EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 11
Issue: 3
339-351
Year: 2024

INFORMATION TECHNIQUE IN BIOSYSTEM BY CHEMINFORMATICS & BIOINFORMATICS: A UNIQUE TOOL IN DRUG DISCOVERY

Dhrubo Jyoti Sen*

School of Pharmacy, Techno India University, Sector-V, Salt Lake City, EM: 4/1, Kolkata-700091, West Bengal, India.

***Corresponding Author: Dr. Dhrubo Jyoti Sen**

School of Pharmacy, Techno India University, Sector-V, Salt Lake City, EM: 4/1, Kolkata-700091, West Bengal, India.

Article Received on 21/01/2024 **Article Revised on 11/02/2024** **Article Accepted on 01/03/2024**

ABSTRACT

Chemoinformatics involves integrating the principles of physical chemistry with computer-based and information science methodologies, commonly referred to as "in silico techniques", in order to address a wide range of descriptive and prescriptive chemistry issues, including applications to biology, drug discovery, and related molecular areas. On the other hand, the incorporation of machine learning has been considered of high importance in the field of drug design, enabling the extraction of chemical data from enormous compound databases to develop drugs endowed with significant biological features. The present review discusses the field of cheminformatics and proposes the use of virtual chemical libraries in virtual screening methods to increase the probability of discovering novel hit chemicals. The virtual libraries address the need to increase the quality of the compounds as well as discover promising ones. On the other hand, various applications of bioinformatics in disease classification, diagnosis, and identification of multidrug-resistant organisms were discussed. The use of ensemble models and brute-force feature selection methodology has resulted in high accuracy rates for heart disease and COVID-19 diagnosis, along with the role of special formulations for targeting meningitis and Alzheimer's disease. Additionally, the correlation between genomic variations and disease states such as obesity and chronic progressive external ophthalmoplegia, the investigation of the antibacterial activity of pyrazole and benzimidazole-based compounds against resistant microorganisms, and its applications in chemoinformatics for the prediction of drug properties and toxicity—all the previously mentioned—were presented in the current review. The sensible use of bioinformatics and chemoinformatics in drug research and discovery holds significant potential for the introduction of novel and more potent pharmaceuticals. This covers how genetic variants affect medication response, which includes enhancing the qualities of currently available drugs as well as the health of people, animals, and plant communities. In order to identify novel compounds, design molecules for selectivity, efficacy, and safety, and generate clinical trial candidates from novel compounds, this also makes use of a broad range of computational scientific approaches. Over the past few years, there has been an increase in the creation of new, accurate computational methods for drug design due to the synergistic interaction between medical chemistry and bioinformatics. The primary objective is the drug, where binding mode, interaction, binding site, and binding energy are estimated using computer-based techniques such molecular docking, de novo design, and other virtual screening methods. Consequently, I will assess the linked product based on contemporary informatics, the sophisticated approach for rational drug design, and the relationship between the drug development process and safety in this study. In drug discovery, bioinformatics enables the efficient analysis and interpretation of large scale biological data, facilitating target identification, lead compound optimization and prediction of drug target interactions.

KEYWORDS: Chemoinformatics; Bioinformatics; Applications; Formulation; Advances.

INTRODUCTION

Bioinformatics plays a major role in drug discovery as well development. It provides efficient analysis of large scale biological data provides facility in identification of target, optimization of lead compound, and prediction of drug-target interactions.

It assists in characterization and identification of potential drug targets, prediction of drug metabolism and pharmacokinetic properties, provides analysis of drug-drug interactions, adverse drug reactions, and personalized medicine approaches. Acceleration of drug discovery process leads to improved success rates, and reduced costs. At the same time it offers insights into safety and efficacy of potential drug candidates.[1]

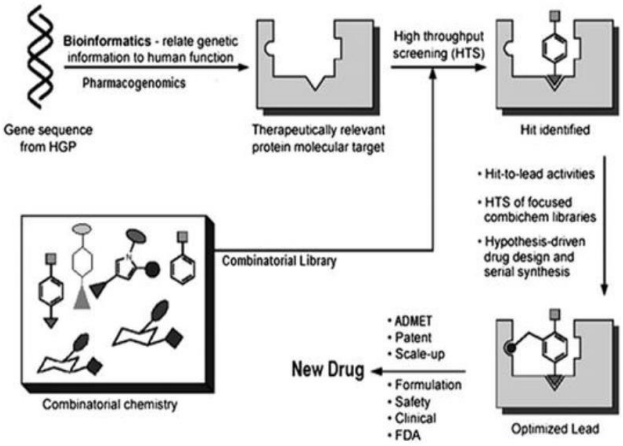

Figure 1: Rational drug design.

Chemoinformatics is chemical information which provides an integrated approach to study the function of chemical systems using pharmacophore modeling, quantitative structure–activity relationship (QSAR), docking, and molecular dynamics (MD) simulations. It is a branch in chemistry that depends on *in-silico* mathematics The informatic methods helps to solve chemical problems", predict possible toxicological properties derived from structural data compared to other previously known chemical, proved experimentally to be toxic. It is the mixing of information resources in order to transform data into information, and information into knowledge, for the purpose of making better decisions faster which leads to optimization and identification of drug. Thus it is a stream that extends beyond the context of drug discovery which includes the application of computational methods to all kinds of chemical data. Specialist working in this sector requires competency from a variety of disciplines, like data mining, statistical methods biological sciences, bioinformatics, database and web technology.[2]

Role of bioinformatics: The in-disciplinary field of science of collecting and analyzing complex biological data and their interpretation.

Fundamental objective-

- Identifying genes and proteins and determine their function to establish their conformation.
- Include human genome project (HGP) and human microbiome project which provide a complete sequencing of human genome to unlock genetic contribution for many diseases: personalized medication.
- Bioinformatics can speed up the identification of therapeutic agents through screening and refinement of those agents.
- Widely applied in the examination of genomics, proteomics, 3D structure modeling of proteins, drug image designing (single-molecule protein sequencing analysis), transcriptomic, epigenetic, cistronic, ribosome profiling data and more.

The science has open a new area of research: One search method is HTS (High Throughput Screening) where targets are screened against databases of small molecule compounds to see which molecule bind strongly to the target, check its affinity, specificity, characterization. If there is a 'hit' with a particular compound, it can be extracted from the database for further testing.[3]

Example- ZINC is a good example of a HTS compound library.

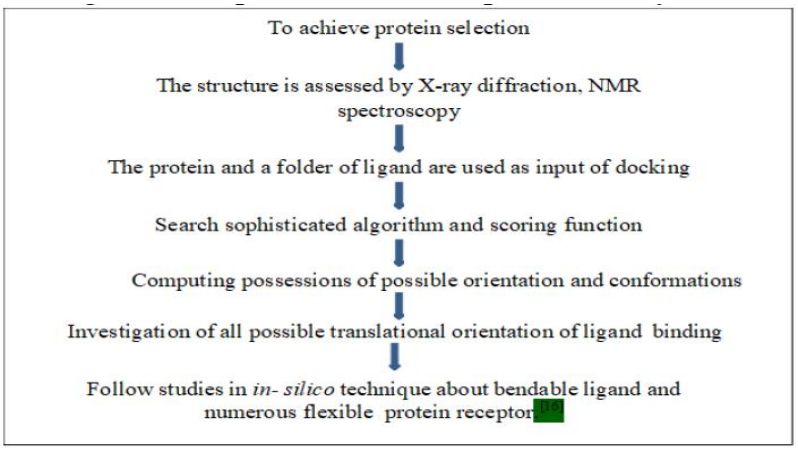

Figure 2: Flow chart of homology modeling.

The process of discovering and crafting new drugs is difficult, risky, time-consuming, and maybe rewarding. It needs a ton of money, human resources, and technological proficiency. A new drug must also pass stringent production and testing criteria rules before it can be sold and used by consumers. In the drug development process, bioinformatics and chemoinformatics facilitate faster drug discovery and reduced expenses for development.[4]

Example- translational bioinformatics approaches to enhance the quality of drug discovery. The bioinformatics and computer aided drug design tool is one of the renovated platforms in current drug discovery. It provides the biological, chemical and toxicological information to streamline early drug discovery.

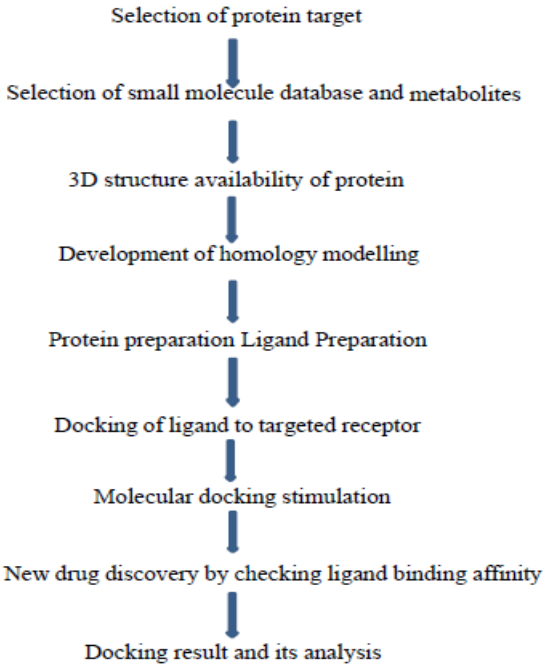

Figure 3: Bioinformatics flow chart.

Bioinformatics, as related to genetics and genomics, is a scientific sub discipline that involves using computer technology to collect, store, analyze and disseminate biological data and information, such as DNA and amino acid sequences or annotations about those sequences. Bioinformatics is about developing new technologies and software tools in the fields of medicine, biological research, and biotechnology. The major scope and application of bioinformatics are: To understand the function of genes.[5]

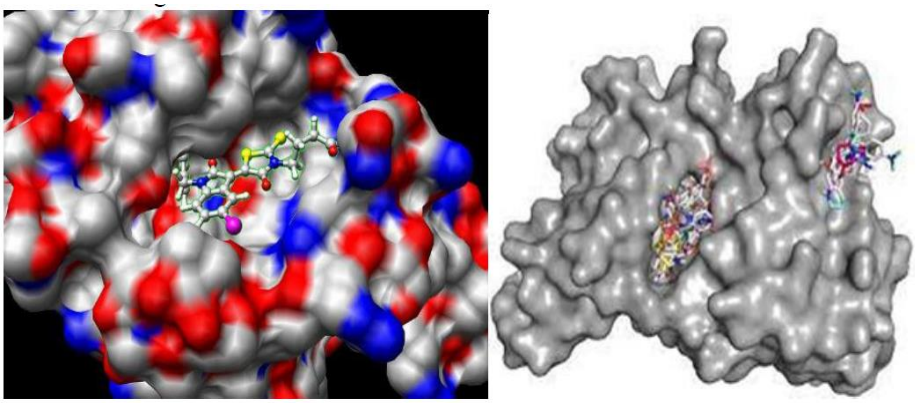

Figure 4: Molecular docking.

Cell organizations and function. Analysis of drug targets. Owing to that required expertise, a career in bioinformatics tends to carry a lucrative salary, a high degree of job security, and a generally strong career outlook. Bioinformatics graduates can teach in government and private colleges, work in the scientific research institutes and manufacturing units of biomedical products.[6]

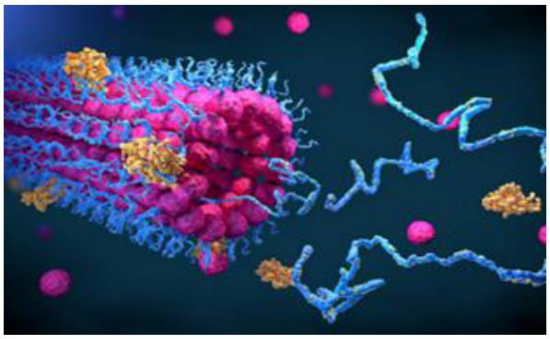

Figure 5: Protein 3D.

Biotechnology firms and Biotechnology research centers recruit students from the bioinformatics background for future research. Bioinformatics is an interdisciplinary field of science that develops methods and software tools for understanding biological data, especially when the data sets are large and complex. Bioinformatics uses biology, chemistry, physics, computer science, computer programming, information engineering, mathematics and statistics to analyze and interpret biological data.[7] The subsequent process of analyzing and interpreting data is referred to as computational biology. Computational, statistical, and computer programming techniques have been used for computer simulation analyses of biological queries.

They include reused specific analysis "pipelines", particularly in the field of genomics, such as by the identification of genes and single nucleotide polymorphisms (SNPs). These pipelines are used to better understand the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations. Bioinformatics also includes proteomics, which tries to understand the organizational principles within nucleic acid and protein sequences.[8]

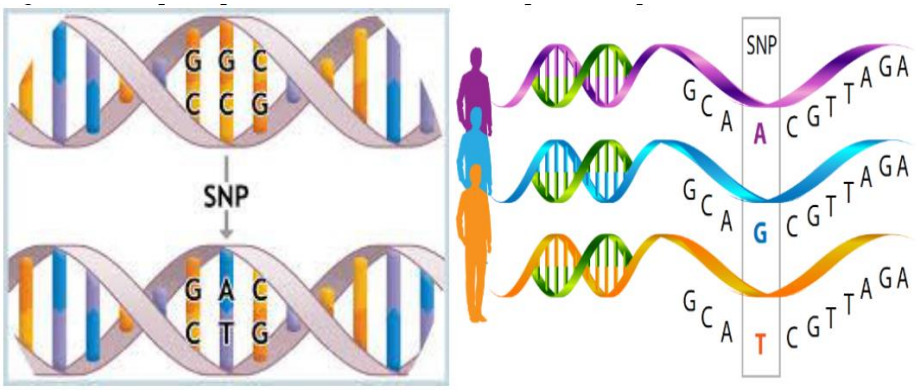

Figure 6: SNP: Single nucleotide polymorphisms.

Image and signal processing allow extraction of useful results from large amounts of raw data. In the field of genetics, it aids in sequencing and annotating genomes and their observed mutations. Bioinformatics includes text mining of biological literature and the development of biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation.[9]

Bioinformatics tools aid in comparing, analyzing and interpreting genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyze and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, RNA, proteins as well as biomolecular interactions. Bioinformatics is used in areas such as genomics, proteomics, and evolutionary biology.[10]

Frank Arthur Brown Jr. (1908–1983) was a leading mid-20th century researcher of biological rhythms and inventor of cheminformatics. He was a professor of biological sciences at Northwestern University and trustee of the Marine Biological Laboratory in Woods Hole, Massachusetts.

Figure 7: Frank Arthur Brown Jr.

Role of cheminformatics: The field of study which works on all aspects of design, creation, organization, representation, management, retrieval, analysis, visualization and use of experimental new drug in chemical information.[11]

Drug discovery through cheminformatics

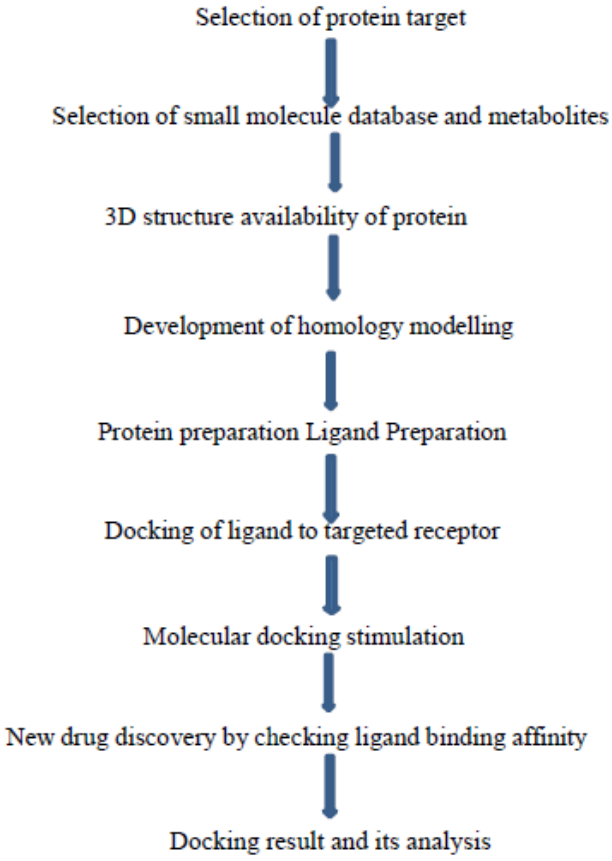

Figure 8: Virtual screening.

Importance-It saves researchers significant amounts of time and money by generating and filtering compound libraries based on their desirable properties, as well as analyzing, visualizing and mining HTS (High Throughput Screening) data. This involves-

- Compound selection
- Virtual library generation
- Virtual HTS
- *In-silico* ADMET

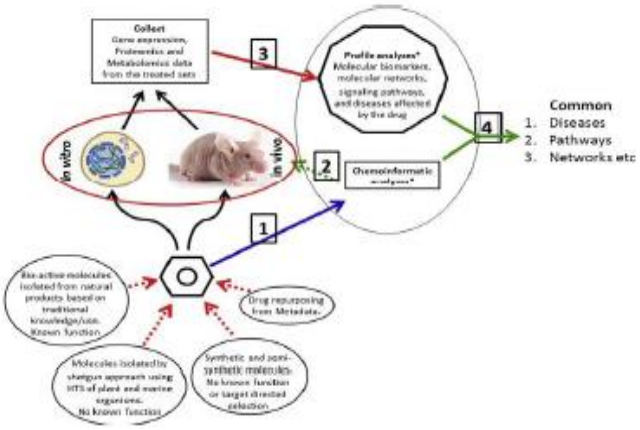

Figure 9: HTS.

- Chemoinformatics gives an opportunity to transform the data obtained via linking the two fields into knowledge which can then be extended to make proper and better decisions in drug discovery, understanding chemical interaction, standardization of drug manufacturing protocol.
- Chemoinformatics play an important role in target identification and validation.
- Chemoinformatics is a tool that facilitates the decision making process across various preclinical stages of drug discovery.
- Chemoinformatics provides better understanding of complex structure of drugs and its rectification.
- Chemoinformatics allows access a huge amount of databases loaded with chemical information otherwise difficult to reach through literature studies.

Cheminformatics, on the other hand, is the application of computer and informational techniques to problems in the field of chemistry. Cheminformatics is a relatively new field of information technology that focuses on the collection, storage, analysis, and manipulation of chemical data.

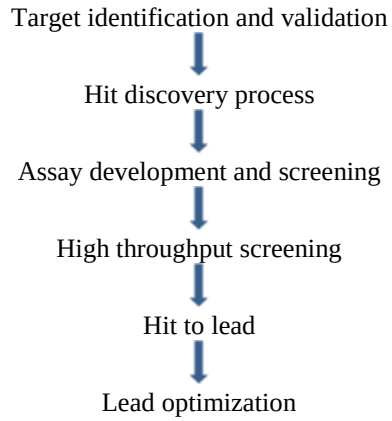

Figure 10: Lead optimization.

A well-established field, cheminformatics focuses on the extraction, plotting, and quick explosion of "big" chemical data from QSAR analysis, combinatorial synthesis, and HTS. It is arranged rationally, giving readers a strong foundation in models and techniques before moving on to applications in drug development and the creation of cheminformatics infrastructures that forecast structure activity relationship, or SAR.[12]

Figure 11: Bioinformatics.

The chemical data of interest typically includes information on small molecule formulas, structures, properties, spectra, and activities (biological or industrial).

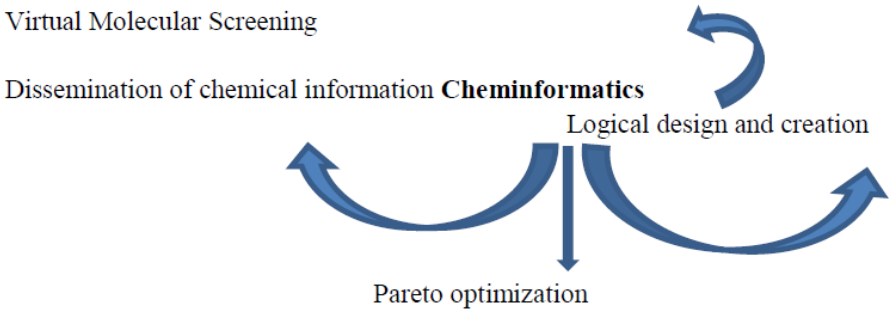

Figure 12: Cheminformatics.

Chemoinformatics can be used for sequential HTS to more efficiently produce screening for ligand-receptor interactions. The stronger these interactions, the more viable the sample is for drug development. Cheminformatics (also known as chemoinformatics) refers to the use of physical chemistry theory with computer and information science techniques—so called "in silico" techniques—in application to a range of descriptive and prescriptive problems in the field of chemistry, including in its applications to biology and related molecular fields.[13]

Such *in-silico* techniques are used, for example, by pharmaceutical companies and in academic settings to aid and inform the process of drug discovery, for instance in the design of well-defined combinatorial libraries of synthetic compounds, or to assist in structure-based drug design. The methods can also be used in chemical and allied industries, and such fields as environmental science and pharmacology, where chemical processes are involved or studied.[14]

Figure 13: Paulien Hogeweg & Ben Hesper.

The first definition of the term bioinformatics was coined by **Paulien Hogeweg** [24 December 1943 (age 80 years), Amsterdam, Netherlands] and **Ben Hesper** in 1970, to refer to the study of information processes in biotic systems. This definition placed bioinformatics as a field parallel to biochemistry (the study of chemical processes in biological systems).

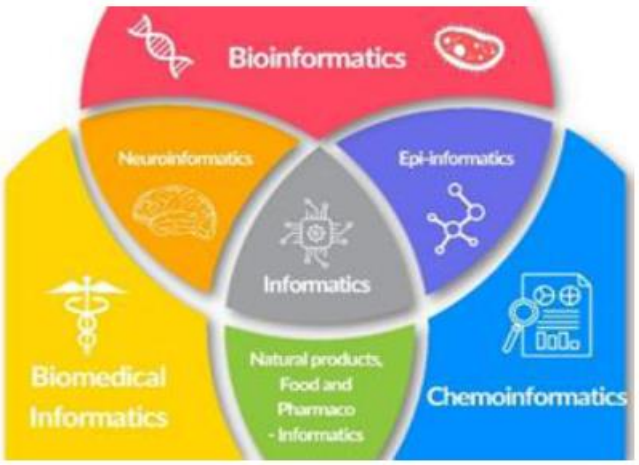

Figure 14: Bioinformatics & Cheminformatics.

Bioinformatics and computational biology involved the analysis of biological data, particularly DNA, RNA, and protein sequences. The field of bioinformatics experienced explosive growth starting in the mid-1990s, driven largely by the Human Genome Project and by rapid advances in DNA sequencing technology. Analyzing biological data to produce meaningful information involves writing and running software programs that use algorithms from graph theory, artificial intelligence, soft computing, data mining, image processing, and computer simulation.

The algorithms in turn depend on theoretical foundations such as discrete mathematics, control theory, system theory, information theory, and statistics. With a large number of prokaryotic and eukaryotic genomes completely sequenced and more forthcoming, access to the genomic information and synthesizing it for the discovery of new knowledge have become central themes of modern biological research. Mining the genomic information requires the use of sophisticated computational tools.[15]

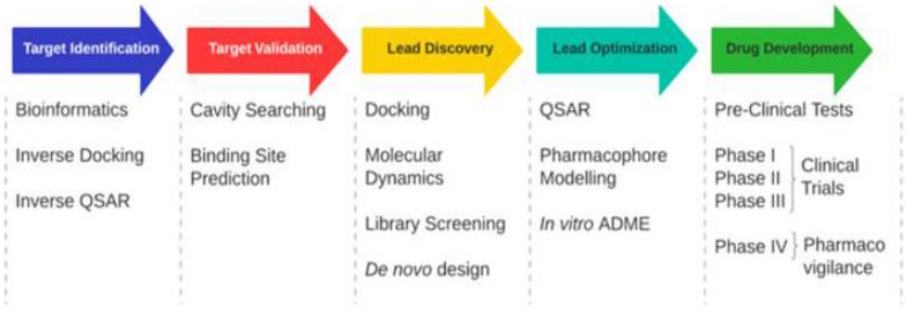

Figure 15: Flowchart of algorithms.

It therefore becomes imperative for the new generation of biologists to initiate and familiarize with a field of study that is concerned with the careful storage, organization and indexing of information in order to tackle the new challenges in the genomic era. Information science has been applied to biology to produce a field is called bioinformatics.

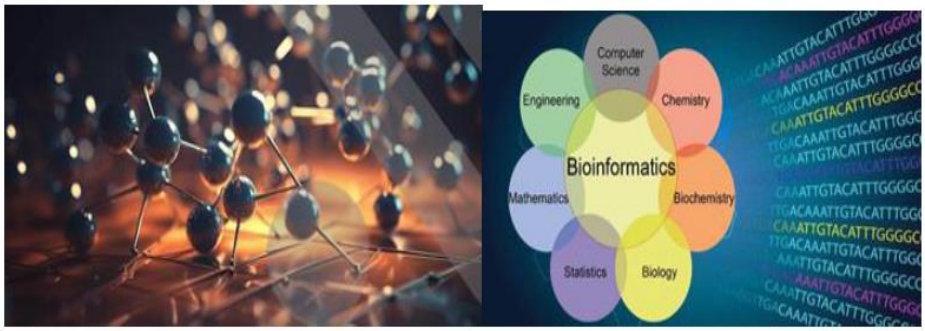

Figure 16: Chem & Bio informatics.

It is concerned with the state of- the-art computational tools available to solve biological research problems. The term bioinformatics was coined by Paulien Hogeweg and Ben Hesper to describe "the study of informatics processes in biotic systems" and it found early use when the first biological sequence data began to be shared.[16]

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. The development of bioinformatics as a field is the result of advances in both molecular biology and computer science over the past 30–40 years.

Figure 17: High Throughput Screening.

As an interdisciplinary field of science, bioinformatics combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret biological data. The key areas of bioinformatics include biological databases, sequence alignment, gene and promoter prediction, molecular phylogenetics, structural bioinformatics, genomics, and proteomics.[17]

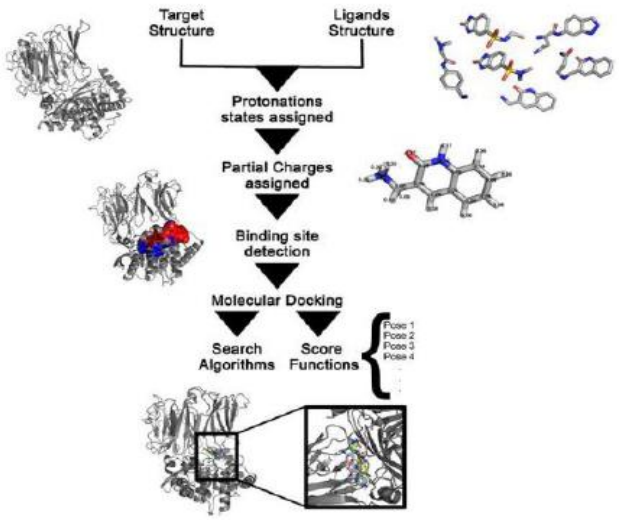

Figure 18: Docking algorithm.

Mechanism

Process

Relation between physically appropriate molecule like protein, peptides, nucleic acid, carbohydrate and lipids play role in signal transduction → Relative orientation of two interaction → A signal is formed (agonists and antagonists) → Docking predict potency and type of signal formed

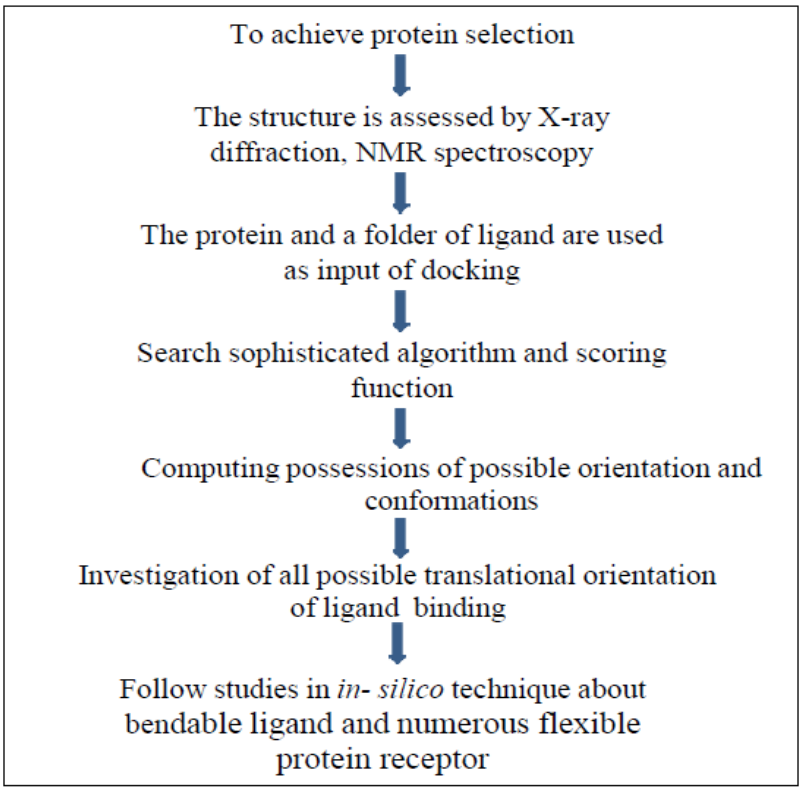

Figure 19: In-silico method.

Principle: Based on structure drug design, capability of binding conformation of small ligand to suitable target site. Fundamental biochemical process.

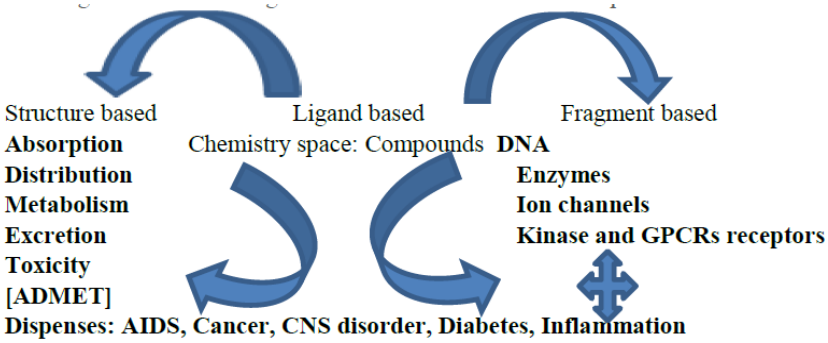

Figure 20: Flow chart of drug discovery.

Drug discovery and development is a complex high risk, time consuming and potentially rewarding process.it requires technological expertise, human resources and literally huge capital.it also requires strict adherence to regulations on testing and manufacturing standards before a new drug comes into market and can be used in general population. Bioinformatics and chemoinformatics in the drug design accelerates the drug discovery process easier and reduces cost in drug development process.
Example- translational bioinformatics approaches to enhance the quality of drug discovery.

The bioinformatics and computer aided drug design tool is one of the renovated platforms in current drug discovery. It provides the biological, chemical and toxicological information to streamline early drug discovery. Characterization of potential drug target.[18]

Combinatorial chemistry importance: These days, one of the most significant advancements in drug research is combinatorial chemistry. It makes it possible for scientists to consistently clone and express target biological receptors, enzymes, and proteins. Numerous novel pharmacological targets for a range of illnesses have now been found using the conventional drug discovery process to screen this large number of targets. Combinatorial chemistry is developing at an opportune time and will surely aid in the search for new medications that will benefit humankind. Since many leads can frequently be generated from a single combinatorial library screen, combinatorial chemistry is especially helpful when used in conjunction with contemporary computational chemistry and molecular modeling approaches. It can be used not only for the initial lead finding but also for medication lead optimization.

Features are

1. Assay formats for rapid screening of compound libraries.
2. New coupling compounds generation.
3. New building blocks and cleavage linkers that are compatible to a variety of organic synthesis and assay systems.

Screening of compounds

High throughput screening method is used to screen a combinatorial library.

1. In solid phase synthesis assays are of two types-
- Direct binding of the molecular target to the bead bound ligand-this binding can be detected by direct visualization (e.g-by a color target such as a dye).
- Detection of functional properties of the bead bound ligand e.g identifying phosphorylation or proteolytic substances.

2. Synthetic or natural products are typically added in a soluble form to each individual well for biological testing during the solution phase of production. Numerous enzymatic assays, cell-based signal transduction assays, competitive receptor binding assays with radiolabelled ligand, antibacterial, antiviral, and anticancer assays are among the different kinds of assays that are accessible.

General approaches

- The 96 well two stage releasable assays.
- In situ releasable assay with immobilized beads.
- In both approaches ligands are attached to solid support via cleavable linkers. The ligands are then released from each bead into solution phase when the biological assays take place.

Determination of the chemical structure of the active ligand

After creation of large amounts of molecular diversity it is necessary to screen the ligands to determine the structure of the ligands primarily three methods are used for structural deconvolution of active compounds:

- Iterative deconvolution
- Positional scanning deconvolution.
- Encoding.

Interactive deconvolution methods: In this case, pools of compounds are created so that every individual pool has a specific building block at one or two positions, and all possible combinations of building blocks are

incorporated at the other positions. The pool with the highest biological activity is chosen to serve as the ideal building block at the designated location. After that, a second cycle of synthesis is carried out, this time with a chosen building block introduced at the first defined position. The biological activities of each pool are assessed in order to determine which construction block is best placed at the additional stated position. Then, same procedure is carried out one more.

Positional scanning (PS) deconvolution method: In order to determine the most significant capabilities at each site of diversity in a library, this approach entails screening distinct single defined positions of synthetic combinatorial libraries. There are 256 chemicals in the libraries displayed. (4*4*4*4) Since each compound is only present in one mixture of each sub-library, the structures of the individual compounds can be ascertained from such a screening.

Encoding: This technique uses a readable chemical tag that is affixed to each bead for every stage involved in the bead's actual molecule's creation by analyzing the tags that have been cleaved by acid hydrolysis or UV irradiation of a photo-labile linker followed by GC, the identities of the active compounds are ascertained.[19]

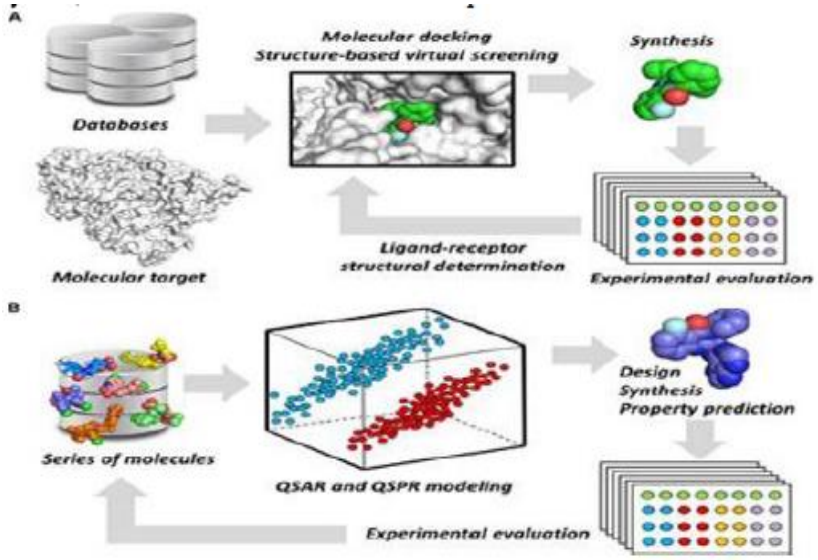

Figure 21: Encoding.

Bioinformatics analysis can not only accelerate drug target identification and drug candidate screening and refinement, but also facilitate characterization of side effects and predict drug resistance. It can explore the causes of diseases at the molecular level ,explain the phenomena of the diseases from the angle of the gene and make use of computer techniques, such as data mining, machine learning and so on, to decrease the scope of analysis and enhance the accuracy of the results so as to reduce the cost and time.

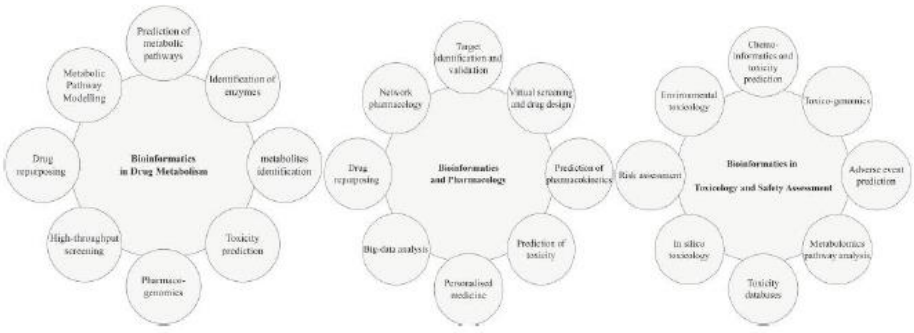

Figure 22: Application of bioinformatics.

Drug discovery by combinatorial synthesis: It is a type of chemical strategy that leads to make large number of chemical variants all at a time and to test them for bioactivity and then to isolate the most promising compound for further development.

It involves the systematic and retentive, covalent connection of a set of different building blocks of varying structures to yield a large array of diverse molecular entities.

Theoretically the number of possible different

compounds prepared by combinatorial synthesis is dependent on two factors

- Number of building blocks available for each step (b).
- The number of synthesis steps in the reaction scheme (x).

If an equal number of building blocks are used in each step then, N=b^x

If the number of building blocks vary in each step (e.g- b, c, d) in a three step synthesis then, N=bcd

This can be done by two process: solution phase synthesis and solid phase synthesis.[20]

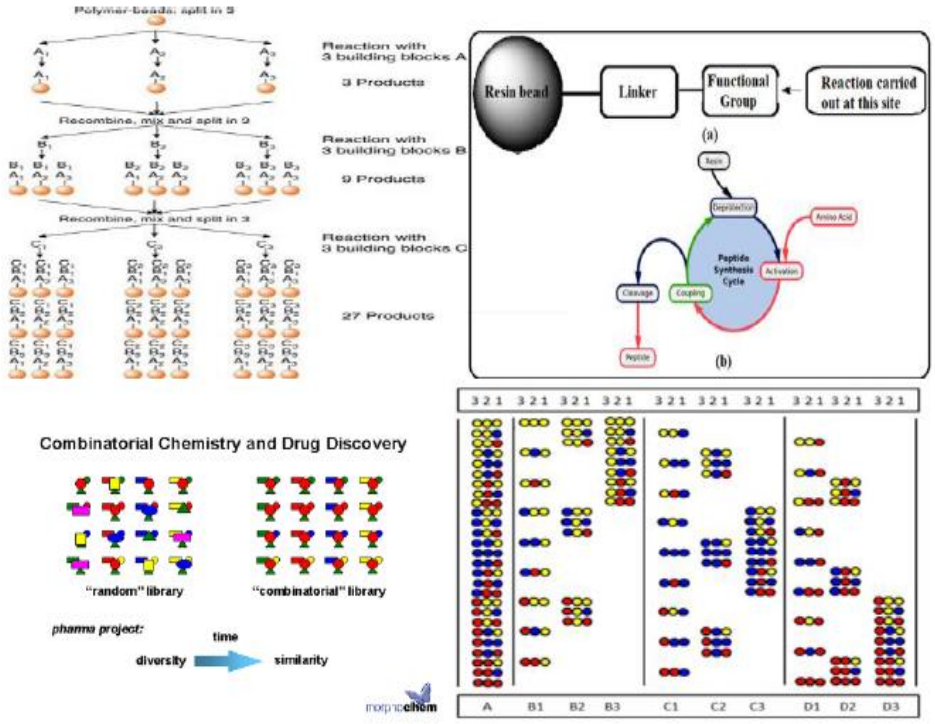

Figure 23: Combinatorial chemistry.

CONCLUSION

Chemoinformatics and Bioinformatics in the Pharmaceutical Sciences brings together two very important fields in pharmaceutical sciences that have been mostly seen as diverging from each other: chemoinformatics and bioinformatics. As developing drugs is an expensive and lengthy process, technology can improve the cost, efficiency and speed at which new drugs can be discovered and tested. This book presents some of the growing advancements of technology in the field of drug development and how the computational approaches explained here can reduce the financial and experimental burden of the drug discovery process. Chemoinformatics provides computer methods for learning from chemical data and for modeling tasks a chemist is facing. The field has evolved in the past 50 years and has substantially shaped how chemical research is performed by providing access to chemical information on a scale unattainable by traditional methods. Bioinformatics is already playing a key role in drug discovery. By analyzing large datasets of genomic and proteomic data, bioinformaticians are able to identify potential drug targets and design drugs that are tailored to specific genetic variants or disease subtypes.

REFERENCES

1. Sim AY, Minary P, Levitt M. "Modeling nucleic acids". Current Opinion in Structural Biology, 2012; 22 (3): 273–8.
2. Dawson WK, Maciejczyk M, Jankowska EJ, Bujnicki JM "Coarse-grained modeling of RNA 3D structure". Methods, 2016; 103: 138–56.
3. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A (2016). "Coarse-Grained Protein Models and Their Applications". Chemical Reviews, 2016; 116(14): 7898–936.
4. Joyce AP, Zhang C, Bradley P, Havranek JJ. "Structure-based modeling of protein: DNA specificity". Briefings in Functional Genomics, 2015; 14(1): 39–49.
5. Spiga E, Degiacomi MT, Dal Peraro M. "New Strategies for Integrative Dynamic Modeling of Macromolecular Assembly". In Karabencheva-Christova T (ed.). Biomolecular Modelling and Simulations. Advances in Protein Chemistry and Structural Biology, 2014; 96: 77–111.
6. Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O, Kmiecik S. "Protein-peptide docking: opportunities and challenges". Drug Discovery Today, 2018; 23(8): 1530–1537.

7. Ouzounis, C. A.; Valencia, A. "Early bioinformatics: the birth of a discipline—a personal view". Bioinformatics, 2003; 19(17): 2176–2190.
8. Hogeweg P Searls DB (ed.). "The roots of bioinformatics in theoretical biology". PLOS Computational Biology, 2011; 7(3): e1002021.
9. Hesper B, Hogeweg P "Bio-informatica: een werkconcept". Kameleon, 1970; 1(6): 28–29.
10. Hogeweg P "Simulating the growth of cellular forms". Simulation, 1978; 31(3): 90–96.
11. Thomas Engel "Basic Overview of Chemoinformatics". J. Chem. Inf. Model, 2006; 46(6): 2267–77.
12. F.K. Brown "Ch. 35. Chemoinformatics: What is it and How does it Impact Drug Discovery". Annual Reports in Medicinal Chemistry, 1998; 33: 375–384.
13. Varnek, A.; Baskin, I. "Chemoinformatics as a Theoretical Chemistry Discipline". Molecular Informatics, 2011; 30(1): 20–32.
14. Weininger, David "SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules". Journal of Chemical Information and Modeling, 1988; 28(1): 31–36.
15. Murray-Rust, Peter; Rzepa, Henry S. "Chemical Markup, XML, and the Worldwide Web. 1. Basic Principles". Journal of Chemical Information and Computer Sciences, 1999; 39(6): 928–942.
16. Kutchukian, Peter; Lou, David; Shakhnovich, Eugene. "FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules occupying Druglike Chemical". Journal of Chemical Information and Modeling, 2009; 49(7): 1630–1642.
17. Sushko, Yurii; Novotarskyi, Sergii; Körner, Robert; Vogt, Joachim; Abdelaziz, Ahmed; Tetko, Igor V. "Prediction-driven matched molecular pairs to interpret QSARs and aid the molecular optimization process". Journal of Cheminformatics, 2014; 6 (1): 48.
18. Soumyajit Mondal, Soumya Chakraborty, Vedansh Upadhayay, Keshav Patwari and Dr Dhrubo Jyoti Sen; Welcome to pharma-world: pharsight of pharmacy: a millennium oath: World Journal of Advance Healthcare Research, 2023; 7(10): 118-128.
19. Villoutreix BO Post-pandemic drug discovery and development: Facing present and future challenges. Frontiers in Drug Discovery, 2021; 1: 72: 69-84.
20. Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. Nature Chemical Biology, 2013; 9(4): 232-240.