Optimization in Computational Chemistry and Molecular Biology, pp. 1-2
C. A. Floudas and P. M. Pardalos, Editors

Methodology for Elucidating the Folding Dynamics of Peptides : Met-enkephalin Case Study

J. L. Klepeis *Department of Chemical Engineering*
Princeton University
Princeton, NJ 08544
john@titan.princeton.edu

C. A. Floudas *Department of Chemical Engineering*
Princeton University
Princeton, NJ 08544
floudas@titan.princeton.edu

Abstract

The ability to characterize the energy surface and the reaction pathways of peptides is an important step in understanding the protein folding process. However, this problem may be intractable for realistic systems, such as those represented by all atom models. In this work, a method for mapping relatively low energy regions through the identification of first order transition states is introduced. The methodology relies on the use of a deterministic global optimization approach for identifying low free energy ensembles of conformers [20]. The technique is applied to both unsolvated and solvated forms of the oligopeptide met-enkephalin.

Keywords: Protein folding, transition states, global optimization, enkephalin.

1 Introduction

A fundamental problem in the area of computational chemistry and molecular biology is how a protein folds to its correct conformation in a reasonable time scale. The complexity of this problem is due to a lack of information regarding the folding pathways along the protein's intricate energy surface. Furthermore, it is believed that a random search through all possible configurations is too slow; that is, certain intermediates are essential in defining the correct folding pathway.

To better understand protein folding it is necessary to examine a protein's energy hypersurface. Such examination often begins with the identification of the most stable energy structure – the global minimum energy conformation. This prediction of a protein's tertiary structure from only sequence information, which is the traditional protein folding problem, is, by itself, a hallmark of computational chemistry. A number of reviews on methods for solving this multiple-minima problem are available [15, 31].

The characterization of the energy surface also includes the identification of other stable and metastable configurations. Mathematically, these structures correspond to stationary points of the energy function. In particular, local minima represent stable conformations, while (first order or higher order) saddle points constitute transition states that connect two stable structures. A folding pathway defines the connection between two stable conformations (local minima) through a series of transition states (saddle points).

Since the folding pathway may include a number of intermediates, a rigorous description of the energy surface would require the identification of all local minima and saddle points of the energy function. These ideas have been used to study a flexible, helical forming tetrapeptide [12, 37]. For large molecular systems, such an approach is computationally infeasible. In this work, a new methodology is outlined which maps out low lying energy regions by first locating low (free) energy minima through a deterministic global optimization technique [20]. These minima are then used to initiate searches for first-order transition states. Connectivity between two minima is established by retracing the paths back down from each transition state.

Folding pathways between any two minima can be identified through the combination of minimum-transition-minimum triples. Once the connectivity of the energy surface has been established, transition rates between minima can also be calculated using Rice-Ramsperger-Kassel-Marcus (RRKM) theory [22, 23]. Finally, following the ideas given in [37], a "rate connectivity graph" can be constructed. This graphical representation of the energy surface is based on the transition rates between minima. Such a graph is similar in nature to the connectivity tree and energy disconnectivity graph concept, which uses finite energy (or temperature) partitioning to define basins of attraction [9, 13].

In this work, the approach for finding low (free) energy minima, as proposed in [20], is outlined in Section 3. A detailed description of the methodology for determining transition states and energy surface connectivity is given in Section 4. Finally, in Section 5, the results for the oligopeptide system of met-enkephalin are given. Both unsolvated and solvated forms of the molecule are studied.

2 Modeling

2.1 Potential energy

To provide a detailed description of the protein system, a semi-empirical all atom forcefield is employed. There exist many parameterizations for molecular potential functions, including AMBER [35, 36], CHARMM [10], ECEPP [27, 28, 29], and MM3 [5]. In this work, the ECEPP/3 (Empirical Conformational Energy Program for Peptides) model, which is the latest installment of the ECEPP forcefield, is utilized [30]. For this force field covalent bond lengths and bond angles are fixed at their equilibrium values, which implies that all residues of the same type have essentially the same geometry in various proteins. Based on these approximations, the conformation is only a function of the dihedral angles. That is, ECEPP/3 accounts for energy interaction terms which can be expressed solely in terms of the independent torsional angles.

The total conformational energy is calculated as the sum of the electrostatic, nonbonded, hydrogen bonded, and torsional contributions, as shown in the following equation :

$$
\begin{aligned}
E_{\text{ECEPP/3}} &= \sum_{(i,j)\in ES} \frac{q_i q_j}{r_{ij}} + \sum_{(i,j)\in NB} F_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \\
&+ \sum_{(i,j)\in HB} \frac{A'_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{10}} + \sum_{k\in TOR} \frac{E_k}{2} \left(1 + c_k \cos\left(n_k \theta_k\right)\right)
\end{aligned}
\tag{1}
$$

Here r_{ij} is the distance between atoms i and j. The sets ES, NB, and HB correspond to those pairs of atoms (i, j) which contribute electrostatic, nonbonded and hydrogen bonded energies, respectively. For the electrostatic term, the dielectric constant and conversion factor have been incorporated into the parameters q_i and q_j. A_{ij}, C_{ij}, A'_{ij} and B_{ij} are parameters that define the well depth and width for a given nonbonded or hydrogen-bonded interaction. F_{ij} is equal to 0.5 for 1–4 interactions and 1.0 for higher order interactions. A torsional term is also included for those dihedral angles in the set TOR. θ_k represents any dihedral angle, while c_k takes a value from -1,1, and n_k reflects the symmetry type for the particular dihedral angle. Additional contributions are calculated for special structural features, such as proline rings and disulfide bridges. The main energy contributions are computed as the sum of terms for each atom pair whose interatomic distance is a function of at least one dihedral angle.

2.2 Solvation energy

Solvation energy can be modeled either explicitly or implicitly. To make the computations tractable, a continuum model based on empirical correlations with solvent accessible volumes is employed [8], which is represented by the following equation :

$$
E_{\text{HYD}} = \sum_i (\text{VHS}_i)(\delta_i) \tag{2}
$$

Here VHS_i corresponds to the volume of the hydration shell, while δ_i is the free energy density parameter for the atom or functional group. The main assumption of this model, and any geometrically parameterized solvation model, is that hydration free energies can be calculated from an average free energy of interaction of each atom (or functional group) with a layer of solvent know as the hydration shell. The total free energy of solvation is taken to be the sum of the free energies for each of the atoms (or function groups) in the peptide.

In general, the analytical form for calculating the volume of hydration shell (VHS_i is not suitable for force field models using pairwise intramolecular potential, such as ECEPP/3. Furthermore, direct truncation at the pairwise double–overlap term would lead to large errors. In this work, the RRIGS (Reduced Radius Independent Gaussian Sphere) approximation is used to efficiently calculate the exposed volume of the hydration shell [8]. This method uses a truncated form (double overlap) for VHS$_i$, while also artificially reducing the van der Waals radii of all atoms other than atom i when calculating VHS_i. These reductions effectively decrease the contribution of the double overlap terms, leading to a cancellation of the error that results from neglecting the triple and higher overlap terms. In addition, the characteristic density of being inside the overlap volume of two intersecting spheres is not represented as a step function, but as a Gaussian function; this provides continuous derivatives of the hydration potential. Therefore, the solvation energy contributions can easily be added at every step of local minimizations since the RRIGS approximation has the same set of interactions as the ECEPP/3 potential. The δ_i are RRIGS specific, and were determined by a least square fitting of experimental free energy of solvation data for 140 small organic molecules [8].

2.3 Free energy

A strict interpretation of Anfinsen's thermodynamic hypothesis requires the global minimization of the conformational free energy to predict the native structure of a protein. In practice, however, most protein models include only potential and solvation effects. One reason for this neglect of including entropic effects is that a rigorous free energy model requires infinite sampling to associate accurate statistical weights with each microstates.

Other approximate calculations exist for estimating these statistical weights (and thus entropic effects). The most simplistic model would rely on only the Boltzmann weight associated with each microstate. A more sophisticated approximation, known as the harmonic approximation, utilizes second derivative information to characterize the basin of attraction. More complex schemes try to mimic the anharmonic trajectory along the energy surface. These quasi-harmonic approximations generally require the use of MC/MD simulations.

In this work, entropic effects are included via the harmonic approximation [17, 18, 14]. The development of this model can be understood physically by first considering the partition function for the system :

$$Z = \exp^{-\frac{(E-TS)}{k_B T}} = \exp^{-\frac{E}{k_B T}} \exp^{\frac{S}{k_B}} \tag{3}$$

In Equation 3 the partition function is the product of the Boltzmann factor ($\exp(-E/k_B T)$) and the number of states available to the system ($\exp(S/k_B)$. At a given stationary point, the harmonic approximation is equivalent to :

$$E(\theta) = \mathbf{E}(\theta_\gamma) + \frac{\mathbf{1}}{\mathbf{2}} (\theta - \theta_\gamma)\, \mathbf{H}(\theta_\gamma)\, (\theta - \theta_\gamma) \tag{4}$$

Here γ identifies the stationary point, and the stationarity condition ($\nabla E(\theta_\gamma) = \mathbf{0}$) is used to eliminate the gradient term. In this way each basin of attraction is characterized properties of its corresponding minima, which include the local minimum energy value, $E(\theta_\gamma)$, and the convexity (Hessian) information around the local minimum, $\mathbf{H}(\theta_\gamma)$. An analogous representation of this system is N_θ independent harmonic oscillators, each its own characteristic vibrational frequency. The minimum is then characterized by the occupation of each normal mode.

To develop an expression for the entropic effect, Equation 4 can be substituted into Equation 3. By summing over all energy states, the partition function becomes :

$$Z_\gamma^{\text{har}} = \exp^{-\frac{E(\theta_\gamma)}{k_B T}} f(T) \prod_i^{N_\theta} \frac{1}{\lambda_i} \tag{5}$$

In Equation 5 $f(T)$ is a function dependent only on temperature, while λ_i represent the eigenvalues of $\mathbf{H}(\theta_\gamma)$. Comparison of Equations 5 and 3 implies that :

$$\exp^{\frac{S}{k_B}} \propto \prod_i^{N_\theta} \frac{1}{\lambda_i} \tag{6}$$

Equation 6 can be rewritten in terms of the harmonic entropic contribution, S_γ^{har} :

$$S_\gamma^{\text{har}} \propto -k_B \ln\left[\text{Det}\left(H(\theta_\gamma)\right)\right] \tag{7}$$

A more rigorous derivation of the harmonic approximation leads to the following expression for the harmonic entropy :

$$S_\gamma^{\text{har}} = -\frac{k_B}{2} \ln\left[\text{Det}\left(H(\theta_\gamma)\right)\right] \tag{8}$$

This can be used to calculate relative free energies via the following equation :

$$F_\gamma^{\text{har}} = E(\theta_\gamma) + \frac{\mathbf{k_B T}}{\mathbf{2}} \ln\left[\text{Det}\left(\mathbf{H}(\theta_\gamma)\right)\right] \tag{9}$$

Finally, each microstate can be assigned a statistical weight (p_γ^{har}) by considering the ratio of the partition function at that microstate (Z_γ^{har}) to the total partition function :

$$p_\gamma^{\text{har}} = \frac{\left[\frac{1}{[\text{Det}(H(\theta_\gamma))]}\right]^{1/2} \exp(-\frac{E(\theta_\gamma)}{k_B T})}{\sum_{i=1}^{N} \left[\frac{1}{[\text{Det}(H(\theta_\mathbf{i}))]}\right]^{1/2} \exp(-\frac{E(\theta_\mathbf{i})}{k_B T})} \tag{10}$$

To develop a meaningful comparison of relative free energies, the total partition function must include an adequate ensemble of low-energy local minima, as well as the global minimum energy conformation. Therefore, an efficient method for identifying low energy ensembles must be employed. It should also be noted that the harmonic approximation does not require the explicit inclusion of a contribution based on the density of states because each local minimizer is accounted for only once (in contrast to counting methods).

3 Locating Low (Free) Energy Minima

3.1 Traditional formulation

In its traditional form, the protein folding problem in dihedral angle space is posed as an unconstrained nonconvex global minimization problem with periodic variable bounds. This is represented by the following formulation :

$$\begin{array}{ll} \min\limits_{\theta} & E(\theta) \\ \text{subject to} & -\pi \quad \leq \theta_i \quad \pi, \quad i = 1, \ldots, N_\theta \end{array} \tag{11}$$

Here θ represents the vector of independent torsion angles. The global minimization of E, which may include both potential and solvation energies, requires efficient global optimization methods due to the extremely rugged nature of the energy hypersurface. Along these lines, the αBB deterministic global optimization approach has been found to be particularly efficient in locating the global minimum energy conformations of isolated and solvated peptides [7, 19, 21]. This branch and bound method addresses a broad class of problems and provides theoretical guarantees of convergence to the global minimum of nonlinear optimization problems with twice-differentiable functions [1, 2, 3, 4, 6].

The αBB algorithm can be summarized in the following manner :

1 The global minimum is bracketed by developing converging sequences of lower and upper bounds.

2 The lower and upper bounds are refined by iteratively partitioning the initial domain.

3 Upper bounds on the global minimum are obtain by local minimizations of the original energy function, E.

4 Lower bounds belong to the set of solutions of the convex lower bounding functions, L, which are constructed (in each subdomain) by augmenting E with the addition of separable quadratic terms.

The development of the convex lower bounding functions, L, ensures the deterministic nature of the algorithm and guarantees convergence to the global minimum. These functions have the following generic form :

$$L \quad = \quad E + \sum_{i}^{N_\theta} \alpha_i \left(\theta_i^L - \theta\right) \left(\theta_i^U - \theta\right) \tag{12}$$

In Equation 12, θ_i^L and θ_i^U refer to the current lower and upper bounds on the variable θ_i. For each separable quadratic term, the α_i correspond to non-negative parameters which must be greater or equal to the negative one-half of the minimum eigenvalue of the Hessian of E over the current domain, as defined by $\left[\theta^{\mathbf{L}}, \theta^{\mathbf{U}}\right]$. The properties and proofs that endow the αBB approach with deterministic guarantees are detailed elsewhere [25].

3.2 Ensembles of low energy conformers

To effectively include entropic contributions via the harmonic approximation, the general protein folding formulation, as given by Equation 11, must be modified. Specifically, in addition to locating the global minimum energy conformer, an ensemble of low energy local minima must be generated. Once this ensemble has been compiled, a free energy ranking can be performed (at a particular temperature) to locate the the relative free energy global minimum.

The nature of the αBB algorithm allows for the rigorous treatment of enumerating all local minima. These approaches are discussed in detail elsewhere [20]. In particular, one formulation involves the solution of a system of nonlinear equations that correspond to the stationarity condition, $\nabla E(\theta) = \mathbf{0}$. Through a reformulation, this problem becomes one of identifying all multiple global solutions. This method has been implemented using the αBB, and used to locate all stationary points (local minima and saddle points) of triatomic molecules and small peptides [37].

Given the astronomical number of local minima for even relatively small oligopeptides, this work addresses the problem of identifying low energy, rather than all, local minima. In particular, the generation of low energy ensembles of conformers is accomplished by algorithmic modifications of the general αBB procedure, rather than reformulation of the original problem. Rigorous implementation of the global optimization algorithm requires the minimization of a *convex* lower bounding function, L, in each domain. Each unique solution can then be used as starting points for the minimization of the original energy function in the current subdomain. By storing these minima a list of low energy conformers can be constructed.

A simplistic method for enhancing the distribution of local minima produced by each subdomain would involve the use of multiple random starting points when minimizing E. This method is better than just using random starting points (over the entire domain), since the subdomains become localized in regions of low energy as the separation between the lower and upper bounding sequences decreases.

However, this approach does not capitalize on all the available information provided by the lower bounding functions. In a strict implementation, each lower bounding function exhibits a single minimum in each subdomain. Since the value of α directly defines the convexity of the lower bounding functions, these parameters can be modified to provide nonconvex underestimators. These lower bounding function, $\hat{L}$ will possess more than one minimum and can be minimized several times in each subdomain. Because the lower bounding functions tend to smooth the upper bounding function, the location of these minima also provide important information on the location of low energy minima on the original energy surface. Therefore, the location of the minima of $\hat{L}$ are used as starting points for local minimization of E, which results in an improved set of lower-energy conformers. In addition, these conformations are also localized in regions with low energy as the size of the subdomains decrease. This approach, which will be referred to as the EDA (Energy Directed Approach), is illustrated in Figure 1.

A second approach enhances the search for free energy local minima by accounting for entropic contributions during the course of the branch and bound algorithm. This is accomplished by calculating and including the relative entropic values at each minima of the upper and lower bounding functions, as shown in the following equation :

$$F^{\mathrm{har}} \quad = \quad U(\theta_{\mathrm{min}}) + \frac{\mathbf{k_B T}}{\mathbf{2}} \ln\left[\mathrm{Det}\left(\mathbf{H}(\theta_{\mathrm{min}})\right)\right] \tag{13}$$

The expression is similar to Equation 9 except that $U(\theta_{\mathrm{min}})$ now represents either the local minima of either E or L. The thermodynamic temperature used in Equation 13 must also be specified as an additional input parameter. A similar modification for Monte Carlo minimization has also been attempted [26].

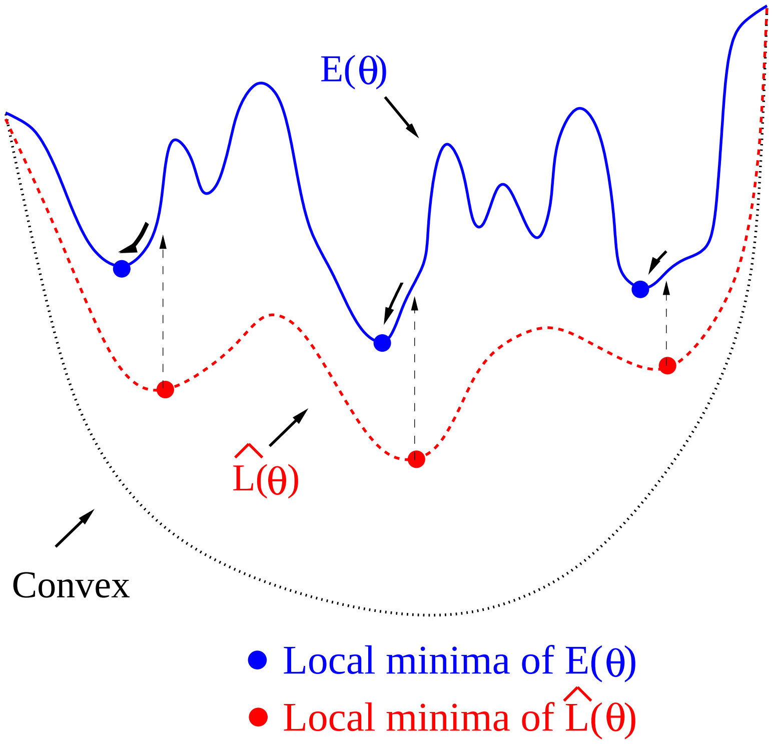


Figure 1: Using multiple minima of the lower bounding function ($\hat{L}$) to find low energy conformers of the upper bounding function (E).

This approach, which will be referred to as the FEDA (Free Energy Directed Approach), can be used to deterministically identify the global minimum free energy conformer at a given temperature. However, since the goal is to identify an ensemble of low (free) energy conformations, the FEDA is implemented using the same modification of the lower bounding functions, $\hat{L}$.

3.3 Algorithmic description

The αBB methodology for locating ensembles of low energy conformers requires interfaces between αBB and several programs. These include the potential and solvation energy modules, ECEPP/3 [30] and RRIGS [8], respectively; a local nonlinear optimization solve, NPSOL [16], which is used to locally solve both the upper and lower bounding problems; and UBC, and upper bound check module that is used to verify the quality of the upper bound solutions.

The algorithmic steps are summarized in the following :

1 Initialize the best upper bound to $+\infty$.

2 Initialize α values for all global variables (variables that will be partitioned). All other variables are treated locally within the periodic interval.

3 If FEDA, specify thermodynamic temperature.

4 Partition the domain along one of the global variables.

5 Construct lower bounding function ($\hat{L}$) in each subdomain. Perform three local minimizations according to the following procedure :

   A Generate 50 random points and perform function evaluations of $\hat{L}$.

   B Use the point with the minimum value as a starting point for local minimization of $\hat{L}$.

   C If FEDA, calculate and add $-TS^{\mathrm{har}}$.

   D Store unique solutions.

   If the minimum valued solution (of $\hat{L}$ in this subdomain) is greater than the current best upper bound, the region is fathomed (discarded from list).

6 The unique local minima of $\hat{L}$ are used as starting points for the minimization of the upper bounding function E in each subdomain. Perform two additional local minimizations according to the following procedure :

   A Generate 50 random points and perform function evaluations of E.

   B Use the point with the minimum value as a starting point for local minimization of E.

   C Check gradient and Hessian conditions (to verify point is local minimum).

   D If FEDA, calculate and add $-TS^{\mathrm{har}}$.

   Update the current best upper bound as the minimum of all upper bounds, including those generated in the current subdomains.

7 Select the subdomain with the lowest lower bound value for further partitioning.

8 Update α values.

9 If the best upper and lower bounds are within the specified tolerance the program will terminate; otherwise it will return to step 4.

4 Identifying Folding Pathways

4.1 Illustrative example

To rigorously characterize the energy surface, and thus the folding pathways, of a given protein, the full set of minima and transition states must be identified. Although this has been attempted for small peptides [37], the extension of such techniques to atomistic level models of oligopeptides is currently intractable. To circumvent these computational limitations, a methodology is developed that builds on the concepts outlined in the previous section.

In particular, the ability for the EDA and FEDA to identify ensembles of low energy minima provides a propitious starting point for further examination of the energy hypersurface. These low energy and free energy minima are used to initiate local searches for first-order transition states. These transition states can then be used to identify minimum-transition-minimum triplets. At the very least, this will provide a robust mapping of low energy regions, which can be enhanced by increasing the number of low energy minima used to initiate transition state searches.

The usefulness of this approach can be illustrated by a small test case. In Figure 2 a two-dimensional contour plot for unsolvated alanine (NH_2–CH–CH_3–C=O–OH) is given. Both the χ and ω dihedral angles are fixed at optimal values. The energy surface exhibits 7 local minima and 12 transition states. When starting a transition state search (and triplet search) with the 3 lowest energy minima, all 7 local minima and 9 of the 12 transition states are identified. The number of found transition states can be increased to 11 (of 12) by initiating the search with 4 low energy minima.

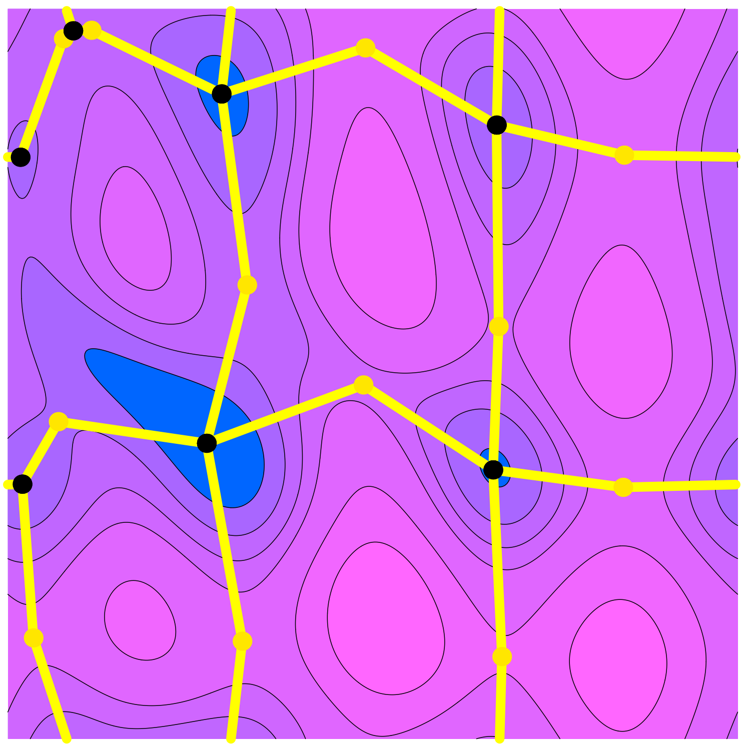


Figure 2: Energy contour plot for unsolvated alanine. The black circles denote local minima, while the grey circles denote transition states.

4.2 Eigenmode following

To implement the transition states search, an efficient method for locating saddle points is essential. In particular, eigenmode methods have been shown to be powerful methods for locating both minima and saddle points [11, 32, 33, 34].

The eigenmode class of local solvers utilize second derivative information and are basically modified implementations of the Newton-Raphson method for locating stationary points. This is best illustrated by first examining the Newton-Raphson step :

$$\Delta\theta = -\mathbf{H}^{-1}\mathbf{g} \tag{14}$$

Here $\mathbf{g}$ and $\mathbf{H}$ are the gradient vector and Hessian matrix at a point on the energy surface. Equation 14 can also be written in terms of eigenvectors and eigenvalues :

$$\Delta\theta = -\sum_{\mathbf{i}}^{\mathbf{N}_\theta} \frac{\mathbf{g_i e_i}}{\lambda_{\mathbf{i}}} \tag{15}$$

Here g_i is the gradient vector along the e_i eigenmode, and λ_i is the corresponding eigenvalue. In actuality, the Newton-Raphson method is only guaranteed to converge to a local minimum when the point is already in the vicinity of a local minimum; that is, all eigenmodes have positive eigenvalues. This is because eigenmodes with positive eigenvalues lead to lower energy, while those with negative eigenvalues will tend toward higher energy. For this reason, the Newton-Raphson search can not be directed towards a particular type of stationary point.

In contrast, the eigenmode methods attempt to influence the step by shifting the eigenvalue. A simple eigenmode implementation which guarantees convergence to a minimum is to use the absolute value of the eigenvalues in Equation 15. The general form for the eigenmode step is given by the following :

$$\Delta\theta = -\sum_{\mathbf{i}}^{\mathbf{N}_\theta} \frac{\mathbf{g_i e_i}}{\lambda_{\mathbf{i}} - \mathbf{L_i}} \tag{16}$$

The eigenvalues, λ_i, are now shifted by the parameters L_i. By controlling the signs of the $\lambda_i - L_i$ expressions, the eigenmode algorithm can be made to converge to stationary points that match the sign of $\lambda_i - L_i$. For example, if one $\lambda_i - L_i$ is enforced to be negative, the algorithm will converge to a first order saddle point. In contrast to Newton-Raphson, eigenmode methods can be used to locate specific classes of stationary points, regardless of the eigenvalues at the starting point. In this particular implementation, these methods can be used to step away from local minima and converge to first order transition states.

A number of formulations have been proposed to determine the shift parameters. In this work the method proposed in [34] is adopted. The rules for choosing the shift parameters depend on the role of the particular eigenmode. If the eigenmode is selected for minimization (decrease in energy), the following rules apply :

$$L_i = \begin{cases} 0 & \text{if} \quad \lambda_i > 0 \\ 2\lambda_i & \text{if} \quad \lambda_i < 0 \end{cases} \tag{17}$$

These conditions provide $\lambda_i - L_I > 0$ as required for minimization. In contrast, when the eigenmode is selected for maximization, the following rules are used :

$$L_i = \begin{cases} \frac{1}{2}(\lambda_i + \sqrt{\lambda_i^2 + 4g_i^2}) & \text{if} \quad \lambda_i > 0 \\ 0 & \text{if} \quad \lambda_i < 0 \end{cases} \tag{18}$$

When $\lambda_i > 0$, the direction of movement is reversed using an empirical shift parameter.

In this work, the eigenmode method is used to perform $2N_\theta$ first order transition states searches from each local minimum. These searches correspond to an initial step away from the minimum in both the positive and negative directions for each eigenmode of the system (low and high). Once the initial step is taken, the particular eigenmode (for maximization) is chosen by checking the largest overlap with the previously followed eigenmode. This procedure is called the eigenmode following method. Another approach (not used here) would be to select the eigenmode for maximization based on the lowest eigenvalue (least eigenvalue method). Each first order transition state can be linked to the two local minima it connects by perturbing the system away from the saddle point along the eigenmode corresponding to the negative eigenvalue (positive and negative step). This provides the minimum-transition-minimum triples that define the connectivity of the energy surface.

4.3 Transition rates

Once the triples are identified, rates of transitions can be calculated using Rice–Ramsperger–Kassel–Marcus (RRKM) theory. If the system is initially at minimum A, the probability that a transition to minimum B will occur (through transition state TS) is proportional to the ratio of the partition functions at TS and A. The rate for this transition is equal to :

$$W_{\mathrm{A}\to\mathrm{TS}\to\mathrm{B}} = \frac{kT}{h}\frac{Z_{\mathrm{TS}}}{Z_{\mathrm{A}}} \tag{19}$$

$\frac{kT}{h}$ is a temperature dependent factor that sets the average rate for transitions (at a given temperature). The ratio of partition functions Z_{TS} and Z_{A} represents the probability that the transition state can be attained. This implies that once the transition state is encountered via thermal fluctuations from minimum A, the transition to minimum B is essentially complete. If the harmonic approximation is again employed (as in Equation 5), Equation 19 becomes :

$$W_{\mathrm{A}\to\mathrm{TS}\to\mathrm{B}} = \frac{\prod_i^{N_\theta} f_{i,\mathrm{A}}}{\prod_i^{N_\theta - 1} f_{i,\mathrm{TS}}} \exp\left[-\frac{(E_{\mathrm{TS}} - E_{\mathrm{A}})}{kT}\right] \tag{20}$$

Here E_{TS} and E_{A} are the energies of the system at the transition state and minimum A, respectively. The f_i represent the vibrational frequencies of the protein, which can be obtained through the solution of the following problem :

$$\left(H - (2\pi f)^2 I\right)\theta = 0 \tag{21}$$

As before, H is the second derivative matrix of the energy function. I now represents the true inertia tensor of the molecular system.

It is important to note that the product of frequencies for the transition state (in the denominator of Equation 20) involve one less mode than the for a local minimum. In particular, this mode is the eigenmode corresponding to the negative eigenvalue and represents the reaction coordinate for the system. For this reason the factors $\frac{kT}{h}$ (one from Equation 19 and one from the extra vibrational frequency for minimum A) negate each other.

A similar relationship can be written for the reverse transition. That is, the rate of going from minimum B through transition state TS to minimum A is equal to the following :

$$W_{\mathrm{B}\to\mathrm{TS}\to\mathrm{A}} \quad = \quad \frac{\prod_{i}^{N_\theta} f_{i,\mathrm{B}}}{\prod_{i}^{N_\theta - 1} f_{i,\mathrm{TS}}} \exp\left[-\frac{(E_{\mathrm{TS}} - E_{\mathrm{B}})}{kT}\right] \tag{22}$$

In addition, the transition rate between any two minima can be calculated by summing the rates for all transition along the pathway(s) connecting these two minima.

4.4 Rate connectivity

To better analyze the connectivity of the minima a "rate disconnectivity graph", as proposed in [37], can also be constructed. This graphical description of the energy surface stems from the concepts originally given in [12], which introduce the idea of an energy tree derived from energy barriers between minima. The ideas were later refined to include a generalization for graphically representing the clustering of minima for finite energy or temperature scales [9]. This "disconnectivity graph" shows how disconnected minima merge as energy or temperature increases.

In [37] this concept was taken a step further by defining a "rate disconnectivity graph", a graphical representation based on transition rates instead of energy barriers. Here the vertical axis represents the transition rate cutoff. If all minima are connected, there will be a frequency low enough so that all minima are grouped under a single node. That is, transitions exist which connect all minima in the graph. The transitions with the lowest frequency are connected near the top of the graph. As the cutoff frequency is increased, certain transitions are eliminated and the graph begins to bifurcate.

Once bifurcation occurs, the graph can be used to check the connectivity between two minima. Minima within the same group are connected by a transition pathway, while minima from different groups are not connected. The maximum point (vertically) along the pathway between two minima identifies the slowest transition. As the rate continues to increase, fewer transitions remain connected and a finer partitioning will appear. Eventually, at a frequency above the highest transition rate, all minima will define their own group.

5 Computational Studies : Met-Enkephalin

Met-enkephalin (H–Tyr–Gly–Gly–Phe–Met–OH) is an endogenous opioid molecule found in the human brain, pituitary, and peripheral tissues and is involved in a variety of physiological processes. The peptide consists of 24 independent torsional angles and a total of

75 atoms, and has played the role of a benchmark molecular conformation problem. All 24 dihedral angles were considered variable, with the 10 dihedral angles of the backbone residues acting as global variables (variables on which branching occurs). Both approaches (EDA and FEDA) for finding low energy conformers were applied to the isolated form of met-enkephalin using the ECEPP/3 forcefield. The EDA was also applied to the RRIGS + ECEPP/3 solvated form of this pentapeptide.

A detailed presentation of the results, including the derivation of thermodynamic quantities and identification of a folding transition, can be found in [20]. In this work the unsolvated FEDA results, with 10 independent runs and initial temperature values varying from 50 to 500 K in 50 K increments, and the solvated EDA results are utilized. The ensemble of local minima generated by these runs are used to identify a set of distinct conformers by checking for repeated and symmetric conformations. An additional criterion, which requires that at least one dihedral angle varies by more the 50^o when comparing each pair of structures, is imposed. The unique conformations are then used to generate ranked distributions according to energy and free energy values. The density of unique conformers is then determined by separating the energy ranking into discrete 0.5 kcal/mol energy bins relative to the global minimum energy structure. The density of states is found to follow a Boltzmann-like distribution.

In the case of isolated met-enkephalin, the 10 FEDA runs generated a total of 87 974 distinct local minima. The potential energy ground state conformation for met-enkephalin consists of a type II' β–bend along the N–C' peptidic bond of Gly^3 and Phe^4, with an energy of -11.707 kcal/mol. The 10 solvated EDA runs produced 72 784 distinct minima. In this case, the ground state exhibits a more extended backbone than the unsolvated form, although the aromatic rings are still proximate. When considering the harmonic free energy, the prediction of the free energy minimum (FEM) can be calculated over a range of temperatures. Table 1 compares the FEM structures at 300 K to the ground state structures for both the unsolvated and solvated systems.

The unsolvated results indicate that the FEGM structure possesses a potential energy contribution 1.808 kcal/mol higher than the ground state structure. The change in dihedral angle values, especially those for the ϕ and ψ angles, also point to an important structural variation between these conformations. In fact, the FEGM conformation exhibits central residues within the α helical region of the Ramachandran plot; which is a significant change from the type II' β turn of the ground state structure. For the solvated form, the relative energy difference between the FEGM and ground state is even larger than for the unsolvated case. In addition, the central residue for this conformation are essentially fully extended – even more so than the ground state structure. These results emphasize the importance that entropic effects have in defining the free energy of the system.

The qualitative effect of adding entropic contributions can be determined by comparing the distribution of distinct minima for the original energy and the free energy (300 K) of the system. Table 2 provides the number of distinct minima in 0.5 kcal/mol (free) energy bins within 8.0 kcal/mol of the FEGM. As previously mentioned, the distributions are Boltzmann-like, with the distributions becoming more dense at 300 K. For the solvated case this increased density is especially dramatic. These results indicate that, as temperature increases, the relative stability of entropic contributions offset substantial differences in energy. At large bin number the number of minima begins to decrease, especially for the

Table 1: Dihedral angle values for the ground state and free energy global minimum (FEGM at 300 K) structures of met-enkephalin. Columns 3 and 4 present the results for the unsolvated form, while columns 5 and 6 are for solvated met-enkephalin. The last two rows indicate the energy and harmonic free energy values (kcal/mol), respectively.

		Unsolvated		Solvated	
Residue	DA	Ground	FEGM (300 K)	Ground	FEGM (300 K)
Tyr_1	ϕ	-83.4	179.8	-168.2	-168.4
	ψ	155.8	-18.2	-30.9	-34.3
	ω	-177.1	-178.1	178.6	-178.9
	χ_1	-173.2	178.2	-173.5	178.7
	χ_2	79.3	81.3	-100.9	-100.8
	χ_3	-166.3	177.3	19.3	179.0
Gly_2	ϕ	-154.3	-59.8	78.5	177.8
	ψ	85.8	-37.6	-86.5	-179.9
	ω	168.5	-178.8	-177.3	180.0
Gly_3	ϕ	83.0	-67.0	162.4	-180.0
	ψ	-75.0	-40.1	92.2	179.9
	ω	-170.0	179.7	172.6	179.7
Phe_4	ϕ	-136.9	-70.9	-150.3	-155.3
	ψ	19.1	-39.5	159.8	147.2
	ω	-174.1	-179.8	-178.1	-176.8
	χ_1	58.9	173.9	65.8	-179.5
	χ_2	94.5	-102.6	-87.4	-111.7
Met_5	ϕ	-163.5	-161.0	-75.0	-78.7
	ψ	160.9	122.1	113.9	-51.1
	ω	-179.8	-178.0	-178.4	179.7
	χ_1	52.9	-174.7	-172.3	-67.2
	χ_2	175.3	174.0	176.1	-178.8
	χ_3	-179.9	179.0	-180.0	-179.9
	χ_4	-178.6	-60.1	60.0	-180.0
F^{har}		-11.707	14.175	-50.060	-28.604
E		-11.707	-9.899	-50.060	-46.030

Table 2: Number of distinct minima in bins for isolated met-enkephalin using FEDA. Each bin represents a 0.5 kcal/mol range above the previous bin. The temperatures are given in the first row.

	Unsolvated		Solvated	
Bin	0 K	300 K	0 K	300 K
1	2	5	10	19
2	3	16	14	98
3	12	42	34	378
4	46	97	117	885
5	47	208	326	1730
6	87	403	717	2812
7	161	846	1440	4451
8	297	1524	2611	5390
9	543	2597	3891	6301
10	828	4032	5567	6736
11	1066	5726	6677	6675
12	1527	7499	7624	6295
13	2244	9315	7650	5756
14	2818	10862	7047	5113
15	3657	12004	6375	4361
16	4472	12167	5534	3437

solvated case. This is most likely due to inadequate sampling of high energy minima, which would fill these free energy bins.

It is also useful to calculate relative free energies for clusters of low energy conformers because it is impossible to determine the population of a given structure class based on a pointwise approximation of entropic effects for individual structures. By clustering conformers into classes, the error associated with the harmonic approximation should also be reduced. In this work, structures are grouped according to the Zimmerman codes for the central residues of the peptide [38]. In particular, two conformers belong to the same cluster if the central three residues possess the same Zimmerman code. The relative free energy of a cluster is defined as :

$$F_{\text{cluster}} = -k_B T \ln \left[\sum_{i \in C} p_i^{\text{har}} \right] \qquad (23)$$

Here, the individual p_i^{har} are calculated by Equation 10, and the set C defines conformations with the same structural classification. A reference free energy is also used to normalize the probabilities. The results for unsolvated and solvated met-enkephalin at 300 K are shown in Table 3.

In general, the structural characteristics of the cluster with the lowest free energy are not necessarily the same as that of the FEGM. The results for unsolvated met-enkephalin indicate that the CD*A class of structures are dominant at 300 K, in part due to the large number of low energy structures with this classification. This group of structures are similar

Table 3: Clustered relative free energies for met-enkephalin. The information provided in this table includes : Zimmerman code[a], number of individual conformers in cluster, total probability ($\sum p_i^{\text{har}}$) and free energy of cluster (F_{cluster}).

Unsolvated				Solvated			
Code	Number	$\sum p_i^{\text{har}}$	F_{cluster}	Code	Number	$\sum p_i^{\text{har}}$	F_{cluster}
CD*A	2128	0.263	0.796	E*EE	148	0.0474	1.818
C*DE	1360	0.125	1.239	EE*E	152	0.0445	1.856
AAA	327	0.111	1.309	D*E*E	149	0.0273	2.147

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

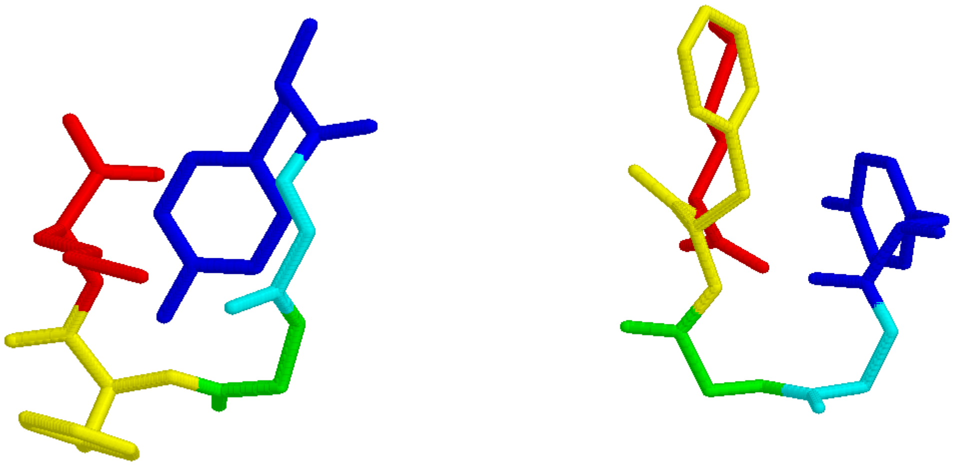
to the ground state conformation, except that the type II' β-bend has shifted from the Gly^3-Phe^4 to the Gly^2-Gly^3 backbone region. Figure 3 compares the ground state structure to the lowest free energy structure within the CD*A cluster. In the case of solvated met-enkephalin, the FEGM structure actually belongs to the dominant cluster type (E*EE). Figure 4 illustrates the differences between the extended FEGM structure class and the ground state conformer.

Based on the results for locating low energy minima, it is intractable to identify the full set of stationary points for the met-enkephalin system. In fact, the number of local minima for met-enkephalin is estimated to be on the order of 10^{11} [24]). Therefore, this system provides an ideal case for testing the methodology for mapping the energy surface and folding pathways of peptides.

To meaningfully characterize the energy surface, an obvious choice is to focus the analysis on connectivity to the ground state conformation. In addition, transitions between the ground state structure and representative structures of the dominant free energy cluster at 300 K were selected for folding pathway determination. For these reasons, the set of low energy conformers used to initialize the transition state searches are based on the union of the 1000 lowest energy minima and the 1000 lowest free energy minima at 300 K. In this way, the connectivity between low free energy structures and the ground state structure could be explored, while maintaining a tractable number of starting structures.

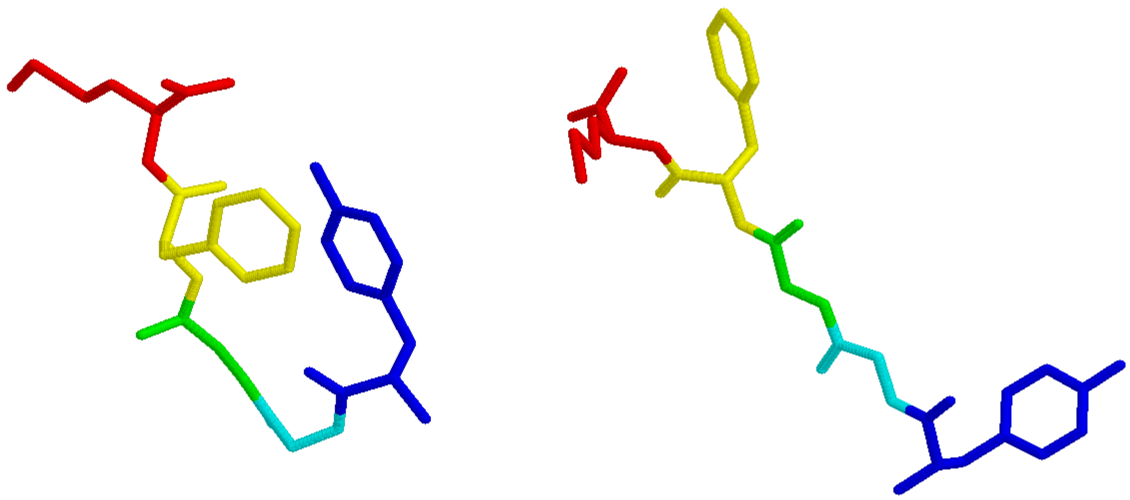
For each unique local minima the transition state search entailed 48 eigenmode following runs; that is, 2 initial step directions for all 24 unique eigenmodes. These searches were implemented in parallel using MPI on NPACI's HP Exemplar. Following these runs, the set of unique first order transition states were identified and ranked. For each transition state, the next step involved the determination of the minimum-transition-minimum connectivity, which required 2 additional eigenmode searches corresponding to the positive and negative step along the reaction coordinate. For unsolvated met-enkephalin, this procedure resulted in the identification of 51272 total stationary points, with 22775 local minima and 28497 first order transition states. The corresponding results for solvated met-enkephalin were 76828 total stationary points, 34722 minima and 42106 first order transition states.

To quantify the connectivity to the ground state, the transition rates were calculated for each minimum-transition-minimum triple. This information can be used to compile the



A B

Figure 3: Ground state structure for unsolvated met-enkephalin (A). Lowest energy unsolvated structure for dominant free energy cluster at 300 K (B). Only heavy atoms are represented.



A B

Figure 4: Ground state structure for solvated met-enkephalin (A). Lowest energy solvated structure for dominant free energy cluster at 300 K (B). Only heavy atoms are represented.

Table 4: Number of local minima connected to the ground state conformation for transition rate cutoffs of 10^7, 10^9 and 10^{11} Hz.

Transition Rate Cutoff (Hz)	Unsolvated # Connected Minima	Solvated # Connected Minima
10^7	15753	25148
10^9	1685	20660
10^{11}	131	1040

rate disconnectivity graph, and thus identify the number of connected minima for various transition rate cutoffs. Table 4 shows the number of minima connected to the ground state conformation for various transition rate cutoffs. Initially, at a 10^7 Hz rate cutoff, a large fraction of the minima are (≈ 70 %) still connected to the unsolvated ground state conformation. However as higher rate cutoffs are applied the node containing the ground state becomes quite small, with only 131 minima at a 10^{11} Hz cutoff. In contrast, the solvated ground state is connected to a larger fraction of minima at all cutoff frequencies. This is particularly evident for the 10^9 rate cutoff. These observations imply that there exist faster transitions which link the ground state conformation to larger regions of the energy surface.

These results can be visualized by plotting the rate disconnectivity graphs for the various rate cutoffs. The corresponding graphs are depicted in Figures 5 and 6. For unsolvated met-enkephalin these graphs reveal that the branch connecting the ground state conformation to the majority of other minima has a relatively low transition, that is, between 10^7 and 10^8 Hz. For this reason, the ground state cluster for 10^9 Hz cutoff is reduced to only a small fraction of the complete graph. In contrast, Figure 6 shows that the solvated ground state conformer is part of the main disconnectivity because of several branches with transitions higher than 10^{10} Hz. Even for a 10^{11} the node containing the ground state conformation is quite dense when compared to the corresponding unsolvated disconnectivity graph. These graphs also illustrate the more liquid-like behavior of the solvated system.

The examination of folding pathways was accomplished by characterizing the transition between the ground state and the lowest free energy minimum of the dominant cluster. In the case of unsolvated met-enkephalin, this transition characterizes the shifting of the type II' β-bend from the Gly^2–Gly^3 to the Gly^3–Phe^4 backbone region. The representative structure for the dominant free energy cluster (denoted as CD*A$_1$) at 300 K exhibits an energy of -9.402 kcal/mol, which is 2.305 kcal/mol higher in energy than the ground state structure. For solvated met-enkephalin, the fully extended to ground state pathway was explored. In this case, the dominant free energy cluster (E*EE) is represented by the FEGM at 300 K. This structure (denoted as E*EE$_1$) has an energy of -46.030 kcal/mol, which is about 4 kcal/mol higher in energy than the ground state conformer. In addition, the second lowest free energy minimum of the E*EE cluster (denoted as E*EE$_2$) was included in this analysis. This minimum exhibits a slightly higher energy of -45.712 kcal/mol.

One way to characterize the folding transition is to examine the pathways connecting two

(a)

(b)

(c)

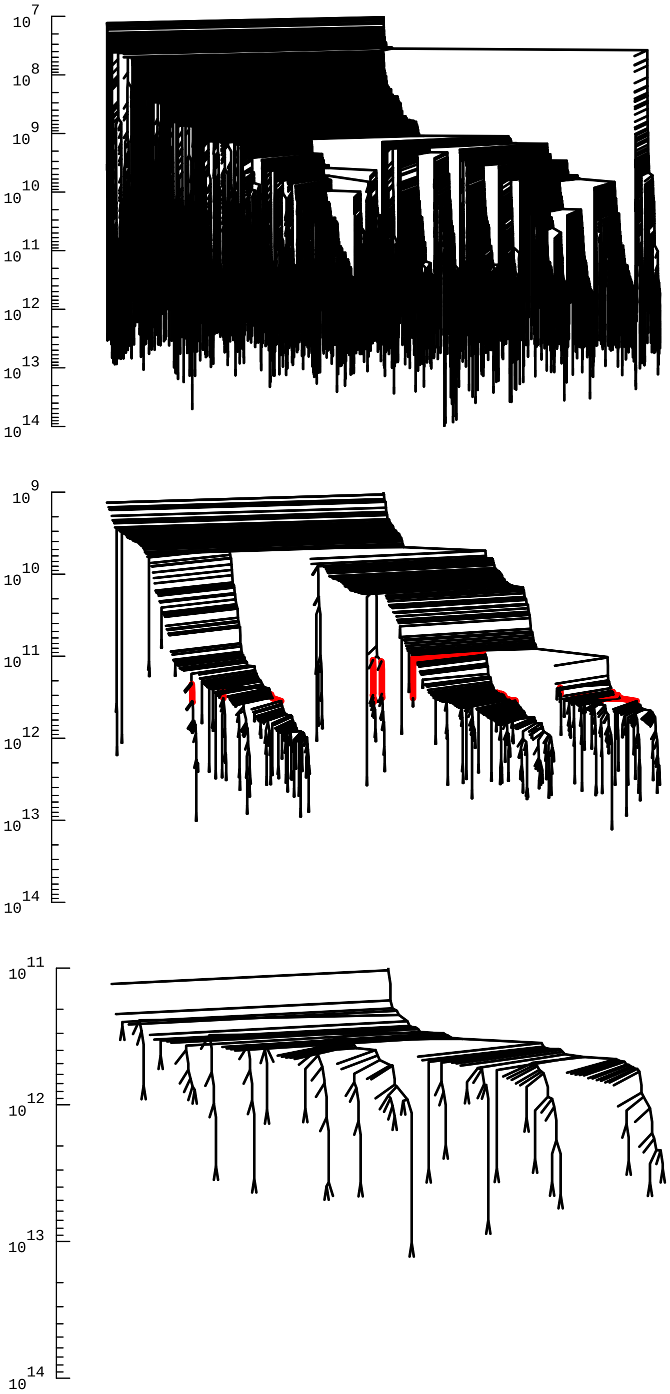


Figure 5: Rate disconnectivity graphs of unsolvated met-enkephalin for transition rate cutoffs of 10^7 (a), 10^9 (b) and 10^{11} (c) Hz. Only the node containing the ground state conformation is shown. The ground state conformer always corresponds to the rightmost branch.

(a)
(b)
(c)

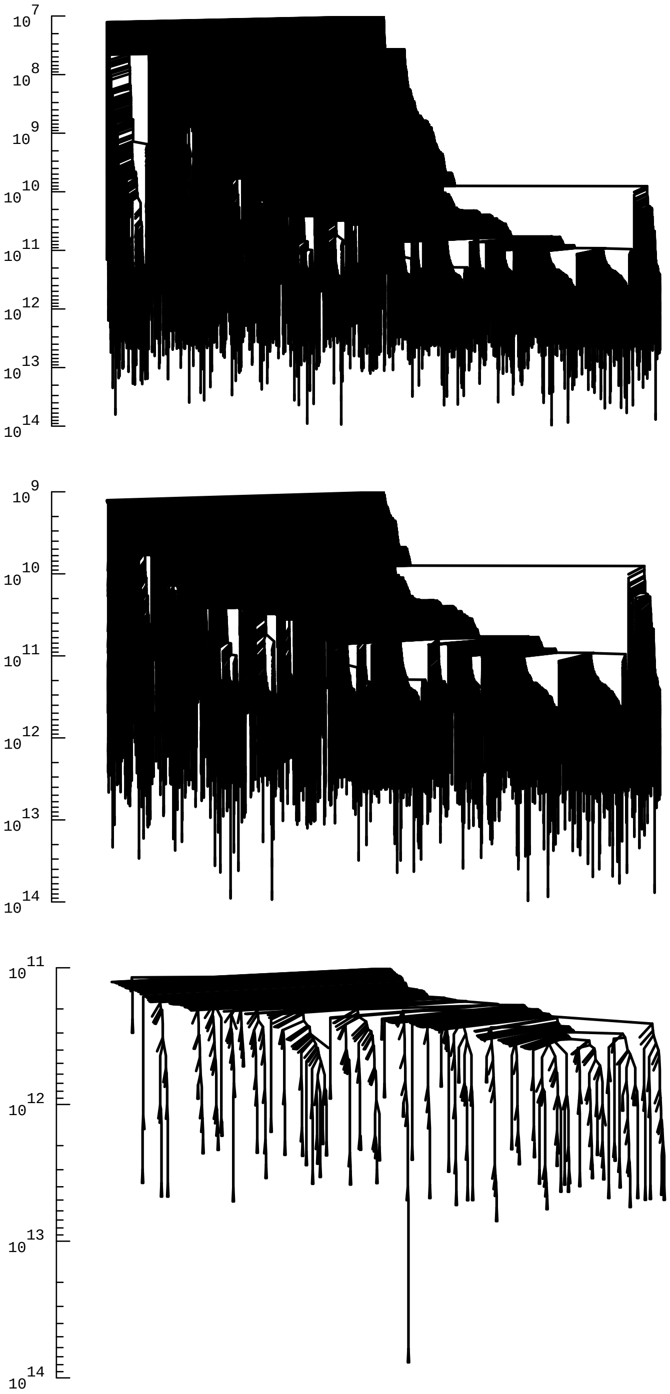


Figure 6: Rate disconnectivity graphs of solvated met-enkephalin for transition rate cutoffs of 10^7 (a), 10^9 (b) and 10^{11} (c) Hz. Only the node containing the ground state conformation is shown. The ground state conformer always corresponds to the rightmost branch.

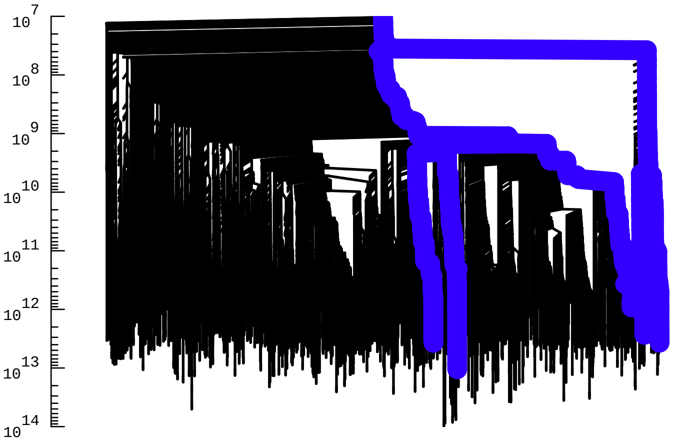
Table 5: Number of unique pathways between two minima. The first column denotes the length of the pathway in terms of the number of transitions. For unsolvated met-enkephalin, the overall transition is between the ground state conformer and the lowest free energy conformer for the dominant free energy cluster (CD*A). For solvated met-enkephalin two transitions with the ground state conformer are considered : The first (E^*EE_1) and second (E^*EE_2) lowest free energy conformers for the dominant free energy cluster (E*EE).

# of Transitions	Unsolvated	Solvated	
	$CD^*A_1 \leftrightarrow$ Ground	$E^*EE_1 \leftrightarrow$ Ground	$E^*EE_2 \leftrightarrow$ Ground
5	–	–	2
6	–	–	–
7	4	13	17
8	18	108	16
9	148	1082	135
10	1228	8583	436
11	10024	62867	2595
12	78936	102124	14154

minima. Through graph theory techniques an exhaustive search can be used to enumerate the full set of pathways that connect these two minima. However, such a search will quickly become impractical when the total number of connected minima is large. To avoid this problem, the enumeration was conducted for pathways comprising less than 12 transitions states. The number of unique pathways for different numbers of transition states (for the minima defined in the preceding paragraph) are given in Table 5. These results indicate a stronger connectivity among the solvated minima, which corroborate the results from the disconnectivity graphs. In particular, when considering the transition between the ground state and the representative structure from the dominant free energy structures, the solvated case exhibits an order of magnitude increase in the number of pathways with an equal number of transitions. In addition when considering the second solvated transition, a number of relatively short pathways can be identified.

Since the length of the pathway is not necessarily proportional to the overall transition rate between the two minima, a more detailed analysis of these folding pathways must be considered. A visual method for accomplishing this is to trace a pathway between the two minima on the rate disconnectivity graph. The overall transition time corresponds to the location of the highest branch (slowest transition) in the connectivity tree. Figure 7 illustrates the tracing of folding pathways for unsolvated and solvated met-enkephalin. For unsolvated met-enkephalin the traced pathway includes 7 transition states. To display the entire pathway, the disconnectivity graph with a 10^7 transition rate cutoff was constructed. This implies that the lowest transition rate is connected by a transition in the range of 10^7 to 10^8 Hz. In the case of solvated met-enkephalin, the shortest pathway connected E^*EE_2 to the ground state, and included only 5 transition states. In addition, the transition rates between these minima are much faster, with a relatively high cutoff of 10^{11} Hz being sufficient for displaying the entire traced pathway.

(a)



(b)

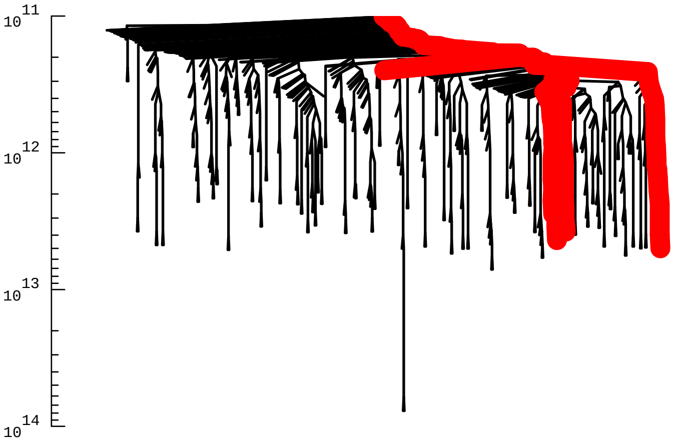


Figure 7: Tracing pathways on rate disconnectivity graphs. (a) shows a 7 transition state pathway (for unsolvated met-enkephalin) connecting CD*A_1 to the ground state on a disconnectivity graph with a 10^7 rate cutoff. (b) shows a 5 transition state pathway (for solvated met-enkephalin) connecting E*EE_2 to the ground state on a disconnectivity graph with a 10^{11} rate cutoff. The traced pathways are indicated by thick grey lines which extend to the root node. The highest branch constitutes the slowest transition. The ground state corresponds to the rightmost branch.

Another method for visualizing the folding pathways is given in Figure 8. In this figure, $\phi - \psi$ conformational plots are used to indicate the structural changes which occur along the pathway. The same unsolvated and solvated pathways shown in Figure 7 are illustrated. In each plot, five lines are drawn, with each line corresponding to one of five residues for met enkephalin. The black dots pinpoint the $\phi - \psi$ values for the transition states along the pathway, while minima are represented by grey dots. The plots imply that along the unsolvated pathway, large structural changes correspond to changes in ϕ values. In contrast, many residues in the solvated pathway undergo only small changes, while two residues display significant variation in both ϕ and ψ variables. Similar plots can be constructed for all dihedral angles, and used to identify important variables along the folding pathway.

6 Conclusions

An important step in better understanding the protein folding process is the characterization of energy surfaces. Rigorously, this characterization requires the identification of all stationary points of the energy function. Along these lines, deterministic methods have been applied to small peptide systems [37]. However, the jump to larger peptide systems with astronomically large numbers of minima is currently intractable.

In this work, a new methodology for analyzing energy surfaces of peptides was introduced. The approach is based on the identification of low (free) energy local minima using a modification of the αBB deterministic global optimization technique [20]. Once this minima have been identified, the low energy regions are explored by transition state searches and the determination of minimum-transition-minimum triples.

The methodology was applied to both unsolvated and solvated forms of met-enkephalin. On the order of 80000 local minima were identified for both forms. In addition to typical forcefield components, entropic contributions are accounted for through a harmonic approximation. The results were used to determine the free energy global minima at 300 K, as well as the dominant free energy cluster at this temperature.

In order to map out the low energy regions, the union of 1000 lowest energy and 1000 lowest free energy minima were compiled and used to initiate transition state searches. First order transition states, as well as minimum-transition-minimum triples, were identified through a parallel implementation of eigenmode following methods. The final set of stationary points included 22775 minima and 28497 first order transition states for unsolvated met-enkephalin and 34722 minima and 42106 first order transition states for solvated met-enkephalin.

Transition rates and connectivity were then calculated for these structures. To explore the connectivity to the ground state conformations, rate disconnectivity graphs were constructed for different rate cutoffs. This analysis indicated higher transition rate connectivity for the solvated form.

In addition, the transitions between specific minima were examined. These transitions were first characterized by the number of pathways (for given number of transition states). Visual interpretation of these results included a pathway tracing on the disconnectivity graph and conformational variations of $\phi - \psi$ variables along the pathway.

In conclusion, the combination of a deterministic global optimization based search for low (free) energy ensembles and eigenmode searches for transition states and minimum-

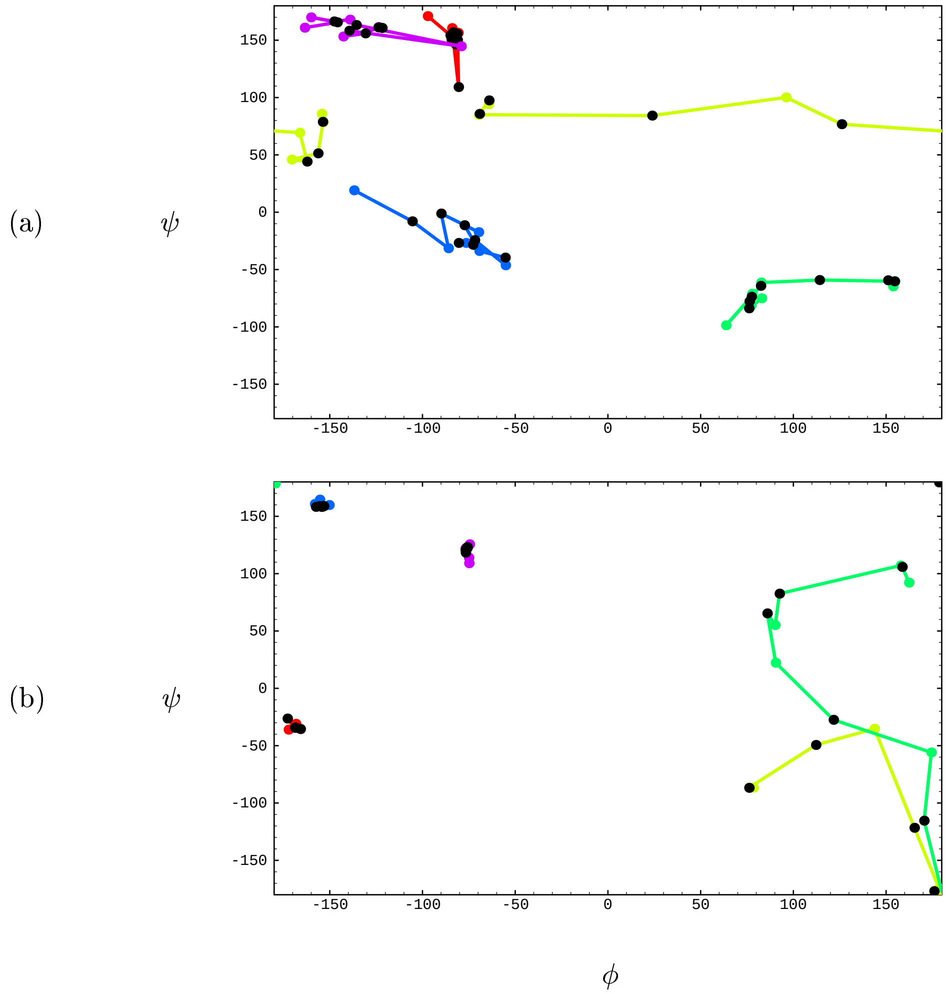


Figure 8: Tracing pathways on $\phi - \psi$ conformational plot. Each line corresponds to one of five residues for met-enkephalin. The black dots represent transition states, while grey dots are minima. The same pathways shown in Figure 7 are plotted. (a) shows a 7 transition state pathway (for unsolvated met-enkephalin) connecting CD*A_1 to the ground state. (b) shows a 5 transition state pathway (for solvated met-enkephalin) connecting E*EE_2 to the ground state.

transition-minimum triples can be used to effectively characterize energy surfaces. The techniques were applied to the unsolvated and solvated forms of met-enkephalin and used to perform a comparative analysis of these systems. A number of visualization techniques were also shown to be useful for further elucidating the folding process for this peptide.

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation, Air Force Office of Scientific Research, the National Institutes of Health (R01 GM52032) and NPACI for parallel computing resources.

References

[1] Adjiman C. S., Androulakis I. P., Maranas C. D. and Floudas C. A. (1996), "A global optimization method, αBB, for process design," *Comp. Chem. Eng. Vol. 20*, S419–S424.

[2] Adjiman C. S., Androulakis I. P. and Floudas C. A. (1997), "Global optimization of MINLP problems in process synthesis and design," *Comp. Chem. Eng. Vol. 21*, S445–S450.

[3] Adjiman C. S., Dallwig S., Floudas C. A. and Neumaier A. (1998), "A global optimization method, αBB, for general twice-differentiable constrained NLPs – I. Theoretical advances," *Comp. Chem. Eng. Vol. 22*, 1137–1158.

[4] Adjiman C. S., Androulakis I. P. and Floudas C. A. (1998), "A global optimization method, αBB, for general twice-differentiable constrained NLPs – II. Implementation and computational results," *Comp. Chem. Eng. Vol. 22*, 1159–1179.

[5] Allinger N. L., Yuh Y. H. and Lii J. H. (1989), "Molecular mechanics. The MM3 force field for hydrocarbons," *J. Am. Chem. Soc. Vol. 111*, 8551–8565.

[6] Androulakis I. P., Maranas C. D. and Floudas C. A. (1995), "αBB : A global optimization method for general constrained nonconvex problems," *J. Global Opt. Vol. 7*, 337–363.

[7] Androulakis I. P., Maranas C. D. and Floudas C. A. (1997), "Global minimum potential energy conformations of oligopeptides," *J. Global Opt. Vol. 11*, 1–34.

[8] Augspurger J. D. and Scheraga H. A. (1996), "An efficient, differentiable hydration potential for peptides and proteins," *J. Comp. Chem. Vol. 17*, 1549–1558.

[9] Becker O. M. and Karplus, M. (1997), "The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics," *J. Chemical Physics, Vol. 106*, 1495–1517.

[10] Brooks B., Bruccoleri R., Olafson B. and States D. (1983), "CHARMM: A program for macromolecular energy minimization and dynamics calculations," *J. Comp. Chem. Vol. 4*, 187–217.

[11] Cerjan C. J. and Miller W. H. (1981), "On finding transition states," *J. Chemical Physics Vol. 75*, 2800–2806.

[12] Czerminski R. and Elber R. (1989), "Reaction path study of conformational transitions and helix formation in a tetrapeptide," *Proc. Natl. Acad. Sci. USA, Vol. 86*, 6963–6967.

[13] Czerminski R. and Elber R. (1990), "Reaction path study of conformational transitions in flexible systems: Application to peptides," *J. Chemical Physics, Vol 92.*, 5580–5601.

[14] Flory, P. J. (1974), "Foundations of rotational isomeric state theory and general methods for generating configurational averages," *Macromolecules Vol. 7*, 381–392.

[15] Floudas C. A., Klepeis J. L. and Pardalos P. M. (1999), "Global Optimization Approaches in Protein Folding and Peptide Docking," *In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 47*, 141–171, American Mathematical Society, Providence, Rhode Island.

[16] Gill P. E., Murray W., Saunders M. A. and Wright M. H. (1986), "NPSOL 4.0 User's Guide," Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, California.

[17] Go N. and Scheraga H. A. (1969), "Analysis of the contribution of internal vibrations to the statistical weights of equilibrium conformations of macromolecules," *J. Chemical Physics Vol. 51*, 4751–4767.

[18] Go N. and Scheraga H. A. (1976), "On the use of classical statistical mechanics in the treatment of polymer chain conformations," *Macromolecules Vol. 9*, 535–542.

[19] Klepeis J. L., Androulakis I. P., Ierapetritou M. G. and Floudas C. A. (1998), "Predicting solvated peptide conformations via global minimization of energetic atom-to-atom interactions," *Comp. Chem. Eng. Vol. 22*, 765–788.

[20] Klepeis J. L. and Floudas C. A. (1999), "Free Energy Calculations for Peptides via Deterministic Global Optimization," *J. Chemical Physics, Vol. 110*, 7491–7512.

[21] Klepeis J. L. and Floudas C. A. (1999), "A comparative study of global minimum energy conformations of hydrated peptides," *J. Comp. Chem. Vol. 20*, 636–654.

[22] Kunz R. E. and Berry R. S. (1995), "Statistical interpretation of topographies and dynamics of multidimensional potentials," *J. Chemical Physics, Vol. 103*, 1904–1912.

[23] Laidler K. J. (1989), "Chemical Kinetics," Harper Collins, New York.

[24] Li Z. and Scheraga H. A. (1988), "Structure and free energy of complex thermodynamic systems," *J. Molecular Structure (Theochem) Vol. 179*, 333–352.

[25] Maranas C. D. and Floudas C. A. (1994), "Global minimum potential energy conformations of small molecules," *J. Global Opt. Vol. 4*, 135–170.

[26] Meirovitch H. and Meirovitch E. (1997), "Efficiency of Monte Carlo minimization procedures and their use in analysis of NMR data obtained from flexible peptides," *J. Comp. Chem. Vol. 18*, 240–253.

[27] Momany F. A., Carruthers L. M., McGuire R. F. and Scheraga H. A. (1974), "Intermolecular potential from crystal data. III.," *J. Physical Chemistry Vol. 78*, 1595–1620.

[28] Momany F. A., Carruthers L. M. and Scheraga H. A. (1974), "Intermolecular potential from crystal data. IV.," *J. Physical Chemistry Vol. 78*, 1621–1630.

[29] Momany F. A., McGuire R. F., Burgess A. W. and Scheraga H. A. (1975), "Energy parameters in polypeptides. VII.," *J. Physical Chemistry Vol. 79*, 2361–2381.

[30] Némethy G., Gibson K. D., Palmer K. A., Yoon C. N., Paterlini G., Zagari A., Rumsey S. and Scheraga H. A. (1992), "Energy parameters in polypeptides. 10.," *J. Physical Chemistry Vol. 96*, 6472–6484.

[31] Neumaier A. (1997), "Molecular Modeling of Proteins and Mathematical Prediction of Protein Structure," *SIAM Rev., Vol. 39*, 407–460.

[32] Nichols J., Taylor H., Schmidt P. and Simons J. (1990), "Walking on potential energy surfaces," *J. Chemical Physics Vol. 92*, 340–346.

[33] Simons J., Jorgensen P., Taylor H. and Ozment J. (1983), "Walking on potential energy surfaces," *J. Physical Chemistry Vol. 87*, 2745–2753.

[34] Tsai C. J. and Jordan K. D. (1993), "Use of an eigenmode method to locate the stationary points on the potential energy surfaces of selected argon and water clusters," *J. Physical Chemistry Vol. 97*, 11227–11237.

[35] Weiner S., Kollman P., Case D. A., Singh U. C., Ghio C., Alagona G., Profeta S. and Weiner P. (1984), "A new force field for molecular mechanical simulation of nucleic acids and proteins," *J. Am. Chem. Soc. Vol. 106*, 765–784.

[36] Weiner S., Kollman P., Nguyen D. and Case D. (1986), "An all atom force field for simulations of proteins and nucleic acids," *J. Comp. Chem. Vol. 7*, 230–252.

[37] Westerberg K. M. and Floudas C. A. (1999), "Locating All Transition States and Studying the Reaction Pathways of Potential Energy Surfaces," *J. Chemical Physics, Vol. 110*, 9259–9295.

[38] Zimmerman S. S., Pottle M. S., Némethy G. and Scheraga H. A. (1977), "Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP," *Macromolecules Vol. 10*, 1–9.