STRUCTURE-BASED DRUG DESIGN: A SIMPLIFIED INSILICO PROTOCOL

Rahul Kumar

University School of Chemical Technology, Guru Gobind Singh Indraprastha University
Dwarka Sector 16C, New Delhi-India

ABSTRACT

Insilico drug discovery is crucial in the current scenario where a large number of compounds are present and testing them individually for each disease or disorder will be highly costly and inefficient. Insilico drug discovery has become the backbone of drug design process because of quick and reliable results and past success stories. With the advancement in technology and era of supercomputing, all the extensive simulations and calculations are being performed in seconds and minutes. Structure-based drug design process is one of the Insilico methods of drug discovery. The review summarises the concept of using ΔG based on the complexation energy of the target protein with a particular chemical compound or protein. Based on ΔG of the complexation energy types of interaction are obtained. Structure-based drug discovery also uses the concept of analogue drug design in order to increase the binding efficiency and increase the total negative ΔG, which in turn increases overall reaction constant (k). The review summarises the algorithm to develop and design a novel drug in a simplified, yet in a sophisticated manner, which requires critical analysis at every point of the algorithm. From selection of target protein to designing of new drug and the following algorithm can be used to improve the efficiency of the existing drugs and their efficiency with their target protein.

Keywords: Insilico, Drug Design, Reaction Constant, Complexation Energy.

INTRODUCTION

The frequency of occurring of new disease or any genetic changes in the present strains has increased dramatically over the past decade. Execution of wet lab experiments for a particular target protein with individual chemical compounds will be highly time-consuming, inefficient and expensive in nature. Discovery of novel medical compounds having high standards of safety and therapeutically efficient is required. This is where Insilico methods of drug discovery come into action providing cost-efficient, time-saving and therapeutically efficient screening of the compounds, which can be considered for further experiments. The increase in the knowledge and availability of biologically detailed structural data has provided a huge opportunity and gained attention across the globe. The foundation and success stories were laid back in 1990-1994, where the peptide-based design for HIV proteinase, an inhibitor of HIV-1 protease and development and design of an effective oral HIV protease inhibitor were major contributions in structure based drug discovery, which corroborates the importance of structure-based drug discovery. During the period, 1990-1994 Insilico methods were a concept and structure-based discovery was executed as wet lab
experiments. Modernization of equipment’s and the era of data have led to a revolution in the design of drugs computationally.

Advantages of using Insilico Drug Design techniques:
1. Shorter time, as compared to wet lab experiments.
2. Act as screening section to identify lead compounds.
3. Inexpensive and efficient depending upon the selection of mathematical model and approach.
4. Large libraries of compounds can be screened further for experiments in minutes and hours.

METHODOLOGY AND ALGORITHM

The process of Insilico structure-based drug discovery is an iterative one. The Process is shown below in Figure 1:

**Identification of target protein**
Target identification is the most crucial. In this step, the identification of the disease and the protein related to the biological activity of it. The target protein is selected for the lead drug design.

**Protein structure retrieval and modelling**
Target protein PDB structure is obtained from the protein database. Protein database available are:
Table 1: Databases to obtain protein structure in PDB format and sequence retrieval

<table>
<thead>
<tr>
<th>Database</th>
<th>Web link</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCSB Protein Data Bank</td>
<td><a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a></td>
</tr>
<tr>
<td>Uniprot</td>
<td><a href="http://www.uniprot.org/">http://www.uniprot.org/</a></td>
</tr>
<tr>
<td>DDBJ</td>
<td><a href="http://www.ddbj.nig.ac.jp/">http://www.ddbj.nig.ac.jp/</a></td>
</tr>
<tr>
<td>EMBL-EBI</td>
<td><a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a></td>
</tr>
</tbody>
</table>

In case, the protein structure is not available, Protein modelling is possible based on homology modelling with the proteins already present in the database.

Protein modelling

Protein Modelling servers are available to model your protein structure based on the input protein sequence. If directly protein sequence is not available, obtain the genomic sequence for the respective protein and translate the sequence into a respective protein sequence. Online servers like Viral Genome ORF (VIGOR) http://www.jcvi.org/vigor/submission.php is available for protein translation based on ORF. Fasta Format of the genome is selected and pasted in the input box as shown in figure 2 and the job is submitted to the server5-6.

![Figure 2: VIGOR Interface](image)

Results were retrieved and the required amino sequence is selected for further modelling. After the sequence is obtained protein structural, modelling is performed based on the percent identical and homology related. PHYRE 2 Fig.3 [http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index ]7 and SWISS-Model Fig.4 ([https://swissmodel.expasy.org/ ]8) is intensively used server for protein modelling. Amino acid sequence of the required protein obtained from the VIGOR is submitted to any of the available online servers for structural modelling. Protein structures will be modelled based on quality score (Q-score) and percent identical Score in the case of SWISS-Model and PHYRE 2 respectively. Higher the Q-score and percent identical score, accurate is the modelled structure.

Validation of Structure by Ramachandran Plot

Modelled structure in the previous step was downloaded in PDB format. In order validate the structure of the generated protein Ramachandran plot is generated which is done by SWISS-PDB Viewer ([http://spdbv.vital-it.ch/disclaim.html# ]10). Dihedral angles plotted for each amino acid residues. RC plot forecasts the stable protein structure with a number of amino acids in allowed regions in terms of phi-φ and psi-ψ for each amino acid residues. More the number of phi-φ and psi-ψ in allowed region for alpha helices and beta sheets more is the stability of the protein structure. This is done to check the plausibility of the modelled structure.
Figure 3: PHYRE 2 Interface

Figure 4: SWISS-MODEL Interface

Figure 5: I GEMDOCK Interface
Energy and Torsional Minimization

After structural validation by RC plot, the modelled protein has energy and torsional strain, which is required to be minimised in order to make the protein achieve the lowest energy state. Lower the energy state or energy more stable is the protein structure. Energy and torsional minimization were performed in SWISS-PDB Viewer. After Energy minimization, the structure is saved in PDB format.

Hit Identification and screening – Docking

The main motive of this step is to identify the backbone structure of the drug molecule i.e. molecule having a maximum affinity with the modelled target protein structure. Select the potential candidates from literature review or else to go for library screening. iGEMDOCK (Figure 5) is the best docking software, which can be installed locally to screen the drug and obtain the amino acid and interaction profile. Download the molecules from the library available and save them in pdf format or select the one mentioned in the literature for further efficiency improvement. Select the target protein and the library of molecules saved locally. Under the drug screening section, select rough docking in order to obtain lead molecules for further intensive drug screening. Start the docking of the molecules with the target protein. Obtain the interaction analysis table showing the complexation energy ΔG. Sort the results in terms of energy i.e. lowest to highest. Lower the complexation energy (ΔG) i.e. more negative more stable will be the complex (protein-molecule). Obtain the top results for further intensive analysis and efficiency improvement.

Select a datum value based on literature review of available drug ΔG value and select the molecules above the datum value for intensive docking simulation. Reselect the target protein and in prepare compounds tab select the top lead molecules above datum value and in the default setting tab select drug screening option and simulate the docking. Interaction analysis will be obtained in terms of hydrogen bonding, Van Der Waal interaction and electrostatic interaction. The absence of electrostatic interaction between the drug molecule and the target signifies weak complex which is the major reason for the failure for the drug in later clinical trials. The absence of strong interaction between the target and the molecule is must for the molecular dynamics of the complex at the later stages after the drug has been introduced. After obtaining the interaction analysis and amino acid profile select the top most drug molecule for efficiency improvement.

Drug Design

Select the top lead molecule for efficiency improvement and in amino acid profile analyse the amino acids and types of interaction that can be introduced with them. Use AVOGADRO in figure 6, (https://avogadro.cc/) to design the drug molecules based on the amino acid profile. This step is also known as analogue drug design step. In this step, we try to develop a different analogue of the molecule by introducing various functional groups, elements and try to induce other properties such as aromaticity in order to increase the efficiency of the molecule in terms of interaction with the target protein. For example, functional groups can be introduced in terms +I and –I effect of the group which will shift the electron cloud during the complex formation. Modify
and design the molecule as required based on amino acid interaction. The main goal of this step is to design a molecule such a manner that during docking simulation electrostatic attraction is generated along with H-Bonding and Vander Waal interaction which will increase the affinity of the drug molecule to multiple folds as the electrostatic interaction is one of the strongest interactions stabilising the complex.

**ADMET Analysis**

Further ADME-Toxicity analysis must be performed after the completion of above steps from PRE-ADMET online server and Drug Likeness prediction from Sanjeevini IIT-D Drug design server.

**CONCLUSION AND FUTURE PROSPECTS**

The author developed the mentioned novel protocol and the approach of study. The main cause of failure of the drug is their incomplete Insilico analysis specifically in terms of interaction profile. Weak interactions between the target and the molecule are the focal cause of failure in drug designing of novel molecules. The following protocol lays a foundation, which can be used to design and develop drug molecules at screening stage where most of the errors occur.

**REFERENCES**