THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

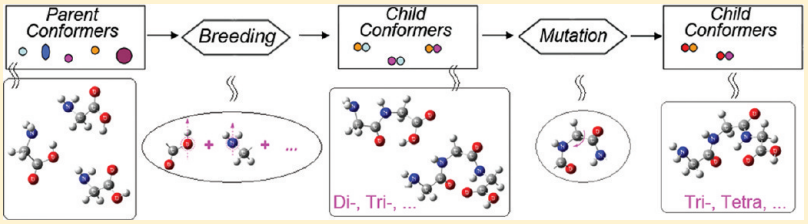
Comprehensive Conformational Studies of Five Tripeptides and a Deduced Method for Efficient Determinations of Peptide Structures

Wenbo Yu,[†,‡,§] Zhiqing Wu,[†] Huibin Chen,[‡] Xu Liu,[‡] Alexander D. MacKerell, Jr.,[§] and Zijing Lin*[,†,‡]

[†]Hefei National Laboratory for Physical Sciences at Microscale and [‡]Department of Physics, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230026, China

[§]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn Street, Baltimore, Maryland 21201, United States

Supporting Information



ABSTRACT: Thorough searches on the potential energy surfaces of five tripeptides, GGG, GYG, GWG, TGG, and MGG, were performed by considering all possible combinations of the bond rotational degrees of freedom with a semiempirical and ab initio combined computational approach. Structural characteristics of the obtained stable tripeptide conformers were carefully analyzed. Conformers of the five tripeptides were found to be closely connected with conformers of their constituting dipeptides and amino acids. A method for finding all important tripeptide conformers by optimizing a small number of trial structures generated by suitable superposition of the parent amino acid and dipeptide conformers is thus proposed. Applying the method to another five tripeptides, YGG, FGG, WGG, GFA, and GGF, studied before shows that the new approach is both efficient and reliable by providing the most complete ensembles of tripeptide conformers. The method is further generalized for application to larger peptides by introducing the breeding and mutation concepts in a genetic algorithm way. The generalized method is verified to be capable of finding tetrapeptide conformers with secondary structures of strands, helices, and turns, which are highly populated in larger peptides. This show some promise for the proposed method to be applied for the structural determination of larger peptides.

1. INTRODUCTION

The properties of a biomolecule often depend on a few stable conformers on its potential energy surface (PES). Finding the most stable conformer is the key to understanding the behaviors of a biomolecule.[1−4] However, the task of reliably determining stable conformers is far from trivial.[5,6] Due to the bond rotational flexibility of biomolecules, there are many possible conformers with varying degrees of stability dictated by different combinations of intramolecular interactions such as hydrogen bonds.[7,8] One way to reliably determine the stable conformers is through a thorough structural search on the PES by considering all trial structures generated by a full combination of internal single-bond rotamers.[9] Unfortunately, the number of trial structures thus generated increases exponentially with the number of rotational degrees of freedom.[10] The thorough search method is feasible only for small biomolecules like amino acids but impractical for larger peptides and proteins.[2,10] Consequently, most conformational searches of peptides and proteins rely on the screening of a subset of possible structures by some constrained sampling methods like Monte Carlo,[11−14] simulated annealing,[15−17] and genetic algorithm[18−21] based methods that generate trial structures with some designated stochastic processes. These sampling methods are usually much faster than the thorough search method, but there is no assurance of their reliabilities due to the lack of trusted approaches for designing the trial parameter sets. Indeed, recent studies on amino acids, dipeptides, and larger peptides show that these sampling methods often miss important conformers and thus may provide false interpretations of the experimental results.[2,10,22] Clearly, reliability is the basic requirement for developing any efficient sampling method.

On the basis of the observed structural connections between stable conformers of dipeptides and amino acids, we have recently put forward a method for sampling the PESs of dipeptides.[22] This method generates the trial dipeptide conformers by a designated combination of the parent amino acid conformers. Performance of the method applied on numerous dipeptides studied before by Monte Carlo or molecular dynamics simulations proved that the method is

Received: August 14, 2011
Revised: December 16, 2011
Published: January 19, 2012

   dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

both highly efficient and reliable by providing the most complete ensembles of dipeptide conformers as well as improved agreements with the available experimental data.

Encouraged by the appealing results of dipeptides, we now move forward to tripeptide systems. Tripeptides are important biomolecules as they are involved in many biological processes. Moreover, the tripeptide is also the smallest peptide where the impacts of neighboring amino acid residues on the middle residue can be examined. As structural information of tripeptide segments is useful in the building of protein structures,[23,24] comparing the structural characteristics of the middle residue in a tripeptide with that in proteins could shed light on the exploration of protein structures.

Small peptides, which served as affordable models to mimic conformational preferences on various properties of proteins, have been continuously studied. However, most peptide systems that were theoretically studied are actually so-called peptide analogues or capped peptides and usually in the form of acetyl and methyl amide blocked amino acids or peptides. Such peptide models were widely studied for their conformational equilibrium,[25,26] solvation effects,[27] ion coordination effects,[28] vibrational absorption and circular dichroism spectra,[29] and excited states.[30] They have also been commonly applied in force field developments and assessments.[31−33] Though less work was performed on natural or real peptides, these studies still cover various aspects, including solvation effects,[34] metal ion coordinations,[35] infrared spectra,[36] aromatic stacking effects,[37] protonation states,[38] and excited states.[39,40] However, detailed conformational studies were even less limited by the high computational costs, and if we focus on tripeptides, only conformations of five tripeptides may be considered to have been carefully searched so far, besides the simplest glycine tripeptide.[41−45] Moreover, these studies were carried out in a constrained way and relied on the screening of a subset of all possible structures by the force field based molecular dynamics. Like their counterpart in the dipeptide studies, the reliability of such tripeptide conformational searches cannot be taken for granted and should be benchmarked with results derived from a systematic approach.

In the present work, we performed thorough searches for the conformers of five tripeptides, GGG, GYG, GWG, TGG, and MGG (G = glycine, Y = tyrosine, W = tryptophan, T = threonine, M = methionine). This set of tripeptides was chosen because the conformers of their constituting amino acids and dipeptides had mostly been thoroughly studied before,[9,22,46−48] allowing a detailed comparison among the conformers of tripeptides and their constituting segments. Similar to our previous studies on dipeptides, close connections between tripeptide conformers and dipeptides or amino acid conformers were found. As a result, an efficient method for locating stable tripeptide conformers is proposed. The method is validated by applications to other tripeptides studied before, and additional low-energy conformers were found. The proposed method is further generalized in the view of the genetic algorithm, and its applicability to larger peptides is discussed in the context of locating typical secondary structures in tetrapeptides.

2. COMPUTATIONAL METHODS

2.1. Method for Exploring the Tripeptide Conformations. There are many local minima within a small-energy range of the global minimum on the PES of a tripeptide due to numerous possible combinations of intramolecular interactions. To ensure a reliable description of the PES and to locate all of the low-energy conformers, the conformational space of a tripeptide should be fully explored by considering all possible combinations of the bond rotational degrees of freedom. The process of generating trial conformers for exploration of tripeptide conformations was the same as that described before for the conformational studies of amino acids,[9] with the exception that the peptide bonds were always kept to be in the trans- configuration as conformers in the cis- form were known to be energetically unfavorable.

The total numbers of trial conformers thus generated for GGG, GYG, GWG, TGG, and MGG were 3072, 36864, 36864, 55296, and 165888, respectively. The trial conformers of GGG were first optimized at the HF/3-21G(d) level of theory, and the unique structures obtained were subjected to further optimization at the BHandHLYP/6-31G(d) level. The final single-point energies were evaluated at the BHandHLYP/6-311++G(d,p) level. To save computational cost, trial conformers of the other four tripeptides were first optimized by the semiempirical method of PM3. Single-point energies at the level of HF/3-21G(d) were calculated for all of the unique structures obtained. Sets of 120 structures with increasing HF/3-21G(d) single-point energies were then successively optimized at the HF/3-21G(d) level until the last set of 120 structures did not produce a new conformer within the energy range of 5 kcal/mol from the located global minimum at this level of theory. Subsequently, sets of 120 structures with increasing optimized HF/3-21G(d) energies were successively reoptimized at the BHandHLYP/6-31G(d) level until the last set of 120 structures did not produce a new conformer within the energy range of 3 kcal/mol from the located global minimum at this level. The final single-point energies were then evaluated at the BHandHLYP/6-311++G(d,p) level. Here, BHandHLYP was chosen as it performed the best among the DFT variants when benchmarked with the CCSD results for amino acids.[49] However, for comparison purposes, single-point energies for some of the most important tripeptide conformers were also calculated at the MP2/aug-cc-pVTZ level. To calibrate BHandHLYP optimized geometries, a modern DFT method M06-L, which includes dispersion corrections and is shown to have good performance,[50] was also used to reoptimize some important tripeptide conformers. Average and standard differences on geometry parameters were then evaluated to judge the necessity of considering a dispersion term for tripeptide systems. It should be noted that such a multistage optimization strategy has been commonly used to search PESs of larger molecular systems. For example, Roy et al.[51,52] used a PM3-HF-B3LYP multistage optimization strategy to study glycolic acid water clusters, and Toroz et al.[53] explored the PES of tyrosine-glycine dipeptide with a HF-B3LYP multistage conformational search strategy.

All of the low-energy tripeptide conformers were verified to be the true local minima by corresponding frequency calculations with the harmonic approximation. Calculated vibrational frequencies and their contributions to the zero-point vibrational energies (ZPVE) and free energies were scaled with a factor of 0.926.[22] Equilibrium conformational distributions were then determined by their relative Gibbs free energies based on the Boltzmann distribution function. To evaluate the impact of anharmonic effects on our results, anharmonic frequency calculations were also performed for glycine tripeptide.

2.2. Connections between Tripeptide, Dipeptide, and Amino Acid Conformations. In order to explore the

inherent connections between conformations of tripeptides, dipeptides, and amino acids, the low-energy tripeptide conformers were analyzed against the conformations of dipeptides and amino acids. The conformers of amino acids used in the analysis are the same as those used in our previous work on dipeptides.[22] All of the low-energy dipeptide conformers except that for GY came from our previous work.[22] Low-energy GY conformers were obtained here by a systematic search while considering only conformers with the *trans*-carboxyl configuration. A total of 1153 trial GY conformers were generated and optimized at the HF/3-21G(d) level, and unique conformers were subsequently reoptimized at the BHandHLYP/6-31G(d) level. For consistency, single-point energies were calculated at the BHandHLYP/6-311++G(d,p) level for all low-energy conformers of dipeptides and amino acids.

Comparative analysis on connections between low-energy conformers of tripeptides, dipeptides, and amino acids leads to a method of finding tripeptide conformers based on the structures of constituting dipeptides and amino acids. The proposed method was applied to locate the stable conformers of five tripeptides studied before in other works. Conformers of the five tripeptides obtained by the new method were presented at the respective level of theories used in the former studies for the convenience of comparison. That is, results were described at the MP2/6-31+G(d)//B3LYP/6-31+G(d) level for YGG[41] and RI-MP2/cc-pVTZ//RI-MP2/cc-pVDZ level for FGG, WGG, GFA, and GGF (A = alanine, F = phenylalanine).[42−45]

RI-MP2 calculations were carried out using the ORCA package of programs,[54] MP2/aug-cc-pVTZ level single-point energy calculations were done with the NWChem package of programs,[55] M06-L calculations were performed using the Gaussian 09 suit of programs,[56] and all other calculations were performed using the Gaussian 03 suit of programs.[57] Trial conformers were generated with the help of our in-house-developed program written in visual basic language.

3. RESULTS AND DISCUSSION

3.1. Important Tripeptide Conformers. After extensive searches on the PESs of the five tripeptides, many stable conformers were found for each tripeptide. Upon comparison among the conformational sets of tripeptides, dipeptides, and amino acids, two characteristic trends concerning the relative conformational energies and stabilities are evident. First, the numbers of stable tripeptide conformers in a given electronic energy range are larger due to the increased number of rotational degrees of freedom. For example, there are 40 stable GWG conformers in a 3 kcal/mol electronic energy range from the global minimum, while there are only 15 stable GW conformers in the same energy range. Therefore, the energy differences between different stable conformers are often smaller for tripeptides than those for dipeptides. Second, the impact of the vibrational energy and entropy on the free energy increases with the number of atoms and is larger in the tripeptide than that in the dipeptide and amino acid. For example, the GWG conformer with the lowest Gibbs free energy at the standard state has an electronic energy of 2.3 kcal/mol above its global minimum. That is, there are substantial differences between the PES and the Gibbs free energy surface for the tripeptide compared to that of the dipeptide and amino acid.

Because the equilibrium conformational populations are determined by their relative free energies and the temperature range around room temperature is of the most interest, a tripeptide conformer is denoted here with a numeral suffix to indicate its relative Gibbs free energy ordering at the standard state. For example, the conformer ggg1 is the most populated conformer among all GGG conformers at the standard state. For conciseness, we focus our discussions on important conformers. A conformer is defined to be important if its percentage share at the standard state is over 3%. The numbers of important conformers for GGG, GYG, GWG, TGG, and MGG are 6, 9, 15, 8, and 7, respectively. As conformers with low electronic energies are also of high interest on their own, two conformers with the lowest electronic energy are also included in the set of important conformers in the following discussion even when their percentage shares at the standard state are less than 3%. Details about the relative energies and equilibrium distributions of conformers at the standard state can be found in the Supporting Information (SI).

Compared to BHandHLYP/6-311++G(d,p) level energy profiles, similar relative electronic energy orders at the MP2/aug-cc-pVTZ level are found for all five tripeptides, as shown in Tables S1−5 in the SI. However, generally speaking, larger energy differences between conformers are seen. That is, fewer important conformers will be present at the MP2/aug-cc-pVTZ level. A similar trend is seen if anharmonic corrections are considered for equilibrium distributions compared to that with only harmonic approximations, as shown in Table S1 in the SI using glycine tripeptide as a test case. As the main purpose of this paper is to develop a method for finding all important conformers, use of BHandHLYP/6-311++G(d,p) level results that have more important conformers as target data can ensure the developed method is more reliable. Also, use of the same level of theory as that used in our previous dipeptide studies can keep the results consistent. However, one should be cautious that the anharmonic corrections will become more important for larger tripeptides and the harmonic approximation may not be sufficient to accurately describe equilibrium distributions.

A modern DFT method with a dispersion correction, M06-L, was also used to optimize a few important tripeptide conformers. The geometry differences with BHandHLYP level optimized conformers were evaluated as shown in Table S11 (SI), and the aligned structures are shown in Figure S1 in the SI. The standard differences on bonds, angles, and dihedrals of all tested tripeptide conformers are less than 0.01 Å, 0.8°, and 8°, respectively. The largest difference is that M06-L optimized structures have shorter intramolecular hydrogen bonding (H-bond) distances than that of BHandHLYP, especially for π-electron-involved interactions as found in GYG. Indeed, GYG conformers have larger root-mean-square distances than other tested tripeptides considering the superimposed structures shown in Figure S1 (SI), where M06-L optimized GYG conformers have closer backbone−side chain contacts compared to the BHandHLYP optimized one. This is consistent with the findings for M06-L and B3LYP optimized YG conformers by Toroz et al.[58] As the search method to be developed is independent of the optimization methods, BHandHLYP optimized tripeptide conformers are used here to be comparable with previously optimized conformers of amino acids and dipeptides. However, a modern DFT method with dispersion correction can be used with our conformation search method in the future to get more accurate structures.

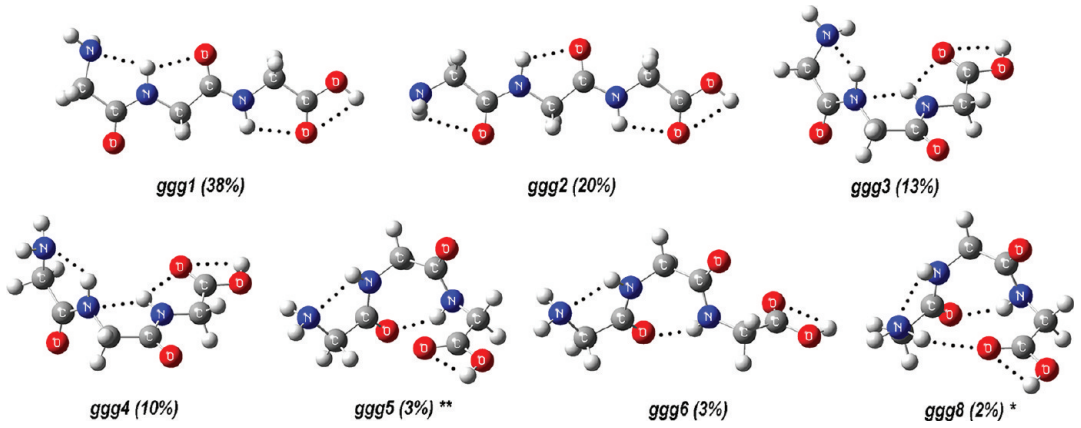


Figure 1. Important conformers of GGG with the equilibrium populations at the standard state given in parentheses. The lowest and the second lowest electronic energy conformers are noted by one and two asterisks, respectively. This notation is used throughout the paper.

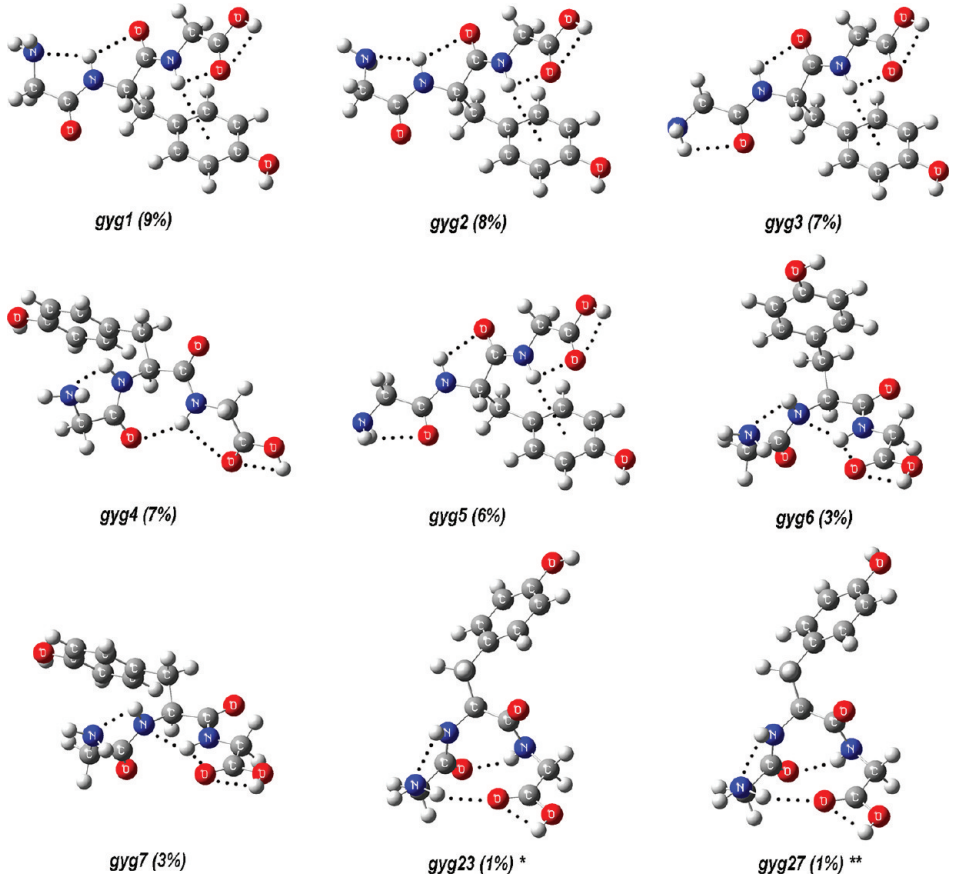


Figure 2. Structures of some important GYG conformers.

Figure 1 shows the important conformers of GGG. Being the smallest tripeptide, GGG serves as a basic model system for understanding tripeptide conformers. As shown in Figure 1, the most important GGG conformer, ggg1, adopts an extended configuration with four major favorable intramolecular interactions, OCO−H···O═COH (*cis*-carboxyl group), OCN−H···O═COH, OCN−H···O═CNH, and OCN−H···NH_2. The structure of ggg2 is very similar to that of ggg1 and differs only in the H-bond interaction formed at the N-terminus, which is NH_2···O═CNH in ggg2. Conformers ggg3 and ggg4 have similar structures and differ only in the placement of the amino groups. Similar to ggg1, conformers ggg3 and ggg4 each have four important intramolecular interactions, with the OCN−H···O═CNH interaction in ggg1 being replaced by the OCN−H···N(H)CO interaction in ggg3 or ggg4. Conformers ggg5 and ggg6 differ in their placement of the carboxyl group, and there are only three important interactions in ggg5 or ggg6. All of the favorable interactions between the peptide bond units in ggg5 or ggg6 and in ggg1 or ggg2 are in the OCN−H···O═CNH form. However, the differences are such that the positive charge donor involved in such an interaction is on the C-terminus side for ggg5 or ggg6 while it is on N-terminus side for ggg1 or ggg2.

dx.doi.org/10.1021/jp207807a | *J. Phys. Chem. B* 2012, 116, 2269−2283

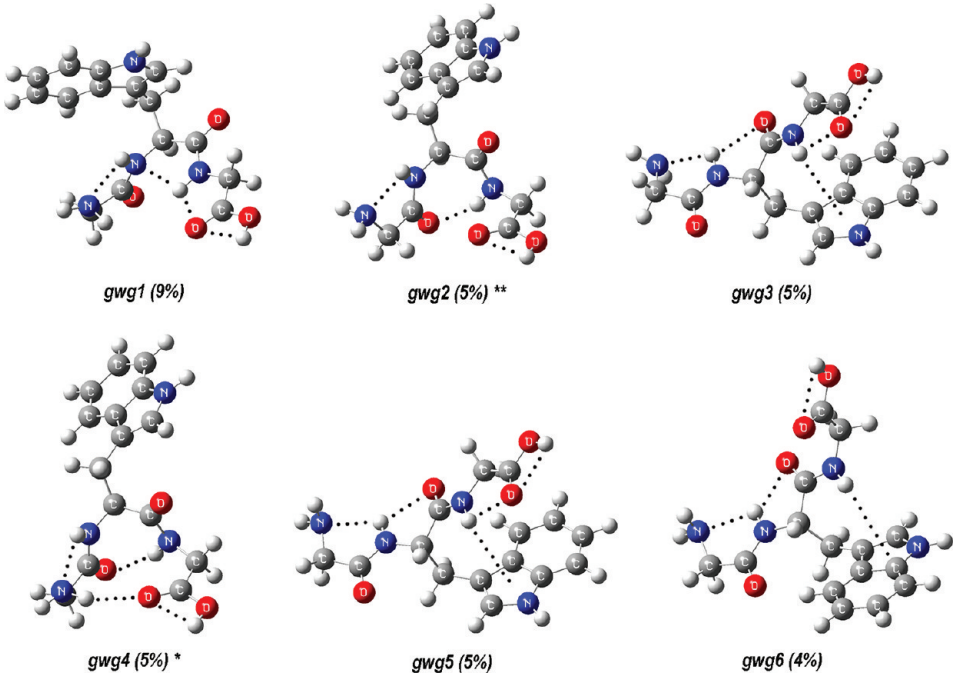


Figure 3. Structures of some important GWG conformers.

As mentioned above, the relative electronic energy ordering can be very different from the relative free energy ordering for tripeptides. Conformers ggg8 and ggg5 adopt relatively compact configurations that are beneficial for forming strong H-bonds, and their electronic energies are the lowest. Due to an unfavorable entropic effect, however, they are relatively unimportant at the standard state when compared with ggg1 or ggg2, which adopt extended configurations. The lowest electronic energy conformer found in this work is the same as that identified by Zhang et al.[59] The result is also consistent with observations given by Wang et al.[60] regarding the lowest-energy conformer obtained by Strittmatter et al.[61] through Monte Carlo simulation that does not corresponds to the global minimum on the electronic energy surface. In fact, the most stable conformer located by Strittmatter et al. corresponds to our conformer ggg9 in the present study, and its electronic energy is 2.5 kcal/mol above the global minimum. Similar to the results for dipeptides,[22] this result once again casts doubt on the reliability of exploring the PES by the Monte Carlo method.

Figures 2 and 3 show the structures of the most important conformers for GYG and GWG, respectively. Compared to GGG, new types of interactions formed between side-chain and backbone atoms are introduced by the aromatic side chain in GYG and GWG. In addition to the important intramolecular interactions found in GGG, there are interactions formed between the backbone OCN−H group and the π-electron of the aromatic side chain in conformers gyg1, gyg2, gyg3, and gyg5 and in conformers gwg3, gwg5, and gwg6. Other than the side-chain difference, similarity in the backbone configurations is often seen among the important conformers of GGG, GYG, and GWG. For example, the backbone configurations of gyg23, gyg27, and gwg4 are basically the same as that of ggg8.

Notice that the OCN−H···π-electron interaction is relatively weak. Moreover, for GYG and GWG, a strong H-bond cannot be formed between the side chain N−H or O−H group and backbone atoms as the side-chain tail is difficult to bend to the backbone direction. Consequently, the energy difference between conformers with and without the backbone−side chain interaction is small. As a result, a more evenly distributed conformer population than that in GGG is seen in GYG and GWG due to the increased number of conformers with similar stabilities.

The relative stabilities of important GGG and GYG (GWG) conformations are well-correlated when compared with the differences among the important GGG and YGG[41] (WGG[43]) conformations. Similarly, the structural characteristics of important FGG,[42] GFA,[44] and GGF[45] conformations are also quite different from that of GGG. Existence of a side chain at the N- or C-terminus will have more impact to the conformational stabilities than that at the middle residue position. That is because, in the first case, the whole tripeptide structure becomes more flexible for forming interactions, including strong H-bonds, between the backbone and side-chain functional groups. Generally speaking, a side-chain substitution taking place at the terminal side is supposed to have a stronger effect on tripeptide structures than that at the middle residue position.

Figure 4 shows structures of the seven important conformers of TGG. Other than the backbone−side chain interactions, there is notable correspondence among the conformations of TGG and GGG. For example, the backbone configurations of the most important conformer tgg1 and the two lowest electronic energy conformers tgg5 and tgg10 are basically the same as their GGG counterparts ggg1, ggg5, and ggg8, respectively. The side-chain hydroxyl group is involved in forming one or two H-bonds with the backbone functional groups in all of the conformers shown in Figure 4. One such H-bond, $C_\beta O-H\cdots O{=}CNH$, is found in conformers tgg1, tgg2, tgg3, tgg4, and tgg10. The side-chain hydroxyl is involved in forming two H-bonds in conformer tgg5 by interacting with the N-terminus amino group and the C-terminus carboxyl group as $HNH\cdots O(H)C_\beta$ and $C_\beta O-H\cdots O{=}COH$. The side-chain hydroxyl group is also involved in forming two H-bonds in

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

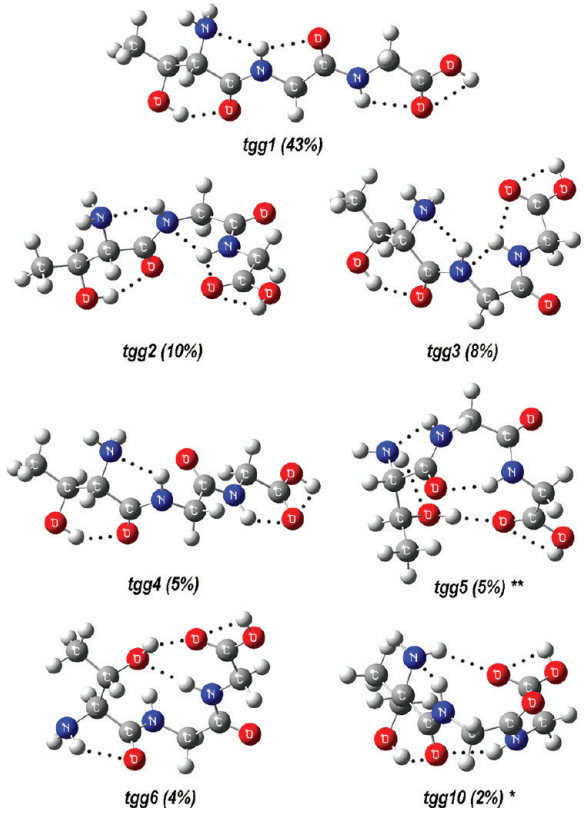


Figure 4. Structures of some important TGG conformers.

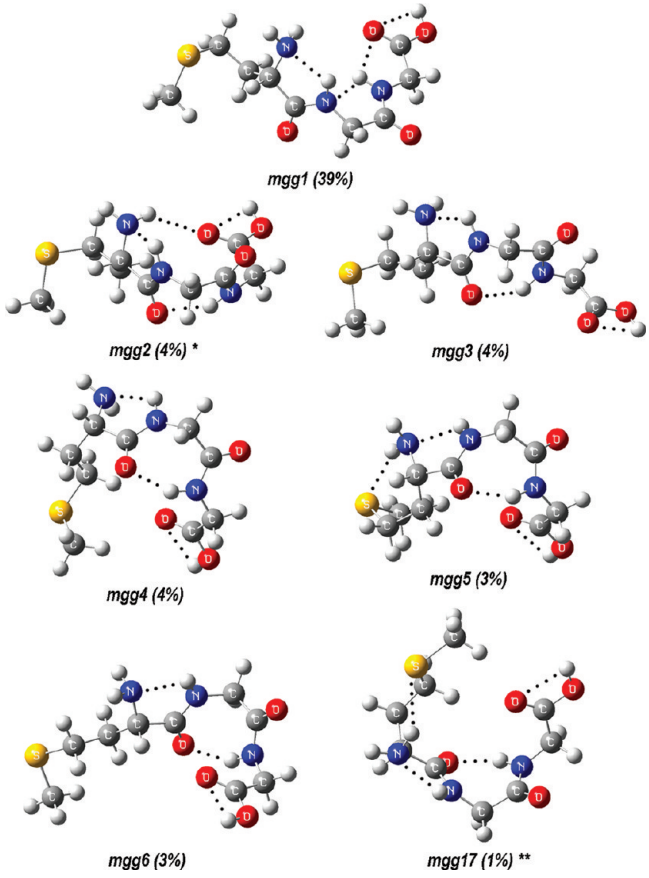


Figure 5. Structures of some important MGG conformers.

conformer tgg6 by interacting with the amido group in the peptide bond unit and the C-terminus carboxyl group as OCN−H···O(H)C_β and C_βO−H···O═COH.

Structures of the seven important conformers of MGG are shown in Figure 5. Similar to MG conformers studied before,[22] the sulfur-involving H-bonds are not much favored as the S···H interaction is relatively weak and requires compact configuration that is associated with reduced entropy. For example, the backbone configurations of mgg4 and mgg5 are quite similar, and mgg5 contains an extra H-bond, HNH···S. However, mgg4 is more important than mgg5 as mgg5 has a more compact configuration.

Overall, it is worth noting that there are at least three important intramolecular interactions in any one of the important conformers shown in Figures 1−5. This can be attributed to the H-bond stabilization effect and the availability of various charge donor and acceptor groups in a tripeptide. Another common feature is that the tripeptide conformers with the lowest electronic energy studied here all adopt compact structures to form more and/or stronger intramolecular interactions. However, the most important tripeptide conformers usually adopt extended structures as such a structure is thermodynamically favorable. Comparing the available results on the conformations of amino acids, dipeptides, and tripeptides shows a clear trend that the more the rotational degrees of freedom a biomolecule has, the less important the lowest electronic energy conformer is. Going through amino acids to dipeptides and to tripeptides, the entropic effect plays an increasingly important role in determining the importance of a conformer. In the absence of very strong interactions such as metal ion bindings, the entropy effect is expected to be dominant in determining the conformations adopted by large biomolecules. Indeed, the importance of the entropy contributions has been well-known for peptide conformational stabilities[62,63] and protein foldings.[64,65]

3.2. Secondary Structures of the Important Conformers. The structural preference of the middle residue in a tripeptide can be studied using the secondary structure language of Ramachandran φ and ψ values defined for residues in proteins.[66] Figure 6 shows a Ramachandran map for middle

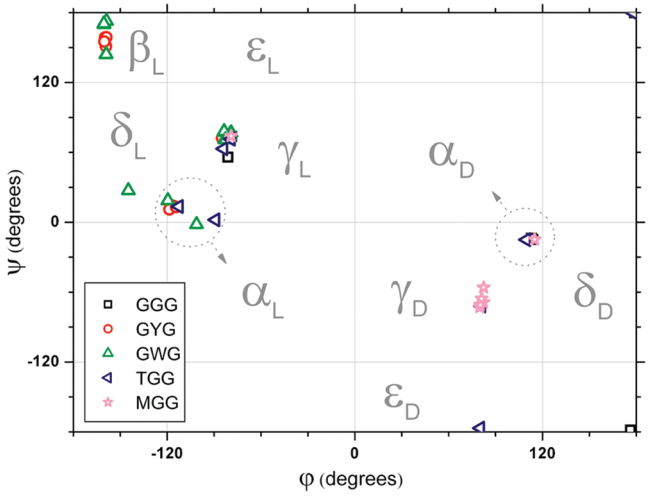


Figure 6. Ramachandran map for the middle residues of important tripeptide conformers studied here. The tripeptide structures represented inside of the two dotted circles are significantly different from those of dipeptides and show a tendency of larger peptides to favor structures in the α region.

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

residues of the most important tripeptide conformers studied here. The nine catchment regions on the Ramachandran map are labeled according to the notation adopted by Császár and Perczel.[67] Clearly, the overall characteristics of the tripeptide secondary structures are quite different from that found for proteins. That is, the structures of isolated tripeptides are far from representative of the structures of protein segments.

However, useful insight can be revealed by comparing the secondary structures of dipeptides and tripeptides. For dipeptides, only ψ for the N-terminus residue and φ for the C-terminus residue really exist. Nevertheless, for the sake of comparison, we may define a ψ value for the C-terminus residue in a dipeptide based on the structural similarity between O−H in a *trans*-carboxyl group and N−H in a peptide bond unit, the same way as adopted in ref 53. For convenience of describing a structure on the Ramachandran map, the structure notation[68] based on the number of atoms involved in a pseudocycle formed with a H-bond that is commonly used for peptide analogues is adopted here for dipeptides and tripeptides. In this notation, the β-strand structure in the β_L region is labeled C5, the γ-turn structure in the middle of the γ_L/γ_D region is labeled C7eq/C7ax, the structure in the lower-left/upper-right corner of the γ_L/γ_D region is labeled β_2/β, the structure in the ε_D region is labeled α_D, the structure in the α_D region is labeled α_L, and the structure in the δ_D region is labeled α'. However, one should be cautious that such structure notation can be confused with Ramachandran map region notation as similar labels are used for different purposes.

Going through the C-terminus residue of the important dipeptide conformers with *trans*-carboxyl configuration studied before,[22] two types of structure such as C5 and C7eq/C7ax can be found. Besides these structures found in dipeptides, β_2/β and α_D structures are also found for tripeptides. The GWG conformer found in the δ_L region is formed due to its specific indole side-chain orientation that pushes φ from the corner of the γ region to the δ_L region and can be treated as a special β_2 structure. These structures are energetically disfavored in dipeptides but are stabilized by the additional residue in tripeptides.

The appearance of β_2/β structures in the lower-left/upper-right corner of the γ_L/γ_D region in tripeptides implies a trend that the structures in the α_L/α_D region will be energetically favored with increasing number of amino acid residues. This is in accordance with the previous results that secondary structures such as helices (in α_L/α_D region) and β-turns may appear when there are more than three residues.[67] The α_D structure in the ε_D region also implies the emergence of a new secondary structure, the β-turn. This is in line with the experimental and theoretical studies by Chin et al[69] that stable β-turn structures exist in capped dipeptides (*trans*-carboxyl tripeptide analogues)[70−72] besides β-strand and γ-turn structures, while only β-strand and γ-turn structures can be found in capped amino acids (*trans*-carboxyl dipeptide analogues).[73]

The φ angles of the middle glycine residue in GGG, TGG, and MGG can be both positive (as in ggg3) and negative (as in ggg5). The φ angles of the tyrosine residue in GYG and the tryptophan residue in GWG adopt only negative values. This is in agreement with the finding that, among the naturally occurring amino acids, only the glycine residue can adopt positive φ angles in proteins due to the laevogyrate chirality of all other naturally occurring amino acids.

3.3. Efficient Determinations of Tripeptide Structures.

Similar to the structural relationships between dipeptide and amino acid conformers,[22] there are structural connections among the conformations of tripeptides and their constituting dipeptides and amino acids. That is, stable amino acid conformers that have similar conformations to the three residues in stable tripeptide conformers can be found. Similarly, stable dipeptide conformers with similar conformations to either N-terminus or C-terminus two-residue fragments in stable tripeptide conformers can be located. Figure 7 shows

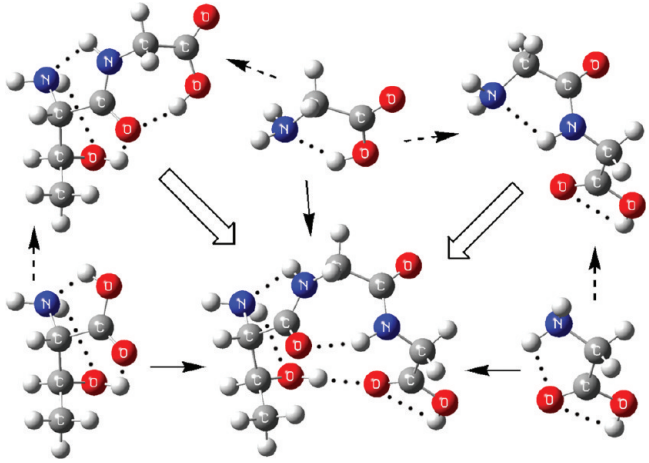


Figure 7. Structural similarities among conformations of tripeptides, dipeptides, and amino acids; similarities between amino acids and dipeptides, between amino acids and tripeptides, and between dipeptides and tripeptides are indicated by dashed arrows, solid arrows, and hollow arrows, respectively.

such structural connections using conformer tgg5 as an example. The observation points to a possible strategy to find all important tripeptide conformers by properly combining structures of parent fragments followed with geometry optimizations, just as the dipeptide structure determination strategy proposed before.[22] Conceivably, the trial tripeptide conformers can be constructed from conformers of its constituting segments in the following ways: (a) combining conformers of the three constituting amino acids (referred to hereafter as a "1+1+1" form); (b) combining conformers of the N-terminus part of the dipeptide and the C-terminus amino acid (referred to hereafter as a "2+1" form); and (c) combining conformers of the N-terminus amino acid and the C-terminus part of the dipeptide (referred to hereafter as a "1+2" form). As dominant tripeptide conformers have a peptide bond in the *trans*-configuration, only amino acid and dipeptide conformers with a *trans*-carboxyl group need to be considered for the N-terminus part of the amino acid or dipeptide and the middle amino acid in the tripeptide. Because the number of eligible amino acid or dipeptide conformers is relatively small, the number of trial tripeptide structures thus generated is much smaller than that in a full conformational search. Therefore, all three combining forms described above will be much more efficient than the thorough search method if they are capable of finding all important tripeptide conformers.

As discussed above in section 3.2, new structural features that are absent in dipeptides are becoming available in tripeptides. Straightforward extension of the dipeptide structure construction method is unlikely to be sufficient for finding all important tripeptide conformers. New concepts that can enable

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

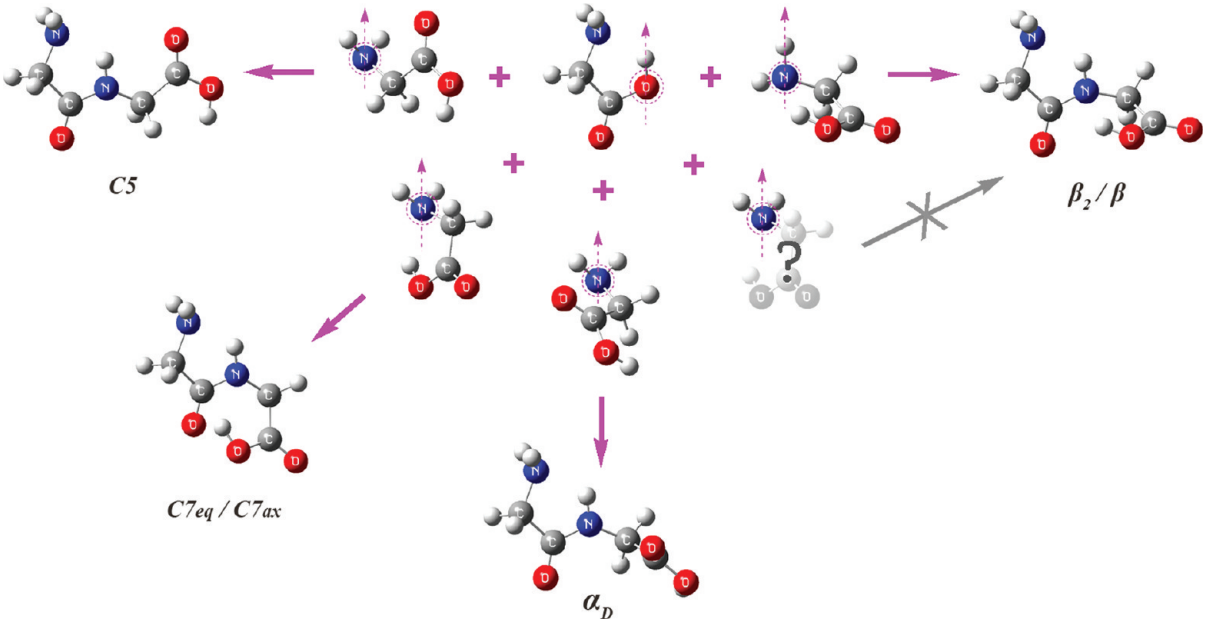


Figure 8. Sketch for constructing different structures of glycine dipeptide.

descriptions of new structural features need to be introduced into the dipeptide structure determination method for it to be applicable for tripeptides.

We may start out by discussing the emerged issues in the 1+1+1 form of the trial conformer construction. In this construction form, the first two amino acid conformers are jointed together to form a dipeptide structure, and then, an additional amino acid conformer is added to form the final trial tripeptide conformer. As the combination methodology determines the placement of one residue relative to the other, the secondary structure of the middle residue in a trial tripeptide conformer is thus determined when the first two amino acid conformers are joined together in a construction. Therefore, here, we will discuss the ability of previous construction methodology to generate dipeptide conformations with different C-terminus secondary structures. For convenience of reference, we summarize the dipeptide structure construction method here. Trial structures for a dipeptide X−Y are constructed by the combination of the low-energy conformers of its parent X (N-terminus) and Y (C-terminus) amino acids. The joining is formed by superposition of the hydroxyl oxygen atom in the X conformer and the amino nitrogen atom in the Y conformer with the hydroxyl O−H bond in the X conformer located in the plane containing the two amino N−H bonds in the Y conformer and bisecting the angle of H−N−H. A trial X−Y dipeptide structure is obtained after eliminating the hydroxyl oxygen atom and the two amino hydrogen atoms in the superposed position, as shown in Figure 8.

Taking glycine dipeptide shown in Figure 8 as an example, the previous used combination method has the ability to generate trial dipeptide conformers whose C-terminus residue adopts C5, C7eq/C7ax, and α_D structures. Furthermore, the combination method was verified to successfully locate all important dipeptide conformers as conformers with other secondary structures were energetically disfavored. However, such a method fails to generate a dipeptide conformer whose C-terminus residue has a β_2/β structure that is needed in order to generate a trial tripeptide conformer whose middle residue has such a structure. This is because the combination method prefers closer contact between the oxygen rather than nitrogen in the first peptide bond unit and the hydrogen in the second peptide bond unit and trends to form a C7eq/C7ax structure for the tripepdide. One alternative way to generate trial conformers with the C-terminus β_2/β structure is by superposition of the hydroxyl O−H bond in the first residue and one N−H bond of the amino group in the second residue and then eliminate the unwanted atoms, as shown in the top right part of Figure 8. Such a combination methodology is successful in finding conformers with β_2/β structure but is not economical as it results in a significant increase of the number of trial conformers required. For example, if N trial conformers are required in order to locate the important dipeptide conformers, then $2N$ more trial conformers will be generated by the new combination method (as there are two N−H bonds in an amino group) in order to locate dipeptide conformers whose C-terminus residue has β_2/β structure. Thus, a different way is needed to keep the method efficient.

Fortunately, after examining all important tripeptide conformers, we find that for a conformer whose middle residue has β_2/β structure, one can always find a corresponding stable conformer whose middle residue has C7eq/C7ax structure while keeping similar structures for other parts (e.g., conformers gyg4 and gyg7). Thus, a conformer whose middle residue has β_2/β structure may be obtained by optimizing a trial conformer generated by rotating the middle C_α−C bond of a corresponding conformer whose middle residue has C7eq/C7ax structure. As only a small number of conformers that were obtained from optimizing N trial conformers generated by the combination method need to be treated with the C_α−C bond rotation to get the β_2/β structures, the combined approach will be more efficient than that of directly optimizing $3N$ trial conformers.

The above analysis is also applicable to the 1+2 form of trial conformer construction. For the 2+1 form of construction, available structures of the middle residue in a tripeptide trial conformer are solely determined by available C-terminus structures of the stable dipeptide conformers used in

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

construction if no other operation is performed. Thus, if one populated tripeptide structure is not present in stable dipeptide conformers, the construction method will fail. Indeed, for some dipeptides, no stable conformers whose C-terminus residues have α_D structure can be found. Thus, in order to get a trial tripeptide conformer whose middle residue has α_D structure, similar rotation operation as that describe above is needed. In summary, despite the different construction forms, all secondary structure features of important tripeptide conformers can be located by optimizing trial conformers generated by the dipeptide construction method proposed before with additional rotation operations performed on the middle C_α−C bond.

The above-discussed middle C_α−C bond rotation operation is an efficient way to find important tripeptide conformers with the desired structural features of the middle residue. It can also be used to increase the efficiency of finding important tripeptide conformers with desired structural features of the N-terminus amino group. For example, each glycine conformer has a mirror symmetric conformer with respect to the placement of the amino group, and these two conformers are energetically degenerated. When forming a peptide with other chiral amino acids, such symmetry breaks down, and thus, both glycine mirror conformers should be considered in the construction of trial conformers. For example, conformers gyg3 and gyg5 shown in Figure 2 differ only in the placement of the amino group, and one has an amino group on the left side of the peptide bond plane while the other one is on the right side. However, if an N-terminus C_α−C bond rotation step is adopted after the construction step, then only one out of two glycine mirror conformers is needed in the construction step, which will efficiently reduce the number of calculations.

In short, the method for finding important tripeptide conformers has two steps. First, the trial tripeptide conformers are generated by joining constituting amino acid(s) or dipeptide conformers based on the dipeptide structure construction strategy proposed before.[22] These trial conformers are then optimized to get the stable tripeptide conformers. Second, a set of low-energy conformers are processed by rotating the N-terminus C_α−C bond, and a subset of conformers whose middle residue has C7eq/C7ax structure are processed by rotating the middle C_α−C bond by suitable degrees. The N-terminus rotation is necessary for chirality consideration, and the middle C_α−C bond rotation is required for finding conformers with β_2/β and α_D structures. Such a new set of trial conformers is then optimized to yield additional important tripeptide conformers.

3.4. Performance of the Proposed Tripeptide Structure Search Method. The trial conformer construction methods in the forms of 1+1+1, 2+1, or 1+2 defined above are verified to produce all important conformers that were obtained by thorough conformational searches of GGG, GYG, GWG, TGG, and MGG. The applicability of the proposed tripeptide conformation search methods to other tripeptides can be rationalized in a way similar to that for the dipeptide conformation search method, as summarized below.

There are basically four types of intramolecular interactions in a tripeptide, the backbone−side chain interaction within one residue, the backbone−backbone interaction between residues, the backbone−side chain interaction between residues, and the side chain−side chain interaction between residues. The four types of interactions should be properly balanced in an important conformer. For an efficient search method, we should use as small amount of trial conformers as possible to get all important conformers. Thus, a good trial structure in this sense should bear some features that are representative of the four types of interaction, which will be well-balanced in the following optimization procedure to target the important conformer. Therefore, next, we will discuss if our method can produce trial conformers with sufficient considerations of the four types of interactions.

First, the backbone−side chain interaction in one residue (or a two-residue fragment) is inherently considered because stable amino acid or dipeptide conformers are used in the trial conformer construction step. Second, information about the backbone−backbone interaction is presented in glycine tripeptide conformers. As all low-energy glycine tripeptide conformers are shown to be obtainable from the constructed trial conformers, it means that the backbone−backbone interaction is fully reflected in the construction process. The argument for having taken the possible backbone−side chain and side chain−side chain interactions between residues into consideration is less rigorous but can be illustrated with some concrete example. For example, conformer tgg5 shown in Figure 7 has a H-bond formed between the C-terminus backbone oxygen atom and the N-terminus side-chain hydroxyl group. This conformer can be obtained by optimizing a trial conformer constructed with a stable threonine conformer shown in Figure 7, which has a H-bond formed between the backbone oxygen and side-chain hydroxyl group. Therefore, optimization of the trial conformer breaks the local H-bond interaction within one residue and forms a new inter-residue interaction. That is, the geometry optimization performed on trial conformers can delicately tune the intramolecular interactions to form a new interaction between residues and produce low-energy configurations.

If we denote the three residues in a tripeptide as X, Y, and Z and the numbers of low-energy conformers required in the trial conformer construction step as $N(X)$, $N(Y)$, and $N(Z)$ for the 1+1+1 form, $N(XY)$ and $N(Z)$ for the 2+1 form, and $N(X)$ and $N(YZ)$ for the 1+2 form, then the total number of trial conformers generated in the construction step, N_{T1}, will be $N(X) \times N(Y) \times N(Z)$, $N(XY) \times N(Z)$, and $N(X) \times N(YZ)$ for the method in the 1+1+1, 2+1, and 1+2 forms, respectively. If the number of the lowest-energy tripeptide conformers used for the rotation step performed on the N-terminus and middle C_α−C bond is N_R and N_{RC} (N_{RC} is the number of conformers with C7eq/C7ax structure among the total of N_R conformers), respectively, then the number of new trial conformers generated would be $N_{T2} = N_R + 4N_{RC} \leq 5N_R$. For simplicity, N_{T2} is taken as $5N_R$ in the following discussions. Therefore, the total number of unique trial conformers generated or the total number of geometry optimizations required is $N_T = N_{T1} + N_{T2} = N_{T1} + 5N_R$. According to the experience gained in the dipeptide structure study, a value of about 20 for $N(X)$, $N(Y)$, or $N(Z)$ is sufficient for small amino acids, and the value may be increased up to 60 for larger amino acids. Because the 2+1 and 1+2 forms of the tripeptide construction method are in the same spirit as those for the dipeptide method, $N(XY \text{ or } YZ)$ is expected to be of the same magnitude of $N(X)$ suggested for larger amino acid as 60. Adding a factor of 2 as the safe margin, a value of 120 should be sufficient for $N(XY)$ or $N(YZ)$. N_R is expected to be similar for either 1+1+1, 2+1, or 1+2 construction forms, and its value should be similar to $N(X)$. Overall, the maximum number of geometry optimizations is expected to be $N_T = 120 \times 60 + 5 \times 60 = 7500$ for the 2+1 or 1+2 construction form. However, on average, N_T is expected to

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

be 60 × 20 + 5 × 20 = 1300 for the 2+1 or 1+2 form and 20 × 20 × 20 + 5 × 20 = 8100 for the 1+1+1 form. Furthermore, if we take some other specific conformer selection criteria into account, such as only conformers with *trans*-carboxyl configuration need to be used for the N-terminus and middle residues, N_T will be much smaller than the above estimation. Generally speaking, the proposed tripeptide conformer search method in all three forms can be orders of magnitude more efficient than the thorough search method. Moreover, the methods in the 2+1 or 1+2 form will be more efficient than that in the 1+1+1 form.

Table 1 shows the performance of our construction method in the three forms, 1+1+1, 2+1, and 1+2, to locate all important conformers of the five tripeptides studied above. N(X or Y or Z), N(XY or YZ), and N_R are determined based on trying to minimize N_T and will not be representative when the thorough searches are not performed (i.e., the results are not known a priori). However, Table 1 does provide some information on the efficiency of the proposed methods. Moreover, Table 1 provides useful information on how to choose between the 2+1 and 1+2 construction forms. As shown in Table 1, the 1+2 form is more efficient than 2+1 for GGG, TGG, and MGG, while it is opposite for GYG and GWG. The results can be understood as the following; only GY or GW conformers with *trans*-carboxyl configurations are required in the 2+1 form of the construction method for GYG and GWG, while both *cis*- and *trans*-carboxyl YG or WG conformers are needed if the 1+2 form of construction method is used. But, for TGG and MGG, using the 1+2 form of the construction method will benefit from the fact that the number of low-energy GG conformers is far smaller than that of TG or MG conformers. Therefore, we recommend one choose the 2+1 or 1+2 form by first trying to divide the tripeptide into fragments with similar sizes, for example, use T+GG (1+2) instead of TG+G (2+1) for TGG, and second adopting the 2+1 form when the resulting dipeptide fragments in the two construction forms are of similar length and complexity, for example, dividing GYG into GY+G (2+1) instead of G+YG (1+2).

Table 1. Numbers of Low-Energy Amino Acid (N(X), N(Y), and N(Z)) and Dipeptide (N(XY) and N(YZ)) Conformers Required in the Trial Conformer Construction Step, Numbers of Low-Energy Tripeptide Conformers (N_R) Required in the Rotation Step, and Total Number of Geometry Optimizations Needed for Producing All Important Tripeptide Conformers

XYZ		GGG	GYG	GWG	TGG	MGG
1+1+1	N(X)	3	3	3	6	9
	N(Y)	3	12	13	3	2
	N(Z)	5	1	5	5	5
	N_R	2	24	31	8	1
	N_T	55	156	350	130	95
2+1	N(XY)	16	51	43	29	20
	N(Z)	5	1	5	5	5
	N_R	0	0	9	1	1
	N_T	80	51	260	150	105
1+2	N(X)	3	3	3	6	9
	N(YZ)	12	38	60	12	8
	N_R	2	51	31	8	1
	N_T	46	369	335	112	77

3.5. Applications of the Proposed Method to the Previously Studied Tripeptides. To further validate the proposed method, the method was used to find conformers of the five tripeptides that had been systematically searched before by other groups.[41−45] For comparison purpose, the constructed trial conformers were optimized at the respective levels of theory used in the former works as well as for the single-point energy calculations (MP2/B3LYP for YGG and RI-MP2 for the other four tripeptides, FGG, WGG, GFA, and GGF).

The trial conformer construction method in the 1+2 form was applied to YGG according to the recommendation proposed above. 600 YGG trial conformers were generated by combining the 20 most stable tyrosine conformers with

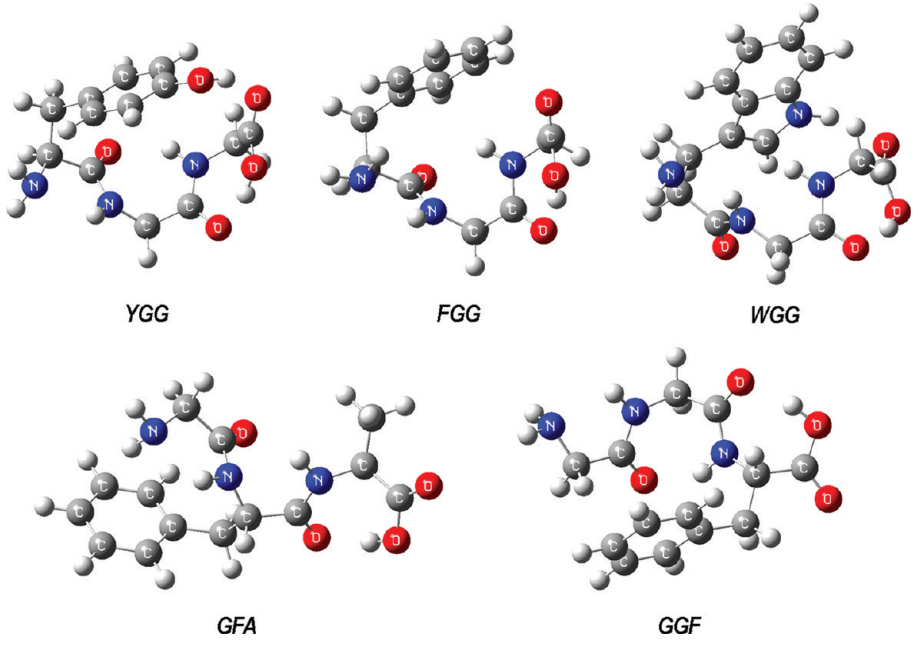


Figure 9. The most stable new conformers of YGG, FGG, WGG, GFA, and GGF tripeptides located by our proposed method.

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269−2283

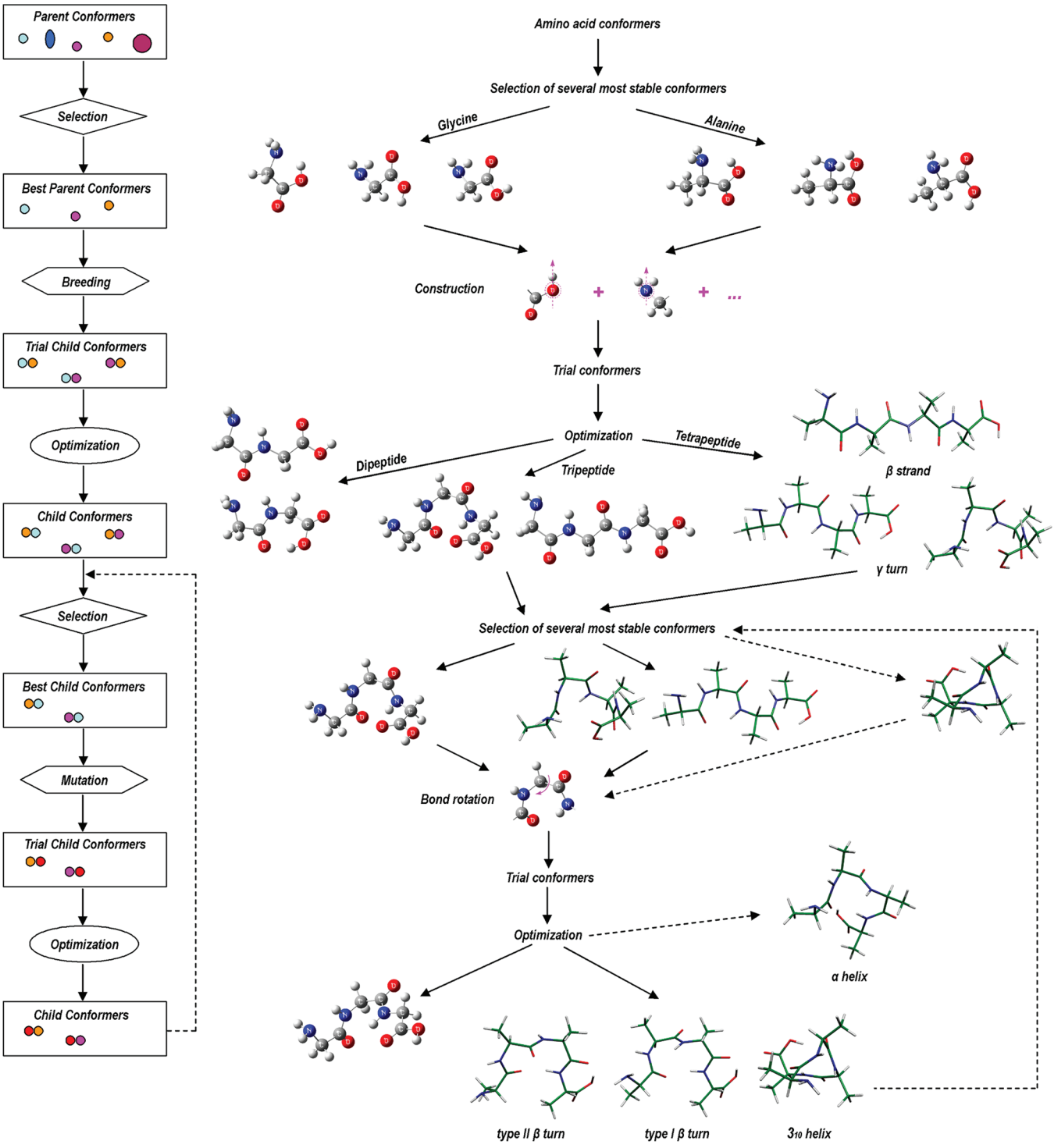


Figure 10. Flowchart for a typical genetic method and its correspondence to our method using glycine dipeptide, glycine tripeptide, and alanine tetrapeptide as examples (the alanine tetrapeptide structure is shown with a stick representation for a clear view of its secondary structure; the second round of mutation is indicated by dashed lines).

trans-carboxyl configurations and the 30 most stable GG conformers (though 60 conformers were recommended for general dipeptide residues, 30 GG conformers are found to be sufficient in all cases due to the simplicity of the GG fragment). After optimization of the 600 trial conformers, the 60 lowest-energy YGG conformers thus obtained were subjected to the rotation operation performed on the N-terminus and middle C_α–C bonds. As a result, 14 conformers were found to be within 4 kcal/mol of the global minimum considering the total energy at the MP2/6-31+G(d)//B3LYP/6-31+G(d) level. In addition to the conformers located previously,[74] three new conformers were found (see Table S6 in the SI for details). The structure of the most stable new conformer is shown in Figure 9. Compared to the large number of geometry optimizations and complicated strategy used in the previous work,[41] our method is verified to be more efficient and more reliable.

dx.doi.org/10.1021/jp207807a | J. Phys. Chem. B 2012, 116, 2269–2283

Similar to YGG, the 1+2 form of construction procedure was used to generate trial conformers of FGG and WGG. 600 trial conformers were generated for WGG, while only 570 trial conformers were generated for FGG as only conformers with *trans*-carboxyl configuration were used for the N-terminus residue and there was only a total of 19 *trans*-carboxyl phenylalanine conformers. After geometry optimizations, the 60 lowest-energy conformers of FGG and WGG were used in the bond rotation procedure to generate new trial conformers. As a result, a total of 14 FGG conformers were found to be within a 3.5 kcal/mol electronic energy range of its global minimum. For WGG, 10 conformers were found in the 2.5 kcal/mol range. In addition to conformers located before,[42,43] two new FGG and three new WGG conformers in the respective energy range were found. Figure 9 shows the structures of the lowest-energy new conformers missed in previous studies. Details about the energies of the 14 FGG conformers and 10 WGG conformers can be found in the SI (Tables S7 and S8). The new results may affect the theoretical interpretation of the experimental observations. For example, the new WGG conformer (conformer 4) shown in Figure 9 has a similar structure as conformer wgg02 identified before. As they only differ in the placement of the amino group, similar IR spectrum characteristics can be predicted. Therefore, both conformer wgg02 and conformer 4 can be assigned to the experimental structure *a* based on the IR spectrum.[43]

The trial conformers of GFA and GGF were generated with the construction method in the 2+1 form according to the above proposed recommendation; 780 trial conformers were generated for GFA with the 60 most stable GF conformers with *trans*-carboxyl configuration and all 13 stable alanine conformers. For GGF, 520 trial conformers were generated with all 26 *trans*-carboxyl GG conformers and the 20 most stable phenylalanine conformers. The 60 most stable GFA and GGF conformers thus obtained were used in the rotation procedure to generate new trial conformers. Finally, 17 GFA conformers and 9 GGF conformers were found in 2 kcal/mol ranges from the global energy minima of GFA and GGF, respectively. In addition to all of the conformers identified before,[44,45] one new conformer is found for both GFA and GGF (Figure 9), and their RI-MP2/cc-pVTZ energies are 1.76 and 1.52 kcal/mol above their respective global minimum (see Tables S9 and S10 in SI).

Summarized briefly, the validations performed on the five tripeptides studied before shows that our proposed method for searching tripeptide conformers has both high efficiency and high reliability. The total number of geometry optimizations required is less than 1000 for all five tripeptides, which is far smaller than that required by the systematic search method or that used in the previous studies. However, the new method is not only capable of finding all conformers identified before but also has the ability to locate low-energy conformers missed in the previous studies.

3.6. Method Generalization and Its Applicability to Larger Peptides. Our proposed method of finding stable tripeptide conformers actually shares the same basic idea with the genetic algorithm.[75,76] The construction and rotation step in our method corresponds to the breeding and mutation operations in a genetic algorithm, respectively. For the convenience of showing the correspondence, flowcharts of a modified version of the genetic algorithm and our proposed method are sketched on the left and right parts of Figure 10, respectively.

The genetic algorithm has been long used for conformational searches of various molecular systems including atomic clusters,[77,78] alkanes,[79] carbohydrates,[80] amino acids,[81] peptides,[18,19] and even proteins.[20,21] A genetic algorithm applied for finding stable conformers can be outlined as follows: The first step is to choose the best parent conformers according to a prescribed rule. The second step is to generate trial child conformers based on the structural information of their parent conformers by a given breeding strategy, and then, stable child conformers are obtained by geometry optimization. Next, some child conformers are selected according to some criteria, and they are mutated by a designated mutation strategy in order to introduce new structural information that is absent in their parent conformers. The breeding, mutation, and geometry optimization processes can be repeated until the results for the lowest-energy conformers are converged. As the results can be improved systematically, the genetic algorithm is conceptually attractive. However, the success of a genetic algorithm in practical applications is rather limited. The reason may be attributed to the lack of knowledge about what structures should be chosen for breeding and mutation and how to perform the breeding and mutation effectively. As there is no generally trustworthy formula to set up the key parameters controlling the whole process, performance of a user-dependent genetic algorithm can differ widely. A broadly successful genetic algorithm should rely on some specific yet generally applicable parameter definitions, which are the basis for its high efficiency and reliability.

As shown in Figure 10, each step in our proposed method (right flowchart) corresponds to a step in the genetic algorithm (left flowchart), and our method can be viewed as a specific application of the general genetic algorithm. Moreover, our method removes the ambiguities when applying the genetic algorithm by providing definite structure selection rules as well as breeding and mutation operations in specified forms. For example, structures used for breeding are the lowest-energy conformers of the parent residues, and more selectively, only conformers with *trans*-carboxyl configurations are chosen for the N-terminus residue. The breeding step is accomplished by a specified method of combining the parent conformers, and no iterating breeding cycles are required. Mutations are performed on the lowest-energy conformers obtained in the previous breeding step, and the mutation specifically refers to the C_α−C bond rotation operations. As the above validations demonstrate the efficiency and reliability of our proposed method, the correspondence between our method and the genetic algorithm that provides critical hints on what and how to breed and mutate may be helpful for the development of a more general genetic algorithm based method that can be applicable to other oligopeptides.

As preliminary exploration of the above idea, we here show how to apply our method to different peptide systems in the context of the genetic algorithm. As shown in Figure 10, a number of low-energy amino acid conformers are served as the parent conformers, with the restriction of using only *trans*-carboxyl conformers for the N-terminus residue. Different parent conformers are joined together to form trial child conformers (breeding), and stable conformers are obtained by the geometry optimizations. For dipeptides, the above process was shown to be enough to reproduce all important conformers in our test cases.[22] For tripeptides, the N- or C-terminus conformers mentioned above could be replaced with the corresponding dipeptide conformers, and the same process

follows. Moreover, the C_α−C bonds of the low-energy child conformers are further rotated to yield new trial structures (mutation) so that all important tripeptide conformers can be located. However, no iterating mutation cycles are required for the tripeptide cases.

To see the feasibility of the method for the structural determination of larger peptides, results for alanine tetrapeptide are described briefly here. For convenience, the trial structures were constructed in the 1+1+1+1 fashion. As indicated in Figure 10, starting from the three most stable alanine conformers with *trans*-carboxyl configuration, a number of trial alanine tetrapeptide structures were constructed (breeding). After optimizing these trial structures, stable alanine tetrapeptide conformers (stick representation) with γ-turn and β-strand secondary structures were located. The two γ-turn conformers were further treated with C_α−C bond rotations to generate new trial structures (mutation). After geometry optimizations, conformers with β-turn (both type I and type II) and 3_{10} helix structures were produced. The α-helix structure, which is highly populated in large peptides, can be found by applying one more rounds of mutation on the 3_{10} helix conformer followed by geometry optimization (shown with dashed arrows). Therefore, peptide conformers with important secondary structures such as strands, helices, and turns can be successfully located by our proposed method performed in a genetic algorithm fashion.

The above results are not difficult to understand. As described above, only breeding is needed to locate all important dipeptide conformers. This can be understood as all structural features of stable dipeptide conformers are already reflected in some way by the joining of parent amino acid conformers. However, complexity grows with the peptide size, and some structural characteristics of tripeptides cannot be fully considered in the breeding step or the joining of parent conformers. Consequently, the mutation step is needed to generate new structural features that are intrinsic for some stable tripeptide conformers but are not recognized in the breeding step. As the complexity grows further for larger peptides, it is natural to expect that more mutation steps will be required to produce various kinds of structural features that are necessary for finding all important peptide conformers. In other words, iterating mutation steps may be required for searching the tetrapeptide or larger peptide conformers. This seems to be true by considering the fact that many highly populated secondary structures such as helices and turns for large peptides are not favored or even not possible for small peptides and cannot be constructed by joining conformers of constituting short fragments. However, these structures will have high stability in large peptides due to the formation of H-bonds between distant residues as accommodated by the conformational flexibilities of the long peptide chains. The significance of our proposed method is to point out an efficient and reliable breeding and mutation framework, though more tests are required.

4. CONCLUSIONS

We have performed thorough searches on the PESs of five tripeptides, and their structural characteristics were systematically analyzed. Entropic energy contribution is found to play a very important role in determining the stabilities of tripeptide conformers. Compared to dipeptide conformers, the tripeptide conformer whose middle residue has β_2/β structure becomes important. This is consistent with the tendency that more secondary structures will emerge in larger peptides.

Similar to dipeptides studied before, connections between structures of tripeptide and parent fragments were found, and a method for searching important tripeptide conformers based on their parent dipeptide and amino acid conformers is proposed. The construction step used to generate trial tripeptide structures in the method can be performed in the 1+1+1, 2+1, and 1+2 forms, and guidance for how to choose between different construction forms for different cases is presented. Validations on previously studied tripeptides demonstrates that the method is not only more efficient but also more reliable than most common approaches as it provides the most complete ensembles of tripeptide conformers.

The proposed method for tripeptide conformational search is an extension of the existing dipeptide method with a new concept of bond rotation being introduced. The new concept links our method to the genetic algorithm. Consequently, our method is generalized under the framework of the genetic algorithm with specified breeding and mutation operations designed for high efficiency and reliability. The new method is further shown to be capable of locating different secondary structures for alanine tetrapeptide, providing preliminary proof that the proposed method may be applicable to larger peptides, though more validations are needed.

■ ASSOCIATED CONTENT

Supporting Information

Details regarding the relative energies and equilibrium distributions of stable conformers for the five currently studied and the five previously studied tripeptides, structure differences between BHandHLYP and M06-L optimized conformers as well as their structure alignments, and full citations of refs 56 and 57. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: zjlin@ustc.edu.cn.

■ ACKNOWLEDGMENTS

Z.L. acknowledges the financial supports from the National Science Foundation of China (11074233) and the State Key Development Program for Basic Research of China (2012CB215405). A.M. acknowledges financial support from the National Institutes of Health (GM051501) and computational support from the Pittsburgh Supercomputing Center and the NSF/TeraGrid computational resources. The authors thank Tanja van Mourik for providing the coordinates of their tripeptide conformers.

■ REFERENCES

(1) Wyttenbach, T.; Liu, D.; Bowers, M. T. *J. Am. Chem. Soc.* **2008**, *130*, 5993−6000.
(2) Xu, X.; Yu, W.; Huang, Z.; Lin, Z. *J. Phys. Chem. B* **2010**, *114*, 1417−1423.
(3) Weisbrich, A.; Honnappa, S.; Jaussi, R.; Okhrimenko, O.; Frey, D.; Jelesarov, I.; Akhmanova, A.; Steinmetz, M. O. *Nat. Struct. Mol. Biol.* **2007**, *14*, 959−967.
(4) Guillaume, M.; Ruud, K.; Rizzo, A.; Monti, S.; Lin, Z.; Xu, X. *J. Phys. Chem. B* **2010**, *114*, 6500−6512.
(5) Zhou, Y.; Duan, Y.; Yang, Y.; Faraggi, E.; Lei, H. *Theor. Chem. Acc.* **2011**, *128*, 3−16.

(6) Moult, J.; Fidelis, K.; Kryshtafovych, A.; Rost, B.; Tramontano, A. *Proteins* **2009**, *77*, 1−4.
(7) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol.* **2003**, *326*, 1239−1259.
(8) Yu, W.; Lin, Z.; Huang, Z. *ChemPhysChem* **2006**, *7*, 828−830.
(9) Huang, Z.; Lin, Z. *J. Phys. Chem. A* **2005**, *109*, 2656−2659.
(10) Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Haranczyk, M.; Gutowski, M. *J. Phys. Chem. A* **2006**, *110*, 12282−12291.
(11) Velikson, B.; Garel, T.; Niel, J.; Orland, H.; Smith, J. C. *J. Comput. Chem.* **1992**, *13*, 1216−1233.
(12) Ozkan, S. B.; Meirovitch, H. *J. Phys. Chem. B* **2003**, *107*, 9128−9131.
(13) Christen, M.; Van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *29*, 157−166.
(14) Schlund, S.; Müller, R.; Graβmann, C.; Engels, B. *J. Comput. Chem.* **2008**, *29*, 407−415.
(15) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Chem. Rev.* **1994**, *94*, 2183−2239.
(16) Corcho, F. J.; Filizola, M.; Pérez, J. J. *Chem. Phys. Lett.* **2000**, *319*, 65−70.
(17) Fujitani, N.; Shimizu, H.; Matsubara, T.; Ohta, T.; Komata, Y.; Miura, N.; Sato, T.; Nishimura, S. *Carbohydr. Res.* **2007**, *342*, 1895−1903.
(18) Le Grand, S. M.; Merz, K. M. Jr. *Mol. Simul.* **1994**, *13*, 299−320.
(19) Meza, J. C.; Judson, R. S.; Faulkner, T. R.; Treasurywala, A. M. *J. Comput. Chem.* **1996**, *17*, 1142−1151.
(20) Pedersen, J. T.; Moult, J. *Proteins* **1995**, *23*, 454−460.
(21) Sakae, Y.; Hiroyasu, T.; Miki, M.; Okamoto, Y. *Pac. Symp. Biocomput.* **2011**, *16*, 217−228.
(22) Yu, W.; Xu, X.; Li, H.; Pang, R.; Fang, K.; Lin, Z. *J. Comput. Chem.* **2009**, *30*, 2105−2121.
(23) Dayalan, S.; Gooneratne, N. D.; Bevinakoppa, S.; Schroder, H. *Bioinformation* **2006**, *1*, 78−80.
(24) Anishetty, S.; Pennathur, G.; Anishetty, R. *BMC Struct. Biol.* **2002**, *2*, 9.
(25) Tobias, D. J.; Brooks, C. L. III. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(26) Ishizuka, R.; Huber, G. A.; McCammon, J. A. *J. Phys. Chem. Lett.* **2010**, *15*, 2279−2283.
(27) Wang, Z.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699−1716.
(28) Fan, J.; Tang, M.; Qiao, L.; Liu, J.; He, L. *Amino Acids* **2010**, *39*, 685−697.
(29) Yang, S.; Cho, M. *J. Chem. Phys.* **2009**, *131*, 135102.
(30) Shemesh, D.; Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4899−4905.
(31) Mackerell, A. D. Jr.; Feig, M.; Brooks, C. L. III. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(32) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27−85.
(33) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.
(34) Soriano-Correa, C.; del Valle, F. J. O.; Munoz-Losa, A.; Galvan, I. F.; Martin, M. E.; Aguilar, M. A. *J. Phys. Chem. B* **2010**, *114*, 8961−8970.
(35) Pingitore, F.; Bleiholder, C.; Paizs, B.; Wesdemiotis, C. *Int. J. Mass Spectrom.* **2007**, *265*, 251−260.
(36) Haber, T.; Seefeld, K.; Kleinermanns, K. *J. Phys. Chem. A* **2007**, *111*, 3038−3046.
(37) Haber, T.; Seefeld, K.; Engler, G.; Grimme, S.; Kleinermanns, K. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2844−2851.
(38) Dunbar, R. C.; Steill, J. D.; Polfer, N. C.; Oomens, J. *Int. J. Mass Spectrom.* **2009**, *283*, 77−84.
(39) Shemesh, D.; Hattig, C.; Domcke, W. *Chem. Phys. Lett.* **2009**, *482*, 38−43.
(40) Clavaguera, C.; Piuzzi, F.; Dognon, J. P. *J. Phys. Chem. B* **2009**, *113*, 16443−16448.
(41) Toroz, D.; van Mourik, T. *Mol. Phys.* **2007**, *105*, 209−220.
(42) Řeha, D.; Valdés, H.; Vondrášek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803−6817.
(43) Valdés, H.; Řeha, D.; Hobza, P. *J. Phys. Chem. B* **2006**, *110*, 6385−6396.
(44) Valdés, H.; Spiwok, V.; Řezáč, J.; Řeha, D.; Abu-Riziq, A.; de Vries, M. S.; Hobza, P. *Chem.—Eur. J.* **2008**, *14*, 4886−4898.
(45) Valdés, H.; Pluháčková, K.; Pitonák, M.; Řezáč, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747−2757.
(46) Zhang, M.; Lin, Z. *J. Mol. Struct.: THEOCHEM* **2006**, *760*, 159−166.
(47) Császár, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9568−9575.
(48) Zhang, M.; Huang, Z.; Lin, Z. *J. Chem. Phys.* **2005**, *122*, 134313−1−7.
(49) Yu, W.; Liang, L.; Lin, Z.; Ling, S.; Haranczyk, M.; Gutowski, M. *J. Comput. Chem.* **2009**, *30*, 589−600.
(50) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
(51) Roy, A. K.; Hart, J. R.; Thakkar, A. J. *Chem. Phys. Lett.* **2007**, *434*, 176−181.
(52) Roy, A. K.; Hu, S.; Thakkar, A. J. *J. Chem. Phys.* **2005**, *122*, 074313.
(53) Toroz, D.; van Mourik, T. *Mol. Phys.* **2006**, *104*, 559−570.
(54) Neese, F. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, in press.
(55) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *181*, 1477−1489.
(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision B.01; Gaussian Inc.: Wallingford, CT, 2010.
(57) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. *Gaussian 03*, revision B.04; Gaussian Inc.: Pittsburgh, PA, 2003.
(58) Cao, J.; van Mourik, T. *Chem. Phys. Lett.* **2010**, *485*, 40−44.
(59) Zhang, K.; Cassady, C. J.; Chung-phillips, A. *J. Am. Chem. Soc.* **1994**, *116*, 11512−11521.
(60) Wang, P.; Wesdemiotis, C.; Kapota, C.; Ohanessian, G. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 541−552.
(61) Strittmatter, E. F.; Williams, E. R. *Int. J. Mass Spectrom.* **1999**, *185*, 935−948.
(62) Möhle, K.; Hofmann, H. *J. Mol. Model.* **1998**, *4*, 53−60.
(63) Oka, M.; Montelione, G. T.; Scheraga, H. A. *J. Am. Chem. Soc.* **1984**, *106*, 7959−7969.
(64) Doig, A. J.; Sternberg, M. J. *Protein Sci.* **1995**, *4*, 2247−2251.
(65) Zhang, J.; Liu, J. S. *PLoS Comput. Biol.* **2006**, *2*, e168.
(66) Ramachandrana, G. N.; Ramakrishnana, C.; Sasisekharana, V. *J. Mol. Biol.* **1963**, *7*, 95−99.
(67) Császár, A. G.; Perczel, A. *Prog. Biophys. Mol. Biol.* **1999**, *71*, 243−309.
(68) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L. III; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989−5997.
(69) Chin, W.; Dognon, J.; Piuzzi, F.; Tardivel, B.; Dimicoli, I.; Mons, M. *J. Am. Chem. Soc.* **2005**, *127*, 707−712.
(70) Chin, W.; Dognon, J.; Canuel, C.; Piuzzi, F.; Dimicoli, I.; Mons, M.; Compagnon, I.; von Helden, G.; Meijer, G. *J. Chem. Phys.* **2005**, *122*, 054317.
(71) Chin, W.; Piuzzi, F.; Dognon, J.; Dimicoli, I.; Mons, M. *J. Chem. Phys.* **2005**, *123*, 084301.
(72) Chin, W.; Piuzzi, F.; Dimicoli, I.; Mons, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1033−1048.
(73) Chin, W.; Mons, M.; Dognon, J.; Mirasol, R.; Chass, G.; Dimicoli, I.; Piuzzi, F.; Butz, P.; Tardivel, B.; Compagnon, I.; von Helden, G.; Meijer, G. *J. Phys. Chem. A* **2005**, *109*, 5281−5288.
(74) Personal communication with Tanja van Mourik for the coordinates of their stable YGG conformers.
(75) Fraser, A. S. *Biometrics* **1959**, *15*, 158−159.
(76) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley Professional: Reading, MA, 1989.
(77) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368−1372.
(78) Deaven, D. M.; Ho, K. M. *Phys. Rev. Lett.* **1995**, *75*, 288−291.

(79) Goodman, J. M.; Nair, N. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 317−320.
(80) Rosen, J.; Miguet, L.; Pérez, S. *J. Cheminform.* **2009**, *1*, 16.
(81) Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11695−11707.