REVIEW ARTICLE

An Overview of Computational Approaches in Structure Based Drug Design

Dipali Singh[*], Anushree Tripathi, Gautam Kumar

[1]Department of Bioinformatics, Indian Institute of Information Technology-Allahabad, India.
[*]Corresponding Address: Indian Institute of Information Technology-Allahabad, 211012, India.
Email: ibi2010002@iiita.ac.in, dipali.d4@gmail.com

Short title: Computational Approaches in SBDD

Abstract

Drug design is a costly and difficult process. Drug must fulfill several criteria of being active, non-toxic and bioavailable. The conventional way of synthesizing drugs is a monotonous process. But computer aided drug design is a proficient way to overcome the tedious process of conventional method. Drugs can be designed computationally by structure or target based drug designing (SBDD). This review summarizes the methods of structure based drug design, usage of related softwares and a case study that explores to find a suitable drug (lead) molecule for the mutated state of H-Ras protein in order to prevent complex formation with Raf protein.

Keywords: computer aided drug design, structure based drug design, Ras-protein

Introduction

Traditionally, new drugs were generated from plants and other natural products through accidental observations and discoveries. Leads for new drug were generated from screening of organic compounds. Increasing information on the three dimensional structure of the biological target has paved path for structure based drug design. The rapid progress in the field of genomic, proteomic, and structural biology has increased the opportunities for future drug lead discovery. The antihypertensive drug, captopril,

an angiotensin-converting enzyme (ACE) inhibitor was the first success story in structure-based drug design [1].

Kubinyi has reviewed success stories of structure based drug design in the search for new, potent and selective HIV protease inhibitors, thrombin inhibitors, neuraminidase inhibitors and integrin receptor antagonists [1]. Anderson in his review paper mentioned that two of the first drugs to reach the market using SBDD were Amprenavir and Nelfinavir developed against HIV protease [2]. Structure-

based drug design can successfully contribute to the discovery process at different stages. It can be used at a very early stage at which no leads are available [3].

1. Overview of the Process

Proteins 3D structures are generally used by SBDD to assist for the development and design of new lead (drug compounds). The overall process of SBDD (figure 1) would be divided mainly into two parts:

a. Docking Ligands

Proteins are flexible molecules and they adjust their shape to place bound ligands through rotation of bonds. SBDD allows to dock ligand/drug molecules into protein active sites and to visualize the movement that occurs in amino acid side chains.

b. Lead Optimization

Lead optimization is a technique of refining 3D structures of drug molecules and it promotes the binding of drug to protein active sites. In this technique, researches gradually modify the structure of the drug compound by docking every specific structure of a drug compound in active site of protein, and calculating their extent of interactions.
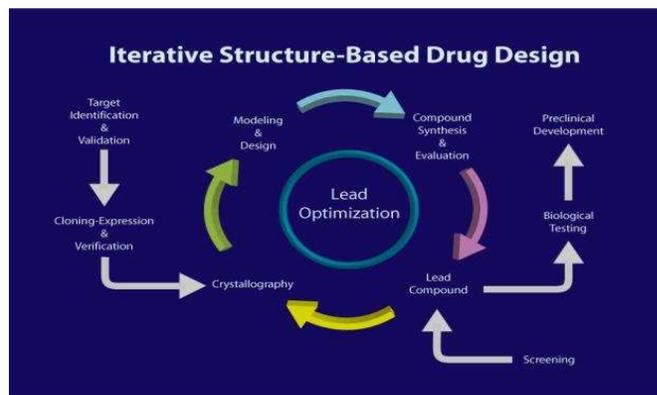

Fig.1: The Outline of Structure-Based Drug Design

2. Design Process

a. Choice of drug target

The target should be closely linked to cause of human disease and binds to a small molecule, generally a protein, in order to carry out a function. Drug target are usually protein having a well-defined binding pocket. SBDD against RNA targets with well-defined secondary structure has also been effective [2].

After the identification of target, structure can be determined following any of the methods:

1. X-ray crystallography
2. Nuclear magnetic resonance Spectroscopy (NMR)
3. Computational methods (Modelling)
4. Atomic Force Field Microscopy (AFM)

3. Drug Design Methods

Once identification of structure and target site is completed, there are number of ways to develop lead based on the structure of the target

which can be categorized as computer aided versus experimental. In experimental method, high-throughput screening is performed with combinatorial chemistry and thousands of molecules are tested for biochemical effects. Computer-aided methods can be classified into 3 categories.

a. Database searching and docking methods
b. De novo drug design methods
c. Ligand binding scoring functions

a. Database searching and docking methods

Widely used computational docking methods are DOCK, CONCORD, AUTODOCK, FLO98 and FLEXX. DOCK systematically attempts to fit each compound from a database to the target structure's binding site in such a way that in the database, three or more atoms of the molecule overlap with a set of predefined site points in the target binding site [7]. The default method for site point generation involves creating an inverse surface of the binding site. This is specified by the set of overlapping spheres that fill the binding site and touch the molecular surface at two points. The sphere centers (for all spheres with radii within a specified range) are used as site points. CONCORD is based on the combination of geometry rules and optimization methods. It selects lowest energy conformer of the molecule then scores on grid using different energy functions. On the basis

of precalculated values for protein, each match is scored on a grid throughout the binding site of target molecule [7].

b. De novo drug design methods

Structure based drug designing methods rely exclusively on ligand optimization approach based on the study of protein active site properties. There are three important categories of computational methods for the de novo design of structure based ligands: fragment positioning methods, molecule growth methods, and fragment methods coupled to database searches [6].

Fragment positioning methods

Basically these methods are based on the selection of structures of individual functional groups or fragments from predefined library which fill the active site of enzyme [5]. Two well-known programs which predict energetically favorable binding site positions for chemical fragments are GRID and MCSS (Multiple Copy Simultaneous Search). GRID calculates protein interaction energies for functional groups on a grid surrounding the target structure. It includes non-bonded interaction like hydrogen bonding, electrostatic and Van der Waals. It is mainly useful for modifying existing lead compounds. Limitation of GRID is that the sphere probe must be

capable of making hydrogen bonds and must not be in a linear arrangement [6]. In MCSS method, the probes are fully flexible and individual atoms are represented by CHARMM potential energy function. The de novo drug designing approach involves three steps [6]. Initially it uses fragment positioning method. Secondly, clustering and connecting the optimally placed molecular fragments to form chemically sensible candidate ligands. Finally, depicts the binding of proposed compounds with another and to existing drugs.

Molecule growth methods

A fragment is fitted in the binding site of the target structure while ligand molecule is successively built by bonding a further fragment to it. There are various molecule growth methods are available, including SMoG (Small Molecule Growth), GrowMol, GroupBuild and GenStar.

SMoG uses simple model for ligand-protein interactions as well as a knowledge-based potential. A large number of structures are statistically analyzed by an efficient Monte Carlo molecular growth algorithm that generates molecules through the adjoining of functional groups directly in the binding region [7].

GrowMol generates ligand structures from a library of atom as well as small functional group types and is scored based on its chemical complementarities with nearby atoms to the binding site of the target. GroupBuild is similar to GrowMol, it uses a predefined library of chemical fragments and scores candidate fragment positions depending on force field to get candidate small molecule ligands fragment by fragement. GenStar generates chemically reasonable structure which fills active site of enzyme. The proposed molecules provide good steric contact with the enzyme and also exist in low energy conformation. These structures consist of sp^3 hybridized carbons which are grown sequentially, but which can also branch or form rings. Atoms are grown from predocked inhibitor core. For each new atom generated by the program, several hundred candidate positions representing a range of reasonable bond lengths, bond angles, and torsion angles are considered. Then, each candidate is scored, with a simple enzyme contact model. From the highest scoring cases, positions are chosen at random. Duplicate structures may be removed applying variety of criteria. Energy of compounds may be minimized and displayed using standard modeling programs.

Fragment methods coupled to database searches

It is an integrated approach for fragment

positioning methods and database searching techniques either to extract those existing molecules from a database that can be docked with the preferred fragments in their most favorable positions into the binding site or for de novo design. HOOK generates a database of molecular skeletons without involving functional groups on the database molecules and then fit molecular skeletons into the target binding site such that two MCSS functional group minima can be hooked by using docking. Then, it undergoes geometrical superposition of two designated hooks in the skeletal molecules and finally using inverted Lennard-Jones type contact potential, the fit of the skeleton in the binding site in two functional group minima, is scored. After validating scores, secondary searches are carried out to attach additional MCSS minima to the skeleton, if fit is acceptable [6]. CAVEAT is comparatively faster method due to consideration of interaction between the skeletal molecule and the binding site in post processing step. It is similar to HOOK in that it involves searching of a database of three-dimensional structures of small cyclic molecules to connect optimally placed fragments in the binding site. In the database, specific bonds of each molecule are represented as vectors, and the molecule is specified as a set of pairwise combinations of bond vectors. It

finds matches between pairs of bond vectors from the fragments of the query molecules and the database molecules [6].

c. Ligand binding scoring functions

The ligands binding scoring functions are major determinant of the accuracy of scoring functions that ranks the lead compounds. Factors which contribute to ligand binding include hydrophobic effect, dispersion interactions, hydrogen bonding, other electrostatic interactions and solvation effects. With increasing complexity, the various approaches for estimating binding affinities include scoring functions based on statistical analysis of known structure of protein ligand complexes, physicochemical properties, force field calculations, force field calculation with added solvation corrections and free energy perturbation (FEP) calculations. SMoG pseudo energy function is a scoring function based on statistical analysis of high resolution X ray structure. Currently knowledge based, regression based and first principle based methods have been developed to rank lead compounds [8].

4. Case Study

In the Ras subfamily, mainly K-RAS, H-RAS and N-RAS codes for those proteins which are

made up of 189 amino acids with molecular weight 21kDA protein [9]. Guanosine nucleotide binding protein or G-proteins works in form of signaling switches with two states that are active and inactive. Usually it is bound to the nucleotide GDP in the inactive state. On the other hand, it is bound to GTP in the active state. Guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs) are mainly used in exchanging the bound nucleotide. Ras is associated with an intrinsic GTPase activity in which it can hydrolyze bound GTP into GDP. But, due to its less efficiency, RasGAP is needed which is formed by binding of Ras and GAP and stabilizes the Ras catalytic residues by releasing inorganic phosphate and ultimately leads to Ras molecule in GDP bound state for Ras inactivation. It has been found that mutations in the Ras family of proto-oncogenes are very commonly observed in 20% to 30% of all human tumors [10].The inappropriate activation of the gene affects malignant transformation, proliferation and signal transduction [11], due to which, the mutated Ras P21 has a structure that disables its ability to bind with GTPase activating protein (GAP) and creates an autophosphorylation site, keeping the Ras P21 in the GTP-bound activate state and contributing to a malignant cell phenotype [12, 13].

In this context, target-based drug discovery is considered to be highly potential. The mutated H-Ras is perceived to be an important target to treat colorectal and pancreatic cancer. A suitable drug (lead) molecule can be searched for the mutated state of H-Ras protein in order to prevent complex formation with Raf protein.

a. Materials and Methods

The protein structures of H-RAS P21 mutant (PDB ID - 521P) and of Ras-binding domain (PDB ID-1WXM) were taken from Protein Data Bank *http://www.rcsb.org/pdb/home/home.do.*

There were two methods used to predict potential binding site.

In the first approach, screening of ligand molecules was carried out through BLAST search engine by submitting the mutated HRas (PDB ID: 521P) protein sequence to DrugBank database:

http://redpoll.pharmacy.ualberta.ca/drugbank/drugBlast.htm. The DrugBank search showed trifluoroethanol, S-oxymathionine and isopropanol as active ligands.

In a second approach, ChemBank ligand entries were downloaded from Ligand in SDF format and entries of ligand was used for virtual screening and docking into effectors region of mutated H-Ras by using Discovery Studio/LigandFit program to identify active

Nepal Journal of Biotechnology. Dec. 2011, Vol. 2, No. 1: 53– 61

Biotechnology Society of Nepal (BSN), All rights reserved

potential drugs. The Ligand Fit docking algorithm produced 10 different hits of ligand, such as YS035, nizatidine, leuhistin, 3-aminopropanesulphonic acid, guanidine, acetamide, methoxamine, urea, aluminum fluoride and hydroxyurea from two different binding site cavities that were encompassed in effectors region of mutated H-Ras.

5. Results and Discussion

Leaving all the other molecules, the 3-aminopropanesulphonic acid was docked with energy of -0.009 kcal/mol and hydroxyurea with -3.014 kcal/mol. These two ligand molecules were also found to obey the Lipinski's rule of five. This Rule evaluates drug ability, or finds a chemical compound with some particular pharmacological properties that can make it an orally active drug in humans.

This result correlates well with earlier experimental results [14, 15] and it depicts that the identified binding conformations of these inhibitors are reliable and produce anti-tumor effects in a variety of solid tumor [16] and leukemia. 3-aminopropanesulfonic acid is a synthetic gammaaminobutyric acid (GABA) analog. Hydroxyurea is an antineoplastic agent that produces anti-tumor effects in animals and man in a various forms of solid tumor and

would be an effective drug to inhibit function of mutant H-Ras P21 protein, which will be able to arrest cell growth and cancer cell proliferation. From this study and previously reported experimental data in literature, we observe that hydroxyurea and 3-aminopropanesulphonic acid would be an effective drug to inhibit function of mutant H-Ras P21 protein, which will in turn arrest the process of cell growth and proliferation of the cancer cell [17]. It was earlier reported that the oral administration of hydroxyurea to 20 patients with chronic myelogenous leukemia, resulted in the decrease count of white blood cell [18].

Conclusion

The major goal of structure-based drug design is to develop an efficient process that involves a high resolution crystal structure of validated biological target molecules and reliably generates an easily synthesized, high affinity small molecule with desirable pharmacological properties. New advancement in the field of structural genomics, proteomics and bioinformatics will enhance variety of approaches for structure based drug design.

Acknowledgement

informative lecture on Computer Aided Drug Design course. We are also thankful to all the faculties of Department of Bioinformatics, Indian   Institute of Information Technology, Allahabad, India for introducing and enlightening us to the bioinformatics field.

References

1.   Kubinyi H: **Structure-based design of enzyme inhibitors and receptor ligands.** *Current Opinion in Drug Discovery and Development* 1998, **1**(1):4-15.

2.   Amy CA: **The Process of Structure-Based Drug Design.** *Chemistry & Biology* 2003, **10**: 787-797.

3.   Shravanti K *et al*: **A Review on Structure Based Drug Design of Protein Tyrosine Phosphatase 1B Inhibitors for Target for obesity and Type 2 Diabetes Mellitus**. *Journal of Pharmacy Research* 2010, **3**(12): 2939-2940

4.   Sistla R, Ghadiyaram C, Srinivasan NC and Subramanya HS: **A Structure based strategy for New Drug Discovery.** *Innovations in Pharmaceutical technology* 2006,August 30, **20**:18-23

5.   Sergio HR, Mark AM. **GroupBuild: A Fragment-Based Method for De Novo Drug Design**. *J.Med. Chem* 1993, **36**: 1700-1710

6.   Robert SD, Eugene IS. **SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence**. *J. Am. Chem. Soc* 1996, **118**: 11733-1174

7.   Diane JM: **Computational approaches to structure-based ligand design.** *Pharmacology & Therapeutics* 1999, **84**: 179–191.

8.   Holger G, Gerhard K: **Statistical potentials and scoring functions applied to protein–ligand binding, Current Opinion in Structural Biology**. *Elsevier* 2001, **11**(2):231-235.

9.   Valencia A, Chardin P, Wittinghofer A, Sander C: **The Ras protein family: Evolutionary tree and role of conserved amino acids**. *Biochemistry* 1991, **30**: 4637-4648.

10.   Bos J: **Ras oncogenes in human cancer: a review**. *Cancer Res* 1989, **49** (17): 4682-4689.

11.   Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J: **Chapter 25, Cancer**. *Molecular cell biology* (4th ed.). San Francisco: W.H. Freeman 2000. ISBN 0-7167-3706-X.

12.   Bos J: **The ras gene family and human carcinogenesis.** *Mutat Res.* 1988, **195**(3): 255-71

13.   Henson ES, Gibson SB: **Surviving cell death through epidermal growth factor   (EGF) signal transduction pathways: Implications for cancer therapy**. *Cell Signal* 2006, **18**: 2089-2097.

14.   Young CW, Schochetman G, Karnofsky DA: **Hydroxyureainduced Inhibition of Deoxyribonucleotide Synthesis: Studies in Intact Cells**. *Cancer Res*

1967, **27**: 526-534.

15. Akerblom L, Ehrenberg A, Graslund A, Lankinen, Reichard P, Thelander L: **Overproduction of the Free Radical of Ribonucleotide Reductase in Hydroxyurea-Resistant Mouse Fibroblast 3T6 Cells**. *Proc Natl ACAD Sci USA* 1981, **78**: 2159-2163.

16. Schwartz HS, Garofalo M, Sternberg SS, Philips FS: **Hydroxyurea: Inhibition of Deoxyribonucleic Acid Synthesis in Regenerating Liver of Rats**. *Cancer Res* 1965, **25**:1867-1870.

17. Krackoff IH, Savel H, Murphy ML: **Phase II studies of hydroxyurea (NSC-32065) in adults: clinical evaluation of Cancer**. *Cancer Chemother Rep*1964, **40**: 53-5.

18. Kennedy BJ: **Hydroxyurea therapy in chronic myelogenous leukemia**. *Cancer* 1972. **29**: 1052-1055.