Introduction

Protein-protein docking algorithms aim to predict how two proteins interact with each other to form a complex. Docking algorithms need to fulfill two main tasks: (1) sampling the possible relative positions of the two proteins and (2) computing the interaction energy at each position. Although the docking problem has been studied for over 25 years, doing this accurately and thoroughly remains a computationally expensive task. We are therefore developing a new algorithm to face the protein docking problem in a novel and efficient way.

We are interested in performing an exhaustive search of the rigid-body docking space, where the positioning of the two proteins is not driven by any prior knowledge (e.g., from homology), while still using an accurate force field interaction energy model. In order to reduce the computational cost of atomistic force-field models, we use a coarse-grained (CG) bead model taken from ATTRACT [1]. However, a naive energy calculation for every trial orientation would still cost $O(N^2)$ energy evaluations in the number of beads. We are therefore developing a method to detect the locations of all possible clashes before performing any energy calculations, thus allowing us to avoid calculating energies for many millions of useless trial orientations.

Based on a preliminary study of protein-protein interfaces in complexes from the Protein Docking Benchmark [2], we found that a large number of interfaces contain at least one pair of CG beads at almost their optimal distance. Therefore, our idea is to perform a series of restricted docking searches in which one surface bead from each docking partner is placed in contact at the coordinate origin. This leaves a 3D rotational search, in which ligand may rotate around a fixed receptor.

Of course, a full docking search requires all possible pairs of surface beads to be placed together. However, within each rotational sub-problem, we can exploit the fact that rotations do not change any distances from the origin for each selected pair of beads, as illustrated on the left.

For the translational part of a docking search, we consider all possible pairs of receptor and ligand surface beads in turn, and for each pair we translate the ligand and receptor so that the selected beads (here shown in yellow) are in perfect contact at the origin with their centres of mass are aligned on the z-axis. This absorbs 3 translational degrees of freedom, and leaves a pure 3D rotational search problem about the origin for each selected pair of beads, as illustrated on the left.

Before calculating ATTRACT bead energies, we use the bead coordinates to calculate a list of rotations that will cause steric clashes. These are rotations that should be avoided during the docking search. A 3D representation of this “clash map” is shown on the right. It can be seen that only a small region of the search space is not forbidden by steric clashes.

Conclusions

We have found that many protein-protein complexes have at least one pair of beads in near-optimal contact. Thus, using pairs of surface beads to initialise a docking search is a valid heuristic for treating 3D translations.

The algorithm described here will allow us to perform an exhaustive recursive docking search of 3D rotation space. Thus, in principle, our approach will never miss any local energy minima, and is guaranteed to find the global minimum of the ATTRACT energy function for each pair of starting beads.

Our algorithm is still under development. Nonetheless, we believe it could be faster than the conventional gradient minimisation that is currently used in ATTRACT. It also promises to be more accurate if we can train a new scoring function to optimise the local minima among the sampled poses.

Eventually, we will test our approach on the Protein Docking Benchmark.

References


Acknowledgements

Maria Elisa Ruiz-Echartea is supported by a doctoral bursary (COBEN-0) from Inria Nancy Grand Est.