Journal of Chemical Physics, accepted for publication, January, 1999

Free Energy Calculations for Peptides via Deterministic Global Optimization

J. L. Klepeis and C. A. Floudas*

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

Abstract

The αBB algorithm is a deterministically based global optimization method that has been successfully used to locate the global minimum energy conformations of peptide systems. The goal of this procedure is to identify the native conformation of a given peptide by identifying the structure possessing the global minimum potential energy. However, a rigorous conformational search should locate the structure exhibiting the global minimum free energy. In this work, novel methods are developed for locating free energy global minimum conformations and clusters of peptides. These methods are based on an harmonic approximation for entropic effects, which requires the ability to generate a dense ensemble of distinct low energy local minima. Two approaches, both based on the general concepts of the αBB branch and bound framework, are used to generate these ensembles. In performing these calculations, potential energy contributions were modeled using an all-atom force field. In addition, hydration effects were also considered by utilizing a solvent–accessible volume of hydration shell model. The free energy analysis was applied to both the unsolvated and solvated forms of met– and leu–enkephalin. It was found that both methods produce dense, Boltzmann-like, distributions of low energy metastable states. The inclusion of entropic effects was also found to influence the prediction of free energy global minima. In addition, a

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

statistical treatment of the thermodynamics of folding showed that the transition temperature, which signified a collapse from high energy, extended structures to a ground-state-like ensemble, could be identified.

1 Introduction

Predicting the native three–dimensional structure of proteins from amino acid information, or the protein folding problem, has been the focus of much research since Anfinsen introduced the hypothesis that native proteins are in their thermodynamically most–stable state [1]. Through the use of empirical potential energy functions, which are used to model the interatomic interactions of the peptide system, the multiple–minima problem has become a hallmark of computational chemistry. Some methods, including metropolis Monte Carlo and molecular dynamics, can be used to search the multidimensional energy landscape, although the results tend to be dependent on the choice of starting conformations. Therefore, a fundamental issue that has been addressed through the protein folding problem is the ability to develop efficient methods for finding the global potential energy minimum from among many local minima [2,3].

However, locating the global minimum potential energy conformation is not sufficient because Anfinsen's thermodynamic hypothesis requires the minimization of the conformational free energy. Specifically, potential energy minimization neglects the entropic contributions to the stability of the molecule. An approximation to these entropic contributions can be developed by using information about low energy conformations. That is, once a sufficient ensemble of low energy minima has been identified, a statistical analysis can be used to estimate the relative entropic contributions, and thus the relative free energy, for conformations in the ensemble.

Therefore, the analysis of the free energy of peptides requires efficient methods for locating not only the global minimum energy structure but also large numbers of low energy conformers. A variety of methods have been used to find such stationary points on potential energy surfaces. For example, periodic quenching during a Monte Carlo or molecular dynamics trajectory can be used to identify local minima [4]. However, a drawback of these approaches is their inherent stochastic nature. In its original form, the αBB *deterministic* global optimization algorithm [5,6,7,8,9] has been shown to be an efficient method for finding the global minimum energy conformation for both unsolvated and solvated peptide systems [10,11,12]. In this work, novel methods are proposed within the framework of the αBB algorithm to optimize the free energy of peptide systems. These modifications facilitate the

generation of ensembles of low energy conformers, which can be used to identify the global minimum free energy conformation, as well as perform detailed free energy rankings.

In the first section, issues related to the energetic modeling of peptide systems are considered. This includes specification of potential and solvation modeling, as well as a discussion on the ability to rigorously and approximately model the entropic contributions to the free energy of these systems. The next section gives an overview of the problem formulation. The third section outlines the theoretical and algorithmic development for the αBB based methods, including their proficiency in locating the global, as well as ensembles of local minimum energy conformations. Computational results are then presented for both the isolated and solvated forms of met– and leu–enkephalin examples. This is followed by a final section devoted to a summary and discussion of the results.

2 Energy Modeling

2.1 Potential Energy Model

Many models have been developed using a classical description of molecules in terms of atomic bonds and effective interactions. Some of these parameterizations of molecular potential functions include ECEPP[13,14,15], AMBER[16,17], CHARMM[18], DISCOVER[19], GROMOS[20], MM3[21], ENCAD[22], ECEPP/2[23] and ECEPP/3[24]. In most cases, these force fields are atom centered potentials from which the total molecular energy is computed as a sum over all pairwise interactions.

In this work, the ECEPP/3 (Empirical Conformational Energy Program for Peptides) potential model is utilized[24]. In this force field, it is assumed that covalent bond lengths and bond angles are fixed at their equilibrium values. Under this assumption, all residues of the same type have essentially the same geometry in various proteins. Therefore, a chain of any sequence can be generated using the fixed geometry specific to each type of amino acid residue in the sequence.

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. 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. The contributing terms to the total potential energy of ECEPP/3 are shown in Figure 1, and the development of the appropriate parameters is discussed and reported elsewhere[24].

2.2 Solvation Energy Model

Solvation contributions are generally believed to be a significant force in stabilizing the native conformations of proteins. Explicit methods include solvation effects by actually surrounding the peptide with solvent molecules. Energetic evaluations require the calculation of both solvent–peptide and solvent–solvent interactions. Although these methods are conceptually simple, explicit inclusion of solvent molecules greatly increases the computational time needed to simulate the peptide system. Therefore, most simulations of this type are limited to local conformational searches. In addition, it is difficult to quantify the effect of hydrophobic interactions that result from the ordering of water molecules.

In contrast, continuum models use a simplified representation of the solvent environment by neglecting the molecular nature of the water molecules. Calculations of solvation free energies using electrostatic continuum models rely on numerical solutions to the Poisson–Boltzmann equation from which dielectric and ionic strength effects are obtained[25]. Other continuum models estimate free energies of solvation as a function of geometric quantities, such as surface areas and volumes.

In this work, solvation contributions are included implicitly using empirical correlations with solvent accessible volumes[26]. The main assumption of these models is that, for each functional group of the peptide, a hydration free energy can be calculated from an averaged free energy of interaction of the group with a layer of solvent known as the hydration shell. In addition, the total free energy of hydration is expressed as a sum of the free energies of hydration for each of the functional groups of the peptide.

2.2.1 Accessible Volume of Hydration Shell

The proportionality between the hydration energy and the solvent–accessible volume of a hydration layer surrounding the peptide can be represented in the form :

$$E_{HYD} \;=\; \sum_{i=1}^{N} (VHS_i)(\delta_i) \tag{1}$$

An additive relationship for the N individual atoms of the peptide is assumed. Here (VHS_i) represents the solvent–accessible volume of hydration shell for each atom i that is exposed to water, and the (δ_i) parameters are empirically determined free energy of hydration densities for these atoms.

The hydration shell is defined by the volume inside a sphere of radius R_i^h but outside a sphere of radius R_i^v, with both radii centered on atom i. The larger radius, R_i^h, corresponds to the radius of the first hydration shell of atom i, while R_i^v is equal to the van der Waals radius. In order to calculate (VHS_i), the volume of a collection of overlapping hard spheres must be computed using :

$$V(\mathbf{R}) \;=\; \sum_i a_i S_i - \sum_{ij} b_{ij} D_{ij} + \sum_{ijk} c_{ijk} T_{ijk} - \sum_{ijkl} d_{ijkl} Q_{ijkl} \tag{2}$$

In Equation (2), S_i signifies the volume of a single sphere, while D_{ij}, T_{ijk} and Q_{ijkl} represent the volume of intersection of two, three and four spheres, respectively. This is sufficient because all higher order overlaps can be decomposed into the three types of intersections included in Equation (2). Therefore, the solvent–accessible volume of hydration can be written as :

$$(VHS_i) \;=\; V(R_i^h) - V(R_i^v) \tag{3}$$

The first term in Equation (3) is calculated using Equation (2) with the radii of all atoms set equal to their van der Waals radii, while the second term is calculated with the radius of atom i equal to R_i^h and the van der Waals radii of all the other atoms. A number of methods to compute hydration shell volumes have been proposed[26,27,28].

The form of Equation (2) is not suitable for force field models using pairwise intramolecular potential, such as ECEPP/3. Furthermore, direct truncation at the 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[26]. This method uses a truncated form of Equation (2) but also artificially reduces 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.

Free energy density parameters for solvent accessible volumes have been developed for nonionic and charged organic solute molecules[29,30,31]. In this work, RRIGS specific (δ_i), which were developed by a least square fitting of experimental free energy of solvation data for 140 small organic molecules[26], are used.

2.3 Free Energy Modeling

Most force fields are used to describe only potential energy and, occasionally, solvation energy contributions for a given peptide system. However, rigorous consideration of Anfinsen's thermodynamic hypothesis[1] requires the global minimization of the conformational free energy. Therefore, the entropic contributions to the stability of the protein must be estimated in order to augment the prediction provided by potential and solvation energy models. In peptide systems, this entropic contribution arises from fluctuations around a local conformational state. There exists a number of procedures, including both exact and approximate calculations, that can be used to determine the entropic contributions, and thus the free energy, of peptide systems.

First, assume that the full conformational space R can be considered as the union of disjoint basins of attraction, and the conformational space associated with a given basin (denoted by γ) is defined by R_γ. The energy, E, is a function of the variable set θ, which corresponds to the set of dihedral angles used to describe the conformational state of the system. Each basin of attraction is characterized by a unique local minimum at position

θ^*_γ, with a corresponding energy E^*_γ. That is, local minimization starting at any point in R_γ will lead to the local minimum at θ^*_γ. It should be noted that this approximation of the conformational space excludes all maxima and saddle point conformations.

For a given temperature, T, the probability that a peptide occupies the conformational space of a given basin (R_γ) can be described by a Gibbs/Boltzmann distribution :

$$p_\gamma \;=\; \frac{\int_{R_\gamma} exp(-\beta E(\theta))d\theta}{\int_R exp(-\beta E(\theta))d\theta} \tag{4}$$

Here β is equivalent to $\frac{1}{k_B T}$. If the numerator is redefined as the partition function (Z_γ) for the basin, Equation (4) can be rewritten as :

$$p_\gamma \;=\; \frac{Z_\gamma}{Z} \tag{5}$$

The total partition function for the entire conformational space is represented by Z. Since this function is described by a disjoint set of basins (R_γ), it is equivalent to the following form :

$$Z \;=\; \sum_\gamma Z_\gamma \tag{6}$$

Once the probability is known, the corresponding free energy, G_γ, associated with each basin can also be calculated :

$$G_\gamma \;=\; -\frac{\ln p_\gamma}{\beta} \tag{7}$$

Using these definitions, a rigorous procedure can be envisioned for calculating the exact probability associated with a given basin. First, a sample of conformations must be generated with initial starting energies E_i, as defined by the total set I. Each structure is minimized to identify its corresponding basin minimum (θ^*_γ). These structures define the set $I(\gamma)$ (i.e., those structures associated with basin γ). As the sampling goes to infinity, the probability associated with basin γ can be calculated by the following expression :

$$p_\gamma^{exact} = \frac{\sum_{i(\gamma)\in I(\gamma)} exp(-\beta E_{i(\gamma)})}{\sum_{i\in I} exp(-\beta E_i)} \tag{8}$$

Obviously, such a method is intractable for large systems, and this is the impetus for developing approximate methods.

2.3.1 Harmonic Approximation

A tractable method for including entropic effects for proteins relies on the concept of the harmonic approximation. Initially, the theoretical development of this approximation for polymer systems generated debate in the literature[32,33,34]. In the work of Go and Scheraga [32] a classical rigid model was used to characterize a partition function based on the fixed bond length and bond angle assumptions. In contrast, Flory[34] derived a different partition function using a classical flexible model. Later analysis by Go and Scheraga[33] actually showed that the flexible model was also applicable to the fixed bond length and bond angle system (i.e., a peptide described by the internal coordinate system).

In either case (i.e., rigid or flexible), entropic contributions can be calculated by employing an harmonic approximation[33]. The fundamental concept is to characterize the basin of attraction (γ) by the properties of its corresponding local minimum (θ^*_γ), and not by a random sampling of conformations. These properties include the local minimum energy value, E^*_γ, and the convexity around the local minimum. Essentially, the convexity measure is used to approximate the basin of attraction region as a hyperparabola centered at the local minimum. Therefore, the anharmonic nature of the true basin, which defines the deviation from approximated harmonic behavior, controls the error associated with this assumption.

At each minimum (θ^*_γ) the harmonic approximation to the entropy can be evaluated using the following expression :

$$S^{approx}_\gamma = -\frac{k_B}{2} \ln \left[Det \left(H_\gamma \right) \right] + \hat{f}(T) \tag{9}$$

Here $Det\left(H_\gamma\right)$ refers to the determinant of the Hessian (second derivative matrix) evaluated at the local minimum θ^*_γ. The function $\hat{f}(T)$ is an additive term that is only dependent on temperature[35]. The approximated free energy can then be calculated by combining the energetic and entropic contributions through the follow expression :

$$G^{approx}_\gamma = E^*_\gamma - TS^{approx}_\gamma + \bar{f}(T) \tag{10}$$

By substituting the harmonic entropic approximation from Equation (9), Equation (10) becomes :

$$G^{approx}_\gamma = E^*_\gamma + \frac{1}{2\beta} \ln \left[Det \left(H_\gamma \right) \right] + \tilde{f}(T) \tag{11}$$

In this equation, it becomes evident that the free energy for a given basin is estimated using only the properties of the corresponding local minimum, that is, the local minimum energy (E^*_γ) and a measure of local convexity ($Det\,(H_\gamma)$). A temperature dependent term, $\tilde{f}(T)$, is included, although it does not affect relative free energy comparisons.

Expressions for the probabilities and partition functions can also be developed. By combining Equations (5), (7) and (11), an approximation for the partition function of a given basin can be written as :

$$\ln Z_\gamma^{approx} = -\beta E^*_\gamma - \frac{\ln\left[Det\,(H_\gamma)\right]}{2} - \beta\tilde{f}(T) + \ln Z \tag{12}$$

A further simplification can be made by realizing that $-\beta\tilde{f}(T)$ and $\ln Z$ are constant for a given temperature (i.e., $f(T) = -\beta\tilde{f}(T) + \ln Z$). Equation (12) can be rewritten as :

$$Z_\gamma^{approx} = \left[\frac{1}{\left[Det\,(H_\gamma)\right]}\right]^{1/2} \exp\left(-\beta E^*_\gamma\right) f(T) \tag{13}$$

Finally, by using Equation (6), an approximate probability associated with a given basin (γ) can be calculated using the following equation :

$$p_\gamma^{approx} = \frac{\left[\frac{1}{\left[Det(H_\gamma)\right]}\right]^{1/2} \exp\left(-\beta E^*_\gamma\right)}{\sum_{i=1}^{N}\left[\frac{1}{\left[Det(H_i)\right]}\right]^{1/2} \exp\left(-\beta E^*_i\right)} \tag{14}$$

As expected, the $f(T)$ term disappears, and the statistical weight becomes a function of only the temperature (through β), the local minimum energy value and the measure of convexity. In order to develop a meaningful comparison of relative free energies, the total partition function (i.e., the denominator of Equation (14)) must include an adequate ensemble of low energy local minima, as well as the global minimum energy conformation.

These probabilities can be used to estimate the occupancy of each individual basin, or summed in order to calculate cumulative probabilities for an ensemble of structures exhibiting similar physical or energetic properties. It should be noted that the determination of free energy using the harmonic approximation does not require the explicit inclusion of a contribution based on the density of states. That is, the harmonic approximation decomposes the energetic states within a basin of attraction into one energetic value represented by the local minimizer of the basin. In contrast to counting methods, which estimate probabilities

based on the density of states, the contribution of each structure should be accounted for only once. Therefore, using the harmonic approximation requires a structural comparison of all local minimizers.

The probabilities obtained through the harmonic approximation can also be used to calculate thermodynamic quantities. Once the set of unique minimizers has been identified, these structures can be ranked according to their free energy values, and then divided into bins of a specified energy width. Probabilities for each bin can be calculated by summing the individual probabilities (as defined in Equation (14)) :

$$P_j^{approx} = \sum_{\gamma=1}^{n_j} p_\gamma^{approx} \tag{15}$$

Here P_j^{approx} signifies the probability for energy bin j. The summation includes the n_j individual probabilities (p_γ^{approx}) belonging to bin j. Average thermodynamic quantities can now be estimated using the equations with the following form :

$$< E >_T = \sum_j P_j^{approx} < E >_j \tag{16}$$

Here the total average energy, $< E >_T$, is calculated by summing the bin probabilities multiplied by the mean energy of bin j, $< E >_j$.

3 Problem Formulation

The energy minimization problem for proteins is formulated as a nonconvex nonlinear optimization problem. Let $i = 1, \ldots, N_{RES}$ be an indexed set describing the sequence of amino acid residues in the peptide chain. There are $\phi_i, \psi_i, \omega_i,\ i = 1, \ldots, N_{RES}$ dihedral angles along the backbone of this peptide. In addition, let $k = 1, \ldots, K^i$ denote the dihedral angles of the side chains for the i^{th} residue and $j = 1, \ldots, J^N$ denote the dihedral angles of the amino end group and $j = 1, \ldots, J^C$ of the carboxyl end group, respectively. Therefore, these angles can be defined in following manner : $\chi_i^k,\ i = 1, \ldots, N_{RES},\ k = 1, \ldots, K^i$ for the side chain dihedral angles; $\phi_j^N,\ j = 1, \ldots, J^N$ and $\phi_j^C,\ j = 1, \ldots, J^C$ for the amino and carboxyl end group dihedral angles, respectively. Using these definitions the optimization problem takes

the following form :

$$\min \qquad E(\phi_i, \psi_i, \omega_i, \chi_i^k, \phi_j^N, \phi_j^C) \tag{17}$$

$$\begin{aligned}
\text{subject to} \quad -\pi &\leq \phi_i \leq \pi, \quad i = 1, \ldots, N_{RES} \\
-\pi &\leq \psi_i \leq \pi, \quad i = 1, \ldots, N_{RES} \\
-\pi &\leq \omega_i \leq \pi, \quad i = 1, \ldots, N_{RES} \\
-\pi &\leq \chi_i^k \leq \pi, \quad i = 1, \ldots, N_{RES}, \; k = 1, \ldots, K^i \\
-\pi &\leq \phi_j^N \leq \pi, \quad j = 1, \ldots, J^N \\
-\pi &\leq \phi_j^C \leq \pi, \quad j = 1, \ldots, J^C
\end{aligned}$$

In general, E represents the total of both the potential and solvation energies. In this work, these functions are modeled using the ECEPP/3[24] and RRIGS[26] force fields. At this point, the usual approach for predicting the native conformation of a protein is to treat Equation (17) as a global optimization problem. Along these lines, a large number of techniques have been developed to locate global minima on complicated energy hypersurfaces, and these approaches have been extensively reviewed[2,3]. The typical limitation is that many methods depend heavily on the supplied initial conformation. As a result, there is no guarantee for global convergence because large sections of the domain space may be overlooked.

It has been found that the αBB global optimization approach[5,6,7,8,9] is particularly efficient for locating the global minimum energy conformations of isolated and solvated peptides [10,11,12,36]. The development of this novel branch and bound method was motivated by the need for an algorithm that could guarantee convergence to the global minimum of nonlinear optimization problems with twice–differentiable functions[2,37]. The application of this algorithm to the minimization of potential energy functions was first introduced for microclusters [38,39], and small acyclic molecules[40,41].

The inclusion of free energy modeling into the protein folding problem does not change the general formulation given by Equation (17). However, an additional condition must be satisfied; that is, an ensemble of local minimum low energy conformations must be generated along with the global minimum energy conformation. Once this ensemble has been compiled,

a free energy ranking can be performed using the harmonic approximation presented in the previous section.

In the following sections, the application of the αBB deterministic global optimization algorithm to the free energy protein folding problem is discussed. For introductory purposes, the general αBB approach, as applied to the prediction of global minimum energy conformations, is outlined in Section 4. The remainder of Section 4 is devoted to the development of rigorous αBB methods for locating ensembles of low energy conformations. Finally, novel αBB based algorithms, which are used to locate low energy conformers in this work, are presented in Section 5.

4 αBB – Deterministic Global Optimization

4.1 Global Minimum Energy Conformation

The αBB global optimization algorithm effectively brackets the global minimum solution by developing converging sequences of lower and upper bounds. These bounds are refined by iteratively partitioning the initial domain. Upper bounds on the global minimum are obtained by function evaluations or local minimizations of the original energy function, E. Lower bounds belong to the set of solutions of the convex lower bounding functions, which are constructed by augmenting E with the addition of separable quadratic terms. The lower bounding function (L) of the energy hypersurface can be expressed in the following manner :

$$
\begin{aligned}
L \;=\; E + \{ \quad & \sum_{i=1}^{N_{RES}} \alpha_{\phi_i} \left(\phi_i^L - \phi_i\right)\left(\phi_i^U - \phi_i\right) + \\
& \sum_{i=1}^{N_{RES}} \alpha_{\psi_i} \left(\psi_i^L - \psi_i\right)\left(\psi_i^U - \psi_i\right) + \\
& \sum_{i=1}^{N_{RES}} \alpha_{\omega_i} \left(\omega_i^L - \omega_i\right)\left(\omega_i^U - \omega_i\right) + \\
& \sum_{i=1}^{N_{RES}} \sum_{k=1}^{K^i} \alpha_{\chi_i^k} \left(\chi_i^{k,L} - \chi_i^k\right)\left(\chi_i^{k,U} - \chi_i^k\right) + \\
& \sum_{j=1}^{J^N} \alpha_{\phi_j^N} \left(\phi_j^{N,L} - \phi_j^N\right)\left(\phi_j^{N,U} - \phi_j^N\right) +
\end{aligned}
\tag{18}
$$

$$\sum_{j=1}^{J^C} \alpha_{\phi_j^C} \left(\phi_j^{C,L} - \phi_j^C\right) \left(\phi_j^{C,U} - \phi_j^C\right) \}$$

Here $\phi_i^L, \psi_i^L, \omega_i^L, \chi_i^{k,L}, \phi_j^{N,L}, \phi_j^{C,L}$ and $\phi_i^U, \psi_i^U, \omega_i^U, \chi_i^{k,U}, \phi_j^{N,U}, \phi_j^{C,U}$ represent lower and upper bounds on the dihedral angles $\phi_i, \psi_i, \omega_i, \chi_i^k, \phi_j^N, \phi_j^C$. The α represent nonnegative parameters which must be greater or equal to the negative one–half of the minimum eigenvalue of the Hessian of E over the defined domain. A number of methods have been developed for estimating these parameters[7,8,40,42]. The overall effect of these terms is to overpower the nonconvexities of the original nonconvex terms by adding the value of 2α to the eigenvalues of the Hessian of E. The convex lower bounding functions, L, possess a number of important properties which guarantee global convergence[41] :

(i) L is a valid underestimator of E;

(ii) L matches E at all corner points (points that uniquely define the multidimensional subregion) of the box constraints (constraints that set the lower and upper variable bounds);

(iii) L is convex in the current box constraints;

(iv) the maximum separation between L and E is bounded and proportional to α and to square of the diagonal of the current box constraints. This property ensures that an ϵ_c convergence tolerance can be reached for a finite size partition element;

(v) the underestimators L constructed over supersets of the current set are always less tight than the underestimator constructed over the current box constraints for every point within the current box constraints.

Once solutions for the upper and lower bounding problems have been established, the next step is to modify these problems for the next iteration. This is accomplished by successively partitioning the initial domain into smaller subdomains. One possible partitioning strategy involves successive subdivision of the original hyper-rectangle by halving on the midpoint of the longest side (bisection). In order to ensure non–decreasing lower bounds, the hyper–rectangle to be bisected is chosen by selecting the region which contains the infimum of the

minima of lower bounds. A non–increasing sequence for the upper bound can be developed by solving the original nonconvex problem, E, locally and selecting it to be the minimum over all the previously recorded upper bounds. Obviously, if the single minimum of L for any subdomain is greater than the current upper bound, this hyper–rectangle can be fathomed (discard region from branch and bound tree because it cannot contain the global minimum). In the worst case scenario the full branch and bound tree would need to be expanded and searched. However, for the protein folding problem and for many problems (in general), the ability to fathom significant portions of the domain space accelerates convergence to the global minimum. A one-dimensional example of this algorithm is illustrated in Figure 2.

4.2 Enumerating All Local Minima : Rigorous Approaches

Several rigorous methods can be envisioned for locating local minimum energy conformations using the αBB deterministic global optimization approach. As an introduction to the ideas used in this work, two rigorous approaches for finding all local minimum energy conformations are discussed.

The first method relies on the introduction of a single inequality constraint to the problem formulation given by (17). The new formulation is :

$$\begin{aligned}
\min \quad & E(\phi_i, \psi_i, \omega_i, \chi_i^k, \phi_j^N, \phi_j^C) \qquad (19)\\
\text{subject to} \quad & (E^* - E) + \epsilon^* \ < \ 0 \\
& -\pi \ \leq \ \phi_i \ \leq \ \pi, \ \ i = 1, \ldots, N_{RES} \\
& -\pi \ \leq \ \psi_i \ \leq \ \pi, \ \ i = 1, \ldots, N_{RES} \\
& -\pi \ \leq \ \omega_i \ \leq \ \pi, \ \ i = 1, \ldots, N_{RES} \\
& -\pi \ \leq \ \chi_i^k \ \leq \ \pi, \ \ i = 1, \ldots, N_{RES}, \ k = 1, \ldots, K^i \\
& -\pi \ \leq \ \phi_j^N \ \leq \ \pi, \ \ j = 1, \ldots, J^N \\
& -\pi \ \leq \ \phi_j^C \ \leq \ \pi, \ \ j = 1, \ldots, J^C
\end{aligned}$$

The additional constraint requires that the objective function values be larger than the energy value at some local (or global) minimum, as denoted by E^*, plus a positive parameter, ϵ^*.

When $\epsilon^* = 0$, the solution of the corresponding global optimization problem will give the best local minimum energy conformation with an energy larger than E^*. The original formulation given by (17) is actually a special case of this problem in which $E^* = -\infty$ and $\epsilon^* = 0$. That is, in (17) no bounds are placed on the value of the objective function, E. The global minimum energy conformation is only required to take some finite value. In order to locate all local minima, a set of global optimization problems must be solved iteratively with updating of the parameter E^*.

The problem of finding all local minimum energy conformations can also be formulated as a single global optimization problem, which can be deterministically solved using the αBB algorithm[43]. This method stems from the idea that all stationary points (i.e., minima, maxima and transition states) of the energy hypersurface satisfy the constraint $\nabla E(\theta) = \mathbf{0}$. This can be written as :

$$\frac{\partial E(\theta)}{\partial \theta_i} = 0 \ , \qquad i = 1, \ldots, N_\theta \tag{20}$$

Here N_θ represents the total number of dihedral angles defined by the variable set θ. The problem of finding local minima is equivalent to finding all solutions of Equation (20) for which the Hessian of E is positive definite.

The problem posed in Equation (20) involves the solution of a system of nonlinear equations. The general case takes the following form :

$$\begin{aligned} h_j(\theta) &= 0 \qquad j = 1, \ldots, N_h \\ g_k(\theta) &\leq 0 \qquad k = 1, \ldots, N_g \\ \theta^{\mathbf{L}} \leq \ \theta \ &\leq \theta^{\mathbf{U}} \end{aligned} \tag{21}$$

Here $h_{j(\theta)}$ represents the N_h equality constraints, while $g_{k(\theta)}$ represents the N_g inequality constraints. The lower and upper bounds on the set of variables (θ) are given by $\theta^{\mathbf{L}}$ and $\theta^{\mathbf{U}}$, respectively.

Furthermore, the formulation given by (20) corresponds to a special case of (21) in which there are no inequality constraints. This problem can first be transformed to a min-max problem and then reformulated as a global optimization problem with the introduction of a single slack variable, s.

$$\min_{\theta, \mathbf{s}} s \tag{22}$$

$$
\begin{aligned}
\text{subject to} \quad \frac{\partial E(\theta)}{\partial \theta_i} \quad &- \quad s \leq 0 \ , \qquad i = 1, \ldots N_\theta \\
-\frac{\partial E(\theta)}{\partial \theta_i} \quad &- \quad s \leq 0 \ , \qquad i = 1, \ldots N_\theta \\
\theta^{\mathbf{L}} \leq \quad \theta \quad &\leq \theta^{\mathbf{U}}
\end{aligned}
$$

For this formulation, the solution requires the global minimization of the slack variable, s, with respect to s and the dihedral angles, θ. The constraints are infeasible for all negative values of the slack variable, s. Therefore, the global solutions for this formulation require $s = 0$ provided that the original set of equations has a feasible solution. If the global optimum exhibits a nonzero value for s, the system of equations given by Equation (20) does not possess a solution. In order to find all local minima for (20), all global minima of the formulation given in (22) must be identified.

The identification of all multiple global solutions requires the use of a deterministic global optimization method. Along these lines, the αBB approach has been applied to a variety of problems generalized by equations in (21)[43]. Successful identification of all solutions hinges on effective subdivision of the conformational space. The possibility of a given subdomain containing a global solution is explored by locating a lower bound on the global minimum in that region. These lower bounds are developed by solving a relaxed optimization problem in each subdomain. Each problem requires the construction of convex underestimators with expanded feasibility regions, which can be accomplished by adding the appropriate separable quadratic terms to the nonlinear constraints. The formulation becomes :

$$
\begin{aligned}
&\min_{\theta, \mathbf{s}} s \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (23) \\
\text{subject to} \quad \frac{\partial E(\theta)}{\partial \theta_i} \quad &+ \quad \sum_{l}^{N_\theta} \alpha_{l,i}^{+} \left(\theta_{\mathbf{l}}^{\mathbf{L}} - \theta_{\mathbf{l}}\right) \left(\theta_{\mathbf{l}}^{\mathbf{U}} - \theta_{\mathbf{l}}\right) - s \leq 0 \ , \qquad i = 1, \ldots N_\theta \\
-\frac{\partial E(\theta)}{\partial \theta_i} \quad &+ \quad \sum_{l}^{N_\theta} \alpha_{l,i}^{-} \left(\theta_{\mathbf{l}}^{\mathbf{L}} - \theta_{\mathbf{l}}\right) \left(\theta_{\mathbf{l}}^{\mathbf{U}} - \theta_{\mathbf{l}}\right) - s \leq 0 \ , \qquad i = 1, \ldots N_\theta \\
\theta^{\mathbf{L}} \leq \quad \theta \quad &\leq \theta^{\mathbf{U}}
\end{aligned}
$$

The α's ($\alpha_{l,i}^{+}$ and $\alpha_{l,i}^{-}$) are selected in order to guarantee the convexity of the constraint functions within the particular subdomain. If the solution of the lower bounding problem results in $s > 0$, the region is fathomed because $s = 0$ cannot be a feasible solution to the

original formulation. In contrast, when $s \leq 0$ for the lower bounding problem, the region must be further subdivided until all regions are considered.

Both methods for rigorously locating all local minimum energy conformations have some disadvantages. On one hand, the first approach should effectively locate low energy conformers in order of increasing energy. However, locating each minimum requires the solution of a full global optimization problem. The second approach avoids this drawback because it can be solved as a single global optimization problem. However, when dealing with a high dimensional search space, the number of necessary subdivisions may be computationally inhibitive. In addition, this method will potentially locate stationary points other than local minima. Therefore, the development of other methods for locating low energy local minimum energy conformations were pursued in this work.

5 Ensemble of Local Minimum Energy Conformations

5.1 Energy Directed Approach (EDA)

Since the number of local minima on a given energy hypersurface may become astronomically large (e.g., the number of local minima for met-enkephalin is estimated to be on the order of 10^{11} [44]), methods that do not necessarily provide all local minima were developed. Specifically, it was determined that the generation of ensembles of low energy conformers is possible through algorithmic modifications of the general αBB procedure. Rigorous implementation of the global optimization algorithm requires the minimization of a *convex* lower bounding function in each domain. The unique solution θ for each lower bounding minimum can then used as a starting point for the minimization (or function evaluation) of the original energy function in the current domain. In the case of local minimization, each partitioned region provides a single minimum energy conformation as the algorithm proceeds. Using this information, along with the global minimum energy conformation, a list of low energy conformers can be constructed.

A method for increasing the number of local minima produced within each subdomain would involve the selection of multiple random starting points for minimizing the upper bounding function. At first, this approach appears to be equivalent to choosing random

points for local minimization. Initially, when the subdomains constitute significant portions of the original domain space, this is the case. However, as the separation between lower and upper bounds decreases, the subdomains are localized in regions of low energy. Therefore, the random point selection is localized in regions which contain low energy local minima.

However, this approach does not take advantage of the information provided by the lower bounding functions. Rigorously, these functions posses a single minimum in each subdomain. Since the choice of α affects the convexity of the lower bounding functions, the α values can be modified to ensure a certain nonconvexity in these functions. In this case, the lower bounding functions possess multiple minima, and these functions can be minimized several times in each domain. In addition, since the lower bounding functions smooth the original energy hypersurface, the location of these multiple minima provide information on the location of low energy minima for the upper bounding function. Therefore, by using the location of the minima of the lower bounding function as starting points for local minimization of the upper bounding function, an improved set of low energy conformations can be identified. As before, these conformations are also localized in those domains with low energy as the subdomains decrease in size. This energy directed approach (EDA) is represented schematically in Figure 3.

5.1.1 Algorithmic Description

The determination of local (and global) minimum energy conformations using αBB requires the interfacing of a number of programs : αBB[5,6,7,8,9], PACK[45], NPSOL[46] and potential and solvation energy modules. PACK, a peptide generation program, is called once directly by αBB in order to initialize the current problem. In subsequent steps PACK is called through NPSOL[46], a local nonlinear optimization solver used to locally solve both the upper and lower bounding problems[47]. PACK internally transforms to and from Cartesian and internal coordinate systems, and provides potential energy and gradient contributions for the ECEPP/3[24] potential model at every step of the local minimizations. When considering solvation energy using solvent accessible volume of hydration layers, the RRIGS module is also used[26]. Finally, an additional module, UBC (Upper Bound Check), is used to verify the quality of the upper bound solutions. The overall interface is shown schematically in

Figure 4.

The basic steps of the algorithm are as follows:

(1) The initial best upper bound is set to an arbitrarily large value. The original domain is partitioned along one of the global variables. α values are initially chosen to be constant ($\alpha = \alpha_0$) for all global variables.

(2) The lower bounding function (L) is constructed in each hyper–rectangle. Three local minimization are performed using the following procedure :

(A) 50 random points are generated and used for function evaluations.

(B) The point with the minimum value is used as a starting point for local minimization of L using NPSOL, with calls (through PACK) to ECEPP/3 and possibly the RRIGS solvation module.

(C) The unique solutions are stored

If the minimum valued solution (of all local minima of L in this subdomain) is greater than the current best upper bound the subdomain is fathomed.

(3) The unique local minima (points) for L are used as initial starting points for local minimizations of the upper bounding function (E) in each hyper–rectangle. Again, the appropriate calls are made to PACK and the potential and solvation energy modules. Two additional minimizations are performed using the following procedure :

(A) 50 random points are generated and used for function evaluations.

(B) The point with the minimum value is used as a starting point for local minimization of E using NPSOL, with calls (through PACK) to ECEPP/3 and possibly the RRIGS solvation module.

In all cases, the UBC (upper bound check) module is also called. UBC checks that the absolute value of each gradient in the objective function gradient vector is below a specified tolerance (10^{-6} kcal/mol/deg). If a gradient does not satisfy this check the corresponding variable bounds are incrementally increased and the problem is solved

with the previous point used as the initial starting point. This process is repeated until the gradient constraints are satisfied or an iteration limit is exceeded. UBC also employs algorithms to calculate the second derivative matrix[48], which is used to verify that the upper bound solution is a local minimum; that is, the Hessian matrix must be positive semi–definite. If the matrix is not positive semi–definite or the gradient checks are not satisfied, the upper bound solution is rejected. All local minima are stored.

(4) The current best upper bound is updated to be the minimum of those thus far stored.

(5) The hyper–rectangle with the current minimum value for L (this is the minimum value of all local minima of L in each subdomain) is selected and partitioned along one of the global variables. All α values are updated according to the following rule:

$$\alpha = \alpha_0 R^L \tag{24}$$

In this equation α_0 refer to the initial values from Step 1. R is a reduction parameter ($0 < R \leq 1$), and L refers to the current level in the branch and bound tree. For $R = 1$ the α values are kept constant at the initial value, α_0.

(6) If the best upper and lower bounds are within the ϵ tolerance, or a maximum iteration limit has been exceeded, the program will terminate, otherwise it will return to Step 2.

5.2 Free Energy Directed Approach (FEDA)

A second approach incorporates free energy information into the branch and bound algorithm. Specifically, harmonic entropic contributions are calculated and included at each minima of the upper and lower bounding functions. In this way, the progression of lower and upper bounds includes a temperature dependent entropic term. A similar modification to the Monte Carlo minimization method has also been proposed[49], and has been shown to be effective in locating low energy conformers of peptides[50,51].

The problem formulation is identical to the one given in (17). That is, the minimization of E and L are still performed using only potential and solvation energy contributions.

However, once local minima have been located, the free energy is calculated by the following expression:

$$G = U_{Min} + \frac{1}{2\beta} \ln \left[Det \left(He_{Min} \right) \right] \tag{25}$$

This equation is similar to Equation (11), although the additive term $f(T)$ has been omitted because it is a function of temperature only. U_{Min} represents the local minimum energy of E or L, and $Det\left(H_{Min}\right)$ is the determinant of the Hessian evaluated at this local minimum. The specification of a thermodynamic temperature ($\beta = \frac{1}{k_B T}$) is required as an additional input parameter.

A single rigorous application of the αBB algorithm to this problem will result in the identification of the global minimum free energy at a given temperature. However, the goal of this work is to identify an ensemble of low energy and, in this case, low free energy conformers so that a free energy ranking and comparison can be made. Therefore, the algorithmic steps for the Free Energy Directed Approach (FEDA) are similar to those for EDA, with the additional evaluation of the free energy (G) at each local minima of E and L. The thermodynamic temperature used in Equation (25) must be specified as an additional input parameter.

6 Computational Studies

6.1 Isolated Met-enkephalin : EDA

The EDA was first applied to the isolated form of met-enkephalin. Met–enkephalin (H–Tyr–Gly–Gly–Phe–Met–OH) is an endogenous opioid pentapeptide 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). For both peptides, the EDA algorithm detailed in Section 5.1.1, was applied 10 times. The input conditions used for these runs are given in Table I[52].

Once the ensemble of local minima had been compiled, a set of distinct conformations was identified by checking for repeated and symmetric conformations. In addition, a conformation was only considered unique if at least one dihedral angle differed by at least 50^o when comparing each pair of conformations. These conformations were then used to generate results and distributions according to energy and free energy values. Energy bins were used to characterize a group of distinct structures between a range of energy values (every 0.5 kcal/mol) relative to the global minimum energy structure. For example, Bin 1 contains structures that are 0.0-0.5 kcal/mol above the global minimum energy structure, Bin 2 contains structures that are 0.5-1.0 kcal/mol above the global minimum energy structure, etc.

In the case of isolated met-enkephalin, the 10 (EDA) runs generated a total of 83908 distinct local minima. The potential energy global minimum (PEGM) conformation for met-enkephalin possesses an energy of -11.707 kcal/mol. This conformation exhibits a type II' β–bend along the N–C' peptidic bond of Gly^3 and Phe^4. Essentially, this structure corresponds to the free energy global minimum (FEGM) conformation for a temperature of 0 K, that is, when entropic contributions are not included. When considering the harmonic free energy, the prediction of the FEGM can be calculated over a range of temperatures. Table II provides information on the FEGM for temperatures ranging from 100 K to 500 K.

As Table II shows, the PEGM persists as the FEGM at a temperature of 100 K. However, at the next three temperature points (i.e., 200 K, 300 K, 400 K) the FEGM exhibits a potential energy contribution 1.808 kcal/mol higher than the PEGM. The ϕ and ψ values for this structure are also significantly different than those for the PEGM. In fact, the conformational code (B*AAAE) indicates that the central residues display an α helical configuration. At a temperature of 500 K, the FEGM structure changes again, while the potential energy difference between the FEGM and PEGM increases to 5.369 kcal/mol. These differences suggest that the inclusion of entropic contributions greatly affects the relative stability of individual low energy structures. In addition, as the temperature increases, the stability offered by entropic contributions offsets substantial differences in potential energy.

Table III provides information on the distribution of distinct low free energy minima within 8.0 kcal/mol of the FEGM for a range of temperatures. For a given temperature the

general trend indicates a large increase in the number of minima as the free energy increases above the FEGM. Several exceptions to this trend occur at high temperature and large bin number. In these cases, the number of minima remains constant or even decreases slightly. This is most likely due to an inadequate sampling of higher potential energy minima. For a given bin, it is also apparent that the clustering of low free energy structures increases with temperature. This increased density of the free energy bins indicates that increases in energy are offset by entropic contributions.

These observations are also supported by the information shown in Figure 5. This plot displays the range of potential energy in free energy bins at temperatures of 250 and 500 K, with the potential energy bins included for comparison. As expected, the potential energy values for the free energy bins increase with increasing temperature. In addition, the range of potential energy values increases in higher free energy bins. It is interesting to note that this occurs because the minimum potential energy is relatively (i.e., within a few kcal/mol of the PEGM) low for each bin, whereas the maximum potential energy value increases in higher bins. The corresponding differences are also greater at higher temperature. For example, at 500 K some bins exhibit a 20 kcal/mol range in potential energy. These trends explain the increased number of low free energy conformers. That is, bins of low free energy contain conformers of relatively high potential energy because of their more stabilizing entropic contributions. The plot also implies that the PEGM appears in bins 3 and 10 for temperatures of 250 and 500 K, respectively.

Relative free energies were also calculated for clusters of low energy conformers. This analysis is useful because it is difficult to capture the true accessibility of individual structures based on a point-wise approximation of entropic effects. That is, the harmonic free energy approximation does not provide a continuous free energy landscape. By clustering structures into larger groups, it is hoped that the error associated with these estimates will average out. Typically, structures are clustered by calculating and comparing root mean squared deviations. In this work, since the enkephalin peptide is relatively small, structures were grouped based on the Zimmerman codes for the central residues of the peptide [53]. Specifically, for met-enkephalin, structures were said to belong to the same cluster if the central 3 residues possessed the same 3 code letters based on the Zimmerman classification

[53]. The relative free energy of a cluster was calculated by the following equation :

$$G_{cluster} = -\frac{\ln \sum_{i \in C} p_i^{approx}}{\beta} \tag{26}$$

In Equation (26) the individual p_i^{approx}, which refers to the statistical weight based on the harmonic approximation, are summed for the set of conformations belonging to a particular cluster (C). These individual probabilities were calculated by normalizing each probability with respect to the overall probability at a given temperature :

$$p_i^{approx} = \frac{\exp[-\beta(G_o^{approx} - G_i^{approx})]}{\sum_j \exp[-\beta(G_o^{approx} - G_j^{approx})]} \tag{27}$$

A reference free energy, G_o^{approx}, was used to normalize the probabilities at each temperature point. All free energies, G_o^{approx}, G_i^{approx} and G_j^{approx}, refer to the harmonic approximation of the free energy as calculated using Equation (11). The denominator, which represents the total probability at a given temperature, is calculated by summing over the set of all conformers.

The relative free energies for clusters of met-enkephalin structures are given in Table IV. At each temperature point the Zimmerman code and corresponding data for the top 3 clusters are listed. The results indicate that the structure exhibiting the individual lowest free energy does not always belong to the cluster with lowest free energy. At 100 and 200 K the DC*B and AAA clusters are consistent with the structures of the FEGM. However, although the FEGM retains the AAA structure at 300 and 400 K, the group of structures possessing the lowest $G_{cluster}$ at these temperatures exhibits a CD*A Zimmerman code. This is, at least in part, attributable to the large number of structures grouped in this cluster. In contrast to the α-helical type structure for the FEGM, the CD*A structures possess elements of a β-turn conformation. Specifically, the lowest free energy conformer exhibiting a CD*A structure at 300 and 400 K, possesses a type II β-bend along the Gly^2–Gly^3 backbone.

6.2 Isolated Met-enkephalin : FEDA

FEDA was also applied to the isolated form of met-enkephalin. For this approach, the thermodynamic temperature appears as an input parameter, and these values had to be specified along with initial α values. Several methods can be envisioned for initializing the

FEDA. For example, if the goal is to characterize the low free energy conformers at a single temperature, a full set of FEDA runs could be performed for that temperature. This type of search should efficiently locate the global and many low free energy conformers for that temperature. However, in this work the goal was to effectively characterize the FEGM and low free energy conformers over a range of temperatures. Therefore each of the 10 (FEDA) runs were conducted at a unique temperature point in the range of 50 to 500 K. The details of the conditions for these runs are given in Table V.

In total, 87974 distinct local minima were found after compiling the results from the 10 (FEDA) runs for isolated met-enkephalin. The PEGM and FEGM found using the FEDA are displayed in Table VI. It should be noted that when comparing PEGM for the EDA and FEDA, both structures possess the same potential energies, but a different set of dihedral angles. However, these structures are actually the same. That is, the different values of χ_2 and χ_3 for Tyr_1 represent a degenerate state for tyrosine, which is generated by rotating both of these dihedral angles by 180^o. An important observation is that at 200 K the FEDA method predicts a slightly lower FEGM. The structure possesses a lower potential energy (-10.547 vs. -9.899 kcal/mol) and exhibits a free energy value that is 0.044 kcal/mol lower than the EDA predicted FEGM. The remaining FEGM predictions are consistent for the two approaches.

An analysis of the distribution of distinct minima, as given by Table VII, reveals that the results are qualitatively consistent with those produced by the EDA. It should be noted that in all cases, the lowest free energy bin is as populated as the corresponding EDA bins, which indicates that each run using the FEDA was able to find a better distribution of low free energy conformers near the FEGM. This is not unexpected, considering that the FEDA runs were conducted at the same discrete temperature points used in the analysis. However, when comparing the populations of higher energy bins at low temperatures, the number of minima are larger for the EDA. Some of this variation, especially near the 150 to 200 K range is probably due to the lower FEGM found by the FEDA. In general, the FEDA seems to provide a denser distribution of distinct minima at higher temperatures and large bin number.

A comparison of the relative efficiencies for the EDA and FEDA to generate low energy

local minima can also be made by examining Figure 6. In this plot the cumulative fraction of conformers, which is equal to the total number of unique conformers within the first 8, 12 and 16 energy bins over the total number of unique conformers, is given as a function of temperature. It is apparent that both approaches are highly efficient. For example, at 400 K approximately 90 percent of the total unique conformations identified are in the top 16 free energy bins, which ranges up to 8 kcal/mol above the FEGM. The lower fractions at lower temperatures indicate that a relatively large number of conformations have high potential energies, and that these energetic differences are not offset by entropic effects at low temperatures. A more subtle comparison can be made by observing that the EDA cumulative fractions are generally higher for temperatures lower than 400 K. Although the total number of unique conformations is slightly lower for the EDA, this trend indicates that the EDA is more efficient at filling low energy bins, especially at lower temperatures.

The results for the cluster analysis of the FEDA met-enkephalin structures are given in Table VIII. There are some differences between the EDA and FEDA cluster free energies, although the overall trend is the same. At all temperatures, excluding 200 K, the cluster exhibiting the lowest cluster free energy is the same as in the EDA analysis. At 200 K, the FEDA predicts the AAA cluster as having a slightly higher free energy than the C*DE cluster , which only appears as the third cluster in Table IV. In both analyses, the transition from the ground state DC*B cluster to the CD*A cluster as temperature increases, is evident.

Since both the EDA and FEDA provide large amounts of statistical information for the peptide system, this data was used to perform a simple thermodynamic analysis of the folding process. It is widely accepted that the folding of peptides progresses successively. The first step of this process is typically associated with a structural collapse, that is, a transition from random extended structures to an ensemble of compact structures. This transition should also be associated by significant changes in the description of the ensemble as temperature changes. For example, a peak in the specific heat at the transition temperature indicates a steep decrease in average potential energy of the ensemble. In order to verify that such a transition occurs for met-enkephalin, the specific heat was calculated using the following expression :

$$C = \frac{\beta^2 \left(< E^2 >_T - < E >_T^2 \right)}{N} \tag{28}$$

Here N refers to the number of amino acid residues in the peptide. The average energy and squared energy ($< E >_T$ and $< E^2 >_T$, respectively) were calculated at 10 temperature points using expressions of the form given in Equation (16). The bin probabilities were based on an energy width of .015625 kcal/mol. In addition, a reference free energy, G_o^{approx} (the lowest free energy), was used to normalize the probabilities at each temperature point.

The results for isolated met-enkephalin are shown in Figure 7. Both the EDA and FEDA predict a transition temperature in the 250-275 K temperature range. This is consistent with the increase in bin density and structural diversity at higher temperatures, and suggests a sharp increase in the average potential energy of the system at this temperature. It also supports the transition from the DC*B ground state (PEGM) cluster to the higher potential energy CD*A cluster in this temperature range.

Similar results for characterizing the folding transitions of enkephalins have also been obtained by multicanonical simulations[54]. This is encouraging because the two methods possess fundamental differences. In contrast to this work, the multicanonical approach does not rely on the identification of low energy local minima or the concepts of the harmonic approximation. Instead, thermodynamic quantities are developed by first generating large ensembles of structures with wide ranging energies and then employing reweighting techniques. In addition, although the multicanonical simulations included detailed atomistic level modeling, only unsolvated systems were considered.

6.3 Isolated Leu-enkephalin : EDA

A similar free energy analysis was performed for isolated leu–enkephalin (H–Tyr–Gly–Gly–Phe–Leu–OH), an endogenous pentapeptide in which the methionine residue of met-enkephalin has been replaced by a leucine residue. The peptide also possess 24 independent torsional angles, and a total of 77 atoms. Like met-enkephalin, the PEGM for isolated leu-enkephalin exhibits a type II' β-bend, although the bend is now shifted to the N–C' peptidic bond of Gly^2–Gly^3. For this example, the 10 (EDA) runs produced a total of 55414 unique local minima, which were then checked and ranked based on the set of distinct minima. Information on the FEGM over a range of temperature values is given in Table IX. The results exhibit some important differences from those for met-enkephalin. Most notably, the PEGM

structure persists as the FEGM over a wider range of temperatures. This indicates that the PEGM for isolated leu-enkephalin occupies a wider basin, which is supported by the value of $ln[Det\,(H)] = 84.66$. In addition, when the FEGM structure becomes different at temperatures of 400 and 500 K, these structures exhibit a β-turn conformation. That is for both FEGM, a type II β-bend occurs around the N–C' peptidic bond of Gly^2–Gly^3.

The data in Table X confirm these observations. Specifically, the distribution of distinct minima indicates that free energy bin populations at low free energy are smaller than the corresponding potential energy bin populations. This is related to the combination of the lowest potential energy contributions with relatively large (less negative) entropic contributions for both the PEGM and a group of related low energy structures. Only at approximately 3 kcal/mol (about 6 bins) does the trend of increasing bin density with increasing temperature become evident. The decrease in bin populations at large bin numbers and high temperature is also more apparent for this example.

Potential energy comparisons also reveal the low free energy characteristics of the PEGM. A plot of the minimum and maximum potential energies for free energy bins at temperatures of 0, 250 and 500 K are given in Figure 8. In contrast to met-enkephalin, the minimum and maximum potential energy values are the same for the first free energy bins at both 0 and 250 K. This indicates that at 250 K the lowest potential energy structures also have the lowest free energy values. Although this is not the case at 500 K, the narrow range of energy values does suggest a small cluster of low free energy structures at this temperature. Other observations are more consistent with the results for met-enkephalin. For example, the range of potential energy values increases with increasing free energy. In addition, a significant number of low energy minima occupy high energy bins, which is evidenced by the relatively low minimum energy values at both 250 and 500 K. This indicates that although the PEGM exhibits a relatively wide minima, other minima with low potential energy are much narrower. For example, a low energy conformer with a potential energy of -9.333 kcal/mol (just 0.16 kcal/mol higher than the PEGM), occupies bin 6 (rather than bin 1 for the PEGM) at 250 K.

In order to determine the overall accessibility of the cluster containing the PEGM conformer, a cluster analysis was also performed for isolated leu-enkephalin. The results are

given in Table XI. The C*DE cluster, which contains the PEGM, persists as the dominant cluster at both 100 and 200 K. However, at 300 K the lowest free energy cluster does not include the FEGM. In fact, although the C*DE cluster still has relatively low free energy values at higher temperatures, the CD*A cluster provides the lowest free energy at 300, 400 and 500 K. This cluster, which includes a type II β-bend along the Gly^2–Gly^3 backbone, is identical to dominant cluster for isolated met-enkephalin at the same temperatures. These results suggest that the CD*A cluster is highly accessible basin for both forms of isolated enkephalin.

6.4 Isolated Leu-enkephalin : FEDA

FEDA runs were also conducted for isolated leu-enkephalin using the protocol outlined in Section 6.1. These runs produced a total of 57723 distinct local minima. At all temperature points between 50 and 500 K the FEGM were found to be identical (with the exception of symmetric shifts in certain χ angles) between the EDA and FEDA. In addition, the FEDA again predicts as many, if not slightly more, distinct minima for the lowest free energy bin. However, in general the EDA still performed better for low temperatures and intermediate bins; although at intermediate temperatures, the methods are comparable. As expected, at high temperatures and at large bin number the FEDA again provides more populated bins. This is evidenced by the data plotted in Figure 9. As with isolated met-enkephalin, both approaches are also efficient in locating low free energy conformers. However, in contrast to met-enkephalin, the higher fractions at lower temperatures indicate that a larger number of conformers have both relatively low potential and free energy at these temperatures.

A clustering analysis of the FEDA low energy minima was also performed, and the results provided qualitative agreement with those from the EDA analysis (shown in Table IV). In order to develop a complete description of the leu-enkephalin system, specific heat calculations were then conducted so as to identify the transition temperature. The results for isolated leu-enkephalin are shown in Figure 10. Both the EDA and FEDA predict a transition temperature near 275 K, which is consistent with the met-enkephalin results shown in Figure 7.

6.5 Solvated Met-enkephalin

The EDA was then applied to the RRIGS solvated form of met-enkephalin using the same protocol and conditions as detailed in Section 6.1. Qualitatively, the PEGM (in this case PEGM refers to potential+solvation) for solvated met-enkephalin exhibits a more extended conformation than that which is observed for the isolated form. As detailed in Table XII, the PEGM structure persists as the FEGM at 100 K. However, at each subsequent temperature, the FEGM structure changes, and this change is accompanied by an increase in total energy (potential and solvation). As with isolated met-enkephalin, the difference in total energy between the PEGM and FEGM at 500 K is greater than 5 kcal/mol. This suggests that entropic effects are important in defining the predicted native structure. When considering individual structures, entropic effects tend to produce more extended FEGM conformations at higher temperatures, especially with regard to the placement of the aromatic rings. It is interesting to note that in a previous study the positioning of aromatic rings was found to be a major difference when considering the ability of solvation models to predict extended PEGM conformations for the solvated enkephalin peptides[11]. The sequence of FEGM structures is illustrated in Figure 11.

The distribution of the 72784 distinct minima for solvated met-enkephalin exhibits some important differences from those results obtained for the isolated form of the peptide. This is evidenced by the information presented in Table XIII and the plot in Figure 12. In particular, the low and intermediate energy bins are much denser than the corresponding bins for isolated met-enkephalin, especially within 4 kcal/mol (8 bins) of the FEGM. In addition, some higher energy bins are actually more populated at lower temperatures. One obvious reason for these differences is the high density of conformers for the original system (at 0 K). This high density of states causes the original energy differences to be relatively small, and the entropic correction tends to induce an even stronger equalization of the free energy values. This equalization is best illustrated by the data plotted in Figure 12, which indicates that the efficiency of locating low free energy conformers is relatively high at all temperatures. In fact, the highest density of states occurs near the middle of the temperature range, rather than at high temperatures as predicted for the isolated peptide. This behavior

may be due to a lack of much higher energy local minima which would probably populate these high temperature, high energy bins.

Similar conclusions can be drawn by examining the data presented in Figure 13, which provides information on the energy extrema for free energy bins at temperatures of 0, 250 and 500 K. As expected, for both 250 and 500 K, the range of energy values increases for higher free energy bins. In addition, for all bins, the minimum energy is relatively low and generally within a few kcal/mol of the PEGM. However, unlike the isolated met-enkephalin results, the maximum energy values do not become larger at higher temperatures. In fact, the curves for maximum energy at 250 and 500 K are almost identical. This indicates that relatively high energy minima may be needed in order to fill out these high temperature bins.

A clustering analysis of the low free energy conformers was also performed for solvated met-enkephalin, and the results are shown in Table XIV. At 100 K, the lowest free energy cluster included the FEGM structure, which is also the PEGM structure. At higher temperatures, the correlation between the extended FEGM structures and the lowest free energy cluster was also evident. In fact, all low energy clusters at 300, 400 and 500 K possess highly extended backbone conformations, with nearly all geometries within the E and E* regions on the Zimmerman conformational map. In fact, although the number of individual structures in each cluster is not excessively large, many of these extended conformers reside in the lowest free energy bins.

A specific heat profile was also derived for solvated met-enkephalin in order to understand how the dominance of these extended cluster geometries affect the folding transition. These results are show in Figure 14. As with isolated met-enkephalin, a folding transition is indicated by the peak in the specific heat, which, in this example, occurs between 275 and 300 K. This represents a significant change in average energy, which accompanies the collapse from an ensemble of extended conformations (EE*E and E*EE clusters) to the more compact ground state cluster. For the solvated met-enkephalin example, this transition is clearly illustrated by the cluster analysis and the structure plots given in Figure 11.

6.6 Solvated Leu-enkephalin

An analogous set of runs, conducted for the RRIGS solvated form of leu-enkephalin, provided a total of 60288 distinct local minima. As with solvated met-enkephalin, the PEGM for solvated leu-enkephalin is characterized by the extended backbone structure and the proximity of its two aromatic rings. Data for the FEGM over the 100-500 K temperature range are given in Table XV. Although the PEGM persists as the FEGM at a temperature of 100 K, the FEGM at higher temperatures adopt more extended structures. In fact, since all central residues for these FEGM reside in the E and E* regions of the Zimmerman conformational map, these conformations are extremely similar to the corresponding solvated met-enkephalin FEGM.

This similarity between the enkephalin results also extends to the information presented in Figure 15. As with solvated met-enkephalin, these data show that the free energy distribution provides a high density of conformers at low and intermediate energy bins. This is also evidenced by relatively high efficiencies for locating low free energy structures at all temperatures, as compared to the results for isolated leu-enkephalin. In addition, the decrease in densities at high temperature and high bin energies is consistent with the solvated met-enkephalin results, and reflects the role of entropic effects in equalizing the free energy values.

The results of the clustering analysis of the low free energy conformers for solvated leu-enkephalin are shown in Table XVI. At 100 K, the lowest free energy cluster included the FEGM structure, which again corresponds to the PEGM. At all higher temperatures the dominant cluster possesses a E*EE or EE*E structure, which are similar to the set of FEGM structures at these temperatures. In contrast to the cluster analysis for solvated met-enkephalin, the the E*EE appears to be the dominant cluster even at 200 K. In spite of these minor differences, the results for both solvated enkephalins follow the same qualitative trends.

Finally, the results for the transition temperature prediction are shown in Figure 16. As with solvated met-enkephalin, this transition is signified by a peak in the specific heat at approximately 275 K. The ability to clearly identify the folding transition for both isolated

and solvated enkephalin systems suggests that the analysis could be generalized to other small peptide systems.

7 Conclusions

In this work a novel method was developed for locating the *free energy* global minimum (FEGM) conformation of peptides within the context of the deterministic global optimization algorithm, αBB. The inclusion of entropic effects was accomplished by employing an harmonic approximation; a method that relies on the ability to generate an adequate ensemble of distinct low energy local minima. In order to generate large numbers of low energy metastable states two independent methods were implemented. The first method, EDA, attempts to use information offered by the lower bounding function to find low energy local minima. The second method, FEDA, expands on the first approach by incorporating free energy information into the overall algorithm. Both approaches are based on the general concepts of the αBB branch and bound framework, and rely on the algorithm's ability to identify regions, rather than points, of low energy. Another important feature is the use of lower bound information in the generation of dense ensembles of low energy local minima.

The EDA and FEDA methods were then tested on isolated and solvated forms of the peptides met- and leu-enkephalin. The results indicated that both approaches produced dense distributions of low energy local minima[55]. These data were then used to perform detailed free energy rankings over a wide range of temperatures. In general, the inclusion of entropic effects was found to influence the prediction of both the individual free energy global minimum structures, as well as the accessibility of clusters of low energy structures. The statistical data generated during these analyses were also used to perform a simple thermodynamic analysis of the folding process. It was found that each system could be characterized by a unique transition temperature, which signified the collapse of relatively high energy, extended structures to an ensemble of ground-state-like structures.

8 Acknowledgments

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

References

[1] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White, *J. Proc. Nat. Acad. Sci. USA*, 47, 1309–1314 (1961).

[2] C. A. Floudas, J. L. Klepeis, and P. M. Pardalos, In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, 1999, (in press).

[3] A. Neumaier, *SIAM Rev.*, 39, 407–460 (1997).

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

[5] C. S. Adjiman, I. P. Androulakis, C. D. Maranas, and C. A. Floudas, *Comput. Chem. Eng.*, 20, S419–S424 (1996).

[6] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas, *Comput. Chem. Eng.*, 21, S445–S450 (1997).

[7] C. S. Adjiman, S. Dallwig, C. A. Floudas, and A. Neumaier, *Comput. Chem. Eng.* (1998), (in press).

[8] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas, *Comput. Chem. Eng.* (1998), (in press).

[9] I. P. Androulakis, C. D. Maranas, and C. A. Floudas, *J. Glob. Opt.*, 7, 337–363 (1995).

[10] I. P. Androulakis, C. D. Maranas, and C. A. Floudas, *J. Glob. Opt.*, 11, 1–34 (1997).

[11] J. L. Klepeis, I. P. Androulakis, M. G. Ierapetritou, and C. A. Floudas, *Comput. Chem. Eng.*, 22, 765–788 (1998).

[12] J. L. Klepeis and C. A. Floudas, *J. Comp. Chem.*, (accepted for publication), 1999.

[13] F. A. Momany, L. M. Carruthers, R. F. McGuire, and H. A. Scheraga, *J. Phys. Chem.*, 78, 1595–1620 (1974).

[14] F. A. Momany, L. M. Carruthers, and H. A. Scheraga, *J. Phys. Chem.*, 78, 1621–1630 (1974).

[15] F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, 79, 2361–2381 (1975).

[16] S. Weiner, P. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, 106, 765–784 (1984).

[17] S. Weiner, P. Kollman, D. Nguyen, and D. Case, *J. Comp. Chem.*, 7, 230–252 (1986).

[18] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.*, 4, 187–217 (1983).

[19] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Proteins*, 4, 31 (1988).

[20] W. F. van Gunsteren and H. J. C. Berendsen, *GROMOS*, Groningen Molecular Simulation, Groningen, The Netherlands, 1987.

[21] N. L. Allinger, Y. H. Yuh, and J. H. Lii, *J. Am. Chem. Soc.*, 111, 8551–8565 (1989).

[22] M. Levitt, *J. Mol. Biol.*, 170, 723–764 (1983).

[23] G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, 87, 1883–1887 (1983).

[24] G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, *J. Phys. Chem.*, 96, 6472–6484 (1992).

[25] B. Honig, K. Sharp, and A. Yang, *J. Phys. Chem.*, 97, 1101–1109 (1993).

[26] J. D. Augspurger and H. A. Scheraga, *J. Comp. Chem*, 17, 1549–1558 (1996).

[27] A. J. Hopfinger, *Macromolecules*, 4, 731–737 (1971).

[28] Y. K. Kang, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.*, 91, 4105–4109 (1987).

[29] Y. K. Kang, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.*, 91, 4109–4117 (1987).

[30] Y. K. Kang, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.*, 91, 4118–4120 (1987).

[31] Y. K. Kang, K. D. Gibson, G. Némethy, and H. A. Scher aga, *J. Phys. Chem.*, 92, 4739–4742 (1988).

[32] N. Go and H. A. Scheraga, *J. Chem. Phys.*, 51, 4751–4767 (1969).

[33] N. Go and H. A. Scheraga, *Macromolecules*, 9, 535–542 (1976).

[34] P. J. Flory, *Macromolecules*, 7, 381–392 (1974).

[35] The temperature (only) dependent terms ($\hat{f}(T)$, $\bar{f}(T)$, $\tilde{f}(T)$, $f(T)$) that appear in Equations 9-13 cancel because the methods involve only relative entropies and free energies. Therefore, the exact forms are not necessary.

[36] C. D. Maranas, I. P. Androulakis, and C. A. Floudas, In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 23, pages 133–150. American Mathematical Society, 1996.

[37] C.A. Floudas, In L.T. Biegler, T.F. Coleman, A.R. Conn, and F.N. Santosa, editors, *Large Scale Optimization with Applications, Part II: Optimal Design and Control*, volume 93, pages 129–184. IMA Volumes in Mathematics and its Applications, Springer–Verlag, 1997.

[38] C. D. Maranas and C. A. Floudas, *J. Chem. Phys.*, 97, 7667–7677 (1992).

[39] C. D. Maranas and C. A. Floudas, *Annals of Operations Research*, 42, 85–117 (1993).

[40] C. D. Maranas and C. A. Floudas, *J. Chem. Phys.*, 100, 1247–1261 (1994).

[41] C. D. Maranas and C. A. Floudas, *J. Glob. Opt.*, 4, 135–170 (1994).

[42] C. S. Adjiman and C. A. Floudas, *J. Glob. Opt.*, 9, 23–40 (1996).

[43] C. D. Maranas and C. A. Floudas, *Journal of Global Optimization*, 7, 153–182 (1995).

[44] Z. Li and H. A. Scheraga, *J. Mol. Struct. (Theochem.)*, 179, 333–352 (1988).

[45] H.A. Scheraga, *PACK: Programs for Packing Polypeptide Chains*, 1996, online documentation.

[46] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *NPSOL 4.0 User's Guide*, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, CA., 1986.

[47] The use of NPSOL is for illustrative purposes only. Any local nonlinear minimization package can be substituted.

[48] T. Noguti and N. Go, *J. Phys. Soc. Japan*, 52, 3685–3690 (1983).

[49] M. Vásquez, G. Némethy, and H. A. Scheraga, *Chemical Reviews*, 94, 2183–2239 (1994).

[50] H. Meirovitch and E. Meirovitch, *J. Comput. Chem.*, 18, 240–253 (1997).

[51] H. Meirovitch and M. Vásquez, *J. Mol. Struct. (Theochem.)*, 398-399, 517–522 (1997).

[52] A single rigorous implementation of the algorithm can be used to verify that the global minimum energy structure has been found. However, for the enkephalin examples, the EDA and FEDA runs also identified the corresponding global energy minima.

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

[54] U. H. Hansmann, M. Masuya, and Y. Okamoto, *Proc. Natl. Acad. Sci. USA*, 94, 10652-10656 (1997).

[55] At a given temperature, the density of distinct metastable states was found to follow a Boltzmann-like distribution within 5 kcal/mol of the free energy global minimum.

Table I: Input parameters used for 10 EDA runs. α_0 refers to the initial α values used for all global variables. R refers to the reduction rate applied at each level of the branch and bound tree.

Run No.	α_0	R
1-5	5	0.90
6-10	10	0.80

Table II: Dihedral angle values for PEGM and FEGM structures of isolated met-enkephalin using EDA. The temperatures are provided in the first row. The last two rows indicate the harmonic free energy (kcal/mol) and the potential energy value (kcal/mol), respectively.

Residue	DA	PEGM	100 K	200 K	300 K	400 K	500 K
Tyr_1	ϕ	-83.4	-83.4	179.8	179.8	179.8	90.2
	ψ	155.8	155.8	-18.2	-18.2	-18.2	149.1
	ω	-177.1	-177.1	-178.1	-178.1	-178.1	177.5
	χ_1	-173.2	-173.2	178.2	178.2	178.2	169.8
	χ_2	79.3	79.3	81.3	81.3	81.3	-108.2
	χ_3	-166.3	-166.3	177.3	177.3	177.3	177.6
Gly_2	ϕ	-154.3	-154.3	-59.8	-59.8	-59.8	-66.1
	ψ	85.8	85.8	-37.6	-37.6	-37.6	87.5
	ω	168.5	168.5	-178.8	-178.8	-178.8	-173.4
Gly_3	ϕ	83.0	83.0	-67.0	-67.0	-67.0	147.2
	ψ	-75.0	-75.0	-40.1	-40.1	-40.1	-36.7
	ω	-170.0	-170.0	179.7	179.7	179.7	175.1
Phe_4	ϕ	-136.9	-136.9	-70.9	-70.9	-70.9	-92.5
	ψ	19.1	19.1	-39.5	-39.5	-39.5	-34.7
	ω	-174.1	-174.1	-179.8	-179.8	-179.8	-179.1
	χ_1	58.9	58.9	173.9	173.9	173.9	179.1
	χ_2	94.5	94.5	-102.6	-102.6	-102.6	74.7
Met_5	ϕ	-163.5	-163.5	-161.0	-161.0	-161.0	-154.7
	ψ	160.9	160.9	122.1	122.1	122.1	135.3
	ω	-179.8	-179.8	-178.0	-178.0	-178.0	179.9
	χ_1	52.9	52.9	-174.7	-174.7	-174.7	-172.6
	χ_2	175.3	175.3	174.0	174.0	174.0	175.1
	χ_3	-179.9	-179.9	179.0	179.0	179.0	179.9
	χ_4	-178.6	-178.6	-60.1	-60.1	-60.1	-60.0
G		-11.707	-2.499	6.151	14.175	22.200	29.592
E		-11.707	-11.707	-9.899	-9.899	-9.899	-6.338

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

Bin	0 K	50 K	100 K	150 K	200 K	250 K	300 K	350 K	400 K	450 K	500 K
1	2	1	2	10	6	3	3	4	16	16	8
2	3	5	13	22	12	9	15	24	18	21	31
3	12	25	36	58	52	42	40	40	59	69	77
4	45	48	55	105	105	100	101	115	164	184	184
5	49	69	120	233	199	206	213	249	309	397	475
6	90	125	263	451	435	403	410	491	726	893	918
7	166	292	467	806	763	765	848	1043	1438	1655	1687
8	303	497	766	1250	1297	1362	1524	1906	2464	2821	2695
9	552	776	1233	1929	2079	2247	2601	3069	3932	4284	4111
10	840	1177	1710	2915	3168	3475	3927	4707	5774	6030	5562
11	1121	1675	2681	3879	4355	4899	5708	6655	7573	7775	7116
12	1618	2467	3526	5303	5935	6572	7364	8333	9437	9448	8721
13	2331	3223	4491	6821	7619	8360	9203	10228	10730	10473	9719
14	2973	4050	6037	8058	8834	9712	10598	11244	11651	11285	10630
15	3747	5250	7258	9031	9821	10585	11504	11939	11915	11396	10745
16	4588	6422	8053	8587	9687	10958	11563	11432	9406	8482	8338

Table IV: Clustered relative free energies for isolated met-enkephalin using the EDA. From left to right, the information provided in this table includes : temperature, Zimmerman code[i], number of individual structures in cluster, total probability ($\sum_i p_i^{approx}$) and free energy of cluster ($G_{cluster}$).

Temp (K)	Code	Number	$\sum_i p_i^{approx}$	$G_{cluster}$
	DC*B	113	0.636	0.0899
100	CC*B	136	0.0794	0.503
	C*DE	557	0.0765	0.511
	AAA	323	0.230	0.585
200	DC*A	1828	0.213	0.615
	C*DE	676	0.192	0.656
	CD*A	2685	0.297	0.723
300	DC*A	1843	0.100	1.372
	AAA	328	0.0990	1.379
	CD*A	2654	0.219	1.209
400	DC*A	1799	0.0452	2.461
	AAA	329	0.0380	2.600
	CD*A	2449	0.112	2.174
500	C*C*A	1361	0.0256	3.640
	C*AE	1463	0.0229	3.752

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

Table V: Input parameters used for FEDA runs. α_0 refers to the initial α values used for all global variables. R refers to the reduction rate applied at each level of the branch and bound tree. T refers to the thermodynamic temperature at which the free energy was calculated.

Run No.	α_0	R	T (K)	Run No.	α_0	R	T(K)
1	5	0.90	50	6	5	0.90	300
2	5	0.90	100	7	5	0.90	350
3	5	0.90	150	8	5	0.90	400
4	5	0.90	200	9	5	0.90	450
5	5	0.90	250	10	5	0.90	500

Table VI: Dihedral angle values for PEGM and FEGM structures of isolated met-enkephalin using FEDA. The temperatures are provided in the first row. The last two rows indicate the harmonic free energy (kcal/mol) and the potential energy value (kcal/mol), respectively.

Residue	DA	PEGM	100 K	200 K	300 K	400 K	500 K
Tyr_1	ϕ	-83.4	-83.4	-163.1	179.8	179.8	-90.2
	ψ	155.8	155.8	-40.5	-18.2	-18.2	149.1
	ω	-177.1	-177.1	-177.7	-178.1	-178.1	177.5
	χ_1	-173.2	-173.2	-172.2	178.2	178.2	169.8
	χ_2	-100.7	-100.7	93.2	81.3	81.3	71.8
	χ_3	13.7	13.7	-177.2	177.3	177.3	-2.4
Gly_2	ϕ	-154.3	-154.3	65.1	-59.8	-59.8	-66.1
	ψ	85.8	85.8	-89.7	-37.6	-37.6	87.5
	ω	168.5	168.5	174.1	-178.8	-178.8	-173.4
Gly_3	ϕ	83.0	83.0	-152.6	-67.0	-67.0	147.2
	ψ	-75.0	-75.0	34.4	-40.1	-40.1	-36.7
	ω	-170.0	-170.0	-178.9	179.7	179.7	175.1
Phe_4	ϕ	-136.8	-136.8	-155.4	-70.9	-70.9	-92.5
	ψ	19.1	19.1	159.8	-39.5	-39.5	-34.7
	ω	-174.1	-174.1	179.2	-179.8	-179.8	-179.1
	χ_1	58.9	58.9	52.1	173.9	173.9	179.1
	χ_2	-85.5	-85.5	82.9	-102.6	-102.6	74.7
Met_5	ϕ	-163.5	-163.5	-79.3	-161.0	-161.0	-154.7
	ψ	160.9	160.9	130.4	122.1	122.1	135.3
	ω	-179.8	-179.8	-178.7	-178.0	-178.0	179.9
	χ_1	52.9	52.9	-66.8	-174.7	-174.7	-172.6
	χ_2	175.3	175.3	179.8	174.0	174.0	175.1
	χ_3	-179.9	-179.9	-179.9	179.0	179.0	179.9
	χ_4	-178.6	-178.6	-60.0	-60.1	-60.1	180.0
G		-11.707	-2.499	6.107	14.175	22.200	29.592
E		-11.707	-11.707	-10.547	-9.899	-9.899	-6.338

Table VII: 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.

Bin	0 K	50 K	100 K	150 K	200 K	250 K	300 K	350 K	400 K	450 K	500 K
1	2	1	3	10	8	5	5	6	17	15	8
2	3	6	14	9	10	11	16	23	19	23	30
3	12	26	38	52	53	43	42	41	56	63	86
4	46	48	55	87	91	100	97	107	156	188	193
5	47	69	116	180	189	205	208	249	324	407	478
6	87	122	259	373	400	391	403	481	721	898	988
7	161	290	470	654	730	758	846	1051	1476	1801	1756
8	297	488	760	1063	1246	1368	1524	1936	2576	2966	3052
9	543	762	1182	1637	1918	2188	2597	3181	4136	4618	4538
10	828	1140	1624	2413	2996	3511	4032	4863	6033	6481	6070
11	1066	1560	2569	3542	4193	4852	5726	6791	8047	8466	7832
12	1527	2404	3433	4735	5785	6616	7499	8630	9989	10069	9426
13	2244	3070	4470	6288	7382	8341	9315	10632	11286	11130	10484
14	2818	4004	5833	7451	8649	9727	10862	11833	12430	11937	11102
15	3657	5064	7075	8723	9617	10818	12004	12606	12358	11968	11238
16	4472	6257	7848	8718	10108	11295	12167	12003	9952	8640	8576

Table VIII: Clustered relative free energies for isolated met-enkephalin using the FEDA. From left to right, the information provided in this table includes : temperature, Zimmerman code[i], number of individual structures in cluster, total probability ($\sum_i p_i^{approx}$) and free energy of cluster ($G_{cluster}$).

Temp (K)	Code	Number	$\sum_i p_i^{approx}$	$G_{cluster}$
	DC*B	107	0.532	0.125
100	C*DE	990	0.232	0.291
	CC*A	1604	0.0636	0.547
	C*DE	1275	0.331	0.439
200	AAA	322	0.209	0.623
	DC*A	1729	0.174	0.694
	CD*A	2128	0.263	0.796
300	C*DE	1360	0.125	1.239
	AAA	327	0.111	1.309
	CD*A	2116	0.192	1.313
400	C*DE	1362	0.0464	2.440
	DC*A	1714	0.0429	2.502
	CD*A	1966	0.0922	2.368
500	C*AE	2088	0.0308	3.459
	C*C*A	1900	0.0279	3.555

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

Table IX: Dihedral angle values for PEGM and FEGM structures of isolated leu-enkephalin using EDA. The temperatures are provided in the first row. The last two rows indicate the harmonic free energy (kcal/mol) and the potential energy value (kcal/mol), respectively.

Residue	DA	PEGM	100 K	200 K	300 K	400 K	500 K
Tyr_1	ϕ	-163.1	-163.1	-163.1	-163.1	-152.2	-150.8
	ψ	-42.3	-42.3	-42.3	-42.3	154.5	155.2
	ω	-177.7	-177.7	-177.7	-177.7	176.2	176.3
	χ_1	-174.8	-174.8	-174.8	-174.8	174.0	175.1
	χ_2	90.2	90.2	90.2	90.2	73.2	-106.8
	χ_3	-177.3	-177.3	-177.3	-177.3	-175.4	179.1
Gly_2	ϕ	65.9	65.9	65.9	65.9	-68.4	-68.8
	ψ	-88.3	-88.3	-88.3	-88.3	81.5	80.1
	ω	174.2	174.2	174.2	174.2	-174.8	-175.0
Gly_3	ϕ	-150.8	-150.8	-150.8	-150.8	153.6	154.9
	ψ	31.9	31.9	31.9	31.9	-38.7	-39.9
	ω	-178.7	-178.7	-178.7	-178.7	174.2	174.1
Phe_4	ϕ	-158.7	-158.7	-158.7	-158.7	-88.7	-86.5
	ψ	157.2	157.2	157.2	157.2	-40.6	-39.4
	ω	178.0	178.0	178.0	178.0	-179.7	180.0
	χ_1	53.2	53.2	53.2	53.2	-179.8	-179.5
	χ_2	84.4	84.4	84.4	84.4	80.3	80.3
Leu_5	ϕ	-77.8	-77.8	-77.8	-77.8	-79.6	-77.0
	ψ	123.3	123.3	123.3	123.3	127.8	-44.5
	ω	-178.7	-178.7	-178.7	-178.7	-178.7	178.5
	χ_1	-179.8	-179.8	-179.8	-179.8	179.8	179.8
	χ_2	64.5	64.5	64.5	64.5	64.3	63.9
	χ_3	172.6	172.6	172.6	172.6	172.3	-67.9
	χ_4	179.4	179.4	179.4	179.4	-60.6	-60.7
G		-9.349	-0.938	7.473	15.885	23.538	30.686
E		-9.349	-9.349	-9.349	-9.349	-5.137	-4.654

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

Bin	0 K	50 K	100 K	150 K	200 K	250 K	300 K	350 K	400 K	450 K	500 K
1	15	11	4	4	4	4	3	8	3	3	3
2	15	9	12	6	5	5	15	16	7	2	2
3	37	30	22	21	26	29	34	29	18	12	7
4	60	47	51	49	48	62	65	79	42	21	38
5	147	114	101	94	102	107	152	187	128	113	109
6	240	239	216	211	210	232	322	414	304	254	226
7	371	362	354	357	394	488	615	778	587	519	591
8	558	561	559	593	720	882	1182	1450	1111	1033	975
9	781	808	896	1040	1190	1412	1856	2298	1951	1676	1495
10	1215	1271	1378	1525	1744	2192	2728	3202	2902	2634	2324
11	1598	1746	1941	2161	2540	3121	3937	4695	4045	3575	3245
12	2072	2295	2566	3007	3619	4286	5263	5971	5404	4906	4382
13	2731	2945	3316	4008	4615	5450	6510	7131	6771	5994	5155
14	3234	3787	4447	4909	5631	6591	7579	8182	7534	6943	6479
15	4166	4702	5173	5736	6618	7570	8509	8963	8393	8061	7452
16	4111	4658	5327	6295	7174	7792	7320	5722	7578	7436	6640

Table XI: Clustered relative free energies for isolated leu-enkephalin using the EDA. From left to right, the information provided in this table includes : temperature, Zimmerman code[i], number of individual structures in cluster, total probability ($\sum_i p_i^{approx}$) and free energy of cluster ($G_{cluster}$).

Temp (K)	Code	Number	$\sum_i p_i^{approx}$	$G_{cluster}$
	C*DE	827	0.814	0.0410
100	DC*A	905	0.133	0.401
	CC*A	1027	0.0296	0.699
	C*DE	979	0.517	0.262
200	DC*A	972	0.224	0.594
	CD*A	1402	0.0709	1.052
	CD*A	1450	0.209	0.932
300	C*DE	1055	0.167	1.068
	DC*A	1030	0.128	1.228
	CD*A	1377	0.163	1.444
400	C*DE	989	0.0543	2.316
	DC*A	917	0.0491	2.396
	CD*A	1110	0.0818	2.487
500	C*AE	1051	0.0311	3.449
	C*DE	792	0.0232	3.737

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

Table XII: Dihedral angle values for PEGM and FEGM structures of solvated met-enkephalin. The temperatures are provided in the first row. The last two rows indicate the harmonic free energy (kcal/mol) and the potential energy value (kcal/mol), respectively.

Residue	DA	PEGM	100 K	200 K	300 K	400 K	500 K
Tyr_1	ϕ	-168.2	-168.2	-170.9	-168.4	-168.4	-152.5
	ψ	-30.9	-30.9	-28.5	-34.3	-34.3	153.2
	ω	178.6	178.6	177.5	-178.9	-178.9	178.5
	χ_1	-173.5	-173.5	178.8	178.7	178.7	-179.0
	χ_2	-100.9	-100.9	61.3	-100.8	-100.8	-101.2
	χ_3	19.3	19.3	-4.1	179.0	179.0	-179.9
Gly_2	ϕ	78.5	78.5	73.8	177.8	177.8	-173.9
	ψ	-86.5	-86.5	47.6	-179.9	-180.0	177.1
	ω	-177.3	-177.3	-179.2	180.0	180.0	-179.8
Gly_3	ϕ	162.4	162.4	167.6	-180.0	-180.0	179.6
	ψ	92.2	92.2	-145.2	179.9	179.9	-179.3
	ω	172.6	172.6	175.2	179.7	179.7	179.6
Phe_4	ϕ	-150.3	-150.3	-149.3	-155.3	-155.4	-155.4
	ψ	159.8	159.8	135.8	147.2	149.5	149.3
	ω	-178.1	-178.1	-176.6	-176.8	-178.3	-178.3
	χ_1	65.8	65.8	177.3	-179.5	-179.5	-179.7
	χ_2	-87.4	-87.4	-108.1	-111.7	-105.6	74.4
Met_5	ϕ	-75.0	-75.0	-85.5	-78.7	-78.7	-78.9
	ψ	113.9	113.9	-41.1	-51.1	113.4	113.5
	ω	-178.4	-178.4	179.9	179.7	-179.1	-179.1
	χ_1	-172.3	-172.3	-65.6	-67.2	-67.4	-67.4
	χ_2	176.1	176.1	-179.6	-178.8	-178.8	-178.8
	χ_3	-180.0	-180.0	-179.4	-179.9	-179.9	-179.9
	χ_4	60.0	60.0	179.5	-180.0	60.0	-60.0
G		-50.060	-41.896	-34.566	-28.604	-22.828	-17.166
E		-50.060	-50.060	-48.676	-46.030	-45.780	-44.797

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

Bin	0 K	50 K	100 K	150 K	200 K	250 K	300 K	350 K	400 K	450 K	500 K
1	10	11	16	17	21	18	19	22	21	21	13
2	14	17	35	122	236	149	98	95	97	94	79
3	34	66	299	542	896	607	378	283	223	195	166
4	117	296	668	1589	2075	1496	885	635	520	412	343
5	326	626	1907	3163	3636	2644	1730	1175	814	678	548
6	717	1582	3324	4902	5438	4256	2812	1957	1418	1047	762
7	1440	2865	5393	6733	6816	5790	4451	3061	2172	1623	1202
8	2611	4521	6906	7692	7569	6730	5390	4376	3123	2299	1705
9	3891	6337	7857	7952	7650	7221	6301	4972	4073	3132	2263
10	5567	7342	8094	7304	6858	7158	6736	5925	4699	3788	2903
11	6677	8090	7193	6612	6320	6374	6675	6232	5426	4453	3501
12	7624	7483	6618	5915	5645	6028	6295	6270	5754	5015	4161
13	7650	6920	5726	4864	4582	5279	5756	5972	5822	5328	4577
14	7047	6106	4680	3875	3645	4280	5113	5546	5689	5387	4879
15	6375	5066	3710	3086	2978	3449	4361	4973	5376	5271	5012
16	5534	4090	2848	2237	2140	2796	3437	4233	4809	5141	4964

Table XIV: Clustered relative free energies for solvated met-enkephalin. From left to right, the information provided in this table includes : temperature, Zimmerman code[i], number of individual structures in cluster, total probability ($\sum_i p_i^{approx}$) and free energy of cluster ($G_{cluster}$).

Temp (K)	Code	Number	$\sum_i p_i^{approx}$	$G_{cluster}$
	C*H*E	139	0.466	0.152
100	C*DF	286	0.224	0.297
	C*G*A	205	0.0991	0.459
	C*A*E	1112	0.0521	1.174
200	A*E*E	393	0.0468	1.217
	E*EE	149	0.0421	1.259
	E*EE	148	0.0474	1.818
300	EE*E	152	0.0445	1.856
	D*E*E	149	0.0273	2.147
	EE*E	151	0.0476	2.419
400	E*EE	145	0.0391	2.575
	EEE	159	0.0266	2.883
	EE*E	149	0.0460	3.059
500	E*EE	142	0.0327	3.397
	EEE	156	0.0299	3.488

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

Table XV: Dihedral angle values for PEGM and FEGM structures of solvated leu-enkephalin. The temperatures are provided in the first row. The last two rows indicate the harmonic free energy (kcal/mol) and the potential energy value (kcal/mol), respectively.

Residue	DA	PEGM	100 K	200 K	300 K	400 K	500 K
Tyr_1	ϕ	-168.4	-168.4	-168.4	-168.4	-168.4	-152.5
	ψ	-30.7	-30.7	-34.3	-34.3	-34.3	153.2
	ω	178.5	178.5	-178.9	-178.9	-178.9	178.6
	χ_1	-173.4	-173.4	178.7	178.7	178.7	-179.0
	χ_2	78.7	78.7	-100.8	-100.8	-100.8	-101.2
	χ_3	-161.1	-161.1	179.0	179.0	179.0	-179.9
Gly_2	ϕ	78.9	78.9	177.8	177.8	177.8	-173.9
	ψ	-87.2	-87.2	-180.0	-180.0	-180.0	177.3
	ω	-177.3	-177.3	180.0	180.0	180.0	-179.8
Gly_3	ϕ	163.2	163.2	-180.0	-180.0	-180.0	179.8
	ψ	91.5	91.5	179.9	179.9	179.9	-179.5
	ω	172.7	172.7	179.7	179.7	179.7	179.7
Phe_4	ϕ	-150.7	-150.7	-155.4	-155.4	-155.4	-155.4
	ψ	161.5	161.5	148.6	148.6	148.6	148.5
	ω	-178.4	-178.4	-178.7	-178.7	-178.7	-178.7
	χ_1	66.7	66.7	-179.7	-179.7	-179.7	-179.8
	χ_2	-86.8	-86.8	73.7	73.7	73.7	-106.3
Leu_5	ϕ	-75.4	-75.4	-76.5	-76.5	-76.5	-76.6
	ψ	105.3	105.3	111.0	111.0	111.0	111.0
	ω	-178.3	-178.3	-179.1	-179.1	-179.1	-179.1
	χ_1	179.5	179.5	179.2	179.2	179.2	179.2
	χ_2	63.8	63.8	63.8	63.8	63.8	63.8
	χ_3	172.2	172.2	172.3	172.3	172.3	172.3
	χ_4	59.3	59.3	179.3	179.3	179.3	179.3
G		-46.565	-38.219	-30.543	-24.671	-18.798	-12.991
E		-46.565	-46.565	-42.287	-42.287	-42.287	-41.301

Table XVI: Clustered relative free energies for solvated leu-enkephalin. From left to right, the information provided in this table includes : temperature, Zimmerman code[i], number of individual structures in cluster, total probability ($\sum_i p_i^{approx}$) and free energy of cluster ($G_{cluster}$).

Temp (K)	Code	Number	$\sum_i p_i^{approx}$	$G_{cluster}$
	C*H*E	149	0.492	0.141
100	C*DF	294	0.192	0.328
	C*G*A	168	0.151	0.376
	E*EE	120	0.0552	1.151
200	A*E*E	271	0.0477	1.209
	C*A*E	759	0.0458	1.225
	E*EE	118	0.0574	1.704
300	EE*E	101	0.0336	2.023
	D*E*E	103	0.0266	2.163
	E*EE	116	0.0451	2.464
400	EE*E	100	0.0361	2.639
	E*DE	104	0.0245	2.948
	E*EE	111	0.0364	3.291
500	EE*E	100	0.0351	3.329
	EEE	138	0.0236	3.723

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

Figure 1: Potential energy terms in ECEPP/3 force field. r_{ij} refers to the interatomic distance of the atomic pair (ij). Q_i and Q_j are dipole parameters for the respective atoms, in which the dielectric constant of 2 has been incorporated. F_{ij} is set equal to 0.5 for 1–4 interactions and 1.0 for 1–5 and higher interactions. A_{ij}, C_{ij}, A'_{ij} and B_{ij} are nonbonded and hydrogen bonded parameters specific to the atomic pair. $E_{o,k}$ and $E_{o,l}$ are parameters corresponding to torsional barrier energies for a given dihedral angle. θ_k represents any dihedral angle. c_k and c_l take the values -1,1, and n_k and n_l refer to the symmetry type for the particular dihedral angle. The cystine loop–closing term is calculated as a penalty term of three distances involved in loop–closing, where r_{il} represents the actual distance and r_{io} represents the required distance. B_i, the penalty parameter, is set equal to 100. Finally, E_p is a fixed internal energy that is added for each proline residue in the protein.

Fiugre 2: One-dimensional illustrative example of the αBB approach. In iteration 1 the overall domain is bisected, the two convex lower bounding functions are created and their unique minima ($L1$ and $L2$) are identified. An upper bound is also identified. Since $L1$ is less than $L2$, the region containing $L1$ is further bisected in iteration 2, while the other region is stored. The minimum of one region ($L3$) is greater than the new upper bound, so this region can be fathomed. The other region is stored. In iteration 3 the region with the next lowest lower bound ($L2$) is bisected and since both new lower bound minima ($L5$ and $L6$) are greater than the current best upper bound, the entire region is fathomed. Finally, by iteration 4, the region containing $L4$ is bisected which results in a region that can be fathomed (containing $L7$) and a convex region whose minimum ($L8$) equals the current upper bound and is the global minimum.

Figure 3: Using multiple lower bound minima to find low energy conformers of the upper bounding function.

Figure 4: Interface for global optimization.

Figure 5: Potential energy comparison for isolated met-enkephalin using EDA. Minimum and maximum potential energies versus bin number are plotted for three temperatures : T = 0 K, 250 K, 500 K.

Figure 6: Plot of cumulative fraction of low energy conformers for isolated met-enkephalin, which is equal to the number of unique conformers within the first 8, 12 and 16 energy bins over the total number unique conformers, versus temperature. Both EDA and FEDA data are plotted.

Figure 7: Plot of specific heat using EDA and FEDA free energy results for isolated met-enkephalin.

Figure 8: Potential energy comparison for isolated leu-enkephalin using EDA. Minimum and maximum potential energies versus bin number are plotted for three temperatures : T = 0 K, 250 K, 500 K.

Figure 9: Plot of cumulative fraction of low energy conformers for isolated leu-enkephalin, which is equal to the number of unique conformers within the first 8, 12 and 16 energy bins over the total number unique conformers, versus temperature. Both EDA and FEDA data are plotted.

Figure 10: Plot of specific heat using EDA and FEDA free energy results for isolated leu-enkephalin.

Figure 11: FEGM structures for solvated met-enkephalin. The top figure is the PEGM and the FEGM for 100 K. The structures at other temperatures (200,300,400,500) are shown left to right, top to bottom.

Figure 12: Plot of cumulative fraction of low energy conformers for solvated met-enkephalin, which is equal to the number of unique conformers within the first 4, 6, 8, 10, 12, 14 and 16 energy bins over the total number unique conformers, versus temperature.

Figure 13: Energy comparison for solvated met-enkephalin. Minimum and maximum potential energies versus bin number are plotted for three temperatures : T = 0 K, 250 K, 500 K.

Figure 14: Plot of specific heat using free energy results for solvated met-enkephalin.

Figure 15: Plot of cumulative fraction of low energy conformers for solvated leu-enkephalin, which is equal to the number of unique conformers within the first 4, 6, 8, 10, 12, 14 and 16 energy bins over the total number unique conformers, versus temperature.

Figure 16: Plot of specific heat using free energy results for solvated leu-enkephalin.

$$E = \sum_{(ij)\in ES} \frac{q_i q_j}{r_{ij}}$$

(Electrostatic)

$$+ \sum_{(ij)\in NB} F_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}$$

(Nonbonded)

$$+ \sum_{(ij)\in HX} \frac{A'_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{10}}$$

(Hydrogen bonded)

$$+ \sum_{k\in TOR} \left(\frac{E_{o,k}}{2}\right)(1+c_k \cos n_k \theta_k)$$

(Torsional)

$$+ \sum_{l\in SS} B_l \sum_{i=1}^{i=3} (r_{il} - r_{io})^2$$

(Cystine Loop-Closing)

$$+ \sum_{p\in PRO} E_p$$

(Proline Internal)

Figure 1:

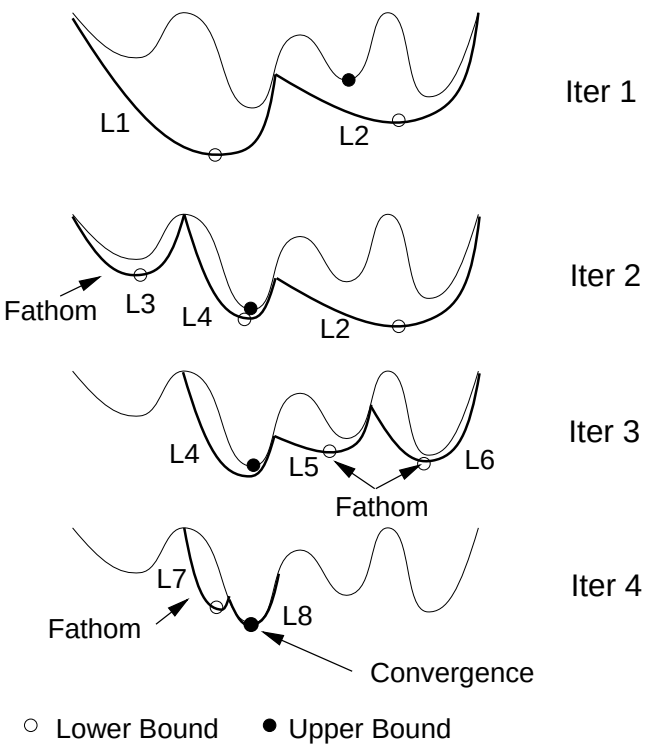
Iter 1
L1
L2
Iter 2
Fathom
L3
L4
L2
Iter 3
L4
L5
L6
Fathom
Iter 4
L7
Fathom
L8
Convergence
Lower Bound
Upper Bound

Figure 2:

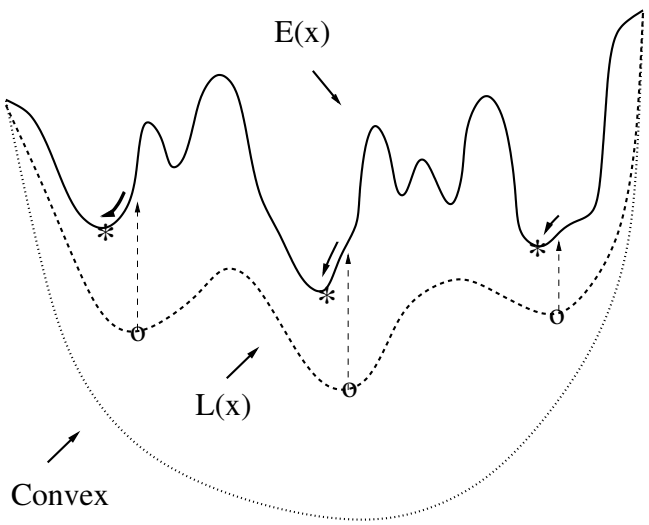

Figure 3:

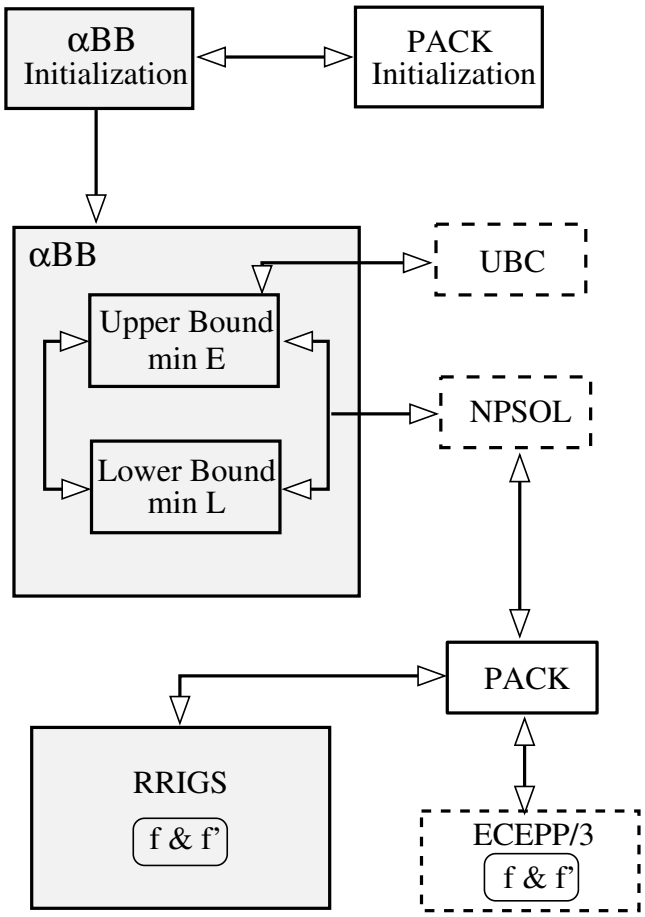
αBB
Initialization
PACK
Initialization
αBB
UBC
Upper Bound
min E
NPSOL
Lower Bound
min L
PACK
RRIGS
f & f'
ECEPP/3
f & f'

Figure 4:

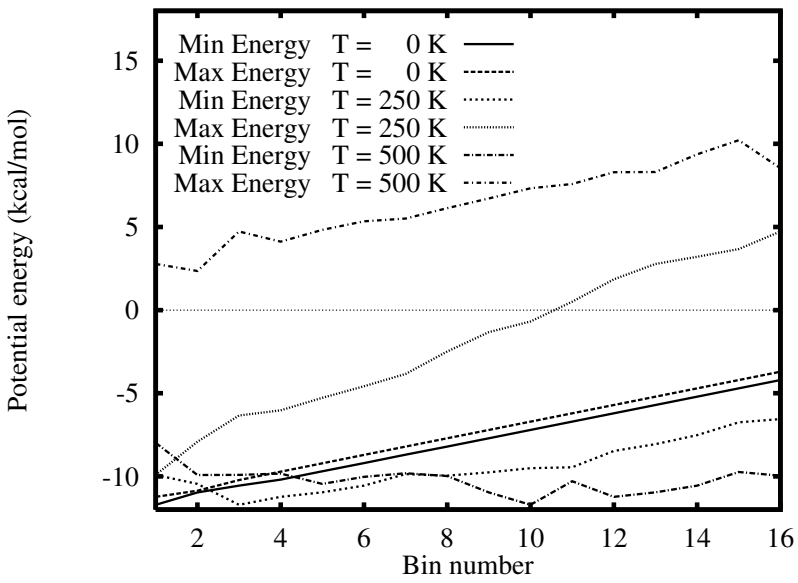
Min Energy T = 0 K
Max Energy T = 0 K
Min Energy T = 250 K
Max Energy T = 250 K
Min Energy T = 500 K
Max Energy T = 500 K
Potential energy (kcal/mol)
15
10
5
0
-5
-10
2
4
6
8
10
12
14
16
Bin number

Figure 5:

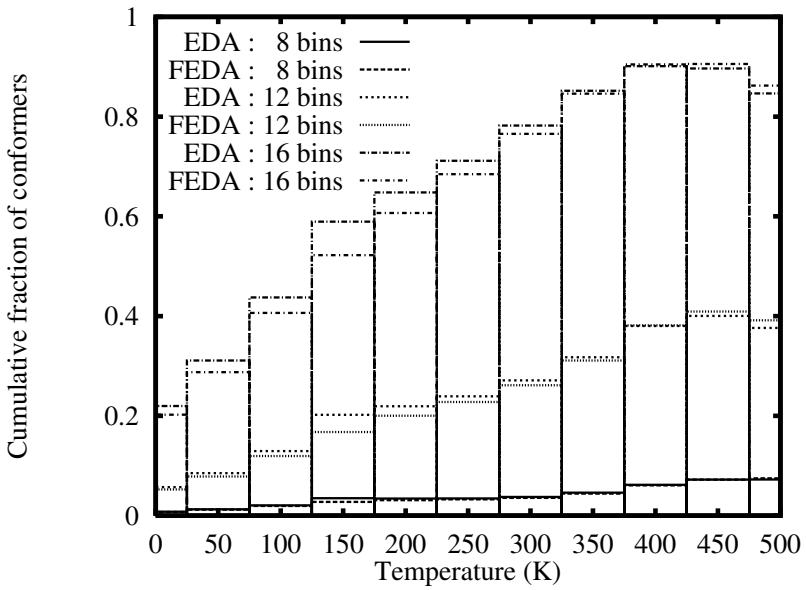
Cumulative fraction of conformers
Temperature (K)
EDA : 8 bins
FEDA : 8 bins
EDA : 12 bins
FEDA : 12 bins
EDA : 16 bins
FEDA : 16 bins
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
250
300
350
400
450
500

Figure 6:

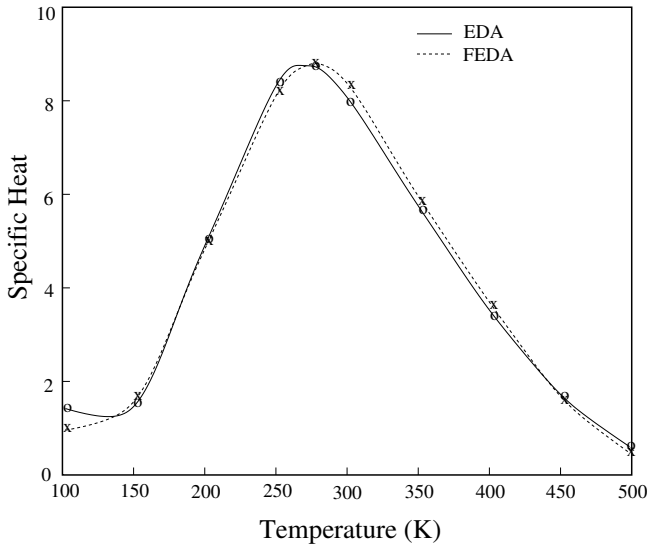
EDA
FEDA
Specific Heat
Temperature (K)
0
2
4
6
8
10
100
150
200
250
300
350
400
450
500

Figure 7:

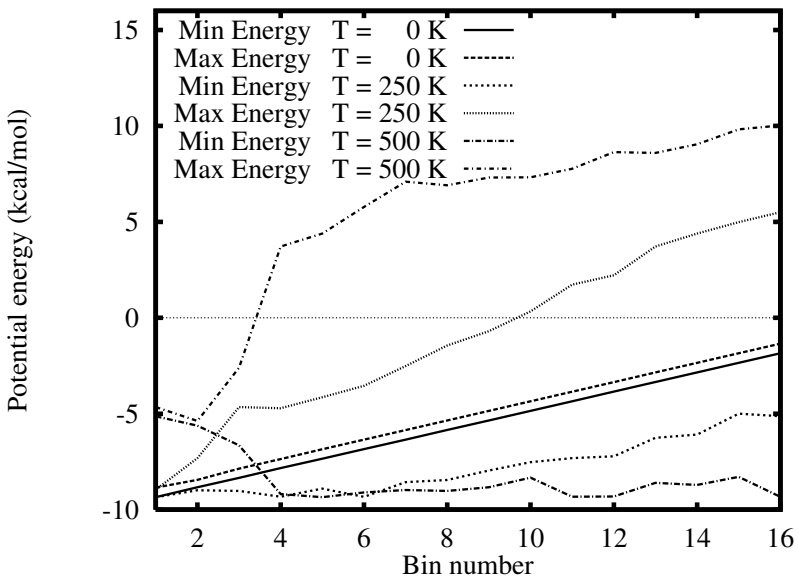
Min Energy T = 0 K
Max Energy T = 0 K
Min Energy T = 250 K
Max Energy T = 250 K
Min Energy T = 500 K
Max Energy T = 500 K
Potential energy (kcal/mol)
Bin number
15
10
5
0
-5
-10
2
4
6
8
10
12
14
16

Figure 8:

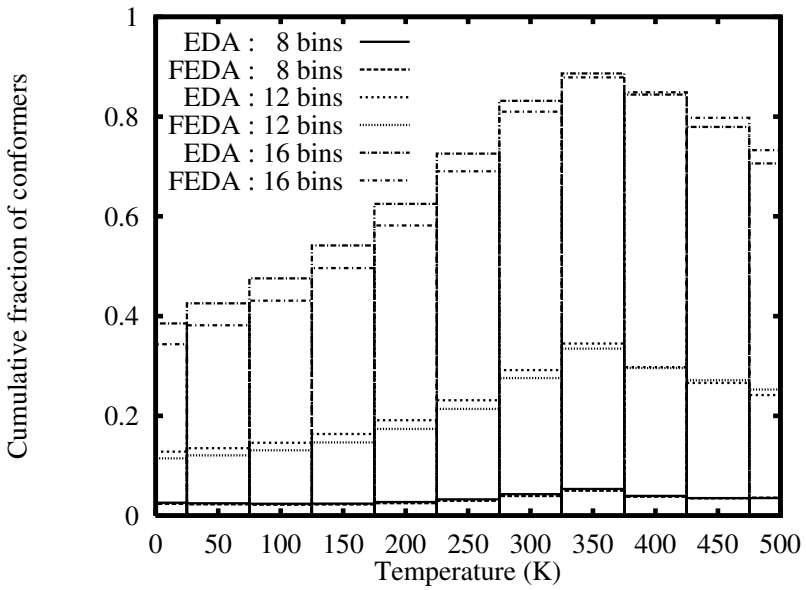
Cumulative fraction of conformers
Temperature (K)
EDA : 8 bins
FEDA : 8 bins
EDA : 12 bins
FEDA : 12 bins
EDA : 16 bins
FEDA : 16 bins
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
250
300
350
400
450
500

Figure 9:

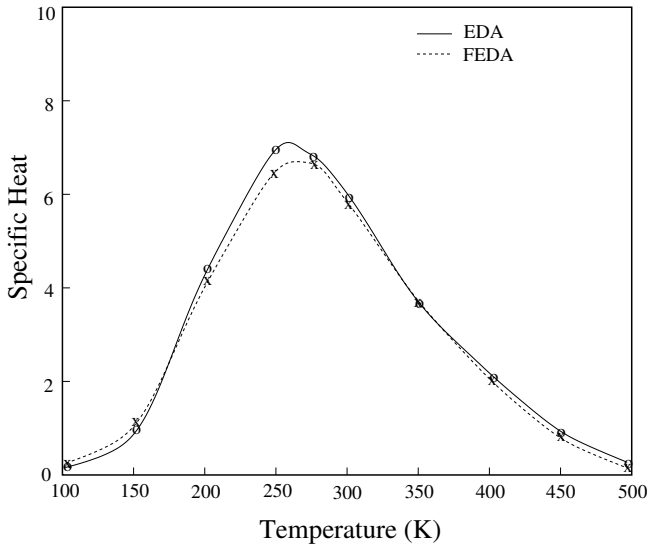
EDA
FEDA
Specific Heat
Temperature (K)
0
2
4
6
8
10
100
150
200
250
300
350
400
450
500

Figure 10:

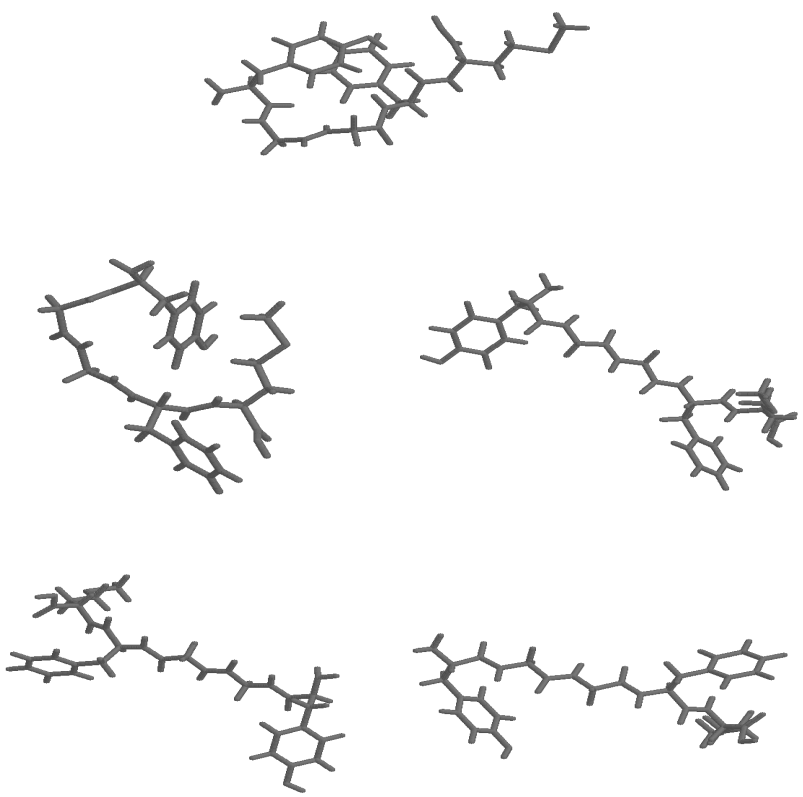

Figure 11:

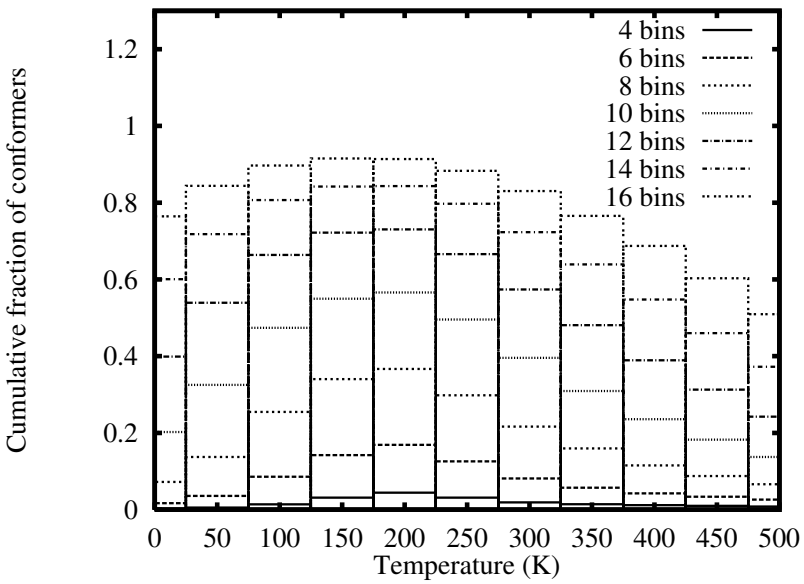
Cumulative fraction of conformers
Temperature (K)
4 bins
6 bins
8 bins
10 bins
12 bins
14 bins
16 bins
0
0.2
0.4
0.6
0.8
1
1.2
0
50
100
150
200
250
300
350
400
450
500

Figure 12:

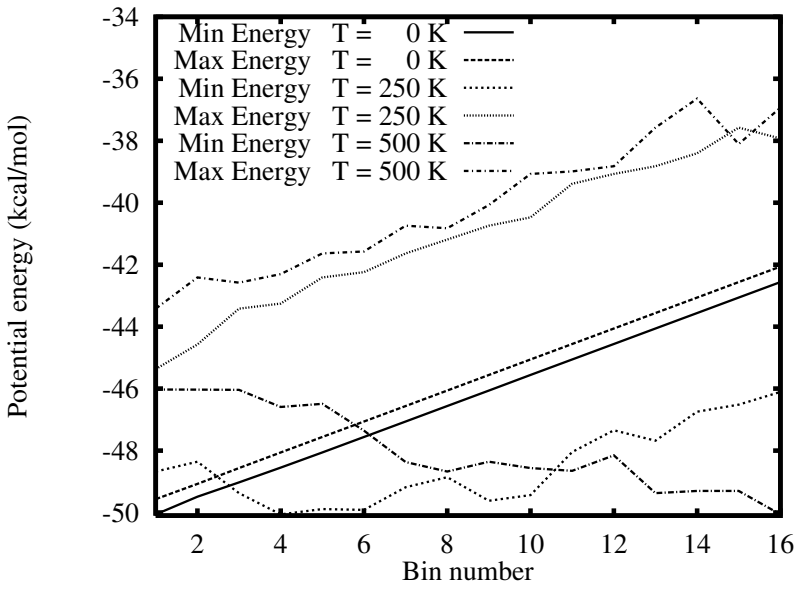
Min Energy T = 0 K
Max Energy T = 0 K
Min Energy T = 250 K
Max Energy T = 250 K
Min Energy T = 500 K
Max Energy T = 500 K
Potential energy (kcal/mol)
-34
-36
-38
-40
-42
-44
-46
-48
-50
2
4
6
8
10
12
14
16
Bin number

Figure 13:

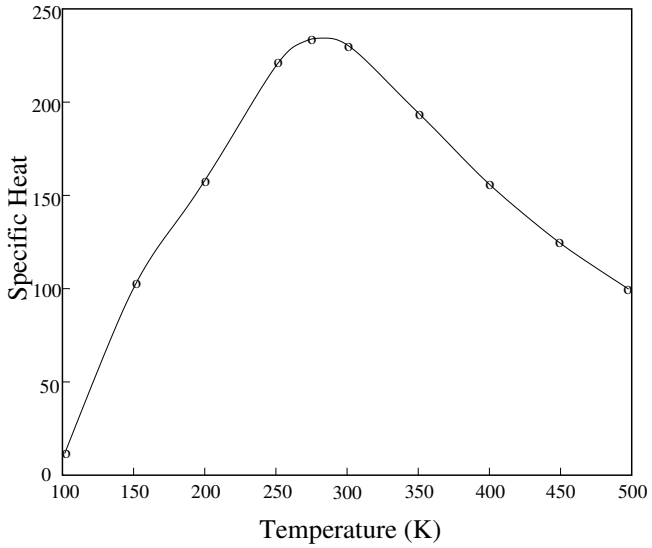
Specific Heat
Temperature (K)
0
50
100
150
200
250
100
150
200
250
300
350
400
450
500

Figure 14:

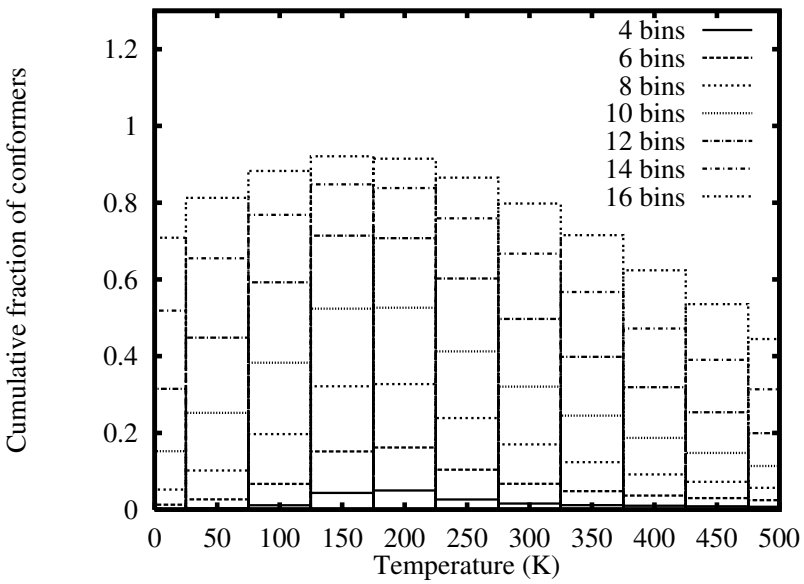
Cumulative fraction of conformers
Temperature (K)
4 bins
6 bins
8 bins
10 bins
12 bins
14 bins
16 bins
0
0.2
0.4
0.6
0.8
1
1.2
0
50
100
150
200
250
300
350
400
450
500

Figure 15:

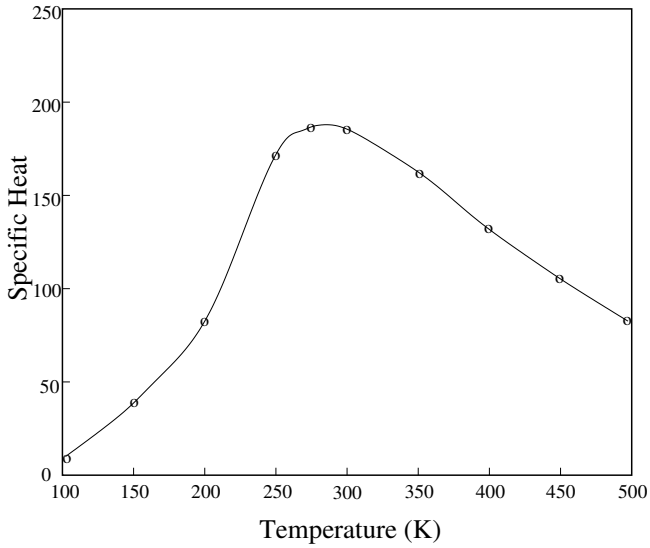
Specific Heat
Temperature (K)
0
50
100
150
200
250
100
150
200
250
300
350
400
450
500

Figure 16: