Comparison of protein-protein docking prediction and optimization methods

Caiwei Wang, Xiaokai Qian, Sean Lander,
Haipei Fan, Puneet Gaddam, Brett Koonce

University of Missouri - Columbia

April 7, 2014

1 Abstract

In this paper we applied three docking tools (Zdock, Cluspro and Hex) to two CAPRI targets (T50 and T53), improved prediction accuracy using docking optimization and assessed the performance using a few complementary tools (RosettaDock and RMSD). We first selected the receptor and ligand PDBs of T50 and T53. After pre-processing we created decoys using Zdock, Cluspro and Hex, and then scored and optimized them using RosettaDock. We then compared the results between before and after optimization by measuring both their RosetttaDock scores and the RMSD of the docking interface. We used these results to analyze performance across the tools we used in order to better understand the tools and how they may complement or hinder one another. Ultimately, we visualized the top model from each decoy group.

2 Introduction

Protein-protein docking is used to predict the structure of a protein complex given a receptor and a ligand. It is very useful in drug creation and disease prevention. Because the computational power needed for predictions can be high, many unique algorithms and methods have been created to both predict and score docked structures.

For this method comparison we will be using three prediction servers, each of which use a variation on the Fast Fourier Transform method for prediction and their own unique methods for rankings. We then standardize these predictions using RosettaDock's scoring function. Once the top 10 are selected we will apply Rosetta's optimization method, which allows for a relaxing of side-chain constraints, before scoring and sorting them a second time. We will then use RMSD, a traditional scoring method when native conformation is known, of the predicted interface in order to analyze and compare the different methods of prediction, optimization, and scoring.

3 Methods

3.1 Pipeline

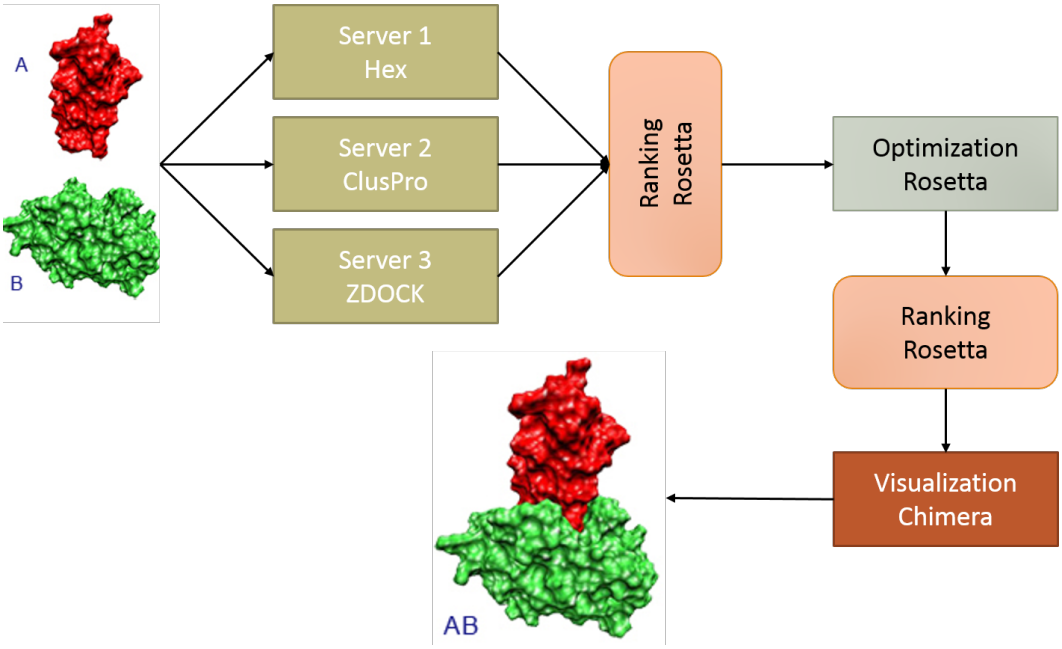

Figure 1: An overview of the comparison pipeline

1. Decoy Creation: Decoys are created using an ensemble of docking servers. This list can be extended beyond the servers listed, as well as include locally created predictions.
2. Ranking: RosettaDock's prediction scoring is used to normalize the results, as each tool uses its own scoring and ranking system. The decoys are then ranked based on their new scores and the top 10 continue on for optimization.
3. Optimization: Once again RosettaDock is used, this time for its optimization capabilities. The top 10 decoys are fed into Rosetta and the results are fed into a second round of scoring.
4. Final Selection and Scoring: The optimized decoys are scored by RosettaDock once again and ranked accordingly. The top scoring model is chosen as the best candidate.
5. Visualization: Chimera is used to visualize the best model overlaid with the native conformation, as well as the top 10 before and after optimization.

3.2 CAPRI Target and Docking Tool Selection

We selected the following template-free targets from the CAPRI database to build a model for: T53, T50. These targets were chosen due to their high predictability by many of the tools used, allowing for a balanced comparison of predictive capacity on easy targets.

The tools selected represent some of the different methods currently employed for docking prediction and rankings - Fast Fourier Transforms, Spherical Polar Fourier, filter by clustering and data-driven rankings - as well as how base decoys compare to their optimized counterparts.

3.3 Pre-Processing

3.4 Decoy Creation

3.4.1 Hex

3.4.2 ClusPro

ClusPro is a web server which uses an automated rigid-body docking and discrimination algorithm that rapidly filters docked conformations, and ranks them based on their clustering properties.

After signing up for an account and logging into the server, you can create a docking job with a receptor and ligand by uploading the PDB files of T50 and T53. You may optionally choose what chains to use in docking. After that, you can view the queue status for the decoys completed. It will generate the number of top models you choose.

3.4.3 ZDOCK

ZDOCK uses a mixture of Fast Fourier Transform and data-driven scoring/ranking in order to create and sorts its decoy. While this allows for very fast decoy creation, it results in a stochastic algorithm, so multiple runs will always generate the same results. The ZDOCK server produces 500 decoys on a run, but only the top 10 of those are available for download, making it hard to build up an appropriately sized sample group.

3.5 Decoy Optimization

1. Decoy Ranking
2. Optimization

4 Results

4.1 Interface Creation

4.2 Scoring Methods - RMSD vs RosettaDock

1. Before Optimization
2. After Optimization

4.3 Visualization

5 Conclusion

6 Citations

We thank the following tools and papers: