이학박사 학위논문

Development of Computational Methods for Predicting Protein Interactions

단백질 상호작용 예측을 위한 계산 방법 개발

2018 년 8 월

서울대학교 대학원

화학부 물리화학 전공

백 민 경

ABSTRACT

Development of Computational Methods for Predicting Protein Interactions

Minkyung Baek

Department of Chemistry

The Graduate School

Seoul National University

Proteins are important components of living organisms and are involved in many biological processes. The biological functions of proteins result from their molecular interactions with other molecules such as metal ions, small organic compounds, peptides, lipids, nucleic acids, or other proteins. Therefore, computational approaches to predict interactions between proteins and other molecules are useful to understand protein functions in molecular level and to design molecules that regulate protein functions. Specifically, ligand binding site prediction methods can be used to identify druggable sites of target proteins while protein-ligand docking techniques can contribute to identifying hit or lead compounds and optimizing lead compounds during structure-based drug discovery process. In addition, because a large fraction of cellular proteins self-assemble to form symmetric homo-oligomers to play their biological roles, computational methods to predict homo-oligomer structures can also contribute to drug discovery

process by providing atomic details of target oligomer interfaces.

In this thesis, three computational methods developed to predict protein interactions are introduced: (1) an improved metal and organic molecule binding site prediction method, (2) a protein-ligand docking method with an improved hybrid scoring function and a sampling algorithm utilizing predicted binding hot spot information, and (3) a protein homo-oligomer modeling method using bioinformatics and physical chemistry approaches. All methods described here show high performances in benchmark tests when compared to other state-of-the-art programs. These benchmark results suggest that computational approaches introduced in this thesis can be applied to *in silico* drug discovery process.

keywords: protein interaction prediction, ligand binding site prediction, protein-ligand docking, docking scoring function, protein homo-oligomer structure prediction

Student Number: 2013-20267

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Proteins are bio-macromolecules consisting of amino acids and have one or several chains. Like other bio-macromolecules such as nucleic acids and polysaccharides, proteins are important components of living organisms and are involved in many biological processes including enzymatic activities, metabolism, and signal transductions (Kristiansen 2004; Negri et al. 2010; Pawson and Nash 2000). The biological functions of proteins result from their molecular interactions with other molecules such as metal ions, small organic compounds, lipids, peptides, nucleic acids, or other proteins. Structural knowledge of protein complexes is required to understand how proteins and various molecules work together to fulfil their tasks. Computational approaches to predict ligand binding site and protein-ligand/protein-protein complex structure are powerful tools to gain such structural knowledge and improve our understanding of protein function, reducing time and labor for investigating protein interactions experimentally. The major scope of this thesis is discussing the developments of computational methods to predict protein interactions, especially to predict ligand binding site, protein-ligand complex structure, and protein homo-oligomer structures.

Typically, proteins interact with other molecules by binding them at specific sites. Therefore, identification of the binding sites on the three-dimensional protein surfaces can be an important step for inferring protein functions (Campbell et al. 2003; Kinoshita and Nakamura 2003) and for designing novel molecules that control protein functions (Laurie and Jackson 2006; Sotriffer and Klebe 2002). In the last few decades, various computational methods have been developed to predict the possible ligand binding sites in proteins (Brylinski and Skolnick 2008, 2011; Hendlich et al. 1997; Heo et al. 2014; Roche et al. 2011; Yang et al. 2013b).

Those methods are based on geometry, energy, evolutionary information, or combinations of them (Tripathi and Kellogg 2010). Methods utilizing available experimentally resolved structures of homologous protein-ligand complexes were proven to be successful in predicting binding sites in the community-wide blind prediction experiments (Gallo Cassarino et al. 2014; Lopez et al. 2009; Lopez et al. 2007; Schmidt et al. 2011). Our group developed a non-metal ligand binding site prediction program called GalaxySite using information from similar protein-ligand complexes in the context of protein–ligand docking (Heo et al. 2014). By combining information from template and molecular docking techniques, GalaxySite made more precise binding site predictions than simple superposition-based methods for small organic molecules. In **Chapter 2**, an extended version of GalaxySite named GalaxySite2 is introduced. In GalaxySite2, a metal binding site prediction is newly incorporated and the overall performance is improved by using BioLiP database (Yang et al. 2013a) with improved ligand and template selection method and introducing a new re-scoring function to select final binding sites.

Computational protein-ligand docking is used to predict ligand binding poses for given binding sites of proteins. The conformational space of the protein-ligand complex is explored in order to compute energetically stable conformations during the protein-ligand docking process. The success of a protein–ligand docking program depends on the program's performance on two famous, but still unsolved problems: (1) scoring binding poses to discriminate near-native binding poses from non-native poses and (2) sampling a wide range of conformations covering conformational space enough.

In **Chapter 3**, a new docking scoring function developed to improve its ability to discriminate correct binding poses is introduced. In many cases, scoring functions trained to reproduce experimental binding affinities have been used to

score both binding poses and binding affinities (Bohm 1998; Eldridge et al. 1997; Huey et al. 2007; Korb et al. 2009; Morris et al. 1998; Trott and Olson 2010; Wang et al. 2002). To improve scoring ability on binding pose discrimination, I adopt a hybrid of different types of scoring functions to take advantage of different scores and optimized weight factors by an iterative parameter optimization procedure that trains the scoring function to favor near-native poses over non-native poses.

A new approach tackling sampling problem in protein-ligand docking, especially for large flexible ligands, is introduced in **Chapter 4**. It has been shown that docking small ligands with 6 or fewer rotatable bonds is in general very accurate (Plewczynski et al. 2011). However, as the dimensionality of the search space increases with large ligands, accurate docking of large flexible ligands becomes very challenging. To tackle this problem, it is essential to develop an efficient conformational space sampling algorithm. I developed GalaxyDock-Frag, a new approach to improve the sampling ability of protein-ligand docking program by utilizing predicted binding hot spot information. By predicting fragment binding hot spot information using FFT-based fragment docking and utilizing such binding hot spot information, the translational and rotational degrees of freedom could be searched efficiently resulting in improvements of docking performance for large flexible ligands.

A large fraction of cellular proteins self-assemble to form symmetric homo-oligomers with distinct biochemical and biophysical properties (Andre et al. 2008; Goodsell and Olson 2000; Poupon and Janin 2010). For example, ligand-binding sites or catalytic sites are located at oligomer interfaces in many proteins (Snijder et al. 1999; Ali et al. 2010; Pidugu et al. 2016), and oligomerization is often necessary for effective signal transduction through membrane receptor proteins (Heldin 1995; Stock 1996) and selective gating of channel proteins

(Clarke and Gulbis 2012). Therefore, knowledge of the homo-oligomer structure is essential for understanding the physiological functions of proteins at the molecular level and for designing molecules that regulate the functions. In **Chapter 5**, a method to predict protein homo-oligomer structures called GalaxyHomomer is described.

The methods described in this thesis can be applied to not only studying proteins' function but also discovering new drug molecules for target proteins. Druggable sites in a target protein can be detected by developed ligand binding site prediction method named GalaxySite2, while structure-based virtual screening can be done by newly developed protein-ligand docking methods (GalaxyDock BP2 Score and GalaxyDock-Frag) which predict protein-ligand complex structures. If the target protein forms a homo-oligomer structure, its structure and interfaces can be predicted by GalaxyHomomer program introduced in this thesis.

Chapter 2. Prediction of Metal and Small Organic Molecule Binding Sites in Proteins

2.1. Introduction to Binding Site Prediction

Proteins are involved in numerous biological processes such as enzymatic activities and signal transductions (Kristiansen 2004; Negri et al. 2010; Pawson and Nash 2000). The biological functions of proteins result from their molecular interactions with other molecules such as metal ions, small organic compounds, lipids, peptides, nucleic acids, or other proteins. Typically, proteins interact with other molecules by binding them at specific sites. Therefore, identification of the binding sites on the three-dimensional protein surfaces can be an important step for inferring protein functions (Campbell et al. 2003; Kinoshita and Nakamura 2003) and for designing novel molecules that control protein functions (Laurie and Jackson 2006; Sotriffer and Klebe 2002) or designing new proteins with desired interaction properties (Damborsky and Brezovsky 2014; Feldmeier and Hocker 2013). Various methods have been developed to predict ligand binding sites of proteins from protein sequences or structures (Brylinski and Skolnick 2008, 2011; Hendlich et al. 1997; Heo et al. 2014; Roche et al. 2011; Yang et al. 2013b). Those methods are based on geometry, energy, evolutionary information, or combinations of them (Tripathi and Kellogg 2010). Methods utilizing available experimentally resolved structures of homologous protein-ligand complexes were proven to be successful in predicting binding sites in the community-wide blind prediction experiments (Gallo Cassarino et al. 2014; Lopez et al. 2009; Lopez et al. 2007; Schmidt et al. 2011). Those methods assume that binding sites and interactions at the binding sites are conserved among homologs and thus predict binding sites of target proteins by transferring the available binding information on homologs. However, such

methods based on evolutionary information may not be sufficient to predict interactions at the binding sites in atomic detail, and physicochemical interactions may have to be considered in addition.

To predict not only the binding site residues but interactions at the binding sites in atomic detail, a program named GalaxySite was developed (Heo et al. 2014). GalaxySite uses information from similar protein-ligand complexes in the context of protein–ligand docking. By combining information from template and molecular docking techniques, GalaxySite made more precise binding site predictions than simple superposition-based methods for non-metal ligand. In this chapter, an extended version of GalaxySite program named GalaxySite2 is introduced. In GalaxySite2, a metal binding site prediction is newly incorporated and the overall performance is improved by using BioLiP database (Yang et al. 2013a) with improved ligand and template selection method and introducing a new re-scoring function to select final binding sites.

GalaxySite2 has been tested on the following test sets: 238 metal-binding proteins with known experimental structures, 46 homology models of metal-binding proteins, and 420 targets of the ligand binding site prediction category from the continuous automated model evaluation server (CAMEO) released between June 13th and August 29th, 2014. In these tests, the performance of GalaxySite2 was superior or comparable to other state-of-the-art prediction methods.

2.2. Methods

2.2.1. Overall Procedure

The GalaxySite2 program predicts metal and small organic molecule-binding sites of a given protein by protein–ligand docking, as shown in **Figure 2.1**. From the BioLiP database (Yang et al. 2013a), up to three ligands are extracted from the protein–ligand complex structures of similar proteins detected by HHsearch (Soding 2005). Ligand binding poses are then predicted by Monte-Carlo search for metal ions or LigDockCSA (Shin et al. 2011) for small organic molecules. After ligand-binding poses are predicted for each selected ligand, the lowest energy ligand binding poses are re-scored by confidence score, and the final results are reported.

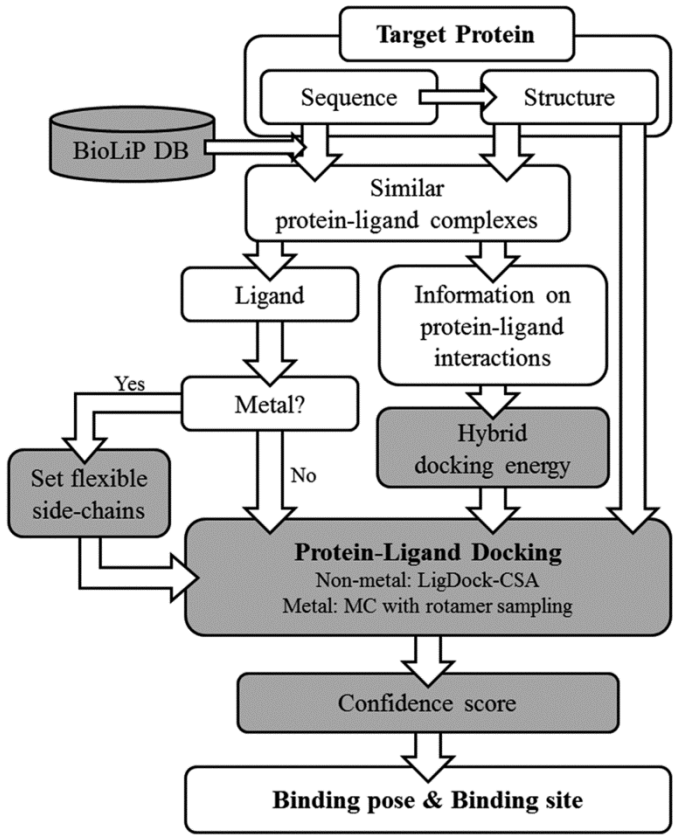


Figure 2.1. The overall procedure of GalaxySite2

2.2.2. Bound Ligand and Template Complex Selection

Ligands to be docked to the target protein structure are selected using experimental structures of template proteins with bound ligands. The template search is performed against the protein structure database 'pdb70' with a maximum mutual sequence identity of 70% using HHsearch. Among top 50 proteins from the HHsearch results, proteins whose structures are very different from that of the target protein are filtered out. Among the remaining proteins, the proteins in BioLiP database, which is the curated database containing biological relevant protein-ligand complex structures, are selected as template candidates.

All the ligands in template candidates are scored according to the Ligand Score as represented in Eq. (2.1) and up to three ligands with the highest score and templates containing selected ligands are used in the docking calculations.

$$\text{Ligand Score} = \sum_{templ} (w_{TM}\,\text{TMscore} + w_{BDT}\,\text{BDTalign} + w_{PSSM}\,\text{PSSMscore}) \quad (2.1)$$

The TMscore (Zhang and Skolnick 2005) is used to measure global structural similarity while BDTalign score (Roche et al. 2012) is used to measure local structural similarity of binding sites between target and template proteins. BDTalign score is calculated using following equation:

$$\text{BDTalign} = \frac{\sum_{i=0}^{N_p} \max(S_{ij})}{\max(N_p, N_t)}, \qquad S_{ij} = \frac{1}{1+(d_{ij}/d_0)^2} \quad (2.2)$$

where N_p and N_t are the number of residues within 5 Å from any ligand atom in target and template proteins, respectively. d_{ij} denotes distance between two aligned residues in target and template proteins and d_0 is set to 3 Å. PSSMscore is used to

measure sequence similarity of binding sites between target and template proteins defined as following:

$$\text{PSSMscore} = \sum_{i=0}^{N_p} PSSM_{res_i}(aa_j) \tag{2.3}$$

where N_p is the number of residues within 5 Å from any ligand atom in the target protein. res_i denotes the position of the i^{th} residue in binding sites, and aa_j is the amino acid of the residue j in template aligned to the i^{th} residue in the target protein. The weight factors were trained on the CAMEO binding site prediction targets released between 16th August and 16th November 2013. The resulting weight factors are as follows: $w_{TM} = 0.3$, $w_{BDT} = 0.3$, $w_{PSSM} = 0.4$.

2.2.3. Binding Pose Refinement Using Molecular Docking

2.2.3.1. Optimization of Small Organic Molecule Binding Poses

LigDockCSA protein-ligand docking program (Shin et al. 2011) is used to optimize binding pose of small organic molecules as described in original GalaxySite paper (Heo et al. 2014). The flexibility of the ligand is considered only during docking. A pool of 30 conformations is first generated by perturbing the initial conformations obtained from superposed template ligand poses. The pool is then evolved by generating trial conformations, gradually focusing on narrower regions of lower energy in the conformational space. Out of the final pool of 30 structures, the pose with the lowest docking energy in the largest cluster is selected as a representative binding pose.

The energy function used for docking is expressed as follows:

$$E = E_{\text{AutoDock}} + 1.1E_{\text{Restraint}} \quad (2.4)$$

where $E_{AutoDock}$ is the same as the AutoDock3 energy function (Morris et al. 1998) except that the maximum energy value for each interacting atom pair is set to 1.0 kcal/mol to tolerate steric clashes that may be caused by inaccurate protein model structures or ligand-unbound structures. The restraint term $E_{Restraint}$ is derived from the template structures that contain the selected ligand.

2.2.3.2. Optimization of Metal Binding Poses

Because side-chain orientations are really important to make proper coordination geometry of a metal ion, conformations of selected side-chains are sampled using preferred discrete rotamers (Dunbrack 2002) during the binding pose optimization step. Side-chains in the binding site are set to be flexible when the angles between two orientation vectors in target and template proteins are larger than 30 degrees. The orientation vector of side-chain is defined by Cα coordinates and the center of mass of the electron donors.

Side-chain conformations and coordinates of metal ion are optimized using Monte Carlo minimization (Monte Carlo simulation at T=0). In each minimization step, one of the flexible side-chains is perturbed randomly followed by optimization of metal coordinates using Monte Carlo minimization. This minimization step is repeated 100 times. From the 30 independent runs, the pose with the lowest docking energy in the largest cluster is selected as a representative binding pose.

The energy function used for docking is expressed as follows:

$$E_{met} = w_{\text{Restraint}} E_{\text{Restraint}} + w_{Coord} E_{Coord} + w_{qq} E_{qq} + E_{clash} \qquad (2.5)$$

where $E_{Restraint}$ is restraint term derived from the template structures that contain selected metal ion. The E_{Coord} term is empirical scoring function to describe coordination geometry, the E_{qq} term is Coulomb interaction energy, and the E_{Clash} is the term for considering clashes between metal ion and side-chains.

2.2.4. Final Model Selection and Binding Site Residue Prediction

After ligand-binding poses are optimized for each selected ligand, the optimized ligands are re-scored based on confidence score as follows:

$$\text{Confidence Score} = w_{lig} S_{lig} + w_{dock} S_{dock} + w_{sim} S_{sim} \qquad (2.6)$$

The confidence score consists of normalized ligand score (S_{lig}), normalized docking energy (S_{dock}) and template similarity score (S_{sim}). Normalized ligand score is calculated by dividing each ligand score used in ligand/template selection step by the maximum ligand score so that it has a range from 0 to 1. To correct size-dependency of the docking energy, docking energy is divided by a cube root of its size before normalizing by maximum value. If docking energy is larger than 0, which means that corresponding ligand binding is energetically unfavorable, that ligand is filtered out. Template similarity score measures how resulting binding poses are similar to those in templates. It is defined using logistic function as Eq. (2.7), where Z means the differences in the shortest distances between binding site residue and ligand atom in model and those in template.

$$S_{sim}(Z) = \frac{1 - 1/2^{1/0.15}}{\left(1 + \exp(-2/|Z|)\right)^{1/0.15}} + \frac{1}{1 - 1/2^{1/0.15}} \quad (2.7)$$

The weight factors, w_{lig}, w_{dock}, and w_{sim} are determined to 0.1, 0.45, and 0.45, respectively. For each ligand having confidence score larger than 0.5, residues are considered in binding site if the distance between any heavy atom in the residue and any ligand atom is less than the sum of the van der Waals radii plus 0.5 Å.

2.2.5. Test Sets and Evaluation Metrics for Binding Site Prediction

Three test sets were used to assess the performance of the GalaxySite2. The 'native metalloprotein set' was compiled from the FINDSITE-metal benchmark set (Brylinski and Skolnick 2011) consisting of 238 protein-metal complexes. It was used to compare the performance of the GalaxySite2 with the state-of-art metal binding site prediction program named FINDSITE-metal. The 'homology model set', containing CASP8 and CASP9 metal binding site prediction targets (16 targets) and a subset of native metalloprotein set (30 targets), was employed to verify the effectiveness of the side-chain sampling during the metal binding site prediction. The 'CAMEO benchmark set', consisting 420 CAMEO binding site prediction targets released between 13th June and 29th August 2014, was used to compare the overall performance of GalaxySite2 with the other binding site prediction servers.

To measure the performance of binding site prediction, three evaluation metrics, accuracy, coverage, and Matthew's correlation coefficient (MCC), are calculated as:

$$\text{Accuracy} = \frac{TP}{TP + FP} \tag{2.8}$$

$$\text{Coverage} = \frac{TP}{TP + FN} \tag{2.9}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{2.10}$$

where TP, TN, FP, and FN denote the number of true positive, true negative, false positive, and false negative predictions in binding site residue predictions. In addition to three metrics, the metal ion displacement in final predicted binding poses are employed to compare the performance of GalaxySite2 with FINDSITE-metal.

2.3. Results and Discussions

2.3.1. Performance of Metal Binding Site Prediction

To test the performance of metal binding site prediction of GalaxySite2, the native metalloprotein set and the homology model set were used as described in **section 2.2.5.** The overall performance of GalaxySite2 on these two test sets are summarized in **Table 2.1**.

Table 2.1. The overall performance of GalaxySite2 in the metal binding site prediction

Evaluation metric	Native metalloprotein set	Homology model set
MCC	0.852	0.829
Coverage	0.857	0.830
Accuracy	0.870	0.857
Metal Displacement	0.614 Å	1.147 Å

The performance of GalaxySite2 in the metal binding pose prediction is compared with that of FINDSITE-metal program in **Figure 2.2**. FINDSITE-metal program predicts metal binding site and the location of metal ion by superposing the detected templates onto the target protein structure. As shown in **Figure 2.2**, GalaxySite2 outperforms FINDSITE-metal program in the metal binding pose prediction. When the rates of the cases in which the predicted metal binding poses are within 1 Å RMSD from the native pose are considered, GalaxySite2 shows the highest success rate of 79% while the success rate of FINDSITE-metal is 30% (**Figure 2.2 (B)**). This result implies that the docking of metal ion using Monte Carlo simulation is helpful to predict better metal binding poses compared to the simple superposition-based method adopted in FINDSITE-metal.

To evaluate the effect of the side-chain sampling in the metal binding site prediction, the homology model set consisting total 46 targets was used. **Figure 2.3** shows the performance differences by the side-chain sampling during the refinement docking step described in **section 2.2.3**. By sampling side-chains during docking, the MCC, coverage, and accuracy measures are improved by 6.18 %, 7.46 %, and 4.98 %, respectively. As shown in **Figure 2.3 (D)**, the environment RMSD, defined as a RMSD between binding residues in the crystal structure and those in the final model structure, are also improved.

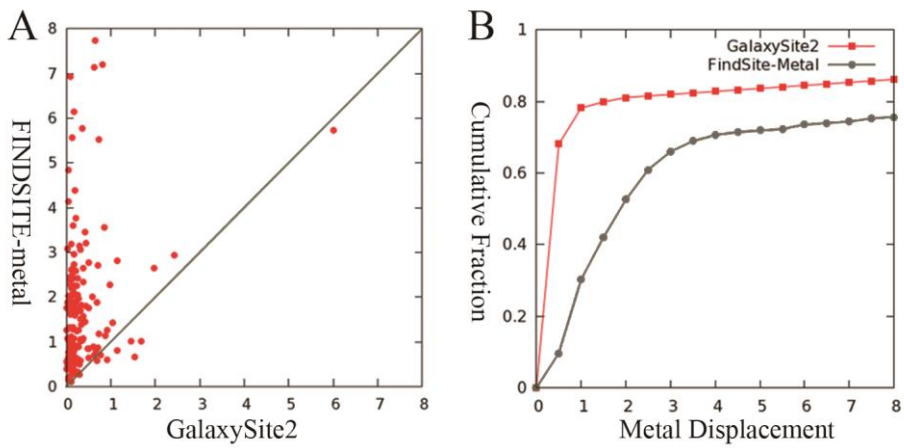


Figure 2.2. Performance comparison between GalaxySite2 and FINDSITE-metal. Head-to-head comparison of the displacement (in Å) of predicted metal ion generated by FINDSITE-metal and GalaxySite2 is shown in panel (A). Panel (B) shows the cumulative fraction of targets with a distance between the metal position in the crystal structure and the closest of the top three predicted binding sites

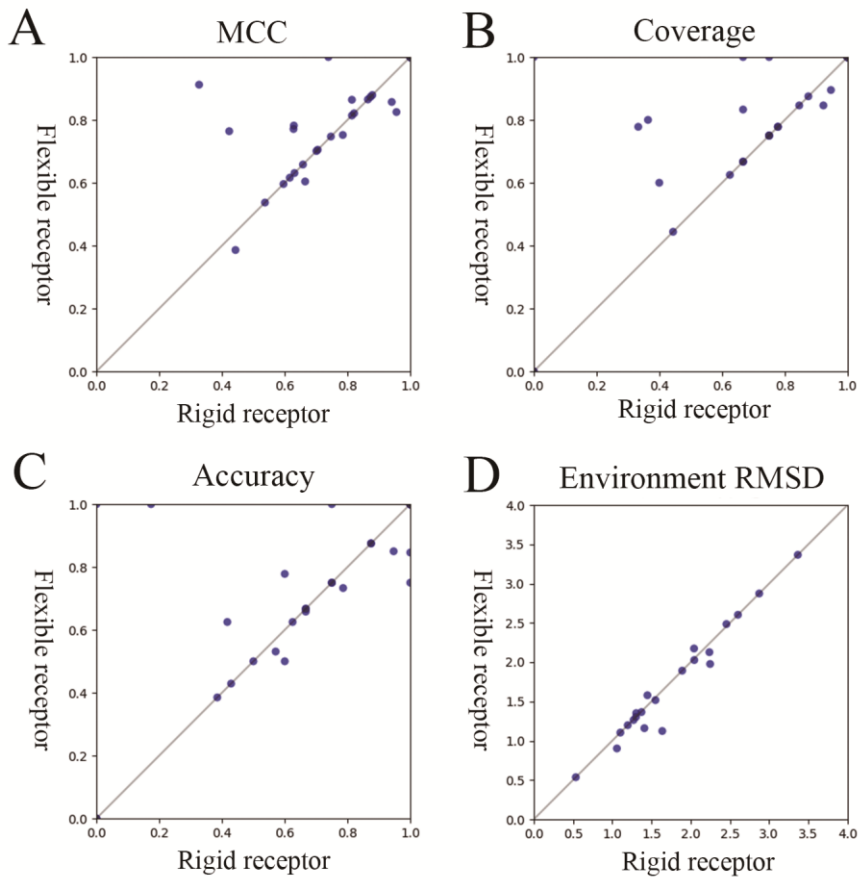


Figure 2.3. Head-to-head comparison of the quality of predicted binding sites between docking with rigid receptor mode and with flexible receptor mode in terms of MCC (A), coverage (B), accuracy (C), and environment RMSD (D).

2.3.2. Improved Ligand and Template Selection with Ligand Score

Unlike the original GalaxySite program, the ligand/template selection method in GalaxySite2 is modified to consider local and global structural similarity as well as sequence similarity. To measure improvements of the ligand/template selection, the qualities of all selected ligands to be docked in GalaxySite and GalaxySite2 were evaluated by MCC of binding site prediction using superposed ligand. As shown in **Figure 2.4**, the ligand/template selection method used in GalaxySite2 shows better performance compared to that used in GalaxySite. By considering both of structure and sequence similarity, the performance of ligand/template selection improved by 27.1 %.

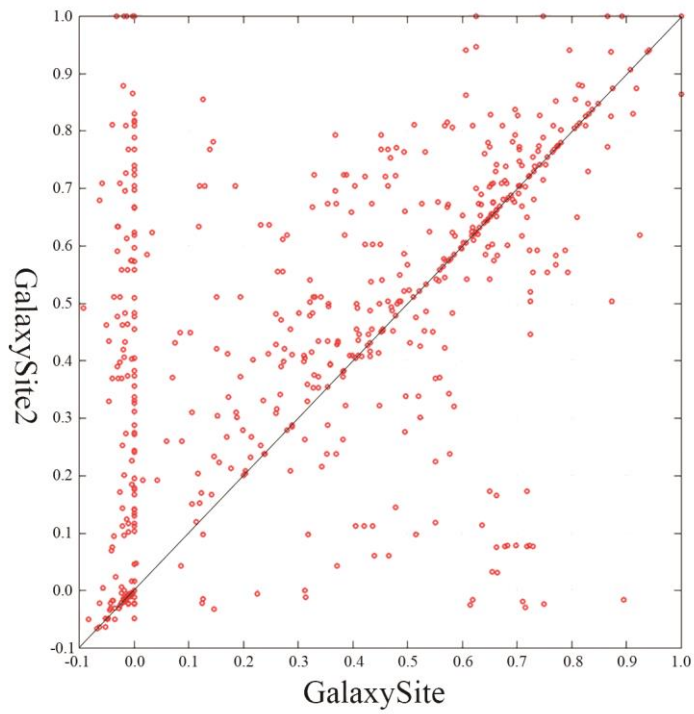


Figure 2.4. Head-to-head comparison of the quality of selected ligand and templates in terms of MCC of the binding site prediction using superposed ligand

2.3.3. Performance Comparison with Other Binding Site Prediction Servers on CAMEO Benchmark Set

The performance of GalaxySite2 was tested on 420 targets of the ligand binding-site prediction category from the continuous automated model evaluation server released between June 13th and August 29th, 2014. In **Figure 2.5**, results of GalaxySite2 for these targets were compared with available results of other servers in terms of median values of MCC. Because the number of predicted targets are different by server by server, only common targets were considered for comparison. The overall results show that the performance of GalaxySite2 is consistently comparable or superior to other available server methods.

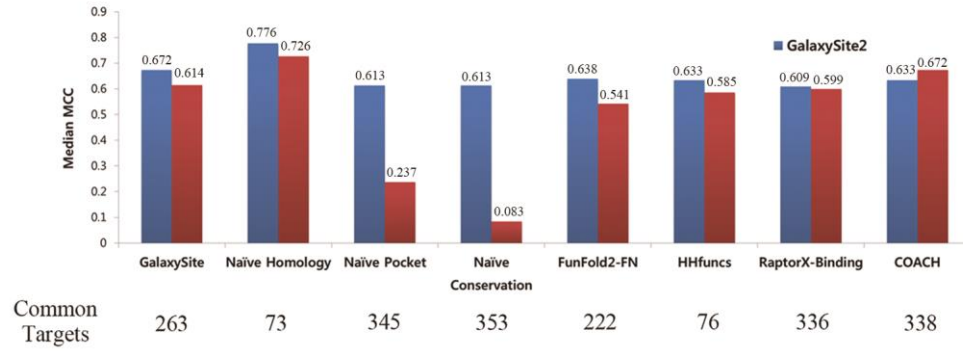


Figure 2.5. Comparison of different binding site prediction methods on the CAMEO benchmark set in terms of median MCC

2.4. Conclusion on Binding Site Prediction

GalaxySite2 predicts a binding site of metal and small organic molecules in proteins using molecular docking guided by information from similar protein-ligand complexes. The performance of GalaxySite2 in the metal binding site prediction is improved by docking metal ions using Monte Carlo simulation with the side-chain sampling. Moreover, the performance in prediction of small organic molecule binding sites is improved by improving the ligand and template selection method and introducing the final re-scoring method. Overall performance of GalaxySite2 is better than or comparable to other state-of-the-art programs according to tests on the CAMEO (Continuous Automated Model Evaluation) benchmark sets. Unlike other available methods that predict binding residues only, GalaxySite2 provides additional predictions on key protein–ligand interactions in terms of optimized 3D coordinates of the protein-ligand complexes. Such specific information would be very useful for computer-aided drug discovery.

Chapter 3. Development of a Hybrid Scoring Function for Accurate Protein-Ligand Docking

3.1. Introduction to Protein-Ligand Docking Score

In previous chapter, the ligand binding site prediction method named GalaxySite2 was introduced. When a ligand binding site of a target protein is known, one can develop drugs targeting that ligand binding site. Computational methods can contribute to identifying hit or lead compounds and to optimizing lead compounds for given protein targets, reducing time and labor during the drug discovery process. A number of protein–ligand docking programs have been applied to virtual screening of compound libraries for such purposes (Abagyan et al. 1994; Morris et al. 1998; Kramer et al. 1999; Verdonk et al. 2003; Friesner et al. 2004; Korb et al. 2006; Trott and Olson 2010; Shin et al. 2011; Shin and Seok 2012; Spitzer and Jain 2012; Shin et al. 2013; Allen et al. 2015).

The success of a protein–ligand docking program depends on the program's performance on two famous, but still unsolved problems: scoring and sampling. Scoring in docking has two aspects, which are scoring different binding poses of a given ligand to a given protein receptor and scoring different ligands by their binding affinities to a given protein. In the docking community, scoring methods have often been developed and tested separately of sampling to simplify the docking problem. In many cases, scoring functions trained to reproduce experimental binding affinities have been used to score both binding poses and binding affinities (Jones et al. 1997; Morris et al. 1998; Trott and Olson 2010; Eldridge et al. 1997; Bohm 1998; Wang et al. 2002; Huey et al. 2007; Korb et al. 2009). These binding affinity-based approaches have shown some success in

binding pose scoring in various benchmark tests (Bursulaya et al. 2003; Wang et al. 2003; Ferrara et al. 2004; Perola et al. 2004; Warren et al. 2006; Zhou et al. 2007; Cheng et al. 2009; Cross et al. 2009; Li et al. 2014a).

In this chapter, I introduce a new docking scoring function, named GalaxyDock BP2 Score, specifically designed to predict binding poses with high accuracy (Baek et al. 2017b). The score can also be used within GalaxyDock2 (Shin et al. 2013), a protein–ligand docking program that employs a global optimization technique called conformational space annealing (CSA) (Lee et al. 2005; Lee et al. 1997). The previous version, GalaxyDock, was developed by combining CSA with the AutoDock3 energy (Shin et al. 2011; Shin and Seok 2012). When the AutoDock3 energy was subject to CSA optimization, it often produced binding poses that had lower energy than those found by the Lamarckian genetic algorithm of AutoDock3 (Morris et al. 1998), but were more distant from the crystal poses, indicating a problem of the AutoDock3 energy (Shin et al. 2011). It was noted that ligand torsion energy is absent in the AutoDock3 energy, so a ligand torsion energy term was added to the AutoDock3 energy by means of the PLP score (Gehlhaar et al. 1995). This resulting energy function is called GalaxyDock BP Score, and it has been shown that the docking performance of GalaxyDock improved with this addition (Shin et al. 2011).

I further improved GalaxyDock BP Score for binding pose prediction by a more systematic energy optimization. I first adopted the more recent AutoDock4 energy (Huey et al. 2007) instead of the AutoDock3 energy (Morris et al. 1998). Although the physics-based energy terms of AutoDock4 are effective in sampling physically realistic binding poses, other empirical or knowledge-based scoring functions perform better in some benchmark tests (Cheng et al. 2009; Li et al. 2014a). This may be because the force field-based energy with an implicit

solvation model does not as effectively account for hydrophobic effects, conformational entropy, water-mediated interactions, etc., that are important for docking. I therefore decided to adopt a hybrid of different types of scoring functions to take advantage of different scores. This type of scoring function was successfully employed in other modeling problems such as protein structure prediction (Zhang et al. 2004; Zhu et al. 2006; Chopra et al. 2010; Park and Seok 2012b; Park et al. 2014b). The new GalaxyDock BP2 Score consists of the AutoDock4 energy (Huey et al. 2007), PLP ligand torsion energy (Gehlhaar et al. 1995), an empirical hydrophobic matching score called HM-score taken from X-Score (Wang et al. 2002), and a knowledge-based atom pair potential that mimics DrugScore (Gohlke et al. 2000). Proper balance of the score terms is crucial, and the weight factors of the score terms were determined by an iterative parameter optimization procedure that trains the scoring function to favor near-native poses over non-native poses.

To improve the docking performance of GalaxyDock, GalaxyDock BP2 Score was implemented in GalaxyDock2, which is a more recent version of GalaxyDock that generates an initial docking poses using a fast, geometry-based docking method that employs a β-complex (Shin et al. 2013). After training on protein–ligand complexes selected from the PDBbind 2013 database (Li et al. 2014b), performance tests were carried out on five other complex sets. First, the decoy discrimination power of GalaxyDock BP2 Score was tested on binding pose sets generated from the 443 complexes compiled from a refined set of PDBbind 2013 database (Li et al. 2014b) and on different pose sets from the 195 complexes of the CASF-2013 benchmark set (Li et al. 2014a). Additionally, GalaxyDock2 with GalaxyDock BP2 Score showed superior performance when compared to other state-of-the-art docking programs not only in the self-docking tests on the 85

target complexes of the Astex diverse set (Hartshorn et al. 2007) and the 64 target complexes of the Cross2009 benchmark set (Cross et al. 2009) but also in a more realistic docking test on the Astex non-native set (Verdonk et al. 2008). Even though GalaxyDock BP2 Score was optimized for scoring binding poses, it shows results comparable to other available methods in scoring binding affinities when tested on CASF-2013 benchmark set and DUD data set (Huang et al. 2006).

3.2. Methods

3.2.1. Components of GalaxyDock BP2 Score

GalaxyDock BP2 Score is a linear combination of multiple components that are taken from physics-based, empirical, and knowledge-based scores as follows:

$$\begin{aligned} E_{\mathrm{BP2}} = {} & E_{\mathrm{vdW,\,PL}} + w_1 E_{\mathrm{hbond,\,PL}} + w_2 E_{\mathrm{qq,\,PL}} + w_3 E_{\mathrm{desolv,\,PL}} \\ & + w_4 E_{\mathrm{vdw,\,L}} + w_5 E_{\mathrm{hbond,\,L}} + w_6 E_{\mathrm{qq,\,L}} + w_7 E_{\mathrm{desolv,\,L}} \\ & + w_8 E_{\mathrm{PLP\_tor}} + w_9 E_{\mathrm{HM}} + w_{10} E_{\mathrm{DrugScore}} \end{aligned} \tag{3.1}$$

where E's are energy terms and w's are weight factors. The first eight terms are physics-based energy terms of AutoDock4 that describe interactions between protein (P) and ligand (L) atoms (from $E_{\mathrm{vdW,PL}}$ to $E_{\mathrm{desolv,PL}}$) and those within ligand (from $E_{\mathrm{vdW,L}}$ to $E_{\mathrm{desolv,L}}$) (Huey et al. 2007). The E_{vdW} terms describe van der Waals dispersion/repulsion interactions with the 12-6 Lennard-Jones potential. The E_{hbond} term describes directional hydrogen bond energy with the 12-10 potential. The E_{qq} term represents screened Coulomb electrostatic energy with a sigmoidal distance-dependent dielectric constant. Gasteiger charges (Gasteiger and Marsili 1980) and

the partial charges of CHARMM22 force field (MacKerell et al. 1998) were assigned to ligand and protein, respectively. The E_{desolv} term represents desolvation free energy based on an estimation of solvation free energy loss for each atom due to the solvent volume occluded by surrounding atoms.

The component $E_{\text{PLP\_tor}}$ is the ligand torsion energy term adopted from the PLP score (Gehlhaar et al. 1995). This term may be considered a physics-based energy term. The term considers intra-ligand torsional strains that are not accounted for in the AutoDock4 energy (Shin et al. 2011). The previous scoring function of GalaxyDock, called GalaxyDock BP Score, consists of the AutoDock3 energy terms (Morris et al. 1998) that correspond to the first six terms of Eq. (3.1) and $E_{\text{PLP\_tor}}$. GalaxyDock BP Score added $0.1E_{\text{PLP\_tor}}$ to the AutoDock3 energy with the original weight factors (Shin et al. 2011). In contrast to GalaxyDock BP Score, within the new GalaxyDock BP2 Score, all weight factors were optimized again including those for two additional terms called HM-score and DrugScore, as described in the next paragraphs.

The hydrophobic matching score E_{HM} is adopted from the X-score (Wang et al. 2002) to describe hydrophobic effects involved in protein–ligand binding. This term, called HM-score, can be considered an empirical score. HM-score scores favorably when a hydrophobic ligand atom is placed in a hydrophobic environment. Hydrophobicity of ligand and protein atoms was defined based on the LogP value predicted by an in-house version of XLogP program (Wang et al. 2000). Although the hydrophobic effect is partially considered by the van der Waals and

the desolvation terms of AutoDock4, more complicated hydrophobic effects that involve both solvent entropy and enthalpy changes are not effectively taken into account. The E_{HM} term therefore supplements this deficiency of AutoDock4.

The last component, $E_{DrugScore}$, is a knowledge-based potential that was derived in-house using the same logic as the DrugScore distance-dependent atom-pair potential (Gohlke et al. 2000). Since this knowledge-based potential is obtained from statistics of the distances between atoms found in the experimentally resolved protein–ligand complex structures, important effects (e.g., water-mediated interactions, solvation free energy, ligand conformational entropy) that are hard to be accounted for by other physics-based or empirical scores are captured implicitly. To derive the potential, the refined set of PDBbind 2013 database (2,959 PDB structures) (Li et al. 2014b) was used instead of the ReLiBase (Hendlich et al. 2003) used for the original DrugScore (Gohlke et al. 2000). Those complexes used for testing obtained energy parameters were excluded from the refined set (see the next two subsections for explanation on energy parameter training). The SYBYL atom types were used as in the original DrugScore. The newly derived $E_{DrugScore}$ was validated by comparing with the distance-dependent part of DSX (Neudert and Klebe 2011), the latest version of DrugScore.

3.2.2. Energy Parameter Optimization Based on the Decoy Discrimination

The procedure used to optimize the weight factors w in Eq. (3.1) is depicted in

Figure 3.1. Near-native and non-native docking poses were first generated, and energy parameters were then trained such that near-native poses were favored over non-native decoy poses, as described in detail below. The optimization procedure started with the original weight parameters of the AutoDock4 energy with the weights of zero for HM-score and DrugScore, and the final set of parameters form the GalaxyDock BP2 Score.

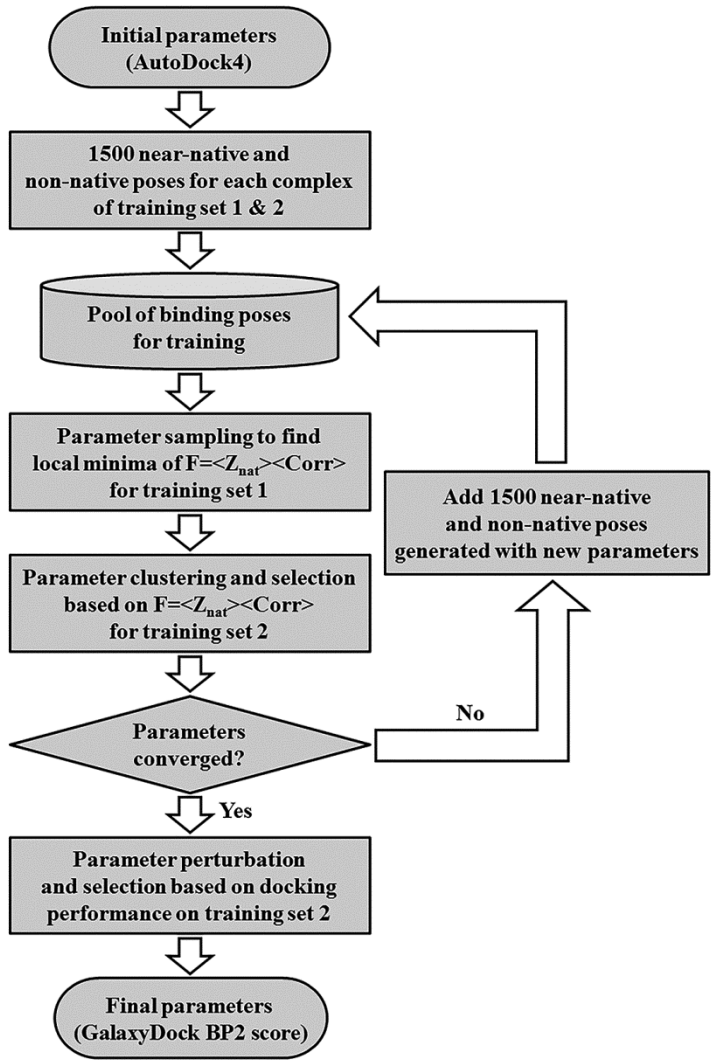


Figure 3.1. Overall procedure of energy parameter optimization

First, a set of docking poses that contains both near-native and non-native poses was constructed for each protein–ligand complex of the training sets 1 and 2 (715 and 415 complexes, respectively). How the training sets were constructed is explained in the **section 3.2.3**. During the pose generation, only ligand translational/rotational/torsional degrees of freedom were sampled, and protein conformation was fixed at the crystal structure using the rigid-receptor mode of GalaxyDock (Shin et al. 2011; Shin and Seok 2012). A total of 1,500 near-native and non-native poses were generated for each complex as follows: (1) Near-native poses were generated by perturbing the crystal ligand pose 2,500 times, following the procedure developed to optimize BP Score (Shin et al. 2011), and minimizing the resulting structures. The poses were clustered into 500 clusters by K-means clustering (Hartigan and Wong 1979), and the 500 cluster centers were used for parameter optimization. (2) More poses were generated by running GalaxyDock ten times and collecting all intermediate poses sampled during global optimization by the conformational space annealing algorithm. The crystal pose was included in the initial set of poses for GalaxyDock runs so that conformational sampling does not drift too far from the native pose. The total number of resulting poses was on the order of 10^5 for each complex. After filtering out high-energy poses, the poses were clustered into 1,000 clusters by K-means clustering, and the 1,000 cluster centers were used for parameter optimization.

Second, sampling was performed in the parameter space to minimize the following objective function F for the pose sets generated as described above (Park

et al. 2011):

$$F = \langle Z_{\text{nat}} \rangle \times \langle \text{Corr} \rangle \tag{3.2}$$

where Z_{nat} is the Z-score of the average energy of the top 10% poses closest to the crystal pose in the energy–RMSD distribution, where RMSD is root-mean-square-deviation of a given ligand conformation from the crystal ligand conformation, Corr is the Pearson correlation coefficient between energy and RMSD, and $\langle\rangle$ denotes an average over the complexes in the training set. Monte Carlo minimization was run 2,000 times to find local minima of F in the parameter space for training set 1. The resulting 2,000 parameter sets were clustered into 10 clusters by K-means clustering, and the centers of the 5 largest clusters were selected. Among the 5 parameter sets, the parameter set that gave the minimum value of the objective function F for the training set 2 was used as the initial parameter set for the next round of iteration.

With a new energy parameter set, the energy landscape can undergo changes in local energy minima. New local minima may appear, and some of old local minima may disappear. Therefore, before each new parameter optimization cycle, a new set of poses was generated by perturbing crystal structures and running GalaxyDock for each complex of the training sets with the new set of energy parameters. The poses were clustered into 1,500 clusters, and the cluster centers were added to the existing pose set. Energy parameters were then sampled with the enlarged pose sets to minimize the objective function. This procedure was

repeated until convergence.

In the final step of parameter optimization, parameters were further refined to optimize docking performance on the training set 2. One hundred parameter sets were generated by perturbing the converged parameter set obtained as described above, and GalaxyDock2 was run with each parameter set for each complex of the training set 2. The parameter set that showed the highest success ratio was selected as the final parameter set. A successful prediction was defined as a prediction in which the lowest-energy docking pose was within 2 Å RMSD from the native pose.

3.2.3. Training and Test Sets

Two sets of protein–ligand complexes were used for training the energy, and five sets for testing the performance of the energy. The two training sets and the first test set were compiled from the refined set of PDBbind 2013 database (2,959 PDB structures) (Li et al. 2014b). The database was randomly split into three subsets, and each subset was clustered with a sequence identity cutoff of 70% to remove redundancy. Up to 3 complexes were selected from each cluster, resulting in 715, 415, and 443 non-redundant complexes. The first two sets of 715 and 415 complexes were used for training the energy parameters (called training set 1 and 2), and the third set of 443 complexes (called test set 1) for testing the ability of the new score to discriminate near-native poses from non-native poses. Decoy binding

poses for test set 1 were generated by the same method used for generating poses for the training set complexes.

The second test set is the core set of PDBbind 2013 database, which consists of 195 protein–ligand complexes. This set is also referred to as CASF-2013 benchmark set because it was used in the comparative assessment of scoring functions (CASF) project in 2013 to compare performances of 20 different scoring functions (Li et al. 2014a). In the current study, this benchmark set is used as an additional test set for assessing the decoy pose discrimination capability of the new GalaxyDock BP2 Score in comparison to other scoring functions tested in CASF-2013. The decoy binding poses used for the CASF project were also used in our test. The poses were generated by selecting up to 100 docking poses from those obtained by running GOLD (Jones et al. 1997; Verdonk et al. 2003), Surflex (Jain 2007), and MOE (Vilar et al. 2008) for each complex.

The third test set comprises 85 protein–ligand complex structures of the Astex diverse set (Hartshorn et al. 2007), and the fourth set contains 64 protein–ligand complex structures taken from Ref. (Cross et al. 2009), referred to as Cross2009 benchmark set. These two sets were used to evaluate docking performance with the new scoring function in comparison to other previously tested docking methods. For a more realistic docking test, Astex non-native set (Verdonk et al. 2008), which includes receptor structures that are not bound to the given ligands, was used as the fifth set.

3.3. Results and Discussions

3.3.1. Results of Energy Optimization

The ten weight factors for the energy components of Eq. (3.1) were determined as described in the **section 3.2.2**. The final energy with the optimized weight factors is called the GalaxyDock BP2 Score [Eq. (3.1)], while the previous energy function of GalaxyDock used for scoring binding poses is called the GalaxyDock BP Score (Shin et al. 2011).

The objective function for parameter training written in Eq. (3.2) was designed to increase the energy-RMSD correlation for various binding poses and at the same time to lower the energy of native-like poses relative to non-native ones. The results of this training can depend heavily upon the quality of the binding pose sets. Therefore, special care was taken to generate the training poses to cover both near-native and non-native regions in the conformational space as much as possible. The near-native region was covered by perturbations of the crystal pose, while the non-native region was covered by the CSA global optimization implemented in the GalaxyDock protein–ligand docking program (Shin et al. 2011; Shin and Seok 2012). In **Figure 3.2**, score values of the three scoring functions, AutoDock4 energy, DrugScore, and GalaxyDock BP2 Score, are plotted against RMSD for the poses generated for 1ajp, a member of training set 2. It can be seen from the figure that the poses cover a broad range of conformational space and that the poses generated by perturbing the crystal structure (red) and CSA global optimization (blue) show some overlap. This particular example illustrates a successful

parameter optimization: GalaxyDock BP2 Score has its putative global energy minimum near the native pose at ~1.5 Å while its components AutoDock4 energy and DrugScore at ~3.5 Å and ~5 Å, respectively.

The parameter training procedure was applied iteratively until convergence. The energy surface can be modulated by changes in parameters, so new poses were generated with the new parameter set at each iteration step to reflect the change in the energy surface. Many parameter sets that give similarly low values for the objective function could be found, and the parameter sets that form the largest clusters were selected for further validation. Here, we assumed that a parameter set located in a broad basin in the parameter space would result in an energy surface more robust to small parameter changes and that such parameter set would be more transferable to other complexes not included in the training set. Parameter values obtained by this strategy indeed converged rapidly, settling to an almost converged parameter set after the second iteration, as displayed in **Figure 3.3**. From iteration 2 to iteration 5, the weight factors and the contributions of the energy components did not show much change, so the iteration was terminated after iteration 5.

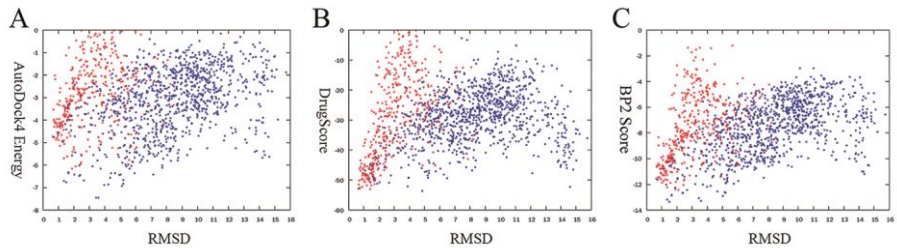


Figure 3.2. The energy versus RMSD scatter plots for AutoDock4 energy (A), DrugScore (B) and GalaxyDock BP2 Score (C). The ligand binding poses (PDB ID: 1AJP) were generated by two independent methods, perturbation of the crystal structure (red) and GalaxyDock docking using CSA (blue).

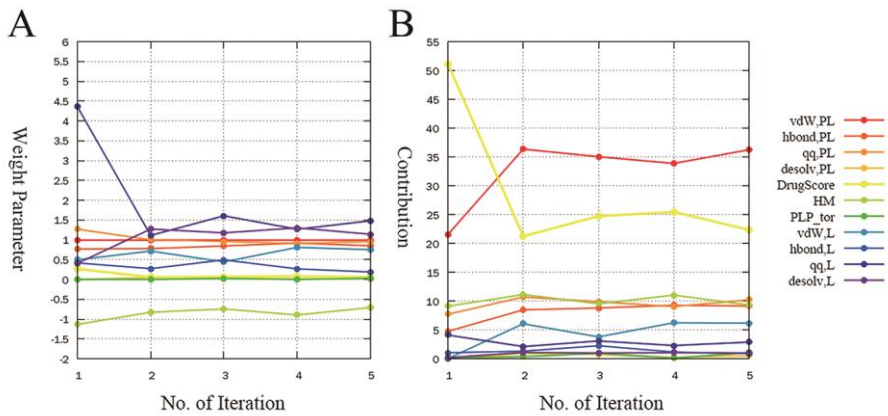


Figure 3.3. Changes of weight factors (A) and contributions of the energy components (B) as a function of iteration number during energy optimization. Contribution of each energy component is defined as its weight multiplied by standard deviation of the energy values for the pose sets for the complexes of the training set 2, averaged over the complexes and normalized over the components.

The final parameters and the contributions of the energy components to the total energy variation are reported in **Table 3.1**. The contribution of each energy component is defined as the standard deviation of the energy values of the component multiplied by the weight factor for each pose set averaged over the complexes in the training set 2 and normalized over the energy components. The largest contributions to GalaxyDock BP2 Score come from the van der Waals term, followed by the DrugScore and the HM-score terms. This result indicates that the newly added knowledge-based potential DrugScore and the empirical hydrophobic interaction score HM-score indeed contribute to the improved binding-pose scoring power of GalaxyDock BP2 Score.

Table 3.1. The optimized weights and the contributions of the energy components to the total energy variation

Energy	Weight	Contribution (%)*
$E_{vdW,PL}$	1.00	35.0
$E_{hbond,PL}$	0.85	8.8
$E_{qq,PL}$	0.93	9.6
$E_{desolv,PL}$	0.12	2.0
$E_{vdW,L}$	0.80	6.4
$E_{hbond,L}$	0.25	1.1
$E_{qq,L}$	1.35	2.6
$E_{desolv,L}$	1.00	0.9
E_{PLP_tor}	0.01	0.3
E_{HM}	-0.80	10.2
$E_{DrugScore}$	0.07	23.1

* Contribution of each energy component is defined as its weight multiplied by standard deviation of the energy values for the pose sets for the complexes of the training set 2, averaged over the complexes and normalized over the components.

3.3.2. Decoy Discrimination Test on the Pose Sets Generated by GalaxyDock

To assess the decoy pose discrimination power of the newly developed GalaxyDock BP2 Score, we employed two types of decoy sets. The first decoy set was generated by the same method as described in the **section 3.2.3**. Near-native and non-native poses were generated by perturbing the crystal poses and by running GalaxyDock multiple times on the 443 complexes of the test set 1. This test set was compiled from the refined set of the PDBbind 2013 database independently of the training sets, as described in the **section 3.2.3**. To minimize bias of decoy conformations by the energy function used for generating them, the following three energy functions were used to generate the poses: AutoDock4 energy, the energy function with the parameter set obtained after the first iteration, and GalaxyDock BP2 Score.

The performance of GalaxyDock BP2 Score in decoy pose discrimination is compared with those of other available scoring functions: AutoDock4 energy (Huey et al. 2007), DSX (Neudert and Klebe 2011), GalaxyDock BP Score (Shin et al. 2011), and X-score^{HM} (Wang et al. 2002) in **Figure 3.4**. This figure compares the rates of the cases in which the top one, two, or three best-scoring poses are within 2 Å RMSD from the native pose for the test set complexes. When the best-scoring pose is picked, GalaxyDock BP2 Score shows the highest success rate of 75% while success rates of AutoDock4 energy, DSX, GalaxyDock BP score, and X-score^{HM} are 72%, 68%, 67%, and 60%, respectively. GalaxyDock BP2 Score also shows the highest success rates for top two and three poses.

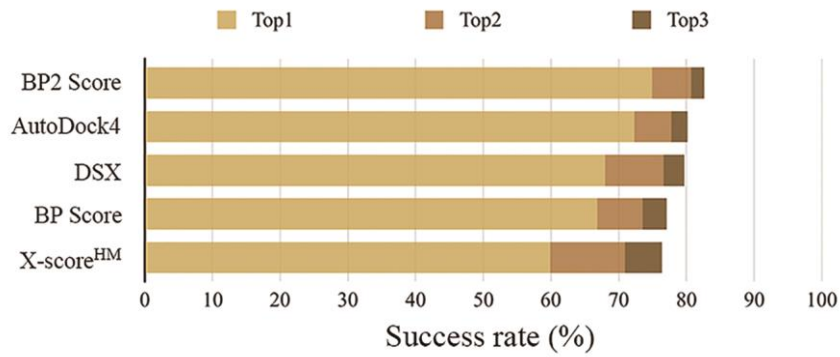


Figure 3.4. Rates of the complexes of the test set for which the top one (light brown bars), two (brown bars), or three (dark brown bars) best-scoring poses are within 2 Å RMSD for five scoring functions. The scoring functions are ordered by the success rate for the top one poses.

3.3.3. Decoy Binding Pose Discrimination Test on the CASF-2013 Benchmark Set

The second decoy pose discrimination test was carried out on the CASF-2013 benchmark set to compare the performance of GalaxyDock BP2 Score with other scoring functions evaluated in the CASF-2013 benchmark study (Li et al. 2014a). The binding pose sets generated by the authors using GOLD, Surflex, and MOE were used. The native poses were also included following the study. Therefore, this is a test on binding pose sets generated independently of GalaxyDock. GalaxyDock BP2 Score was evaluated after local energy minimization as well as for the given poses. Local minimization was performed to remove steric clashes because GalaxyDock BP2 Score contains energy terms sensitive to atomic clashes such as van der Waals energy and directional hydrogen bond energy. Local minimization improved the success rate prediction of the top 1 pose by ~2%. It was also discussed previously that rescoring predefined poses can be misleading when assessing docking performance of scoring functions (Korb et al. 2012). The success rates of GalaxyDock BP2 Score with and without minimization, AutoDock4 energy, DSX, GalaxyDock BP score, X-score^{HM}, and the top 10 scoring functions in the CASF-2013 benchmark test are compared in **Figure 3.5**. When the best-scoring pose is considered, GalaxyDock BP2 Score shows the highest success rate of 84.1% (86.2% when minimized), followed by ChemPLP@GOLD (Korb et al. 2009) (81.0%), DSX (Neudert and Klebe 2011) (80.5%) and GlideScore-SP (Friesner et al. 2004; Halgren et al. 2004) (78.5%). When the three best-scoring poses are considered, DSX (Neudert and Klebe 2011) shows the best success rate of 91.3%,

followed by GalaxyDock BP2 Score (89.7%; 91.3% when minimized), ChemPLP@GOLD (Korb et al. 2009) (89.7%), and ChemScore@GOLD (Eldridge et al. 1997) (88.2%).

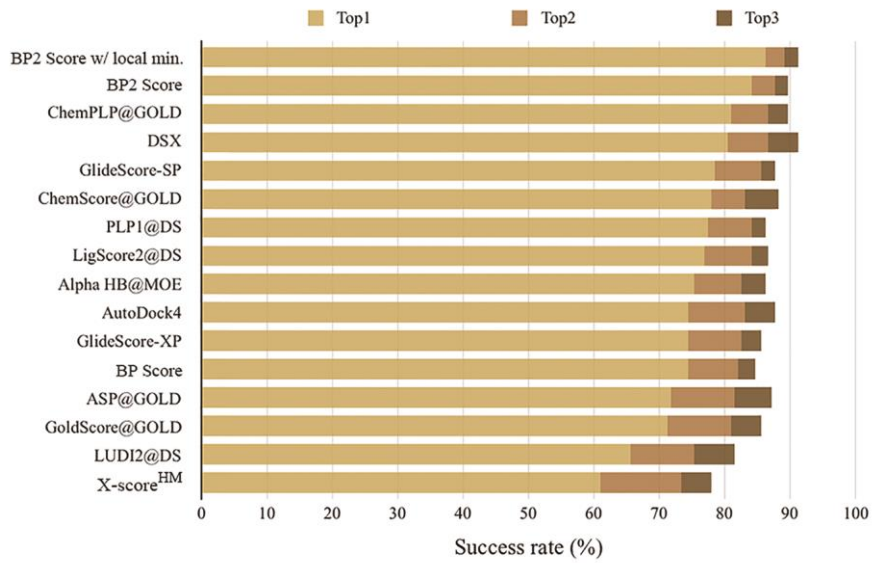


Figure 3.5. Success rates of 15 scoring functions on the CASF-2013 benchmark set when the top one (light brown bars), two (brown bars), or three (dark brown bars) best-scoring poses are within 2 Å RMSD. The scoring functions are ordered by the success rate for the top one poses.

To find clues about which features of GalaxyDock BP2 Score contribute to its improved performance compared to the existing scoring functions, results on nine subsets of the CASF-2013 benchmark set with different ligand or interaction types defined by the authors (Li et al. 2014a) were examined. The subsets were constructed by classifying the benchmark set complexes into three groups each based on three criteria: the number of rotatable bonds of ligand (subsets A1-A3), the percentage of the solvent-accessible surface area of ligand buried upon binding (subsets B1-B3), and the hydrophobic scale of the binding pocket of protein measured by logD (subsets C1-C3). Statistics of the subsets are provided in **Table 3.2**.

Success rates of the scoring functions mentioned above on the nine subsets are illustrated in color scale in **Figure 3.6**. Most scoring functions show higher success rates on the subsets A1 and A2 than on A3, on B3 than on B1 and B2, and on C1 or C2 than on C3. Interestingly, GalaxyDock BP2 Score shows relatively high performance on difficult subsets A3, B1, and B2 compared to other scoring functions. The members of subset A3 overlap with those of subsets B1 and B2, and this is understandable considering that ligands in A3 have relatively large number of rotatable bonds, and those in B1 and B2 bury relatively smaller fraction of surface area upon binding. Therefore, the 32 targets of A3 are also members of either B1 (13 targets) or B2 (19 targets).

Table 3.2. Subsets of the CASF-2013 benchmark set

Classification Standard	Range	No. of complexes	Subset Symbol
No. of rotatable bonds of ligand	1 ~ 3	116	A1
	4 ~ 8	47	A2
	> 8	32	A3
Fraction of ligand solvent-accessible surface area buried upon binding	0.0 ~ 0.65	37	B1
	0.65 ~ 0.85	117	B2
	> 0.85	41	B3
Hydrophobic scale of the binding pocket measured by log D	< -0.50	42	C1
	-0.50 ~ 0.00	116	C2
	> 0.00	37	C3

This table is taken from (Li et al. 2014a).

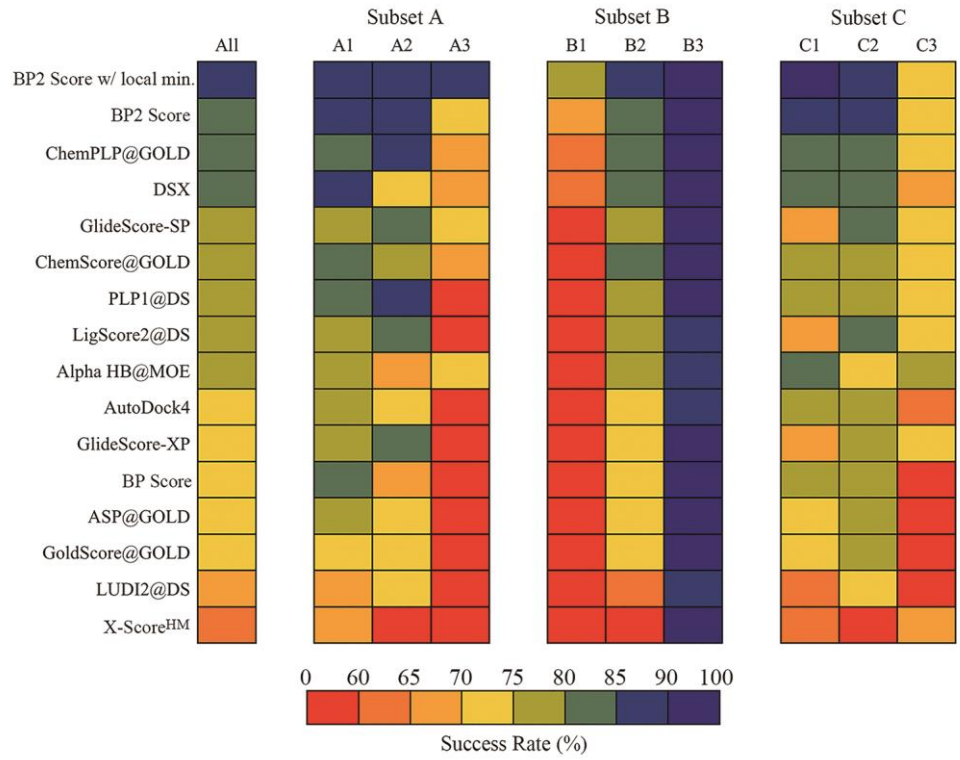


Figure 3.6. Success rates of 15 scoring functions on the nine subsets of CASF-2013 benchmark set. Performance on the entire test set is displayed on the left as a reference.

The improved performance of GalaxyDock BP2 Score on the subsets A3, B1, and B2 implies that this score can describe protein–ligand interactions in relatively solvent-exposed environments more accurately than other scoring functions. We attribute this success partially to its effective treatment of water-mediated interactions by the knowledge-based component DrugScore. Evidence of this was found when we examined the cases in which GalaxyDock BP2 Score succeeded while AutoDock4 energy failed in decoy discrimination. For example, the target 1h23 has a relatively open binding pocket and involves many water-mediated interactions, as shown in **Figure 3.7 (A)**. The AutoDock4 energy selects a conformation with 2.4 Å RMSD from the crystal pose which make many direct contacts with protein atoms instead of water-mediated interactions (**Figure 3.7 (B)**). GalaxyDock BP2 Score scores the crystal pose the highest, and the high score is mainly due to the DrugScore term. DrugScore favors the crystal pose over the pose selected by AutoDock4 energy by 1.62 energy unit, while the score terms from AutoDock4 disfavor the crystal pose by 1.11 energy unit. The hydrophobic matching score, HM-score, favors the crystal pose by 0.21 energy unit, which is a much smaller contribution than that by DrugScore.

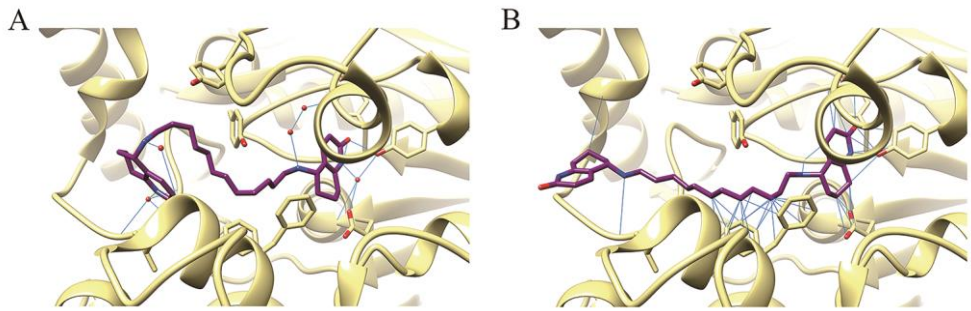


Figure 3.7. The best-scoring pose for the target 1H23 by GalaxyDock BP2 Score, which is the same as the crystal pose (A) and that by AutoDock4 scoring function (RMSD = 2.4 Å) (B). Water-mediated interactions are present in A.

Targets with more hydrophobic binding pockets (subset C3) are the most difficult to score even with GalaxyDock BP2 Score. It seems that the combination of HM-score and the orientation-dependent hydrogen bond energy term of AutoDock4 score mainly contribute to the slightly better performance of GalaxyDock BP2 Score. For example, the crystal pose of the target 3u9q makes many hydrophobic contacts and forms a specific hydrogen bonding network. AutoDock4 energy picks a binding pose with 4.1 Å RMSD that makes a few strong Coulombic interactions because Coulomb interaction favors this pose over the crystal pose by 0.5 energy unit. GalaxyDock BP2 Score selects the crystal pose over the decoy pose because HM-score favors the crystal pose by 1.5 energy unit. When compared to other decoy poses with similar HM-score values, GalaxyDock BP2 Score favors the crystal pose mainly due to the hydrogen bond energy term.

3.3.4. Comparison with Energy Parameter Optimization Based on Binding Affinity Data

To demonstrate the benefits of the current energy optimization strategy based on decoy discrimination over a strategy based on binding affinity data, a different set of energy parameters was obtained by optimizing the Pearson correlation coefficient between the available experimental binding affinity data and the score values of the crystal poses averaged over the same training set targets. The resulting weight factors are as follows: w_1 = 0.68, w_2 = 0.76, w_3 = 0.20, w_4 = 1.45, w_5 = 0.23, w_6 = 1.77, w_7 = 1.91, w_8 = 0.25, w_9 = -3.16, w_{10} = 0.70.

When the decoy pose discrimination ability of this binding affinity-based scoring function was evaluated, it showed worse performance than GalaxyDock BP2 Score. The success rate of the binding affinity-based scoring function on the pose sets generated in this study for the test set of 443 complexes compiled from the PDBbind 2013 database is 63% when the lowest energy conformation is picked. GalaxyDock BP2 Score shows a success rate of 75% for the same set. When the binding affinity-based scoring function tested on the decoys for the 195 complexes of CASF-2013 benchmark set, the success rate is 77% for top 1 pose, compared to 84% with GalaxyDock BP2 Score. The decreased performance of the binding affinity-based scoring function, which has the same functional form as GalaxyDock BP2 Score but has a different parameter set, implies that the overall success of GalaxyDock BP2 Score is mainly due to the energy parameter optimization strategy. The current example clearly shows that a scoring function optimized for better scoring of binding poses has better decoy pose discrimination power than a scoring function optimized for better binding affinity prediction. This score trained on affinity data did not perform better than BP Score in a preliminary virtual screening test as described in **section 3.3.6**. Therefore, I think that a more sophisticated training method than presented here is necessary for developing a function that predicts binding affinity in the context of virtual screening.

3.3.5. Improved Docking Performance of GalaxyDock2 with GalaxyDock BP2 Score

So far, I have evaluated the performance of GalaxyDock BP2 Score in scoring poses generated in advance. The improved score, or energy function, may be used to guide conformational sampling towards the global minimum of the given energy function, providing scores *in situ* while sampling is performed. This may result in improved docking performance.

GalaxyDock BP2 Score has been implemented in the GalaxyDock2 protein–ligand docking program, which employs a global optimization technique called conformational space annealing. GalaxyDock2 with the new energy function was tested on the 85 complexes of Astex diverse set (Hartshorn et al. 2007) and the 64 complexes of Cross2009 benchmark set (Cross et al. 2009) for which other state-of-the-art docking programs were tested previously. The results on the two benchmark sets are summarized in **Table 3.3** and **Table 3.4**, respectively. On the Astex diverse set, GalaxyDock2 with the new score (GalaxyDock BP2 Score) succeeded in generating protein–ligand complex structures with RMSD better than 2 Å as the lowest energy models in 76 of 85 cases (success rate = 89.4%). This is a better performance than those of GalaxyDock2 with the old score (Shin et al. 2013) (GalaxyDock BP Score) (81.2%), AutoDock3 (Morris et al. 1998) (81.7%), GOLD (Jones et al. 1997; Verdonk et al. 2003) (80.5%), and Surflex-Dock (Jain 2007) (80.0%). The average RMSD values of the lowest energy models also improved compare to other docking methods.

Table 3.3. Performance of docking programs on the Astex diverse set

Docking Program	Success Rate (%)[1)]	RMSD (Å)[2)]
GalaxyDock2 w/ BP2 Score	89.4	1.10
GalaxyDock2 w/ BP Score	81.2	1.69
AutoDock3[3)]	81.7	1.60
GOLD[4)]	80.5	-
Surflex-Dock[5)]	80.0	1.66

[1)] Percentage of the cases in which RMSD of the best scoring pose from the crystal pose is less than 2 Å.

[2)] RMSD of the top scoring pose from the crystal structure averaged over the targets in the Astex diverse set.

[3)] Taken from (Shin et al. 2011).

[4)] Taken from (Bursulaya et al. 2003). Average RMSD is not reported in the reference.

[5)] Taken from (Spitzer and Jain 2012).

Table 3.4. Performance of docking programs on the Cross2009 benchmark set

Docking Program	Success Rate (%)[1)]	RMSD (Å)[2)]
GalaxyDock2 w/ BP2 Score	89.1	1.11
Glide-XP[3)]	84.4	1.20
Glide-SP[3)]	76.6	1.42
ICM[3)]	75.0	1.80
DOCK[3)]	62.5	2.76
Surflex[3)]	62.5	2.04
FRED[3), 4)]	60.9	-
PhDock[3)]	57.8	2.58
GalaxyDock2 w/ BP Score	54.7	3.26
FlexX[3)]	50.0	3.46

[1)] Percentage of the cases in which RMSD of the top scoring pose from that of the crystal structure is less than 2 Å.

[2)] RMSD of the top scoring pose from the crystal pose averaged over the targets in the set.

[3)] Taken from (McGann 2011).

[4)] Average RMSD is not reported here because RMSD values of FRED predicted results higher than 2 Å were reported as 2.1 Å in (McGann 2011).

On the Cross2009 benchmark set, GalaxyDock2 with the new score predicted poses within 2 Å RMSD for 57 out of 64 targets (success rate = 89.1%). Average RMSD of the best scoring function was 1.11 Å. The performance of GalaxyDock2 with BP2 Score is superior to that of the eight other docking programs reported in Ref. (Cross et al. 2009) and that of GalaxyDock2 with the old score. The success rates of the other top three methods in **Table 3.4** are 84.4% (Glide-XP (Friesner et al. 2006)), 76.6% (Glide-SP (Friesner et al. 2004; Halgren et al. 2004)), and 75.0% (ICM (Abagyan et al. 1994)). RMSD distributions of the lowest energy conformations obtained by each docking method for the targets of Cross2009 benchmark set are shown in **Figure 3.8**. Although GalaxyDock2 with the new score shows higher median RMSD values than Glide-XP and ICM, it also shows a narrower distribution and fewer outliers.

Additional docking test of GalaxyDock2 with BP2 Score was performed on the Astex non-native set in which ligands were docked into the non-native protein conformations that are unbound or bound to different ligands. The results on this benchmark set are summarized in **Table 3.5**. GalaxyDock2 with BP2 Score showed much better performance than GalaxyDock2 with the old BP Score and the four other docking programs reported in (Gaudreault and Najmanovich 2015). This improved performance is due to the fact that GalaxyDock BP2 Score can tolerate some conformational errors in the binding pocket with additional knowledge-based and empirical score terms. These results highlight the strength of combining an improved hybrid scoring function with an efficient sampling algorithm.

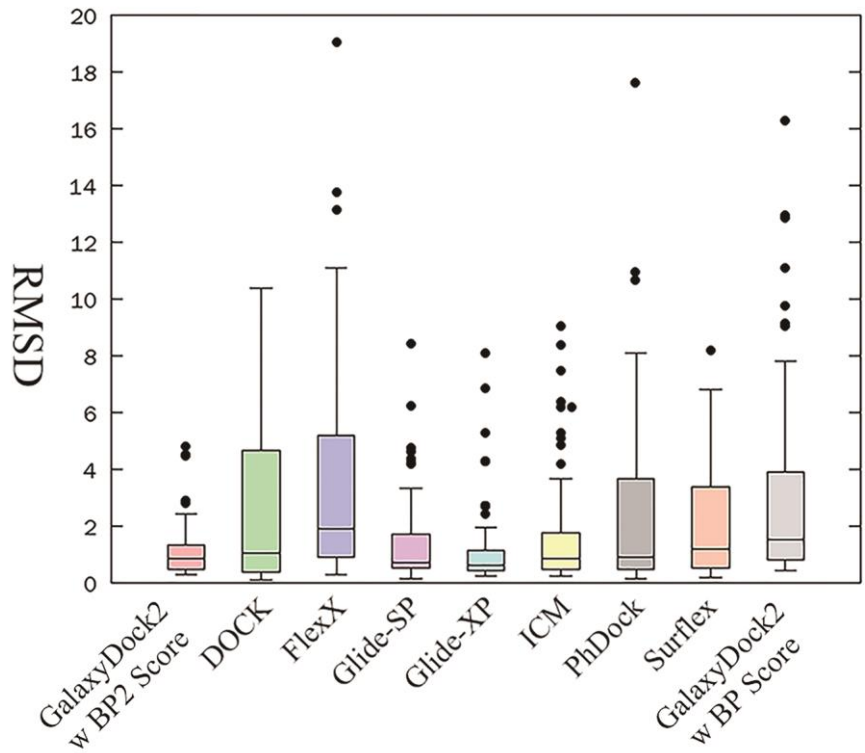


Figure 3.8. Distribution of RMSDs of the lowest energy conformations selected by GalaxyDock2 with BP2 Score, DOCK, FlexX, Glide-SP, Glide-XP, ICM, PhDock, Surflex, and GalaxyDock2 with BP score for the Cross2009 benchmark set.

Table 3.5. Performance of docking programs on the Astex non-native set

Docking Program	Success Rate (%)[1]	
	Ideal subset[2]	Largest common subset[3]
GalaxyDock2 w/ BP2 Score	61.4 (1106)	66.4
rDock[4]	54.7 (1106)	55.6
GalaxyDock2 w/ BP Score	53.7 (1106)	58.3
AutoDock-Vina[4]	41.7 (999)	44.4
FlexX[4]	41.3 (925)	41.7
FlexAID[4]	38.5 (775)	41.7

[1] Percentage of the cases in which RMSD of the top scoring pose from that of the crystal structure is less than 2 Å .

[2] The ideal subset is the largest subset in which a given program is able to run without technical errors; The number of targets (in parentheses) in ideal subset is a different for each program.

[3] The largest common subset contains 669 structures representing 36 unique targets.

[4] Taken from (Gaudreault and Najmanovich 2015).

3.3.6. Scoring, Ranking, and Screening Power Test on the CASF-2013 Benchmark Set and DUD Data Set

To see how the scoring function optimized to discriminate binding poses works for predicting binding affinity and ranking compounds, we carried out scoring, ranking and screening power test of GalaxyDock BP2 Score on CASF-2013 benchmark. The scoring power was evaluated by Pearson correlation coefficient between predicted and experimental binding affinity data. The results of scoring power test are summarized in **Table 3.6**. Even though GalaxyDock BP2 Score was not optimized to predict binding affinity, it showed a performance comparable to other available scoring function. When we adopted the free-ligand correction strategy in which the score for the free ligand is subtracted from the interaction energy, as described in (Shin et al. 2013), the correlation coefficient increased from 0.570 to 0.590.

The ranking power was quantitatively evaluated by the success rate of ranking the compounds in the order of binding affinity (high-level) and the success rate of picking the compound of highest binding affinity (low-level). GalaxyDock BP2 Score showed comparable performance to other scoring functions, as shown in **Table 3.7**. Similar to scoring power test, the free-ligand correction improved ranking ability of the scoring function from 52.3% to 53.8% in the high-level and from 64.6% to 67.7% in the low-level.

Table 3.6. Performance of 15 scoring functions in the scoring power test on CASF-2013 benchmark set

Scoring Function	N[1)]	Corr.[2)]	SD[3)]
X-Score^{HM} 4)	195	0.614	1.78
ChemPLP@GOLD[4)]	195	0.592	1.84
GalaxyDock BP2 Score	195	0.570	1.85
DSX	195	0.570	1.85
PLP1@DS[4)]	195	0.568	1.86
GalaxyDock BP Score	195	0.565	1.86
ASP@GOLD[4)]	195	0.556	1.88
ChemScore@GOLD[4)]	189	0.536	1.90
AutoDock4	193	0.527	1.92
Alpha-HB@MOE[4)]	195	0.511	1.94
LUDI3@DS[4)]	195	0.487	1.97
GoldScore@GOLD[4)]	189	0.483	1.97
LigScore2@DS[4)]	190	0.456	2.02
GlideScore-SP[4)]	169	0.452	2.03
GlideScore-XP[4)]	164	0.277	2.18

[1)] The number of complexes that show favorable binding scores by the scoring function.

[2)] The Pearson correlation coefficient between the experimental binding data and the computed binding scores.

[3)] The standard deviation in the linear correlation between the experimental binding data and the computed binding scores.

[4)] Taken from (Li et al. 2014a).

Table 3.7. Performance of 15 scoring functions in the ranking power test on CASF-2013 benchmark set

Scoring Function	Success Rate (%)	
	High-level[1)]	Low-level[2)]
X-Score^{HM} [3)]	58.5	72.3
ChemPLP@GOLD[3)]	58.5	72.3
PLP2@DS[3)]	55.4	76.9
GoldScore@GOLD[3)]	55.4	76.9
DSX	55.4	72.3
LUDI1@DS[3)]	52.3	69.2
Alpha-HB@MOE[3)]	52.3	66.2
GalaxyDock BP2 Score	52.3	64.6
LigScore1@DS[3)]	52.3	61.5
AutoDock4	49.2	63.1
ASP@GOLD[3)]	47.7	72.3
ChemScore@GOLD[3)]	46.2	63.1
GalaxyDock BP Score	44.6	63.1
GlideScore-SP[3)]	43.1	56.9
GlideScore-XP[3)]	35.4	47.7

[1)] Ranking three complexes in a cluster as the best > the median > the poorest

[2)] Ranking the best complex in a cluster as the top one

[3)] Taken from (Li et al. 2014a).

The screening power was evaluated by enrichment factor for top 1%, 5%, and 10% of database. Decoy poses provided by CASF-2013 benchmark were used to predict binding poses of ligands on target proteins. The screening results are summarized in **Table 3.8**. Unlike scoring and ranking power test, GalaxyDock BP2 Score showed quite poor performance on this screening power test. We think that this screening test based on pre-generated poses is rather limited to assess the full potential of this score, so we tried additional screening test which involves docking as a tool for generating possible poses, as discussed below.

To evaluate the virtual screening power of GalaxyDock BP2 Score with GalaxyDock2 protein-ligand docking program in more realistic circumstances, its performance on DUD data set (Huang et al. 2006) was compared to previously tested docking programs (Cross et al. 2009). The screening performance was evaluated by ROC enrichment factor (ROC EF) (Nicholls 2008) and area under curve of ROC (ROC AUC). ROC EF is used to evaluate early recovery of active compounds while ROC AUC is used to evaluate whether screening performance is better than random prediction or not. The results are summarized in **Table 3.9**. GalaxyDock2 with BP2 Score showed comparable performance in both of ROC EF and ROC AUC to other methods. Especially, GalaxyDock2 with BP2 Score showed good early recovery of active compounds (ROC EF value 18.0 at 0.5% false positive rates) following GLIDE-HTVS (18.9) and DOCK6.1 (18.8).

Table 3.8. Performance of 15 scoring functions in the screening power test on CASF-2013 benchmark set

Scoring Function	Enrichment Factor		
	Top 1%	Top 5%	Top 10%
GlideScore-SP[1)]	19.54	6.27	4.14
ChemScore@GOLD[1)]	18.90	6.83	4.08
GlideScore-XP[1)]	16.81	6.02	4.07
LigScore2@DS[1)]	15.90	6.23	3.51
ChemPLP@GOLD[1)]	14.28	5.88	4.31
LUDI1@DS[1)]	12.53	4.28	2.80
ASP@GOLD[1)]	12.36	6.23	3.79
GoldScore@GOLD[1)]	7.95	4.52	3.16
GalaxyDock BP2 Score	7.31	3.81	2.89
PLP1@DS[1)]	6.92	4.28	3.04
DSX	6.92	3.95	2.86
AutoDock4	5.90	4.37	3.00
Alpha-HB@MOE[1)]	4.87	3.23	1.32
GalaxyDock BP Score	3.85	2.47	2.51
X-Score	2.31	2.14	1.41

[1)] Taken from (Li et al. 2014a).

Table 3.9. Screening performance of docking programs on the DUD data set

Docking Program	ROC EF[1)]				ROC AUC[2)]
	0.5%	1.0%	2.0%	5.0%	
GLIDE-HTVS[3)]	18.9	14.8	10.7	6.5	0.72
DOCK 6.1[3)]	18.8	12.3	8.2	4.7	0.55
GalaxyDock2 with BP2 Score	18.0	12.3	8.6	4.8	0.61
ICM[3)]	16.9	12.7	8.0	4.6	0.63
PhDock[3)]	16.9	11.3	7.7	4.1	0.59
Surflex[3)]	14.3	11.1	7.9	4.9	0.66
FlexX[3)]	13.7	9.8	7.2	4.4	0.61

[1)] Mean ROC enrichment factors were calculated for early false positive rates

[2)] Mean ROC AUC of 0.5 indicates random performance

[3)] Taken from (Cross et al. 2009).

3.4. Conclusion on Protein-Ligand Docking Score

In this chapter, I introduced a newly developed docking scoring function named GalaxyDock BP2 Score by combining physics-based, knowledge-based, and empirical scoring functions. The scoring function can be used for scoring binding poses generated by other docking programs or as a scoring component of a protein–ligand docking program. The scoring function was optimized to have high energy-RMSD correlation and high discrimination power of near-native poses from non-native ones in the conformational space represented by pre-generated binding poses. The poses were generated iteratively during energy parameter optimization since the energy landscape can change with different parameters. I showed that the new score shows better decoy discrimination power than other available scoring functions when tested on two types of binding pose sets. Moreover, the new score combined with GalaxyDock2 outperforms other state-of-the-art docking methods when tested on the Astex diverse set, the Cross2009 benchmark set, and the Astex non-native set. This success in docking tests indicates that the new score has strength not only in decoy discrimination, but also in guiding conformational sampling during docking.

Interestingly, GalaxyDock BP2 Score showed a reasonably good performance in scoring binding affinities even though it was optimized for scoring poses, as reported in **section 3.3.6**. However, I think that a separate hybrid score must be developed in the future for the purpose of predicting binding affinity in virtual screening. Difficulties related to such development is that designing an

objective function for the purpose is not straightforward and the training procedure may take even more computer time than spent here. I envision that the current GalaxyDock BP2 Score can be applied to virtual screening of a compound database for a given target receptor by separating the stages of binding pose prediction and binding affinity prediction. Accurate binding pose prediction is a pre-requisite of accurate binding affinity prediction because binding affinity score is evaluated for the predicted binding pose. Further study in this direction is underway.

Chapter 4. Improving Docking Performance of Large Flexible Ligands Using Hot Spot Information Predicted by Fragment Docking

4.1. Introduction to Docking of Large Flexible Ligands

Computational protein-ligand docking is a technique that explores the conformational space of the protein-ligand complex in order to compute energetically stable conformations that model the structure of the complex. The success of a protein–ligand docking program depends on the program's performance on two famous, but still unsolved problems: scoring and sampling. In previous chapter, I discussed about the scoring problem in protein-ligand docking and introduced the improved docking scoring function named GalaxyDock BP2 Score. In this chapter, I will focus on improving sampling algorithm which is the other essential part of the protein-ligand docking.

It has been shown that docking small ligands with 6 or fewer rotatable bonds is in general very accurate (Plewczynski et al. 2011). However, as the dimensionality of the search space increases with large ligands, prediction of correct binding poses for large flexible ligands becomes very challenging. Tackling the challenge of docking large ligands is important for designing putative drug compounds that have many rotatable bonds like peptides or peptidomimetics (Mandal et al. 2011; McMurray 2008). The field of drug design based on the peptides or peptidomimetics is rapidly growing in the pharmaceutical industry because of their low toxicity and high specificity (Bellmann-Sickert and Beck-Sickinger 2010; Vlieghe et al. 2010; Lau and Dunn 2018). Development of an accurate docking program for large flexible ligands would be very useful to design

peptide or peptidomimetic drug compounds.

In this chapter, GalaxyDock-Frag, a new approach to improve the sampling ability of GalaxyDock protein-ligand docking program for large flexible ligands, is introduced. This approach is based on an assumption that binding "hot spots" of a specific receptor protein could be reasonably predicted for each fragment of a given ligand. This assumption has been widely tested in druggable site detection (Brenke et al. 2009), pharmacophore-based docking (Goto et al. 2004; Hu and Lill 2014), and fragment-based drug discovery (Sheng and Zhang 2013), and it has turned out that the assumption is quite reasonable.

Utilizing the predicted hot spot information might lead to efficient conformational search by reducing the conformational space that should be sampled. To identify "hot spots" for each fragment, target ligands were fragmented into rigid fragments, and up to five fragments were docked into the receptor protein using fast Fourier transformation (FFT)-based rigid-body docking. The detected fragment binding hot spot information was used to generate initial conformations and further trial conformations during conformational space annealing (CSA) global optimization. With this approach, higher quality initial conformations were obtained and the binding pose sampling of large flexible ligands was improved compared to the previous GalaxyDock2 program with GalaxyDock BP2 Score.

4.2. Methods

4.2.1. Overall Procedure

The improved version of protein-ligand docking method using fragment binding hot spots named GalaxyDock-Frag can be summarized as in **Figure 4.1**. First, a

given ligand molecule splits into rigid fragments. Among the rigid fragments, up to five fragments are selected based on their size and hydrophobicity. For each selected rigid fragment, millions of docked conformations are sampled by the rigid body docking that will be discussed in more details in **section 4.2.2**. Up to 20 binding hot spots per fragment are predicted based on fragment docking results. The detected fragment binding hot spot information is used to generate initial conformations and further trial conformations during conformational space annealing (CSA) global optimization.

Initial pool of seed ligand binding poses used in conformational space annealing global optimization are generated by enumerating all possible multiple-points matches between binding hot spots and corresponding rigid fragments of an ensemble of ligand conformations. The sampled binding poses are then optimized and selected as an initial pool using GalaxyDock BP2 Score. Starting with the initial pool of ligand binding poses, the ligand binding poses are optimized using very efficient and powerful global optimization algorithm named conformational space annealing (CSA).

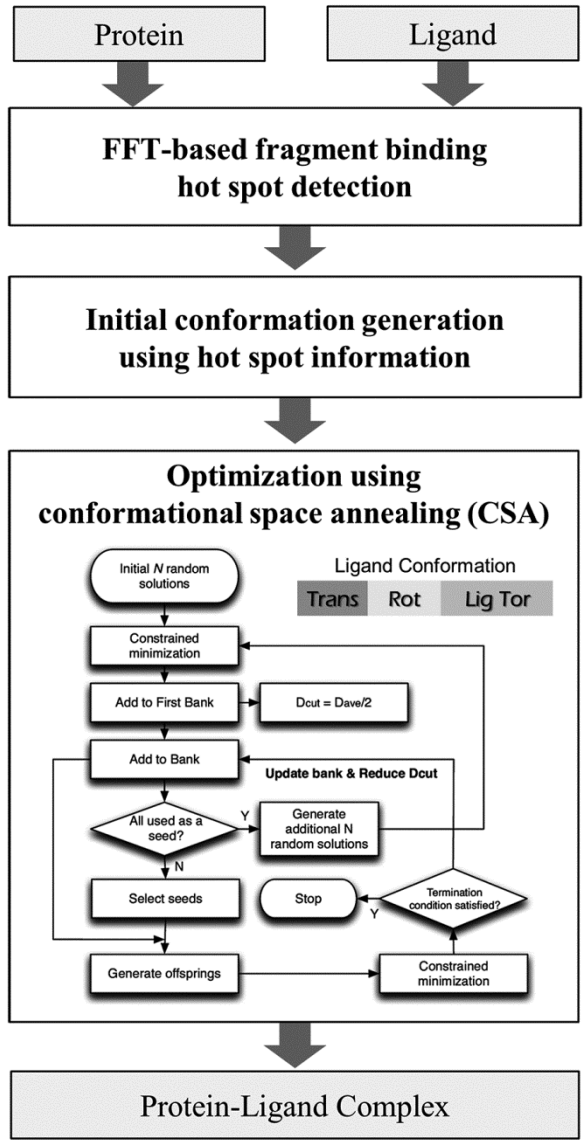


Figure 4.1. Flowchart of GalaxyDock-Frag protocol

4.2.2. Fragment Binding Hot Spot Detection Using FFT-based Fragment Docking

4.2.2.1. Ligand Fragmentation

A given ligand molecule splits into rigid fragments by breaking all possible rotatable bonds. All rigid fragments are then sorted by their sizes and hydrophobicities calculated using XLogP program (Wang et al. 2000). All the fragments having hydrophobicity ranged from -0.3 to 0.3 are ignored. Up to five rigid fragments are selected to detect binding hot spots by FFT-based rigid body docking algorithm.

4.2.2.2. Fragment Docking Based on the FFT Correlation Approach

Each fragment is docked into binding pockets in a receptor protein by evaluating fragment binding energy exhaustively in the discretized 6-dimensional space. The translational space is represented as a grid of 0.375 Å displacements of the fragment center of mass, and the rotational space is sampled using 300 rotations based on spherical Fibonacci point sets. The energy function describing the receptor-fragment interactions is defined on the grid and is expressed as the sum of P correlation functions for all possible translations α, β, γ of the fragment at a given rotation:

$$E(\alpha,\beta,\gamma)=\sum_{P}\sum_{l,m,n}R_P(l,m,n)L_P(l+\alpha,m+\beta,n+\gamma) \qquad (4.1)$$

where R_P and L_P are the components of the correlation function defined on the receptor and the fragment, respectively. This expression can be efficiently calculated using P forward and one inverse FFTs by reducing computational cost

from $O(N^6)$ to $O(N^3 \log N^3)$.

For each rotational orientation, which is taken consecutively from the set of rotations, the fragment is rotated and the L_P function is calculated on the grid. The correlation function of L_P with the pre-calculated R_P function is calculated using FFT. Three lowest energy translations for the given rotation are stored. Finally, results from different rotations are collected and sorted by their interaction energy.

4.2.2.3. Energy Function Used in FFT-based Fragment Docking

A grid version of GalaxyDock BP2 Score (Baek et al. 2017b) is used in FFT-based fragment docking. It includes the van der Waals energy (E_{vdW}), the hydrogen bond energy (E_{Hbond}), Coulomb interaction energy (E_{qq}), knowledge-based pairwise potential ($E_{DrugScore}$), and hydrophobicity matching score (E_{HM}).

The van der Waals interactions are expressed using 12-6 Lennard-Jones potential. To calculate van der Waals energy on grid space, R_P and L_P function designed as follows:

$$
\begin{aligned}
E_{vdW} &= \sum_{\text{atom type J}}^{\text{ligand}} R_P(l,m,n) L_P(l+\alpha, m+\beta, n+\gamma) \\
R_P(l,m,n) &= \sum_{i_r} \left(\frac{A}{r^{12}} - \frac{B}{r^6}\right), \quad r = \left| R_{rec}(i_r) - R_{grid}(l,m,n) \right| \\
L_P(l,m,n) &= \begin{cases} 1 & \text{if } (l,m,n) \text{ overlaps with ligand atom} \\ 0 & \end{cases}
\end{aligned}
\tag{4.3}
$$

where (l,m,n) is the grid point, and i_r denotes index of atoms in receptor protein.

The hydrogen bonding interactions on the grid space are described based on similar formula to Eq. (4.3). The only difference is hydrogen bonding interactions are described using 12-10 Lennard-Jones potential instead of 12-6 Lennard-Jones potential. The Coulomb interactions are expressed as in Eq. (4.4) on the grid space.

$$
\begin{aligned}
E_{qq} &= \sum_{i_l} R_P(l,m,n) L_P(l+\alpha, m+\beta, n+\gamma) \\
R_P(l,m,n) &= \sum_{i_r} \frac{\varepsilon(r) q_{i_r}}{r}, \quad r = \left| R_{rec}(i_r) - R_{grid}(l,m,n) \right| \\
L_P(l,m,n) &= \begin{cases} \mathrm{q}_{\mathrm{i_l}} & \text{if } (l,m,n) \text{ overlaps with ligand atom } i_l \\ 0 \end{cases}
\end{aligned}
\tag{4.4}
$$

where q_{i_r} and q_{i_l} are the partial charges of atom i_r and i_l in receptor and ligand, respectively. The knowledge-based pairwise potential is formulated as in Eq. (4.5).

$$
\begin{aligned}
E_{DrugScore} &= \sum_{\text{atom type J}}^{\text{ligand}} R_P(l,m,n) L_P(l+\alpha, m+\beta, n+\gamma) \\
R_P(l,m,n) &= \sum_{i_r} V_{DrugScore}(r), \quad r = \left| R_{rec}(i_r) - R_{grid}(l,m,n) \right| \\
L_P(l,m,n) &= \begin{cases} 1 & \text{if } (l,m,n) \text{ overlaps with ligand atom} \\ 0 \end{cases}
\end{aligned}
\tag{4.5}
$$

The final term, hydrophobicity matching score can be expressed on the grid space as following:

$$
\begin{aligned}
E_{HM} &= \sum_{i_l}^{\text{ligand}} R_P(l,m,n) L_P(l+\alpha, m+\beta, n+\gamma) \\
R_P(l,m,n) &= \begin{cases} 1 & \text{if } \mathrm{env_{hp}}(l,m,n) > -0.5 \\ 0 \end{cases} \\
L_P(l,m,n) &= \begin{cases} \log P(i_l) & \text{if } (l,m,n) \text{ overlaps with ligand atom} \\ 0 \end{cases}
\end{aligned}
\tag{4.6}
$$

where $env_{hp}(l, m, n)$ is the environment hydrophobicity defined as a sum of the hydrophobicities of receptor atoms within 6 Å from the grid point (l, m, n).

4.2.2.4. Fragment Binding Hot Spot Detection

For each fragment, all docked fragment binding poses in **section 4.2.2.2** are clustered based on its center of mass using DBSCAN algorithm (Ester et al. 1996). Clusters are sorted by its size, and up to 20 fragment binding hot spots per fragments are defined as geometric centers of each cluster.

4.2.3. Initial Ligand Binding Poses Generation Using Predicted Hot Spot Information

An initial pool of ligand binding poses used in further CSA optimization are generated by the process depicted in **Figure 4.2**. First, the ligand conformations without internal clashes are sampled by perturbing all rotatable angles randomly. The generated ligand conformation are then placed into the binding site based on a Bron-Kerbosch clique detection algorithm (Bron and Kerbosch 1973) that enumerates all possible multi-points matches of fragment and its corresponding binding hot spots as described in next paragraph.

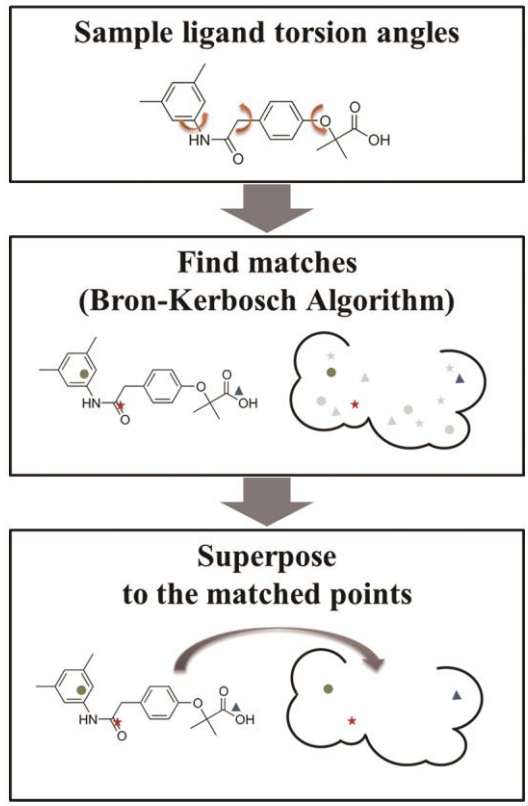


Figure 4.2. Ligand binding pose generation method using predicted binding hot spot information

All edge lengths, defined as distances between selected fragments in the ligand, are calculated. The edge lengths between all fragment binding hot spots are also calculated. All ligand fragment edges that match the fragment binding hot spots edges based on the fragment types of their vertices and edge lengths are identified. Throughout the matching process, a tolerance of 0.5 Å for the edge lengths is allowed. The matching process can be represented by a graph in which each node represents a matching fragment and corresponding binding hot spot pair. The Bron-Kerbosch clique detection algorithm then identifies all the completely connected subgraphs from this graph. All the vertices in the completely connected subgraph (clique) are matched points which can be used to place ligand into binding pockets.

The ligand is placed into binding sites by superposing the ligand fragments to its matched binding hot spots in each clique followed by local optimization using simplex algorithm (Nelder and Mead 1965). The GalaxyDock BP2 Score is used in local optimization. If the minimized ligand binding pose has energy lower than 500.0, it is added to the initial pool of ligand binding poses used in CSA. When the number of ligand binding poses in the initial pool is reached 50, the initial binding pose generation step is terminated.

4.2.4. Global Optimization Using Conformational Space Annealing

The initial pool of ligand binding poses is further optimized by conformational space annealing (CSA) global optimization method. CSA is an efficient global optimization technique that has been applied to protein-ligand docking successfully (Shin et al. 2011; Shin and Seok 2012; Shin et al. 2013; Baek et al. 2017b). In CSA, a relatively small number of ‘bank’ conformations are evolved by gradually

reducing the effective size of the conformational space explored by each bank member. A distance measure in the conformational space is introduced as an annealing parameter for this purpose. During evolution of the bank, trial conformations are generated by crossovers and mutations as in a genetic algorithm.

In this work, the conformational space is formed by three translational, three rotational, and N_{tor} torsional degrees of freedom for ligand. The translational degrees of freedom are represented by the Cartesian coordinate of the ligand center atom, and the rotational degrees of freedom by the quaternion for the ligand orientation. The number of conformations in bank is set to 50.

At each iteration step of CSA, trial conformations are generated by three kinds of operators: (1) crossovers and mutations of translational and rotational degrees of freedom, (2) crossovers and mutations of torsional degrees of freedom, and (3) pose sampling using fragment binding hot spot information as described in **section 4.2.3** after perturbing torsion angles by crossovers. The operators are represented in **Figure 4.3**. The ratio between three operators is 3:5:2.

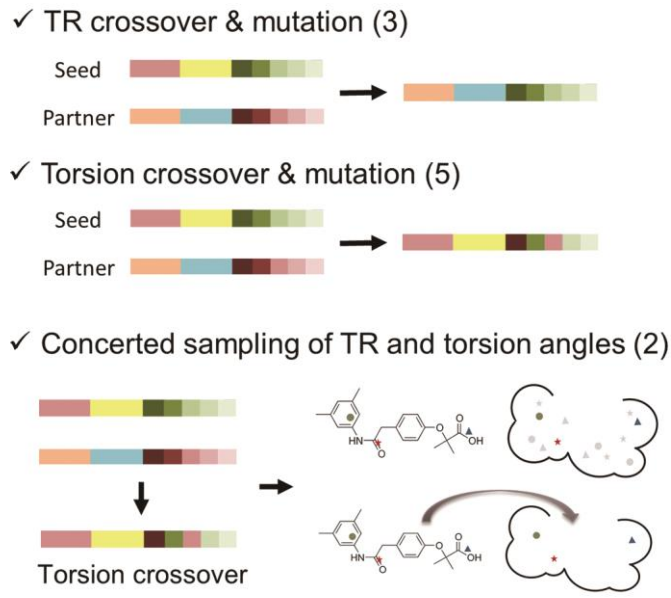


Figure 4.3. Operators used to sample binding poses during conformational space annealing (CSA)

After trial conformations are generated, the bank is updated considering structural diversity and energy of the current bank and the trial conformations. The essence of CSA is to focus on narrower conformational space of lower energy gradually as the iteration proceeds. For this purpose, a measure of distance between two conformations is required, and RMSD between conformations are used as the distance measure. In CSA, the effective size of the conformational space represented by each bank member is controlled by the distance parameter D_{cut} at the stage of bank update. If a trial conformation within D_{cut} from a bank conformation has lower energy than the bank conformation, it replaces the bank conformation. If a trial conformation has distances greater than D_{cut} from all the current bank conformations and has lower energy than the highest energy bank conformation, it replaces the highest energy bank conformation. The parameter D_{cut} is gradually reduced as CSA iteration proceeds, and therefore, conformational search focuses on narrower spaces of lower energy. When all bank members are used as seed, one round of CSA terminates. Two rounds of CSA are executed in this study.

4.2.5. Benchmark Test Sets

Two sets of protein–ligand complexes were used for testing the performance of the developed method. The first test set was compiled from the refined set of PDBbind 2013 database (Li et al. 2014b). The database was clustered with a sequence identity cutoff of 30% to remove redundancy resulting in 331 protein-ligand complexes. The second test set comprises 53 protein-peptide complexes of LEADS-PEP benchmark set (Hauser and Windshugel 2016).

4.3. Results and Discussions

4.3.1. The Effect of Utilizing Predicted Hot Spot Information in Generating Initial Ligand Binding Poses

To evaluate the correctness of predicted fragment binding hot spot information used in GalaxyDock-Frag, a rigid ligand docking experiment was performed. The native ligand conformation was docked into the binding sites using the binding hot spot information predicted by FFT-based fragment docking as described in the **section 4.2.3**. The result is shown in **Figure 4.4**. It shows the success rate of 82.5%, 89.4%, and 93.4% when the success of docking is defined as a prediction having the ligand RMSD lower than 1.0 Å, 2.0 Å, and 3.0 Å, respectively. This result implies that, in most cases, it is a reasonable assumption that binding "hot spots" of a specific receptor protein could be reasonably predicted for each fragment of a given ligand.

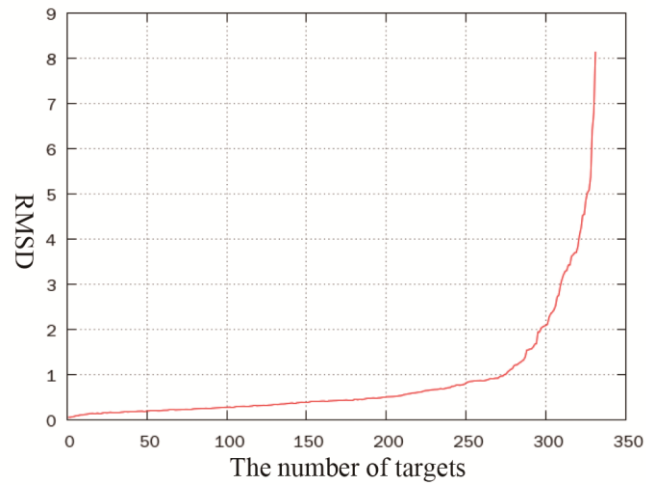


Figure 4.4. The cumulative number of targets within various RMSD cutoffs when the native ligand conformation docked into binding sites

GalaxyDock searches ligand conformational space using CSA, a population-based global optimization technique, during the docking. It has been shown that performances of population-based methods tend to depend on the quality of the initial population (Kazimipour et al. 2013). Therefore, it is highly expected that the performance of GalaxyDock can be improved by improving the quality of the initial pool of ligand binding poses. The performance of initial ligand binding pose generation based on predicted binding hot spots was compared to that of a geometry-based docking method named BetaDock employing beta-complex, derived from the Voronoi diagram, used in GalaxyDock2. The performance of initial population generation was evaluated using the success rate when the lowest RMSD ligand binding pose is considered. The result is summarized in **Table 4.1**. When the docking success is defined as a prediction having the ligand RMSD lower than 1.0 Å, 2.0 Å, and 3.0 Å, respectively, the success rates of binding hot spot-based method introduced in this study are 54.1%, 76.7%, and 87.2% while those of BetaDock are 40.6%, 70.6%, and 86.1%, respectively.

Table 4.1. The success rate of initial binding pose sampling method used in GalaxyDock-Frag and GalaxyDock2 with various RMSD cutoffs when the lowest RMSD conformation among sampled binding poses is considered

Success rate	< 1.0 Å	< 2.0 Å	< 3.0 Å
GalaxyDock-Frag	54.1%	76.7%	87.2%
GalaxyDock2	40.6%	70.6%	86.1%

4.3.2. The Docking Performance Comparison on the PDBbind Set

To assess the docking performance of the newly developed GalaxyDock-Frag, it was tested on the compiled PDBbind set consisting of 331 targets having various torsional degrees of freedom ranging from 1 to 30. The docking performance of the GalaxyDock-Frag was compared to the performance of GalaxyDock2 developed by our group previously. The success rate of GalaxyDock-Frag is 79.2% while that of GalaxyDock2 is 77.3%. When the average RMSD of the lowest energy binding poses is considered, the performance differences between GalaxyDock-Frag and GalaxyDock2 become more dramatic (1.86 Å and 2.48 Å, respectively). The GalaxyDock-Frag predicted ligand binding poses more accurately compared to the GalaxyDock2.

Detailed comparison of GalaxyDock-Frag and GalaxyDock2 is shown in **Figure 4.5**. In general, GalaxyDock-Frag could predict the ligand binding poses having lower energy and lower RMSD compared to the GalaxyDock2. Especially, GalaxyDock-Frag predicted better binding poses of large flexible ligands than GalaxyDock2 program. This tendency implies that utilizing fragment binding hot spot information is indeed helpful to reduce the search space by reducing the translational and rotational degrees of freedom during the docking procedure.

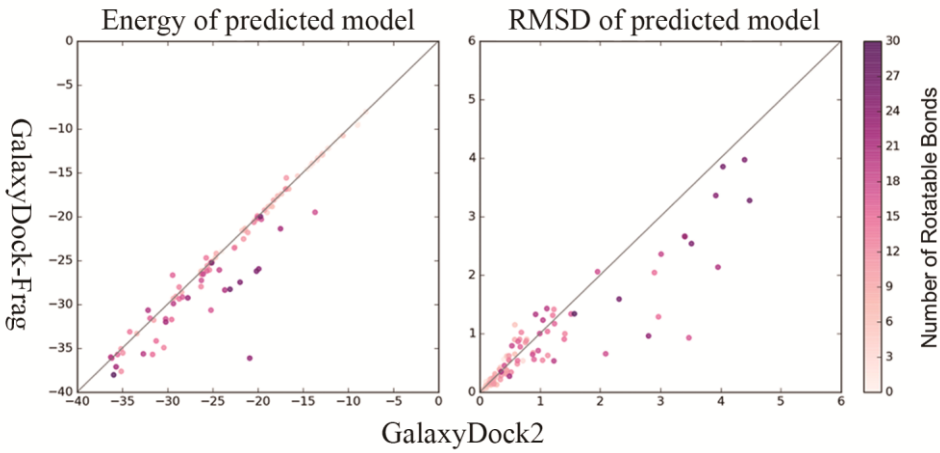


Figure 4.5. Performance comparison between GalaxyDock-Frag and GalaxyDock2. The energy and RMSD values of the lowest energy binding poses are shown in left and right panel, respectively.

4.3.3. The Peptide Docking Performance Test on the LEADS-PEP Benchmark Set

To evaluate the docking ability of GalaxyDock-Frag for large flexible ligand, the docking performance test was conducted on the LEADS-PEP benchmark set consisting of 53 protein-peptide complexes having residues ranging from 3 to 12 residues. As in LEADS-PEP paper (Hauser and Windshugel 2016), a docking pose was considered as near-native conformation once its backbone RMSD is ≤2.5 Å. The RMSD of top-scored docking poses predicted by GalaxyDock2 and GalaxyDock-Frag were calculated and were compared with the results of the other state-of-the-art protein-ligand docking programs. This benchmark result is summarized in **Table 4.2** and **Figure 4.6**.

Considering the median RMSD over the whole benchmark data set for the top-scored pose, GalaxyDock-Frag utilizing predicted hot spot information revealed as most accurate docking approach (3.4 Å), followed by GOLD (4.6 Å), GalaxyDock2 (4.7 Å), and Surflex (5.0 Å). The median RMSD of AutoDock4 and AutoDock-Vina are 6.9 Å and 7.2 Å, respectively. When the success rates of the docking methods were considered, the method developed in this study showed the highest success rate of 43.4%, followed by Surflex (32.1%) and GOLD (30.2%).

As shown in **Table 4.2**, most docking programs were capable to predict binding poses of shorter peptides (3−4 residues) quite accurately. When the medium peptides (5~7 residues) are only considered, GalaxyDock-Frag shows the highest success rate of 41.2%, followed by GOLD (29.4%) and Surflex (23.5%). For longer peptides (8~12 residues), Surflex shows the best performance (24.0%), followed by GalaxyDock-Frag (16.0%) and GOLD (12.0%).

Table 4.2. Peptide Docking Performance in terms of RMSD (in Å) as Measured by Best Scored Binding Modes

PDB	Res.	AutoDock4*	AutoDock-Vina*	Surflex*	GOLD*	GalaxyDock2	GalaxyDock-Frag
1B9J	3	1.1	1.0	0.4	0.4	0.4	0.3
2OY2	3	0.5	7.2	7.1	0.4	0.6	0.6
3GQ1	3	2.5	1.6	0.9	4.4	2.0	1.7
3BS4	3	0.5	0.7	0.4	0.9	0.5	0.4
2OXW	3	3.4	6.8	7.1	6.8	1.5	1.4
2B6N	3	8.4	7.8	7.9	8.6	0.8	0.5
1TW6	4	1.3	1.0	0.9	0.4	0.6	0.5
3VQG	4	2.8	0.6	0.7	0.7	0.7	0.8
1UOP	4	0.6	6.4	6.5	0.4	0.7	0.7
4C2C	4	1.0	0.6	0.7	1.0	1.3	1.3
4J44	4	1.0	0.8	0.9	0.8	0.8	0.6
2HPL	5	6.9	2.8	7.4	7.3	2.4	0.9
2V3S	5	3.9	5.6	11.2	1.6	3.9	2.3
3NFK	5	4.2	8.4	6.7	3.4	4.7	1.1
1NVR	5	7.3	9.0	9.1	5.2	2.8	2.9
4V3I	5	9.7	6.4	3.1	7.3	2.0	1.6
3T6R	5	4.4	7.2	7.4	0.7	2.4	1.4
1SVZ	6	5.1	8.0	6.5	6.4	4.8	3.9
3D1E	6	9.4	10.6	9.5	9.6	4.0	4.7
3IDG	6	6.3	7.2	5.0	9.7	4.0	3.8
3LNY	6	7.4	11.3	11.3	3.9	9.1	3.4
4NNM	6	9.7	0.8	3.3	1.5	3.3	2.8
4Q6H	6	10.7	9.9	8.7	2.8	3.9	3.9
3MMG	7	10.4	1.2	2.1	1.3	9.7	2.4
3Q47	7	6.7	9.9	7.7	7.7	4.9	2.4
3UPV	7	4.7	4.9	2.5	4.6	3.2	3.3
4QBR	7	8.6	11.3	1.2	1.9	5.4	3.4
3NJG	7	2.5	2.6	0.4	2.7	5.6	3.3
1ELW	8	3.2	9.2	2.5	3.5	7.6	6.3
3CH8	8	7.8	5.4	5.4	6.5	6.9	5.9
4WLB	8	5.3	6.3	5.0	6.5	5.2	4.4
1OU8	8	10.2	7.5	4.7	3.9	3.9	4.1
1N7F	8	11.9	14.3	9.2	8.4	4.6	5.0
3OBQ	9	12.5	14.5	2.2	5.5	6.1	5.9
4BTB	9	14.7	15.5	8.6	9.7	13.5	7.4
2W0Z	9	11.3	14.3	4.4	14.3	2.6	0.8
4N7H	9	7.4	12.1	6.8	2.2	5.0	5.6

2QAB	9	4.3	4.4	4.5	4.8	7.5	5.9
1H6W	10	13.9	3.2	2.6	1.5	3.7	4.4
3BRL	10	11.3	4.4	3.1	2.5	5.8	5.3
1NTV	10	4.9	4.7	15.3	13.9	3.4	4.1
4DS1	10	5.5	17.5	1.6	5.4	6.4	1.8
2O02	10	5.0	4.9	12.0	4.0	5.2	5.6
1N12	11	12.6	16.8	1.3	4.5	10.0	5.0
2XFX	11	15.6	2.0	1.4	7.0	9.7	1.9
3BFW	11	11.5	18.5	0.4	19.8	7.4	1.2
4EIK	11	5.3	7.8	4.6	4.1	7.8	5.6
3DS1	11	5.8	5.2	12.6	8.0	4.2	5.2
4J8S	12	7.3	11.4	13.9	14.2	6.3	6.0
2W10	12	15.1	15.4	4.8	5.6	10.7	8.9
3JZO	12	5.8	6.1	13.4	9.8	5.9	5.8
4DGY	12	9.3	8.9	7.8	8.9	7.0	6.7
2B9H	12	13.2	9.5	10.2	4.3	8.5	6.8

* Data taken from (Hauser and Windshugel 2016)

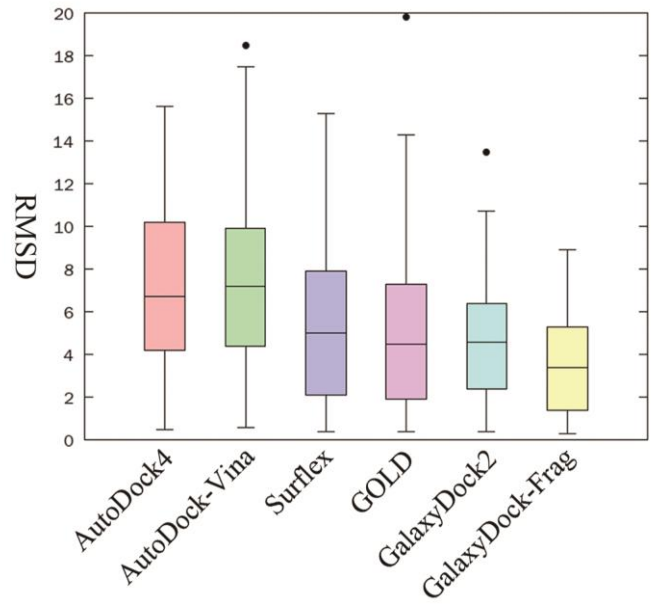


Figure 4.6. Distribution of RMSDs of the lowest energy conformations selected by AutoDock4, AutoDock-Vina, Surflex, GOLD, GalaxyDock2 with BP2 Score, and GalaxyDock-Frag for the LEADS-PEP benchmark set

Figure 4.7 shows two successful cases of GalaxyDock-Frag, 2HPL and 3NFK. In both cases, the fragment binding hot spots depicted transparent sphere were predicted correctly, and it guided the correct binding pose prediction resulting in RMSD 0.92 Å and 1.14 Å, respectively. It implies that the fragment binding hot spots could be predicted well, and utilizing this predicted binding hot spot information could lead to a correct prediction of binding poses for large flexible ligands.

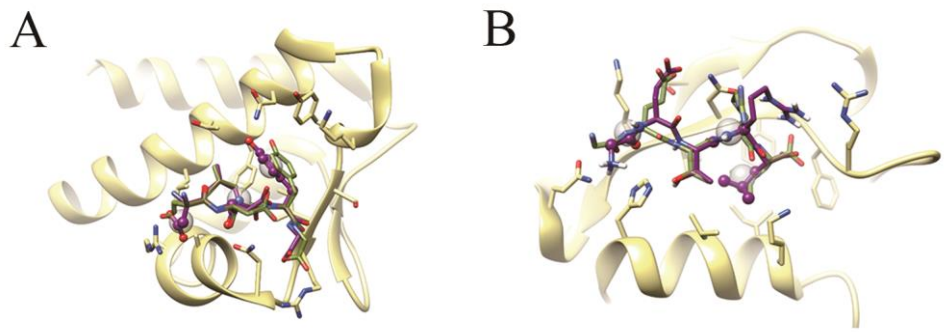


Figure 4.7. Successful examples of GalaxyDock-Frag. Predicted binding poses of GalaxyDock-Frag (dark magenta) are compared to native binding poses dark green) for two examples, 2HPL (A) and 3NFK (B). The predicted binding hot spots are presented in gray spheres.

4.4. Conclusion on Docking of Large Flexible Ligands

In this chapter, I introduced a newly developed protein-ligand docking program name GalaxyDock-Frag with an improved sampling algorithm based on predicted binding hot spot information. By FFT-based fragment docking, binding hot spots for each fragment could be predicted well, and it could reduce the search space by sampling translational and rotational degrees of freedom using graph-based sampling algorithm utilizing binding hot spot information. GalaxyDock-Frag showed better performance than previous GalaxyDock2 with BP2 Score version when it tested on the PDBbind set. GalaxyDock-Frag could find more near-native conformations with lower energy compared to GalaxyDock2, especially for large flexible ligands having more than 20 rotatable bonds.

GalaxyDock-Frag showed a superior performance to other protein-ligand docking programs on LEADS-PEP benchmark set consisting of 53 protein-peptide complexes. It showed better performance for peptides with medium length (5~7 residues). The docking performance of GalaxyDock-Frag for larger peptides (8~12 residues) is comparable to Surflex and GOLD, but it has a room for improvement. In current GalaxyDock-Frag method, only translational and rotational degrees of freedom are sampled using predicted hot spot information. Therefore, only 6 dimensions or search space are reduced, and it is not sufficient to dock larger peptides having more than 8 residues efficiently. If torsional degrees of freedom could be sampled with predicted hot spot information, the search space could be reduced more. The further study in this direction is now underway.

The GalaxyDock-Frag method, introduced in this chapter, can be further extended to information-driven protein-ligand docking program. As the number of protein-ligand complex structures increases, it becomes easier to get interaction

information from the structure database. If similar protein-ligand complexes already exist in structure database, hot spot information might be extracted from the database instead of predicting by *ab initio* docking.

Chapter 5. Prediction of Protein Homo-oligomer Structures

5.1. Introduction to Homo-oligomer Structure Prediction

A large fraction of cellular proteins self-assemble to form symmetric homo-oligomers with distinct biochemical and biophysical properties (Andre et al. 2008; Goodsell and Olson 2000; Poupon and Janin 2010). For example, ligand-binding sites or catalytic sites are located at oligomer interfaces in many proteins (Snijder et al. 1999; Ali et al. 2010; Pidugu et al. 2016), and oligomerization is often necessary for effective signal transduction through membrane receptor proteins (Heldin 1995; Stock 1996) and selective gating of channel proteins (Clarke and Gulbis 2012). Therefore, knowledge of the homo-oligomer structure is essential for understanding the physiological functions of proteins at the molecular level and for designing molecules that regulate the functions.

Methods for predicting the protein homo-oligomer structure can be divided into two categories: those that use templates selected from the protein structure database and others that dock monomer structures *ab initio*, without using template information. Usually, template-based methods require a sequence as input, whereas docking methods require a monomer structure as input. The latter requirement can be more restrictive for the user if the monomer structure has to be predicted by another method, but it may be preferred if an experimentally resolved monomer structure is available. It is generally expected that template-based

methods produce more accurate predictions under a situation in which similar proteins forming oligomers exist in the structure database. Docking methods may be more useful when proper oligomer templates are not available but the monomer structure is reliable. Several protein–protein docking methods have been reported to date (Dominguez et al. 2003; Gray et al. 2003; Comeau et al. 2004; Pierce et al. 2005; Schneidman-Duhovny et al. 2005; Tovchigrechko and Vakser 2006; Macindoe et al. 2010; Torchala et al. 2013; Lensink et al. 2016), and some of these are available as public web servers for predicting homo-oligomer structures. M-ZDOCK (Pierce et al. 2005) and GRAMM-X (Tovchigrechko and Vakser 2006), which use *ab initio* docking based on fast Fourier transformation (FFT), are two such examples. The oligomeric state must be provided as input in these servers. However, relatively few web servers that use template-based methods have been reported. ROBETTA (Kim et al. 2004; DiMaio et al. 2011) and SWISS-MODEL (Biasini et al. 2014) are two web servers that predict the homo-oligomer structure from an amino acid sequence. GalaxyGemini (Lee et al. 2013) predicts the homo-oligomer structure from a monomer structure. These servers predict the oligomeric state automatically. Depending on the availability of information on the oligomeric state, the user may or may not prefer to specify the oligomeric state. Here, I introduce a new method called GalaxyHomomer that predicts the homo-oligomer structure from either the amino acid sequence or from the monomer structure (Baek et al. 2017a). It can perform both template-based oligomer modeling and *ab initio* docking. It returns five model structures and automatically decides how many models are generated by which method depending on the existence of proper

oligomer templates.

Oligomer structures predicted by template-based methods may have errors due to sequence differences between the target and template proteins. Those predicted by docking methods may have inaccuracy if structural change of the monomer induced by oligomerization is not considered. In the previous CASP experiment conducted in 2014 in collaboration with CAPRI, we showed that such errors in predicted oligomer structures could be reduced by re-modeling inaccurately predicted loops or termini and by relaxing the overall structure (Lee et al. 2016). GalaxyHomomer incorporates such state-of-the-art model refinement methods to improve the accuracy of homo-oligomer models generated by both template-based modeling and *ab initio* docking.

According to the assessment of the recent blind prediction experiment CASP12 conducted in 2016, GalaxyHomomer, participated as "Seok-assembly", ranked second among the servers participated in the assembly category (Lensink et al. 2018; Lafita et al. 2018). When I tested GalaxyHomomer on 136 targets from PISA benchmark set, 47 targets from a membrane protein set, 20 targets from CASP11 experiments, and 89 targets from CAMEO protein structure prediction category, it showed a performance better than or comparable to that of other available homo-oligomer structure prediction methods.

5.2. Methods

5.2.1. Overall Procedure

The overall pipeline of GalaxyHomomer is presented in **Figure 5.1**. Either a sequence or structure (experimental or predicted structure) of the monomer can be provided as input. If the oligomeric state is not specified by the user, possible oligomeric states are predicted first. Five homo-oligomer structures with the given oligomeric states are then generated by template-based modeling and *ab initio* docking. Oligomer templates required by template-based modeling are detected based only on the sequence as well as with additional structure information. The models are further refined by loop/terminus modeling using GalaxyLoop (Lee et al. 2010; Park and Seok 2012a; Park et al. 2014a) and by overall relaxation using GalaxyRefineComplex (Heo et al. 2016).

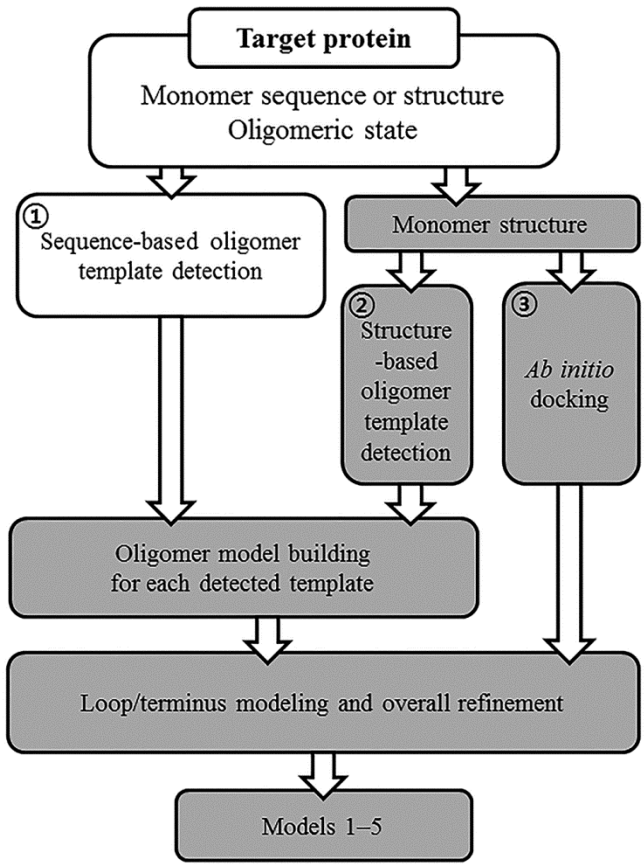


Figure 5.1. Flowchart of the GalaxyHomomer algorithm. The homo-oligomer structure prediction methods based on sequence similarity, structure similarity and ab initio docking are attempted in the order in which they are numbered until five homo-oligomer models are generated. When the monomer structure is given as input, only shaded procedures are executed.

5.2.2. Prediction of the Oligomeric State

Possible oligomeric states are predicted from the input sequence by a similarity-based method as follows. First, HHsearch (Soding 2005) is run in the local alignment mode to detect proteins that are similar to the target in the protein structure database 'pdb70', with a maximum mutual sequence identity of 70%. The oligomeric states of the database proteins were assigned according to the biological units described in 'REMARK 350'. Second, the proteins are re-ranked by a score S, which combines the HHsearch sequence score, HHsearch secondary structure score, and sequence identity between target and templates (Ko et al. 2012). Next, the S scores of the proteins in the same oligomeric states are summed for the top 100 proteins, and the ratios of different oligomeric states are determined in proportion to the S sums. Finally, oligomeric states for five models are assigned according to the oligomeric state ratios.

5.2.3. Template-based Oligomer Modeling

The same top 100 proteins described above are considered as candidates for oligomer templates. If a sequence is provided as input, up to five proteins are selected as templates based on the ranking of S among those with S greater than 0.2 times the highest S overall and those greater than 0.7 times the highest S for the given oligomeric state. If the number of detected templates using this sequence-based method is less than five, additional templates are selected using the monomer structure predicted by the template-based modeling program GalaxyTBM (Ko et al. 2012). Structure-based templates are then selected according to the ranking of S among those with monomer structures similar to the given monomer structure (TM-score calculated using TM-align (Zhang and Skolnick 2005) > 0.5) and in the

given oligomeric state. If a structure is provided as input, only the structure-based template detection is used with the monomer structure provided by the user.

For each oligomer template detected by the sequence-based method, an oligomer structure is built using the in-house model-building program GalaxyCassiopeia, a component of the most recent version of GalaxyTBM (Ko et al. 2012). GalaxyCassiopeia builds models from the sequence alignment and template structure by the VTFM optimization used in MODELLER (Sali and Blundell 1993) but with FACTS solvation free energy (Haberthur and Caflisch 2008), knowledge-based hydrogen bond energy (Kortemme et al. 2003), and dipolar-DFIRE (Yang and Zhou 2008) in addition to molecular mechanics bonded and non-bonded energy terms and template-derived restraints. For each template detected by the structure-based method, an oligomer structure is built by superimposing the monomer structure onto the oligomer template.

5.2.4. *Ab initio* Docking

If less than five oligomer templates are detected by the two template detection methods described above, the remaining homo-oligomer models with the given oligomeric states are generated using the in-house *ab initio* docking program GalaxyTongDock_C. GalaxyTongDock_C predicts C_n-symmetry homo-oligomer structures from the monomer structure using a grid-based FFT docking method similar to M-ZDOCK (Pierce et al. 2005) considering C_n-symmetry. The top 200 homo-oligomer structures generated by FFT are clustered using NMRCLUST (Kelley et al. 1996), and the clusters are ranked according to the cluster size. From each of the highest ranking clusters, the highest-score structure is selected.

5.2.5. Structure Refinement Using Loop Modeling and Global Optimization

Less reliable loop or terminal regions are re-modelled using GalaxyLoop (Lee et al. 2010; Park and Seok 2012a; Park et al. 2014a) considering symmetry of the homo-oligomer structure for the first model for those regions predicted to be unreliable if a sequence is provided as input, and for all five models for user-specified regions if a structure is provided as input. GalaxyRefineComplex (Heo et al. 2016) is subsequently run to further relax the overall structure.

5.3. Results and Discussions

5.3.1. Overall Performance of GalaxyHomomer Method

The GalaxyHomomer server was tested on 25 targets in CASP12 in a blind fashion, and this server, named "Seok-assembly", ranked second among the servers participated in the assembly category (Lensink et al. 2018; Lafita et al. 2018). In CASPs, the oligomeric state is provided by the organizers. The server was also tested on three benchmark sets for which the oligomeric state is given as input (136 homo-oligomer proteins from the PISA benchmark set (Ponstingl et al. 2003), 47 homo-oligomer membrane proteins compiled from the PDB, and 20 homo-oligomer proteins among the targets of CASP11 held in 2014 in collaboration with CAPRI (Lensink et al. 2016)) and on a set for which the oligomeric state is not provided as input (89 homo-oligomer proteins among CAMEO (Haas et al. 2013) targets released from August 13, 2016 to November 11, 2016). In these tests, the performance of GalaxyHomomer was better than or comparable to that of other methods for which performance data are available for the sets in terms of the CAPRI accuracy criterion, as summarized in **Table 5.1**. Note that some methods take only the structure as input. The CAPRI criterion reflects the biological

relevance of the model structures, and model qualities are classified as high (***), medium (**), acceptable (*), and incorrect considering the ligand root mean-square deviation (L-RMSD) and interface RMSD (I-RMSD) from the experimental structure and the fraction of predicted native contacts (F_{nat}) (Lensink and Wodak 2010)**.**

Table 5.1. Performance comparison of homo-oligomer structure prediction methods in terms of the CAPRI accuracy criteria

Benchmark Set	Prediction Methods	Input	Up to 5 models[1)]	Top 1 Model[1)]
PISA (136 targets)[2)]	GalaxyHomomer	Sequence	62/5***/38**	57/3***/39**
	HH+MODELLER[3)]	Sequence	61/3***/38**	45/1***/26**
Membrane proteins (47 targets)[2)]	GalaxyHomomer	Sequence	19/1***/14**	19/1***/9**
	HH+MODELLER	Sequence	18/0***/6**	14/0***/4**
CASP11 (20 targets)[2)]	GalaxyHomomer	Sequence	12/0***/8**	12/0***/5**
	HADDOCK	Structure	14/0***/10**	13/0***/9**
	ClusPro	Structure	14/0***/7**	10/0***/5**
	BAKER-ROSETTASERVER	Sequence	9/0***/8**	9/0***/7**
	SwarmDock	Structure	9/0***/3**	8/0***/3**
	GalaxyGemini[4)]	Structure	Not available	7/0***/5**
	GRAMM-X	Structure	5/0***/1**	3/0***/1**
CAMEO (89 targets)	GalaxyHomomer	Sequence	44/6***/25**	35/3***/25**
	Robetta	Sequence	28/4***/17**	26/4***/15**
	SWISS-MODEL[4)]	Sequence	Not available	23/3***/16**

[1)] Data represent the numbers of targets for which the best of up to five predicted models were of acceptable or higher/high accuracy (***) and medium accuracy (**); values for model 1 are shown.

[2)] Oligomeric state of target protein is given as an input.

[3)] Up to five homo-oligomer models were generated by MODELLER based on the templates detected by HH-search

[4)] Data for up to five models were not provided for GalaxyGemini and SWISS-MODEL because they generated only single models.

It has to be noted that GalaxyHomomer does not consider the lipid bilayer environment of membrane proteins explicitly in terms of energy or geometry during energy-based optimization and docking. However, the results on membrane proteins in **Table 5.1** are quite promising, implying that membrane environment was effectively taken into account in an implicit manner by using the database structures of membrane proteins as templates. GalaxyHomomer showed better performance than GalaxyGemini (Lee et al. 2013), a previous homo-oligomer structure prediction server developed by us, on the CASP11 benchmark set, as summarized in **Table 5.1**. The difference in the performance is mainly due to the cases in which predicted monomer structures are not accurate enough. In such cases, oligomer structures built directly from the sequence using sequence-based templates (method 1 in **Figure 5.1**) tended to be more accurate than those obtained by superimposing the predicted monomer structures on the structure-based templates (method 2 in **Figure 5.1**). GalaxyGemini builds oligomer models using only method 2. Additional model refinement performed by GalaxyHomomer also improved the model accuracy.

5.3.2. The Effect of Loop Modeling and Global Refinement on Homo-oligomer Model Quality

Improvement of the predictions achieved by additional ULR modeling and global refinement was analyzed using the benchmark test results on CASP11 benchmark set. ULR modeling was performed on 11 of the 13 targets for which initial models were generated based on proper templates. Both L-RMSD and I-RMSD were, on average, improved by ULR modeling as shown in **Figure 5.2**. Therefore, ULR modeling contributed to enhancing the prediction accuracy of the oligomer

structure interface.

Structures of the two example cases, T85 and T90, in which I-RMSD was improved significantly by interface loop modeling, are presented in **Figure 5.3**. In both cases, improved loop structures lead to improved interface structure, measured by I-RMSD, and overall docking accuracy, measured by L-RMSD.

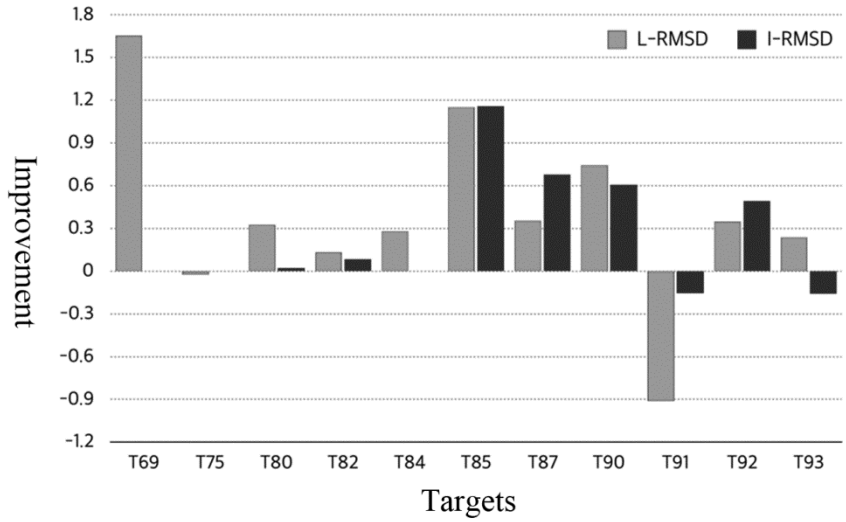
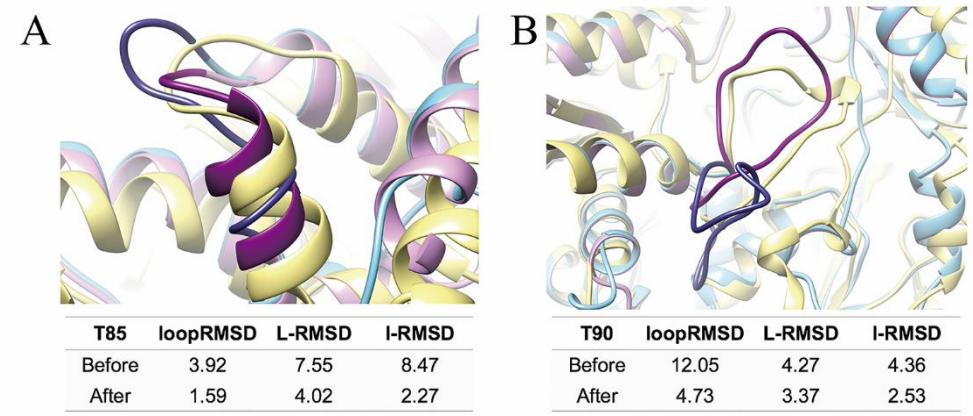


Figure 5.2. Improvement in L-RMSD and I-RMSD by ULR modeling for the 11 targets in CASP11 benchmark set for which ULR modeling was performed



T85	loopRMSD	L-RMSD	I-RMSD
Before	3.92	7.55	8.47
After	1.59	4.02	2.27

T90	loopRMSD	L-RMSD	I-RMSD
Before	12.05	4.27	4.36
After	4.73	3.37	2.53

Figure 5.3. Improvement by loop modeling for (A) T85 and (B) T90 as measured by loop RMSD, L-RMSD, and I-RMSD changes by loop modeling (in Å). Loop regions are colored in dark blue (before) and dark magenta (after), and the remaining regions are colored in sky blue (before) and pink (after). They are compared with the experimental structure, shown in yellow.

Finally, the performance of model refinement carried out by GalaxyRefineComplex at the last stage of GalaxyHomomer was analyzed. Refinement could improve models in all four accuracy measures, F_{nat}, F_{nonnat}, L-RMSD, and I-RMSD (**Figure 5.4**). Among them, native interface contacts covered by model structure, as measured by F_{nat}, shows the most improvement. This result can be understood by considering that the refinement procedure can optimize local interactions at the interface by repetitive repacking of interfacial side chains. Overall docking pose can also be adjusted during short relaxation simulations performed after each side chain repacking, leading to small but consistent improvement in L-RMSD and I-RMSD. An example case of T85 in which refinement performed after loop modeling improves model quality is illustrated in **Figure 5.5**.

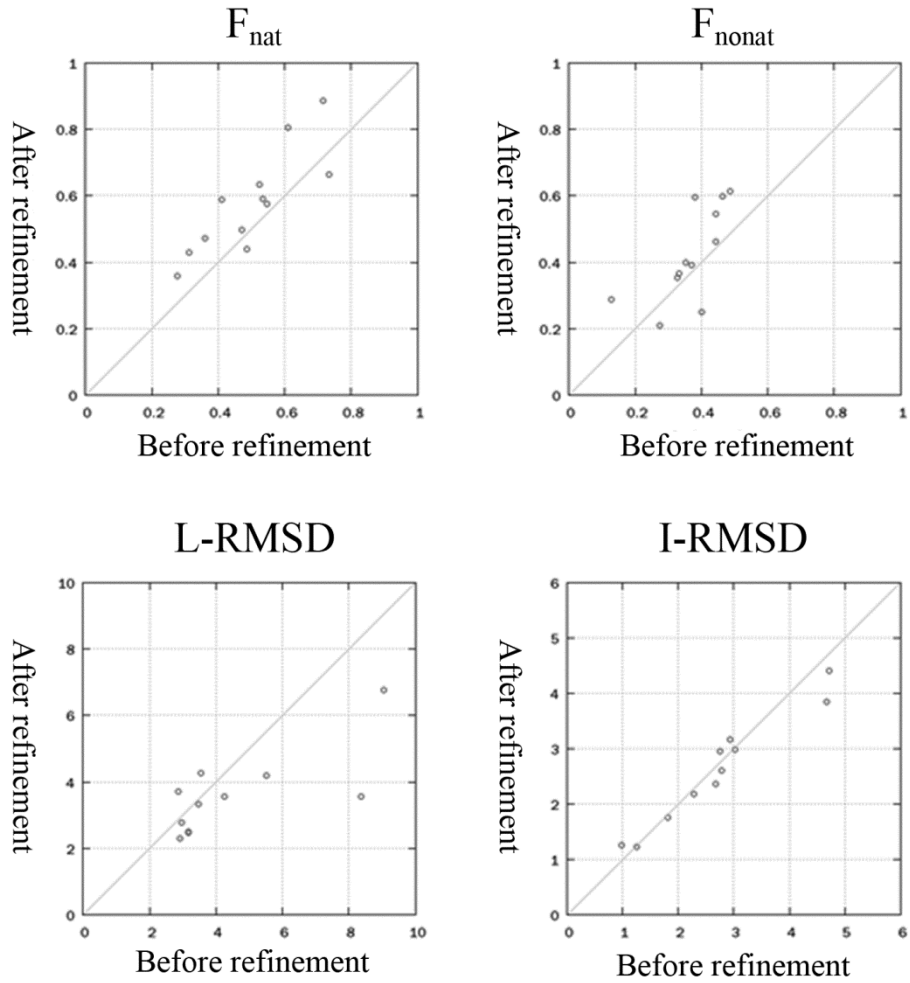
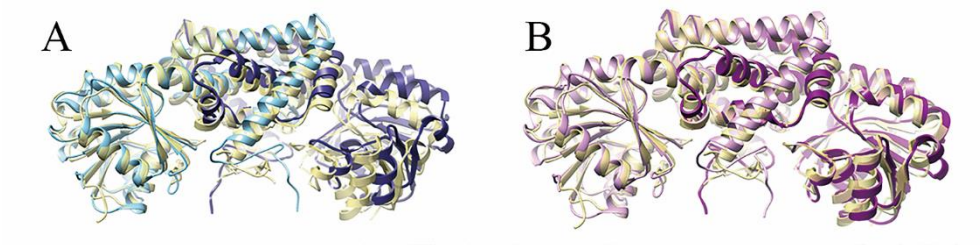


Figure 5.4. Model accuracy measured by F_{nat}, F_{nonnat}, L-RMSD, and I-RMSD before and after refinement.



T85	F_{nat}	F_{nonnat}	L-RMSD	I-RMSD
Before	0.60	0.32	4.02	2.27

T85	F_{nat}	F_{nonnat}	L-RMSD	I-RMSD
After	0.66	0.34	2.31	2.07

Figure 5.5. Improvement of model quality by refinement for T85. L-RMSD and I-RMSD are measured in Å. (A) The initial structure (sky blue for chain A and dark blue for chain B) and (B) the refined structure (pink for chain A and dark magenta for chain B) are compared to the experimental structure (yellow).

5.4. Conclusion on Homo-oligomer Structure Prediction

The GalaxyHomomer method predicts the homo-oligomer structure of a target protein from a sequence or monomer structure. It performs both template-based modeling and *ab initio* docking, and adopts additional model refinement that can consistently improve model quality. The server provides different options that can be chosen by the user depending on the availability of information on monomer structure, oligomeric state, and locations of unreliable/flexible loops or termini.

Modeling of loops/termini at the oligomer interface by GalaxyLoop improved interface RMSD, and refinement driven by side chain repacking by GalaxyRefineComplex improved the fraction of native contacts. By combining additional refinement based on loop modeling and overall structure refinement, GalaxyHomomer may generate more precise homo-oligomer models that can be useful for further applications such as for drug design targeting protein homo-oligomer interfaces.

Chapter 6. Conclusion

Proteins perform their biological functions by interacting with other molecules. Therefore, predicting protein interactions in atomic detail is very important to understand protein functions and to design drugs regulating proteins' activities. In this thesis, three computational methods for predicting ligand binding sites, protein-ligand complex structures, and protein homo-oligomer structures have been introduced.

In **Chapter 2**, a program for metal and small organic molecule binding site prediction named GalaxySite2 was described. GalaxySite2 is an extended version of GalaxySite which combines evolutionary information with protein-ligand docking technique. In GalaxySite2, a metal binding site prediction is newly added, and the selection method of putative binding ligand is improved by considering not only sequence similarity but also global and local structural similarity between target and template proteins. Because GalaxySite2 provides additional predictions on key protein–ligand interactions in terms of optimized 3D coordinates of the protein–ligand complexes, the results of GalaxySite2 would be very useful for locating cofactors before docking ligands into proteins. Moreover, based on the interactions observed in prediction results of GalaxySite2, it would be possible to get a clue to design principles of molecules targeting predicted binding sites.

Next, I developed protein–ligand docking programs which can predict protein–ligand complex structures when ligands and target binding sites are given. The two major components of a protein–ligand docking program are sampling and scoring. In **Chapter 3**, an improved scoring function, GalaxyDock BP2 Score, was introduced. By combining different types of scoring functions and balancing them

to have high discrimination power of near-native poses from non-native ones, the performance of GalaxyDock BP2 Score was improved in decoy discrimination tests. When GalaxyDock BP2 Score was applied to docking benchmark tests by implementing into GalaxyDock2 program, it showed better performance than other state-of-the-art programs in self-docking tests on Astex diverse set and Cross2009 benchmark set as well as in more realistic docking tests on Astex non-native set. This improved performance on Astex non-native set is due to the fact that GalaxyDock BP2 Score can tolerate some conformational errors in the binding pocket with additional knowledge-based and empirical score terms. Even if GalaxyDock BP2 Score showed a reasonably good performance in scoring binding affinities, a separate hybrid score must be developed in the future for the purpose of predicting binding affinity in virtual screening for more practical use.

With the improved docking scoring function, an efficient sampling algorithm was also developed as described in **Chapter 4**. The sampling algorithm was improved by utilizing fragment binding hot spot information predicted by fragment docking. By sampling translational and rotational degrees of freedom based on predicted binding hot spot information, the new sampling algorithm, GalaxyDock-Frag, could reduce the search space efficiently resulting in improvements of docking performance for large flexible ligands having more than 20 rotatable bonds. GalaxyDock-Frag was also applied to predict protein-peptide complex structures, and it showed good performance for peptides with medium length (5~7 residues). For larger peptides (8~12 residues), it shows comparable performance to other docking programs, but it has a room for improvement. In current GalaxyDock-Frag method, only translational and rotational degrees of freedom are sampled using predicted hot spot information. Therefore, only 6 dimensions of search space are reduced, and it is not sufficient to dock larger

peptides having more than 8 residues efficiently. This can be tackled in the future by developing torsion angle sampling method with predicted hot spot information.

In **Chapter 5**, the GalaxyHomomer method to predict protein homo-oligomer structure was introduced. It performs both template-based modeling and *ab initio* docking, and adopts additional model refinement which can consistently improve model quality. Modeling of loops/termini at the oligomer interface by GalaxyLoop improved interface RMSD, and refinement driven by side chain repacking by GalaxyRefineComplex improved the fraction of native contacts. By combining additional refinement based on loop modeling and overall structure refinement, GalaxyHomomer may generate more precise homo-oligomer models that can be useful for further applications such as for drug design targeting protein homo-oligomer interfaces.

The methods described in this thesis can be applied to *in silico* structure-based drug design. Druggable sites in a target protein can be detected by GalaxySite2 program, while structure-based virtual screening can be done by GalaxyDock-Frag method with GalaxyDock BP2 Score. If the target protein forms homo-oligomer, its structure and interfaces can be predicted by GalaxyHomomer program. Although all the methods have rooms for improvement in order to be applicable to various challenging problems, I hope that the issues mentioned above are resolved and the programs will be used in interesting functional and design studies in the future.

Bibliography

Abagyan R, Totrov M, Kuznetsov D (1994) ICM - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. J Comput Chem 15 (5):488-506. doi:DOI 10.1002/jcc.540150503

Ali A, Bandaranayake RM, Cai Y, King NM, Kolli M, Mittal S, Murzycki JF, Nalam MN, Nalivaika EA, Ozen A, Prabu-Jeyabalan MM, Thayer K, Schiffer CA (2010) Molecular Basis for Drug Resistance in HIV-1 Protease. Viruses 2 (11):2509-2535. doi:10.3390/v2112509

Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, Case DA, Kuntz ID, Rizzo RC (2015) DOCK 6: Impact of New Features and Current Docking Performance. J Comput Chem 36 (15):1132-1156. doi:10.1002/jcc.23905

Andre I, Strauss CE, Kaplan DB, Bradley P, Baker D (2008) Emergence of symmetry in homooligomeric biological assemblies. Proceedings of the National Academy of Sciences of the United States of America 105 (42):16148-16152. doi:10.1073/pnas.0807576105

Baek M, Park T, Heo L, Park C, Seok C (2017a) GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. Nucleic Acids Res 45 (W1):W320-W324. doi:10.1093/nar/gkx246

Baek M, Shin WH, Chung HW, Seok C (2017b) GalaxyDock BP2 score: a hybrid scoring function for accurate protein-ligand docking. J Comput Aided Mol Des 31 (7):653-666. doi:10.1007/s10822-017-0030-9

Bellmann-Sickert K, Beck-Sickinger AG (2010) Peptide drugs to target G protein-coupled receptors. Trends Pharmacol Sci 31 (9):434-441. doi:10.1016/j.tips.2010.06.003

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research 42 (Web Server issue):W252-258. doi:10.1093/nar/gku340

Bohm HJ (1998) Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. J Comput Aid Mol Des 12 (4):309-323. doi:Doi 10.1023/A:1007999920146

Brenke R, Kozakov D, Chuang GY, Beglov D, Hall D, Landon MR, Mattos C, Vajda S (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. Bioinformatics 25 (5):621-627. doi:10.1093/bioinformatics/btp036

Bron C, Kerbosch J (1973) Algorithm 457: finding all cliques of an undirected graph. Communications of the ACM 16 (9):575-577

Brylinski M, Skolnick J (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc Natl Acad Sci U S A 105 (1):129-134. doi:10.1073/pnas.0707684105

Brylinski M, Skolnick J (2011) FINDSITE-metal: integrating evolutionary information and machine learning for structure-based metal-binding site prediction at the proteome level. Proteins 79 (3):735-751. doi:10.1002/prot.22913

Bursulaya BD, Totrov M, Abagyan R, Brooks CL (2003) Comparative study of

several algorithms for flexible ligand docking. J Comput Aid Mol Des 17 (11):755-763. doi:DOI 10.1023/B:JCAM.0000017496.76572.6f

Campbell SJ, Gold ND, Jackson RM, Westhead DR (2003) Ligand binding: functional site location, similarity and docking. Curr Opin Struct Biol 13 (3):389-395

Cheng T, Li X, Li Y, Liu Z, Wang R (2009) Comparative assessment of scoring functions on a diverse test set. J Chem Inf Model 49 (4):1079-1093. doi:10.1021/ci9000053

Chopra G, Kalisman N, Levitt M (2010) Consistent refinement of submitted models at CASP using a knowledge-based potential. Proteins-Structure Function and Bioinformatics 78 (12):2668-2678. doi:10.1002/prot.22781

Clarke OB, Gulbis JM (2012) Oligomerization at the membrane: potassium channel structure and function. Advances in experimental medicine and biology 747:122-136. doi:10.1007/978-1-4614-3229-6_8

Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics (Oxford, England) 20 (1):45-50

Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu YB, Humblet C (2009) Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. J Chem Inf Model 49 (6):1455-1474. doi:10.1021/ci900056c

Damborsky J, Brezovsky J (2014) Computational tools for designing and engineering enzymes. Curr Opin Chem Biol 19:8-16. doi:10.1016/j.cbpa.2013.12.003

DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I (2011) Modeling symmetric macromolecular structures in Rosetta3. PloS one 6 (6):e20450. doi:10.1371/journal.pone.0020450

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society 125 (7):1731-1737. doi:10.1021/ja026939x

Dunbrack RL, Jr. (2002) Rotamer libraries in the 21st century. Curr Opin Struct Biol 12 (4):431-440

Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aid Mol Des 11 (5):425-445. doi:Doi 10.1023/A:1007996124545

Ester M, Kriegel H-P, Sander J, Xu X A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd, 1996. vol 34. pp 226-231

Feldmeier K, Hocker B (2013) Computational protein design of ligand binding and catalysis. Curr Opin Chem Biol 17 (6):929-933

Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL (2004) Assessing scoring functions for protein-ligand interactions. Journal of medicinal chemistry 47 (12):3032-3047. doi:10.1021/jm030489h

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of medicinal chemistry 47 (7):1739-1749. doi:10.1021/jm0306430

Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. Journal of medicinal chemistry 49 (21):6177-6196. doi:10.1021/jm051256o

Gallo Cassarino T, Bordoli L, Schwede T (2014) Assessment of ligand binding site predictions in CASP10. Proteins 82 Suppl 2:154-163. doi:10.1002/prot.24495

Gasteiger J, Marsili M (1980) Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. Tetrahedron 36 (22):3219-3228. doi:Doi 10.1016/0040-4020(80)80168-2

Gaudreault F, Najmanovich RJ (2015) FlexAID: Revisiting Docking on Non-Native-Complex Structures. J Chem Inf Model 55 (7):1323-1336. doi:10.1021/acs.jcim.5b00078

Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST (1995) Molecular Recognition of the Inhibitor Ag-1343 by Hiv-1 Protease - Conformationally Flexible Docking by Evolutionary Programming. Chem Biol 2 (5):317-324. doi:Doi 10.1016/1074-5521(95)90050-0

Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295 (2):337-356. doi:10.1006/jmbi.1999.3371

Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. Annual review of biophysics and biomolecular structure 29:105-153. doi:10.1146/annurev.biophys.29.1.105

Goto J, Kataoka R, Hirayama N (2004) Ph4Dock: pharmacophore-based protein-ligand docking. J Med Chem 47 (27):6804-6811. doi:10.1021/jm0493818

Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. Journal of molecular biology 331 (1):281-299

Haas J, Roth S, Arnold K, Kiefer F, Schmidt T, Bordoli L, Schwede T (2013) The Protein Model Portal--a comprehensive resource for protein structure and model information. Database : the journal of biological databases and curation 2013:bat031. doi:10.1093/database/bat031

Haberthur U, Caflisch A (2008) FACTS: Fast analytical continuum treatment of solvation. J Comput Chem 29 (5):701-715. doi:10.1002/jcc.20832

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. Journal of medicinal chemistry 47 (7):1750-1759. doi:10.1021/jm030644s

Hartigan JA, Wong MA (1979) Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society Series C (Applied Statistics) 28 (1):100-108

Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WT, Mortenson PN, Murray CW (2007) Diverse, high-quality test set for the validation of protein-ligand docking performance. Journal of medicinal chemistry 50 (4):726-741. doi:10.1021/jm061277y

Hauser AS, Windshugel B (2016) LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. J Chem Inf Model 56 (1):188-200. doi:10.1021/acs.jcim.5b00234

Heldin CH (1995) Dimerization of cell surface receptors in signal transduction.

Cell 80 (2):213-223

Hendlich M, Bergner A, Gunther J, Klebe G (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. J Mol Biol 326 (2):607-620

Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model 15 (6):359-363, 389

Heo L, Lee H, Seok C (2016) GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. Sci Rep 6:32153. doi:10.1038/srep32153

Heo L, Shin WH, Lee MS, Seok C (2014) GalaxySite: ligand-binding-site prediction by using molecular docking. Nucleic Acids Res 42 (Web Server issue):W210-214. doi:10.1093/nar/gku321

Hu B, Lill MA (2014) PharmDock: a pharmacophore-based docking program. J Cheminform 6:14. doi:10.1186/1758-2946-6-14

Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. Journal of medicinal chemistry 49 (23):6789-6801. doi:10.1021/jm0608356

Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. J Comput Chem 28 (6):1145-1152. doi:10.1002/jcc.20634

Jain AN (2007) Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J Comput Aid Mol Des 21 (5):281-306. doi:10.1007/s10822-007-9114-2

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267 (3):727-748. doi:DOI 10.1006/jmbi.1996.0897

Kazimipour B, Li X, Qin AK Initialization methods for large scale global optimization. In: Evolutionary Computation (CEC), 2013 IEEE Congress on, 2013. IEEE, pp 2750-2757

Kelley LA, Gardner SP, Sutcliffe MJ (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 9 (11):1063-1065

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic acids research 32 (Web Server issue):W526-531. doi:10.1093/nar/gkh468

Kinoshita K, Nakamura H (2003) Protein informatics towards function identification. Curr Opin Struct Biol 13 (3):396-400

Ko J, Park H, Seok C (2012) GalaxyTBM: template-based modeling by building a reliable core and refining unreliable local regions. BMC Bioinformatics 13:198. doi:10.1186/1471-2105-13-198

Korb O, Stutzle T, Exner TE (2006) PLANTS: Application of ant colony optimization to structure-based drug design. Lect Notes Comput Sc 4150:247-258

Korb O, Stutzle T, Exner TE (2009) Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. J Chem Inf Model 49 (1):84-96. doi:10.1021/ci800298z

Korb O, Ten Brink T, Victor Paul Raj FR, Keil M, Exner TE (2012) Are predefined

decoy sets of ligand poses able to quantify scoring function accuracy? J Comput Aided Mol Des 26 (2):185-197. doi:10.1007/s10822-011-9539-5

Kortemme T, Morozov AV, Baker D (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J Mol Biol 326 (4):1239-1259

Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 37 (2):228-241

Kristiansen K (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. Pharmacol Ther 103 (1):21-80. doi:10.1016/j.pharmthera.2004.05.002

Lafita A, Bliven S, Kryshtafovych A, Bertoni M, Monastyrskyy B, Duarte JM, Schwede T, Capitani G (2018) Assessment of protein assembly prediction in CASP12. Proteins-Structure Function and Bioinformatics 86:247-256. doi:10.1002/prot.25408

Lau JL, Dunn MK (2018) Therapeutic peptides: Historical perspectives, current development trends, and future directions. Bioorgan Med Chem 26 (10):2700-2707. doi:10.1016/j.bmc.2017.06.052

Laurie AT, Jackson RM (2006) Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Curr Protein Pept Sci 7 (5):395-406

Lee H, Baek M, Lee GR, Park S, Seok C (2016) Template-based modeling and ab initio refinement of protein oligomer structures using GALAXY in CAPRI round 30. Proteins. doi:10.1002/prot.25192

Lee H, Park H, Ko J, Seok C (2013) GalaxyGemini: a web server for protein homo-oligomer structure prediction based on similarity. Bioinformatics (Oxford, England) 29 (8):1078-1080. doi:10.1093/bioinformatics/btt079

Lee J, Lee D, Park H, Coutsias EA, Seok C (2010) Protein loop modeling by using fragment assembly and analytical loop closure. Proteins 78 (16):3428-3436. doi:10.1002/prot.22849

Lee J, Scheraga HA, Rackovsky S (1997) New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. J Comput Chem 18 (9):1222-1232. doi:Doi 10.1002/(Sici)1096-987x(19970715)18:9<1222::Aid-Jcc10>3.0.Co;2-7

Lee K, Czaplewski C, Kim SY, Lee J (2005) An efficient molecular docking using conformational space annealing. J Comput Chem 26 (1):78-87. doi:10.1002/jcc.20147

Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ (2018) The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. Proteins 86 Suppl 1:257-273. doi:10.1002/prot.25419

Lensink MF, Velankar S, Kryshtafovych A, Huang SY, Schneidman-Duhovny D, Sali A, Segura J, Fernandez-Fuentes N, Viswanath S, Elber R, Grudinin S, Popov P, Neveu E, Lee H, Baek M, Park S, Heo L, Rie Lee G, Seok C, Qin S, Zhou HX, Ritchie DW, Maigret B, Devignes MD, Ghoorah A, Torchala M, Chaleil RA, Bates PA, Ben-Zeev E, Eisenstein M, Negi SS, Weng Z, Vreven T, Pierce BG, Borrman TM, Yu J, Ochsenbein F, Guerois R, Vangone A, Rodrigues JP, van Zundert G, Nellen M, Xue L, Karaca E, Melquiond AS, Visscher K, Kastritis PL, Bonvin AM, Xu X, Qiu L, Yan C, Li J, Ma Z, Cheng J, Zou X, Shen Y, Peterson LX, Kim HR,

Roy A, Han X, Esquivel-Rodriguez J, Kihara D, Yu X, Bruce NJ, Fuller JC, Wade RC, Anishchenko I, Kundrotas PJ, Vakser IA, Imai K, Yamada K, Oda T, Nakamura T, Tomii K, Pallara C, Romero-Durana M, Jimenez-Garcia B, Moal IH, Fernandez-Recio J, Joung JY, Kim JY, Joo K, Lee J, Kozakov D, Vajda S, Mottarella S, Hall DR, Beglov D, Mamonov A, Xia B, Bohnuud T, Del Carpio CA, Ichiishi E, Marze N, Kuroda D, Roy Burman SS, Gray JJ, Chermak E, Cavallo L, Oliva R, Tovchigrechko A, Wodak SJ (2016) Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. Proteins 84 Suppl 1:323-348. doi:10.1002/prot.25007

Lensink MF, Wodak SJ (2010) Docking and scoring protein interactions: CAPRI 2009. Proteins 78 (15):3073-3084. doi:10.1002/prot.22818

Li Y, Han L, Liu Z, Wang R (2014a) Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. J Chem Inf Model 54 (6):1717-1736. doi:10.1021/ci500081m

Li Y, Liu Z, Li J, Han L, Liu J, Zhao Z, Wang R (2014b) Comparative assessment of scoring functions on an updated benchmark: 1. Compilation of the test set. J Chem Inf Model 54 (6):1700-1716. doi:10.1021/ci500080q

Lopez G, Ezkurdia I, Tress ML (2009) Assessment of ligand binding residue predictions in CASP8. Proteins 77 Suppl 9:138-146. doi:10.1002/prot.22557

Lopez G, Rojas A, Tress M, Valencia A (2007) Assessment of predictions submitted for the CASP7 function prediction category. Proteins 69 Suppl 8:165-174. doi:10.1002/prot.21651

Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW (2010) HexServer: an FFT-based protein docking server powered by graphics processors.

Nucleic acids research 38 (Web Server issue):W445-449. doi:10.1093/nar/gkq311

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102 (18):3586-3616. doi:10.1021/jp973084f

Mandal PK, Gao F, Lu Z, Ren Z, Ramesh R, Birtwistle JS, Kaluarachchi KK, Chen X, Bast RC, Jr., Liao WS, McMurray JS (2011) Potent and selective phosphopeptide mimetic prodrugs targeted to the Src homology 2 (SH2) domain of signal transducer and activator of transcription 3. J Med Chem 54 (10):3549-3563. doi:10.1021/jm2000882

McGann M (2011) FRED Pose Prediction and Virtual Screening Accuracy. J Chem Inf Model 51 (3):578-596. doi:10.1021/ci100436p

McMurray JS (2008) Structural basis for the binding of high affinity phosphopeptides to Stat3. Biopolymers 90 (1):69-79. doi:10.1002/bip.20901

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19 (14):1639-1662. doi:Doi 10.1002/(Sici)1096-987x(19981115)19:14<1639::Aid-Jcc10>3.0.Co;2-B

Negri A, Rodriguez-Larrea D, Marco E, Jimenez-Ruiz A, Sanchez-Ruiz JM, Gago F (2010) Protein-protein interactions at an enzyme-substrate interface: characterization of transient reaction intermediates throughout a full catalytic cycle

of Escherichia coli thioredoxin reductase. Proteins 78 (1):36-51. doi:10.1002/prot.22490

Nelder JA, Mead R (1965) A simplex method for function minimization. The computer journal 7 (4):308-313

Neudert G, Klebe G (2011) DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. J Chem Inf Model 51 (10):2731-2745. doi:10.1021/ci200274q

Nicholls A (2008) What do we know and when do we know it? J Comput Aid Mol Des 22 (3-4):239-255. doi:10.1007/s10822-008-9170-2

Park H, Ko J, Joo K, Lee J, Seok C, Lee J (2011) Refinement of protein termini in template-based modeling using conformational space annealing. Proteins 79 (9):2725-2734. doi:10.1002/prot.23101

Park H, Lee GR, Heo L, Seok C (2014a) Protein loop modeling using a new hybrid energy function and its application to modeling in inaccurate structural environments. PLoS One 9 (11):e113811. doi:10.1371/journal.pone.0113811

Park H, Lee GR, Heo L, Seok C (2014b) Protein Loop Modeling Using a New Hybrid Energy Function and Its Application to Modeling in Inaccurate Structural Environments. Plos One 9 (11). doi:ARTN e113811

10.1371/journal.pone.0113811

Park H, Seok C (2012a) Refinement of unreliable local regions in template-based protein models. Proteins 80 (8):1974-1986. doi:10.1002/prot.24086

Park H, Seok C (2012b) Refinement of unreliable local regions in template-based protein models. Proteins-Structure Function and Bioinformatics 80 (8):1974-1986.

doi:10.1002/prot.24086

Pawson T, Nash P (2000) Protein-protein interactions define specificity in signal transduction. Genes Dev 14 (9):1027-1047

Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins-Structure Function and Bioinformatics 56 (2):235-249. doi:10.1002/prot.20088

Pidugu LSM, Mbimba JCE, Ahmad M, Pozharski E, Sausville EA, Emadi A, Toth EA (2016) A direct interaction between NQO1 and a chemotherapeutic dimeric naphthoquinone. Bmc Struct Biol 16. doi:ARTN 1

10.1186/s12900-016-0052-x

Pierce B, Tong W, Weng Z (2005) M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. Bioinformatics 21 (8):1472-1478. doi:10.1093/bioinformatics/bti229

Plewczynski D, Lazniewski M, Augustyniak R, Ginalski K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem 32 (4):742-755. doi:10.1002/jcc.21643

Ponstingl H, Kabir T, Thornton JM (2003) Automatic inference of protein quaternary structure from crystals. J Appl Crystallogr 36:1116-1122. doi:10.1107/S0021889803012421

Poupon A, Janin J (2010) Analysis and prediction of protein quaternary structure. Methods in molecular biology (Clifton, NJ) 609:349-364. doi:10.1007/978-1-60327-241-4_20

Roche DB, Buenavista MT, McGuffin LJ (2012) FunFOLDQA: a quality

assessment tool for protein-ligand binding site residue predictions. PLoS One 7 (5):e38219. doi:10.1371/journal.pone.0038219

Roche DB, Tetchner SJ, McGuffin LJ (2011) FunFOLD: an improved automated method for the prediction of ligand binding residues using 3D models of proteins. BMC Bioinformatics 12:160. doi:10.1186/1471-2105-12-160

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234 (3):779-815. doi:10.1006/jmbi.1993.1626

Schmidt T, Haas J, Gallo Cassarino T, Schwede T (2011) Assessment of ligand-binding residue predictions in CASP9. Proteins 79 Suppl 10:126-136. doi:10.1002/prot.23174

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic acids research 33 (Web Server issue):W363-367. doi:10.1093/nar/gki481

Sheng C, Zhang W (2013) Fragment informatics and computational fragment-based drug design: an overview and update. Med Res Rev 33 (3):554-598. doi:10.1002/med.21255

Shin WH, Heo L, Lee J, Ko J, Seok C, Lee J (2011) LigDockCSA: protein-ligand docking using conformational space annealing. J Comput Chem 32 (15):3226-3232. doi:10.1002/jcc.21905

Shin WH, Kim JK, Kim DS, Seok C (2013) GalaxyDock2: protein-ligand docking using beta-complex and global optimization. J Comput Chem 34 (30):2647-2656. doi:10.1002/jcc.23438

Shin WH, Seok C (2012) GalaxyDock: protein-ligand docking with flexible protein

side-chains. J Chem Inf Model 52 (12):3225-3232. doi:10.1021/ci300342z

Snijder HJ, Ubarretxena-Belandia I, Blaauw M, Kalk KH, Verheij HM, Egmond MR, Dekker N, Dijkstra BW (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. Nature 401 (6754):717-721. doi:10.1038/44890

Soding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21 (7):951-960. doi:10.1093/bioinformatics/bti125

Sotriffer C, Klebe G (2002) Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. Farmaco 57 (3):243-251

Spitzer R, Jain AN (2012) Surflex-Dock: Docking benchmarks and real-world application. J Comput Aid Mol Des 26 (6):687-699. doi:10.1007/s10822-011-9533-y

Stock J (1996) Receptor signaling: dimerization and beyond. Current biology : CB 6 (7):825-827

Torchala M, Moal IH, Chaleil RA, Fernandez-Recio J, Bates PA (2013) SwarmDock: a server for flexible protein-protein docking. Bioinformatics (Oxford, England) 29 (6):807-809. doi:10.1093/bioinformatics/btt038

Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein-protein docking. Nucleic acids research 34 (Web Server issue):W310-314. doi:10.1093/nar/gkl206

Tripathi A, Kellogg GE (2010) A novel and efficient tool for locating and characterizing protein cavities and binding sites. Proteins 78 (4):825-842.

doi:10.1002/prot.22608

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31 (2):455-461. doi:10.1002/jcc.21334

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52 (4):609-623. doi:10.1002/prot.10465

Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW (2008) Protein-ligand docking against non-native protein conformers. J Chem Inf Model 48 (11):2214-2225. doi:10.1021/ci8002254

Vilar S, Cozza G, Moro S (2008) Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. Curr Top Med Chem 8 (18):1555-1572. doi:Doi 10.2174/156802608786786624

Vlieghe P, Lisowski V, Martinez J, Khrestchatisky M (2010) Synthetic therapeutic peptides: science and market. Drug Discov Today 15 (1-2):40-56. doi:10.1016/j.drudis.2009.10.009

Wang RX, Gao Y, Lai LH (2000) Calculating partition coefficient by atom-additive method. Perspect Drug Discov 19 (1):47-66. doi:Doi 10.1023/A:1008763405023

Wang RX, Lai LH, Wang SM (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aid Mol Des 16 (1):11-26. doi:Doi 10.1023/A:1016357811882

Wang RX, Lu YP, Wang SM (2003) Comparative evaluation of 11 scoring

functions for molecular docking. Journal of medicinal chemistry 46 (12):2287-2303. doi:10.1021/jm0203783

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. Journal of medicinal chemistry 49 (20):5912-5931. doi:10.1021/jm050362n

Yang J, Roy A, Zhang Y (2013a) BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res 41 (Database issue):D1096-1103. doi:10.1093/nar/gks966

Yang J, Roy A, Zhang Y (2013b) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29 (20):2588-2595. doi:10.1093/bioinformatics/btt447

Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 72 (2):793-803. doi:10.1002/prot.21968

Zhang C, Liu S, Zhou YQ (2004) Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. Protein Sci 13 (2):391-399. doi:10.1110/ps.03411904

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33 (7):2302-2309. doi:10.1093/nar/gki524

Zhou ZY, Felts AK, Friesner RA, Levy RM (2007) Comparative performance of several flexible docking programs and scoring functions: Enrichment studies for a

diverse set of pharmaceutically relevant targets. J Chem Inf Model 47 (4):1599-1608. doi:10.1021/ci7000346

Zhu J, Xie L, Honig B (2006) Structural refinement of protein segments containing secondary structure elements: Local sampling, knowledge-based potentials, and clustering. Proteins-Structure Function and Bioinformatics 65 (2):463-479. doi:10.1002/prot.21085

국문초록

단백질은 생체 내의 중요한 구성 요소로 다양한 생물학적 반응에 관여한다. 단백질은 금속 이온부터 유기 분자, 펩타이드, 지질, 핵산, 단백질까지 다양한 분자와 상호작용하며 그 기능을 수행한다. 따라서 단백질과 다른 분자 사이의 상호작용을 예측하는 계산 방법들은 단백질의 기능을 분자 수준에서 이해하고 단백질의 기능을 제어하는 신약 물질을 개발하는데 유용하게 쓰일 수 있다. 특히 단백질에 존재하는 리간드 결합 자리를 예측하는 방법은 신약개발에서 표적으로 하는 단백질의 약물 결합 자리를 찾아내는데 활용될 수 있으며, 단백질-리간드 도킹 방법은 신약 개발 단계에서 신약 후보 물질을 찾고 이를 최적화하는데 기여할 수 있다. 또한 대다수의 단백질들이 그 생물학적 기능을 수행하기 위해 호모 올리고머 구조를 이루기 때문에, 이러한 단백질의 호모 올리고머 구조를 예측하는 방법 역시 올리고머 결합 자리를 표적으로 하는 신약 개발에 큰 도움을 줄 수 있다.

본 논문은 단백질의 상호작용을 예측하기 위한 세가지 계산 방법의 개발 내용과 그 결과를 담고 있다. 첫 번째 방법은 단백질에 존재하는 금속 이온 및 유기 분자 결합 자리를 예측하는 방법으로 곁가지 구조의 유연성을 고려한 금속 이온 도킹 방법과 새로운 주형 구조 선택 방법을 도입함으로써 그 성능을 향상시켰다. 두 번째 방법은 단백질과 리간드의 결합 구조를 예측하는 단백질-리간드 도킹 방법으로, 물리 기반 에너지와 통계 기반 에너지를 혼합한 형태의 향상된 평가 함수와 리간드 조각 구조의 결합 정보를 예측하여 활용하는 새로운 구조

샘플링 방법을 도입하였다. 마지막으로, 생물정보학적 접근 방법과 물리화학적 접근 방법을 혼합하여 단백질의 호모 올리고머 구조를 예측하기 위한 방법을 개발하였다. 본 논문에서 기술하고 있는 세 방법 모두 다른 최신 프로그램들과 비교하였을 때 좋은 성능을 보여주었다. 이 방법들을 활용한다면 단백질의 기능을 이해하는 것뿐만 아니라 컴퓨터 계산을 활용한 신약 물질 개발에도 큰 도움을 줄 수 있을 것으로 기대된다.

주요어: 단백질 상호작용 예측, 리간드 결합 자리 예측, 단백질-리간드 도킹, 도킹 평가 함수, 단백질 호모 올리고머 구조 예측

학 번: 2013-20267