Pairwise Structure Alignment Specifically Tuned for Surface Pockets and Interaction Interfaces [*]

Xuefeng Cui
Computer, Electrical and Mathematical Sciences and Engineering Division
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia

Hammad Naveed
Computer, Electrical and Mathematical Sciences and Engineering Division
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia

Xin Gao [†]
Computer, Electrical and Mathematical Sciences and Engineering Division
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia

Abstract

To detect and evaluate the similarities between the three-dimensional (3D) structures of two molecules, various kinds of methods have been proposed for the pairwise structure alignment problem [6, 9, 7, 11]. The problem plays important roles when studying the function and the evolution of biological molecules. Recently, pairwise structure alignment methods have been extended and applied on surface pocket structures [10, 3, 5] and interaction interface structures [8, 4]. The results show that, even when there are no global similarities discovered between the global sequences and the global structures, biological molecules or complexes could share similar functions because of well conserved pockets and interfaces. Thus, pairwise pocket and interface structure alignments are promising to unveil such shared functions that cannot be discovered by the well-studied global sequence and global structure alignments.

State-of-the-art methods for pairwise pocket and interface structure alignments [4, 5] are direct extensions of the classic pairwise protein structure alignment methods, and thus such methods share a few limitations. First, the goal of the classic protein structure alignment methods is to align single-chain protein structures (i.e., a single fragment of residues connected by peptide bonds). However, we observed that pockets and interfaces tend to consist of tens of extremely short backbone fragments (i.e., three or fewer residues connected by peptide bonds). Thus, existing pocket and interface alignment methods based on the protein structure alignment methods still rely on the existence of long-enough backbone fragments, and the fragmentation issue of pockets and interfaces rises the risk of missing the optimal alignments. Moreover, existing interface structure alignment methods focus on protein-protein interfaces, and require a "blackbox preprocessing" before aligning protein-DNA and protein-RNA interfaces. Therefore, we introduce the PROtein STucture

Alignment (PROSTA) family of pairwise structure alignment methods [1, 2] that address the fragmentation issue of pockets and interfaces, and automatically align interfaces between any types of biological complexes.

Our PROSTA structure alignment methods have two critical advantages comparing to existing structure alignment methods. First, our methods are completely sequence order independent, which is critical to the success of pairwise pocket and interface structure alignments. This is achieved by introducing contact groups that are not limited to backbone fragments, and by employing a maximum weighted bipartite matching solver from the beginning of the alignment process. In addition, our methods incorporate similarities of sequentially and structurally remote residues that potentially model the topology of the global structure. Comparing to existing methods that focus on local structure or whole sequence similarities, topological similarities are more reliable to find near-optimal structure alignments in the initial alignment state. As a result, a significant number of similar pockets and interfaces are newly discovered, and literatures also support that similar functions are shared between biological complexes in our cases studies.

The PROSTA web-server and source codes are publicly available at "http://www.cbrc.kaust.edu.sa/prosta/".

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Surface Pocket, Pocket Structure Alignment, Protein-Protein Interaction, Protein-DNA Interaction, Protein-RNA Interaction, Interface Structure Alignment

Keywords

Protein, DNA, RNA, Biological Complex, Pocket, Interaction, Interface, Structure, Alignment

1. REFERENCES

[1] X. Cui, S. C. Li, D. Bu, and M. Li. Towards reliable automatic protein structure alignment. In *WABI*, pages 18–32, 2013.

[2] X. Cui, H. Naveed, and X. Gao. Finding optimal interaction interface alignments between biological complexes. *Bioinformatics*, 31(12):i133–i141, 2015.

[3] J. Dundas, L. Adamian, and J. Liang. Structural signatures of enzyme binding pockets from order-independent surface alignment: a study of metalloendopeptidase and nad binding proteins. *Journal of Molecular Biology*, 406(5):713–729, 2011.

[4] M. Gao and J. Skolnick. ialign: a method for the structural comparison of protein-protein interfaces. *Bioinformatics*, 26(18):2259–2265, Sep 2010.

[5] M. Gao and J. Skolnick. Apoc: Large scale identification of similar protein pockets. *Bioinformatics*, page btt024, 2013.

[6] L. Holm and C. Sander. Protein Structure Comparison by Alignment of Distance Matrices. *Journal of Molecular Biology*, 233(1):123–138, Sept. 1993.

[7] E. Krissinel and K. Henrick. Secondary-structure matching (ssm), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D: Biological Crystallography*, 60(12):2256–2268, 2004.

[8] V. Pulim, B. Berger, and J. Bienkowska. Optimal contact map alignment of protein-protein interfaces. *Bioinformatics*, 24(20):2324–2328, Oct 2008.

[9] I. N. Shindyalov and P. E. Bourne. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering*, 11(9):739–747, Sept. 1998.

[10] L. Xie and P. E. Bourne. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc Natl Acad Sci U S A*, 105(14):5441–5446, Apr 2008.

[11] Y. Zhang and J. Skolnick. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic Acids Research*, 33(7):2302–2309, 2005.