

MAPPIS: Multiple Alignment of Protein-Protein Interfaces

Alexandra Shulman-Peleg¹, Maxim Shatsky¹, Ruth Nussinov^{2,3} and Haim J. Wolfson¹

Keywords: protein-protein interfaces, physico-chemical properties, structural alignment

1 Introduction.

Association and dissociation of protein molecules are crucial for most of the cellular processes. A *protein-protein interface* (PPI) is defined by a pair of regions of two interacting protein molecules that are linked by non-covalent bonds. Analysis and classification of PPIs may help to recognize certain interface binding organizations shared by different protein families, which may be important for the formation and stability of protein-protein complexes. These features may constitute targets for drug discovery and assist in predicting side effects.

Many functionally similar protein binding sites share no similar sequential patterns, thus, they cannot be recognized by commonly used methods like PROSITE. A sequence order independent structural alignment method was used by Keskin et al.[2] to classify all known PPIs according to C_α patterns. However, side chains play an important role in the interactions between molecules. Thus, a more accurate, atomic-level, analysis is required[4]. Recently, we have developed a method for alignment between a pair of PPIs[6]. We have applied it to classify[3] the interfaces of all the protein-protein complexes currently available in the PDB.

Here we present a novel method for multiple structural alignment of a set of PPIs. To the best of our knowledge this is the first method that can align and detect common patterns of interaction shared by a set of PPIs. Our main motivation is similar to the multiple sequence alignment thesis, a feature common to a number of proteins is probably functionally more significant than a similar feature found only between a pair of proteins. We present the method and show its application to the dataset created by Mintz et al.[3].

2 MAPPIS Method and Results.

We define a protein-protein *interface* (PPI) as an unordered pair of interacting binding sites from two non-covalently linked protein molecules. Each binding site is represented by a set of pseudocenters[4] which are points in 3D space that represent centers of potential interactions (hydrogen bond donors and acceptors, hydrophobic aliphatic and aromatic interactions). Only surface exposed pseudocenters that are within 4Å from the surface of the binding partner are considered (see Fig. 1).

Given a set of PPIs our goal is to find a set of 3D transformations that maximize the similarity of their physico-chemical and geometrical properties as well as the number of interactions between them. An *interaction* is defined by a pair of pseudocenters, one from

¹School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: {shulmana,maxshats,wolfson}@cs.tau.ac.il.

²Sackler Inst. of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: ruthnu@post.tau.ac.il.

³Basic Research Program, SAIC-Frederick, Inc, Lab. of Experimental and Computational Biology, Bldg. 469, Rm. 151, Frederick, MD21702, USA

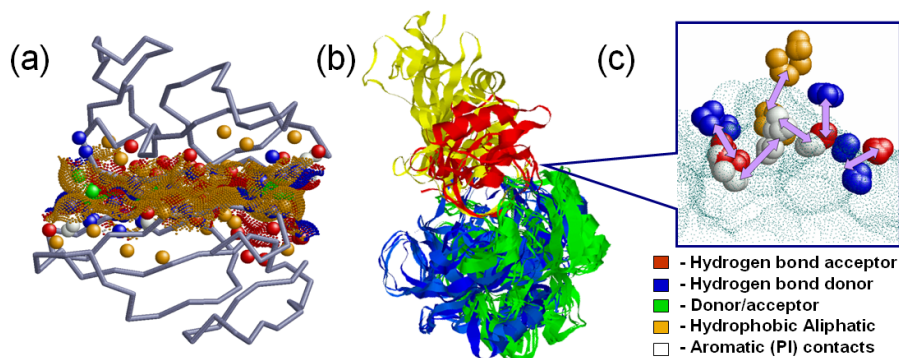


Figure 1: (a) Interface representation. Surfaces are in dots. Pseudocenters are colored as in (c). Backbones are not a part of the input to MAPPIS. (b) Alignment of 7 interfaces by MAPPIS: Trypsin-like serine proteases (4sgb, 1ppf, 1acb, 1an1, colored blue and red) and Subtilisin-like (1cse, 2sic, 1oyv, colored green and yellow). (c) The common interactions recognized by MAPPIS.

each side of the interface, with complementary properties within a distance that will allow their interaction. The problem of maximization of the number of similar interactions is NP-hard even for a pair of interfaces (this is similar to the 3D k -partite matching[5]). Our method includes two major computational steps. First, we generate candidate 3D transformations that align the interfaces. The transformation is defined between a pair of interactions from one interface and a pair of interactions from the other. In the second step, our aim is to find a combination of interface alignments that gives the highest score. This problem is NP-Hard[1]. Our solution is based on a branch-and-bound approach which in practice is very efficient, thus, the running times are usually several minutes (on a 2Ghz PC) on the tested examples. The overall scheme guarantees to find an approximate solution for the geometrical problem. However, in practice we devised a scoring function that measures similarity of the physico-chemical properties between the matched pseudo-centers as well as the matched interactions (manuscript in preparation). We have applied our method to analyze the interactions shared by PPI clusters created by Mintz et al[3]. Fig. 1 presents the interactions shared by the interfaces of serine proteases, which is the most well known example of functionally similar interfaces created by proteins with different folds.

References

- [1] T. Akutsu and M. M. Halldorson. On the approximation of largest common subtrees and largest common point sets. *Theoretical Computer Science*, 233:33–50, 2000.
- [2] A. Keskin, C. H. Tsai, H. J. Wolfson, and R. Nussinov. A new, structurally non-redundant, diverse dataset of protein-protein interfaces and its implications. *Prot. Sci.*, 13(4):1043–55, 2004.
- [3] S. Mintz, A. Shulman-Peleg, H. J. Wolfson, and R. Nussinov. Generation and analysis of a protein-protein interface dataset with similar chemical and spatial patterns of interactions. *Proteins*, *accepted*.
- [4] S. Schmitt, D. Kuhn, and G. Klebe. A new method to detect related function among proteins independent of sequence or fold homology. *J. Mol. Biol.*, 323:387–406, 2002.
- [5] M. Shatsky, A. Peleg-Shulman, R. Nussinov, and H.J. Wolfson. Recognition of binding patterns common to a set of protein structures. In *RECOMB 2005*.
- [6] A. Shulman-Peleg, S. Mintz, R. Nussinov, and H.J. Wolfson. Protein-protein interfaces: Recognition of similar spatial and chemical organizations. In I. Jonassen and J. Kim, editors, *Workshop on Algorithms in Bioinformatics*, pages 194–205. 2004. LNCS, 3240.