

Conformational analysis of peptides using Monte Carlo simulations combined with the genetic algorithm

Junmei Wang^a, Tingjun Hou^a, Lirong Chen^b, Xiaojie Xu^{a,*}

^a Department of Chemistry, Peking University Jiuyuan Molecular Design Laboratory, Peking University, Beijing 100871, China

^b Department of Technical Physics, Peking University, Beijing 100871, China

Abstract

In this paper, genetic algorithm (GA) was used to both sample the conformational spaces and thoroughly search the global conformations of peptides. δ -conotoxin PVIA, a peptide of 29 amino acids, was used to test our procedure. The results indicated that this procedure could not only successfully explore a set of the conformational spaces using a hybrid Monte Carlo simulations/GA minimization but also find the global conformations for most of peptides using pure GA minimization. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Genetic algorithm; Conformational analysis; Peptide; Monte Carlo simulations

Contents

1. Introduction	348
2. Method	348
2.1. Modeling conformation	348
2.2. Evaluating energy	348
2.3. Conformation searching	348
3. Results and discussion	350
Acknowledgements.	351
References	351

* Corresponding author

1. Introduction

Conformational analysis now becomes an important research branch in the field of computational chemistry. Most methods of conformational analysis can be grouped into two classes: one is systematic searching, another is random searching including Monte Carlo simulations and genetic algorithm (GA) minimization. Molecular dynamics and distance geometry are also being widely used now. It is believed that molecular dynamics and GA are the most efficient methods to explore the global conformation [1]. In order to sample the conformational spaces efficiently, some hybrid methods are proposed, such as Monte Carlo simulations/simulated annealing and hybrid Monte Carlo simulations/molecular dynamics. GA, which is an intelligent method between the stochastic and gradient methods, is now being widely used in conformational analysis [2–5]. It has not only overcome the potential barriers easily as random search, but also has good partial solutions survived. The strength of GA lies in its ability to handle large and diverse sets of variables. It is especially suitable for GA to explore the global conformations of peptides and proteins which have a great deal of variables (twist angles).

In this paper, we offered a procedure to explore the global conformation of peptides using GA minimization. A hybrid Monte Carlo simulation/GA minimization to sample the conformational spaces of peptides and proteins is also proposed. δ -conotoxin PVIA [6,7], a peptide of 29 amino acids was used to test our procedure and the results are encouraging.

2. Method

2.1. Modeling conformation

First, the initial structure of the peptide is modeled using 20 standard amino acids. Each amino acid has an index to inform the program which form of the secondary structure this residue prefers. These information can be obtained easily using some secondary-structure prediction programs. The initial conformation can also be modeled using randomly generated twist angles.

2.2. Evaluating energy

The following energetic terms are included in our program: Van der Waals energy, electrostatic energy, H-bond energy, torsional energy, disulfide bond energy and distance restrained energy. Other energetic terms are omitted for simplification. The restrained energy is calculated using a harmonic potential function. The force field is AMBER.

2.3. Conformation searching

A population of ‘chromosomes’ is first generated and the ‘gene’ is twist angle. Each ‘chromosome’ represents a conformation. The fitness score is the conformational energy; the lower the energy, the higher the fitness score. The three operations of GA which are crossover mutation and selection are repeatedly performed. The crossover operator randomly chooses two ‘chromosomes’ as parents to breed a pair of children maintaining some of their characteristics by dividing both parents at a randomly chosen point and then joining the pieces together. If the mutation probability exceeds a randomly generated number, mutation operator is performed by replacing a certain ‘gene’ in a ‘chromosome’ with a randomly generated one. The selection operator chooses ‘chromosomes’ in a manner which is weighted according to their fitness values; the fitter a ‘chromosome’, the more chances it has of being selected. Furthermore, a set of best-performing ‘chromosomes’ from one generation is frequently saved and survive unchanged into the next generation (the elitist strategy). Although GA is an outstanding stochastic search algorithm, it is still common that GA staggers in the local optima, a modification with the normal GA is that for each generation, a set of the lowest ‘chromosomes’ are replaced with randomly generated ones. We hope this combined method can efficiently overcome the deficiency of GA.

The convergence criteria of GA can be grouped into two classes: the inter-population and intra-population convergence. The inter-population convergence criterion monitors how much the average fitness of the population has changed since the last check. If it has not changed by user-defined itera-

tions then the GA is deemed to have converged [8]. The convergence criterion of our program belongs to this class. When the average fitness score of a population is not improved after certain iterations, say 100 iterations, the convergence is achieved. For a peptide with 50 residues, 1000–2000 iterations are enough to find the global conformation.

As for the hybrid Monte Carlo simulations/GA minimization procedure, certain steps of Monte Carlo simulation are primarily performed to equilibrate the system, then, GA is easily introduced into Monte Carlo simulations through performing GA minimization every each Monte Carlo steps.

In the case of δ -conotoxin PVIA, the randomly generated conformation was first submitted to the 2000-step GA minimization with an aim to explore the global conformations. Then a hybrid Monte Carlo simulation/GA minimization for conotoxin PVIA with one or two disulfide bonds restrained, respectively, was performed in order to understand the folding and deforming processes of this peptide. The force constant of the restrained energetic term is $100 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$. The Monte Carlo simulation steps to-

tal 50,000 and the first 10,000 were performed simply to equilibrate the system, then a 300-step GA minimization was performed every 2000 Monte Carlo simulation steps. The lowest energetic conformation of each GA minimization was stored, and finally, there was a total of 20 conformations saved. The native conformation is the lowest energetic conformation of the 20 ones.

As for GA, the population size and elite size of 'chromosomes', the mutation ratio, crossover ratio as well as the simulation steps play main roles in the minimization. The population size is usually three-fold or fourfold the number of variables. Elite size is about 5% to 10% of population size. In the case of δ -conotoxin PVIA, the population size is 300, the elite size is 15. The mutation ratio and crossover ratio are 0.05 and 0.35 in our procedure. By analogy with nature, mutation operator is generally invoked with a low frequency, the ratio is usually smaller than 0.1. The simulation steps may be changed for different molecules. For a typical peptide with 50 residues, the steps of pure GA minimization are about 1500. However, 300 steps are enough to perform a hybrid

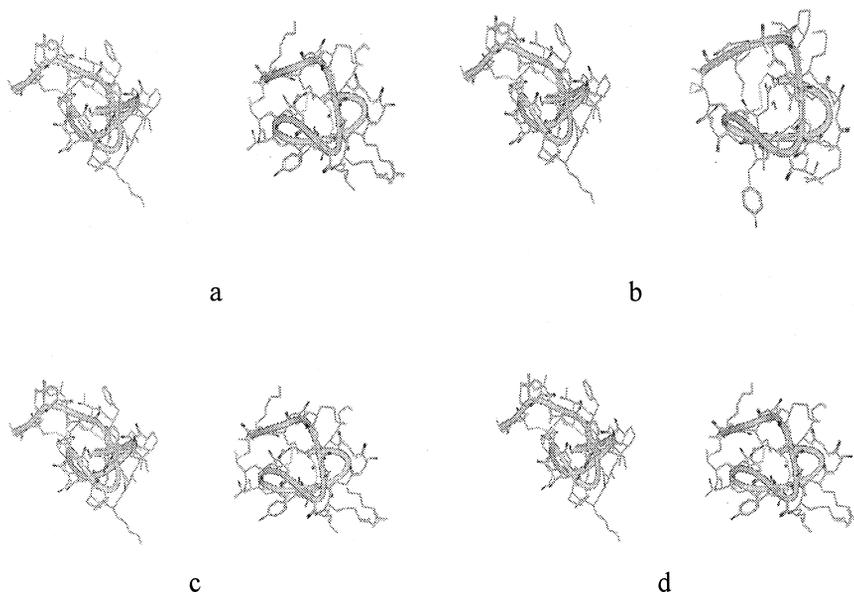


Fig. 1. Amino acid sequence of conotoxin PVIA. The natural form of this peptide can form three disulfide bonds (3–18, 20–22, 17–27). *O stand for hydroxyproline.

- a: IOMN (omega-conotoxin M V II C) [5], rmsd = 2.880 b: ICCO (omega-conotoxin GVIA), rmsd = 2.547
 c: IOMC (omega-conotoxin GXIA)[6], rmsd = 2.778 d: IOMG (omega-conotoxin MV II A)[7], rmsd = 2.879

Fig. 2. The superimposed conformations of the natural form with its homologous proteins. For each case, the left picture is the native conformation of δ -conotoxin PVIA, the right one is its homologous proteins. (a) IOMN (omega-conotoxin M V II C) [5], RMSD = 2.880; (b) ICCO (omega-conotoxin GVIA), RMSD = 2.547; (c) IOMC (omega-conotoxin GXIA) [6], RMSD = 2.778; (d) IOMG (omega-conotoxin MV II A) [7], RMSD = 2.879.

Monte Carlo simulation/GA simulation. The above mentioned parameters can be changed by the users.

3. Results and discussion

Fig. 1 shows the amino acid sequence of δ -conotoxin PVIA, the natural form of this peptide can form three disulfide bonds. Fig. 2 shows the superimposed conformations of the natural form with its homologous proteins, for example omega-conotoxin M V II C [9], omega-conotoxin GVIA, omega-con-

otoxin GXIA [10] and omega-conotoxin MV II A [11]. All of the root mean square distance (RMSDs) are reasonable. Our procedure of GA minimization to explore the global conformation is really a successful one. We also carried out hybrid Monte Carlo simulations/GA for conotoxin PVIA with partial disulfide bonds formed. Fig. 3 shows the global conformations of conotoxin PVIA with partial disulfide bonds formed. The first three configurations form only one disulfide bond, however the last three form two. We can presume the deforming process of this peptide according to the conformational energies of cono-

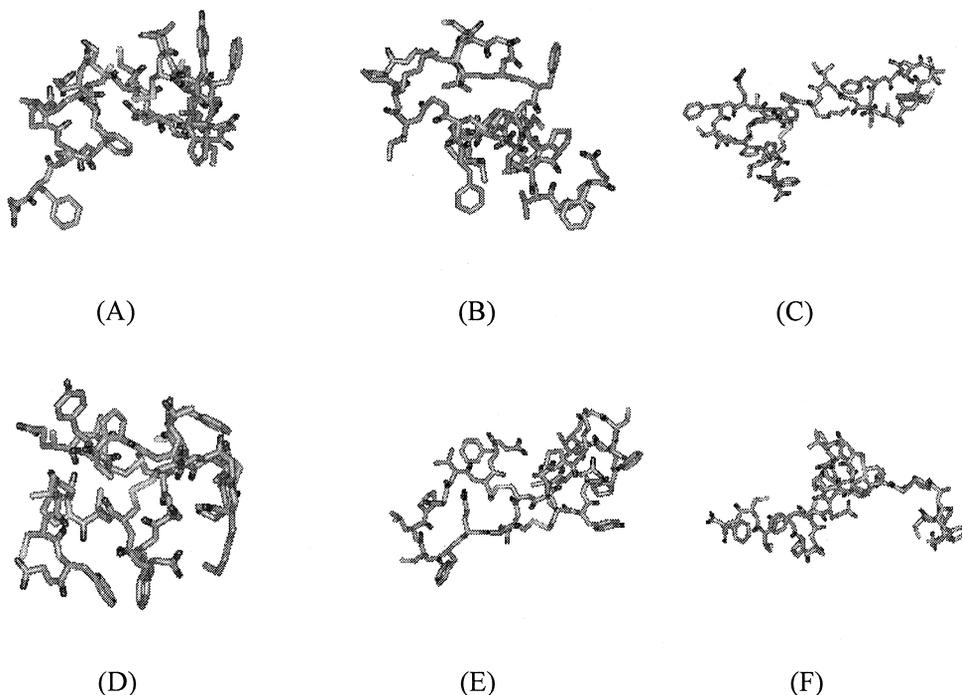


Fig. 3. The global conformations of conotoxin PVIA with partial disulfide bonds formed. (A) 3–18 -688.4 kJ mol $^{-1}$; (B) 10–22 -871.1 kJ mol $^{-1}$; (C) 17–27 -772.5 kJ mol $^{-1}$; (D) 3–18, 10–22 -539.2 kJ mol $^{-1}$; (E) 3–18, 17–27 -619.9 kJ mol $^{-1}$; and (F) 10–22, 17–18 -562.7 kJ mol $^{-1}$.

toxin PVIA with different disulfide bonds formed. In other words, from the energies of the global conformations and other low energetic conformers, we can suppose which disulfide bond will break first and which one will follow when the peptide deforms. Among the configurations which has two disulfide bonds formed, configuration E has the lowest energy, which is $-619.9 \text{ kJ mol}^{-1}$, so the disulfide bond between the cystine 10 and cystine 22 may be broken first. The energies of other conforms also support this deduction. Configuration A which has disulfide bond formed between cystine 3 and 18 has energy of $-688.4 \text{ kJ mol}^{-1}$. However, the energy of configuration C which has one disulfide bond between cystine 17 and 27 is more than 80 kJ mol^{-1} lower than the previous one. So the disulfide bond between cystine 3 and 18 may be broken more easily than the one between cystine 17 and 27. In short, if the deformed δ -conotoxin PVIA revives again, a disulfide bond between cystine 17 and 27 will be formed, followed by the bond between the cystine 3 and 18, and then by the disulfide bond between cystine 10 and 20. The information from the peptide deforming and reforming procedures can guide us to find practical synthesis paths.

Acknowledgements

This project was supported by NSFC 29573095.

References

- [1] D. Dasgupta, Z. Michalewicz, *Evolutionary Algorithms in Engineering Applications*, Springer-Verlag, Berlin, 1997.
- [2] S. Fraga, J.M. Parker, *Computer Simulations of Protein Structure and Internations*, Springer-Verlag, Berlin, 1995, p. 164.
- [3] Z. Michalewicz, *Genetic Algorithm + Data Structure = Evolution Programs*, Springer-Verlag, Berlin, 1994, p. 13.
- [4] J. Lee, H.A. Scheraga, S. Rackovsky, *J. Comput. Chem.* 18 (1997) 1222.
- [5] W.J. Pullan, *J. Comput. Chem.* 18 (1997) 1096.
- [6] K.J. Shon, A. Hasson et al., *Biochemistry* 33 (1994) 11420.
- [7] K.J. Shon et al., *Biochemistry* 34 (1995) 4913.
- [8] D.R. Westhead, D.E. Clark et al., *J. Computer-Aided Molecular Design* 9 (1995) 139.
- [9] J.S. Farr, G.P. Miljanich, L. Nadasdi, *J. Mol. Biol.* 248 (1993) 106.
- [10] J.H. Davis, E.K. Bradley, G.P. Miljanich, L. Nadasdi, *Biochemistry* 32 (1993) 7396.
- [11] T. Kohno, J.I. Kim, K. Kobayashi, *Biochemistry* 34 (1995) 10256.