Protein docking algorithm based on structural homology with protein-protein interfaces

T. Kirys, A.V. Tuzikov

United Institute of Informatics Problems, National Academy of Sciences of Belarus,

Surganova, 6, Minsk, tuzikov@newman.bas-net.by

An original protein docking algorithm based on structural similarity with experimentally determined protein-protein interfaces is proposed. A pair of proteins is aligned to an interface from a database of protein-protein interfaces. The alignment is performed by dynamic programming on distance matrices of protein and interface. The idea of algorithm is shown in fig. 1.

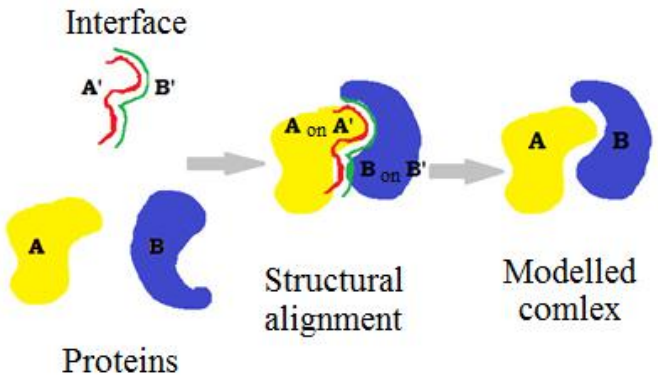

Fig 1. The modeled complex AB predicted based on interface $A'B'$ is obtained after superposition of protein A on interface A' and protein B on interface B'

The assumption behind this approach is that interaction information can be extrapolated from one complex structure to the homologous of interacting proteins. In protein-protein interface there are patches of continuous segments of protein polypeptide chain. This observation is used for finding similar parts in proteins. We also assume that if two structures are similar then their distance matrices (containing distances between all *C*α-atoms of proteins) are correlated.

Continuous segments (patches) with length not less than predefined value ($l \geq 3$) were selected in interface A': $\overline{A'} = \bigcup_{i-1}^{k} A'_i$. To align protein A with interface A', patches $A_i \subset A$,

such that length of A_i is equal to the length of A_i' and correlation between their distance matrices is maximum are found:

$$\text{Correlation}\,(M_{\bigcup_{i=1}^{k} A_i}, M_{\bigcup_{i=1}^{k} A_i'}) \to \max.$$

To do so a heuristic based on dynamic programming is proposed. When the alignment if found, protein and interface are superimposed based on the movement minimizing distance between corresponding points [1]. The similarity between the aligned protein A and interface A' is calculated with TM-score [2].

The proposed algorithm was tested on protein benchmark set 3.0 [3]. For each protein pair ten best models with the highest TM-score were selected. The number of good predictions based on CAPRI [4] criteria was higher on 20% compared to ZRANK [5], tested on a similar benchmark.

Finding particularly spatial similar continuous segments reduces searching time drastically in comparison with hashing algorithms and adequately reflects the nature of protein interactions. The role of the proposed algorithm based on structural homology with protein-protein interfaces will increase with the increase of number of experimentally determined protein complexes.

1. R. Horn (1985) Matrix Analysis (Cambridge University Press).

2. Y. Zhang, J. Skolnick (2004) Scoring function for automated assessment of protein structure template quality, *Proteins*, **57:**702–710.

3. Y. Gao, D. Douguet, A. Tovchigrechko, I.Vakser (2007) Dockground system of databases for protein recognition studies: Unbound structures for docking, *Proteins*, **69**: 845-851.

4. S. Wodak, R. Méndez (2004) Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications, *Curr Opin Struct Biol*, **14**(2): 242-249.

5. B. Pierce, Z. Weng (2007) ZRANK: Reranking protein docking predictions with an optimized energy function, *Proteins: Structure, Function, and Bioinformatics*, **67**: 1078-1086.