Locating All Transition States and Studying the Reaction Pathways of Potential Energy Surfaces

K. M. Westerberg and C. A. Floudas*

Department of Chemical Engineering
Princeton University,
Princeton, New Jersey 08544-5263

Abstract

We propose a new method for calculating 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 $\nabla V = 0$. We apply this method to Murrel-Sorbie analytic potential energy surfaces of HCN, HSiN, HBO, and CS_2, and to the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential energy surfaces of alanine, alanine dipeptide, and tetra-alanine. For alanine, alanine dipeptide, and tetra-alanine, we proceed to analyze the topography of the potential energy surface by calculating reaction pathways, transition rate matrices, time-evolution of occupation probabilities, and rate disconnectivity graphs.

*Author to whom all correspondence should be addressed. Tel: (609)258-4595; Fax: (609)258-0211; email: floudas@titan.princeton.edu.

1 Introduction

Stationary points of potential energy surfaces (i.e., where $\nabla V = 0$) play an important role in computational chemistry. The local minima of a potential energy surface represent stable molecular configurations, and the first order saddle points generally correspond to transition states which connect two such configurations. A chemical reaction, isomerization process, or a protein-folding process can be thought of as a transition between two local minima through a transition state on the potential energy surface, or a series of such transitions.

Determining the location of the stationary points on a potential energy surface is the first step in understanding how the potential energy surface is connected (i.e., understanding its topography). Once these stationary points are known, the possible reaction pathways between two given configurations can be enumerated and transition rates can be calculated. Even higher order saddle points may play an important role in transition state theory, as these states provide alternative reaction pathways. Hence, a method of generating all stationary points of a given potential energy surface would be extremely useful in elucidating the pathways selected.

There are a number of methods for generating stationary points of potential energy surfaces already in use. The most obvious method is applying the Newton-Raphson method to $\nabla V = 0$. The Newton-Raphson method tends to yield a solution whenever the initial guess is close to a stationary point and the Hessian matrix has the appropriate signature for the type of stationary point desired (minima, first-order saddle, etc). It cannot be used, for example, to walk away from a local minimum towards a first-order saddle point.

The various "eigenmode-following" methods are sophisticated variants of the Newton-Raphson method [1, 2, 3, 4, 5, 6]. The Hessian is diagonalized, and a modified Newton-Raphson step is generated by "shifting" some of the eigenvalues of the Hessian, from positive to negative or vice versa, before applying its inverse. These methods allow one to step away from local minima in search of transition states, and vice versa. Further details can be found in Appendix A.

There are a number of stochastic methods used to find stationary points [7]. Local minima can be obtained by frequent quenching of a constant energy (or temperature) trajectory [8]. Simulated annealing by running a constant temperature trajectory simulation, slowly reducing the temperature to zero in the process, can sometimes lead to good candidates for the global minimum. The method of "slowest slides" [9] has been used to search for transition states connecting two given local minima: a constant energy trajectory is followed during a transition from one local minimum to the other, and the maximum along that trajectory is taken as an initial guess for the transition state.

Other methods exist for searching for the global minimum of a potential energy surface. Diffusion equation and distance scaling methods have been applied to the problem of finding the global minimum of a potential energy surface [10]. Smoothing transformations are applied to the potential energy surface to eliminate the irrelevant local minima. The remaining minima are tracked back to the original potential energy surface as the transformations are gradually removed. Another method involves obtaining a large sample of local minima and forming a "convex global underestimator" of the potential energy surface based on those sample points [11]. The global minimum of the original potential energy surface is sought in the vicinity of the global minimum of the convex global underestimator.

All of these methods, good in their own right, share one very important drawback: there is no guarantee that all (or even the most important) local minima and first or higher-order transition states will be found.

In this paper, we propose a method of finding all stationary points of a given potential energy surface based on the αBB deterministic branch and bound global optimization algorithm [12, 13, 14, 15, 16], which is described in detail in Sections 2–4. This is an application of the more general method of finding all solutions to systems of non-linear equations described in [12]. We applied this method to the Murrel-Sorbie analytic potential energy surfaces of triatomic molecules, such as HCN, HSiN, HBO, and CS_2, and to the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential energy surfaces of alanine, alanine dipeptide, and tetra-alanine. The details of this search will be discussed in Sections 5, 6, 7, and 9.

Once the minima and saddles have been located, the topography of the potential energy surface can be further analyzed. In Section 8, we present the results of our analysis for alanine and alanine dipeptide. Section 9 presents the computational studies of tetra-alanine and the analysis of the pathways.

It is simple matter to 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. Since we have 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 Rice–Ramsperger–Kassel–Marcus (RRKM) theory [17, 18].

Recently, Becker and Karplus [19] proposed a graphical representation of the topography of a potential energy surface. 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 an "energy (temperature) disconnectivity graph". The shape of the disconnectivity graph reveals an enormous wealth of dynamical information. We extended their idea by constructing a "rate disconnectivity graph" based on transition rates rather than energy levels or barrier heights.

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 constraint $\nabla V = 0$, which can be written out more fully as

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

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

Equation (1) is a special case of the general class of non-linearly constrained systems of algebraic equations:

$$\begin{aligned} f_i(\mathbf{x}) &= 0 \; , \qquad i = 1, \ldots N_f \\ g_j(\mathbf{x}) &\leq 0 \; , \qquad j = 1, \ldots N_g \\ \mathbf{x}^L \leq &\, \mathbf{x} \leq \mathbf{x}^U \end{aligned} \tag{2}$$

where $f_i(\mathbf{x})$ represent the equality constraints (N_f is the number of such constraints) and $g_j(\mathbf{x})$ represent the inequality constraints (N_g is the number of such constraints). Indeed, (2) can be reduced to (1) by assigning $f_i(\mathbf{x}) = \partial V / \partial x_i$ for $i = 1, \ldots, N_f = N_x$ and $N_g = 0$.

Equation (2) 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:

$$\begin{aligned} &\min_{\mathbf{x},s} s \\ \text{subject to} \quad f_i(\mathbf{x}) - s &\leq 0 \; , \qquad i = 1, \ldots N_f \\ -f_i(\mathbf{x}) - s &\leq 0 \; , \qquad i = 1, \ldots N_f \\ g_j(\mathbf{x}) &\leq 0 \; , \qquad j = 1, \ldots N_g \\ \mathbf{x}^L \leq &\, \mathbf{x} \leq \mathbf{x}^U \end{aligned} \tag{3}$$

The following two facts are self-evident:

- If $s < 0$, the constraints in (3) are infeasible.
- If $s = 0$, the constraints in (3) reduce to the original problem (2).

It follows that $s = 0$ is the *global minimum* of (3) (provided that (2) has solutions), and that there is a one-to-one correspondence between global minima $(\mathbf{x}^*, s^*)$ of (3) and solutions $\mathbf{x}^*$ of the original problem (2). Therefore, the problem of finding all solutions to (2) can be reformulated as the problem of finding all global minima of (3).

In the next section, we will explain how the αBB global optimization algorithm [12, 13, 14, 15, 16] can be used to find all global minima of (3), and hence, all solutions to (2).

3 The αBB global optimization approach

In this section, we describe the αBB global optimization algorithm [12, 13, 14, 15, 16], as it is applied to the general problem of determining all solutions to a system of algebraic constraints (2). This adaptation is based on the correspondence between solutions of (2) and global minima of (3) with $s = 0$. The αBB algorithm can be applied to any problem involving constraints which are twice continuously differentiable (C^2). Thus, the only necessary assumptions we need to make are that $f_i(\mathbf{x})$ and $g_j(\mathbf{x})$ are C^2 functions for $i = 1, \ldots, N_f$ and $j = 1, \ldots, N_g$, respectively.

The algorithm proceeds by exploring the configuration space for solutions to (2). 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 solution to (2) can possibly exist there. This is accomplished by finding a lower bound of the global minimum of (3) over the region in question. If the lower bound is positive, then $s = 0$ cannot lead to a feasible point of (3), and hence no solution to (2) 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 solution to (2) in that region. In this case, further subdivision and testing will be necessary. If the region size becomes small enough and the region is still active (i.e., its lower bound is negative or zero), then a solution to (2) is obtained within that region by a local search. The algorithm terminates when all regions have been fully processed.

Lower bounds of the global minimum of (3) are determined by solving a modified problem, called the *lower bounding problem*, over the given region. The goal here is to replace (3) with a new problem which is convex and has an expanded feasibility region. Both goals can be accomplished by replacing the LHS of each constraint in (3) with *convex underestimators* of those functions. If $\hat{f}_i^+(\mathbf{x}) \leq f_i(\mathbf{x})$, $\hat{f}_i^-(\mathbf{x}) \leq -f_i(\mathbf{x})$, and $\hat{g}_j(\mathbf{x}) \leq g_j(\mathbf{x})$ are all convex functions, then a valid lower bounding problem would be to minimize s subject to the convexified constraints:

$$\begin{aligned} \hat{f}_i^+(\mathbf{x}) - s &\leq 0\ , \qquad i = 1, \ldots, N_f \\ \hat{f}_i^-(\mathbf{x}) - s &\leq 0\ , \qquad i = 1, \ldots, N_f \\ \hat{g}_j(\mathbf{x}) &\leq 0\ , \qquad j = 1, \ldots, N_g \end{aligned} \tag{4}$$

Since the constraints are all convex functions, any local optimization package should be able to locate its global minimum. Furthermore, every feasible point of (3) is also a feasible point of (4) since these functions are *underestimators* of the original functions. It follows that the global minimum of (4) is a valid lower bound of the global minimum of (3).

The crux of the αBB algorithm is finding valid convex underestimators, $\hat{f}_i^\pm(\mathbf{x})$ and $\hat{g}_j(\mathbf{x})$, for the functions $\pm f_i(\mathbf{x})$ and $g_j(\mathbf{x})$, respectively, over a given region. An important consideration is that the convex underestimators be as tight (i.e., close in value to the original constraint functions) as is reasonably possible, as tighter underestimators lead to better lower bound estimates. It is important to be able to fathom regions as quickly as possible if they do not contain any solutions to (2). However, this cannot always be done: it frequently occurs that a region contains no solution to (2) (i.e., the global minimum of (3) over that region is positive), but the lower bound obtained from (4) for that region is negative. Such regions obviously must be explored further, until positive lower bounds are obtained. A better lower bound estimate can lead to significant improvement in the efficiency of the algorithm.

It should be noted that the partitioning of a region into subregions naturally permits tighter convex underestimators, as the restrictions placed on the convex underestimators are reduced (they only have to be convex and underestimate the original constraint functions over the region of interest). This is critical to the eventual termination of the αBB algorithm, since it is important that regions which contain no solutions ultimately be fathomed. If the lower bounds on those regions were never to improve to a point where they all would become positive, then the algorithm would never terminate.

The most generally applicable method for the generation of convex underestimators, which we employ in our search for all stationary states, is to subtract off a quadratic term from the original constraints:

$$\hat{f}_i^+(\mathbf{x}) \;=\; f_i(\mathbf{x}) - \alpha_i^{f,+} \sum_k (x_k^U - x_k)(x_k - x_k^L)$$

$$
\begin{aligned}
\hat{f}_i^-(\mathbf{x}) &= -f_i(\mathbf{x}) - \alpha_i^{f,-} \sum_k (x_k^U - x_k)(x_k - x_k^L) \\
\hat{g}_j(\mathbf{x}) &= g_j(\mathbf{x}) - \alpha_j^g \sum_k (x_k^U - x_k)(x_k - x_k^L)
\end{aligned} \tag{5}
$$

where $\mathbf{x}^U$ and $\mathbf{x}^L$ are the upper and lower bounds of the region in question [12]. The coefficients $\alpha_i^{f,+}$, $\alpha_i^{f,-}$ and α_j^g must be positive and sufficiently large so as to guarantee the convexity of $\hat{f}_i^+(\mathbf{x})$, $\hat{f}_i^-(\mathbf{x})$ and $\hat{g}_j(\mathbf{x})$ over the region $[\mathbf{x}^L, \mathbf{x}^U]$. A function is convex if and only if its Hessian matrix is positive semi-definite at all points $\mathbf{x} \in [\mathbf{x}^L, \mathbf{x}^U]$. This observation leads to the following lower limit on the choice of each α[1]:

$$
\begin{aligned}
\alpha_i^{f,+} &\geq -\frac{1}{2} \min_{\mathbf{x}\in[\mathbf{x}^L,\mathbf{x}^U]} \{\lambda_k(H_{f_i}(\mathbf{x})), 0\} \\
\alpha_i^{f,-} &\geq +\frac{1}{2} \max_{\mathbf{x}\in[\mathbf{x}^L,\mathbf{x}^U]} \{\lambda_k(H_{f_i}(\mathbf{x})), 0\} \\
\alpha_j^{g} &\geq -\frac{1}{2} \min_{\mathbf{x}\in[\mathbf{x}^L,\mathbf{x}^U]} \{\lambda_k(H_{g_j}(\mathbf{x})), 0\}
\end{aligned} \tag{6}
$$

The α underestimators $\hat{f}_i^+(\mathbf{x})$, $\hat{f}_i^-(\mathbf{x})$ and $\hat{g}_j(\mathbf{x})$ given by (5) have the following properties:

- They are valid underestimators of the original constraint functions and are convex over the region $[\mathbf{x}^L, \mathbf{x}^U]$ (provided, of course, that (6) is satisfied).

- They match the constraint functions at all corner points (i.e., $\hat{f}_i^\pm(\mathbf{x}) = \pm f_i(\mathbf{x})$ and $\hat{g}_j(\mathbf{x}) = g_j(\mathbf{x})$ at all corner points).

- The maximum separation between an α underestimator $\hat{f}(\mathbf{x})$ and the function it underestimates $f(\mathbf{x})$ always occurs at the center of the rectangle $\mathbf{x}^{\mathrm{mid}} = (\mathbf{x}^L + \mathbf{x}^U)/2$ and has the value
$$\max_{\mathbf{x}\in[\mathbf{x}^L,\mathbf{x}^U]} f(x) - \hat{f}(x) = \frac{1}{4}\alpha \left\|\mathbf{x}^U - \mathbf{x}^L\right\|^2$$
This value decreases as the region is partitioned.

- The α underestimator constructed over the region $[\mathbf{x}^L, \mathbf{x}^U]$ will always be more tight than an α underestimator constructed over a larger region. In other words, as the regions are partitioned into smaller region, the underestimators will more closely approximate the functions they are underestimating.

[1] It is at this point that the assumption that all constraint functions are C^2 come into play. It is important that the Hessian of each constraint exist and have bounded eigenvalues over the region of interest, or else the inequalities (6) may be meaningless or impossible to satisfy.

The latter two properties insure that the lower bounds follow a monotonically non-decreasing sequence as the algorithm progresses, and that it will eventually terminate.

Calculating values of α according to (6) is difficult in general because the Hessian matrices H_{f_i} and H_{g_j} depend on $\mathbf{x}$. If the selected values of α 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 α is to start with small values of α (e.g., $\alpha_i^{f,+} = \alpha_i^{f,-} = \alpha_j^g = 5$) and increase the values of α until no new solutions are found. This can be a practical solution to many problems where the correct values of α are difficult to determine, or lead to poor performance. 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 of alanine by setting $\alpha_i^{f,\pm} = 5$. Further discussion of alanine can be found in Section 6.

A more robust method involves calculating the Hessian matrices H_{f_i} and H_{g_j} at various grid points to get a sample of required α values. First we select a grid, $\{\mathbf{x}^k\}$. Then we evaluate the Hessian for each constraint at each grid point, $H_{f_i}(\mathbf{x}^k)$ and $H_{g_j}(\mathbf{x}^k)$, and use (6) to determine precomputed values of $\alpha_i^{f,+}(\mathbf{x}^k)$, $\alpha_i^{f,-}(\mathbf{x}^k)$ and $\alpha_j^g(\mathbf{x}^k)$ at each grid-point. During the αBB run, appropriate values of α for a given region are determined by selecting the maximum α over all grid-points contained in the region. The search is expedited by computing the $\alpha(\mathbf{x}^k)$ values in a well-defined order and making sure each α value is greater than its predecessors. That way, we can be sure that one particular grid point, the one which was calculated last, will indeed contain the maximum value of $\alpha(\mathbf{x}_k)$ over all grid points in the entire region.

The "grid" method for the calculation of α was used to determine values of α for the triatomic molecules. While the absolute theoretical guarantee is still lacking, since searching over a grid is not the same as searching over *all* points in a region, the grid alphas are quite practical and rather effective for such problems. It should be noted that the generation of values of α that are theoretically rigorous can be made based on the recent work of [15, 16].

4 Geometrical interpretation

In this section, we give a geometric illustration of how the αBB algorithm works by showing how it would locate all of the solutions of a single equation $f(x) = 0$ over the interval $x \in [0, 4]$. The function we use for our illustration is

$$f(x) = -2\cos\frac{\pi}{3}(x + 0.05) + e^{-20(x-0.2)^2} - e^{-20(x-1.6)^2} + e^{-20(x-2.4)^2} - e^{-20(x-3.5)^2}$$

A graph of this function is given in Figure 1. There are three solutions to $f(x) = 0$ in this interval. They are

$$x_{\text{sol}} \in \{0.59014, 1.82399, 3.27691\}$$

The corresponding global optimization problem is obtained by introducing a slack variable s and minimizing s subject to the constraints

$$f(x) - s \leq 0 \leq f(x) + s$$

The feasibility region for fixed s is determined by intersecting the region of space between $f(x) - s$ and $f(x) + s$ with the x-axis. This procedure is shown graphically in Figure 2. For $s > 0$, the feasibility region forms intervals around the actual solution to $f(x) = 0$. Minimizing s subject to the constraints above has the effect of pushing the two graphs together until they both meet at $f(x)$ (at $s = 0$). At $s = 0$, the feasibility region reduces to the solution set for $f(x) = 0$ (each interval reduces to a point). For $s < 0$, the graphs cross and the feasibility region is empty. $s = 0$ is clearly the global minimum whenever $f(x) = 0$ has solutions.

In order to set up the lower bounding problem, we need to find convex underestimators for $\pm f(x)$ for each interval under consideration. We begin with the complete interval $[0, 4]$. The function $f(x)$, and a valid set of convex underestimators $\hat{f}^{\pm}_{[0,4]}(x)$ are plotted in Figure 3[2].The

[2]We constructed $\hat{f}^{\pm}_{[a,b]}(x)$ by augmenting

$$L_{f,[a,b]}(x) = f(a)\frac{b-x}{b-a} + f(b)\frac{x-a}{b-a}\ ,$$

the linear interpolation of $f(x)$ over the interval $[a, b]$, by linear combinations of concave functions of the general form

$$\psi_{[a,b],k_1,k_2}(x) = \frac{1 - e^{-k_1(x-a)}}{1 - e^{-k_1(b-a)}} - \frac{e^{k_2(x-a)} - 1}{e^{k_2(b-a)} - 1}\ ,$$

which equal zero at both endpoints. For $[0, 4]$, we used

$$\hat{f}^{+}_{[0,4]}(x) = L_{f,[0,4]}(x) - 3.85\,\psi_{[0,4],1/4,7/4}(x)$$

$$-\hat{f}^{-}_{[0,4]}(x) = L_{f,[0,4]}(x) + 3.68\,\psi_{[0,4],2,1/4}(x) + 0.05\,\psi_{[0,4],1/10,100}(x)$$

convex underestimators $\hat{f}^{\pm}_{[0,4]}(x)$ essentially envelop the graph of $f(x)$ in a convex region. This convex region contains all the points $\hat{f}^{+}(x) \leq y \leq -\hat{f}^{-}(x)$, and its intersection with the x-axis is given by $\hat{f}^{+}(x) \leq 0 \leq -\hat{f}^{-}(x)$. All solutions to $f(x) = 0$ in the region $x \in [0, 4]$ must lie in this intersection region because $\hat{f}^{+}(x)$ and $-\hat{f}^{-}(x)$ surround the function $f(x)$ (see Figure 3). If this region had been empty, then no solution to $f(x) = 0$ could possibility exist in the interval $[0, 4]$. This is not the case, but see later on when we discuss the interval $[2, 3]$.

Determining whether or not the feasibility region of $\hat{f}^{+}(x) \leq 0 \leq -\hat{f}^{-}(x)$ is empty involves introducing a slack variable and minimizing it subject to

$$\hat{f}^{+}(x) - s \leq 0 \leq -\hat{f}^{-}(x) + s \tag{7}$$

This is the lower bounding problem. For $s = 0$, (7) reduces to $\hat{f}^{+}(x) \leq 0 \leq -\hat{f}^{-}(x)$. For $s \neq 0$, the feasibility region of (7) is determined by shifting the enveloping functions $\hat{f}^{+}(x)$ and $-\hat{f}^{-}(x)$ by an amount s — away from each other if $s > 0$, and towards each other if $s < 0$ (see Figure 4). Graphically, minimizing s subject to (7) involves expanding or shrinking the region between the underestimators by adjusting s until the region between $\hat{f}^{+}(x) - s$ and $-\hat{f}^{-}(x) + s$ intersects the x-axis at a single point. For the interval $[0, 4]$, this requires moving $\pm\hat{f}^{\pm}(x)$ *towards* each other, implying $s_{\min} < 0$ (in fact, the value is $s_{\min} = -2.135$). The fact that $s_{\min} < 0$ indicates that there might be solutions to $f(x) = 0$ in this interval: we will be forced to explore this region further. Note that the lower bounding problem is a *convex problem*, and so any local optimization package should reach this unique global minimum.

We therefore subdivide the interval $[0, 4]$ into two subintervals, $[0, 2]$ and $[2, 4]$, and explore each interval for solutions just as we did for $[0, 4]$. The convex underestimators for each interval, $\hat{f}^{\pm}_{[0,2]}(x)$ and $\hat{f}^{\pm}_{[2,4]}(x)$, are shown in Figure 5[3]. Note that each pair of underestimators envelops the corresponding portion of the function $f(x)$, and that the underestimators have improved: they are closer to the function $f(x)$. This will continue to happen as the intervals become narrower.

Again, the question we ask in each interval is: can a solution to $f(x) = 0$ exist there? The question is answered by solving the lower bounding problem. In both cases, the region

[3]We used

$$\begin{aligned}
\hat{f}^{+}_{[0,2]}(x) &= L_{f,[0,2]}(x) \\
-\hat{f}^{-}_{[0,2]}(x) &= L_{f,[0,2]}(x) + 4.70\,\psi_{[0,2],1,2}(x) \\
\hat{f}^{+}_{[2,4]}(x) &= L_{f,[2,4]}(x) - 3.50\,\psi_{[2,4],2,1}(x) \\
-\hat{f}^{-}_{[2,4]}(x) &= L_{f,[2,4]}(x) + 0.40\,\psi_{[2,4],1/2,10}(x)
\end{aligned}$$

$\hat{f}^+(x) \leq 0 \leq -\hat{f}^-(x)$ does intersect the x-axis (see Figure 5), indicating possible solutions in each interval. This fact is established by minimizing s subject to (7) within each interval. In both cases, $s_{\min} < 0$[4], suggesting that $\hat{f}^{\pm}(x)$ must move towards each other to reduce the feasibility region to a point (see Figures 6 and 7). Both intervals must be explored further.

So we subdivide again, and look at the intervals $[0,1]$, $[1,2]$, $[2,3]$ and $[3,4]$. The underestimators $\hat{f}^{\pm}_{[n,n+1]}(x)$ are plotted in Figure 8[5]. For the intervals $[0,1]$, $[1,2]$, and $[3,4]$, the story is the same: $s = 0$ yields feasible points, $s_{\min}$ is negative[6], and so we must subdivide those intervals further. But something new happens for $[2,3]$. The convex envelope $\hat{f}^{\pm}_{[2,3]}(x)$ completely isolates $f(x)$ from the x-axis. The lower bounding problem (7) is *infeasible* for $s = 0$. The region between $\hat{f}^{+}_{[2,3]}(x)$ and $-\hat{f}^{-}_{[2,3]}(x)$ must be *expanded* before it touches the x-axis (see Figure 9), and thus $s_{\min}$ will be greater than zero (in fact, $s_{\min} = +0.479$). We have rigorously concluded that no solution to $f(x) = 0$ can exist in the interval $[2,3]$, and so we do not need to explore this interval any further. The ability to fathom regions like this is what distinguishes αBB from a straight grid-search.

Exploration will continue with the intervals $[0,1]$, $[1,2]$, and $[3,4]$. These intervals will be subdivided and further tested. As the algorithm progresses, most intervals will eventually be fathomed. A few intervals (three, in fact) will survive. Each of these intervals surrounds a solution point, which will be located by a local search once the interval size is small enough.

The underestimators for our illustration were handpicked to illustrate the αBB algorithm at its best. Underestimators are usually not that tight. If the underestimators are not tight enough, what can happen is that an interval which does not contain any solution to $f(x) = 0$ might not be successfully fathomed. Let's look again at the interval $[2,3]$ of our example,

[4] $s^{[0,2]}_{\min} = -1.189$ and $s^{[2,4]}_{\min} = -1.150$.

[5] We used

$$
\begin{aligned}
\hat{f}^{+}_{[0,1]}(x) &= L_{f,[0,1]}(x) - 1.50\,\psi_{[0,1],1,2}(x) \\
-\hat{f}^{-}_{[0,1]}(x) &= L_{f,[0,1]}(x) + 0.55\,\psi_{[0,1],10,2}(x) \\
\hat{f}^{+}_{[1,2]}(x) &= L_{f,[1,2]}(x) - 0.30\,\psi_{[1,2],1,4}(x) \\
-\hat{f}^{-}_{[1,2]}(x) &= L_{f,[1,2]}(x) + 1.25\,\psi_{[1,2],6,2}(x) \\
\hat{f}^{+}_{[2,3]}(x) &= L_{f,[2,3]}(x) - 1.10\,\psi_{[2,3],4,10}(x) \\
-\hat{f}^{-}_{[2,3]}(x) &= L_{f,[2,3]}(x) \\
\hat{f}^{+}_{[3,4]}(x) &= L_{f,[3,4]}(x) - 1.10\,\psi_{[3,4],1,1}(x) \\
-\hat{f}^{-}_{[3,4]}(x) &= L_{f,[3,4]}(x) + 0.80\,\psi_{[3,4],3,9}(x)
\end{aligned}
$$

[6] $s^{[0,1]}_{\min} = -0.464$, $s^{[1,2]}_{\min} = -0.253$, and $s^{[3,4]}_{\min} = -0.353$.

except this time we use inferior underestimators $\hat{f}^{\text{inf},\pm}_{[2,3]}(x)$ (see Figure 10)[7]. Notice how the region between $\hat{f}^{\text{inf},+}_{[2,3]}(x)$ and $-\hat{f}^{\text{inf},-}_{[2,3]}(x)$ intersects the x-axis, even though $f(x)$ does not, suggesting that there might be solutions in that interval. In solving the lower bounding problem, $s^{\text{inf}}_{\text{min}}$ will be negative[8]. In this case, the interval will have to explored further.

The underestimators we used to find stationary states of potential energy surfaces were obtained by subtracting off a sufficiently large quadratic term from $\pm f(x)$. For functions of a single variable, the appropriate formula is

$$\hat{f}^{\alpha,\pm}_{[x^L,x^U]}(x) = \pm f(x) - \alpha^{\pm}(x^U - x)(x - x^L)$$

The values of α^+ and α^- must be positive, and chosen so that $\hat{f}^{\alpha,+}(x)$ and $\hat{f}^{\alpha,-}(x)$ are both convex over the interval $x \in [x^L, x^U]$. This condition is very simple for functions of a single variable: we must choose $\alpha^{\pm}$ so that

$$\begin{aligned} \alpha^+ &\geq \max_{x\in[x^L,x^U]}\{0, -\frac{1}{2}\frac{d^2 f}{dx^2}\} \\ \alpha^- &\geq \max_{x\in[x^L,x^U]}\{0, \frac{1}{2}\frac{d^2 f}{dx^2}\} \end{aligned} \tag{8}$$

This guarantees that $d^2\hat{f}^{\alpha,+}/dx^2 = d^2f/dx^2 + 2\alpha^+$ and $d^2\hat{f}^{\alpha,-}/dx^2 = -d^2f/dx^2 + 2\alpha^-$ will both be positive for all x in the interval in question.

For our example problem, $d^2f/dx^2 \in [-30.415, +32.305]$ over the complete interval $x \in [0, 4]$ (see Figure 11), and so minimum acceptable values of alpha are $\alpha^+ = 15.206$ and $\alpha^- = 16.152$. In Figure 12, we plot $f(x)$ and $\hat{f}^{\alpha,\pm}(x)$ for various intervals, starting with the complete interval $[0, 4]$, and subdividing.

The underestimators for the large interval $[0, 4]$ are extremely loose, but as the interval is subdivided, the underestimators improve dramatically. The maximum deviation of the underestimators from the function $f(x)$ decreases with decreasing interval size $x^U - x^L$ according to

$$\max_{x\in[x^L,x^U]} |\hat{f}^{\alpha,\pm}(x) - f(x)| = \frac{1}{4}\alpha^{\pm}|x^U - x^L|^2$$

When $x^U - x^L = \frac{1}{2}$, the underestimators are sufficiently tight that 3 out of the 5 intervals not containing any solutions are successfully fathomed (see Figure 12). When $x^U - x^L = \frac{1}{4}$,

[7] We used

$$\begin{aligned} \hat{f}^{\text{inf},+}_{[2,3]}(x) &= \hat{f}^{+}_{[2,3]}(x) - 2.0\psi_{[2,3],1,1}(x) \\ -\hat{f}^{\text{inf},-}_{[2,3]}(x) &= -\hat{f}^{-}_{[2,3]}(x) + 4.5\psi_{[2,3],1,1}(x) \end{aligned}$$

[8] $s^{\text{inf}}_{\text{min}} = -0.407$.

another 6 intervals are trimmed away, leaving just four intervals: three of them containing solutions. As the intervals are subdivided further, it will generally be the case that three intervals survive (those containing the solutions) and the other intervals will be fathomed. Once the interval size is small enough, αBB essentially becomes a binary search targeted around each solution.

The α-underestimators can be further improved for some of the smaller intervals by reducing the value of α^+ and α^-. The maximum in (8) need only be taken over the interval in question, so it may be possible to reduce them from the appropriate values over the interval $[0, 4]$. For example, it can be seen from Figure 11 that $d^2f/dx^2 \in [-10.001, 19.986]$ over the interval $x \in [\frac{1}{2}, 1]$. This permits us to use smaller alphas when generating the convex underestimators for this interval: $\alpha^+_{[\frac{1}{2},1]} = 5.000$ and $\alpha^-_{[\frac{1}{2},1]} = 9.993$, thus improving the overall quality of the underestimators. Although it obviously makes no difference for this particular interval (it cannot be fathomed whatever we do because a solution to $f(x) = 0$ is contained there), substantial savings in CPU time may result from using values of alpha computed over each region in question.

It may also be possible to improve the algorithm performance by performing "bound updates". As an illustration of this procedure, let's look again at the interval $[0, 2]$ (see Figure 5). Recall that all solutions to $f(x) = 0$ which lie in the interval $[0, 2]$ must also lie in the interval determined by $\hat{f}^+_{[0,2]}(x) \le 0 \le -\hat{f}^-_{[0,2]}(x)$, which is the intersection of the x-axis and the convex region enveloped by the functions $\hat{f}^+(x)$ and $-\hat{f}^-(x)$. If this interval can be determined, then we lose nothing by reducing the interval under consideration from $[0, 2]$ to this smaller interval. This can lead to an enhanced performance of the algorithm. The procedure is illustrated in Figure 13[9].

The new bounds for the interval are determined computationally by solving the "bound-update problem"

$$\begin{aligned} &\min_x/\max \; x \\ &\text{subject to} \;\; \hat{f}^+(x) \le 0 \le -\hat{f}^-(x) \end{aligned} \tag{9}$$

These are the same constraints as the lower bounding problem (with $s = 0$), although the objective function is different. Two optimization runs are required (one for each bound on

[9]The updated interval is $[x^L, x^U] = [0.310, 1.938]$. The convex underestimators for this interval are given by

$$\hat{f}^+_{\text{bu}}(x) = L_{f,[x^L,x^U]}(x)$$
$$-\hat{f}^-_{\text{bu}}(x) = L_{f,[x^L,x^U]}(x) + 5.05\,\psi_{[x^L,x^U],1,1.5}(x)$$

the variable), so there is a cost associated with bound updates[10]. However, if the convex envelope around $f(x)$ intersects the x-axis over a relatively small interval, reducing the interval under consideration may be well worth the extra effort. Not much is gained in the case of $[0, 2]$ (about a 20% reduction, see Figure 13), but for the interval $[1, 2]$ (see Figure 8), the reduction is substantial.

Note that the bound update procedure can be repeated, but eventually the tradeoff will not be worthwhile. For example, if we repeat the bound update procedure for the new interval $[0.310, 1.938]$ (Figure 13), we would solve (9) with $\hat{f}^{\pm}(x)$ replaced by $\hat{f}^{\pm}_{\text{bu}}(x)$, and obtain the new bounds $[x^L, x^U] = [0.404, 1.897]$ (this is where $-\hat{f}^{-}_{\text{bu}}(x)$ intersects the x-axis: see Figure 13). However, these bounds are only a slight improvement (about 8%): definitely not worth the extra effort to perform two local optimization runs. Do note that no matter how many times we update the bounds, the resulting interval must contain both solutions to $f(x) = 0$, so there is a limit to what this procedure can do for us.

The αBB algorithm may be applied to the problem of finding all stationary points of a given potential energy surface. It is especially suitable for small molecules or large molecules with a limited number of degrees of freedom. In the next few sections, we shall apply this method to triatomic molecules (3 degrees of freedom), alanine (4 degrees of freedom), alanine dipeptide (7 degrees of freedom) and tetra-alanine (8 degrees of freedom).

5 Computational Studies: Triatomic Molecules

We applied the αBB algorithm to the problem of finding all stationary points of potential energy surfaces of several triatomic molecules: HCN, HSiN, CS_2, and HBO [20, 22, 23, 21]. We used the analytic potential energy surfaces introduced by Murrell and Sorbie [24] for each of the molecules (see Tables I–V). In each case, the triatomic molecule geometry was described using the three interatomic distances R_{AB}, R_{AC}, and R_{BC}. We solved the following system of equations and inequalities:

$$\begin{aligned} \partial V/\partial R_{AB} &= 0 \\ \partial V/\partial R_{AC} &= 0 \\ \partial V/\partial R_{BC} &= 0 \end{aligned}$$

[10]It is an interesting exercise to compare the bound update procedure with solving the lower bounding problem. Using the bound update procedure for the *sole purpose* of determining the feasibility of the lower bounding problem for $s = 0$ would be inefficient because two local optimizations are needed instead of one. The benefit of performing bound updates comes from reducing the size of the interval. Solving the lower bounding problem and finding $s_{\min} < 0$ only tells us that there may be solutions to $f(x) = 0$ in the interval: the lower bounding problem does not enable us to reduce the size of this interval.

$$
\begin{aligned}
R_{AB} &\leq R_{AC} + R_{BC} \\
R_{AC} &\leq R_{AB} + R_{BC} \\
R_{BC} &\leq R_{AB} + R_{AC}
\end{aligned} \tag{10}
$$

The three equality constraints impose the stationary condition $\nabla V = 0$. The three inequality constraints impose the triangle inequality. Since the inequality constraints are linear, convex underestimators for those constraints will not be necessary: they are simply added to the lower bounding problem.

We selected as our initial region $0.7\,\text{Å} \leq R_i \leq 5.0\,\text{Å}$. The lower bound was selected by calculating forces for various choices of (R_{AB}, R_{AC}, R_{BC}) for HCN. It was found that the three atoms strongly repelled each other when the interatomic spacings were set to $0.7\,\text{Å}$, making it extremely unlikely that any important minima or transition states would exist for $R_i < 0.7\,\text{Å}$. Since interatomic distances for stable configurations of these molecules tend to be in the 1–2 Å range, an upper bound of 5.0 Å seems reasonable.

The constraints (10) turn out to be inappropriate for molecular configurations which are collinear. When the molecule is collinear, one of the triangle inequalities becomes an equality constraint. In this case, the R_i can no longer vary independently in all three directions, and it makes no sense to require that $\partial V/\partial R_i$ vanish in all three directions. For example, if atom B is in the middle, collinearity implies that $R_{AC} = R_{AB} + R_{BC}$. In this case, a simultaneous increase of R_{AC} and decrease of R_{AB} and R_{BC} is geometrically impossible. The computed values of $V(R_{AB}, R_{AC}, R_{BC})$ may very well change to first order when R_i is varied as such, and we are still at a physical stationary state. The proper way to handle collinear molecules is to first impose the collinearity condition and replace the third variable by its expression in terms of the other two: for example, replace $V(R_{AB}, R_{AC}, R_{BC})$ by $V(R_{AB}, R_{AB}+R_{BC}, R_{BC})$, and *then* impose the stationary conditions with respect to the two remaining variables:

$$
\begin{aligned}
\frac{d}{dR_{AB}} V(R_{AB}, R_{AB} + R_{BC}, R_{BC}) = 0 \\
\frac{d}{dR_{BC}} V(R_{AB}, R_{AB} + R_{BC}, R_{BC}) = 0
\end{aligned}
$$

The triangle inequality constraints are no longer needed.

It is therefore necessary to solve four problems for each triatomic molecule: one problem for the non-collinear case and three problems for the collinear case, one for each atom in the middle.

The values for $\alpha_i^{\pm}$ were selected by use of a $13 \times 13 \times 13$ grid (or a 13×13 grid for collinear molecules). The values of R_i were partitioned according to the set

$$
\{0.7, 0.8, 0.9, 1.0, 1.1, 1.3, 1.6, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0\}
$$

and Hessian values $H_{\partial V/\partial R_i}$ were computed for each constraint at each grid point. Note that the Hessian calculation involves *third* derivatives of V: the Hessian for $\partial V/\partial R_i$ contains elements of the form $\partial^3 V/\partial R_i \partial R_j \partial R_k$. These derivatives were evaluated analytically. Since the potential energy terms contain factors which decrease exponentially with interatomic spacing, it is expected that higher alpha values occur for *smaller* interatomic spacings. We made sure of this by maximizing the alpha values obtained for a given grid point over all other grid points with higher interatomic spacings. Appropriate values of alpha for the region

$$[R_{AB}^L, R_{AB}^U] \times [R_{AC}^L, R_{AC}^U] \times [R_{BC}^L, R_{BC}^U]$$

are obtained by rounding down the lower bounds of each interval to the nearest grid point and using the precomputed grid alphas stored there.

The results of our search for the molecules HCN, HSiN, CS_2, and HBO are summarized in Tables VI–X. The minima and first-order saddles and their connectivity are depicted in Figures 14–18. Note that two potential energy surfaces for HBO were used, since they were provided in the literature. "+++++" indicates the long side of a collinear molecule in the tables.

The results for HCN are listed in Table VI and shown in Figure 14. min.1 and min.2 are, the collinear HCN and HNC isomers, respectively. The two minima are connected via the transition state 1st.1. It is clear that the transition between HCN and HNC takes place via transfer of the hydrogen from one side of the molecule to the other. This transition is discussed in [20]. In addition to these stationary points, there are two other minima and three other transition states. min.3 is shaped like a small triangle and is connected by a very shallow barrier to min.2 via the transition state 1st.2. min.4 corresponds to a weak hydrogen-bonded structure $C\cdots HN$ (also discussed in [20]). It can make a transition to min.2 via a simple rotation of the hydrogen around the nitrogen atom. min.4 can also make a transition to min.1 via a slightly more complicated transition involving a hydrogen exchange between N and C along with a rotation around the carbon atom. The escape barriers for min.3 and min.4 are relatively shallow, indicating that they play very little role in the dynamics of HCN. min.3 appears to be nothing more than a wrinkle in the potential energy surface. There is a good chance that it is an artifact of the construction of the analytic potential energy surface, and therefore does not correspond to any physical state.

The results for HSiN are listed in Table VII and shown in Figure 15. min.1 and min.2 are the collinear HNSi and HSiN isomers, respectively. Note the reversal in priority of these two minima as compared with the analogous HCN molecule: HNSi is the ground state. These two configurations are connected via the transition state 1st.4. The transition between HNSi and HSiN is clearly analogous to the transition between HCN and HNC. This transition is discussed in [20]. There are four additional first-order transition states. Two of them (1st.1 and 1st.5) connect min.1 and min.2 respectively to the dissociated state SiN+H. The remaining two first-order transition states (1st.2 and 1st.3) connect min.2 and min.1 respectively to an artificial minimum of extremely low energy corresponding to SiH $\cdots$ N (this minimum does not appear in Table VII because the solution is outside of the bounds $0.7\,\text{Å} \leq R_i \leq 5.0\,\text{Å}$: the nitrogen atom is about 6 angstroms from SiH). This anomalous region of the potential energy surface can be explained by close examination of the analytic formula in Table II. The exponential cutoff of the 3-body interaction term V_I is governed by the γ_i parameters. Note that γ_2 and γ_3 are assigned very small values. This means that V_I is not cut off properly when R_{SiN} and R_{NH} take on large values. This allows one of the terms in V_I (namely, $-1.0421\rho_2^2$) to make a large negative contribution to the potential energy. It is clear that this region of the potential energy surface is an artifact of its construction and is not based on any physical reality. Note that both barriers of entry into this anomalous region (1st.2 and 1st.3) are lower than the barrier between HSiN and HNSi (1st.4), making this potential energy surface a highly dubious choice in any kinetic study of HSiN.

The results for CS_2 are listed in Table VIII and shown in Figure 16. min.1 corresponds to the collinear SCS conformation. min.1 is connected to itself via the transition state 1st.2. The transition involves a 360° rotation of one of the sulfur atoms around the CS. The other transition state (1st.1) is the transition structure for the reaction $C + S_2 \rightleftharpoons CS + S$. Careful analysis reveals that min.1 is also connected to the dissocated state S+CS without a barrier. There is very little discussion of CS_2 in [21] since their primary concern was with the tetra-atomic molecule OCS_2.

The results for HBO (PES1) are listed in Table IX and shown in Figure 17. min.1 corresponds to the collinear HBO minimum. It is connected to itself via the transition state 1st.2. The transition involves a 360° rotation of the hydrogen atom around the boron atom (the oxygen atom gets out of the way). The other transition state (1st.1) connects min.1 to the dissociated state H+BO. This potential energy surface was discussed in [22], where it was claimed that there is a second minimum corresponding to a "bent" HOB structure. We were unable to locate this minimum or any transition state which might lead to this minimum using αBB. We also did not locate this minimum using an eigenmode-following algorithm (Eigenmode III, described in Appendix A), using Figure 1 of [25] to obtain a reasonably close initial guess. Our conclusion is that this second minimum does not exist on the potential energy surface described in [22]. We suspect that there may be a typographical error in that potential energy surface[11].

The same authors also provided a second potential energy surface for HBO (called PES2) in [23]. Our results for this potential energy surface are listed in Table X and shown in Figure 18. We found both the collinear HBO (min.1) and collinear HOB (min.2) minima, as well as the transition between them (1st.1). The other transition state (1st.2) connects min.1 to itself, with the same 360° rotation of the hydrogen atom around the boron atom as found in PES1.

6 Computational Studies: Alanine

We have also applied the proposed approach to the problem of finding all stationary states of alanine. We used the ECEPP/3 potential energy surface [26], fixing all bond lengths and bond angles to their equilibrium values, and allowing just the four relevant dihedral angles, ϕ, ψ, ω, and χ to vary (see Figure 19). Note that the dihedral angle χ specifies the rotation of a methyl group, and therefore has a 120° rotational symmetry. The other angles are allowed to vary over a complete 360° interval.

We performed our search for stationary states using the proposed method by trying various fixed values of $\alpha_i^{\pm}$. The results of our searches, which were performed on an HP-132B workstation, are summarized in Table XI. The 17 local minima and 62 first-order transition states can be found in Tables XII and XIII, respectively.

We appear to have found all of the minima and first-order transition states by setting $\alpha_i^{\pm} = 5$. This conclusion is based on the fact that increasing α to 10, 15, and finally 20 did not locate any additional states. As a final check, we conducted a brute force Eigenmode III search for minima and first-order transition states. In each case, we laid down a $9 \times 9 \times 9 \times 9$ grid of starting points and allowed Eigenmode III to take up to 100 steps in the direction of the stationary state. The Eigenmode III search located the same 17 local minima and 62 first-order transition states, and took about 229 CPU seconds[12].

[11] Communication with the original authors was unsuccessful.

[12] A couple of points should be made regarding the Eigenmode III grid search. First of all, although the Eigenmode III grid search appears to be more (time) efficient than αBB in locating stationary states, there is no guarentee, even theoretical, that all solutions will be found by the grid search. Increasing the size of the grid will increase the likelihood that any given stationary point will be found, but at a cost of increasing the CPU time required for the search. It should also be pointed out that the time required for the grid search is exponential in the number of degrees of freedom of the molecule, and so is only practical for small molecules.

The ψ–ϕ plot for alanine is shown in Figure 20. It shows the 17 local minima (large numbered dots) and 62 first-order transition states (small unnumbered dots), as well as the connections between them. Further discussion of alanine can be found in Section 8.

7 Computational Studies: Alanine Dipeptide

We have also applied the αBB approach to alanine dipeptide (also known as "terminally blocked alanine"). We again employed the ECEPP/3 potential energy surface [26], fixing all bond lengths and bond angles to their equilibrium values, and allowing the 7 most relevant dihedral angles $(\theta_1, \theta_2, \phi, \psi, \omega, \chi, \theta_3)$ to vary (see Figure 21). The dihedral angles θ_1, χ, and θ_3 describe the orientation of a methyl group, and therefore exhibit a 120° rotational symmetry. The other dihedral angles can in principle vary over a complete 360° interval.

Unlike alanine, it is both impractical and uninteresting to allow all of the dihedral angles to vary completely over their domain. For example, angles such as ω and θ_2 which control the orientation of the CO double bonds are usually confined to values near 180°, at least for the lower energy conformers. We determined the bounds to use for each angle by first conducting an Eigenmode III search for minima and first-order transition states on a 6^7 point grid, and then examining the range of values taken by each angle for the lowest energy minima and first-order transition states. We came up with the following bounds[13]:

$$\begin{aligned}
\theta_1 &\in [108.9°, 189.1°] \\
\theta_2 &\in [166.2°, 189.1°] \\
\phi &\in [0°, 360°] \\
\psi &\in [0°, 360°] \\
\omega &\in [166.2°, 189.1°] \\
\chi &\in [108.9°, 229.2°] \\
\theta_3 &\in [108.9°, 189.1°]
\end{aligned}$$

After selecting the bounds for all of the variables, we performed an αBB run with $\alpha_i^{\pm} = 30$. After 550958 iterations and 62204 seconds of CPU time on an HP-132B workstation, we found a total of 203 stationary points, including 10 local minima, 38 first-order saddles, 60 second-order saddles, 54 third-order saddles, 30 fourth-order saddles, 9 fifth-order saddles,

[13] The reason why the bounds are not nice round numbers (when expressed in degrees) is because we expressed them in radians when we selected them.

and 2 sixth-order saddles. The 10 local minima are listed in Table XIV and the 38 first-order transition states are listed in Table XV.

The ψ–ϕ plot for the alanine dipeptide is shown in Figure 22. Unlike alanine, the alanine dipeptide molecule has been extensively studied in the literature [27, 28, 29, 30, 31, 32, 33, 34, 35]. More common names have been assigned to the various stationary points and/or regions of the ψ-ϕ plot for alanine dipeptide. The C_7^{eq} conformation (so named because of the 7-atom ring closed by a hydrogen bond [27]) is widely regarded as the global minimum, and indeed corresponds well with min.1. Its mirror image, C_7^{ax}, corresponds to both min.9 and min.10. C_5 (or C_5^{ext}) corresponds to min.2, which is an extended conformation. The "polyproline conformation", P_{II}, corresponds to min.5. α_{R} refers to the conformations which tend to form right-handed alpha-helix structures in larger peptides, and correspond to min.3, and in some cases, min.6. Its mirror image, α_{L}, corresponds to min.7. One reference, [32], indicates the existence of a "β_2" structure which is fairly close to min.4, as well as an "α_d" which corresponds to min.8.

None of the references above acknowledged the existence of all 10 minima. Of course, it must be acknowledged that most of these studies used potential energy surfaces other than ECEPP/3, which may account for the difference. Reference [30] includes a study of alanine dipeptide using ECEPP/2 (which is identical to ECEPP/3 for this molecule). In Table V of [30], they list 5 minima, while in Figure 10 of [30], it is clear from the contour plot that at least 3 additional minima exist, and there is strong evidence for the existence of the remaining two minima.

Further discussion of alanine dipeptide can be found in Section 8.

8 Reaction pathways

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.

First we determine the connectivity between the local minima and the first-order transition states. We followed each first-order transition state to the two local minima it connects by using the Eigenmode III algorithm. The transition state is perturbed slightly in each of the two directions along the reaction coordinate (i.e., along the one eigenmode corresponding to a negative eigenvalue), and then Eigenmode III is used to find a local minimum from that starting point. This gives us a list of (minimum, transition state, minimum) triples.

The lists of triples for alanine and alanine dipeptide are given in Tables XVI and XVII, respectively. These triples define the connectivity between pairs of minima through transition states, and were used in the construction of the ψ–ϕ plots for alanine and alanine dipeptide which were given in Figures 20 and 22, respectively.

Once the triples are determined, we can calculate the transition rate matrix using Rice–Ramsperger–Kassel–Marcus (RRKM) theory [17, 18]. The basic assumption behind RRKM theory is that the behavior of the molecule near the transition state can be treated thermodynamically, even though the transition state itself is not stable. A transition from one minimum to another through a given transition state is considered to have taken place when the transition state is reached via thermal fluctuation. The probability that this occurs is proportional to the ratio of the partition functions around the transition state and initial minimum respectively (see Appendix B for details). The rate associated with this transition is given by

$$W_{j' \to \mathrm{ts} \to j} = \frac{kT}{h} \frac{Q_{\mathrm{ts}}}{Q_{j'}} \tag{11}$$

The partition functions are evaluated using the harmonic approximation around a given stationary point. The result is

$$Q = e^{-E_0/kT} \prod_i \frac{kT}{h\nu_i} \tag{12}$$

where E_0 is the potential energy and ν_i are the vibrational frequencies of the molecule around the stationary point in question. The product is taken over all vibrational modes (i.e., with positive ν_i^2). Note that for a transition state, this is one less mode than for a local minimum since the reaction coordinate direction is excluded. Putting (11) and (12) together, we obtain

$$W_{j' \to \mathrm{ts} \to j} = \frac{\prod_i \nu_i^{j'}}{\prod_{i \neq \mathrm{r.c.}} \nu_i^{\mathrm{ts}}} e^{-(E_{\mathrm{ts}} - E_{j'})/kT} \tag{13}$$

Note that the vibrational frequencies take into account the vibrational entropy of the system. It is important to remember that for $T \neq 0$, it is the free energy surface which governs the reaction process.

The transition rate matrix is defined as

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

It can be calculated for a given molecule once we have determined all of the local minima, transition states, and their connectivity. Vibrational frequencies are computed by solving the generalized eigenvalue problem:

$$(H - (\nu/2\pi)^2 I)\, x = 0 \tag{14}$$

where H is the Hessian of the potential energy function, and I is the generalized inertia tensor[14]. The transition rate matrix for alanine at room temperature $T = 300\,\mathrm{K}$ is given in Table XVIII. This can be converted into a transition probability matrix by normalizing each column so that they add up to 1. The transition probability matrix for alanine is given in Table XIX.

The time-evolution of occupation probabilities for each of the 17 stable configurations is determined as follows:

$$\begin{aligned}\frac{dP_j}{dt} &= \left(\begin{matrix}\text{overall rate of transitions}\\ \text{into state } j\end{matrix}\right) - \left(\begin{matrix}\text{overall rate of transitions}\\ \text{out of state } j\end{matrix}\right)\\ &= \sum_{j'\neq j} W_{jj'}P_{j'}(t) - (\sum_{j'\neq j} W_{j'j})P_j(t)\end{aligned}$$

which can be rewritten in a more compact form as

$$\frac{dP_j}{dt} = \sum_{j'} w_{jj'}P_{j'}(t) \tag{15}$$

where $w_{jj'}$ is equal to $W_{jj'}$, except that the diagonal elements are replaced by

$$w_{jj} = -\sum_{j'\neq j} W_{j'j}$$

Note that the sum over each column of $w_{jj'}$ is zero.

Coupled linear differential equations like (15) are solved by diagonalizing the matrix $w_{jj'}$, so that

$$\sum_{j'} w_{jj'}u^{(i)}_{j'} = \lambda^{(i)}u^{(i)}_j$$

Substituting $P_j(t) = \sum_i p_i(t)u^{(i)}_j$ into (15) yields a system of decoupled differential equations

$$\dot{p}_i = \lambda^{(i)}p_i(t)$$

which can easily be solved: $p_i(t) = p_i(0)e^{\lambda^{(i)}t}$. The general solution to (15) can therefore be written

$$P_j(t) = \sum_i p_i(0)e^{\lambda^{(i)}t}u^{(i)}_j$$

where the coefficients $p_i(0)$ are determined by the initial conditions $P_j(0)$.

One of the eigenvalues $\lambda^{(0)}$ is zero (this is a consequence of $\sum_j w_{jj'} = 0$). The associated eigenvector corresponds to the equilibrium probability distribution:

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

[14]The generalized inertia tensor I is defined so that the kinetic energy of the molecule is $\frac{1}{2}\sum_{i,j}\dot{x}_iI_{ij}\dot{x}_j$.

All other eigenvalues are negative, and correspond to transient probabilities. The decay time is given by $\tau^{(i)} = -1/\lambda^{(i)}$. The probability eigenvectors and corresponding decay times for alanine at $T = 300\,\mathrm{K}$ are given in Table XX.

The decay times range from $1.6 \times 10^{-13}\,\mathrm{s}$ to $5.9 \times 10^{-11}\,\mathrm{s}$. All of the interesting dynamics takes place over the time interval $10^{-14}\,\mathrm{s} \le t \le 10^{-9}\,\mathrm{s}$. A logarithmic plot of $P_j(t)$ over this time range for each of the 17 "pure state" initial conditions is given in Figures 23–25.

We make the following observations regarding alanine. The energies and molecular configurations for the 17 local minima are summarized in Table XII, and the connectivity is summarized in the ψ–ϕ plot in Figure 20. There appear to be two main groups of minima, min.01–min.08, and min.09–min.17, the energy gap between the two groups being much larger than the energy gaps within the two groups. At room temperature equilibrium, alanine occupies the lower group of minima almost exclusively. This is because energies within the lower group are within kT (0.597 kcal/mole) of the ground state energy, but the higher group is 6–10 times kT higher in energy.

Because transition rates between the lower 8 states are so large (most of the barriers are smaller than kT), it is difficult to partition this group any further. It is worth observing that min.01 is only the third most likely occupied state at equilibrium. min.07 and min.08 have higher equilibrium probabilities. Even though these two states are almost kT higher in energy, they have a large partition function because one of the vibrational frequencies is small (see (12)). This is particularly noticeable for min.07 (the state with the highest occupation probability), where one of the eigenvalues of the Hessian is very nearly zero.

In contrast with the lower 8 states, the upper 9 states can be partitioned further. Based on the transition probability matrix in Table XIX, we discover that min.09–min.17 partition into 3 groups: (min.09, min.10, min.11), (min.12, min.14, min.15), and (min.13, min.16, min.17) (see Figure 26). The transitions between states within each group are considerably more likely than transitions between states in different groups. Only in the case of the third group, 13-16-17, is there a significant probability of a transition to the lower 8 states, indicating a "slow" decay from the upper 9 to the lower 8 states. Actually, "slow" may be a bit of a misnomer, since equilibrium is achieved from any initial probability distribution in less than $10^{-9}\,\mathrm{s}$.

The groupings come to life when we examine the solutions to the Master equation (15). For example, if we follow the time-evolution of alanine starting in state min.09 (i.e., $P_9(0) = 1$, all others 0), we see that for the first $10^{-12}\,\mathrm{s}$, P_{10} and P_{11} increase to significant values, indicating transitions into those states. Only after $10^{-11}\,\mathrm{s}$ do we begin to see significant amounts of min.01–min.08. The other two groups (the 12-14-15 and 13-16-17 groups) see very little action, indicating a low probability of transition into these groups from 9-10-11. The time-evolutions corresponding to starting in min.10 and min.11 are similar.

If we start in the 12-14-15 group, the effect is very similar, although transitions from 12-14-15 to 9-10-11 is somewhat more likely than the reverse (no doubt this is energy-related). It seems that the path from 12-14-15 to the lower group is as likely as not to pass through one of the 9-10-11 states first.

The 13-16-17 group is the most interesting. Notice that min.13 is very stable by comparison to the other upper 9 minima because the barriers to escape from this particular minimum are very high. If alanine starts in min.13, it is very likely to make a transition directly from min.13 to the lower group. Examining the transition probability matrix, we see that min.13 also has a fairly high likelihood of making a transition to min.16 or min.17. But the time-evolution of min.13 indicates a maximum of about 1% occupation probability of occupying either min.16 or min.17. This is resolved by realizing that min.16 and min.17 both have *much* shorter lifetimes, about 100 times shorter than min.13. Thus, alanine will spend very little time in either of these states, preferring instead to make a transition back to min.13, or to the lower group. This fact is also evident in the time-evolution of min.16 and min.17. In both cases, the occupation of min.13 peaks out around $t = 10^{-12}\,\mathrm{s}$ with an occupation probability around 80%. This implies that min.16 and min.17 will very likely make a transition to min.13, where alanine will spend a relatively long time, before reaching the lower group.

This analysis is further facilitated by constructing a "rate disconnectivity graph" for alanine, in which the connectivity of the minima are classified according to the transition rates.

Minima may be classified into groups according to their connectivity. Transitions between minima which fall below a certain cutoff rate are ignored. The remaining transitions (the "fast" ones) define the rate-dependent connectivity of the potential energy surface. The minima are then partitioned into connection components according to this connectivity: minima within the same group are connected via a reaction pathway, minima within different groups are not connected. A larger rate cutoff results in fewer transitions being considered, and therefore a finer partition will result. For example, if the rate cutoff exceeds the highest transition rate, the partition will be the discrete partition (each minimum will fall into its own group, separate from all other minima). If the rate cutoff is lower than the lowest transition rate, all minima will fall into one group.

This partitioning into groups may be visualized by drawing the rate disconnectivity graph. The vertical axis represents the transition rate cutoff, with lower cutoffs appearing higher in the graph. If we start at the top of the graph, all minima fall into one group which is represented by a single node. At some cutoff value, a critical transition gets eliminated which disconnects the minima into two groups. This is represented graphically 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. Finally, the rate cutoff is sufficiently high that there are no transitions left: each minimum will lie in a group by itself. The minima are identified at the base of the graph.

The rate disconnectivity graph for alanine at room temperature is shown in Figure 27. The partitioning of the upper nine states into the 3 subgroups 9-10-11, 12-14-15, and 13-16-17 is now transparently obvious. The rate of transitions between two minima is indicated on the rate disconnectivity graph by how far up the tree they are connected. Minima within the same group are connected at a much lower point on the tree than minima within different groups, suggesting that transitions within the same group are much faster.

Let's focus for a moment on the 9-10-11 group of alanine. From the rate disconnectivity graph, we can readily see that transitions within the group occur at a rate in excess of 10^{12} Hz, whereas transitions in or out of the group occur at a rate of around 10^{11} Hz, which is about 10 times slower. It is likely that if alanine is prepared in (or makes a transition to) one of these states, it will make several transitions within the group before finally making a transition to another state outside the group. Similar observations hold for the 12-14-15 and 13-16-17 groups.

Upon close examination of the rate disconnectivity graph, it appears that the states min.01–min.08 do in fact partition into subgroups. It appears that (min.01, min.03, min.04) form one group, which is somewhat separated from (min.02, min.05, min.06, min.08). Evidence for these groupings does exist in the solutions to the Master equation, if one looks carefully enough. For example, in the time evolution of min.01, P_3 and P_4 rise slightly above the rest of the minima (with the exception of P_7), indicating slightly higher transition rates among that group of states. Similar observations may be made for the time evolution of min.02–min.06 and min.08 (min.07 seems to be doing its own thing).

The partitioning of min.01–min.08 is much less pronounced than the partitioning of min.09–min.17 because transitions amongst all of the states in the lower group min.01–min.08 are fast to begin with. Nevertheless, the rate disconnectivity graph reveals these groupings very clearly.

Let us compare the kinetic analysis above with the ψ–ϕ plot in Figure 20. The first thing to note is that there is no evidence of the partitioning into the two main groups (min.01–

min.08 and min.09–min.17) in the ψ–ϕ plot. However, this is because the large energy gap is not caused by rotations of ϕ or ψ, but is instead caused by rotations of ω. The lower energy group corresponds to $\omega \sim 180°$ and the upper energy group corresponds to $\omega \sim 0°$ (see Table XII). This is consistent with the well-known fact that $\omega \sim 180°$ is strongly energetically favored in peptides in general.

Evidence of further partitioning of the two main groups into subgroups can be found by examining the ψ–ϕ plot. Each of the subgroups (1-3-4, 2-5-6-8, 9-10-11, 12-14-15, and 13-16-17) lies in horizontal "strips". This corresponds to free variation in ϕ, but little variation in ψ. The implication of our kinetic analysis is that variations in ϕ occur more readily (i.e., with lower barriers) than variations in ψ in alanine. We will shortly find that quite the opposite is true in the longer peptides.

We can also perform a detailed analysis of alanine dipeptide using the same techniques as with alanine. As a reminder, the list of triples is given in Table XVII and the ψ–ϕ plot is given in Figure 22. We calculated the transition rate matrix, which is given in Table XXI. The probability eigenvectors and decay times were also calculated, and are given in Table XXII. The time-evolution of the occupation probabilities are shown in Figures 28 and 29. Finally, the rate disconnectivity graph is shown in Figure 30.

The first thing to notice is the partitioning of the 10 minima into two connection components, min.01–min.06 and min.07–min.10, characterized by the fact that *there are no transitions between these two groups*. This can be confirmed by examining the ψ–ϕ plot, or by examining the transition rate matrix which will be in block diagonal form, as shown in Table XXI. In fact, examination of the ψ–ϕ plot reveals that the two groups are separated according to their ϕ values (for min.01–min.06, ϕ lies in the interval $[200, 290]$, and for min.07–min.10, ϕ lies in the interval $[50, 80]$). This suggests that for alanine dipeptide, the ψ degree of freedom is flexible, but the ϕ degree of freedom is totally inflexible. This is quite the opposite of what we observed for alanine, and the contrast between the two degrees of freedom is much more extreme in this case[15].

Because there are no transitions from min.01–min.06 to min.07–min.10, both of these groups of minima evolve independently of one another. The transition rate matrix will partition into block diagonal form. Also, the probability eigenvectors will partition into two classes, one class describing the dynamics within the group min.01–min.06 (eigenvectors

[15] In fact, it turns out there *are* transition states which connect min.01–min.06 and min.07–min.10 provided that we are willing to relax the bounds we placed on the other degrees of freedom (i.e., allowing all variables to range from 0° to 360°). However, all of the resulting barriers are extremely high, and so the transitions between these two groups of minima are slow by comparison. The transition from min.07 down to min.01 (the most relevant transition) takes about 0.01 s.

$u^{(0-5,a)}$) and the other class describing the dynamics within the group min.07–min.10 (eigenvectors $u^{(0-3,b)}$). There will be two distinct equilibrium probability distributions, which is apparent from the list of probability eigenvectors, and also from the time-evolution of the occupation probabilities (e.g., compare the long time behavior of the system initially occupying one of min.01–min.06 to the same system initially occupying one of min.07–min.10). The rate disconnectivity graph also partitions into two trees with no connection between them.

We will focus on each group separately. The transition rate matrix for min.01–min.06 takes on a peculiar shape. Minima in min.01–min.03 can make direct transitions only to minima in min.04–min.06, and vice versa. This is manifested in the rate disconnectivity graph which separates min.01, min.02, and min.03 from each other, as well as min.04, min.05, and min.06. It can also be inferred from the occupation probabilities. For example, if the system initially occupies min.01, then P_2 and P_3 take longer than P_4 and P_5 to attain appreciable values. Similar behavior is observed for min.02–min.06.

The dynamics of min.07–min.10 also has some interesting characteristics. The equilibrium probability distribution is concentrated on the state min.07 because the energy levels are so widely spread out (the four minima cover an energy range of more than $8kT$). According to the rate disconnectivity graph, min.07–min.10 partitions further into (min.07, min.08) and (min.09, min.10). This is actually quite evident from the transition rate matrix (from which the graph is constructed), and appears in the time-evolution of min.07–min.10. Apparently, there is only one transition from (min.09, min.10) to (min.07, min.08) (the transition state is 1st.37), and it has a rather high barrier (almost $7kT$). The transition out of min.09 is particularly slow, since it must first climb to min.10 before making the transition through 1st.37 down to min.08.

9 Computational Studies: Tetra-alanine

We also studied tetra-alanine, employing the ECEPP/3 potential energy surface [26], fixing all bond lengths and bond angles to their equilibrium values, as well as most of the dihedral angles, and allowing only the 8 (ϕ, ψ) variables to vary (see Figure 31).

Tetra-alanine is one of the smallest peptides which can exhibit a full alpha-helical turn (corresponding to $(\phi_i, \psi_i) = (300°, 300°)$), as well as an extended conformation (a beta sheet conformation corresponding to $(\phi_i, \psi_i) = (300°, 120°)$) [35]. The isobutyryl-(ala)$_3$-NH-methyl (IAN) tetra-peptide, a variant of tetra-alanine, has been studied in Ref. [19, 35]. They found 139 local minima and 502 first-order transition states.

We conducted a brute force Eigenmode III search for the minima and first-order transition states. In each case, we laid down a rather coarse grid (4^8 points) of starting points and allowed Eigenmode III to take up to 100 steps in the direction of the stationary state. The Eigenmode III search located 7543 minima and 18902 first-order transition states. When we followed each of the first-order saddles to the minima it connects, we located an additional 8582 minima, for a grand total of 16125 minima. It is very likely that there are even more minima and transition states that we have not located yet.

We searched the stationary states for alpha-helical and extended conformation minima. For alpha-helical states, we searched the following region for minima

$$270^\circ \leq \phi_i \leq 330^\circ \qquad 270^\circ \leq \psi_i \leq 335^\circ$$

and found a single state, the ground state min.1, which satisfied these constraints. For extended conformation states, we searched the following region for minima

$$270^\circ \leq \phi_i \leq 330^\circ \qquad 90^\circ \leq \psi_i \leq 150^\circ$$

and found a single state, min.1583, which satisfied these constraints. These two minima are listed in Table XXIV along with other minima. We will be focussing on the transition from min.1583 to min.1.

The first step towards understanding how a protein might proceed from an extended state (min.1583) to a folded state (min.1) is to enumerate the reaction pathways between these two states. This enumeration can be done 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 triple which connects two minima. The set of all paths from one given 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 (so that the most relevant pathways would be found). We can restrict the set of pathways in any number of ways: (1) select pathways with monotonically decreasing energies, (2) select pathways whose length (number of visited minima) is less than some cutoff value, or (3) select pathways whose highest energy barrier is below some cutoff.

To analyze tetra-alanine, we chose method (2): restricting the path length. We began by searching for pathways of increasing length until we finally found one. The shortest pathways have a length of 10 (i.e., 10 total minima, including min.1583 and min.1, and therefore 9 transitions). The number of pathways found for a given length limit is shown in Table XXV for length limits up to 15. The 43 pathways corresponding to a length limit of 11 are listed in Table XXVI and displayed in graph form in Figure 32. One of the pathways (the first one in Table XXVI) is shown pictorially in Figure 33. Table XXIV contains a list of all minima found in these pathways. The minima are classified according to Table XXIII.

In order to prioritize the pathway list, we estimated the amount of time it would take for tetra-alanine to follow a particular pathway from min.1583 to the ground state. This was accomplished by solving the Master equation for a reduced system consisting only of the minima and transition states which lie along the given pathway in question. The decay time of the longest lived transient probabilities was used as an estimate of the overall transition time.

Most transition times lie between $10^{-10}\,$s and $10^{-8}\,$s. However, there were 7 pathways which involved very slow transitions (over a second), which corresponded to very long transition times[16]. In each case, it is possible through examination of the transition rate matrix to determine exactly where the bottle-necks occur. Very slow transitions (indicative of very high barriers) are indicated in Table XXVI by an asterisk between the two minima. These long-lived pathways are very unlikely to play any serious role in the dynamics of protein folding, since they occur so slowly.

It therefore makes sense to eliminate these pathways by effectively removing the very slow transitions. We applied a rate cutoff of $10^6\,$Hz, eliminating triples corresponding to transition rates below that cutoff, and recalculated the pathways. The number of pathways remaining after applying the cutoff for each length limit considered earlier is also listed in Table XXV. The 7 pathways marked “long” in Table XXVI were eliminated when the rate cutoff was applied, but the other 36 pathways of length 11 or less survived, as expected.

We next determine the time required overall by tetra-alanine to find its ground state from min.1583. We expect the overall transition time to be on the order of $10^{-10}\,$s, the estimated transition time for the fastest pathways. Unfortunately, there are over 16000 minima, and even if we only pay attention to those minima which are connected to min.1, we still have over 12000 minima to work with. Solving the Master equation involves finding eigenvectors and eigenvalues of a 12000×12000 matrix, which does not fit in our computer's memory. However, if we again apply the same rate cutoff, $10^6\,$Hz, this reduces the number of minima still connected to min.1 to 2173, which is a number we can handle. In effect, we are focussing our attention to a subtree of the rate disconnectivity graph containing min.1 and with a root node at around $10^6\,$Hz. This subtree contains 2173 minima.

[16]The actual value of the transition time is meaningless, since the eigenvalue of the transition rate matrix which corresponds to the long transition time is on the order of the machine epsilon, and will therefore be dominated by round-off error.

The Master equation was solved over the 2173 minima, resulting in Figure 34. Evidently, tetra-alanine leaves min.1583 after on the order of 10^{-12} s and eventually finds its way to min.1 after 10^{-10} s. The time in between is obviously spent passing through intermediate states.

We next analyze the pathways listed in Table XXVI in terms of the changes in the underlying geometry of the tetra-alanine molecule as it proceeds from min.1583 to min.1. We do this by first classifying the minima according to its (ϕ_i, ψ_i) values. Each pair of (ϕ, ψ) values is classified by a single letter optionally followed by one or two primes, as summarized in Table XXIII. The geometry of a given minimum is therefore described by four such letters. Note particularly that the letter "a" (without a prime) indicates values of (ϕ, ψ) which are within 30° of the nominal values expected for an alpha-helical structure, and similarly, "b" indicates values of (ϕ, ψ) which are within 30° of the values expected for a beta-sheet structure. Therefore, min.1 is classified as "aaaa" and min.1583 is classified as "bbbb". The classification of each minimum is included in the list of minima (Table XXIV) as well as the list of pathways (Table XXVI).

During the transition from min.1583 to min.1, each of the four ψ angles must rotate from "b" to "a". Examining Table XXVI reveals that this process takes place in small well-defined steps. If we look at the fastest pathway, we see that in this case, each ψ angle rotates about half-way (i.e., b $\rightarrow$ i) until an "ii′ii" structure is reached, and then each ψ angle rotates the rest of the way (i.e., i $\rightarrow$ a). This pathway is illustrated in Figure 33. Other pathways approach the alpha-helical structure more creatively. For example, the second fastest pathway begins by rotating 3 out of the 4 ψ angles half-way, reaching an "ii′bi" structure, before rotating ψ_2 all the way to "a". Then, ψ_3 is finally rotated from "b" to "i", but not before going through "j" first. Finally, the remaining ψ variables are rotated the rest of the way to the alpha-helical structure of min.1.

We make the following general observations regarding the rotation of the ψ angles during the transitions from min.1583 to min.1:

- Each ψ angle normally progresses in the sequence b $\rightarrow$ i $\rightarrow$ a or b $\rightarrow$ j $\rightarrow$ i $\rightarrow$ a.

- No direct b $\rightarrow$ a transitions are observed, indicating that a rotation of ψ from beta-sheet to alpha-helical values is too large for a single transition.

- The ψ angle rarely proceeds backwards, which is to be expected since these are the shortest pathways. However, cases do exist where this happens (e.g., min.34 $\rightarrow$ min.74 in the second very slow pathway in Table XXVI).

- Usually only one ψ angle changes at a time. The exceptions tend to occur in the slower pathways.

- There is a tendency for most of the ψ angles to rotate half-way (b $\rightarrow$ i) before any of them rotate the rest of the way (i $\rightarrow$ a). In other words, tetra-alanine tends to visit intermediate states consisting of 3 or 4 i's during its progress from min.1583 to min.1.

The rotation of ϕ angles plays much 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. Furthermore, as is true for alanine dipeptide, rotations in the ϕ angle are far less flexible than rotations in the ψ angle. To understand this, consider the very slow transitions shown in Table XXVI. The bottle-necks tend to occur whenever a ϕ angle rotates from within the interval $[180°, 330°]$ (indicated by one or zero primes) to outside this interval (indicated by a double-prime). In fact, it is always the case that whenever such a ϕ rotation takes place, the transition is a slow one.

We next construct the rate disconnectivity graph for tetra-alanine. Although it is not difficult to construct the complete graph, it is difficult to display it in a readable format since there are 16125 minima. Even if we focus our attention on those minima connected to min.1, there are still 12373 minima to deal with. We were still able to make several comments about the graph just from wandering about it and examining various subgraphs.

The first thing we noticed is that most nodes split very unevenly: often one of the two branches consists of a single state, or a very small number of states. A graph of this sort of shape (single minima breaking off from the main branch) suggests a funnel-type potential energy surface [19].

The graph begins at the slowest transition rate, which is 1.29×10^{-17} Hz. The graph involves small groups breaking away from the central branch until a transition rate of around 1 Hz at which point the main branch still contains 7972 minima. Between 1 Hz and 12 Hz, several groups of several hundred states each break away from the main branch. At 12 Hz, 2176 minima remain. This suggests a great deal of interesting dynamics occuring on a time-scale of about one second.

It is interesting to note that all 2176 minima which are connected to min.1 by transitions faster than 12 Hz satisfy the constraints $180° \leq \phi_i \leq 330°$. Conversely, most of the minima which satisfy these constraints are among the 2176 minima connected to min.1 by transitions faster than 12 Hz. This demonstrates unequivocally what we asserted earlier, that any ϕ rotation from within $[180°, 330°]$ to outside this interval was guarenteed to be slow. Apparently there is a very high barrier between these two regions.

Between 12 Hz and 10^{11} Hz, very little happens along the main branch. At 10^{10} Hz, the main branch still contains 2166 minima (just 10 minima broke away), and at 10^{11} Hz, the main branch still contains 2044 minima.

Beginning at around 10^{11} Hz, the graph becomes interesting again. min.1 and min.1583 remain connected until the rate cutoff reaches 2.24×10^{11} Hz, which is consistent with the transition times of the pathways listed in Table XXVI, and with Figure 34. The subgraph under the node which splits min.1 and min.1583 is shown in Figure 35. It contains 1112 minima. The rate disconnectivity graph ends at around 10^{13} Hz.

We followed the graph down to min.1 and min.1583 respectively, and then proceeded up the graph a few levels just to see which minima were most tightly connected to these minima. The subgraphs are shown in Figures 36 and 37, respectively. We can see that several of the minima in the vicinity of min.1583 are the same minima which appear along the shortest length pathways from min.1583 to min.1 (Table XXVI). min.1114, min.1410, min.1095, and min.633 are all strongly connected to min.1583. In contrast, the minima most strongly connected to min.1 do not appear to have anything to do with the pathways from min.1583. Only min.11 appears both in Table XXVI and Figure 36.

However, this asymmetry should not be a surprise. The minima which are closely connected to min.1583 are those minima which min.1583 can make fast transitions to. These transitions out of min.1583 inevitably lead to the ground state, so it makes sense that many of these minima will appear in a list of pathways to the ground state. On the other hand, minima which are closely connected to min.1 represent states which can make fast transitions to min.1. There is no reason to believe that many, or even any of these states making a transition to min.1 would have originally came from min.1583. Following the potential energy surface down from min.1583 will inevitably lead to min.1, but following the potential energy surface up from min.1 will not necessarily reach min.1583.

After completing our analysis of tetra-alanine using the minima and first-order saddles we obtained from the Eigenmode III search, we conducted αBB runs on select regions of the potential energy surface. A sampling of our results is given in Table XXVII. We began our search with $\alpha = 20$, and if necessary, increased α until we found all of the stationary points found by the Eigenmode III search. In all cases, modest values of α (i.e., less than 100) were sufficient to locate all minima and first-order saddles found by Eigenmode III. In many cases, we found additional first-order saddles.

To give a firmer test base, we repeated the Eigenmode III search on a larger grid (6^8 grid points), and searched for first-order saddles. We then generated the minima by following each first-order saddle back to the two minima it connects. Combining the results of the 4^8

grid and 6^8 grid gives us a grand total of 62372 minima and 212938 first-order saddles. The Eigenmode III results listed in Table XXVII include these additional stationary points.

A Eigenmode following algorithms

A very effective class of algorithms used for finding stationary points of a potential energy surface are the "eigenmode-following" algorithms. Essentially, they are sophisticated variations on the Newton-Raphson method applied to the equation $\nabla V = 0$.

The Newton-Raphson step is given by

$$\Delta \mathbf{x} = -H^{-1}\mathbf{g}$$

where $\mathbf{g} = \nabla V$ and $H = \nabla\nabla V$ are the gradient and Hessian matrix of the potential energy surface respectively. If we diagonalize the Hessian

$$H\mathbf{e}_i = b_i \mathbf{e}_i$$

and decompose the gradient vector into components along the "eigenmodes" $\mathbf{e}_i$

$$\mathbf{g} = \sum_i g_i \mathbf{e}_i$$

the Newton-Raphson step can be rewritten as

$$\Delta \mathbf{x} = -\sum_i \frac{g_i}{b_i}\mathbf{e}_i$$

The Newton-Raphson algorithm tends to locate stationary points which have the same signature (i.e., number of negative eigenvalues) as the Hessian matrix at the starting point. The reason for this is that the step $\Delta\mathbf{x}$ tends to increase V along the eigenmode directions corresponding to $b_i < 0$ (the step is along $+g_i\mathbf{e}_i$) and decrease V along eigenmode directions corresponding to $b_i > 0$ (the step is along $-g_i\mathbf{e}_i$). Thus, the Newton-Raphson method can never be used to walk away from a local minimum towards a first-order saddle point. It would always head back towards the local minimum.

In an eigenmode-following algorithm, the Newton-Raphson step is replaced by a slightly more complicated step:

$$\Delta \mathbf{x} = -\sum_i \frac{g_i}{b_i - \lambda_i}\mathbf{e}_i$$

The eigenvalues of the Hessian b_i are effectively shifted by an amount λ_i. This algorithm converges towards a stationary point whose signature matches the signs of $b_i - \lambda_i$ (i.e., if exactly one $b_i - \lambda_i$ is negative, then the algorithm will converge towards a first-order saddle point). This is because V tends to increase in directions corresponding to $b_i - \lambda_i < 0$ and decrease in directions corresponding to $b_i - \lambda_i > 0$. Eigenmode-following methods involve making a selection of the shift parameters λ_i which allows one to control the signature of the target stationary point, regardless of the signature of the Hessian at the starting point, by effectively shifting b_i to change its sign as required. Eigenmode-following algorithms can be used to walk away from local minima towards first-order transition states, etc.

A very effective eigenmode-following algorithm is Eigenmode III given in [1]. The shift parameters are given by

$$\lambda_i = \begin{cases} 0 & \text{actual } b_i > 0, \text{ desired } b_i > 0 \\ 2b_i & \text{actual } b_i < 0, \text{ desired } b_i > 0 \\ \frac{1}{2}(b_i + \sqrt{b_i^2 + 4g_i^2}) & \text{actual } b_i > 0, \text{ desired } b_i < 0 \\ 0 & \text{actual } b_i < 0, \text{ desired } b_i < 0 \end{cases}$$

We used Eigenmode III to generate a list of stationary states as a test of our own algorithm. We also used Eigenmode III to follow each first-order transition state back to the two minima it connects, so as to determine the reaction pathways.

B RRKM theory

In this Appendix, we discuss calculating reaction rates using RRKM theory in great detail, and prove (11), (12), and (14).

Let's begin with calculating the transition rate from one minimum j' to another minimum j through a particular transition state. According to RRKM theory [17], the rate in which the transition occurs for an isolated system of (constant) energy E is

$$\kappa(E) = \frac{1}{h}\frac{N_{\text{ts}}(E)}{\rho_{j'}(E)} \tag{16}$$

where $\rho_{j'}(E)$ is the density of states near the minimum j' and $N_{\text{ts}}(E) = \int_{-\infty}^{E} \rho_{\text{ts}}(\epsilon)\, d\epsilon$ is the total number of states near the transition state which are available to the system at energy E[17].

The transition rate associated with a system in contact with a thermal reservoir is related to the constant energy transition rate by integration over the probability distribution

$$\kappa(T) = \int_{-\infty}^{\infty} \kappa(E) \left(\frac{\rho_{j'}(E) e^{-E/kT}\, dE}{\int_{-\infty}^{+\infty} \rho_{j'}(\epsilon) e^{-\epsilon/kT}\, d\epsilon} \right) \tag{17}$$

[17]It pays to pay attention to the units to make sure they work out. $N_{\text{ts}}(E)$ is dimensionless, but $\rho_{j'}(E)$ has units of E^{-1}. This implies that the proposed formula for the rate constant has units of E/h, which are frequency units. These are appropriate units for a transition rate.

The denominator is the partition function $Q_{j'}$ of the system around the initial minimum j'. In the numerator, $\kappa(E)\rho_{j'}(E)$ can be replaced by $N_{\rm ts}(E)/h$ (equation (16)). The calculation then proceeds via integration by parts:

$$\begin{aligned}
\kappa(T) &= \frac{1}{h}\frac{1}{Q_{j'}}\int_{-\infty}^{\infty} N_{\rm ts}(E)e^{-E/kT}\,dE \\
&= \frac{1}{h}\frac{1}{Q_{j'}}\left(N_{\rm ts}(E)(-kTe^{-E/kT})|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\rho_{\rm ts}(E)(-kTe^{-E/kT})\,dE\right) \\
&= \frac{kT}{h}\frac{1}{Q_{j'}}\int_{-\infty}^{\infty}\rho_{\rm ts}(E)e^{-E/kT}\,dE \\
&= \frac{kT}{h}\frac{Q_{\rm ts}}{Q_{j'}}
\end{aligned}$$

which is (11)[18].

Equation (11) can be understood physically. The partition function Q can be expressed in the form

$$Q = e^{-F/kT} = e^{-(E-TS)/kT} = e^{S/k}e^{-E/kT} \tag{18}$$

which is the product of $e^{S/k}$, the number of quantum states available to the system, and $e^{-E/kT}$, the Bolzmann factor. Thus, (11) can be expressed in words

$$\kappa(T) = \frac{kT}{h}\left(\frac{(\text{number of quantum states near ts})e^{-E_{\rm ts}/kT}}{(\text{number of quantum states near } j')e^{-E_{j'}/kT}}\right) \tag{19}$$

The ratio in parenthesis is the fraction of time spent near the transition state in an ergodic system, or the probability that a random transition out of quantum state near j' will reach a quantum state in the vicinity of the transition state (at which point the transition from j' to j is assumed to occur). If we now interpret the factor kT/h as the average rate in which transitions between quantum states occur, we end up with the sensible interpretation that $\kappa(T)$, the rate in which the transition state is reached from minimum j', is equal to the product of the transition rate between quantum states and the probability that such a transition would lead to the transition state. In other words, we can think of the system as hopping randomly among states near the minimum j' until, by chance, one of those hops takes the system in the vicinity of the transition state, shortly after which the complete transition from j' to j will occur.

We need to evaluate the partition function Q around a given stationary point. We will treat the potential energy surface in the harmonic approximation, by truncating its Taylor

[18] Note that the boundary terms both vanish in the integration by parts: at $E \to \infty$ because of the exponential factor $e^{-E/kT}$, and at $E \to -\infty$ because $N_{\rm ts}(E)$ is certainly zero whenever E is less than the global minimum of the potential energy surface.

series expansion to second order:

$$V(\mathbf{x}) = V(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0) \cdot H(\mathbf{x} - \mathbf{x}_0) \tag{20}$$

where H is the Hessian matrix (recall that $\mathbf{g} = \nabla V = 0$ at a stationary point, so there will be no linear term).

Such a potential energy surface describes a system of n independent harmonic oscillators with vibrational frequencies $\nu_1, \ldots, \nu_n$. The stationary states are each characterized by a list of integers $(k_1, \ldots, k_n)$ $(k_i = 0, 1, \ldots)$, which are the "occupation numbers" of each normal mode. The energy of such states is given by

$$E_{(k_1,\ldots,k_n)} = E_0 + \sum_{i=1}^{n} h\nu_i(k_i + \frac{1}{2}) \tag{21}$$

where $E_0 = V(\mathbf{x}_0)$ is the value of the potential energy at the stationary point.

The partition function is calculated by summing up the Boltzman factor for each fixed energy state available to the system.

$$Q = \sum_{k_1,\ldots,k_n=0}^{\infty} \exp(-E_{(k_1,\ldots,k_n)}/kT) \tag{22}$$

For a harmonic oscillator system, this can be calculated exactly

$$\begin{aligned}
Q &= \sum_{k_1,\ldots,k_n=0}^{\infty} \exp(-E_0/kT + \sum_{i=1}^{n} h\nu_i(k_i + \frac{1}{2})/kT) \\
&= e^{-E_0/kT} \sum_{k_1,\ldots,k_n=0}^{\infty} \prod_{i=1}^{n} \exp(-h\nu_i(k_i + \frac{1}{2})/kT) \\
&= e^{-E_0/kT} \prod_{i=1}^{n} \sum_{k_i=0}^{\infty} \exp(-h\nu_i(k_i + \frac{1}{2})/kT) \\
&= e^{-E_0/kT} \prod_{i=1}^{n} e^{-h\nu_i/2kT} \sum_{k_i=0}^{\infty} \exp(-h\nu_i/kT)^{k_i} \\
&= e^{-E_0/kT} \prod_{i=1}^{n} \frac{e^{-h\nu_i/2kT}}{1 - e^{-h\nu_i/kT}} \\
&= e^{-E_0/kT} \prod_{i=1}^{n} \frac{1}{2\sinh(h\nu_i/2kT)}
\end{aligned}$$

If we take the classical limit $h\nu_i/kT \ll 1$, we obtain (12).

It remains for us to calculate the vibrational frequencies of a system around a given stationary point. It is fortuitous that quantum mechanics and classical mechanics give the same answer as far as calculating vibrational frequencies for a harmonic oscillator, so we will present a classical derivation of (14).

The configuration of the system is determined by a set of *generalized coordinates*, which are simply the variables $\mathbf{x} = (x_1, \ldots, x_n)$. The Lagrangian of this system is given by

$$L = K - V = \frac{1}{2}\dot{\mathbf{x}} \cdot I\dot{\mathbf{x}} - \frac{1}{2}\mathbf{x} \cdot H\mathbf{x} \tag{23}$$

where H is the Hessian of the potential energy surface, and I is the generalized inertia tensor (without loss of generality, we assume that $\mathbf{x}_0 = 0$ and $E_0 = 0$).

The motion of such a system is determined by the Euler-Lagrange equation

$$\frac{\partial L}{\partial x_i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} \tag{24}$$

which leads to the differential equation

$$I\ddot{\mathbf{x}} + H\mathbf{x} = 0 \tag{25}$$

The normal modes of the system are determined by requiring harmonic motion around the equilibrium configuration

$$\mathbf{x}(t) = e^{-i\omega t}\mathbf{x}^{(0)} \tag{26}$$

where $\mathbf{x}^{(0)}$ is the normal mode and $\nu = \omega/2\pi$ is the frequency associated with that mode. If we substitute (26) into (25), we obtain

$$(H - \omega^2 I)\,\mathbf{x}^{(0)} = 0 \tag{27}$$

which is equivalent to (14). The vibrational frequencies are calculated by solving the generalized eigenvalue problem (27) (or (14)).

Note that the frequency calculation requires calculating the inertia tensor I. This is accomplished by calculating the kinetic energy of the system in terms of the generalized coordinates $\mathbf{x}$, and then comparing the result to (23).

For this, we need the cartesian coordinates of all of the atoms as a function of the generalized coordinates $\mathbf{r} = \mathbf{r}(\mathbf{x})$, where $\mathbf{r} = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots)$ is the complete list of cartesian coordinates. The velocities can be written in terms of $\dot{\mathbf{x}}$ by the chain rule

$$\frac{dr_j}{dt} = \sum_{i=1}^{n} \frac{\partial r_j}{\partial x_i}\frac{dx_i}{dt} = \sum_{i=1}^{n} \mathcal{X}_{ji}\frac{dx_i}{dt}, \qquad j = 1, \ldots, 3N_a \tag{28}$$

where $\mathcal{X}_{ji} = \partial r_j/\partial x_i$ is the $3N_a \times n$ rectangular matrix of derivatives of cartesian coordinates with respect to the generalized coordinates. The kinetic energy is

$$\begin{aligned} K &= \frac{1}{2}\sum_{j=1}^{3N_a} m_j \dot{r}_j^2 \\ &= \frac{1}{2}\sum_{j=1}^{3N_a} m_j (\sum_{i=1}^{n} \mathcal{X}_{ji}\dot{x}_i)(\sum_{i'=1}^{n} \mathcal{X}_{ji'}\dot{x}_{i'}) \\ &= \frac{1}{2}\sum_{i,i'=1}^{n} \dot{x}_i (\sum_{j=1}^{3N_a} m_j \mathcal{X}_{ji}\mathcal{X}_{ji'})\dot{x}_{i'} \end{aligned}$$

where $m_1 = m_2 = m_3 =$ the mass of the first atom, $m_4 = m_5 = m_6 =$ the mass of the second atom, etc. Comparing this result to (23), we find

$$I_{ii'} = \sum_{j=1}^{3N_a} m_j \mathcal{X}_{ji} \mathcal{X}_{ji'} = \sum_{j=1}^{3N_a} m_j \frac{\partial r_j}{\partial x_i} \frac{\partial r_j}{\partial x_{i'}} \tag{29}$$

The ECEPP/3 modeling system [26] provides us with the coordinate functions $\mathbf{r}(\mathbf{x})$. We calculate the rectangular matrix $\mathcal{X}_{ji}$ first by finite differencing, and then apply (29) to obtain I.

One must be very careful with the cartesian coordinate functions $\mathbf{r}(\mathbf{x})$ provided by ECEPP/3 since these coordinates are usually assigned in a manner inconsistent with momentum and angular momentum conservation. A molecule can undergo overall translations and rotations without affecting its structure. This introduces six additional degrees of freedom which must be fixed in order to evaluate the cartesian coordinates $\mathbf{r}(\mathbf{x})$. Usually this is accomplished by fixing one atom at the origin, another atom along a fixed axis, and a third atom within a fixed plane. Unfortunately, these constraints introduce external forces and torques on the molecule which do not really exist, and these extra forces and torques have an effect on the vibrational frequencies.

One solution is to enforce conservation of momentum and angular momentum on the system by imposing more sensible restrictions on the cartesian coordinates. For example, if the three translational degrees of freedom are eliminated by fixing the center of mass of the system to the origin, rather than fixing the position of one particular atom, then momentum will be conserved along any possible trajectory $\mathbf{r}(\mathbf{x}(t))$. This can be accomplished by a simple adjustment of the coordinate functions received from ECEPP/3. However, figuring out the conditions which impose *angular* momentum conservation is much trickier.

A more elegant solution (and one which we adopt) is to augment the generalized coordinates $\mathbf{x}$ with 6 additional coordinates, three translation variables (a_x, a_y, a_z) and three rotation angles $(\theta_x, \theta_y, \theta_z)$. Cartesian coordinates can then be calculated in terms of the extended variable set as follows

$$\mathbf{r}(\mathbf{x}, a_i, \theta_i) = \mathcal{R}_x(\theta_x)\mathcal{R}_y(\theta_y)\mathcal{R}_z(\theta_z)(\mathbf{r}(\mathbf{x}) + \mathbf{a}) \tag{30}$$

The effect of introducing the additional coordinates is that the artificial constraint forces imposed on the system are removed since the system is now free to translate and rotate. The six additional columns and rows added to H will all be zero because the potential V does not depend on these coordinates. However, the six additional columns and rows added to I

will *not* be zero because pure translation and rotation does contribute to the kinetic energy of the system.

If we solve (27) for the vibrational frequencies, we will discover 6 zero frequency modes, which correspond to overall translation and rotation (only the last six components of the normal mode will be non-zero). The remaining normal modes will in general have non-zero frequencies, and involve motions within the molecule (i.e., changes in $\mathbf{x}$), as well as some overall translation and rotation component (whatever is necessary to insure overall momentum and angular momentum conservation). The vibrational frequencies computed this way will be physically correct, and appropriate for calculating the partition function Q using (12). By augmenting the variable set in this manner, we no longer need to worry about the cartesian coordinate functions $\mathbf{r}(\mathbf{x})$ that we receive from ECEPP/3.

Acknowledgement

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

References

[1] C. J. Tsai and K. D. Jordan, *J. Phys. Chem.* **97** (1993), 11227.

[2] C. J. Cerjan and W. H. Miller, *J. Chem. Phys.* **75**(6) (1981), 2800.

[3] J. Simons, P. Jorgensen, H. Taylor, J. Ozment, *J. Phys. Chem.* **87** (1983), 2745.

[4] A. Banerjee, N. Adams, J. Simons, R. Shepard, *J. Phys. Chem.* **89** (1985), 52.

[5] D. O'Neal, H. Taylor, J. Simons, *J. Phys. Chem.* **88** (1984), 1510.

[6] P. Culot, G. Dive, V. H. Nguyen, J. M. Ghuysen, *Theor. Chim. Acta* **82** (1992), 189.

[7] R. S. Berry, *Chem. Rev.* **93** (1993), 2379.

[8] F. H. Stillinger, T. A. Weber, *J. Chem. Phys.* **80** (1984), 4434.

[9] R. S. Berry, H. L. Davis, T. L. Beck, *Chem. Phys. Lett.* **147** (1988), 13.

[10] R. J. Wawak, J. Pillardy, A. Liwo, K. D. Gibson, H. A. Scheraga, *J. Phys. Chem.* A **102**(17) (1998), 2904.

[11] K. A. Dill, A. T. Phillips, J. B. Rosen, in *IMA Volumes in Mathematics and its Applications*, Vol. **94**, Springer-Verlag (1997), 1.

[12] C. D. Maranas and C. A. Floudas, *J. Global Opt.* **7** (1995), 143.

[13] I. P. Andoulakis, C. D. Maranas, C. A. Floudas, *J. Global Opt.* **7**(4) (1995), 337.

[14] C. Adjiman and C. A. Floudas, *J. Global Opt.* **9** (1996), 23.

[15] C. Adjiman, S. Dallwig, C. A. Floudas, A. Neumaier, *Computers and Chem. Eng.* **22**(9) (1998), 1137.

[16] C. Adjiman, I. P. Androulakis, C. A. Floudas, *Computers and Chem. Eng.* **22**(9) (1998), 1159.

[17] M. F. Jarrold, in *Clusters of Atoms and Molecules.*, ed. H. Haberland (Springer, Berlin, 1994), Chap. 2.7, 163.

[18] R. E. Kunz and R. S. Berry, *J. Chem. Phys.* **103**(5) (1995), 1904.

[19] O. M. Becker, M. Karplus, *J. Chem. Phys.* **106**(4) (1997), 1495.

[20] J. N. Murrell, S. Carter, A. J. C. Varandas, *Molec. Phys.* **35**(5) (1978), 1325.

[21] A. Aguilar, M. González, F. Illas, J. Rubio, R. Sayós, *Chem. Phys.* **161** (1992), 99.

[22] M. Alberti, R. Sayós, A. Solé, A. Aguilar, *J. Chem. Soc. Faraday Trans.* **87**(8) (1991), 1057.

[23] X. Grande, M. Alberti, X. Giménez, J. M. Lucas, A. Aguilar, *J. Chem. Soc. Faraday Trans.* **89**(10) (1993), 1587.

[24] K. S. Sorbie and J. N. Murrell, *Molec. Phys.* **29** (1975), 1387.

[25] M. Alberti, A. Solé, A. Aguilar, *J. Chem. Soc. Faraday Trans.* **87**(1) (1991), 37.

[26] G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, H. A. Scheraga, *J. Phys. Chem.* **96** (1992), 6472.

[27] C. L. Brooks III and D. A. Case, *Chem. Rev.* **93** (1993), 2487.

[28] D. J. Tobias and C. L. Brooks III, *J. Phys. Chem.* **96** (1992), 3864.

[29] R. Kaschner and D. Hohl, *J. Phys. Chem.* **A102** (1998), 5111.

[30] I. K. Roterman, M. H. Lambert, K. D. Gibson, H. A. Scheraga, *J. Biomol. Struct. Dyn.* **7** (1989), 421–453.

[31] S. S. Zimmerman, M. S. Pottle, G. Némethy, H. A. Scheraga, *Macromolecules* **10** (1977), 1–9.

[32] F. Herrmann and S. Suhai, *J. Comp. Chem.* **16**(11) (1995), 1434.

[33] E. Neria, S. Fischer, M. Karplus, *J. Chem. Phys.* **105**(5) (1996), 1902.

[34] B. M. Pettitt and M. Karplus, *Chem. Phys. Lett.* **121**(3) (1985), 194.

[35] R. Czerminski and R. Elber, *J. Chem. Phys.* **92**(9) (1990), 5580.

[36] S. Bell and J. Crighton, *J. Chem. Phys.* **80**(6) (1984), 2464.

[37] P. Jorgensen, H. J. A. Jensen, T. Helgaker, *Theor. Chim. Acta* **73** (1988), 55.

[38] L. R. Pratt, *J. Chem. Phys.* **85**(9) (1986), 5045.

[39] R. Elber and M. Karplus, *Chem. Phys. Lett.* **139** (1987), 375.

Table I: Potential energy surface for HCN [20]. Distances are in Å and energies are in eV.

$$
\begin{aligned}
V &= V_{\mathrm{CH}} + V_{\mathrm{CN}} + V_{\mathrm{NH}} + V_{\mathrm{I}} \\
V_{\mathrm{CH}} &= -2.977\,(1.0 + 2.130r - 3.381r^2 + 4.694r^3)\,e^{-2.130r} \\
r &= R_{\mathrm{CH}} - 1.0784 \\
V_{\mathrm{CN}} &= -7.929\,(1.0 + 5.244r + 7.339r^2 + 4.976r^3)\,e^{-5.244r} \\
r &= R_{\mathrm{CN}} - 1.1718 \\
V_{\mathrm{NH}} &= -3.994\,(1.0 + 3.070r)\,e^{-3.070r} \\
r &= R_{\mathrm{NH}} - 1.037 \\
V_{\mathrm{I}} &= 4.582\,(1.0 - 0.0246\rho_1 + 6.2973\rho_2 - 0.0737\rho_3 - 0.1741\rho_1\rho_2 + 0.3521\rho_1\rho_3 \\
&\quad + 0.1559\rho_2\rho_3 - 0.0804\rho_1^2 + 18.2174\rho_2^2 - 0.6147\rho_3^2 + 0.2020\rho_1^3 + 19.1358\rho_2^3 \\
&\quad - 0.3955\rho_3^3 - 0.3026\rho_1\rho_2\rho_3) \times \prod_{i=1}^{3}(1 - \tanh\gamma_i\rho_i/2) \\
\rho_1 &= R_{\mathrm{CH}} - 1.479,\;\; \rho_2 = R_{\mathrm{CN}} - 1.479,\;\; \rho_3 = R_{\mathrm{NH}} - 1.479 \\
\gamma_1 &= 1.617,\;\; \gamma_2 = 5.729,\;\; \gamma_3 = 1.935
\end{aligned}
$$

Table II: Potential energy surface for HSiN [20]. Distances are in Å and energies are in eV.

$$
\begin{aligned}
V &= V_{\mathrm{SiH}} + V_{\mathrm{SiN}} + V_{\mathrm{NH}} + V_{\mathrm{I}} \\
V_{\mathrm{SiH}} &= -2.122\,(1.0 + 2.929r)\,e^{-2.929r} \\
r &= R_{\mathrm{SiH}} - 1.5224 \\
V_{\mathrm{SiN}} &= -5.708\,(1.0 + 3.642r + 2.646r^2 + 1.241r^3)\,e^{-3.642r} \\
r &= R_{\mathrm{SiN}} - 1.5718 \\
V_{\mathrm{NH}} &= -3.994\,(1.0 + 3.070r)\,e^{-3.070r} \\
r &= R_{\mathrm{NH}} - 1.037 \\
V_{\mathrm{I}} &= 5.018\,(1.0 + 0.6408\rho_1 + 0.0915\rho_2 + 0.2329\rho_3 + 0.5917\rho_1\rho_2 + 0.7692\rho_1\rho_3 \\
&\quad + 0.1314\rho_2\rho_3 + 0.2698\rho_1^2 - 1.0421\rho_2^2 - 0.1567\rho_3^2) \times \prod_{i=1}^{3}(1 - \tanh\gamma_i\rho_i/2) \\
\rho_1 &= R_{\mathrm{SiH}} - 1.993,\;\; \rho_2 = R_{\mathrm{SiN}} - 1.993,\;\; \rho_3 = R_{\mathrm{NH}} - 1.993 \\
\gamma_1 &= 2.4,\;\; \gamma_2 = 0.4,\;\; \gamma_3 = 0.2
\end{aligned}
$$

Table III: Potential energy surface for CS_2 [21]. Distances are in Å and energies are in eV.

$$\begin{aligned}
V &= V_{\mathrm{CS}} + V_{\mathrm{CS'}} + V_{\mathrm{SS'}} + V_{\mathrm{I}} \\
V_{\mathrm{CS}} &= -7.434\,(1.0 + 3.445r + 2.371r^2 + 1.239r^3)\,e^{-3.445r} \\
r &= R_{\mathrm{CS}} - 1.5349 \\
V_{\mathrm{CS'}} &= -7.434\,(1.0 + 3.445r + 2.371r^2 + 1.239r^3)\,e^{-3.445r} \\
r &= R_{\mathrm{CS'}} - 1.5349 \\
V_{\mathrm{SS'}} &= -3.954\,(1.0 + 3.954r + 4.312r^2 + 2.332r^3)\,e^{-3.954r} \\
r &= R_{\mathrm{SS'}} - 1.8892 \\
V_{\mathrm{I}} &= 3.438\,(1.0 + 1.874\rho_1 + 1.874\rho_2 - 0.950\rho_3 + 9.327\rho_1^2 + 9.327\rho_2^2 + 1.833\rho_3^2 \\
&\quad + 19.613\rho_1\rho_2 - 9.093\rho_1\rho_3 - 9.093\rho_2\rho_3) \times (1 - \tanh\gamma(\rho_1 + \rho_2 + \rho_3)) \\
\rho_1 &= R_{\mathrm{CS}} - 1.5525,\;\; \rho_2 = R_{\mathrm{CS'}} - 1.5525,\;\; \rho_3 = R_{\mathrm{SS'}} - 3.1050 \\
\gamma &= 1.6
\end{aligned}$$

Table IV: Potential energy surface for HBO (PES1) [22]. Distances are in Å and energies are in eV.

$$\begin{aligned}
V &= V_{\mathrm{BH}} + V_{\mathrm{BO}} + V_{\mathrm{OH}} + V_{\mathrm{I}} \\
V_{\mathrm{BH}} &= -2.3939\,(1.0 + 3.3508\rho_1)\,e^{-3.3508\rho_1} \\
V_{\mathrm{BO}} &= -7.7486\,(1.0 + 3.6800\rho_2)\,e^{-3.6800\rho_2} \\
V_{\mathrm{OH}} &= -4.1470\,(1.0 + 3.7197\rho_3)\,e^{-3.7197\rho_3} \\
V_{I} &= -2.3888\,(1.0 - 1.7893\rho_1 - 0.0640\rho_2 + 3.2157\rho_3 + 0.2661\rho_1^2 - 2.6342\rho_1\rho_2 \\
&\quad - 2.2514\rho_1\rho_3 + 0.4440\rho_2^2 + 3.5274\rho_2\rho_3 + 3.4489\rho_3^2 - 4.3204\rho_1^3 \\
&\quad + 2.0899\rho_3^3 - 16.1359\rho_1\rho_2\rho_3) \times \prod_{i=1}^{3}(1 - \tanh\gamma_i\rho_i/2) \\
\rho_1 &= R_{\mathrm{BH}} - 1.16485,\;\; \rho_2 = R_{\mathrm{BO}} - 1.18542,\;\; \rho_3 = R_{\mathrm{OH}} - 2.35027 \\
\gamma_1 &= 0.425,\;\; \gamma_2 = 3.925,\;\; \gamma_3 = 2.40
\end{aligned}$$

Table V: Potential energy surface for HBO (PES2) [23]. Distances are in Å and energies are in eV.

$$
\begin{aligned}
V &= V_{\mathrm{BH}} + V_{\mathrm{BO}} + V_{\mathrm{OH}} + V_{\mathrm{I}} \\
V_{\mathrm{BH}} &= -2.3939\,(1.0 + 3.3508\rho_1)\,e^{-3.3508\rho_1} \\
V_{\mathrm{BO}} &= -7.7486\,(1.0 + 3.6800\rho_2)\,e^{-3.6800\rho_2} \\
V_{\mathrm{OH}} &= -4.1470\,(1.0 + 3.7197\rho_3)\,e^{-3.7197\rho_3} \\
V_I &= -2.3888\,(1.0 - 1.1485\rho_1 - 0.3768\rho_2 + 3.8907\rho_3 - 0.5429\rho_1^2 - 2.0869\rho_1\rho_2 \\
&\quad - 1.1469\rho_1\rho_3 + 0.0339\rho_2^2 + 2.0365\rho_2\rho_3 + 6.8852\rho_3^2 + 0.0027\rho_1^3 \\
&\quad + 2.6051\rho_2^3 + 7.8303\rho_3^3) \times \prod_{i=1}^{3}(1 - \tanh\gamma_i\rho_i/2) \\
\rho_1 &= R_{\mathrm{BH}} - 1.16485,\ \rho_2 = R_{\mathrm{BO}} - 1.18542,\ \rho_3 = R_{\mathrm{OH}} - 2.35027 \\
\gamma_1 &= 1.50,\ \gamma_2 = 3.25,\ \gamma_3 = 3.75
\end{aligned}
$$

Table VI: Stationary states for HCN

Name	Energy (eV)	R_{CN} (Å)	R_{CH} (Å)	R_{NH} (Å)
min.1	−13.592216	1.153199	1.065499	+++++
min.2	−12.972507	1.159151	+++++	0.993337
1st.1	−11.444169	1.117973	1.053920	1.387751
min.3	−11.379411	0.857321	0.980846	0.989053
1st.2	−11.345399	0.929066	1.039138	1.041348
min.4	−5.548224	+++++	2.332871	1.038901
1st.3	−5.249953	2.344235	2.980408	1.044279
1st.4	−3.102484	2.582865	1.081559	2.737336
2nd.1	−1.937592	2.311896	1.792855	2.327696
2nd.2	8.094669	+++++	0.857573	0.806900

Table VII: Stationary states for HSiN

Name	Energy (eV)	R_{SiN} (Å)	R_{SiH} (Å)	R_{NH} (Å)
min.1	−9.358509	1.523293	+++++	0.998205
min.2	−6.098599	1.529589	1.459587	+++++
1st.1	−5.666608	1.575921	+++++	2.969229
1st.2	−5.148745	2.006322	1.361586	+++++
1st.3	−4.908996	2.394222	2.137985	0.974497
1st.4	−3.717495	1.461352	1.634576	2.093709
1st.5	−3.144739	1.523964	2.426269	+++++
2nd.1	−2.876954	1.501908	2.309780	3.069649
2nd.2	−0.728138	2.155742	1.473093	2.044809
2nd.3	1.109602	2.778075	2.617596	+++++
max.1	1.720516	2.647092	2.415996	3.498876

Table VIII: Stationary states for CS_2

Name	Energy (eV)	$R_{\mathrm{SS'}}$ (Å)	R_{CS} (Å)	$R_{\mathrm{CS'}}$ (Å)
min.1	−12.004549	+++++	1.552423	1.552423
1st.1	−1.668827	2.695109	2.761780	+++++
2nd.1	−0.049003	3.978689	4.171035	4.171035
1st.2	97.485407	1.417729	0.909824	+++++
2nd.2	103.740893	1.813412	0.949957	+++++

Table IX: Stationary states for HBO (PES1)

Name	Energy (eV)	R_{BO} (Å)	R_{BH} (Å)	R_{OH} (Å)
min.1	−16.678316	1.185029	1.165505	+++++
1st.1	−7.598281	1.187663	3.264083	+++++
1st.2	−6.556670	+++++	1.162756	2.349431
2nd.1	−0.216647	2.554093	+++++	3.688902

Table X: Stationary states for HBO (PES2)

Name	Energy (eV)	R_{BO} (Å)	R_{BH} (Å)	R_{OH} (Å)
min.1	−16.678851	1.184168	1.168947	+++++
min.2	−11.305022	1.192048	+++++	2.383483
1st.1	−11.134196	1.185979	2.906576	2.398381
1st.2	−6.639250	+++++	1.154136	2.344208

Table XI: Stationary points of alanine.

	$\alpha_i^{\pm} = 5$	$\alpha_i^{\pm} = 10$	$\alpha_i^{\pm} = 15$	$\alpha_i^{\pm} = 20$
Iterations	17306	30137	43735	58046
CPU time (sec)	576	1152	1854	2656
Local minima	17	17	17	17
1st-order saddles	62	62	62	62
2nd-order saddles	71	82	83	83
3rd-order saddles	30	45	46	46
Local maxima	3	5	7	7
Total solutions	183	211	215	215

Table XII: 17 local minima of alanine.

Minimum	E (kcal/mole)	ϕ (deg)	ψ (deg)	ω (deg)	χ (deg)
min.01	2.677051	-80.917	-33.304	179.667	180.340
min.02	2.861981	-73.536	139.081	180.259	180.486
min.03	2.892502	60.085	-46.408	179.323	179.469
min.04	3.009668	187.686	-54.033	179.430	178.797
min.05	3.015236	61.899	123.574	180.549	179.606
min.06	3.142731	186.523	104.847	180.512	178.161
min.07	3.220681	-57.605	22.982	180.258	183.134
min.08	3.309932	214.440	171.296	179.980	180.905
min.09	6.236178	65.377	66.425	-3.223	179.919
min.10	6.764486	-79.789	72.511	-2.937	179.312
min.11	6.940900	184.014	75.412	-2.343	178.160
min.12	7.014898	-78.138	159.567	3.224	181.398
min.13	7.084476	185.147	-55.619	5.577	174.675
min.14	7.387601	56.137	157.767	5.131	179.913
min.15	7.408455	207.597	162.271	-0.217	181.208
min.16	8.772372	-26.959	-44.694	-19.985	179.562
min.17	8.850583	-27.168	-62.739	16.715	178.631

Table XIII: 62 first-order transition states of alanine.

Saddle	E (kcal/mole)	ϕ (deg)	ψ (deg)	ω (deg)	χ (deg)
1st.01	3.156555	205.936	-22.612	179.863	180.122
1st.02	3.187229	188.969	67.250	180.668	179.061
1st.03	3.220681	-57.695	22.538	180.253	183.117
1st.04	3.245969	-62.257	54.156	180.482	181.574
1st.05	3.311030	222.458	170.910	179.987	180.934
1st.06	3.311484	207.652	165.482	180.025	180.535
1st.07	3.614353	-1.518	-5.620	179.844	181.279
1st.08	3.640605	64.066	22.725	180.210	182.231
1st.09	3.729918	193.090	233.646	179.493	181.187
1st.10	3.805265	-63.292	233.512	179.665	182.014
1st.11	3.846946	124.089	-55.131	179.253	179.600
1st.12	3.953160	124.257	108.300	180.634	179.282
1st.13	4.013967	-1.651	161.635	180.194	181.140
1st.14	4.173786	66.577	224.589	179.698	183.051
1st.15	6.223994	-82.041	-32.479	179.669	120.402
1st.16	6.417347	-75.500	141.615	180.246	120.526
1st.17	6.610153	193.207	-48.342	179.488	119.217
1st.18	6.617442	62.318	-45.515	179.309	119.703
1st.19	6.756069	62.652	129.574	180.508	119.756
1st.20	6.787633	193.607	128.744	180.328	119.295
1st.21	7.078400	229.659	69.796	-0.564	179.378
1st.22	7.159186	129.534	67.075	-3.181	179.030
1st.23	7.411289	218.209	162.965	-1.113	181.403
1st.24	7.675726	-4.949	68.951	-0.924	180.595
1st.25	8.064792	-0.726	162.114	-0.245	181.568
1st.26	8.501975	125.121	158.201	4.795	180.456
1st.27	8.872613	-24.942	-56.299	3.397	178.450
1st.28	9.128055	-71.909	135.390	-35.050	180.453
1st.29	9.279329	58.866	137.308	-34.323	179.525
1st.30	9.320592	-72.315	-47.805	-34.950	170.150
1st.31	9.354995	-66.836	100.753	33.914	179.258
1st.32	9.431146	63.312	101.295	33.817	178.848
1st.33	9.540317	191.382	136.590	-34.503	179.472
1st.34	9.542205	187.232	102.092	34.567	178.075
1st.35	9.640362	-78.449	-79.254	30.564	167.858
1st.36	9.818227	24.174	-47.920	-30.903	170.220
1st.37	10.171651	191.949	-54.376	-1.771	116.061
1st.38	10.186598	26.140	-75.566	25.336	167.903
1st.39	10.301416	-80.260	163.588	3.294	121.411
1st.40	10.880595	60.711	162.073	4.561	120.410
1st.41	11.009258	183.917	-61.049	87.367	178.203
1st.42	11.086837	72.367	-50.273	82.793	180.875
1st.43	11.124220	-86.290	-28.501	-86.074	179.967
1st.44	11.181241	69.208	-23.889	71.028	181.004
1st.45	11.210376	64.456	98.459	-87.746	178.877
1st.46	11.213608	213.229	-3.257	-86.087	180.787
1st.47	11.268395	-68.548	113.054	-87.000	179.801
1st.48	11.309974	-81.980	-5.516	81.495	180.976
1st.49	11.310009	-53.860	-60.472	80.875	180.436
1st.50	11.389593	182.369	90.523	-86.697	177.576
1st.51	11.484439	57.483	-33.757	-86.483	178.862
1st.52	11.494613	213.154	24.176	85.592	182.301
1st.53	11.509215	-72.840	136.179	86.403	180.607
1st.54	11.602675	-69.368	92.193	-20.946	117.112
1st.55	11.606165	65.102	88.618	-19.685	116.863
1st.56	11.712331	61.326	124.430	86.501	179.598
1st.57	11.732850	159.961	-11.986	66.565	177.907
1st.58	11.793153	192.980	91.439	-20.137	116.206
1st.59	11.800318	79.959	19.480	-68.216	183.655
1st.60	11.839776	187.902	113.001	86.308	178.601
1st.61	11.982450	79.570	-53.377	-14.238	118.667
1st.62	12.098725	-45.796	-51.352	-20.126	116.985

Table XIV: 10 minima of alanine dipeptide.

Minimum	E (kcal/mole)	θ_1 (deg)	θ_2 (deg)	ϕ (deg)	ψ (deg)	ω (deg)	χ (deg)	θ_3 (deg)
min.01	-5.180369	179.567	181.346	279.636	75.818	178.849	180.715	180.318
min.02	-4.469561	180.167	179.776	205.253	157.194	180.098	179.237	179.953
min.03	-4.373772	179.869	180.335	286.362	325.115	180.062	181.608	179.912
min.04	-4.076014	180.373	179.396	209.269	45.565	180.044	181.042	179.976
min.05	-4.058307	179.916	180.199	284.458	139.059	180.608	181.916	179.859
min.06	-3.445010	180.006	180.023	201.722	302.536	179.063	174.413	180.315
min.07	-2.815686	180.287	179.342	54.669	46.016	180.278	186.879	180.046
min.08	-0.112282	180.369	179.137	63.665	185.251	177.831	200.847	180.851
min.09	2.065108	180.857	177.019	77.757	295.756	179.986	207.508	180.088
min.10	2.188548	179.267	182.567	57.636	271.806	188.766	163.886	177.140

Table XV: 38 first-order saddles of alanine dipeptide.

Saddle	E (kcal/mole)	θ_1 (deg)	θ_2 (deg)	ϕ (deg)	ψ (deg)	ω (deg)	χ (deg)	θ_3 (deg)
1st.01	-4.331961	119.502	181.467	279.585	75.788	178.850	180.716	180.318
1st.02	-4.052569	179.918	180.229	285.078	130.222	180.451	181.710	179.949
1st.03	-3.978339	179.568	181.340	279.683	75.929	178.692	180.709	120.473
1st.04	-3.635211	180.165	179.524	209.052	108.130	179.557	178.142	180.161
1st.05	-3.623680	120.123	179.752	205.241	157.200	180.098	179.237	179.953
1st.06	-3.529282	179.916	180.288	237.932	60.216	179.784	180.756	180.078
1st.07	-3.518100	119.858	180.399	286.376	325.062	180.062	181.604	179.909
1st.08	-3.305410	180.005	180.475	245.192	149.006	180.367	181.303	179.867
1st.09	-3.265688	180.167	179.776	205.252	157.201	180.112	179.237	119.953
1st.10	-3.224577	120.290	179.356	209.223	45.585	180.046	181.037	179.976
1st.11	-3.205613	119.913	180.236	284.477	138.899	180.608	181.911	179.860
1st.12	-3.166873	179.865	180.329	286.317	325.163	180.082	181.612	119.943
1st.13	-3.114901	179.516	181.724	255.119	10.378	179.882	182.899	180.017
1st.14	-2.872860	180.374	179.396	209.269	45.571	180.050	181.041	119.978
1st.15	-2.851601	179.917	180.198	284.398	138.885	180.681	181.916	119.768
1st.16	-2.597710	119.997	180.024	201.715	302.546	179.060	174.418	180.316
1st.17	-2.340470	180.274	179.367	239.218	302.774	179.124	177.239	180.294
1st.18	-2.241695	180.007	180.023	201.725	302.591	178.923	174.420	120.414
1st.19	-1.961786	120.279	179.264	54.689	46.027	180.280	186.870	180.048
1st.20	-1.602898	180.291	179.345	54.696	45.950	180.270	186.903	119.956
1st.21	-1.000853	180.261	179.486	203.337	159.698	180.151	119.207	179.930
1st.22	0.013990	180.178	179.482	200.855	303.145	178.457	115.139	180.527
1st.23	0.476551	179.419	181.975	265.114	103.799	179.087	121.641	180.321
1st.24	0.518612	179.717	181.236	272.179	326.978	180.130	124.351	179.934
1st.25	0.739801	120.358	179.051	63.691	185.232	177.833	200.831	180.851
1st.26	0.787970	180.671	178.632	220.681	27.648	181.910	124.575	179.338
1st.27	1.090129	180.369	179.137	63.647	185.305	177.524	200.811	120.999
1st.28	2.011870	179.757	180.993	69.743	127.642	180.397	203.690	179.857
1st.29	2.235108	180.105	180.081	57.013	34.932	181.155	129.311	179.838
1st.30	2.910394	121.122	176.561	77.963	295.819	179.979	207.449	180.091
1st.31	3.036065	119.046	182.928	57.489	271.826	188.750	163.861	177.144
1st.32	3.270683	180.857	177.026	77.751	295.751	179.981	207.509	120.043
1st.33	3.693830	180.235	179.152	67.815	283.213	187.576	186.788	177.622
1st.34	4.043629	179.872	180.900	60.407	178.274	177.748	139.602	180.896
1st.35	4.557812	178.562	185.042	65.673	283.817	182.556	124.496	179.286
1st.36	5.990539	180.290	179.409	205.195	237.066	182.285	181.648	179.244
1st.37	6.272659	179.918	180.448	53.354	240.740	174.271	181.764	181.541
1st.38	6.705320	179.952	179.853	271.009	238.764	178.227	181.828	180.618

Table XVI: Triples for alanine.

	Lower minimum		Higher minimum	
Saddle	Minimum	Barrier (kcal/mole)	Minimum	Barrier (kcal/mole)
1st.01	min.01	0.479505	min.04	0.146888
1st.02	min.01	0.510179	min.06	0.044499
1st.03	min.01	0.543631	min.07	0.000001
1st.04	min.02	0.383988	min.07	0.025288
1st.05	min.02	0.449050	min.08	0.001099
1st.06	min.06	0.168753	min.08	0.001552
1st.07	min.01	0.937302	min.03	0.721851
1st.08	min.03	0.748103	min.05	0.625369
1st.09	min.04	0.720251	min.08	0.419987
1st.10	min.01	1.128215	min.02	0.943285
1st.11	min.03	0.954444	min.04	0.837279
1st.12	min.05	0.937925	min.06	0.810430
1st.13	min.02	1.151987	min.05	0.998731
1st.14	min.03	1.281284	min.05	1.158550
1st.15	min.01	3.546959	min.01	3.546944
1st.16	min.02	3.555384	min.02	3.555366
1st.17	min.04	3.600515	min.04	3.600486
1st.18	min.03	3.724991	min.03	3.724940
1st.19	min.05	3.740885	min.05	3.740834
1st.20	min.06	3.644938	min.06	3.644903
1st.21	min.10	0.313914	min.11	0.137500
1st.22	min.09	0.923009	min.11	0.218287
1st.23	min.12	0.396391	min.15	0.002835
1st.24	min.09	1.439549	min.10	0.911240
1st.25	min.12	1.049894	min.14	0.677192
1st.26	min.14	1.114375	min.15	1.093521
1st.27	min.16	0.100242	min.17	0.022030
1st.28	min.10	2.363569	min.12	2.113157
1st.29	min.09	3.043152	min.14	1.891729
1st.30	min.13	2.236116	min.16	0.548221
1st.31	min.10	2.590510	min.12	2.340097
1st.32	min.09	3.194969	min.14	2.043546
1st.33	min.11	2.599418	min.15	2.131863
1st.34	min.11	2.601306	min.15	2.133751
1st.35	min.13	2.555886	min.17	0.789779
1st.36	min.13	2.733751	min.16	1.045856
1st.37	min.13	3.087355	min.13	3.087175
1st.38	min.13	3.102122	min.17	1.336015
1st.39	min.12	3.286518	min.12	3.286445
1st.40	min.14	3.492995	min.14	3.492952
1st.41	min.04	7.999591	min.13	3.924782
1st.42	min.03	8.194336	min.13	4.002361
1st.43	min.01	8.447169	min.13	4.039744
1st.44	min.09	4.945064	min.13	4.096765
1st.45	min.05	8.195140	min.09	4.974199
1st.46	min.01	8.536558	min.13	4.129132
1st.47	min.02	8.406415	min.10	4.503910
1st.48	min.01	8.632923	min.10	4.545489
1st.49	min.01	8.632959	min.17	2.459426
1st.50	min.01	8.712542	min.11	4.448693
1st.51	min.03	8.591937	min.13	4.399963
1st.52	min.01	8.817562	min.11	4.553713
1st.53	min.02	8.647235	min.12	4.494317
1st.54	min.10	4.838233	min.10	4.838190
1st.55	min.09	5.370058	min.09	5.369988
1st.56	min.05	8.697095	min.14	4.324731
1st.57	min.11	4.791950	min.13	4.648374
1st.58	min.11	4.852294	min.11	4.852254
1st.59	min.09	5.564141	min.13	4.715843
1st.60	min.06	8.697045	min.11	4.898876
1st.61	min.13	4.898154	min.13	4.897975
1st.62	min.16	3.326618	min.16	3.326354

Table XVII: Triples for alanine dipeptide.

	Lower minimum		Higher minimum	
Saddle	Minimum	Barrier (kcal/mole)	Minimum	Barrier (kcal/mole)
1st.01	min.01	0.848409	min.01	0.848382
1st.02	min.01	1.127800	min.05	0.005738
1st.03	min.01	1.202047	min.01	1.202030
1st.04	min.02	0.834349	min.04	0.440803
1st.05	min.02	0.845881	min.02	0.845848
1st.06	min.01	1.651087	min.04	0.546732
1st.07	min.03	0.855672	min.03	0.855646
1st.08	min.02	1.164151	min.05	0.752897
1st.09	min.02	1.203885	min.02	1.203873
1st.10	min.04	0.851437	min.04	0.851405
1st.11	min.05	0.852694	min.05	0.852667
1st.12	min.03	1.206913	min.03	1.206899
1st.13	min.03	1.258871	min.04	0.961113
1st.14	min.04	1.203167	min.04	1.203154
1st.15	min.05	1.206712	min.05	1.206706
1st.16	min.06	0.847300	min.06	0.847271
1st.17	min.03	2.033302	min.06	1.104540
1st.18	min.06	1.203327	min.06	1.203316
1st.19	min.07	0.853900	min.07	0.853867
1st.20	min.07	1.212788	min.07	1.212787
1st.21	min.02	3.468708	min.02	3.468706
1st.22	min.06	3.459121	min.06	3.458999
1st.23	min.01	5.657017	min.01	5.656919
1st.24	min.03	4.892453	min.03	4.892383
1st.25	min.08	0.852082	min.08	0.852046
1st.26	min.04	4.864000	min.04	4.863983
1st.27	min.08	1.202419	min.08	1.202410
1st.28	min.07	4.827556	min.08	2.124151
1st.29	min.07	5.050969	min.07	5.050793
1st.30	min.09	0.845286	min.09	0.845254
1st.31	min.10	0.847518	min.10	0.847489
1st.32	min.09	1.205599	min.09	1.205575
1st.33	min.09	1.628720	min.10	1.505280
1st.34	min.08	4.156048	min.08	4.155910
1st.35	min.09	2.493324	min.10	2.369264
1st.36	min.02	10.460099	min.06	9.435548
1st.37	min.08	6.384940	min.10	4.084111
1st.38	min.03	11.079090	min.05	10.763625

Table XVIII: Transition rate matrix for alanine at $T = 300\,\mathrm{K}$. Table entries are base 10 logarithms of the actual transition rates expressed in Hz.

	From 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	10.6	11.1	11.9	12.3		11.6	11.1			9.7	10.3		10.8				11.7
2	11.0	10.5			11.8		10.9	12.1		10.4		10.2					
3	11.5		10.2	11.9	11.6								10.5				
4	12.0		12.0	10.5				10.9					10.5				
5		11.6	11.6		10.2	11.9			10.1					10.4			
6	11.5				12.0	10.3		11.7			10.1						
7	11.6	11.5															
To 8		12.3		11.3		11.9											
9					7.3				9.1	11.8	12.2		9.4	11.2			
10	6.4	7.2							11.6	9.5	12.3	11.0					
11	7.0					6.9			12.0	12.3	9.3		8.8		10.7		
12		6.8								10.8		10.8		12.1	12.3		
13	7.2		7.3	7.2					9.0		8.5		10.7			12.5	12.2
14					6.8				10.5			11.7		10.4	11.4		
15											10.6	12.3		11.8			
16													10.8			10.7	12.3
17	6.4												10.5			12.3	

Table XIX: Transition probability matrix for alanine at $T = 300\,\mathrm{K}$.

	From 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.0189	.0452	.3710	.6417		.2130	.6140			.0022	.0061		.2363				.1241
2	.0456	.0127			.3225		.3861	.6736		.0099		.0057					
3	.1627		.0079	.2734	.1831								.1298				
4	.4348		.4225	.0098				.0511					.1097				
5		.1521	.1987		.0067	.3559			.0078					.0113			
6	.1389				.4878	.0100		.2754			.0034						
7	.1994	.1243															
To 8		.6659		.0752		.4212											
9					8e-6				.0010	.2374	.4640		.0099	.0732			
10	1e-6	5e-6							.2587	.0013	.5144	.0327					
11	5e-6					4e-6			.7124	.7248	.0007		.0024		.0248		
12		2e-6								.0247		.0226		.5890	.8572		
13	7e-6		9e-6	5e-6					.0007		9e-5		.1798			.5733	.3685
14					3e-6				.0197			.1928		.0127	.1181		
15											.0116	.7464		.3141			
16													.2175			.0094	.5075
17	1e-6												.1150			.4175	

Table XX: Probability eigenvectors and decay times for alanine at $T = 300\,\mathrm{K}$.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\tau^{(0)} = \infty$ (equilibrium)																		
$u_j^{(0)}$	=	.120	.091	.051	.057	.055	.081	.406	.145	.001	.001	.001	.001	.001	.001	.001	.001	.001
$\tau^{(1)} = 5.92 \times 10^{-11}$ s																		
$u_j^{(1)}$	=	−.059	−.043	−.025	−.028	−.025	−.038	−.216	−.069	.141	.090	.093	.071	.002	.029	.078	.001	.001
$\tau^{(2)} = 8.15 \times 10^{-12}$ s																		
$u_j^{(2)}$	=	.007	.038	−.003	−.002	.016	.025	.262	.057	−.044	−.027	−.028	.039	−.384	.016	.044	−.008	−.007
$\tau^{(3)} = 7.88 \times 10^{-12}$ s																		
$u_j^{(3)}$	=	.002	−.002	.002	.002	−.002	−.003	.002	−.004	−.219	−.133	−.141	.176	.051	.069	.197	.001	.001
$\tau^{(4)} = 3.90 \times 10^{-12}$ s																		
$u_j^{(4)}$	=	−.044	−.079	−.035	−.033	−.061	−.090	.500	−.161	.001	.001	.001	.001	.001	.001	.001	.001	.001
$\tau^{(5)} = 1.64 \times 10^{-12}$ s																		
$u_j^{(5)}$	=	−.236	.119	−.127	−.138	.015	.037	.118	.213	.001	.001	−.001	−.001	.001	−.001	−.001	.001	.001
$\tau^{(6)} = 7.08 \times 10^{-13}$ s																		
$u_j^{(6)}$	=	.083	.113	−.004	.068	−.216	−.230	−.053	.238	−.001	.001	.001	.001	−.001	−.001	.001	.001	.001
$\tau^{(7)} = 4.89 \times 10^{-13}$ s																		
$u_j^{(7)}$	=	−.226	.117	.215	−.017	.145	−.231	.026	−.027	.001	−.001	−.001	−.001	.001	.001	−.001	−.001	−.001
$\tau^{(8)} = 4.70 \times 10^{-13}$ s																		
$u_j^{(8)}$	=	−.021	.010	.019	−.004	.014	−.013	.003	−.010	−.337	.251	.098	.031	.001	−.118	.077	.001	.001
$\tau^{(9)} = 4.47 \times 10^{-13}$ s																		
$u_j^{(9)}$	=	−.006	−.003	.006	.001	−.001	.003	.002	.004	.266	−.195	−.068	.053	−.001	−.230	.169	−.001	−.001
$\tau^{(10)} = 4.30 \times 10^{-13}$ s																		
$u_j^{(10)}$	=	−.005	.068	.010	−.027	.078	−.098	−.009	−.040	.003	−.002	−.001	.001	−.321	−.001	.001	.154	.189
$\tau^{(11)} = 3.84 \times 10^{-13}$ s																		
$u_j^{(11)}$	=	.183	.156	−.168	.014	.150	−.157	−.048	−.129	.001	−.001	−.001	.001	.001	−.001	.001	−.001	−.001
$\tau^{(12)} = 2.86 \times 10^{-13}$ s																		
$u_j^{(12)}$	=	−.174	−.178	−.093	.233	.087	−.056	.036	.146	−.001	.001	.001	−.001	.001	−.001	.001	.001	.001
$\tau^{(13)} = 2.70 \times 10^{-13}$ s																		
$u_j^{(13)}$	=	.167	−.236	.018	−.153	.085	−.113	.004	.229	−.001	−.001	.001	−.001	.001	.001	.001	−.001	−.001
$\tau^{(14)} = 2.49 \times 10^{-13}$ s																		
$u_j^{(14)}$	=	−.004	.012	.001	.003	−.004	.004	−.001	−.009	−.027	−.024	.053	−.434	.001	.059	.373	−.001	−.001
$\tau^{(15)} = 2.20 \times 10^{-13}$ s																		
$u_j^{(15)}$	=	−.004	.005	−.001	.002	.001	−.002	.001	−.002	−.158	−.307	.463	.032	−.001	−.002	−.029	.001	.001
$\tau^{(16)} = 1.60 \times 10^{-13}$ s																		
$u_j^{(16)}$	=	−.046	.002	.001	.010	−.001	.003	.003	−.002	−.001	.001	−.001	−.001	.114	.001	.001	−.453	.370

Table XXI: Transition rate matrix for alanine dipeptide at $T = 300\,\mathrm{K}$. Table entries are base 10 logarithms of the actual transition rates expressed in Hz.

		From 1	2	3	4	5	6	7	8	9	10
	1	12.4			12.2	11.7					
	2		12.4		11.8	11.4	5.9				
	3			12.4	11.4	4.1	11.7				
	4	11.8	11.7	11.2	12.4						
To	5	11.8	11.7	4.4		12.4					
	6		4.8	10.5			12.4				
	7							12.4	10.7		
	8							8.5	12.4		10.0
	9									12.4	11.5
	10								7.9	11.6	12.2

Table XXII: Probability eigenvectors and decay times for alanine dipeptide at $T = 300\,\mathrm{K}$.

	1	2	3	4	5	6	7	8	9	10
$\tau^{(0,a)} = \infty$ (equilibrium)										
$u_j^{(0,a)} =$	.251	.158	.161	.106	.314	.011	.000	.000	.000	.000
$\tau^{(1,a)} = 6.05 \times 10^{-12}\,\mathrm{s}$										
$u_j^{(1,a)} =$	.142	.098	$-.450$	.029	.231	$-.050$	.000	.000	.000	.000
$\tau^{(2,a)} = 2.02 \times 10^{-12}\,\mathrm{s}$										
$u_j^{(2,a)} =$	.028	.030	$-.478$	$-.022$	.114	.328	.000	.000	.000	.000
$\tau^{(3,a)} = 9.31 \times 10^{-13}\,\mathrm{s}$										
$u_j^{(3,a)} =$	.291	$-.486$	.004	$-.014$	.205	.000	.000	.000	.000	.000
$\tau^{(4,a)} = 7.84 \times 10^{-13}\,\mathrm{s}$										
$u_j^{(4,a)} =$	.317	.014	$-.040$	.167	$-.460$	.002	.000	.000	.000	.000
$\tau^{(5,a)} = 2.93 \times 10^{-13}\,\mathrm{s}$										
$u_j^{(5,a)} =$	.347	.123	.031	$-.386$	$-.113$	.000	.000	.000	.000	.000
$\tau^{(0,b)} = \infty$ (equilibrium)										
$u_j^{(0,b)} =$	.000	.000	.000	.000	.000	.000	.993	.007	.000	.000
$\tau^{(1,b)} = 1.88 \times 10^{-10}\,\mathrm{s}$										
$u_j^{(1,b)} =$	.000	.000	.000	.000	.000	.000	$-.500$	.054	.202	.245
$\tau^{(2,b)} = 2.16 \times 10^{-11}\,\mathrm{s}$										
$u_j^{(2,b)} =$	.000	.000	.000	.000	.000	.000	$-.499$	.500	.000	.000
$\tau^{(3,b)} = 1.36 \times 10^{-12}\,\mathrm{s}$										
$u_j^{(3,b)} =$	.000	.000	.000	.000	.000	.000	$-.028$	.069	.431	$-.472$

Table XXIII: Classification scheme for (ϕ, ψ) pair.

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

Table XXIV: Some relevant minima of tetra-alanine, including the alpha-helical ground state (min.1) and the extended conformation (min.1583). Table is continued on the next page.

Minimum	Class	E (kcal/mole)	ϕ_1 (deg)	ψ_1 (deg)	ϕ_2 (deg)	ψ_2 (deg)	ϕ_3 (deg)	ψ_3 (deg)	ϕ_4 (deg)	ψ_4 (deg)
min.1**	aaaa	−6.6427	291.088	324.888	288.232	323.652	291.110	320.053	291.100	319.917
min.2	aaai	−4.6014	290.594	331.498	296.407	320.177	283.311	312.544	272.011	67.416
min.3	aaab′	−3.7123	289.741	330.166	289.774	325.541	283.982	316.464	206.928	102.188
min.11	aai′a′	−2.2950	293.732	327.039	294.150	324.732	260.545	50.035	199.309	304.202
min.14	aaab	−1.8670	290.042	325.945	286.426	326.265	288.257	321.662	289.283	144.673
min.17	j′aai	−1.5702	203.340	175.491	302.231	315.665	284.649	308.041	273.059	73.339
min.21	i′aaa	−1.3500	215.034	36.412	287.996	329.027	293.890	321.499	284.431	319.025
min.22	aaai″	−1.3494	289.971	324.220	290.397	323.466	287.173	324.820	54.204	47.013
min.23	iaai′	−1.3344	280.912	75.518	293.578	330.091	299.642	327.048	244.500	36.758
min.25	ai′a′i	−1.1691	296.516	317.533	263.439	53.102	198.981	303.747	278.303	73.169
min.29	iaaa	−0.8911	278.421	68.085	271.089	308.381	283.423	325.858	284.583	330.163
min.31	ia′aa	−0.8004	279.257	63.605	194.069	304.378	273.089	311.542	277.316	313.388
min.33	ia′aa	−0.5982	280.566	68.270	198.965	302.454	280.251	319.576	279.515	328.443
min.34	iaai	−0.5532	280.776	75.704	292.887	332.344	297.638	322.884	276.076	70.001
min.35	aai′i	−0.5506	288.392	324.138	288.590	325.096	252.129	49.902	279.822	75.130
min.36	aai′i	−0.5480	293.923	314.343	283.786	315.939	268.756	64.240	279.564	75.341
min.40	ia′ai′	−0.3703	279.771	65.649	195.270	302.493	277.843	328.026	236.284	27.847
min.43	aai′a	−0.2937	288.403	325.983	287.642	328.594	251.206	42.855	284.408	325.160
min.60	iaii	0.2089	280.659	74.810	293.340	317.836	277.284	70.527	279.615	75.181
min.63	ai′i′i	0.2624	296.354	324.322	251.358	28.730	203.421	41.462	279.775	75.482
min.64	i′aai	0.2775	213.296	40.242	292.361	316.562	285.937	317.365	272.128	66.233
min.74	jaai	0.4873	283.353	153.193	295.352	326.348	294.049	322.643	275.108	68.654
min.79	aai′b	0.5774	287.361	324.511	288.710	325.221	250.999	49.459	284.039	144.497
min.82	iaii′	0.6078	280.737	74.867	290.365	320.928	278.578	71.853	208.745	41.400
min.84	ij′ai′	0.6904	280.198	75.042	204.528	160.723	298.647	325.376	249.197	38.321
min.91	iiii	0.7689	280.921	75.765	280.927	75.267	280.768	75.110	280.048	75.397
min.95	iaij′	0.8342	280.667	74.953	291.054	320.776	279.092	74.585	206.929	159.006
min.98	ai′i′a	0.8575	296.117	324.339	251.263	28.030	204.158	40.043	283.932	325.227
min.112	ai′i′j′	0.9968	297.773	324.921	250.435	33.232	204.529	43.945	205.589	159.141
min.116	iai′i	1.0649	279.344	71.274	284.771	318.539	207.040	40.669	278.903	75.073
min.132	ai′i′i′	1.1718	296.674	326.686	253.105	30.532	206.490	40.207	208.083	43.002
min.137	iaib	1.1983	280.699	74.974	293.509	317.540	277.481	70.810	283.286	147.302
min.138	ai′j′i	1.1999	289.121	326.880	214.597	38.209	202.901	161.278	281.728	77.102
min.148	ai′i′b	1.2984	296.237	323.823	250.337	28.976	203.036	41.710	283.017	144.502
min.151	ai′ai′	1.3162	290.833	326.730	246.372	42.212	291.456	328.274	216.916	41.079
min.159	ai′ii	1.3879	291.265	323.531	249.176	49.310	280.647	75.002	280.161	75.486
min.173	iia′i	1.4927	280.883	75.859	282.187	73.426	199.681	304.197	279.480	74.667
min.177	aaji	1.5261	287.639	326.979	288.441	326.290	281.875	160.095	281.193	74.727
min.179	ai′ii	1.5324	289.537	330.985	219.202	51.987	281.038	75.302	280.196	75.645
min.180	iiii″	1.5337	281.275	76.297	279.362	340.949	281.781	73.712	51.398	60.822
min.187	i′a′ai′	1.5853	254.065	19.303	205.733	304.060	293.358	328.309	222.447	35.408
min.199	iaj′i	1.6603	280.683	74.994	287.146	329.197	207.453	159.003	280.368	75.870
min.201	ibai	1.6833	281.095	76.231	285.565	145.947	293.017	320.213	277.303	72.427
min.209	ii′ii	1.7245	280.240	73.245	207.981	41.535	280.738	75.510	280.058	75.447
min.221	ij′ii	1.7956	281.107	76.490	205.670	160.385	281.141	75.101	279.988	75.387
min.235	ibii	1.8712	281.085	76.101	286.094	146.431	280.901	75.070	279.896	75.298
min.239	iabi	1.8963	280.600	74.982	289.492	321.029	279.827	142.372	281.013	76.420

Minimum	Class	E (kcal/mole)	ϕ_1 (deg)	ψ_1 (deg)	ϕ_2 (deg)	ψ_2 (deg)	ϕ_3 (deg)	ψ_3 (deg)	ϕ_4 (deg)	ψ_4 (deg)
min.251	iibi	1.9322	280.763	75.646	280.974	75.432	285.583	145.736	280.407	75.851
min.267	ii′ai′	2.0218	280.303	73.253	208.637	41.492	298.750	324.633	248.244	37.194
min.316	ii′ai′	2.1937	280.285	73.201	208.964	41.125	292.110	329.845	219.195	41.771
min.325	i′aia	2.2266	209.841	47.550	294.183	317.171	276.396	68.436	283.600	324.392
min.326	iai′b	2.2275	279.311	71.485	284.412	318.924	206.296	42.003	282.084	142.317
min.352	i′aii′	2.3051	209.854	48.419	291.256	319.922	278.076	71.752	209.510	41.284
min.358	ii′j′i	2.3318	280.169	73.172	207.982	40.964	204.695	158.623	280.186	75.617
min.385	ij′ii′	2.4042	280.657	75.518	205.244	159.503	280.204	72.490	207.274	42.239
min.387	iaai″	2.4080	280.328	74.153	291.009	326.554	290.249	326.407	53.229	47.300
min.401	ii′ii′	2.4461	280.336	73.374	207.887	41.765	280.350	73.729	207.799	42.924
min.404	iaa′a	2.4562	279.478	71.386	281.740	315.463	199.331	301.857	278.776	315.899
min.417	ibii′	2.5007	281.010	75.961	286.380	146.832	280.489	73.273	207.748	42.681
min.466	ii′ij′	2.6164	280.183	73.298	208.123	41.445	280.586	76.200	205.952	157.751
min.512	ii′ib	2.7183	280.304	73.456	208.125	41.462	280.967	76.153	285.384	145.483
min.523	ibj′i	2.7397	281.001	76.235	287.728	143.591	205.183	158.937	280.098	75.594
min.531	ii′i′i	2.7567	280.210	72.964	209.229	40.244	208.778	43.064	279.936	75.602
min.547	i′ia′i	2.7970	209.002	45.699	282.348	74.164	200.200	304.095	279.727	74.962
min.551	biij′	2.8116	285.536	140.720	280.950	75.411	280.458	75.459	205.826	157.618
min.562	ii′bi	2.8513	280.069	73.107	207.664	41.618	285.047	144.447	280.476	76.038
min.568	ij′j′i	2.8676	280.744	76.664	206.237	158.969	205.393	158.426	280.208	75.835
min.600	biib	2.9507	285.711	139.655	280.849	75.211	280.778	75.440	285.184	144.851
min.613	bbii	2.9765	285.318	141.350	286.323	145.310	280.911	75.115	279.849	75.303
min.618	iii′i″	2.9871	281.138	76.121	279.165	67.669	200.719	48.084	54.528	44.496
min.623	ii′i′i	2.9917	280.087	72.716	251.070	15.100	210.756	41.654	280.005	75.492
min.630	ii′j′j′	3.0096	280.366	73.666	208.081	41.340	204.934	158.813	205.464	158.129
min.633	ibbi	3.0132	280.999	76.263	286.077	146.175	286.025	144.576	280.441	76.008
min.663	bibi	3.1032	285.337	140.991	281.207	75.803	285.347	146.696	280.412	75.874
min.697	ibi′i	3.1691	281.204	76.434	287.072	144.627	209.720	43.487	279.948	75.628
min.730	ii′i′j′	3.2451	280.158	73.025	208.933	40.187	208.476	42.407	204.721	157.156
min.822	ij′j′j′	3.4907	280.716	76.757	206.209	159.098	205.407	159.017	205.518	158.279
min.869	ibbj′	3.5841	280.947	76.257	285.750	146.635	287.388	143.407	205.488	159.016
min.959	bbij′	3.7821	285.009	142.273	285.638	146.402	280.831	75.657	205.710	157.134
min.975	ii′i′b	3.8019	280.185	73.032	209.269	40.394	208.413	43.532	284.451	143.135
min.1095	bbib	4.0125	285.167	140.777	285.741	145.585	281.152	75.833	284.850	145.157
min.1114	bbbi	4.0442	284.987	142.700	285.984	146.123	286.157	145.937	280.376	75.923
min.1253	bj′j′j′	4.2862	286.361	141.712	205.525	159.880	205.410	158.862	205.503	158.399
min.1305	bj′i′j′	4.3852	286.584	142.079	205.645	160.522	211.090	40.653	204.113	157.106
min.1338	bbj′j′	4.4576	284.689	142.123	286.990	145.084	205.378	159.673	205.440	157.973
min.1410	bbbj′	4.6113	284.739	142.857	285.569	146.610	287.414	144.872	205.411	158.785
min.1473	iaa′i″	4.7402	280.256	74.012	285.665	332.880	206.028	302.858	53.097	47.534
min.1497	bi′i′j′	4.7648	285.615	140.902	213.960	38.185	207.003	43.025	204.742	157.080
min.1583**	bbbb	4.9159	284.726	142.633	285.597	147.139	285.907	146.801	285.359	145.270
min.2210	aai′a″	6.4112	295.754	329.033	300.723	326.911	245.860	38.554	65.320	283.281
min.2482	ba″i′i	7.4542	296.215	117.472	63.760	281.189	204.359	36.378	277.629	73.127
min.9305	aaia″	36.4913	291.329	328.388	299.971	322.799	284.344	87.086	124.069	321.733
min.9309	iaia″	36.6860	280.932	75.379	300.330	322.262	285.105	87.140	124.004	321.676

Table XXV: Number of pathways from min.1583 to min.1 with a given length restriction.

Maximum length	Pathways	Pathways (10^6 Hz cutoff)
9	0	0
10	4	3
11	43	36
12	325	235
13	2023	1376
14	11895	6698
15	62292	28298

Table XXVI: Pathways from min.1583 (extended conformation) to min.1 (alpha-helical conformation) consisting of 11 minima or less, in increasing order of overall transition time. An asterisk indicates an extremely slow transition (over 1 s). Table is continued on the next page.

Transition time (s)	Pathway										
8.79×10^{-11}	1583	1114	663	562	209	60	36	35	2	1	
	bbbb	bbbi	bibi	ii′bi	ii′ii	iaii	aai′i	aai′i	aaai	aaaa	
1.06×10^{-10}	1583	1114	663	562	239	199	60	36	35	2	1
	bbbb	bbbi	bibi	ii′bi	iabi	iaj′i	iaii	aai′i	aai′i	aaai	aaaa
1.10×10^{-10}	1583	1114	663	562	209	179	159	36	35	2	1
	bbbb	bbbi	bibi	ii′bi	ii′ii	ai′ii	ai′ii	aai′i	aai′i	aaai	aaaa
1.15×10^{-10}	1583	1114	663	562	358	199	60	36	35	2	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	iaj′i	iaii	aai′i	aai′i	aaai	aaaa
1.16×10^{-10}	1583	1114	633	251	91	209	60	36	35	2	1
	bbbb	bbbi	ibbi	iibi	iiii	ii′ii	iaii	aai′i	aai′i	aaai	aaaa
1.18×10^{-10}	1583	1114	633	235	91	209	60	36	35	2	1
	bbbb	bbbi	ibbi	ibii	iiii	ii′ii	iaii	aai′i	aai′i	aaai	aaaa
1.18×10^{-10}	1583	1114	663	251	91	209	60	36	35	2	1
	bbbb	bbbi	bibi	iibi	iiii	ii′ii	iaii	aai′i	aai′i	aaai	aaaa
1.20×10^{-10}	1583	1114	663	562	358	138	63	35	2	1	
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ai′j′i	ai′i′i	aai′i	aaai	aaaa	
1.24×10^{-10}	1583	1114	613	235	91	209	60	36	35	2	1
	bbbb	bbbi	bbii	ibii	iiii	ii′ii	iaii	aai′i	aai′i	aaai	aaaa
1.50×10^{-10}	1583	1114	663	562	358	138	63	132	98	43	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ai′j′i	ai′i′i	ai′i′i′	ai′i′a	aai′a	aaaa
1.51×10^{-10}	1583	1114	633	235	221	209	60	36	35	2	1
	bbbb	bbbi	ibbi	ibii	ij′ii	ii′ii	iaii	aai′i	aai′i	aaai	aaaa
1.55×10^{-10}	1583	1114	663	562	358	138	177	36	35	2	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ai′j′i	aaji	aai′i	aai′i	aaai	aaaa
1.61×10^{-10}	1583	1114	613	235	221	209	60	36	35	2	1
	bbbb	bbbi	bbii	ibii	ij′ii	ii′ii	iaii	aai′i	aai′i	aaai	aaaa
2.09×10^{-10}	1583	1114	663	562	358	531	63	35	2	1	
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ii′i′i	ai′i′i	aai′i	aaai	aaaa	
2.20×10^{-10}	1583	1410	869	417	385	84	267	316	151	2	1
	bbbb	bbbj′	ibbj′	ibii′	ij′ii′	ij′ai′	ii′ai′	ii′ai′	ai′ai′	aaai	aaaa
2.28×10^{-10}	1583	1114	633	523	568	358	138	63	35	2	1
	bbbb	bbbi	ibbi	ibj′i	ij′j′i	ii′j′i	ai′j′i	ai′i′i	aai′i	aaai	aaaa
2.34×10^{-10}	1583	1114	663	562	358	531	63	132	98	43	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ii′i′i	ai′i′i	ai′i′i′	ai′i′a	aai′a	aaaa
2.70×10^{-10}	1583	1114	663	562	358	138	63	132	98	11	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ai′j′i	ai′i′i	ai′i′i′	ai′i′a	aai′a′	aaaa
3.17×10^{-10}	1583	1114	633	523	568	358	531	63	35	2	1
	bbbb	bbbi	ibbi	ibj′i	ij′j′i	ii′j′i	ii′i′i	ai′i′i	aai′i	aaai	aaaa
3.18×10^{-10}	1583	1114	663	562	358	531	63	132	98	11	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ii′i′i	ai′i′i	ai′i′i′	ai′i′a	aai′a′	aaaa
3.55×10^{-10}	1583	1114	663	562	358	531	623	116	35	2	1
	bbbb	bbbi	bibi	ii′bi	ii′j′i	ii′i′i	ii′i′i	iai′i	aai′i	aaai	aaaa

Transition time (s)	Pathway										
4.78×10^{-10}	1583	1095	600	512	137	975	148	79	14	2	1
	bbbb	bbib	biib	ii′ib	iaib	ii′i′b	ai′i′b	aai′b	aaab	aaai	aaaa
7.59×10^{-10}	1583	1095	600	512	137	34	23	29	33	31	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaai′	iaaa	ia′aa	ia′aa	aaaa
8.46×10^{-10}	1583	1095	600	512	137	34	23	316	151	2	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaai′	ii′ai′	ai′ai′	aaai	aaaa
9.68×10^{-10}	1583	1095	600	512	137	34	23	40	187	2	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaai′	ia′ai′	i′a′ai′	aaai	aaaa
1.00×10^{-9}	1583	1114	633	523	697	201	173	547	25	2	1
	bbbb	bbbi	ibbi	ibj′i	ibi′i	ibai	iia′i	i′ia′i	ai′a′i	aaai	aaaa
1.15×10^{-9}	1583	1095	600	512	137	34	60	36	35	2	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaii	aai′i	aai′i	aaai	aaaa
1.58×10^{-9}	1583	1410	959	551	466	401	82	352	325	21	1
	bbbb	bbbj′	bbij′	biij′	ii′ij′	ii′ii′	iaii′	i′aii′	i′aia	i′aaa	aaaa
1.71×10^{-9}	1583	1095	600	512	466	401	82	352	325	21	1
	bbbb	bbib	biib	ii′ib	ii′ij′	ii′ii′	iaii′	i′aii′	i′aia	i′aaa	aaaa
2.08×10^{-9}	1583	1410	959	551	466	95	82	352	325	21	1
	bbbb	bbbj′	bbij′	biij′	ii′ij′	iaij′	iaii′	i′aii′	i′aia	i′aaa	aaaa
2.19×10^{-9}	1583	1095	600	512	466	95	82	352	325	21	1
	bbbb	bbib	biib	ii′ib	ii′ij′	iaij′	iaii′	i′aii′	i′aia	i′aaa	aaaa
3.42×10^{-9}	1583	1410	1338	1253	1305	1497	730	112	3	2	1
	bbbb	bbbj′	bbj′j′	bj′j′j′	bj′i′j′	bi’i’j’	ii′i′j′	ai′i′j′	aaab′	aaai	aaaa
3.43×10^{-9}	1583	1410	1338	1253	822	630	730	112	3	2	1
	bbbb	bbbj′	bbj′j′	bj′j′j′	ij′j′j′	ii′j′j′	ii′i′j′	ai′i′j′	aaab′	aaai	aaaa
3.84×10^{-9}	1583	1095	600	512	137	34	74	17	64	2	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	jaai	j′aai	i′aai	aaai	aaaa
1.31×10^{-8}	1583	1095	600	512	137	975	148	112	3	2	1
	bbbb	bbib	biib	ii′ib	iaib	ii′i′b	ai′i′b	ai′i′j′	aaab′	aaai	aaaa
7.62×10^{-4}	1583	1095	600	512	137	34	23	29	404	31	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaai′	iaaa	iaa′a	ia′aa	aaaa
long	1583	1095	600	512	137	326	* 2482	* 116	35	2	1
	bbbb	bbib	biib	ii′ib	iaib	iai′b	ba″i′i	iai′i	aai′i	aaai	aaaa
long	1583	1095	600	512	137	34	74	* 387	22	* 2	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	jaai	iaai″	aaai″	aaai	aaaa
long	1583	1095	600	512	137	34	* 387	1473	22	* 2	1
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaai″	iaa′i″	aaai″	aaai	aaaa
long	1583	1095	600	512	137	34	* 387	22	* 2	1	
	bbbb	bbib	biib	ii′ib	iaib	iaai	iaai″	aaai″	aaai	aaaa	
long	1583	1114	633	523	697	* 618	* 9309	9305	* 2210	* 43	1
	bbbb	bbbi	ibbi	ibj′i	ibi′i	iii′i″	iaia″	aaia″	aai′a″	aai′a	aaaa
long	1583	1114	663	562	209	60	34	* 387	22	* 2	1
	bbbb	bbbi	bibi	ii′bi	ii′ii	iaii	iaai	iaai″	aaai″	aaai	aaaa
long	1583	1114	663	562	239	* 180	* 60	36	35	2	1
	bbbb	bbbi	bibi	ii′bi	iabi	iiii″	iaii	aai′i	aai′i	aaai	aaaa

Table XXVII: Selected results from our αBB tetra-alanine runs.

Region	Type	Eigenmode III	αBB	α	CPU (s)	Iterations
aaaa	min	1	1	25	2341	1733
bbbb	min	1	1	20	235916	284127
	1st	4	4			
	2nd	6	6			
	3rd	4	4			
	4th	1	1			
bibi	min	1	1	20	21136	25632
	1st	1	2			
	2nd	0	1			
bbbj′	min	2	2	20	395041	469618
	1st	8	9			
	2nd	4	17			
	3rd	3	16			
	4th	2	7			
	5th	0	1			
aai′i	min	2	2	80	484865	362337
	1st	1	1			

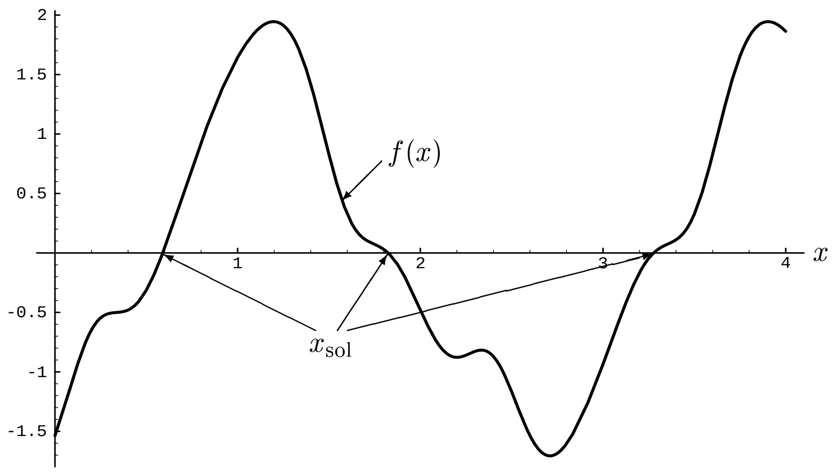

Figure 1: Plot of $f(x)$ for $x \in [0, 4]$.

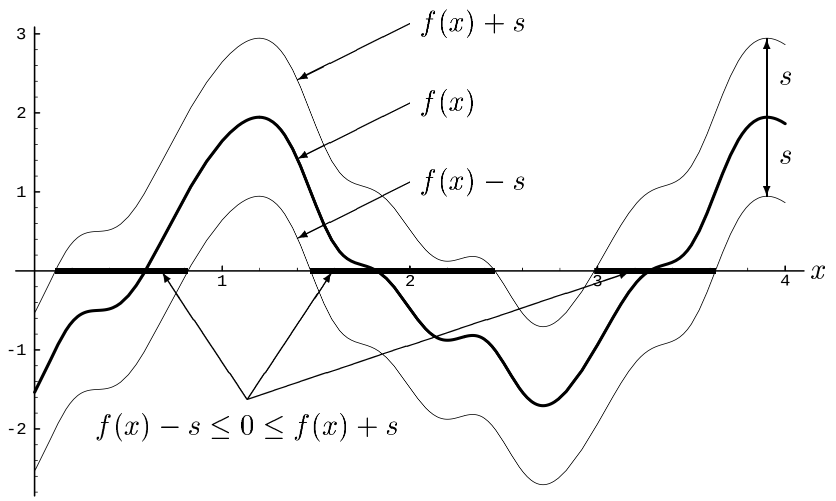

Figure 2: $f(x)$ is shifted by a positive slack variable $s = 1$. Note that the feasibility region of $f(x) - s \leq 0 \leq f(x) + s$ forms intervals around the solutions to $f(x) = 0$.

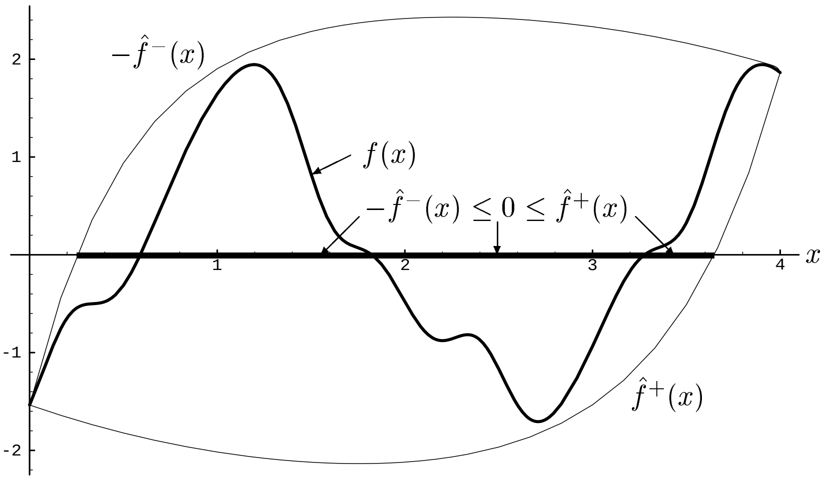

Figure 3: The functions $\hat{f}^{\pm}(x)$ are convex underestimators of $\pm f(x)$ over the interval $[0, 4]$. Note how $\hat{f}^{+}(x)$ and $-\hat{f}^{-}(x)$ form a convex envelope around $f(x)$.

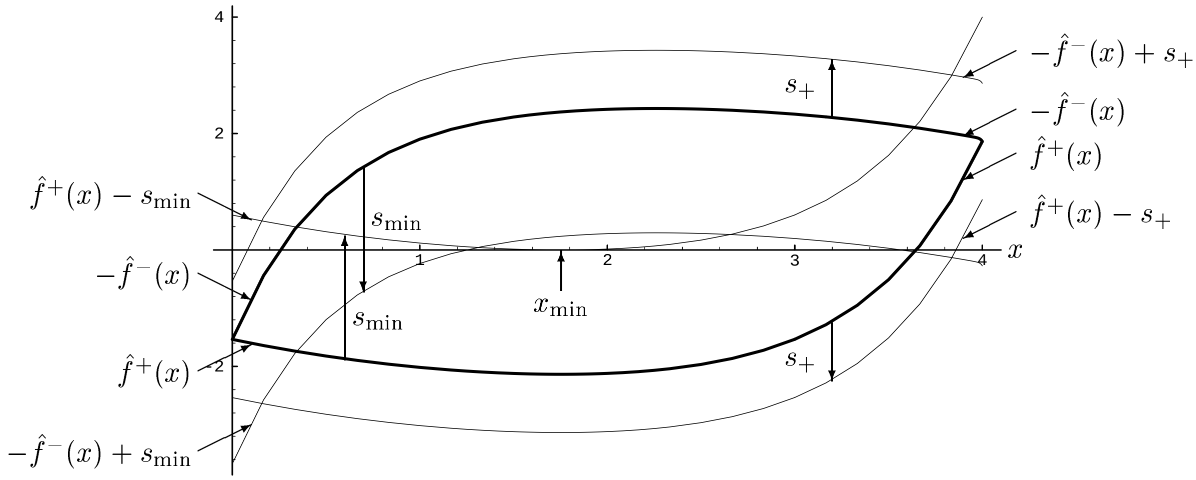

Figure 4: During the solution to the lower bounding problem, the convex underestimators $\hat{f}^{\pm}(x)$ are shifted by a slack variable. Two different shifts are shown above, one positive $s_{+} = 1$, and the other negative $s_{\min} = -2.135$. $s_{\min}$ represents the global minimum to the lower bounding problem: the feasibility region of the lower bounding problem is reduced to a single point $x_{\min} = 1.754$, shown above.

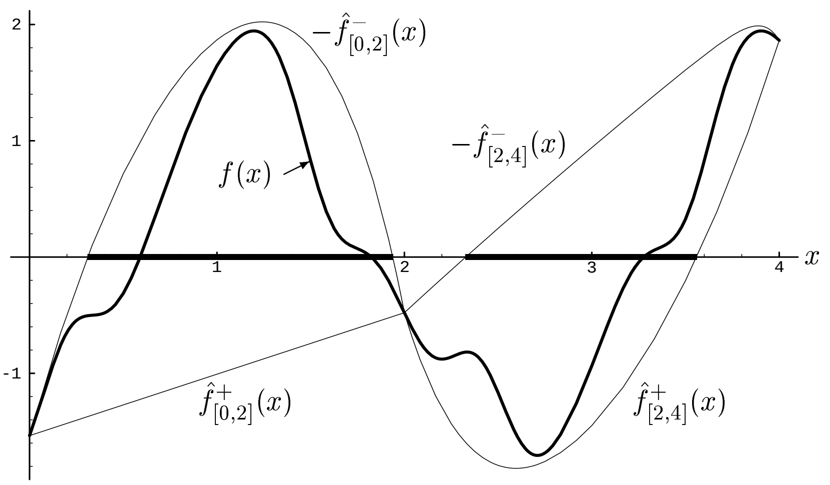

Figure 5: The interval $[0,4]$ has been subdivided into $[0,2]$ and $[2,4]$. The convex underestimators $\hat{f}^{\pm}(x)$ for each subinterval, shown above, form a convex envelope around $f(x)$. As the intervals get smaller, the envelope gets tighter.

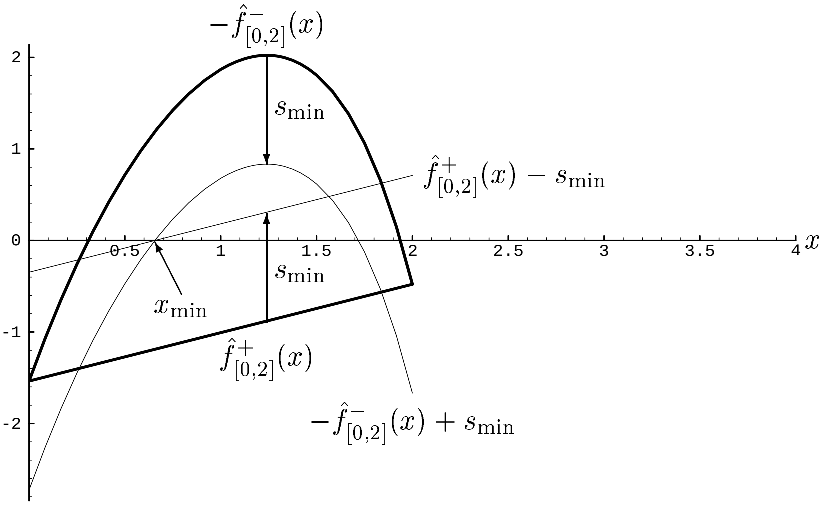

Figure 6: This figure represents the solution to the lower bounding problem in the interval $[0,2]$. $(x_{\min}, s_{\min}) = (0.656, -1.189)$.

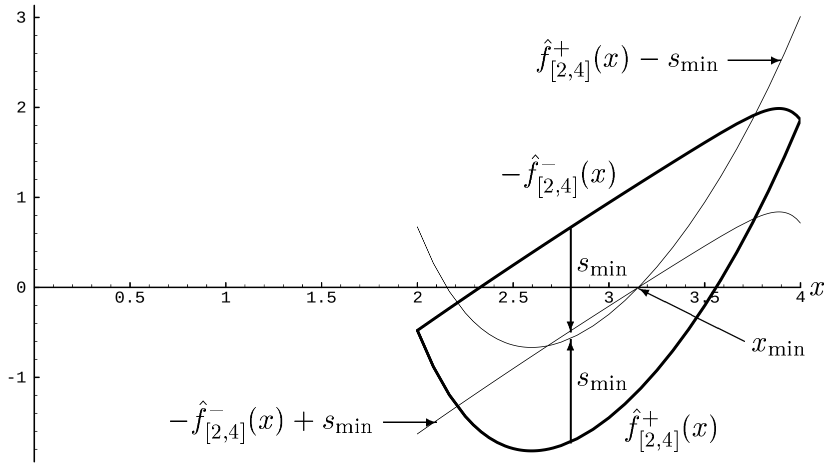

Figure 7: This figure represents the solution to the lower bounding problem in the interval $[2,4]$. $(x_{\min}, s_{\min}) = (3.154, -1.150)$.

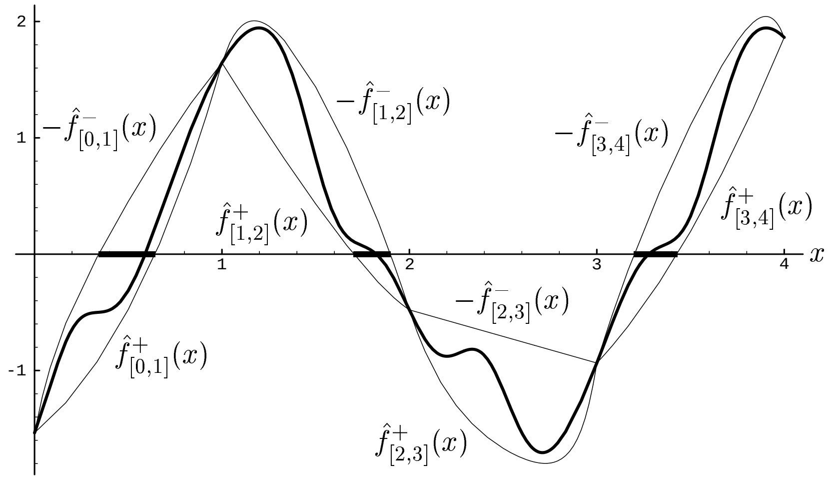

Figure 8: The intervals $[0,2]$ and $[2,4]$ have been further subdivided into $[0,1]$, $[1,2]$, $[2,3]$, and $[3,4]$. Shown above are the convex envelopes around $f(x)$ formed by convex underestimators in each of these intervals. Note that the convex envelopes for $[0,1]$, $[1,2]$, and $[3,4]$ intersect the x-axis, but the convex envelope for $[2,3]$ does not. This will allow us to conclude rigorously that no solutions to $f(x) = 0$ exist in $[2,3]$.

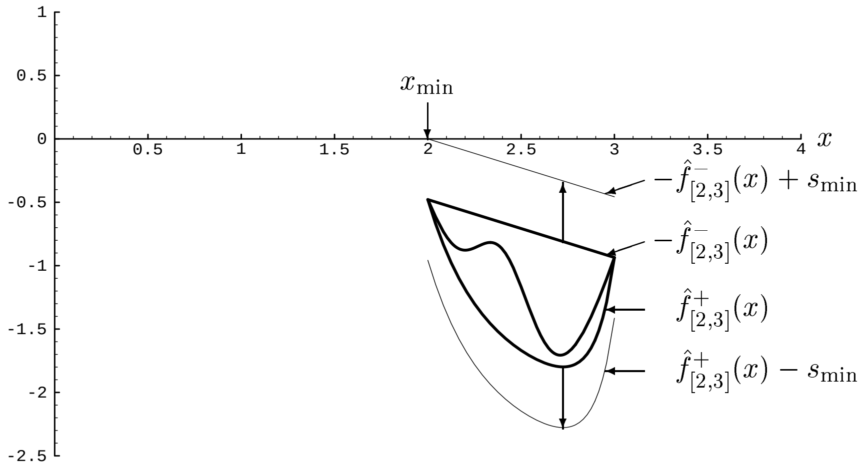

Figure 9: The lower bounding problem for the interval $[2,3]$ is solved. Note that the convex envelope must be *expanded* before it touches the x-axis resulting in a positive value for $s_{\min}$. This interval will be fathomed. $(x_{\min}, s_{\min}) = (2, +0.479)$.

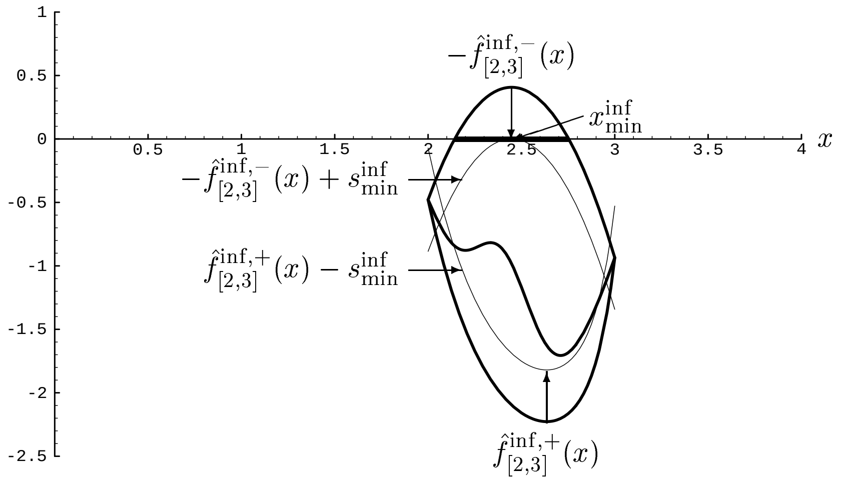

Figure 10: This is the same situation as Figure 9, except that the underestimators $\hat{f}^{\text{inf},\pm}_{[2,3]}(x)$ are less tight. As a result, the convex envelope now crosses the x-axis and $s^{\text{inf}}_{\min}$ will be negative. This interval will *not* be fathomed, in spite of the fact that no solutions to $f(x) = 0$ exist there. $(x^{\text{inf}}_{\min}, s^{\text{inf}}_{\min}) = (2.447, -0.407)$.

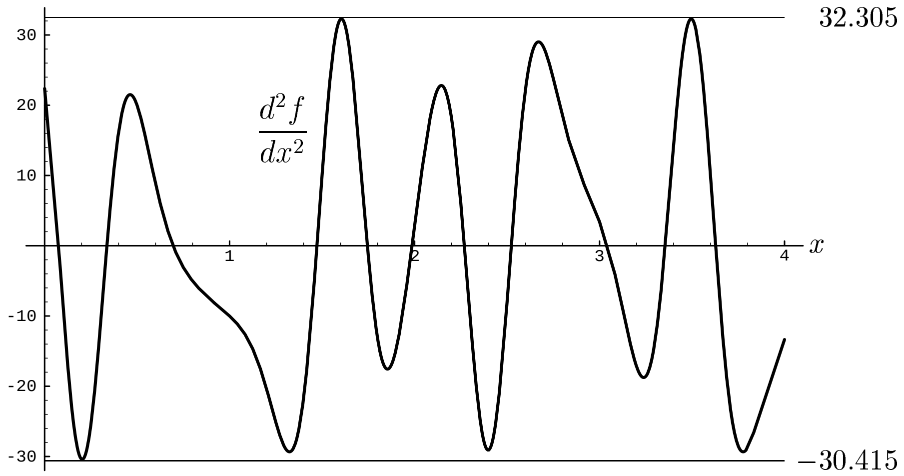

Figure 11: This is a plot of d^2f/dx^2. The maximum and minimum values of d^2f/dx^2 are useed to calculate adequate values of $\alpha^{\pm}$.

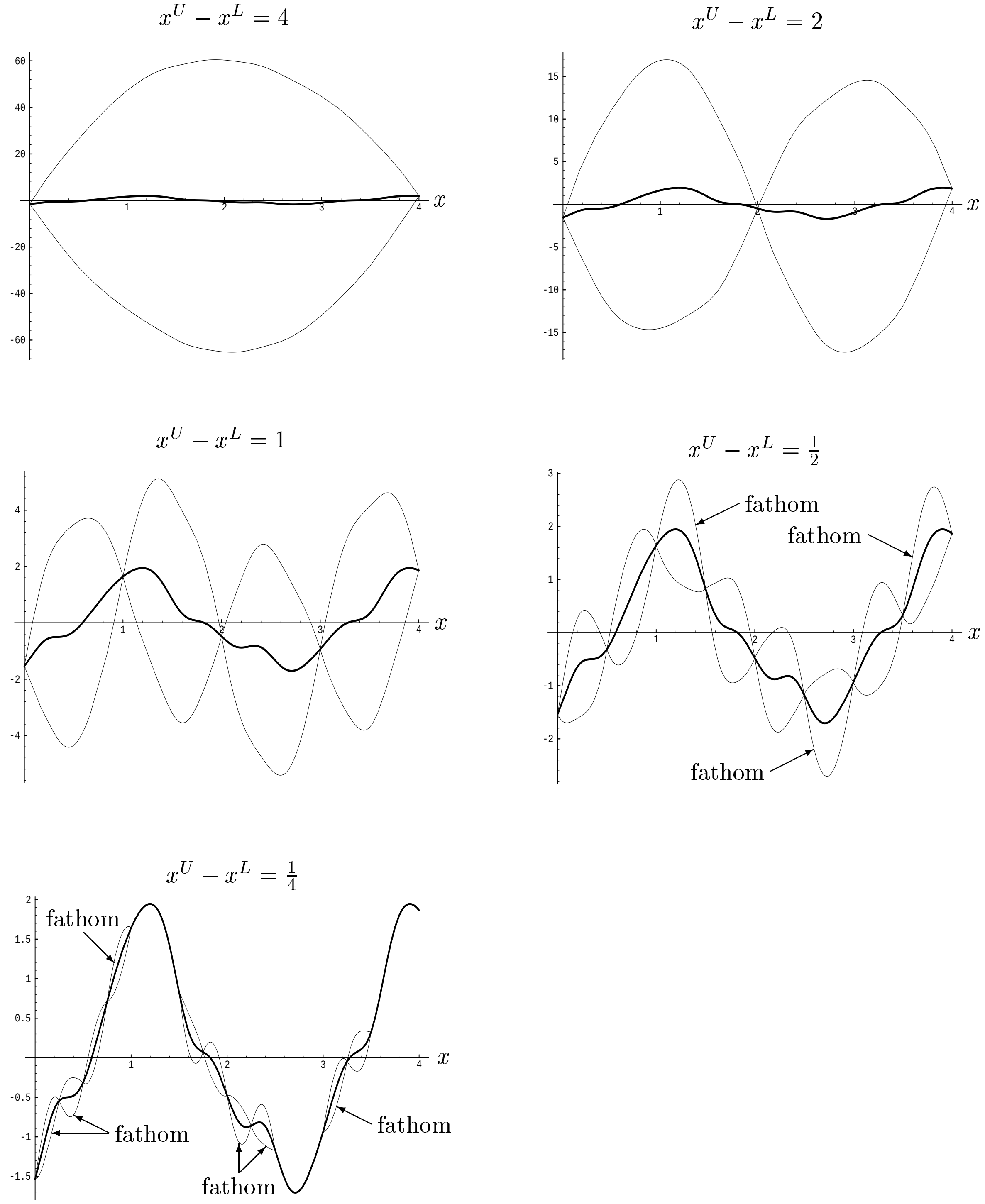

Figure 12: Shown above are the convex envelopes formed by α underestimators for various interval sizes $x^U - x^L$, starting with the entire interval $[0, 4]$, and subdividing. The values of alpha are fixed to $\alpha^+ = 15.206$ and $\alpha^- = 16.152$. The convex envelopes which don't cross the x-axis are designated as "fathomed" to indicate that these intervals would be fathomed by the αBB algorithm.

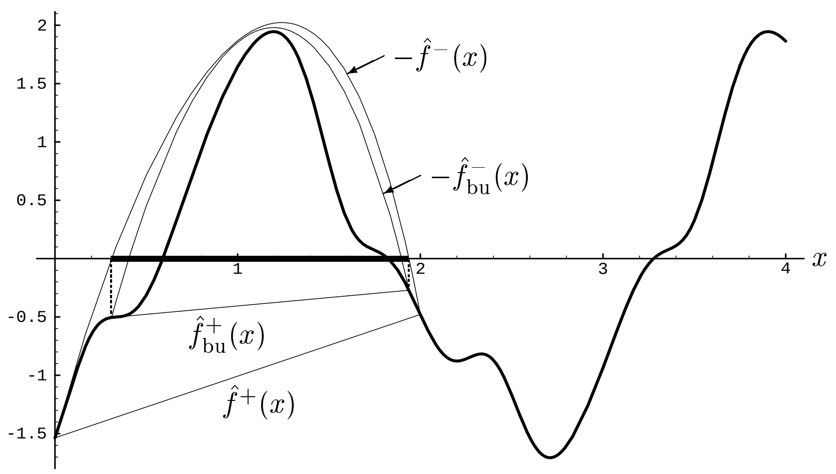

Figure 13: Using the bound update procedure, the interval $[0, 2]$ can be reduced to $[x^L, x^U] = [0.310, 1.938]$ (highlighted above). This allows us to use tighter underestimators $\hat{f}^{\pm}_{\mathrm{bu}}(x)$ and at the same time, reduce the interval we are considering. This procedure can be repeated, although in this case, not much would be gained.

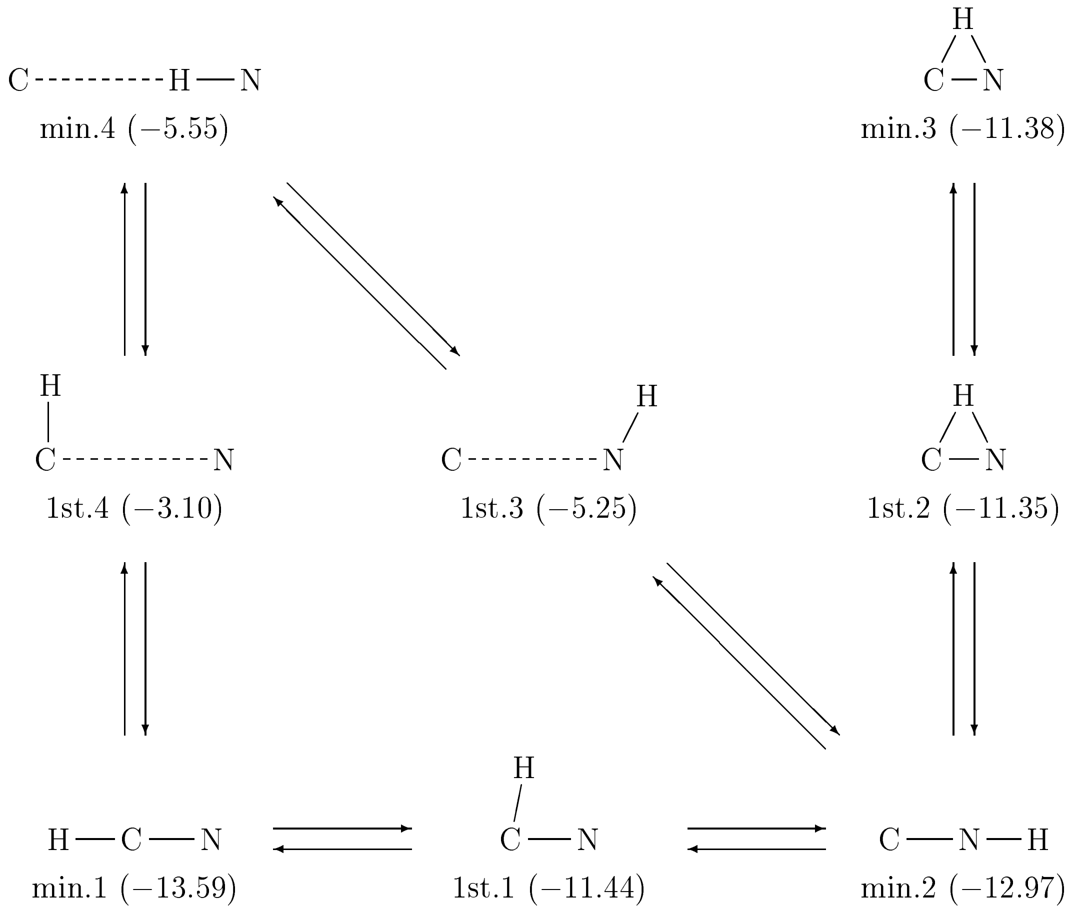

Figure 14: Minima and first-order transition states of HCN.

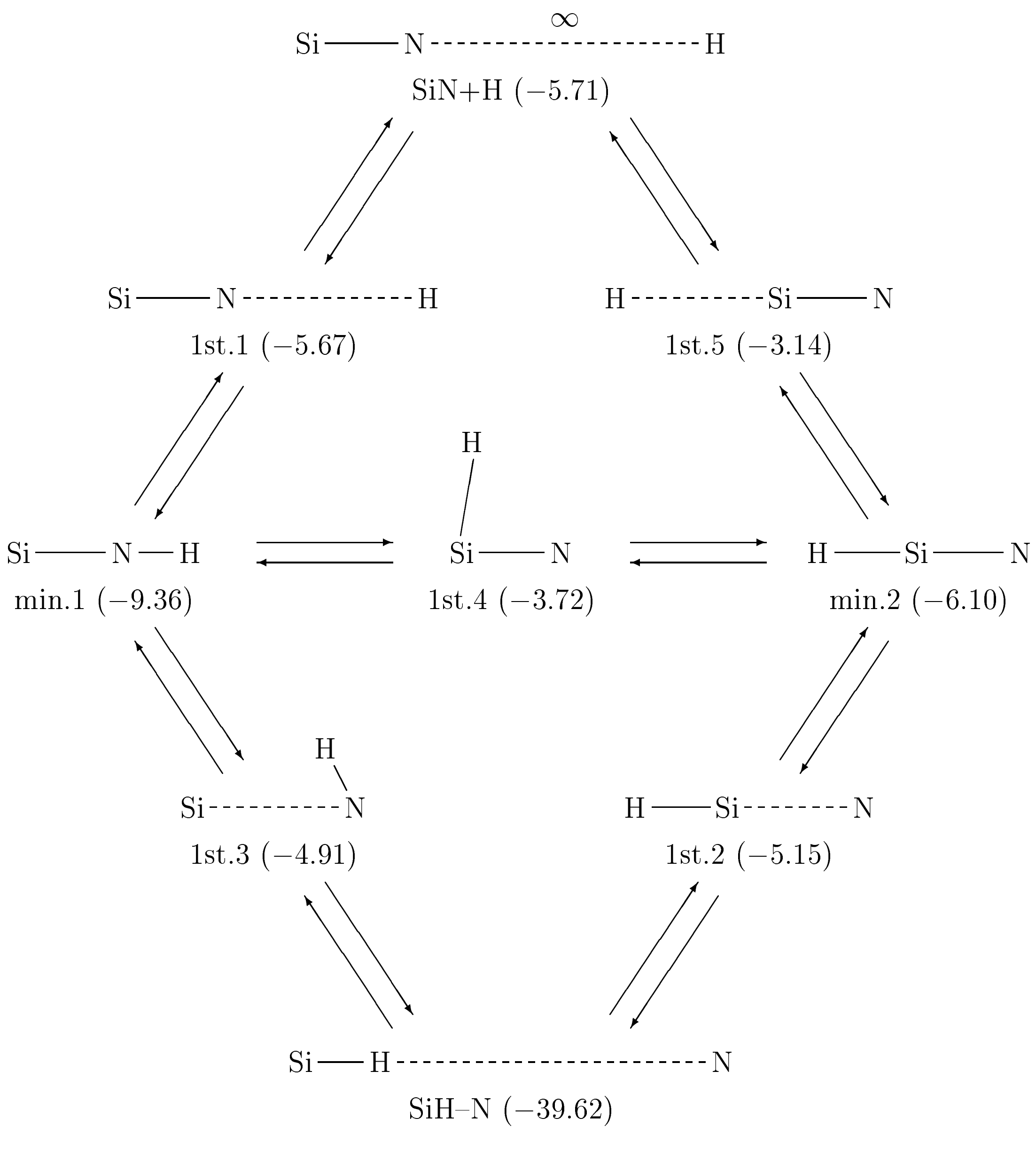

Figure 15: Minima and first-order transition states of HSiN.

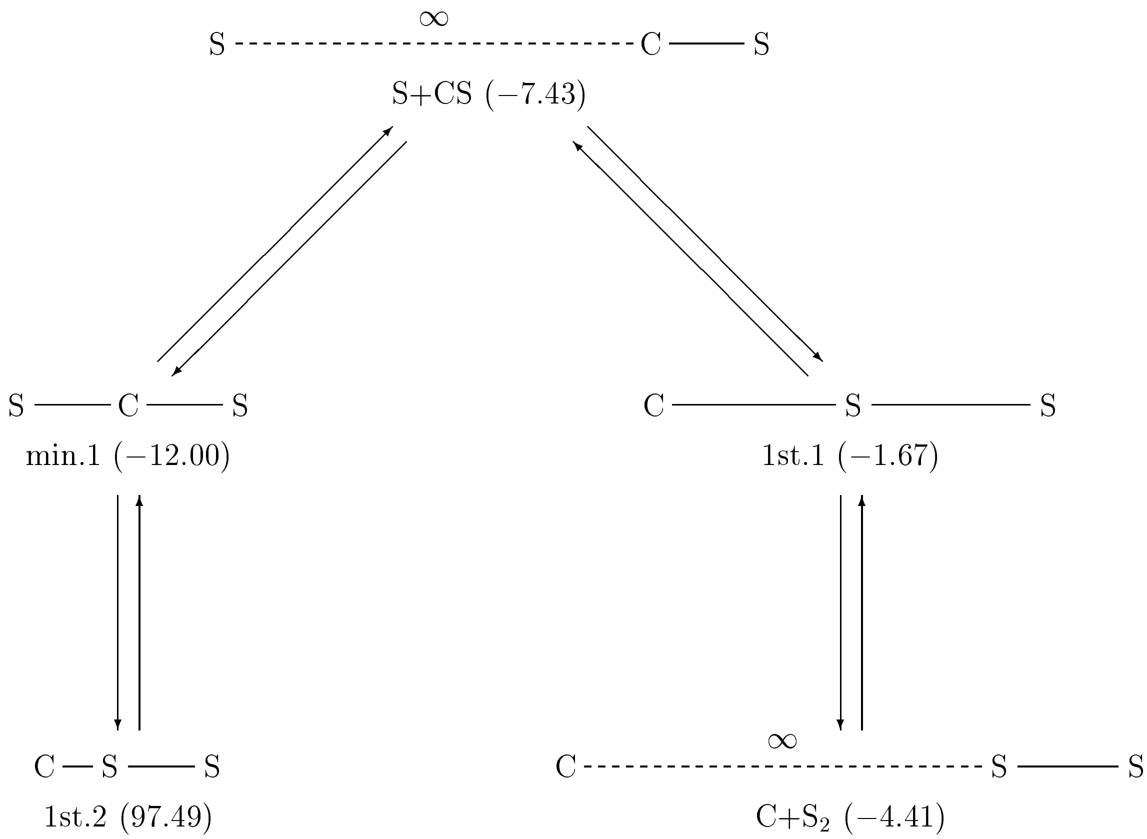

Figure 16: Minima and first-order transition states of CS_2.

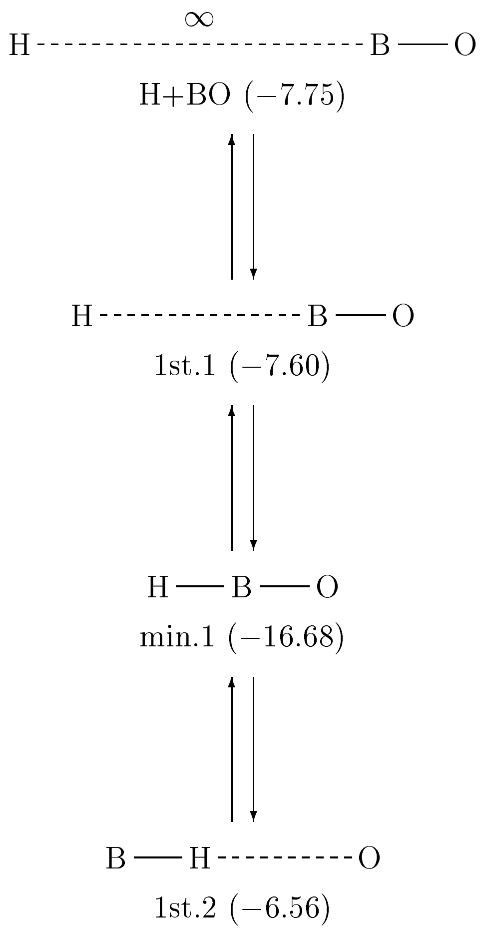

Figure 17: Minima and first-order transition states of HBO (PES1).

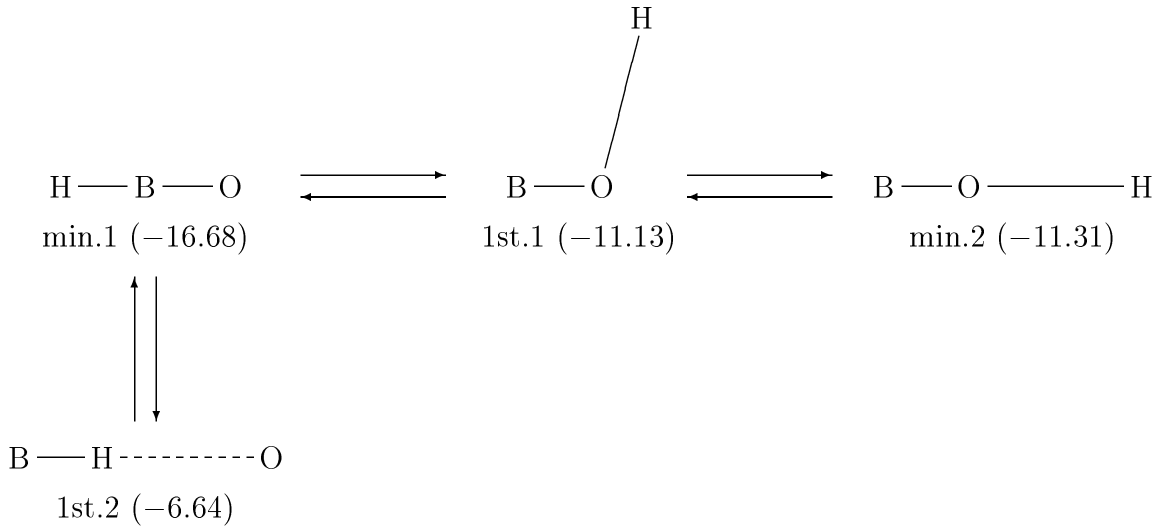

Figure 18: Minima and first-order transition states of HBO (PES2).

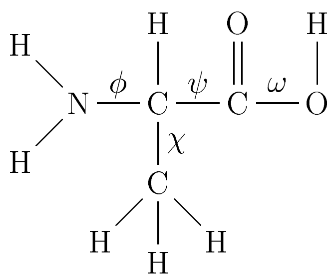

Figure 19: Alanine

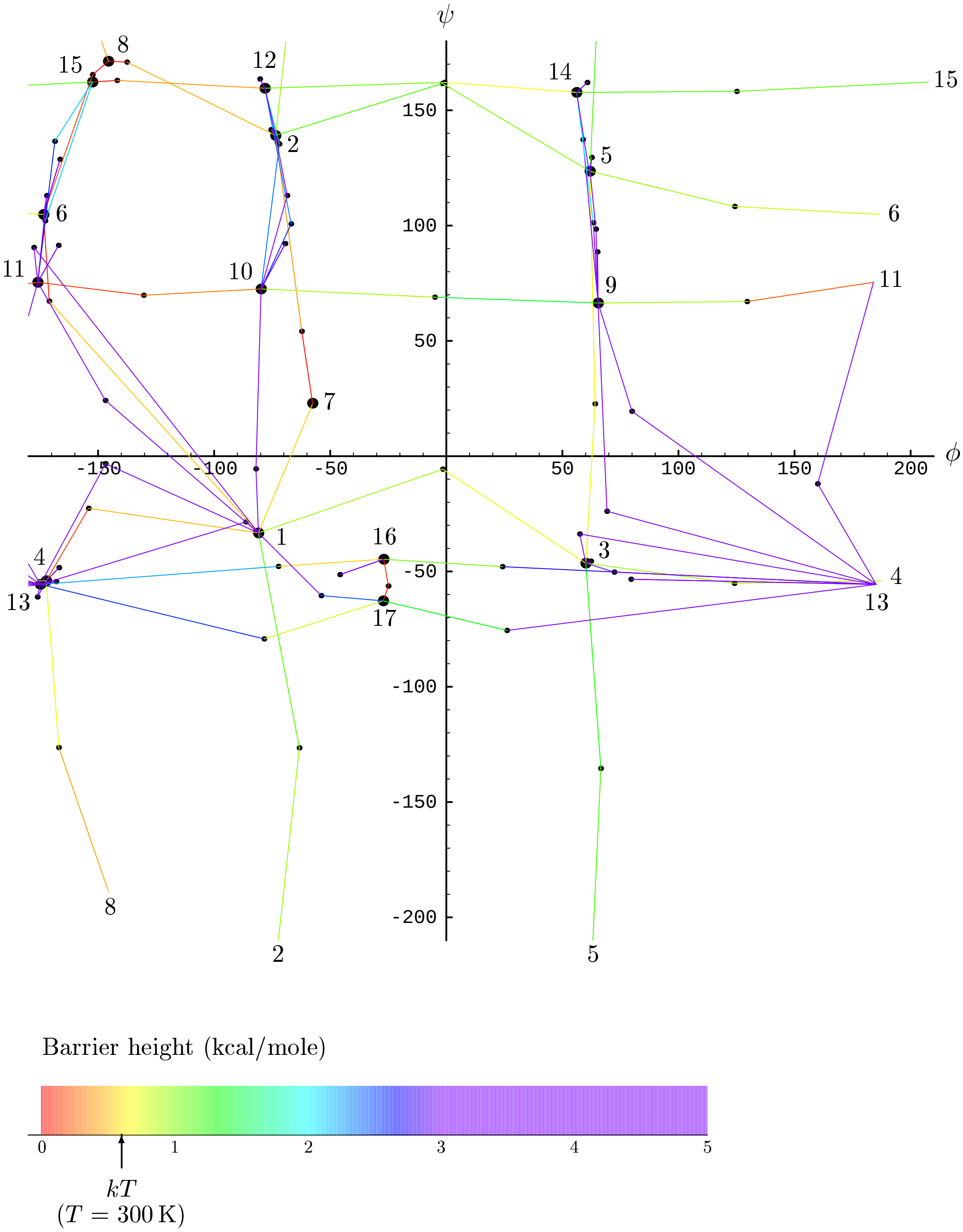

Figure 20: ψ–ϕ plot for alanine.

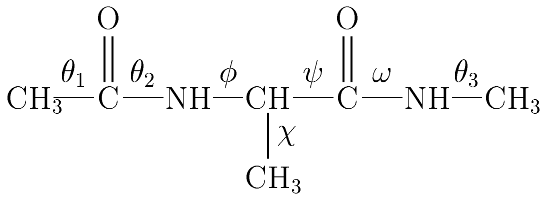

Figure 21: Alanine dipeptide

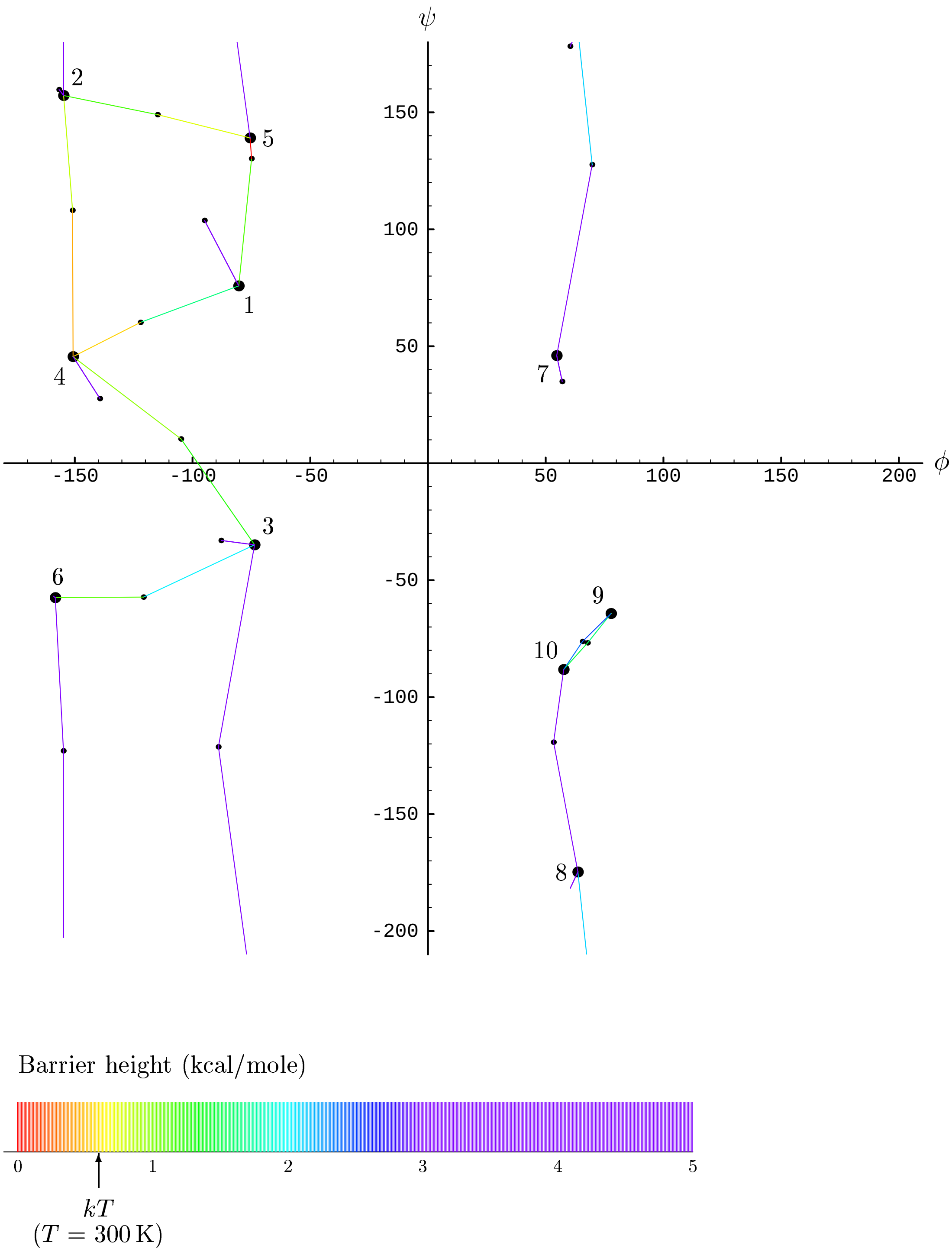

Figure 22: ψ–ϕ plot for alanine dipeptide.

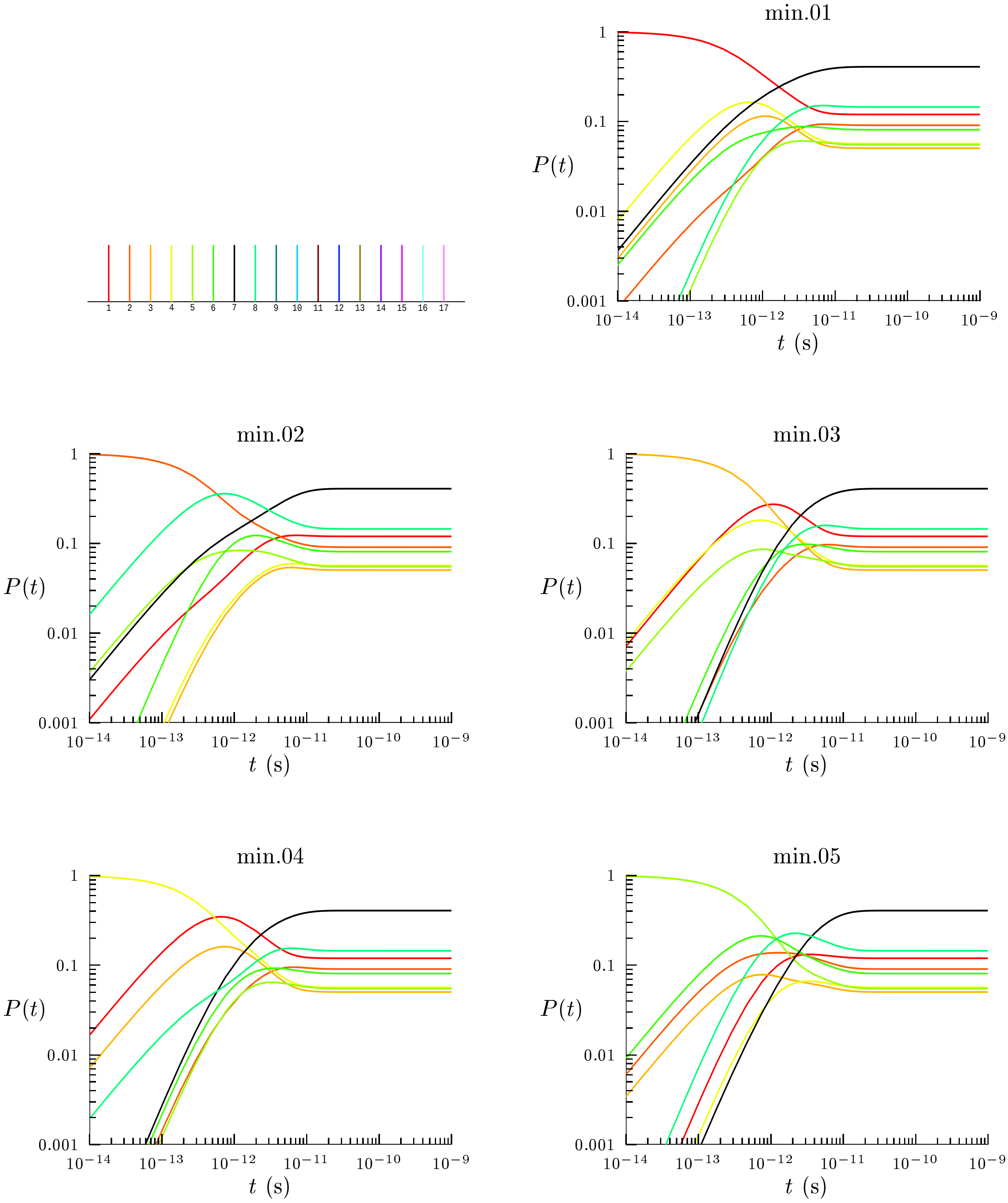

Figure 23: Time evolution of min.01–min.05 for alanine.

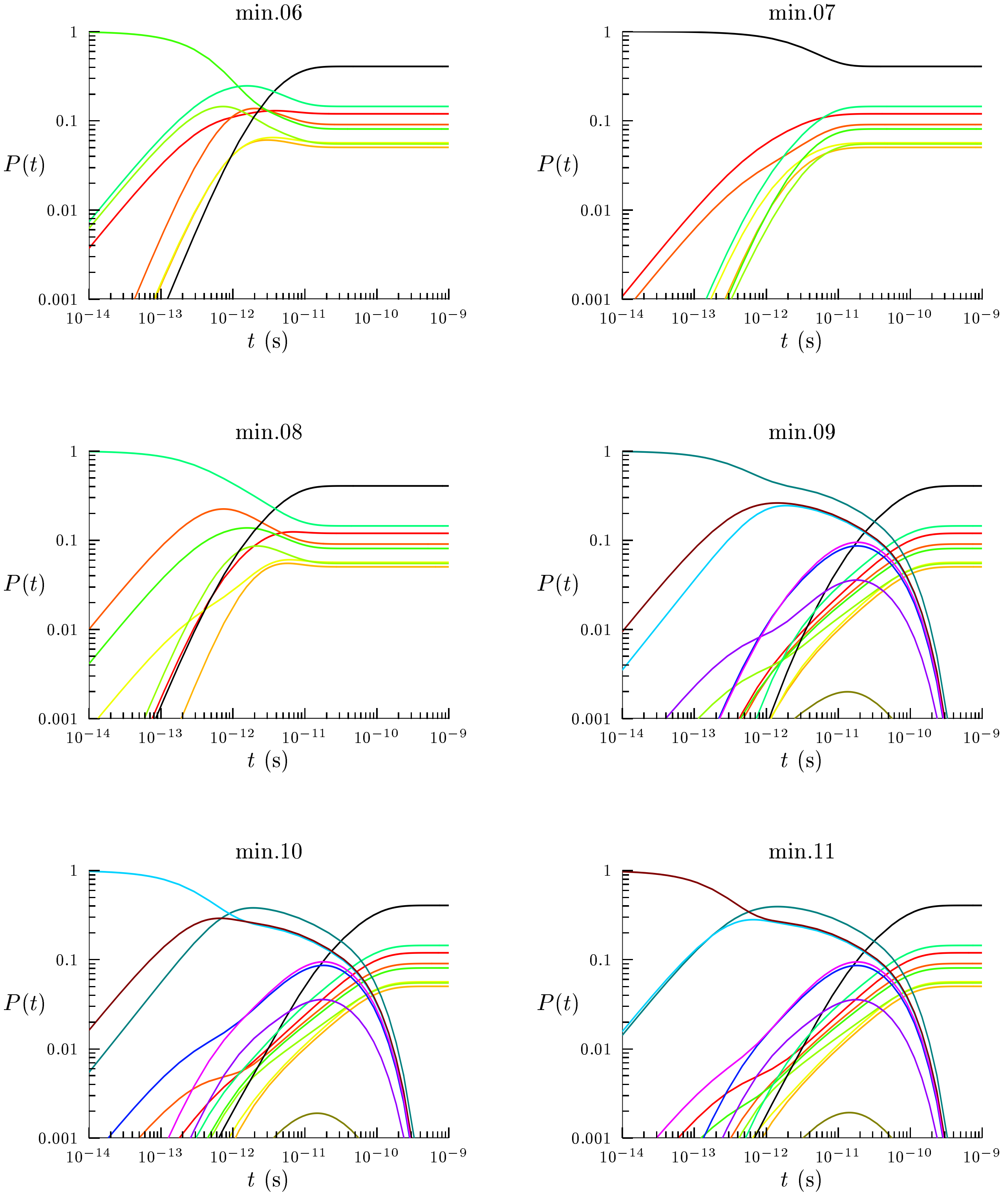

Figure 24: Time evolution of min.06–min.11 for alanine.

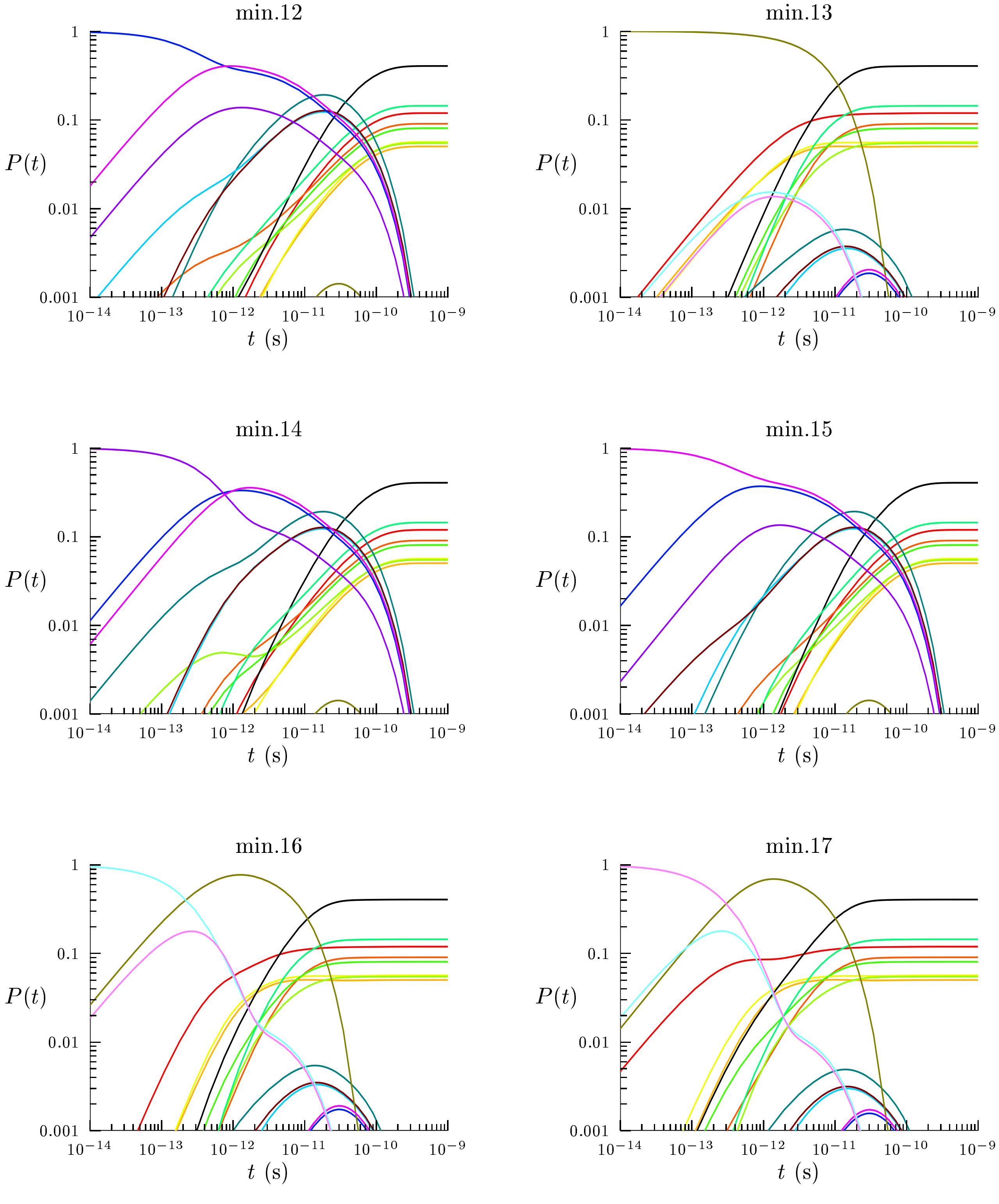

Figure 25: Time evolution of min.12–min.17 for alanine.

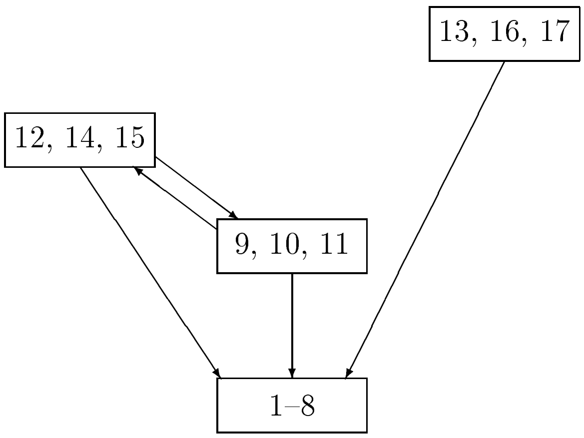

Figure 26: Schematic drawing showing how min.01–min.17 of alanine are connected.

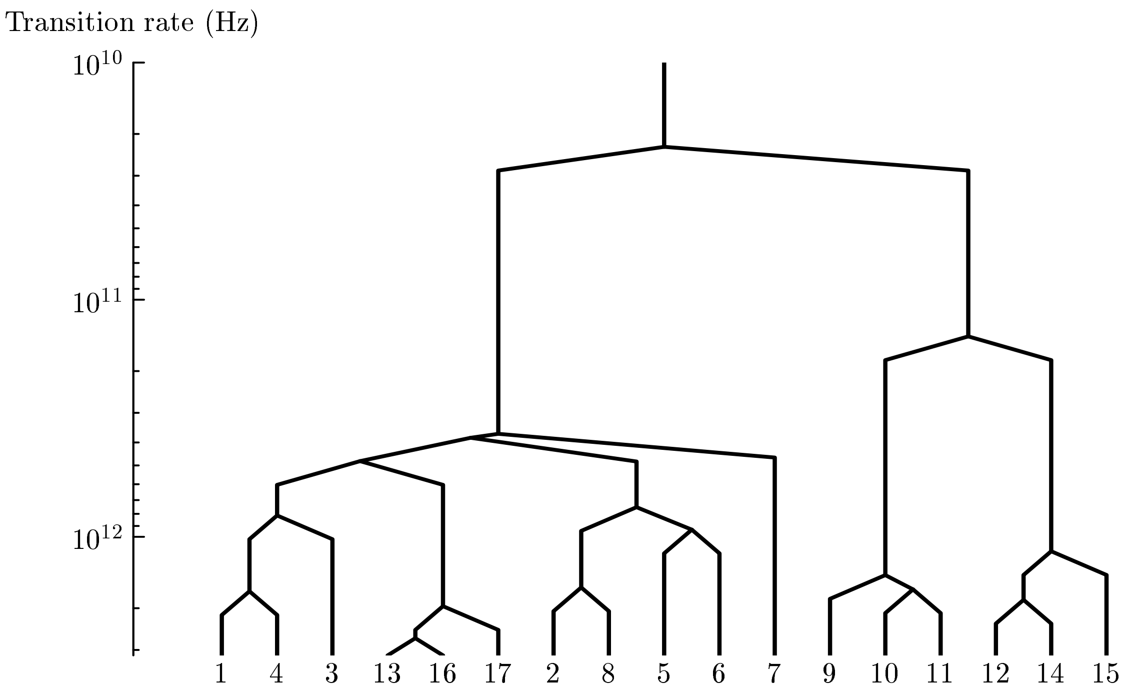

Figure 27: Rate disconnectivity graph for alanine at $T = 300\,\mathrm{K}$.

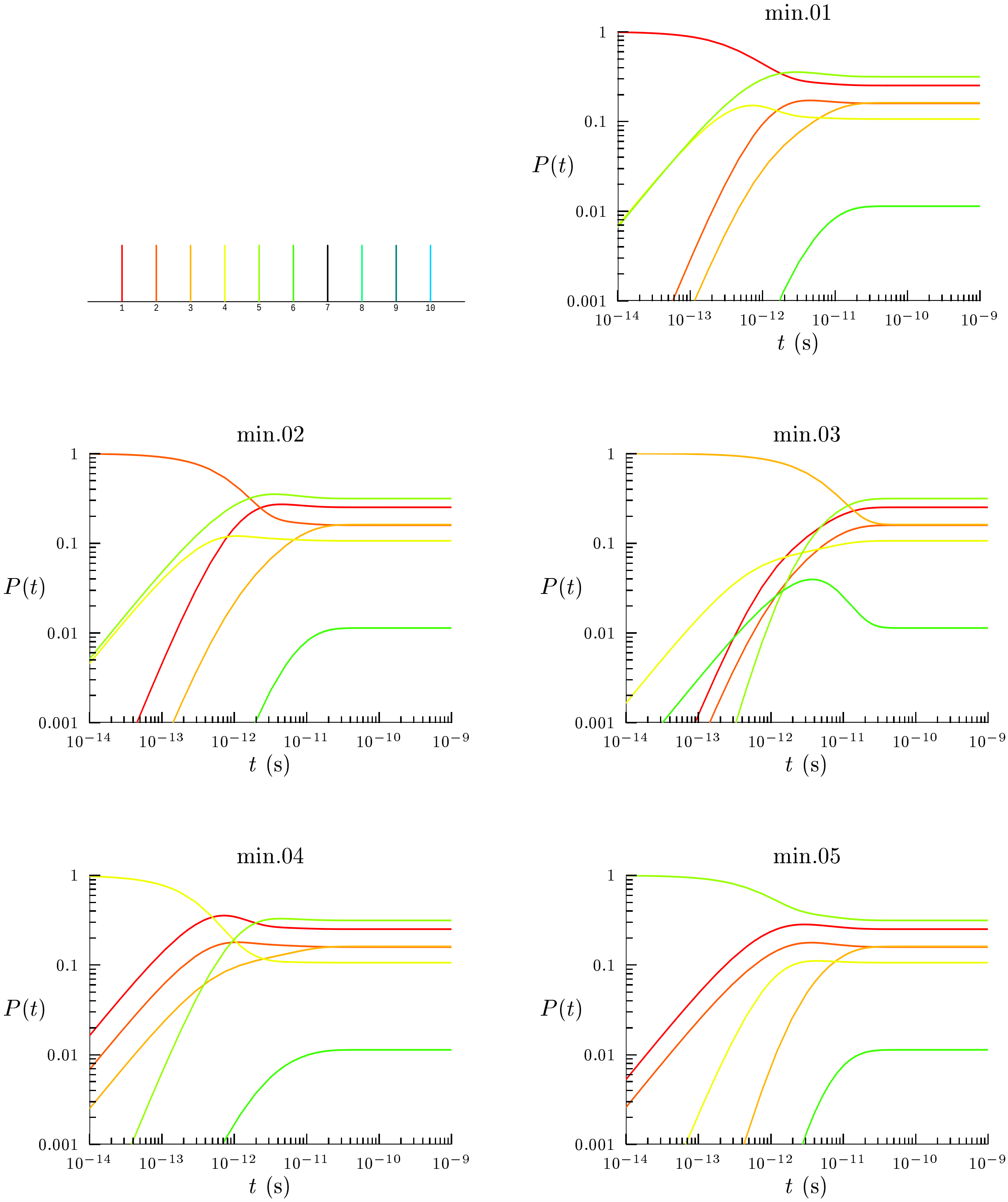

Figure 28: Time evolution of min.01–min.05 for alanine dipeptide

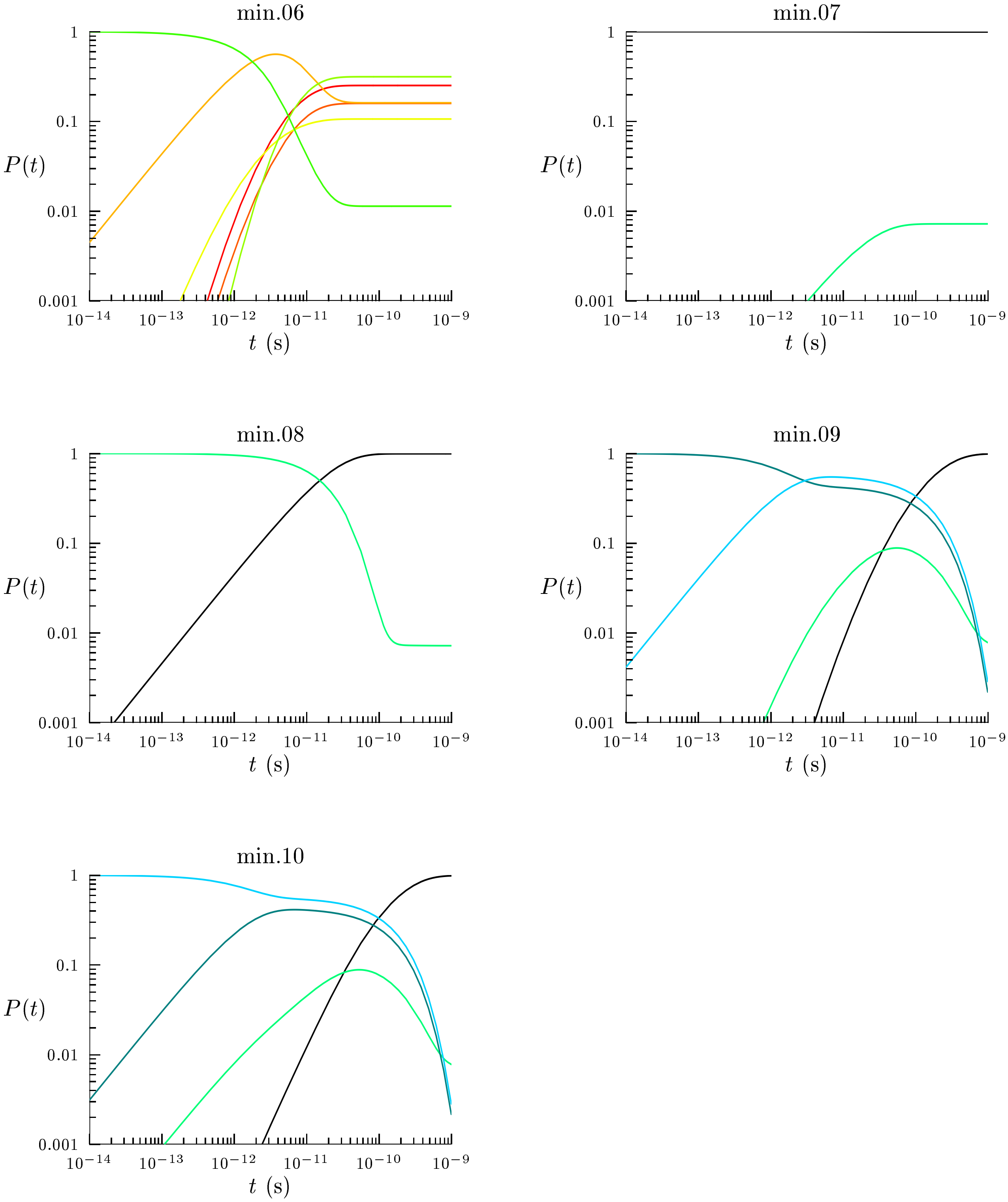

Figure 29: Time evolution of min.06–min.10 for alanine dipeptide

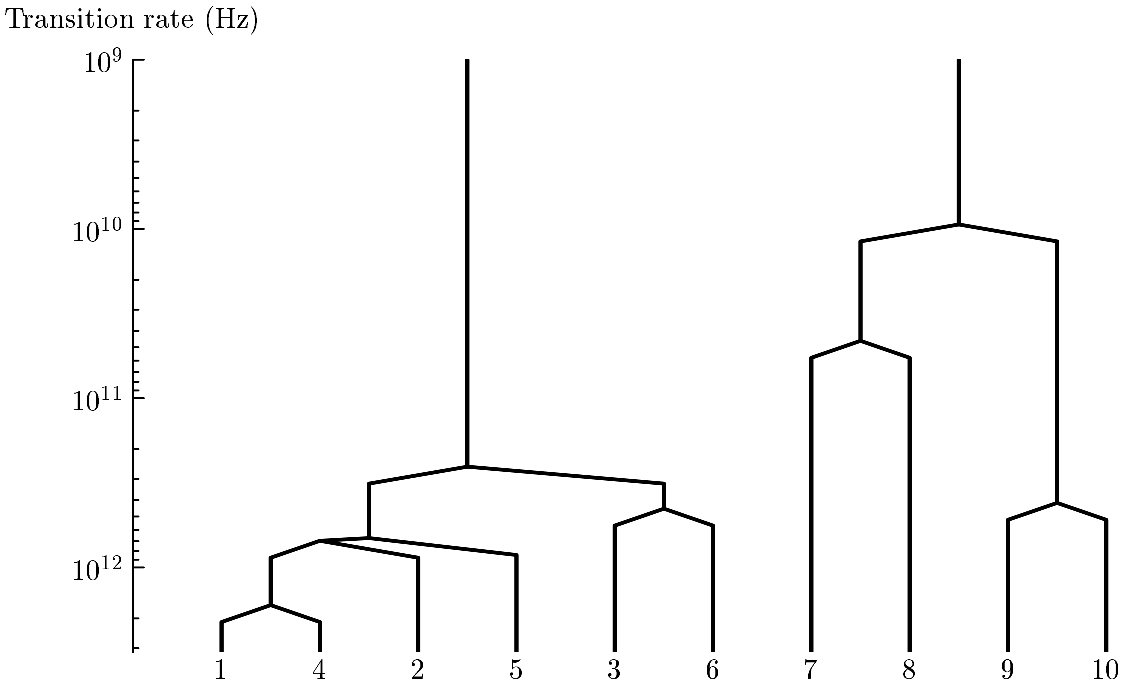

Figure 30: Rate disconnectivity graph for alanine dipeptide at $T = 300\,\mathrm{K}$.

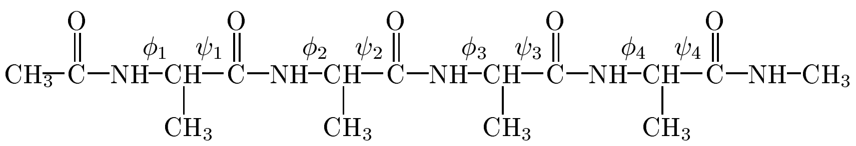

Figure 31: Tetra-alanine

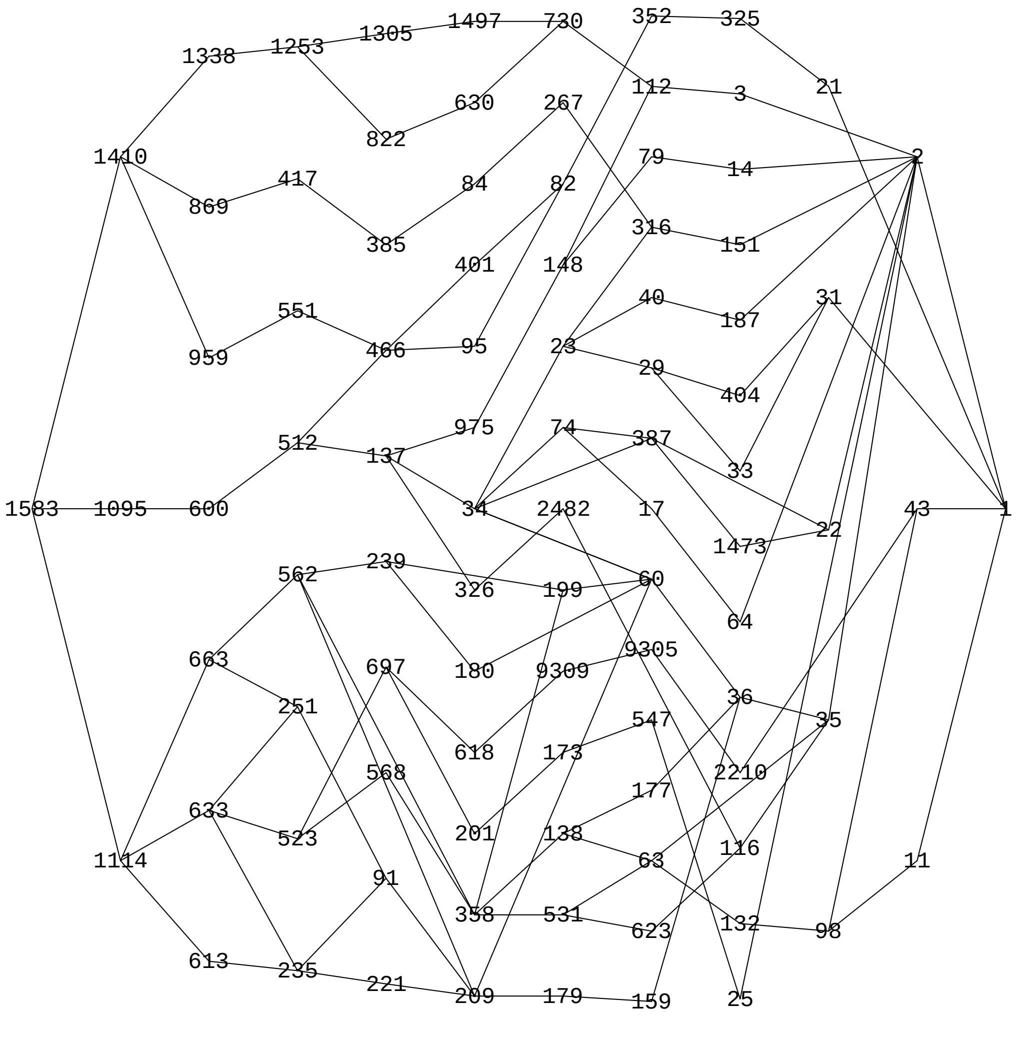

Figure 32: Graphical display of the reaction pathways which are enumerated in Table XXVI.

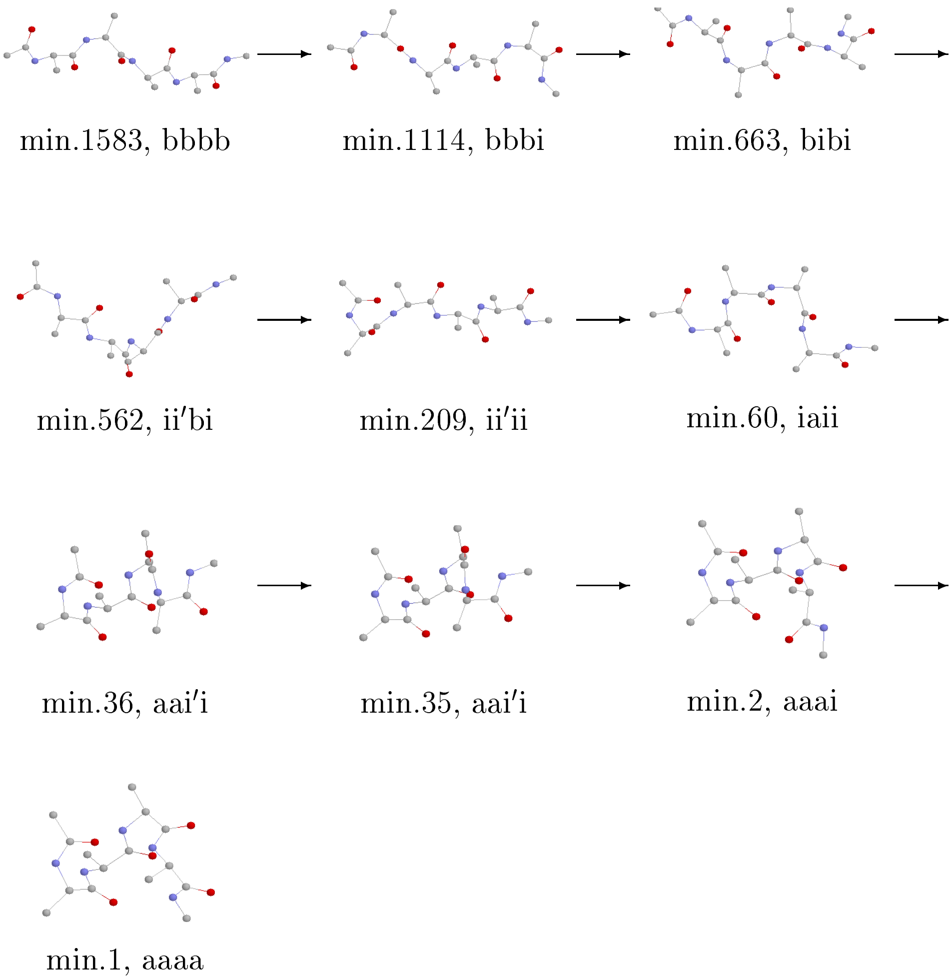

Figure 33: One possible pathway from min.1583 to min.1.

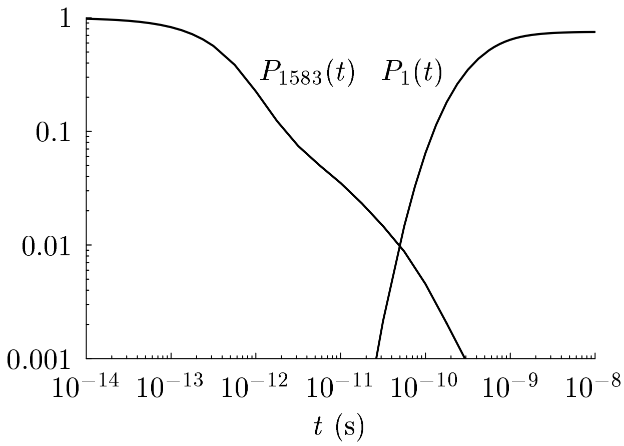

Figure 34: Occupation probabilities for min.1583 and min.1 for tetra-alanine at $T = 300\,\mathrm{K}$, given that the system is in min.1583 initially.

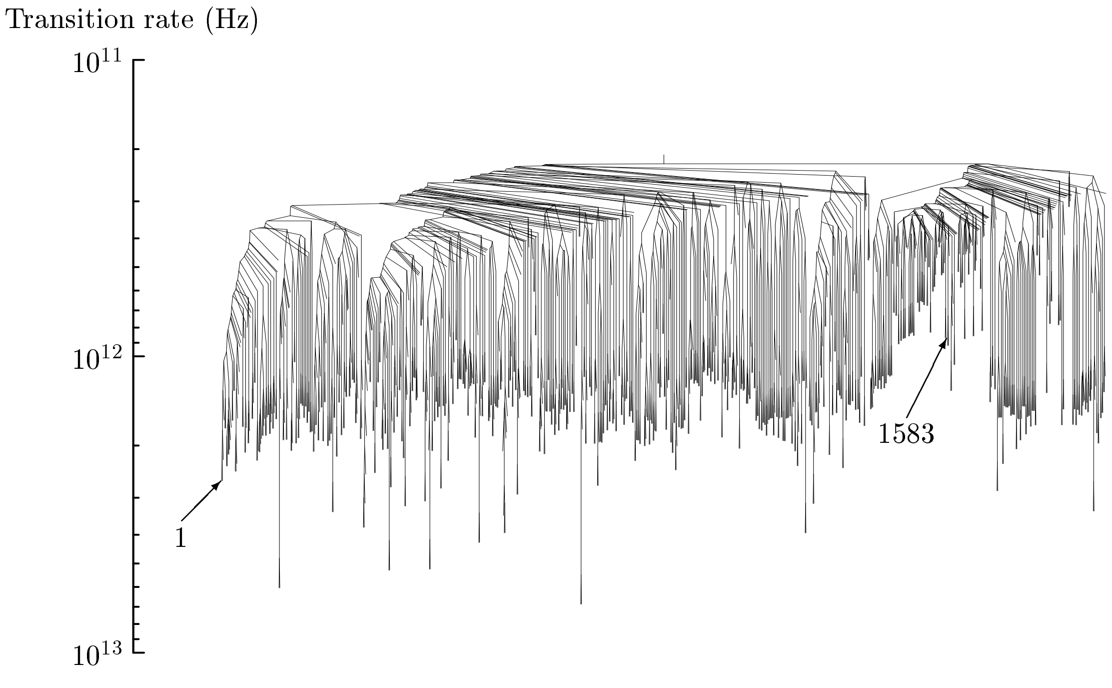

Figure 35: Partial rate disconnectivity graph for tetra-alanine. The root node is the node which connects min.1583 to min.1 in the full rate disconnectivity graph. It contains 1112 minima.

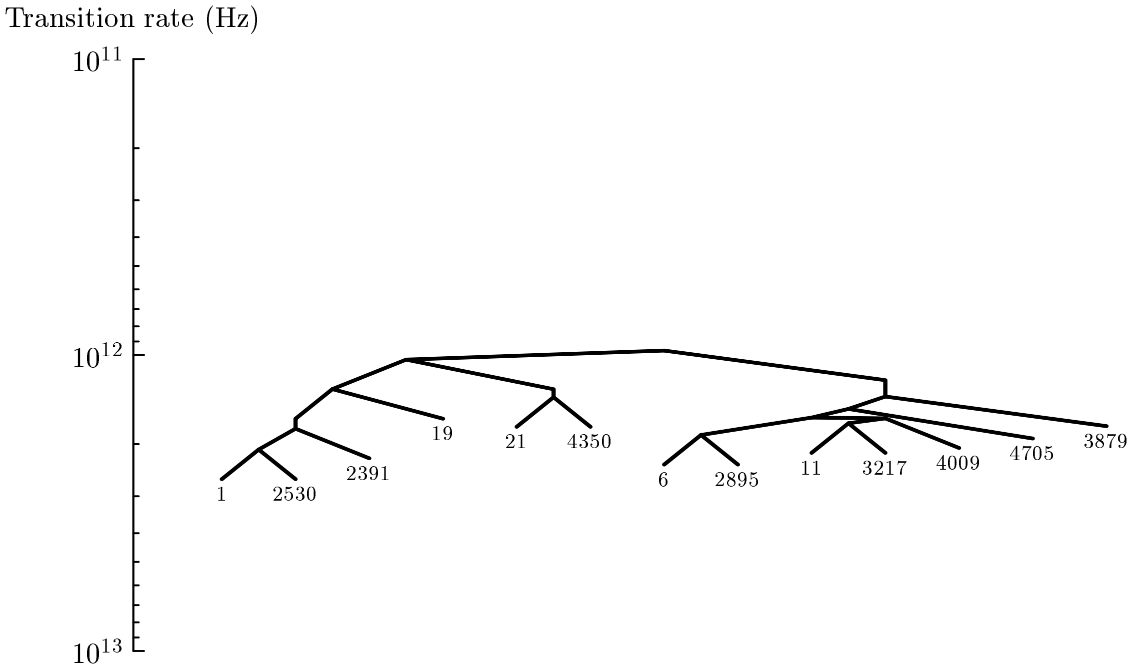

Figure 36: Partial rate disconnectivity graph for tetra-alanine, in the vicinity of min.1.

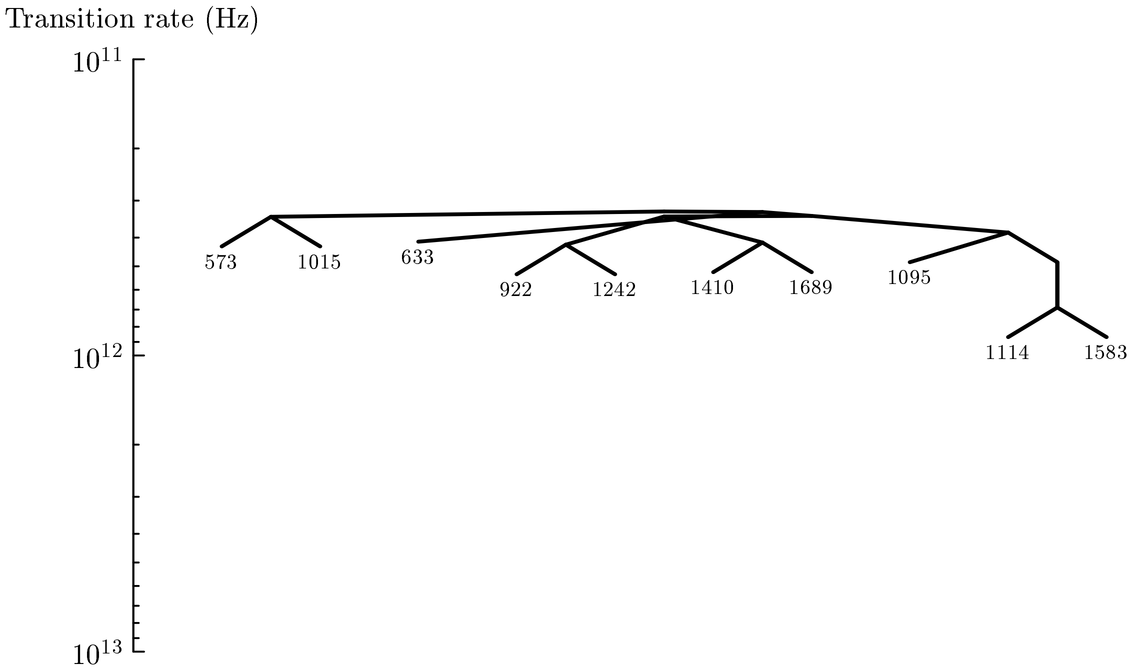

Figure 37: Partial rate disconnectivity graph for tetraalnine, in the vicinity of min.1583.