Modeling Protein Loop Structure by Cyclic Coordinate Descent-based Approach

Haiou Li\textsuperscript{1}, Lu Sun\textsuperscript{2}, Sheng Luo\textsuperscript{1}, Xiaoyan Xia\textsuperscript{1,3}, Qiang Lyu\textsuperscript{1,3*}, Peide Qian\textsuperscript{1,3}

1 School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu, China
2 School of Electronic Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada
3 Jiangsu Provincial Key Lab for Information Processing Technologies, Suzhou, Jiangsu, China

* E-mail: qiang@suda.edu.cn

In both predicting protein loop structure and modeling flexible areas in a complex, sampling of local loop conformation plays a vital role, as the variability of loop regions carries out biological functionality such as molecular recognition and signal transduction. We address the challenge with a topological manipulation method based on cyclic coordinate descent (CCD) algorithm. The protein loop sampler (LMbCCD\textsuperscript{2}) presented here allows us to drive some key points of the loop towards some specific positions. We first test the method’s performance on sets of loops, and the results show that LMbCCD\textsuperscript{2} accurately controls the loops topology. Then we used TIM as an example to test how LMbCCD\textsuperscript{2} rotates the flexible peptide segment from unbound status to bound status.

Since protein loop regions follow a kind of complex geometric constraint on its movement, it is difficult to make some special changes on their specific topology. However, the protein in native state sometimes shows a specific movement pattern, how to simulate this movement and control the specific topology has become a difficult aspect in protein loop prediction. We expand the origin CCD method and divide the loop refine process into two mainly stages (CCD\textsuperscript{2}). The first stage is called “shape-stage”, which we try to control topology configuration of the loop. During this stage we randomly select some pivot atoms and push the initial loop close to the target loop in topology configuration. The second stage can be regarded as a “closure-stage”. We use the scheme of origin CCD method to close the loop to the rest of the protein residues.

The goal of the method is to move the pivot atoms to get as close as possible to their target atoms by changing each of dihedral pair (\(\phi\) and \(\psi\)). As show in Figure 1, the pivot atoms are marked by \(P_1, P_2, ..., P_m\), the target of pivot atoms are marked by \(T_1, T_2, ..., T_m\), and the position of pivot atoms after twisting are marked by \(Q_1, Q_2, ..., Q_m\). Moving the pivot atoms toward the target atoms is equivalent to minimizing the objective function as below:

\[
S = \sum |T_i - Q_i|^2
\] (1)

In this study we focus on the backbone flexibility. Base on CCD\textsuperscript{2} algorithm, we propose a loop sampling method called LMbCCD\textsuperscript{2}, it is a kind of random walk simulation. There are many simulation programs based on molecular dynamics or quantum chemistry\textsuperscript{[1]}. These methods are aimed at employing physical or chemical rules to determine the movement of each atom at each
step, while the geometrical constraints, e.g. bond length and bond angle, do not keep fixed and are adjusted to the energy landscape. This often introduces a large number of DoFs, which make the simulation cumbersome and computationally expensive. In our loop sampling method, all of the geometrical constraints are kept fixed, therefore pruning most of DoFs whilst maintaining the closure of polypeptide chain.

**Single loop sampling** Our loop sampling method was applied to 5 hard loop targets of length 12 residues proposed by Canutescu and Dunbrack[2]. These loops were used to test the performance of several loop sampling algorithms before[3, 4, 5]. For each of the loops in the test set, the minimum backbone RMSDs from the crystal structure among 5000 conformations sampled by the following methods are compared in Table 1. Our method performs better than these methods in all 12-residue loop targets. With LMB CCD method the minimum RMSD improves from 2 Å to 1.54 Å, from 2.1 Å to 1.62 Å, and from 2.18 Å to 1.84 Å. The current method is different from other above methods in two respects. First, the LMB CCD loop sampling method does not contain any other constraints. Second, some methods like CCD-based Rosetta protocol has been used with Rosetta for loop modeling is guided by the energy function.

**From Unbound to Bound By Only One Spin** Triose Phosphate Isomerase(TPI or TIM) is one of the first proteins for which large conformational changes in the loop region[6]. TIM loop 6 is a well-characterized example for fragment hinge motion. TIM contains a segment(V167-A176) that undergoes significant conformational change but maintains with relatively fixed torsion angles. It is observed that rotating the open conformation(1YPI) approximately 50° about the axis between the two ends of the segment produces a conformation similar with the closed loop(2YPI). It indicates the main difference between the unbound and bound state is the hinge motion, while
Table 1: The minimum backbone RMSD values of the loops sampled by the method LMbCCD, CCD, CJSD, SOS, and FALCm

<table>
<thead>
<tr>
<th>Loop</th>
<th>LMBCCD</th>
<th>CCD</th>
<th>CJSD</th>
<th>SOS</th>
<th>FALCm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1CRU_358</td>
<td>1.54</td>
<td>2.54</td>
<td>2</td>
<td>2.39</td>
<td>2.07</td>
</tr>
<tr>
<td>1CTQ_26</td>
<td>1.13</td>
<td>2.49</td>
<td>1.86</td>
<td>2.54</td>
<td>1.66</td>
</tr>
<tr>
<td>1DS1_282</td>
<td>1.62</td>
<td>3.04</td>
<td>3.1</td>
<td>2.33</td>
<td>2.1</td>
</tr>
<tr>
<td>1QLW_31</td>
<td>1.61</td>
<td>3.38</td>
<td>2.32</td>
<td>1.73</td>
<td>2.2</td>
</tr>
<tr>
<td>1QOP_178</td>
<td>1.84</td>
<td>4.57</td>
<td>2.18</td>
<td>2.21</td>
<td>2.36</td>
</tr>
</tbody>
</table>

the internal conformation keep relatively fixed. But such a purely rotation breaks the bond angle constraints at the two ends of the loop. We used TIM as an example to show the rotation of the peptide segment from unbound status to bound status. Here we divide the rotation path with an interval $1^\circ$, shown in Figure 2(a), therefore reducing the shape-bias. The rotation trajectories from the open conformation (1YPI) to the closed conformation (2YPI) are shown in Figure 2(b), and the loop after the spin is compared in Figure 2(c). The original $\text{C}_\alpha$RMSD between the open and closed conformation is $5.18\AA$ while the $\text{C}_\alpha$RMSD after this rotation drops to $1.16\AA$. And the internal coordinates of the rotated loop have a high agreement with the ones of the closed conformation.

![Figure 2](a)(b)(c)

Figure 2: (a) The trajectories of TPAs are illustrated by cuboids, the colors of which change from light red to deep red. Not all of the produced TPAs are shown. The main chain atoms of closed and open conformations are shown in gray and green, respectively. The cyan arrow indicates the rotation vector. (b) It shows the trajectory of rotated loop. As the same with the trajectories of TPAs, the color of trajectory of loop is also gradual. (c) The red is the loop rotated $-55^\circ$ about the rotation vector. The conformation keeps high agreement with the open one, both in topology and torsion angles.
Acknowledgements  This work was supported by National Science Foundation of China under the grant number 61170125.

References


