

# Conformational Ensemble of B Chain in T<sub>6</sub> Human Insulin Based on the Landau Free Energy

YANLIN LEI<sup>a</sup>, JIANFENG HE<sup>a,\*</sup>, JIAOJIAO LIU<sup>a</sup> AND JING LI<sup>b</sup>

<sup>a</sup>School of Physics, Beijing Institute of Technology, Beijing 100081, P.R. China

<sup>b</sup>Beijing Genetech Pharmaceutical Co. Ltd., Beijing 102299, P.R. China

(Received October 4, 2017; in final form October 21, 2017)

Insulin is an important peptide hormone for the glucose metabolism. The structural flexibility of insulin B chain attracts a lot of our attention for understanding the biological activity. Our work carried out the extensive sampling to statistically clarify the structural changes of isolated T<sub>6</sub> human insulin B chain. We introduced the Landau free energy to describe the isolated insulin B chain whose experimental structure locates a local energy minimum. Its trained model was subjected to thousands of heating and cooling circles between the high and low temperatures. Six typical structure clusters were found by classifying the final generated structures with RMSD and radius of gyration. The structures in clusters indicate the potential deformations of insulin B chain at residues 5–8 of N-terminus, residues 9–12 of central helix and residues 24–29 of C-terminus, which agrees with the experimental results.

DOI: [10.12693/APhysPolA.133.1261](https://doi.org/10.12693/APhysPolA.133.1261)

PACS/topics: insulin, structural flexibility, free energy, conformational ensemble

## 1. Introduction

Insulin plays a critical role in the glucose metabolism as a peptide hormone secreted by the pancreatic islet  $\beta$ -cell. Insulin and its analogs are widely used for treatment of type I diabetes mellitus. In the different environment, insulin can exist in the form of monomer, dimer, hexamer and so on [1–4]. For example, insulin is synthesized and stored in the pancreas as the hexamer, however its biologically active form is monomer [5]. The insulin monomer is composed of a 21-residue A chain and a 30-residue B chain, linked together by disulfide bonds. It is supposed that insulin B chain may keep much of its conformation independently of A chain [6, 7]. The flexibility of B chain is important for comprehending insulin activity, but it is not well understood.

Structures of insulin have been resolved by X-ray diffraction and NMR spectroscopy [1–8]. The secondary structure of insulin B chain generally consists of the N-terminus (residues 1–8), central  $\alpha$ -helix and extended C-terminus. Structures of insulin are denoted by T or R state according to whether the N-terminus of B chain has an extended or an  $\alpha$ -helical conformation. The T-state conformation is believed to be associated with the biological activity of insulin. Our study employed the structure of T<sub>6</sub> human insulin with 1.0 Å resolution determined by cryo-EM (PDB entry: 1mso) [1]. Its B chain contains two helix (residues 8–20, 21–23) and an extended C-terminal structure (residues 24–26).

Molecular dynamics (MD) is a valuable complement to experimental methods for providing insights into the

protein folding and conformation changes. Some MD researches have reported the behavior of insulin monomer, dimer and hexamer in solvent [9–12]. Since the calculation amount is large for considering the all-atom interactions and solving Newtonian equations, the limited number of trajectories is often computed in those MD simulations. The sufficient sampling of conformation becomes important to understand the macroscopic behaviour of protein. It motivates us to develop the new model and theory of the protein dynamics. In the present work, we built the coarse-grained model of skeletal C <sub>$\alpha$</sub>  chain, and introduced the universal energy for the protein in the native state from the intrinsic gauge symmetry [13–16]. The structural self-assembly of insulin B chain was modelled by the Monte Carlo protocol, based on the trained free energy. This novel approach consumed very fewer computation amount than MD, which allowed us to execute plenty of repeatedly simulations for the protein folding. Thus, the structural landscape of insulin B chain was statistically investigated by thousands of simulations in this work.

## 2. Theoretical and computational method

In our model, the positions of C <sub>$\alpha$</sub>  atoms in a protein were described by the coordinates  $\mathbf{r}_i$  ( $i = 1, 2, \dots, N$ ). As shown in Fig. 1, we introduced the unit tangent vector  $\mathbf{t}_i$ , unit binormal vector  $\mathbf{b}_i$ , and unit normal vector  $\mathbf{n}_i$  as follows:

$$\mathbf{t}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|}, \quad \mathbf{b}_i = \frac{\mathbf{t}_{i-1} \times \mathbf{t}_i}{|\mathbf{t}_{i-1} \times \mathbf{t}_i|}, \quad \mathbf{n}_i = \mathbf{b}_i \times \mathbf{t}_i. \quad (1)$$

The right-handed orthogonal triplet  $(\mathbf{t}_i, \mathbf{n}_i, \mathbf{b}_i)$  was the discrete Frenet frame at the position  $\mathbf{r}_i$ . The bond and torsion angles were formed by linking the continuous C <sub>$\alpha$</sub>  atoms, shown in Fig. 2. The bond angle of C <sub>$\alpha$</sub>  atoms at the positions  $\mathbf{r}_i, \mathbf{r}_{i+1}$  and  $\mathbf{r}_{i+2}$  was

\*corresponding author; e-mail: [hjf@bit.edu.cn](mailto:hjf@bit.edu.cn)

$$\kappa_i \equiv \kappa_{i+1,i} = \arccos(\mathbf{t}_{i+1} \cdot \mathbf{t}_i). \quad (2)$$

The torsion angle of  $C_\alpha$  atoms at  $\mathbf{r}_{i-1}$ ,  $\mathbf{r}_i$ ,  $\mathbf{r}_{i+1}$  and  $\mathbf{r}_{i+2}$  was

$$\tau_i \equiv \tau_{i+1,i} =$$

$$\text{sgn}((\mathbf{b}_i \times \mathbf{b}_{i+1}) \cdot \mathbf{t}_i) \arccos(\mathbf{b}_{i+1} \cdot \mathbf{b}_i). \quad (3)$$

Obviously, all of  $(\kappa_i, \tau_i)$  angles along the backbone  $C_\alpha$  atoms can be obtained by the discrete Frenet frames. Conversely, if these angles were given, the Frenet frames at the positions of backbone  $C_\alpha$  atoms can be iteratively constructed by the following discrete Frenet equation:

$$\begin{pmatrix} \mathbf{n}_{i+1} \\ \mathbf{b}_{i+1} \\ \mathbf{t}_{i+1} \end{pmatrix} = \begin{pmatrix} \cos \kappa \cos \tau & \cos \kappa \sin \tau & -\sin \kappa \\ -\sin \tau & \cos \tau & 0 \\ \sin \kappa \cos \tau & \sin \kappa \sin \tau & \cos \kappa \end{pmatrix}_{i+1,i} \begin{pmatrix} \mathbf{n}_i \\ \mathbf{b}_i \\ \mathbf{t}_i \end{pmatrix}. \quad (4)$$

After we had the Frenet frames, the coordinates  $\mathbf{r}_k$  ( $k = 1, 2, \dots, N$ ) of  $C_\alpha$  atoms were determined by

$$\mathbf{r}_k = \sum_{i=0}^{k-1} |\mathbf{r}_{i+1} - \mathbf{r}_i| \cdot \mathbf{t}_i. \quad (5)$$

where  $\mathbf{r}_0 = 0$ ,  $\mathbf{t}_0$  pointed along the positive  $z$ -axis and  $\mathbf{t}_1$  was in the  $y$ - $z$  plane. The distance between two neighbouring  $C_\alpha$  atoms was assigned the constant value of 3.80 Å [17].

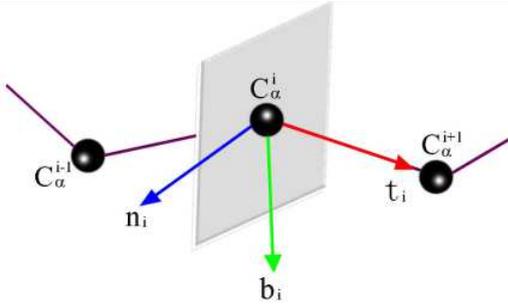


Fig. 1. Frenet frame  $(\mathbf{t}_i, \mathbf{n}_i, \mathbf{b}_i)$  at the  $i$ -th  $C_\alpha$  atom.  $\mathbf{t}_i$  points from  $i$ -th  $C_\alpha$  towards  $(i+1)$ -th  $C_\alpha$  atom,  $\mathbf{n}_i$  is in the plane determined by  $(i-1)$ -th,  $i$ -th,  $(i+1)$ -th  $C_\alpha$  atoms and vertical to  $\mathbf{t}_i$ , and  $\mathbf{b}_i$  is perpendicular to the plane  $(\mathbf{t}_i, \mathbf{n}_i)$ .

We noted that the vectors  $(\mathbf{n}_i, \mathbf{b}_i)$  in the normal plane were not explicitly relevant to Eq.(5). Any new frame can be easily built by rotating the two vectors any angle around the vector  $\mathbf{t}_i$ ,

$$\begin{pmatrix} \mathbf{n}_i \\ \mathbf{b}_i \\ \mathbf{t}_i \end{pmatrix} \rightarrow \begin{pmatrix} \cos \theta_i & \sin \theta_i & 0 \\ -\sin \theta_i & \cos \theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{n}_i \\ \mathbf{b}_i \\ \mathbf{t}_i \end{pmatrix}. \quad (6)$$

Here  $\theta_i$  was the arbitrarily local rotation angle. Since  $\mathbf{t}_i$  was held tightly, the local  $SO(2)$  transformation did not change the coordinates of  $C_\alpha$  atoms according to Eq.(5). It provided other framing to describe the model of skeletal

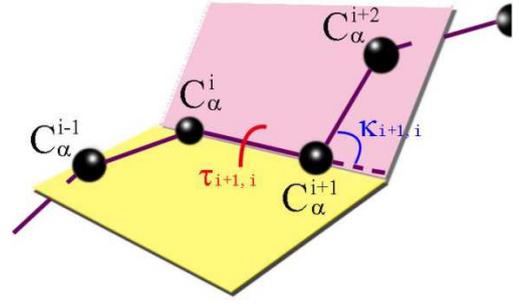


Fig. 2. Virtual bond and torsion angles  $(\kappa_{i+1}, \tau_{i+1})$  defined in Eqs. (2) and (3) have been shown in the schematic diagram.

tal  $C_\alpha$  chain instead of the Frenet frame. Furthermore, it indicated that our model of protein had the rotation invariance of  $(\mathbf{n}_i, \mathbf{b}_i)$  vectors. Correspondingly, the bond and torsion angles had the transformation

$$\kappa_i \rightarrow e^{i\theta_i} \kappa_i, \tau_i \rightarrow \tau_i + \theta_{i-1} - \theta_i. \quad (7)$$

The ranges of bond and torsion angles were  $\kappa_i \in [0, \pi]$ ,  $\tau_i \in [-\pi, \pi]$  from Eqs.(2) and (3). We found it was convenient to extend the range of  $\kappa_i$  into  $[-\pi, \pi]$  [13,17]. Here the double covering was compensated by introducing the following discrete  $\mathbb{Z}_2$  gauge transformation:

$$\kappa_l \rightarrow -\kappa_l, \tau_i \rightarrow \tau_i - \pi \forall l \geq i. \quad (8)$$

This was the special case of Eqs.(6) and (7) with  $\theta_i = \pi$  and  $\theta_{i-1} = 0$ . The  $\mathbb{Z}_2$  transformation had no effect on the  $C_\alpha$  coordinates and left the skeletal  $C_\alpha$  chain intact.

The protein structure in the thermal equilibrium locates at a local minimum of Helmholtz free energy [18] which is a function of interatomic distances. Following our model of skeletal  $C_\alpha$  chain, all  $(\kappa_i, \tau_i)$  angles can also constitute a complete set of structural order parameters for the free energy. Angles  $(\kappa_i, \tau_i)$  possessed the  $\mathbb{Z}_2$  invariance in Eq.(8), which indicated the protein can be described by the Landau free energy [19,20]. The deformations of protein kept slow and small around the energy minimum, thus the following expansion of free energy was permitted:

$$\begin{aligned} E = & - \sum_{i=1}^{N-1} 2\kappa_{i+1}\kappa_i + \sum_{i=1}^N [2\kappa_i^2 + c(\kappa_i^2 - m^2)^2] \\ & + \sum_{i=1}^N [b\kappa_i^2\tau_i^2 + d\tau_i + e\tau_i^2 + q\kappa_i^2\tau_i] + \dots \end{aligned} \quad (9)$$

Here  $c, m, b, d, e$  and  $q$  were parameters. They were determined by training a conformation of local energy minimum to model the protein backbone, detailed analysis in Refs. [13,15].

We simulated the movement of isolated insulin B chain by subjecting the Landau free energy (9) which was trained with the crystallographic structure (1ms0). The dynamical process with increasing and decreasing temperature was executed according to Monte Carlo (MC) protocol [21,22], which was abided by the Arrhenius law. The transition probability  $p(i \rightarrow j)$  from a conformation

$i$  to another  $j$  was estimated by

$$p(i \rightarrow j) = \frac{1}{1 + \exp\left(\frac{\Delta E_{ji}}{kT}\right)}. \quad (10)$$

Here  $\Delta E_{ji} = E_j - E_i$  was the energy difference between two conformations evaluated by Eq.(9), and  $kT$  was the MC temperature factor. The scale of unit in  $kT$  can be determined by the renormalization of the energy function (9) with the experimentally measured temperatures. The conversion of the temperature into Kelvin scale was exhibited in detail in Ref. [23], using the transition temperature between the folded phase and completely flexible chain (random walk) phase of protein.

In the case of isolated insulin B chain,  $kT_{low} = 10^{-15}$  and  $kT_{high} = 10^{-2}$  were taken as the values of low and high temperatures in the heating and cooling cycle. A single cycle had  $5 \times 10^7$  MC steps and was carried out as follows: the initial structure of protein was thermally equilibrated at the very low temperature  $kT_{low}$  by  $5 \times 10^6$  steps, the protein was warmed up by linearly increasing the temperature from  $kT_{low}$  to  $kT_{high}$  with  $1 \times 10^7$  steps, and then  $kT_{high}$  was holden in the following  $2 \times 10^7$  steps to ensure the protein fully thermalized, the protein was cooled back to the low temperature  $kT_{low}$  with the reversed strategy of the heating process. At each step,  $(\kappa_i, \tau_i)$  angles were perturbed with the amplitude:  $\kappa_i \rightarrow \kappa_i + 0.05r$  and  $\tau_i \rightarrow \tau_i + 0.05r$ , where  $r$  was a random number between 0 and 1 with Gaussian distribution. The steric exclusion in the conformation was checked at each MC step. 3696 repeated heating and cooling cycles were performed in this work, and we obtained good statistical properties for isolated insulin B chain.

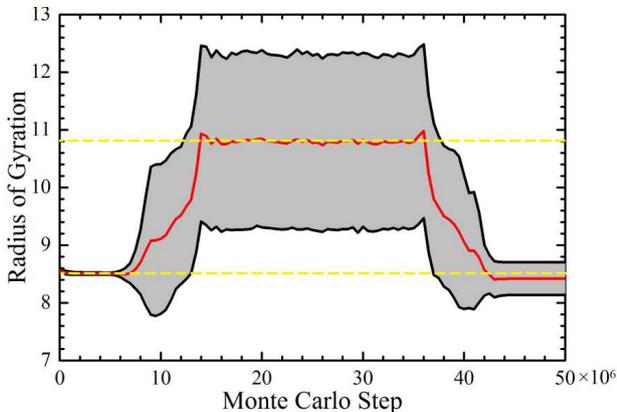


Fig. 3. Evolution of  $R_g$  (in Å) with the Monte Carlo step during the heating and cooling cycle. The red line is the average of  $R_g$  and the gray band is the deviation distance from the average value.

### 3. Results and discussion

For simplifying the structural analysis, the high flexible residues were not included and the fragment 5-29 of insulin B chain was considered in our simulations. We

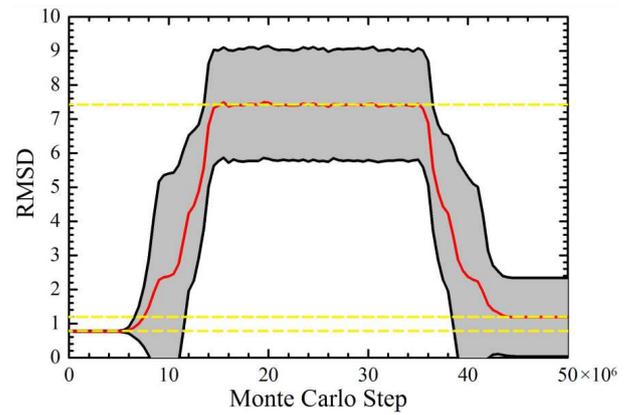


Fig. 4. Evolution of RMSD (in Å) between the generated structure and the crystal structure with the Monte Carlo step, during the heating and cooling cycle. The red line is the average of RMSD and the gray band is the deviation distance from the average value.

employed the following equation to evaluate the radius of gyration ( $R_g$ ):

$$R_g = \frac{1}{N} \sqrt{\frac{1}{2} \sum_{i,j} (\mathbf{r}_i - \mathbf{r}_j)^2}. \quad (11)$$

Here  $\mathbf{r}_i$  and  $\mathbf{r}_j$  were the coordinates of backbone  $C_\alpha$  atoms,  $N$  was the number of  $C_\alpha$  atoms in the protein. The  $R_g$  of generated structure at each MC step was computed for all 3696 heating and cooling cycles. Then we took the average of  $R_g$  and estimated the  $R_g$  deviation distance from the average value at every MC step. Figure 3 displayed that the averages of  $R_g$  were constant at the low and high temperatures. The average value was 8.5 Å at the entire initial low temperature stage of heating and cooling cycle. However,  $R_g = 8.4$  Å with the 0.3 Å deviation fluctuation at the final low temperature stage. It indicated that some of the final structures were more compact than the structures at the initial low temperature stage. The average of  $R_g$  kept the value of 10.8 Å at the high temperature, accompanied by the 1.5 Å fluctuation. It made clear that the structure was fully opened and became very loose after being heated. Combining with the high fluctuation of  $R_g$ , we were confident that the protein was in the random walk regime at high temperatures by the phase transition [20]. The profile of  $R_g$  evolution can reveal that the structure of insulin B chain was fully thermalized in the low and high temperatures, and the process that insulin B chain unfolded and refolded back to the impact form.

The generated structures of 5-29 fragment in insulin B chain were aligned and compared with the crystal structure (1ms0). The root mean square deviation (RMSD) of skeletal  $C_\alpha$  can be easily obtained by

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i^g - \mathbf{r}_i^i)^2}. \quad (12)$$

Here  $\mathbf{r}_i^i$  and  $\mathbf{r}_i^g$  were the coordinates of  $i$ -th  $\text{C}_\alpha$  atoms in the crystal and generated structures. We also calculated the average of RMSD and the RMSD deviation from the average value at each MC step for all heating and cooling cycles, as shown in Fig. 4. Like the  $R_g$  evolution, the averages of RMSD were constant at both low and high temperatures. The average of RMSD was 0.77 Å at the initial low temperature stage during the heating and cooling cycle. RMSD = 1.19 Å with the 1.15 Å deviation fluctuation at the final low temperature region. We inferred that some of the final structures were very different from the crystal structure. In the mean time, there was an obvious difference among the final structures. The average of RMSD had the value of 7.40 Å with the 1.65 Å fluctuation at the high temperature. It stated that the structures of 5–29 fragment at the high temperature stage had the diverse spatial shapes. The profile of red line in Fig. 4 further revealed the unfolding and refolding process from the perspective of structural evolution.

From Figs. 3 and 4, two obvious structural transitions can be observed in both heating and cooling periods. At the first one,  $R_g = 9.17$  Å and RMSD = 2.45 Å. The averages of  $R_g$  and RMSD were 9.5 Å and 4.24 Å at another transition, respectively. It addressed that the insulin B chain had two potential intermediate states during its unfolding and refolding processes, and the folding dynamics of isolated insulin B chain may be divided into three phases.

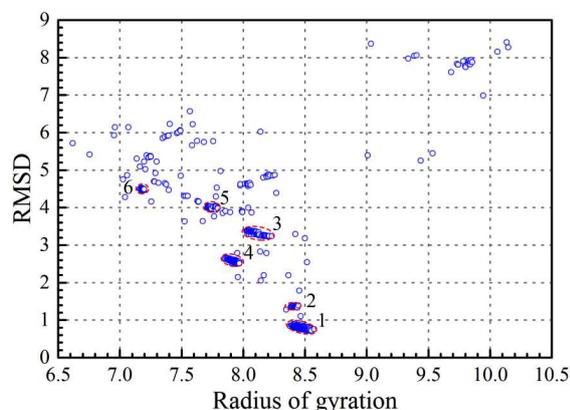


Fig. 5. The RMSD vs.  $R_g$  (in Å) distribution of final generated structures in a run with 3696 repeated heating and cooling simulations. Each final structure is represented by one blue cycle. The six major structural clusters are encircled by a red ellipse dash line with the number 1–6.

In order to reveal the spatial shape difference, the final structures were classified in terms of their  $R_g$  and RMSD from the crystal structure. Figure 5 exhibited the distribution of final structures for our 3696 repeat heating and



Fig. 6. Ten structures randomly selected among the ones of cluster 1.

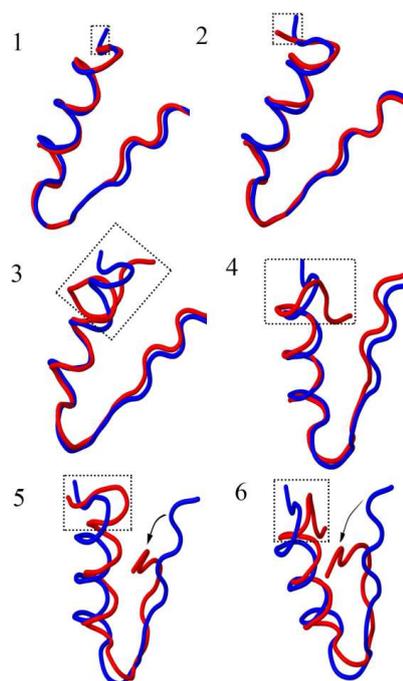


Fig. 7. The representative structure of clusters 1–6 denoted in Fig. 5. Each plot displays the superimposition of final generated structure (red) in the corresponding cluster and the crystal structure (blue).

cooling simulations with the selected low and high temperatures. It was clear that the  $R_g$  values had a general tendency to reduce from 8.6 Å to 6.5 Å. This declared the insulin B chain could collapse to more compact conformation after it was isolated from the binding state with A chain. It was also identified that most of blue cycles in Fig. 5 converged into six local regions separated from each other, where about 95% of structures occupied. The structures in a single region, called structural cluster, existed very small deviations of RMSD and  $R_g$  between each other. Thus the structures in the same clus-

ter were almost identical except for a little fluctuation. Figure 6, as an example, shows ten structures randomly picked from cluster 1. We believed that the clusters corresponded to the local minima of energy function constructed for isolated insulin B chain.

Figure 5 showed that the structures in cluster 1 had the smallest values of  $R_g$  and RMSD from the crystal structure than the ones in other clusters. Their RMSDs were less than 1.0 Å, and  $R_g$  were nearly identical to the one of crystal structure. We believe the structures in cluster 1 fell into the vicinity of the native state where the experimental structure (1ms0) located. The cluster 2 was very close to cluster 1 with 0.3 Å RMSD deviation, which implied their structures similar. From cluster 3 to 5, their  $R_g$  were sequentially reduced and the interval of their RMSD were almost equivalent. The RMSD and  $R_g$  of cluster 6 were about 4.5 Å and 7.2 Å, respectively. It indicated that cluster 6 possessed the most compact conformation and the largest structural difference from the crystal structure among the clusters.

For visually exposing the structural changes, Fig. 7 presented one representative structure of each cluster superimposed with the crystal structure. Here the residue 5–29 fragment in isolated insulin B chain we considered contained the secondary structures, two helix (residues 8–20, 21–23) and the extended C-terminus (residues 24–26). From Fig. 7, we observed that the structures of cluster 1 and 2 very well overlapped the crystal structure, except for a twist at residue 5. This was in line with the deduction that the structures of clusters 1–2 were similar with the crystal structure by Fig. 5. In clusters 3–6, the fragment of residues 5–8 turned around and faced the different directions, which uncovered that the N-terminus was more flexible than the central helix and the C-terminus of insulin B chain. In cluster 3 and 5, there was a deformation at residues 9–12 which situated in the forefront of  $\alpha$ -helix. We noticed that the structure of residues 24–29 in the C-terminus also had an obvious distortion in clusters 5–6. Apparently, the structural difference in clusters 1–6 was consistent with the change trend of RMSD and  $R_g$  shown in Fig. 5. It also agreed well with the experimental results illustrated by the  $C_\alpha$  displacement of insulin B chain [1].

#### 4. Conclusion

The structural variation can reveal the inherent flexibility of insulin B chain. Our results presented the information about structural changes of isolated T<sub>6</sub> human insulin B chain. We performed more than three thousand heating and cooling simulations by the temperature disturbance. In contrast with MD simulation, our sampling was more sufficient and the results statistically illuminated the structure variability of isolated insulin B chain. We found six potential structure clusters for isolated insulin B chain which were believed to locate at the local energy minima. The structural analysis of clusters revealed that the N-terminus of insulin B chain had

a higher flexibility. There was also a strong mutability of structure in the front of central  $\alpha$ -helix and the C-terminus.

#### Acknowledgments

We are grateful to Prof. Antti J. Niemi of Uppsala University for the discussion on the theoretical method. We also thanks the support of the international cooperation project of Beijing Institute of Technology.

#### References

- [1] G.D. Smith, W.A. Pangborn, R.H. Blessing, *Acta Crystallogr. D* **59**, 474 (2003).
- [2] M. Adams, T. Blundell, E. Dodson, et al, *Nature* **224**, 491 (1969).
- [3] T.L. Blundell, J.F. Cutfield, S.M. Cutfield, et al, *Nature* **231**, 506 (1971).
- [4] N.C. Kaarsholm, H.C. Ko, M.F. Dunn, *Biochemistry* **28**, 4427 (1989).
- [5] H.B. Olsen, S. Ludvigsen, N.C. Kaarsholm, *Biochemistry* **35**, 8836 (1996).
- [6] B. Hawkins, K. Cross, D. Craik, *Int. J. Pept. Protein Res.* **46**, 424 (1995).
- [7] F.Y. Dupradeau, T. Richard, G. Le Flem, et al, *J. Pept. Res.* **60**, 56 (2002).
- [8] J. Badger, M.R. Harris, C.D. Reynolds, et al, *Acta Crystallogr. B* **47**, 127 (1991).
- [9] A.E. Mark, H.J. Berendsen, W.F. van Gunsteren, *Biochemistry* **30**, 10866 (1991).
- [10] V. Zoete, M. Meuwly, M. Karplus, *J. Mol. Biol.* **32**, 913 (2004).
- [11] F.S. Legge, A. Budi, H. Treulein, et al, *Bio-phys. Chem.* **119**, 146 (2006).
- [12] A. Papaioannou, S. Kuyucak, Z. Kuncic, *PLOS ONE* **10**, e0144058 (2015).
- [13] J.F. He, J. Dai, J. Li, et al, *J. Chem. Phys.* **142**, 045102 (2015).
- [14] J. Dai, A.J. Niemi, J.F. He, et al, *Phys. Rev. E* **93**, 032409 (2016).
- [15] J. Dai, A.J. Niemi, J.F. He, *J. Chem. Phys.* **145**, 045103 (2016).
- [16] J.J. Liu, J. Dai, J.F. He, et al, *Phys. Rev. E* **95**, 032406 (2017).
- [17] X. Peng, A. Chenani, S. Hu, et al, *BMC Struct. Biol.* **14**, 27 (2014).
- [18] C. Anfinsen, H.A. Scheraga, *Adv. Protein Chem.* **29**, 205 (1975).
- [19] L.D. Landau, *Zh. Eksp. Teor. Fiz.* **7**, 19 (1937).
- [20] M. Chernodub, S. Hu, A.J. Niemi, *Phys. Rev. E* **82**, 011916 (2010).
- [21] R.J. Glauber, *J. Math. Phys.* **4**, 294 (1963).
- [22] A.B. Bortz, M.H. Kalos, J.L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- [23] A. Krokhotin, A.J. Niemi, X. Peng, *J. Chem. Phys.* **142**, 0451032 (2013).