^2) \rangle$) increases with the size of the peptide. The results from Table VI suggest an additional model, in which the coefficients kept fixed to an average value which we will take from the $(Ala)_{12}$ results for model 1. This makes a nonpolarizable model, which has the potential parameters of model 1 (see Tables III and IV) but with $C_A = \sqrt{0.44}$ for all residues. This is simply AMBER4.1 with slightly different parameters for the peptide group charges and dihedral force constant ($V_2 = 13.4$ kcal/mol). This model will allow us to examine how deviations from the average change the structure and dynamics.

The free energy difference between the helix and coil configurations can be found from

$$\Delta G = -kT \ln \frac{\langle N_H \rangle}{\langle N_C \rangle}, \tag{24}$$

where $\langle N_H \rangle$ is the average number of configurations for which the fraction helix is greater than 0.5, and $\langle N_C \rangle$ is number of configurations in the coil state $(1 - \langle N_H \rangle)$.[60] The energy difference can be found from $\Delta E = \langle E_H \rangle - \langle E_C \rangle$, where $\langle E_H \rangle$ and $\langle E_C \rangle$ are the average potential energies in the helix and coil states, respectively. The entropic difference is $T\Delta S = \Delta E - \Delta G$. In Table VII, ΔG, ΔE and $T\Delta S$ are listed

for the four potential models (model 4 is model 1 with C_A fixed at $\sqrt{0.44}$). The polarizable models (1–3) show a constant decrease in ΔG as a function of N. For the polarizable models, the change in $\Delta G(\Delta\Delta G = \Delta G_N - \Delta G_{N-3})$ is about 1.7 kcal/mol or about 0.6 kcal/mol per residue. For the nonpolarizable model, $\Delta\Delta G = -1.9$ kcal/mol for $N=12$ and -1.0 kcal/mol for $N=15$, so the free energy gain decreases for larger peptides. The differences between the polarizable models and the nonpolarizable model are due to differences in the energy and entropy changes. For the polarizable models, the $\Delta\Delta E$ decreases from about -3 kcal/mol at $N=12$ to -4 kcal/mol at $N=15$ and $T\Delta\Delta S$ increases from 1 to 2 kcal/mol. For nonpolarizable model, $\Delta\Delta E$ increases from -2.6 to -1.6 kcal/mol and the entropic part stays constant at about 0.6 kcal/mol. Among the polarizable models, a larger difference in the dihedral force constant between the two states (see Table III) causes a larger entropy decrease, since the formation of an alpha helix causes a shift towards a tighter ω-angle force constant. Also given in Table VII is the fraction helix, which increases with the number of residues.

The free energy data can be used to estimate Zimm–Bragg parameters. In the Zimm–Bragg theory, the equilibrium constant for propagating an α-helix by one residue is given by s and the equilibrium constant for initializing a helix—forming a helix of one residue or 1–4 hydrogen bond—is given by σs.[73] In this model, helix formation begins with a difficult initiation step followed by a series of helix propagation steps with the same equilibrium constant. The theory then assumes the equilibrium constants for both propagation and initiation are independent on the peptide length. The equilibrium constant for propagation was a constant for the polarizable models, but not for the nonpolarizable model. The free energy difference between coil and helix for N-residues is then

$$\Delta G_N = -kT \ln \sigma - (N-2)kT \ln s. \tag{25}$$

The free energy increases as $(N-2)kT \ln(s)$, since it takes at least three residues to form the first 1–4 hydrogen bond with the oxygen atom on the acetyl blocking group. The data from Table VIII can be used to estimate σ and s (see Table VIII). The nucleation parameter, σ, shows a strong dependence on the flexibility of the ω-torsion of state A. Flexibility makes it

J. Chem. Phys., Vol. 112, No. 11, 15 March 2000

The nonplanarity of the peptide group   5239

TABLE VIII. Zimm–Bragg parameters, s and σ, and helix formation rate constants for $(AlA)_{12}$ in 1/psec.

Model	s	σ	k
1	2.5(2)	0.007(5)	0.000 23
2	2.7(1)	0.0004(2)	0.000 32
3	2.3(1)	0.0016(7)	0.000 18
4	2.1(2)	0.010(9)	0.000 34

FIG. 6. Values of C_A^2 (solid line, left axis), fraction helix (dotted line, left axis) and energy (dashed line, right axis) vs. time, t, for trajectories starting in a fully extended chain geometry. The average values for the simulations starting in an α-helix geometry are given for C_A^2 (square), fraction helix (triangle), and energy (diamond).

more difficult to initially form a helix and the model with the smallest V_2^A (model 2) has the smallest σ. The more rigid models (model 3 with $V_2^A = 10$ kcal/mol and model 4 with $V_2^A = 13.4$ kcal/mol) have the largest σ-values. The values of the propagation parameter, s, are greater for the polarizable models than for the nonpolarizable models (model 4). Nonadditive hydrogen bond strengths make the formation of an additional helical residue more favorable. The values of s and σ are similar to other calculated values, both from gas-phase and aqueous studies.[55,57,59–61] Values of s are mostly in the range of 1 to 3, with slightly smaller values for the calculations done with water. Values of σ are more widely distributed. They vary over orders-of-magnitude and depend strongly on the method used to calculate them. Experimental estimates for polyalanine, in water, are 0.002 to 0.008 for σ and 1.1 to 2.2 for s.[74–76]

B. Beginning from a fully extended chain

An additional set of simulations for 12 residue polyalanine starts from a fully extended chain (FEC) conformation, with ϕ and ψ equal to 180°. In these simulations, equilibration consisted of 10 ps with restraints to keep the configuration in a FEC configuration as described at the end of Sec. II. After equilibration, the restraints are removed and the peptide is allowed to fold. Five hundred trajectories starting from different initial velocities were run until they formed a helix or until they reached 2.0 ns, whichever came first. Of these, two hundred were run for at least 500 ps in order to calculate averages over this length of time. Constant temperature dynamics is necessary to avoid heating since the energy can decrease by as much as -76 kcal/mol during the transition (see Table V). If half this energy goes into kinetic energy, then the temperature will rise by 144 K (12 residue polyalanine has 132 atoms and 131 bond length constraints for a total of 265 degrees of freedom). Transport properties are in general sensitive to the method used to keep the temperature constant and it is preferable to use constant E,V,N dynamics. However, the Nosé–Hoover method provides for a gentle rescaling of the velocities and can reproduce dynamical properties well.[77]

For all models, the fully extended chain converts quickly to a lower energy structure within 50 ps (see Fig. 6). Overall the energy drops by 70 kcal/mol, the value of $\langle C_A^2 \rangle$ drops from 0.61 to 0.43, and the fraction helix increases from 0 to almost 0.5. The energy and $\langle C_A^2 \rangle$ reach the average values from the simulations starting from an α-helix but the fraction helix is much different indicating that for this property there is a long approach to equilibrium. The other models show similar changes in energy, fraction helix, and $\langle C_A^2 \rangle$. The first
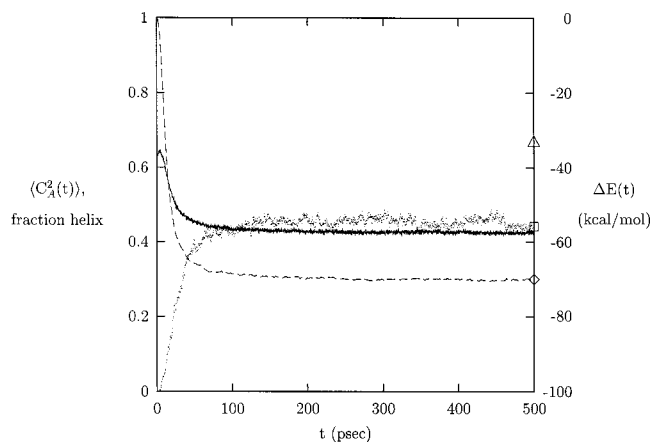
hydrogen bonds to form are 1–2 hydrogen bonds, between adjacent residues. The average time to form the first 1–2 hydrogen bond is about 1 ps for all models. Following the formation of the 1–2 hydrogen bonds is the formation of 1–3 hydrogen bonds, which occurs in 6 ps, in average. The first 1–4 hydrogen bonds form in about 12 ps. Hydrogen bonds form first at the ends, usually the N-terminal end. This is consistent with other simulations of polyalanine that find that the N-terminus is more helical than the C-terminus, although neither end is as helical as the center.[55,59,60] Helix formation tends to begin at the ends of the peptide and propagate to the center reaching a final state where the center is helical and the ends are not. For model 2, 4% of the ω-angles become cis in the first 10 ps and with time the percentage of cis peptide bonds decreases to about 1% after 500 ps. For model 1, 0.3% of the peptide bonds are cis at all times. For models 3 and 4, no cis peptide bonds form. Starting in the high energy FEC state provides the energy for the $trans–cis$ isomerization. From the simulations starting in an α-helix geometry no cis peptide bonds formed. The models with the more flexible dihedral parameters for state A undergo the $trans–cis$ isomerization, even though states A and B remain, on average, equally weighted with $\langle C_A^2 \rangle$ varying from 0.6 to 0.4. With an equal mix of states, the dihedral force constant for all models are about the same. Therefore, in order for the models to be different, fluctuations in the coefficients are necessary.

Figure 7 shows the time it takes to form a helix for three of the models. Model 3 is similar to model 1 and is not shown. Plotted is the fraction of trajectories that have formed a helix by time t.[64] At long times, after 200 ps, the fraction of nonhelix forming trajectories is assumed to decay exponentially as Ae^{-kt}. The rate constants are given in Table VIII and the exponential fits are shown on Fig. 7. Among the polarizable models (1–3), the rate at which helices are formed increases with the amount of flexibility in state A. Comparing model 3 with models 1 and 2 indicates that fluctuations in the dihedral flexibility (which are not present in model 3) increase the rate constant. Electronic polarizability
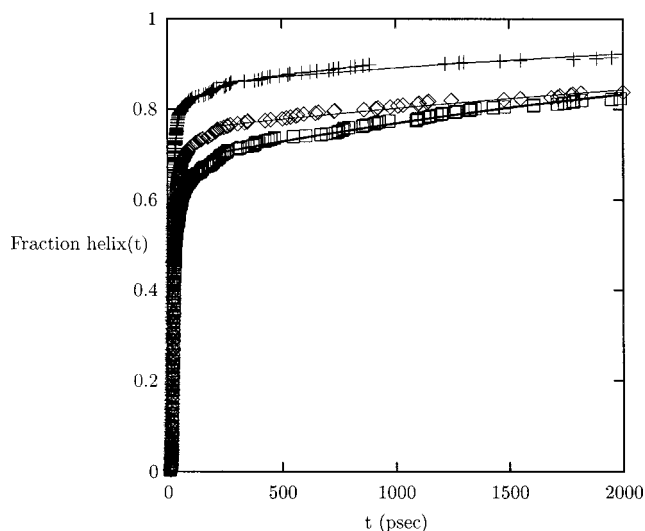
FIG. 7. Fraction of trajectories that have formed an α-helix by time, t, for models 1 (diamonds), 2 (squares), and 4 (+). The lines are the fit to $1 - A\exp(-kt)$.

decreases the rate constant, as a comparison of models 1 and 4 shows. If fluctuations in the charges and in the dihedral force constants are turned off, then the rate constant increases. Polarizability may lead to stronger hydrogen bonds which take more time to break in order to convert from a coil to a helix. Using a similar analysis on 155 trajectories, but with the long-time behavior estimated, Bertsch *et al.* calculate a rate constant of 0.0048 ps^{-1}, an order-of-magnitude larger than our results.[64] That study was done at a higher temperature (450 K) and used a distant dependent dielectric constant and fixed bond angles, both of which can change the dynamical properties. Our results give a half-life for forming a helix, $1/k$, of 3 to 6 nanoseconds and a time scale to initiate an α-helix by forming the first 1–4 hydrogen bond of 12 ps. These two results suggests that the rate limiting step in helix formation may not be helix initiation but propagation. This same result is also indicated by the results of Bertsch *et al.*[64] The coil to helix rate constant for a 21 residue alanine-based peptide has been measured to be 6×10^{-5} ps^{-1} (Ref. 78) and 10^{-4} ps^{-1} (Ref. 72) near 300 K. In the folding of the protein apomyoglobin, a fast component of the folding is found, at 333 K, with a rate constant of 10^{-5} ps^{-1} which was interpreted as the rate for helix formation.[79] The rates determined by our simulations are consistent with these experimental rates. The simulated rates are larger which is to be expected because the simulations are for smaller peptides in the absence of solvent.

V. CONCLUSIONS

The two-state model for the peptide bond is a new method for treating polarizability, in which not only the electrostatic but the bonded interactions are sensitive to the environment. The importance of environmental influences on the flexibility of the ω-angle was demonstrated in an analysis of atomic resolution protein structures. This analysis showed that the distribution of ω-angles, $P(\omega)$, is dependent on secondary structure with α-helix residues having a more narrow

distribution and coil residues having a broader distribution (Fig. 2). Antiparallel β-sheets also have a broad distribution with the mean shifted from 180° (Fig. 3). Lower resolution protein structures have a much narrower $P(\omega)$ than the atomic resolution structures, which were refined without using restraints on the ω-angles (Fig. 1). Ramachandran plots reveal differences in the distribution of ϕ and ψ angles as well (Figs. 4 and 5). The ϕ and ψ angles from the atomic resolution structures are nearer the allowed regions than the angles from the 2.2 Å structures. Overconstraining the ω-angles in the refinement of the lower resolution structures may be compensated by broadening the ϕ and ψ distributions.

The two-state model introduces no new interactions into the common potentials used to simulate proteins and, additionally, a standard 1 fs time step can be used. Therefore, this model does not require much more computational time. Rather than introducing new polarizable interactions into the potential, the model allows the interactions to respond to their environment. In this application, the model just includes polarization of four peptide group charges and the ω-angle dihedral force constant, but it could be easily extended to include changes in bond lengths, bond angles, and atomic polarization. The changes were easily inserted into a standard force field (AMBER4.1)[45] with no other modifications of potential parameters necessary. An interesting application of this model will be in the refinement of structures by x-ray crystallography using low and medium resolution data. The application of this model to the helix–coil transition of poly-alanine reveals some important effects due to polarizability. Compared to a nonpolarizable model, the helix–coil free energy change increases more rapidly for the polarizable model, by 0.4 kcal/mol, as a function of the number of residues, from 9 to 15. The helix–coil energy change, ΔE, is different by over 2 kcal/mol between the polarizable and nonpolarizable models. The energy differences are partially compensated by entropy. Relative to the nonpolarizable model, the change in the dihedral flexibility upon formation of an α-helix leads to energetically stronger, lower entropy hydrogen bonds (Table VII). Polarizability influences dynamical properties as well. Fluctuations in the peptide bond dihedral force constants change the rate constant of the coil to helix transition by a factor of two. Model 2, which has the largest difference in dihedral flexibility between the two states has a rate constant twice as large as the rate constant of model 3, which has no difference in dihedral flexibility for the two states.

The two-state picture has relevance to *cis–trans* isomerization as well. *Cis–trans* isomerization is catalyzed by hydrogen bonds to the amide nitrogen, which would stabilize the single bond state and lower the rotational barrier of the peptide bond.[80] On the other hand, hydrogen bonds to the carbonyl oxygen, which would stabilize the double bond state, raise the barrier.[81] Our simulations found that *cis–trans* isomerization only occurred if the dihedral flexibility fluctuates. Additionally, in the *cis* conformation the oxygen and the amide hydrogen atoms are closer. The attractive electrostatic interactions will stabilize state B relative to state A and previous *ab initio* calculations for NMA have found that the

electrostatic potential (ESP) fitted charges on the oxygen and hydrogen atoms are greater in magnitude for the *cis* conformer.[50] The importance of conformationally dependent charges on the solvation free energy difference between *cis* and *trans* NMA has been shown in a number of studies.[27,82,83] Calculations using conformationally independent charges do not get the correct free energy difference. If different charges are assigned by hand[82,83] or the charges adjust through a fluctuating charge Hamiltonian,[27] then the calculated free energy difference is in agreement with experiment. The two-state model is an alternative method for introducing conformationally dependent charges.

ACKNOWLEDGMENTS

[1] L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, NY, 1960).

[2] G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, Heidelberg, 1991).

[3] M. W. MacArthur and J. M. Thornton, J. Mol. Biol. **264**, 1180 (1996).

[4] E. J. Dodson, G. J. Davis, V. S. Lamzin, G. N. Murshudov, and K. S. Wilson, Structure **6**, 685 (1998).

[5] I. M. Klotz and J. S. Franzen, J. Am. Chem. Soc. **84**, 3461 (1962).

[6] A. Wada, Adv. Biophys. **9**, 1 (1976).

[7] P. T. van Duijnen and B. T. Thole, Biopolymers **21**, 1749 (1982).

[8] H. Guo and M. Karplus, J. Phys. Chem. **98**, 7104 (1994).

[9] R. Ludwig, F. Weinhold, and T. C. Farrar, J. Chem. Phys. **107**, 499 (1997).

[10] F. H. Stillinger and C. W. David, J. Chem. Phys. **69**, 1473 (1978).

[11] P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, Nature (London) **282**, 459 (1979).

[12] T. P. Lybrand and P. A. Kollman, J. Chem. Phys. **83**, 2923 (1985).

[13] J. A. C. Rullman and P. T. van Duijnen, Mol. Phys. **63**, 451 (1988).

[14] M. Sprik and M. L. Klein, J. Chem. Phys. **89**, 7556 (1988).

[15] P. Ahlström, A. Wallqvist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).

[16] P. Cieplak, P. Kollman, and T. Lybrand, J. Chem. Phys. **92**, 6755 (1990).

[17] S. Kuwajima and A. Warshel, J. Phys. Chem. **94**, 960 (1990).

[18] L. X. Dang, J. Chem. Phys. **97**, 2659 (1992).

[19] R. E. Kozack and P. C. Jordan, J. Chem. Phys. **96**, 3120 (1992).

[20] A. Wallqvist and B. J. Berne, J. Phys. Chem. **97**, 13841 (1993).

[21] D. N. Bernardo, Y. Ding, K. Krogh-Jespersen, and R. M. Levy, J. Phys. Chem. **98**, 4180 (1994).

[22] J. W. Caldwell and P. A. Kollman, J. Phys. Chem. **99**, 6208 (1995).

[23] J. Blodholt, M. Sampoli, and R. Vallauri, Mol. Phys. **86**, 149 (1996).

[24] A. A. Chialvo and P. T. Cummings, J. Chem. Phys. **105**, 8274 (1996).

[25] S.-B. Zhu, S. Singh, and G. W. Robinson, J. Chem. Phys. **95**, 2791 (1991).

[26] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).

[27] S. W. Rick and B. J. Berne, J. Am. Chem. Soc. **118**, 672 (1996).

[28] B.-C. Perng, M. D. Newton, F. O. Raineri, and H. L. Friedman, J. Chem. Phys. **104**, 7153 (1996).

[29] I. M. Svishchev, P. G. Kusalik, P. G. Wang, and R. J. Boyd, J. Chem. Phys. **105**, 4742 (1996).

[30] M. J. Field, Mol. Phys. **91**, 835 (1997).

[31] Y.-P. Liu, K. Kim, B. J. Berne, R. A. Friesner, and S. W. Rick, J. Chem. Phys. **108**, 4739 (1998).

[32] J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne, and R. A. Friesner, J. Chem. Phys. **110**, 741 (1999).

[33] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).

[34] S. Nosé, Mol. Phys. **52**, 255 (1984).

[35] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[36] A. Warshel and R. M. Weiss, J. Am. Chem. Soc. **102**, 6218 (1980).

[37] J. Lobaugh and G. A. Voth, J. Chem. Phys. **104**, 2056 (1996).

[38] D. E. Sagnella and M. E. Tuckerman, J. Chem. Phys. **108**, 2073 (1998).

[39] R. Vuilleumier and D. Borgis, J. Phys. Chem. B **102**, 4261 (1998).

[40] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. **112**, 535 (1977).

[41] W. A. Hendrickson, Methods Enzymol. **115**, 252 (1985).

[42] A. T. Brünger, J. Kuriyan, and M. Karplus, Science **235**, 458 (1987).

[43] D. E. Stewart, A. Sarkar, and J. E. Wampler, J. Mol. Biol. **214**, 253 (1990).

[44] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nuclaic Acids* (Cambridge University Press, Cambridge, 1987).

[45] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[46] G. Gilli, V. Bertolasi, F. Bellucci, and V. Ferretti, J. Am. Chem. Soc. **108**, 2420 (1986).

[47] D. van Belle, M. F. G. Lippens, and S. J. Wodak, Mol. Phys. **77**, 239 (1992).

[48] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[49] D. K. Remler and P. A. Madden, Mol. Phys. **70**, 921 (1990).

[50] D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. Goddard III, and B. Honig, J. Am. Chem. Soc. **116**, 11875 (1994).

[51] *CRC Handbook of Chemistry and Physics*, edited by R. C. Weast (CRC Press, Boca Raton, FL, 1985-1986), Vol. 66.

[52] J. Applequist, J. R. Carl, and K.-K. Fung, J. Am. Chem. Soc. **94**, 2952 (1972).

[53] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[54] G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).

[55] W. S. Young and C. L. Brooks III, J. Mol. Biol. **259**, 560 (1996).

[56] J. A. McCammon, S. H. Northrup, M. Karplus, and R. M. Levy, Biopolymers **19**, 2033 (1980).

[57] V. Daggett, P. A. Kollman, and I. D. Kuntz, Biopolymers **31**, 1115 (1991).

[58] F. M. DiCapua, S. Swaminathan, and D. L. Beveridge, J. Am. Chem. Soc. **113**, 6145 (1991).

[59] V. Daggett and M. Levitt, J. Mol. Biol. **223**, 1121 (1992).

[60] Y. Okamoto and U. H. E. Hansmann, J. Phys. Chem. **99**, 11276 (1995).

[61] L. Wang, T. O'Connell, A. Tropsha, and J. Hermans, Biopolymers **39**, 479 (1995).

[62] Y. Wang and K. Kuczera, J. Phys. Chem. B **101**, 5205 (1997).

[63] M. Takano, T. Takahashi, and K. Nagayama, Phys. Rev. Lett. **80**, 5691 (1998).

[64] R. A. Bertsch, N. Vaidehi, S. I. Chen, and W. A. Goddard III, Proteins **33**, 343 (1998).

[65] S. Samuelson and G. J. Martyna, J. Chem. Phys. **109**, 11061 (1998).

[66] C. L. Brooks III, Curr. Opin. Struct. Biol. **8**, 222 (1998).

[67] J. G. Savin and P. G. Wolynes, J. Mol. Biol. **257**, 199 (1996).

[68] D. Suckau, Y. Shi, S. C. Beu, M. W. Senko, J. P. Quinn, F. M. Wampler, and R. W. McLafferty, Proc. Natl. Acad. Sci. USA **90**, 790 (1993).

[69] B. A. Collings and D. J. Douglas, J. Am. Chem. Soc. **118**, 4488 (1996).

[70] D. S. Gross, P. D. Schnier, S. E. Rodriguez-Cruz, C. K. Fagerquist, and E. R. Williams, Proc. Natl. Acad. Sci. USA **267**, 1483 (1996).

[71] Y. Mao, J. Woenckhaus, J. Kolafa, M. A. Ratner, and M. F. Jarrold, J. Am. Chem. Soc. **121**, 2712 (1999).

[72] P. A. Thompson, W. A. Eaton, and J. Hofrichter, Biochemistry **36**, 9200 (1997).

[73] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).

[74] J. Wojcik, K.-H. Altman, and H. A. Scheraga, Biopolymers **30**, 121 (1990).

[75] A. Chakrabartty, T. Kortemme, and R. L. Baldwin, Protein Sci. **3**, 843 (1994).

[76] A. Chakrabartty and R. L. Baldwin, in *Protein Folding: In Vivo and In Vitro*, edited by J. Cleland and J. King (American Chemical Society, Washington, DC, 1993), p. 166.

[77] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego, 1996).

[78] S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, Biochemistry **35**, 691 (1996).

[79] R. Gilmanshin, S. Williams, R. H. Callender, W. H. Woodruff, and R. B. Dyer, Proc. Natl. Acad. Sci. USA **94**, 3709 (1997).

[80] C. Cox, V. G. Young, Jr., and T. Lectka, J. Am. Chem. Soc. **119**, 2307 (1997).

[81] S. Scheiner and C. W. Kern, J. Am. Chem. Soc. **99**, 7042 (19).

[82] W. L. Jorgensen and J. Gao, J. Am. Chem. Soc. **110**, 4212 (1988).

[83] P. Cieplak and P. Kollman, J. Comput. Chem. **12**, 1232 (1991).