

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231394657>

New Theoretical Methodology for Elucidating the Solution Structure of Peptides from NMR Data. 1. The Relative Contribution of Low-Energy Microstates to the Partition Function

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY · MARCH 1995

Impact Factor: 2.78 · DOI: 10.1021/j100013a065

CITATIONS

29

READS

16

3 AUTHORS, INCLUDING:



Jooyoung Lee

Korea Institute for Advanced Study

120 PUBLICATIONS 3,929 CITATIONS

SEE PROFILE

New Theoretical Methodology for Elucidating the Solution Structure of Peptides from NMR Data. 1. The Relative Contribution of Low-Energy Microstates to the Partition Function

Hagai Meirovitch,* Eva Meirovitch,* and Jooyoung Lee

Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-4052

Received: September 26, 1994[⊗]

Under certain experimental conditions peptide solutions can generate sufficient NMR data for structure determination. Yet, unlike proteins, peptides typically prevail as ensembles of interconverting structures, and therefore, the experimental variables (which are intensities of cross peaks in nuclear Overhauser enhancement spectroscopy (NOESY) spectra, or NOEs) are average quantities. The process of structure elucidation is complex, and many related questions are still open to debate. We have developed a new theoretical methodology for treating ensembles of interconverting conformations which is based on purely thermodynamic grounds. The peptide is described by a potential energy function, and its conformational space is viewed as a collection of microstates, which are local conformational regions around energy minima. The overall methodology enables one to identify the thermodynamically most stable microstates, determine their populations, calculate the individual microstate NOEs, and obtain the overall NOEs as averages over the individual contributions, weighted by the microstate populations. In this paper we develop for the first time theoretical methods for calculating the relative contribution of microstates to the partition function, as their minimum energy is increased above the global energy minimum (GEM). This is necessary for determining which microstates should be considered in detail in the next stage of the analysis. The new methods presented herein are applied to the pentapeptide Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH) described by the potential energy function ECEPP, with calculations carried out at 280 K. We find that the microstates contained within relatively small ranges of 2 to 3 kcal/mol above the GEM (or above the lowest harmonic free energy) constitute 0.59 to 0.75 of the partition function. This should be compared with energy ranges up to 15 kcal/mol set arbitrarily in previous studies. A detailed comparison of theoretical predictions with experimental data, obtained from a cryoprotective Leu-enkephalin solution at 280 K, will be carried out in the next paper of this series.

I. Introduction

Multidimensional nuclear magnetic resonance (NMR) became in the last 15 years the only physical technique that can generate three-dimensional (3D) structures of biomolecules in solution.^{1,2} This technique relies on a relaxation phenomenon (the nuclear Overhauser enhancement (NOE) effect) and provides the *dynamic structure* of the system investigated.^{3–5} The experimentally measured variables are intensities of cross peaks in nuclear Overhauser enhancement spectroscopy (NOESY) spectra, denoted as NOEs, which are in direct proportion to r^{-6} , where r is the spatial distance between two protons. The method is readily applicable to globular proteins^{1,2} that are moderately flexible, i.e., span only a limited region of conformational space around their native structure (such a region is called microstate, to be discussed in detail in section II). In this case intense long range NOEs, which are generated by pairs of constituents that are close in space, but distant along the chain, will be observed in the spectrum. On the other hand, if the system is highly flexible—such as an aqueous solution of a small linear peptide—a multitude of backbone conformations, spanning a large portion of the conformational space, will coexist. The long range NOEs will be weak and eventually undetectable.⁶ This stems from the fact that the distribution of spatial distances r between two atoms that are separated by t Å along the chain is peaked around the average value $\langle r^2 \rangle^{1/2}$ that increases with t as $\langle r^2 \rangle^{1/2} \sim t^\nu$, where $\nu = 0.5$ and ~ 0.59 for a polymer in a θ solvent and a good solvent, respectively.^{7,8} Thus, the probability of closing long range loops is low. The average $\langle r^{-6} \rangle$ will also decrease

with increasing t but more slowly than $t^{-6\nu}$, since its main contribution comes from relatively small distances r , with r smaller than $\langle r^2 \rangle^{1/2}$ (see refs 9 and 10).

There exist, however, systems of intermediate flexibility (such as peptides of protein folding initiation sites,^{11a,b} immunogenic peptides^{11c–f}, and peptides dissolved in membrane-like, organic, or cryoprotective solvents^{11g}), which give rise to sufficiently intense cross-peaks generated by long range NOEs. Because of the r^{-6} dependence pointed out above, the mere detectability of a well-resolved NOESY spectrum suggests the prevalence of a small number of dominant backbone conformations (microstates) in thermodynamic equilibrium. In this case, however, the measured NOE intensities (and the related distances) represent averages over the contributions of these microstates. The analysis of such NMR data is complex, since it requires identifying the dominant microstates, determining their populations, and calculating their individual contribution to the (average) NOE intensities.

The translation of NMR data obtained from globular proteins into their dynamic structure is, at present, an accomplished and widely used methodology.^{1,2} Conversely, the existing tools for linear peptides still leave several basic questions unsolved.

Previous approaches for determining the NMR structure of peptide solutions fall into several categories. A commonly used approach is to acknowledge the prevalence of multiconformational equilibria and interpret the experimental data in qualitative ways.^{11a–f,12} Quantitative methods based on restrained distance geometry,¹³ or molecular dynamics,^{14,15} have been used to determine the 3D structure of cyclic peptides. Commonly the NMR data are interpreted in terms of one distinct conformation;

[⊗] Abstract published in *Advance ACS Abstracts*, March 1, 1995.

in some instances two interconverting structures have been invoked,^{17,18} but populations have not been calculated theoretically.^{16–18}

A new method, called a multiconformational evaluation of distance information using a stochastically constrained minimization algorithm (MEDUSA), has been set forth recently by Ernst.^{19,20} The underlying hypothesis is that individual conformations may violate some of the NOE distance restraints, which are fulfilled only by the entire dynamic set of substates. The scheme includes a restrained conformational search based on a sensible, but otherwise arbitrary, upper-bound energy range. All possible pairs are tested as potential exchange systems, with probabilities determined by minimizing a target function. A statistical picture, depicting the number of structurally similar pairs corresponding to a given error in the target function, is obtained. MEDUSA has been applied so far successfully to the cyclic decapeptide antamanide.

A few additional approaches, which treat the problem of conformational multiplicity explicitly, have been developed.^{21,22} The technique of Nikiforovich and co-workers²¹ has been applied to a large number of systems. In general, these methods consist of two stages. First, a set of reasonable conformations is determined, using any of the currently available empirical potential energy functions. Conformations are retained if their energies do not exceed an arbitrarily set threshold, they differ structurally, and they are consistent with a subset of the NMR restraints. In a second stage the respective populations, viewed as free parameters, are determined using statistical best-fit procedures.

The main disadvantages of the previously developed approaches is that, to a large extent, the set of conformations used is chosen arbitrarily. Therefore the best-fit populations are not thermodynamically meaningful, and the structural information derived cannot possibly be unique.

We have developed a new methodology for treating ensembles of interconverting conformations which is based on purely thermodynamic grounds and hence largely free of the arbitrariness and limitations inherent in the “best-fit” methods. It enables one to identify the thermodynamically most stable microstates, determine their populations, and calculate their individual contribution to the (average) NOE intensities. Recently we have reported on two methods for searching the conformational space of a peptide, where the search is carried out with respect to the energy²³ and the harmonic free energy,²⁴ respectively. These methods, together with other available techniques,^{25–28} enable one to generate a comprehensive set of low-energy structures of a small peptide.

In the present study we proceed by developing theoretical methods and criteria for determining the most stable microstates, i.e., those which contribute most significantly to the partition function. In particular, these methods determine the energy range above the global energy minimum (GEM) to which the most stable microstates pertain. Values up to 15 kcal/mol have been assigned arbitrarily to this range in previous studies.^{21,22,29–31} In this work we find a much smaller range, on the order of 2–3 kcal/mol, when our methods are applied to the ECEPP³²-modeled pentapeptide Leu-enkephalin at 280 K. The magnitude of this range is extremely important, since it determines the set of microstates to be studied in detail in the next stage of our methodology (and, for that matter, to be used in any theoretical analysis of NOEs or other experimental variables). In the next stage,³³ a comprehensive conformational search is carried out using the methods mentioned above,^{23–25a} and a set of *different* energy-minimized structures is obtained within the relevant energy range. Each member of this set is used as a “seed” for a Monte Carlo (MC)³⁴ simulation, which spans a relatively large

spatial region around the seed, called “MC microstate” (to be distinguished from the “microstate” concept defined and discussed in this paper). The free energy (hence, the population) of each MC microstate is calculated using the local states (LS) method.^{35,36} The MC sample is also used to calculate the respective MC microstate NOEs. Finally, averages over the NOEs of the individual MC microstates, weighted by their populations, are calculated and compared with the experimental data.³⁷

II. Theory

Let us assume that a conformation of a peptide is defined by a set of internal variables ω which span the entire conformational space Ω and that the energy $E(\omega)$ is given by a potential energy function such as ECEPP.³² We define a microstate $\Omega_i(\omega_i)$ as the ensemble of conformations within the basin of attraction of a *local* energy minimum E_i^m . Namely, a conformation ω pertains to microstate i if local energy minimization starting from ω leads to ω_i . The contribution Z_i of microstate i to the total partition function Z is

$$Z_i = \int_{\Omega_i} \exp(-E(\omega)/k_B T) d\omega \quad (1)$$

where T is the absolute temperature and k_B is the Boltzmann constant. One can view Ω as a collection of microstates $\{\Omega_i\}$, where in the limit of a complete set of local energy minima $\Omega = \cup \Omega_i$, and Z can be obtained as $\sum Z_i$. This argument assumes a negligible contribution to Z from individual conformations corresponding to maxima and saddle points, which cannot naturally be assigned to a particular basin of attraction, or inherent structure.³⁸

In order to estimate the relative contribution of microstates to Z , one selects a large energy range (say, 10 kcal/mol) above (and including) the GEM and divides it into f small intervals (bins) j ($1 \leq j \leq f$) of energy $\Delta E = 10/f$. The energy E_j of interval j is represented by its mean energy value. A microstate i pertains to interval j if its minimal energy E_i^m falls within the interval limits.

II.1. Exact Procedure. One can define an exact procedure for estimating $p_j(\text{exact})$, the relative contribution to Z of the microstates that belong to bin j . A large number of conformations l , with energies E_t , $1 \leq t \leq l$, is generated at random. The individual structures are energy minimized, in order to identify the microstate i , and the related bin j , to which they pertain. The probability $p_j(\text{exact})$ is obtained as

$$p_j(\text{exact}) \approx \sum_{t(j)=1}^{l_j} \exp(-E_{t(j)}/k_B T) / \sum_{t=1}^l \exp(-E_t/k_B T) \quad (2)$$

where l_j is the number of conformations in the sample that pertain to j , and the equal sign is justified only for an infinite sample. Notice that this is a sampling with repetitions, i.e. a conformation is allowed to be selected any number of times. The mean energy of the bin, E_j , is not used here. We shall also be interested in the average exact energy of each bin $\bar{E}_j(\text{exact})$, given by

$$\bar{E}_j(\text{exact}) \approx \sum_{t(j)} E_{t(j)} \exp(-E_{t(j)}/k_B T) / \sum_{t(j)} \exp(-E_{t(j)}/k_B T) \quad (3)$$

where $t(j)$ runs over all the energies pertaining to bin j .

This procedure is only feasible for systems with very few degrees of freedom, where the probability for locating low-energy structures at random is not too low; for larger systems, one must resort to approximate procedures. We have devised two such methods, as outlined below.

II.2. Counting Method. n conformations are generated at random and energy minimized, as described above. Approximate probabilities $p_j(\text{count})$ are calculated as

$$p_j(\text{count}) \approx n_j \exp(-E_j/k_B T) / \sum_j n_j \exp(-E_j/k_B T) \quad (4)$$

where n_j is the number (with repetitions) of microstates in the sample that pertain to j ($n = \sum n_j$) and E_j is the mean energy value of bin j . Note that this procedure assigns the mean energy of the bin E_j , which is close to minimum energy value E_i^m , to all the members of microstate i . Therefore, at low enough temperatures (where differences in Boltzmann factors are substantial), this method is expected to overestimate the probability of the low energy bins.

II.3. Harmonic Approximation. An alternative approximate procedure has been devised. Conformations are selected at random and are energy minimized, as before. At each energy minimum one calculates S_i^{har} , the entropy in the harmonic approximation,³⁹ as

$$S_i^{\text{har}} = -\frac{k_B}{2} \ln[\text{Det}(\text{Hessian})] + S^{\text{har}}(T) \quad (5)$$

where Det stands for determinant and Hessian is the matrix of second derivatives of the energy with respect to the internal variables. $S^{\text{har}}(T)$ is an additive term that only depends on T ; it cancels out in the calculation of the probabilities and can therefore be omitted. The qualitative meaning of eq 5 is that wide minima (with small values of Det) will have less negative entropies than narrow minima (with large values of Det). The harmonic free energy is

$$F_i^{\text{har}} = E_i^m + Nk_B T/2 - TS_i^{\text{har}} \quad (6)$$

where N is the number of variables. The average harmonic energy \bar{E}_i^{har} of microstate i is

$$\bar{E}_i^{\text{har}} = E_i^m + Nk_B T/2 \quad (7)$$

From F_i^{har} one can obtain the harmonic partition function Z_i^{har} as

$$Z_i^{\text{har}} = \exp(-F_i^{\text{har}}/k_B T) \quad (8)$$

With this method, unlike the previous ones, the contribution of each microstate should be taken into consideration only once, which requires performing a structural comparison of all the minimized conformations. The probability in the harmonic approximation is

$$p_j(\text{har}) \approx \sum_{i=1}^{m_j} Z_{i(t)}^{\text{har}}(j) / \sum_{i=1}^m Z_{i(t)}^{\text{har}} \quad (9)$$

where m and m_j are the number of different microstates in the entire sample and in bin j , respectively. $Z_{i(t)}^{\text{har}}(j)$ is the harmonic partition function of microstate i that pertains to bin j and has been obtained in step t of the process. The total average harmonic energy $\bar{E}^{\text{har}}(\text{tot})$ is

$$\bar{E}^{\text{har}}(\text{tot}) = \sum_j p_j(\text{har}) \bar{E}_j^{\text{har}} \quad (10)$$

where $\bar{E}_j^{\text{har}} = E_j + Nk_B T/2$, and E_j is the mean energy of bin j .

The classical harmonic approximation is adequate at low enough temperatures (however, not close to the absolute zero, where quantum effects become important^{39a,40,41}), where most of the contribution to Z_i comes from the bottom region of the

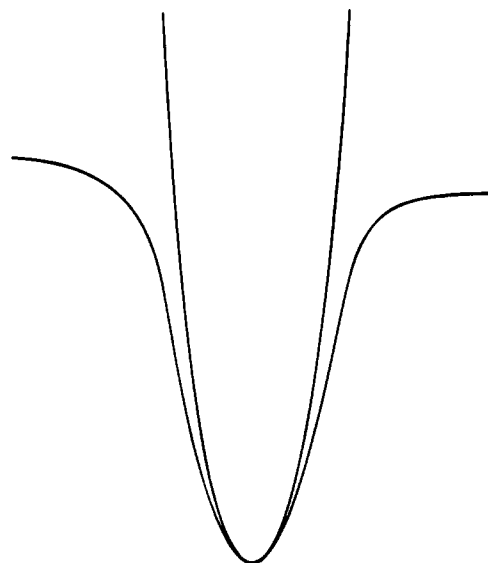


Figure 1. A typical one-dimensional potential well with broadened anharmonic branches, approximated by a parabola.

potential energy well. For a one-dimensional system, this region can be approximated by a parabola, as illustrated in Figure 1. At higher temperatures, contributions to Z_i also come from the higher energy branches of the true potential well, which typically have a broadened anharmonic shape (see Figure 1). It is thus expected that $D = Z_i - Z_i^{\text{har}} > 0$. Correspondingly, the exact entropy, and for strong anharmonicity, also the exact average energy, will be underestimated by their harmonic counterparts.⁴² In general, higher energy wells are expected to be shallower than low-energy ones (this is supported by harmonic entropy data presented in Table 3, to be discussed in section III.2). For a shallow well, significant contribution to Z_i^{har} [through $S^{\text{har}}(T)$ (eq 5) and $\bar{E}^{\text{har}}(T)$ (eq 7)] will arise at high temperatures from parts of the parabola which extend above the true well. This suggests that D is small and might even become negative, for high enough T . Therefore, the p_j^{har} values of the low (high) energy bins are expected to underestimate (overestimate) the exact probabilities.⁴³ In this respect the harmonic approximation and the counting method display opposite trends, implying that their average $p_j(\text{ave})$

$$p_j(\text{ave}) = [p_j(\text{count}) + p_j(\text{har})]/2 \quad (11)$$

is a better approximation than either $p_j(\text{count})$ or $p_j(\text{har})$.

III. Results and Discussion

The test case used in the development of our overall methodology is Leu-enkephalin, for which experimental NMR data, obtained from cryoprotective solutions at 280 K, are available in the literature.³⁷ We shall use this system to test the validity of our theoretical methods.

The peptide is modeled by ECEPP/2, which is a relatively simple force-field based on constant bond angles and lengths.³² The peptide bond angles ω are fixed at 180° and no explicit solvent is considered. Therefore a conformation is defined solely by the 10 backbone dihedral angles ϕ and ψ and the nine side chain dihedral angles χ . Energy minimizations are carried out with the Newton Raphson method (based on first and second analytical derivatives of the energy), using the program FANTOM.⁴⁴ The second derivatives of the energy are also used to calculate the harmonic entropy (eq 5). In order to obtain reliable results for the latter, the energy minima should be obtained with high precision. This we ensured by setting an absolute value of 10^{-8} (kcal/mol)/deg for the gradient as tolerance for stopping

the minimization and a Cholesky decomposition parameter $\delta = 10^{-4}$. The multiplication factor of the unit matrix that is added to a negative definite second derivative matrix was set equal to zero if it became negative during minimization. The parameter ρ for the first test in the line search was set equal to 0.2, and the parameter σ for the second test in the line search to 0.9. Finally, the factor τ , by which the step size is reduced in the line search after failure of the first test, was set equal to 0.1 (for further details see ref 44). The entropy was calculated only for true energy minima, i.e., where the Hessian was positive definite.

III.1. Results for the four-Variable Model. We first test the validity of the approximate procedures by conducting a comparison with the exact method. This, however, is only feasible for a small number of variables (dihedral angles). To generate a viable model, we have fixed 15 out of the 19 dihedral angles at the values of the global energy minimum (GEM) structure (defined in Table 1 of ref 23), allowing only the remaining four angles to vary. Two quartets were chosen: set 1 consisted of ψ , χ^1 , χ^2 , and χ^6 of Tyr¹ and set 2 of ψ of Gly³ and ϕ , χ^1 , and χ^2 of Phe.⁴ Because of the small number of variables, only three low-energy potential wells which practically determine the partition function are obtained. For the sake of convenience, we shall identify the corresponding bins by means of the energy minima of these potential wells.

In order to calculate $p_j(\text{exact})$ (eq 2), two samples of 17 and 15 million random conformations were generated for set 1. For the first sample only conformations with energy below 1 kcal/mol were taken into account. Each conformation was subjected to energy minimization, in order to identify the potential well E_i^m to which it belonged. A total of 576 190 conformations satisfied this energy criterion, with $\sim 91\,000$, $41\,000$, and $45\,000$ pertaining to bins 1, 2, and 3, respectively (the rest belonged to higher energy bins, which contributed negligibly to the partition function). For the second sample we used a lower energy cutoff of -3 kcal/mol. The number of conformations retained was 21 634, with $\sim 10\,000$, 4300 , and 5000 pertaining to bins 1, 2, and 3, respectively. We found that up to 320 K, the two samples generate practically identical results. Therefore we used the -3 kcal/mol cutoff to produce a sample of 7 million random conformations for the second set of 4 variables. A total of 9714 conformations, with a distribution of ~ 1800 , 4200 , and 3700 in bins 1, 2, and 3, respectively, were retained.

The results for $p_j(\text{count})$ (eq 4) were calculated from samples of $\sim 19\,700$ and 8000 randomly selected, and subsequently energy minimized, conformations obtained for sets 1 and 2, respectively. The distribution of conformations in bins 1, 2, and 3 was 2680, 1500, and 1192 for set 1 and 3050, 2940, and 2000 for set 2. These samples were also used to calculate $p_j(\text{har})$ (eq 9). Since the calculation of $p_j(\text{har})$ requires that the minimized conformations be structurally different, we have adopted a variance criterion whereby two structures are considered different if at least two corresponding dihedral angles differ by 1° or more.⁴⁵ Note that according to this criterion, two structures that differ by a symmetrical operation (e.g., rotation by 180° of χ^2 of Phe) are considered different.

Results obtained for the first set of four variables at several temperatures $150\text{ K} \leq T \leq 320\text{ K}$ are shown in Table 1 for the three bins defined by the three lowest energy minima of -9.7 (the GEM), -8.9 , and -7.2 kcal/mol. The corresponding probabilities p_1 , p_2 , and p_3 behave as expected (see section II). Thus, at all temperatures $p_1(\text{har}) < p_1(\text{exact}) < p_1(\text{count})$, while the "counting" and "harmonic" values of p_2 and p_3 of the higher energy bins satisfy the opposite relation. In all cases $p_j(\text{ave})$ (eq 11) provides a better approximation to $p_j(\text{exact})$ than both $p_j(\text{count})$ and $p_j(\text{har})$. As expected, p_1 dominates at all tem-

TABLE 1: Results at Several Temperatures T for the Exact Probability, $p_j(\text{exact})$ (eq 2) and the Approximate Probabilities $p_j(\text{har})$ (eq 9), $p_j(\text{count})$ (eq 4), and $p_j(\text{ave})$ (eq 11)^a

$T(\text{K})$	approx	p_1	p_2	p_3	\bar{E}_1	\bar{E}_2	\bar{E}_3	$\bar{E}(\text{tot})$
150	exact	0.93	0.07	0.00	-9.0	-8.2	-6.6	-8.9
	har.	0.91	0.09	0.00	-9.1	-8.3	-6.6	-9.0
	count	0.96	0.04	0.00				
	av	0.93	0.07	0.00				
240	exact	0.81	0.16	0.03	-8.4	-7.7	-6.3	-8.2
	har.	0.75	0.20	0.05	-8.7	-8.0	-6.3	-8.3
	count	0.89	0.11	0.00				
	av	0.82	0.15	0.03				
280	exact	0.76	0.19	0.05	-8.0	-7.4	-6.1	-7.8
	har.	0.67	0.22	0.10	-8.6	-7.8	-6.1	-8.1
	count	0.87	0.12	0.01				
	av	0.77	0.17	0.06				
320	exact	0.71	0.20	0.08	-7.7	-7.1	-5.9	-7.4
	har.	0.60	0.23	0.16	-8.4	-7.6	-6.0	-7.7
	count	0.84	0.15	0.01				
	av	0.72	0.19	0.09				

^a The calculations refer to the first set of 4 variables ψ , χ^1 , χ^2 , and χ^6 of Tyr¹ (the other 15 dihedral angles are kept fixed in the values of the global energy minimum — see ref 23). Only three energy minima contribute significantly to the probabilities; their values, $E_1 = -9.7$, $E_2 = -8.9$, and $E_3 = -7.2$, define the three corresponding bins (j). The average energies $\bar{E}_j(\text{exact})$ (eq 3) and $\bar{E}_j(\text{har})$ (eqs 7 and 10) for each bin, and the total average energies $\bar{E}(\text{tot})$ for the exact and harmonic (eq 10) calculations, are also provided. All the energies are in units of kcal/mol. The statistical errors in $p_j(\text{count})$ and $p_j(\text{exact})$ are not larger than ± 0.02 and ± 0.03 , respectively. The statistical errors in \bar{E}_j and $\bar{E}(\text{tot})$ are smaller than ± 0.06 kcal/mol. All the energy values are rounded to one decimal place.

peratures; it decreases as a function of increasing temperature, whereas p_2 and p_3 increase. Table 1 also features results for $\bar{E}_j(\text{har})$ (eq 7) and $\bar{E}_j(\text{exact})$ (eq 3), the average harmonic, and average exact energies of each bin. It can be seen that at 150 K, the harmonic energy is only slightly smaller than the exact energy (they are equal for $j = 3$). As expected, this implies that anharmonic effects are minimal. For $j = 1$ and 2, the anharmonicity becomes more pronounced at higher temperatures, where the harmonic energy is significantly smaller than the exact energy. For $j = 3$, however, the harmonic and exact energies are practically identical at all temperatures, suggesting that this potential well is nearly harmonic. On the whole, the total average harmonic energy of the three potential wells $\bar{E}^{\text{har}}(\text{tot})$ (eq 10) is smaller than the corresponding average exact energy (obtained by substituting $\bar{E}_j(\text{exact})$ (eq 3) for $\bar{E}_j(\text{har})$ in eq 10), and the difference between the two increases with increasing temperature.

Analogous results are presented in Table 2 for the second set of four variables. The three energy minima of -9.7 , -9.2 , and -9.0 kcal/mol are much more closely spaced than the corresponding energy minima of set 1. Therefore the relations among the probabilities p_j , predicted by the theoretical section II and borne out by the results shown in Table 1, are expected to manifest themselves in a less pronounced manner in Table 2. The harmonic and exact values of \bar{E}_j for the three potential wells are practically identical; $p_j(\text{har})$ and $p_j(\text{exact})$ are very similar at all the temperatures. This suggests that the three wells are approximately harmonic.

The expected inequalities between the various probabilities are satisfied for set 2, as they were for set 1. However, $p_j(\text{count})$ deviates from $p_j(\text{exact})$ at all temperatures, much more than $p_j(\text{har})$. Therefore, while $p_j(\text{ave})$ (eq 11) constitutes a better approximation to $p_j(\text{exact})$ than $p_j(\text{count})$, it is significantly inferior to $p_j(\text{har})$. This, however, is not expected to be the typical case, since in general the potential wells deviate from harmonicity.⁴⁶ Another reason for the unexpectedly large deviation of $p_j(\text{count})$ from $p_j(\text{exact})$ stems from the fact that

TABLE 2: As in Table 1, for the Second Set of Four Variables, ψ of Gly³, ϕ , χ^1 , and χ^2 of Phe⁴ ^a

<i>T</i> (K)	approx	<i>p</i> ₁	<i>p</i> ₂	<i>p</i> ₃	\bar{E}_1	\bar{E}_2	\bar{E}_3	$\bar{E}(\text{tot})$
150	exact	0.54	0.33	0.13	-9.1	-8.5	-8.4	-8.8
	harm.	0.47	0.37	0.16	-9.1	-8.6	-8.4	-8.6
	count	0.78	0.16	0.06				
	av	0.63	0.26	0.11				
240	exact	0.36	0.42	0.22	-8.7	-8.2	-8.0	-8.4
	harm.	0.31	0.45	0.24	-8.7	-8.3	-8.1	-8.2
	count	0.66	0.24	0.11				
	av	0.48	0.34	0.18				
280	exact	0.32	0.43	0.25	-8.6	-8.1	-7.9	-8.2
	harm.	0.28	0.46	0.26	-8.6	-8.1	-7.9	-8.2
	count	0.62	0.25	0.13				
	av	0.45	0.36	0.19				
320	exact	0.29	0.44	0.27	-8.4	-8.0	-7.7	-8.0
	harm.	0.25	0.46	0.28	-8.4	-7.9	-7.7	-8.0
	count	0.59	0.27	0.14				
	av	0.42	0.37	0.21				

^a The definition of the various parameters, and the statistical errors, appear in the footnote of Table 1.

(unlike the results for set 1) for $T \geq 240$ K $p_2(\text{exact})$ is significantly larger than $p_1(\text{exact})$. For example, at $T = 280$ K we find that $p_2(\text{exact}) = 0.43$ and $p_1(\text{exact}) = 0.32$.

It will be shown below that $p_j(\text{har})$ for the complete set of 19 variables, calculated in bins of 0.5 kcal/mol above the GEM, decreases monotonically up to 1.5 kcal/mol, from 0.25 for bin 1 to 0.04 for bin 3. Then it fluctuates slightly around 0.06 for the next nine bins (see Table 3 below). These observations and the fact that we are interested in $p_j^a(\text{ave})$, the accumulated (over increasing energy ranges above the GEM) values of the probabilities, suggest that, in general, $p_j^a(\text{ave})$ provides a better estimate for $p_j(\text{exact})$ than either $p_j^a(\text{har})$ or $p_j^a(\text{count})$.

III.2. Results for the 19-Variable Model. In the previous section we examined the properties of the approximate "counting" and "harmonic" methods by comparing them to the exact procedure, for limited sets of four variables. It was shown that the exact populations of the low-energy microstates are overestimated by $p_j(\text{count})$ and underestimated by $p_j(\text{har})$, whereas the opposite holds true for the higher energy microstates. For the typical case (set 1), the arithmetic average of the "counting" and "harmonic" populations was found to approximate closely the exact populations of the relevant microstates.

It is therefore justified to apply these approximations to an ECEPP model of Leu-enkephalin in which all the 19 variables are allowed to vary. Two large samples (1 and 2) of 6×10^5 and 6.8×10^5 random conformations were generated and their energies minimized. To calculate the probabilities, we considered 81 794 (11 235) conformations with minimized energy that did not exceed 1 (-3) kcal/mol for sample 1 (2). The generation of two independent samples enabled us to estimate the statistical error.

The results shown in Tables 3 and 4 were obtained with set 1 (1 kcal/mol cutoff). Practically identical results were obtained with set 2 (-3 kcal/mol cutoff) in those cases where microstates were sorted into bins according to their minimum energy (Table 3). The higher energy cutoff was necessary when microstates were sorted into bins according to their harmonic free energy (Table 4). The generation and minimization of 10^4 conformations required 7 h of CPU time on the IBM 375 RISC 6000 workstation.

Results obtained at 280 K are presented in Table 3 for different energy bins j of width 0.5 kcal/mol as they increase above the GEM (first column). The results for the probabilities p_j^a in the last three columns are for accumulated bins, defined in column 7.

In the second column are shown the values of n_j , the number of energy minima E_j^m (with repetitions) obtained for each bin,

which are used to calculate $p_j(\text{count})$ according to eq 4. The n_j values increase strongly with increasing energy, while the corresponding values of $p_j(\text{ave})$ decrease as the energy increases. This means that on the average the individual probability of a microstate pertaining to a higher energy bin will be very low. It should be noted that the GEM was obtained only twice; on the average ~60 000 minimizations are required to obtain an energy minimum that pertains to the first bin.

In order to calculate $p_j(\text{har})$ (eq 9), the energy minima of different conformations should be considered, i.e., no repetitions are allowed. The m_j values shown in column 3 were obtained using the variance criterion (defined previously) whereby two structures are considered different if at least two corresponding dihedral angles differ by 1° or more. As expected, they are smaller than the corresponding n_j values. The m_j values did not change upon increasing the variance criterion from 1° to 20°.

It is of interest to characterize the functional dependence of m_j on E_j . First, if the statistically more reliable data of the higher energies can be fitted to an analytic function, it will be possible to interpolate to the less reliable low-energy regime, which contributes most significantly to medium and long range NOEs. Such a fit will also enable estimating the error. Schaumann et al.⁴⁴ carried out an extensive study of energy minima for Met-enkephalin described by ECEPP and obtained a Gaussian-shaped histogram (Figure 5) for the distribution of the ~13 000 different energy minima found vs the energy. In the first region of this histogram m_j increases strongly from the GEM up to approximately 8 kcal/mol above the GEM; in the second region m_j levels off for ~3.5 kcal/mol and then decreases approximately to zero within the next 8.5 kcal/mol. We found a similar behavior for Leu-enkephalin in a preliminary study; however, the results in Table 3 pertain to the first region only. We fitted all the data and parts of them to exponential, Gaussian, and linear functions and obtained the best fits for the exponential function. However, these latter fits were unstable (i.e., adding or subtracting points always changed the results significantly), and therefore it was not possible to reliably interpolate to low energies or estimate the errors from these studies (the errors were determined in a different way; see below).

Columns 4–6, which depict $p_j(\text{count})$ (eq 4), $p_j(\text{har})$ (eq 9), and their average values $p_j(\text{ave})$ (eq 11), indicate that the lowest energy bins are the most populated ones. $p_j(\text{count})$ is significantly larger than $p_j(\text{har})$ for the first four bins, and the opposite relation prevails for bins 6 and above. This indicates that the decrease in $p_j(\text{count})$ with increasing energy is more pronounced than the decrease in $p_j(\text{har})$. Similar trends are illustrated by the accumulated probabilities.

The important result is that at 280 K the contribution to the partition function of the low-energy microstates is very significant. The microstates contained within 2 kcal/mol above the GEM contribute $p_j^a(\text{count}) = 0.75$ (expected to be an overestimation) and $p_j^a(\text{har}) = 0.44$ (expected to be an underestimation). Their arithmetic average $p_j^a(\text{ave})$ is 0.59. The accumulated probabilities within 4 kcal/mol are 0.95, 0.72, and 0.84; within 6 kcal/mol all have values above 0.9. Note that if one is interested in the accumulated probabilities at higher energies, it is appropriate to generate samples which are smaller than those used to obtain the results of Table 3. This is so since the values of p_j^a are high and the fluctuations in their increments are small. For example, we found with sample 2 that the p_j^a values for 2.5, 3, and 4 kcal/mol, obtained from samples of ~10 000, 5700, and 2000 minimized conformations, deviate by only 3% from the corresponding data in Table 3.

In Table 3 we also present the average harmonic entropy T_j^{har} of bin j , obtained by averaging the harmonic entropy at

TABLE 3: Results at 280 K for Leu-Enkephalin Described by the Entire Set of 19 Variables^a

bin(<i>j</i>)	<i>n_j</i>	<i>m_j</i>	<i>p_j</i> (count)	<i>p_j</i> (har)	<i>p_j</i> (ave)	$-T\bar{S}_i^{\text{har}}$	bin ^a	<i>p_j</i> ^a (count)	<i>p_j</i> ^a (har)	<i>p_j</i> ^a (ave)
0.0–0.5	11	10	0.40	0.25	0.33	17.9 ± 0.5	0.5	0.40	0.25	0.33
0.5–1.0	8	8	0.14	0.08	0.11	17.9 ± 0.3	1.0	0.54	0.34	0.44
1.0–1.5	15	14	0.09	0.04	0.07	18.2 ± 0.6	1.5	0.64	0.37	0.51
1.5–2.0	47	42	0.11	0.07	0.09	18.5 ± 1.0	2.0	0.75	0.44	0.59
2.0–2.5	78	73	0.08	0.11	0.09	18.2 ± 1.0	2.5	0.83	0.55	0.69
2.5–3.0	149	141	0.06	0.06	0.06	18.1 ± 0.9	3.0	0.89	0.61	0.75
3.0–3.5	238	226	0.04	0.05	0.05	17.8 ± 0.8	3.5	0.93	0.66	0.80
3.5–4.0	401	378	0.03	0.05	0.04	17.9 ± 0.9	4.0	0.95	0.72	0.84
4.0–4.5	660	630	0.02	0.06	0.04	17.6 ± 1.0	4.5	0.97	0.78	0.87
4.5–5.0	1007	957	0.01	0.05	0.03	17.4 ± 1.0	5.0	0.98	0.82	0.90
5.0–5.5	1546	1469	0.01	0.05	0.03	17.5 ± 1.2	5.5	0.99	0.87	0.93
5.5–6.0	2297	2175	0.00	0.05	0.03	17.2 ± 1.2	6.0	0.99	0.92	0.96
6.0–6.5	2802	2654	0.00	0.03	0.01	17.1 ± 1.2	6.5	1.00	0.95	0.98
6.5–7.0	3676	3512	0.00	0.02	0.01	17.1 ± 1.2	7.0	1.00	0.97	0.99
7.0–7.5	4974	4719	0.00	0.01	0.01	17.0 ± 1.2	7.5	1.00	0.98	0.99
7.5–8.0	6198	5871	0.00	0.01	0.00	16.8 ± 1.2	8.0	1.00	0.99	0.99

^a The results are given for bins *j* of 0.5 kcal/mol as the energy is increased above the global minimum (−9.706 kcal/mol). *n_j* is the number of energy-minimized conformations (with repetitions) found in bin *j*. *m_j* is the number of *different* energy-minimized conformations that pertain to *j*. *p_j*(count) (eq 4) and *p_j*(har) (eq 9) are approximate probabilities, and *p_j*(ave) (eq 11) is their average. In the last three columns are presented the accumulated values of the probabilities (denoted by “a”) for bins of increasing size (energy). The statistical errors decrease, because *n_j* and *m_j* increase, with increasing *j*. They are not larger than ±0.04 for the first 6 bins and decrease rapidly to ±0.01 for *j* > 6. Notice that these errors are not independent. Results at *T* = 280 K are also given for \bar{S}_i^{har} , the harmonic entropy of bin *j* (in kcal/mol) obtained by averaging over the individual harmonic entropies (first term of eq 5) of the *m_j* different microstates.

TABLE 4: Results at 280 K for Leu-Enkephalin Described by the Entire Set of *N* = 19 Variables^a

bin(<i>j</i>)	<i>g_j</i>	<i>p_j</i> (har)	<i>E_j</i> ^{min}	<i>E_j</i> ^{max}	bin ^a	<i>p_j</i> ^a (har)
0.0–0.5	9	0.27	−9.46	−7.22	0.5	0.27
0.5–1.0	10	0.12	−9.70	−8.10	1.0	0.39
1.0–1.5	26	0.10	−9.46	−6.72	1.5	0.49
1.5–2.0	49	0.09	−8.94	−5.97	2.0	0.59
2.0–2.5	174	0.11	−7.87	−3.17	2.5	0.70
2.5–3.0	343	0.10	−8.53	−3.30	3.0	0.79
3.0–3.5	695	0.08	−8.16	−2.32	3.5	0.87
3.5–4.0	1233	0.06	−8.04	−1.27	4.0	0.93
4.0–4.5	1714	0.03	−7.75	−0.01	4.5	0.96
4.5–5.0	2254	0.02	−7.18	0.39	5.0	0.98
5.0–5.5	3331	0.01	−6.91	0.90	5.5	0.99
5.5–6.0	4504	0.01	−6.83	0.99	6.0	0.99
6.0–6.5	6090	0.00	−6.16	1.00	6.5	1.00

^a They are presented for bins *j* of 0.5 kcal/mol, as the harmonic free energy (eq 6) increases above the global harmonic free energy minimum. *g_j* is the number of *different* microstates with harmonic free energy pertaining to bin *j*. *p_j*(har) (eq 9) is the harmonic probability. *E_j*^{min} and *E_j*^{max} are the lowest and highest minimized energy values (in kcal/mol) obtained for each harmonic free energy bin. In the last column are presented the accumulated values of the probabilities (denoted by “a”) for bins of increasing size (free energy). The statistical error in *p_j*(har) decreases, because *g_j* increases, with increasing *j*. It is not larger than ±0.04 for the first four bins and decreases rapidly to ±0.01 for *j* > 4. Notice that these errors are not independent.

280 K (first term of eq 5) of the *m_j* (different) microstates of bin *j*. It can be seen that \bar{S}_i^{har} increases from −18.5 to −16.8 cal/mol in going from the fourth to the last bin (the results for the first three bins are based on poor statistics). Although the observed 1.7 kcal/mol increase in \bar{S}_i^{har} is smaller than the corresponding 6 kcal/mol increase in energy, it demonstrates that on average the microstates associated with higher energy are also associated with larger harmonic entropy, and probably with larger exact entropy. It should be noted, however, that the fluctuations in the harmonic entropy for each bin are relatively large, which means that some microstates deviate from the average picture.

It would be of interest to examine the validity of the harmonic approximation for the present case of 19 variables. This would require carrying out several Metropolis MC³⁴ simulations starting from the minimum energy structures (“seeds”) and comparing the average energy and entropy calculated from the MC samples with their harmonic counterparts. However,

estimation of the entropy is not trivial and would require using procedures such as the quasiharmonic technique⁴⁰ or the local states (LS) method.^{35,36} We have carried out this investigation in a separate study (to be reported shortly³³), using the LS method. In this work we only compare the overall average harmonic energy $\bar{E}^{\text{har}}(\text{tot})$ (eq 10) to the overall average energy calculated by other simulation techniques.

Calculation of $\bar{E}^{\text{har}}(\text{tot})$ (eq 10) based on relatively narrow bins of 0.015 625 kcal/mol leads to -1.78 ± 0.03 kcal/mol, a value which, according to the corresponding result for set 1 of 4 variables, is expected to underestimate the exact overall average energy. We carried out four relatively long MC simulations³⁴ at 280 K, starting from the GEM structure (1), an energy minimum structure of −7.7 kcal/mol (2), and two random conformations (3 and 4). At each MC step, one out of 19 angles was picked at random, and a trial value was selected at random from a range of ±10° around the current value. The usual MC acceptance criterion was used. We have carried out 2 million MC steps (sweeps) per angle. The average energy of each run seemed to converge, but the energy values generated by the four simulations were found to be significantly different (−1.54 ± 0.04, −0.13 ± 1.0, 13.4 ± 0.3, and −1.47 ± 0.3 kcal/mol for simulations 1–4, respectively). These calculations are, therefore, not conclusive and indicate that each simulation spanned only a relatively limited region of conformational space around the initial structure, because of restrictive high energy barriers.

In view of these limitations, we also calculated the average energy with the multicanonical MC method of Berg and co-workers,²⁷ as implemented in ref 47. With this procedure the system is simulated with a probability distribution where (unlike the Boltzmann probability) all energies are equally probable. Therefore the system can cross energy barriers, and the entire energy range is visited quite effectively. The relative population of each energy, which is proportional to its degeneracy, is known, making possible the calculation of the entropy and the average energy.

A series of 20 simulations were carried out each of 600 000 MC (multicanonical) steps per dihedral angle. At each such step, one out of 19 angles was picked at random, and a trial value was selected at random (with equal probability) from the whole range [−180°, 180°], or from a narrow region of ±5° around the current value. In order to achieve efficient coverage

of the relevant energy range, the simulation was restricted to the region of $[-7.5, 20]$ kcal/mol. We found that during the simulation the system moved 280 times back and forth between these limits. The average energy was -0.23 ± 0.09 kcal/mol, which, as expected, is larger than the average harmonic energy. The ~ 1.55 kcal/mol difference between the exact and harmonic energies is larger than the analogous difference of ~ 0.4 obtained for the first set of four variables, suggesting that the extent of anharmonicity in the potential energy wells of the 19-variable model exceeds that of the 4-variable model. This supports the validity of the $p_j(\text{ave})$ (eq 11) approximation and constitutes some measure to its quality: the larger the anharmonicity, the better the approximation.

We have also examined the contribution of Z_i^{har} (eq 8) to $p_j(\text{har})$ as a function of the harmonic free energy F_i^{har} (eq 6), rather than the energy. The pertinent results, obtained at 280 K, are presented in Table 4 for bins of 0.5 kcal/mol above the lowest harmonic free energy value. We find that for the lower free energy bins, the $p_j(\text{har})$ values (second column Table 4) are larger than the corresponding values in Table 3. Hence, the p_j^a values in Table 4 are larger than their counterparts in Table 3. The dominance of the lower free energy bins is also reflected by the values of g_j (the number of conformations in the free energy bin j) in Table 4, which exceed the corresponding m_j values in Table 3.

The GEM structure appears in the second bin. Table 4 also depicts E_j^{max} and E_j^{min} , the maximum and minimum values of the energy (E_i^{m}) within each free energy bin. The highest energy of -5.97 kcal/mol found in the first four bins, which contribute 0.59 of the total population, is only 3.7 kcal/mol above the GEM. Since the harmonic approximation overestimates the free energy of the higher energy microstates, the E_j^{max} values for bins populated according to the exact free energy are expected to be even lower than -5.97 kcal/mol, implying that low free energy microstates are also relatively low-energy microstates.

These results suggest that the partition function Z can be expressed as the sum of two terms, Z_a and Z_b . Z_a contains the relatively small number of the lowest energy (and free energy) microstates and provides the dominant contribution to Z . The probability of each microstate is relatively large, and its contribution to the NOE intensities is significant. On the other hand, not only is Z_b smaller than Z_a , but the number of structures pertaining to Z_b is extremely large (see the strong increase in m_j and g_j with increasing energy, illustrated in Tables 3 and 4), which means that the probability of individual microstates pertaining to Z_b is small, and their (individual) contribution to the NOE intensities is negligible. Moreover, in view of their multiplicity, the conformations pertaining to Z_b necessarily display a much larger structural diversity than those pertaining to Z_a . Therefore it is possible to regard the Z_b conformations as an ensemble of random coil structures. As outlined previously, such an ensemble will contribute negligibly to long range NOEs, irrespective of the population $p_b = Z_b/(Z_a + Z_b)$. Although p_b does not appear explicitly in the theoretical formulation, it affects the comparison between theory and experiment through the relation $p_a = 1 - p_b$. Since p_a is "NOE-visible" whereas p_b is not, higher sensitivity of the NOESY experiment correlates with small values of p_b .

The $p_j(\text{ave})$ results obtained for the 19-variable model, and the considerations outlined above, indicate that in order to explain the experimental results, it is appropriate to consider only 2–3 kcal/mol above the GEM. This is a very important conclusion, recalling that minimized structures within energy ranges of 5–10 kcal/mol,^{21b,c,29–31} and at times extending up to 15 kcal/mol³¹ above the GEM, have been used previously in

the analysis of NMR data obtained from peptide solutions. In some cases, the size of the conformational set considered has been determined arbitrarily. The latter approach has been adopted in an earlier study of Leu-enkephalin.^{21d,e}

It should be pointed out that with ECEPP the number of energy minima is relatively large even within a limited range above the GEM, and therefore it is difficult to identify and take all of them into account. Many of these microstates are small (i.e., extend over a very limited region of conformational space) and often cluster into dense groups which comprise similar structures.^{38c} For example, a nearly exhaustive set of low-energy Met-enkephalin conformations was generated in a recent study using ECEPP.⁴⁵ It was found that the majority of conformations (74 out of 77) within 2.5 kcal/mol above the GEM display the GEM backbone motif.

In order to circumvent this problem, we have introduced the concept of an MC microstate, which will be presented in detail in a forthcoming paper.³³ Low-energy structures are identified by an exhaustive conformational search, and the complete set of *significantly* different structures within 2 kcal/mol above the GEM is determined. The members of this set are used as "seeds" for MC simulations.³⁴ The latter span relatively large regions of conformational space around the seeds. Such a region—called an MC microstate—consists of a large number of the microstates defined in this paper. The free energies of the MC microstates (hence, their populations) are calculated using the LS method. Comparison between the respective LS populations and their harmonic counterparts enables testing the validity of the harmonic approximation.

IV. Summary

We have proposed to view the conformational space as a collection of microstates, which are local conformational regions around energy minima. Methods for calculating the relative contribution of microstates to the partition function, as their minimum energy is increased above the GEM, have been developed.

These methods provide for the first time a global classification of the stability of different regions of conformational space. This is necessary for determining which microstates should be considered in detail in the next stage of our methodology.

For the ECEPP model of Leu-enkephalin used in this study, we found that the microstates contained within the relatively small ranges of 2 and 3 kcal/mol above the GEM (or above the lowest harmonic free energy) constitute 0.59 to 0.75 of the partition function, respectively. Since the individual populations of the abundant higher energy microstates are expected to be very low, and their structural diversity very large, the microstates residing within 2–3 kcal/mol should suffice for explaining the experimental results. In general, the energy (or harmonic free energy) cutoff will depend on the number of structurally different microstates and the available computer resources.

In the case of well-resolved NOE data, the present methods enable one to evaluate different models, with the model that leads to a lower cutoff energy being probably the more adequate. For such studies it should be sufficient to compare the corresponding values of $p_j(\text{count})$, which are easier to calculate than $p_j(\text{har})$. We are currently carrying out such calculations for Leu-enkephalin, using the program FANTOM,⁴⁴ which includes solvation energy terms based on the solvent accessible surface area.

Acknowledgment. We are grateful to Dr. Max Vásquez for valuable discussions. We thank Professor W. Braun for providing us with the program FANTOM and Dr. T. C. Oppe for his help in installing this program. We acknowledge support

from the Florida State University Supercomputer Computations Research Institute, which is partially funded by the US Department of Energy under contract number DE-FC05-85ER250000.

References and Notes

- (1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.
- (2) Clore, G. M.; Gronenborn, A. M. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 29–63.
- (3) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95–117.
- (4) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1.
- (5) Masefski, W., Jr.; Bolton, P. H. *J. Magn. Reson.* **1985**, *65*, 526–530.
- (6) Hruby, V. *Biopolymers* **1993**, *33*, 1073–1082.
- (7) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University: Ithaca, NY, 1953.
- (8) De Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University: Ithaca, NY, 1985.
- (9) Meirovitch, H. *Macromolecules* **1985**, *18*, 563–569.
- (10) Guttman, C. M.; McCrackin, F. L.; Han, C. C. *Macromolecules* **1982**, *15*, 1205–1207.
- (11) (a) Wright, P. E.; Dyson, H. J.; Lerner, R. A. *Biochemistry* **1988**, *27*, 7167–7175. (b) Shin, H.-C.; Merutka, G.; Waltho, J. P.; Wright, P. E.; Dyson, H. J. *Biochemistry* **1993**, *32*, 6348–6355. (c) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161–200. (d) Dyson, H. J.; Norrby, E.; Hoey, K.; Parks, D. E.; Lerner, R. A.; Wright, P. E. *Biochemistry* **1992**, *31*, 1458–1463. (e) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **1993**, *3*, 60–65. (f) Alexandrescu, A. T.; Exans, P. A.; Pitkeathly, M.; Baum, J.; Dobson, C. M. *Biochemistry* **1993**, *32*, 1707–1718. (g) Motta, A.; Tancredi, T.; Temussi, P. A. *FEBS Lett.* **1987**, *215*, 215–218.
- (12) (a) Stradley, S. J.; Rizo, J.; Bruch, M. D.; Stroup, A. N.; Gierasch, L. M. *Biopolymers* **1990**, *29*, 263–287. (b) Bruch, M. D.; Dhingra, M. M.; Gierasch, L. M. *Proteins* **1991**, *10*, 130–139.
- (13) Peishoff, C. E.; Ali, F. E.; Bean, J. W.; Calvo, R.; D'Ambrosio, C. A.; Eggleston, D. S.; Hwang, S. M.; Kline, T. P.; Koster, P. F.; Nichols, A.; Powers, D.; Romoff, T.; Samanen, J. M.; Stadel, J.; Vasko, J. A.; Kopple, K. D. *J. Med. Chem.* **1992**, *35*, 3962–3969.
- (14) Sautilis, J.; Mierke, D. F.; Byk, G.; Gilon, H.; Kessler, H. *J. Am. Chem. Soc.* **1992**, *114*, 4818–4827.
- (15) Young, J. K.; Hicks, R. P. *Biopolymers* **1994**, *34*, 611–623.
- (16) (a) Kessler, H.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W. F.; Berendsen, J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393–3396. (b) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994**, *116*, 1042–1049.
- (17) Kessler, H.; Matter, H.; Gemmeker, G.; Kottenhahn, M.; Bats, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 4805–4818.
- (18) Bogusky, M. J.; Naylor, A. M.; Pitzenberger, S. M.; Nutt, R. F.; Brady, S. F.; Colton, C. D.; Sisko, J. T.; Anderson, P. S.; Veber, D. F. *Int. J. Pept. Protein Res.* **1992**, *39*, 63–76.
- (19) Bruschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3–11.
- (20) Bruschweiler, R. *Structural Dynamics in Biomolecules Monitored by Nuclear Magnetic Resonance Relaxation*. Ph.D. Thesis, ETH, Zurich, 1991.
- (21) (a) Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 3399–3406. (b) Nikiforovich, G. V.; Hruby, V. J.; Prakash, O.; Gehrig, C. A. *Biopolymers* **1991**, *31*, 941–955. (c) Shenderovich, M. D.; Nikiforovich, G. V.; Sautilis, J. B.; Chipens, G. I. *Biophys. Chem.* **1988**, *31*, 163–173. (d) Vesterman, B.; Sautilis, J.; Betins, J.; Liepins, E.; Nikiforovich, G. V. *Biochem. Biophys. Acta* **1989**, *998*, 204–209. (e) Betins, J.; Nikiforovich, G. V.; Chipens, G. J. *Mol. Struct.* **1986**, *137*, 129–132.
- (22) Landis, C.; Allured, V. S. *J. Am. Chem. Soc.* **1991**, *113*, 9493–9499.
- (23) Meirovitch, H.; Meirovitch, E.; Michel, A. G.; Vázquez, M. *J. Phys. Chem.* **1994**, *98*, 6241–6243.
- (24) Vázquez, M.; Meirovitch, E.; Meirovitch, H. *J. Phys. Chem.* **1994**, *98*, 9380–9382.
- (25) (a) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615. (b) Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150–3152. (c) Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371–4378. (d) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
- (26) (a) Meirovitch, H. *J. Phys. A* **1982**, *15*, L735. (b) Meirovitch, H.; Vázquez, M.; Scheraga, H. A. *J. Chem. Phys.* **1990**, *92*, 1248–1257.
- (27) (a) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333–1338. (b) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249–253.
- (28) (a) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. *Tetrahedron Lett.* **1988**, *29*, 4373–4376. (b) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671–680.
- (29) Sumner, S. C. J.; Jiang, S.-P.; Jernigan, R. L.; Ferretti, J. A. *J. Biomol. Struct. Dyn.* **1992**, *10*, 429–439.
- (30) Peishoff, C. E.; Ali, F. A.; Bean, J. W.; Calvo, R.; D'Ambrosio, C. A.; Eggleston, D. S.; Hwang, S. M.; Kline, T. P.; Koster, P. F.; Nichols, A.; Powers, D.; Romoff, T.; Samanen, J. M.; Stadel, J.; Vasko, J. A.; Kopple, K. D. *J. Med. Chem.* **1992**, *35*, 3962–3969.
- (31) Peishoff, C. A.; Bean, J. W.; Kopple, K. D. *J. Am. Chem. Soc.* **1991**, *113*, 4416–4421.
- (32) (a) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361–2381. (b) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1833–1887. (c) Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231–6233.
- (33) Meirovitch, E.; Meirovitch, H. New Theoretical Methodology for Elucidating the Solution Structure of Peptides from NMR Data. II. Free Energy of the Dominant Microstates of Leu-Enkephalin and Ensemble Averaged NOE Intensities, submitted.
- (34) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (35) (a) Meirovitch, H. *Chem. Phys. Lett.* **1977**, *45*, 389–392. (b) Meirovitch, H. *Phys. Rev. A* **1985**, *32*, 3709–3715.
- (36) (a) Meirovitch, H.; Vázquez, M.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 651–671. (b) Meirovitch, H.; Kitson, D. H.; Hagler, A. T. *J. Am. Chem. Soc.* **1992**, *114*, 5386–5399. (c) Meirovitch, H.; Koerber, S. C.; Rivier, J.; Hagler, A. T. *Biopolymers* **1994**, *34*, 815–839.
- (37) Picone, D.; D'Ursi, A.; Motta, A.; Tancredi, T.; Temussi, P. A. *Eur. J. Biochem.* **1990**, *192*, 433–439.
- (38) (a) Stillinger, F. H.; Weber, T. A. *J. Phys. Chem.* **1983**, *87*, 2833–2840. (b) Stillinger, F. H.; Weber, T. R. *Science* **1984**, *225*, 983–989. (c) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318–321.
- (39) (a) Gö, N.; Scheraga, H. A. *J. Chem. Phys.* **1969**, *51*, 4751–4767. (b) Gibson, K. D.; Scheraga, H. A. *Physiol. Chem. Phys.* **1969**, *1*, 109–126. (c) Gö, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535–542.
- (40) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
- (41) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. J. *Am. Chem. Soc.* **1979**, *101*, 6842–6852.
- (42) If the anharmonicity of a potential well is increased, for example, by tilting outward the upper part of one of its branches, the free energy will decrease. Correspondingly both the entropy and average energy are expected to increase. On the other hand, the average energy in the harmonic approximation is constant for a given T ($Nk_B T/2$) irrespective of how steep the parabola is. Therefore, as anharmonicity becomes more pronounced, the exact average energy is expected to exceed its harmonic counterpart.
- (43) These relationships are further enhanced by our random selection of conformations (rather than microstates), which gives somewhat higher preference to high-energy shallow microstates, over their low-energy steeper counterparts. Note, however, that if all the microstates would have been identified (which never occurs in a practical simulation), this bias would disappear.
- (44) (a) Schaumann, T.; Braun, W.; Wüthrich, K. *Biopolymers* **1990**, *29*, 679–694. (b) von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1993**, *14*, 510–521. (c) von Freyberg, B.; Richmond, T. J.; Braun, W. *J. Mol. Biol.* **1993**, *233*, 275–292. (d) von Freyberg, B.; Schaumann, T.; Braun, W. *FANTOM User's Manual and Instructions*; ETH Zürich: Zürich, 1993.
- (45) von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1991**, *12*, 1065–1076.
- (46) (a) Mao, B.; Pear, M. R.; McCammon, J. A.; Northrup, S. H. *Biopolymers* **1982**, *21*, 1979–1989. (b) Levy, R. M.; Perahia, D.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 1346–1350. (c) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571–6575. (d) Gö, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696–3700. (e) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423–447. (f) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.
- (47) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211–214, 2353(E).

JP942620Q