Stony Brook University

Computational Study of protein-Ligand Interaction with Application in Drug Discovery

A Thesis Presented

by

Yuanqing Guo

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Chemistry

Stony Brook University

December 2007

Stony Brook University

The Graduate School

Yuanqing Guo

We, the thesis committee for the above candidate for the

Master of Science degree,

Hereby recommend the acceptance of this thesis.

Jin Wang, Ph. D., Advisor
Department of Chemistry, Stony Brook University

George Stell, Ph. D., Chairperson
Department of Chemistry, Stony Brook University

Dale G. Drueckhammer, Ph. D., Third Member
Department of Chemistry, Stony Brook University

This thesis is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Thesis

Computational Study of protein-Ligand Interaction

with Application in Drug Discovery

by

Yuanqing Guo

Master of Science

In

Chemistry

Stony Brook University

2007

Protein-ligand binding is a critical issue in uncovering fundamental mechanism of biological systems and in developing lead compounds to new drugs in practice. It has attracted much interest from scientists within the last decade. There are two crucial factors in determining a binding process. One is affinity and the other is specificity. In previous studies, only affinity has been fully taken into account by specialists in this field. However, high affinity cannot always guarantee high specificity. Therefore, it is necessary to apply the other equivalent critical factor, specificity, together with affinity to make a new two dimensional drug screening criterion. In this study, first of all, a new measurement of specificity, intrinsic specificity ratio (ISR), is deduced from energy landscape theory. Secondly, a series experiments to a biological system, FKBP12, are carried out to quantify ISR as a potential complement to affinity in the drug screening. In addition, a few new drug candidates for SmpB were predicted by applying this new two dimensional drug screening criterion.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Protein –Ligand Binding

Protein-ligand binding is a process to find the "best" fit between proteins and ligands. It is controlled by noncovalent bonds and can be completely understood by considering thermodynamic, kinetic and structural aspects [2]. Scientists have invested tremendous efforts and made great progresses in this field due to its potential applications in pharmaceutical industry. With the advances in development of computational chemistry [3], it is feasible to understand and predict protein-ligand binding by theoretical methods.

In computational scheme binding problems, also referred to as docking problems, apply structure-based strategies based on assumption that significant structural and chemical complementarity must be present between ligands and their target receptors [4]. Structure-based approach falls into two categories: receptor-based and ligand-based. Compared to sufficient structural information of small ligands, there is deficient knowledge about large receptors. Therefore, ligand-based approach is always adopted.

There are two current challenges in docking problem, which include efficient search algorithm and accurate scoring function [5]. In the search algorithm, searching space, flexibility and binding site are three issues to be addressed. Regarding to search space there also exists Levinthal paradox [6] in docking problem: docking process occurs in a biological time scale while searching all possible conformational space costs a cosmological time [1]. Therefore, a robust algorithm is critical for docking problem, which searches only certain space rather than the entire possible states and can find native structure in an affordable computational time. Native structure is defined as the structure of a molecule in its X-ray crystallography structure or NMR structure.

Because two docking partners can be able to change their conformations upon docking, it can definitely enhance the chance to find native structure of a complex by including conformational flexibility of both ligand and receptor. However, it is computationally expensive and therefore it becomes an important issue to be addressed. There are three levels of complication regarding to the extent of flexibility of docking partners included into the algorithm [5]:

(1) Rigid docking--- both ligand and receptor are treated as rigid bodies;

(2) Semi-flexible docking--- only small ligand is regarded as flexible while relative large receptor is rigid;

(3) Flexible docking--- both ligand and receptor are considered flexible. However, there are still some limitations such as only side chains are set flexible while backbone is rigid.

The third issue in search algorithm is the location of binding site. Based on availability of binding sites, there are two kinds of docking jobs: "bound" docking and "unbound" docking". In "bound" docking, the ligand is extracted from its co-crystal

structure of the receptor and the ligand, and reconstruction of the complex is followed. Nonetheless, “unbound” docking, also referred to as “blind” docking, is more challenging and more time-consuming without any information about binding sites [7]. Several possible predictions about the location of binding sites must be made first and a series of computational comparison based on sophisticated algorithm are followed to select most probable binding site.

Another bottleneck in docking problem is scoring function, also referred to as energy function or potential function, which can be used to identify the “true” docking mode of a given ligand and also to be able to rank one ligand relative to another. However, in most time scoring function cannot work so effectively. And hence it is inevitable to optimize the scoring function and make it more reliable to predict the performance of docking job. One of approaches to solve the above-mentioned problem is to involve more descriptors and weigh them properly. The following descriptors are always used in scoring function: geometric complementarity, intramolecular and intermolecular overlap, hydrogen bonds, contact area, pairwise amino acid and atom-atom contacts, electrostatic interactions, solvation energy and binding site information [5]. Another popular approach is consensus scoring [8], which allows combining different docking programs with a variety of scoring functions [9].

1.2 Intrinsic Speficity Ratio (ISR)

In the stage of drug screening of virtual ligand database in pharmaceutical industry, there are two equivalently crucial factors underlying a docking process. One is affinity which measures the stability of associating two molecules together. The other is specificity which measures the ability to distinguish among “good” native binding state from “bad” non-native states even if they have roughly the same affinity. In previous studies, only affinity has been fully taken into account by specialists in this field. However, high affinity cannot always guarantee high specificity. Some of drugs produced by the criterion of only affinity can cause serious side effects to patients. Therefore, it is urgent to apply the other specificity together with affinity to make a new two dimensional drug screening criterion.

Affinity is measured by the free energy difference (ΔG) in the process of association of two docking partners; while conventionally specificity is measured as the affinity difference ($\Delta\Delta G$) between two different receptors bound with the same ligand. However, the affinity-based measurement for specificity is not suitable to be a criterion for determining the docking process. First of all, binding affinity estimates are strongly restricted by limitations of computational approaches for docking [10]. The conventional definition of specificity is derived from affinity and therefore is not accurate either. Secondly, even the affinity is accurate, it still cannot always guarantee specificity. Because many different binding modes can have the same binding affinity and only one or none of them can distinguish or discriminate different protein receptors. In addition,

from practical point of view, it is not easy to find receptor's competitors with similar binding site to be used as a specificity marker. Furthermore, much great binding affinity, which leads to very tight association, does not favor signal transduction, gene expression and other cellular activities. Therefore, the conventional measurement of specificity is not proper and a new and precise measurement of specificity is on demand.

To resolve the above problem, intrinsic specific ratio (ISR) is proposed to be a quantitative measurement for specificity or selectivity based on energy landscape theory. Binding can be viewed as a similar process of folding and they represent intermolecular and intramolecular recognition respectively. The presence or absence of the chain connectivity is the main difference between them [6]. Previous studies indicates that hydrophobic interactions are the key driving forces during these processes while electrostatic forces are long-range and play an important role in guiding and steering docking partners together. Similar to folding, the resulting binding energy landscape is also funnel-shaped which is crucial in successful structure prediction (Figure1.1) [1].

In principle, three thermodynamic phases exist in the process of binding:

(1) native state--- it is unique and corresponds to global minimum
(2) non-native states--- they are non-native binding modes
(3) glass or trapping phase--- it corresponds to local minimum with multiple meta-stable states

The definition of intrinsic specificity ration (ISR) is the ratio of the energy gap (δE) between the native state and the average of non-native state versus the dispersion or variance of the non-native states (ΔE). And hence, ISR =δE/ΔE and maximized ISR becomes a novel criterion for the specificity of binding [1] process in order to optimize the binding process. Furthermore, based on the Bolzman weight $P\sim e^{-\beta E}$, large ISR prefers the thermodynamic stable binding modes to be dominant in population and distincts or discriminates from the rest of the binding modes. Therefore, the specificity is realized by maximized ISR. The new measurement of specificity is based on structural matching or fit compared with conventional measurement of specificity based on affinity.

1.3 AutoDock Software

AutoDock 3.0.5 is the latest version of AutoDock and have significant strengths compared with other docking packages in the following aspects:

(1) AutoDock takes into account the factors of Van der Waal forces, hydrogen bonding, electrostatic, desolvation and torsions in its scoring function as follows:

$\Delta G_{binding} = \Delta G_{vdw} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$

The improved scoring function is believed to be able to discriminate positive and false positive solutions.

(2) AutoDock 3.0.5 is semi-rigid docking which can automatically dock flexible ligand to rigid protein in contrast to other docking software which only deals with rigid docking. And hence it significantly increases the accuracy of the docking job and

hopefully can handle flexible docking in its higher version---AutoDock 4.0.

Degrees of Freedom (DOF) are defined as the sum of degrees of position or translation, degrees of orientation or quaternion and degrees of rotatable bonds or torsions. AutoDock 3.0.5 can deal with flexible ligands with larger number of DOF.

(3) AutoDock uses grid maps to sample force field for macromolecules. The entire binding site is covered by grid maps and each ligand atom type has each corresponding map by putting a probe atom into the grid map. By the means of this, computational time can be saved more than 100 times.

(4) AutoDock 3.0.5 has a novel algorithm for sampling. In any docking scheme, a tradeoff must exist between accurate prediction and reasonable computational time. There is no exception for AutoDock that has to adopt stochastic search with much iteration to enhance the chance to success rather than systematic and exhaustive search of all possible configuration states. There are quite a lot algorithm can be used such as Monte Carlo Simulated Annealing (SA), genetic algorithm including genetic algorithm (GA) and evolutionary programming (EP) or local search Solis and Wets (SW) and pseudo-Solis and Wets (pSW) to search configuration states for flexible ligands.

In AutoDock 3.0.5, a new hybrid global –local algorithm, referred to as Lamarckian Genetic Algorithm (LGA), is included. LGA is a hybrid of GA and SW, and possesses the advantages of both global search algorithm and local algorithm. GA starts with a random population and performs 2- point crossover (two parents give two children) and random mutation (each individual gives one mutant child). It automatically decides the number of individuals in a population which can be survived into next generation. It also works as a series of cycles and stops when total energy evaluations or maximum generations reached [11]. SW and pSW belong to local search algorithm and the main difference is fixed or flexible variances they use.

AutoDock also has its weakness and is not suitable when no 3 dimensional structures available, too flexible protein or ligands with too many atoms, types and torsions. However, basically AutoDock is one of the best docking tools for study of protein-ligand docking so far.

1.4 Tables and Figures

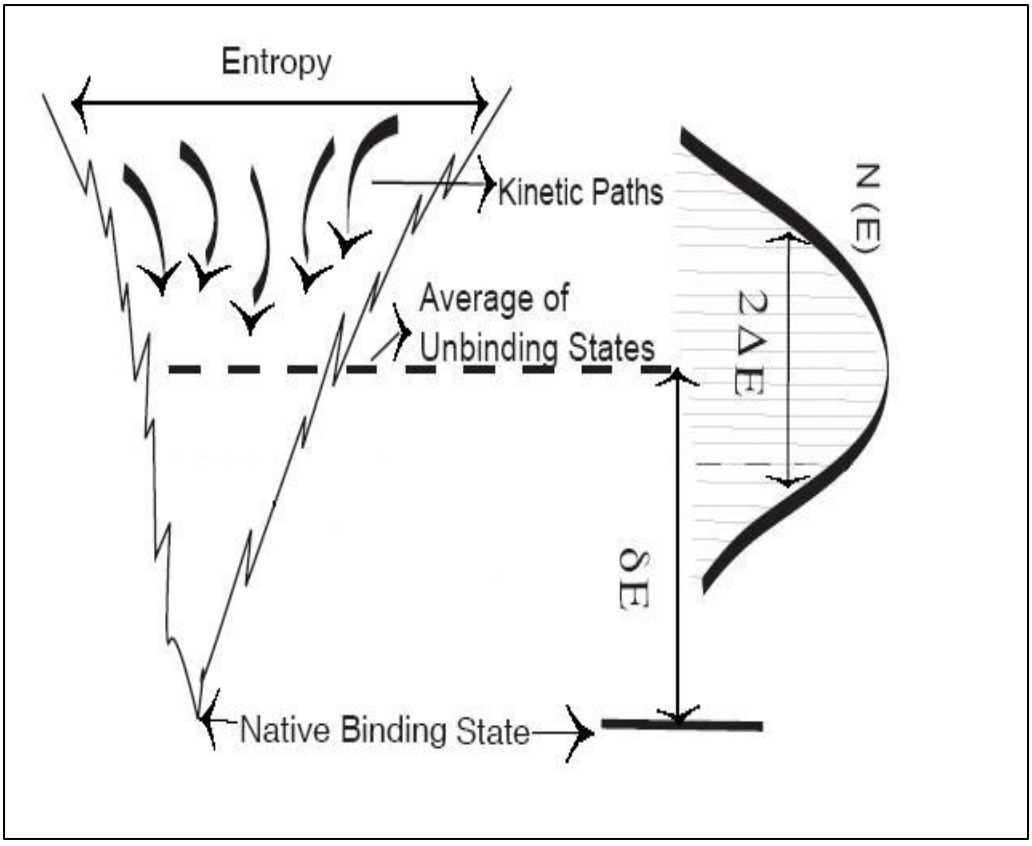


Figure 1.1 Funneled Energy Landscape of Biomolecular Binding

Chapter 2: Qualification of ISR in Drug Discovery

2.1 Introduction of FKBP12

Natural products---FK506 [12] [13] and rapamycin [14] [15] [16] are potent immunosuppressive drugs, which can dampen T cell activation and growth by interfering with signal transduction pathways required by T cell. FK506, rapamycin and their analogs have provided a promising future prospect for the treatment of stroke sequelae, central and peripheral nerve injury and neutron degenerative disorders such as Alzheimer's disease and Parkinson's disease [17].

Previous studies showed that FK506 and rapamycin play an important role as "molecular glue inhibitors" to simultaneously bind to two different proteins. The portion of FK506 which interacts with FKBP12 [18] (FK506 binding protein) is defined as" binding domain" and the other portion is defined as "effector domain". The"binding domain" of rapamycin is similar as FKBP12 but its "effector domain" is different. FK506 and rapamycin function by forming complexes with FKBP12 and other proteins simultaneously. Both of FK506 and rapamycin can bind to FKBP12 while FK506–FKBP12 complex and rapamycin–FKBP12 complex can bind to different proteins. The FKBP12-FK506 complex interacts with calcineurin (CaN), which inhibits the production of the cytokine interleukin (IL)-2. The FKBP12-rapamycin complex specially binds and inhibits mTOR (mammalian target of rapamycin), which inhibits the cell's response to IL-2. Approximately the binding site of FKBP12 is occupied by half of the whole FK506 or rapamycin molecule while the other half is penetrated into the binding site of CaN or mTOR.

The binding site of FKBP12 constitutes Tyr82, Ile92, Phe36, Phe99, Tyr26, Phe46, Phe48, Val55, Ile56, and Trp59, and can only accommodate a ring with five-member or six –member [17]. As seen from graphs above (Figure2.1 and Figure2.2), FK506 and rapamycin have similar "binding" domain to FKBP12 and both of them have a pipecolinyl moiety (Figure2.3) which contains a six-membered ring in their binding domains.

The pipecolinyl moiety is critical in the docking process for FK506 and rapamycin to FKBP12 and is referred to as "core structure", "motif" or "molecular anchor", which is intolerant to any mutation or substitution. The active site of FKBP12 can be divided into two portions: one is relatively rigid and cannot be induced. Core structure specially binds to that portion and provides major specificity and part of affinity. The other portion is relatively flexible and the remainders of the ligand bind to that portion and confer main affinity [19] [20]. Based on this assumption, in this study the core structure of FK506 and rapamycin, pipecolinyl moiety, was first docked into the binding site of FKBP12. Subsequently, the NCI (National Cancer Institute)-Diversity-Database was screened to obtain statistical data to analyze the specificity of the small ligands bound to FKBP12.

Many new potential FKBP12 inhibitors are designed with the starting point of

pipecolinyl moiety due to its essentiality in the docking event [17] [21]. A series of FKBP12 inhibitors, FK506, rapamycin, GPI (Figure2.4), 107(Figure2.5), and 308 (Figure2.6), are studied by energy landscape theory, which all possess pipecolinyl moiety or other similar core structure.

2.2 Experimental Section

The Protein Data Bank (PDB) provides the coordinates of X-ray crystal structure for binding complex. The protein and ligand pdb files can be extracted from complex file downloaded from PDB. For macromolecules, polar hydrogen, Kollman charge and solvation parameters are assigned by AutoDock Tool (ADT). For ligands, hydrogen, partial charge, root and rotatable bonds are generated by PRODRG 2 sever [22].

To use AutoGrid to generate grid map for macromolecules, three key parameters [23] [24] [25] need adjustment: box size, spacing and grid center. In this study box size is altered from (40, 40, 40) to (120,120,120) and spacing is changed from 0.375Å to 0.55 Å considering the relatively large size of ligand under investigation. And grid center is set as the center of protein. In general, grid center can be set as the center of ligand or by selecting any atom of ligand.

Docking parameter file for ligands is made by using AutoDock and Lamarckian Genetic algorithm is chosen in AutoDock 3.0.5. In this stage three important parameters are used as default value: popsize is 50, evaluation numbers is 250000 and generation numbers is 27000 while number of run in a job is altered from 10 to 100. The other two parameters, crossover rate and mutation rate, are not critical [23] and set as default value as 0.8 and 0.02 respectively.

The semi-flexible docking was performed on RAM workstation and NCI-Diversity-Database was screened. Binding affinity and specificity for each small ligand were measured and a series of analysis were made based on those data.

2.3 Results and Discussion

2.3.1 Reproduction of Binding Process

From data in table 2.1, we can see AutoDock can reproduce the docking process to form FKBP12-ligand complexes with low RMSD value less than 1 Å except for GPI. All entries except entry for GPI in the table have six-membered ring in core structure while the core structure of GPI is a five-membered ring (Figure2.7). I think the difference in

RMSD value is due to the difference in core structure. And from the data the six-membered ring is assumed to be the better core structure than five-membered rings.

The specificity is quite significant with large ISR, which can be seen from their energy spectra with obvious energy gap (Figure2.8). The pipecolinyl moiety possesses similar ISR to FK506, rapamycin, 107 and 308. With regard to affinity, pipecolinyl moiety provides less affinity than inhibitors containing it. These results are consistent with the idea [20] that core structure provides main specificity and part of affinity while the periphery groups in the ligands only contribute to affinity. This can also quality the ISR as a good measurement of specificity in a binding process.

From the data of screening the NCI-Diversity database, the statistical distribution of free energy binding spectrum or density of states for each molecule with FKBP12 was calculated. Figure2.9 shows a plot of the energy spectrum, as well as the underlying binding energy landscape and the corresponding structures of three representative small molecules binding to FKBP12 with high, medium and low ISR. The molecule with a high ISR value of 4.11 shows pipecolinyl moiety essential for selective inhibition of FKBP12. The medium and low ISR compounds do not contain such a pipecolinyl moiety and thus are not expected to be specific inhibitors of FKBP12. The high ISR compound corresponds to a smooth energy landscape where the energy gap, which measures the steepness of the funnel towards the native state, is significantly larger than the spread of the spectrum. It effectively discriminates the native state from the rest of the local minimum. The medium ISR compound corresponds to rougher energy landscape where the energy gap is not significantly larger than the spread of the spectrum. The small ISR compound corresponds to very rough binding energy landscape where the energy gap is comparable to the spread of the spectrum. Therefore, the compounds with small or medium ISR are hard to discriminate the native state from the local minima. Only the compounds with high ISR can specifically bind to certain protein.

In figure2.10, energy spectra and corresponding structures are made for different small molecules with high ISR. Those three small molecules all present similar core structure. These data confirmed ISR as a reliable measurement for specificity in drug screening again.

Figure2.11 is the plot between affinity and ISR for 1773 small molecules in the NCI-Diversity-Database and shows no significant correlation between them. From the graph, we can see the high affinity region, the ISR spread from 1.25 to 4.5. That means only affinity cannot always guarantee the specificity. At the meanwhile, in the high specificity region, there is also a distribution for affinity from -5.5 to -9.7 kcal/mol. That proves that only specificity cannot ensure affinity either. Therefore in order to do a complete drug screening we need to consider both dimensions of affinity and specificity.

2.3.2 Distributions of affinity and specificity

The distribution of the physical variables obtained is in general universally revealing the common features among different bio-molecular binding complexes. It helps the understanding of the evolution and function. The values of the parameters in the distribution charactering the properties of the underlying energy landscape may be

different for different receptor-ligand complexes. They can be inferred from the experiments.

Experiments on random sequence protein folding and protein design have implied the statistical distributions of physical observable. Experimental and computational studies on bio-molecular binding have also shown evidences of distribution of physical relevant quantities. Since Log K is proportional to the free energy difference between the native and non-native states termed the stability or affinity, this infers the free energy also has a distribution. In the free energy distribution, the native (strong) binding state should appear in the low end of the tails where the density of these binding free energies becomes discrete. In parallel to protein folding, one therefore expects the similar physical principles underline the protein folding should also be responsible for binding. Different sequences of ligands or small molecules will have different specificity for binding with a specific receptor. So the specificity should also have a statistical distribution. The distribution of specificity should be the same as the one of free energy. This reflects different degrees of binding specificity. There is a small group of high specificity ligands among all the available ones. This is rare and lies in the high end tail of the distribution of the specificity. These are obviously the targets we are looking for. Characterizing their properties will be important to unravel the underlying functions.

The theoretical studies on the exact functional form of the statistical distributions of the physical variables to uncover the properties of the underlying binding free energy landscape are currently unexplored. Thus it is the purpose of this study to fill the gap.

External Distribution

From the data of screening the NCI-Diversity database, the statistical distribution of the affinity of 1773 small molecules with FKBP12 was calculated. Figure 2.12 and figure 2.13 show a distribution of affinity. As easily seen, near the center or the mean, the distribution can be fitted well with a Gaussian. Near the tail, the distribution of the affinity can be fitted well with exponential. This confirms the analytical we discussed above. Most of the ligands are with relatively small affinity with a receptor. A small number of the ligands have high affinity to a receptor and they are crucial for the biological function.

In figure 2.14 and figure 2.15, we show the statistical distribution of specificity characterized by ISR. We also see that the distribution of the specificity (ISR) can be fitted well with Gaussian distribution near the mean and can be fitted well with the exponential distribution at the tails. This means most of the ligands bind with a receptor non-specifically with low specificity. A few ligands bind with a receptor specifically with high specificity. They are at the high end tail of the specificity distribution.

In sum we have confirmed the analytical form of the distribution functions with 1773 diversified small molecules binding with a specific receptor, FKBP12. This statistical methodology and approach based on energy landscape theory is quite general, one expects to apply not only to protein-protein binding, but also protein-RNA, protein-DNA and RNA-DNA bindings.

Internal Distribution

We have performed the investigation of the significance and implications of flexible docking of ligands with receptor target FKBP12. Initially a diverse set of 1773 small molecules were selected from the NCI-Diversity database. All conformers of each of the 1773 selected molecules were docked with FKBP12 using AutoDock to generate a binding energy spectrum for each. From this data, the statistical distribution of free energy binding spectrum or density of states for each molecule with FKBP12 was calculated. Figure2.16 shows energy spectra corresponding to high, medium and low ISR cases. We can see that only high ISR spectrum with large energy gap can effectively discriminate the native state from the rest of local minimum.

The distributions of the free energy corresponding to the three different cases are shown. They are all Gaussian distributed at the center (Figure 2.17, 2.18, 2.19) and exponentially distributed at the tail (Figure 2.20, 2.21, 2.22). This is quite consistent with the results from analytical studies above.

The gaps between native and average as well as widths for the distributions are different for different ligands as shown in figure2.23 and figure2.24. We can see that the width of the distribution however is different for each case. For high specific binding (ISR), the width of the distribution is small relative to the gap; while for low specific binding specificity (ISR), the width of the distribution is more spread and comparable to the energy gap. The width of the distribution is a measure of the roughness of the binding free energy landscape. Rougher energy landscape has a larger width or variance in binding free energy.

In sum we have confirmed the distribution of free energy spectrum corresponding to the three different cases is universally Gaussian distributed near the mean and exponentially distributed near the tail. In this work, to the first order approximation, we have ignored correlations between different cases. It is expected that the correlations will influence the tail properties of the statistical distribution of the physical relevant variables quantitatively. It will be interesting to extend the current study to incorporate this effect.

2.4 Tables and Figures

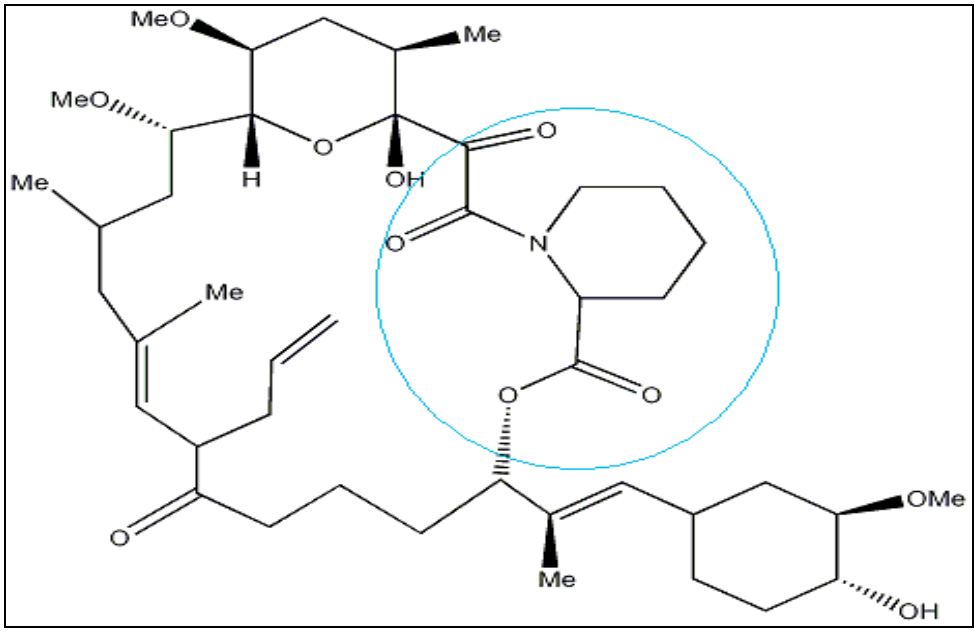


Figure 2.1 Structure of FK506

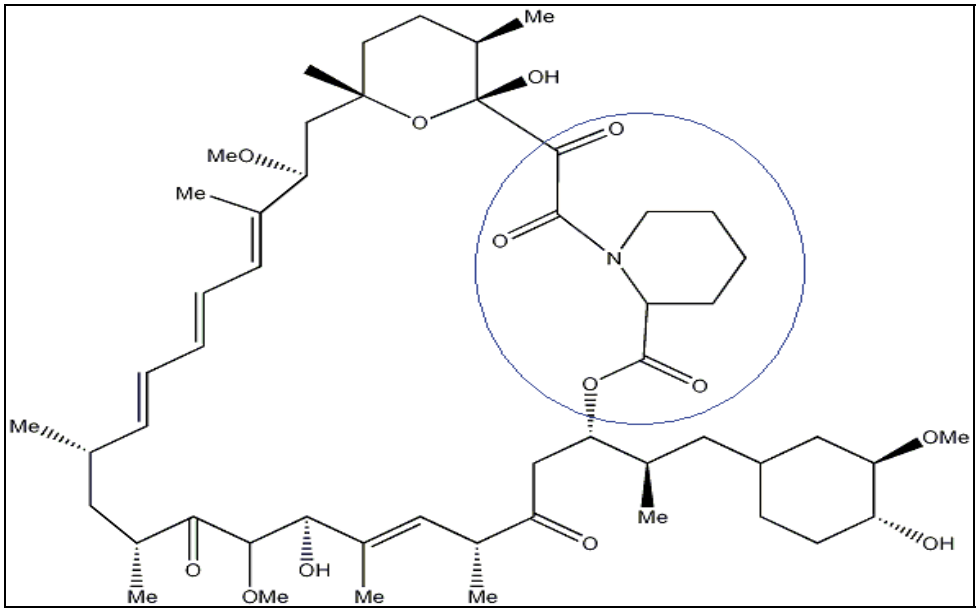


Figure 2.2 Structure of Rapamycin

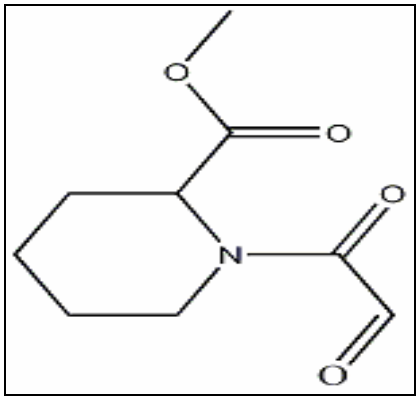


Figure 2.3 Structure of pipecolinyl moiety

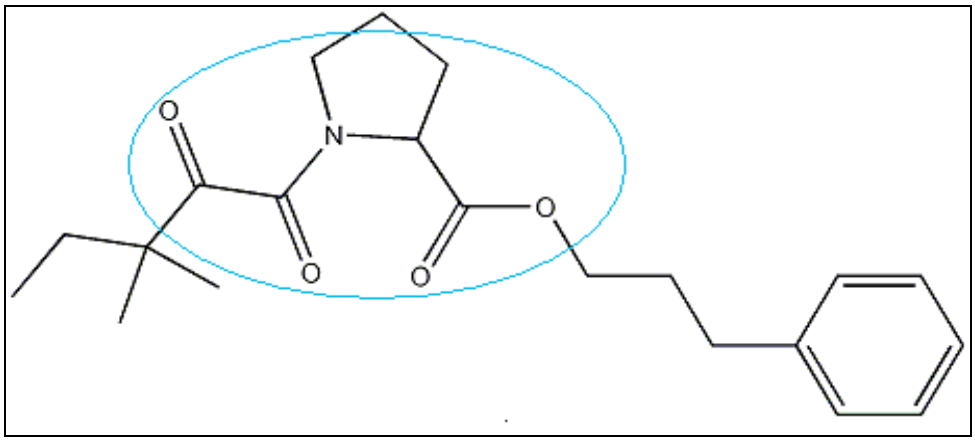


Figure 2.4 Structure of GPI

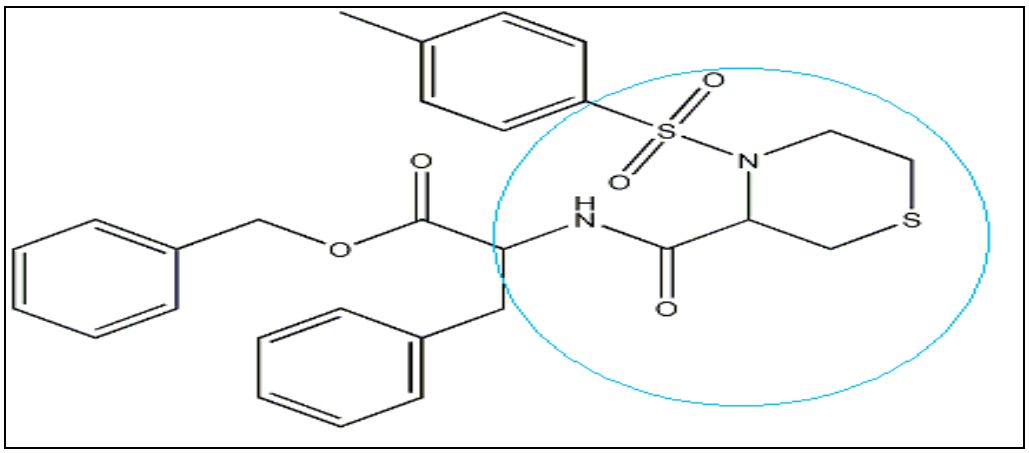


Figure 2.5 Structure of 107

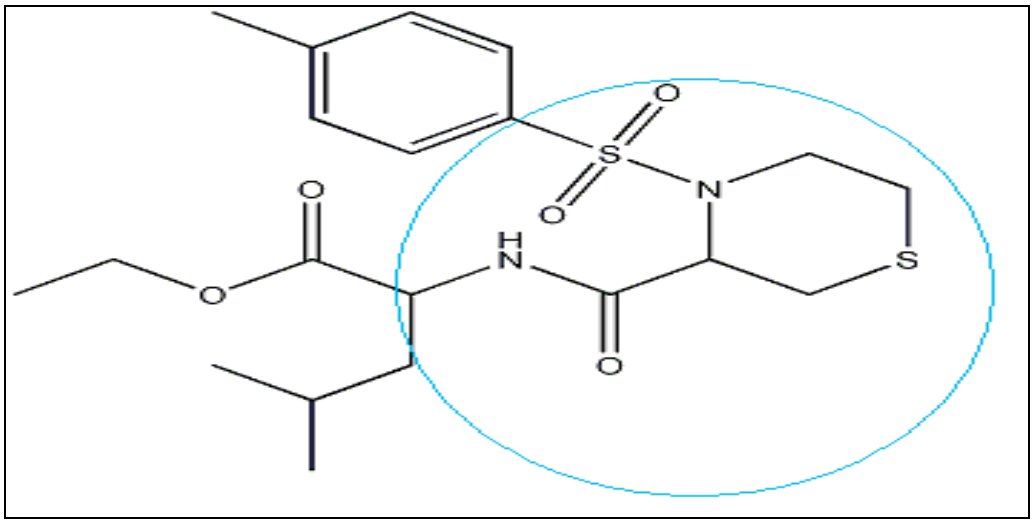


Figure 2.6 Structure of 308

Compound	Calculated Binding Affinity (kcal/mol)	Intrinsic Specificity Ratio (ISR)	RMSD(Å)
pipecolinyl moiety	-5.66	4.31	0.59
FK506	-10.18	4.21	0.67
rapamycin	-12.55	4.30	0.48
GPI	-6.00	3.08	1.26
107	-7.03	4.04	0.76
308	-6.31	4.26	0.76

Table 2.1 Affinity, ISR and RMSD for FKBP12 Inhibitors with Similar Core Structure

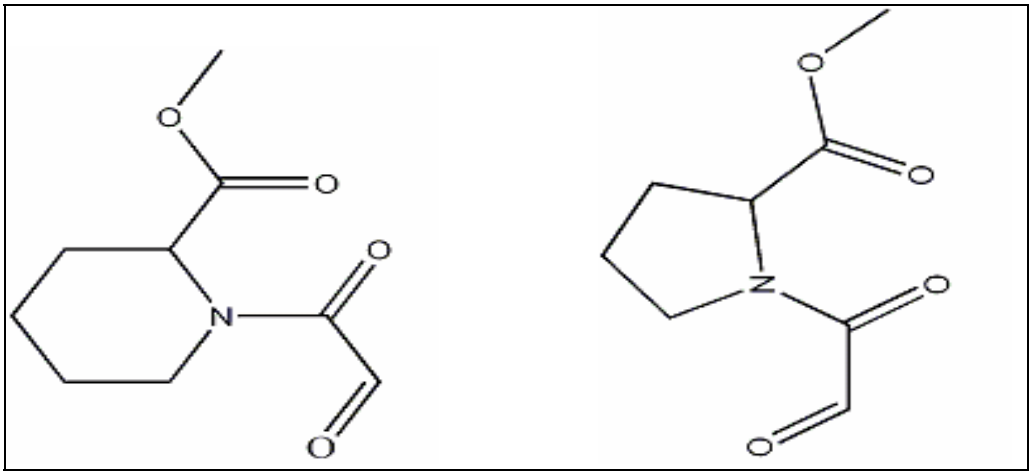


Figure 2.7 Core structure for FK506 and GPI

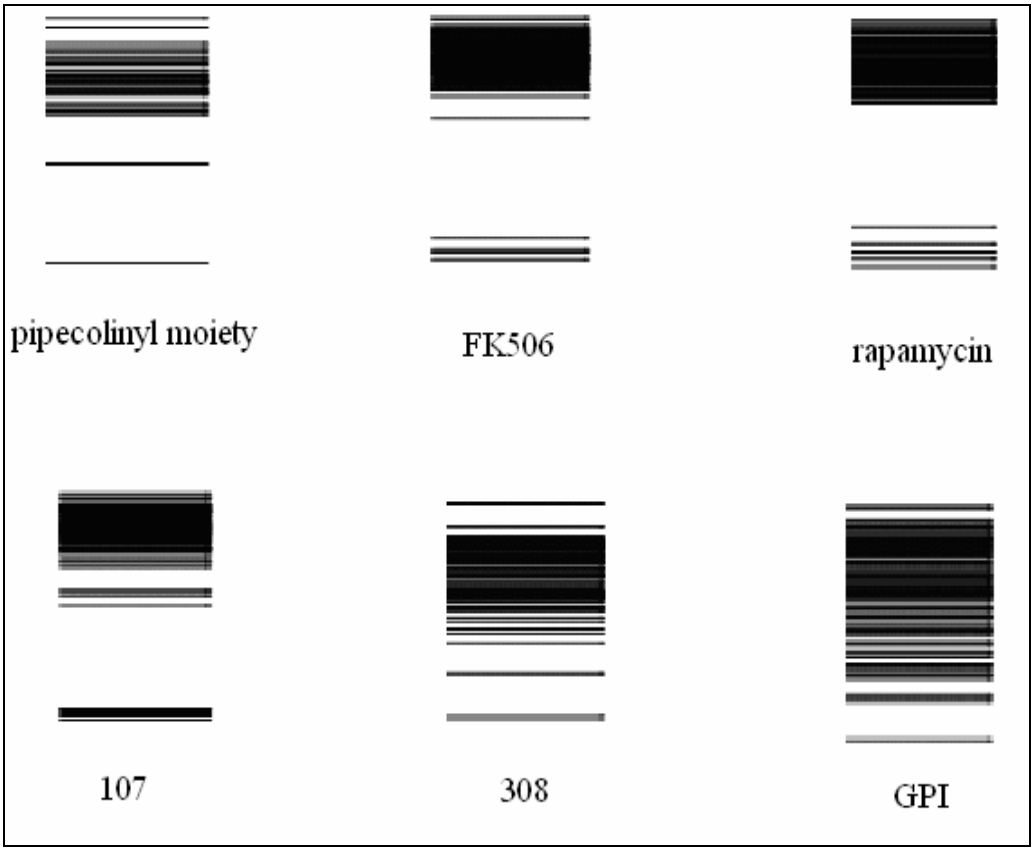
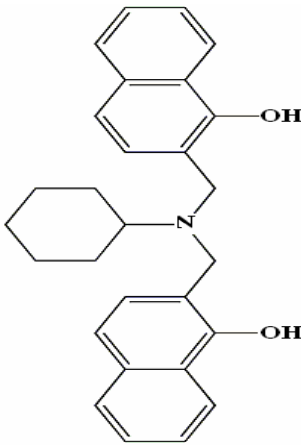
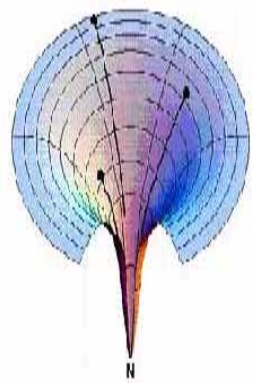
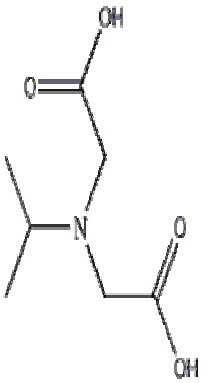


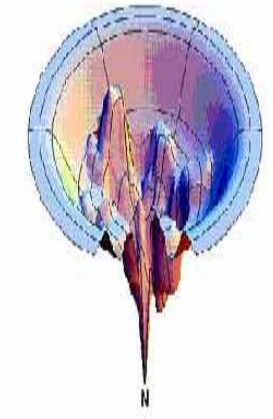
Figure 2.8 Energy Spectra for different inhibitors



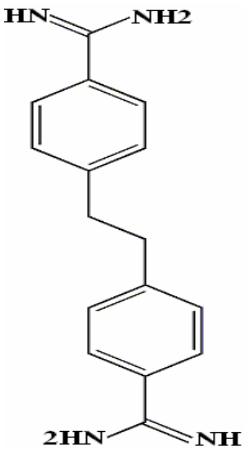


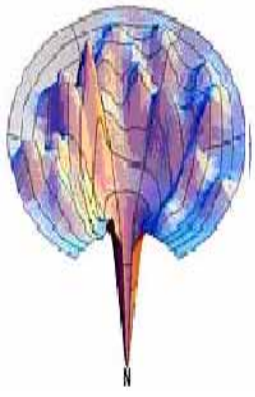
ISR=4.11





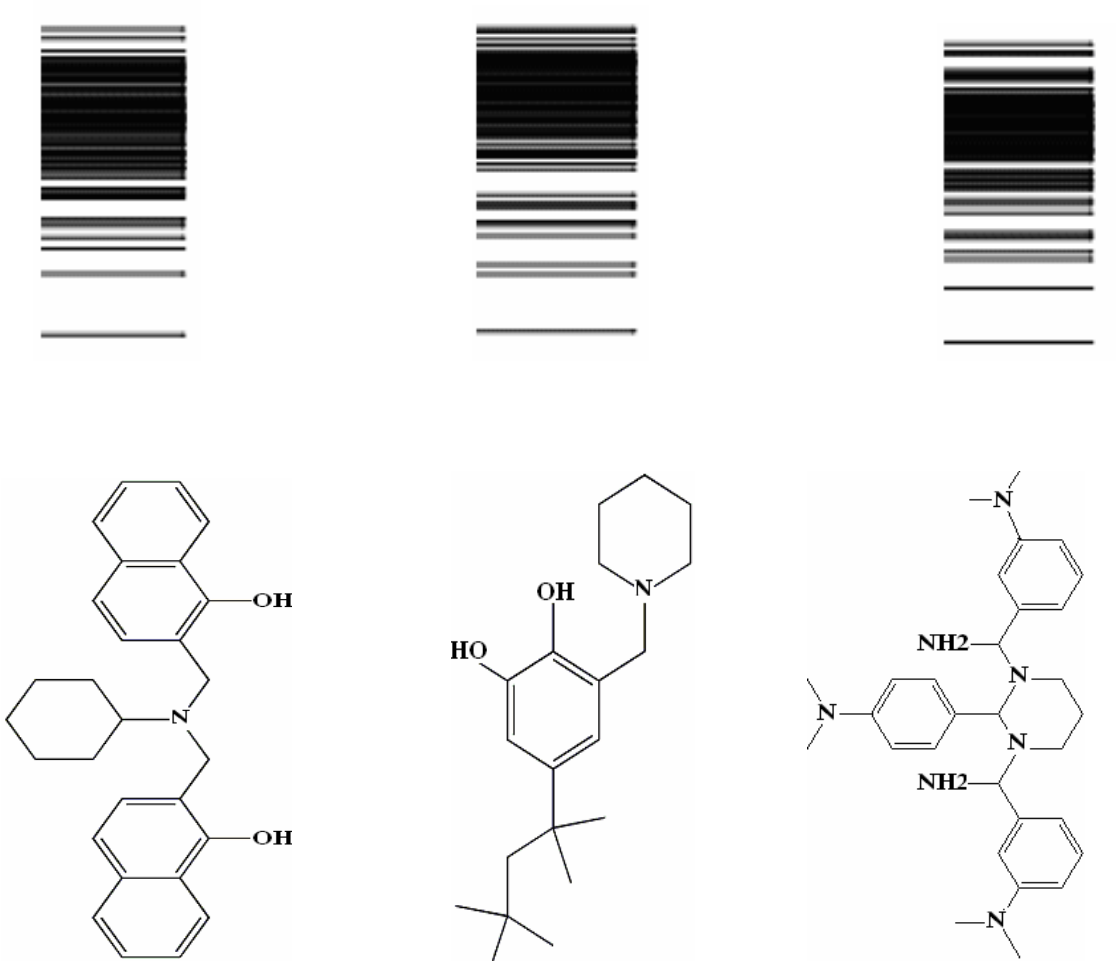
ISR=2.56





ISR=1.56

Figure 2.9 Energy Spectra for small molecules with different ISR



ISR=4.11 **ISR= 4.07** **ISR=4.09**

Figure 2.10 Energy Spectra for small molecules with high ISR

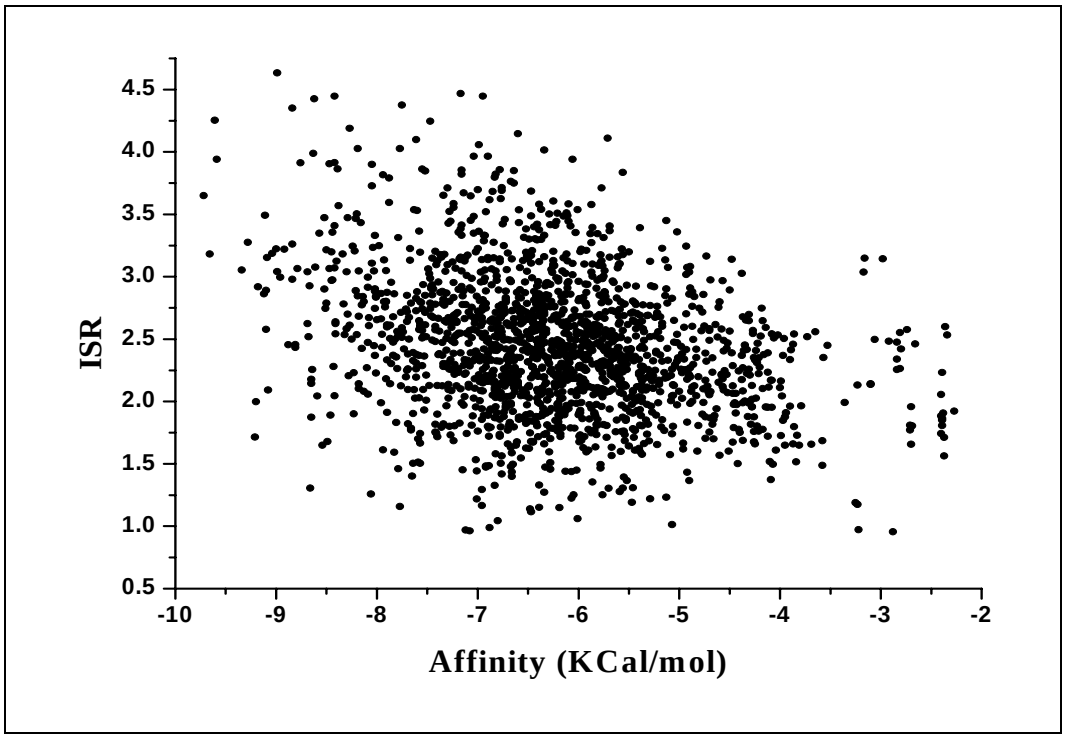


Figure 2.11 Affinity versus ISR

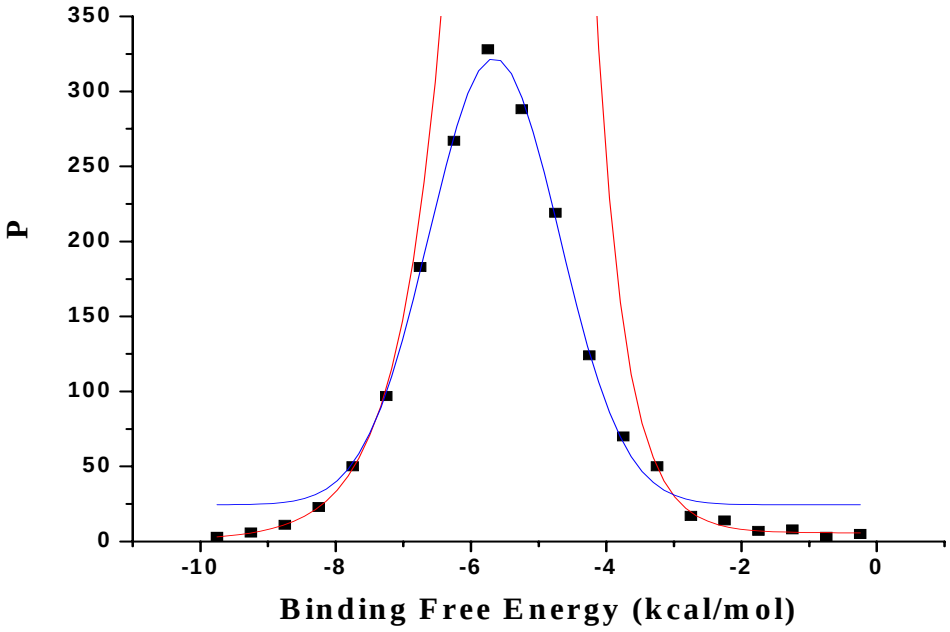


Figure 2.12 External Affinity Distribution of FKBP12

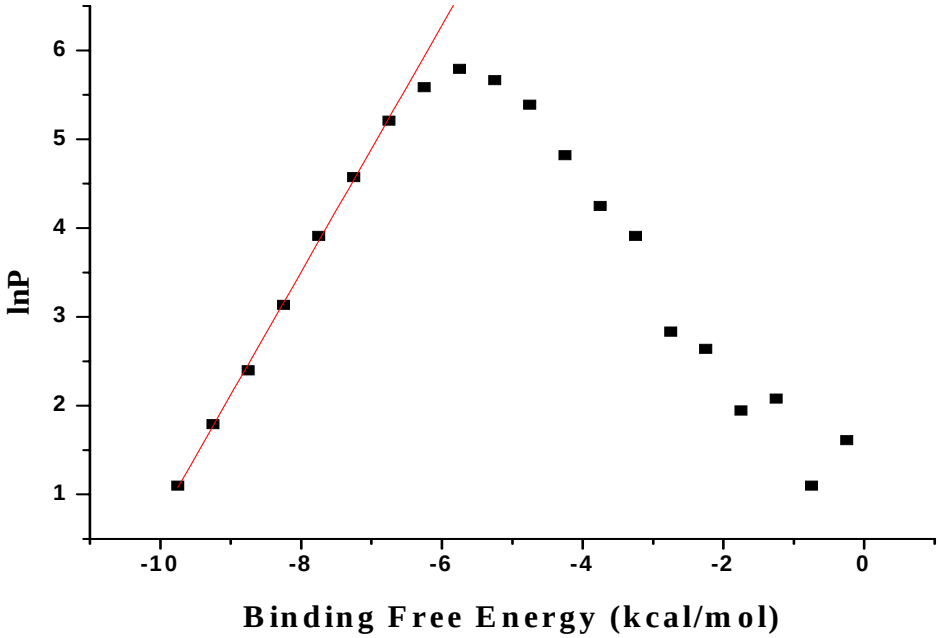


Figure 2.13 External Affinity Tail Distribution of FKBP12

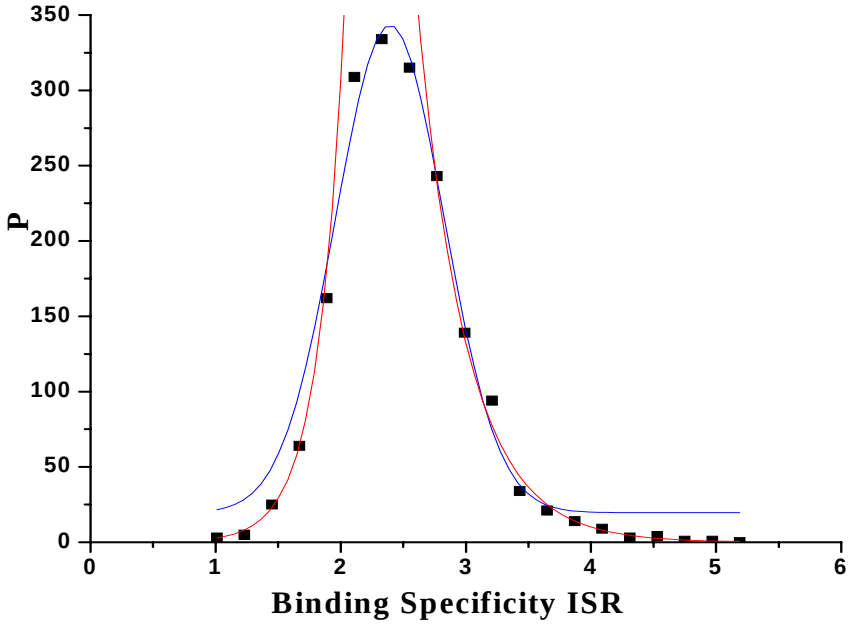


Figure 2.14 External Specificity Distribution of FKBP12

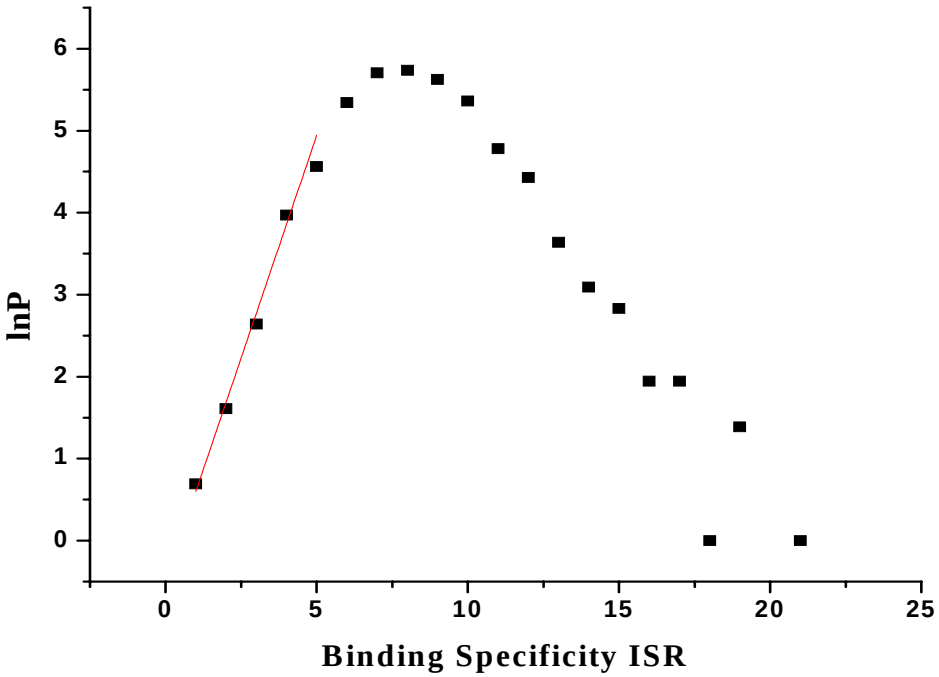
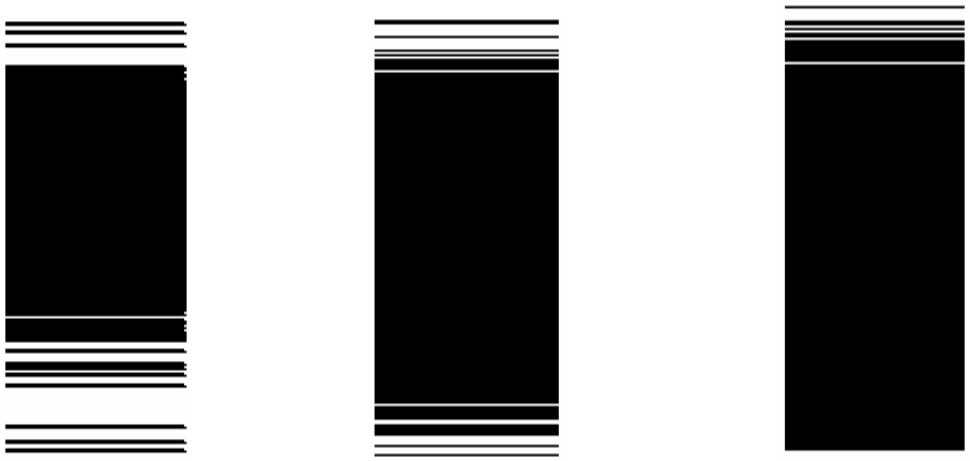


Figure 2.15 External Specificity Tail Distribution of FKBP12



ISR=4.50 **ISR=2.51** **ISR=1.95**

Figure 2.16 Energy Spectra of FKBP12 for Three Cases

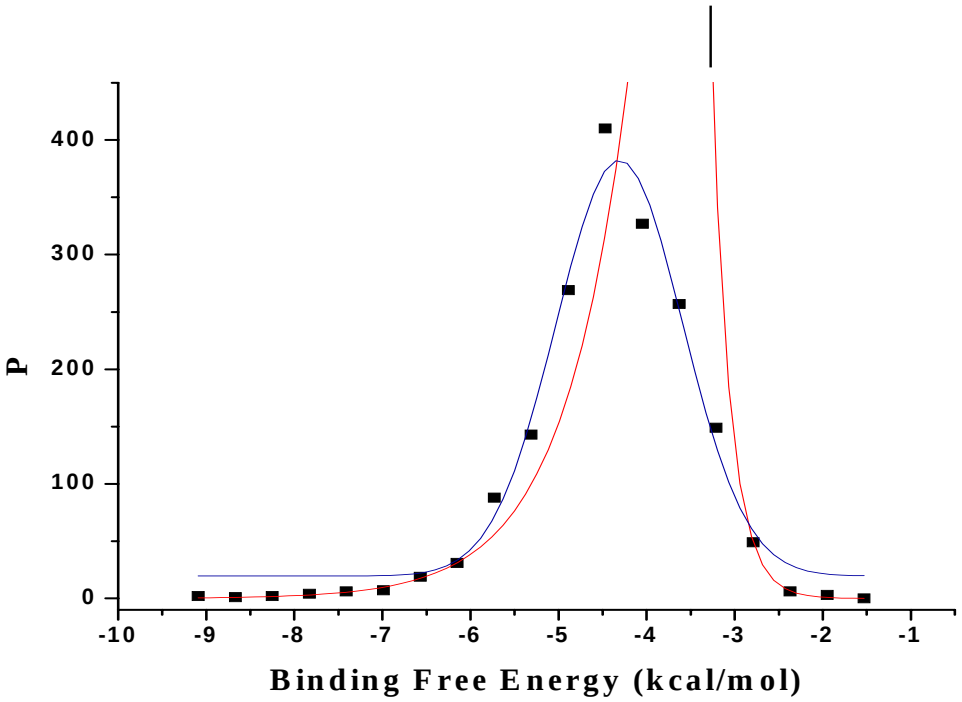


Figure 2.17 Internal Affinity Distribution of FKBP12 (ISR=4.50)

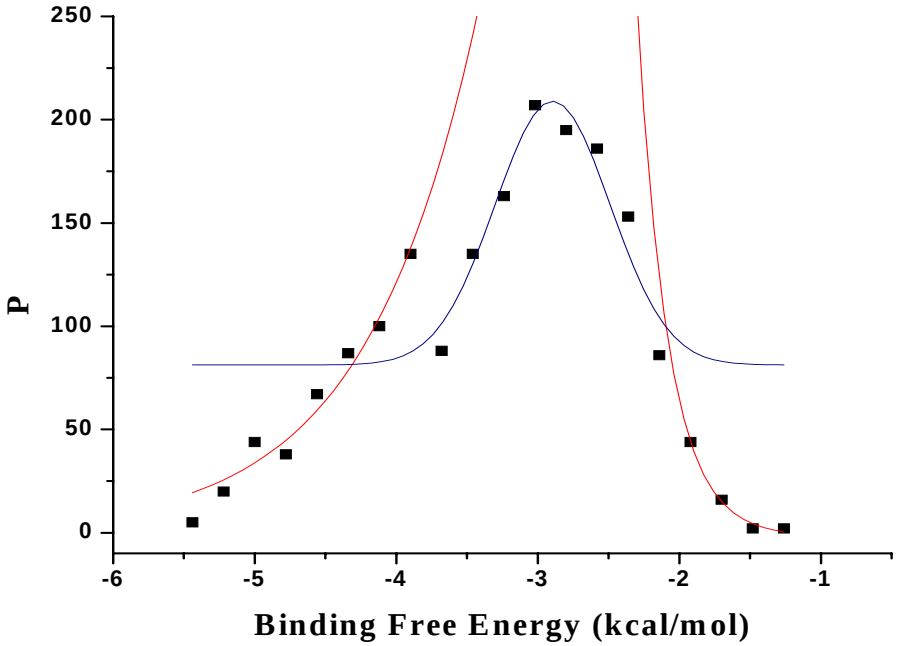


Figure 2.18 Internal Affinity Distribution of FKBP12 (ISR=2.51)

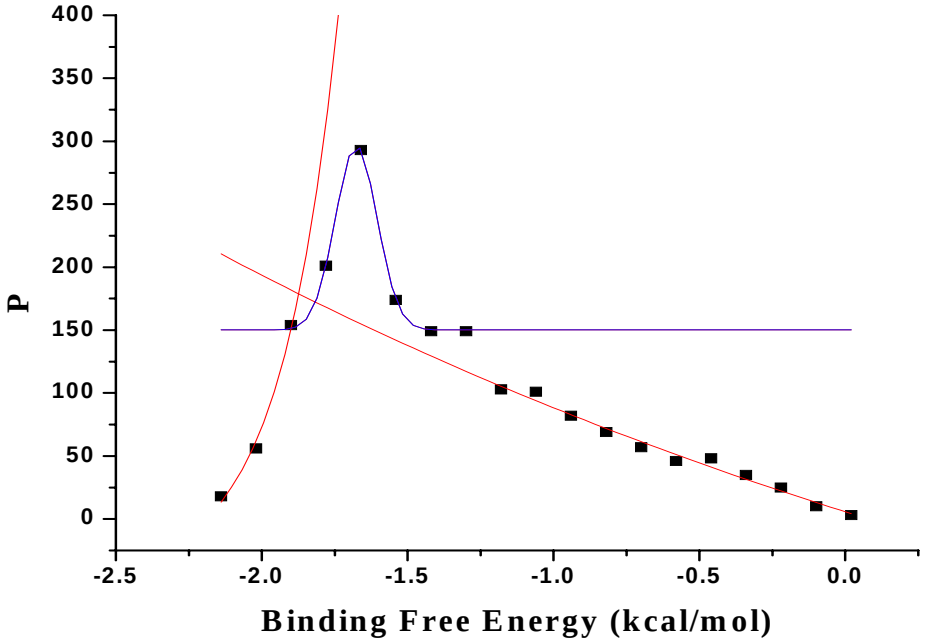


Figure 2.19 Internal Affinity Distribution of FKBP12 (ISR=1.48)

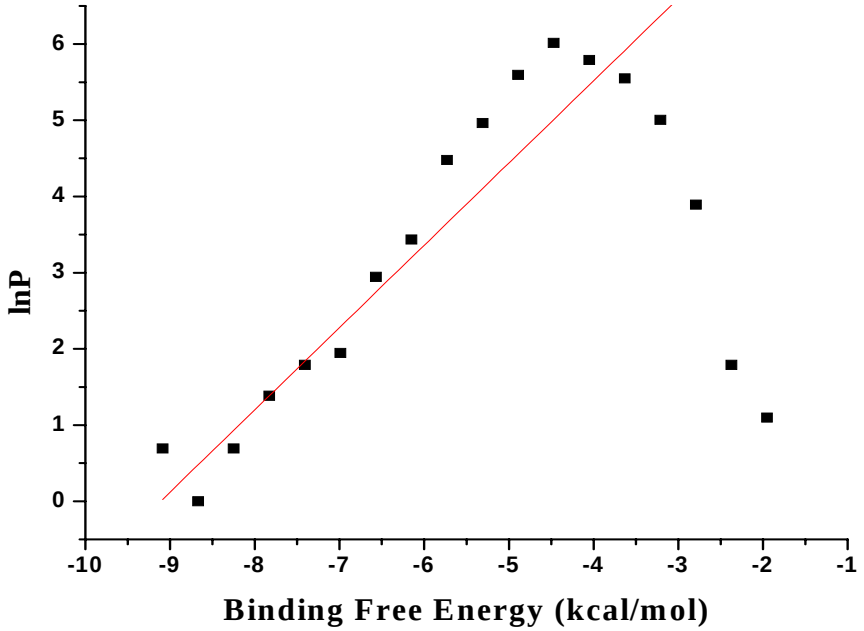


Figure 2.20 Internal Affinity Tail Distribution of FKBP12 (ISR=4.50)

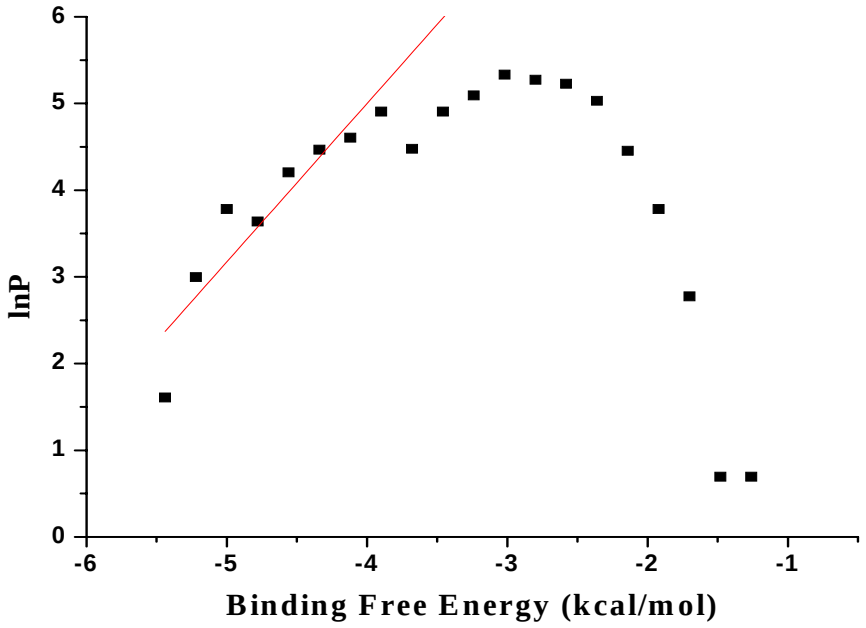


Figure 2.21 Internal Affinity Tail Distribution of FKBP12 (ISR=2.51)

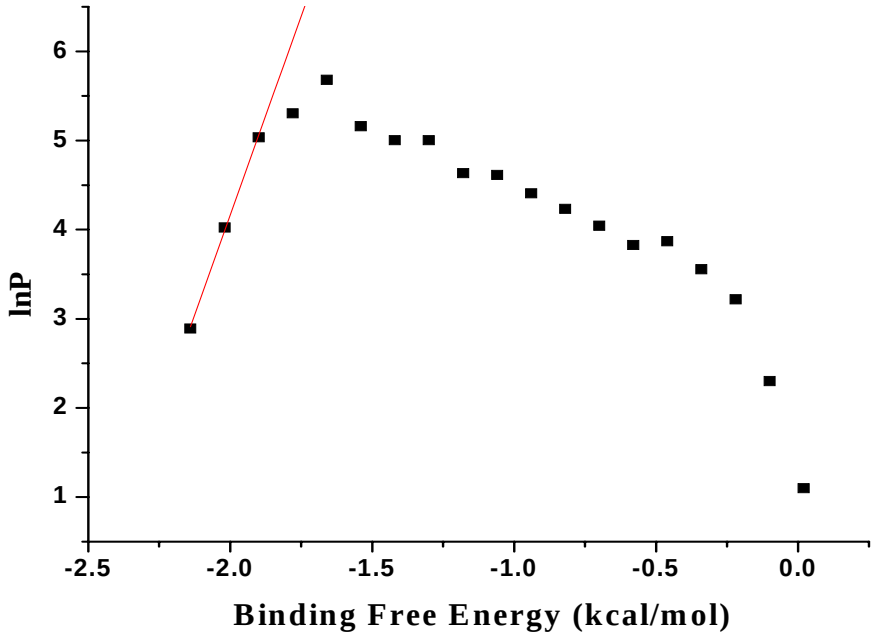


Figure 2.22 Internal Affinity Tail Distribution of FKBP12 (ISR=1.48)

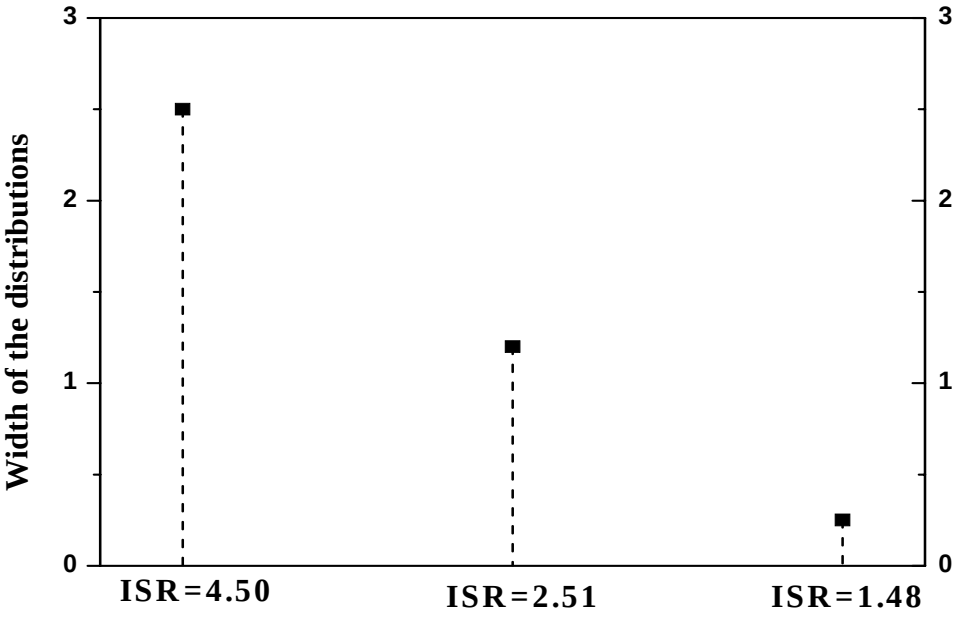


Figure 2.23 Comparison of Width of Internal Distribution for Three Cases

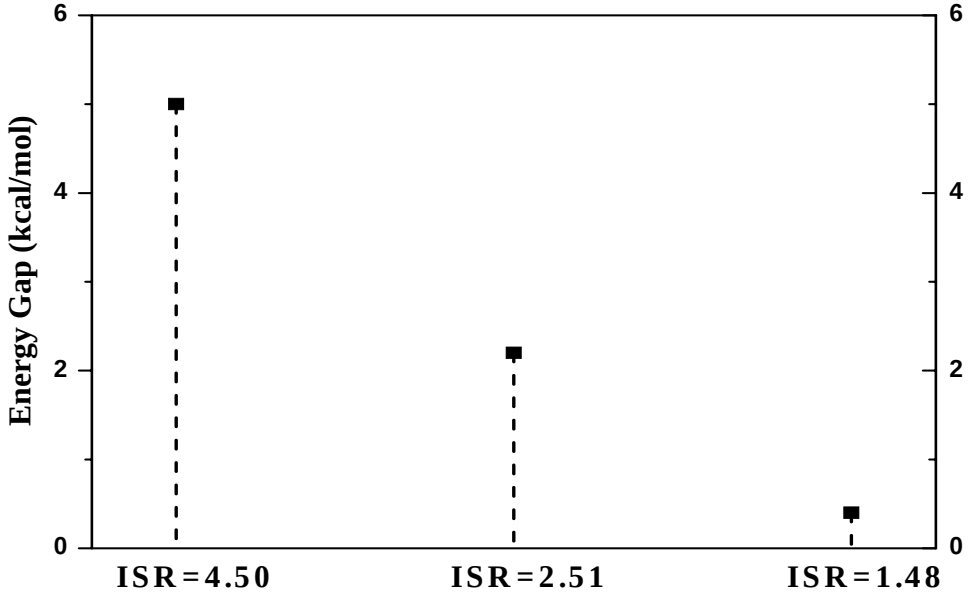


Figure 2.24 Comparison of Energy Gap for Three Cases

Chapter 3: New Inhibitors Prediction for SmpB

3.1 Introduction of SmpB

Small protein B (SmpB) is an essential component of SsrA.SmpB system which is highly conserved in all bacteria and vital for bacterial survival. SsrA is a remarkable RNA molecule, also known as SsrcRNA, tmRNA, or 10SaRNA, and functions as both a transfer RNA (tRNA) and a messenger RNA (mRNA). The tRNA-like domain of SsrA is proposed to be the site of SmpB binding. Upon SmpB bound to SsrA with high affinity, dual biological functions of releasing stalled ribosome from damaged messenger RNAs and targeting incompletely synthesized protein fragments for degradation [26], [27], [28], and [29].

Since the SmpB is significantly involved into the survival of bacterial, it has become a hot target for drug discovery and design. Unfortunately, by now there is no drug with significant affinity and specificity to SmpB to block its function. Recently, Professor Wali Karzai's group found that the hotspot in SmpB is the pocket composed by E26, L86, H88, K89 and K119. It really provides the most useful information for binding site and therefore makes feasible the docking job by virtual screening.

3.2 Experimental Section

The Protein Data Bank (PDB) provides the coordinates of X-ray crystal structure for binding complex (1P6V). The protein and ligand pdb files are extracted from complex file downloaded from PDB. For macromolecules, polar hydrogen, Kollman charge and solvation parameters are assigned by AutoDock Tool (ADT). For ligands, hydrogen, partial charge, root and rotatable bonds are generated by PRODRG 2 sever.

The three key parameters in the stage of AutoGrid are: box size (120,120,120), spacing (0.55 Å) and grid center (the center of protein). In the next stage of AutoDock, all important parameters are used as default value: pop size is 50, evaluation numbers is 250000, generation numbers is 27000, number of run in a job is 100, crossover rate and mutation rate are 0.8 and 0.02 respectively.

After the successful reproduction of the protein-ligand complex, the set of parameters are adopted to screen the NCI-Diversity-Database on RAM workstation. After screening the databases and using the two dimensional drug screening criterion, the top 25 binders in the NCI-Diversity-Database were selected and visually inspected via the VMD software to ensure that they are right inside the hotspot of SmpB.

3.3 Results and Discussion

In the top 25 binders, 2 of them are inside the hotspot of SmpB and selected as potential lead compounds. Figure 3.1 and figure 3.2 show the chemical structures of two compounds for experimental testing. We can see both of them present high affinity and high specificity. Figure3.3 and figure3.4 show their relative position in the binding pocket of SmpB. From the graphs, we can see clearly that two small molecule penetrate deeply inside the expected binding pocket. All these experiment results suggest that these two small molecules are likely to be potential lead compounds to SmpB.

However, the number of potential lead compounds is too little due to the relative small size of the database. We expect more potential molecules to find some common properties from chemical structures. Now another database ZINC which contains about 2 million molecules is screening against SmpB with exactly the same parameter set as NCI-Diversity-Database on RAM workstation. If we can find more potential lead compounds, we may identify some common properties from them. Furthermore we can improve their affinity, specificity, solubility and etc. as needs and even design some new ones base on those common properties. After we select the most possible small molecules, we can carry out a series of in vitro experiments by the cooperation with Professor Wali Karzai's group.

3.4 Tables and Figures

NCI-Diversity-Database ID: Diversity0465
Binding Energy: -10.28 kcal/mol, ISR=3.54
Formula: $C_{28}H_{38}N_2O$

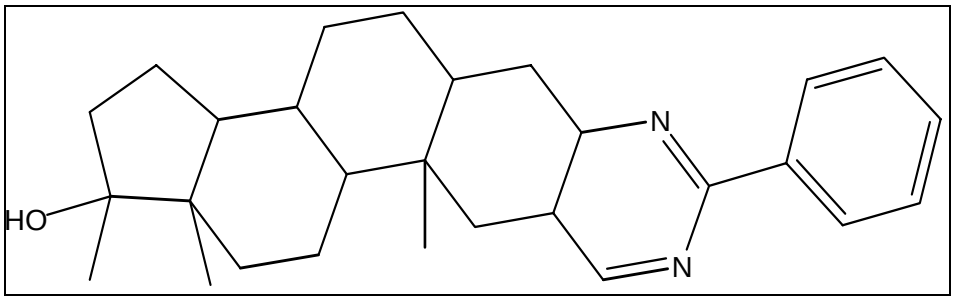


Figure 3.1 Structure for Diversity0465

NCI-Diversity-Database ID: Diversity1117
Binding Energy: -11.05 kcal/mol, ISR=2.91
Formula: $C_{42}H_{52}N_2$

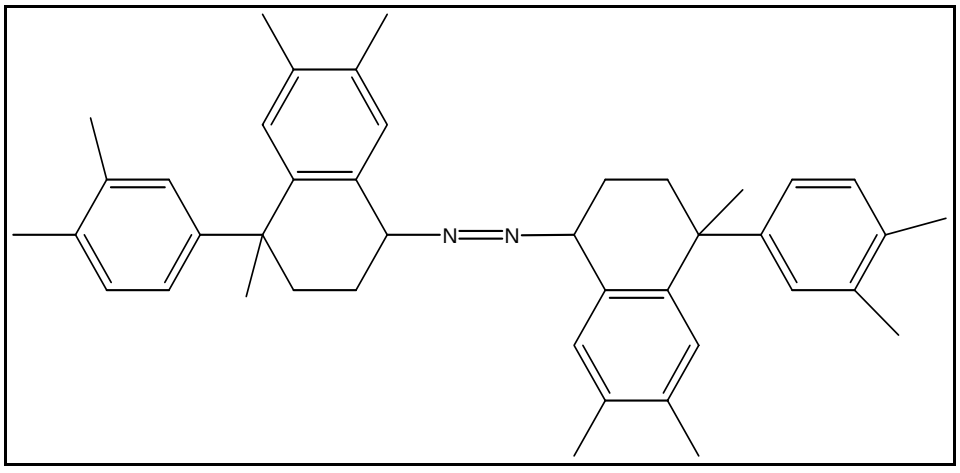


Figure 3.2 Structure for Diversity1117

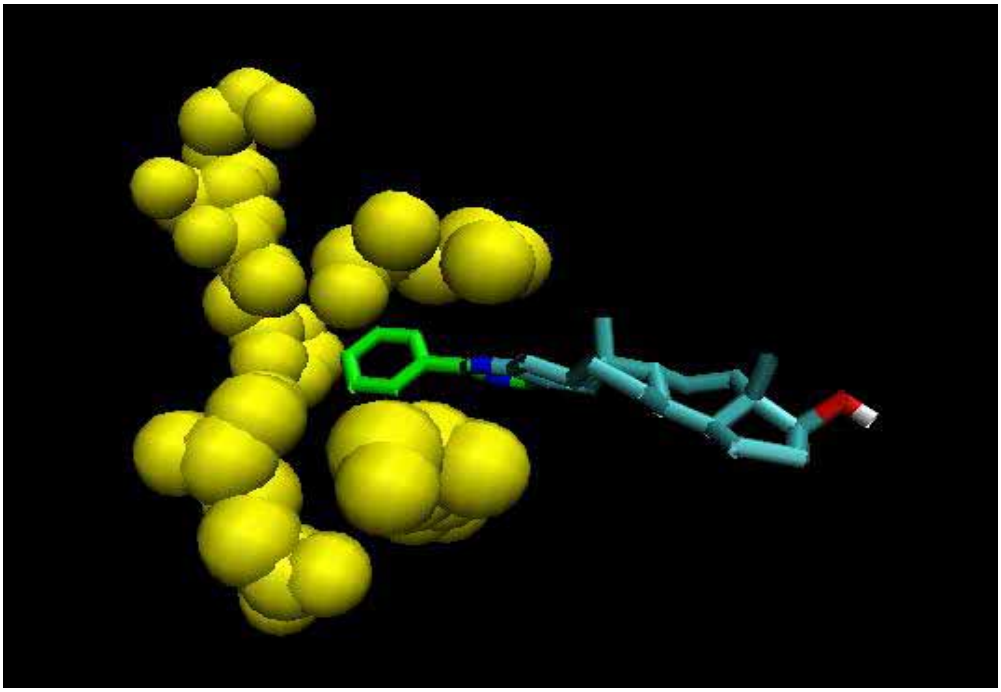


Figure 3.3 Diversity0465 in the binding pocket of SmpB

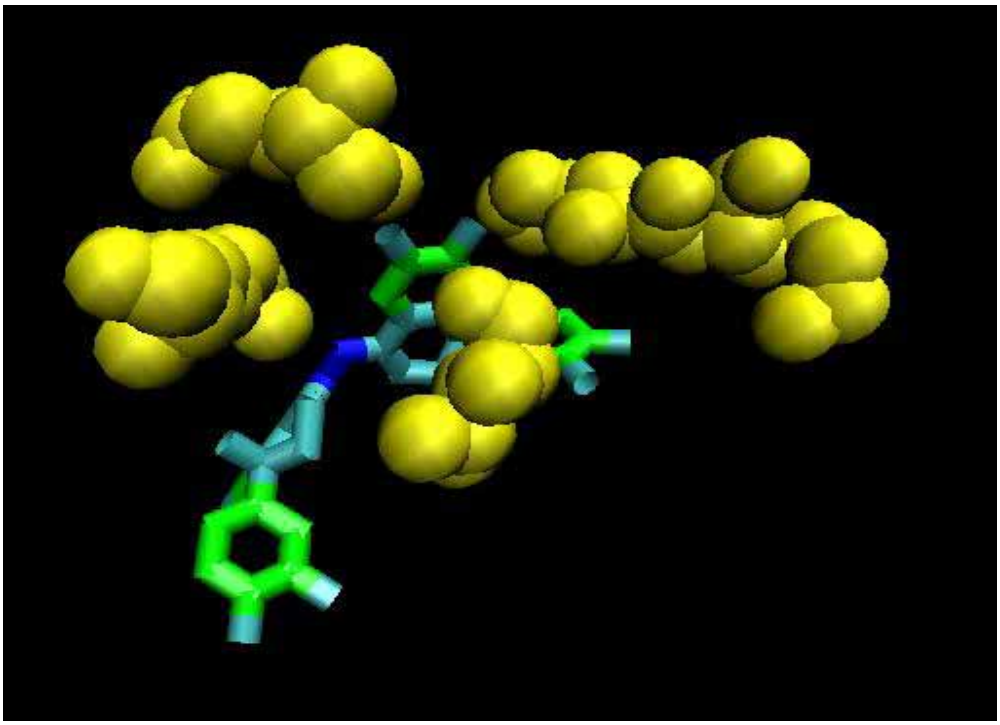


Figure 3.4 Diversity1117 in the binding pocket of SmpB

References

[1] Jin Wang and Gennady M. Verkhivker (2003) Energy Landscape Theory, funnels, Specificity, and Optimal Criterion of Biomolecular Binding, Physical Review Letters, Volume 90, Number18, 2003.

[2] J Andrew McCammon (1998) Theory of biomolecular recognition, Current Opinion in Structural Biology 8: 245-249.

[3] Tami J. Marrone, James M. Briggs, and J. Andrew McCammon (1997) Structure-Based Drug Design: Computational Advances, Annu. Rev. Pharmacol. Toxicol, 37: 71-90.

[4] Irwin D. Kuntz (1992) Structure-based Strategies for Drug Design and Discovery Science Vol. 257, 1078-1082.

[5] Inbal Halperin, Buyong Ma, Haim Wolfson, and Ruth Nussinov (2002) Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions PROTEINS: Structure, Function, and Genetics 47:409-443.

[6] P. DeBrunner, J. Tsibris and E. Munck (1969) C. Levinthal, in Mossbauer spectroscopy in biological systems, University of Illinois Press, Urbana P.22.

[7] P.J.Goodford (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J.Med.Chem. 28, 849-857.

[8] Paul S. Charifson, Joseph J. Corkery, Mark A. Murcko, and W. Patrick Walters (1999) Consensus scoring: A method for obtaining improved hit rates from Docking databasesof three-dimensional structures into proteins, J. Med. Chem. 42, 5100-5109.

[9] Emanuele Perola, W. Patrick Walters and Paul S. Charifson (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance, Proteins: Structure, Function, and Bioinformatics 56:235-249.

[10] Paul A. Bush, U. Chandra Singh, Frank K.Brown, Robert Langridge and Peter A. Kollman (1987) Calculation of the relative change in binding free energy of a protein-inhibitor complex , Science, Vol. 235.

[11] Scott C. -H. Pegg, Jose J. Haresco,and Irwin D.Kuntz (2001) A genetic algorithm for structure-based de novo design Journal of Computer-Aided Molecular Design,15:911-933.

[12] Steffan Ho, Nell Clipstone, Luika Timmerann, Jeffrey Northrop, Isabella Graef, David Fiorentino, Jame Nouse, and Gerald R. Carbtree (1996) The Mechanism of Action of Cyclosporine A And FK506, Clinical Immunology and Immunopathology, Vol. 80, No.3, Spetember, pp.S40-S45.

[13] Fancis J.Dumont (2000) FK506, An Immunosuppressant Targeting Calcineurin Function, Current Medicinal Chemistry, 7, 731-748.

[14] Robert T. Abraham and Gregory J.Wiederrecht (1996) Immunopharmacology of Rapamycin, Annu.Rev. Immunol. 14:483-510.

[15] Eric J,Brown, Mark W. Alberts, Tae Bum Shln, Kazuo Ichlkawa, Curtls T. Kelth, Willam S. Lane, and Stuart L. Schrelber (1994), A mammalian protein targeted by G1-arresting rapamycin-receptor complex , Letters to Nature, Vol.369.

[16] Jungwon Choi, Jie CHen, Stuart L. Schreiber and Jon Clardy (1996) Structure of

the FKBP12-Rapamycin complex interacting with the binding domian of human FRAP , Science Vol.273.

[17] Fei Sun, Pengyun Li, Yi Ding, Liwei Wang, Mark Bartlam, Culin Shu, Beifen Shen, Hualiang Jiang, Song Li and Zihe Rao, Design and Structure-based Study of New Potential FKBP12 Inhibitors, Biophysical Journal Volume 85 November 203 3194-3201.

[18] Robert E.Babine and Steven L. Bender (1997) Molecular Recognition of Protein-ligand Complexes: Applications to Drug Design", Chem Rev. 97, 1359-1472.

[19] Paul A. Rejto and Gennady M. Verkhivker (August 1996) Unraceling priciples od lead discovery : Fromm unfrustrated energy landscapes to novel molecular anchors, Proc. Natl. Acad. Sci. USA Vol. 93, pp.8945-8950,.

[20] Nirav K. Shah, Paul A. Rejto, and Gennady M. Verkhivker (1997) Structural consensus in ligand-protein docking identifies recognition peptide motifs that bind streptavidin PROTEINS: Structure, Function, and Genetics 28:421-433.

[21] Robert E. Babine, T.M.Bleckman, C.R.Kissinger, R.Showalter, L.A.Pelletier, C.Lewis, Kathleen Tucker, Eleen Moomaw, H.E.Parge and J. ernest Villafranca (1995), Design, Synthesis and X-ray Crystallographic Studies of Novel FKBP-12 Ligands, Bioorganic & Mecinal Chemistry Letters, Vol. 5, No. 15, pp. 1719-1724.

[22] A.W. Schuettelkopf; D.M.F. Van Aalten Acta Crystallographica (2004) 1635-1663.

[23] Yuan-ping Pang, Norberto D. Silva, Jr., Christopher Haydock, and Franklyn G. Prendergast (1997) Docking Studies on the Complexed and Uncomplexed FKBP12 Structures with Bound and Unbound Ligands: An Implication of a Conformational Selection Mechanism for Binding, J. Mol. Model. 3, 240-248.

[24] Csaba Hetenyi and David Van Der Spoel (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site, protein Science 11: 1729-1737.

[25] Wesley E. Stites (1997) Protein-protein interactions: interface structure, binding thermodynamics and mutational analysis, 97, 1233-1250.

[26] Gang Dong, Jacek Nowakowski and David W. Hoffman (2002) Structure of small protein B: the protein component of the the tmRNA-SmpB system for ribosome rescue, The EMBO Journal, Vol. 21, No.7, pp1845-1854.

[27] Jacek Wower, Christian w. Zwieb, David W. Hoffman, and Iwona K. Wower (2002) SmpB: A protein that binds to Double-Stranded Segments in tmRNA and tRNA, Biochemistry 41, 8826-8836.

[28] A Wali Karzai and Robert T. Sauer (2001) Protein factors associated with the SsrA-SmpB tagging and ribosome rescue complex, 3040-3044, PNAS, March 13, 2001 Vol.98, No.6.

[29] A Wali Karzai, Eric D. Roche and Robert T. Sauer (June 2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue, Nature Structural biology, Volume 7, number 6.