http://dx.doi.org/10.5012/bkcs.2012.33.3.770

Conformational Sampling of Flexible Ligand-binding Protein Loops†

Gyu Rie Lee, Woong-Hee Shin, Hahnbeom Park, Seokmin Shin, and Chaok Seok*

*Department of Chemistry, Seoul National University, Seoul 151-747, Korea. *E-mail: chaok@snu.ac.kr*
Received September 30, Accepted November 10, 2011

Protein loops are often involved in diverse biological functions, and some functional loops show conformational changes upon ligand binding. Since this conformational change is directly related to ligand binding pose and protein function, there have been numerous attempts to predict this change accurately. In this study, we show that it is plausible to obtain meaningful ensembles of loop conformations for flexible, ligand-binding protein loops efficiently by applying a loop modeling method. The loop modeling method employs triaxial loop closure algorithm for trial conformation generation and conformational space annealing for global energy optimization. When loop modeling was performed on the framework of ligand-free structure, loop structures within 3 Å RMSD from the crystal loop structure for the ligand-bound state were sampled in 4 out of 6 cases. This result is encouraging considering that no information on the ligand-bound state was used during the loop modeling process. We therefore expect that the present loop modeling method will be useful for future developments of flexible protein-ligand docking methods.

Key Words : Loop modeling, Ligand binding, Conformational change, Conformational sampling

Introduction

Protein loop modeling concerns solving the problem of predicting three-dimensional atomic structures for protein loops. Protein loops refer either to segments that link secondary structure elements (α-helices or β-strands) when viewed from a structural perspective, or to gaps or insertions in sequence alignments of homologous proteins from a homology modeling perspective. These regions are not well conserved during protein evolutions, so it is difficult to predict their structures from the evolutionary information alone.[1-4] Nonetheless, this variance among homologous proteins is what determines the specificity of each protein. Functional differences among homologous proteins usually result from relatively small degree of variability in protein structures. Loops often correspond to such variable regions that determine the specificity of protein function, making the issue of understanding protein loop structures in atomic detail more critical.[5-9] For example, loops are involved in enzyme active sites,[10] DNA-binding or ligand-binding sites,[11,12] or antibody complementary determining regions.[13]

When loops are involved in ligand binding, they often undergo conformational changes. This phenomenon has been of particular interest because of its relevance to structure-based drug discovery.[14-18] However, predicting the conformational changes of protein loops is still a challenging problem. Many methods have been developed to study loop flexibility, and some of them have been applied to protein-ligand docking and computational drug design.[19-21]

In this work we apply a newly developed loop modeling method[22] to tackle the problem of sampling flexible protein loop conformations. The loop modeling protocol employs a powerful global optimization process called conformational space annealing[23] in the space of closed loops. These loops are generated by the triaxial loop closure algorithm[24] to be geometrically consistent with the rest of the protein. While the protocol was originally developed to refine template-based models, here we use it to obtain ensembles of loop conformations to assess its applicability to studies of protein loop flexibilities and flexible protein-ligand docking.

The test set of protein loops in this study are borrowed from the study of Wong *et al.*[25] The selected protein loops show loop-latching motion, a dynamic loop motion covering up the ligand upon binding. Therefore, the complex including the loop tends to be closed in the holo (ligand-bound) form and open in the apo (ligand-free) form. We tested whether the loop conformation of the holo (or apo) form can be sampled by using the apo (or holo) structure only. Predicting the holo structure from a known apo form is especially of significant importance because it has direct relevance to the prediction of ligand-bound structures in computer-aided drug design.[25] Encouraging results obtained from the procedures discussed above are presented and compared with those of Wong *et al.*, and possible improvements that can be made are discussed.

Methods

Test Set of Ligand-binding Protein Loops. The proteins used in the loop modeling study by Wong *et al.*[25] were employed as the test set to assess the applicability of the loop modeling protocol[22] to sampling of flexible ligand-binding protein loops. The selected proteins show loop latching motions upon ligand binding and also have several apo and holo crystal structures available. They are Yersinia protein-tyrosine phosphatase (PTP), L-Ala-D-Glu epimerase (AEE), triose-

†This paper is to commemorate Professor Kook Joe Shin's honourable retirement.

Table 1. List of the test set protein loops used in this study

Protein	PDB ID	Loop residues
Yersinia protein-tyrosine phosphatase (PTP)	1YPT (apo) 1YTN (holo)	350-361
L-Ala-D-Glu epimerase (AEE)	1JPM (apo) 1TKK (holo)	13-29
Triosephosphate isomerase (TIM)	1R2T (apo) 2YPI (holo)	165-178
Enolase (ENO)	1EBH (apo) 1EBG (holo)	34-46
Phosphoribosyl glycinamide formyltransferase (GART)	1ZLX (apo) 1CDE (holo)	140-147
Spermidine synthase (SRM)	1INL (apo) 1JQ3 (holo)	170-181

phosphate isomerase (TIM), enolase (ENO), phosphoribosyl-glycinamide formyltransferase (GART), and spermidine synthase (SRM), as listed in Table 1. For each of the apo and holo forms, the rest of the protein structure other than the protein loop, which we call "framework", was fixed and only the loop was modeled on it. The loop length was extended by one residue on each side compared to the loop definition of Wong *et al.*[25] In this way, the variance of the framework structures between apo and holo forms could be reduced near the anchor points of the loop.

Generation of Initial Conformations by FALC and Clustering. The loop modeling protocol was used to generate initial conformations for each protein loop as follows. First, a set of 1,000 conformations at local energy minima were generated by using the FALC (Fragment Assembly and analytical Loop Closure) method.[4] This method collects fragment structures from proteins of similar sequence features in the structure database and randomly assembles the fragments to construct model loops. The triaxial loop closure (TLC)[24] is then applied to adjust backbone torsion angles so that the loop conformations fit into the rest of the protein body. The resulting models for closed loops are then clustered into 30 groups, and the cluster centers are selected as the initial conformations for global energy minimization.

Global Energy Optimization to Sample Loop Conformations. The 30 loop conformations obtained as above form the initial bank for a global optimization procedure by conformational space annealing (CSA).[23] In CSA, the bank of 30 conformations is evolved iteratively until convergence by gradually reducing the size of the conformational space that each bank member covers (represented by a parameter called D_{cut}). The energy function optimized for loop modeling was developed by combining knowledge-based terms (such as distance-dependent atomic pair energy derived from the structure database[26]) with physics-based terms (such as bonded energy, Coulomb energy, and van der Waals energy). At each iteration step, trial conformations are generated by cross-over and mutation, and they are considered together with the current bank members to update the bank into a new set of 30 representative, low-energy conformations for a given D_{cut}. The D_{cut} parameter is a cut-off distance between two conformations and is used to control the approximate size of the conformational space represented by each bank member. Whenever trial loop conformations are generated, they are closed by TLC and the side-chains are remodeled. After that, local energy minimization and a short molecular dynamics (MD) simulation followed by another minimization are performed. This iteration continues until convergence by gradually reducing the D_{cut} value.

We analyze the 30 conformations in the final bank obtained at the end of the conformational space annealing iteration. A sound aspect of the protocol is that it searches for low-energy minima effectively by incorporating TLC to generate loops with geometric consistency. Although the number of generated conformations for each loop (30 in this study) is not large, we found that they contain important conformations in different functional states (i.e., ligand-bound and unbound states) in some test cases.

Results and Discussion

The loop sampling results are summarized in Table 2.

Table 2. Loop sampling results

Protein	Loop length	Loop RMSD[a], crystal apo *vs* holo (Å)	Sampling Results Best loop RMSD (Å) (E rank, RMSD (Å) of 1st rank)[b]			
			Modeled on the "apo" framework and compared with		Modeled on the "holo" framework and compared with	
			crystal apo	crystal holo	crystal holo	crystal apo
PTP	12	3.1	1.2 (9, 1.8)	1.5 (17, 2.2)	0.68 (3, 1.2)	2.1 (9, 2.6)
AEE	17	6.2	5.2 (17, 7.4)	3.2 (12, 5.4)	1.9 (24, 3.5)	4.8 (23, 7.5)
TIM	14	4.1	1.6 (8, 1.9)	1.7 (15, 3.3)	0.80 (6, 1.7)	3.2 (1, 3.2)
ENO	13	4.9	3.8 (27, 6.1)	4.1 (27, 5.3)	1.7 (22, 3.5)	3.4 (6, 3.5)
GART	8	0.72	2.2 (25, 2.8)	2.2 (25, 3.0)	2.0 (21, 5.0)	2.1 (21, 5.5)
SRM	12	3.7	2.3 (22, 3.5)	2.8 (2, 5.5)	2.3 (11, 4.6)	2.6 (11, 4.8)

[a]Backbone root-mean-square deviation of loop residues between the crystal apo structure and the crystal holo structure. [b]Backbone loop RMSD of the loop conformation closest to the comparison target (crystal apo or holo structure) among the 30 final bank conformations. The energy rank of this conformation and the backbone loop RMSD of the lowest-energy loop conformation in the final bank are shown in parenthesis.

They are discussed below in two parts, one about sampling native-like loop conformations and the other dealing with prediction of conformational changes.

Native Reconstruction. First, we assess whether the native-like loop conformations are sampled, for example, whether loop conformations close to the crystal structure (apo loop structure) are found when loop modeling is performed on the framework structure for apo forms (taken from the crystal apo structure). As can be seen from Table 2, loop conformations within 3 Å RMSD from the native structure were found in 10 out of 12 cases. This result is encouraging considering the fact that the target loops are rather long, ranging from 8 to 17 residues, and ligand molecules are not explicitly taken into account during loop modeling of holo structures. Long loops (longer than 8 residues) are hard to model accurately because the possible number of conformations grows exponentially as the loop length increases.[27] Six loops show particularly successful results, with RMSD below 2 Å from the native structure. For example, a loop structure with RMSD of 0.68 Å from the native was found when modeled in the holo framework of PTP, and a loop with RMSD of 0.80 Å from the native when modeled in the holo framework of TIM. These loop conformations are compared with their native structures in Figure 1. Although the main purpose of this work is to sample functionally different forms of the protein loop, the above result is also meaningful, reconfirming the capability of the loop modeling protocol on this test set.

Meanwhile, loops modeled on the apo framework of AEE and that of ENO show especially high RMSD values. The best loop conformation modeled on the apo framework of AEE has RMSD of 5.2 Å from the native, while the best RMSD on the holo framework is 1.9 Å, as shown in Figure 1(c) and (d). The same trend is found for ENO. The crystal structures of the loops of AEE and ENO in the apo state are very much extended towards solvent. This implies the difficulty of modeling protruding loops, probably because the solvation free energy is not considered accurately in the current energy function.

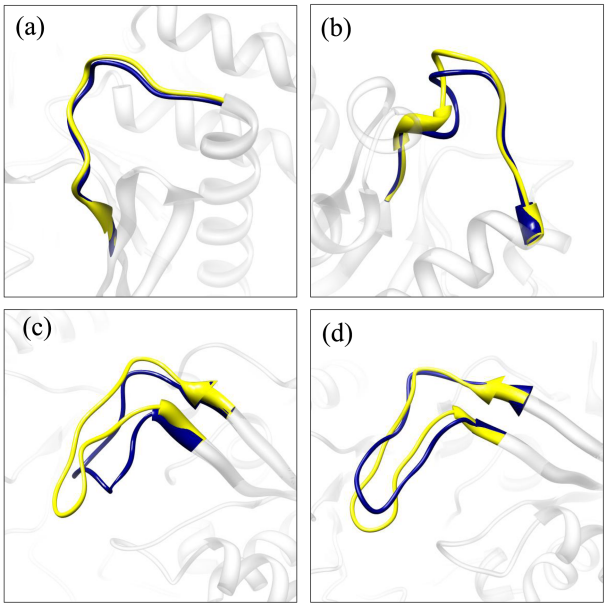

Figure 1. The best loop conformation (navy blue) obtained from loop modeling in the fixed native framework is compared with the native structure (yellow): (a) holo form of PTP (0.68 Å), (b) apo form of TIM chain A (1.6 Å), (c) apo form of AEE (5.2 Å), and (d) holo form of AEE (1.9 Å). Those loops largely extended outwards tend not to be sampled well currently, as in the case of (c).

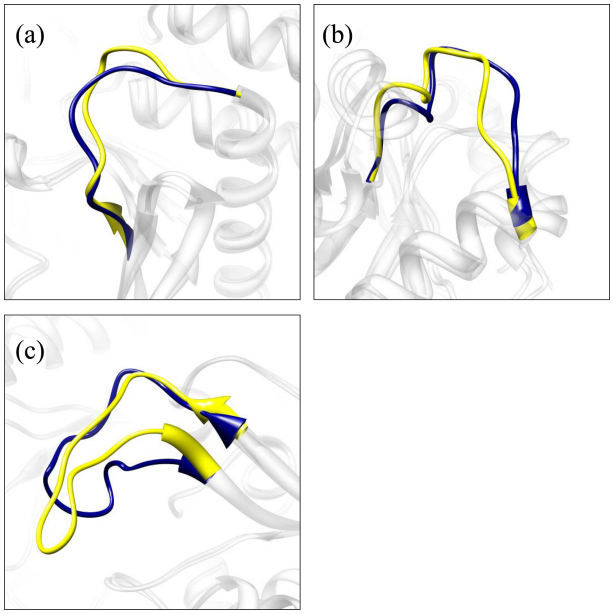

Figure 2. The best loop conformation (navy blue) obtained from loop modeling in the fixed framework in a different functional state is compared with the native structure (yellow): (a) modeled on the apo framework of PTP and compared to the crystal holo structure (1.5 Å) (b) modeled on the apo framework of TIM and compared to the crystal holo structure (1.7 Å) (c) modeled on the holo framework of AEE and compared to the crystal apo structure (4.8 Å).

Table 2 also presents the energy rank of the best loop conformations among the 30 conformations in the final bank. The average energy rank of the 12 reconstruction cases is 18, meaning that the current energy function cannot select the best conformation. This again illustrates the problem of the current energy function that has to be improved in future works. Nevertheless, the energy function used is reasonable enough to sample many native-like conformations in the final bank of 30 members, which is an extremely small size considering the extents of the overall conformational space.

Prediction of Conformational Changes. The main purpose of this paper is to assess the capability of the loop modeling protocol described above in sampling different functional forms. In this subsection, we examine whether holo (or apo) loop structures can be sampled when loop modeling is performed on the apo (or holo) framework. In Wong *et al.*'s study,[25] loop conformations were sampled in 3 stages: REMD (Replica Exchange Molecular Dynamics) simulations (5-ns production runs with 10 replicas), clustering of the simulation snapshots in the trajectories, and additional loop modeling around the cluster centers using PLOP (Protein Local Optimization Program). Restraints had to be applied

during the REMD simulations to have the loop structures stay near the binding pockets. They could find loop structures similar to the crystal holo structures when the crystal apo structures were used as initial structures, implying that "conformational selection" works. Here we explore whether a straightforward application of our protein loop modeling method alone can be used to sample conformations in different functional states. Our loop modeling method requires much less computational efforts (the number of energy evaluations is 8.8×10^5 on average) than those in Ref. 25 (the number of energy evaluations is 5×10^7 if only the REMD simulation step is considered).

It can be seen from Table 2 that in 4 out of 6 cases the best model loop conformations are within 3 Å RMSD from the holo crystal loop structures when loop modeling is performed on the apo framework. The case of PTP is one of the most successful cases in which the best conformation is 1.5 Å away from the crystal structure of the holo crystal loop structure when loops are modeled on the apo framework structure. Holo loops of TIM were also sampled well, with the best RMSD of 1.7 Å. These successful examples are illustrated in Figure 2(a) and (b).

For AEE and ENO, however, conformations in different functional forms were not sampled very well. The apo loop structure of AEE was not modeled accurately when modeled on the holo framework, with the best RMSD of 4.8 Å, as shown in Figure 2(c). This probably originated from the same kind of energy problem mentioned above, related to consideration of the solvation free energy for protruding loops.

Sampling ligand-bound structures starting from an unbound structure is especially meaningful because it implies that predicting ligand-bound conformations would be possible by applying the sampling method even when only the unbound form is known.[25] In Table 3, we compare the values of RMSD for the best loop conformations of the holo form when loops are modeled on the apo framework with those reported in Wong *et al.*[25] Our loop modeling result is encouraging: RMSD is better or comparable to the result from Wong *et al*. in 4 out of 6 cases. In 5 out of 6 cases, the RMSD is better than the RMSD between the crystal apo and holo structures. This result implies that the loop modeling protocol can be useful for flexible ligand docking. In protein-ligand docking studies, conformational flexibilities of receptor proteins are often taken into account by docking ligands onto multiple receptor conformations taken from crystal structures bound to different ligands or sampled from molecular dynamics simulations.[18,28,29] The current loop modeling method can be directly applied to such approaches. In addition to the big advantage of computational efficiency, our loop modeling method provides a selective pool of most promising loop conformations, and makes taking an additional step to choose representative structures from long simulation trajectories unnecessary. Therefore, it will be worthwhile to combine the loop modeling method used in this study with ligand docking methods to develop an efficient flexible protein-ligand docking program.

Table 3. Comparison of the sampling performance of our method and that of Wong *et al.*[23] when the loop conformations sampled on the apo framework are compared with the crystal holo loop structure

Protein	Loop RMSD, crystal apo *vs* holo (Å)	Best loop RMSD (Å)	
		This work	Wong *et al.*
PTP	3.1	1.5	1.4
AEE	6.2	3.2	1.3
TIM	4.1	1.7	1.9
ENO	5.0	4.1	1.8
GART	0.72	2.2	1.6
SRM	3.7	2.8	2.9

Conclusions

In this study we modeled loops on the frameworks of protein crystal structures by using a loop modeling method. The test proteins have flexible loops showing latching motions upon ligand binding. Therefore, we were able to investigate whether loop conformations close to those in a different functional state (for example, ligand-bound state) can be sampled by modeling on the framework of another state for which experimental structure is available (for example, ligand-free state). The results are encouraging, showing success in several cases even with a few orders of magnitude less computational efforts than a previous study. Overall, by using the loop modeling method, both apo and holo forms of ligand binding loops could be sampled reasonably well, implying the capability of predicting conformational changes of loops.

However, loops extended outwards in their apo conformation were not modeled well. This implies that further improvement of the solvation free energy term in the current energy function is necessary. A study to improve the current protocol is ongoing.

Since the loop modeling protocol is highly efficient, further studies making use of the protocol as a component can be pursued. Successful loop modeling methods which can sample conformations of different functional states can be applied to predicting conformational changes related to biological functions. In particular, developing flexible protein-ligand docking programs by combining the loop modeling method used in this study with existing ligand docking methods seems promising.

Acknowledgments. This work was supported by the KISTI Supercomputing Center (No. 305-20110004) and the Center for Marine Natural Products and Drug Discovery (CMDD), one of the MarineBio21 programs funded by the Ministry of Land, Transport, and Maritime Affairs. GRL acknowledges technical help from Lim Heo.

References

1. Rohl, C. A.; Strauss, C. E.; Chivian, D.; Baker, D. *Proteins* **2004**, *55*, 656.
2. Soto, C. S.; Fasnacht, M.; Zhu, J.; Forrest, L.; Honig, B. *Proteins*

2008, *70*, 834.
3. Zhu, K.; Pincus, D. L.; Zhao, S.; Friesner, R. A. *Proteins* **2006**, *65*, 438.
4. Lee, J.; Lee, D.; Park, H.; Coutsias, E. A.; Seok, C. *Proteins* **2010**, *78*, 3428.
5. Fiser, A.; Do, R. K.; Sali, A. *Protein Science: A Publication of the Protein Society* **2000**, *9*, 1753.
6. Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. *Annual Review of Biophysics and Biomolecular Structure* **2000**, *29*, 291.
7. Fernandez-Fuentes, N.; Zhai, J.; Fiser, A. *Nucleic Acids Research* **2006**, *34*, W173.
8. Fernandez-Fuentes, N.; Oliva, B.; Fiser, A. *Nucleic Acids Research* **2006**, *34*, 2085.
9. Daga, P. R.; Patel, R. Y.; Doerksen, R. J. *Current Topics in Medicinal Chemistry* **2010**, *10*, 84.
10. Taylor, J. C.; Takusagawa, F.; Markham, G. D. *Biochemistry* **2002**, *41*, 9358.
11. Murre, C.; McCaw, P. S.; Vaessin, H.; Caudy, M.; Jan, L. Y.; Jan, Y. N.; Cabrera, C. V.; Buskin, J. N.; Hauschka, S. D.; Lassar, A. B. *et al. Cell* **1989**, *58*, 537.
12. Auffinger, P.; Bielecki, L.; Westhof, E. *Chemistry & Biology* **2003**, *10*, 551.
13. Weliky, D. P.; Bennett, A. E.; Zvi, A.; Anglister, J.; Steinbach, P. J.; Tycko, R. *Nature Structural Biology* **1999**, *6*, 141.
14. Mendieta, J.; Ramírez, G.; Gago, F. *Proteins* **2001**, *44*, 460.
15. Ragona, L.; Fogolari, F.; Catalano, M.; Ugolini, R.; Zetta, L.; Molinari, H. *The Journal of Biological Chemistry* **2003**, *278*, 38840.
16. Carlson, H. A.; McCammon, J. A. *Molecular Pharmacology* **2000**, *57*, 213.
17. Teague, S. J. *Nature Reviews Drug Discovery* **2003**, *2*, 527.
18. Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. *Journal of the American Chemical Society* **2005**, *127*, 9632.
19. Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *Journal of Molecular Biology* **2004**, *337*, 227.
20. Shehu, A.; Clementi, C.; Kavraki, L. E. *Proteins* **2006**, *65*, 164.
21. Murphy, P. M.; Bolduc, J. M.; Gallaher, J. L.; Stoddard, B. L.; Baker, D. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 9215.
22. Park, H.; Seok, C. in revision. (Modeling of Unreliable Local Regions in Template-based Protein Models)
23. Lee, J.; Scheraga, H. A.; Rackovsky, S. *Journal of Computational Chemistry* **1997**, *18*, 1222.
24. Coutsias, E. A.; Seok, C.; Jacobson, M. P.; Dill, K. A. *Journal of Computational Chemistry* **2004**, *25*, 510.
25. Wong, S.; Jacobson, M. P. *Proteins* **2008**, *71*, 153.
26. Yang, Y.; Zhou, Y. *Proteins* **2008**, *72*, 793.
27. Zhao, S.; Zhu, K.; Li, J.; Friesner, R. A. **2011**, in press.
28. Zhao, Y.; Sanner, M. F. *Proteins* **2007**, *68*, 726.
29. Wong, C. F.; Kua, J.; Zhang, Y.; Straatsma, T. P.; McCammon, J. A. *Proteins* **2005**, *61*, 850.