Hierarchical Conformational Scanning Using Potential Smoothing. Application to Cycloheptadecane.

Rohit V. Pappu §, Reece K. Hart and Jay W. Ponder *
Department of Biochemistry and Molecular Biophysics
Washington University School of Medicine
St. Louis, Missouri 63110

* To Whom Correspondence should be Addressed

§ Current Address: Department of Biophysics and Biophysical Chemistry,
Johns Hopkins Univ. School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205

Abstract.

We describe a generalized potential smoothing method for conformational search problems. This new method, called *conformational scanning,* surveys the entire conformational space and locates a set of low energy structures that span the full space. The procedure is applied to structural analysis of cycloheptadecane, a benchmark problem for conformational search. Conformational scanning is defined as a two step protocol. 1) Mapping: A set of major basins of attraction are initially identified. This set of basins covers the full extent of conformational space for the given molecule and each basin represents a distinct region of conformational space. 2) Local Sampling: Each basin is sampled locally to identify the lowest energy conformer within the basin giving a complete set of low energy conformers over the entire space. The result represents a generalized unbiased sampling of all of conformational space as opposed to a search directed toward the global minimum or to specific low energy conformers confined to small regions of the potential surface. The two step procedure is facilitated by a mapping of the energy basins on a series of deformed potential energy surfaces. A multidimensional potential energy surface is analytically transformed by reducing the number of unique minima and the heights of energy barriers. This comes about due to a spatial averaging over local high frequency features of the surface. The complexity of the search is hierarchically reduced with increase in deformation of the potential energy surface. An iterative procedure for exhaustive mapping of basins or local minima on potential energy surfaces is presented. This mapping procedure becomes extremely efficient on smoother surfaces. The search effort, characterized by the number of minimizations required to find a set of basins covering all of conformational space, is reduced by upwards of three orders of magnitude on deformed surfaces. Use of an efficient local sampling method for smoothed surfaces constitutes the second half of the conformational scanning protocol. For cycloheptadecane, a set of 70 basins spans the full MM2 conformational space. Within these 70 basins, we efficiently locate all 10 of the lowest energy MM2 structures and 15 of the twenty lowest energy conformers. The physical basis of hierarchical partitioning on deformed surfaces is explored by comparing local structural patterns of cycloheptadecane conformers within basins. Conformational scanning should be a useful paradigm for conformational analysis and computation of extensive properties of large, flexible molecules of biological and biochemical interest.

Introduction.

Efficient algorithms for extensive conformational sampling of systems such as proteins, molecular clusters, cyclic molecules and glasses are required to compute their dynamic and thermodynamic properties. Highly flexible multidimensional systems demonstrate complex hierarchical structures characterized by rugged potential energy surfaces[1-3]. The nature of these potential energy surfaces leads to dynamical events on vastly different time scales presenting serious challenges for sampling protocols.

In recent years the highly flexible cycloheptadecane molecule has emerged as a benchmark problem for conformational analysis[4-6]. Locating low energy conformers for cycloheptadecane is a challenging problem with parallels to structure prediction problems in molecular biology and biochemistry[4,5]. While not a particularly large molecule, cycloheptadecane presents a difficult challenge due to its flexibility and the close energy spacing of low lying minima. Searching conformational space to identify each of the minimum energy conformers is computationally intensive. The search problem essentially grows exponentially as the size of the system[6,7]. Potential energy surfaces for flexible molecules like cycloheptadecane may possess underlying structure, thus lending themselves to hierarchical partitioning into a smaller set of macrostates or basins of attraction. The presence of hierarchical structures on potential energy surfaces has been established for a variety of molecular systems including clusters, glasses and polypeptide chains[1-3,9-19].

The extent of conformational space to be sampled for cycloheptadecane or any other large flexible molecule is determined by the type of information one seeks. Location of the single lowest energy conformer or a set of low energy conformers requires a nearly exhaustive search of conformational space. Methods for conformational searching can be classified into four overlapping categories: (1) heuristic methods; (2) space covering methods such as molecular

dynamics;[16] (3) stochastic methods; and (4) identifying representative low energy conformers within a set of basins which span all of conformational space, *i.e.*, *conformational scanning*.

Heuristic search algorithms include general methods such as distance geometry[20-22] and energy embedding[23] or specialized methods such as pseudosystematic searches,[5] torsional tree searches,[24] penalty function methods such as poling,[25] ''build up'' procedures,[26] systematic searching[27] and complimentarity methods[6]. All specialized heuristic methods are tailored to the specific system under consideration and some of these[5,6] are restricted to cyclic molecules such as cycloheptadecane.

Examples of random search methods include the so called ''random kick'' method,[28] and Monte Carlo Minimization (MCM)[29]. Random search methods become inefficient for larger molecules because the number of fruitless searches increases rapidly with the system.

The goal in the current study is to map out the important basins of attraction on the potential energy surface of cycloheptadecane. Each basin can then be sampled individually to identify the lowest energy conformer within the basin. The end result is a basis set of low energy conformers which spans the entire conformational space of the molecule. We refer to this two step algorithm as *conformational scanning* because it serves to identify low energy conformers spanning the extent of conformational space.

Our algorithm for conformational scanning is based on the concept of spatial averaging or hypersurface deformation which results in smoother potential energy surfaces for enhanced conformational sampling. Spatial averaging refers to methods which have a coarse graining effect on the potential energy function leading to a potential energy surface wherein low lying regions are exposed and the barriers between these regions are lowered. Variants of coarse graining or spatial averaging have been used to develop mean field methods and hypersurface deformation methods for improved sampling conformational space. Some of these include time dependent Hartree trajectory methods such as local enhanced sampling,[30-32] self-guided

molecular dynamics,[33] hypersurface deformation methods such as Monte Carlo Minimization (MCM)[29] and ''basin hopping'',[34] Gaussian transform smoothing methods[35] such as the diffusion equation method,[36,37] adiabatic and non-adiabatic Gaussian density annealing,[38,39] packet annealing[40] and other integral transform methods such as the method of bad derivatives[41]. In potential smoothing the potential function is analytically transformed to reduce the number of minima and lower the heights of barriers. A variety of potential smoothing algorithms have been used with varying degrees of success as tools for global optimization of molecular conformation[36-43].

Figure 1a shows an idealized smoothing protocol for global optimization. Unstable, high lying shallow minima merge into a single catchment region as the level of smoothing, denoted by the parameter λ, is increased. At an intermediate level of smoothing mergers of high lying minima lead to two prominent basins of attraction, reflecting the underlying structure of the potential energy surface. If deformation of the potential energy surface is continued, a convex surface with a single minimum or a degenerate set of minima related to the global minimum will result. A reversal protocol consisting of iterative adiabatic reduction of the deformation followed by local minimization will lead to the global minimum. Smoothing succeeds at winnowing down the number of local minima by ''projecting''[44] out catchment regions related to the deepest, broadest basins on the undeformed potential energy surface.

A smoothed potential function derives from a coarse grained spatially averaged form, *i.e.,* the original *d*-dimensional potential function, $U(\{r\})$ where $\{r\}$ denotes the conformational coordinates is replaced by $\langle U(\{r\})\rangle_\lambda$:

$$\langle U(\{r\})\rangle_\lambda = \int \frac{1}{(4\pi\lambda)^{d/2}} \exp\left\{ \frac{-\left\|\{r\}-\{r\}_o\right\|^2}{4\lambda} \right\} U(\{r\})\, d\{r\}$$

$$= \int K_{\lambda}\left(\{r\}, \{r_o\}\right) U(\{r\})\, d\{r\} \qquad [1]$$

The kernel $K_{\lambda}\left(\{r\}, \{r_o\}\right)$ is a two parameter smoothing function that averages over the spatial extent of the original potential function $U(\{r\})$. In equation [1] $K_{\lambda}\left(\{r\}, \{r_o\}\right)$ is an isotropic Gaussian of width λ and center $\{r\}_o$. λ is the spatial scale over which the potential function is averaged and if it spans the distance between distinct minima, the transformation in equation [1] will result in a potential energy surface with fewer minima. This form of potential smoothing is also called Gaussian averaging because the kernel involved is a Gaussian function. The Gaussian kernel can be replaced by other kernels[41] which will lead to qualitatively similar deformation of a potential to that illustrated in Figure 1a. Andricioaei and Straub[41] have shown that Gaussian averaging by adiabatically setting the deformation to λ leads to smoothing out spatial fluctuations of frequency $1/\lambda$.

In practice potential smoothing or Gaussian averaging is accompanied by certain features illustrated in Figure 1b[42]. Typical potential energy surfaces are manifestly anisotropic leading to different widths of minimum energy basins and different heights for barriers along different directions. Using spherical isotropic Gaussian transforms for spatial averaging can lead to a reordering of the relative energies of some basins on smooth surfaces. This is shown in Figure 1b where two unique minima *A* and *B* from the original potential energy surface with conformational energies E_A and E_B are such that $E_A < E_B$. For some level of smoothing a reordering of minima can lead to a crossing of relative energies, *i.e.,* $E_A > E_B$ for $\lambda = \lambda_2$. Reordering can occur if the deepest basin on the undeformed surface is not necessarily the broadest one and the energy gap between the deepest basin and the second lowest minimum is not pronounced.

It is clear that both Figures 1a and 2b illustrate the conceptual basis for using potential smoothing for conformational scanning. Higher lying minima separated by lower energy barriers are subsumed into a basin of attraction on a deformed potential energy surface. This is shown in Figure 1b for $\lambda = \lambda_1$ or the $\lambda = \lambda_2$ levels. For each of these levels the underlying structure of the potential energy surface, *i.e.*, two major basins separated by a large barrier, is exposed. Reduction in the number of minima as a function of increased deformation leads to a drastic simplification of the search problem. Generating a mapping of minima on a deformed potential energy surface is computationally less intensive than mapping all of the minima on an undeformed surface. In Figure 1b the major basins of attraction can be identified or mapped on either the $\lambda = \lambda_1$ or the $\lambda = \lambda_2$ levels. The smoothing transformation can be reversed and each basin can be sampled locally to identify the lowest energy conformer within the basin. The end result would be two low energy conformers which form a basis set spanning the extent of conformational space as all other higher lying minima can be accessed by activation out of these two conformers. This is the premise of *conformational scanning* and can be further extended to use extensive local sampling within individual basins to construct accurate estimates of thermodynamic averages and state functions.

Hierarchical partitioning of potential energy surfaces.

Wales, *et al.*[1] have constructed extremely useful pictorial representations known as disconnectivity graphs to depict hierarchical structures of potential energy landscapes for molecular clusters. Their analysis derives from quantitating the energies of transition states with respect to the lowest lying minimum and the network of pathways connecting local minima through transition states. Some characteristic landscapes include a ‘‘gentle funnel’’ which reflects low lying minima separated by large barriers and no pronounced separation of energies. Some other landscapes for clusters are ‘‘steep funnels’’ with large negative gradients toward low lying regions and lower barriers connecting these low lying minima. Such landscapes are

characterized by strong correlations between the energy gaps and relative distances of higher lying minima relative to the low energy ground state. Furthermore barriers separating the lowest energy conformer from the rest of the potential energy surface are smaller and easily traversed. ‘‘Gentle funnels’’ and ‘‘steep funnels’’ lend themselves to hierarchical partitioning wherein clusters of local minima can be connected by barriers of increasing energy and the number of prominent basins decreases with increase in barrier height. A third kind of potential energy surface would be a ‘‘rough’’ landscape with highly degenerate ground states and random distribution of very high energy barriers between low lying minima. In a hierarchical partitioning of a potential energy surface a characteristic time scale τ_i can be defined such that $\tau_i^{\text{local}} \ll \tau_i \ll \tau_i^{\text{global}}$ where τ_i^{local} is the relaxation time for transitions between local minima within a basin M and τ_i^{global} is the relaxation time for transitions between other basins and M[11]. This kind of separation of time scales is possible for ‘‘gentle’’ or ‘‘steep’’ funnels alone. Long simulations using generic methods for sampling conformational space such as molecular dynamics[16] or Metropolis Monte Carlo sampling[17] can be expected to give reliable thermodynamic averages within the spatial extent of a macrostate M. However the extent of sampling obtained using these methods is inadequate for computing thermodynamic averages over spatial extents greater than a basin M[18,19].

A number of studies have also demonstrated the hierarchical structure of potential energy surfaces for molecules of biological and biochemical interest. Such evidence comes from the work of Stillinger and Weber[9,10] who first introduced the idea of kinetic partitioning of conformational space in the context of molecular dynamics simulations. Selected points along a molecular dynamics trajectory were quenched to the nearest local minimum on the energy landscape. For a single long trajectory this method exposes any ‘‘inherent structure’’ such as valleys or predominant basins on the potential energy surface. Becker and Karplus[12] used graph

theory methods to show that basins of attraction rather than individual local minima are the actual determinants of dynamics for a small peptide. They partitioned the potential energy surface of a tetrapeptide isobutyryl-(ala)3-NH-methyl (IAN) into energy dependent and temperature dependent basins of attraction to show that the landscape of this molecule has a funnel-like structure. Straub and Thirumalai[19] who used a ''fluctuation metric'' to quantitate sampling in molecular dynamics simulations of ribonuclease S-peptide, showed that a potential energy surface for a polypeptide can be partitioned into conformationally distinct basins. Similarly Velikson, *et al.*[14] have shown that conformationally distinct local minima of poly-*L*-alanine chains can be clustered into specific regions of conformational space in a hierarchical fashion using a so-called ''ultrametricity'' criteria to distinguish basins of attraction. Noguti and Go[15] have done extensive work to delineate protein substates using a hierarchical basis set derived from the length scales of motions in proteins. All of these developments suggest that the energy landscapes of many important macromolecules possess an underlying structure which may be exploited to develop tools for conformational scanning.

Objectives of the current work. We use the highly flexible seventeen membered ring structure of cycloheptadecane in all our calculations. The first part of *conformational scanning* is the mapping of basins of attraction or local minima. If the mapping is performed on undeformed surfaces, all local minima on the potential energy surface can be identified. If the mapping is performed on deformed surfaces, *i.e.*, potential energy surfaces described by $\langle U(\{r\})\rangle_\lambda$ for some $\lambda > 0$, then the mapping of minima reduces to mapping basins of attraction as shown in Figure 1b.

In the first part of the paper we describe an iterative basin hopping algorithm for mapping minima on a potential energy surface. This was used to enumerate all the local minima within 25 kcal/mole of the global minimum for cycloheptadecane. Each of these minima was tracked as a function of increasing potential function averaging to determine the hierarchy of basins on the

potential energy surface. We also demonstrate that partitioning of potential energy surfaces as a function of increased deformation represents physically meaningful clustering of minima from the undeformed surface. This is shown by comparing structures within a purported basin to structures across all other basins.

The second half of the *conformational scanning* protocol is used after a mapping has been completed on deformed surfaces. Each basin from the mapping is sampled locally to identify the lowest energy conformer within the basin. A straightforward approach is to follow the reversal algorithm illustrated in Figure 1b to locate the lowest energy conformer within each basin. If there are no energy crossings preceding the mergers of minima into a unique basin the algorithm pictured in Figure 1b will, in fact, locate the global minimum within each basin. However in most cases energy crossings are the norm rather than an exception. We describe a trajectory and quench method for sampling the local extent of a basin to locate the lowest energy conformer therein. The general conformational scanning algorithm applied to cycloheptadecane comprises a λ-level mapping combined with trajectory and quench (TQ) sampling to identify a family of low energy conformers that span the extent of conformational space.

In the next section we outline highlights of the smoothing procedure and parameterization. This is followed by a description of our algorithm for hopping between minima. Next we discuss enumeration of all minima within 25 kcal/mole of the global minimum on the MM2 energy surface of cycloheptadecane. This is followed by validation of the partitioning of the potential energy surface and a discussion of the mapping procedure applied to identify basins on deformed surfaces for cycloheptadecane. We then discuss a trajectory-quench method to sample basins of attraction and locate the lowest energy conformer within a basin. We conclude with a discussion summarizing this work in the context of ongoing developments for improved conformational searching.

Parameters and force field. All calculations were performed using the TINKER[46] modeling package, which implements a self-contained force field engine providing access to several molecular mechanics force fields. We use a modified version of the MM2 force field[47] traditionally used for cycloheptadecane.

In Cartesian space the MM2 potential includes a covalent bond stretch term, an angle bending term, a stretch-bend coupling term, torsional potentials representing 1-, 2- and 3-fold barriers and an exp-6 function for van der Waals interactions. The exp-6 function is approximated as a sum of two Gaussians of opposite sign centered about the origin. The functional form for the Gaussian approximation is: $U_{\mathrm{vdw}}(r_{ij}) \approx \sum_{k=1}^{2} a_{k(ij)} \exp(-b_{k(ij)} r_{ij}^2)$. Parameters for the two Gaussian approximation are chosen to provide the best fit to the original MM2 function, $U_{\mathrm{MM2}}(r_{ij}) = \varepsilon \left\{ \exp\left(-\frac{12.5\, r_{ij}^*}{r_{ij}}\right) - 2.25\left(\frac{r_{ij}^*}{r_{ij}}\right)^6 \right\}$. We use values of ($a_1$, b_1) = (3423.562 kcal/mole, 9.692Å) and (a_2, b_2) = (-6.503 kcal/mole, 1.585Å) for r_{ij}^* = 1Å and ε = 1.0 kcal/mole. The pairwise Gaussian parameters are scaled according to the individual r_{ij}^* and ε values for each pairwise interaction prescribed by the force field.

Functional forms for potential smoothing. The deformed Gaussian van der Waals function is of the form $U_{\mathrm{vdw}}(r_{ij},\lambda) = \sum_{k=1}^{2} \frac{a_{k(ij)}}{(1+4\lambda)^{3/2}} \exp\left(\frac{-b_{k(ij)} r_{ij}^2}{1+4\lambda}\right)$, where λ is the deformation parameter. For the torsion potential, the deformed functional form is: $U_{\mathrm{tor}}(\omega,\lambda) = \sum_j V_j \left\{ 1 + \cos(j\omega + \phi) \exp(-j^2\lambda) \right\}$. Here ω is the torsional angle value, j is the periodicity, V_j is the half height of the rotational barrier, ϕ is a phase factor, and λ is the deformation parameter.

The covalent bond stretch, angle bending and stretch-bend terms are not smoothed. The unsmoothed functional forms for these terms are part of the total deformed potential and impose penalties for deviations from covalent geometries just as they would on the undeformed surface.

"Normal mode" iterative basin hopping. The methodology used to a generate a self-consistent mapping of minima on a potential energy surface is based on an iterative scheme of hopping between minima.

1. A random starting conformation is first minimized in Cartesian space to a convergence criterion of 1.0 ´ 10^{-6} kcal/mole/Å in the *rms* gradient. This conformation is the starting seed for an iterative search procedure to sample the potential energy surface.
2. Next we compute the torsional space Hessian matrix and diagonalize it to obtain the eigenvalues and eigenvectors.[48]
3. The system is moved out of the local minimum in fixed steps along each eigenvector. A step along an eigenvector $\mathbf{X}_\omega$ can be written as $k\mathbf{X}_\omega$ where k = 1....65. The energy is computed at each of the points k along the eigenvector.[49]
4. If the conformational energy E_k at point k along an eigenvector $\mathbf{X}_\omega$ satisfies the inequalities $E_{k-1} > E_k$ and $E_{k-1} > E_{k+1}$, where the E's denote conformational energy values, it is chosen to be the point from which to start a new minimization. Both inequalities in this condition suggest downhill progress on the potential energy surface.[50] If the eigenvector scan does not yield a point satisfying the inequalities described above no new minimizations are performed.
5. If the minimization leads to a new minimum, the energy for which is within a predefined threshold E_{cut} from the lowest minimum in the current list the new minimum is added to the list of minima and return to step 2. Steps 2-5 are repeated until the ‘‘normal mode’’ search is carried out of every minimum found on the potential energy surface.

We refer to the general search procedure as a ‘‘normal mode’’ search because it uses the eigenvectors of the Hessian matrix as search directions. The method is not restricted to cycloheptadecane; rather it can be adapted to any molecular conformation problem in Cartesian, torsional, rigid body or mixed coordinate systems; the number of search directions scales linearly as the number of degrees of freedom and therefore does not necessarily become unwieldy for larger problems. The algorithm always iterates to self consistency, in that a search is carried out from every minimum located on the potential energy surface within some cutoff energy of the lowest known minimum. For cycloheptadecane we chose E_{cut} = 25 kcal/mole since inspection showed all minima with energies exceeding this cutoff value to be physically unrealistic artifacts of the MM2 potential function.

Results.

Mapping the MM2 potential energy surface. The algorithm for hopping between minima was applied to completely enumerate the set of minimum energy conformers of cycloheptadecane. The objective was to generate the density of states for on the MM2 potential energy surface within 25 kcal/mole of the global minimum. The starting structure for the mapping procedure was the purported global minimum for cycloheptadecane which has an MM2 energy of 19.068015 kcal/mole. Structures found using the ‘‘normal mode’’ search procedure are unique minima to an *rms* gradient per atom of 1.0×10^{-6} kcal/mole/Å. The mapping procedure finds a total of 20469 unique minima for cycloheptadecane with an energy distribution shown in Figure 2a.

Each cycle of a ‘‘normal mode’’ search can lead to a maximum of 34 possible minimizations corresponding to 34 search directions for energy scans from 17 eigenvectors. If all eigenvector scans lead to a minimization, the algorithm would require 695946 minimizations to

complete the self-consistent iterative procedure to map 20469 minima. In practice, the ‘‘normal mode’’ search for undeformed cycloheptadecane required 509719 minimizations to generate the density of states, which translates to approximately 25 minimizations to locate each new minimum. Of the 20469 minima we find 261 within 3 kcal/mole of the global minimum. This is nominally two less than the number found by Ngo and Karplus,[5] however, one or both of their additions may be due to slight energy rounding differences resulting from their different force field engine. Similarly there are 1635 minima within 5 kcal/mole of the global minimum, 4788 within 7 kcal/mole, 14347 within 11 kcal/mole, 19360 within 15 kcal/mole, 20069 within 17 kcal/mole, 20412 within 21 kcal/mole and 20469 within 25 kcal/mole. Increase in the number of new minima is very slow for energies larger than 15 kcal/mole from the global minimum. Based on the slow growth in the number of minima with increased energy distance from the ground state we estimate to have a nearly exhaustive enumeration of minima for cycloheptadecane on the MM2 energy surface. Using an efficient truncated Newton minimization method,[51] generation of the full distribution required roughly 13 days of CPU time on a 250MHz DEC Alpha workstation. The global minimum has an MM2 energy of 19.068015 kcal/mole. A second minimum lies only 0.01 kcal/mole above the global minimum and has an MM2 energy of 19.077409 kcal/mole. These two structures are separated by about 0.4 kcal/mole from the third best structure. The low energy tail of the distribution is shown in Figure 2b.

The density of states shown in Figure 2a is of particular interest in light of the recent work of Ngo and Karplus[5]. They estimated, based on fitting the low energy spectrum to a Gaussian function for the density of states, that there should be approximately 900 unique minima for cycloheptadecane on the MM2 energy surface with a mean energy of 22.6 kcal/mole. Furthermore they used their estimates as evidence for the validity of random energy models[52] for mesoscopic systems. We find 20469 minima within 25 kcal/mole of the global minimum, approximately 95% of which are within 15 kcal/mole of the ground state. The mean energy for

the distribution shown in Figure 2a is 28.507751 kcal/mole which is roughly 6 kcal/mole higher than the value estimated by Ngo and Karplus. Random energy models may very well be valid for macromolecular systems, but it can be misleading to estimate such a characteristic based entirely on the energy distribution of a small set of low energy conformers.

Hierarchical classification of the cycloheptadecane potential energy surface. As a first step we start with the extensively enumerated undeformed surface for cycloheptadecane. Each minimum on the MM2 energy surface is minimized on the undeformed surface using the Gaussian approximation to the MM2 exp-6 van der Waals potential. All minimizations were performed to 1.0 x 10^{-6} kcal/mole/Å in the *rms* gradient using a truncated Newton method[51]. Figure 3 shows the correlation of energies for all 20469 minima on the MM2 energy surface with the energies on the undeformed surface relative to their respective ground states. The correlation coefficient is approximately 0.9999.

We tracked the evolution of each minimum on the undeformed surface as a function of adiabatically increased deformation. The value for the deformation parameter is increased according to a cubic schedule between $\lambda = 0$ and $\lambda = 25$. For a schedule with 150 points, the deformation corresponding to level i is chosen from the formula $\lambda_i = 25.0\left(\frac{i}{150}\right)^3$. For each local minimum the deformation is increased, followed by a local minimization on the deformed surface. Two unique minima on the undeformed surface can merge into a common basin at some level of smoothing. This leads to a gradual reduction in the number of minima as a function of increased deformation. In the case of cycloheptadecane we start with 20469 unique minima on the undeformed surface, all of which merge into a single convex basin on the $\lambda = 25$ surface. Reduction in the number of minima as a function of increased smoothing for cycloheptadecane is shown in Figure 4.

Clustering on deformed surfaces. Based on a series of mergers of minima from the undeformed surface, deformed surfaces are partitioned into a set of basins. The number of basins decreases as a function of increased deformation. Clustering can be interpreted using tree diagrams to show the evolution of minima and the partitioning into basins on deformed energy surfaces. A sample sketch of a cluster tree is shown in Figure 5. The deformation parameter λ is the clustering ''distance''. A basin is denoted as a node on a tree diagram, for a given level of deformation λ. The labeling convention for nodes of the tree is as follows: we assume two minima m_1 and m_2 merge at a level λ_1 and at a level λ_2 minimum m_3 merges into this same basin. If the conformational energies on the undeformed surface are $E(M_1) < E(M_2) < E(M_3)$, then the remaining basin at level λ_2 which clusters these three structures is labeled m_1 to reflect the lowest energy undeformed conformer within the basin.

Each minimum for cycloheptadecane is labeled from 1 to 20469 based on its MM2 conformational energy vis a vis the global minimum, *i.e.*, the global minimum is referred to as 1, the second lowest minimum 2 and the highest energy conformer is labeled 20469. Figure 6 shows the partitioning of the potential energy surface on the levels from $\lambda = 0$ to 25. For the $\lambda = 1.942$ level, 20469 minima from the undeformed surface cluster into 62 basins. The tree diagram illustrates the reduced complexity for conformational searching on smooth surfaces. Locating the 62 basins on the $\lambda = 1.942$ level is equivalent to obtaining the basis set of basins which spans all of conformational space for cycloheptadecane. The reduction of complexity with increased deformation is hierarchical in that 20469 unique minima cluster into montonically decreasing numbers of basins.

It can be shown that clustering follows a pattern of first merging minima connected via low energy barriers, before merging the basins separated by higher energy barriers[53]. The clustering would therefore be roughly equivalent to the ''temperature dependent basins of

attraction'' of Becker and Karplus[12] or the disconnectivity graphs of Wales, *et al.*[1] We explore this connection further in work to be reported elsewhere[53]. In the following sections we will show that clustering in terms of hierarchical increase of potential smoothing is physically meaningful.

Structural similarities between conformations within a cluster. There are two generic ways to partition conformers from the undeformed surface into conformational clusters. Clustering can be done in terms of geometric considerations such as Euclidean distance based on *rms* superposition of pairs of structures[45] or it can reflect the clustering based on energetic considerations. Conformational energies of minima on smooth surfaces are the determinants of clustering patterns as a function of increased smoothing. In Figure 1b minima within the basin labeled *A* are not all similar in energy on the undeformed surface, rather they are connected via lower intra-basin energy barriers compared to the higher inter-basin energy barriers separating minima in the basin labeled *B*. If the deformation is adiabatically increased from $\lambda = 0$ to $\lambda = \lambda_1$, as shown in Figure 1b, and each of the structures minimized on this level, the result would be two unique minima, *i.e.*, all structures derived from the basin labeled *A* would merge into a single basin with a unique conformation and energy. Therefore the level of deformation is the clustering distance and conformational energies of degenerate minima on a deformed surface determine basin memberships. This method of partitioning is meaningful only if there is an underlying structure to the undeformed surface.

For cycloheptadecane, it can be shown that partitioning the potential energy surface into basins of attraction as a function of increased deformation is based on weakly correlated structural patterns between members of a basin. This is especially true of structures from the low energy spectrum. We analyzed the structural basis of smoothing-based hierarchical partitioning by studying clusters from different basins on the $\lambda = 1.6813$ level.

It is difficult to compare conformers of cycloheptadecane based on Euclidean distances between conformers because of the 68 equivalent permutations of this cyclic molecule required for each pairwise distance calculation. A complete conformational distance map would require $68 \times 20469 \times 10234 = 1.3047 \times 10^{10}$ distinct superposition calculations. Instead we compare the frequency of occurrence of local dihedral angle patterns across conformers spanning a basin to look for structural similarities. We follow the work of Saunders, *et al.*[4] and divide all dihedral angles into three groups: positive (p), negative (n), and *trans* (t) . Any n_s-letter dihedral string will have 4 equivalent permutations. For example the 11-letter string {nnntnnttppt} is equivalent to {ppptppttnnt}, {tppttnntnnn} and {tnnttpptppp}. In order to analyze a cluster generated on a level λ, we establish the following set of rules. A cluster with n_c members is a hard cluster if there exists an n_s-letter string or a few n_s-letter strings to describe all structures within the cluster but does not occur in structures across any of the other clusters.

Table 1 summarizes the results of local structure comparisons for clusters on the $\lambda = 1.6813$ level. Local structural patterns are derived from the undeformed surface conformations. For completeness we indicate the set of patterns which span all structures within a cluster. We restrict ourself to strings of 9 or more letters because shorter patterns are not adequate discriminants of structure. Often there is no single local structural pattern to fully describe a cluster. For instance the C1 cluster groups the nine structures 1, 35, 45, 577, 696, 865, 11054, 13383 and 13859 into the same basin. The 11-letter string {nntnntppt} is found in six of these structures: S1 ≡[1, 35, 45, 696, 865, 11054]. The same 11-letter string is found in only 22 of the remaining 20460 structures on the potential energy surface. The remaining three structures, 577, 13383 and 13859 have different 11-letter local structural patterns in common with smaller sets of conformers from the C1 basin. The string {ntnnntnnttp} is found in S2 ≡[1, 35, 45, 577, 865]; the string {tnntnptppt} in S3 ≡[35, 577, 13383] and the string {nntnntnptpp} in S4 ≡[35, 13383,

13859]. A union of the four sets S1, S2, S3 and S4 spans the cluster C1. The strings {ntnnntnnttp}, {tnntnptppt} and {nntnntnptpp} are found in 57, 198 and 52 structures respectively from the 20460 structures outside C1 basin. In Table 1 we provide detailed information on the set of strings which span structures across a given cluster. Table 1 also includes information on the extent of overlap between the individual sets and statistics of occurrence across the rest of the energy surface.

Our results, typified by the C1 cluster, suggest that clustering is especially strong for lower energy conformers. The clustering of higher energy conformers is semi-random. As described above, higher energy structures such as 13383 and 13859 have only one 11-letter string each in common with one of the lower energy conformers. The relationship with the lower energy conformer 35 determines the basin for these two structures. Figure 7 shows representative structures for cycloheptadecane from basins C1 and C6 and the clustering mechanisms on three different levels of deformation.

The sizes of basins on the $\lambda = 1.6813$ level vary from 3 for C3 to 680 for C120. Clusters C1, C2, C4 and C10 are among the smaller clusters. The n_s-letter dihedral angle strings spanning a cluster tend to become shorter for larger clusters. Also for many of the largest clusters it is not possible to find a single n_s-letter dihedral angle string of reasonable length to span the cluster. Rather we find a few sets of structures described by different 9-letter strings. These sets of structures within a cluster are of roughly equivalent size and the union of these sets span the cluster with partial overlap. Table 1 summarizes all of these results for clusters C1 through C10. The breakdown of strong clustering is almost entirely due to higher lying minima which can have few structural features in common with lower lying minima.

In addition to comparing local structural patterns between structures within a cluster and structures across clusters, we computed the distance between a sample set of clusters using a pairwise superposition method. For any two conformers the 17 carbon atoms from each

conformer can be superposed in 68 possible ways accounting for the different ways of traversing a ring and mirror images due to achiral centers. We computed the smallest superposition distance for all pairwise combinations of structures within the cluster C1. For the 9 conformers within the cluster there are 36 such distances. Similarly we computed the superposition distance between individual conformers within the C1 cluster and conformers in each of the other clusters located on the $\lambda = 1.6813$ level. Comparing structures across the C1 and C2 clusters requires computing the distances between each of the 9 conformers that are part of the C1 cluster and the 13 conformers which are part of the C2 cluster. Figure 8 shows the distribution of superposition distances between pairs of structures within the C1 cluster. Also shown are the distribution of superposition distances between structures from C1 and structures from other clusters. The lowest average inter-structure distance is for structures within the C1 cluster. As the lowest energy conformer within a cluster increases, as identified by the cluster label, the average pairwise superposition distance for structures vis a vis the C1 cluster increases. These results are to be regarded with some caution because they are influenced by two parameters, the size of clusters being compared to C1 and the fact that for cyclic molecules there is not going to be a dramatic difference in superposition distances between conformers. However, figure 8 does provide supportive evidence in favor of partitioning the potential energy surface in terms of hierarchically increased deformation, especially if the deformation is not too large.

***Ab initio* mapping of basins on a deformed surface.** In the previous sections we used the iterative basin hopping procedure to extensively enumerate minima on the MM2 energy surface. Each of 20469 were tracked as a function of increased deformation to show a hierarchical clustering of structures into a reduced number of basins. In establishing the hierarchical basis for partitioning a potential energy surface we relied exclusively on *a priori* knowledge of the many-to-few mapping of all of the 20469 minima on to a λ-level deformed surface. While this calculation is instructive it has little value for generalized *conformational scanning* because it

requires an initial enumeration of all minima on the undeformed surface. However, the clustering itself suggests that the search problem is greatly simplified on smooth surfaces. For $\lambda = 1.6813$ there are 244 unique basins, *i.e.*, the many-to-few mapping reduces 20469 minima to 244 basins.

We show that a deformed potential energy surface can be mapped using the ''normal mode'' basin hopping procedure. Table 2 summarizes results from application of our ''normal mode'' basin hopping algorithm for surfaces ranging from $\lambda = 0.8714$ to $\lambda = 12.8$. The mapping procedure is largely successful in locating basins which span all of conformational space for a given level of deformation. In all of our calculations we set E_{cut} for the ''normal mode'' basin hopping algorithm to be 25 kcal/mole. For the $\lambda = 1.7653$ level, of the 152 basins derived from the 20469 minima on the undeformed surface 149 are within 25 kcal/mole of each other. Table 2 shows that the energy gap between basins on smooth surfaces remains strongly conserved. For narrow deep canyons separated by very steep barriers on the undeformed surface the energies on deformed surfaces increase with respect to broader basins. This increase in energy on deformed surfaces can be in excess of the 25 kcal/mole threshold for the mapping procedure. By construction the mapping procedure cannot locate basins outside of a pre-defined ceiling. We found that increasing the value for the ceiling energy E_{cut} does not necessarily lead to a complete mapping primarily because basins that exceed the generous 25 kcal/mole threshold are very narrow and will in all likelihood be missed by a method which seeks out broad catchment regions.

The ''normal mode'' search procedure for iterative basin hopping locates better than 85% of the accessible basins on deformed surfaces as shown in Table 2. The number of minimizations required to map the potential energy surface drops precipitously with increase in smoothing, *i.e.*, the search procedure is greatly simplified with increase in deformation. For instance the basin hopping procedure requires only 1016 minimizations, or 15 minimizations per minimum, on the $\lambda = 1.8522$ surface to map a basis set of basins that spans 98.6% of conformational space. This is

to be contrasted with the mapping of minima on the MM2 energy surface which requires 509719 minimizations for exhaustive mapping. Increased reduction in complexity of the search process as a function of increased deformation is shown in Table 2. Deforming the potential energy surface reduces the number of minimizations required for exhaustive sampling by upwards of 3 orders of magnitude. The many-to-few mapping of local minima as a function of increased deformation leads to a simplification of the search process. More importantly our method for iterative basin hopping becomes increasingly efficient on deformed surfaces and succeeds in enumerating a basis set of basins which span better than 85% of conformational space. Each of the minimum energy basins obtained in the mapping procedure are part of the map generated by tracking all of the minima from the undeformed surface as a function of increased deformation, indicating the completeness of the mapping procedure.

Following each of the basins back to the undeformed surface will lead to vastly different regions of conformational space as indicated in Figures 1b, 7a and 7b. Because the explicit relationship between all of the minima from the undeformed surface and basins of attraction on a deformed surface is built into the many-to-few mapping, the ability to generate a near exhaustive sampling of basins spanning a potential energy surface with very little computational effort represents a significant advantage for conformational scanning.

Local sampling within a basin derived from a deformed surface. In this section we discuss the second half of the *conformational scanning* algorithm as applied to cycloheptadecane. All calculations in this section are done using the $\lambda = 1.6813$ level as the seed.

Each of the 244 basins on the $\lambda = 1.6813$ level can be sampled locally to locate the lowest energy conformer within the basin. A trivial way to do this is to follow the schematic of Figure 1b and apply a reversal schedule to each of the 244 basins. However, on deformed surfaces broad wells from the undeformed surface tend to be energetically favored over deep narrow canyons.

The proposal indicated in Figure 1b for recovering the lowest energy conformer within each basin will not succeed in general, due to intra-basin energy crossings that precede mergers.

The parameter λ determines the spatial extent of a basin. A basin may be sparsely populated, as in the C1 basin or densely populated as in the C6 basin. We use a trajectory-quench (TQ) method in order to sample the local confines of a basin generated at a given level of smoothing. The algorithm for TQ sampling is outlined below.

1. For a given basin a reversal schedule, shown in Figure 1b, is used to locate a representative minimum on the undeformed surface.
2. This minimum is used as the starting conformation for generating a short molecular dynamics (MD) trajectory. The parameters used for this trajectory are a time step of 1 fs and 10000 steps of MD with no equilibration. Velocities are reshuffled by weak coupling to a bath at T = 300K using the method of Berendsen, *et al.*[54]
3. Hundred equal spaced points along the trajectory are selected for local minimizations using a fast Newton method. [51] Should the quench locate a lower energy conformer we first check to see if the new structure belongs in the current basin. This is done by increasing the level of smoothing adiabatically up to λ = 1.6813 followed by minimization. If the new structure tracks to the parent of the current basin on the λ = 1.6813 level, we have found a lower energy conformer within the basin. However if structure tracks to the parent of a different basin, it is discarded as a possible low energy alternative.
4. At the end of this procedure if the quench has found a lower energy conformer, steps 2 and 3 are repeated out of the new conformer. The procedure is iterated until a lower alternative structure within the basin cannot be found.

There are very good reasons for choosing a TQ protocol. First of all, space covering methods such as molecular dynamics are extremely local in terms of the extent of conformational space they sample. This is precisely our objective because the local region to be sampled has already been delineated by the first half of the *conformational scanning* algorithm. Secondly, short MD trajectories are sufficient for our purposes making the protocol computationally efficient.

In implementing the TQ method for local sampling of a basin we make the following observations: (i) for reasonable temperatures $T < 400$K the sampling is strictly local, *i.e.*, at step 3 in the algorithm we seldom find an excursion outside the basin delineated by the mapping procedure. This reflects the validity of basin partitioning on the smooth surface especially if the surface mapping is done at reasonable values of λ; (ii) for $T = 300$K we obtain local sampling but a 10 ps trajectory is inadequate for locating alternate low energy conformers, especially for basins of larger volume. Increasing the length of the trajectory is not a satisfactory solution; (iii) increasing the bath temperature to increase intrabasin sampling has the undesirable effect of nonlocal sampling which leads to numerous inconsequential minimizations at step 3. The transition region between nonlocal sampling and local intrabasin sampling is extremely sharp.

For small values of λ the only mergers on a deformed surface involve predominantly high lying minima. On such surfaces any underlying structure of the potential energy surface will be preserved. Such a surface would also present lower energy barriers which may be easily traversed. We reconfigure the parameters for the MD trajectory at step 2, and instead of generating a trajectory on the undeformed surface we generate a trajectory using the seed structure on a moderately deformed surface. We choose $\lambda = 0.2662$ as the level on which an MD trajectory is generated. Points from this trajectory are quenched on the undeformed surface in accordance with steps 3 and 4 described above.

The TQ sampling method was applied to all 244 basins found on the $\lambda = 1.6813$ level. Results for 70 of these basins derived from 261 low energy conformers within 3 kcal/mole of the global minimum on the undeformed surface are summarized in Table 3. We find the lowest energy conformers within 55 of the 70 basins using default parameters of 10 ps trajectories, 100 minimizations per trajectory and MD sampling on the $\lambda = 0.2662$ surface as discussed above. We find the second lowest energy conformer for 7 of the 70 basins. In cases where the protocol fails to find the lowest energy conformer within a basin it is almost always due to the volume of the target basin, *i.e*, a large number of conformations need to be sampled in order to locate the lowest energy conformer. Alternatively, the lowest energy conformer may be a deep narrow well not easily accessible to an MD trajectory.

The two step procedure for conformational scanning successfully identifies low energy conformers which are representatives of distinct regions of conformational space. The method identifies all 10 lowest energy conformers and 15 of the twenty lowest energy conformers as being distinct representatives of different regions of the potential energy surface. The 70 basins which have lowest energy conformers derived from the 261 minima within 3 kcal/mole also encompasses 8251 of the 20469 minima on the potential energy surface, *i.e*., high lying minima map onto basins derived from lower energy conformers.

In implementing the TQ sampling protocol we used MD trajectories on slightly deformed surfaces rather than trajectories on the undeformed potential energy surface. Figure 9 validates this usage, showing the enhanced sampling obtained on slightly deformed surfaces. We used 10 ps MD trajectories for temperatures between $T = 150$K and $T = 600$K and a seed structure from the C6 basin on the undeformed surface and the $\lambda = 0.2662$ surface. For each temperature the number of minima sampled is quantified by quenching along 100 equally spaced points from a 10 ps trajectory. On the undeformed surface very few minima are sampled for temperatures $T <$ 350 K. Also this sampling is local in that all of the minima sampled are confined to the C6 basin.

The number of minima sampled goes through a sharp increase above T = 350 K due to extensive nonlocal and stochastic sampling representing a change in ensembles due to higher temperature. The very sharp transition region precludes the possibility of selecting an appropriate sampling temperature on the undeformed surface to ensure covering the spatial extent of a chosen basin. The situation changes on the deformed surface as illustrated by the increased local sampling for lower temperatures. The transition from low and intermediate temperatures to the high temperature regime is relatively gradual. The increased local sampling within a basin is due to the enhancement obtained on a deformed surface which presents lower energy barriers easily negotiated by a space covering method like MD. The idea of using smooth energy surfaces to enhance local sampling is in the general spirit of the local enhanced sampling algorithm of Elber and coworkers[30-32].

Discussion.

In this work we have shown how potential smoothing can be used as a robust method for hierarchical conformational search. This is established for cycloheptadecane by applying the *conformational scanning* algorithm to span the relevant conformational space. We selected this problem because the work of Saunders, *et al.*[4] has established this as a benchmark problem for conformational search.

Highlights of the *conformational scanning* technique include: (i) a ‘‘normal mode’’ basin hopping algorithm to map minima on a potential energy surface. This iterative algorithm generates a self-consistent mapping which in many cases spans the entire conformational space. (ii) The ‘‘normal mode’’ basin hopping algorithm has been applied to map basins on deformed surfaces, where the deformation parameter is set adiabatically. As shown in Table 2 the basis set of minimum energy basins which span conformational space can be generated in an extremely efficient manner. (iii) Clustering on deformed surfaces, where the deformation parameter is the

clustering ‘‘distance’’, shows weak correlations between structures within a cluster. We have analyzed the structural basis of clustering by comparing local dihedral angle patterns for structures within a cluster to structures across all clusters. Statistics indicate that a single string, for small clusters, or a few set of strings, for larger clusters, may be found to effectively describe all structures within a cluster. (iv) The mapping of minimum energy basins on a deformed surface, which forms a basis set of basins spanning conformational space, can be sampled locally in order to generate representative low energy conformers on the undeformed surface. The TQ method locates the lowest energy conformer for most of the basins identified at the $\lambda = 1.6813$ level for cycloheptadecane. Trajectories generated on deformed surfaces fully explore, but can be restricted to, the basins delineated by the mapping procedure. This is additional proof in favor of partitioning a potential energy surface based on potential smoothing. (v) Local sampling of basins is enhanced and accelerated by generating trajectories on slightly deformed surfaces. The enhancement comes about due to lower energy barriers on deformed surfaces.

Gaussian transform of a potential energy function. The basis of the current work is derived from the theoretical framework of Shalloway[11,13] and Straub[38,39] who have shown that coarse-grained potentials are a consequence of a Gaussian phase packet approximation to the Gibbs / Boltzmann distribution function. The normalized equilibrium Boltzmann distribution integrated over the momenta for an *N*-atom system at temperature $\beta = 1 / RT$ can be written in a Gaussian phase packet approximation as:[11]

$$\rho_{eq}(\{r\},\beta) = \frac{\exp[-\beta U(\{r\})]}{\int \exp[-\beta U(\{r\})]\, d\{r\}} \approx \sum_{i=1}^{n_p} P_i \rho^G_{\Lambda(i)}, \qquad [2]$$

$$\text{where } \rho^{G}_{\Lambda(i)} = \prod_{k=1}^{N} \psi^{i}_{k}(r^{i}_{k}) \text{ and } \psi^{i}_{k}(r^{i}_{k}) = \frac{1}{(2\pi\lambda^{2}_{k(i)})^{3/2}} \exp\left[-\frac{1}{2}\left(\frac{r^{i}_{k} - r^{o(i)}_{k}}{\lambda_{k(i)}} \right) \right], \qquad [3]$$

i.e., the equilibrium Boltzmann distribution is approximated as a sum of n_{P} macrostate Gaussian phase packets; each macrostate packet i of radius Λ_i and packet occupation probability P_i is written as a Hartree product of single atom Gaussian packets of width $\lambda_{k(i)}$ and packet centers position $r^{o(i)}_{k}$. For $\beta \to \infty$ or $T \to 0$, $n_P = 1$, $\Lambda_1 \to 0$ and the occupation probability of the global minimum is nearly unity. For $\beta \to 0$ or $T \to \infty$ or , $n_P = 1$, $\Lambda_1 \to \infty$ and the occupation probabilities for each of the local minima within a single macrostate are nearly equal. Between and $\beta \to \infty$ and $\beta \to 0$ the system can be partitioned into macrostates characterized by $n_P \geq 2$. Generating a hierarchical classification of a potential energy surface requires computing individual macrostate packet parameters including their radii Λ_i, number of packets, n_P, individual packet occupation probabilities P_i, and individual packet centers $\{r\}^{o}_{i}$ [11,30,40]. These equations become unwieldy for larger problems.

The coarse-grained description of the classical distribution function leads to a spatially averaged potential function or free energy, *i.e.*, for values of $\Lambda > 0$ spatial averaging transforms a function $F(\{r\})$ to $\langle F(\{r\}) \rangle_{\Lambda}$ where

$$\langle F(\{r\}) \rangle_{\Lambda} = \int F(\{r\})\, \rho^{\mathrm{G}}_{\Lambda}(\{r\},\{r^{o}\})\, d\{r\} \qquad [4]$$

Here $F(\{r\})$ could be the potential function $U(\{r\})$ or the unnormalized Gibbs free energy $\exp\left(\frac{-\beta U(\{r\})}{2} \right)$.

In our work we set the single atom packet widths in equation [3] to be equal to each other and $F(\{r\})$ is chosen to be the potential function $U(\{r\})$ in d-dimensions, $\Lambda^2 = 2d\lambda$, and we assume no explicit temperature dependence for the deformation parameter λ. For a potential function in d-dimensions this formalism reduces the coarse grain description of the potential function to the form:[38]

$$\langle U(\{r\})\rangle_\lambda = \int \frac{1}{(4\pi\lambda)^{d/2}} \exp\left\{ \frac{-\left\| \{r\} - \{r^o\} \right\|^2}{4\lambda} \right\} U(\{r\})\, d\{r\} \qquad [5]$$

which is precisely the form in equation [1]. The control parameter λ is set adiabatically to generate a smooth potential function. Mapping minima on a deformed surface is equivalent to mapping the macrostates described by the Gaussian packet approximation to the classical distribution function[11,13]. Our methods are a simplification of the work of Shalloway, *et al.*[11,13] and a generalization of earlier algorithms for potential smoothing such as the diffusion equation method[36,37] and local search aided potential smoothing[42,43].

Monte Carlo Minimization[29] has also been used with an adiabatic smoothing protocol to find the global energy minimum for crystal packing of small molecules[55]. In principle this method is similar to the ideas developed in our work, *i.e.*, deformed surfaces are sampled to generate basins of conformers which are tracked back to the undeformed potential energy surface to isolate families of low energy conformers. We have shown that potential smoothing facilitates efficient conformational search of a potential energy surface. Conformational scanning aided by potential smoothing can easily be adapted to larger molecules such as proteins which are likely to show a more pronounced hierarchical organization of basins of attraction on their energy landscapes[56].

Acknowledgments. The authors would like to thank Dr. Enoch S. Huang and Prof. Garland R. Marshall for useful discussions. This work was supported by a grant from the DOE Environmental Science Management Program. All programs used for calculations reported in this work are part of the TINKER modeling package, available via anonymous ftp from dasher.wustl.edu or WWW from the site http://dasher.wustl.edu/tinker/.

References and Notes.

1. Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature*, **1998**, *394*, 758.
2. Ball, K. D.; Berry, R. S.; Kunz., R. E.; Li, F. Y.; Proykova, A.; Wales, D. J. *Science*, **1996**, *271*, 963.
3. Berry, R. S.; Breitengaser-Kunz, R. *Phys. Rev. Lett*., **1995**, *74*, 3951.
4. Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. *J. Am. Chem. Soc*., **1990**, *112*, 1419.
5. Ngo, J. T.; Karplus, M. *J. Am. Chem. Soc*., **1997**, *119*, 5657.
6. Wang, C. S. *J. Comput. Chem*., **1997**, *18*, 277.
7. Leach, A. R. *Ann. Rev. Comput. Chem*., **1991**, *2*, 1.
8. Vasquez, M.; Némethy, G.; Scheraga, H. A. *Chem. Rev*., **1994**, *94*, 2183.
9. Weber, T. A; Stillinger, F. H. *J. Chem. Phys*., **1984**, *80*, 2742.
10. Stillinger, F. H.; Weber, T. A. *J. Chem. Phys*., **1988**, *88*, 5123.
11. Oresic, M.; Shalloway, D. *J. Chem. Phys*., **1994**, *101*, 9844.
12. Becker, O. M.; Karplus, M. *J. Chem. Phys*., **1997**, *106*, 1495.
13. Shalloway, D. *J. Chem. Phys*., **1996**, *105*, 9986.
14. Velikson, B.; Bascle, J.; Garel, T.; Orland, H. *Macromolecules*, **1993**, *26*, 4791.
15. Noguti, T.; Go, N. *Proteins: Struc. Func. Genet*., **1989**, *5*, 97,104,113,125,132.
16. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*. Oxford University Press, New York, **1987**.
17. Binder, K.; Heerman, D. W. *Monte Carlo Methods in Statistical Physics. An Introduction.* 3rd Edition, Springer-Verlag, Berlin, **1986**.
18. Berne, B. J.; Straub, J. E. *Curr. Opin. Struc., Biol.*, **1997**, *7*, 181.
19. Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci*., **1993**, *90*, 809.
20. Crippen, G. M.; Havel, T. M. *Distance Geometry and Molecular Conformation*. John Wiley & Sons, New York, **1988**.
21. Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *Bull. Math. Biol*., **1983**, *45*, 665.
22. Kuntz, I. D.; Thomason, J. F.; Oshiro, C. M. *Methds. Enzymol*., **1989**, *177*, 159.

23. Crippen, G. M. *J. Phys. Chem*., **1987**, *91*, 6341.
24. Bruccoleri, R. E.; Karplus, M. *Biopolymers*, **1987**, *26*, 137.
25. Smellie, A.; Teig, S. L.; Towbin, P. J. *J. Comput. Chem*., **199**, *16*, 171.
26. Vasquez, M.; Scheraga, H. A. *Biopolymers*, **1985**, *24*, 1437.
27. Dammkoehler, R. A.; Karasek, S. F.; Shands, E. B. F.; Marshall, G. R. *J. Comput. Aid. Molec. Desgn*., **1989**, *3*, 3.
28. Saunders, M. *J. Comput. Chem*., **1989**, *10*, 203.
29. Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA.,* **1987**, *84*., 6611.
30. Elber, R.; Karplus, M. *J. Am. Chem. Soc*., **1990**, *112*, 9161.
31. Ulitsky, A.; Elber, R. *J. Chem. Phys*., **1993**, *98*, 3380.
32. Simmerling, C.; Miller, J. L.; Kollman, P. A. *J. Am.Chem. Soc*., **1998**, *120*, 7149.
33. Wu, X.; Wang, S. *J. Phys. Chem. B*., **1998**, *102*, 7238.
34. Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A*., **1997**, *101*, 5111.
35. Moré, J.; Wu, Z. In *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding* (DIMACS Vol, *23*) Ed., P. M. Pardalos, D. Shalloway and G. Xue, Am. Math. Soc., Providence, **1996**, 151.
36. Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem*., **1989**, *93*, 3339.
37. Kostrowicki, J.; Piela, L.; Cherayil, B. J.; Scheraga, H. A. *J. Phys. Chem*., **1991**, *95*, 4113.
38. Ma, J.; Straub, J. E. *J.Chem. Phys*., **1994**, *101*, 533.
39. Straub, J. E. in *Recent Developments in Theoretical Studies of Proteins*, R. Elber, Ed., World Scientific, Singapore, **1996**.
40. Shalloway, D. *J. Global Optim*., **1992**, *2*, 281.
41. Andricioaei, I.; Straub, J. E. *J. Comput. Chem*., **1998**, *19*, 1445.
42. Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B*., **1998**, *102*, xxxx. *(in press)*
43. Nakamura, S.; Hirose, H.; Ikeguchi, M.; Doi, J. *J. Phys. Chem.*, **1995**, *99*, 8374.
44. The term ‘‘projection’’ as used here and in reference 42 does not imply a reduction in dimensionality of conformational space. Rather it refers to the relationship of catchment regions on smooth surfaces to specific regions of conformational space, *i.e.*, a ‘‘projected’’

catchment region may or may not be related to the basin of the global minimizer (see Figures 1a and 1b).

45. Shenkin, P.; McDonald, D. Q. *J. Comput. Chem.*, **1994**, *15*, 899.
46. Ponder, J. W. TINKER: Software Tools for Molecular Design, Version 3.7, Washington University School of Medicine, October 1998, (http://dasher.wustl.edu/tinker/).
47. Allinger, N. L. *J. Am. Chem. Soc*., **1977**, *99*, 8127.
48. For cycloheptadecane there are 17 eigenvalues and eigenvectors in torsional space corresponding to the 17 rotatable bonds.
49. If each step along an eigenvector were to represent change of a single torsion, the magnitude of this change would be approximately 6°.
50. It should be noted that the energy scan along the eigenvector is done in torsional space, while all minimizations are done in Cartesian space using a truncated Newton method with a preconditioned linear conjugate gradient solution of Newton's equations[51].
51. Ponder, J.W.; Richards, F.M. *J. Comput. Chem*., **1987**, *8*, 1016.
52. Derrida, B. *Phys. Rev. B.*, **1981**, *24*, 2613.
53. Hart, R. K.; Pappu, R. V.; Ponder, J. W. *J. Comput. Chem*., **1999**, *20*, xxxx. *(submitted)*
54. Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys*., **1984**, *81*, 3684.
55. Wawak, R. J.; Pillardy, J.; Liwo, A.; Gibson, K. D.; Scheraga, H. A. *J. Phys. Chem. A*., 1998, 102, 2904.
56. Dill, K. A.; Chan, H. S. *Nature Struc. Biol.*, **1997**, *4*, 10.

Figure Captions

Figure 1. (a) One dimensional schematic of an idealized potential smoothing protocol for global optimization. The original PES is transformed by successive application of the smoothing operator. The extent of smoothing is governed by a control parameter λ. As the surface is deformed higher lying minima merge into the basin of low lying broad wells and barriers separating minima are also lowered. Open circles are starting points for minimization on each level of smoothing and solid circles denote the result of local minimizations. Dashed arrows show the result of local optimizations and solid arrows represent adiabatic movement from a local minimum on one surface to the corresponding point on a different surface. **(b)** Schematic of a more realistic potential smoothing protocol for conformational searching. At intermediate levels of smoothing the underlying basin structure is ''projected'' out by filtering high frequency unstable minima within each basin. In this schematic the PES has two prominent basins labeled *A* and *B*. A potential smoothing protocol with the deformation parameter λ adiabatically set to either λ_1 or λ_2 reveals the underlying PES structure of two basins of attraction or macrostates. The parameter λ sets the length scale of PES deformation, *i.e.*, all minima under the Gaussian envelope of width 2λ merge into a common basin. A plausible search protocol would identify the two basins and hence the region of conformational space they span. A reversal protocol comprising of adiabatically decreasing deformation, denoted by solid arrows, followed by local minimization, denoted by dashed arrows, may be used to generate low energy conformational representatives within each basin. Additionally each basin identified by a representative structure through the reversal protocol can be sampled extensively using the representative structure as a seed. The superposition of extensively sampled basins can lead to accurate estimation of space covered thermodynamic averages.

Figure 2. (a) Distribution of energies for the 20460 unique minima for cycloheptadecane on the MM2 energy surface. All energies are within 25 kcal/mole of the global minimum. The number of minima found in each 0.1 kcal/mole energy bin is plotted as a function of increasing MM2 energy. **(b)** Low energy tail of Figure 2(a) showing the distribution of energies within 3 kcal/mole of the global minimum.

Figure 3. Plot showing the correlation of energies for 20469 minima on the MM2 energy surface and the undeformed Gaussian energy surface used to mimic the MM2 PES. Points on the plot represent the energy distance of a given minimum from the global minimum on the corresponding energy surfaces. The correlation coefficient for relative spacing of minima between the two potential energy surfaces is 0.9999.

Figure 4. Reduction in the number of minima for cycloheptadecane as a function of increased deformation. 20469 unique minima on the undeformed surface merge into a single convex basin on the $\lambda = 25$ PES. The figure shows a plot of the evolution of $\log_{10}$ of the number of unique minima as a function of the deformation parameter λ.

Figure 5. Schematic of a tree diagram to illustrate the partitioning of minima from the undeformed surface into clusters of minima on deformed surfaces. Each node on a given level of smoothing represents a region of conformational space from the undeformed surface. The clustering distance is the deformation parameter λ which measures the spatial extent of PES clustered into a single basin. The figure shows five minima labeled $m_1 ... m_5$ such that $E(m_1) < E(m_2) < E(m_3) < E(m_4) < E(m_5)$ on the undeformed surface. The abscissa represents a pseudo conformational coordinate which places members of a conformational basin close to each other. The ordinate represents increasing levels of surface deformation. For $\lambda = \lambda_1$ the energies of m_4

and m_5 are equal. This implies a merger of these two minima into a single basin denoted as node labeled m_4. Similarly m_1 and m_2 merger on the $\lambda = \lambda_2$ level where $\lambda_2 > \lambda_1$. For $\lambda = \lambda_3 > \lambda_2$ the basins labeled m_4 and m_1 are similar in energy and merge into a common basin labeled m_1. Finally for $\lambda = \lambda_4 > \lambda_3$ we obtain a single minimum denoted by a single node on the tree diagram implying a unimodal convex surface. It should be clarified that the label m_1 for the single surviving basin on the $\lambda = \lambda_4$ does not imply that this basin is related to the global minimum on the undeformed surface, because energy crossings could result in the ''projection'' of a catchment region related to a higher lying region on the undeformed surface. The label reflects the goal of an algorithm to search for the lowest lying conformer. Clearly the $\lambda = \lambda_3$ level illustrates the goal of *conformational scanning*. The energy surface on this level can be mapped to obtain three basins of attraction using the efficient iterative basin hopping scheme. Each basin can be tracked down to the undeformed surface to locate the lowest energy conformers within the basin. For a mapping generated on the $\lambda = \lambda_3$ level *conformational scanning* will obtain three unique minima on the undeformed surface, minima m_1, m_3 and m_4 which as shown in the diagram are low energy representatives spanning the entire extent of the available conformational space.

Figure 6. Tree diagram showing the many-to-few mapping of 20469 minima into 70 unique basins on the $\lambda = 1.9418$ level. Only the partitioning of structures using the 261 low energy conformers is shown. In reality the sizes of each of the clusters is larger representing the clustering of conformers across the potential energy surfaces. The labels indicate the ranks of energies on the MM2 energy surface.

Figure 7. Hierarchical clustering based on potential smoothing. The picture shows two representative structures each from the C1 and C6 clusters on the undeformed surface. Structures for each of the conformers are tracked as a function of increasing deformation λ to illustrate the structural significance of clustering.

Figure 8. Distributions of superposition distances for between pairs of structures from the C1 cluster and structures across the C2 through C8 clusters derived from partitioning on the λ = 1.6813 level. The C1−C1 distribution measures the smallest distances for superposition of structures within the C1 cluster. The C1−C*i* distributions for *i* = 2 ... 8 measure distances for pairs of structures from the C1 and C*i* clusters. Also shown are the mean superposition distance for each set of pairwise distances. The smallest average distance is for the C1−C1 set of distances indicating that structures within the cluster C1 are on average closer to each other compared to their distances to structures across other clusters.

Figure 9. Enhanced local sampling in a molecular dynamics trajectory is feasible on a slightly deformed surface. The figure shows a plot of the number of minima sampled as a function of increasing simulation temperature. The open circles are numbers obtained from TQ sampling on the undeformed surface for a trajectory generated using a seed minimum out of the C6 cluster. For T = 350 K there is a sharp enhancement in sampling noted by the increased number of minima seen using the TQ method. However a very large fraction of the minima sampled are nonlocal to the C6 basin. The transition region between the local and non-local regions is extremely sharp. The data for TQ sampling on the slightly deformed PES (λ = 0.2662) shows a gradual transition into the nonlocal sampling regime. The transition temperature is still T = 350 K noted by the change in sampling properties. However for temperatures less than 350 K there is an enhancement in the number of minima sampled, all of which are local to the C6 basin.

TABLE 1: Local structural patterns for cycloheptadecane conformers within some of the clusters identified on the λ = 1.6813 level.*

Cluster	n_c (N_s)	N_s-letter strings as cluster descriptors	n_m (n_{ov})	$\%_{ic}$	f_{ac} ($\%_{ac}$)
C1	9 (11)	S1: nnntnnttppt	6 (0)	67	22 (0.1)
		S2: ntnnntnnttp	5 (4)	56	57 (0.28
		S3: tnntnnptppt	3 (2)	33	198 (0.97)
		S4: nntnntnptpp	3 (2)	33	52 (.025)
C2	13 (11)	S1: tnntnnttppt	7 (0)	54	66 (0.32)
		S2: tnnttnttppt	4 (2)	31	32 (0.16)
		S3: tnnttntnntt	3 (1)	23	47 (0.23)
		S4: tnntnttnntt	3 (1)	23	36 (0.18)
C4	17 (9)	S1: ntnntnntt	8 (0)	47	364 (1.7)
		S2: nntnttnnt	6 (1)	35	204 (1.2)
		S3: ntnntnnpt	6 (2)	35	246 (1.2)
C5	203 (9)	S1: ntntnnttp	36 (0)	17.7	148 (0.73)
		S2: nntntnntt	35 (9)	17.2	160 (0.79)
		S3: nntnpptpt	31 (1)	15.3	105 (0.52)
		S4: ntnpptptp	29 (2)	14.3	72 (0.35)
		S5: nntntnnpt	26 (3)	12.8	167 (0.82)
		S6: ntnpptttp	24 (1)	11.8	74 (0.36)
		S7: nntttnptp	23 (6)	11.3	205 (1.01)
		S8: tnnpttnnt	22 (3)	10.8	125 (0.62)
C6	503 (9)	S1: ntnntnptp	77 (0)	15.3	200 (1.00)
		S2: ntnptpttp	68 (12)	13.5	196 (0.98)
		S3: nntnttnpt	64 (9)	12.7	234 (1.2)
		S4: tnntttnpt	61 (2)	12.1	206 (1.03)
		S5: ntnpttptp	59 (9)	11.7	140 (0.7)
		S6: ntnptpptt	59 (13)	11.7	273 (1.4)
		S7: ntnnttptp	58 (6)	11.5	202 (1.01)
		S8: nptttpptp	56 (4)	11.1	174 (0.9)
		S9: ntnttnptt	54 (6)	10.7	416 (2.1)
		S10: ntnttnttp	50 (1)	9.9	222 (1.1)

TABLE 1: Local structural patterns for cycloheptadecane conformers within some of the clusters identified on the λ = 1.6813 level.*

Cluster	n_c (N_s)	N_s-letter strings as cluster descriptors	n_m (n_{ov})	$\%_{ic}$	f_{ac} ($\%_{ac}$)
C7	117 (9)	S1: ntnnttppt	30 (0)	25.6	248 (1.2)
		S2: nptpttppt	16 (3)	13.6	171 (0.84)
		S3: ttnntnptt	15 (5)	12.8	289 (1.4)
		S4: tnntnttpt	14 (5)	11.9	222 (1.1)
		S5: ntnnttptt	14 (2)	11.9	297 (1.4)
		S6: ntnttpntt	13 (2)	11.1	344 (1.7)
		S7: ntnttnntt	13 (4)	11.1	159 (0.8)
C8	488 (9)	S1: ntnnttpnt	70 (0)	14.3	372 (1.9)
		S2: ntnttpnnt	56 (7)	11.5	352 (1.8)
		S3: nttpnttnt	51 (12)	10.4	314 (1.6)
		S4: npttntnnt	50 (10)	10.2	227 (1.1)
		S5: nnpttntnn	49 (6)	10.0	153 (0.8)
		S6: nppptpnnt	46 (3)	9.4	72 (0.4)
		S7: ntnnpttnt	45 (8)	9.2	255 (1.3)
		S8: ntnttnptt	41 (8)	8.4	429 (2.2)
		S9: nttnpttnt	40 (5)	8.2	225 (1.1)
		S10: nnntnpptp	36 (4)	7.4	146 (0.73)
C9	88 (9)	S1: nntnttpnt	24 (0)	27.3	270 (1.3)
		S2: nttnntnpt	16 (4)	18.2	205 (1.0)
		S3: ntnnttntn	15 (5)	17.0	325 (1.6)
		S4: nptpttppt	14 (4)	15.9	173 (0.8)
		S5: nptppttpt	12 (3)	13.6	151 (0.7)
		S6: nptntnntn	11 (1)	12.5	112 (0.5)
C10	28 (9)	S1: ntnnttppt	11 (0)	39.2	267 (1.3)
		S2: nntnpptpt	6 (2)	21.4	130 (0.6)
		S3: nttptppnt	5 (2)	18.0	242 (1.2)
		S4: ntnpptptt	5 (2)	18.0	125 (0.6)
		S5: ntnpptpnt	5 (1)	18.0	154 (0.8)

* Key for Table 1:

Cluster: Label which denotes the lowest energy conformer within the cluster.
n_c: Number of structures within the cluster.

N_s: Length of the dihedral angle string used to characterize local structural patterns within the cluster.
n_m: Number of structures within a cluster for which a given string occurs.
n_{ov}: Number of structures from the set of n_m structures which overlap with the largest set within the cluster.
$\%_{ic}$: Intra-cluster percentage occurrence of a dihedral angle string.
f_{ac}: frequency of occurrence of a string across all other clusters
$\%_{ac}$: Percentage occurrence of a string across all other clusters.

TABLE 2: Results from the *ab initio* mapping of deformed potential energy surfaces using the ‘‘normal mode’’ basin hopping method.

Deformation λ	Number of basins on the smooth PES	Number of minima within 25 kcal/mole threshold	Number of minima found via mapping protocol	Percentage of accessible PES covered by the mapping	Number of minimizations required for the mapping
0.0000					
0.8714	2123	2123	1919	90.4	35166
0.9259	1885	1885	1693	89.8	30509
0.9826	1648	1648	1486	90.2	25755
1.0415	1402	1402	1252	89.3	21401
1.1027	1187	1186	1061	89.5	17749
1.1664	996	995	876	88.0	14319
1.2324	854	852	706	82.9	11322
1.3008	723	713	577	80.9	9235
1.3718	626	607	508	83.7	8154
1.4452	546	510	439	86.1	7058
1.5213	459	423	351	83.0	5661
1.6000	364	339	273	80.5	4329
1.6813	248	240	191	79.6	2911
1.7653	152	149	130	87.2	1942
1.8522	71	71	70	98.6	1016
1.9418	63	63	59	93.7	835
2.0342	63	63	56	88.9	792
2.1296	63	63	56	88.9	810
2.2278	63	62	58	93.5	856
2.3291	63	59	57	96.6	804
2.4334	63	55	53	96.4	752
2.5407	63	55	53	96.4	746
2.6511	63	55	52	94.5	717
2.7648	63	55	53	96.4	737
2.8816	63	55	54	98.2	766
3.0016	63	55	52	94.5	724
3.2516	63	55	53	96.4	716
3.5152	63	55	54	98.2	727
3.6251	63	55	52	94.5	773
3.7925	63	55	52	94.5	725

TABLE 2: Results from the *ab initio* mapping of deformed potential energy surfaces using the ‘‘normal mode’’ basin hopping method.

Deformation λ	Number of basins on the smooth PES	Number of minima within 25 kcal/mole threshold	Number of minima found via mapping protocol	Percentage of accessible PES covered by the mapping	Number of minimizations required for the mapping
3.9366	63	55	52	94.5	731
4.0842	63	55	53	96.4	747
4.3904	63	55	54	98.2	805
4.5490	63	55	54	98.2	781
4.8778	47	39	35	89.7	519
5.0479	40	32	31	96.9	452
5.2219	34	27	24	88.9	363
5.4000	31	25	18	72.0	256
5.5820	29	25	16	64.0	241
5.7680	28	25	21	84.0	322
5.9582	26	24	21	87.5	321
6.1524	25	23	17	73.9	254
7.1874	20	19	15	78.9	229
12.8000	6	6	4	66.7	76

TABLE 3: Results of trajectory quench (TQ) sampling to locate lowest energy conformers within each basin. The name of the cluster points to the lowest energy conformer within the basin. For instance C1 denotes the basin containing structure 1, C2 the basin containing structure 2, *etc*. In cases where the TQ sampling does not converge to the lowest energy conformer we indicate the rank of the conformer found within the basin in parentheses.

Cluster	Number of minima within the cluster	ID and (intracluster rank) of structure found using TQ sampling	MM2 energy in kcal/mole of structure found using intracluster TQ sampling
C1	9	[1]	19.068015
C2	13	[2]	19.077409
C3	3	[3]	19.437156
C4	17	[4]	19.450893
C5	203	[5]	19.657099
C6	503	[6]	19.729217
C7	117	[7]	19.732780
C8	423	[8]	19.840039
C9	88	[9]	19.878884
C10	28	[10]	19.957866
C11	34	[11]	19.986874
C15	45	[15]	20.201107
C18	23	[18]	20.273776
C20	18	[20]	20.348724
C21	135	[21]	20.378235
C27	225	[43] (2)	20.725894 (2)
C30	244	[38] (2)	20.667246 (2)
C31	8	[31]	20.565374
C33	13	[33]	20.578204
C34	28	[34]	20.628833
C39	80	[39]	20.686371
C42	32	[42]	20.723730
C44	340	[89] (2)	21.186665
C49	3	[49]	20.775245
C51	43	[51]	20.796096
C52	71	[52]	20.816618
C54	27	[54]	20.836406
C56	156	[56]	20.899172
C62	24	[62]	20.981552

TABLE 3: Results of trajectory quench (TQ) sampling to locate lowest energy conformers within each basin. The name of the cluster points to the lowest energy conformer within the basin. For instance C1 denotes the basin containing structure 1, C2 the basin containing structure 2, *etc*. In cases where the TQ sampling does not converge to the lowest energy conformer we indicate the rank of the conformer found within the basin in parentheses.

Cluster	Number of minima within the cluster	ID and (intracluster rank) of structure found using TQ sampling	MM2 energy in kcal/mole of structure found using intracluster TQ sampling
C63	4	[63]	20.987552
C67	287	[67]	21.047738
C71	57	[71]	21.071477
C75	64	[283] (11)	22.138811
C81	58	[211] (5)	21.846210
C88	36	[88]	21.178008
C105	93	[105]	21.282901
C107	15	[107]	21.300658
C109	43	[109]	21.305627
C110	165	[110]	21.308996
C112	36	[112]	21.319145
C117	64	[476] (4)	22.589420
C118	23	[118]	21.380500
C120	893	[381] (2)	22.389978
C121	261	[144]	21.530022
C141	633	[141]	21.516036
C143	40	[279] (3)	22.132350
C157	103	[157]	21.611177
C160	104	[160]	21.645035
C162	32	[162]	21.652816
C164	19	[164]	21.659881
C165	188	[165]	21.664341
C166	145	[688] (10)	21.672975
C170	195	[170]	21.694756
C173	199	[173]	21.704688
C178	160	[178]	21.730460
C179	49	[209] (2)	21.839144
C187	81	[187]	21.766724
C203	149	[365] (2)	22.362882

TABLE 3: Results of trajectory quench (TQ) sampling to locate lowest energy conformers within each basin. The name of the cluster points to the lowest energy conformer within the basin. For instance C1 denotes the basin containing structure 1, C2 the basin containing structure 2, *etc*. In cases where the TQ sampling does not converge to the lowest energy conformer we indicate the rank of the conformer found within the basin in parentheses.

Cluster	Number of minima within the cluster	ID and (intracluster rank) of structure found using TQ sampling	MM2 energy in kcal/mole of structure found using intracluster TQ sampling
C205	70	[521] (2)	24.695445
C212	138	[599] (3)	22.816736
C225	101	[1462] (9)	23.884296
C226	22	[226]	21.923409
C227	273	[624]	22.311807
C230	96	[232]	21.941825
C239	121	[239]	21.960682
C241	30	[241]	21.977562
C246	34	[582] (3)	22.789037
C251	269	[943] (4)	23.350588
C255	39	[566] (2)	22.758141
C260	51	[260]	22.067869

Figure 1

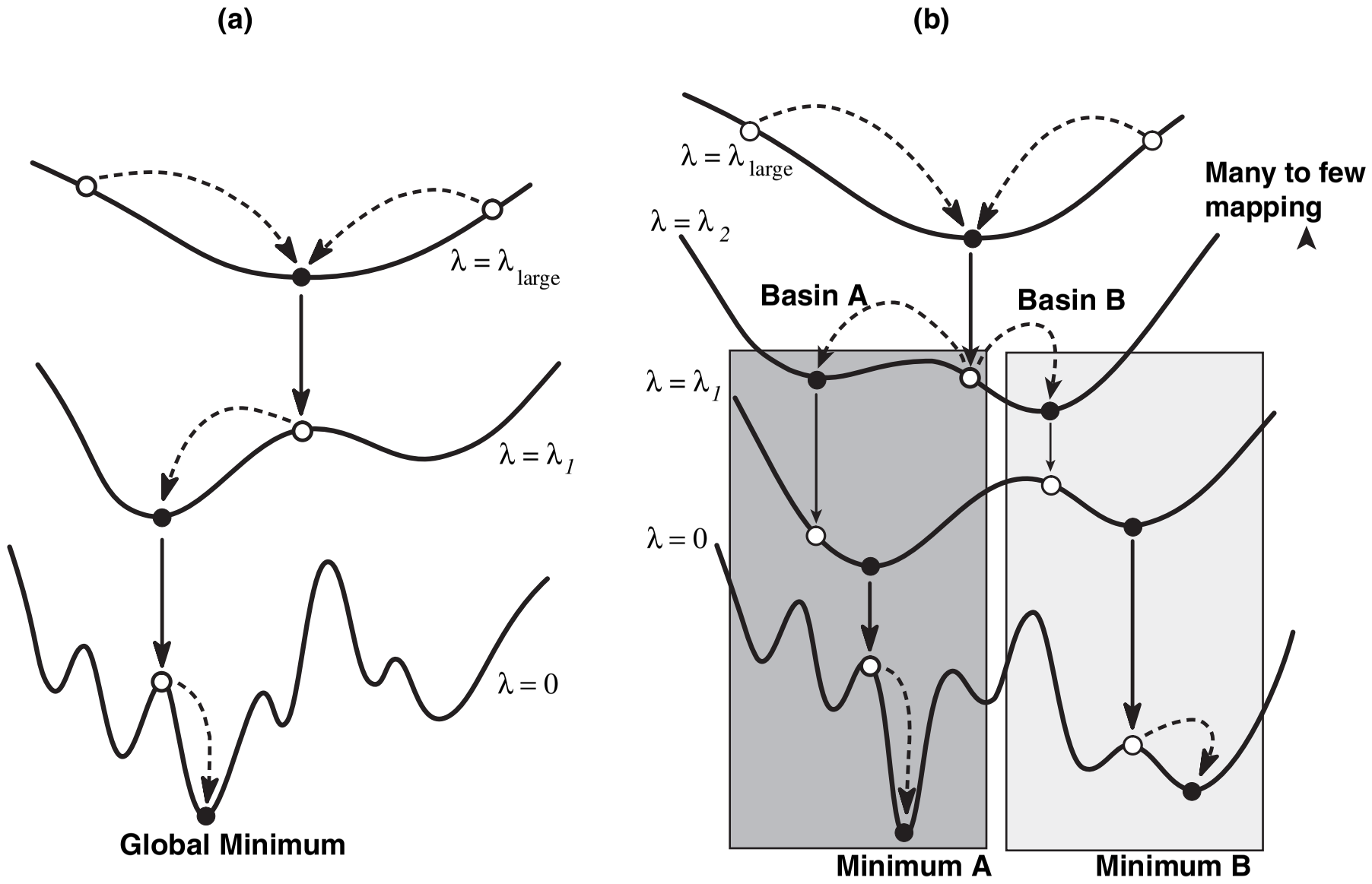

(a)
λ = λ large
λ = λ 1
λ = 0
Global Minimum
(b)
λ = λ large
λ = λ 2
Many to few mapping
Basin A
Basin B
λ = λ 1
λ = 0
Minimum A
Minimum B

Figure 2a

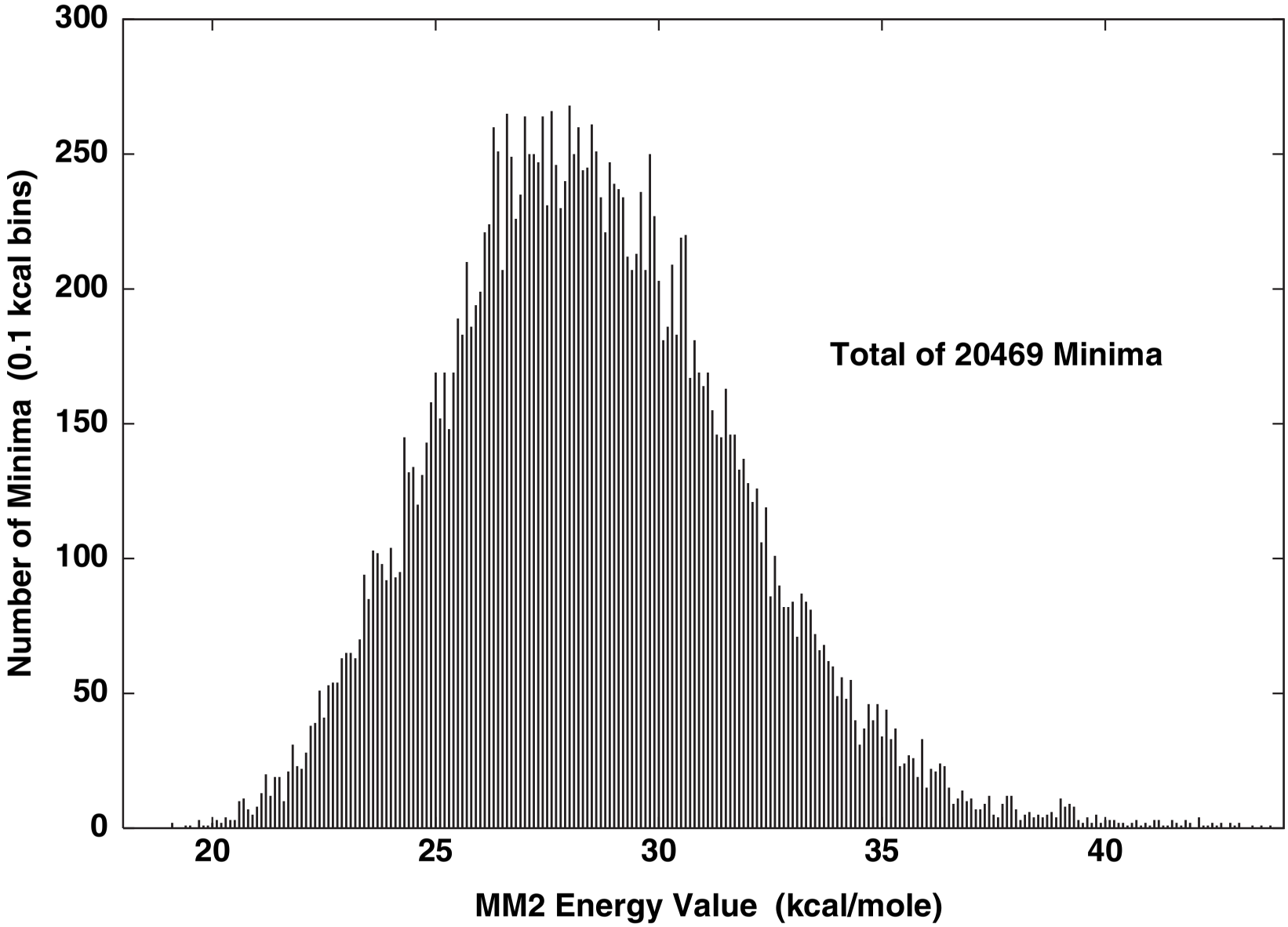

300
250
200
150
100
50
0
Number of Minima (0.1 kcal bins)
Total of 20469 Minima
20
25
30
35
40
MM2 Energy Value (kcal/mole)

Figure 2b

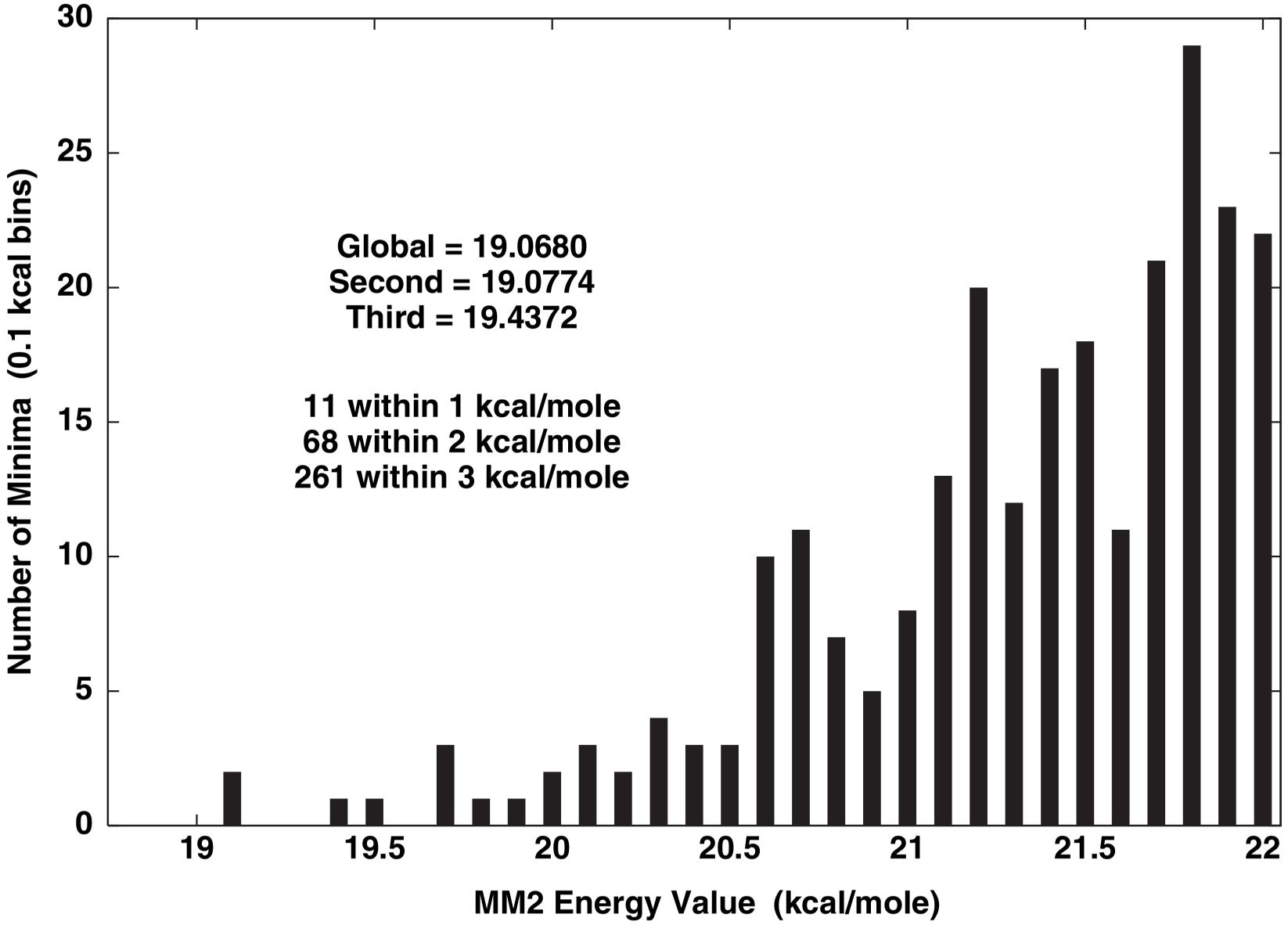

Number of Minima (0.1 kcal bins)
30
25
20
15
10
5
0
Global = 19.0680
Second = 19.0774
Third = 19.4372
11 within 1 kcal/mole
68 within 2 kcal/mole
261 within 3 kcal/mole
19
19.5
20
20.5
21
21.5
22
MM2 Energy Value (kcal/mole)

Figure 3

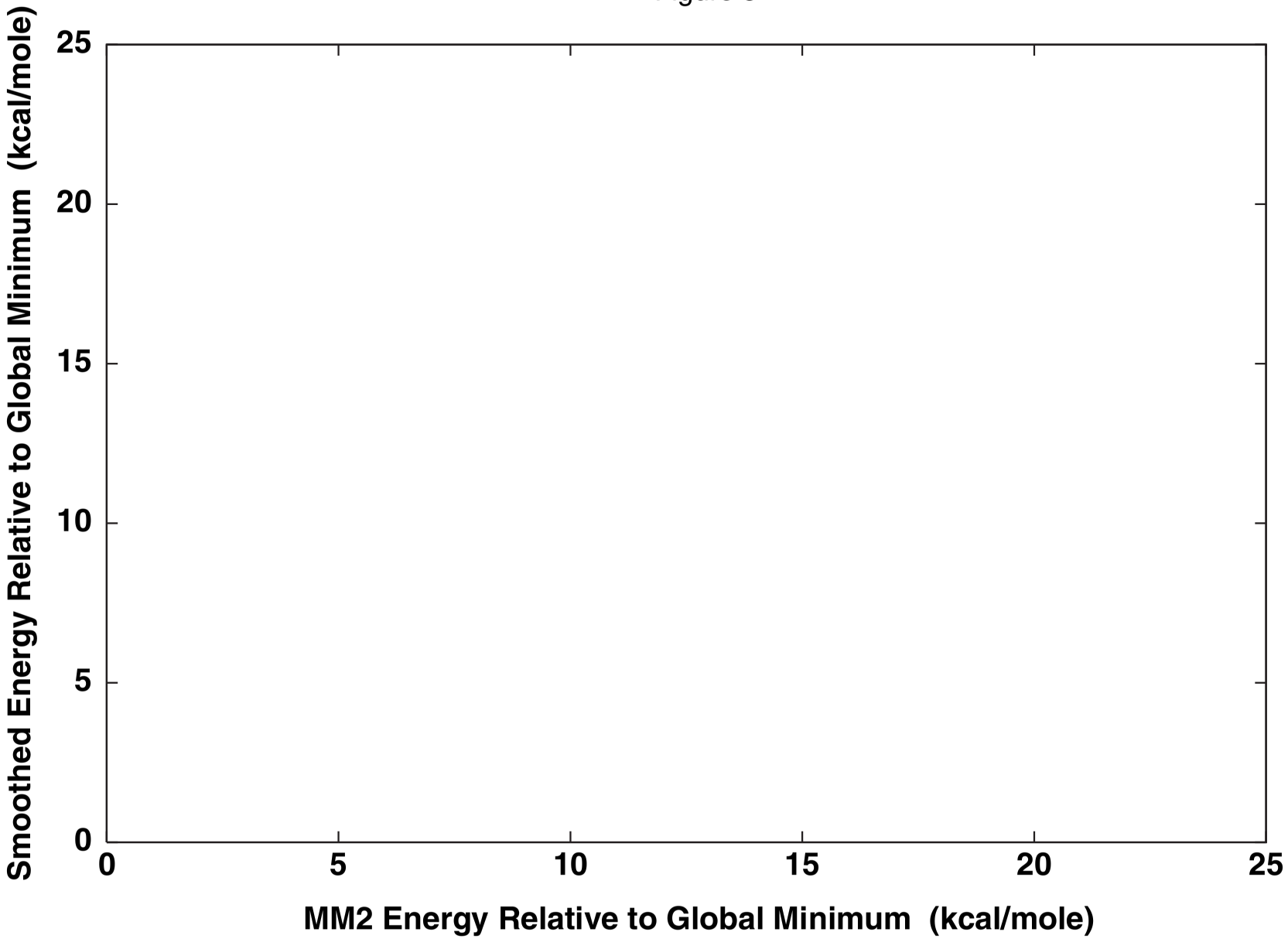

Smoothed Energy Relative to Global Minimum (kcal/mole)
25
20
15
10
5
0
0
5
10
15
20
25
MM2 Energy Relative to Global Minimum (kcal/mole)

Figure 4

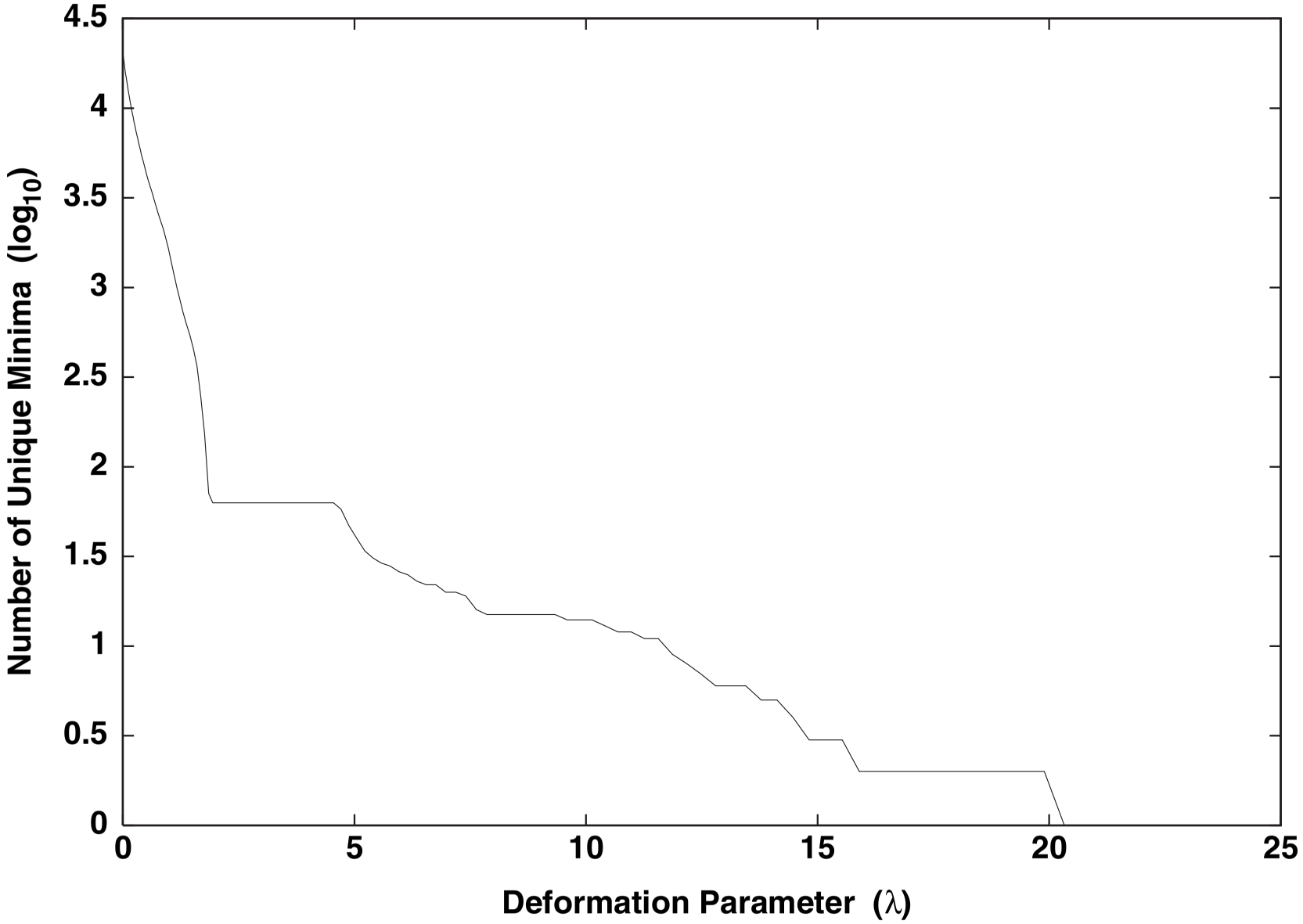

Number of Unique Minima (log10)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
5
10
15
20
25
Deformation Parameter (λ)

Figure 6

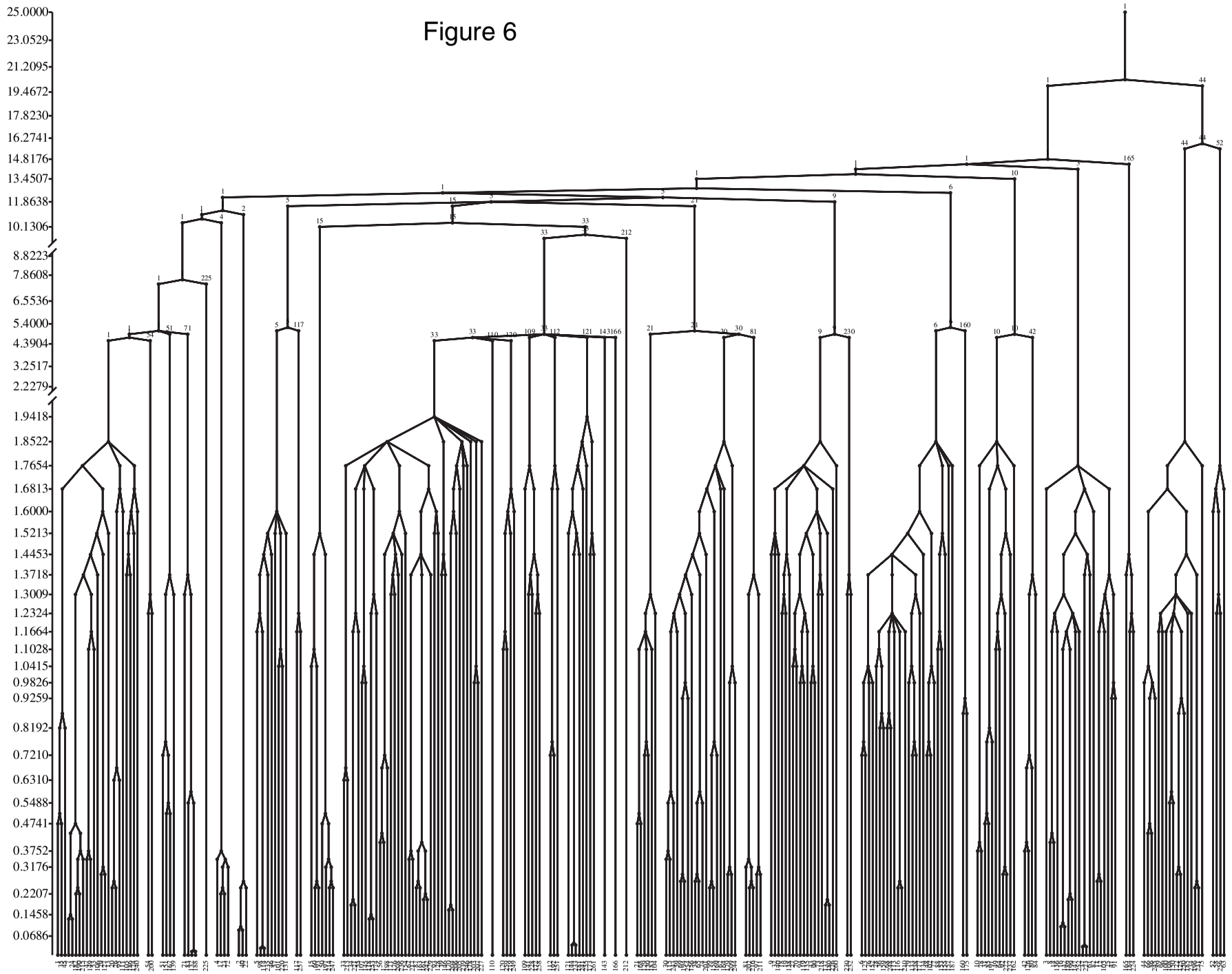
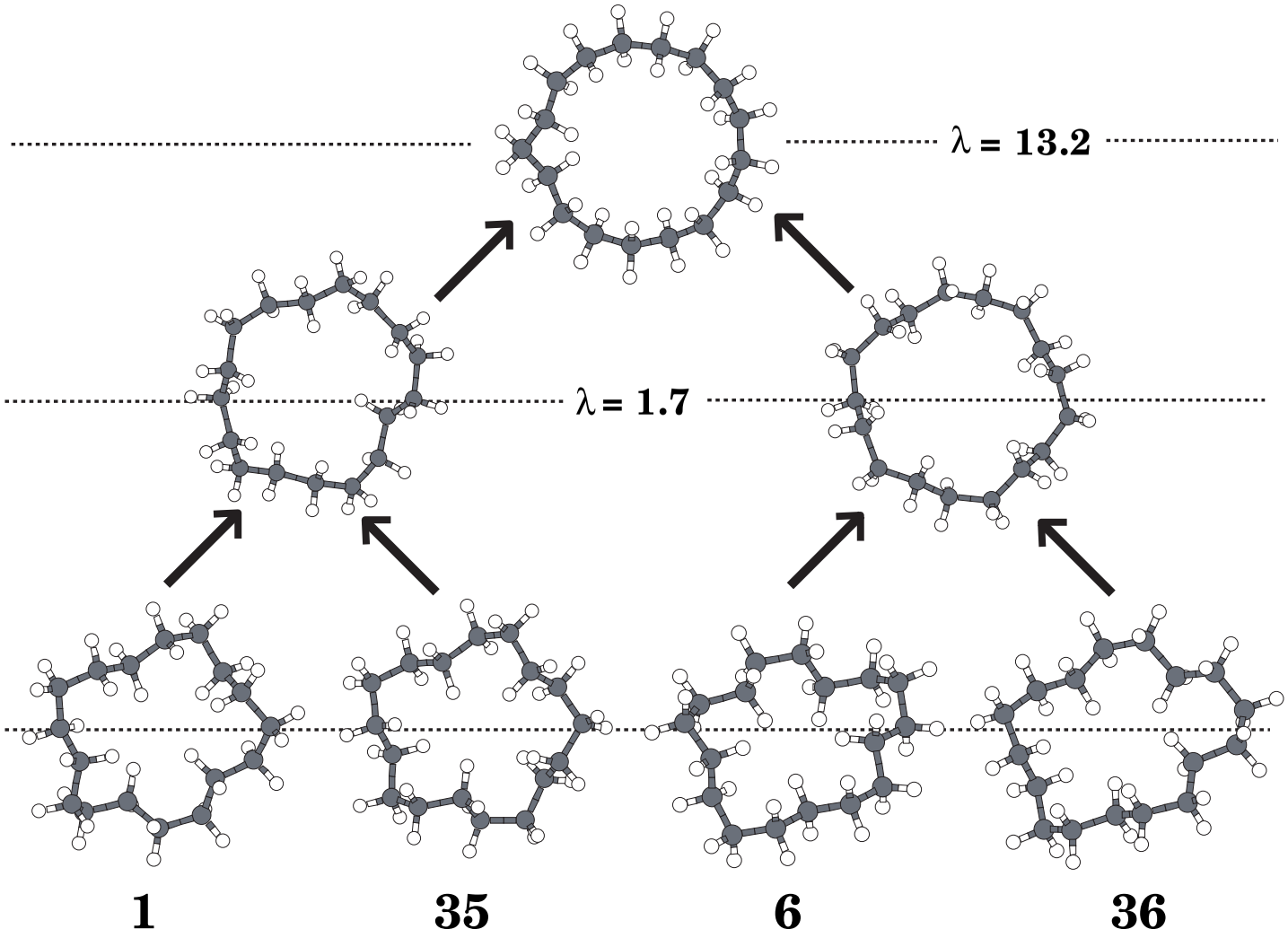

25.0000
23.0529
21.2095
19.4672
17.8230
16.2741
14.8176
13.4507
11.8638
10.1306
8.8223
7.8608
6.5536
5.4000
4.3904
3.2517
2.2279
1.9418
1.8522
1.7654
1.6813
1.6000
1.5213
1.4453
1.3718
1.3009
1.2324
1.1664
1.1028
1.0415
0.9826
0.9259
0.8192
0.7210
0.6310
0.5488
0.4741
0.3752
0.3176
0.2207
0.1458
0.0686

Figure 7



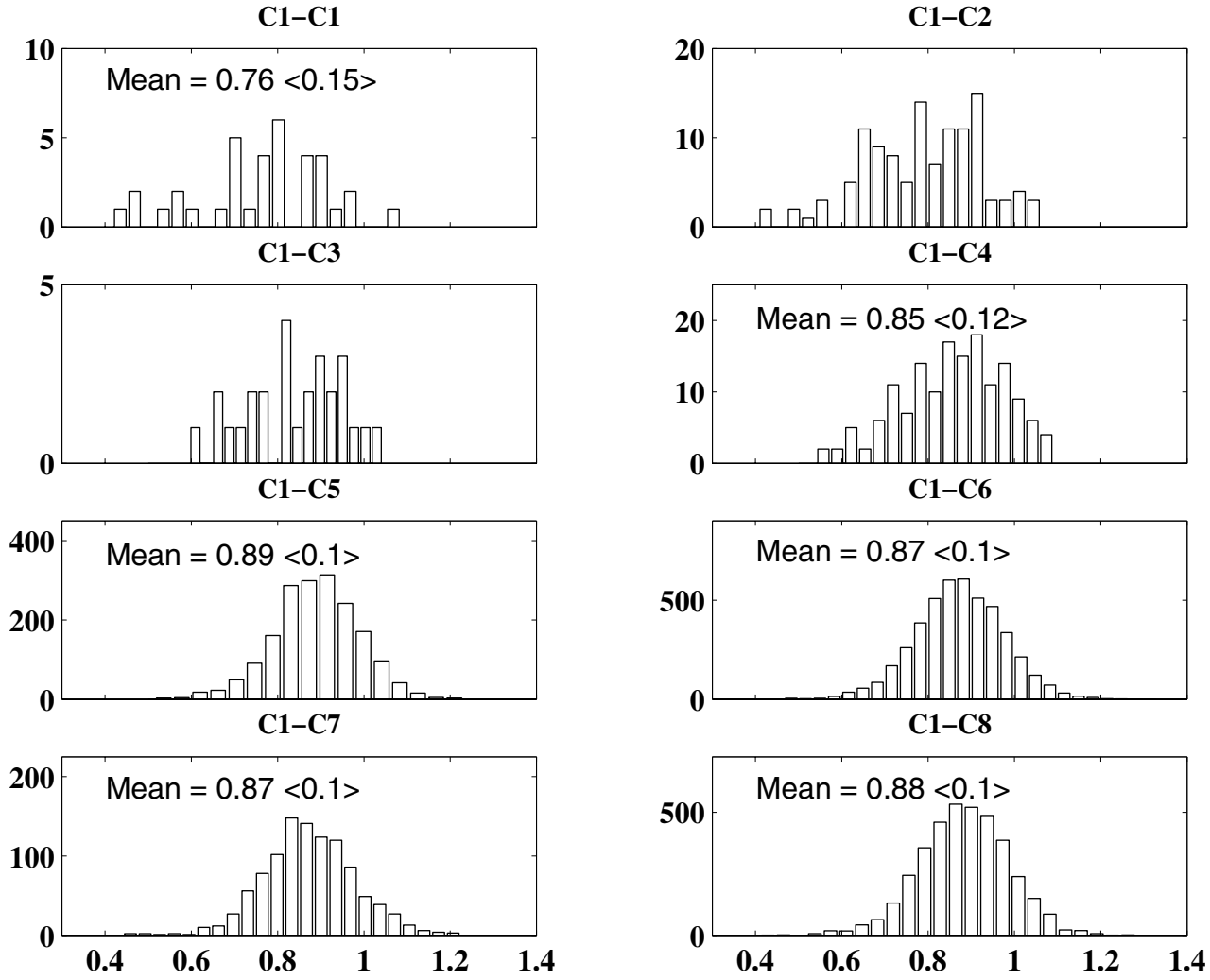
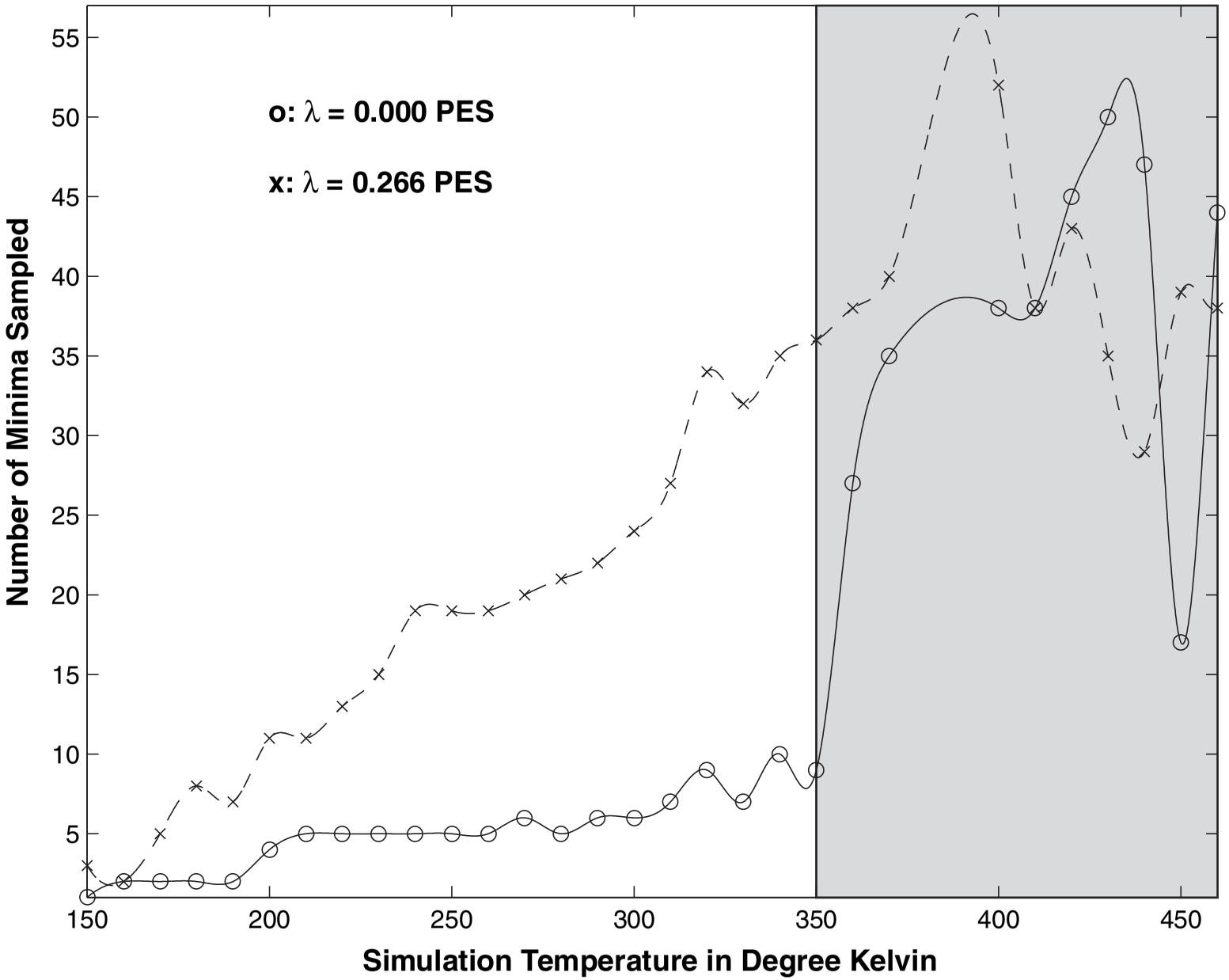

C1–C1
Mean = 0.76 <0.15>
C1–C2
C1–C3
C1–C4
Mean = 0.85 <0.12>
C1–C5
Mean = 0.89 <0.1>
C1–C6
Mean = 0.87 <0.1>
C1–C7
Mean = 0.87 <0.1>
C1–C8
Mean = 0.88 <0.1>
0.4
0.6
0.8
1
1.2
1.4

Figure 8

Figure 9


o: λ = 0.000 PES
x: λ = 0.266 PES
Number of Minima Sampled
Simulation Temperature in Degree Kelvin
55
50
45
40
35
30
25
20
15
10
5
150
200
250
300
350
400
450