Journal of Global Optimization **15:** 261–297, 1999.

Dynamics of Peptide Folding: Transition States and Reaction Pathways of Solvated and Unsolvated Tetra-Alanine

K.M. WESTERBERG and C.A. FLOUDAS
Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544, USA

(Accepted in original form 5 July 1999)

Abstract. A new approach is proposed for enclosing all stationary states, including saddle points of all orders, of a potential energy surface based on the αBB deterministic branch and bound global optimization algorithm. This method is based on rigorous optimization methods and offers a theoretical guarentee of enclosing all solutions to the equation $\nabla V = 0$. This method is applied to the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential energy surfaces of unsolvated and solvated tetra-alanine. By analyzing the topography of the potential energy surfaces, we calculate reaction pathways, transition rate matrices, time-evolution of occupation probabilities, and rate disconnectivity graphs, and we identify appropriate criteria for the selection of a reaction coordinate.

Key words: Global Optimization, Dynamics, Peptide folding, Tetra-alanine

1. Introduction

The protein folding problem is a very important problem in computational chemistry and molecular biology. The ability of a protein to function properly within the cell depends on its tertiary structure. Considering how precisely and reliably a protein shapes itself to perform its specific task, very little is understood about the mechanism of protein folding. Better understanding and insight on the mechanism of protein folding are of major importance.

A promising approach to understanding protein folding is the study of its potential energy surface. The first step in the study of any potential energy surface is the identification of stationary points (local minima and saddle points), since these points play a crucial role in defining the topography of the surface. The local minima represent stable configurations of the protein molecule, and the first-order saddle points generally correspond to transition states which connect two such configurations. A protein-folding process can be thought of as a transition between two local minima through a transition state, or a series of such transitions.

We study further a recently proposed approach [1] of finding all stationary points of a given potential energy surface based on the αBB deterministic branch and bound type global optimization algorithm [2–6]. This is an application of the

more general method of finding all solutions to systems of non-linear equations described in [2], which we will describe in Sections 2 and 3.

Once the minima and saddles have been located, we follow each saddle point back to the two minima it connects. It is worth noting that searching downhill from transition states towards minima is always more reliable than searching uphill from a given minimum towards a first-order transition state. If one has already located all of the first-order transition states, it is not necessary to start at the minima and proceed uphill. After determining the connectivity of the potential energy surface, transition rates between minima can be calculated using the Rice–Ramsperger–Kassel–Marcus (RRKM) theory [7, 8].

The transition rates can be used to calculate occupation probabilities for each state as a function of time given some specific initial condition (such as the system occupying a particular state with probability 1). This gives us a direct indication of how long it takes for a protein prepared in a given unfolded (excited) state to reach its folded (ground) state. It is also possible to calculate the time evolution of other quantities, such as average energies, atomic distances, and (dihedral) angles.

The connectivity of the potential energy surface can be understood by enumerating the reaction pathways. Reaction pathways are obtained by joining minimum–saddle–minimum 'triples' together in chains. These pathways represent the protein folding mechanism and are very important in understanding how the protein folds.

Becker and Karplus proposed a graphical representation of the topography of a potential energy surface [9] based on the connectivity tree originally introduced by Czerminiski and Elber [10]. They define a finite energy (temperature) generalization of the 'catchment region'. As the energy (temperature) is increased, regions that were once disconnected by high barriers begin to merge. This coalescence process is described by means of a 'energy (temperature) disconnectivity graph'. The shape of the disconnectivity graph reveals an enormous wealth of dynamical information. We extended this idea in [1] by constructing a 'rate disconnectivity graph' which is based on transition rates, rather than energy levels or barrier heights.

We have applied these methods to the Murrel-Sorbie analytic potential energy surfaces of triatomic molecules and to the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential energy surfaces of alanine, alanine dipeptide, and unsolvated and solvated tetra-alanine. A full analysis of the triatomic molecules, alanine, alanine dipeptide, and unsolvated tetra-alanine can be found in [1]. In Sections 4 and 5 of this paper, we study the unsolvated and solvated tetra-alanine.

2. Problem formulation

Stationary points of all orders (i.e., minima, maxima, first order and higher order transition states) of a given potential energy surface $V(\mathbf{x})$ are determined by the

constraints

$$\frac{\partial V}{\partial x_i} = 0\,, \qquad i = 1, \ldots, N_x \tag{1}$$

where N_x is the number of variables: $\mathbf{x} = (x_1, \ldots, x_{N_x})$.

Equation (1) can be re-expressed as a global optimization problem by introducing a slack variable s and minimizing its value over an augmented variable set $(\mathbf{x}, s)$ subject to a set of relaxed constraints:

$$\begin{aligned} & \min_{\mathbf{x},s} s \\ \text{subject to} \quad & \partial V/\partial x_i - s \leqslant 0\,, \qquad i = 1, \ldots N_f \\ & -\partial V/\partial x_i - s \leqslant 0\,, \qquad i = 1, \ldots N_f \\ & \mathbf{x}^L \leqslant \mathbf{x} \leqslant \mathbf{x}^U \end{aligned} \tag{2}$$

Note that (2) is infeasible for $s < 0$ and reduces to (1) for $s = 0$. It follows that $s = 0$ is the *global minimum* of (2), provided of course that (1) has solutions, and that there is a one-to-one correspondence between global minima of (2) and solutions to (1). Therefore, the problem of finding all solutions to (1) can be reformulated as the problem of finding all global minima of (2). We will exploit this fact in the next section where we explain how the αBB global optimization algorithm [1–6] can be used to find all solutions to (1).

3. The αBB global optimization approach

In this section, we describe the αBB global optimization algorithm [1–6] as it applies to the problem of finding all stationary points of a potential energy surface. This adaptation is based on the correspondence between solutions of (1) and global minima of (2) with $s = 0$. The αBB algorithm can actually be used to find all solutions to any system of non-linear equations, assuming only that the constraints are twice continuously differentiable (C^2). For a more complete explanation of the αBB algorithm, see [1].

The algorithm proceeds by exploring the configuration space for stationary points (i.e., solutions to (1)). We begin with the full region $\mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U]$, and subdivide regions into smaller regions. Each region is tested before it is divided to see if a stationary point can possibly exist there. This is accomplished by finding a lower bound of the global minimum of (2) over the region in question. If the lower bound is positive, then $s = 0$ cannot lead to a feasible point of (2), and hence no solution to (1) can exist in the given region. The region will be fathomed (i.e., eliminated from further consideration). On the other hand, if the lower bound is negative or zero, there may or may not be a stationary point in that region. In this case, further subdivision and testing will be necessary. Once the region size becomes small enough, the stationary point is sought by performing a local search in that region. The algorithm terminates when all regions have been fully processed.

Lower bounds of the global minimum of (2) are determined by solving the *lower bounding problem* over the given region. In the lower bounding problem, each constraint in (2) is replaced by a *convex underestimator*, which is obtained from the original constraint by subtracting off a sufficiently large quadratic term.

$$\begin{aligned} &\min_{\mathbf{x},s} s \\ \text{subject to} \quad &\partial V/\partial x_i - \alpha_i^+ \sum_k (x_k^U - x_k)(x_k - x_k^L) - s \leqslant 0 \\ &-\partial V/\partial x_i - \alpha_i^- \sum_k (x_k^U - x_k)(x_k - x_k^L) - s \leqslant 0 \\ &\mathbf{x}^L \leqslant \mathbf{x} \leqslant \mathbf{x}^U \end{aligned} \tag{3}$$

If the coefficients $\alpha_i^{\pm}$ are chosen to be positive and sufficiently large, the global minimum of (3) will be a valid lower bound of the global minimum of (2), and can be obtained by any local optimization package since (3) is a convex problem.

The basis of the αBB algorithm is the selection of valid values of $\alpha_i^{\pm}$. The lower bounding problem will be convex only if the Hessian matrix associated with each constraint in (3) is positive definite. The quadratic terms were introduced to overpower the non-convexities of the original constraints in (2). This is accomplished provided that $\alpha_i^{\pm}$ satisfy the following inequalities:

$$\begin{aligned} \alpha_i^+ &\geqslant -\frac{1}{2} \min_{\mathbf{x}\in[\mathbf{x}^L,\mathbf{x}^U]} \{\lambda_k(H_{\partial V/\partial x_i}(\mathbf{x})), 0\} \\ \alpha_i^- &\geqslant +\frac{1}{2} \max_{\mathbf{x}\in[\mathbf{x}^L,\mathbf{x}^U]} \{\lambda_k(H_{\partial V/\partial x_i}(\mathbf{x})), 0\} \end{aligned} \tag{4}$$

Calculating values of $\alpha_i^{\pm}$ according to (4) is difficult in general because the Hessian matrices $H_{\partial V/\partial x_i}$ depend on $\mathbf{x}$. If the selected values of $\alpha_i^{\pm}$ are too small (the inequalities are not satisfied), the 'convex underestimators' will not be convex after all, and thus there is no guarantee that using a local optimization solver will yield the global minimum of the lower bounding problem. Lower bounds generated this way may not be valid. On the other hand, if the selected values are too large, the underestimators will be convex, but they will be very loose. This may lead to poor computational performance of the algorithm.

A simplified method of calculating $\alpha_i^{\pm}$ is to start with small values of $\alpha_i^{\pm}$ (e.g., $\alpha_i^{\pm} = 5$) and increase the values of $\alpha_i^{\pm}$ until no new solutions are found. This can be a practical solution to many problems where the correct values of $\alpha_i^{\pm}$ are difficult to determine. However, this method has the one serious drawback in that it sacrifices the theoretical guarantee of finding *all* solutions. In spite of this fact, we were able to identify all minima and first-order transition states using modest values of $\alpha_i^{\pm}$ for alanine, alanine dipeptide, and tetra-alanine. The tetra-alanine results are described in Sections 4 and 5.

A more robust method involves calculating the Hessian matrices $H_{\partial V/\partial x_i}$ at various grid points to get a sample of required $\alpha_i^{\pm}$ values. First we select a grid, $\{\mathbf{x}^k\}$.

Table 1. Eigenmode III results for unsolvated tetra-alanine

	4^8 grid	6^8 grid
Local minima	16125	62373
1st-order saddles	18902	212938

Then we evaluate the Hessian for each constraint at each grid point, $H_{\partial V/\partial x_i}(\mathbf{x}^k)$, and use (4) to determine precomputed values of $\alpha_i^{\pm}(\mathbf{x}^k)$ at each grid-point. During the αBB run, appropriate values of $\alpha_i^{\pm}$ for a given region are determined by selecting the maximum $\alpha_i^{\pm}$ over all grid-points contained in the region. This method of generating $\alpha_i^{\pm}$ was used when we studied triatomic molecules. The results are discussed in [1].

It should be noted that the generation of values of $\alpha_i^{\pm}$ that are theoretically rigorous can be made based on the recent work of [5, 6].

4. Computational studies: Unsolvated tetra-alanine

We first studied tetra-alanine in vacuum. We used the ECEPP/3 potential energy surface [11], fixing all bond lengths and bond angles to their equilibrium values, and allowing only the eight (ϕ, ψ) dihedral angles to vary.

Tetra-alanine is one of the smallest peptides which can exhibit a full alpha-helical turn (corresponding to $(\phi_i, \psi_i) = (300^\circ, 300^\circ)$), as well as an extended conformation (a beta sheet conformation corresponding to $(\phi_i, \psi_i) = (300^\circ, 120^\circ)$). We studied unsolvated tetra-alanine in [1], and in this paper we present results on both solvated and unsolvated tetra-alanine.

We first obtained a testbed of minima and first-order saddles by applying a brute force eigenmode-following search (Eigenmode III [12–17]) using a grid of starting points. The results of our search are summarized in Table 1.

We first laid down a 4^8 grid of starting points and performed minimum and first-order saddle searches from each point. The transition states were then followed down to the minima they connect, resulting in additional minima found. The analysis of the 16125 minima and 18902 first-order saddles obtained from the 4^8 grid can be found in [1].

In this work, we considered a 6^8 grid of starting points and performed first-order saddle searches from each point. The minima were obtained by following each transition state down to the minima they connect. The results were then merged with the 4^8 grid results. The analysis of the 62373 minima and 212938 first-order saddles obtained from the 6^8 grid is presented in this paper.

We subdivided the (ϕ, ψ) plane into regions and classified them according to Table 2. Values of (ϕ, ψ) corresponding to alpha-helix formation are classified as 'a', and values of (ϕ, ψ) corresponding to beta-sheet formation are classified as

Table 2. Classification scheme for (ϕ, ψ) pair

Symbol	ψ	Decoration	ϕ
a	$270^\circ \leqslant \psi \leqslant 335^\circ$	No prime	$270^\circ \leqslant \phi \leqslant 330^\circ$
i	$335^\circ \leqslant \psi$ or $\psi \leqslant 90^\circ$	Prime	$180^\circ \leqslant \phi \leqslant 270^\circ$
b	$90^\circ \leqslant \psi \leqslant 150^\circ$	Double prime	otherwise
j	$150^\circ \leqslant \psi \leqslant 270^\circ$		

Table 3. Ground state and extended conformation of unsolvated tetra-alanine

Minimum	Classification	E (kcal/mole)	F (kcal/mole)
min.1	aaaa	−6.643	−11.798
min.1587	bbbb	4.916	−5.549

Table 4. Selected results from αBB tetra-alanine runs

Region	Type	Eigenmode III	αBB	α
aaaa	min	1	1	25
bbbb	min	1	1	20
	1st	4	4	
	2nd	6	6	
	3rd	4	4	
	4th	1	1	
bibi	min	1	1	20
	1st	1	2	
	2nd	0	1	
bbbj′	min	2	2	20
	1st	8	9	
	2nd	4	17	
	3rd	3	16	
	4th	2	7	
	5th	0	1	
aai′i	min	2	2	80
	1st	1	1	

'b'. Each conformation of tetra-alanine is characterized by four (ϕ, ψ) pairs, and hence can be classified by a concatination of four symbols.

Of the 62373 minima, we found one alpha-helical conformation, min.1 (aaaa), and one extended conformation, min.1587 (bbbb). Their potential energy and free energy values can be found in Table 3. The alpha-helix conformation is the lowest energy conformation of tetra-alanine. We will be concentrating on the folding process from the extended conformation to the ground state.

We checked the αBB algorithm against the Eigenmode III search for stationary points by conducting αBB runs on selected regions of the potential energy surface. Selected results are given in Table 4. We started with a constant value of $\alpha = 20$, and then increased α in subsequent runs until we found all stationary points located by the Eigenmode III search. In all cases, modest values of α (less than 100) were sufficient to locate all minima and first-order saddles found by Eigenmode III. In many cases, additional saddle points were located.

Having now identified the local minima and first-order transition states, we are now in a position to enumerate the reaction pathways between states and calculate transition rates. The connectivity between the various minima is determined by following each transition state back to the minima they connect. This is accomplished by perturbing the transition state slightly in each of the two directions along the reaction coordinate, and then using Eigenmode III to locate a local minimum from that starting point. This gives us a list of (minimum, transition state, minimum) triples.

We can then calculate the transition rate matrix using Rice-Ramsperger-Kassel-Marcus (RRKM) theory (see [1] for a more complete discussion). According to RRKM theory, the transition rate for a single transition is given by

$$W_{j' \to \mathrm{ts} \to j} = \frac{kT}{h} \frac{Q_{\mathrm{ts}}}{Q_{j'}}. \tag{5}$$

The partition functions at the minima and first-order saddles are related to the free-energies of those stationary points, and can be evaluated using the harmonic approximation

$$Q = e^{-F/kT} = e^{-E/kT} \prod_i \frac{kT}{h\nu_i}, \tag{6}$$

where E and F are the potential energy and free energy, respectively, of the stationary point, and ν_i are the vibrational frequencies of the molecule around the stationary point. The product over frequencies takes into account the vibrational entropy of the system. Substituting (6) into (5) yields

$$W_{j' \to \mathrm{ts} \to j} = \frac{\prod_i \nu_i^{j'}}{\prod_{i \neq \mathrm{r.c.}} \nu_i^{\mathrm{ts}}} e^{-(E_{\mathrm{ts}} - E_{j'})/kT}.$$

Summing over all transition states connecting two particular minima yields the transition rate matrix

$$W_{jj'} = \sum_{\text{ts}} W_{j' \to \text{ts} \to j}.$$

The time-evolution of occupation probabilities can be calculated by solving the Master equation

$$\frac{dP_j}{dt} = w_{jj'} P_{j'}(t), \tag{7}$$

where

$$w_{jj'} = \begin{cases} W_{jj'} & \text{if } j \neq j' \\ -\sum_{j''} W_{j''j} & \text{if } j = j'. \end{cases}$$

Coupled differential equations like (7) are solved by diagonalizing the matrix $w_{jj'}$, so that

$$\sum_{j'} w_{jj'} u_{j'}^{(i)} = \lambda^{(i)} u_j^{(i)}.$$

The general solution to (7) can be written in the form

$$P_j(t) = \sum_i a_i e^{\lambda^{(i)} t} u_j^{(i)}, \tag{8}$$

where the coefficients a_i are determined by the initial probability distribution at $t = 0$.

One of the eigenvalues $\lambda^{(0)}$ is zero. The associated eigenvector corresponds to the equilibrium ($t = \infty$) probability distribution,

$$u_j^{(0)} = P_j(+\infty) = Q_j / \sum_{j'} Q_{j'}.$$

All other eigenvalues are negative, and correspond to transient probabilities with a decay time of $\tau^{(i)} = -1/\lambda^{(i)}$.

The time-evolution of occupation probabilities for the extended conformation and the 3 lowest free energy states of unsolvated tetra-alanine at room temperature $T = 300\,\text{K}$, starting with the extended conformation at $t = 0$ (i.e., $P_{\text{bbbb}}(0) = 1$, all other $P_j(0) = 0$) is given in Figure 1. It takes tetra-alanine about 10^{-10} s to reach the ground state from the extended conformation.[1]

Details of the folding process can be determined by enumerating the pathways from the extended conformation to the ground state. A pathway is defined as a sequence of minima joined together by transition states.

initial state → ts → min → ts → min → ... → min → ts → final state

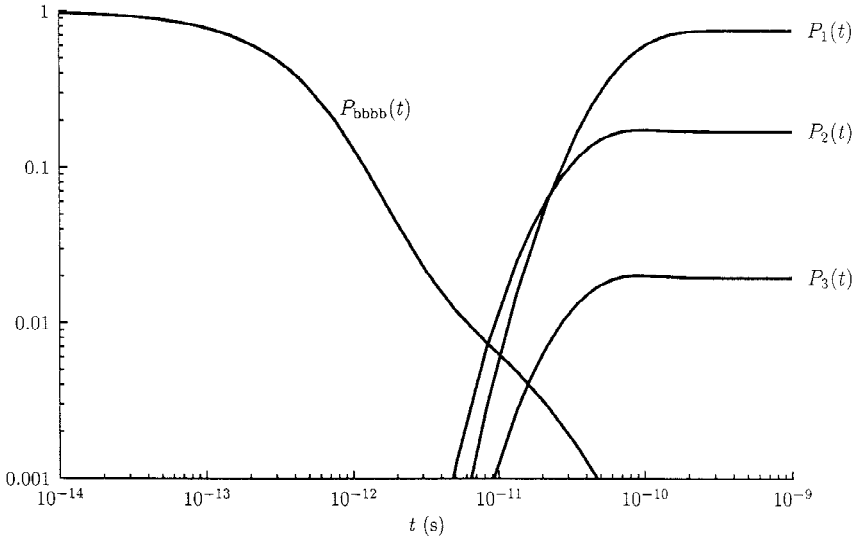

Figure 1. Time evolution of the extended conformation and the 3 lowest free energy states of unsolvated tetra-alanine at $T = 300$ K.

Pathways between these two states can be enumerated using graph-theory techniques. We construct a graph where each node in the graph represents a minimum and each edge in the graph represents a transition state which connects two minima. The set of all pathways which from one minimum to another can be generated by an exhaustive search.

If we conduct this exhaustive search without restriction, we would generate an enormous number of pathways. It is important for us to be able to restrict the pathways we generate in a sensible manner. We selected pathways based on two criteria: (1) we restrict the length of the pathway (i.e., number of minima) to be less than or equal to some prescribed maximum length, and (2) we also apply a transition rate cutoff, effectively ignoring transitions whose rates fall below the cutoff value. The number of pathways from the extended conformation to the ground state of unsolvated tetra-alanine at $T = 300$ K for various length and rate cutoffs is given in Table 5. The total number of minima and transition states involved in such pathways are given in Table 6.

These two criteria were applied in an attempt to find the most relevant pathways. Since the faster pathways are likely to be the most important ones, it makes sense to eliminate pathways which involve one or more slow transitions (i.e., transitions which fail to meet the rate cutoff). The length cutoff is chosen for more practical reasons. Even with a transition rate cutoff, the number of pathways increases exponentially with the length cutoff (about a factor of 10 for each additional minimum). An exhaustive pathway search would be intractable if we did not impose a length cutoff. It is assumed that the fastest pathways are also among the shortest in length.

Table 5. Number of pathways from extended conformation to ground state with given length restriction and rate cutoff

Maximum length	No cutoff	10^6 Hz	10^7 Hz	10^8 Hz	10^9 Hz	10^{10} Hz	10^{11} Hz
6							
7	4						
8	38						
9	999	421	421	421	421	285	130
10	19963	10836	10828	10828	10733	7443	2099
11	297974	150831	150396	149391	146493	92216	21004
12	4132256	1868821	1859469	1832692	1768736	1002874	221592

Table 6. Number of minima / transition states involved in pathways from the extended conformation to ground state with given length restriction and rate cutoff

Maximum length	No cutoff	10^6 Hz	10^7 Hz	10^8 Hz	10^9 Hz	10^{10} Hz	10^{11} Hz
6							
7	12/ 14						
8	26/ 42						
9	236/ 488	96/ 183	96/ 183	96/ 183	96/ 183	86/ 160	65/ 114
10	886/ 2339	339/ 952	339/ 951	339/ 951	332/ 932	287/ 790	188/ 466
11	2817/ 8341	664/2177	663/2173	657/2152	651/2120	526/1696	357/1044
12	6403/21316	943/3405	938/3388	922/3341	913/3291	754/2622	509/1699

Although we have no proof of this,[2] we will see evidence later on that suggests that we have found the most relevant pathways.

We examined in detail the pathways of length 9 and 10 with a transition rate cutoff of 10^6 Hz. An example pathway of length 9 is given in Figure 2.

For each such pathway, we estimated the amount of time it would take for tetra-alanine to proceed from the extended conformation to the ground state along that particular pathway by solving the Master equation for a reduced system consisting only of the minima and transition states involved in the pathway. The decay time of the longest lived transient probabilities was used as an estimate of the overall transition time. The fastest transition times were on the order of 5×10^{-11} s, and most of the 10836 pathways we looked at had transition times less than 1×10^{-9} s. Clearly there is no single most important pathway: there are many pathways which are all equally important. We also found that the pathways of length 9 tended to be among the fastest of the pathways of length 10 or less, suggesting that shorter pathways tend to be faster.

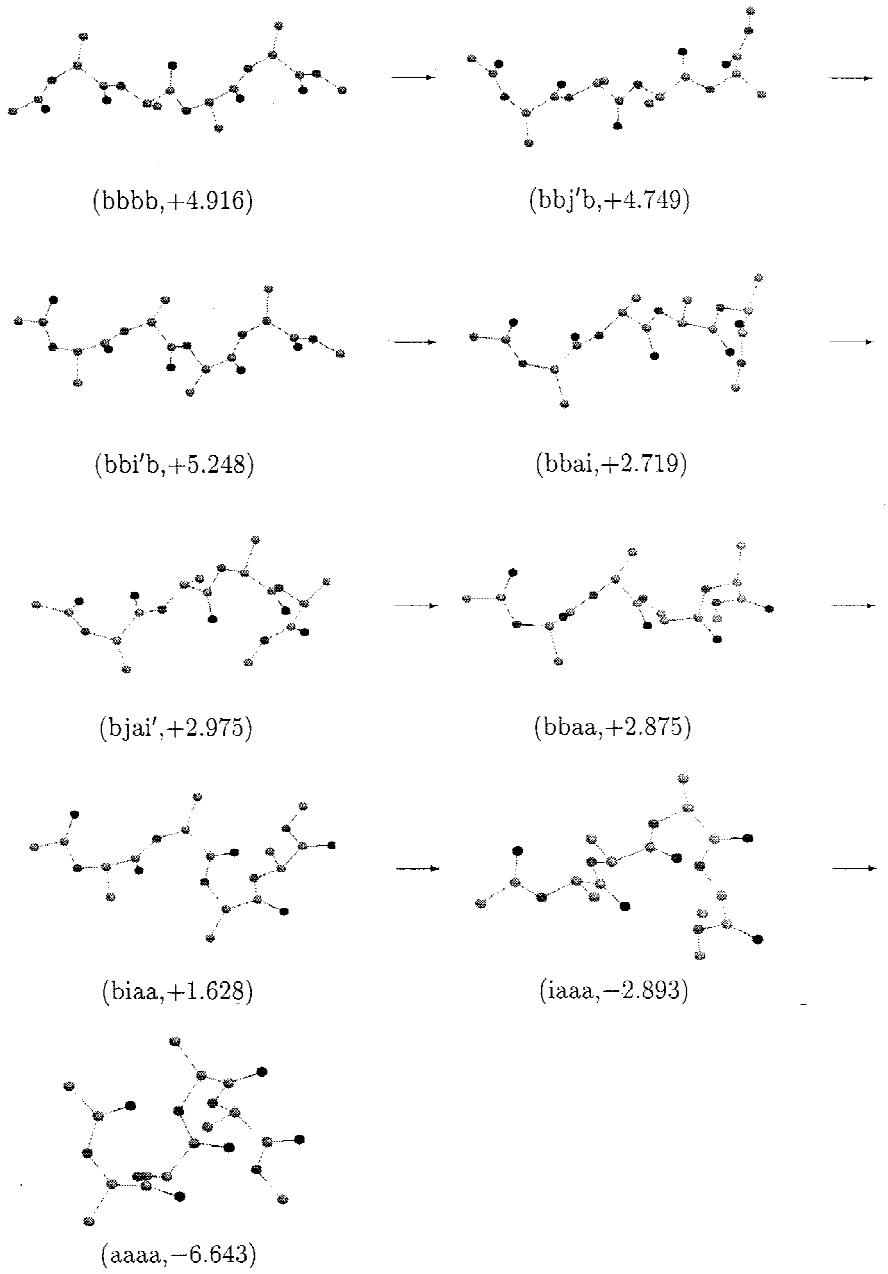

Figure 2. One possible pathway from the extended conformation to the ground state of unsolvated tetra-alanine.

We also studied the pathways of length 10 or less in terms of changes in the ϕ and ψ angles. Each (ϕ, ψ) pair is classified according to Table 2. In proceeding from the extended conformation to the ground state, each of the four (ϕ, ψ) pairs must proceed from 'b' to 'a'. We observed that this process tends to follow regular patterns.

We make the following general observations regarding the rotation of the ψ angles:

1. Each ψ angle normally progresses in the sequence b $\rightarrow$ i $\rightarrow$ a or b $\rightarrow$ j $\rightarrow$ i $\rightarrow$ a.
2. No direct b $\rightarrow$ a transitions are observed,[3] indicating that a rotation of ψ from beta-sheet to alpha-helical values is too large for a single transition.
3. Most pathways of length 10 or less involve at least one transition where more than one ψ angle changes (cooperative motion).
4. A wide variety of cooperative motion is possible, but the two most common types are as follows:

   bi $\rightarrow$ ia 36%

   bj $\rightarrow$ ii 14%

5. There is a tendency for one half of the molecule to fold (nearly) completely followed by the other half (e.g., bbbb $\rightarrow$ bbaa $\rightarrow$ aaaa).[4]

We can analyze the pathway given in Figure 2 in terms of these observations. The individual ψ angles proceed as follows:

ψ_1 : b $\rightarrow$ i $\rightarrow$ a

ψ_2 : b $\rightarrow$ j $\rightarrow$ b $\rightarrow$ i $\rightarrow$ a

ψ_3 : b $\rightarrow$ j $\rightarrow$ i $\rightarrow$ a

ψ_4 : b $\rightarrow$ i $\rightarrow$ a

Except for a slight backtrack in ψ_2, this pathway is consistent with (1) and (2). This pathway also exhibits 3 transitions which involve cooperative motion. Two of them are in the form bi $\rightarrow$ ia, which is the most common form observed. The other cooperative motion, ji $\rightarrow$ ba (non-adjacent alanines), has also been observed but is not nearly as common as the two forms listed above. Finally, it should be remarked that this pathway does pass through a bbaa minimum. In other words, the right side (the carboxyl terminus) folds completely before the left side (the amino terminus) folds at all. Not all pathways follow this rule strictly, although we have found that tetra-alanine tends to fold its right side most of the way before its left side makes significant progress.

The rotation of the ϕ angles plays less of a role in the folding process than rotation of ψ angles. ϕ takes on similar values for alpha-helical and beta-sheet conformations. We found that the very slowest transitions (on the order of 100 Hz or less) tend to involve rotations of the ϕ angles from inside to outside of the range $180^\circ \leqslant \phi \leqslant 330^\circ$ and vice versa. In fact, none of the minima involved in pathways

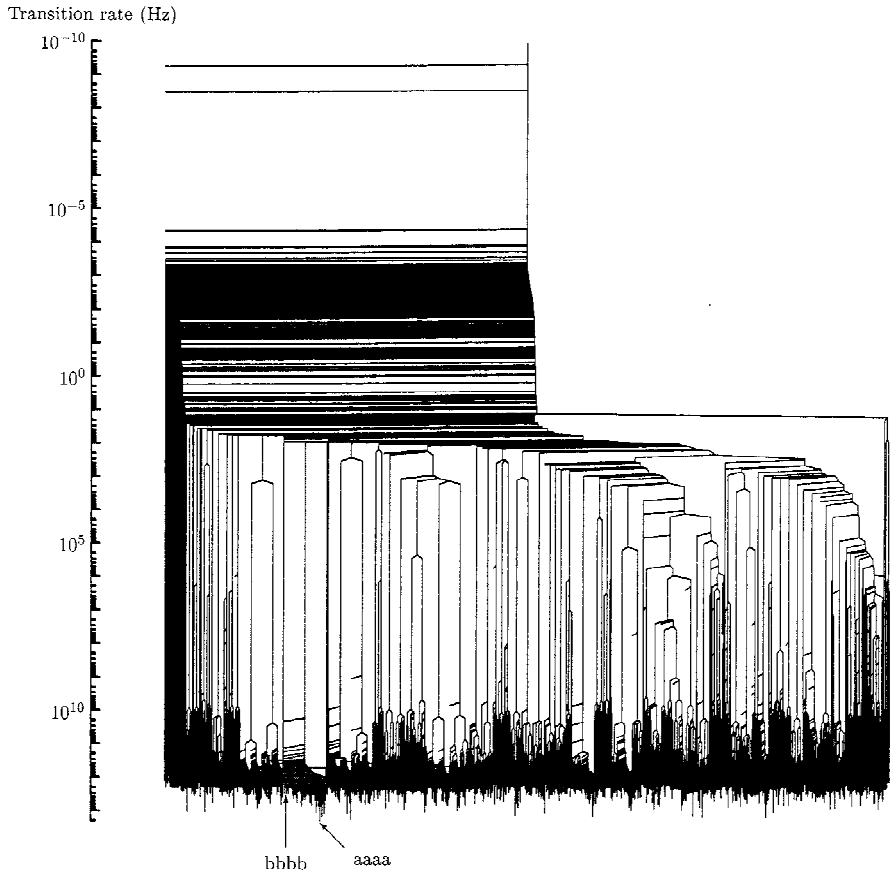

Figure 3. Complete rate disconnectivity graph for unsolvated tetra- alanine at $T = 300$ K. The alpha-helical ground state and the extended conformation both lie in the highlighted subtree.

of any length with a rate cutoff of 10^6 Hz involves ϕ angles outside this range (they would indicated in our classification scheme by a double-prime). This can be proved rigorously by examination of the rate disconnectivity graph, which we will discuss next.

We constructed the rate disconnectivity graph for tetra-alanine at $T = 300$ K. It is shown in Figure 3. The rate disconnectivity graph provides us with the rate-dependent connectivity of the potential energy surface [1, 9, 10]. If we begin at the top of the graph, with a very small rate cutoff, all of the minima fall into one group which is represented by a single node. As we increase the rate cutoff, transitions get eliminated. At some point, a critical transition gets eliminated which disconnects the minima into two groups. This is represented by the node splitting into two at the rate cutoff value. As the rate cutoff is increased further, more and more transitions are eliminated and the graph continues to bifurcate as the groups of minima further

subdivide. At the base of the graph, no transitions remain, and each minimum falls into its own group. The minima can be identified at the base of the graph.

The rate disconnectivity graph for tetra-alanine shown in Figure 3 covers 23 orders of magnitude in transition rates, and contains 62357 minima.[5] Starting at the top, we see that a relatively small number of minima break away as the rate cutoff is increased to around 10 Hz. Between 10 Hz and 100 Hz, a number of large groups of minima (several thousand minima each) break away from the main branch, indicating a great deal of interesting dynamics occuring on a time scale of about 0.1 s. Between 10^2 Hz and 10^{10} Hz relatively little happens. There seems to be two well-separated time scales with characteristic times roughly 0.1 s and 10^{-10} s.

The highlighted section of the rate disconnectivity graph contains a total of 3713 minima, including the extended conformation and the alpha-helical ground state. If we apply a transition rate cutoff anywhere between 10^2 Hz and 10^{10} Hz, we would find that all of the minima in the highlighted region would be connected to one another, and disconnected from all of the rest. In other words, it would take about 10^{-10} sec to make transitions between two minima within this group, and about 10^{-2} sec to make transitions out of this group. This is consistent with our solution of the master equation (see Figure 1).

We looked for a distinguishing characteristic of the minima within this group. We found that all 3713 minima in this group satisfy the constraints

$$180^\circ \leqslant \phi_i \leqslant 330^\circ$$

for all four ϕ angles. Conversely, we found that all except for one minimum which satisfies these constraints on all four ϕ angles lies within this group. This leads us to the following conclusions:

1. Transitions involving large changes in ϕ (from within $[180^\circ, 330^\circ]$ to outside this range, or vice versa), tend to be very slow, requiring longer than 0.01 sec (sometimes much longer). This is no doubt a result of very high barriers separating these two regions of configuration space.
2. Transitions involving small changes in ϕ (i.e., those which stay within the range $[180^\circ, 330^\circ]$), and arbitrary changes in ψ, tend to be much faster, typically on the order of 10^{-10} sec. The folding of tetra-alanine from its extended conformation (bbbb) to the ground state (aaaa) falls into this catagory.

Another way of obtaining an overall picture of the folding process of tetra-alanine is to study the time-evolution of averages of certain quantities, such as energy, dihedral angles, or distances between specific atoms. If q_j is the value of some quantity at minimum j, then $\langle q \rangle$, the average value of q, and σ_q, the standard deviation, can be calculated as a function of time with the help of the Master equation:

$$\begin{aligned} \langle q \rangle(t) &= \sum_j P_j(t) q_j = \sum_{i,j} a_i e^{\lambda^{(i)} t} u_j^{(i)} q_j \\ &= \sum_i a_i \Big(\sum_j u_j^{(i)} q_j \Big) e^{\lambda^{(i)} t} \end{aligned} \tag{9}$$

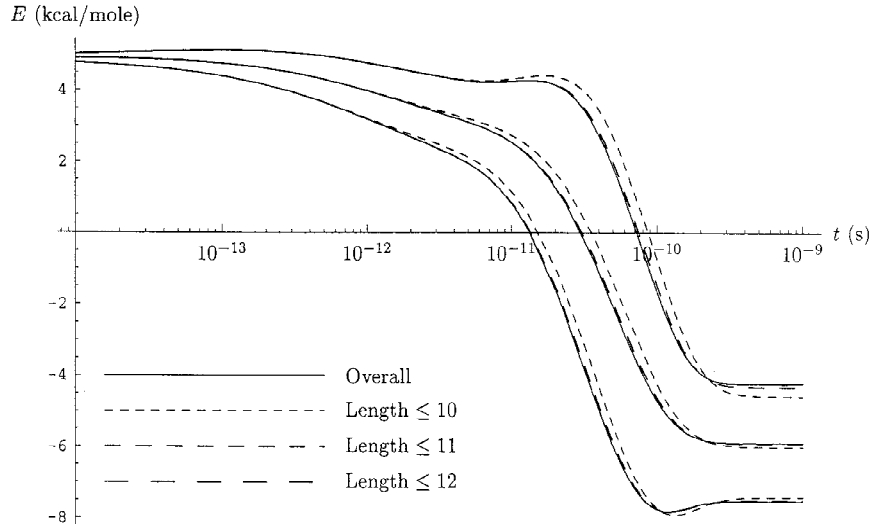

Figure 4. Time-evolution of E as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

$$\langle q^2 \rangle(t) = \sum_i a_i \Big(\sum_j u_j^{(i)} q_j^2 \Big) e^{\lambda^{(i)} t} \tag{10}$$

$$\sigma_q(t) = \sqrt{\langle q^2 \rangle(t) - \langle q \rangle^2(t)} \tag{11}$$

Plots of $\langle q \rangle$ and $\langle q \rangle \pm \sigma_q$ as functions of time for $q = E, \phi_i$, and ψ_i are given in Figures 4–12.

To obtain the correct time evolution of $\langle q \rangle$ and σ_q, it is necessary to solve the Master equation over all of the minima.[6] We can also calculate the approximate time evolution of $\langle q \rangle$ and σ_q by restricting our attention to only a certain subset of pathways. This is accomplished by restricting the minima and transition states we use to solve the Master equation to those which are visited by the selected pathways.

In Figures 4–12, we compare the overall time evolution of E, ϕ_i and ψ_i with the time evolution obtained by restricting our attention to pathways with various length restrictions. The deviations are rather large for a length cutoff of 10, but are much smaller for a length cutoff of 11 or 12. It appears that applying a length cutoff of 11 will yield most of the relevant pathways.

We can also determine the effect of various transition rate cutoffs on the time evolution of E, ϕ_i, and ψ_i. In Figures 13–21, we compare the overall time evolution of E, ϕ_i, and ψ_i with that obtained by restricting our attention to pathways with a length cutoff of 11 and various transition rate cutoffs. We find significant deviation from the overall time evolution only when the transition rate cutoff is increased to 10^{11} Hz. It appears that the most significant pathways are those of length 11 or less

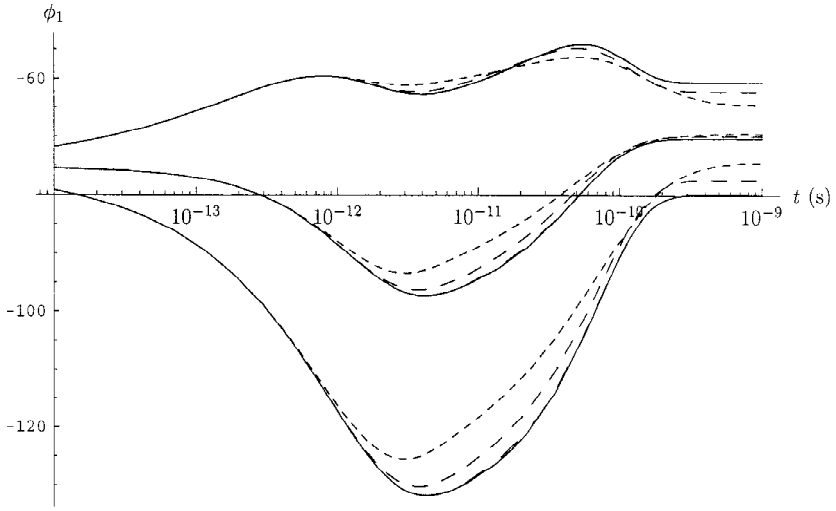

Figure 5. Time-evolution of ϕ_1 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

which satisfy a transition rate cutoff of 10^{10} Hz. There are 92216 such pathways, and they involve only 526 minima and 1696 transition states. This is significantly less than the 62373 minima and 212938 transition states we started with.

It would be useful to characterize the folding process by means of determining a viable reaction coordinate. A reaction coordinate is a quantity which accurately measures the progress from the initial state to the final state. Ideally, it should be monotonic and proceed at a uniform rate along each individual pathway. If we examine the time evolution of E, ϕ_i, and ψ_i, we see that the energy and the ψ angles seem to make reasonable reaction coordinates, but the ϕ angles definitely do not. However, these plots only reveal the average progress of these quantities. What we would really like to know is which, if any, of these quantities proceeds monotonically and uniformly for *each* pathway.

To help answer this question, we developed two 'reaction coordinate indicators', one which measures the monotonicity of the reaction coordinate, and the other which measures the uniformity of the reaction coordinate. For a given pathway of length N

$$\min_1 \rightarrow \min_2 \rightarrow \ldots \rightarrow \min_N$$

a certain quantity q takes on values

$$q_1 \rightarrow q_2 \rightarrow \ldots \rightarrow q_N.$$

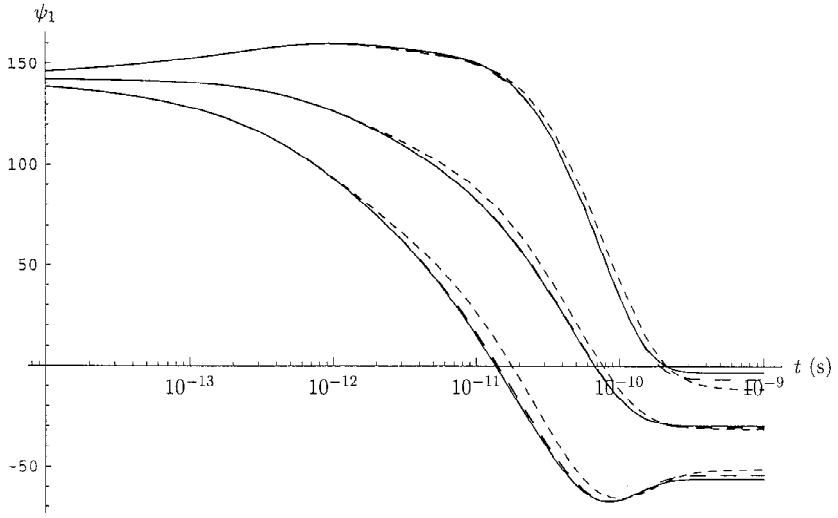

Figure 6. Time-evolution of ψ_1 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

The two reaction coordinate indicators are d/D and D^2/S, where

$$d = \left| \sum_{i=1}^{N-1} (q_{i+1} - q_i) \right| \qquad \text{(displacement)}$$

$$D = \sum_{i=1}^{N-1} |q_{i+1} - q_i| \qquad \text{(distance)}$$

$$S = (N-1) \sum_{i=1}^{N-1} |q_{i+1} - q_i|^2 \qquad \text{(squared distance)}$$

d/D measures the monotonicity of q along the given pathway, and D^2/S measures the uniformity of q along the given pathway. Both indicators take the value 1 in the ideal case.

For each of the quantities E, ϕ_i and ψ_i, we tabulated the average value and standard deviation of these two reaction coordinate indicators over the 92216 relevant pathways in Table 7. As expected, the ϕ angles perform poorly on the monotonicity test (d/D is very small), whereas the energy and the ψ angles perform reasonably well on the monotonicity test. However, none of the quantities do very well on the uniformity test: the average value of D^2/S is around 0.30 for each of the dihedral angles and around 0.48 for the energy. This suggests that changes in a given dihedral angle tend to occur in a small number of big steps, rather than in a large number of small steps. This is consistent with our earlier pathway analysis, where we found that the ψ angles tend to change one or two at a time.

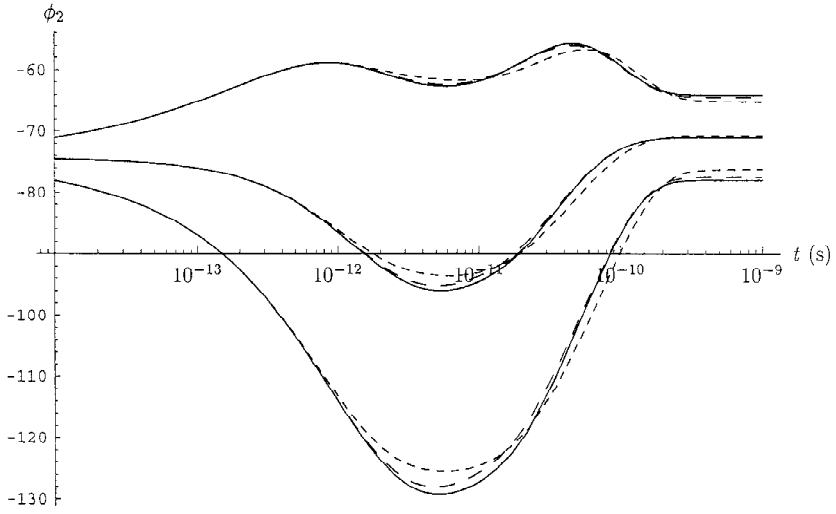

Figure 7. Time-evolution of ϕ_2 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

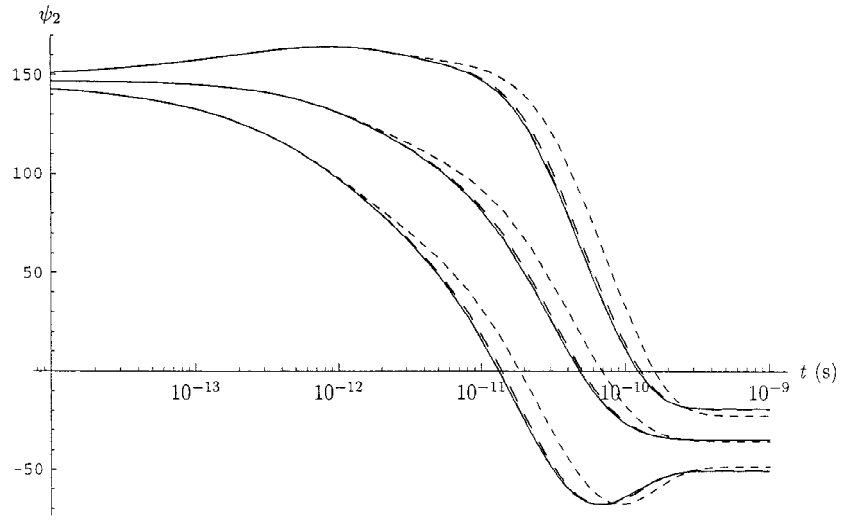

Figure 8. Time-evolution of ψ_2 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

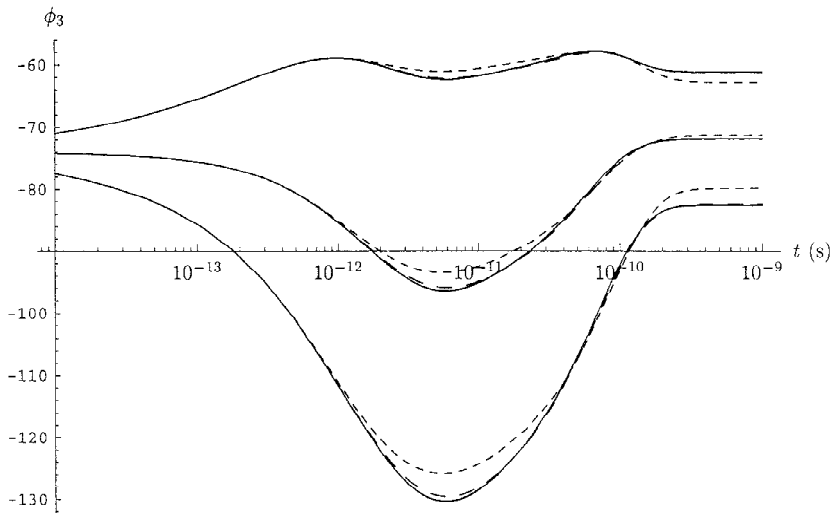

Figure 9. Time-evolution of ϕ_3 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

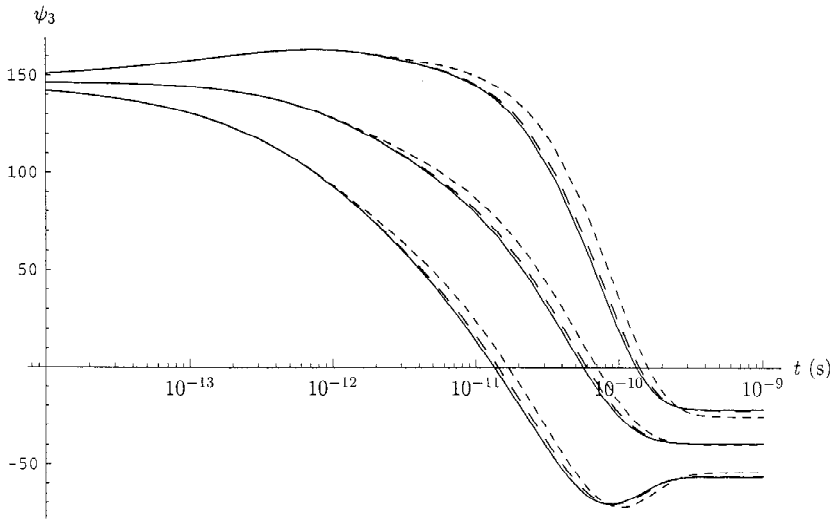

Figure 10. Time-evolution of ψ_3 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

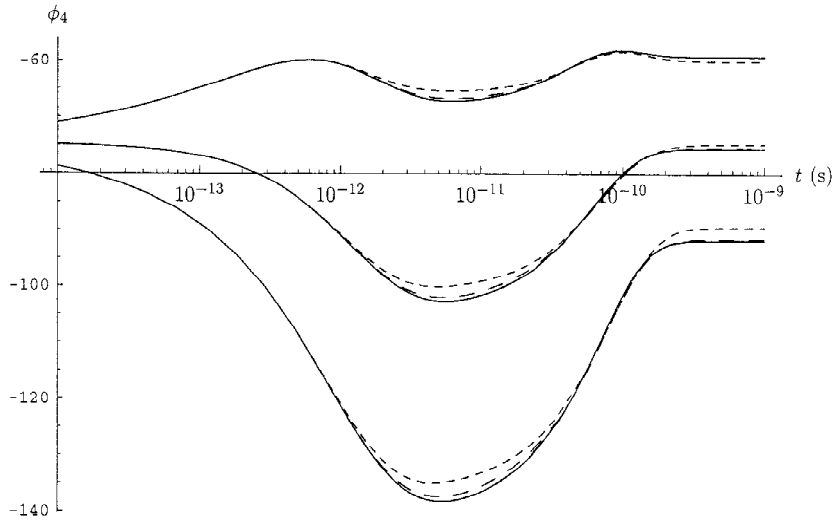

Figure 11. Time-evolution of ϕ_4 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

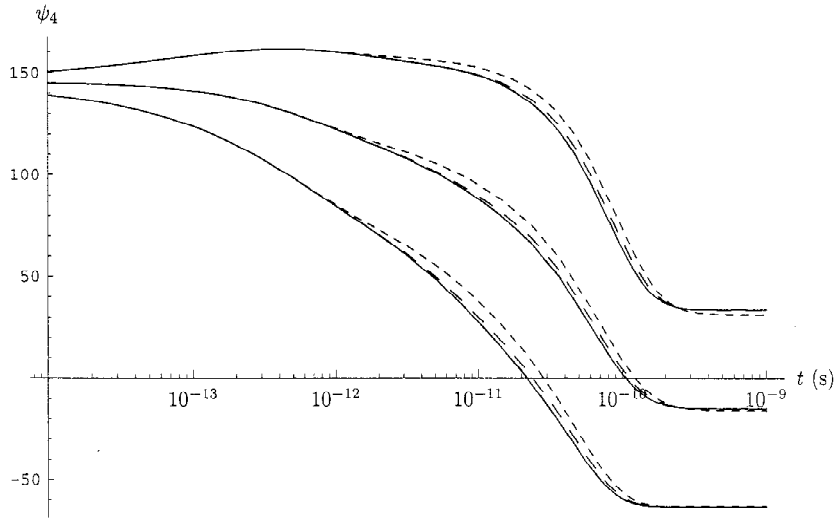

Figure 12. Time-evolution of ψ_4 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. Various pathway length cutoffs are employed.

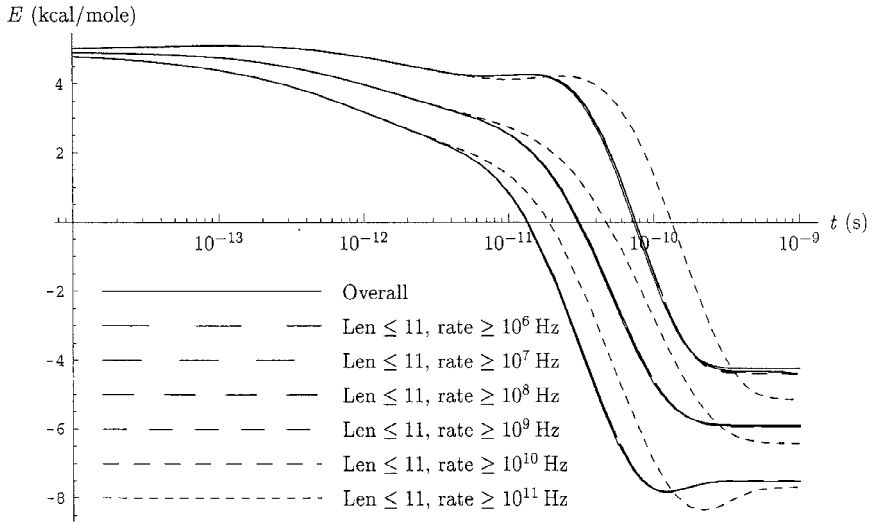

Figure 13. Time-evolution of E as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

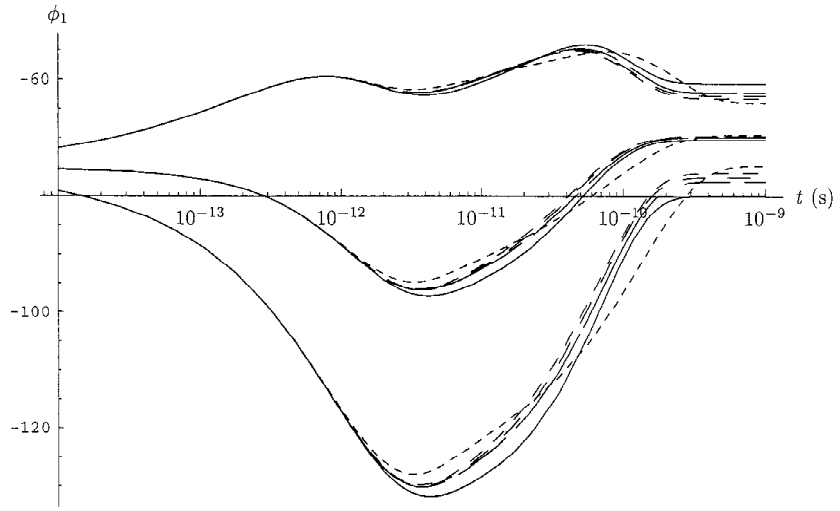

Figure 14. Time-evolution of ϕ_1 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

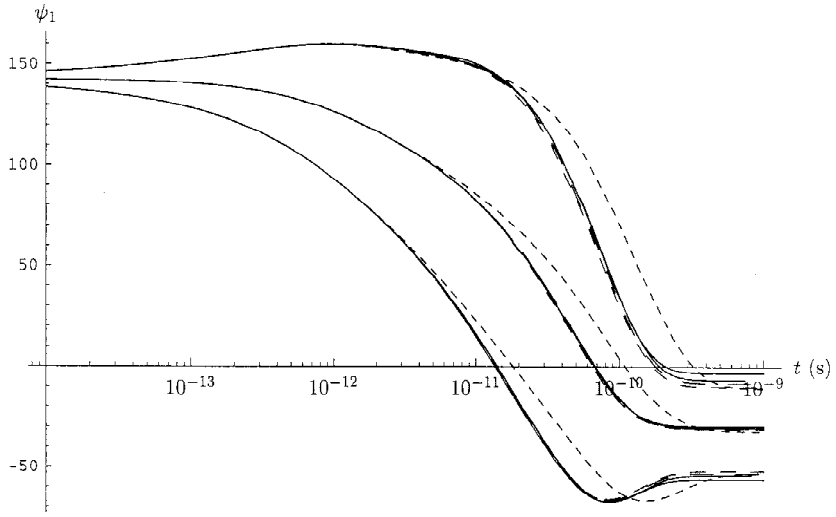

Figure 15. Time-evolution of ψ_1 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

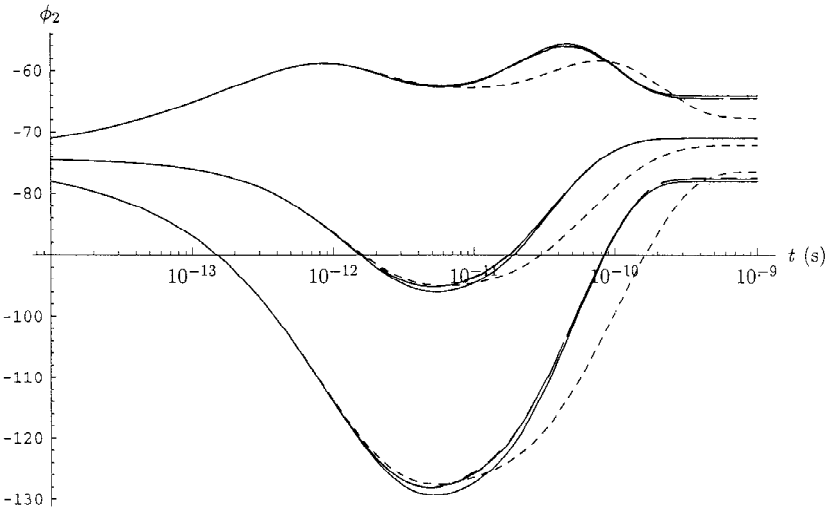

Figure 16. Time-evolution of ϕ_2 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

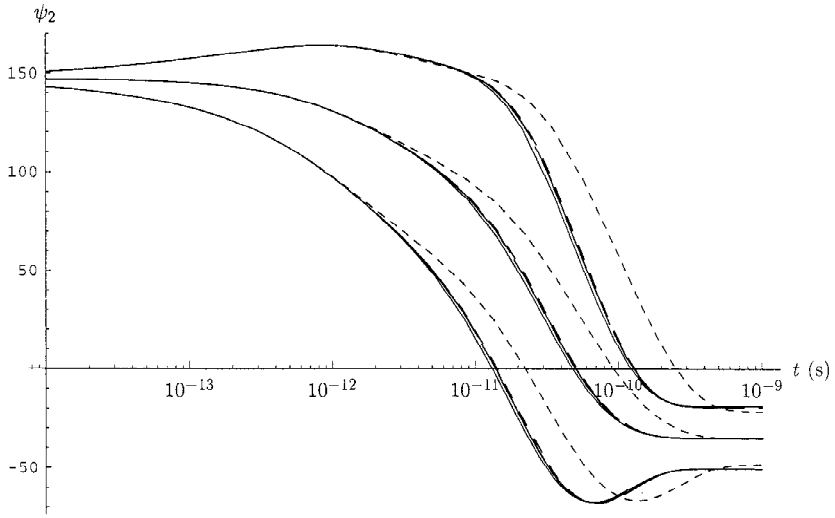

Figure 17. Time-evolution of ψ_2 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

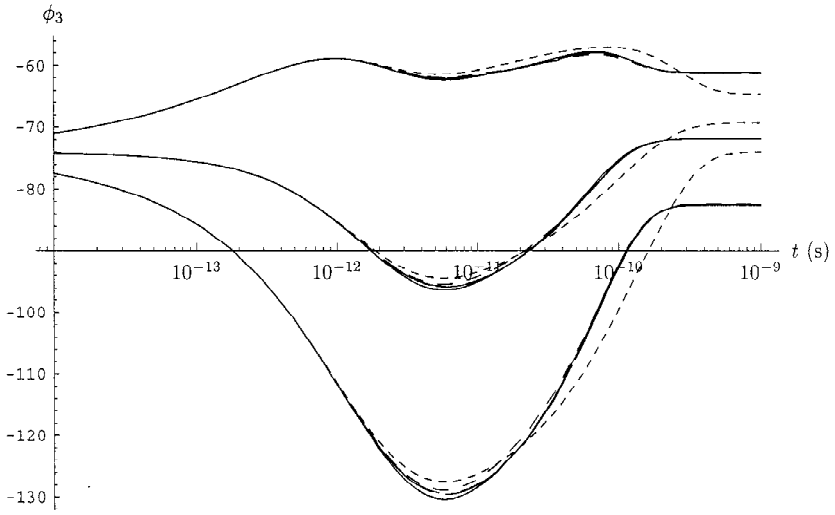

Figure 18. Time-evolution of ϕ_3 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

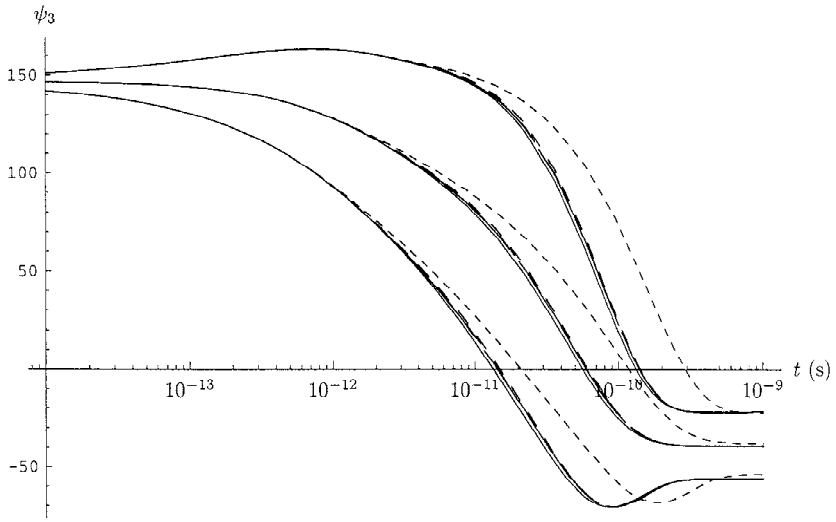

Figure 19. Time-evolution of ψ_3 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

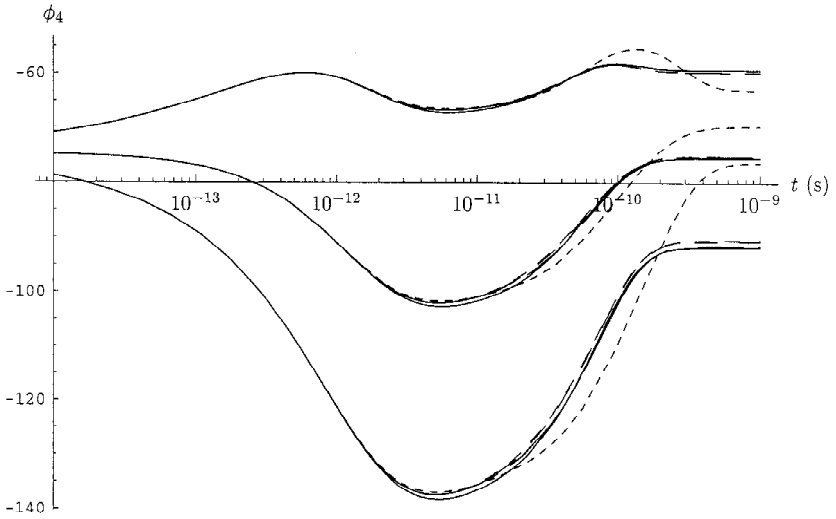

Figure 20. Time-evolution of ϕ_4 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

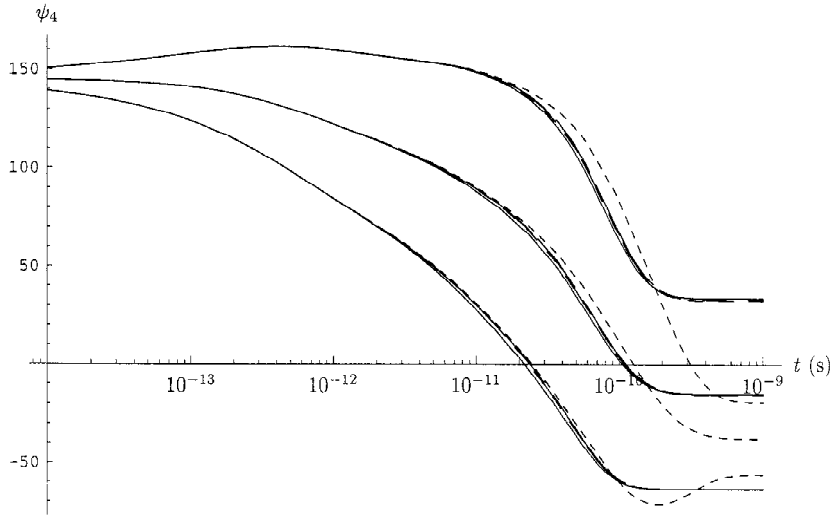

Figure 21. Time-evolution of ψ_4 as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. A pathway length limit of 11, along with various transition rate cutoffs, are employed.

Table 7. Average and standard deviation values of the reaction coordinate indicators d/D and D^2/S for various quantities over all pathways of length 11 or less with transition rates exceeding 10^{10} Hz from the extended conformation to the ground state of unsolvated tetra-alanine

	d/D		D^2/S	
Quantity	average	std	average	std
E	0.796	0.099	0.482	0.144
ϕ_1	0.224	0.138	0.291	0.080
ψ_1	0.899	0.120	0.256	0.060
ϕ_2	0.032	0.034	0.304	0.077
ψ_2	0.850	0.100	0.283	0.051
ϕ_3	0.081	0.081	0.332	0.084
ψ_3	0.867	0.129	0.298	0.071
ϕ_4	0.046	0.038	0.302	0.075
ψ_4	0.849	0.132	0.293	0.059
$\sum_i \psi_i$	0.927	0.066	0.749	0.066
d_{α_1,α_4}	0.674	0.138	0.355	0.115
d_1	0.762	0.129	0.467	0.098
d_2	0.712	0.111	0.523	0.142
$d_1 + d_2$	0.818	0.103	0.587	0.133

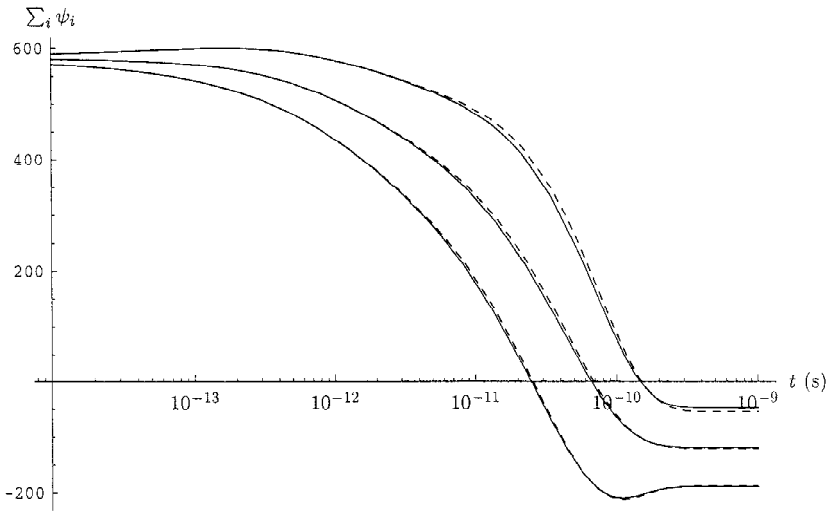

Figure 22. Time-evolution of $\sum_i \psi_i$ as a function of time (average $\pm$ one standard deviation), given that the system occupies the extended conformation at $t = 0$ s. solid curve shows the overall time evolution, and dotted line shows time evolution with a pathway length limit of 11 and a transition rate cutoff of dotted 10^{10} Hz.

It is clear that progress towards the alpha-helical ground state should not be measured in terms of a single ψ angle, but should reflect the progress of *all* ψ angles. This suggests that we might look at $\sum_i \psi_i$ as a reaction coordinate. The time evolution of $\sum_i \psi_i$ is plotted in Figure 22, and the average value and standard deviation of the reaction coordinate indicators are given in Table 7. The average value of the reaction coordinate indicators, $d/D = 0.927$ and $D^2/S = 0.749$, both indicate very strongly that $\sum_i \psi_i$ makes a good reaction coordinate. To confirm this, we constructed a scatter plot of D^2/S vs. d/D for each of the 92216 pathways, shown in Figure 23. For most of the pathways, the reaction coordinate indicators are both near 1, further suggesting that $\sum_i \psi_i$ makes a good reaction coordinate.

Further insight into the folding process may be gained by looking for a simpler reaction coordinate. An alpha-helix is stabilized by the formation of hydrogen bonds between the i and $i + 3$ residues. Since these residues tend to be farthest apart in the extended conformation, and must be brought close together to form the hydrogen bond, it makes sense to use the hydrogen bonding distance as a reaction coordinate.

We first tried d_{α_1,α_4}, the distance between the first and fourth α-carbons. This distance is indicated in Figure 24. This distance varies from 9.079 Å in the extended conformation to 4.998 Å in the ground state. The alpha-helical ground state is not the only conformation with $d_{\alpha_1,\alpha_4} < 5.0$ Å. Of the 526 minima involved in the 92216 relevant pathways, 26 of them satisfy this inquality.

The distance between α-carbons is only a crude measure of hydrogen bonding. A more direct measure is the distance between the nitrogen-bonded hydrogen atom

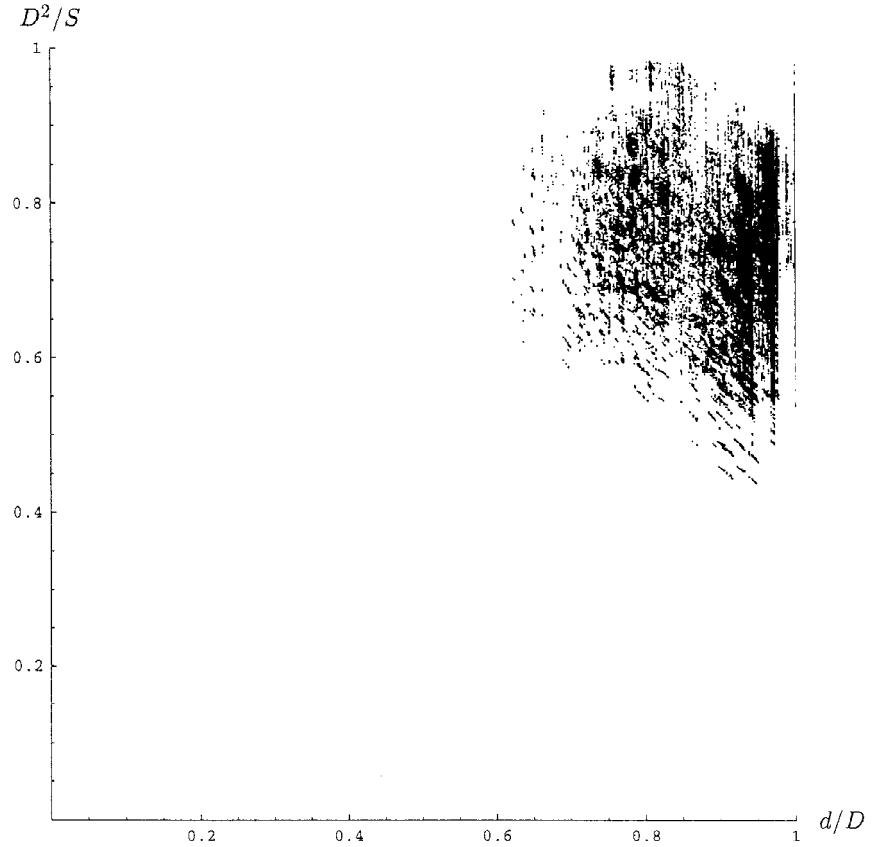

Figure 23. Scatter plot of reaction coordinate indicators for $\sum_i \psi_i$ for each pathway. Only pathways of length 11 or less with all transition rates exceeding 10^{10} Hz are used (92216 pathways).

and the oxygen atom which shares it. It turns out there are two candidate hydrogen bonding distances, as indicated in Figure 24. These distances in the ground state are $d_1 = 1.934\,\text{Å}$ and $d_2 = 1.921\,\text{Å}$. It turns out that neither distance alone uniquely determines the ground state. Of the 526 relevant minima, 9 of them satisfy $d_1 < 2\,\text{Å}$ and 7 of them satisfy $d_2 < 2\,\text{Å}$. However, only the ground state satisfies both inequalities. Apparently there are two hydrogen bonds which stabilize the alpha-helix in tetra-alanine.

We plotted four distance parameters related to hydrogen bonding, d_{α_1,α_4}, d_1, d_2, and $d_1 + d_2$, as functions of time in Figures 25–28, and tabulated the average value and standard deviation of the reaction coordinate indicators in Table 7. The motivation of including $d_1 + d_2$ among the distance parameters is similar to that of including $\sum_i \psi_i$. Since there are two hydrogen bonds to form, it makes sense that reaction progress should be measured by *both* hydrogen bond distances. Any

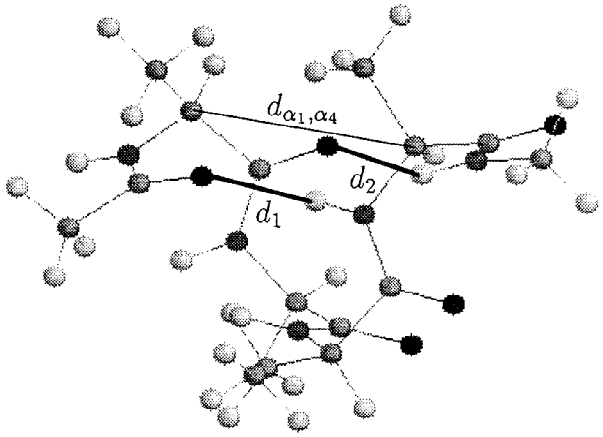

Figure 24. Alpha-helical ground state of unsolvated tetra-alanine, with the hydrogen bonds indicated.

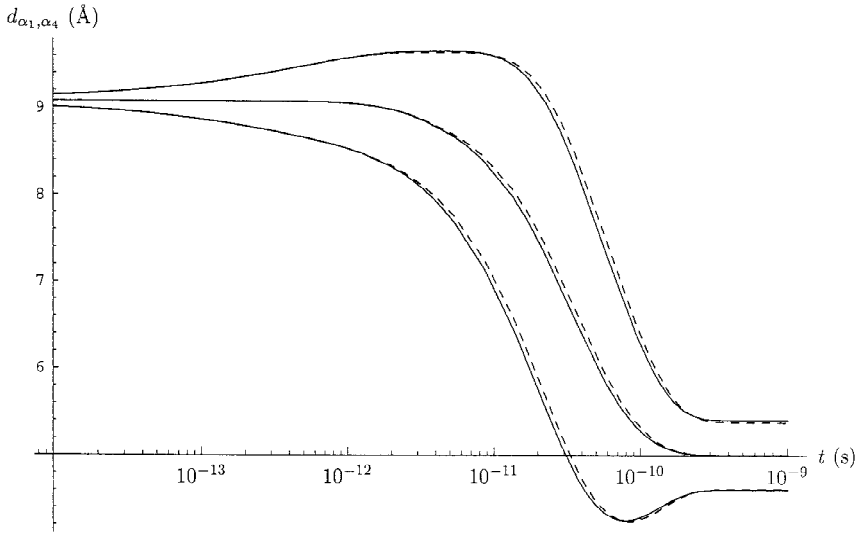

Figure 25. Time-evolution of d_{α_1,α_4} as a function of time (average $\pm$ one standard deviation). Given that the system occupies the exended conformation at $t = 0$ s. Solid curve shows the overall time evolution, and dotted line shows time evolution with a pathway length limit of 11 and a transition rate cutoff of 10^{10} Hz.

of the four distance parameters would make a reasonable reaction coordinate, but $d_1 + d_2$ is clearly the best with $d/D = 0.818$ and $D^2/S = 0.587$. A scatter plot of D^2/S vs. d/D for $d_1 + d_2$ is given in Figure 29.

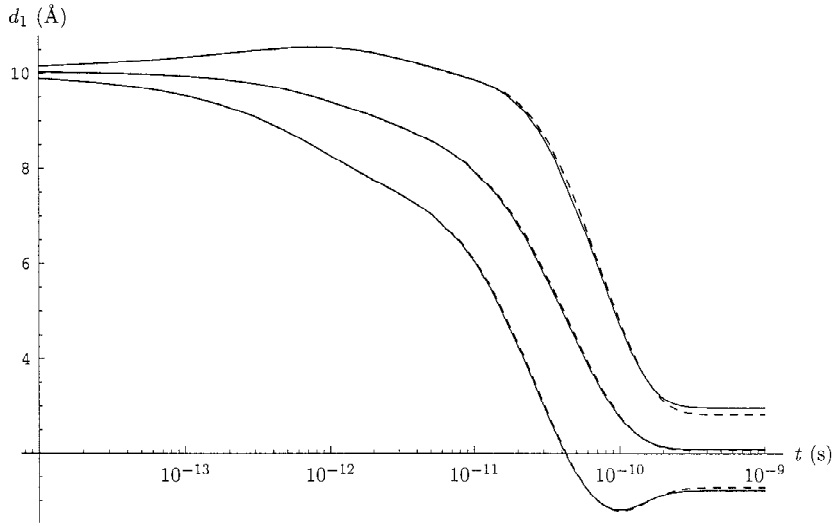

Figure 26. Time-evolution of d_1 as a function of time (average $\pm$ one standard deviation). Given that the system occupies the exended conformation at $t = 0$ s. Solid curve shows the overall time evolution, and dotted line shows time evolution with a pathway length limit of 11 and a transition rate cutoff of 10^{10} Hz.

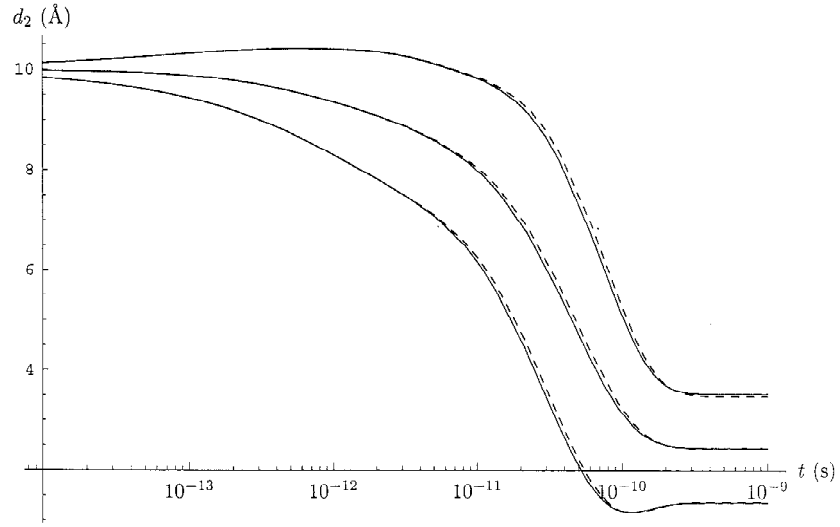

Figure 27. Time-evolution of d_2 as a function of time (average $\pm$ one standard deviation). Given that the system occupies the exended conformation at $t = 0$ s. Solid curve shows the overall time evolution, and dotted line shows time evolution with a pathway length limit of 11 and a transition rate cutoff of 10^{10} Hz.

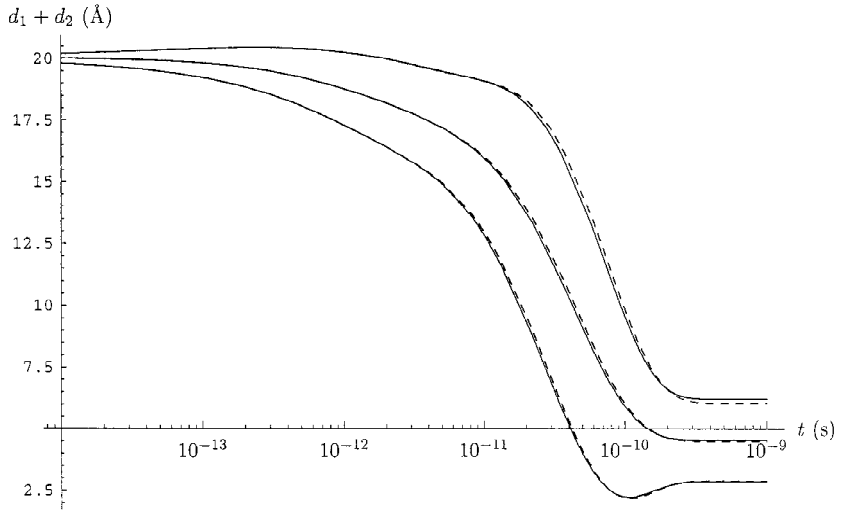

Figure 28. Time-evolution of d_1+d_2 as a function of time (average $\pm$ one standard deviation). Given that the system occupies the exended conformation at $t = 0$ s. Solid curve shows the overall time evolution, and dotted line shows time evolution with a pathway length limit of 11 and a transition rate cutoff of 10^{10} Hz.

Table 8. Eigenmode III results for solvated tetra-alanine

	6^8 grid
Local minima	66228
1st-order saddles	195639

5. Computational studies: Solvated tetra-alanine

We next studied tetra-alanine in solvation. We used the ECEPP/3 potential energy surface [11] coupled with the volume method for calculating solvation energies using the Reduced Radius Independent Gaussian Sphere (RRIGS) approximation [18].

We determined the minima and first-order saddles by applying a brute force eigenmode-following search (Eigenmode III) with a 6^8 grid of start points, just

Table 9. Ground state and extended conformation of solvated tetra-alanine

Minimum	Classification	E (kcal/mole)	F (kcal/mole)
min.1	aaaa	−35.249	−40.741
min.874	bbbb	−30.823	−41.194

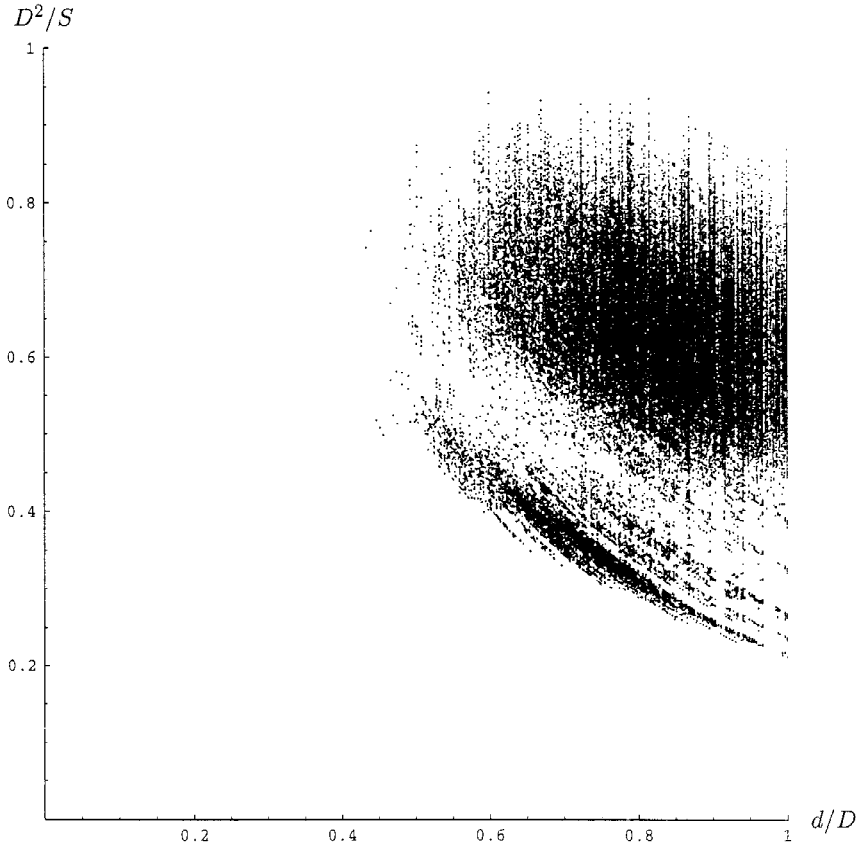

Figure 29. Scatter plot of reaction indicators for $d_1 + d_2$ for each pathway. Only pathways of length 11 or less with all transition rates exceeding 10^{10} Hz are used (92216 pathways).

as we did for unsolvated tetra-alanine. The results of this search can be found in Table 8.

Of the 66228 minima, we found one alpha-helical conformation, min.1 (aaaa), and one extended conformation, min.874 (bbbb). The potential energy (which includes the solvation energy) and free energy (which includes the vibrational entropy) of these two states can be found in Table 9.

The first thing to notice is that, although the alpha-helical conformation has the lowest potential energy (and hence lowest free energy at $T = 0\,\mathrm{K}$), the extended conformation has a lower free energy at room temperature ($T = 300\,\mathrm{K}$) than the ground state. The result of adding solvation energy reduces the energy gap from 11.6 kcal/mole to 4.4 kcal/mole. The entropic term in the free energy is more than enough to overpower this energy gap and reduce the free energy of the extended conformation below that of the alpha-helical ground state. This has significant implications.

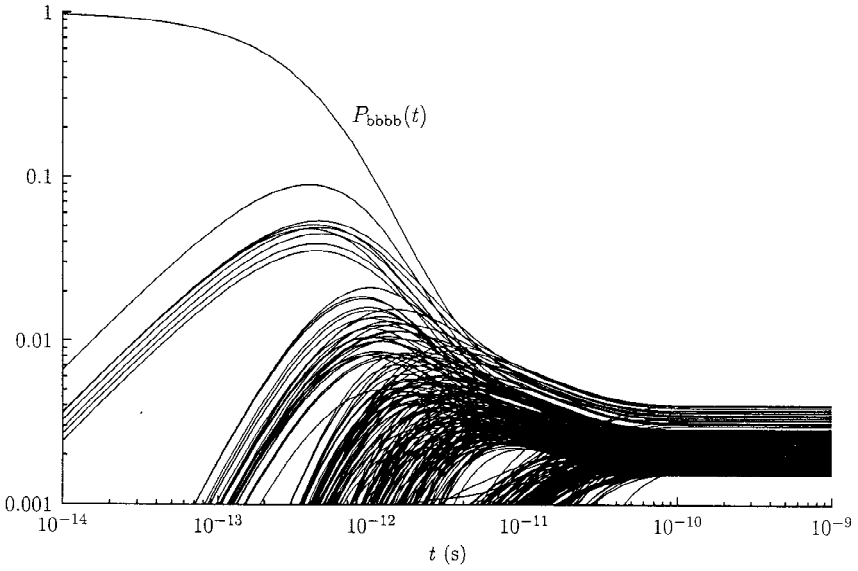

Figure 30. Time evolution of the extended conformation and the 300 lowest free energy states of solvated tetra-alanine at $T = 300$ K. No single state has an equilibrium probability which exceeds 0.004.

We calculated the free energies of all the minima using the harmonic approximation (this is discussed in the previous section) in order to determine the equilibrium probability distribution. We found that the several hundred lowest free energy minima have about the same free energy, and that no single minimum has an equilibrium occupation probability which exceeds 0.004. This is in stark contrast with unsolvated tetra-alanine, where the ground state had an equilibrium occupation probability of 0.748, and the lowest 3 potential energy states accounted for 0.936 of the total equilibrium probability.

As a check, we calculated the transition rate matrix for solvated tetra-alanine at $T = 300\,\mathrm{K}$, and solved the Master equation starting with the extended conformation at $t = 0\,\mathrm{s}$. We plotted the time evolution of the occupation probabilities of the 300 lowest free energy states. That plot is given in Figure 30. The equilibrium probability distribution is achieved in about 10^{-10} s.

It seems likely that solvated tetra-alanine exhibits liquid-like behavior at $T = 300\,\mathrm{K}$. To be sure, we need to verify that the several hundred minima which share the equilibrium probability distribution do not occupy the same region of configuration space. If that were the case, the potential energy surface would have one deep basin with a rough bottom. The true characteristics of a liquid-like molecule is that it randomly (and quickly) samples widely distinct configurations. In Figure 31, we plotted the distribution of minima on four (ϕ, ψ) plots. Our conclusion is that the minima which share the equilibrium probability distribution do occupy distinct regions of configuration space.

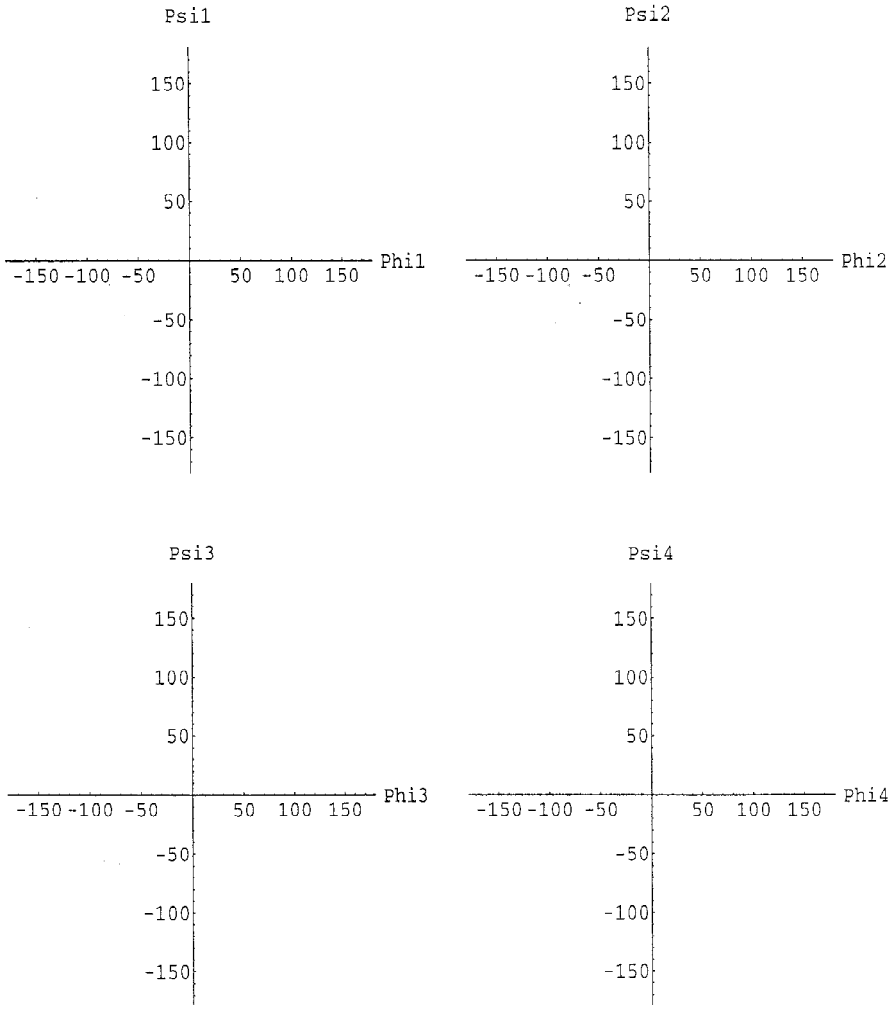

Figure 31. Scatter plot for solvated tetra-alanine. Dot size is proportional to equilibrium probability.

If solvated tetra-alanine is to be liquid-like at $T = 300\,\mathrm{K}$, then there must be a phase transition. This should show up as a peak in the heat capacity versus temperature plot. The heat capacity can be calculated by calculating energy fluctuations at equilibrium

$$C_v = \frac{d}{dT}\langle E\rangle_{\mathrm{eq}} = \frac{\langle E^2\rangle_{\mathrm{eq}} - \langle E\rangle_{\mathrm{eq}}^2}{kT^2}$$

where equilibrium averages may be calculated from free energies

$$\langle q\rangle_{\mathrm{eq}} = \frac{\sum_i q_i e^{-F_i/kT}}{\sum_i e^{-F_i/kT}}$$

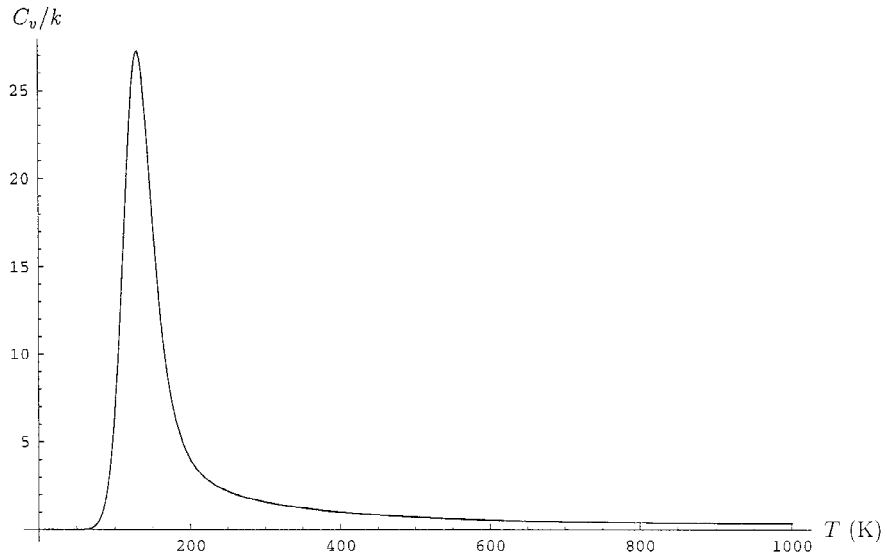

Figure 32. Heat capacity as a function of temperature for solvated tetra-alanine.

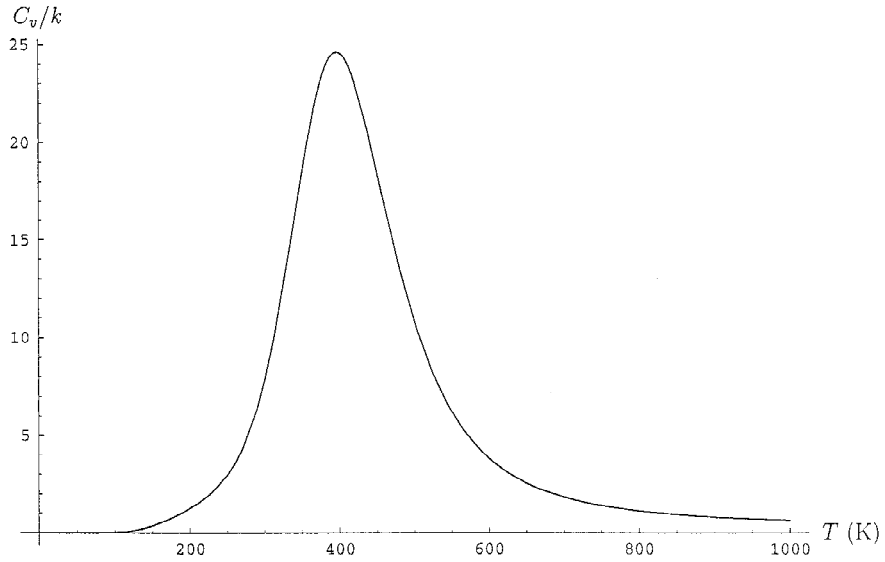

Figure 33. Heat capacity as a function of temperature for unsolvated tetra-alanine.

We calculated C_v as a function of T for temperatures ranging from (just above) 0 K to 1000 K for both solvated and unsolvated tetra-alanine. The plots are given in Figures 30 and 31. The transition temperatures are given by

$$T_{\text{sol-liq}}^{\text{solv}} = 130\,\text{K} \qquad T_{\text{sol}-\text{liq}}^{\text{unsolv}} = 395\,\text{K}$$

The lower transition temperature for solvated tetra-alanine can be traced back to the reduction in the energy gap between the alpha-helical ground state conformation and the other higher energy states, including the extended conformation, and indeed do explain the appearance of liquid-like behavior for solvated tetra-alanine (but not for unsolvated tetra-alanine) at $T = 300\,\mathrm{K}$.

6. Conclusions

Determining all stationary points of a potential energy surface is the critical first step in analyzing its topography. The proposed αBB algorithm provides a theoretical guarentee of finding all stationary points of a potential energy surface. We explored several regions of the potential energy surface of tetra-alanine using both a brute force eigenmode-following search and the proposed αBB algorithm. In all cases, αBB found all of the minima and first and higher-order saddles located by the eigenmode-following search, and in many cases, found additional transition states.

Once we have all of the minima and transition states, we analyzed the potential energy surfaces of both unsolvated and solvated tetra-alanine. We have developed procedures for this analysis, which involves calculating min-saddle-min triples, transition rates, time-evolution of occupation probabilities and average quantities, pathways, and the rate disconnectivity graph. We also developed two reaction coordinate indicators to help us determine whether a given quantity makes a good reaction coordinate or not.

For unsolvated tetra-alanine, we found a total of 62373 minima and 212938 first-order transition states, including an alpha-helical structure and an extended conformation. The alpha-helical structure is the ground state.

We found that most of the interesting dynamics can be reproduced by restricting our attention to the pathways between the extended conformation and the ground state which are of length 11 or less with transition rates exceeding 10^{10} Hz. This limits us to 92216 pathways which involve only 526 minima and 1696 transition states, a substantial reduction from the 62373 minima and 212938 transition states we started out with.

We were able to characterize the most important pathways and determined that most of these pathways involved cooperative motion, and that most pathways pass through a state where the carboxyl side of the molecule is mostly folded and the amino side of the molecule is mostly extended.

We also constructed the rate disconnectivity graph and found that unsolvated tetra-alanine involves transitions on two widely separated time scales: large rotations in ϕ angles account for the slow transitions (on the order of 0.01 s or longer), and small rotations in ϕ account for fast transitions (on the order of 10^{-10} s).

We applied our two reaction coordinate indicators to several quantities, including energy, the ϕ and ψ angles, as well as distances between atoms which characterize the hydrogen bonds formed in the alpha-helical ground state.

Of all the quantities we looked at, we determined that $\sum_i \psi_i$ provided the best reaction coordinate, most accurately representing the progress from the extended conformation to the ground state along each of the 92216 most important pathways. The sum of the two hydrogen bond distances, $d_1 + d_2$, also would make a suitable reaction coordinate.

For solvated tetra-alanine, we found a total of 66228 minima and 195639 first-order transition states. We found that, unlike unsolvated tetra-alanine, several hundred minima (occupying widely varying regions of configuration space) share the equilibrium probability distribution at $T = 300\,\mathrm{K}$, which is characteristic of a liquid-like molecule. We confirmed this by plotting C_v vs. T for both solvated and unsolvated tetra-alanine.

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation and the National Institutes of Health (R01 GM52032 and 1 F32 GM20007-01).

Notes

1. This is about a factor of 10 faster than we concluded in [1]. This speed-up is most likely due to the 11-fold increase in the number of transition states located.
2. In fact, the very shortest pathways are all very slow, as indicated by Table 5.
3. This has been checked rigorously for all pathways length 11 or less with a rate cutoff of 10^6 Hz. What we have in fact found is that there are transition states which connect two minima b $\rightarrow$ a, but either the transition itself is very slow, or else the minima are so high in energy that it seems unlikely that a fast pathway (of *any* length) could pass through it. Our conclusion is that b $\rightarrow$ a is not observed for all but the very slow pathways.
4. The last 3 observations are in stark contrast with the observations we made about the pathways of tetra-alanine in [1], which were based on the minima and transition states obtained from the 4^8 grid search. In that paper, we found very little cooperative motion and had also concluded that most of the alanines fold halfway before any of them folds the rest of the way (e.g., bbbb $\rightarrow$ iiii $\rightarrow$ aaaa). This illustrates very clearly the importance of finding *all* minima and transition states, or at least the most important minima and transition states, before analyzing the potential energy surface. Still, we were surprised with how different the pathways were when we analyzed the results of the 6^8 grid. Since very few low-lying minima were added by that search, the changes must be mostly due to the additional transition states available to the system. The transition states allowing for cooperative motions of the different amino groups must have been systematically overlooked as a result of the particular 4^8 grid we selected.
5. The remaining 16 minima are not connected to the main group by any transition states at all.
6. Actually, we only solve the Master equation over the 3713 minima in the highlighted region of the rate disconnectivity graph shown in Figure 2. This is necessary because solving the Master equation for all 62373 minima would require diagonalizing a 62373×62373 matrix which does not fit in computer memory. Fortunately, it is also sufficient since the other minima are unreachable during times on the order of 10^{-9} sec.

References

1. Westerberg, K.M. and Floudas, C.A. (1999), *J. Chem. Phys.* 110(18): 9259.
2. Maranas, C.D. and Floudas, C.A. (1995), *J. Global Opt.* 7 143.
3. Andoulakis, I.P., Maranas, C.D. and Floudas, C.A. (1995), *J. Global Opt.* 7(4): 337.
4. Adjiman, C. and Floudas, C.A. (1996), *J. Global Opt.* 9: 23.
5. Adjiman, C., Dallwig, S., Floudas, C.A. and Neumaier, A. (1998), *Computers and Chem. Eng.* 22(9): 1137.
6. Adjiman, C., Androulakis, I.P. and Floudas, C.A. (1998), *Computers and Chem. Eng.* 22(9): 1159.
7. Jarrold, M.F. (1994), in *Clusters of Atoms and Molecules*, H. Haberland (ed.). Springer, Berlin, Chap. 2.7, 163.
8. Kunz R.E. and Berry, R.S. (1995), *J. Chem. Phys.* 103(5): 1904.
9. Becker, O.M. and Karplus, M. (1997), *J. Chem. Phys.* 106(4): 1495.
10. Czerminski, R. and Elber, R. (1990), *J. Chem. Phys.* 92(9): 5580.
11. Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. and Scheraga, H.A. (1992), *J. Phys. Chem.* 96: 6472.
12. Tsai, C.J. and Jordan, K.D. (1993), *J. Phys. Chem.* 97: 11227.
13. Cerjan, C.J. and Miller, W.H. (1981), *J. Chem. Phys.* 75(6): 2800.
14. Simons, J., Jorgensen, P., Taylor, H. and Ozment, J. (1983), *J. Phys. Chem.* 87: 2745.
15. Banerjee, A., Adams, N., Simons, J., Shepard, R. (1985), *J. Phys. Chem.* 89: 52.
16. O'Neal, D., Taylor, H. and Simons, J. (1984), *J. Phys. Chem.* 88: 1510.
17. Culot, P., Dive, G., Nguyen, V.H. and Ghuysen, J.M. (1992), *Theor. Chim. Acta* 82: 189.
18. Augspurger, J.D. and Scheraga, H.A. (1987), *J. Phys. Chem.* 91: 4105.