

# 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

## 2 Introduction

## 3 Methods

### 3.1 Pipeline

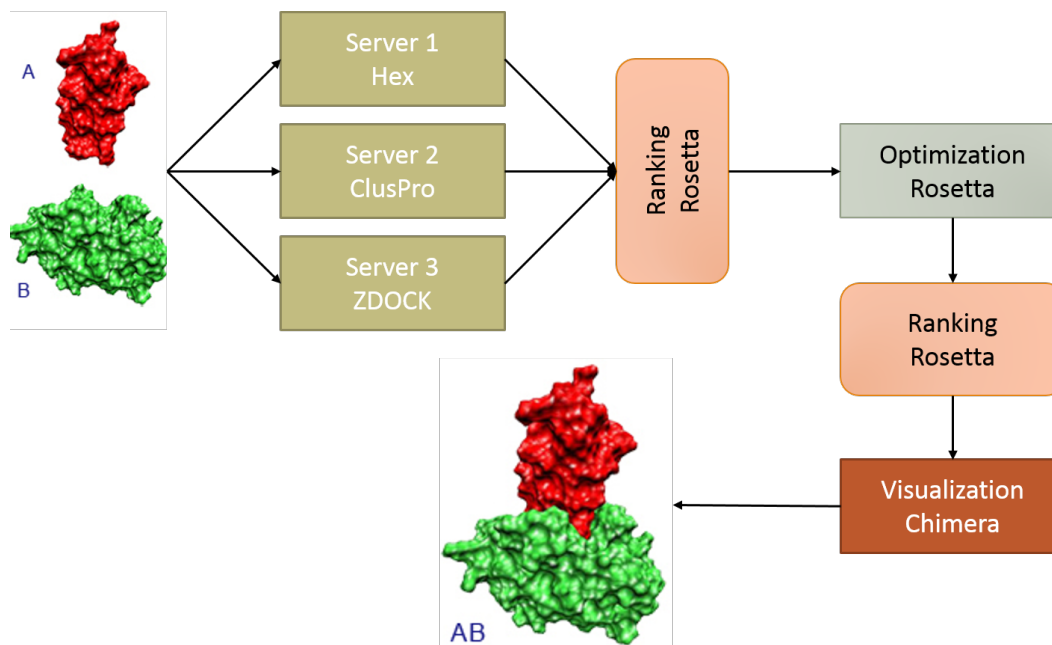


Figure 1: An overview of the comparison pipeline

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**

#### **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: