

Review Article- Computer Aided Drug Design

Bhakti Thorat¹, Akshada Darekar², Sanishwar Darekar³, Santosh Waghmare⁴, Dr . H.V.Kamble⁵

^{1, 2, 3, 4}Dept of pharmaceutical chemistry

⁵ Dept of pharmacology

^{1, 2, 3, 4, 5} Loknete Shri Dadapatil Pharate College of Pharmacy , Mandavgan pharata.

Abstract- Computer aided drug design (CADD) is an evolving cascade of research area encompassing many facets. Computer-aided drug design (CADD) is an exciting and diverse discipline where various aspects of applied and basic research merge and stimulate each other. The theoretical basis of CADD involves quantum mechanics and molecular modeling studies like structure based drug design; ligand-based drug design; database searching and binding affinity based on the knowledge of a biological target. In this present review we present the areas where CADD tools support drug discovery process.

Keywords- Computer aided drug design, Ligand- Based, Molecular modelling, Biological target, Drug discovery process.

I. INTRODUCTION

Computer aided drug design (CADD) provides several tools and techniques that helps in various stages of drug design thus reducing the cost of research and development time of the drug. Drug discovery and developing a new medicine is a long, complex, costly and highly risky process that has few peers in the commercial world. This is why computer-aided drug design (CADD) approaches are being widely used in the pharmaceutical industry to accelerate the process. The cost benefit of using computational tools in the lead optimization phase of drug development is substantial. The cost and time invested by the pharmacological research laboratories are heavy during the various phases of drug discovery, starting from therapeutic target identification, candidate drug discovery, drug optimization through pre-clinical and extensive clinical experiments to assess the and safety of newly developed drugs. The major pharmaceutical companies have invested heavily in the routine ultra-High Throughput Screening (uHTS) of vast numbers of 'drug-like' molecules. In parallel with this, drug design and optimization increasingly uses computers for virtual screening. Recent advancements in DNA microarray experiments explore thousands of genes involved in a disease can be used for gaining in depth knowledge about the disease targets, metabolic pathways and toxicity of the drugs. The theoretical tools include empirical molecular mechanics, quantum mechanics and, more recently, statistical mechanics. This

latest advance has permitted explicit solvent effects to be incorporated. All this work is the availability of high quality computer graphics, largely supported on workstations.

Two distinct categories of research are clearly distinguishable

- 1) Crystallography, NMR or homology modelling. A detailed molecular structure of the target macromolecule, the drug receptor, is known from x-ray.
- 2) Variable activity of otherwise similar molecules.

The target receptor binding site has properties which can only be inferred from a knowledge of the both these types of approach.

Drug Discovery Process

Drug discovery is a series of processes which when followed identify the drug compounds for the effective treatment or control of disease targets. It starts with the screening of large number of chemical compounds to optimize the disease targets. insight information about the structure of the drug receptor so that the drug molecules can be adjusted to the binding site to the binding site

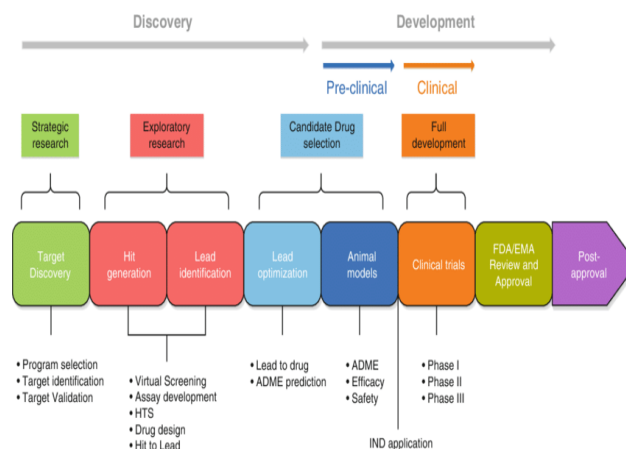


Fig No 1: Drug Discovery Process

Drug Discovery Process

Drug discovery process starts with understanding the disease for which the drug to be designed. It consists of the following steps.

1. Candidate Drug Discovery

- Selection of Therapeutic Target
- Lead Discovery.
- Lead Optimization

2. Pre clinical and clinical trials to evaluate the safety, efficacy and adverse effects of the drug

- Animal Studies
- Clinical Trials

3. FDA approval process for the newly discovered drug and bringing the drug to market for public use.

- Additional post marketing testing
- Further improvement of the drug.

In general, it takes 3-6 years for new drug discovery and pre-clinical development. The clinical trials can last up to 10 years or more before the product reaches the market. Approximately it takes 12-15 years and costs more than \$1.3 billion to bring a successful drug to market. On an average, among the 5000-10000 screened compounds about 250 compounds are selected for preclinical trials. From them only 5 survive to enter into clinical trials while only one approved by the FDA after strenuous review of the newly discovered drug.

CADD Strategies in the Drug Discovery Process .

Strategies for CADD vary depending on the extent of structural and other information available regarding the target (enzyme/receptor) and the ligands. —Direct and —indirect design are the two major modeling strategies currently used in the drug design process. In the indirect approach the design is based on comparative analysis of the structural features of known active and inactive compounds. In the direct design the three-dimensional features of the target (enzyme/receptor) are directly considered.

Working of CADD

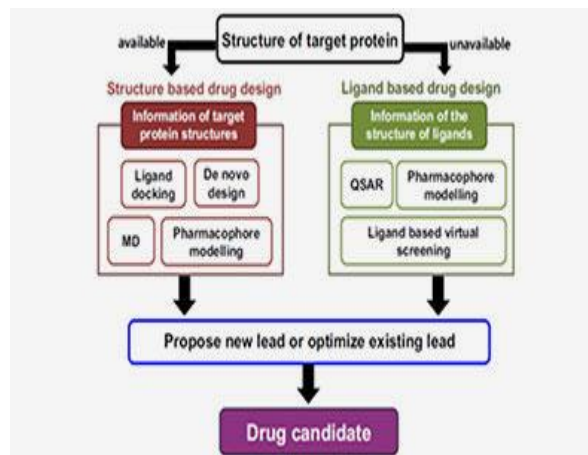


Fig no 2: Working Of CADD

Preparation of a Target Structure

Success of virtual screening depends upon the amount and quality of structural information known about both the target and the small molecules being docked. The first step is to evaluate the target for the presence of an appropriate binding pocket. This is usually done through the analysis of known target-ligand co-crystal structures or using in-silico methods to identify novel binding sites. A target structure experimentally determined through X-ray crystallography or NMR techniques and deposited in the PDB is the ideal starting point for docking. Structural genomics has accelerated the rate at which target structures are being determined. In the absence of experimentally determined structures, several successful virtual screening campaigns have been reported based on comparative models of target proteins

Homology Modeling

In the absence of experimental structures, computational methods are used to predict the 3D structure of target proteins. Comparative modeling is used to predict target structure based on a template with a similar sequence, leveraging that protein structure is better conserved than sequence, i.e., proteins with similar sequences have similar structures. Homology modeling is a specific type of comparative modeling in which the template and target proteins share the same evolutionary origin. Comparative modeling involves the following steps

1. identification of related proteins serve as template structures,
2. sequence alignment of the target and template proteins,
3. copying coordinates for confidently aligned regions
4. constructing missing atom coordinates of target structure

5. model refinement and evaluation.

Molecular dynamics-based detection

The dynamic nature of biomolecules sometimes makes it insufficient to use a single static structure to predict putative binding sites. Multiple conformations of target are often used to account for structural dynamics of target. Classic molecular dynamic (MD) simulations can be used for obtaining an ensemble of target conformations beginning with a single structure. The MD method uses principles of Newtonian mechanics to calculate a trajectory of conformations of a protein as a function of time. Classic MD methods tend to get trapped in local energy minima. To overcome this, several advanced MD algorithms such as targeted-MD, conformational folding simulations, temperature accelerated MD simulations, and replica exchange MD, have been implemented for traversing multiple minima energy surface of proteins

Monte Carlo Search with Metropolis Criterion (MCM)

MCM samples conformational space faster than molecular dynamics in that it requires only energy function evaluation and not the derivative of the energy functions. Although traditional MD drives a system toward a local energy minimum, the randomness introduced with Monte Carlo allows hopping over the energy barriers, preventing the system from getting stuck in local energy minima. MCM simulations have been adopted for flexible docking applications such as in MCDOCK.

Genetic Algorithms

Genetic algorithms introduce molecular flexibility through recombination of parent conformations to child conformations. In this simulated evolutionary process, the —fittestl or best scoring conformations are kept for another round of recombination. In this way, the best possible set of solutions evolves by retaining favorable features from one generation to the next. In docking, a set of values that describe the ligand pose in the protein are state variable. State variables may include set of values describing translation, orientation, conformation, number of hydrogen bonds, etc. The state corresponds to the genotype; the resulting structural model of the ligand in the protein corresponds to the phenotype, and binding energy corresponds to the fitness of the individual. Genetic operators may swap large regions of parents genes or randomly change (mutate) the value of certain ligand states to give rise to new individuals. Genetic Optimization for Ligand Docking (GOLD) explores full ligand flexibility with partial target flexibility using a genetic algorithm.

Scoring Functions for Evaluation of Protein Ligand Complexes

Docking applications need to rapidly and accurately assess protein-ligand complexes, i.e., approximate the energy of the interaction. A ligand docking experiment may generate hundreds of thousands of target-ligand complex conformations, and an efficient scoring function is necessary to rank these complexes and differentiate valid binding mode predictions from invalid predictions.

Force-Field or Molecular Mechanics-Based Scoring Functions

Force-field scoring functions use classic molecular mechanics for energy calculations. These functions use parameters derived from experimental data and ab initio quantum mechanical calculations. The binding free energy of protein-ligand complexes are estimated by the sum of van derWaals and electrostatic interactions. DOCK uses the AMBER force fields in which van der Waals energy terms are represented by the Lennard-Jones potential function while electrostatic terms are accounted for by coulombic interaction with a distance-dependent dielectric function.

Empirical Scoring Functions

Empirical scoring functions fit parameters to experimental data. An example is binding energy, which is expressed as a weighted sum of explicit hydrogen bond interactions, hydrophobic contact terms, desolvation effects, and entropy. Empirical function terms are simple to evaluate and are based on approximations. The weights for different parameters are obtained from regression analysis using experimental data obtained from molecular data. Empirical functions have been used in several commercially available docking suits like LUDI, FLEXX[and SURFLEX.

Knowledge-Based Scoring Function

Knowledge based scoring functions use the information contained in experimentally determined complex structures. They are formulated under the assumption that interatomic distances occurring more often than average distances represent favourable contacts. On the other hand, interactions that are found to occur with lower frequencies are likely to decrease affinity. Several knowledge based potentials have been developed to predict binding affinity like potential of.

Consensus-Scoring Functions

Consensus approaches rescore predicted poses several times using different scoring functions. These results can then be combined in different ways to rank solutions. Some strategies for combining scores include (1) weighted combinations of scoring functions, (2) a voting strategy in which cut-offs established for each scoring method is followed by decision based on number of poses a molecule has, (3) a rank by number strategy ranks each compound by its average normalized score values and (4) a rank by rank method sorts compounds based on average rank determined by individual scoring functions.

Structure-Based Virtual High-Throughput Screening

Structure-based virtual high-throughput screening (SB-vHTS), the in silico method for identifying putative hits out of hundreds of thousands of compounds to the targets of known structure, relies on a comparison of the 3D structure of the small molecule with the putative binding pocket. SB-vHTS selects for ligands predicted to bind a particular binding site as opposed to traditional HTS that experimentally asserts general ability of a ligand to bind, inhibit, or allosterically alter the proteins function. To make screening of large compound libraries within finite time feasible. SB-vHTS often uses limited conformational sampling of protein and ligand and a simplified approximation of binding energy that can be rapidly computed. The key steps in SB-vHTS are:

1. preparation of the target protein and compound library for docking
2. determining a favorable binding pose for each compound
3. ranking the docked structures.

Ligand-Based Computer-Aided Drug Design

The ligand-based computer-aided drug discovery (LBDD) approach involves the analysis of ligands known to interact with target of interest. These methods use a set of reference structures collection from compounds known to interact with the target of interest and analyse their 2D or 3D structures. The overall goal is to represent these compounds in such a way that the physicochemical properties most important for their desired interactions are retained, whereas extraneous information not relevant to the interactions is discarded. It is considered as an indirect approach to the drug discovery in that it does not necessitate knowledge of the structure of the target of interest. The two fundamental approaches of LBDD are (1) selection of compounds based on chemical similarity to known actives using some similarity measure or (2) the construction of a quantitative structure activity relationship (QSAR) model that predicts biological

activity from chemical structure. The methods are applied for in silico screening for novel compounds possessing the biological activity of interest, hit-to-lead and lead-to drug optimization, and also for the optimization of DMPK/ADMET properties. LBDD is based on the similar property principle which states that molecules that are structurally similar are likely to have similar properties. LBDD approaches in contrast to SBDD approaches can also be applied when the structure of the biological target is unknown. Additionally, active compounds identified by ligand-based virtual high-throughput screening (LB-vHTS) methods are often more potent than those identified in SB-Vhts.

Molecular Descriptors

Molecular descriptors can include properties such as molecular weight, geometry, volume, surface areas, ring content, rotatable bonds, interatomic distances, bond distances, atom types, planar and nonplanar systems, molecular walk counts, electro negativities, polarizabilities, symmetry, atom distribution, topological charge indices, functional group composition, aromaticity indices, solvation properties, and many others. These descriptors are generated through knowledge-based, graph-theoretical methods, molecular mechanical, or quantum-mechanical tools and are classified according to the dimensionality of the chemical representation from which they are computed: 1-dimensional (1D), scalar physicochemical properties such as molecular weight; 2D, molecular constitution-derived descriptors; 2.5D, molecular configuration-derived descriptors; 3D, molecular conformation-derived descriptors. These different levels of complexity, however, are overlapping with the more complex descriptors, often incorporating information from the simpler ones. Software for General Purpose Molecular Modeling.

II. CONCLUSION

Computer aided drug design (CADD) is a multidisciplinary field attracting the researchers from information technology, medicine, pharmacology etc. to discover new tools and techniques or enhance the available tools and techniques to assist in drug discovery process. These techniques proved to be effective in various stages of drug discovery process thus reducing both cost and time taken for developing a drug than conventional methods. Various CADD tools that assist during the process of drug development are provided with few examples of the drugs that are available in the market which were successfully designed using these tools. These tools can be used, enhanced to assist the various phases of drug discovery.

REFERENCES

- [1] Whittaker. The role of bioinformatics in target validation. Drug Discovery To-Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Medical Journal of Australia, 2004; 181: 293-4.
- [2] Lengauer. Bioinformatics. From Genomes to Drugs. Wiley- VCH, Weinheim, Germany, 2002.
- [3] Lipinski. Lead and drug-like compounds: the rule-of-five revolution. Drug Discovery Today: Technologies, 2004; 1(4): 337-341.
- [4] Leeson P D, Davis A M, Steele J. Drug-like properties: Guiding principles for design or chemical prejudice? Drug Discovery Today: Technologies, 2004; 1(3): 189-195.
- [5] Hou T. Xu X. Recent development and Application of Virtual Screening in Drug Discovery: An Overview. Current Pharmaceutical Design, 2004; 10: 1011-1033.
- [6] Klebbe G. Lead Identification in Post-Genomics: Computers as a Complementary Alternative. Drug Discovery Today: Technologies, 2004; 1(3): 225-215.
- [7] Gisbert Schneider. Uli Fechner. Computer-based de novo design of drug-like molecules. Nature.Reviews. Drug Discovery, 2005; 4(8): 649-663.
- [8] Butte A. The use and analysis of microarray data. Nature Reviews Drug Discovery, 1(12): 951-960.
- [9] Richards W. G. Computer-Aided Drug Design. Pure and Applied Chemistry, 1994; 6(68): 1589-1596.
- [10] Kitchen D B. Decornez H. Furr J R. Bajorath J. Docking and scoring in virtual screening for drugdiscovery: methods and applications. Nature reviews in drug discovery, 2004; 3: 935-949.
- [11] DiMasi J A. Grabowski H G. The cost of biopharmaceutical R&D: is biotech different? Managerialand Decision Economics, 2007; 28: 469-479.
- [12] Hajduk PJ, Huth JR, and Tse C. Predicting protein druggability. Drug. Discov. Today, 2005; 10:1675-1682
- [13] Fauman EB, Rai BK, and Huang ES. Structure-based druggability assessment-identifying suitable targets for small molecule therapeutics. Curr. Opin. Chem. Biol., 2011; 15: 463-468
- [14] Laurie AT, Jackson RM. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Curr. Protein. Pept. Sci., 2006; 7: 395-406.
- [15] Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, Cheruku S, Heifetz A, Mohanty P Fichman M, Sharadendu A. An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. J. Med. Chem., 2006; 49: 3116-3135
- [16] Warner SL, Bashyam S, Vankayalapati H, Bearss DJ, Han H, Mahadevan D, Von Hoff DD, HurleyLH. Identification of a lead small-molecule inhibitor of the Aurora kinases using a structure-assisted, fragment-based approach. Mol. Cancer. Ther., 2006; 5: 1764-1773.
- [17] Budzik B, Garzya V, Walker G, Woolley-Roberts M, Pardoe J, Lucas A, Tehan B, Rivero RA, andLangmead CJ. Novel N-substituted benzimidazolones as potent, selective, CNS-penetrant, and orallyactive M(1) mAChR agonists. Med. Chem. Lett., 2010; 1: 244-248.
- [18] Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, and Jones DT. Protein annotationand modelling servers at University College London. Nucleic. Acids. Res., 2010; 38: 563-568.
- [19] Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, and Sali A. Comparative proteinstructure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct., 2000; 29: 291-325.
- [20] Schlitter J, Engels M, Krüger P. Targeted molecular dynamics: a new approach for searchin pathways of conformational transitions. J. Mol. Graph, 1994; 12: 84-89.
- [21] Grubmüller H. Predicting slow structural transitions in macromolecular systems:
- [22] Conformational flooding. Phys. Rev. E. Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics, 1995; 52:2893-2906. 22. Abrams CF, Vanden-Eijnden E. Large-scale conformational sampling of proteins usingtemperature-accelerated molecular dynamics. Proc. Natl. Acad. Sci. U.S.A., 2010; 107: 4961-4966.
- [23] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem.Phys. Lett., 1999; 314: 141-151.
- [24] Liu M, Wang SM. MCDOCK: a Monte Carlo simulation approach to the molecular dockingproblem. J. Comput. Aided. Mol. Des., 1999; 13: 435-451.
- [25] Jones G, Willett P, Glen RC. A genetic algorithm for flexible molecular overlay andpharmacophore elucidation. J. Comput. Aided. Mol. Des., 1995; 9: 532-549.
- [26] Halgren TA. Merck molecular force field. 1. Basis, form, scope, parameterization, and performanceof MMFF94. J. Comput. Chem., 1996; 17: 490-51
- [27] Böhm HJ. The computer program LUDI: a new method for the de novo design of enzyme inhibitors.J. Comput. Aided. Mol. Des., 1992; 6: 61-78.
- [28] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incrementalconstruction algorithm. J. Mol. Biol., 1996; 261: 470-489.
- [29] Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-basedsearch engine. J. Med. Chem., 2003; 46: 499-511.

- [30] Shimada J, Ishchenko AV, Shakhnovich EI. Analysis of knowledge-based protein-ligand potentials using a self-consistent method. *Protein. Sci.*, 2000; 9: 765-775.
- [31] Velec HFG, Gohlke H, Klebe G. DrugScore (CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.*, 2005; 48: 6296-6303.
- [32] DeWitte RS, Shakhnovich E. SMOG: De novo design method based on simple, fast and accurate free energy estimates. *J. Am. Chem. Soc.*, 1997; 119: 4608-4617
- [33] Mitchell JBO, Laskowski RA, Alex A, Forster MJ, Thornton JM. BLEEP-Potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comput. Chem.*, 1999; 20: 1177-1185.
- [34] Feher M. Consensus scoring for protein-ligand interactions. *Drug. Discov. Today*, 2006; 11: 421-428.
- [35] O'Boyle NM, Liebeschuetz JW, Cole JC. Testing assumptions and hypotheses for rescoring success in protein-ligand docking. *J. Chem. Inf. Model*, 2009; 49: 1871-1878.
- [36] Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, Cheruku S, Heifetz A, Mohanty P, and selective amidosulfonamide 5-HT_{1A} agonist (PRX-00023) for the treatment of anxiety and depression. *J. Med. Chem.*, 2006; 49: 3116-3135.
- [37] Johnson MA, Maggiora GM. *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990.
- [38] Stumpfe D, Bill A, Novak N, Loch G, Blockus H, Geppert H, Becker T, Schmitz A, Hoch M, Kolanus W. Targeting multifunctional proteins by virtual screening: structurally diverse cytohesin inhibitors with differentiated biological functions. *Chem. Biol.*, 2010; 5: 839-849
- [39] Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, 1988; 110: 5959-5967.
- [40] Kiaris H, Spandidos DA. Mutations of Ras Genes in Human Tumors. *International Journal of*