JCTC Journal of Chemical Theory and Computation

Comparison of Protocols for Calculation of Peptide Structures from Experimental NMR Data

Marc Fuhrmans,[†] Alexander G. Milbradt,[†] and Christian Renner*[,†,‡]

Max-Planck-Institut für Biochemie, Martinsried, Germany, and School of Biomolecular and Natural Sciences, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom

Received August 13, 2005

Abstract: In a comparison of structure calculation protocols we clearly demonstrate the need for generating independent starting structures, which is for peptides most efficiently achieved by distance geometry (DG) methods. Our test set consisted of 20 peptides with 7−9 amino acid residues additionally constrained by backbone cyclization and/or the presence of a disulfide bridge. Small peptides usually adopt defined conformational properties only upon introduction of additional constraints, such as cyclization. Therefore, we believe the results of our comparison to be applicable to a large and important class of molecules. The problems associated with the use of restrained molecular dynamics (MD) for conformational searching in the context of structure calculation consist in energy barriers that derive mainly but not exclusively from the experimental NOE constraints. A valid alternative to the DG approach, although for peptides computationally less efficient, is MD simulated annealing starting from random structures as commonly performed in the protein structure calculation from NMR data. As a consequence of our study it must be expected that a considerable fraction of published peptide structures are artificially well-defined or even wrong. Given the relevance of peptide studies for both drug development and protein folding we regard it highly important that structure calculations of peptides are performed with more consideration.

Introduction

In living organisms the blueprints for building proteins are stored in the form of amino acid sequences on genes. The translation of this primary structure into a folded and functional protein is effected by the inherent properties of the amino acids and the cellular environment including chaperones, folding adjuvants, and proper folding conditions. In this highly efficient way only a small amount of information needs to be stored (primary sequence) for generating macromolecules with oftentimes very complex structural and dynamic properties. Unfortunately, the translation rules are exceedingly complicated making protein folding one of the most challenging problems of biochemistry. For these reasons considerable experimental effort needs to be invested for determining three-dimensional structures of proteins although the primary sequence is already available or can be determined in a straightforward manner for a given gene of interest. Whereas X-ray crystallography dominates the structure determination of proteins,[1] for small peptides NMR spectroscopy is the method of choice because most peptides do not crystallize. Technically, peptide NMR is less complicated than the highly sophisticated multidimensional heteronuclear experiments used in protein NMR.[2] The differences in approaching small peptides or proteins by NMR are easily understood as peptides comprise a much smaller number of resonances, and therefore problems of overlap are of lesser concern. However, due to the generally higher flexibility of peptides and the smaller number of experimental constraints obtainable from the NMR spectra, conversion of NMR data to three-dimensional

* Corresponding author phone: +44-115-848-3522; fax: +44-115-848-6636; e-mail: christian.renner@ntu.ac.uk.

[†] Max-Planck-Institut für Biochemie.

[‡] Nottingham Trent University.

10.1021/ct050203j CCC: $33.50
Published on Web 12/06/2005

structures, i.e., structure calculation, is of higher concern for peptides.[3−10] For well-structured proteins the conformational space available to the molecule is so restricted as to leave less room for variations depending on the details of the structure calculation protocol. Contrarily, in peptides often multiple conformations occur, and their correct representation in the final structural ensemble might be sensitive to the calculation strategy.[4−6,10]

The two most widely used approaches to NMR conformational analysis of peptides are distance geometry (DG) based and pure molecular dynamics (MD) protocols, both used since the beginning of peptide NMR. For calculation of three-dimensional structures sufficient sampling of the conformational space is possibly the most important factor.[10] While MD simulations should in principle detect any possible conformation of the molecule according to the ergodicity theorem, the required time might exceed the current computational limits by far. Alternatively, more direct methods for sampling the conformational space can be used: different DG methods[11−16] and related approaches, e.g. refs 17−19, have been devised for structure calculation based on experimental distance constraints as obtained by NMR. Again good sampling properties are vital for the performance of these strategies.[14,15,20−22] Despite the existing discussion in the literature about conformational sampling, pure MD protocols are frequently used in published NMR studies of peptides. The applicability of the specific protocol and the possibility of incomplete sampling of the conformational space are usually not addressed. Comparisons of structure calculation methods have been reported before,[3,13,23−25] but only for one or two molecules in each case. We want to demonstrate the shortcomings of the simple MD method on a larger set of molecules. Furthermore, we compare for our set of peptides the results from the typical peptide protocols to those obtained with the structure calculation protocols that are commonly used for proteins.

Methods

General. Distance geometry[12] and molecular dynamics-simulated annealing (MD-SA)[26] calculations were performed with the INSIGHTII (version 2000) software package (Accelrys, San Diego, CA) on Silicon Graphics O2 R5000 computers (SGI, Mountain View, CA). In each calculation 100 structures were generated either by distance geometry, by assigning random values to the coordinates, or by a molecular dynamics run of 1 ns, where one structure was saved each 10 ps. In all cases the 100 structures were refined with a short MD-SA protocol: After an initial minimization, 5 ps at 300 K were simulated followed by exponential cooling to ~0 K during 10 ps. The refinement for the random structures included an additional 2 ps at 500 K prior to the 5 ps at 300 K. The cooling phase was reduced to 8 ps in this case resulting in the same overall length of 15 ps for the refinement step. This modification was introduced, because releasing residual strain was found to be more difficult for the structures, which were derived from random coordinates. The final structures were sorted according to their final energy, and the 20 energy-lowest were analyzed. A time step of 1 fs was used with the CVFF force field[27] while simulating the solvents DMSO and H_2O with dielectric constants of 46.7 and 80.0, respectively. For some examples additional calculations were performed in an identical manner but using the AMBER force field.[28] The experimental constraints were applied at every stage of the calculations with the same force constants as for the published structures.[29−32]

Distance Geometry protocol ("DG/MD"). One hundred structures were generated from distance-bound matrices.[12] Triangle-bound smoothing and prospective metrization were used. The structures were generated in four dimensions, then reduced to three dimensions, and optimized with a simulated annealing step according to the standard protocol of the DG II package of INSIGHT II. DG calculations generally result in poor covalent geometry (bond length, angles, etc). Therefore, in addition to the coarse optimization of the standard protocol a subsequent MD-SA refinement with DISCOVER was performed (see section *General* above).

Simple MD Protocol ("Pure MD"). All molecular dynamics calculations were performed with the DISCOVER module of INSIGHTII. As the starting point for the generation of 100 structures the energy-lowest structure of the published NMR ensemble[29−32] was used for each molecule. Velocities for this starting structure were generated at 10 K, and the system was then heated to 1000 K during 50 ps (temperature bath coupling with a 5 ps time constant). During the following 1 ns production run at 1000 K one structure was saved each 10 ps for further refinement (see above). In many cases an additional structure calculation was performed with a second starting structure corresponding to a low energy structure of the published ensemble that was conformationally dissimilar to the first starting structure and was not sampled in the first MD run.

MD Protocol with Reduced Force Field during the Conformational Sampling ("Scaled MD"). The only difference to the previous protocol (pure MD) was that during the conformational sampling at 1000 K nonbonded interactions (van der Waals and Coulomb) were reduced to 10%, while the through-bond interactions were scaled down to 50%. For the annealing step the force field was applied with full strength.

MD Protocol Starting from Random Coordinates ("Random MD"). Starting from random coordinates the first step consisted in achieving approximately reasonable covalent geometry: van der Waals and electrostatic interactions were scaled down to 1%, while the force constants for bond lengths, angles, and dihedrals were only reduced to 50%. After minimization nonbonded interactions were scaled up to 10% (as for the scaled MD). Another minimization followed before 10 ps could be simulated at 1000 K. Finally, the force field was restored to its normal strength for the simulated annealing step.

Results

Multiple structure calculations were performed for 20 peptides (Figure 1) that have been investigated recently in our laboratory (refs 29−32 and unpublished results) according to different structure calculation protocols (Figure 2). Most of the 20 peptides are similar in that they contain a peptide

Cyclic backbone with *cis/trans* azo group

Disulfide bridge		No disulfide bridge	
c[APB-ACATCDGF]	(1 c/t)	c[AMPB-KARGDfV]	(7 c/t)
c[AMPB-ACATCDGF]	(2 c/t)	c[APB-ACATCDGF]	(8 c/t)
c[AMPB-KCATCDKK]	(3 c/t)	StBu StBu	
c[AMPB-KCGHCKKK]	(4 c/t)	c[AMPB-ACATCDGF]	(9 c/t)
c[APB-ACATCDGFF]	(5 t)	StBu StBu	
c[AMPB-KCATCDKKK]	(6 t)	c[AMPB-KSATSDKK]	(10 c/t)

Linear

Disulfide bridge		No disulfide bridge	
ACATCDGF	(11)	GWGQPHGG	(12)

Figure 1. Peptides for which NMR data were determined previously in our group. Azobenzene containing peptides (**1**−**10**) can occur in two isomeric forms cis and trans of the azo moiety indicated by (c/t). Whereas trans is the ground state, also the cis isomer has a lifetime long enough to perform NMR experiments. Because the geometry of trans and cis azobenzene is completely different, the two isomers are treated as separate molecules (e.g. **1 cis** and **1 trans**) for the purpose of the present work. For **5** and **6** the structures were only determined for the trans isomer. Note: f = d-Phe, APB = 4-(amino)phenylazobenzoic acid, AMPB = 4-(aminomethyl)-phenylazobenzoic acid, tBu = *tert*-butyl.

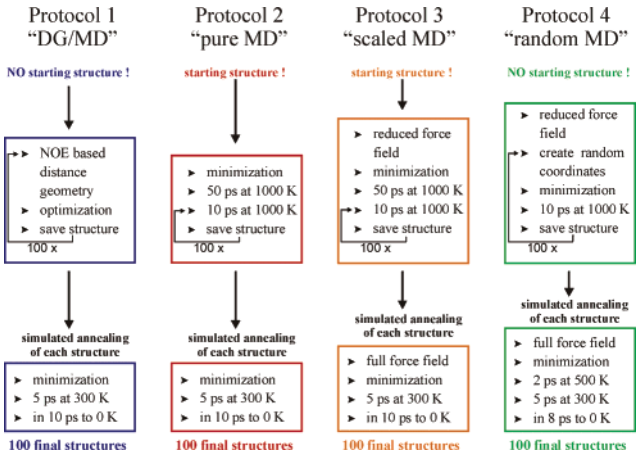

Figure 2. Flowcharts for the different protocols that were used for calculating structural ensembles using NMR data. The final structures were sorted according to energy and the 20 energy-lowest were analyzed.

stretch of 7−9 amino acid residues backbone-cyclized by 4-(amino)phenylazobenzoic acid or 4-(aminomethyl)phenylazobenzoic acid. This similarity allows for the systematic investigation of the influence of ring size and additional cyclization (i, i+3 disulfide bridge). For reference a peptide sequence with the i, i+3 disulfide bridge (**11**), but not backbone-cyclized, as well as a linear unconstrained octapeptide (**12**) were used. Figure 2 shows that each protocol consists of a first part for generating structures and a second part consisting of a simulated annealing step. The annealing step is identical for all protocols except for the fourth where a short 500 K phase was inserted before the simulation at 300 K. To achieve the same overall length the final cooling period is shortened correspondingly. In the first part the

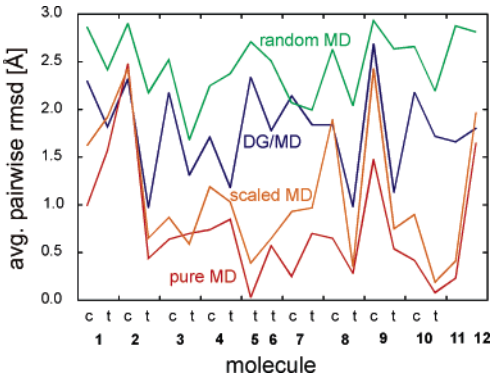

Figure 3. Conformational variability among the 20 energy-lowest structures obtained with different protocols for each molecule. Only backbone atoms were used for calculation of the rmsd.

structure generation consists of 10 ps at 1000 K per structure for protocols 2−4 that do not use distance geometry. In this way the computational costs for protocols 2−4 are nearly identical, and variations between protocols are kept to a minimum. Note that protocol 1 is the one that was used for generating the published structures.[29−32] Our protocols do not necessarily represent the most common or optimized versions of the respective approach. Instead they were chosen in such a fashion that they are as similar as possible to allow for interpretation of differences in the resulting structural ensembles. Specifically, we have placed an emphasis on exploration of conformational space represented by the first part of our protocols. Published MD protocols often employ multiple annealing cycles of a few picoseconds each for improved sampling and location of energy minima. While we used only one final annealing step for location of the energy minima, our conformational search comprised 1000 ps at 1000 K and thus clearly surpasses common MD strategies in terms of conformational sampling. With each protocol 100 structures were calculated for each of the 20 peptides. The 20 energy-lowest of each structural ensemble were analyzed with respect to NOE violations, final energy, and conformational variability as expressed by the average pairwise rmsd. The average pairwise rmsd is calculated as the average of the rmsd values of all possible pairs i and j with $i \neq j$ being two structures of the respective structure ensemble. For comparing two ensembles the pairwise rmsd can be calculated with pairs i and j, where i is from one ensemble and j is from the other. All pairwise rmsds can be displayed in so-called cluster graphs that are 3D diagrams where x and y define the number of the structures i and j and the corresponding rmsd(x, y) constitutes the third dimension. Analysis of cluster graphs calculated for two concatenated structural ensembles allows for detailed comparison of both. We consider two structures distinctly different, if their rmsd is greater than 2 Å. This limit was found useful for defining and differentiating conformational families. The average pairwise rmsd for one ensemble results in slightly larger values as the more familiar average rmsd to the average structure. However, the latter is not as easily generalized to a comparison of two ensembles.

Figure 3 compares the conformational homogeneity or heterogeneity of structural ensembles obtained using the

Table 1. Comparison of DG/MD and Pure MD Approach for 20 Peptides[c]

peptide [a]	homogeneous ensemble			DG ≈ MD	MD ≠ MD (1st) (2nd) energy barrier
	DG	MD	MD [b] (2nd)		
1 *cis*		X	X		
1 *trans*				X	
2 *cis*				X	
2 *trans*	X	X		X	
3 *cis*		X	X		
3 *trans*	X	X		X	
4 *cis*		X	X		X
4 *trans*	X	X		X	
5 *trans*		X	X		X
6 *trans*		X	X		X
7 *cis*		X	X		X
7 *trans*		X	X		X
8 *cis*		X	X		X
8 *trans*	(X)	X	X	(X)	
9 *cis*					X
9 *trans*	X	X		X	
10 *cis*		X	X		
10 *trans*		X	X		
11		X	X		X
12	X	X		X	

[a] Color-coding of peptides: green: DG and MD result in similar ensembles; red: energy barrier in MD simulation; black: no energy barrier, but MD ensemble incomplete. [b] Second MD from a second starting structure that is contained in the DG ensemble but not sampled in the first MD calculation. [c] Peptides **1**−**6** (bicyclic) and peptides **7**−**10** (monocyclic) are sorted according to increasing ring size.

protocols of Figure 2 for all 20 peptides. Obviously, generation of structures by molecular dynamics results mostly in more homogeneous ensembles than the calculation of independent structures either by DG or random starting coordinates. This result is expected, because sampling of conformational space might be impeded in molecular dynamics by high energy barriers (protocols 2 and 3), while protocols 1 and 4 directly start from distinct points in conformational space and, thus, circumvent the problem of high energy barriers (but not that of low barriers and general roughness of the energy surface, as will be seen below). The fact that our DG ensembles capture a larger part of the accessible conformational space compared to the MD ensembles, however, does not prove or even indicate that sampling of the DGII method as implemented in the INSIGHT2 software is ideal or complete (see refs 20−22 for a comparison of various DG methods). A detailed comparison of the structural ensembles of DG/MD and pure MD reveals that in cases when the DG ensemble consists of only one conformational family the same family is also found with the pure MD protocol albeit with partially lower rmsds. However, when more than one conformational family is present in the DG ensemble, often the pure MD reproduces only one of them. In these cases an additional MD calculation (according to protocol 2) was performed starting from a conformation that was present in the DG ensemble but not sampled in the first MD run. For this purpose the DG structure was compared to all 100 structures of the first MD ensemble, not only the 20 energy-lowest. Table 1 summarizes the results of the additional MD calculations. For almost half of the peptides well defined, but dissimilar ensembles were obtained with the same calculation protocol (#2), documenting a strong dependence on the starting structure. In Table 1 peptides are sorted according to constraints imposed by cyclization with bicyclic peptides and small ring sizes at the top and less constrained peptides at the bottom. Inspection of the last column of Table 1 reveals the occurrence of an energy barrier with little correlation to ring size or presence of the additionally constraining disulfide bridge. Therefore, it seems not possible to anticipate the presence or absence of energy barriers in MD structure calculation based on the given chemical structure of the peptide. Although the presence of energy barriers should always be considered, the severity of this problem even at a simulation temperature of 1000 K might surprise those who have so far relied on a variation of the pure MD protocol. Figure 4 exemplifies that apparent structural similarity is also no clue to the presence of energy barriers: Four peptides are shown for which protocol #2 (pure MD) was performed with two starting structures that both belong to the published NMR structural ensemble. While the two starting structures are quite similar for peptides **7 cis** and **8 trans**, those of **7 trans** and **10 trans** exhibit pronounced differences. The resulting ensembles were homogeneous and reasonably well-defined (see Table 1, structures not shown). Structure calculations have to be independent of the starting structure so that the same final ensembles should be obtained for both starting structures. While this was indeed observed for **8 trans** and **10 trans**, shown on the right side in Figure 4, for **7 trans** and **7 cis** the result depended strongly on the starting structure. For the four peptides of Figure 4 obviously no correlation exists between the apparent similarity of the two starting structures and the presence of a considerable energy barrier. For peptide **10 trans** the second MD calculation resulted in basically the same ensemble as the first MD, although the starting structure was different. In these cases the question remains, whether the second starting structure that is contained in the DG, but not in the MD ensemble is a realistic conformation of the peptide or not. We think that every conformation that satisfies the experimental constraints and is compatible with the force field (i.e. low energy) has to be considered as realistic.

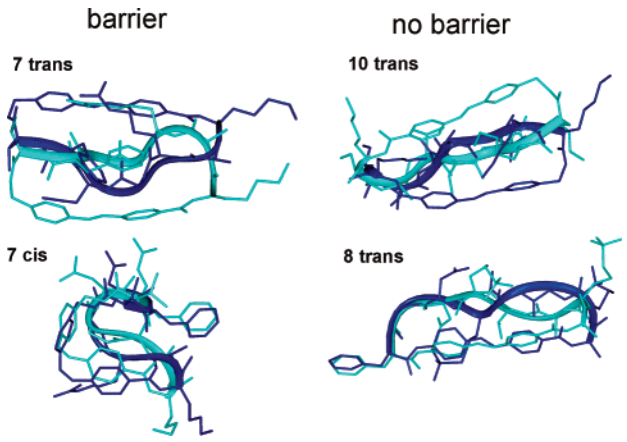

Figure 4. Comparison of the starting structures of two different MD calculations for selected peptides (**7 trans, 10 trans**: starting structures with pronounced difference; **7 cis, 8 trans**: similar starting structures). For peptides on the right very similar results were obtained from the MD simulations, whereas for peptides on the left results depended heavily on the starting structure pointing to the existence of an energy barrier.

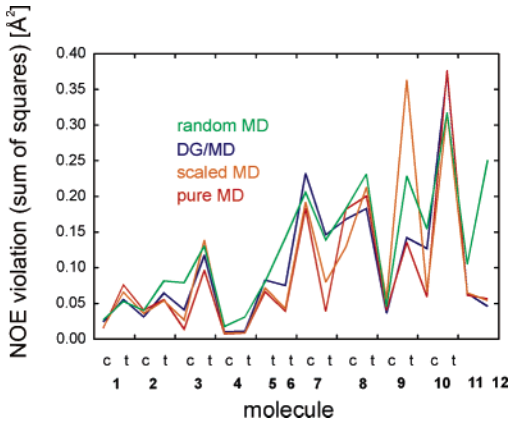

Figure 5. Squared NOE violations were summed up for each structure and averaged over the 20 energy-lowest structures. Results obtained with different protocols are compared for each molecule.

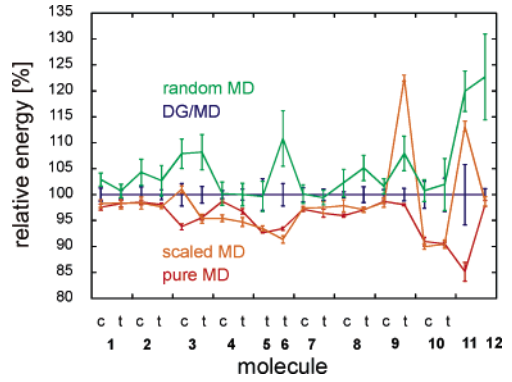

Figure 6. Average energy of the 20 energy-lowest structures relative to the average energy obtained with the DG/MD protocol. Error bars indicate the relative standard deviation, i.e., standard deviation divided by average.

NOE violations are compared in Figure 5. For this graph the squared NOE violations were summed up for each structure, and the resulting sums were averaged over the 20 energy-lowest structures of an ensemble. Multiplication of the sums of squared NOE violations with a force constant of 50 kcal $Å^{-2}$ would yield the corresponding energy penalty. It is seen from Figure 5 that experimental constraints were fulfilled in the different ensembles to similar extents. However, for the final energies clear trends are visible in Figure 6. As the absolute values of the energies have no real meaning, final energies relative to those obtained with protocol 1 are depicted. Clearly, protocol 4 (random MD) results in the most unfavorable energies, whereas the pure MD ensemble exhibits the smallest values. This comparison indicates that the structures of the random MD ensemble and, also to some extent, those of the DG ensemble are not fully relaxed. For the DG ensemble we tested in additional calculations that for most peptides more extensive simulated annealing can indeed reduce the final energies to values close to those of the corresponding pure MD ensemble (data not shown). While sufficiently long or iterated steps of simulated annealing should result in very similar final energies for all four protocols, we have purposely chosen annealing steps of only moderate length for comparing the efficiency of the first step of the protocols where structures are generated (DG vs MD vs random coordinates). We consider a structure generating method more efficient, if the resulting ensemble can be refined more easily to agree with the MD force field as indicated by lower energies. Extensive minimization either by steepest descent or conjugate gradient methods did not significantly reduce the energies of either the DG/MD or pure MD ensemble. Apparently small energy barriers and the general roughness of the energy landscape complicate localization of conformations with minimal energy during the annealing step. For some of the peptides it has even been shown experimentally by ultrafast time-resolved spectroscopy that their energy landscapes are surprisingly complex despite their small sizes.[33−36]

Discussion

Although NMR is the only technique to date for determining the structures of flexible or semiflexible peptides experimentally with atomic resolution, certain limitations inherent in this kind of spectroscopy have to be taken into account. In addition to aspects of correct identification and quantification of geometric constraints[37] the relation between the amount of experimental data and conformational heterogeneity is of concern. The lack of NOE constraints is a consequence of conformational flexibility[38] However, spectral overlap and special amino acid sequences can also prevent extraction of sufficient NOEs for an unambiguous well-defined structure (e.g. poly-proline II helix). NMR dynamics studies or boots-trap methods in structure calculation can help to separate conformational dynamics from the lack of experimental data. A more fundamental and not easy to overcome problem is the time and ensemble averaging during the NMR measurement.[4−9,39] NOE intensities are time averages of tens of milliseconds and ensemble averages of roughly 10^{17} molecules. Clearly, it is impossible to fully incorporate this averaging in a structure calculation. In almost all cases constraints are applied to one molecule, sometimes with some kind of picosecond time averaging. Ensemble calculations with a small number of molecules have been performed, but only in rare cases (ref 6 and references therein). The "averaging problem" is well-known in the NMR literature,[4,38,39] but there has also been a renewed discussion in recent years.[8,9] By comparing for peptides **2**, **8**, and **9** very elaborate unconstrained MD simulations with experimental NMR distance data, we have found that the bias of the NOE or ROE toward shorter distances can lead to suppression of more open and less compact structures in NMR structure calculations when the peptide is quite flexible.[40] However, the purpose of the present paper is not to validate or falsify the NMR structural ensembles for any or every peptide but to compare the performance of commonly used approaches to calculation of peptide structures taking the derived NOE distances for peptides **1**−**20** as given. The results presented above have pointed to the fact that conformational sampling is the most important aspect that needs to be considered. The objective of this work can, therefore, be stated as the following: Do the protocols investigated here explore the conformational space sufficiently. The aim is to test the performance of the computational approach, not the validity

of the individual NOE data or, even, final structures. Among our set of test molecules are some that are known, or likely, to exhibit internal motions and some that do not. Some of either kind displayed energy barriers that render the MD approach inappropriate. The lack of correlation between the presence of an energy barrier and characteristics such as ring size, additional constraints or even similarity of structures suggests that our findings are quite general. They are also in agreement with an emerging consensus that for peptides and proteins the energy landscape in conformational space is rough and rugged. Our test set of molecules comprises almost exclusively mono- or bicyclic peptides of moderate size. However, unconstrained linear peptides usually are mostly unstructured in solution, unless they experience the structuring influence of, for instance, a membrane environment (as does peptide **12**) or of a binding partner. Further, experimental NMR restraints will act to constrain the molecule when used in the structure calculation. Therefore, we believe that investigation of constrained peptides does not constitute a limitation, which is too severe. For the discussion of the performance of the various calculation protocols we will begin with those that use a given starting structure and generate the ensemble by conformational searching during a molecular dynamics simulation (protocols #2 and #3). It is clear that the energy barriers that impede proper sampling are the main problem of the conformational search. Table 1 shows that for many peptides even a simulation temperature of 1000 K was insufficient for overcoming the barriers. In our cases all calculations started from low energy structures fulfilling the NMR restraints and, thus, ensembles were found that represented subsets of the corresponding DG ensembles. However, in the real case the initial structure might be far from the correct ensemble in conformational space, and, thus, trapping of the conformational search by barriers could result in completely artificial results not corresponding to lowest energy structures. Furthermore, in additional calculations with a different force field (AMBER) we found a dependence of the results and also the existence of energy barriers on the force field used (data not shown). This indicates that for a given peptide the problem of energy barriers in MD conformational sampling might depend on the force field and MD software used. Unfortunately, peptide structure determinations based on protocols similar to our protocol 2 are seen quite often in the literature. With our demonstration of the shortcomings of these protocols we would like to discourage the use of them. In protocol 3 scaling of the nonbonded interactions partially alleviates the problems associated with protocol 2. However, in a few cases conformational sampling is still insufficient, and sometimes performance in terms of NOE violations or final energies is not satisfactory. Scaling of force field interactions is more typical for protein structure determination than for peptide studies. Although our study is limited to peptides, we generally recommend against the use of protocol 3 for the structure calculation of peptides or proteins.

Having seen that in many cases MD does not adequately explore conformational space, the question arises whether the restrictions originate from the force field or the NMR constraints that are applied. As mentioned above energy barriers seem dependent on the force field (CVFF vs AMBER). On the other hand, structure calculations without NMR restraints resulted in similar ensembles for protocols 1 and 2 for all peptides except **4 cis** (data not shown). The structures obtained without NMR data were distinctly different from those resulting from calculations with NMR restraints demonstrating that our molecules are not conformationally trivial in the sense that their conformation would already be determined by steric requirements of the amino acid residues and the intramolecular cyclization. Because NOEs contribute to the energy of the molecule they also modify the energy landscape and can create barriers. Apparently NMR constraints are more important than the force field with regard to the presence of energy barriers, although both contribute. The importance of the NOE constraints and the fact that they are very similarly implemented in the various programs for NMR structure calculation suggests that simply moving to another program or another force field might not solve the problems discussed here.

It might be argued that free MD combined with subsequent structure selection based on NMR data represents the optimum solution and in fact this method is also frequently seen in published studies. Aside from the remaining danger of incomplete sampling, as seen for peptide **4 cis** in our study, the yield in low energy structures conforming to the NMR constraints is usually reduced, so that many more structures have to be calculated to obtain a statistical ensemble.

The DG based approach is not restricted by the presence of energy barriers as independent structures are generated by a direct geometrical method.[12] This does not a priori guarantee that every conformation that is compatible with the experimental distance constraints will be found, i.e., that the conformational sampling is sufficient. But the results shown here clearly indicate that DG performs much better than high-temperature MD with regard to sampling properties. In fact, the agreement between the structural ensembles obtained by the DG/MD protocol and the random MD protocol suggests that likely no solutions to the structure determination problems were overlooked by the DG method (see below). Of course, a proper choice of parameters, or algorithms in the case of DG, is a prerequisite for every calculation strategy. We have used reasonable implementations of each method rather than ideal ones, because we intended to focus on the applicability for the nonexpert user.

As DG structures exhibit poor covalent geometry, subsequent refinement by MD simulated annealing steps is indispensable.[41−43] A change of force field (AMBER instead of CVFF) has only little consequences for the resulting structures (data not shown), because the molecular dynamics in this approach only serve to establish a correct local (covalent) geometry with corresponding relaxation of potential energies We noticed that the final energies of the DG ensembles depend somewhat on the extent of annealing that is performed (see above) and are partially higher than those obtained with the pure MD protocol (Figure 6). Although energy constitutes the sorting criterion in our analysis, it cannot be expected that values obtained at our level of sophistication (homogeneous dielectric constant, no cross

terms between potentials, etc.) will be accurate enough for a quantitative discussion. Still it might be of interest that such small peptide models seem to exhibit energy landscapes with pronounced roughness complicating the annealing procedure. Certainly an advantage of the DG/MD protocol is that the DG part that generates (and roughly optimizes) the structural ensemble requires for peptides much less computational time than the other protocols.

For proteins, on the other hand, the DG/MD approach was found to be less efficient,[41,42] so that the typical strategy for protein structure calculation is Protocol 4 that consists of iterative annealing of random starting structures in the presence of NMR restraints. The random coordinates require initially strongly reduced force field interactions that are restored to their normal value in the course of the MD steps. Protocol 4 gave quite similar results as protocol 1 (DG/MD) for our peptides. No additional conformational families were found with the random MD approach compared to the DG based calculations. The fact that the same conformational families are detected by the DG method and the randomization of coordinates suggests that for our peptides both procedures achieve sufficient sampling of the conformational space. The somewhat higher conformational variability of the random MD ensembles (Figure 3) is probably related to the higher overall energies of the final structures (Figure 6). If an ensemble cannot be adequately refined by the annealing step, that is, higher energies are found for the same overall conformations and similar degrees of NOE violation, we assume that the initially generated structures were less compatible with the force field. In a real structure determination one would need to achieve a fully relaxed conformational ensemble by extended or repeated simulated annealing. The increased computational costs compared to the other protocols lead us to consider protocol 4 as less effective. The higher efficiency of the DG based methods in this regard is intuitively understood, because DG generates independent, but not arbitrary structures, that exhibit roughly correct covalent geometry and satisfy the experimental NMR data. Random structures have to acquire these properties during the MD simulated annealing, while the DG structures are merely refined during the MD part.

Conclusions

MD simulations employing experimental NMR restraints are not well suited for conformational searching in a structure calculation protocol for peptides. We observed artificially well-defined structural ensembles for our test set of 20 peptides often accompanied by energy barriers that could not be overcome during 1000 ps simulation runs at 1000 K. The results of the pure MD method for structure calculation were found to depend markedly on the force field used. Distance geometry based protocols on the other hand explore the conformational space more thoroughly and are quite insensitive to changes in force field parameters as they utilize molecular dynamics only for refinement. We recommend the combined DG/MD approach as the most efficient method for the calculation of peptide structures.

Acknowledgment. Dr. Marion Götz is acknowledged for her help in the preparation of the manuscript. This study was supported in part by the Deutsche Forschungsgemeinschaft.

References

(1) The PDB team, The Protein Data Bank. In *Structural Bioinformatics*; Bourne, P. E., Weissig H., Eds.; John Wiley & Sons: Hoboken, NJ, 2002; pp 181−198.

(2) Wider, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 193−275.

(3) Lee, S. C.; Russell, A. F.; Laidig, W. D. *Int. J. Pept., Prot. Res.* **1990**, *35*, 367−377.

(4) Constantine, K. L.; Mueller, L.; Andersen, N. H.; Tong, H.; Wandler, C. F.; Friedrichs, M. S.; Bruccoleri, R. E. *J. Am. Chem. Soc*. **1995**, *117*, 10841−10854.

(5) Melacini, G.; Zhu, Q.; Goodman, M. *Biochemistry* **1997**, *36*, 1233−1241.

(6) Cuniasse, P.; Raynal, I.; Yiotakis, A.; Dive, V. *J. Am. Chem. Soc.* **1997**, *119*, 5239−5248.

(7) Daura, X.; Antes, I. van Gunsteren, W. F.; Thiel, W.; Mark, A. E. *Proteins* **1999**, *36*, 542−555.

(8) Bürgi, R.; Ritera, J.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *19*, 305−320.

(9) Peter, Ch.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297−310.

(10) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol*. **2003**, *13*, 168−174.

(11) Havel, T. F.; Wüthrich, K. *J. Mol. Biol*. **1985**, *182*, 281−294.

(12) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press: Somerset, England, and John Wiley: New York, 1988.

(13) Berndt, K. D.; Guntert, P.; Wüthrich, K. *Proteins* **1996**, *24*, 304−313.

(14) Wells, C.; Glunt, W.; Hayden, T. L. *J. Mol. Struct*. **1994**, *308*, 263−271.

(15) Hodsdon, M. E.; Ponder, J. W.; Cistola, D. P. *J. Mol. Biol.* **1996**, *264*, 585−602.

(16) Reams, R.; Chatham, G.; Glunt, W.; McDonald, D.; Hayden, T. *Comput. Chem.* **1999**, *23*, 153−163.

(17) Vankampen, A. H. C.; Buydens, L. M. C.; Lucasius, C. B.; Blommers, M. J. J. *J. Biomol. NMR* **1996**, *7*, 214−224.

(18) Huang, E. S.; Samudrala, R.; Ponder, J. W. *J. Mol. Biol*. **1999**, *290*, 267−281.

(19) Xu, H. F.; Izrailev, S.; Agrafiotis, D. K. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1186−1191.

(20) Havel, T. F. *Biopolymers* **1990**, *29*, 1565−1585.

(21) Edwards, J.; Chatham, G.; Glunt, W.; McDonald, D.; Wells, C.; Hayden, T. *Comput. Chem*. **1997**, *21*, 115−124.

(22) Vankampen, A. H. C.; Beckers, M. L. M.; Buydens, L. M. C. *Comput. Chem*. **1997**, *21*, 281−297.

(23) Liu, Y.; Zhao, D.; Altman, R. Jardetzky, O. *J. Biomol. NMR* **1992**, *2*, 373−388.

(24) Beckers, M. L. M.; Buydens, L. M. C.; Pikkemaat, J. A.; Altona, C. *J. Biomol. NMR* **1997**, *9*, 25−34.

(25) Karimi-Nejad, Y.; Warren, G. L.; Schipper, D.; Brünger, A. T.; Boelens, R. *Mol. Phys.* **1998**, *95*, 1099−1112.

(26) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27−38.

(27) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J. Genest, M.; Hagler, A. T. *Proteins* **1988**, *4*, 31−47.

(28) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem*. **1986**, *7*, 230−252.

(29) Renner, C.; Behrendt, R.; Spörlein, S.; Wachtveitl, J.; Moroder, L. *Biopolymers* **2000**, *54*, 489−500.

(30) Renner, C.; Cramer, J.; Behrendt, R.; Moroder, L. *Biopolymers* **2000**, *54*, 501−514.

(31) Renner, C.; Behrendt, R.; Heim, N.; Moroder, L. *Biopolymers* **2002**, *63*, 382−393.

(32) Schütt, M.; Krupka, S.; Milbradt, A. G.; Deindl, S.; Sinner, E.-K.; Oesterhelt, D.; Renner, C.; Moroder L. *Chem. Biol*. **2003**, *10*, 487−490.

(33) Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7998−8002.

(34) Bredenbeck, J.; Helbing, J.; Sieg, A.; Schrader, T.; Zinth, W.; Renner, C.; Behrendt, R.; Moroder, L.; Wachtveitl, J.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6452−5457.

(35) Bredenbeck, J.; Helbing, J.; Behrendt, R.; Renner, C.; Moroder, L.; Wachtveitl, J.; Hamm, P. *J. Phys. Chem. B* **2003**, *107*, 8654−8660.

(36) Wachtveitl, J.; Spörlein, S.; Satzger, H.; Fonrobert, B.; Renner, C.; Behrendt, R.; Oesterhelt, D.; Moroder, L.; Zinth W. *Biophys. J*. **2004**, *86*, 2350−2362.

(37) Schneider, T. R.; Brünger, A. T.; Nilges, M. *J. Mol. Biol*. **1999**, *285*, 727−740.

(38) Neuhaus, D.; Williamson, M. *The nuclear Overhauser effect in Structural and Conformational Analysis;* VCH Publisher: New York, 1989.

(39) Nanzer, A. P.; Poulsen, F.; van Gunsteren, W. F.; Torda, A. E. *Biochemistry* **1994**, *33*, 14503−14511.

(40) Carstens, H.; Renner, C.; Milbradt, A. G.; Moroder, L.; Tavan, P. *Biochemistry* **2005**, *44*, 4829−4840.

(41) Holak, T. A.; Kearsley, S. K.; Kim, Y.; Prestegard, J. H. *Biochemistry* **1988**, *27*, 6135−6142.

(42) Holak, T. A.; Nilges, M.; Oschkinat, H. *FEBS Lett*. **1989**, *242*, 218−224.

(43) Knegtel, R. M.; Boelens, R.; Ganadu, M. L.; Kaptein, R. *Eur. J. Biochem*. **1991**, *202*, 447−458.

CT050203J