IJournals: International Journal of Research in Drug & Pharmaceutical Science (IJRDPS)
ISSN: XXXX-XXXX
Volume 1 Issue 1 June 2017

A REVIEW ON INSILICO APPROACHES IN NEW DRUG DEVELOPMENT

JOLLY SAMU[1]; DEEPU RAVINDRAN[2]; AISWARYA K[3]; ANJU THOMAS[4]; ADITHYA E.K[5]

Rajiv Gandhi institute of Pharmacy, Trikaripur, Kasrargod. India.

jollyssamu@gmail.com

ABSTRACT

In **silico pharmacology** (also known as computational therapeutics, computational **pharmacology**) is a rapidly growing area that globally covers the development of techniques for using software to capture, analyse and integrate biological and medical data from many diverse sources. This article is a review on insilico approaches in new drug development.

Keywords: In silico pharmacology , drug design

1. INTRODUCTION

Drug discovery and development is a complex, lengthy process and failure of a candidate molecule can occur as a result of combination of reasons such as poor pharmacokinetics, lack of efficacy, Side effect and commercial reasons .Most drugs are discovered by either modifying the structure of known drugs, by screening compound libraries or by developing proteins as therapeutic agents. With the advent of genomics, proteomics, bioinformatics and technologies like crystallography, NMR, the structures of more and more protein targets are becoming available. So there is a need for computational tools that can identify and analyze active sites and suggest potential drug molecule that can bind to these sites. *In silico* models fill this research lacuna. Studies right from molecular docking, molecular dynamics, quantum mechanics, QSAR to ADMET prediction including dissolution studies are performed *in silico*. Availability of huge database of drugs from drug bank, protein data bank coupled with recent advances in technology further fuel the use of *in silico* models.

2. HISTORY

The expression *in silico* was first used in public in 1989 in the workshop "Cellular Automata: Theory and Applications" in Los Alamos, New Mexico. Pedro Miramontes, a mathematician from National Autonomous University of Mexico (UNAM) presented the report "DNA and RNA Physicochemical Constraints, Cellular Automata and Molecular Evolution". In his talk, Miramontes used the term "*in silico*" to characterize biological experiments carried out entirely in a computer. The work was later presented by Miramontes as his PhD dissertation.

The phrase "*in silico*" originally applied only to computer simulations that modeled natural or laboratory processes (in all the natural sciences), and did not refer to calculations done by computer generically.

3. IN SILICO DRUG DESIGN

In Silico Drug Design In silico is a term that means “computer aided”. The phrase was coined in 1989 as an analogy to the Latin phrases in vivo, in vitro, and in situ. So in silico drug design means rational design by which drugs are designed/discovered by using computational methods. Current trend in drug discovery is shifted from discovery to design, which needs understanding the biochemistry of the disease, pathways, identifying disease causative proteins and then designing compounds that are capable of modulating the role of these proteins. This has become common practice in biopharmaceutical industries. Both experimental and computational methods play significant roles in the drug discovery and development and most of the times run complementing each other. The main aim of computer aided drug design (CADD) is to bring the best chemical entities to experimental testing by reducing costs and late stage attrition. CADD involves:

a. Computer based methods to make more efficient drug discovery and development process.

b. Building up chemical and biological information databases about ligands and targets/proteins to identify and optimize novel drugs.

c. Devising in silico filters to calculate drug likeness or pharmacokinetic properties for the chemical compounds prior to screening to enable early detection of the compounds which are more

likely to fail in clinical stages and further to enhance detection of promising entities.

There are various computational techniques which are capable of producing the desired effect at various stages of the drug discovery process. The two major disciplines of CADD which can manipulate modern day drug discovery process and which are capable of accelerating drug discovery are bioinformatics and cheminformatics.

4. OVERVIEW OF THE PROCESS

The process of in silico drug design is an iterative one and often proceeds through multiple cycles before an optimized lead goes into clinical assay. The first cycle includes the cloning, purification and structure determination of the target protein or nucleic acid by one of three principal methods: X-ray crystallography, nuclear magnetic resonance (NMR), or homology modeling. Using computer algorithms, compounds or fragments of compounds from a database are positioned into a selected region of the structure (docking). These compounds are scored and ranked based on their steric and electrostatic interactions with the target site, and the best compounds are tested with biochemical assays. In the second cycle, structure determination of the target in complex with a promising lead from the first cycle, one with at least micromolar inhibition in vitro, reveals sites on the compound that can be optimized to increase potency. Additional cycles include synthesis of the optimized lead, structure determination of the new target: lead complex, and further optimization of the lead compound. After several cycles of the drug design process, the optimized compounds usually show marked improvement in binding and, often, specificity for the target .

5. *In silico* drug discovery process comprises of 3 stages

Stage 1-It involves Identification of a therapeutic target and building a heterogeneous small molecule library to be tested against it. This is followed by the development of a virtual screening protocol initialized by either docking of small molecules from the library or building these structures in the active site by employing *De novo* design methods.

Stage 2- These selected hits are checked for specificity by docking at binding sites of other known drug targets.

Stage 3-These selected hits are subjected to detailed in silico ADMET profiling studies and those molecules that pass these studies are termed as leads

6. STRATEGIES OF IN SILICO DESIGN

In silico drug design can be applied by either of two strategies of design depending on the knowledge of the target, presence of the primary sequence and 3D structure. These strategies are:

7. Structure Based Drug Design

Structure based drug design (SBDD) is one of the earliest techniques used in drug design. Drug targets are typically key molecules involved in a specific metabolic or cell signaling pathway that is known, or believed, to be related to a particular disease state. Drug targets are most often proteins and enzymes in these pathways. Drug compounds are designed to inhibit, restore or otherwise modify the structure and behavior of disease related proteins and enzymes. SBDD uses the known 3D geometrical shape or structure of proteins to assist in the development of new drug compounds. The 3D structure of protein targets is most often derived from X-ray crystallography or NMR techniques. X-ray and NMR methods can resolve the structure of proteins to a resolution of a few angstroms. However structure based drug design is not a single tool or technique. It is a process that incorporates both experimental and computational techniques. This is generally the preferred method of drug design, since it has the highest success rate. In the drug design stage of SBDD, docking is the preferred tool for giving a computational prediction of compound activity . The following steps are mostly used in SBDD:

Target Determination

7.1. Drug Target

Drug Target is a biomolecule which is involved in signaling or metabolic pathways that are specific to a disease process. Biomolecules play critical roles in disease progression by communicating through either protein–protein interactions or protein–nucleic acid interactions leading to the amplification of signaling events and/or alteration of metabolic processes. In structure based drug design, a known 3D structure of the target is the initial step in target identification. This is usually determined either by X-ray crystallography or by NMR to identify its binding site, the so called active site.

7.2. Homology Modelling

If crystallographic coordinates or a 2D NMR models are not available, then a homology model is usually the next best way for determining the protein structure. A homology model is a three-dimensional protein structure that is built up from fragments of crystallographic models. Thus, the shape of an α-helix may be taken from one crystal structure, the shape of a β-sheet taken from another structure, and loops taken from other structures. These pieces are put together and optimized to give a structure for the complete protein. Often, a few residues are exchanged for similar residues, and some may be optimized from scratch. Homology models may be very accurate or very marginal, depending upon the degree of identity and similarity that the protein bears to other proteins with known crystal structures. Since the homology model building process is dependent upon utilizing crystal structure coordinates for similar proteins (called the template), a crucial factor to consider is how similar the unknown sequence should be to the template protein. A number of metrics have been suggested for this. One of the most conservative metrics suggests that there should be over 70% sequence identity (not similarity) with the template, in order to get a homologous model that can be trusted. Other metrics suggest having over 30% or 40% sequence identity with the template. With higher sequence identities, the percent of error is decreased, where as many as 10% of homology models may have a root mean square deviation (RMSD) greater than 5A° (which represent error cutoff). Rost carried out an extensive study looking at how much sequence identity is needed to get a good homology model as a function with the number of aligned residues. For a small sequence of 25 aligned residues, 60% identity was necessary. For a large region of 250 aligned residues, templates with over 20% identity could give good homology models . The metrics used by Rost are somewhat less conservative than some of the other metrics. Rost's results also reflect improvements in homology model software and methodology compared with earlier work. Percent similarity is also a useful metric to examine. If several potential templates have essentially the same percent identity, then the one with the highest percent similarity may be chosen.

7.3. **Protein Folding**:

Another method for target identification is protein folding. This is a difficult process which starts with the primary sequence only and runs a calculation that tries an incredibly large number of conformers. This is an attempt to compute the correct shape of the protein based on the assumption that the correct shape has the lowest energy conformer. This assumption is not always correct, since some proteins are folded to conformers that are not at the lowest energy with the help of chaperones. It is also difficult to write an algorithm that can determine when disulfide bonds should be formed. So, sometimes protein folding gives an accurate model, and sometimes it gives a rather poor model. The real problem with protein folding is that there is no reliable way to tell whether it has given an accurate model. There are only some checks that provide some circumstantial evidence that the model might be good or bad. For example, one can check if hydrophilic residues are on the exterior of the protein and hydrophobic residues are on the interior. Homology model building has two important advantages over protein folding. First, it is more accurate on average. Second, and more importantly, the researcher can get a better estimate of whether the homology model is likely to be qualitatively and quantitatively accurate, based on the degree of similarity to a known structure. The role and reliability of homology model building is increasing as the number of available crystal structures increases. Knowing the three-dimensional structure of a protein is only the beginning of understanding it. It is also important to understand the mechanism of chemical reactions involving that protein, where it is expressed in the body, the pharmacophoric description, and the mechanism of binding with chemical inhibitors.

7.4. Molecular Docking

Docking is an automated computer algorithm that determines how a compound will bind in the active site of a protein. This includes determining the orientation of the compound, its conformational geometry, and the scoring (Fig. 1). The scoring may be a binding energy, free energy, or a qualitative numerical measure. In some way, every docking algorithm automatically tries to put the compound in many different orientations and conformations in the active site, and then computes a score for each. Some programs store the data for all of the tested orientations, but most only keep a

number of those with the best scores . In general, there are two key components of molecular docking, as follows: a. Accurate pose prediction or binding conformation of the ligand inside the binding site of the target protein. b. Accurate binding free energy prediction, which later is used to rank the order of the docking poses.

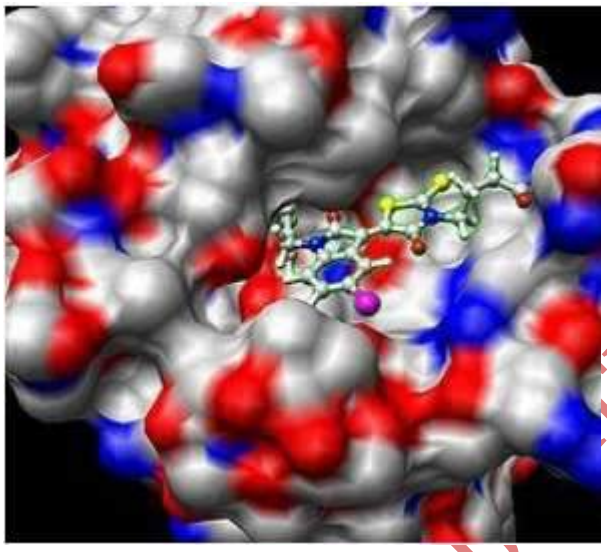

Fig1: Example of molecular docking between chemical compound and a protein.

The docking algorithm usually carries out the first part of the docking (predicting binding conformation) and the scoring function associated with the docking program carries out the second part that is binding free energy calculations.

7.5. Descriptor-based methods

A key aspect in QSAR is the use of molecular descriptors as numerical representations of chemical structures. The number and type of molecular descriptors is large and varied and thus procedures to select those that are most relevant for modelling the biological effect of interest are extremely important. Molecular descriptors are usually classified according to the dimensionality of the chemical representation from which they are computed. On this basis, one-dimensional descriptors encode numerically generic properties such as molecular weight, molar refractivity and octanol/water partition coefficient, offering a fair reflection of the size, shape and lipophilicity of molecules. Despite their low dimensionality, some of these descriptors have been associated with the drug-like character of molecules and are thus found often as biologically relevant descriptors in QSAR equations. On the other hand, 2D descriptors are computed from topological representations of molecules. The models constructed from these descriptors are habitually referred to as 2D-QSAR, a methodology widely established both in predicting physico-chemical properties as well as in providing quantitative

estimates of various biological effects.

8. ***In silico* ADMET (absorption, distribution, metabolism, excretion, toxicity) prediction**

Studies indicate that poor pharmacokinetics and toxicity are the most important causes of costly late-stage failures in drug development and it has become widely appreciated that these areas should be considered as early as possible in the drug discovery process. Combinatorial chemistry and high-throughput screening have significantly increased the number of compounds for which early data on absorption, distribution, metabolism, excretion (ADME) and toxicity (T) are needed. With use of *in silico* tools it is possible to model the most relevant pharmacokinetic, Metabolic and toxicity endpoints, thereby accelerating the drug discovery process.

9. ***In silico* prediction of drug safety**

There is considerable interest in computational models to predict drug safety in drug discovery and development. Significant adverse toxicological findings for a drug in late-stage clinical trials or postmarketing can cause enormous financial losses and place patients at risk. The earlier such molecules are identified and the drug development process halted the better. In addition, insights into the toxicological potential of a scaffold or series of structures early on in the drug discovery process could help medicinal chemists to prioritize particular scaffolds or hits. Finally, computational toxicity models could be used to help understand pre-clinical toxicity data and select appropriate experimental end-points for further studies during clinical candidate selection and early clinical studies. There are tools to predict toxicities like

(1) Genotoxicity,
(2) Liver toxicity,
(3) CYP450 inhibition and
(4) Cardiotoxicity.

10. ***In silico* prediction of drug-drug interactions**

Recently, metabolic drug–drug interactions (M-DDI) have raised some high-profile problems in drug development resulting in restricted use, withdrawal or non approval by regulatory agencies. The use of in vitro technologies to evaluate the potential for M-DDI has become routine in the drug development process. Nevertheless, in the absence of an integrated approach, their interpretation and value remains the subject of debate, and the vital distinction between a

useful ‘‘simulation’’ and a precise ‘‘prediction’’ is not often appreciated. Various *in silico* software are now available for the simulation of M-DDI.One such software is SIMCYP.

11.Applications of in silico pharmacology

The number of proteins with a known 3D structure is increasing rapidly and structures produced by structural genomics are being made available and openly accessible for use in drug discovery and development processes. In silico methods have proved useful for successful rational drug design of amyloid aggregation inhibitors as potential Alzheimer ’ s disease drugs, topoisomerase inhibitors, angiogenesis inhibitors and a Csn-B-derived Nur77 agonist as cancer therapeutics, 5-HT2C-selective ligands as schizophrenia drugs , and countless other promising examples , which are able to show how computational tools have become an integral aspect of optimizing drug design and development. A successful virtual screening and docking exercise that can be cited is the discovery of disrupting agents of MIF (macrophage migration inhibitory factor, a cytokine involved in inflammatory diseases and cancer) to its receptor CD74. To discover small molecule inhibitors of the biological activity of MIF, virtual screening was performed by docking 2.1 million compounds from the ZINC and Maybridge databases into the MIF tautomerase active site.

CONCLUSION

In the selection of new drug candidates, many efforts are focused on the early elimination of compounds that might cause several side effects or interact with other drugs. *In silico* techniques help in this regard and they are going to become a central issue in any rigid drug discovery process. In silico technology alone cannot guarantee the identification of new safe and effective lead compound but more realistically future success depend on the proper integration of new promising technologies with the experience and strategies of classical medicinal chemistry.

12.REFERENCES

[1] Ahlberg C. Visual exploration of HTS databases: bridging the gap between chemistry and biology. Drug Discov Today. 1999;4:370–376.

[2] Akamatsu M. Current state and perspectives of 3D-QSAR. Curr Top Med Chem. 2002;2:1381–1394.

[3] Albert A. Relations between molecular structure and

biological activity: stages in the evolution of current concepts. Ann Rev Pharmacol. 1971;11:13–36.

[4] Albert A. Selective Toxicity. The Physcico-Chemical Basis of Therapy. Chapman and Hall: London; 1985.

[5] Aradi I, Erdi P. Computational neuropharmacology: dynamical approaches in drug discovery. Trends Pharmacol Sci. 2006;27:240–243.

[6] Arïens EJ. Receptors: from fiction to fact. Trends Pharmacol Sci. 1979;1:11–15.

[7] Aronov AM, Munagala NR, Kuntz ID, Wang CC. Virtual screening of combinatorial libraries across a gene family: in search of inhibitors of *Giardia lamblia* guanine phosphoribosyltransferase. Anti microb Agents Chemother. 2001;45:2571–2576.

[8] Bajorath J. Integration of virtual and high-throughput screening. Nat Rev Drug Disc. 2002;1:882–894.

[9] Balakin KV, Ivanenkov YA, Savchuk NP, Ivaschenko AA, Ekins S. Comprehensive computational assessment of ADME properties using mapping techniques. Curr Drug Disc Tech. 2005;2:99–113.[

[10] Barnum D, Greene J, Smellie A, Sprague P. Identification of common functional configurations among molecules. J Chem Inf Comput Sci. 1996;36:563–571.

[11] Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets. Proteins. 2003;50:5–25.

[12] Bleicher KH, Bohm HJ, Muller K, Alanine AI. Hit and lead generation: beyond high-throughput screening. Nat Rev Drug Discov. 2003;2:369–378.

[13] Button WG, Judson PN, Long A, Vessey JD. Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics. J Chem Inf Compu Sci. 2003;43:1371–1377.

[14] Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgenson WL, Lins RD, et al. Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem. 2000;43:2100–2114.

