Structural and Ultrametric Properties of (L-Alanine)$_{20}$

B. Velikson

Section de Biophysique des Protéines et des Membranes, DBCM, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette Cedex, France

J. Bascle,* T. Garel, and H. Orland

Service de Physique Théorique, CEA-Saclay, 91191 Gif-sur-Yvette, France

Received March 22, 1993; Revised Manuscript Received May 27, 1993

ABSTRACT: We study local energy minima of (L-alanine)$_{20}$. The minima are generated using high-temperature molecular dynamics and chain-growth Monte Carlo simulations, with subsequent minimization. We find that the lower-energy configurations are α-helices for a wide range of dielectric constant values (ϵ = 1, 10, and 80) and that there is no noticeable difference between the distribution of energy minima in $\phi\psi$ space for different values of ϵ. Ultrametricity tests show that lower-energy (α-helical) ϵ = 1 configurations form a set which is ultrametric to a certain degree, providing evidence for the presence of fine structure among those minima. We put forward a heuristic argument for this fine structure. We also find evidence for ultrametricity of a different kind among ϵ = 10 and ϵ = 80 energy minima. We analyze the distribution of lengths of α-helical portions among the minimized configurations and find a persistence phenomenon for the ϵ = 1 ones, in qualitative agreement with previous studies of critical lengths.

1. Introduction

There are a number of reasons for studying the structural properties of energy minima of poly(L-alanines). We shall cite here just a few:

1. It has been recently suggested[1] that poly(L-alanine) may serve as a good model for many proteins, namely, that the native backbone geometry for a polypeptide or protein of M residues very often has a closely related metastable analog in a poly(L-alanine) of the same number M of residues. The proteins for which this hypothesis has been successfully checked include bovine pancreatic trypsin inhibitor (M = 58), crambin (M = 46), ribonuclease A (M = 124), and superoxide dismutase (M = 158).

2. Our studies of (L-alanine)$_7$[2] showed that all energy minima group themselves into dense clusters in configuration space, and the extension of this behavior to longer molecules is quite plausible and supported, indeed, by the results of the present paper. On the other hand, there is a general belief that the global minimum of poly(L-alanine) in the vacuum is an α-helix for sufficiently long molecules. Therefore, we have reasons to expect that there are completely or partially α-helical clusters in the low-energy configurations of any poly(L-alanine), and it would be interesting to find out more about the cluster fine structure, if any.

3. If α-helices are in fact the lowest-energy conformations, why is that so? Is it due mostly to steric or electrostatic effects? What is the influence of the dielectric constant?

In this paper we study these and related questions, taking as an example (L-alanine)$_{20}$. This molecule is sufficiently long that α-helices become the lowest-energy configurations (contrary to the (alanine)$_7$ case[2]) but short enough for an extensive computational study. The energy-minimized configurations are obtained by computer simulation and minimized at three different values of dielectric constant: ϵ = 1, 10, and 80. We find that, indeed, there is evidence for fine structure among the ϵ = 1 α-helical (lowest) minima: the set of such minima seems *ultrametric*. We also find evidence for ultrametricity (of a different kind) in the ϵ = 10 and 80 cases. We study the α-helical content of minimized configurations and find that for ϵ = 1 there is "persistence" of an α-helical structure once it has reached a certain length; this is consistent with earlier studies of critical length for α-helix formation and distribution of α-helical portions of longer chains.[3-6] We show that α-helices are the lowest-energy structures independently of ϵ but that the reason for this property—the relative role of different terms in energy—does in fact depend on ϵ. We also note that the value of the dielectric constant does not much affect the position of energy minima in the $\phi\psi$ space: the $\phi\psi$ space clusters of energy minima stay very much in place.

The paper is organized as follows. In section 2, we briefly describe the model and the sampling methods. We have generated Boltzmann-distributed configurations of (L-alanine)$_{20}$ at various temperatures. These configurations were obtained by two methods: (1) the Langevin molecular dynamics simulations run with CHARMM[7,8] and (2) the Monte Carlo chain-growth method described in ref 2. Subsequent minimization was done with the adopted basis-set Newton–Raphson algorithm (ABNR)[7] under CHARMM.

Section 3 is dedicated to the results. In particular, ultrametricity (the evidence for fine structure among lower-energy minima) is discussed in section 3.3. The statistics concerning α-helical content among the minima is presented in section 3.4, and the role of various energy terms in stabilizing the helical conformations is discussed in section 3.5. We end with a short conclusion.

2. The Model and the Methods

In this section we describe the model and the sampling methodology.

Exactly as we did for (L-alanine)$_7$,[2] the (L-alanine)$_{20}$ molecule is represented as $CO_2CH(CH_3)NH(COCH(CH_3)NH)_{18}COCH(CH_3)NH_3$. The atom numbering scheme and chemical structure are shown in Figure 1.

The charges on the carboxyl and ammonium groups are set to zero so as to avoid dominating charge–charge interactions. In this way the principal electrostatic interactions will result in hydrogen bonding between backbone peptide groups.

0024-9297/93/2226-4791$04.00/0

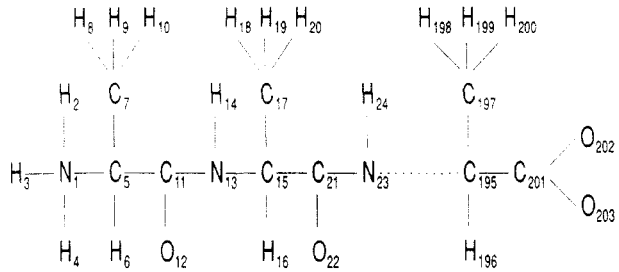

Figure 1. Numbering of atoms for (alanine)$_{20}$. The period for the intermediate (- - -) region is ten atoms long; one finds, e.g., nitrogen atoms at 33, 43, 53, ..., 153, ...

2.1. The Energy Function. For both methods mentioned above, we use the same energy function as described previously:[8,9]

$$E = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi(1 + \cos[n\phi - \delta]) + \sum_{\text{impropers}} k_\nu(\nu - \nu_0)^2 + \sum_{i<j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i,j} \frac{332 q_i q_j}{\epsilon r_{ij}}[e14(i,j)] \quad (1)$$

where distances are in angstroms, angles are in radians, and E is in kcal/mol.

In eq 1 b, θ, ϕ, and ν are the bond lengths, angles, dihedrals, and improper torsions and b_0, θ_0, and ν_0 are the reference values for these properties. The force constants associated with these terms are k_b, k_θ, k_ϕ, and k_ν. For the intrinsic torsions n is the symmetry number of the rotor (e.g., 3 for a methyl group) and δ is the phase angle.

The nonbonded terms in eq 1 are pairwise additive and consist of a 12–6 Lennard-Jones van der Waals term and a Coulombic term representing electrostatic interactions between atomic point partial charges. These nonbonded interactions are included for 1,4 (vicinal) or higher-order atom pairs. The quantities r_{ij}, ϵ_{ij}, σ_{ij}, q_i, and q_j are the nonbonded distance, the Lennard-Jones well depth, the Lennard-Jones diameter, and the charges for atom pairs i and j. The factor $[e14(i,j)]$ in eq 1 is 0.5 for (1,4) pairs and 1 for all other pairs of atoms. A distance-independent relative dielectric constant ϵ of 1, 10, or 80 is assumed. The parameters used in the energy function can be found, e.g., in ref 2.

2.2. Sampling Methods. We made use of two sampling methods here: (1) traditional molecular dynamics and (2) the Monte Carlo chain-growth method.[2] Notice that our aim was to obtain as many as possible different configurations belonging to different basins of attraction of the minimizing procedure rather than to obtain good-quality Boltzmann ensembles. This is why it was quite appropriate to use finite-temperature configurations produced with one value of ϵ as starting points for an energy minimization at this and *other* values of ϵ. Heuristically, this idea was due to the following consideration: in the approximation in which the molecule's energy is dominated by electrostatics, changing ϵ is equivalent to changing the temperature.

Changing the value of ϵ for the purposes of minimization may seem a rather unusual way of handling a sampling procedure. One may worry that this could result in an undesired biasing of the ensemble of minima. Such a consideration would be appropriate if, in doing so, we could *lose* information. In the case at hand, however, we could certainly only gain in finding *more* minima, wherever they come from. That is, without using this procedure, one obtains the deepest energy minima at all values of ϵ in the form of α-helices. However, the statistics is too poor, and one would like to increase it. For this purpose, we observe that all (ϵ = 1)-generated configurations converging to a helix upon minimization belong to the basins of attraction of ϵ = 10 and ϵ = 80 α-helical minima as well. In fact, this procedure gave us as many ϵ = 10 and ϵ = 80 α-helices as the standard procedure did: half of those, in our simulation, were produced by a minimization starting from ϵ = 1 strongly-biased Monte Carlo generated configurations.

Thus, in this way we obtained a better representation of the minima population for each value of ϵ without having to perform more high-cost finite-temperature MC or MD simulations.

(1) MC Chain-Growth Method. A detailed analysis of the Monte Carlo Chain-growth method can be found in 2. In this procedure, an ensemble of chains (molecules) is grown atom by atom, so as to generate an ensemble which (in the large-ensemble limit) obeys the Boltzmann distribution. One advantage of this method is that it does not assume any particular guess for the initial state but rather builds up the chain according to Boltzmann weights. Several such Monte Carlo runs at various temperatures (300K, 1000K, and 6000K) and values of ϵ were performed, giving a population of 86 different configurations produced at ϵ = 1, 36 configurations at ϵ = 10, and 16 configurations at ϵ = 80.

In one of the ϵ = 1 MC simulations, yielding 16 different configurations, we used a biased (guiding field) version of the chain-growth method.[10] Minimization of these at all three values of ϵ invariably produced completely α-helical configurations.

A typical CPU time necessary to produce one configuration was 0.7 s on a Cray 2.

(2) Molecular Dynamics. MD simulations were performed using the program CHARMM[7] with the Langevin dynamics algorithm.[11] In this approach, the equation of motion contains a frictional term and a Gaussian external force. In a one-dimensional representation the equation is

$$m\ddot{x} = -U'(x) - \gamma\dot{x} + f(t) \quad (2)$$

where the friction coefficient γ and the force correlation function $\langle f(t)f(0)\rangle = C\delta(t)$ are related by $C = 2\gamma k_B T$. In our simulations, the values of γ/m = 40 s^{-1} and T = 1000 K were used. We performed two such simulations, with different starting points (α-helix and β-sheet), for each of the three values of ϵ. For ϵ = 1, one simulation of 400 ps and one of 297 ps were performed. For ϵ = 10, the two runs were 100 and 150 ps long. For ϵ = 80, they were 150 and 200 ps long. The configurations to be minimized were taken each 1 ps of the MD trajectory.

A typical 100-ps MD simulation took an average of 37 h of CPU time on an Alliant VFX-80 machine.

(3) Minimization. As we have already mentioned, we used configurations produced at each value of ϵ for minimization both at that and other ϵ values.

In this way, we obtained 799 ϵ = 1 configurations, 272 ϵ = 10 configurations, and 452 ϵ = 80 configurations to be minimized. The minimization was done in two stages. To save CPU time and because we are more interested in low-lying minima, a preliminary minimization of 2000 cycles of the adopted basis-set Newton–Raphson (ABNR) procedure was performed on all of the above configurations. These partly converged minima were ordered in energy, and a set of two-dimensional $\phi\psi$ plots was obtained. Then a number of lower-energy configurations (see Table I) were

Table I. Number of Configurations of (Alanine)$_{20}$ Obtained during Different Simulations[a]

	partially minimized				fully minimized			
	MD	MC	MC$_\epsilon$	total	taken for minimization	nonidentical	taken for ultrametr tests	energy range of fully-minimized conf (kcal/mol)
$\epsilon = 1$	697	86	16	799	226	133	133	−591.202 to −571.371
$\epsilon = 10$	150	36	86	272	84	69	69	−116.285 to −53.715
$\epsilon = 80$	350	16	86	452	32	18	61	−68.519 to −59.537

[a] All obtained configurations were at least partially minimized (2000 cycles of ABNR). Of these, a lower-energy fraction were fully minimized. MC$_\epsilon$ denotes those MC-generated configurations which were generated at a different value of ϵ than the one used during minimization. In the $\epsilon = 80$ case, the 18 fully-minimized nonidentical configurations were complemented by 43 partially-minimized configurations for the ultrametricity test purposes.

chosen to be minimized until the rms energy derivative was 10^{-6} kcal/mol Å or smaller. (The partly minimized configurations were so close in configuration space to actual minima that there was no noticeable difference between the $\phi\psi$ plots of partially and completely minimized configurations.) This second stage of minimization took from 3000 to 15 000 cycles of the ABNR. Naturally, some configurations converged to identical minima. The total count of initial and minimized configurations is given in Table I.

Typically, minimization took about 3 h of CPU time per 1000 ABNR steps on an Alliant VFX-80 machine, i.e., 6 h for the first, 2000-step stage, with an additional 9–45 h during the minal minimization stage.

3. Results

3.1. $\phi\psi$ Distribution and Classification of Minimum-Energy Structures. The obtained $\phi\psi$ maps are very similar among themselves. There is not much difference between the maps of different $\phi\psi$ planes nor between the maps corresponding to different values of ϵ; we present two such maps, for $\epsilon = 1$ and $\epsilon = 80$ (Figure 2a,b). All of the maps show clusters of points located in the same regions as those previously obtained for (L-alanine)$_7$.[2] Moreover, we know that their localization corresponds to the energy minima of an alanine dipeptide.[2] This means that the long-distance cross-residue effects do not affect the location of the minima in the first approximation.

This fact allows us to use the techniques introduced earlier in ref 2 and characterize a configuration by a set of 18 digits varying from 1 to 9. Thus, first we consider the configuration as belonging to the 36-dimensional space spanned by the 18 inner pairs of ϕ and ψ backbone angles. This makes our study much easier than if we considered the whole $(3N - 6) = 603$-dimensional coordinate space. Then we divide each $\phi\psi$ plane into rectangles as shown in Figure 3. The division is such that individual clusters of points on the two-dimensional maps (Figure 2) do not share rectangles. This classification divides the 36-dimensional space into 9^{18} rectangular "boxes", each of which can be labeled by a sequence of 18 integers referring to the consecutive $\phi\psi$ planes, e.g., {5 6 1 5 7 6 1 1 3 2 9 7 2 4 4 9 6 8} etc., so that we can label all minima as belonging to this or that box. This classification provides an easily observable characteristic of a structure, so that, e.g., one can see at a glance the change that a given configuration undergoes during minimization (Table II). In our opinion, such a representation is often preferable to two-dimensional molecular tracings. While, looking at the latter, it is quite difficult to observe that a given structure possesses, say, two α-helical pieces joined by a non-α-helical one, one can immediately see this fact if one encounters a structure labeled {1 2 2 2 2 1 7 2 2 2 2 2 2 2 2 2 2 2}. Moreover, we shall often simplify the notation by writing the latter structure as {1 2$^{(4)}$ 1 7 2$^{(11)}$}, as we did in ref 10, so that a complete α-helix becomes {2$^{(18)}$}.

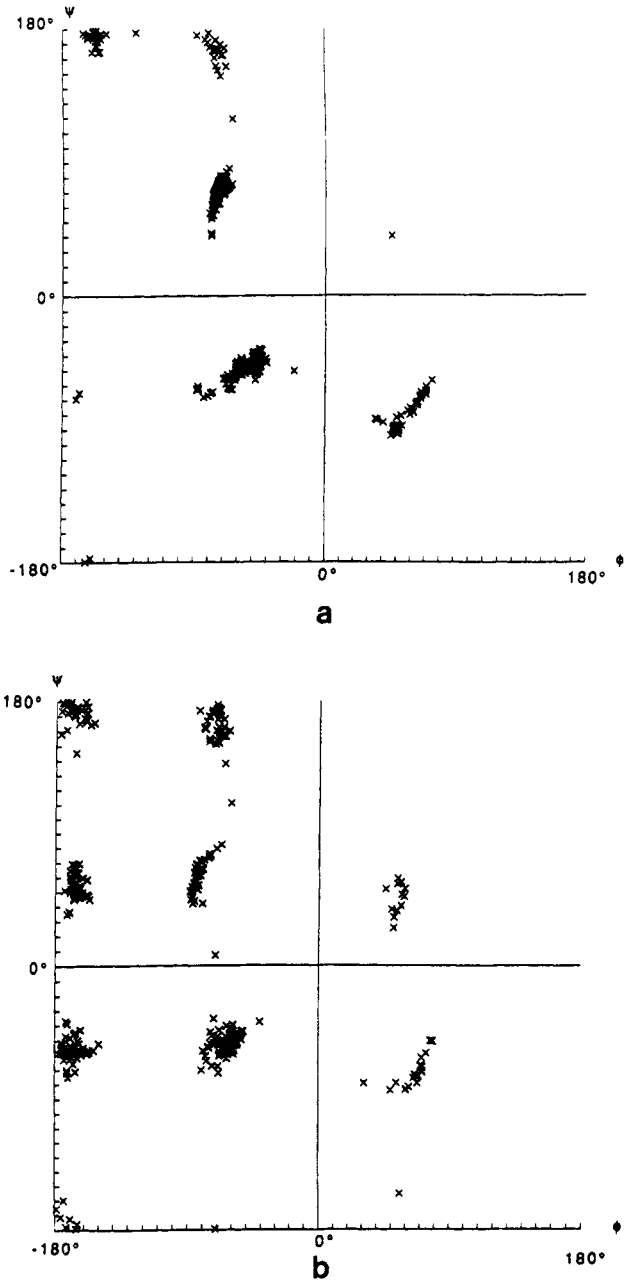

Figure 2. Energy minima of (alanine)$_{20}$ on a typical $\phi\psi$ plane: (a) $\epsilon = 1$; (b) $\epsilon = 80$.

3.2. Characteristics of the Low-Energy Minimized Configurations. For each value of the dielectric constant, we have considered the obtained energy-minimized configurations. As we have just said, the $\phi\psi$ distribution of minima is not much affected by the value of ϵ. Contrary to the case of (L-alanine)$_7$, the lowest-energy configurations always are complete α-helices. This is illustrated in Tables

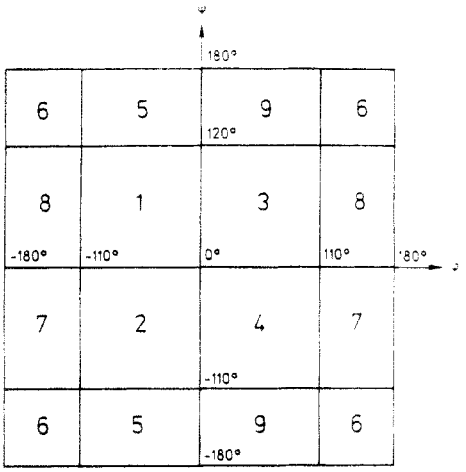

Figure 3. $\phi\psi$ plane splitting generating splitting of the 36-dimensional $\phi\psi$ space into "boxes".

Table II. Example of a Structural Change during Minimization[a]

	structure	energy (kcal/mol)
initial	2221765112661155 26	-387.2929
minimized	2221265111661151 26	-496.8968

[a] The integers refer to the labeling introduced in Figure 2.

Table III. Minimized Configurations, ϵ = 1[a]

structure	energy (kcal/mol)
22222222222222222222	-591.202
22222222222222222222	-591.099
22222222222222222222	-590.401
22222222222222222222	-590.135
22222222222222222222	-589.997
22222222222222222222	-589.791
22222222222222222222	-589.509
22222222222222222222	-589.437
22222222222222222222	-589.163
22222222222222222222	-588.895
22222222222222222222	-588.827
42222222222222222222	-588.748
42222222222222222222	-588.743
22222222222222222222	-588.690
...	...
14222222222222222221	-586.820
...	...
14172222222222222222	-572.014
...	...
33333333333333333333	-551.657
...	...
12626511667575511 1	-476.187

[a] The left-handed helix ($3^{(18)}$) has about 40 kcal/mol higher energy than the lowest α-helix.

III–V for ϵ = 1, 10, and 80, respectively. We included in Table III an artificially obtained left-handed helical structure ($3^{(18)}$). One can see that its energy is much higher (by about 40 kcal/mol) than those of α-helices.

The chain tracing of the lowest-energy ϵ = 1 structure is given in Figure 4. The hydrogen bonds are given as broken lines. One should remember that there is no explicit H-bond term in the CHARMM energy function: H-bonds are simulated by a combination of the electrostatic and Lennard-Jones terms.

Of the 19 ϵ = 1 complete α-helices obtained after minimizing the 799 initial configurations, 5 are the result of MC simulations, of which 3 were the biased ϵ = 1 MC runs and the other 2 were nonbiased MC ϵ = 80 runs minimized at ϵ = 1. The remaining 14 configurations are the result of ϵ = 1 MD simulations starting from an α-helix.

All the 12 ϵ = 10 complete α-helices obtained after minimizing the 272 initial configurations are the result of

Table IV. Minimized Configurations, ϵ = 10

structure	energy (kcal/mol)
22222222222222222222	-113.285
22222222222222222222	-112.635
22222222222222222222	-112.380
22222222222222222222	-111.798
22222222222222222222	-111.748
22222222222222222222	-111.731
22222222222222222222	-111.555
22222222222222222222	-111.493
22222222222222222222	-110.843
22222222222222222222	-110.135
22222222222222222222	-109.234
22222222222222222222	-109.168
27222222222222222222	-107.925
27222222222222222222	-106.445
...	...
17642172172222217 7	-100.896
...	...
52222887222222223 5	-95.879
...	...
56157611329724496 8	-53.715
...	...

Table V. Minimized Configurations, ϵ = 80

structure	energy (kcal/mol)
22222222222222222222	-68.5190
22222222222222222222	-67.8760
22222222222222222222	-67.7040
22222222222222222222	-67.6130
22222222222222222222	-67.5160
22222222222222222222	-67.4910
22222222222222222222	-67.4890
22222222222222222222	-66.8720
22222222222222222222	-66.7250
22222222222222222222	-66.4860
22222222222221222222	-64.8130
22228787222222227	-63.5790
22172222172222187 2	-61.6840
22217222222172222 12	-61.5060
...	...
11261115161111425 8	-52.4166
12222222211512135 5 5	-50.0045
...	...
21322121872491115 1	-35.7261
...	...
56157611329724495 8	-17.6246
56157611329724496 8	-17.3373
...	...

MC simulations, of which 6 were the biased ϵ = 1 MC runs minimized at ϵ = 10 and the other 6 were nonbiased MC ϵ = 10 runs.

All the 10 ϵ = 80 complete α-helices obtained after minimizing the 452 initial configurations are the result of MC simulations, of which 5 were the biased ϵ = 1 MC runs minimized at ϵ = 80 and the other 5 were nonbiased MC ϵ = 80 runs.

The energy differences between the highest- and lowest-energy α-helices ($2^{(18)}$ configurations) are 11.4 (ϵ = 1), 4.1 (ϵ = 10), and 2.0 kcal/mol (ϵ = 80).

3.3. Ultrametricity. We find dense clusters of minima in configuration space. This phenomenon has already been studied for small folded globular proteins using MD[12] and MC[13] techniques. A thorough analysis of the clusterization phenomenon was conducted in our previous paper.[2] It has already been noted more than once (e.g., ref 12) that the complexity of the configuration space and the existence of multiple minima with small energy differences bear resemblence to the case of spin glasses. It comes to mind, therefore, that it would be useful to verify if our distribution of minimized conformations shares with spin glasses one of their interesting properties, i.e., ultrametricity.[14] We recall that a simplistic notion of ultrametricity can be made

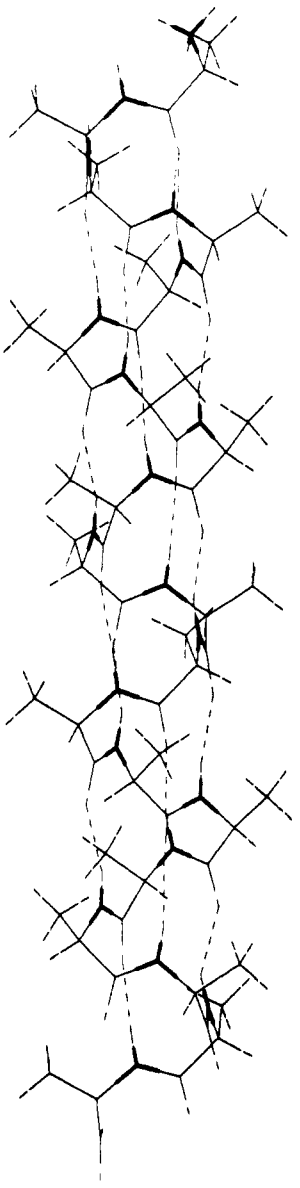

Figure 4. Chain tracing of the lowest-energy $\epsilon = 1$ configuration. The hydrogen bonds are given as broken lines.

by imagining a series of clusters within clusters: they form a treelike structure, so that clusters are present at an arbitrary scale. A preliminary analysis for the case of folded myoglobin can be found in ref 12, where no direct evidence for a tree structure was found. More rigorously,[14] a space (E) endowed with a distance d is called ultrametric if, for any triplet (A,B,C) of points belonging to (E), the distance d satisfies the ultrametric inequality:

$$d(\mathrm{A,C}) \leq \mathrm{Max}\{d(\mathrm{A,B}); d(\mathrm{B,C})\}$$

Note that the above inequality is stronger than the usual triangular inequality

$$d(\mathrm{A,C}) \leq d(\mathrm{A,B}) + d(\mathrm{B,C})$$

and implies that, in an ultrametric space, all triangles are either equilateral or isoceles with a small base.

In our previous paper on (L-alanine)$_7$,[2] we looked into the question whether the minimized configuration set forms an ultrametric ensemble. No evidence for such a property was found there. We are posing the same question for the present case of (L-alanine)$_{20}$. In the $\epsilon = 1$ case, the answer remains negative if one considers the whole set of energy-minimized configurations, and it is positive if one restricts the set to the lower, completely or partially

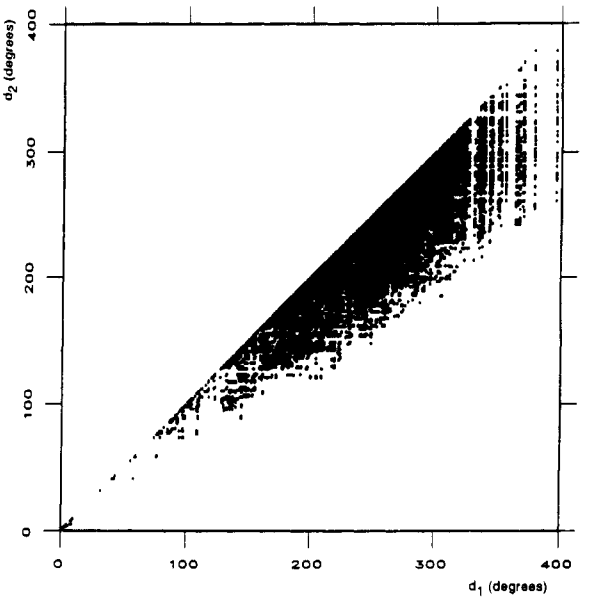

Figure 5. Ultrametricity test for 67 out of 133 lower-energy minima, $\epsilon = 1$. The graph consists of 47 905 points: for each triangle in configuration space formed by three energy minima, the length of the second longest side d_2 is plotted against that of the longest side d_1. In the ideally ultrametric case, all points would lie on the diagonal $y = x$. A subset of only 67 minima was chosen for this graph due to the printer's memory limitations; actual computation was done for 383 306 points generated by 133 minima.

α-helical configurations. In the $\epsilon = 10$ and $\epsilon = 80$ cases, the situation is somewhat different.

To test if the (alanine)$_{20}$ distribution is ultrametric, we first have to choose a distance d between the minima. There is no unique choice of such a distance. A natural choice for peptides is the distance in 36-dimensional $\phi\psi$ space, defined as follows:

$$d = (\sum_{1}^{18}[(\phi_i - \phi_i')^2 + (\psi_i - \psi_i')^2])^{1/2}$$

corrected to take into account the periodicity of ϕ and ψ coordinates: if $|\phi_i - \phi_i'| > 180°$, it is substituted by $(360° - |\phi_i - \phi_i'|)$, and the same applies to ψ. Ultrametricity will be tested below with this choice of distance d. The reader is referred to ref 14 for details concerning the robustness of ultrametricity with respect to the choice of distance.

Considering all possible triplets of energy minima and forming triangles with vertices at each minimum, one then plots the length of the second longest side of each triangle against the length of the longest side. In the ideal ultrametric case, the points will lie on the diagonal $y = x$. If no ultrametricity is present, the points will fill the triangle between the diagonal and the line $y = x/2$. (The lower bound is a trivial result of the triangle inequality.) If ultrametricity is present to some extent, the density of points will become greater toward the diagonal.

In Figures 5–7 we present such plots for the $\epsilon = 1$, $\epsilon = 10$, and $\epsilon = 80$ cases. The computations were done for samplings of 133 out of 799, 69 out of 272, and 61 out of 452 configurations, respectively (see Table I). Thus, in the $\epsilon = 1$ and $\epsilon = 10$ cases, we took all fully minimized configurations for this analysis. In the $\epsilon = 80$ case, we felt that 18 fully minimized configurations would give us too poor a statistics, so that we added 43 upper, partially minimized configurations.

In the $\epsilon = 1$ case, due to the printer's memory limitations, we could not present all the $C_{133}^3 = 383\,306$ triplets on the graph, so that Figure 5 shows a subset of $C_{67}^3 = 47\,905$

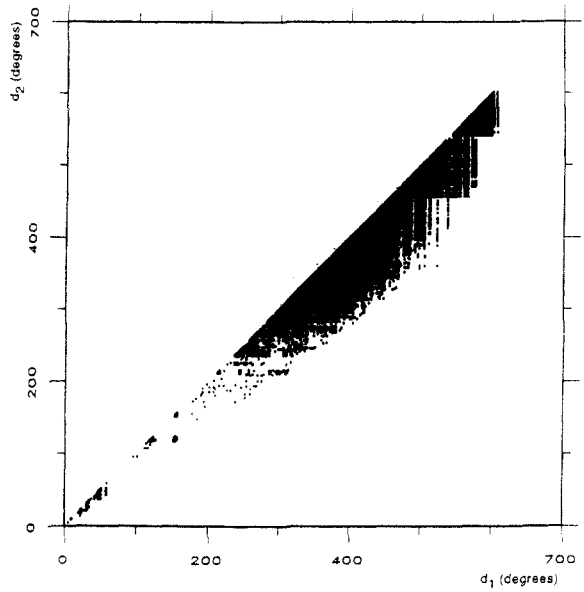

Figure 6. Ultrametricity test for 69 lower-energy minima, ϵ = 10. The graph consists of 52 394 points: for each triangle in configuration space formed by three energy minima, the length of the second longest side d_2 is plotted against that of the longest side d_1. In the ideally ultrametric case, all points would lie on the diagonal $y = x$.

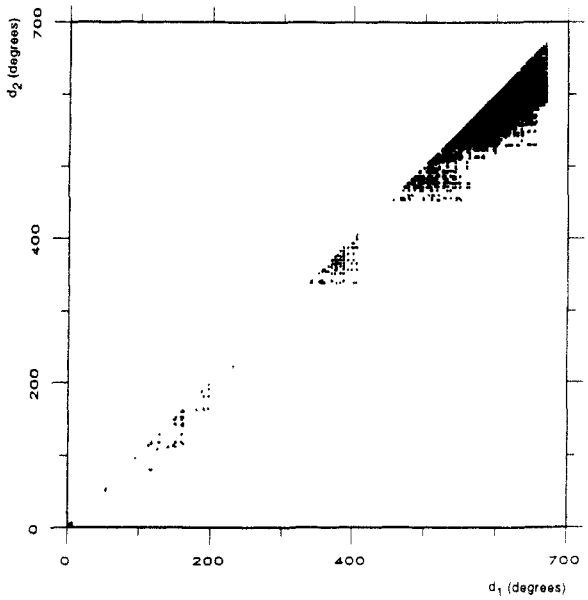

Figure 7. Ultrametricity test for 61 lower-energy minima, ϵ = 80. The graph consists of 35 990 points: for each triangle in configuration space formed by three energy minima, the length of the second longest side d_2 is plotted against that of the longest side d_1. In the ideally ultrametric case, all points would lie on the diagonal $y = x$.

triplets generated by taking each second configuration of the 133 configurations used in computation. The center of inertia of the actual distribution of 383 306 points is at (x = 241.0378, y = 211.8622), i.e., on the line $y = 0.88x$. The ratio of the moments of inertia of the point distribution along the principal axes is 18.0834.

For ϵ = 10, the plot consists of C_{69}^3 = 52 394 triplets. Its center of inertia is at (x = 450.6481, y = 423.6821), i.e., on the line $y = 0.94x$. The ratio of the moments of inertia of the point distribution along the principal axes is 72.107.

For ϵ = 80, the plot consists of C_{61}^3 = 35 990 triplets. Its center of inertia is at (x = 599.8134, y = 577.8925), i.e., on the line $y = 0.96x$. The ratio of the moments of inertia of the point distribution along the principal axes is 29.005.

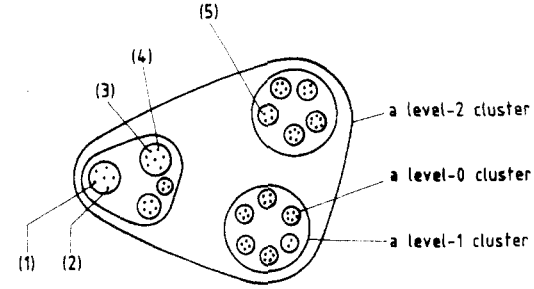

Figure 8. Sketch of the hierarchy of energy minima in $\phi\psi$ space corresponding to a change in a given turn of the helix, in two turns, etc.

These results are to be compared with a similar point distribution generated by a random set of points in the 36-dimensional $\phi\psi$ space. In that case, a numerical analysis gives the ratio of the moments of inertia of 4.0067. This seems to constitute a clear evidence for calling the above distributions of lower minima ultrametric to a large degree.

At the same time, we have to stress the difference between the case of ϵ = 1 lower-energy configurations and the two other cases.

As is well known from the theory of spin glasses,[14,15] one can define roughly two types of ensembles with ultrametric properties: (i) for the first type (which corresponds to the high-temperature phase in the spin glass case), the triangles in the configuration space are mostly equilateral; (ii) the second type (which corresponds to the low-temperature phase in the spin glass case) has, in addition, an isoceles triangle component.

For ϵ = 1, the 133 lower configurations are mostly α-helical with the third, nonplotted side of the triangle being usually much smaller than the other two (second type ultrametricity). By contrast, for ϵ = 10 and 80, only a small fraction of the configurations taken for ultrametricity analysis are helical, since, as mentioned above, our statistics in these two cases is rather poor. In these cases the majority of triangles have three comparable sides (first type ultrametricity). To check this point, we considered separately 495 upper ϵ = 1 configurations with no α-helices present at all. In this case, we obtained a plot of C_{495}^3 = 20 092 215 triplets of points, with the center of gravity at (x = 507.1641, y = 469.5677) (lying on the line $y = 0.93x$), with the ratio of the moments of inertia of 21.6; furthermore, we found that the third nonplotted side of the triangle is comparable to the first two. We thus conclude that for all values of ϵ there is a first type ultrametricity among the non-α-helical configurations.

Much more interesting is the ultrametricity among highly-ordered, α-helical configurations that we see in the lower-energy ϵ = 1 case. One may try to put forward a qualitative argument why ultrametricity could be expected here. Suppose that an independent change in the $\phi\psi$ coordinates corresponding, say, to one turn of the helix can drive the molecule from one minimum to another one with a much smaller change in other $\phi\psi$ coordinates. Consider one particular energy minimum corresponding to an α-helix (minimum 1 in Figure 8). Now consider another minimum corresponding to a small change in the position of one turn of the helix (minimum 2 of Figure 8). This minimum is very close in $\phi\psi$ space to minimum 1. Still another minimum can be obtained by a small change in the position of *two* turns of the helix (minimum 3 of Figure 8). This minimum is further away from minimum 1 or 2, but there will be other minima closer to it obtained from minimum 3 by a small change in the position of only one turn of the helix (minimum 4 of Figure 8). Changing *three* turns of the helix, we obtain a minimum still further

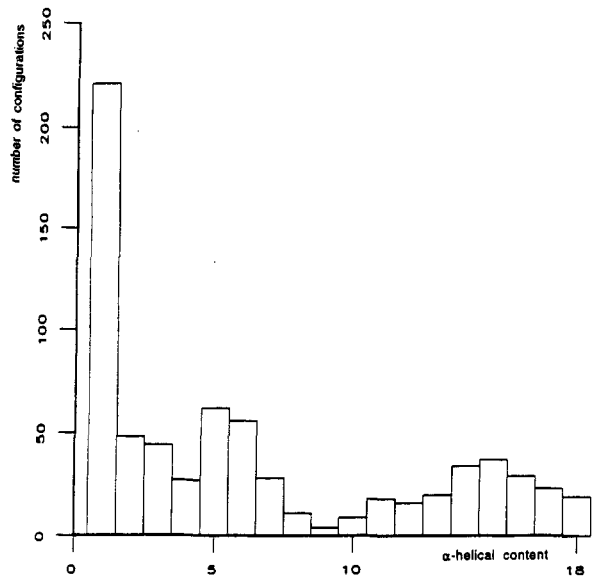

Figure 9. Frequencies of encountering an α-helical portion of length n among the $\epsilon = 1$ minimized configurations. ($n = 18$ corresponds to a complete α-helix.)

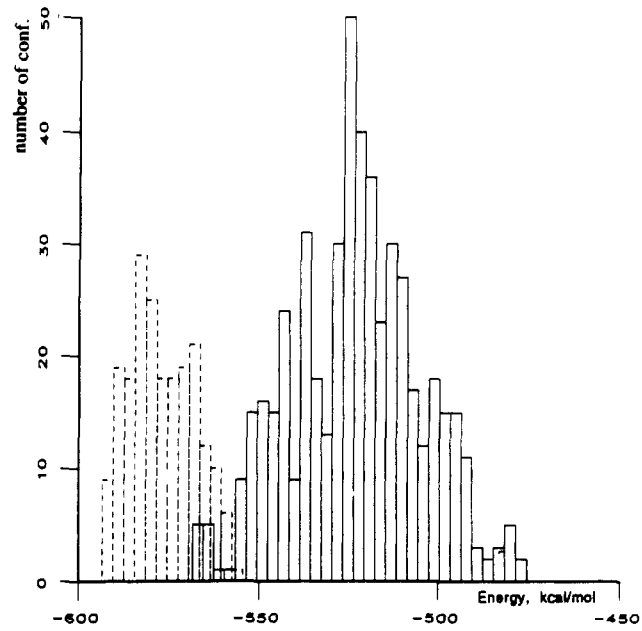

Figure 10. Energy histograms for $\epsilon = 1$ minimized configurations: the y axis represents the number of configuration falling into each 3 kcal/mol energy bin. Dashed line: configurations with helical content $n < 9$; solid line: configurations with helical content $n > 9$.

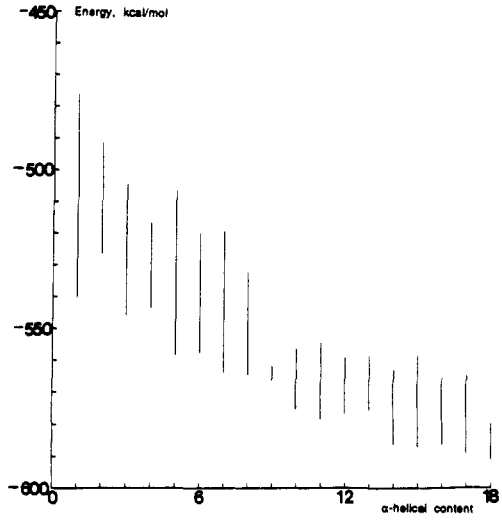

Figure 11. Energies spanned by configurations characterized by a given helical content n.

away (minimum 5 of Figure 8). In this fashion, we obtain a hierarchy of clusters (called level 0, level 1, level 2, etc. cluster in Figure 8), corresponding to a change inside a given turn of the helix, in one more turn, in two more turns, etc. Even though this model is necessarily simplistic and the reality may be organized in a somewhat more complicated way, it provides an insight into the reasons for existence of a fine structure among this type of configurations.

3.4. α-Helical Content: Persistence of α-Helical Structure. A definition of what is an α-helical portion of a secondary structure differs from one author to another.[6,16-18] In our approach, it is natural to adopt the definition stemming from our way of labeling configurations (section 3.1). That is, a configuration is said to have an α-helical piece of length n if its label contains a $2^{(n)}$ part. We shall also say that such a configuration has an α-helical content of n. The reason for choosing this definition rather than a definition based on H-bonds or on explicit variation of ϕ and ψ angles (as in ref 18) is that due to the clustering phenomenon, all such definitions are equivalent, and ours is much easier to use.

(1) $\epsilon = 1$ Case. We analyzed the relative distribution of configurations with different α-helical content n. In Figure 9 we show how many $\epsilon = 1$ minimized configurations contain α-helical pieces of given n. (Our convention is such that, e.g., a chain labeled $\{2^{(2)}$ 1 5 7 3 4 8 4 $2^{(3)}$ 3 7 8 $2^{(3)}\}$, which contains one 2-chain of twos and two 3-chains of twos, will contribute once into the count for 2-chains and once into the count for 3-chains. Other conventions could be envisioned.) We see that the distribution has a minimum at $n \simeq 9$.

This means that in this strong electrostatics case of $\epsilon = 1$, we encounter the phenomenon of persistence of the α-helical structure: once it has reached $n = 10$, there is a tendency for the α-helical chain to become longer, but there is no such tendency for smaller n.

These results qualitatively agree with the known existence of a critical length l_c, below which an α-helix is not the lowest-energy state of a polyalanine. Using a potential different from ours, an l_c of 16 was found in ref 3.

The analysis can be pushed somewhat further. Dividing all configurations into those which contain α-helical pieces with $n > 9$ and those with $n < 9$, we plot two energy histograms, marking ther number of configurations of each kind falling in a given energy bin. These histograms are plotted in Figure 10. We can see that there is a very small overlap of the two histograms: the two kinds of configurations are quite well separated in energy. Another way to see this energy separation is to plot the entire energy range spanned by each set of configurations of a given helical content n. This plot is presented in Figure 11.

(2) Other Values of ϵ. The helical content behavior differs significantly for other values of ϵ. The helical length histograms similar to that of Figure 9 are given in Figures 12 and 13. Since no change in the chain character ever occurred during the second minimization stage and in order to increase the statistics, we took all minimized chains, rather than only the fully minimized ones, for this analysis. This may have resulted in a multiple count for those partially minimized chains which after full minimization converged to identical minima. This does not influence the qualitative character of the histogram. Here it is especially interesting to look at the helical content distribution *with* and *without* taking into account the chains produced by the biased MC method, which searched

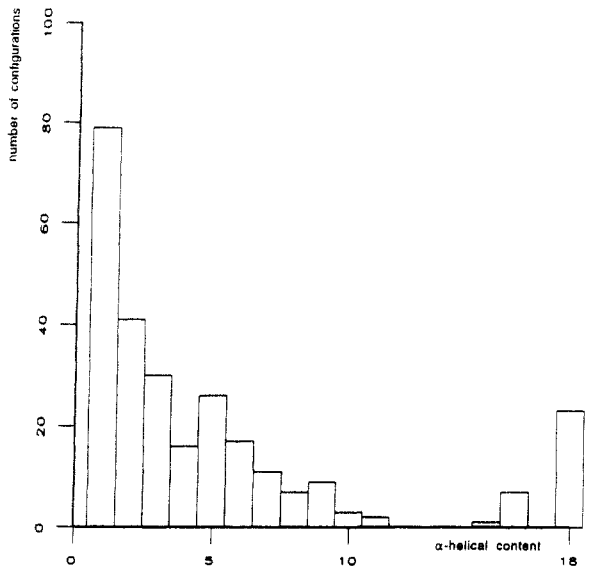

Figure 12. Frequencies of encountering an α-helical portion of length n among the $\epsilon = 10$ minimized configurations. ($n = 18$ corresponds to a complete α-helix.)

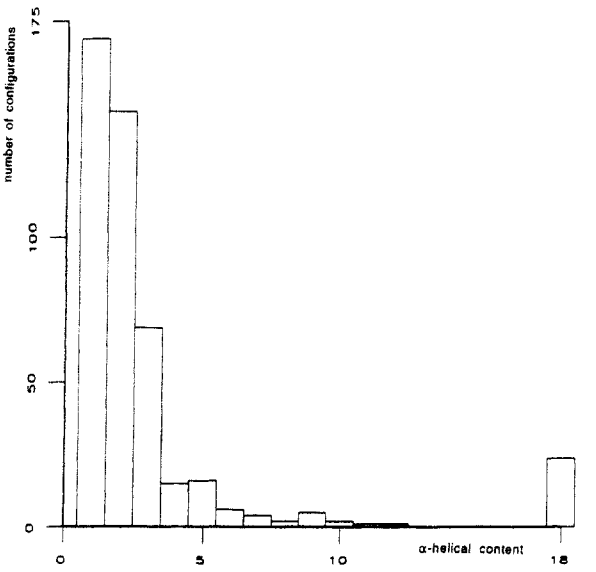

Figure 13. Frequencies of encountering an α-helical portion of length n among the $\epsilon = 80$ minimized configurations. ($n = 18$ corresponds to a complete α-helix.)

specifically for α-helices. Unlike in the $\epsilon = 1$ case, where complete α-helices ($n = 18$) could easily be found by whichever method, here only very few (6 out of 23) are found by nonbiased methods for $\epsilon = 10$, and 6 out of 24 for $\epsilon = 80$. (Remember that this count is different from the count of section 3.2, where all configurations converging to identical minima were identified.) The helical content distribution falls rapidly for $\epsilon = 80$, with some hints of a rise around $n = 16$ for $\epsilon = 10$. Nevertheless, the $n = 18$ full α-helices remain the deepest minima in all cases: it remains to find them. This means that while the position of the deepest minimum cluster does not change much with ϵ, its width falls rapidly.

Energywise, the deepest minimum found is −113.285 kcal/mol for $\epsilon = 10$, with reasonably helical configurations running up to −105.599 kcal/mol, and −68.519 kcal/mol for $\epsilon = 80$, with reasonably helical configurations running up to −63.525 kcal/mol.

3.5. Role of Various Energy Terms in Stabilizing Configurations. The energy components for a sample of 11 lower-energy minimized configurations are given in Table VI. These configurations are completely or almost

Table VI. Energy Components (kcal/mol) for the Lower-Energy (Mostly α-Helical) $\epsilon = 1$ Configurations

energy	bonds	angles	dihedrals	impropers	vdW	Coulomb
−591.202	0.277	4.070	5.567	1.160	−59.021	−543.255
−591.099	0.275	3.928	5.642	1.201	−58.517	−543.628
−590.401	0.315	4.025	5.863	1.217	−58.223	−543.599
−590.135	0.297	3.988	5.659	1.273	−57.841	−543.511
−589.997	0.289	3.889	5.628	1.254	−58.263	−542.795
...	...	...	...	...	...	...
−572.478	0.452	4.712	6.755	1.418	−50.313	−535.501
−572.430	0.445	5.197	6.878	1.634	−55.455	−531.129
−572.314	0.462	4.538	6.528	1.518	−55.084	−530.275
−572.244	0.465	4.536	6.508	1.516	−55.013	−530.256
−572.014	0.446	7.293	6.650	1.662	−54.686	−533.378
−571.371	0.447	4.864	7.218	1.760	−54.144	−531.515
Values Averaged over 133 Lower Minima						
−580.967	0.398	4.733	6.652	1.440	−55.008	−539.182

Table VII. Energy Components for Typical Lower-Energy (α-Helical) $\epsilon = 10$ and $\epsilon = 80$ Configurations

ϵ	energy	bonds	angles	dihedrals	impropers	vdW	Coulomb
10	−111.748	0.626	5.153	4.440	0.132	−70.719	−51.379
10	−111.555	0.616	4.710	4.177	0.116	−70.752	−50.421
10	−111.493	0.560	4.739	4.278	0.122	−70.750	−50.441
	...	...	...	...	...	...	...
10	−54.822	1.220	7.558	4.813	0.155	−26.431	−42.137
10	−54.034	1.242	7.831	5.039	0.170	−26.148	−42.168
10	−53.715	1.243	7.815	4.848	0.155	−25.639	−42.137
80	−67.491	0.635	4.988	3.791	0.086	−70.700	−6.291
80	−68.519	0.616	5.108	4.039	0.095	−72.016	−6.361
80	−67.613	0.671	5.018	4.025	0.089	−71.154	−6.261
	...	...	...	...	...	...	...
80	−60.446	0.847	5.386	5.662	0.050	−66.622	−5.770
80	−59.537	0.997	6.641	8.864	0.143	−70.464	−5.717
80	−59.132	0.997	5.920	7.881	0.113	−68.225	−5.817

α-helical. We can see that the Coulombic interaction is about 0.92 of the total energy, and the van der Waals interaction about 0.09 of the total. The other components can almost be neglected. Even though the absolute value of energy in CHARMM has no physical sense and could be arbitrarily shifted, it is still true that the Coulombic forces give the main contribution to the energy differences between configurations (11.74 kcal/mol out of 19.83 kcal/mol between the lowest and the highest configurations of Table VI). So in this case, we could think that the helical shape is created mainly by electrostatics.

The situation, however, is not that simple: exactly the same form of the helical lowest minimum is stabilized in the case of larger ϵ values mostly by nonelectrostatic components.

Looking at Table VII, where we give the same kind of splitting the energy into components for $\epsilon = 10$ and 80, we can see that in the $\epsilon = 10$ case the energy difference of 58.03 kcal/mol between the lowest and highest configurations of Table VII is due mostly to the Lennard-Jones forces (45.08 kcal/mol). In the $\epsilon = 80$ case, the energy difference is 8.36 kcal/mol and consists mostly of Lennard-Jones (2.47 kcal/mol) and dihedral (4.09 kcal/mol) contributions. The Coulombic contribution is negligible (0.4 kcal/mol). So in this case, the reason for obtaining the same shape of the molecule is quite different. In all cases we can speak about *cooperativity* of various energy components in producing a helical shape. The Ramachandran plot for a single residue, with no dipole–dipole interactions, already gives us a setup for producing a helix; in the case of strong electrostatics, as the molecule becomes longer, Coulombic forces contribute to maintaining an α-helical shape. This result is in agreement with previous studies.[6,19]

4. Conclusions

We studied here the energy-minimized structures of (L-alanine)$_{20}$. We confirmed that the lowest-energy structures are α-helices for a wide range of the dielectric constant. We found, however, that the role of different energy components in stabilizing these structures is quite different. At the same time, the value of ϵ plays little role in the minima distribution in the configuration $\phi\psi$ space. Our main results bear upon the evidence for ultrametric behavior of lowest-energy, α-helical structures and, independently, of higher-energy, disordered structures. We give a qualitative argument for ultrametricity of lower-energy configurations.

We also find, in agreement with previous studies of the critical length for polyalanine and other polypeptide molecules, that the α-helical structure becomes favorable once it has reached a certain length (around ten residues).

Acknowledgment. We would like to thank J. Smith for useful discussions and suggestions.

References and Notes

(1) Head-Gordon, T.; Stillinger, F. H.; Wright, M. H.; Gay, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 11513.
(2) Velikson, B.; Garel, T.; Niel, J.-C.; Orland, H.; Smith, J. C. *J. Comput. Chem.* **1992**, *13*, 1216.
(3) Arridge, R. G.; Cannon, C. G. *Proc. R. Soc. London* **1964**, *A278*, 91.
(4) Piela, L.; Scheraga, H. A. *Biopolymers* **1987**, *26*, S33.
(5) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283.
(6) Rapaport, D. C.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 1238.
(7) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, B.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(8) Smith, J. C.; Karplus, M. *J. Am. Chem. Soc.* **1992**, *114*, 801.
(9) Reiher, W. Ph.D. Thesis, Harvard University, Cambridge, MA, 1986.
(10) Bascle, J.; Garel, T.; Orland, H.; Velikson, B. Preprint SPhT/92-153, Saclay, 1992, accepted for publication in *Biopolymers*.
(11) Van Kempen, N. G. *Stochastic Processes in Physics and Chemistry*; North-Holland: Amsterdam, 1981, and references therein.
(12) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318.
(13) Noguti, T.; Go, N. *Proteins* **1989**, *5*, 104.
(14) Rammal, R.; Toulouse, G.; Virasoro, M. A. *Rev. Mod. Phys.* **1986**, *58*, 765 and references therein.
(15) Mézard, M.; Parisi, G.; Sourlas, N.; Toulouse, G.; Virasoro, M. *Phys. Rev. Lett.* **1984**, *52*, 1156; *J. Phys. (Paris)* **1984**, *45*, 843.
(16) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(17) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.
(18) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 1115.
(19) Ooi, T.; Scott, R. A.; Vanderkool, G.; Scheraga, H. A. *J. Chem. Phys.* **1967**, *46*, 4410.