Chemical Physics Letters 610–611 (2014) 303–309

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

A fragment based step-by-step strategy for determining the most stable conformers of biomolecules

Hongbao Li[a,b], Zijing Lin[a,*], Yi Luo[a,b,**]

[a] Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui 230026, China
[b] Department of Theoretical Chemistry and Biology, School of Biotechnology, Royal Institute of Technology, S-106 91 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 8 May 2014
In final form 22 July 2014
Available online 27 July 2014

ABSTRACT

For biomolecules of increased size and flexibility, more efficient and reliable strategies are always needed to determine their stable low-energy conformers. Here, we propose a fragment based step-by-step strategy to search for the full conformational space of biomolecules. In this strategy, the molecule is divided into several fragments and each of them is systematically optimized in a step-by-step fashion. It can significantly reduce the computational cost without losing any accuracy as demonstrated by the conformer search of several representative di-/tri-/tetra-peptides. Such an approach will be very useful for finding the stable conformers of large biomolecules.

1. Introduction

Biomolecules, such as short peptide and protein, can have many local minima because of their very flexible conformations. The determination of the global or local minima has attracted considerable attention over the years. At the thermal equilibrium, due to the small energy difference among these local minima, many conformers can co-exist with different proportions [1]. It is found that these optimized structures in gas phase can occasionally be a reasonable alternative to mimic the structures in the continuum medium [2]. More importantly, the interpretation of different spectra of these biomolecules requires a correct description of these low-energy stable conformers [3–10], which have often been served as the references to understand the behavior and functionalities of more complicated systems.

For small molecules, like the 20 natural α-amino acids, the potential energy surface can be easily studied by the systematic search method that fully explores the entire conformational space [11–13]. It is expected that such an approach will become too time-consuming for larger systems with many internal rotational degrees of freedom. The large number of trial structures and the relatively small energy differences among them have made the theoretical prediction of the global minimum a very difficult task in general.

Most theoretical studies on large molecular systems mainly involve the stochastic search methods, such as the Monte-Carlo [14–17], the genetic algorithm [18,19], the simulated annealing [20–22] and the basin-hopping [23,24] methods. These are no doubt very effective methods, however, due to the random sampling in the generation of the trial structures, it is very difficult to make sure that the global minimum has been found by such methods [24]. One example is the conformational search for arginine molecule (Arg), which demonstrated that neither the "simple genetic algorithm" [19] nor the "force-field-based strategy" [25] was able to find all the most stable conformers. Instead, a so-called "step-by-step" strategy was employed to localize the stable conformers of gaseous arginine [26] and offered the most accurate structures up to now.

A similar stepwise approach was also used in the structure determination of gas-phase dipeptide Tyr-Gly [27] (Tyr = tyrosine, Gly = glycine). The main procedure of such a strategy contains several key steps: (1) The trial structures were generated by rotating only two or three rotational degrees of freedom of the molecule and then optimized; (2) New trial structures were generated by rotating the next rotational degree of freedom based on all the stable geometries obtained in the previous step and further optimized; (3) Repeating the whole process until all the left degrees of freedom were scanned. But there was no good explanation for the number of the internal rotational degrees of freedom chosen at the first step. Moreover, the whole procedure involves many trial structures and is quite difficult to operate in practice. The step-by-step approach

* Corresponding author.
** Corresponding author at: Department of Theoretical Chemistry and Biology, School of Biotechnology, Royal Institute of Technology, S-106 91 Stockholm, Sweden. Fax: +46 8 553 78 590.
E-mail addresses: zjlin@ustc.edu.cn (Z. Lin), luo@theochem.kth.se (Y. Luo).

http://dx.doi.org/10.1016/j.cplett.2014.07.054
0009-2614/

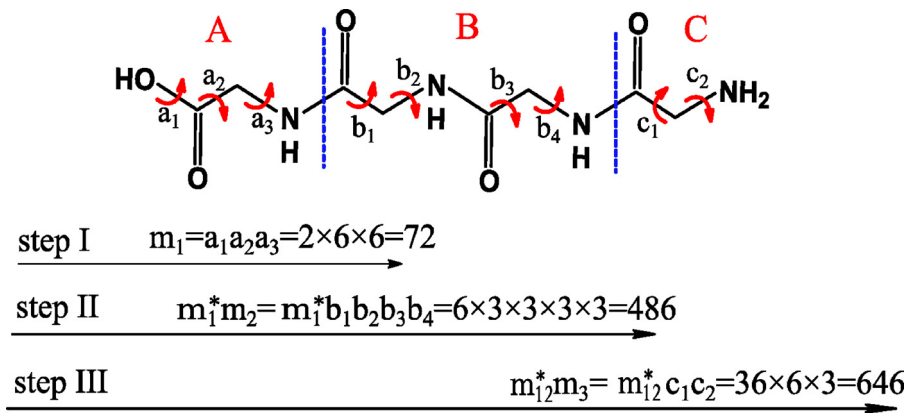

Figure 1. The fragment based step-by-step strategy applied to tetrapeptide Gly-Gly-Gly-Gly. Three fragments A–C, are created by cutting through two peptide bonds. Numbers a_1–a_3; b_1–b_4; c_1–c_2 refer to the internal rotational degrees of freedom of the three fragments with $a_1 = 2$; b_1–b_4 and $c_2 = 3$ and all others equal to 6. Labels m_1, m_2 and m_3 are numbers of trial structures in each fragment that can be used to generate the total trial structures. $m_1{}^{*\bullet}$ and $m_{12}{}^{*\bullet}$ are numbers of effective structures that selected out from the optimized results at step I and step II.

has the ambition of covering all possible structures with the brutal force, but its applicability is limited by the size of the molecules. One way to improve its efficiency is to introduce the concept of fragmentation.

The fragmentation methods have been extensively employed to effectively calculate electronic and geometric structures of molecules, as nicely summarized in a recent review [28]. The basic idea is to divide the molecule into several fragments and to optimize each fragment separately. The same philosophy can be implemented within the step-by-step approach. Instead of searching the entire system, we optimize each fragment in full conformation space. It is important to note that the key intramolecular interactions in and between different parts can be taken into account by the optimization process. In this letter, we demonstrate how this fragment based step-by-step approach works by using a tetrapeptide molecule as the example. The generality of this fragmentation method is also examined by applying it for extensive conformer search of several systems, including three dipeptides (Leu-Gly, Phe-Gly and Tyr-Gly), two tripeptides (Gly-Gly-Gly and Gly-Tyr-Gly) and two tetrapeptide (Gly-Gly-Gly-Gly and Gly-Tyr-Gly-Arg), where Leu = leucine and Phe = phenylalanine. Most of the results are compared with what were obtained before [27,29,30]. It is shown that the fragment based step-by-step strategy is not only more efficient, but also enables to find more stable conformers of large biomolecules than other methods.

2. Computational methods

Based on our experience, the initial structure of the molecules used for further manipulation should be relaxed in an unfolding manner. This means that only short-range intramolecular interactions are considered within the individual functional groups. All possible long-range interactions between different functional groups will be naturally introduced in the following structural rotating and optimizing processes.

The key problem associated with fragment based methods is how to properly divide the object molecule into different fragments. For large molecules in gas phase, the weak intramolecular interactions can have significant impact on the final structures of the molecule. Usually, as we mentioned above, only the short-range interaction needs to be considered in the initial structure. In this case, one can take special functional groups as independent fragments, especially the groups that have great importance in determining the structure and physicochemical properties of the molecule. For example, the amino group and the carboxyl group in a single amino acid should always be in the same fragment, since they are often connected by a hydrogen bond and responsible for some unique spectral features in the related IR or soft X-ray spectra [7–9]. The first rotation of the bonds also starts from such a fragment.

The basic idea behind the fragment based step-by-step strategy is simple. Once the fragments are decided within the molecule, we start to rotate every single bond in the fragments following a pre-defined order. However, in order to unambiguously present this strategy, we would like to use a tetrapeptide (Gly-Gly-Gly-Gly) molecule as an example to describe the whole operating procedure. As depicted in Figure 1, since the trans-peptide bonds are more favored in many studies [29,30], we can divide the molecule into three fragments, marked by A, B and C, respectively, through two peptide bonds. The rotational degrees of freedoms for each fragment are labeled as a_1–a_3, b_1–b_4 and c_1–c_2, respectively. The definition of the internal rotational degrees of freedom can be found in previous studies [11–13]. Generally, a dihedral angle can vary from 0° to 360° and in order to guarantee a complete search, usually an increment of 60° (120°) for the asymmetrical (symmetrical) dihedral angle is required [31,32]. For the C-OH group, only syn- and anti-periplanar arrangements, corresponding to 0° and 180° torsion, respectively, should be considered. The following steps are employed for the entire optimization.

(1) The three internal rotational degrees of the freedom in fragment A (a_1–a_3) are rotated first, while the fragments B and C keep unchanged. All generated trial structures ($m_1 = a_1a_2a_3 = 2 \times 6 \times 6 = 72$) are then optimized at HF/3-21G(d) level. Because according to our previous study on a series of dipeptides and tripeptides [29,30], the HF method correlated much better than the semiempirical method (PM3 or AM1) with the energy ordering results by higher levels of theoretical methods. Finally the unique structures (totally 6) from the fragment A are selected out and used as the effective structures ($m_1{}^{*\bullet} = 6$) for the next step.

(2) With each effective structure ($m_1{}^{*\bullet}$), we start to manipulate the structure of the fragment B by rotating b_1, b_2, b_3 and b_4 with the fragments A and C fixed. The generated structures ($m_1{}^{*\bullet}m_2 = m_1{}^{*\bullet}b_1b_2b_3b_4 = 6 \times 3 \times 3 \times 3 \times 3 = 486$) can be optimized at HF/3-21G (d) level to pick out the unique structures (totally 241) for the fragment B. Here we choose the low-energy conformers within 30 kJ/mol as the effective structures ($m_{12}{}^{*\bullet} = 36$) for the next step.