

# Study on the C-terminal beta-hairpin of protein G in AB heteropolymer model

**Seung-Yeon Kim**

School of Liberal Arts and Sciences, Korea National University of Transportation, Chungju  
27469, South Korea

E-mail: sykimm@ut.ac.kr

**Abstract.** The off-lattice AB heteropolymer model, consisting of the hydrophobic (A) and hydrophilic (B) polymers, is one of popular protein models. Its energy function includes the bending energy and the van der Waals interaction energy. The properties and the energy landscape of the C-terminal beta-hairpin of protein G are studied in the off-lattice AB heteropolymer model with conformational space annealing, a powerful global optimization method.

## 1. Introduction

Proteins [1] are fundamental components of all living organisms. There are about one hundred thousand kinds of proteins in human body, controlling human's biological activities. Information on the three-dimensional tertiary native structure of a protein is vital in unveiling its biological function and role. Understanding the folding of a protein into its three-dimensional tertiary native structure only with its one-dimensional amino-acid sequence information (the so-called primary structure) is one of the most important problems in modern science. The failure of a proper protein folding results in the malfunction of a biological system, leading to fatal diseases. Proteins are linear polymers, built of twenty different amino acids (or residues), with a defined residue sequence. An amino acid consists of the intrinsic side-chain and the common backbone. Because different amino acids have a common backbone structure, their side-chains determine their nature - hydrophilic or hydrophobic. Hydrophobic amino acids of an unfolded protein in water drive it to fold into its three-dimensional tertiary native structure. Ten residues, glycine (G, Gly), alanine (A, Ala), valine (V, Val), leucine (L, Leu), isoleucine (I, Ile), cysteine (C, Cys), methionine (M, Met), phenylalanine (F, Phe), tyrosine (Y, Tyr), and tryptophan (W, Trp), are usually classified as hydrophobic amino acids. The other ten residues, proline (P, Pro), serine (S, Ser), threonine (T, Thr), asparagine (N, Asn), glutamine (Q, Gln), aspartic acid (D, Asp), glutamic acid (E, Glu), lysine (K, Lys), arginine (R, Arg), and histidine (H, His), belong to hydrophilic amino acids, where aspartic acid and glutamic acids are negatively charged but lysine and arginine are positively charged.

Simplified model proteins have been popularly used to understand the physical principles of protein folding because studying the protein-folding problem is too difficult. The HP model [2, 3] is the most simplified model for protein folding, consists of two kinds of amino acids, the hydrophobic (A) and hydrophilic (B) polymers, and is configured as self-avoiding walks on lattices. A more generalized off-lattice AB heteropolymer model [4, 5, 6, 7, 8, 9] is based on



a realistic potential energy including the bending energy and the van der Waals interaction energy. In this work, we investigate the properties and the energy landscape of the C-terminal beta-hairpin of protein G in the off-lattice AB heteropolymer model using conformational space annealing [10]. Protein G is an immunoglobulin binding protein (PDB ID: 2GB1). Its C-terminal  $\beta$ -hairpin fragment [11, 12, 13] with the amino-acid sequence, Gly-Glu-Trp-Thr-Tyr-Asp-Asp-Ala-Thr-Lys-Thr-Phe-Thr-Val-Thr-Glu, has been frequently studied to understand protein folding. This sequence corresponds to ABABABBABBBABABB in the off-lattice AB heteropolymer model.

## 2. Off-lattice AB heteropolymer model and global optimization

The energy function of the off-lattice AB heteropolymer model [4, 5, 6, 7, 8, 9], consisting of two energy terms (the bending energy and the van der Waals interaction energy), is defined by

$$E = \frac{1}{4} \sum_{i=2}^{N-1} (1 + \cos \theta_i) + \sum_{i=1}^{N-2} \sum_{j=i+2}^N \left( \frac{4}{r_{ij}^{12}} - \frac{V_{ij}}{r_{ij}^6} \right). \quad (1)$$

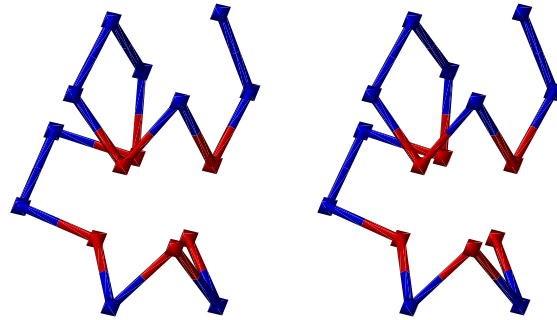
Here,  $\theta_i$  is the bond angle defined by three residues  $i-1$ ,  $i$ , and  $i+1$ , and  $r_{ij}$  is the distance between residues  $i$  and  $j$ . In addition, the van der Waals constant  $V_{ij}$  is 4, 2, and  $-2$  for, respectively, AA, BB, and AB pairs, thus, yielding a strong attraction between AAs, a weak attraction for BBs, and a weak repulsion between ABs. All bond lengths are fixed to the unity, and all physical quantities are unitless.

Global optimization is important in various fields of science and engineering. Global optimization problems are hard to solve because many of them belong to the NP-complete class. The protein-folding problem is one of the well-known classic examples. A powerful global optimization method called conformational space annealing (CSA) was proposed [10], and applied to the protein-folding problem [14, 15, 16, 17]. The CSA method unifies the essential ingredients of the three global optimization methods, simulated annealing, genetic algorithm, and basin-hopping algorithm. Here, we describe the concept of CSA briefly. The details of CSA are well explained in Ref. [10]. First, as in basin-hopping algorithm, we consider only the phase space of local minima, that is, all conformations are energy-minimized by a local minimizer. Secondly, as in genetic algorithm, we consider many conformations (called a bank in CSA) collectively, and we perturb a subset of bank conformations (seeds) using the information in other bank conformations. This procedure is similar to mating in genetic algorithm. Finally, as in simulated annealing, we introduce an annealing parameter  $D_{\text{cut}}$  (a cutoff distance in the phase space of local minima), which plays the role of temperature in simulated annealing. The diversity of sampling is directly controlled in CSA by introducing a distance measure between two conformations and comparing it with  $D_{\text{cut}}$ , whereas there is no such systematic control in simulated annealing. The value of  $D_{\text{cut}}$  is slowly reduced just as in simulated annealing, hence the name conformational space annealing.

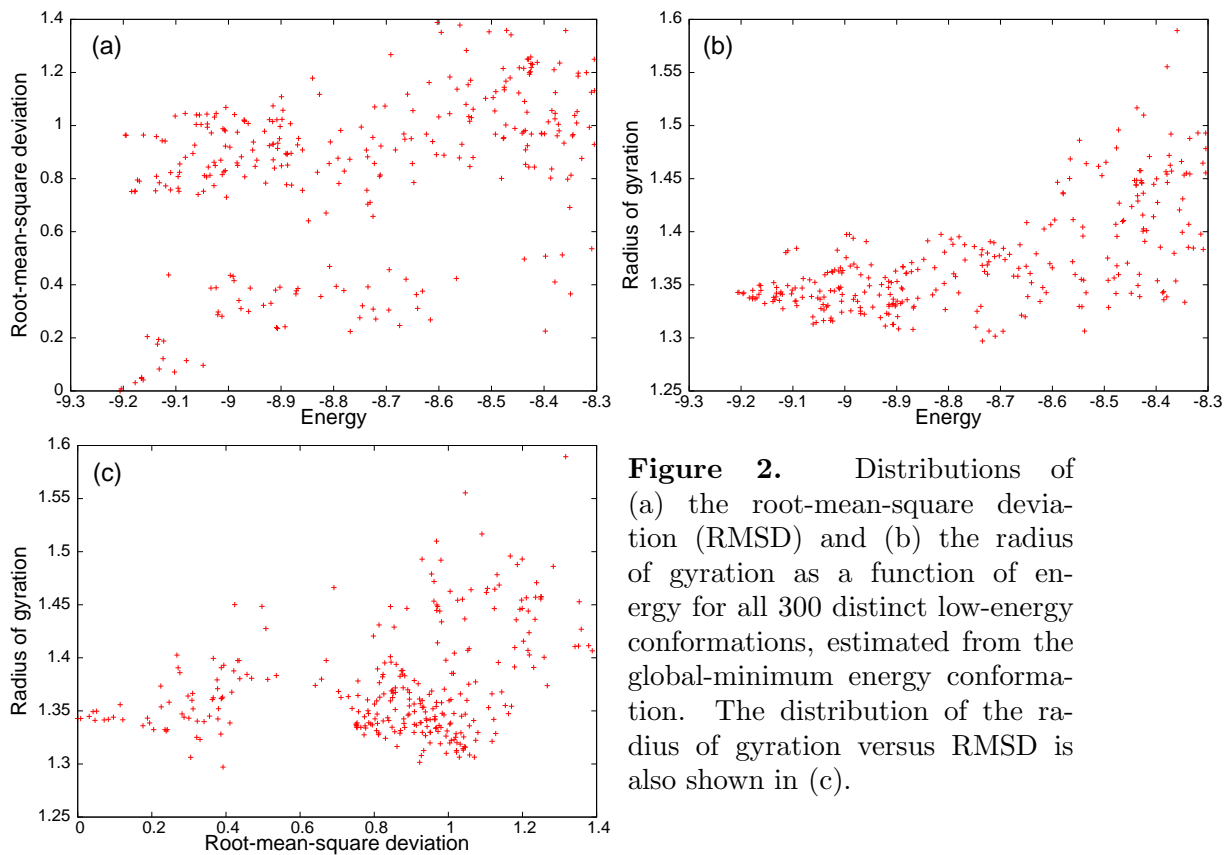
## 3. Beta-hairpin in AB heteropolymer model

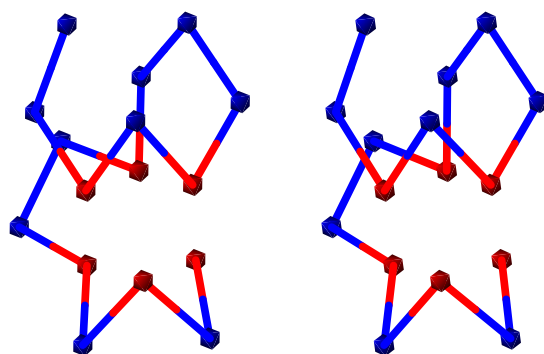
We study the C-terminal  $\beta$ -hairpin fragment of protein G in the off-lattice AB heteropolymer model using CSA, and obtain 300 distinct low-energy conformations between energies  $-9.20604$  and  $-8.30429$ . The lowest-energy conformation with  $E_1 = -9.20604$  can be called the (putative) global-minimum energy conformation. Figure 1 shows the stereographic view of the global-minimum energy conformation whose all hydrophobic residues constitute a well-formed hydrophobic core.

To investigate the energy landscape of the  $\beta$ -hairpin fragment, we measure the root-mean-square deviation (RMSD) of all low-energy conformations from the global-minimum energy conformation. Figure 2(a) depicts the distribution of the RMSD as a function of energy, showing



**Figure 1.** Stereographic view for the global-minimum energy conformation ( $E_1 = -9.20604$ ) whose radius of gyration is 1.3428. The red and blue spheres indicate hydrophobic (A) and hydrophilic (B) polymers, respectively.





**Figure 3.** Stereographic view for the lowest-energy conformation ( $E_3 = -9.19626$ ) of the second group whose RMSD is 0.9638 and radius of gyration is 1.3416.

that these low-energy conformations are classified as two groups: the first group with  $\text{RMSD} \leq 0.5358$  and the second group with  $\text{RMSD} \geq 0.6410$ . Only 64 low-energy conformations belong to the first group. On the other hand, most of the low-energy conformations (amounting to 236 distinct conformations) belong to the second group whose minimum energy is  $E_3 = -9.19626$  (the third lowest energy) with a tiny energy difference  $\Delta E = E_3 - E_1 = 0.00978$ , implying a rugged energy landscape.

The radius of gyration ( $R_g$ ) for all 300 low-energy conformations are also estimated. Figure 2(b) shows the distribution of the radius of gyration as a function of energy, and figure 2(c) shows the distribution of the radius of gyration versus RMSD. The first group and the second group can not be distinguished in figure 2(b), but they are again distinguished in figure 2(c). As shown in figure 2(b), for  $E \leq -8.6$ , the low-energy conformations are relatively compact since  $R_g \leq 1.4086$ . However, for  $E > -8.6$ , the low-energy conformations show a wide range of  $R_g$  values ( $1.3064 \leq R_g \leq 1.5554$ ).

Figure 3 shows the stereographic view of the lowest-energy conformation of the second group whose radius of gyration is 1.3416 and RMSD is 0.9638 from the global-minimum energy conformation. Also, it shows a well-formed hydrophobic core, similar to the global-minimum energy conformation.

### Acknowledgments

This was supported by Korea National University of Transportation in 2016.

## References

- [1] Creighton W E 1993 *Proteins: Structures and Molecular Properties*, 2nd edition (New York: W H Freeman and Company)
- [2] Lau K F and Dill K A 1989 *Macromolecules* **22** 3986
- [3] Lee J H, Kim S-Y and Lee J 2015 *AIP Advances* **5** 127211
- [4] Stillinger F H, Head-Gordon T and Hirshfeld C L 1993 *Phys. Rev. E* **48** 1469
- [5] Kim S-Y, Lee S B and Lee J 2005 *Phys. Rev. E* **72** 011916
- [6] Kim J, Straub J E and Keyes T 2006 *Phys. Rev. Lett.* **97** 050601
- [7] Lee J W, Joo K H, Kim S-Y and Lee J 2008 *J. Comput. Chem.* **29** 2479
- [8] Li B, Li Y and Gong L 2014 *Engineering Applications of Artificial Intelligence* **27** 70
- [9] Kim S-Y 2016 *Chaos, Solitons and Fractals* in press (<http://dx.doi.org/10.1016/j.chaos.2016.04.010>)
- [10] Lee J, Scheraga H A and Rackovsky S 1997 *J. Comput. Chem.* **18** 1222
- [11] Munoz V, Thompson P A, Hofrichter J and Eaton W A 1997 *Nature* **390** 196
- [12] Lee I-H, Kim S-Y and Lee J 2005 *Chem. Phys. Lett.* **412** 307
- [13] Lee J 2013 *Phys. Rev. Lett.* **110** 248101
- [14] Lee J, Kim S-Y, Joo K H, Kim I and Lee J 2004 *Proteins* **56** 704
- [15] Kim S-Y, Lee W and Lee J 2006 *J Chem Phys* **125** 194908
- [16] Kim S-Y 2010 *J Chem Phys* **133** 135102
- [17] Joo K H, Joung I, Lee J H, Lee J W, Lee W, Brooks B, Lee S J and Lee J 2015 *Proteins* **83** 2251