Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 38 (2012) 1264 – 1275
Procedia Engineering
www.elsevier.com/locate/procedia

International Conference on Modeling Optimisation and Computing – (ICMOC – 2012)

A Study on Cheminformatics and its Applications on Modern Drug Discovery

B.Firdaus Begam[a] and Dr. J.Satheesh Kumar[b]

[a]*Research Scholar, Bharathiar University, Coimbatore, India, firdh_2002@yahoo.com*

[b]*Assistant Professor, Bharathiar University, Coimbatore, India, jsathee@ieee.org*

Abstract

Discovering drugs to a disease is still a challenging task for medical researchers due to the complex structures of biomolecules which are responsible for disease such as AIDS, Cancer, Autism, Alzimear etc. Design and development of new efficient anti-drugs for the disease without any side effects are becoming mandatory in the recent history of human life cycle due to changes in various factors which includes food habit, environmental and migration in human life style. Cheminformaticds deals with discovering drugs based in modern drug discovery techniques which in turn rectifies complex issues in traditional drug discovery system. Cheminformatics tools, helps medical chemist for better understanding of complex structures of chemical compounds. Cheminformatics is a new emerging interdisciplinary field which primarily aims to discover Novel Chemical Entities [NCE] which ultimately results in design of new molecule [chemical data]. It also plays an important role for collecting, storing and analysing the chemical data. This paper focuses on cheminformatics and its applications on drug discovery and modern drug discovery techniques which helps chemist and medical researchers for finding solution to the complex disease.

Keywords: Cheminformatics; Drug discovery; Drug development; Lead identification; Lead Optimization.

1. Introduction

The term "Cheminformatics" also referred as Chemoinformatics/Chemiinformatics/Chemical information/Chemical informatics has been recognised in recent years as a distinct discipline in computational molecular sciences. Cheminformatics is also known as interface science as it combines Physics, Chemistry, Biology, Mathematics, Biochemistry, Statistics and informatics [3][5][24].

The primary focus of cheminformatics is to analyse/simulate/modelling/manipulate chemical information which can represented either in 2D structure or in 3D structure. Industry sectors such as,

B.Firdaus Begam, Email Address: firdh_2002@yahoo.com and Dr.J.Satheesh Kumar, Email Address: jsathee@rediffmail.com

1877-7058 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.06.156

agrochemicals, food and pharmaceutical are distinct areas where cheminformatics plays significant role in the recent history of molecular sciences [5][14][15].

Cheminformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization, and use of chemical information [3]. According to F.K Brown "The use of information technology and management has become a critical part of the drug discovery process. Cheminformatics is the mixing of information resources to transform data into information and information into knowledge which is collectively referred as inductive learning as shown in Fig.1. for the intended purpose of making better decisions faster in the areas of drug lead identification and organization" [1]. From J. Gasteiger and T. Engel perception, cheminformatics can be viewed as "The application of informatics methods to solve chemical problems" [2]. M. Hann and R. Green coined cheminformatics as "a new name for an old problem" [5].

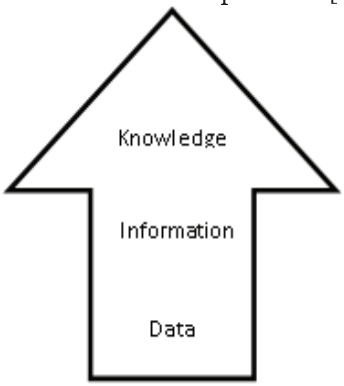


Fig. 1. Cheminformatics transformation

Cheminformatics has mainly dealt with small molecules, whereas bioinformatics addresses genes, proteins, and other larger chemical compounds (shown in Fig. 2). Chem and Bioinformatics complements each other for bimolecular process, like structure and function of proteins, the binding of a ligand to its binding site, the conversion of a substrate within its enzyme receptor, and the catalysis of a biochemical reaction by an enzyme.

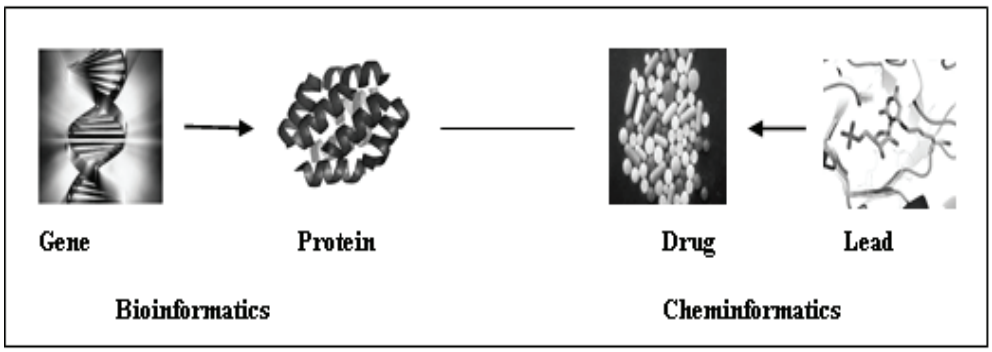


Fig. 2. The Cooperation of Bioinformatics and Cheminformatics

Different tools and methods are available to represent chemical structure, database to store chemical data, to perform searching process, Quality Structure- Activity Relationship(QSAR), Quality Structure-

Property Relationship(QSPR), to predict physical, chemical and biological properties of a molecule[6][5][8][13].

1.1 NEED AND IMPORTANCE OF CHEMINFORMATICS

Cheminformatics plays a key role to maintain and access enormous amount of chemical data, produced by chemist (more than 45 million chemical compounds are known and the number may increase in million every year,) by using a proper database. Also, the field of chemistry needs a novel technique for knowledge extraction from data to model complex relationships between the structure of the chemical compound and biological activity or the influence of reaction condition on chemical reactivity [6][16]. Cheminformatics has wider range of application and Fig. 3. shows influence if cheminformatics in some specific research areas.

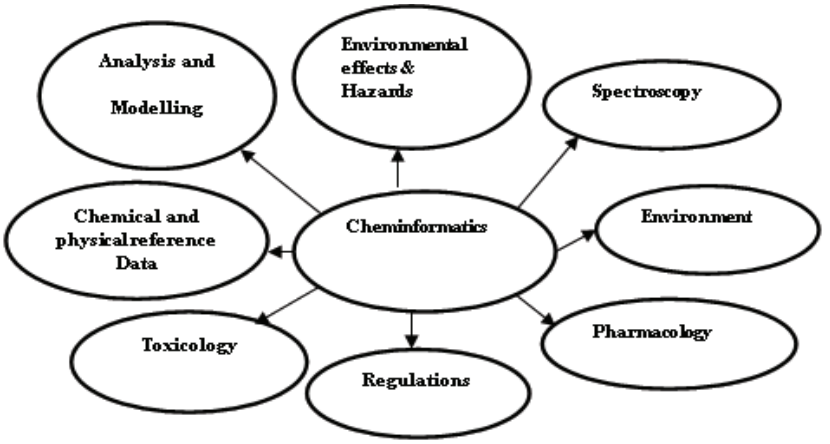


Fig. 3. Need for Cheminformatics

Three major aspects of Cheminformatics are;

i) Information Acquisition, is a process of generating and collecting data empirically (experimentation) or from theory (molecular simulation)
ii) Information Management deals with storage and retrieval of information and
iii) Information use, which includes Data Analysis, correlation, and application to problems in the chemical and biochemical sciences [24]

This paper is organized as follows. Section 2 deals with significant applications of Cheminformatics on various research areas. Section 3 specifies different tools available for analysis, visualize and interpret the properties of chemical information. Section 4 reviews the role of cheminformatics for drug discovery and Section 6 concludes the paper.

2. Cheminformatics and its Applications

Cheminformatics is a significant application of information technology to help chemists for investigating new problems, organize, analyse, and understand scientific data in the development of novel compounds, materials and processes. Primary modules of cheminformatics are Computer-Assisted Synthesis Design, Structure representation and chemmetrics, shown in Fig. 4. [3][4][5][7][8][14][15].

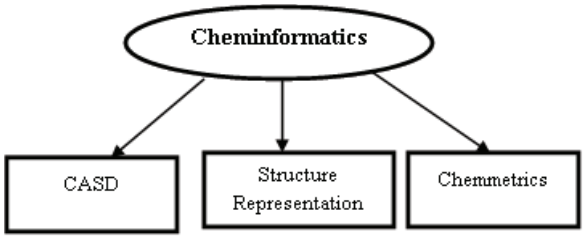
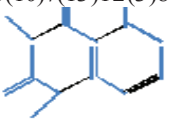
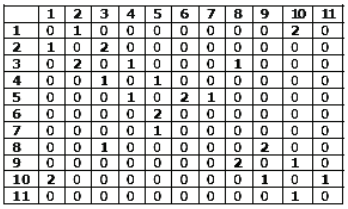


Fig. 4. Different areas focused in Cheminformatics

Computer-Assisted Synthesis Design (CASD) is applied mainly where artificial intelligence technique can be applied. This technique is applied in various applications which included pharmaceutical, food industry, textile industry and agro industry.

Various forms of machine readable chemical representation play basic property to design chemical database where the chemical information are stored for analysis and manipulation. The chemical structure representations can be linear, 2D or in 3D format. Some of the chemical structure representations are shown in Table 1. SMILES (Simplified Molecular Input Line Entry Specification) is one of the linear chemical notation format which is widely used among chemist [38] for various clinical and analysis purpose. Structure representation deals with Reaction Representation, Structure Descriptors, Molecular Modelling, Structure Searching, and Computer-Assisted Structure Elucidation (CASE) as shown in Fig.5.

Table 1: Some of the Chemical Structure Representation

Representation	Name
Caffine	Common Name
trimethylxanthine coffeine, theine, mateine,	Synonyms
$C_8H_{10}N_4O_2$	Empirical formula
3,7-dihydro-1,3,7-trimethyl-1H-purine-2,6-dione	IUPAC Name
58-08-2	CAS Registry Number
T56 BN DN FNVNVJ B1 F1 H1	WLN Notation
CN1C=NC2=C1C(=O)N(C(=O)N2C)C	SMILES
1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3	Inchl
	Markush Structure
	Connection Table
[OH]c1cccc1	Fragment Code
0000100110100111	Fingerprint
5244987098423150	Hash Code

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	0	0	0	0	0	2	0
2	1	0	2	0	0	0	0	0	0	0	0
3	0	2	0	1	0	0	0	1	0	0	0
4	0	0	1	0	1	0	0	0	0	0	0
5	0	0	0	1	0	2	1	0	0	0	0
6	0	0	0	0	2	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0
8	0	0	1	0	0	0	0	0	2	0	0
9	0	0	0	0	0	0	0	2	0	1	0
10	2	0	0	0	0	0	0	0	1	0	1
11	0	0	0	0	0	0	0	0	0	1	0

Reaction Representation helps to understand the basic chemical models, quantify chemical reactivity and extract knowledge from the reaction information [23]. Molecular modelling is a method includes a variety of computational schemes which are aimed at stimulating molecular structures, their properties and "in-silico" behaviour [8].

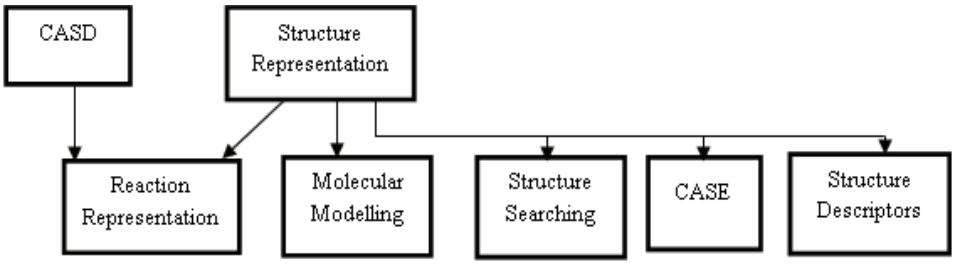


Fig. 5. CASD and Structure representation

Structure Searching involves in determination of features like bond orders, rings and aromaticity. It includes searching the whole structure, substructure, structure similarity and diversity. CASE builds on information obtained from various spectroscopic methods like IR, NMR, MS, etc. Structure Descriptor used to identify the physical, chemical and biological properties of chemical compound and relationship between two structures. The descriptors fall into four classes such as, i) Topological, ii) Geometrical, iii) Electronic and iv) Hybrid or 3D Descriptors [14][23].

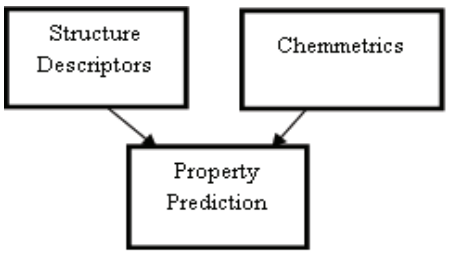


Fig. 6. Structure Representation and Chemmetrics

Chemmetrics is used for quantitative analyse of the chemical data by using mathematical and statistical methods [45]. It also deals with property prediction of chemical information (refer Fig. 6.).

2.1 APPLICATIONS OF CHEMINFORMATICS

The range of applications of cheminformatics is rich indeed; any field of chemistry can profit from its methods. The following lists different areas of chemistry and indicates some typical applications of cheminformatics.

a) Storing data generated through experiments or from molecular simulation Retrieval of chemical Structures from chemical database (Software libraries).
b) Prediction of physical, chemical and biological properties of chemical compounds.
c) Elucidation of the structure of a compound based on spectroscopic data.
d) Structure, Substructure, Similarity and diversity searching from chemical database [14][15].

e) High Throughput Screening (HTS) is the integration of technologies (laboratory automation, assay technology, micro plate based instrumentation, etc.) to quickly screen chemical compounds in search of a desired activity [19].
f) Docking - Interaction between two macromolecules [17][18].
g) Drug Discovery [11][12][15][24].
h) Molecular Science, Materials Science, Food Science (nutraceuticals), Atmospheric chemistry, Polymer chemistry, Textile Industry, Combinatorial organic synthesis (COS) [4].

3. Tools Used for cheminformatics

The development of software and tools for computer – assisted organic synthesis are under vast development. This has resulted in many tools and representations for chemical structures. Some of the tools are listed below [3][13][15][21].

- ISIS-Draw is a chemical structure drawing program for Windows, published by MDL Information Systems. It is the interfacial software to ISIS/Base database [20].
- ChemDraw is a molecule editor developed by the cheminformatics company CambridgeSoft. ChemDraw is, along with Chem3D and ChemFinder, part of the ChemOffice suite of programs and is available for Macintosh and Microsoft Windows [20].
- ChemWindow, is a chemical structure drawing program with several template. The template can be created by the customer can be saved in template folder and opened in preference dialogue box [20].
- ChemSketch, is a chemical structure drawing program with predefined templates are available for drawing and it is more powerful and user friendly tool for structure analysis [20].
- ChemReader is a software developer toolkit for translating digital raster images of chemical structures into standard, chemical file formats that can be searched and analyzed with other open source or commercial cheminformatic software [36].
- JME Molecular Editor is a Java applet which allows to draw / edit molecules and reactions (including generation of substructure queries) and to depict molecules directly within an HTML page [37].
- LogCHEM, an Inductive Logic Programming (ILP) based tool for discriminative interactive mining of chemical fragments [22].
- PLSR (PLS-Regression), a simple chemmetrics tool, which relates two matrix X and Y through linear multivariate model and has the ability to analyse data with many, noisy, collinear, and even incomplete variables in both X and Y [25].
- Wendi (Web Engine for Nonobvious Drug Information), a web based integrative data mining tool. It attempts to find non-obvious relationships between a query compound and scholarly publications, biological properties, genes and diseases using multiple information sources [26].
- ChemMine tool is an online service for small molecule analysis. It provides an interface between cheminformatics and data mining tools for various analytical analyses in chemical genomics and drug discovery [27].
- CML (Chemical Markup Language) is degined as combination of semantic text and non-textual information of chemical strucutre on the internet. It acts like HTML pages [39][40].
- MyChemise (My Chemical Structure Editor) is a new 2D structure editor. It is designed as a Java applet that enables the direct creation of structures in the Internet using a web browser. MyChemise saves files in a digital format (.cse) and the import and export of .mol files using the appropriate connection tables is also possible [41].
- PubChem is an open repository for small molecules and their experimental biological activity. It integrates and provides search, retrieval, visualization, analysis, and

programmatic access tools in an effort to maximize the utility of contributed information [43].

- Open Babel is a chemical tool box which interconverts chemical structures between different formats, over 110 formats [42].
- AmberTools is used Biomolecular simulation and analysis of polymers, nucleotides, and synthetic organic structures [44].

Some other tools such as, CAS Draw, DIVA (Diverse Information, Visualization and Analysis), Structure Checker Accord, DS Accord Chemistry Cartridge, MarvinSketch PowerMV, TINKER, APBS, ArgusLab, Babel, ioSolveIT, ChemTK, Chimera, CLIFF, Dragon, gOpenMol, Grace, JOELib, Jmol, IA_LOGP, Lammps, MIPSIM, Mol2Mol, AMSOL, MOLCAS, Molexel, ICM-Pro, ORTEP, Packmol, Polar, XLOGP,PREMIER Biosoft, Q-chem, ALOGPS, Qmol, SageMD, ChemTK Lite, Transient, CLOGP,TURBOMOLE, UNIVIS, VMD, WHATIF, GCluto, COSMOlogic, KOWWIN are also used for similar kind of applications mentioned above.

4. Role of Cheminformatics in Morden Drug Discovery

Recent chemical developments for drug discovery are generating a lot of chemical data which is referred as information explosion. This has created a demand to effectively collect, organize, analyse and apply the chemical information in the process of modern drug discovery and development [24]. The drug discovery process is aimed at discovering molecules that can be very rapidly developed for effective treatments to meet medical needs.

The entanglement of chemistry and information management started in the mid of 1970s, applying in the area of prediction of protein structure, Fourier transform of X-ray crystallography, enzyme and chemical kinetics, analyse various types of spectroscopy data and binding of chemical compounds. During early 1980s, computer technology is considered as the core component by the medical chemist to solve chemical problems [4][10]. For example, collecting crystal structures of small molecules in Cambridge Structural Database (CSD) provides a fertile resource for geometrical data on molecular fragments for calibration of force fields and validation of results from computational chemistry. The need of storing macromolecular data results in Protein Data Base (PDB). The needs and refinement on these approaches result in several tools and upgrading the process of solving the problems.

The traditional drug discovery process starts with a particular Disease, Identification of target, Identification of molecule effective against target and Preclinical testing. Identification of target and synthesis the molecule to increase their suitability takes more amounts of time and cost (in millions) which is done in “WET Lab”. This is the area where the chemical informatics plays its major role in discovery process of the drug. The development process starts with human clinical trials, approval from authority and delivers the product in the market. This process takes about 10-15 years to discover, develop and bring drug to the market [3][12].

The modern pharmaceutical drug discovery and development pipeline process, as shown in Fig. 7, starts with Disease selection, Target identification, Lead identification, Lead Optimization, Pre-clinical trial testing, Clinical trial testing, Approval and circulation (Drug in market). In traditional drug discovery phase, the process which cost more time and money is replaced with lead identification and lead optimization process in modern drug discovery system. Each phase has an interaction component that transfers data, knowledge and information to one another (shown in Fig. 8).

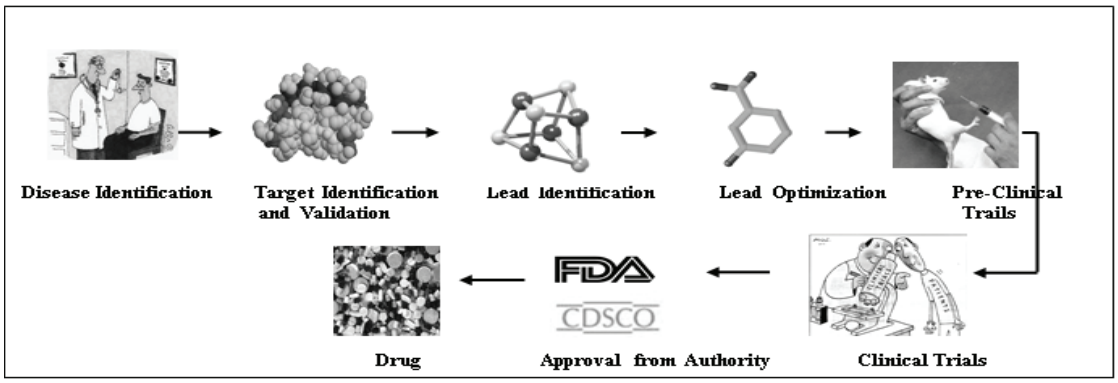


Fig. 7. Modern Drug Discovery and Development Life Cycle

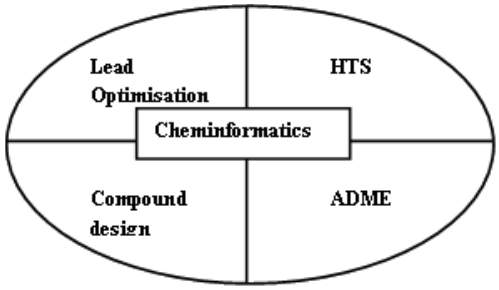


Fig. 8. Interaction Process

4.1 Pre –Drug Discovery Process (Disease identification)

Before the discovery process starts, it is to understand the disease by knowing, how the genes can be altered, how it affects the protein, how these protein will react with each other in living organism, how the affected cells can change the specific tissues and how the disease affects the patient. All the phases in Pre-Discovery, Discovery and Development of Drug is shown in Fig. 9.

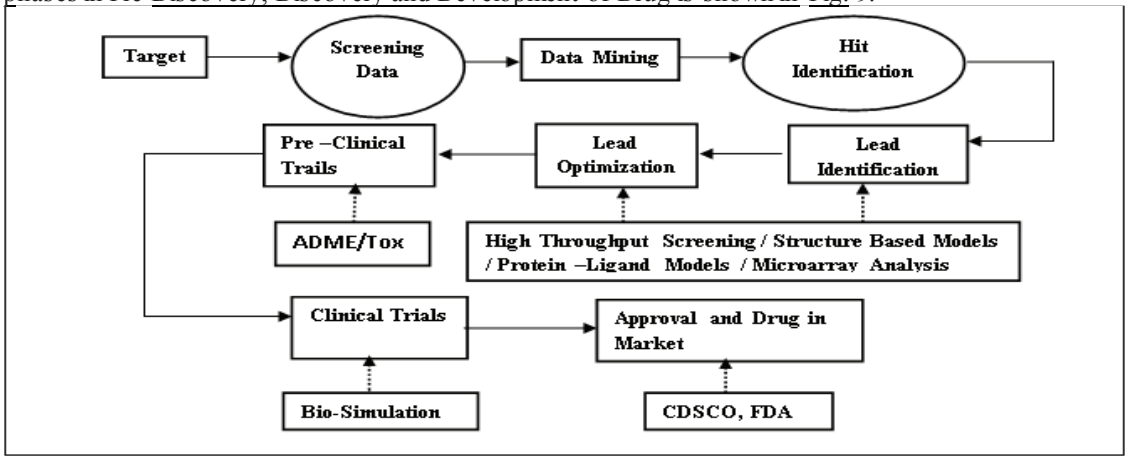


Fig. 9. Drug Discovery and Development Process

4.2 Morden Drug Discovery Process

The discovery process includes four important processes such as, target identification and validation, lead identification, lead optimization and pre-clinical trials.

a) Target Identification & Validation

Cheminformatics is used to identify target molecule which can be either gene or protein and could be a potential drug for the disease (Gene/Protein analysis). The Identified protein is separated, crystallized and ligand binding processes are done. Some approaches will inhibit the disease functionality by making the key molecule stop functioning. Another approach is by promoting specific molecule in the normal way which may have affected in the disease state. These approaches and different databases can be applied for the discovery of drug targets. After target Identification, validation phase starts by determining whether the modulation of the target will yield a desired clinical outcome. This is based on the results obtained between the cellular location and disease/health condition, potential expression and protein binding state [30-32].

b) Lead Identification.

Target identification like protein, gene, leads to "Hit-to-Lead" phase. High Throughput Screening (HTS) technique is applied where the protein targets are automatically screened against database of small-molecule or cell-based assay compounds. Lead identification also helps to see which molecules bind strongly to the target [35]. Several similarity and diversity techniques can be applied for lead identification [33-34].

c) Lead Optimization

This phase results in finding the drug candidate from the lead identified compound. The goal is a process of refining the chemical structure of a confirmed hit to improve its drug characteristics. Several docking techniques can be applied to optimize the lead structures for target affinity and selectivity [30-32].

Different techniques and methods are used for Lead identification and Optimization process where some of the methods Virtual Screening, Molecular Database, Data mining, High-Throughput Screening (HTS), QSAR, Protein – Ligand Models, Structure Based Models, Microarray analysis, Property Calculation and ADMET(adsorption, distribution, metabolism and elimination and Toxicity) [9][10][12].

d) Pre- Clinical Trial

The preclinical stage is an important phase to check whether the compound can be made into a drug to treat specific disease which is not toxic and has minimum side effects. Toxicity tests are undertaken to show safety while pharmacokinetics testing is done to provide data on how a drug is absorbed, distributed, metabolised and excreted (ADME) from the body. Pre-clinical studies and testing can be done with or without animal testing method. In-vitro is a study, based on the test done in the clinical lab and the analysis based on living cell cultures and animal model can be referred as in-vivio method. This phase will be designed in a way such that it achieves risk-free, unproblematic and economic transition from pre-clinical to clinical trial in medical product development [15][30-32].

4.3 Development Process

Development process is another significant stage in the life cycle of finding new drug. This stage consists of three major phase such as clinical trial, approval from the authority and drug in market.

i) Clinical Trial

Clinical trial is the primary phase which will be fastest and safest way to find treatments which acts as the best solution for challenging health disease of human being. Patient with specific disease will be considered for clinical trial and relevant data will be collected with respect to the time. Trials can be done in five ways such as, prevention trials, screening trails, diagnostic trails, treatment trial and quality of life trials [15][30][31].

ii) Approval from the Authority and Drug in Market

Based on the rules and regulation for new drug development in the country as well as in international market, research authorities check the safety and other parameters to approve the drug for marketing. Central Drugs Standard Control Organization (CDSCO) in India and Food and Drug Administration (FDA) in U.K. approves a new pharmaceutical compound for sales and marketing [9-13][15].

5. Conclusion

Average life span of Human being is gradually decreasing in the recent medical history due to the higher influence of new diseases. Identifying and understanding structural and functional behaviour of chemical compounds/biomolecules are one of the challenging issues for medical researchers. Cheminformatics is an emerging field which is used for better understanding of biomolecules. This paper primarily focuses on cheminformatics and its applications on drug discovery, issues of traditional discovery and importance of modern drug discovery system. This in turn helps chemists and researchers for developing drugs without side effects.

Acknowledgements

Authors are thankful to the faculty members of the department of computer applications, Bharathiar University, for their support. Authors are also thankful to the reviewers and editors for their valuable comments to improve the quality of the paper.

References

[1] Brown FK. Chemoinformatics: what is it and how does it impact drug discovery? Ann Rep Med Chem; 1998, 33:375–384.

[2] Hann, M., Green, R. Chemoinformaticss A new name for an old problem. Curr. Opin. Chem. Biol.; 1999, 3, 379-383.

[3] V.Arulmozhi, R.Rajesh. Chemoinformatics – A Quick Review; IEEE 2011,978-1-4244-8679-3/11.

[4] Peter Willett, Chemoinformatics: a history; John Wiley & Sons, L td. January / February 2011, Volume 1.

[5] Thomas Engel. Basic Overview of Chemoinformatics, *J.* Chem. Inf. Model, 2006, 46, 2267-2277.

[6] Nathan Brown. "Chemoinformatics- An Introduction to computer Scientists", ACM Computing survey, vol 41, Np. 2, Article 8, Publication date: Feburary 2009.

[7] Dr. Jurgen Bajorath. Chemoinformatics and its Role in Pharmaceutical Research, Business Briefiing: Pharmatech 2004.

[8] J. Polanski. Chemoinformatics, Elsevier B.V.; 2009.

[9] Peter Ertl, Wolfgang Miltz, Bernhard Rohde and Paul Selzer. Web-based cheminformatics for bench chemists, Drug Discovery World Fall 2000.

[10] Oleg Ursu, Anwar Rayan, Amiram Goldblum and Tudor, Oprea. Understanding drug-likeness, Wiley & Sons, L td.; September /Oc tober 2011,Volume 1.

[11] Jeffrey Augen. The evolving role of information technology in the drug discovery process, research focus; March 2002, Vol. 7, p. 5.

[12] Garland R. Marshall. Introduction to Chemoinformatics in Drug Discovery – A Personal View, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim; 2004, ISBN: 3-527-30753-2.

[13] Sreenivasa Reddy P, T. Sreenivasulu Reddy, G Saayi Krushna. Advantages of Bio and Cheminformatics tools in drug design, Current Research in Drug Targeting, Urpjournals, Current Research in Drug Targeting; 2011, 1(2): 6-17.

[14] Andrew R. Leach, Valerie J. Gillet. An Introduction to Cheminformatics, Springer Science Business Media Inc, India; 2009.

[15] Jurgen Bajorath. Chemoinformatics: Concepts, Methods, tools, for Drug Discovery (methods in Molecular Biology), Humana Press; 2009.

[16] Rajarshi Guha, Kevin Gilbert, Geoffrey Fox, Marlon Pierce, David Wil§, Huapeng Yuan. Advances in Cheminformatics Methodologies and Infrastructure to Support the Data Mining of Large, Heterogeneous Chemical Datasets.

[17] Venkatraman Mohan, Alan C. Gibbs, Maxwell D. Cummings, Edward P. Jaeger and Renee L.DesJarlais. Docking: Successes and Challenges, Current Pharmaceutical Design; 2005, 11, 323-333.

[18] Graham R Smith and Michael JE Sternberg. Prediction of protein–protein interactions by docking methods, Elsevier Science Ltd, Current Opinion in Structural Biology; 2002, 12:28–35.

[19] James Inglese, Ronald L Johnson, Anton Simeonov, Menghang Xia, Wei Zheng, Christopher P Austin & Douglas S Auld. High-throughput screening assays for the identification of chemical probes, Nature Chemical Biology; August 2007, Vol 3, p. 8.

[20] Li, Z, Wan. H, Shi, Y, Ouyang, P. Personal Experience with Four Kinds of Chemical Structure Drawing Software: Review on ChemDraw, ChemWindow, ISIS/Draw and ChemSketch, J. Chem.Inf. Comput. Sci.; 44 (5): 1886–1890. doi:10.1021/ci049794h.PMID 15446849.

[21] Noel M O'Boyle, Rajarshi Guha, Egon L Willighagen, Samuel E Adams, Jonathan Alvarsson, Jean-Claude Bradley, Igor V Filippov, Robert M Hanson, Marcus D Hanwell, Geoffrey R Hutchison, Craig A James, Nina Jeliazkova, Andrew SID Lang, Karol M Langner, David C Lonie, Daniel M Lowe,Jérôme Pansanel, Dmitry Pavlov, Ola Spjuth, Christoph Steinbeck, Adam L Tenderholt, Kevin J Theisen and Peter Murray-Rust. Open Data, Open Source and Open Standards in chemistry: The Blue Obelisk five years on, O'Boyle et al. Journal of Cheminformatics; 2011, 3:37.

[22] V´ıtor Santos Costa, Nuno A. Fonseca, Rui Camacho. LogCHEM: Interactive Discriminative Mining of Chemical Structure, IEEE International Conference on Bioinformatics and Biomedicine;2008, 978-0-7695-3452-7/08.

[23] Johann Gasteiger and Thomas Engel. Cheminformatics, Wiley-VCH.

[24] Deepak Bharati, Jagtap RS, Kanase KG, Sonawame SA, Undale VR and Bhosale AV. Chemoinformatics: Newer Approach for Drug Development, Asian J. Research Chem. 2(1); Jan.-March, 2009, ISSN 0974-4169.

[25] Svante Wold, Michael Sjostroma, Lennart Eriksson. PLS-regression: a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems, Elsevier Science B.V.58; 2001, p. 109–130.

[26] Qian Zhu1, Michael S Lajiness, Ying Ding, David J Wild. WENDI: A tool for finding non-obvious relationships between compounds and biological properties, genes, diseases and scholarly Publications, Zhu et al. Journal of Cheminformatics ; 2010, 2:6.

[27] Tyler W. H. Backman, Yiqun Cao, and Thomas Girke. ChemMine tools: an online service for analyzing and clustering small molecules, Nucleic Acids Res.; 2011 July 1; 39(Web Server issue): W486–W491.

[28] Steinbeck C, Han YQ, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): An open- source Java library for chemo- and bioinformatics, Journal of Chemical Information and Computer Sciences; 2003, 43(2):493-500.

[29] Kim Sweeny. Technology Trends in Drug Discovery and Development: Implications for the Development of the Pharmaceutical Industry in Australia.

[30] V. Srinivasa Rao and K. Srinivas. Modern drug discovery process: An *in silico* approach, Journal of Bioinformatics and Sequence Analysis;June 2011,Vol. 2(5), p. 89-94.

[31] Lifeng Kang,Bong, Geun Chung, Robert Langer and Ali Khademhosseini. Microfludics for drug discovery and development: From target selection to product lifecycle management, Drug Discovery Today; January 2008, vol. 13.

[32] Konrad H. Bleicher, Hans-Joachim Bohm, Klaus Muller and Alexander I. Alanine. Hit and Lead Generation: Beyond High-Throughput Screening, Nature Reviews, Drug Discovery; May 2003, vol. 2, p. 369.

[33] Robert A. Goodnow Jr. Hit and lead identification: Integrated technology-based approaches, Elsevier Ltd; 2006.

[34] Alberto Ambesi-Impiombato and Diego di Bernardo; Computational Biology and Drug Discovery: From Single-Target to Network Drugs, Current Bioinformatics, Bentham Science Publishers Ltd.; 2006,vol. 1, p. 3-13.

[35] S Ekins, J Mestres and B Testa,In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling, Br J Pharmacol; September 2007, 152(1), p. 9–20.

[36] Jungkap Park, Gus R Rosania,Kerby A Shedden, Mandee Nguyen, Naesung Lyu, and Kazuhiro Saitou. Automated extraction of chemical structure information from digital raster images, Chem Cent J; 2009, p. 3: 4.

[37] Peter Ertl. Molecular structure input on the web, J Cheminformtics; 2010, 2: 1.

[38] Richard G. A. Bone, Michael A. Firth, and Richard A. Sykes. SMILES Extensions for Pattern Matching and Molecular Transformations:Applications in Chemoinformatics, *J. Chem. Inf. Comput. Sci.;* 1999, *39,* 846-860.

[39] Peter Murray-Rust and Henry S Rzepa. CML: Evolution and design, Journal of Cheminformatics; 2011, 3:44.

[40] Syed Ahsan, Muhammad Shahbaz, Syed Athar Masood. STYX : A CML based chem.-informatics facility, Journal of American Science;2011,7(8).

[41] Jörg-Hubertus Wilhelm. MyChemise: A 2D drawing program that uses morphing for visualisation purposes, Journal of Cheminformatics; 2011, 3:53.

[42] Noel M O'Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch and Geoffrey R Hutchison. Open Babel: An open chemical toolbox, O'Boyle et al. Journal of Cheminformatics ;2011, 3:33.

[43] Evan E Bolton, Jie Chen, Sunghwan Kim, Lianyi Han, Siqian He, Wenyao Shi, Vahan Simonyan, Yan Sun, Paul A Thiessen, Jiyao Wang, Bo Yu, Jian Zhang and Stephen H Bryant, PubChem3D: a new resource for scientists, Bolton et al. Journal of Cheminformatics; 2011, 3:32.

[44] Christopher C. Roberts. Ligand Parameterization and Simulation with AMBER and AmberTools

[45] Svante Wold. Chemometrics; what do we mean with it, and what do we want from it?, Chemometrics and Intelligent Laboratory Systems, Elsevier Science B.V; 1995,30, p. 109-115.