DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5411

ISSN No. 0976-5697

Volume 9, No. 1, January-February 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

SEARCH FOR IN-SILICO APPLICATIONS IN DRUG DISCOVERY AND APPLICATIONS OF DIFFERENT DISCIPLINES IN IT: A SURVEY

Parthajit Roy
Department of Computer Science
The University of Burdwan
West Bengal, India-713104

Swati Adhikari
Department of Computer Science
The University of Burdwan
West Bengal, India-713104

***Abstract*:** The present paper surveys on different areas in designing of a new drug that can be performed by means of in-silico methods. Diverse applications of different subjects like Biology, Chemistry, Mathematics, Statistics, Physics etc. in different stages of human drug designing process have also been reviewed in this paper from computational points of view.

Keywords: Biology, Chemistry, Mathematics, Statistics, Physics etc.

I. INTRODUCTION

Designing drugs for a specific disease is a process whose outcome is a new suitable drug for that disease. The overall drug designing process is time consuming, requires huge spaces to store biological data and very costly. It takes several years, approximately ten to fifteen years, for a drug to be available in the market. Modern drug development strategies try to minimize this time and also try to make the process space and cost effective by applying computational techniques with traditional methods. Application of computer science in drug development processes helps in addressing these issues.

The notion of drug development relates to the field of biology and chemistry, in general. But, the research areas in drug design can not only be limited to these two fields. It can be broadened to other fields like mathematics, statistics, physics and moreover computer science. When information system is combined with these fields, they are collectively known as bioinformatics, cheminformatics, pharmacology, etc. Computer aided drug development process assembles many scientists from different subject areas to work collaboratively.

The overall drug development process is composed of several stages. Modern drug development process starts with identification of drug targets followed by validation of these targets, discovery of lead drugs and optimization of lead drugs. After this, optimized lead drugs go for pre-clinical and clinical testing. Finally, the new drugs are brought to the market.

Aforesaid fields of study can be applied to each of these stages of drug design to gear up the process. The present paper discusses about the contribution of these fields in each and every stage towards design of a new drug only from computational end and also highlights the areas of human drug designing process where in-silico methods can be applied in order to turn the whole development process time and cost effective.

The rest of the paper is arranged as follows: Section II defines the basic biological terminologies. Section III presents detailed discussion of the computer-aided drug design (CADD) process. The computational tasks those are to be performed by CADD process are given in Section IV. Contributions of different disciplines in CADD process are discussed in Section V. Conclusion is drawn on Section VI. Next the references are given.

II. BASIC BIOLOGICAL TERMINOLOGIES

Definitions and descriptions of some of the terms related to the context of this survey paper are given in this section.

Chromosome: The genetic information is packed into a thread-like structure in nucleus of each cell of any living organism. This structure is known as chromosome. The major components of each chromosome are nucleic acids and proteins.

Nucleic Acid: Transfer of genetic information from one generation to the next generation is carried out through nucleic acids. There exist two types of nucleic acids, namely, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Nucleic acids are composed of nucleotides whose constituents are four types of nitrogen bases, a phosphate group and a 5-carbon sugar. Among the four types of nitrogen bases in DNA, adenine (A) and guanine (G) are purines with two fused rings; cytosine (C) and thymine (T) are pyrimidines with a single ring. The nitrogen bases of RNA are same as that of DNA except for thymine which is uracil (U) in this case.

There are two strands of DNA, namely, 5'-3' and 3'-5' which run in opposite direction and hold together by base pairing of the hydrogen bonds between purines and pyrimidines. According to the base pairing rule, an A and a G in one strand are always paired with a T and a C on the other strand of the DNA respectively. Base pairing between a DNA strand and a RNA strand is performed in the same way as that of base pairing between two strands of DNA; only difference is that an A in DNA strand is paired with a U in RNA strand. That means, each strand of DNA form a character string consisting of A, T, G and

C; ordering of these bases in the string is known as DNA sequence. RNA sequence is series of A, U, G and C. These two sequences are responsible for transformation of genetic information into proteins in chromosomes. Fig. 1 shows parts of two strands of DNA sequence and one strand of RNA sequence.

Protein: Protein is a sequence of twenty types of amino acids. The transformation of information from DNA to protein is a two step process, namely, transcription and translation. In the first step, the information stored in DNA transferred to messenger RNA followed by translation of this information into proteins in the second step. Complementary base pairing between bases of DNA and transcribed RNA are done in transcription. Conversion of genetic information into amino acids is done with the help of code table known as codon. The procedure for synthesis, modification and regulation of proteins is known as protein expression.

Gene: In each chromosome, the genetic information lies in the form of genes which are made up of DNA. The overall procedure of transcription and translation of genes into proteins is known as gene expression. Not all parts of a gene are translated to proteins. It has two parts fragmented throughout the sequence as protein coding and non-coding parts. Numbers of bases in a gene sequence vary in sizes from some hundreds to some millions. Number of genes in humans is also huge.

Gene Mapping: Gene maps are used to get information about how genes are arranged on a chromosome. The specific location of genes on a chromosome is called its locus. It is used to find distances between genes on a chromosome. Gene mapping is a technique which is used to find locus of a gene. It is also helpful for prediction of inheritance types of distinguishing features of a living organism. This helps in understanding of disease related characteristics.

Gene Expression Profile: A test is done to identify all genes in a cell that are taking part into generation of messenger RNA as it is responsible for translating genetic information into proteins. This test is known as gene expression profile. Analysis of this profile has many therapeutic uses. It is used for diagnosis of a disease and also helps in checking the response of body towards a treatment.

Protein Structure: There are four levels in protein structure which are primary, secondary, tertiary and quaternary.

Protein sequences are available with a chain like structure (polypeptide chain) in which lots of amino acids (20 types) are joined together. A water molecule is lost due to the joining of amino acids. That means, protein sequence is actually formed with amino acid residues. In a protein sequence there are more than 50 amino acid residues. Two ends of polypeptide chain are known as N-terminus or amino terminus (on the left end) and C-terminus or carboxyl terminus (on the right end). This structure of protein is considered to be the primary structure. In this structure amino acids are joined together by means of peptide bonds. Fig. 2 shows the fraction of primary protein structure where three letter abbreviation for amino acid residues like Cys for Cysteine, Ala for Alanine, Glu for Glutamate etc. are used.

Secondary protein structures are of two types – α-helix and β-sheet. Hydrogen bonds between the hydrogen atom of N-terminus and oxygen atom of C-terminus on main peptide chain are used to define the secondary structure.

Three-dimensional structures of protein are its tertiary and quaternary structures. To carry out biological function, proteins are needed to be in their three-dimensional shape. Protein folding is a method which brings proteins into their final three-dimensional functional shape. All the levels of protein structure are essential for protein folding.

The methods X-ray crystallography, Nuclear Magnetic Resonance (NMR), etc. are used to determine structure of proteins.

Knowledge of protein structure is required for development of drug compound.

Sequence Alignment: It is a technique to find similar sequences that helps in identifying sequences with similar functionality or identical structures. In this technique, sequences are compared base by base in case of nucleic acids and amino acid residues are compared in case of proteins. Matching between any two sequences may be partial (local alignment) or they may be fully matched (global alignment). Sequence alignment is used to recover structural similarity by finding similar sequences as that of the sequence in which part of the sequence is missing. This is beneficial for treatment of diseases.

Sequence Analysis: It is a process through which it is possible to understand features, structural information, functionality and evolutionary characteristics of nucleic acid and protein sequences. The task of sequence analysis is to know the order of nucleotides or amino acid residues in a DNA, RNA or protein sequence, searching of biological databases and sequence alignment. It is also helpful for creation of a DNA sequence by aligning and joining number of DNA pieces. Sequence analysis has many applications. One of them is in treatment of genetic diseases.

III. COMPUTER AIDED DRUG DESIGN (CADD)

Drug design is an innovative process that finds new medicines or drugs for diseases. This process uses knowledge of biological target in designing new drugs. Extraction of drugs from plants is a traditional way. Modern drug discovery methods consider drug as a chemical or biological substance that has medicinal uses.

The traditional drug design techniques are based on the study of molecular biology [1], system biology [2] [3] and cell biology [4] and all of these techniques are full of time consuming, expensive physical experiments. Modern drug discovery methods have replaced these physical experiments with computational search and as a result the cost and time of the overall process are also become lower.

Computerization of drug designing process has many contributions in modern drug designing field [5] [6]. While developing new drugs, the development process needs to handle lots of raw biological data. The volume of

these data is huge. Some computational approaches towards biological data or computerization of these data also help to store and maintain them in manageable form.

A. *Biological Target (Drug Target)*

Biological target, also referred to as drug targets, are those small molecules in host organisms in which pathogens continue to live. Any disturbance in the functioning of these molecules will destroy the survival environment of the pathogens. These target molecules may be receptor, enzyme, protein and nucleic acid and are responsible for progression of disease with some desirable therapeutic functions or unwanted harmful functions. Drugs are designed in an aim to change the behavior of the target molecules by binding the drug to the target and as a result the pathogens will be no more.

B. *Different Types of CADD Process*

CADD process comes with two forms. These are structure based or direct drug design and ligand based or indirect drug design [7]. When the three dimensional structure of drug targets are available, structure based designing process is used, otherwise, ligand based designing process is used.

1. Structure Based Drug Design Process: This process is decomposed into number of stages [8]. Two major stages prior to any pre-clinical and clinical tests are target selection and lead compound selection (potential drug). Different stages of structure based drug design process are shown in Fig. 3.

Target Selection: This stage is used to select probable drug target for a specific disease. This stage is basically consisting of two steps – target identification and target validation. Numbers of in-silico methods are available to perform these two steps successfully [9].

Target Identification: In this step, molecular targets those are causes for progress of disease are identified. This step performs lots of computations on biological sequences like query processing, sequence alignment and sequence analysis. This step also performs gene selection related to disease, analyzes the genes those are related to drug action, screens poisonous side effect of genes, does functional prediction, gene and protein annotation and prioritization, collects structural information or data about gene and protein expression, compares two or more sequences, analyzes gene expression profiles, maps information and differentiates between healthy and diseased cells. X-ray crystallography, nuclear magnetic resonance (NMR), homology modeling and protein folding methods are used in target identification to determine three dimensional structures of proteins as well as their binding or active sites. Homology Modeling is a technique that generates a model for three-dimensional structure of target protein sequences based on structural similarity with known protein sequences. It is an iterative process.

Target Validation: In the second step, identified targets are verified for their therapeutic advantages for patients and are selected as targets for drug i.e. it is tested to see whether the identified targets are capable of producing desired clinical results or not. It is an improvement or a reduction step. All of the identified targets are not selected as drug target. Some percentages of them are selected based on their priority. Some of the computation tasks to be performed in this step are mapping of genetic network, protein-protein interactions, predictions of sub-cellular localization etc.

Lead Compound Selection: The selection process of lead compound is again a two steps process namely, lead identification and lead optimization.

Lead Identification: In this step, a chemical compound is identified that shows biological or pharmacological behaviour towards a drug target and that compound is medicinally beneficial. Computer-aided techniques like protein crystallography, nuclear magnetic resonance (NMR), de novo design and computerized searches of structural databases to study the existing drug's pharmacophore, known as virtual screening, helps in identifying suitable lead compounds. Virtual screening is used in the scoring, prioritizing and filtering of a numbers of structures that use computer programmes. The task of de novo design is to design new molecules based on three-dimensional structure of a target.

Lead Optimization: In this step, identified lead compounds are tested for their effectiveness, toxicity and absorption power towards a disease and corrective steps are taken accordingly followed by which potential drug is selected. Molecular docking, a computer algorithm, is used to determine how a lead compound will bind to the active site of a target protein [7]. It is used to test how two molecular structures, one for lead compound and the other for target protein, fit together i.e. protein-drug interaction.

2. Ligand Based Drug Design Process: In this process, the knowledge of molecular structure of small molecules those are responsible for biological or pharmacological functioning of the molecules (known as pharmacophore of the molecules) when tied up with the drug targets is considered. The small molecule is termed as ligand. This process is useful for deriving a model (pharmacophore model) for drug target (when structural properties of target are missing) based on the structural information of the molecule that binds to the target. Ligand based drug design process starts with the identification of pharmacophore of a ligand after selection of drug target. Next, based on this pharmacophore, the molecular structure of the ligand is modified iteratively so that it is best fitted for the biological target and treated as potential drug. After which pre-clinical and clinical tests are performed on this potential drug compound. Fig. 4 shows the steps of ligand based drug design process in-between drug target selection and pre-clinical test.

3D-QSAR (three-dimensional quantitative structure-activity relationships) is a computational method that is used in the process of ligand based drug design. The quantitative relationship between the favourable or unfavourable effects of a group of compounds and their three-dimensional features are studied by using the 3D-QSAR technique that facilitate in understanding of the new chemical compound to be treated as drug.

Another method that is used in ligand based drug design process is based on the structural and physical similarities between ligand and known drugs in an expectation to have similar binding properties of the ligand as that of known drugs.

After successful selection of potential drug, it undergoes for pre-clinical (animal) and clinical (human) testing followed by prediction of drug-drug- interaction. The results of interaction between two or more drugs are required when they are applied together. By their combined application one drug may influence the activities of another drug to a great extent. After having positive results out of these steps, the final product of drug is obtained and it is marketized.

C. *ADMET Properties of Drugs*

Prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of a drug is very essential in drug development process. Determination of optimal ADMET properties in pre-clinical test of potential drugs allows concentrating on limited number of them and assures their success in clinical test.

Absorption: Due to absorption, it is possible to know how a drug dissolves in blood after entering into the body.

Distribution: Distribution makes it possible to determine the movement of drug from organ to organ through the blood.

Metabolism: Due to metabolism, the chemical structure of a drug is altered inside the body.

Excretion: Excretion relates to the removal of drug from the body.

Toxicity: Toxicity means poisonous effects of a drug.

These are required to determine the proper dose and timing of a drug. Different in-silico tools for ADMET prediction have also been available [10] [11].

D. *Drug Repurposing*

The inputs of existing drugs are also taken into consideration while designing new drugs. Among the existing drugs, some drugs may have more side effects than others. So, they may not be used in treatment of their intended diseases but may be safe and suitable for treatment of new diseases. Drug repurposing is a technique in which knowledge of existing drugs is studied thoroughly in designing drugs for new diseases. This technique searches for those drugs among the existing ones that can be reused with slight alterations in their structures, doses and timings.

E. *Biomarker*

Biomarker (Biological Marker) is a feature of a biological molecule that can be present in blood or in other body fluids or in tissues and shows the normal or diseased condition of the body. This molecule can be genetic or biochemical characteristic or any other substance that helps in identification of a disease. Biomarker is used to observe the response of a body towards a certain treatment or in evaluation of normal or pathogenic processes. It is used in all of the above mentioned phases of CADD process. It has many clinical uses.

IV. COMPUTATIONAL TASKS TO BE PERFORMED BY CADD PROCESS

A. *Database Management in CADD*

Computerization of biological data results in creation of various biological and chemical databases.

Computer aided drug design process also needs to handle biomedical data and drug data [12]. Biomedical data are received from different pharmaceutical companies, hospitals, nursing homes and clinical laboratories in large volume and with higher dimensionality. These data may also be received from any public network. Drug data may include sequence data, gene expression data, protein-drug or protein-protein or drug-drug interaction data and data of some other types like patients record either in the form of electronic data or in the form of report. That is big data analytics are associated with these data for their systematic management [12].

Numbers of software, tools and databases are available to facilitate the drug development process. Database preparation is a fundamental and an important task in all the stages of drug design process. Bioinformatics and cheminformatics have made it possible to create databases for storing structural information and for various biological sequences of different organisms as well as for biomedical or drug data. The overall process needs to handle different types of biological and chemical databases for genomes, proteins, amino acids or nucleic acid, different types of databases for storing annotation, sequence, structural and functional information or some other types of information. Different information obtained from pre-clinical and clinical studies of potential drugs are also made available in later time by creation of respective databases. Some of these databases are GenBank, EMBL, GEO (Gene Expression Omnibus), etc. All these databases can be retrieved from the server of National Centre for Biotechnology Information (NCBI) [13]. One such web server is developed by Ying Liu, *et al*. that performs biological sequence alignment [14]. List of some public domain databases in medicinal field can be found in [15]. Drug design process starts after identification of a disease by searching disease databases. KEGG [16] and MalaCards [17] are two databases for storing information related to human diseases. Bioinformatics and cheminformatics tools are available to create different medicinal databases. SWEETLEAD is an cheminformatics database [18] whereas ChEMBL is an bioinformatics database [19] to be used for drug designing purpose. Some online databases like BindingDB and ChEMBL [20] are also available for the same purpose. Creations of more advanced databases and web servers using bioinformatics or cheminformatics tools have become an important research area in the field of drug discovery.

B. *Use of Database Searching Tools in CADD*

With the development of databases, there is also need for generation of database searching tools. BLAST is one such tool [21]. BLAST finds local alignment between sequences. There are different types of BLAST tools like Nucleotide BLAST, Protein BLAST etc. Some of the research works regarding construction of robust searching tools for different biological databases can be found in [22] - [25].

C. *Big Data Analytics in CADD*

Big data problem associated with these databases has also been able to draw the attention of the researchers and remedy of this problem is again use of bioinformatics and cheminformatics algorithms. Some of the existing works relating to solution to this problem are given in [26] – [29]; these works are using machine learning approach, artificial intelligence, pattern matching and the software Hadoop to solve the big data problem. Big data also helps in selecting drug target [30] and in virtual screening [31]. A method for identification of drug target path on biomedical data has been given in [32].

The computerization of drug design process is not only restricted to storing and maintenance of biological data. It is supposed to do any kind of task that needs computation. It is applicable in diverse functioning of different phases of drug design process as stated in subsection B. It is also applicable in prediction of ADMET properties of drug, identification of biomarkers and drug repurposing. Fig. 5 depicts different in-silico tasks of CADD process.

V. CONTRIBUTION OF DIFFERENT DISCIPLINES IN CADD

Each and every step of the drug development process opens a new door to the research zone in computer aided drug design. The subject areas covered by this process mainly include bioinformatics [33], cheminformatics [34], pharmacokinetics [35] and pharmacodynamics [36].

Bioinformatics is a field of study that analyzes biological data using mathematics, statistics and computer science [37] whereas cheminformatics is the field of study that uses computer and information systems to solve chemical problems [38]. Bioinformatics concentrates on collection, storage, inspection and controlling of biological data whereas cheminformatics does the same for chemical data. Bioinformatics is used to select drug target and helps in the screening process of the candidate drug; not only that, it also helps in determining side effects of a drug and in predicting drug resistance [33]. Structural bioinformatics, a branch of bioinformatics, also have a large contribution in drug discovery. It becomes helpful in analysis and prediction of three dimensional structures of proteins or nucleic acids. The research work of D. K. Brown and O. T. Bishop discusses the role of structural bioinformatics in drug discovery that uses computational SNP analysis [39]. The use of cheminformatics tools to select lead compounds has been discussed in [34].

Both of bioinformatics and cheminformatics need to do the tasks of pattern recognition and data mining for clinical data throughout the entire process of CADD. Different machine learning approaches are used to identify drug target [40]. Performances of machine learning techniques applied in solving protein folding problem are measured in [41]. Algorithms for ligand based virtual screening using machine learning approaches are discussed in [42]. Clustering techniques are also applied on chemical data for the purpose of analyzing these data [43]. Graph theoretic approaches are also applicable in the process of drug design. These techniques have been applied in ligand based drug design [44] and structural analysis of protein active sites [45]. Sometimes bioinformatics techniques are integrated with cheminformatics techniques, known as biochemoinformatics, to design more robust drug [46].

Computational biology has many applications in drug discovery. Bioinformatics tools have made it easy to use the features of computational biology and this has been discussed in [47]. Not only computational biology, other subjects like biophysics, sociology, biotechnology have many applications in computational drug design which will be discussed later in this section.

Pharmacokinetics is the study of predicting ADME power of a drug over time i.e. reaction of body on a drug. It helps in deciding the proper dose and safety of a drug [48]. On the other hand, pharmacodynamics is the study of a drug effects on its targets i.e. to the body which depends on dose and time of the drug. Pharmacology is the study of drugs that combines the areas of pharmacokinetics and pharmacodynamics. Pharmacokinetics study is generally used at the end of drug development process. Recently, prediction of ADME properties is performed at the very beginning of the drug development process in order to eliminate molecules with low ADME properties. Pharmacological studies are performed in target and ligand screening, drug repourposing and clinical testing of a drug. Some of the computer based methods regarding these uses of pharmacology can be found in [49] – [52].

Omics studies such as genomics, proteomics, transcriptomics and metabolomics play significant roles in the above mentioned stages of drug development as well as in the stages of pre-clinical and clinical testing [53]. It is also used in drug repurposing and in identification of biomarker. Each of these fields is an individual research area. The demand for computational omics studies, either structure based or function based, in drug designing is increasing rapidly. Several bioinformatics and cheminformatics tools are available to assist in different functioning of computational omics studies. Some of the research works in this area have been shown in [54] – [65]. The method given in [54] helps in prediction of drug-protein interaction and the methods presented in [55] use data mining techniques in analyzing protein-protein interaction data collected from biological studies. In [56], computational methods are used in structural genomics and in [57], the methods of computational proteomics or lipidomics to be used for drug design are reviewed. Computational methods for predicting protein structure to be applicable in drug design process are discussed in [58]. Different aspects of proteomics in CADD process have been discussed in [59] [60]. The work as given in the paper [61] shows the use of transcriptomics in identification of a lead compound. Application of metabolomics in drug designing process has been shown in [62] [63]. Application of machine-learning in metabolomics and identifying drug-drug interaction has been shown in [64]. Omics data mining has also an application in drug repurposing [65].

Another field of study, known as pharmacogenomics, has also become effective in selection of optimal drug, its dose, its treatment process and its side effects [66]. It is the combined study of pharmacology and genomics i.e. it can be said that it relates to the part that a genome plays in response to a drug. The task of

computational pharmacogenomics has been outlined in a chapter of the book [67].

Biophysics also has a great contribution in developing a new drug. Biophysical technologies like X-ray crystallography, nuclear magnetic resonance spectroscopy, surface plasmon resonance spectroscopy etc. are considered to be key components of drug discovery [68]. Biophysics is also involves in automated drug design by predicting automated structure and annotation of proteins [69]. Numbers of research works have already been done in this area. Some of them are given in [70] – [72] which are capable of predicting protein-protein interaction and drug-target interaction.

Uses of sociological studies have several advantages in drug development [73]. The biotechnological and genetic engineering methods also play important role in pharmaceutical industry which is responsible for final marketization of newly discovered drugs along with other different kind of tasks. The research works as described in [74] and [75] show these features. Prediction of protein structure and protein sub cellular localization can be performed very well using bioinformatics tools and these are also major tasks in biotechnology. That means, bioinformatics tools are capable of performing biotechnological functions and functions of pharmaceutical science in turn.

So, from the above discussions it is clear that the process of drug discovery does not constitute a single discipline rather it is combination of numbers of disciplines like biology and biophysics – used to identify biological targets; pharmacology and chemistry – used for prioritization or validation of drug targets and for selecting lead compounds; all of chemistry, pharmacology and toxicology - used for pre-clinical testing; pharmaceutical science – used to produce final medicinal product of drug; medicine – used for clinical testing of a drug. These disciplines are assisted by other disciplines that perform computational tasks like mathematics, statistics, computer science and information system. Fig. 6 shows this fact.

VI. CONCLUSION

While keeping in mind the contributions of bioinformatics, cheminformatics, pharmacology, biophysics and sociology in drug development process, other fields like biotechnology, genetic engineering, medicinal industry, and pharmaceutical industry have also adapted these fields for discovery of new drugs. That is, none of the subject can solely claim the ownership of discovery of new drugs. It is the product of combined effect of all of these subjects and this makes this field a challenging research area.

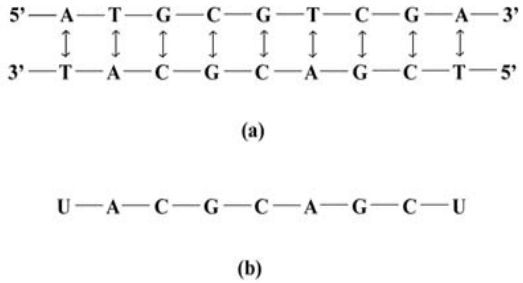

Fig. 1: (a) Parts of two strands of DNA sequence showing complementary base pairing between nucleotides; (b) One strand of RNA sequence paired with 5'-3' DNA strand (on top) in (a).

Fig. 2: Part of primary protein sequence

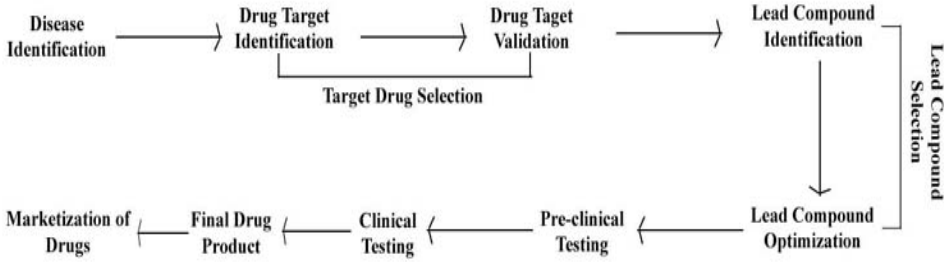

Fig. 3: Stages of structure based DD process

Fig. 4: Stages of ligand based DD process

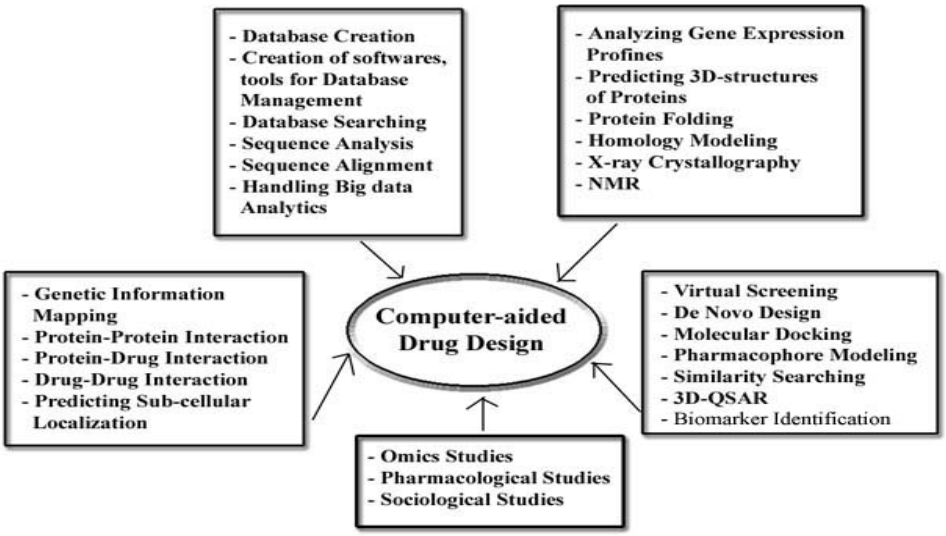

Fig. 5: In-silico tasks to be performed in CADD process

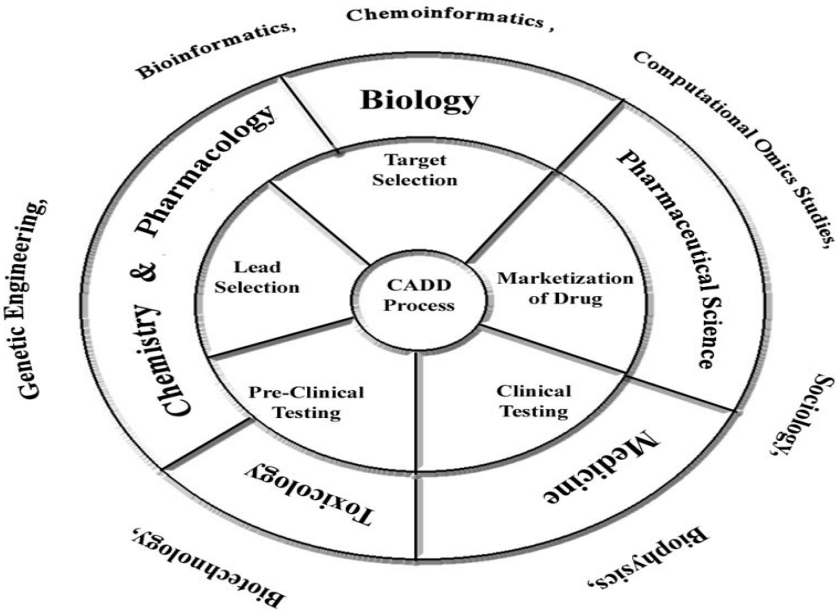

Fig. 6: Different disciplines applied in CADD process

ACKNOWLEDGMENT

Authors express their gratitude to the Department of Computer science, The University of Burdwan.

REFERENCE

[1] Belfield G. P. and Delaney S. J., "The impact of molecular biology on drug discovery," Biochem Soc Trans., vol. 34, no. Pt 2, 313-6, April 2006.

[2] Eugene C Butcher, Ellen L Berg and Eric J Kunkel, "Systems biology in drug discovery," Nature Biotechnology, vol. 22, no. 10, October 2004.

[3] Qing Yan (ed.), "Systems biology in drug discovery and development: methods and protocols," Methods in Molecular Biology, vol. 662, edn. 1, Humana Press, Springer Science-Business Media, LLC 2010

[4] Sorger PK and Schoeberl B., "An expanding role for cell biologists in drug discovery and pharmacology," Drubin DG, ed. Molecular Biology of the Cell. vol. 23, no. 21, pp. 4162-4164. 2012.

[5] Stephani Joy Y. Macalino, Vijayakumar Gosu, Sunhye Hong and Sun Choi, "Role of computer-aided drug design in modern drug discovery," Archives of Pharmacal Research, vol. 38, no. 9, pp. 1686 - 701, September 2015.

[6] Srivastava P. and Tiwari A., "Critical role of computer simulations in drug discovery and development," Curr Top Med Chem, vol. 17, no. 21, pp. 2422 – 2432, 2017.

[7] Le Anh Vu, Phan Thi Cam Quyen and Nguyen Thuy Huong, "In-silico drug design: prospective for drug lead discovery," International Journal of Engineering Science Invention, vol. 4, no. 10, pp. 60-70, October 2015.

[8] Amy C. Anderson, "The process of structure-based drug design," Chemistry & Biology, vol. 10, pp. 787 – 797, September 2003.

[9] Darryl Leon and Scott Markel, "In-silico technologies in drug target identification and validation," CRC Press, Taylor & Francis Group, Boca Raton, FL, 2006.

[10] Gautier Moroy, Virginie Y. Martiny, Philippe Vayer, Bruno O. Villoutreix and Maria A. Miteva, "Toward in-silico Structure-Based ADMET prediction in drug discovery," vol. 17, no. 1-2, pp. 44-45, January 2012.

[11] Alessandra Roncaglioni, Andrey A. Toropov, Alla P. Toropova and Emilio Benfenati, "In-silico methods to predict drug toxicity," Current Opinion in Pharmacology, vol. 13, no. 5, pp. 802 – 806, October 2013.

[12] Keith C.C. Chan, "Big data analytics for drug discovery," in Proceedings of the 2013 IEEE International Conference on Bioinformatics and Biomedicine, Shanghai, China, pp. 1 - 1, 2013, doi: 10.1109/BIBM.2013.6732448

[13] https://www.ncbi.nlm.nih.gov/

[14] Ying Liu, Khaled Benkrid, AbdSamad Benkrid and Server Kasap, "An FPGA-based web server for high performance biological sequence alignment," in Proceedings of the IEEE Nasa/ESA Conference on Adaptive Hardware and Systems, 2009. AHS 2009, San Francisco, CA, USA, 2009.

[15] George Nicola, Tiqing Liu and Michael Gilson, "Public domain databases for medicinal chemistry," Journal of Medicinal Chemistry, vol. 5, no. 16, pp. 6987 – 7002, August 2012.

[16] Kanehisa M. and Goto S., "KEGG: Kyoto Encyclopedia of Genes and Genomes," Nucleic Acids Research, vol. 28, no. 1, pp. 27 – 30, January 2000.

[17] Noa Rappaport, Michal Twik, Inbar Plaschkes, Ron Nudel, Tsippi Iny Stein, Jacob Levitt, Moran Gershoni, C. Paul Morrey, Marilyn Safran and Doron Lancet, "MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search," Nucleic Acids Research, vol. 45, no. Database Issue, pp. D877 – D887, January 2017.

[18] Novick PA, Ortiz OF, Poelman J, Abdulhay AY and Pande VS, "SWEETLEAD: An in-silico database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery," PLoS ONE, vol. 8, no. 11, e79568, November 2013.

[19] Anna Gaulton, Louisa J. Bellis, A. Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Ivonne Light, Shaun McGlinchey, David Michalovich, Bissan-Al-Lazikani and John P. Overington, "ChEMBL: a large scale bioactivity database for drug discovery," Nucleic Acids Research, vol. 40, no. Database Issue, pp. D1100 – D1107, January 2012.

[20] Wassermann A. M. and Bajorath J., "BindingDB and ChEMBL: online compound databases for drug discovery," Expert Opin Drug Discovery, vol. 6, no. 7, pp. 683 – 7, July 2011.

[21] Altschul S. F., Gish W., Miller W., Myers E.W., Lipman D.J., "Basic local alignment search tool," J Mol Biol, vol. 215, pp. 403 – 410, 1990.

[22] Francesco Napolitano, Roberto Tagliaferri and Pierre Baldi, "A scalable reference-point based algorithm to efficiently search large chemical databases," in Proceedings of the IEEE 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, pp. 1 – 6, 2010.

[23] Anna L. Buczak, Charles Wan and Glenn Petry, "SmartPortal for Biomedical data mining," in IEEE Symposium on Computational Intelligence and Data Mining, 2007. CIDM 2007, Honolulu, HI, USA, pp. 221 – 227, 2007.

[24] Bashir Sadjad and Zsolt Zsoldos, "Toward a robust search method for the protein-drug docking problem," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 8, no. 4, pp. 1120 – 1133, July-August 2011.

[25] Fang X. and Wang S., "A web-based 3D-database pharmacophore searching tool for drug discovery," J Chem Inf Comput Sci, vol. 42, no. 2, pp. 192 – 8, March – April 2002.

[26] S. K. Halgamuge, "Knowledge discovery in biological big data using near unsupervised learning: keynote presentation," in Proceedings of the 2014 IEEE 9th International Conference on Industrial and Information Systems (ICIIS), Gwalior, pp. 1 – 1, 2014.

[27] X. Chen, H. Chen, N. Zhang, J. Chen and Z. Wu, "OWL reasoning over big biomedical data," in Proceedings of the 2013 IEEE International Conference on Big Data, Silicon Valley, CA, pp. 29 - 36, 2013.

[28] S. H. Cheng, Y. S. Chiu, S. Y. Dai and H. I. Hsiao, "Duplicate drug discovery using Hadoop," in Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, pp. 24 – 26, 2014.

[29] B. Al Kindhi and T. A. Sardjono, "Pattern matching performance comparisons as big data analysis recommendations for Hepatitis C Virus (HCV) sequence DNA," in Proceedings of the 2015 IEEE 3rd

International Conference on Artificial Intelligence, Modelling and Simulation (AIMS), Kota Kinabalu, pp. 99 - 104, 2015.

[30] B. Chen and A. J. Butte, "Leveraging big data to transform target selection and drug discovery," Clinical Pharmacology & Therapeutics, vol. 99, no. 3, pp. 285 – 297, March 2016.

[31] R. Mennour and M. Batouche, "Drug discovery for breast cancer based on big data analytics techniques," in Proceedings of the 2015 IEEE 5th International Conference on Information & Communication Technology and Accessibility (ICTA), Marrakech, pp. 1 – 6, 2015.

[32] F. Du, T. Li, Y. Shi, L. Song and X. Gu, "Drug target path discovery on semantic biomedical big data," in Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, pp. 3381 – 3386, 2016.

[33] Xuhua Xia, "Bioinformatics and drug discovery," Curr Top Med Chem, vol. 17, no. 15, pp. 1709 – 1726, Jun 2017.

[34] Firdaus Begam and J. Satheesh Kumar, "A study on chemoinformatics and its applications on modern drug discovery," Procedia Engineering, in Proceeding of the International Conference on Modeling Optimization and Computing – (ICMOC – 2012), vol. 38, pp. 1264 – 1275, 2012.

[35] Ruiz-Garcia, Bermeio M., Moss A. and Casabo V. G., "Pharmacokinetics in drug discovery," J. Pharm Sci, vol. 97, no. 2, pp. 654 – 90, February 2008.

[36] James M. Gallo, "Pharmacokinetic/Pharmacodynamic-driven drug development," Mt Sinai J. Med, vol. 77, no. 4, pp. 381 – 388, July – August 2010.

[37] Can T, "Introduction to Bioinformatics," Methods Mol Biol, vol. 1107, pp. 51 – 71, 2014.

[38] Prakash N and Gareja DA, "Cheminformatics," J Proteomics Bioinform, vol. 3, pp. 249 - 252, August 2010.

[39] David K. Brown and Ozlem Tastan Bishop, "Role of structural bioinformatics in drug discovery by computational SNP analysis: analyzing variation at the protein level," Global Heart, vol. 12, no. 2, pp. 151 – 161, June 2017.

[40] K. Y. Hsin, H. Kitano, Y. Matsuoka and S. Ghosh, "Application of machine leaning approaches in drug target identification and network pharmacology," in Proceedings of the IEEE 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, pp. 219 – 219, 2015.

[41] M. A. Khan, M. A. Khan, Z. Jan, H. Ali and A. M. Mirza, "Performance of machine learning techniques in protein fold recognition problem," in Proceedings of the IEEE 2010 International Conference on Information Science and Applications, Seoul, Korea (South), pp. 1-6, 2010.

[42] K. Babaria, S. Ambegaokar, S. Das and H. Palivela, "Algorithms for ligand based virtual screening in drug discovery," in Proceedings of the IEEE 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, pp. 862 - 866, 2015.

[43] M. G. Malhat, H. M. Mousa and A. B. El-Sisi, "Clustering of chemical data sets for drug discovery," in Proceedings of the 2014 IEEE 9th International Conference on Informatics and Systems, Cairo, pp. DEKM-11-DEKM-18, 2014.

[44] H. Palivela, C. R. Nirmala and D. R. Kubal, "Application of various graph kernels for finding molecular similarity in ligand based drug discovery," in Proceedings of the 2017IEEE 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, pp. 1-8, 2017.

[45] N. Weskamp, E. Hullermeier, D. Kuhn and G. Klebe, "Multiple graph alignment for the structural analysis of protein active sites," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 4, no. 2, pp. 310-320, April-June 2007.

[46] Leming Shi, Zhenqiang Su, Aihua Xie, Chenzhong Liao, Wei Qiao, Dajie Zhang, Song Shan, Desi Pan, Zibin Li, Zhiqiang Ning, Weiming Hu and Xianping Lu, "An integrated Biochemoinformatics System for drug discovery," in: Xing W.L. and Cheng J. (eds) Frontiers in Biochip Technology. Springer, Boston, MA, pp. 191 – 206, 2006, ISBN: 978-0-387-25568-2.

[47] "Bioinformatics and Computational Biology in drug discovery and development," William T. Loging (ed.), Cambridge University Press, 2016, ISBN: 9780521768009

[48] Simon Pacey, Paul Workman and Debashis Sarkar, "Pharmacokinetics and Pharmacodynamics in drug development," M. Schwab (ed.), Encyclopedia of Cancer, Springer-Verlag, Berlin, Heidelberg, pp. 2845 – 2848, 2011.

[49] Honorio K. M., Moda T. L. and Andricopulo A. D., "Pharmacokinetic properties and in-silito ADME modeling in drug discovery," Med Chem, vol. 9, no. 2, pp. 163 – 76, March 2013.

[50] S. Ekins, J. Mestres and B. Testa, "In-silico pharmacology for drug discovery: Applications to Targets and Beyond," Br J Pharmacol, vol. 152, no. 1, pp. 21 – 37, September 2007.

[51] S. Ekins, J. Mestres and B. Testa, "In-silico pharmacology for drug discovery: methods for virtual ligand screening and profiling," Br J Pharmacol, vol. 152, no. 1, pp. 9 – 20, September 2007.

[52] Hodos R. A., Kidd B. A., Shameer K., Readhead B. P. and Dudley J. T., "In-silico methods for drug repurposing and pharmacology," Wiley Interdiscip Rev Syst Biol Med, vol. 8, no. 3, pp. 186 – 210, May 2016.

[53] Majid Y. Moridani, Robyn P. Araujo, Caroline H. Johnson and John C. Lindon, "The –omics in drug development," Bonate P. and Howard D. (eds.), Pharmacokinetics in Drug Development, Springer, Boston, MA, 2011, ISBN: 978-1-4419-7936-0.

[54] Q. Gu, Y. Ding, T. Zhang and T. Han, "Prediction drug-target interaction networks based on semi-supervised learning method," in Proceedings of the 2016 IEEE 35th Chinese Control Conference (CCC), Chengdu, pp. 7185 - 7188, 2016.

[55] Z. Nafar and A. Golshani, "Data mining methods for protein-protein interactions," in Proceedings of the IEEE 2006 Canadian Conference on Electrical and Computer Engineering, Ottawa, Ont., pp. 991 - 994, 2006.

[56] Sharon Goldsmith-Fischman and Barry Honig, "Structural genomics: computational methods for structure analysis," Protein Science, vol. 12, no. 9, pp. 1813 – 1821, September 2003.

[57] Mishra NK and Shukla M, "Application of computational proteomics and lipidomics in drug discovery," J Theor Comput Sci, vol. 1:105, no. 1, 2014.

[58] Nishant T, Sathish Kumar D and VVL Pavan Kumar A, "Computational methods for protein structure prediction and its application in drug design," J Proteomics Bioinform, vol. 4, pp. 289 - 293, 2011.

[59] Dubey R. D., Paroha S., Wani V. K., Pandey A. K., Verma S., Daharwal S. J.,, Dewangan D. and Prasad Reddy S. L. N., "Proteomics in computer-assisted molecular design," International Journal of PharmTech

Research, vol. 3, no. 1, pp. 50 – 57, January-March 2011.

[60] Goh W. W. and Wong L., "Computational proteomics: designing a comprehensive analytical strategy," Drug Discovery Today, vol. 19, no. 3, pp. 266 – 74, March 2014.

[61] Bie verbist, Gunter Klambauer, Liesbet Vervoort, William Talloen, The QSTAR Consortium, Ziv Shkedy, Olivier Thas, Andreas Bender, Hinrich W. H. Gohlmann and Sepp Hochreiter, "Using transcriptomics to guide lead optimization in drug discovery Projects: Lessons Learned from the QSTAR Project," Drug Discovery Today, vol. 20, no. 5, pp. 505 – 513, May 2015.

[62] David S. Wishart, "Emerging applications of metabolomics in drug discovery and precision medicine," Nature Reviews Drug Discovery, vol. 15, pp. 473 – 484, March 2016.

[63] Cuperlovic-Culf M. and Culf M. S., "Applied metabolomics in drug discovery," Expert Opin Drug Discov, vol. 11, no. 8, pp. 759 – 70, August 2016.

[64] S. K. Halgamuge, "Deep near unsupervised learning for data analysis in metabolomics, drug-drug interaction discovery and human gait recognition," in Proceedings of the 2016 IEEE 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, pp. 5-6, 2016.

[65] Zhang M, Luo H, Xi Z and Rogaeva E "Drug repositioning for diabetes based on 'omics' data mining," PLoS ONE, vol. 10, no. 5, e0126082

[66] Pramod Katara, "Role of bioinformatics and pharmacogenomics in drug discovery and development process," Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 2, no. 4, pp. 225 – 230, December 2013.

[67] Enrique Hermandez-Lemus, "Computational pharmacogenomics," Omics for Personalized Medicine, Springer India, New Delhi, pp. 163 – 186, 2013.

[68] Jean-Paul Renaud, Chun-wa Chung, U. Helena Danielson, Ursula Egner, Michael Hennig, Roderick E. Hubbard and Herbert Nar, "Biophysics in drug discovery: impact, challenges and opportunities," Nature Reviews Drug Discovery, vol. 15, pp. 679 – 698, August 2016.

[69] M. Q. Yang, J. Y. Yang and O. K. Ersoy, "Computational intelligence – a broad initiative in automated learning from sequences," in Proceedings of the IEEE 2005 ICSC Congress on Computational Intelligence and Applications, Istambul, Turkey, pp. 6, December 2005.

[70] M. Kimura, S. Aoki and H. Ohwada, "Predicting radiation protection and toxicity of p53 targeting radio protectors using machine learning," in Proceedings of the 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Manchester, pp. 1-6, 2017.

[71] N. Goodacre, N. Edwards, M. Danielsen, P. Uetz and C. Wu, "Predicting nsSNPs that disrupt protein-protein interactions using docking," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, no. 5, pp. 1082-1093, Sept.-Oct. 1 2017.

[72] L. Peng, B. Liao, W. Zhu, Z. Li and K. Li, "Predicting drug–target interactions with multi-information fusion," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 2, pp. 561-572, March 2017.

[73] Lixia Yao, James A. Evans and Andrey Rzhetsky, "Novel opportunities for computational biology and sociology in drug discovery," Trends Biotechnol, vol. 28, no. 4, pp. 161 – 170, 2010.

[74] Gaisser S. and Nusser M., "The role of biotechnology in pharmaceutical drug design," Z Evid Fortbild Qual Desundhwes, vol. 104, no. 10, pp. 732 – 7, 2010.

[75] Agnieszka Stryjewska, Katarzyna Kiepura, Tadeusz Libroski and Stanistaw Lochynski, "Biotechnogy and genetic engineering in the new drug development. Part I. DNA technology and recombinant proteins," Pharmacological Reports, vol. 65, pp. 1075 – 1085, 2013.

