JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Searching Peptide Conformational Space

Julie Grouleff and Frank Jensen*

Department of Chemistry, Aarhus University, DK-8000 Aarhus, Denmark

Supporting Information

ABSTRACT: We have performed a near complete analysis of the conformational space in terms of minima and transition structures for four small peptide models with a force field energy function. There is a clear trend that minima having a large difference in structure, as measured by the distance in torsional space, are rarely connected by a single transition structure. There is a similar trend that activation energies for conformational transitions correlate with structure differences, such that small conformational changes occur with low energy barriers and vice versa. This suggests that a systematic search for low energy conformational transition structures should focus on pairs of minima that are structurally similar. Eigenvectors from diagonalization of force constant matrices at minima are better at describing conformational transitions than vibrational normal modes, as verified both by overlaps with geometry difference vectors and results from biased molecular dynamics simulations.

INTRODUCTION

Parameterized force field energy functions are at the core of all biomolecular simulations and determine the ultimate accuracy of the results.[1] The purpose of a molecular dynamics (MD) or Monte Carlo simulation is to sample the phase space, which for biomolecules is determined primarily by the torsional degrees of freedom. The force fields' ability to accurately represent the conformational space near minima and connecting transition structures (TSs) is thus an essential component for providing accurate simulation results. Commonly used force fields like MM2,[2] MM3,[3] MMFF,[4] AMBER,[5,6] OPLS,[7] and CHARMM27[8] have been parametrized to reproduce the conformational minima for a set of small reference systems, but we envision that the next generation of force fields will require parametrization against a much larger training set of not only conformational minima but also connecting TSs.[9−15] It is experimentally difficult to obtain such reference data, but it is relatively easy to generate quite accurate results by using electronic structure calculations.

The conformational minima on a given potential energy surface (PES) can be located by a random or systematic generation of many trial structures followed by minimization,[16] but more sophisticated global search methods are also available.[17] A brute force approach will for small systems enable the location of all minima. However, locating all connecting TSs is more difficult, and it is virtually impossible to establish that all TSs have been found. Several groups have developed methods aimed at establishing transition networks corresponding to a tabulation of minima and connecting saddle points.[18] Saddle points are typically located by attempted uphill walking from a minimum[19] or by using a chain-of-state method for each pair of minima.[20] The question of whether a given pair of minima is connected by a first order saddle point is of general interest, as it is unlikely that all pairs of minima are connected by a TS, and the fraction of minima pairs connected by a TS will decrease with increasing system size. In the present study, we investigate the conformational space of four peptide models by locating all minima and a large fraction of all of the connecting TSs by using a combination of three TS search algorithms, of which one has the potential to locate all TSs. These data allow us to probe whether it is possible to deduce the existence of a TS connecting two conformations, using only information from the two minima, the efficiency of various algorithms for locating such TSs, and whether normal mode directions can be used to guide TS searches or bias MD simulations to achieve conformational transitions. For large systems, a complete enumeration of minima and saddle points is infeasible and unnecessary, and the focus is instead on an adequate sampling of the relevant phase space. Small systems with a well characterized energy surface, however, provide a platform for calibrating different methods for sampling the phase space.

All of the results have been generated using the OPLS force field, and other methods may generate a PES with a different topology, but the conclusions regarding connections between minima and TSs and associated properties should remain generally valid.

COMPUTATIONAL DETAILS

All force field calculations have been carried out using the Tinker[21] and Macromodel[22] programs and the OPLS[23] force fields using a dielectric constant of 1.0. Vibrational analyses have been performed in natural internal coordinates[24] using a locally modified version of the Gamess-US program package.[25] Internal coordinates offer a significant advantage over Cartesian coordinates for describing conformational transitions, as these primarily correspond to changes in torsional coordinates. Following a low-frequency normal mode in internal coordinates effectively only changes the torsional coordinates, while following the same normal mode in Cartesian coordinates also significantly changes bond lengths and angles.

The vibrational analysis corresponds to diagonalizing the force constant matrix in mass-weighted coordinates, which can be

Received: March 8, 2011
Published: May 17, 2011

either Cartesian or internal coordinates, and these eigenvectors are orthogonal. Transforming the eigenvectors from mass-weighted coordinates to non-mass-weighted coordinates, or from internal to Cartesian coordinates, destroys the orthogonality. In order to facilitate the data analysis, it is advantageous to reorthogonalize the transformed eigenvectors, which has been done by a frequency-weighted Löwdin procedure. In a traditional nonweighted Löwdin orthogonalization, a set of vectors is transformed by $\mathbf{S}^{-1/2}$, where $\mathbf{S}$ is the overlap matrix between the nonorthogonal vectors, and this ensures the least change of all of the vectors. As we are primarily interested in the low frequency modes, we have employed a frequency-weighted version[26] where the transformation matrix is given by $\mathbf{W}(\mathbf{WSW})^{-1/2}$, and $\mathbf{W}$ is a diagonal matrix containing the weights $w_i = \exp(-\upsilon_i/1000)$ based on the frequency υ_i (in cm^{-1}). This ensures that the low frequency modes are changed as little as possible by the orthogonalization. For glycine, a typical overlap between modes before and after orthogonalization is 0.7.

Transition structure optimizations have been done using three different methods: quadratic synchronous transit (QST), growing string (GS), and scaled hypersphere search (SHS). The QST method[27] searches for a saddle point between two minima by a sequence of minimizations and maximizations using a quadratic interpolation between the two end points until a low energy intermediate geometry is located.

The GS method[28] is a version of the nudged elastic band method[29] where the points on the string connecting the two minima are placed sequentially and optimized before additional points are added. Both the QST and GS methods provide only a guess for the TS, which is refined using a Newton−Raphson algorithm. The combination of a nudged elastic band and a Newton−Raphson-based algorithm for locating saddle points has been used extensively by Wales and co-workers.[30] The images along the initial reaction path are usually generated by interpolation between the two end-points,[31] and if used as a black-box method, it is only able to locate one connecting saddle point between each pair of minima.

The SHS[32] method searches for TSs by performing a series of constrained optimizations on hyperspheres with increasing radii and the center at a single minimum, where the initial search direction is generated by perturbing along Hessian eigenvectors. The SHS has been claimed to be able to locate essentially all TSs on a given PES,[33] although numerical issues may prevent this in practice. To our knowledge, the present work is the first application of the SHS method for locating conformational TSs on a force field energy surface. The nature of all minima and TSs has been confirmed by diagonalization of the Hessian matrix as having zero or one negative eigenvalue, respectively.

MD simulations have been done in Cartesian coordinates at a temperature of 310 K using the NAMD program.[34] The bias force is taken as a normal mode in internal coordinates, and since the conversion from internal to Cartesian coordinates is nonlinear, this necessitates a look-up table approach. Prior to a simulation, a table is constructed which links geometries along the bias normal mode in internal and Cartesian coordinates. At a given point in the simulation, the geometry change relative to the starting minimum along the bias normal mode is calculated in internal coordinates, and the bias force is the tangent direction in Cartesian coordinates, calculated as a finite difference from the tabulated values. The magnitude of the biasing force needs to be large enough for the bias to have effect but small enough that the temperature of the system does not increase significantly. This depends on the frequency of the given mode and thus has to be determined for each mode, which has been done by running test simulations with different force magnitude settings. In cases where the frequency is relatively high, the required magnitude of force is sufficiently large that the temperature of the system has to be damped additionally, which is done by increasing the Langevin damping coefficient for the system. The escape from the starting minimum is detected by collecting structures every 10th femtosecond and subsequently minimizing them in order to determine to which basin they belong. The first structure not equal to the starting minimum is the one listed in Table 3 along with the time for the transition.

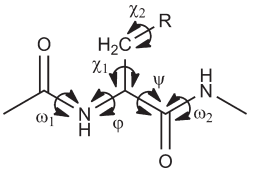

Figure 1. Notation for torsional angles.

Table 1. Calculated Relative Energies and Torsional Angles for Minima for Ace-Gly-NHMe

EQ*i*	class	E_{rel}(kJ/mol)	ω_1	φ	ψ	ω_2
1	TT	0.0	−175.8	124.9	−143.0	−179.9
2	TT	3.6	176.4	109.7	−60.4	179.6
3	TC	13.3	−174.6	127.3	−157.5	0.2
4	CT	14.0	0.3	115.7	−11.2	179.7
5	CT	22.6	4.9	126.6	−145.2	180.0
6	TT	26.1	−180.0	0.0	0.0	180.0
7	CC	34.8	5.5	128.2	−161.4	−0.5
8	CC	43.9	−1.0	−119.9	−78.7	−1.4
9	CC	64.0	12.6	28.3	65.4	0.9

RESULTS AND DISCUSSION

We have selected the glycine, alanine, serine, and cysteine amino acids as our trial set of systems, with side chains corresponding to H, CH_3, CH_2OH, and CH_2SH, respectively. Acetyl and N-methyl groups were added to the N- and C-termini to mimic the environment in longer peptide chains. The notation for the torsion angles is depicted in Figure 1. Conformations for peptides are commonly discussed in terms of a Ramachandran map with the backbone φ and ψ angles as the variables, where a positive sign indicates a clockwise rotation. The side chain and peptide bond torsional angles are labeled χ_i and ω_i, respectively.

Conformational minima have been located by an exhaustive Monte Carlo search. The OPLS PES for glycine contains nine minima, described in Table 1, as well as mirror images of eight of these. The minima are labeled according to energy such that EQ1 is the global minimum. On the basis of the configuration of the two peptide bonds, the minima can be separated into four groups. EQ1, EQ2, and EQ6 all have a *trans* configuration of both bonds and belong to a class labeled TT, and this is the configuration commonly found in peptides and proteins. The *trans,cis* class (TC) only consists of EQ3 while the *cis,trans* class (CT) contains EQ4 and EQ5. The structures with *cis,cis* configurations (CC) have the highest relative energies and include EQ7, EQ8, and EQ9. It should be noted that EQ9 is a very shallow minimum with a transition barrier to the mirror image of EQ8 of only 0.004 kJ/mol.

MP2 calculations for the TT class suggest that only EQ1 and EQ2 are genuine minima, while different force fields disagree on the existence of EQ6.[10] With the employed OPLS force field, EQ6 is a weakly defined minimum with a barrier of only 1.6 kJ/mol for transition to EQ2.

The three TS search algorithms located a total of 66 unique TSs on the PES for glycine. TSs corresponding to methyl rotations are not included in these 66 TSs and are not part of the analysis. Graphical representations of a number of two-dimensional projections of the PES were examined to verify the existence of the found TSs, and these plots allowed us to locate an additional six TSs. Four of these correspond to transitions where two different TSs connect the same two minima, corresponding to short and long reaction pathways depending on the direction of rotation around the torsional angles. The four manually found TSs all belong to the long pathways and in all cases have the three search-algorithm located TSs corresponding to the short pathway. An additional two TSs were constructed by mirroring two already located TSs. On the basis of our analyses of two-dimensional projections of the PES, we believe that the 74 located TSs (Table S1, Supporting Information) represent all of the first order saddle point on the PES, although this is difficult to prove rigorously.

The three search methods display different performances. The GS algorithm locates only 16 TSs. The QST finds 42 TSs, while SHS is capable of locating 49 TSs. That is, none of the methods are able to locate all TSs even when all minima are known. The energy barriers for the found transitions are in the range of 0.004−122 kJ/mol, and all three methods locate TSs corresponding to both low and high energy transitions. None of the methods are particularly better than the others at locating the low energy TSs, which usually are the most interesting. While the SHS method in principle should be able to locate all TSs,[33] the numerical issues regarding the detection of branching points is quite delicate, and the present work suggests that there is still room for improvements.

Out of the 136 possible TSs connecting the 17 minima (including mirror images, assuming only one TS exists between each pair of minima), we have located 50, which indicates that approximately only one-third of the possible TSs exist for this system. In 24 cases, there are two different TSs connecting the same two minima, and including these higher energy dual-TSs accounts for a total of 74 TSs. The corresponding values for only the nine symmetry unique minima are 36 possible TSs, of which we have located 17 unique. The QST and SHS algorithms locate 15 and 14 of these TSs, respectively, while GS only finds eight TSs.

The QST algorithm only optimizes a single structure and is the computationally cheapest method. The GS algorithm performs a sequential optimization of an increasing number of structures (up to 21 in the present cases) and requires approximately two orders of magnitude more computer time. The SHS algorithm requires second derivatives and performs a sequence of constrained optimizations along a set of normal modes and is, further, one order of magnitude more expensive computationally. For the present small systems and simple energy function, the computational times vary from less than 1 s to minutes for each saddle point search.

Wales and co-workers have advocated the combination of a nudged elastic band and a Newton−Raphson-based algorithm for locating saddle points.[30] In our case, the GS variant of the nudged elastic band algorithm is the method displaying the poorest performance, but other implementations and/or use of other coordinates may display better performance. Dual TSs, however, cannot be located by methods relying on interpolating between the two minima but can be located by the SHS algorithm.

Table 2. Calculated Relative Energies and Torsional Angles for Minima for Ace-Ala-NHMe

EQi	E_{rel}(kJ/mol)	class	ω_1	φ	ψ	ω_2
1	0.0	TT	−179.4	−79.4	61.6	179.9
2	4.1	TT	177.8	−152.0	158.3	179.8
3	10.4	TT	175.9	75.4	−48.9	179.8
4	12.8	TT	−177.8	−133.8	40.7	−179.6
5	16.1	CT	−2.8	−83.8	−6.1	179.8
6	18.8	TC	178.2	−151.1	151.7	−1.4
7	20.9	TC	175.5	−90.9	122.3	−3.2
8	21.3	CT	−1.2	−138.3	22.1	−179.5
9	25.1	CT	−1.6	−148.9	155.5	179.9
10	25.8	TT	176.1	−159.0	−49.6	178.9
11	28.0	CT	−2.5	82.3	6.7	−179.9
12	29.2	CT	−2.8	−88.8	137.0	−179.9
13	38.9	CC	−1.6	−148.5	150.3	−0.9
14	42.1	CC	−2.7	−88.5	141.4	0.1
15	44.0	CC	−2.0	−142.9	64.7	1.6
16	46.1	CT	−3.1	83.4	160.1	179.9
17	46.7	TC	176.9	−152.3	−44.8	−5.8
18	46.4	TC	179.9	−70.9	−33.2	−5.0
19	48.1	CC	−3.6	−71.5	−36.5	−3.5
20	51.2	TC	−178.3	98.8	−97.3	2.4
21	51.2	TC	−179.5	102.9	−67.4	−0.7
22	53.1	CC	−1.0	−145.5	−42.6	−5.3
23	54.2	TC	−174.6	93.8	162.4	−0.7
24	55.9	TC	178.0	71.8	43.4	4.8
25	56.7	CC	−1.3	74.6	52.1	2.7
26	61.2	CC	−1.8	85.2	143.8	−1.0
27	80.0	CC	−2.8	95.5	−121.0	1.0

In terms of potential energy, the 32 TSs having both the lowest absolute energies and the lowest activation energies all belong to transitions within each of the four classes, meaning that only the φ and ψ torsional angles change significantly. The remaining 42 TSs all correspond to transitions between classes. In the TT, CT, and CC classes, the lowest transition corresponds to a rotation dominated by the ψ angle. This indicates that the rotation barrier for this torsion is lower than for φ, ω_1, and ω_2, respectively. As expected, the data also show that rotation around the φ angle is softer than rotation around either of the peptide bonds. Nine pairs of minima (including mirror images) are connected by two different TSs, corresponding to rotation in the two possible directions. The shortest pathway in all cases has the lowest energy barrier. Not surprisingly, the three TS search methods are much better at locating the TS corresponding to the shortest pathway than the longer one.

The alanine system has 27 minima, shown in Table 2, and the conformational space is spanned by the same four torsional angles as for the glycine system. The PES for glycine contains seven well-defined minima (EQ1, EQ2, EQ3, EQ4, EQ5, EQ7, EQ8), and minima similar to all of these can be found on the PES for alanine. The introduction of the side chain methyl group

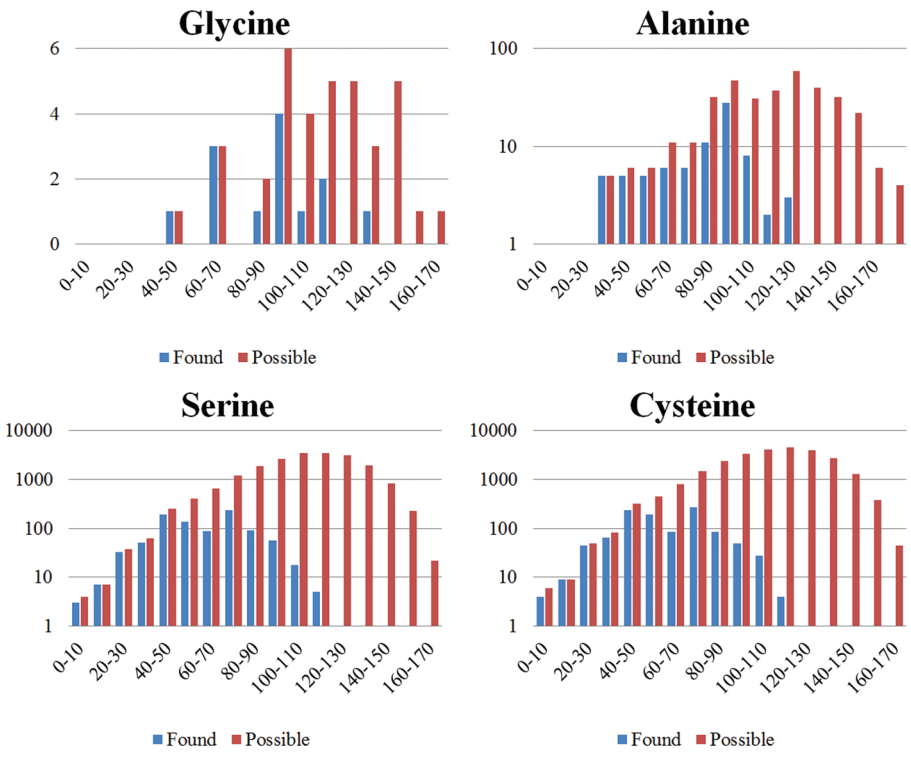

Figure 2. Comparison between the number of possible TSs and those actually located as a function of torsional RMS difference between the two connected minima.

causes some of the glycine energy minima to split up into two distinct minima for alanine. The presence of the methyl side chain introduces additional electrostatic and van der Waals interactions that results in a number of alanine minima which do not have a glycine equivalent.

MP2 calculations for the TT class suggest that there are seven minima,[10] while the presently employed OPLS force field has five minima. These five structures are in agreement with the OPLS free energy surface obtained by metadynamics simulations.[35]

The combination of the QST, GS, and SHS methods located a total of 97 TSs (excluding methyl group rotations) on the PES for alanine (Table S2, Supporting Information), of which 15 are dual TSs. On the basis of the results for the glycine system, we expect that there may be a few additional TSs which have not been found by the three automated methods, of which some represents additional dual TSs. Out of the 351 possible TSs connecting the 27 minima, the results thus suggest that only approximately one-fourth actually exist. In analogy with the glycine system, the 40 lowest energy TSs only connect conformations with a given subclass defined by two amide bonds. In terms of activation energy, the 44 lowest energy TSs correspond to transitions inside a given class, while the rest are transitions between subclasses. Out of the possible 10 TSs within the TT subclass, we have found seven.[36]

The serine system has 202 minima, and we have located a total of 1108 TSs, while the cysteine system has 227 minima, and we have found 1251 TSs (Tables S3−S6, Supporting Information). On the basis of the results for glycine, we expect that these represent a large fraction of all of the possible TSs, and the numbers can be compared to the combinatorially possible 20 301 and 25 651 TSs for the two systems, respectively; i.e., only ∼5% of the possible TSs actually exist. Within the TT subclass, the number of minima for the two systems is 26 and 35, which can be compared to 39 and 47 found at the MP2 level of theory.[10] Within this subclass, there are 325 and 595 possible TSs, respectively, and we have located 73 and 103.

For serine, the 497 lowest energy TSs all belong to transitions within each of the four classes, while the 671 TSs with lowest activation energies correspond to in-class transitions. The results for cysteine are similar: 584 lowest TSs if sorted by energy and 762 lowest TSs if sorted by activation energy. For both serine and cysteine, the TSs with a higher energy correspond to a mixture of transitions inside and between the classes.

Existence of Transition Structures. It seems intuitively reasonable that the probability of finding a TS between a pair of minima is higher for a pair of structurally similar minima than for a pair of structurally very different minima. To probe this hypothesis, we have calculated the torsional root-mean-square (RMS) deviation for all pairs of minima for all four systems and compared these values with the ratio between found and combinatorially possible TSs. In this analysis, the mirror images of the minima for glycine were not included. For glycine and alanine, the torsional RMS is calculated from the four backbone angles (ω_1, φ, ψ, ω_2). For serine and cysteine, the side chain torsions (χ_1, χ_2) are also included in the analyses. As can be seen from Figure 2 (note the logarithmic scale), there is a high possibility of finding a TS if the torsional RMS for the pair of minima is <50°. The likelihood of finding a TS between the two minima falls significantly if the torsional RMS is greater than 80°, and this is especially clear for the cysteine and serine systems.

Figure 3 shows the connection between the torsional RMS deviation between two minima and the activation energy calculated relative to the lowest energy minimum. As expected, there is a general correlation between the two, such that transitions

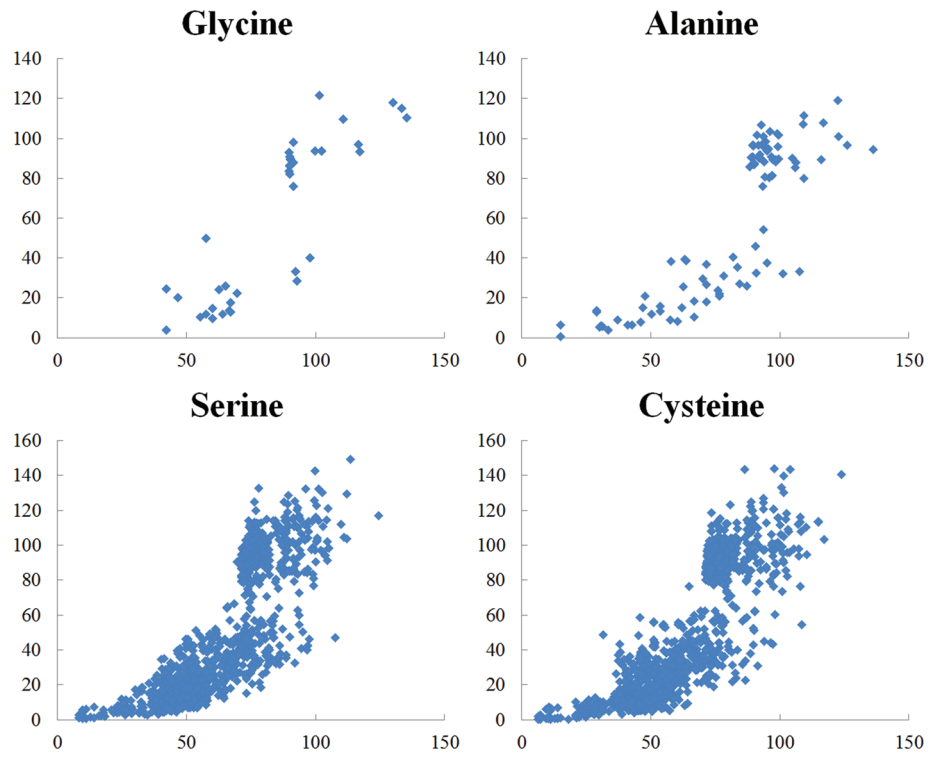

Figure 3. Correlation between activation energy (kJ/mol) and torsional RMS difference (degrees) between the two connecting minima.

involving large structural changes occur by TSs having relatively large activation energies.

The results shown in Figure 2 suggest that large structural changes rarely occur in a single step but rather as a sequence of transitions involving smaller changes. Figure 3 shows that transitions involving large structural changes that do occur in a single step tend to have high activation energies and are therefore less important from a dynamical point of view. The combination of these results suggests that a search for low energy conformational TSs can be focused on pairs of minima that are structurally similar. For the present systems, a torsional RMS less than ~80° would appear as a reasonable criterion, but such a threshold value is likely to be dependent on the system size.

Use of Normal Modes for Predicting Transition Structures. The eigenvectors from diagonalization of the force constant matrix contain information about the curvature of the PES, where the eigenvector associated with the lowest eigenvalue points in the direction where the energy increases least. If the force constant matrix is mass-weighted before diagonalization, the eigenvectors are the vibrational normal modes. It is commonly assumed that the direction of the low-frequency normal modes leads to low-energy TSs, and this is used for example in the eigenvector-following method for locating saddle points.[37] In the following, we will denote the eigenvectors from diagonalization of the force constant matrix as f-vectors and the eigenvectors from diagonalization of the mass-weighted force constant matrix as g-vectors.

While vibrational analysis traditionally is performed in Cartesian coordinates, it is for the present application more convenient to use internal coordinates[24] since the conformational space is effectively spanned only by the torsional coordinates. The f-vectors in internal coordinates are orthogonal, but this is not the case for the g-vectors. In order to facilitate the analysis, we have orthogonalized the g-vectors by a frequency-weighted Löwdin procedure, where the orthogonalization is carried out such that the low-frequency normal modes are perturbed as little as possible.[26]

The reaction path can be approximated either as a difference vector in internal coordinates between the two minima (ΔR_{EQ}) or as a difference vector between the minimum and the TS (ΔR_{TS}). For each system, we have calculated the overlap between these two reaction paths and the normal modes, both f- and g-vectors, for each of the two minima. The normal modes can be ordered either according to energy (frequency) or by the largest overlap with the reaction path. The average number of modes required to obtain an 80% overlap with the reaction path for the four test systems is given in Table 3. Transitions corresponding to methyl rotation have, as previously mentioned, been neglected, and all overlaps with modes corresponding to methyl rotation have been set to 0.

Table 3 shows that 5−6 of the lowest energy f-vectors are required to represent 80% of the reaction path, while only 2−3 vectors are required if the vectors are arranged according to overlap. This is understandable, as two vectors representing terminal methyl group rotations are among the low energy ones, and these make no contributions to the reaction path. The g-vectors, which are the vibrational normal modes, show the same trend, but significantly more modes are required to represent 80% of the reaction path. These results suggest that using the direction of f-vectors as a search bias will generally be more the efficient than using the g-vectors, at least for the present small systems.

To probe whether low energy modes point in the direction of low energy TSs, Figure 4 shows the correlation between the number of modes needed to obtain at least an 80% overlap with ΔR_{EQ} and the activation energy. The f-vectors provide a reasonable correlation,

Table 3. Average Number of Modes Required to Obtain at Least 80% Overlap with the Transition Vector ΔR_{EQ} or ΔR_{TS}

	ΔR_{EQ}				ΔR_{TS}			
	sorted by energy		sorted by overlap		sorted by overlap		sorted by energy	
	f	g	f	g	f	g	f	g
Gly	5.1	11.0	2.3	5.2	5.0	10.8	2.6	5.8
Ala	5.3	11.0	2.4	5.4	6.0	12.4	3.2	6.6
Ser	5.5	12.5	2.3	5.9	6.9	13.9	3.8	7.3
Cys	5.6	12.4	2.3	5.9	6.6	12.8	3.4	6.8

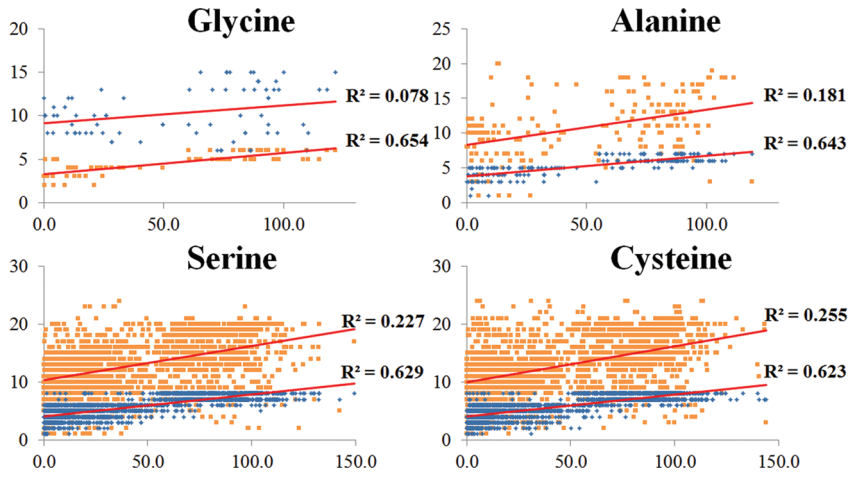

Figure 4. Correlation between the activation energy for a transition (kJ/mol) and the number of modes required to represent at least 80% of the reaction vector.

Table 4. MD Simulations Starting from EQ1 Biased with the Six Lowest Normal Modes in Form of f- and g-Vectors[a]

		f-vectors						g-vectors					
	no bias	1	2	3	4	5	6	1	2	3	4	5	6
EQ1*	5.9	6.4	6.3	43.5	68.6	5.5		41.0	5.6	18.7	7.3	6.7	7.1
EQ2	43.5	41.7	42.1	8.1	31.1	36.3	97.5	7.2	42.2	33.0	40.3	42.0	40.7
EQ4						7.7							
EQ5						48.4							
no change	51.1	51.9	51.6	48.3		1.3	2.5	52.0	52.2	48.2	52.4	52.0	52.2
time (ps)	4.8	4.8	4.9	4.3	1.4	1.0	1.2	4.3	4.9	4.9	4.6	4.7	4.8

[a] The numbers refer to the percentage of simulations where a conformational change from EQ1 to EQ*i* is observed. Transitions that occur in less than 1% of the simulations are not included. The time is the average time before a transition occurs. EQ1* indicates the mirror image of EQ1.

with correlation coefficients of ~0.64, while a much lower correlation is found for the g-vectors. This suggests that using non-mass-weighted normal modes (Hessian eigenvectors) may be more efficient than using vibrational normal modes to guide a search for low energy TSs.

Using Normal Modes to Bias MD Simulations. Isin et al. have recently used vibrational normal modes to bias MD simulations in order to simulate conformational transitions that otherwise would be too slow to be computationally feasible.[38] The correlation between geometry changes and normal modes in Table 3, and the corresponding correlation with activation energies in Figure 4, suggests that eigenvectors from diagonalization of the force constant matrix may have advantages over vibrational normal modes for biasing MD simulations for the present systems. In order to test this, we have performed a series of MD simulations on the glycine system with bias forces along the six lowest normal modes. Starting from the global minimum, we run 1000 simulations for 10 ps each and analyze the trajectories as described in the Computational Details to obtain statistics for which other minima the molecule escapes to during the simulation. The potential energy surface around EQ1 in terms of the φ and ψ angles is depicted in Figure 5, with the directions of f-vectors 3 and 4 and g-vectors 1 and 3 indicated.

The results of MD simulations starting from EQ1 can be seen in Table 4, where the percentage of changes to a given minimum is listed along with the average time for the transition. When no bias is applied, the majority of conformational changes is from EQ1 to EQ2, while a smaller fraction of the transitions are from

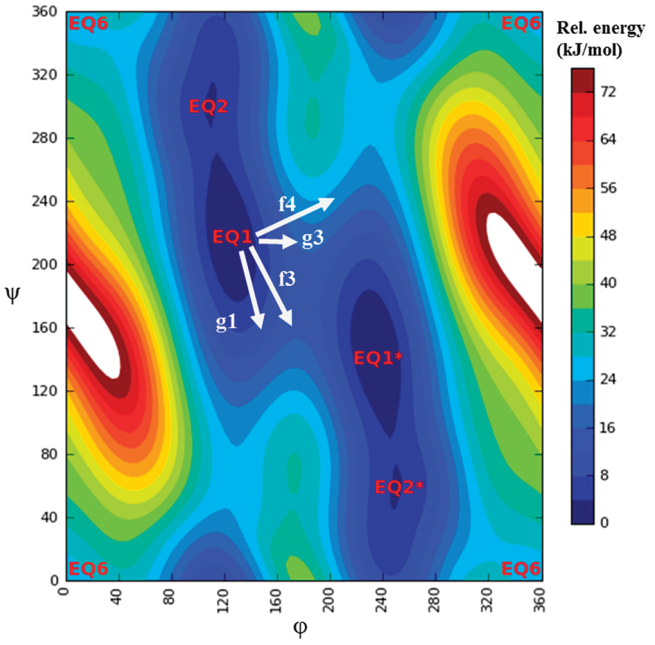

Figure 5. Two-dimensional cut of the PES for glycine showing the three minima (EQ1, EQ2, and EQ6) for the TT class. EQ1* and EQ2* denote mirror images of EQ1 and EQ2.

EQ1 to the mirror image of EQ1 (denoted as EQ1*). This is in line with activation energies of 3.9 and 13.5 kJ/mol for the two transitions (Table S1, Supporting Information). The use of either of the two f-vectors with lowest frequencies as a bias does not change this distribution significantly, which is not a surprising since these two vectors mainly correspond to the rotation of the terminal methyl groups. When f-vector 3 or 4 is used as a bias, the result is a better sampling of the EQ1 to EQ1* transition, whereas using f-vector 6 as a bias almost exclusively induces EQ1 to EQ2 transitions. The use of f-vector 5 leads to sampling of two new transitions, namely, EQ1 to EQ4 and EQ1 to EQ5, i.e. transitions corresponding to a rotation of one of the amide bonds. Using f-vectors 4, 5, and 6 as bias reduce the percentage of simulations where no conformational change is observed to $\leq$2.5%. This is also reflected in the average transition times shown in the last row of Table 4. Without an applied bias, the average time before a conformational change occurs is $\sim$5 ps, which is reduced to $\sim$1 ps when f-vectors 4, 5, or 6 are applied as a bias. The f-vectors 3, 4, and 5 all have reasonable overlaps (27−77%) with ΔR_{EQ} for the transitions that they enhance. For f-vector 6, however, the overlap with ΔR_{EQ} between EQ1 and EQ2 is almost 0%. The observed effect of using this vector as a bias is therefore most likely due to the increase in system energy. Of the six g-vectors used as a bias in this test, only the lowest normal mode leads to significant changes and increases the sampling of the EQ1 to EQ1* transition. None of the g-vectors reduce the percentage of simulations where no conformational change is observed significantly. These results thus support the conclusion from Figure 4 that f-vectors have advantages over g-vectors for biasing MD simulations for these systems.

SUMMARY

The present study has performed a near exhaustive search for conformational minima and transition structures for four small peptide systems using a force field energy function. Analysis of the results suggests that large structural changes rarely occur in a single step, and if they do, they are associated with a relatively large energy barrier. This indicates that a systematic search for low energy conformational transition structures should focus on pairs of minima that are structurally similar, for example, quantified by their torsional angles. None of the three employed algorithms are able to selectively locate low energy transition structures. The number of possible transition structures increases quadratically with the number of minima, but for the present systems, the actual relationship is close to linear, with typically 5 times as many transition structures as minima.

An analysis of overlaps with geometry vectors describing structural changes and molecular dynamics simulations indicates that eigenvectors obtained by diagonalization of the force constant matrix for minima are better at describing directions for low-energy conformational transitions than vibrational normal modes for these systems. Most of the conformational changes correspond to rotation around one or a few torsional angles, and the superiority of the force constant eigenvectors can be rationalized by the observation that these vectors primarily correspond to pure rotation around only one or two torsional angles. Vibrational normal modes, on the other hand, are spread out over more atoms and often involve both rotation and bending degrees of freedom. Large-scale conformational changes in proteins are often global domain movements, which typically require rotation around several torsions, and vibrational normal modes may here be better descriptors.

ASSOCIATED CONTENT

Supporting Information. Tables with structural and energetic information for minima and TSs for glycine, alanine, serine, and cysteine. The material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: frj@chem.au.dk.

ACKNOWLEDGMENT

This work was supported by grants from the Danish Natural Science Research Council, the Danish Center for Scientific Computing, and the Novo Scholarship program. We thank Prof. K. Ohno for help with the GRRM program used for the SHS optimizations.

REFERENCES

(1) Schlick, T. *Molecular modeling and simulation*; Springer: New York, 2002.
(2) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(3) Allinger, N. L.; Yan, L. *J. Am. Chem. Soc.* **1993**, *115*, 11918.
(4) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
(5) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(6) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(7) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(8) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86.

(9) Gundertofte, K.; Liljefors, T.; Norrby, P. O.; Pettersson, I. *J. Comput. Chem.* **1996**, *17*, 429. Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2004**, *126*, 698. Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; Mackerell, A. D., Jr. *Biophys. J.* **2006**, *90*, L36. Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, *27*, 781.

(10) Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774.

(11) Gresh, N.; Tiraboschi, G.; Salahub, D. R. *Biopolymers* **1998**, *45*, 405.

(12) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 9255.

(13) Möhle, K.; Hofmann, H.-J. *J. Mol. Struct.* **1995**, *339*, 57.

(14) Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U. *J. Comput. Chem.* **2003**, *24*, 161.

(15) Yang, Z. Z.; Zhang, Q. *J. Comput. Chem.* **2006**, *27*, 1.

(16) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379. Chandrasekhar, J.; Saunders, M.; Jorgensen, W. L. *J. Comput. Chem.* **2001**, *22*, 1646.

(17) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011.

(18) Noé, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154. Wales, D. J. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3. Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766. Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.

(19) Mortenson, P. N.; Wales, D. J. *J. Chem. Phys.* **2001**, *114*, 6443. Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363. Wei, G.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2002**, *117*, 11379. Wei, G.; Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2003**, *119*, 6403. Yun, M.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2007**, *126*, 105101. St-Pierre, J.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2008**, *128*, 045101.

(20) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *119*, 9947. Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *121*, 1080. Carr, J. M.; Wales, D. J. *J. Chem. Phys.* **2005**, *123*, 234901.

(21) Ponder, J. W. *Tinker*, version 5.1; Washington University School of Medicine: St. Louis, MO. http://dasher.wustl.edu/tinker/ (accessed 1/1/2010).

(22) *MacroModel*, version 9.7; Schrödinger, LLC: New York, 2009.

(23) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.

(24) Jensen, F.; Palmer, D. S. *J. Chem. Theory Comput.* **2011**, *7*, 223. Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. *J. Am. Chem. Soc.* **1992**, *114*, 8191.

(25) Schmidt, M. W.; Baldridge, K. K; Boatz, J. A; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S; Matsunaga, N.; Nguyen, K. A; Su, S; Windus, T. L; Dupuis, M; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(26) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

(27) The saddle option in the Tinker program is an implementation of the ideas in: Halgren, T. A.; Lipscomb, W. N. *Chem. Phys. Lett.* **1977**, *49*, 225. Bell, S.; Crighton, J. S. *J. Chem. Phys.* **1984**, *80*, 2464.

(28) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2004**, *120*, 7877.

(29) Henkelman, G; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978.

(30) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082. Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969.

(31) Bauer, M. S.; Strodel, B.; Fejer, S. N.; Koslover, E. F.; Wales, D. J. *J. Chem. Phys.* **2010**, *132*, 054101.

(32) Ohno, K.; Maeda, S. *Chem. Phys. Lett.* **2004**, *384*, 277. Maeda, S.; Ohno, K. *J. Phys. Chem. A.* **2005**, *109*, 5742. The SHS method as implemented in the GRRM program developed by K. Ohno and co-workers has been used.

(33) Ohno, K.; Maeda, S. *J. Phys. Chem. A* **2006**, *110*, 8933. Yang, X.; Maeda, S.; Ohno, K. *J. Phys. Chem. A* **2007**, *111*, 5099. Watanabe, Y.; Maeda, S.; Ohno, K. *Chem. Phys. Lett.* **2007**, *447*, 21.

(34) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

(35) Liu, Z.; Ensing, B.; Moore, P. B. *J. Chem. Theory Comput.* **2011**, *7*, 402.

(36) Westerberg and Floudas found 10 minima and 38 TSs for the TT-subclass with the ECEPP/3 force field: Westerberg, K. M.; Floudas, C. A. *J. Chem. Phys.* **1999**, *110*, 9259.

(37) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52. Culot, P.; Dive, G.; Nguyen, V. H.; Ghuysen, J. M. *Theor. Chem. Acc.* **1992**, *82*, 189. Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800. Baker, J. *J. Comput. Chem.* **1986**, *7*, 385. Baker, J. *J. Comput. Chem.* **1987**, *8*, 563. Baker, J. *J. Comput. Chem.* **1992**, *13*, 240.

(38) Isin, B.; Schulten, K.; Tajkhorshid, E.; Bahar, I. *Biophys. J.* **2008**, *95*, 789.