Chapter 6

IS-IT?

In Silico Inhibitor Identification Tool
6.1 Introduction

Virtual screening (referred as VS) has become an important part of most drug discovery programs. Its application has led to several success stories where small chemical compounds exhibiting lower to high-range affinities for the target have been identified (Muegge & Oloff 2006). The author was principally involved in the development of two web servers in the host lab that are important from the point of virtual screening and selection of targets from pathogenic organisms.

This chapter deals with the development of IS-IT?: a web-server for evaluating protein-ligand interactions using the academically freely available AUTODOCK 4.0 program as the backend. The ongoing dominant paradigm in drug discovery is the 'one gene, one drug, one disease' paradigm (Drews, 2000). The concept in this paradigm is to design maximally selective ligands to act on individual drug targets. However, many effective drugs act via modulation of multiple proteins rather than single targets. Advances in systems biology are revealing a phenotypic robustness and a network structure in biological systems. The network analysis predicts that if, in most cases, deletion of individual nodes has little effect on disease networks, modulating multiple proteins may be required to perturb robust phenotypes. This finding led to the emergence of a different drug discovery approach called network pharmacology where the idea is to target multiple proteins in biological networks with a single small-molecule inhibitor (Hopkins, et al., 2008). The rational design of polypharmacology faces considerable challenges in the need for new methods to validate target combinations and optimize multiple structure-activity relationships while maintaining drug-like properties. The IS-IT?: web server has been developed to address this issue. The server can help in design of a 'superligand' that is capable of blocking several proteins from a single or multiple pathways of a biological system using virtual screening as a strategy.

VS is a highly CPU-intensive task and is hampered by series of problems like errors in PDB files; induced fit mechanisms; etc. As expected, there are very few freely accessible web-based tools available to the academic user interested in the evaluation of ligand-protein interactions through molecular docking approaches. There is a need for a simple web-based interface that can allow even relatively novice users to perform virtual
screening. A major difficulty that novice users of docking programs face is the optimization of the parameters necessary for a particular protein target. The learning curve is steep compared to the simple YES/NO answer being asked by the scientific worker eg. a medicinal chemist who wants to know whether a compound is likely to be an inhibitor of a particular protein.

IS-IT?: is a simple menu driven web-server based on AUTODOCK4 (Morris, et al., 1998), in which the user has only to input a small molecule in the widely used PDB format and select any of over 110 proteinaceous targets implemented in the server. Results can be retrieved through a number generated at the time of job submission. The docking parameters have been optimized for these drug targets and for quick comparison the results also consist of a precalculated score based on known ligand or inhibitors. AUTODOCK (Morris, et al., 1998) is a robust and widely used docking program that is freely available to academics and has been tested by several groups (http://autodock.scripps.edu/). There are a couple of front-end applications available like BDT (Vaqué, et al., 2006) and DOVIS (Zhang, et al., 2008). The latter permit automated screening of compounds, but are not web-based and consequently involve implementing the programs at the local level. Additionally optimization of the docking parameters is a hurdle to be overcome by new users of these applications.

6.2 Description of the web interface

The IS-IT?: server uses the docking engine of the AUTODOCK4 at the back end. The front-end web interface is written in both PHP and HTML and the PHP scripts were used to integrate the front-end with the AutoDock docking scripts. The server prepares the input files automatically and initiates the AutoGrid/AutoDock routines in a sequential manner. In the first instance, the submitted ligand files along with the list of selected target protein(s) are stored with a unique job identification number. Then the PHP script prepares the ligand file, picks up the protein file and respective pre-calculated docking parameters and initiates the AutoGrid and AutoDock run. Later, these scripts also prepare the result file; by both the preparation of ligand-protein docked complex as well as the extraction of the lowest docking energies as output. All the hydrogen atoms and
Chapter 6  

Is-it?: Web Server

Gasteiger charges are assigned during ligand input file preparation. The web-server is hosted with the help of XAMPP server (www.apachefriends.org/en/xampp.html) on Redhat V4 Linux workstation. The snapshot of the IS-IT?: homepage window is shown in Figure 6.1.

The IS-IT?: homepage window has five self-explanatory modules. In the first module called IS-IT?:, the user can submit a ligand of interest in the common Protein Data Bank (Berman, et al., 2000) file format. Subsequently the choice of targets to be tested has to be selected and job has to be submitted in the server queue. A job-number is given to the user to enable retrieval of the calculations from the Results module. The calculations are run according to pre-optimized parameters for the selected target and takes approximately 15-20 minutes per target. The optimization process has been carried out using known inhibitors/ligands of the given protein and includes fixing the grid size and grid center to define the active site of the target. The snapshot of the IS-IT?: window is shown in Figure 6.2.

The MyDock module allows the user to input both the protein and ligand coordinates in the PDB format for the docking calculations. As before a unique job identification number allows for the retrieval of the results. The user, in this case, can optionally specify the grid center as well as the grid size where the ligand will be docked in the receptor. The coordinates of an atom in the center of the active site can be specified as grid center. If user doesn't specify any grid size the default size of 40x40x40 points is chosen that covers a box of dimensions approximately 15 Å. The snapshot of the Mydock window is shown in Figure 6.3.

The Result module as the name suggests returns result for the docking jobs on submission of unique job identification number. The snapshot of the Result window along with the job identification number is shown in Figure 6.4 and a sample output window is shown in Figure 6.5. The Help module is self-explanatory and includes a link to sample results. The drug targets implemented in the server are described in the Target module.
Fig 6.1 Snapshot of the *IS-IT?*: server Home page. The server has five modules: IS-IT?: for ligand docking against a set of well-known drug targets, Mydock for user specified ligand-protein docking, Result module for retrieving docking results, Target for details of targets implemented in IS-IT?: module and lastly Help module.
Fig 6.2 Snapshot of the IS-IT?: browser window. The ligand file can be uploaded in PDB file format while drug targets can be chosen by selecting check boxes. The drug targets are classified into essential pathways. The docking jobs can be queued in the server by ‘Submit’ button.
This module allows to perform docking calculations of user's interest. The protein and ligand PDB format file has to be submitted and grid size and grid center can be specified. The xys coordinates of an atom in the center of the active site can be chosen as grid center. Grid box should be large enough to allow ligand to rotate freely. The default distance between grid points is 0.375 Å. A 40 × 40 grid size means a box of 15 Å. If user doesn’t specify any grid size the default size of 40 × 40 points is chosen.

Please specify a Ligand: 

Please specify a Protein: 

Please specify Grid Size: 

Please specify Grid Centre:

If you encounter any unexpected behaviour, please let us know!!

**Fig 6.3** Snapshot of the Mydock window. The user has to submit ligand and protein file of choice in PDB format; the grid size and grid centre parameters can be provided optionally. This module will automatically prepare all the input files and perform docking using Lamarckian Genetic Algorithm.
Fig 6.4 The server provides a job identification ID on each docking job submission as shown in upper panel. This job ID can be submitted to result module to retrieve the results (shown in lower panel).
Fig 6.5 Snapshot of the sample output window. The server provides lowest docked energy conformation of the ligand-protein complex with corresponding docking energies as output. The corresponding control docking scores are also listed for comparison. The acceptable docking energy should be less than or around control docking energy. The acceptable docking scores are green colored.
6.2.1 Potential Drug Target Dataset

The major source of 3D coordinates of target structures available in Is-it?: is the Protein Data Bank (PDB). Redundant entries of same proteins were incorporated to model different conformations of the same active site to account the protein flexibility and a potential ligand-induced fit. The proteins for which 3D coordinates were not available were modeled using Modeller9v2 (Marti-Renom, et al. 2000). Comparative modeling based on more than 30% sequence identity is now approaching its natural template-based limits and further improvements require the development of effective refinement techniques capable of driving models toward native structure (Barcellos, et al., 2008). Currently, the parameters of about 110 selected potential drug targets have been optimized. For *M. tuberculosis* these have been selected by performing a metanalysis on the available datasets and are described in the Target module of the server. The *M. tuberculosis* drug targets implemented in this server for docking were deduced based on various studies encompassing the drug targets uniqueness, i.e., absence in host (Anishetty, et al., 2005); various knockout studies; and prioritization based on persistence and metabolic chokepoint analysis (Hasan, et al., 2006) and essentiality in terms of *in-vitro* and *in-vivo* growth and survival (Sassetti, et al., 2003) of the bacterium during different stages of infection with in its host. The datasets (unique, prioritized and essential) were collected and common proteins were extracted as final potential set of drug targets. This set of around 126 proteins was then searched for 3-dimensional structures in PDB (http://www.rcsb.org/pdb). The proteins whose structures were not present in PDB were modeled based on the homologous structures. Others proteins with no homologous structure present, were excluded from the final set. Missing residues and atoms of each protein structure were repaired using the Biopolymer module of Sybyl 7.1 (M/s. Tripos Inc.), and polar hydrogens and Kollman charges were assigned to the protein(s) using ADT GUI (http://mgltools.scripps.edu/). The proteins implemented in the server can be grouped in different pathways as shown in Figure 6.6.

For bacterial, fungal, malarial and other pathogenic protozoan targets, these have been largely polled from the Potential Drug Target Database (Zhenting, et al., 2002) and Therapeutic Target Database (Chen, et al., 2002) databases.
Fig 6.6 The *M. tuberculosis* drug targets implemented in the server. The proteins are grouped into the related pathways. The ‘Other class’ groups those proteins which are essential but cannot be grouped into any pathway currently. The sources of these protein structures are either X-ray or NMR solved structures or homology models. Redundant or duplicate conformations of these proteins either with different ligands or inhibitors and open or closed forms of the enzymes were incorporated as well to model the protein flexibility of the enzyme.

6.2.2 Analyzing Output

The output is a table/list of lowest docking energy along with the corresponding docked complex for the submitted jobs and a sample output is show in Figure 6.5. The results also includes a precalculated comparative score vis-à-vis known ligand or inhibitors, based on Autodock scoring function, to evaluate theoretical affinity of the resulting complexes. Only the lowest energy conformation calculated for each compound is displayed. Compounds that bind equally well as the positive control or better can be considered as a potential hit. In the absence of known inhibitors/ligands, the compounds with the lowest energy can be considered as a potential hit. Protein-ligand interactions
Chapter 6

Is-it?: Web Server

can be visualized through any of the freely available molecular visualizers like SPDBV (Guex, et al., 1999) or Rasmol (Sayle, et al., 1995).

The Docking energies are calculated using free energy scoring function that is based on a linear regression analysis calibrated with protein-ligand complexes with known inhibition constants, with molecular mechanics terms from AMBER force field. This scoring function can be given by following equation.

\[
\Delta G_{\text{bind}}^{\text{eq}} = W_{\text{vdW}} \sum_{i=1}^{n} \sum_{j>i}^{n} \left( \frac{A_{ij}}{r_{ij}^6} - \frac{B_{ij}}{r_{ij}^8} \right) + W_{\text{hbond}} \sum_{i=1}^{n} \sum_{j>i}^{n} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{\text{elec}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{\text{tor}} N_{\text{tor}} + W_{\text{sol}} \sum_{i=1}^{n} \sum_{j>i}^{n} \left( S_i V_j + S_j V_i \right) e^{-r_{ij}/2\sigma^2},
\]

where \( W_{\text{vdW}}, W_{\text{hbond}}, W_{\text{elec}}, W_{\text{tor}}, \) and \( W_{\text{sol}} \) are weighting factors of van der Waals, hydrogen bond, electrostatic interactions, torsional term, and desolvation energy of inhibitors, respectively. \( r_{ij}, A_{ij}, \) and \( C_{ij}, \) and \( B_{ij} \) and \( D_{ij} \) represent the interatomic distance, the depths of energy well, and the equilibrium separations between the two atoms, respectively. The hydrogen bond term has an additional weighting factor, \( E(t), \) representing the angle-dependent directionality. With respect to the distant-dependent dielectric constant, \( \varepsilon(r_{ij}), \) a sigmoidal function proposed was used in computing the interatomic electrostatic interactions between a receptor protein and its ligands. In the entropic term, \( N_{\text{tor}} \) is the number of sp\(^3\) bonds in the ligand. In the desolvation term, \( S_i \) and \( V_i \) are the solvation parameter and the fragmental volume of atom \( i, \) respectively (Park, et al. 2006).

The standard error in docking energy calculation is about 2.5 kcal/mol. This is enough to discriminate between leads with milli-, micro- and nanomolar inhibition constants.

6.3 Web server evaluation

This server is based on Autodock4 which is a validated and widely used docking tool. To test the reliability of the Is-it?: server, the candidate binding proteins were searched for a
series of hydroxymates and aryl amino derivatives. The results were confirmed by \textit{in vitro} and \textit{in vivo} inhibition assays.

\subsection*{6.3.1 Potential Binding proteins for Aryl Amino Derivative}

The ligand 3D structure was prepared and optimized using Builder module of InsightII (M/s. Accelrys Inc.). The structure of the compound is shown in Figure 6.7. The ligand file was then uploaded into \textit{Is-it?}: and docking jobs were queued by selecting all the 85 drug targets from \textit{M. tuberculosis} enlisted in the server. The top candidates identified by \textit{Is-it?}: server, ranked by interaction energies, included 10 out of the 12 targets from FasII pathway (Fig 6.8). It is thus likely that Aryl Amino Derivatives may interact with multiple proteins including fabG or 1UZL (star marked in the Fig 6.8), Enoyl-acp reductase (1P44, 2NSD, inhA) and fabD (2QC3) from FasII pathway. The compound also showed high affinity with another protein Gyrase or gyrB (another star marked in the figure). The assays carried out by Dr. B N Singh’s group, Division of Microbiology, Central Drug Research Institute, Lucknow, India; confirmed that the compound is an inhibitor of FasII pathway.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{fig6_7.png}
\caption{Structure of aryl amino compound: 3-(4-Chloro-phenyl)-1-[4-(3-dimethylamino-propylamino)-phenyl]-propenone.}
\end{figure}
Fig 6.8 The plot of docking energy calculated by Autodock for aryl amino derivative and the protein targets. Only acceptable scores are used for plot. The most potential binders are 1UZL (a fasII pathway enzyme) and gyrB (Gyrase B) and are denoted by a star.

6.3.2 Potential binding proteins for Hydroxymates

The hydroxymates are known inhibitors of both *P. falciparum* peptide deformylase (PDB: 1RL4) and *Lactococcus lactis* 6-Phosphogluconate dehydrogenase (PDB: 2IZ0) enzymes (Robien, *et al.*, 2004; Sundaramoorthy, *et al.*, 2007). A virtual library of aromatic hydroxymates was designed based on these known inhibitors using FragDB, a fragment library described in Chapter 3. Docking calculations were performed by submitting jobs to *Is-it?:* server against all the protein targets. Despite of their small molecular weights, these compounds showed significant affinity in terms of docking energies with almost 50% of the drug targets. The docking energies for a single compound (H3 of Chapter 2) calculated by *Is-it?:* server are shown in Figure 6.9. The top candidates identified by the server, ranked by interaction energies, were then experimentally assessed for inhibition with NAD⁺-dependent DNA Ligase (LigA) from *M. tuberculosis*, *Plasmodium falciparum* peptide deformylase (PfPDF) and *Trypanosoma brucei* 6-Phosphogluconate dehydrogenase (TbPDH) assays. The compound synthesis
was carried out by Dr. R. P. Tripathi's group, the LigA enzymatic assays were carried out by our group while the PfPDF and TbPDH assays were carried out by Dr. R. Tripathi's group, Division of Parasitology, Central Drug Research Institute, Lucknow, India. These enzyme assays confirmed the inhibition and some of the hydroxymate compounds (H series) has already been described as LigA inhibitors in Chapter 2.

6.4 Conclusion

Is-it?: is a valuable web server that can rapidly predict potential inhibitors for a given drug target as well as the target pathway. This approach of docking against a set of proteinaceous targets from different pathways is very useful in identifying potential chemical scaffolds that can inhibit multiple targets. This information so obtained can provide a framework for filtering and designing the chemical libraries for virtual screening experiments.

6.5 Availability

Fig. 6.9 The docking energies better than known ligand or inhibitor calculated for a single molecule of hydroxymate derivatives are shown. The stars are marked for ligA, 1RL4 and 1PGJ for which this compound has been assayed enzymatically and confirmed inhibition.

The molecule is showing highest affinities for malarial and protozoan drug targets which are plotted in the extreme right of the graph viz. 1TV5, 1ONP, 1U5C, 1RL4 as compared with other drug targets.
Chapter 6

6.6 References


Sybyl 7.1 (2004) TRIPOS Inc, 1699 South Hanley Road, St. Louis, Missouri 63144, USA

