AMERICAN JOURNAL OF PHARMTECH RESEARCH

Journal home page: http://www.ajptr.com/

Pharmacoinformatics: A Tool for Drug Discovery

Narendra Nyola*[1], G. Jeyablan[1], M. Kumawat[1], Rajesh Sharma[1], Gurpreet Singh[1], N. Kalra[1]

1. Alwar Pharmacy College, Alwar, Rajasthan, INDIA. 301030

ABSTRACT

Pharmacoinformatics is new emerging information technologies like neuroinformatics, immunoinformatics, bioinformatics, Metabolomics, chemo-informatics, toxico-informatics, cancer informatics, genome informatics, proteome informatics, biomedical informatics are basic tools provided for the purpose of drug discovery. There is an increasing recognition that information technology can be effectively used for drug discovery. The work in pharmacoinformatics can be broadly divided into two categories - scientific aspects and service aspects. The scientific component deals with the drug discovery and development activities whereas the service oriented aspects are more patient centric. The compelling drivers for the pharmaceutical industry are minimizing the time between a drug's discovery and its delivery to the marketplace and maintaining high productivity in the manufacturing processes. During a product's lifecycle many complex decisions must be made to achieve these goals. To better support the development and manufacturing processes at each stage, we have proposed a new epitome to facilitate the management and transfer of data information and knowledge. In future these information technology efforts are expected to grow both in terms of their reliability and scope. Thus, this emerging technology (pharmacoinformatics) is becoming an essential component of pharmaceutical sciences.

Keywords: - Informatics, Pharmaceutical informatics, Drug development, Chemoinformatics, Bioinformatics.

*Corresponding Author Email: narennyola2@gmail.com

Received 3 April 2012, Accepted 13 April 2012

Please cite this article in press as: Nyola N. *et al.*, Pharmacoinformatics: A Tool for Drug Discovery. American Journal of PharmTech Research 2012.

INTRODUCTION:

Pharmacoinformatics is the study, invention and effectuation of discipline where technology with any aspect of drug delivery, from the basic sciences (e.g. drug development or pharmacogenomics/pharmacogenetics) to the clinical use of medications in individuals and populations. It a subset of pharmacoinformatics, typically refers to the interface of technology with the practice of pharmacy. Includes pharmacy technologies involved in the preparation, delivery, and management of medication use within health care delivery systems.[1]

A flow chart showing the current status of the activities in pharmacoinformatics are given in figure 1.

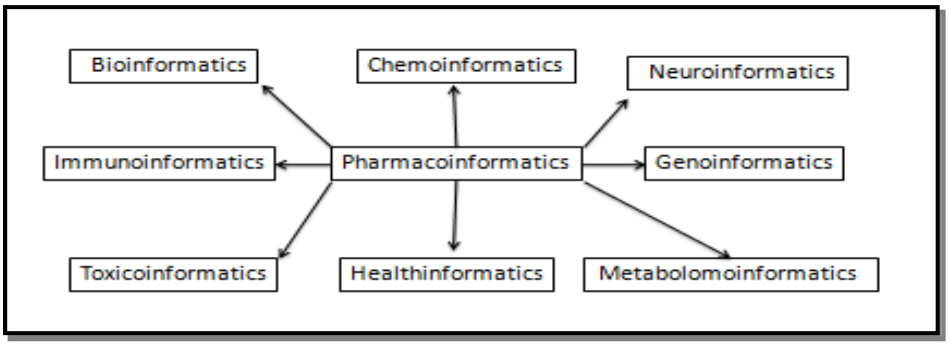


Figure. 1 Classification of Pharmacoinformatics

Bioinformatics:-

The term bioinformatics first came into use in the 1990s and was originally synonymous with the management and analysis of DNA, RNA and protein sequence data. Computational tools for sequence analysis had been available since the 1960s, but this was a minority interest until advances in sequencing technology led to a rapid expansion in the number of stored sequences in databases such as Genbank. [2] Now, the term has enlarged to incorporate many other types of biological data, for example protein structures, gene expression profiles, protein interactions, microarrays (DNA chips) ,functional analysis of biomolecules and drug designing. Each of these areas requires its own set of databases, algorithms and statistical methods.[3]

(Molecular) bio – informatics: bioinformatics is conceptualizing biology in terms of molecules (in the sense of physical chemistry) and applying "informatics techniques" (derived from disciplines such as applied maths, computer science and statistics) to understand and organize the information associated with these molecules, on a large scale. In short, bioinformatics is a management information system for molecular biology and has many practical applications.[4]

Bioinformatics and medical informatics (BIOMI) are multidisciplinary fields at the intersection of computing and informatics, mathematics and statistics, biology, chemistry, and

engineering.[5]Bioinformatics is the combination of biology and information technology. The discipline encompasses any computational tools and methods used to manage, analyze and manipulate large sets of biological data.

The National Center for Biotechnology Information (NCBI 2001) defines Bioinformatics as [6]"Bioinformatics is the field of science in which biology, computer science, and information technologies merge into a single discipline. There are three important sub-disciplines within Bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information."

Basically, bioinformatics has three components: [3]

1. The creation of databases allowing the storage and management of large biological data sets.
2. The development of algorithms and statistics to determine relationships among members of large data sets.
3. The use of these tools for the analysis and interpretation of various types of biological data, including DNA, RNA and protein sequences, protein structures, gene expression profiles, and biochemical pathways.

Bioinformatics and internet

Bioinformatics is largely, although not exclusively, a computer-based discipline. Computers are important in bioinformatics for two reasons: First, many bioinformatics problems require the same task to be repeated millions of times. For example, comparing a new sequence to every other sequence stored in a database or comparing a group of sequences systematically to determine evolutionary relationships. In such cases, the ability of computers to process information and test alternative solutions rapidly is indispensable. [7]

Second, computers are required for their problem-solving power. Typical problems that might be addressed using bioinformatics could include solving the folding pathways of protein given its amino acid sequence, or deducing a biochemical pathway given a collection of RNA expression profiles. Computers can help with such problems, but it is important to note that expert input and robust original data are also required.(fig 2)

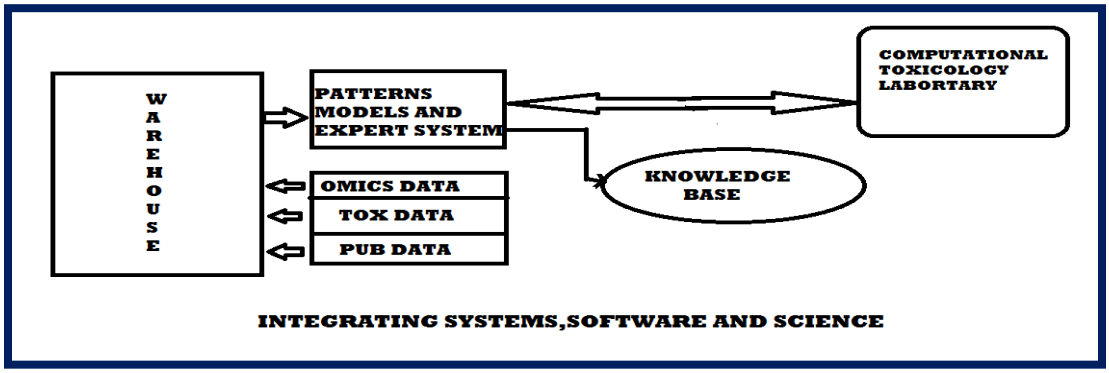


Figure.2 Bioinformatics and internet

This problem of how to digitize phenotypic data such as behavior, electrocardiograms, and crop health into a computer readable form offers exciting challenges for future bioinformatics. [8]Internet plays an important role to retrieve the biological information. Bioinformatics emerging new dimension of biological science includes computer science, mathematics and life science. The Computational part of bioinformatics use to optimize the biological problems like (metabolic disorder, genetic disorders). [9]

The Internet provides various facilities for Bioinformatics such as;

- Bioinformatics research
- Courses
- Resources
- Biological databases
- Construction tools
- Software resources
- WWW search tools
- Courses of Bioinformatics
- Advanced topics in Bioinformatics
- Scientific databases
- Electronic journals
- Asking queries from the librarian in online manner
- News events and activities such as; announcement for Bioinformatics interest group, meetings on federated databases, molecular biosciences and technology seminars.
    - **World Wide Web (WWW) Virtual Library:** This directory, provided by Cato Research Ltd., contains over 1000 URLs specific to biotechnology, pharmaceutical development, and related fields. The emphasis is on product development and the delivery of products and services. [10]

- **Subject Specific Sites:-**These sites are more likely to concentrate on a particular area of Bioinformatics. These sites are further divided into the various areas of Bioinformatics e.g., Codon usage, and Genome analysis/Genomic comparisons, Phylogenetics etc.
- **General Bioinformatics Web Sites:-**Many of the sites are offering the same sorts of links and many to other Bioinformatics sites; many have links to a Sequence Retrieval System or other facilities for sequence retrieval. These are categorized as under:
  - Academic Sites
  - Corporate/Government Sites

- **Access to Journals:-** Providing access to journals such as; Nature, Science, Molecular biology and Evolution, Nucleic Acids Research, Bioinformatics, The Journal of Molecular Biology, Genetics, New Scientist, Online Journal of Bioinformatics, Internet Science Journal.
- **As a Centre for Biotechnology Information:-**One can explore extensive sites of resources and including newsletters, Bioinformatics databases, and links to the major medical bibliographic databases. It not only connects to textual databases but also to Protein Structure Servers. These include 3DB browser, biomolecular modeling and structural classification of proteins etc. Biotechnologists can reach any Bioinformatics centers on Internet. From DNA Databank Japan to European Bioinformatics Centre can be reached by using Internet. One can search databases on protein, nucleotides, and protein structure.

News Services:-Various news services can be categorized as following:

- **Via e-mail:** It can serve as an important tool for being aware of new developments. The Scientist is providing this service with frequent coverage of Bioinformatics news. [11]
- **Directories of news sites:** Some directories of news sites are focusing in Bioinformatics. Southwest Biotechnology and Informatics Center (SWBIC) is a good "launch pad" to news sites. [12]
- **Headline news service:** Genomics Today is a daily headline news service that provides links to genomics news in other sites. It picks the relevant headlines from a wide variety of sources including wire services, newspapers, Yahoo, selected web sites, and university news sites. [13]

- **News Network:** Genome News Network is a good source for news on scientific, as opposed to business, aspects of Bioinformatics. In addition to news there are also featured articles and a few educational links. [41]
- **Conferences:** Internet also informs about conferences and other platforms where designers and architects of high-performance computer chips, software and systems can share and ideas for developing more sophisticated and useful tools for Internet for Bioinformatics. IT provides detailed information about conferences, symposiums like Intelligent Systems for Molecular Biology[15] and Pacific Symposium on Biocomputing. [16]

SEARCHING DATABASE

Types of Biological Databases Accessible

There are many different types of database but for routine sequence analysis, the following are initially the most important.

1. Primary databases
2. Secondary databases
3. Composite databases

- Primary Database (Nucleic Acid Protein):- EMBL, Genbank, DDBJ, SWISS-PROT, TREMBL, PIR.
- Secondary databases: - PROSITE, Pfam.
- Composite databases:-Combine different sources of primary databases. Composite database's NRDB OWL.

Some bioinformatics research and service centers:-

- National Center for Biotechnology Information (NCBI) in the USA;
- European Bioinformatics Institute (EBI) in the UK;
- Swiss Institute of Bioinformatics (SIB);
- Australian National Genome Information Service (ANGIS);
- Canadian Bioinformatics Resource (CBR);
- Peking Center of Bioinformatics (CBI);
- Singapore Bioinformatics Centre (BIC);
- South-African National Bioinformatics Institute (SANBI).

Biological Databases Metabolic pathways Enzymes/ metabolic pathways Classification of proteins and identifying domains Protein families, domains and functional sites Protein information Protein databases Gene level information Genomic databases DNA information

Nucleic acid databases Classification Taxonomic databases Literature Bibliographic databases Information Contain Type of databases

- **GenBank:** GenBank (Genetic Sequence Databank) is one of the fastest growing repositories of known genetic sequences. It has a flat file structure that is an ASC II text file, readable by both humans and computers. In addition to sequence data, GenBank files contain information like accession numbers and gene names, phylogenetic classification and references to published literature. [17-20]
- **EMBL:** The EMBL Nucleotide Sequence Database is a comprehensive database of DNA and RNA sequences collected from the scientific literature and patent applications and directly submitted from researchers and sequencing groups. [21]
- **Swiss Prot:** This is a protein sequence database that provides a high level of integration with other databases and also has a very low level of redundancy (means less identical sequences are present in the database).
- **Prosite:** The PROSITE dictionary of sites and patterns in proteins prepared by Amos Bairoch at the University of Geneva.
- **Ec-Enzyme:** The 'ENZYME' data bank contains the following data for each type of characterized enzyme for which an EC number has been provided: EC number, recommended name, Alternative names, Catalytic activity, Cofactors, Pointers to the SWISS-PROT entire(s) that correspond to the enzyme, Pointers to disease(s) associated with a deficiency of the enzyme.
- **PDB:** The X-ray crystallography Protein Data Bank (PDB), compiled at the Brookhaven National Laboratory.
- **GDB:** The GDB Human Genome Data Base supports biomedical research, clinical medicine, and professional and scientific education by providing for the storage and dissemination of data about genes and other DNA markers, map location, genetic disease and locus information, and bibliographic information.
- **DDBJ (DNA Data Bank of Japan)** began DNA data bank activities in earnest in 1986 at the National Institute of Genetics (NIG) with the endorsement of the Ministry of Education, Science, Sport and Culture.[22]
- **The Center for Information Biology at NIG** was reorganized as the Center for Information Biology and DNA Data Bank of Japan (CIB-DDBJ) in 2001. The new center

is to play a major role in carrying out research in information biology and to run DDBJ operation in the world.

- **OMIM:** The Mendelian Inheritance in Man data bank (MIM) is prepared by Victor Mc Kusick with the assistance of Claire A. Francomano and Stylianos E. Antonarakis at John Hopkins University.
- **PIR-PSD:** PIR (Protein Information Resource) produces and distributes the PIR-International Protein Sequence Database (PSD). It is the most comprehensive and expertly annotated protein sequence database.[23-25]
- **Genethon Genome Databases**
  a) PHYSICAL MAP: computation of the human genetic map using DNA fragments in the form of YAC contigs.
  b) GENETIC MAP: production of micro-satellite probes and the localization of chromosomes, to create a genetic map to aid in the study of hereditary diseases.
  c) GENEXPRESS (cDNA): catalogue the transcripts required for protein synthesis obtained from specific tissues, for example neuromuscular tissues.
- **21 Bdb: LBL's Human Chr 21 database:** This is a W3 interface to LBL's ACeDB-style database for Chromosome 21, 21Bdb, using the ACeDB gateway software developed and provided by Guy Decoux at INRA.
- **MGD: The Mouse Genome Databases:** MGD is a comprehensive database of genetic information on the laboratory mouse.
- **ACeDB (A Caenorhabditis elegans Database):** Containing data from the Caenorhabditis Genetics Center (funded by the NIH National Center for Research Resources), the C. elegans genome project (funded by the MRC and NIH), and the worm community.
- **Medline:** MEDLINE is NLM's premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, and the preclinical sciences. Journal articles are indexed for MEDLINE, and their citations are searchable, using NLM's controlled vocabulary, MeSH (Medical Subject Headings). [26,27]

Database Searching Algorithms

- FASTA (European Bioinformatic Institute) [28-30]
- BLAST (NCBI) [31]
- Smith-Waterman

Fasta: - Suite of programs for database searching by homology each program is launched by typing its name.

Search a nucleotide sequence database with a nucleotide query sequence and a protein sequence database with a protein query sequence .Compare a DNA sequence to a DNA sequence database and protein sequence to a protein sequence database .Fasta algorithm is used to compare a DNA sequence to a protein sequence database *fastx3* and *fasty3* algorithms are used compare the translated DNA sequence in three frames and allow gaps, Compare A protein sequence to a DNA sequence database *tfastx3* and *tfasty3* algorithms are used. Compare the translated DNA sequence in three frames and allow gaps.

fasta algorithm

- Derives from the logic of **Dot Plot**
- Computes best diagonals from all frames of alignment
- use word matching (figure 3,4)[57]

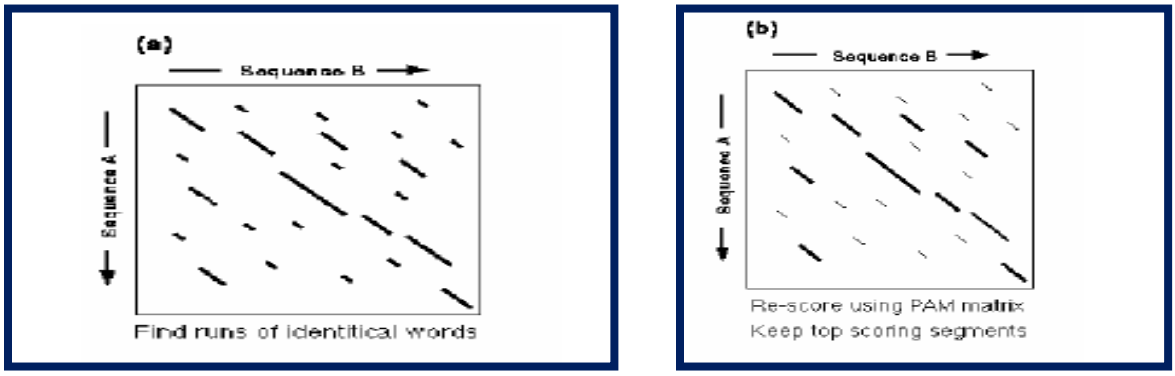


Figure 3 Fasta

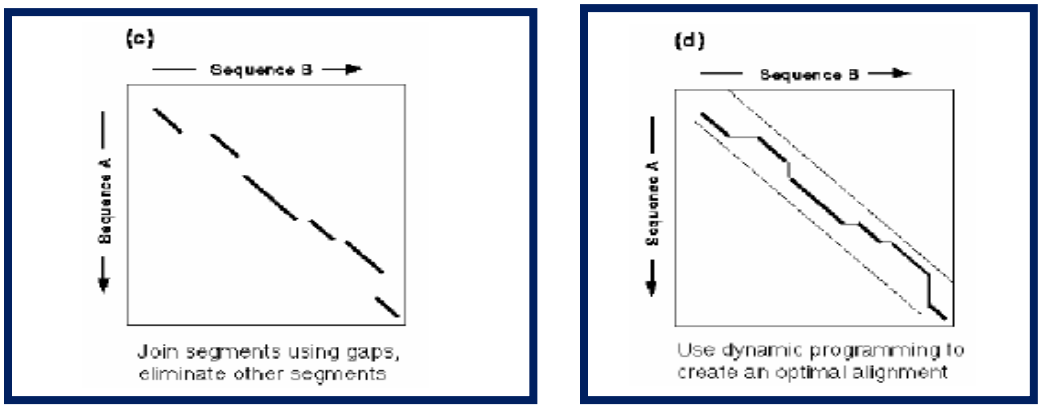


Figure 4 Fasta[57]

a) Fasta locates regions of query sequences and the search set sequence that have high densities of exact word matches.

b) The 10 highest scoring regions are re-scored using a scoring matrix. The score of the highest scoring initial region is saved as the init1 score.

c) Fasta determinates if one of the initial regions from different diagonals may be joined together to form an approximate alignment with gaps Fast determinates the best segment of similarity between the query sequence and the search set sequence.

Blast: - Basic Local Alignment Search Tool use a heuristic search algorithm use word matching ,Compare A nucleotide sequence to a nucleotide sequence database blatsn algorithm is used, Compare A protein sequence to a protein sequence database blatsp algorithm is used, Compare A translated nucleotide sequence to a protein sequence database blatsx algorithm is used ,Compare A protein sequence to a translated nucleotide sequence database tblatsn algorithm is used, Compare A translated nucleotide sequence to a translated nucleotide sequence database tblatsnx algorithm is use.

blast algorithm: - A sequence comparison algorithm used to search sequence databases for optimal local alignments to a query The initial search is done for a word of length "W" that scores at least "T" when compared to the query using a substitution matrix Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S" The "T" parameter dictates the speed and sensitivity of the search. (Figure 5)

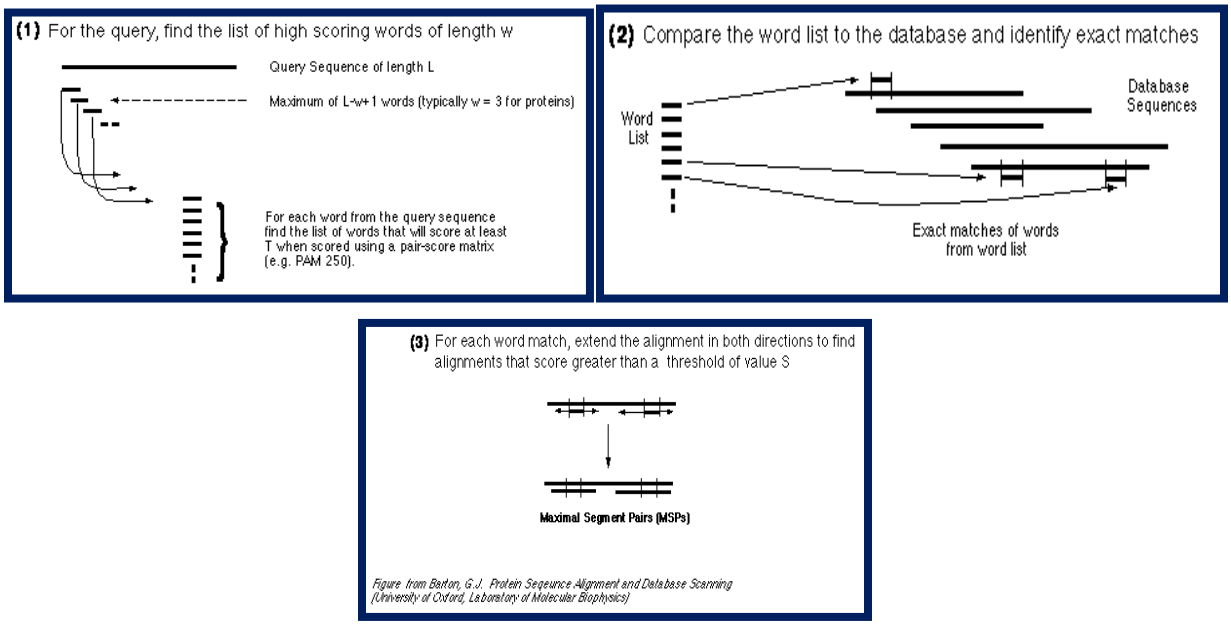


Figure. 5 blast algorithm[57]

Smith-waterman: - more sensitive search much slower uses dynamic programming dynamic programming is applicable when a large search space can be structured into a succession of stages, such that:

- The initial stage contains trivial solutions to sub-problems.
- Each partial solution in a later stage can be calculated by repeating a fixed number of partial solutions in an earlier stage the final stage contain the overall solution.[33]

Table 1 Comparison between Blast, Fast and Smith-Waterman

	Blast	Fast	Smith-waterman
Speed	High	Medium	Very low
Sensitivity	Low	Medium	High
Statistics	Probabilities	Significance fore given dataset	

2.Genome Informatics:

This is a relatively well-known topic being closely related to bioinformatics through sequence analysis. Genome informatics as a field encompasses the various methods and algorithms for analyzing and extracting biologically relevant information from the rapidly growing biological and essential sequence databases. This has lead to a new data driven research paradigm for post genomic biomedical research, which has been claimed for replacing the traditional hypothesis driven paradigm in which experiments are carefully designed to address a specific prior hypothesis. A major component of these efforts is the development and use of annotation standards such as ontologies, which provides conceptualizations of domains of knowledge and facilitates both communication between researchers and the use of domain knowledge by computers for multiple purposes. One of these kinds is the gene ontology database with AmiGO, QucikGO, GOst browsers to facilitate its access. Genome informatics is helpful in drug discovery process at a number of steps starting from the optimization of target selection, unveiling the complexity of gene expression, resolving the genetic variation at the genomic and cellular level, etc. Recent efforts in genome informatics can be categorized as genome sequence analysis, genome expression analysis, Tools for the visualization of gene network, algorithms for recognition of the coding and splicing regions, etc. A number of online resources and servers are available that assist in genome informatics research. Few of them are - FlyBase, KEGG (Kyoto Encyclopedia of Genes and Genomes), and Ensemble Compara Database, cis RED database, genome SCOUT gene RAGE, CoGenT^{++}. In India, Institute of Genomics and Integrated Biology (IGIB) is one of the leading institutes working in the field of genome informatics.

Personalized medicine is the idealized medical practice to give right drugs to right patients at right times. Finding SNPs is considered as a premise for this practice but it is by no means the sufficient effort. Good practice must be supported by well trained medical professionals who can easily access relevant data and knowledge. Such an informational environment would be called the infrastructure for personalized medicine.[37]

3. Immunoinformatics:

The immune system recognizes foreign agents (antigens) to the host organism and raises appropriate responses. Foreign includes viruses, bacteria, parasites, fungi, tumors, and transplants. The application of information technology to the study of immunologically important processes is known as immunoinformatics. It facilitates the understanding of immune function by modeling the interactions among immunological components.

Major Immunoinformatics developments include - (i) immunological databases (ii) sequence analysis and structure modeling of antibodies (iii) modeling of the immune system (iv) simulation of laboratory experiments (v) statistical support for immunological experimentation (vi) immunogenomics, etc. Over 15 immunological databases have appeared over the past few years - ex. MHCPEP (Database of MHC-Binding Peptides), FIMM (Database of Functional Immunology), KABAT (Database of Immunological Proteins), AntiJen (a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data).

The field of immunoinformatics has direct influence in the following areas: (a) improve transplantation outcomes (b) identify novel genes involved in immunological disorders (c) decipher the relationship between antigen presentation pathways and human disease (d) predict allergenicity of molecules including drugs (e) personalized medicine (f) vaccine development.[36]

4. Neuroinformatics:

In this discipline work is focused on the integration of neuroscientific information from the level of the genome to the level of human behavior. A major goal of this new discipline is to produce digital capabilities for a web-based information management system in the form of databases and associated data management tools. The databases and software tools are being designed for the benefit of neuroscientists, behavioral scientists, clinicians and educators in an effort to better understand brain structure, function, and development. Some of the databases developed in Neuroinformatics are Surface Management System (SuMS), the fMRIDC, BrainMap, BrainInfo, X-Anat, The Brain Architecture Management System (BAMS), The Ligand Gated Ion Channel database (LGICdb), ModelDB and Probabilistic atlas and reference system for the human brain. Most of these databases are freely available and can be accessed through internet. They provide the particular information in detail at one place and help in the neuroscience research. Some of the generally used neuroinformatic software tools include GENESIS, NEURON, Catacomb, Channelab, HHsim, NEOSIM, NANS, SNNAP, etc. The data sharing in neuroscience is not the only application of neuroinformatics, it is much more. The computational modeling of ion

channels, various parts of neurons, full neurons and even neural networks helps to understand the complex neural system and its working. This type of modeling greatly overlaps with system biology and also gets benefit from bioinformatics databases. In India neuroinformatics research is mainly being carrying out presently at National Brain Research Centre, Gurgaon under the department of biotechnology, government of India. The computational modeling of various processes related to neurosciences helps in understanding of brain functions in normal and various disorder states. Several efforts in this direction are also in progress.[36]

5. Toxic Informatics:

Toxicoinformatics involves the use of information technology and computational science for the prediction of toxicity of chemical molecules in the living systems. There is a growing need for computational methods which can predict toxicological profiles. There are essentially two basic approaches being used in toxic informatics.

(a) Based on modeling Structure Activity Relationship (SAR).

(b) Rule based methods.

The toxicity predictive systems using this approach include TOPKAT, MULTICASE, COMPACT, etc. The software packages DEREK, HazardExpert, OncoLogic, etc. are the rule based toxicoinformatic systems. TOPKAT (Toxicity Prediction by Komputer Assisted Technology) uses Quantitative Structure Toxicity Relationship (QSTR) regression models developed using electrotopological descriptors like electronic properties (charge, electron density, residual electronegativity and effective polarisability), connectivity descriptors, shape descriptors (kappa shape indices) and substructure descriptors from a library of 3000 molecular fragments. The predicted toxicological endpoints include: rodent carcinogenicity, Ames mutagenicity, developmental toxicity potential, skin and eye irritation, acute oral toxicity LD_{50}, acute inhalation toxicity LC_{50}, acute toxicity LC_{50}, acute toxicity EC_{50}, maximum tolerated dose (MTD), chronic lowest observable adverse effect level (LOAEL), skin sensitisation, and log P. Deductive Estimation of Risk from Existing Knowledge (DEREK) is a knowledge-based system. In the package Hazardexpert, the endpoints predicted are mutagenicity, carcinogenicity, teratogenicity, irritation, sensitisation, immunotoxicity, and neurotoxicity. It contains a knowledge base consisting of toxicophores based on literature in the QSAR field. Oncologic is a knowledge-based expert system for the prediction of chemical carcinogenicity. [36]

6. Metabolome Informatics:

Metabolomics is an emerging new ominous science analogous to genomics, transcriptomics, proteomics, etc. Metabolomics is the lesser-known cousin to genomics and proteomics. Just as

genomics is concerned with the high-throughput, global measurement of all the genes in the genome, metabolomics is concerned with the high throughput, global measurement of all the small-molecule metabolites in the metabolome. The metabolome is formally defined as the collection of all small-molecule metabolites (endogenous or exogenous) that can be found in a living cell or living organism. Metabolomics is a relatively new term, having been coined less than 5 years ago .[39] Metabolomics is also known as Metabolomics [40] or metabolic profiling. [41]

- **Metabolomics:-**The study of the total metabolite pool (metabolome), metabolic regulation and fluxes in individual cells or cell types. Can be achieved through a wide spectrum of technologic methods including LC-MS, GC-MS, and nuclear magnetic resonance (NMR).
- **Metabolomics:-**The study of the systemic biochemical profiles and regulation of function in whole organisms by analyzing a metabolite pool (metabolome) in biofluids and tissues. Usually implies that the study is done specifically through nuclear magnetic resonance profiling.
- **Metabolome:-**The quantitative complement of all the low molecular weight molecules present in cells in a particular physiological or developmental state.
- **Biofluids:-**A fluid sample obtained from a living system. The donor might typically be a human or an animal. Fluids can be excreted (such as urine, sweat), expressed or secreted (such as milk, bile), obtained by intervention (such as blood plasma, serum or cerebrospinal fluid), develop as a result of a pathological process (such as blister or cyst fluid), or be applied and collected (such as dialysis fluid).

Advantages of Metabolomics:-

- Identification of target organ, severity, onset, duration and reversal of the effects (time-course)
- Classify sample as “normal” vs. “abnormal”
- Determine mechanisms of action within the organ
- Potential for identifying novel biomarkers of toxic effect
- Non-invasive
- No a priori decisions about samples need be made
- No sample processing necessary other than cold collection
- Complete time course data can readily be obtained
- Minimization of compound requirements

- Relatively fast analysis (200-300 samples/day)
- Useful tool for modeling physiological variation and exposure conditions in animals and humans.
- Metabolite measurements have been part of organ transplant monitoring .[42]

The progress in the field is closely integrated with progress in information technology. In the field drug discovery metabolome informatics can contribute to target identification, mechanism of action, and pathways of drug toxicity. Information technologies are being used in performing (i) metabolite target analysis, (ii) metabolite profiling, (iii) metabolic fingerprinting, etc. The efforts in this field can be broadly divided into two categories drug metabolism informatics and metabolism pathway informatics. Metabolic databases generally contain the following types of information: Information about biofluids, cellular and tissue-specific metabolomes defining amino acids, vitamins, anti-oxidants, etc in them. Presently there is no complete metabolome database of any species. But different organizations are stepping towards this goal - for example, human metabolome project, Glom metabolome database, human natural products database, metabolite mass spectral database etc.

Limitations of Metabolomics:-

1. Specialized equipment is required
2. Extensive expertise is required
3. Information is limited to time- and dose-points taken
4. High risk of false positive data:-a compound causes significant metabolism changes without associated toxicity
5. Difficulty in separation of physiological (adaptive) and toxicological (adverse) effects
6. Sensitivity of the assay
7. Certain pathological states have negligible effects on biofluids:-liver fibrosis may go undetected until damage is severe.
8. Availability of biofluids for certain organ toxicity: CNS vs. Urine.
9. Distinguishing effects of multi-organ toxicants:-biomarkers in different biofluids are different and in one bio-fluid are inter-mixed.

Future Directions in Metabolomics:-

1. Develop comprehensive metabonomic database.
2. Expand metabonomics applications to many species.
3. Evaluate cryoprobe technology for increased sensitivity or increased throughput.
4. Expand technology to novel targets:

– Cardiac toxicity

– Adrenal toxicity

5. "Grand Unification" of Genomic/Proteomic and Metabonomic technologies.

Fields of application of Metabolomics

- Plant breeding and assessment of crop quality
- Food assessment and safety
- Toxicity assessment
- Nutrition assessment
- Medical diagnosis and assessment of disease status
- Pharmaceutical drug development
- Yield improvement in crops and fermentation
- Biomarker discovery
- Technological advances in analytical chemistry
- Genotyping
- Environmental adaptations
- Gene-function elucidation
- Integrated systems biology

7. Healthcare Informatics:-

Biomedical Informatics is an emerging discipline that has been defined as the study, invention, and implementation of structures and algorithms to improve communication, understanding and management of medical information." Medical informatics is more concerned with structures and algorithms for the manipulation of medical data, rather than with the data itself. This suggests that one difference between bioinformatics and medical informatics as disciplines lies with their approaches to the data there are bioinformatics interested in the theory behind the manipulation of that data and there are bioinformatics scientists concerned with the data itself and its biological implications.[36]

Medical informatics, biomedical informatics, clinical informatics, nursing informatics, etc. come under the service-oriented sectors. Other topics like cancer informatics, diabetes informatics are specific therapeutic area based information technology topics. These topics are also related to pharmacoinformatics as a whole because the information obtained from these subjects leads to decision making in pharmaceutical industry. For example, medical informatics deals with medicines and health care. The databases associated with this filed include the feedback received

from the patient care. Analysis of the data can be applied in deciding the trends in the patient response to a drug. Thus, future designing of the drugs can be made to suit the needs of the patients. Electronic health record (EHR) systems, Hospital Information Systems (HIS), Decision Support Systems (DSS), etc. are the major components of healthcare informatics.

Medical Information Science is the science of using system-analytic tools to develop procedures (algorithms) for management, process control, decision making and scientific analysis of medical knowledge - *Ted Shortliffe*[43]

Medical Informatics comprises the theoretical and practical aspects of information processing and communication, based on knowledge and experience derived from processes in medicine and health care - *Jan van Bemm.*[43]

Medical Informatics (MI) is the study of information processing as it is used in healthcare. It might have been called medical computing, but the French-derived term informatics is more commonly used internationally and probably conveys a broader set of concerns, including the uses and flows of information that may have little to do with computers. Like many engineering fields, MI has scientific aspects that focus on the description, modeling and interpretation of how information is actually generated, disseminated and used, and underlying constraints or natural laws that govern these activities. MI is also deeply concerned with design of appropriate medical information processing systems, with tradeoffs in their implementation, and with ways to evaluate their effectiveness. Some have suggested health informatics as a better, broader term, meant to encompass aspects of health care that are not traditionally the focus of medicine, such as preventive care, nutrition, patient education, epidemiology, etc. Related terms include bioinformatics, which is the study of information processing in biological sciences. Opinion currently varies on whether bioinformatics is part of medical informatics, or-if it forms a distinct discipline—how it relates. Most expect that progress in understanding the molecular basis of disease will bring these fields closer together, if not to merger. Telemedicine (or the recent European coinage telematique) focuses on one aspect of MI, access to and use of medical information at a distance. At MIT, in line with our traditions of institutional flexibility, we have no official organization that does medical informatics, but a number of small foci around the research and teaching interests of faculty in different Departments and Laboratories.—MIT[44]

Pharmacoinformatics preventing adverse drug reactions in hospital patients.[44,45] Health informatics is concerned with the systematic processing of data, information and knowledge in medicine and healthcare, increasingly delivered by a mix of public and private organisations. Health informatics is delivered by operational health practitioners, academic researchers and

educators, scientists and technologists in operational, commercial and academic domains. The ultimately focus is to improve patient safety and organisational effectiveness to achieve better outcomes.

Nursing informatics (NI)

is a specialty that integrates nursing science, computer science, and information science to manage and communicate data, information, knowledge, and wisdom in nursing practice. NI supports consumers, patients, nurses, and other providers in their decision making in all roles and settings. This support is accomplished through the use of information structures, information processes, and information technology opportunities in Health broadly cover the following facets, sometimes in combination and with grey boundaries between them:

- **Information Management –** to collect, collate, analyse and present information to many audiences – professional, management and lay
- **Knowledge Management –** to handle and evaluate a complex evidence base, coming from many (inter)national and academic/commercial and operational sources.
- **Research and Development –** where new tools are brought to market, innovative theories are tested and applied and, with emerging standards, are evaluated for future implementation.
- **ICT –** where efforts are focussed on ensuring solutions operate efficiently, users can get the best out of their use, and new technologies and systems are robustly tested before deployment.
- **Specialist Clinical Informatics** - where effective informatics includes deep knowledge of clinical conditions, disease knowledge and close involvement in direct patient care.

8. Chemo informatics:-

Chemo informatics is the application of informatics methods to solve chemical problems. All areas of chemistry from analytical chemistry to drug design can benefit from chemo informatics methods. And there are still many challenging chemical problems waiting for solutions through the further development of chemo informatics. [46, 47]

The term "Chemoinformatics" appeared a few years ago and rapidly gained widespread use. Workshops and symposia are organized that are exclusively devoted to Chemoinformatics, and many job advertisements can be found in journals. The first mention of Chemoinformatics may be attributed to Frank Brown. [48]

Chemo informatics is the arrangement of information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and organization. So chemo informatics is helpful in drug design, Greg Paris came up with a much broader definition. [52]**Chemical Data →Storage in Databases →Data Information →Data Retrieval →Analysis**

The current schema of chemoinformatics in drug designing is given below: Analysis of predesigned drug structure structural property prediction (QSAR) property prediction by smiles format perform some modification in prior drug again predict the drug property if variation occurs in novel structure save that structure and design a fragment library.Chemoinformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization, and use of chemical information.[53]

The need for chemoinformatics

Recent chemical developments for drug discovery are generating a lot of chemical data. These developments are combinatorial chemistry and high-throughput screening. Some scientists have described this situation as a chemical information explosion. This has created a demand to effectively collect, organize, and apply the chemical information. [54, 55]

Chemo informatics which deals with the information of the molecules, chemical reaction informatics also plays an important role in the field of pharmacoinformatics. Chemical reaction informatics enable a chemist to explore synthetic pathways, quickly design and record completely new experiments from scratch or by beginning with reactions found in the reaction databases.

Chemical reaction informatics a database consists of the following information –

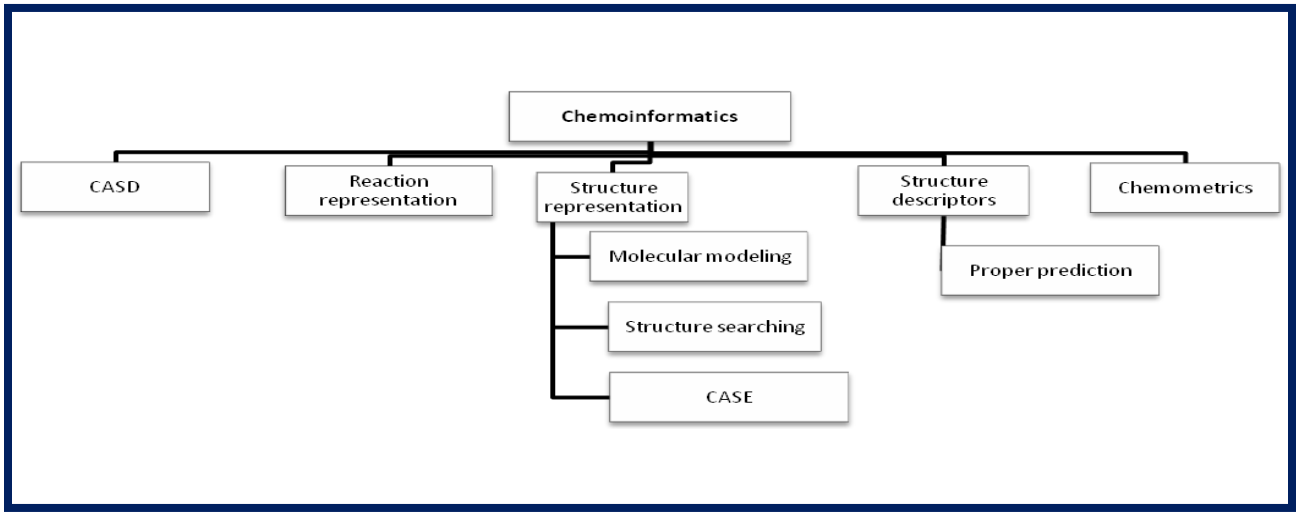
- Reactants and products
- Atom mapping, which allows you to tell which atom, becomes which product atom through the reaction.
- Information regarding reacting center(s)
- The catalyst used
- The atmosphere, including pressure and composition
- The solvent used
- Product yield
- Optical purity
- References to literature.

The chemical reaction informatics would essentially assist the chemist in giving access to reaction information, in deriving knowledge on chemical reactions, in predicting the course and outcome of chemical reactions, and in designing syntheses.

Specifically, the following tasks can be accomplished by analysis tools in chemical reaction informatics –

a) Storing information on chemical reactions
b) Retrieving information on chemical reactions
c) Comparing and analyzing sets of reactions
d) Defining the scope and limitations of a reaction type
e) Developing models of chemical reactivity
f) Predicting the course of chemical reactions
g) Analyzing reaction networks
h) Developing methods for the design of syntheses, etc.

Area of Chemo informatics



Source: Lipinski, C.A *et.al.* (1997)

1. Representation of Chemical Compounds

All chemical compounds and there structures are arranged in the form of computer representation and coded. Special methods had to be advised to uniquely represent a chemical structure, to comprehend features such as rings, aromaticity, stereochemistry, 3D structures and molecular surfaces.[59]

2. Representation of Chemical Reactions

Chemical reactions are represented by the starting materials and products as well as by the reaction conditions. On top of that, one also has to indicate the reaction site, the bonds broken

and made in a chemical reaction. Furthermore, the stereochemistry of reactions has to be handled.

3. Data in Chemistry

Much of our chemical knowledge has been derived from data. Chemistry offers a rich range of data on physical, chemical, and biological properties: binary data for classification, real data for modeling, and spectral data having a high information density. These data have to be brought into a form amenable to easy exchange of information and to data analysis.[56]

4. Data sources and Databases

The enormous amount of data in chemistry has led quite early on to the development of databases to store and disseminate these data in electronic form. Databases have been developed for chemical literature, for chemical compounds, for 3D structures, for reactions, for spectra, etc. The internet is increasingly used to distribute data and information in chemistry.[56]

5. Structure Search Methods

In order to retrieve data and information from databases, access has to be provided to chemical structure information. Methods have been developed for full structure, for substructure, and for similarity searching. A connection table is essentially a representation of the molecular graph. Therefore, for storing a unique representation of a molecule and for allowing its retrieval, the graph isomorphism problem had to be solved to define from a set of potential representations of a molecule a single one as the unique one. The first solution was the Morgan algorithm for numbering the atoms of a molecule in a unique and unambiguous manner.[63]This provided the basis for full structure searching. Then, methods were developed for substructure searching, for similarity searching, and for 3D structure searching.

6. Quantitative Structure Activity / Property Relationship (QSAR/QSPR)

Building on work by Hammett and Taft in the fifties, Hansch and Fujita showed in 1964 that the influence of substituents on biological activity data can be quantified.[61]In the last 40 years, an enormous amount of work on relating descriptors derived from molecular structures with a variety of physical, chemical, or biological data has appeared. These studies have established Quantitative Structure–Activity Relationships (QSAR) and Quantitative Structure-Property Relationships (QSPR) as fields of their own, with their own journals, societies, [62] and conferences.

7. Chemometrics

Initially, the quantitative analysis of chemical data relied exclusively on multi linear regression analysis. However, it was soon recognized in the late sixties that the diversity and complexity of

chemical data need a wide range of different and more powerful data analysis methods. Pattern recognition methods were introduced in the seventies to analyze chemical data. In the nineties, artificial neural networks gained prominence for analyzing chemical data.[63]The growing of this area led to the establishment of chemometrics as a discipline of its own with its own society, [64] journals, and scientific meetings.

8. Methods for Calculating Physical and Chemical Data

A variety of physical and chemical data of compounds can directly be calculated by a range of methods. Foremost are quantum mechanical calculations of various degrees of sophistication. However, simple methods schemes can also be used to estimate a variety of data with reasonable accuracy.[56]

9. Calculation of structure descriptors

In most cases, however, physical, chemical, or biological properties cannot be directly calculated from the structure of a compound. In this situation, an indirect approach has to be taken by, first, representing the structure of the compound by structure descriptors, and, then, to establish a relationship between the structure descriptors and the property by analyzing a series of pairs of structure descriptors and associated properties by inductive learning methods. A variety of structure descriptors has been developed encoding 1D, 2D, or 3D structure information or molecular surface properties.[56]

10. Molecular modeling

In the late sixties, R. Langridge and coworkers developed methods for visualizing 3D molecular models on the screens of Cathode Ray Tubes. At the same time, G. Marshall started visualizing protein structure on graphic screens. The progress in hardware and software technology, particularly as concerns graphics screens and graphics cards, has led to highly sophisticated systems for the visualization of complex molecular structures in great detail. Programs for 3D structure generation, for protein modeling, and for molecular dynamics calculations have made molecular modeling a widely used technique.

11. Data Analysis Methods

A variety of methods for learning from data, of inductive learning methods is being used in chemistry, statistics, pattern recognition methods, artificial neural networks, genetic algorithms. These methods can be classified into unsupervised and supervised learning methods and are used for classification or quantitative modeling.[56]

12. Computer-Assisted Structure Elucidation (CASE)

The elucidation of the structure of a chemical compound, be it a reaction product or a compound isolated as a natural product, is one of the fundamental tasks of a chemist. Structure elucidation has to consider a wide variety of different types of information mostly from various spectroscopic methods, and has to consider many structure alternatives. Thus, it is an ambitious and demanding task. It is therefore not surprising that chemists and computer scientists had taken up the challenge and had started in the 1960 's to develop systems for computer-assisted structure elucidation (CASE) as a field of exercise for artificial intelligence techniques. The DENDRAL project, initiated in 1964 at Stanford University gained widespread interest.[65] Other approaches to computer-assisted structure elucidation were initiated in the late sixties by Sasaki at Toyohashi University of Technology[66] and by Munk at the University of Arizona. [67]

13. Computer-Assisted Synthesis Design (CASD)

The design of a synthesis for an organic compound needs a lot of knowledge about chemical reactions and on chemical reactivity. Many decisions have to be made between various alternatives as to how to assemble the building blocks of a molecule and which reactions to choose. Therefore, computer-assisted synthesis design (CASD) was seen as a highly interesting challenge and as a field for applying artificial intelligence techniques.

In 1969 Corey and Wipke presented their seminal work on the first steps in the development of a synthesis design system.[68] Nearly simultaneously several other groups such as Ugi and coworkers,[69] Hendrickson[70], and Gelernter[71] reported on their work on CASD systems. Later also at Toyohashi work on a CASD system was initiated.[72]

Extracting knowledge from chemical information lots of data (structure, activities, genes, etc) i.e. called as inductive learning. When we extract data from knowledge, it is called as deductive learning.[59] (figure 6)

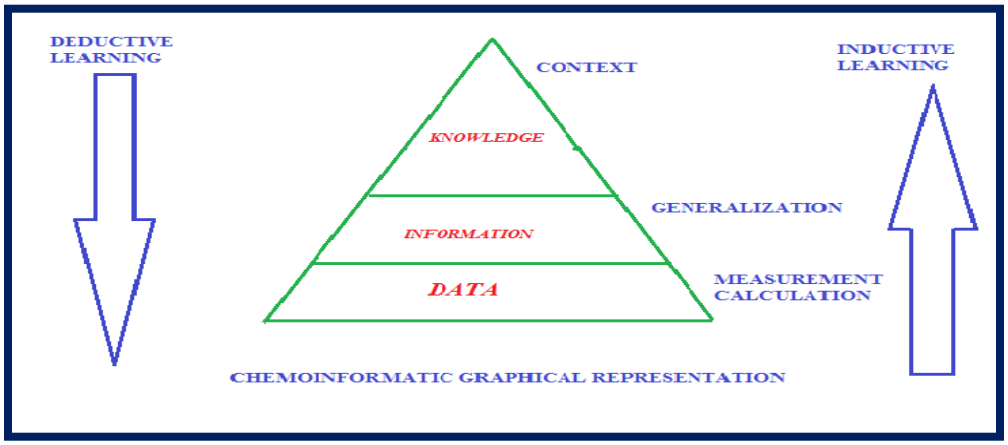


Figure. 6 Chemoinformatics in graphical form

Applications of Chemo informatics [46]

1. Chemical Information

- storage and retrieval of chemical structures and associated data to manage the flood of data
- dissemination of data on the internet
- cross-linking of data to information

2. All fields of chemistry

- Prediction of the physical, chemical, or biological properties of compounds.

3. Analytical Chemistry

- analysis of data from analytical chemistry to make predictions on the quality, origin, and age of the investigated objects
- elucidation of the structure of a compound based on spectroscopic data

4. Organic Chemistry

- prediction of the course and products of organic reactions
- design of organic syntheses

5. Drug Design

- identification of new lead structures
- optimization of lead structures
- establishment of quantitative structure activity relationships
- comparison of chemical libraries
- definition and analysis of structural diversity
- planning of chemical libraries
- analysis of high-throughput data
- docking of a ligand into a receptor
- de novo design of ligands
- modeling of ADME-Tox properties
- prediction of the metabolism of xenobiotics
- analysis of biochemical pathways

6. Textile Industry:

Combinatorial organic synthesis (COS), high through-put screening (HTS), and chemo informatics (CI) are highly efficient and cost-effective tools to develop novel, state-of-the-art,

non-toxic chemicals (*e.g.* dyes, colorants, finishes, pigments, surfactants, etc.) of commercial importance to the textile industry.

CONCLUSION:-

As discussed above there is several information technology efforts related to the pharmaceutical sciences which are useful for drug discovery. In future, these efforts are expected to grow both in terms of their reliability and scope. Thus, this emerging technology (pharmacoinformatics) is becoming an essential component of pharmaceutical sciences.

REFERENCE:-

1. What is Pharmacoinformatics? Pharmacoinformatics Working Group American Medical Informatics Association. Available from: http://archive.amia.org.
2. Focus on Pharmaco Informatics. [Cited 2005 Oct 30]. Available from: http://www.amia.org/mbrcenter/wg/pharmi/focus.asp.
3. "Bioinformatics ";080415 ,Bibliotheca Alexandrina ,Updated by Mariam Salib & Marwa Abdelrassoul
4. Luscombe NM, Greenbaum D, Gerstein M. Review "What is bioinformatics? An introduction and overview. In Yearbook of Medical Informatics. New Haven (USA): Department of Molecular Biophysics and Biochemistry Yale University; 2001; 83-99.
5. Bioinformatics and Medical Informatics Bioinformatics and Medical Informatics in the College of Sciences OFFICE: Geology/Mathematics/Computer Science 41, SDSU GRADUATE BULLETIN 2008-2009/363-64.
6. www.cats.ucsc.edu/
7. David R, Westhead, J, Howard P, Richard MT, Bioinformatics. Oxford: BIOS; 2002.
8. Fox J. What is Bioinformatics? University of British Columbia: UBC Bioinformatics Centre. Available from: http://bioinformatics.ubc.ca/node/368/print.
9. Kanaujia S,Bioinformatics and Internet: New Paradigm to Disciplines. SRELS Journal of Information Management 2004 Mar; 41(1):43-55.
10. www.cato.com/biotech/
11. www.the-scientist.com/
12. www.nbif.org/links/1.20.php
13. www.genomics.phrma.org/today/
14. www.gnn.tigr.org/main.shtml
15. www.ebi.ac.uk/ismb-97/papers2.html

16. www.cgl.ucsf.edu/psb/
17. Huang S. The Genetic Equidistance Result of Molecular Evolution is Independent of Mutation Rates. J Comput Sci Syst Biol 2008; 1:092-102.
18. Mudunuri SB, Rao AA, Pallams etty S, Mishra P, Nagarajaram HA. VMD: Viral Microsatellite Database-A Compre-hensive Resource for all Viral Microsatellites. J Comput Sci Syst Biol 2009; 2:283-86.
19. Ingale A. In Silico Homology Modeling and Epitope Prediction of Nucleocapsid Protein region from Japanese Encephalitis Virus. J Comput Sci Syst Biol 2010;3: 53-58.
20. Nath M, Goel A, Taj G, Kumar A. Molecular Cloning and Comparative In silico Analysis of Calmodulin Genes from Cereals and Millets for Understanding the Mechanism of Differential Calcium Accumulation. (2010) , J Proteomics Bioinform 2010;3:294-301.
21. Garg N, Pundhir S, Prakash A, Kumar A. PCR Primer Design: DREB Genes. J Comput Sci Syst Biol 2008; 1:21-40.
22. Chandra SRA, Somayajulu DVLN. Influenza Classification from Nucleotide Sequence Database. J Comput Sci Syst Biol 2011; 4:77-80.
23. Anandakumar S, Shanmughavel P. Computational Annotation for Hypothetical Proteins of Mycobacterium Tuberculosis. J Comput Sci Syst Biol 2008; 1:50-62.
24. Galperin MY. The molecular biology database collection. Nuc Acids Res 2006; 34: 3-5.
25. Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvear J. The protein information resource. Nuc Acids Res 2003; 31:345–47.
26. Sequence Similarity Searching [Internet]
Available from: www.med.nyu.edu/rcr/rcr/course/PPT/similarity.ppt
27. Stuart M. Brown, Bioinformatics Tools. www.asmusa.org/edusrc/Ugconf02PPT/ StuartBrown%5B2%5D.pdf
28. Nilges, Linge JP. Bioinformatics – a definition. Available from:www.pasteur.fr/recherche/unites/Binfs/definition/ bioinformatics_definition.pdf .
29. Stuart M. Brown, Multiple Alignment www.med.nyu.edu/rcr/nccu/align.ppt.
30. http://www.ebi.ac.uk/
31. www.ncbi.nlm.nih.gov/blast/
32. Nishant T, Arun K, Sathish Kumar D, Vijaya Shanti B. Biological Databases-Integration of Life Science Data. J Comput Sci Syst Biol 2011:4:87-92.

33. Manikandakumar K, Kumaran MS, Srikumar R. Matrix Frequency Analysis of Oryza Sativa (japonica cultivar-group) Complete Genomes. J Comput Sci Syst Biol 2009;2: 159-66.
34. Pandarinath P, Shashi M, Appa Rao A. Bioinformatic Approach for the Identification of Hepatitis B Viral Insert in the Exon Region of Human Genome. J Comput Sci Syst Biol 2010; 3:89-90.
35. .Razia M, Raja KR, Padmanaban K, Sivaramakrishnan S, Chellapandi P. A Phylogenetic Approach for Assigning Function of Hypothetical Proteins in Photorhabdus luminescens Subsp. Laumondii TT01 Genome. J Comput Sci Syst Biol 2010; 3:21-29.
36. Bharatam PV. Pharmacoinformatics - Expanding Horizons [cited 2006 Jan 10]. Available from: www.pharmabiz.com
37. Genome Informatics [internet] Pharmacoinformatics Infrastructure for Genome Based Personalized Medicine Kotoko Nakata1 Tatsuya Nakano PHII/Pharmaceutical Information Infrastructure: [cited 2000; 11:392–93.]
38. Wishart DS. Mini review Metabolomics: The Principles and Potential Applications to Transplantation. American Journal of Transplantatio*n* 2005; 5:2814-20.
39. Drysdale R, Bayraktaroglu L. Current awareness. Yeast 2000; 17:159–66.
40. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 1999; 29:1181–89.
41. Thompson JA, Markey SP. Quantitative metabolic profiling of urinary organic acids by gas chromatography-mass spectrometry: comparison of isolation methods. Anal Chem 1975; 47:1313–21.
42. Salaman JR. Monitoring of rejection in renal transplantation. Immunol Lett 1991; 29: 139–42.
43. Moco S, Bino RJ, De Vos RCH, Vervoort J. Metabolomics technologies and metabolite identification. Trends in Analytical Chemistry 2007; 26:9.
44. Maojo V, Iakovidis I, Martin-Sanchez F, Crespo J, Kulikowski C. Medical Informatics and Bioinformatics: European Efforts to Facilitate Synergy. Journal of Biomedical Informatics 2002; 10:1042.

45. Evans RS, Pestotnik SL, Classen DC, Horn DS, Bass SB, Burke JP. Pharmacoinformatics preventing adverse drug reactions in hospital patients. The Annals of Pharmacotherapy 1994 Apr; 8:523.
46. Gasteiger J, Funatsu K. Chemoinformatics – An Important Scientific Discipline. J Comput Chem Jpn 2006; 5(2):53–58.
47. Gasteiger J. Chemoinformatics: a new field with long tradiation. Anal bioanal Chem 2006; 384:57-64.
48. BROWN F. Chemoinformatics: What is it and How does it impact drug discovery. Annu Rep Med Chem 1998; 33:375–84.
49. Tudor I editor. Introduction to Chemoinformatics in Drug Discovery – A Personal View Garland R. Marshall Chemoinformatics in Drug Discovery. ISBN: 3-527-30753-2
50. Canadian Nurses Association. CNA hosts social justice session [cited 2003 Aug]. Available from: http://www.cnanurses.ca/CNA/documents/pdf/publications/Access_August_2003_e.pdf
51. Commission on Social Determinants of Health. Action on the social determinants of health: Learning from previous experiences. World Health Organizations [cited 2005] Available from: http://www.who.int/social_determinants/strategy/en/CSDH_socialdet_backgrounder.pdf
52. Paris G, Warr W. American Chemical Society [2009 Aug]. Available from: http://www.warr.com/warrzone.htm
53. Shukla1 S, Choubey1 SK, Srivastava1 P, Gomase VS. Chemoinformatics-An emerging field for Computer Aided Drug Design. Journal of Biotechnology Letters 2010;1(1):10-14.
54. Jeffery Loo NLM. Introduction to Chemoinformatics. Avaiable from: jloo1@jhmi.edu.
55. Gasteiger J, Engel T, editors. Chemoinformatics: A Textbook. Weinheim (Germany): wiley-VCH; 2003; 649.
56. Aktar MW, Murmu S. Chemoinformatics: Principles and Applications. Available from: www.shamskm.com
57. Geoffrey JB ,Protein Sequence Alignment and Database Scanning Protein Structure prediction - a practical approach, Edited by M. J. E. Sternberg, IRL Press at Oxford University Press, 1996, ISBN 0 19 9634963.

58. Tate FA. Annual Review of Information Science and Technology. Ann Rev Inf Sci Technol 1967; 2:285-309.
59. Dyson GM, Lynch MF, Morgan HL. A modified IUPAC-Dyson notation system for chemical structures Inform. Inf Storage Retriev 1968; 4:27-83.
60. Morgan HL. Computer handling of generic *chemical* structures. J Chem Docum 1965; 5:107-13.
61. Hansch C, Fujita T. J Am Chem Soc 1964; 86:856-64.
62. QSAR and Modelling Society. Available from: http://www.qsar.org
63. Zupan J, Gasteiger J. Neural Networks in Chemistry and Drug Design. 2nd ed. Weinheim: Wiley-VCH; 1999.
64. International Chemometrics Society: Available from: http://www.mamics.nysaes.cornell.edu/chem-society.html.
65. Lindsay RK, Buchanan BG, Feigenbaum EA, Lederberg J. Applications of Artificial Intelligence for Organic Chemistry; the Dendral Project. New York: McGraw-Hill; 1980.
66. Sasaki SI, Abe H, Ouki T, Sakamoto M, Ochiai S. Automated structure elucidation of several kinds of aliphatic and alicyclic compounds. Anal. Chem 1968; 40:2220-23.
67. Shelley CA, Hays TR, Munk ME, Roman RV. An Approach to Automated Partial Structure Expansion Anal Chim Acta 1978; 103:121-132.
68. Corey EJ, Wipke WT. Computer-assisted design of complex organic syntheses. Science. 1969 Oct 10; 166(3902):178–192.
69. Blair J, Gasteiger J, Gillespie C, Gillespie PD, Ugi I. Representation of the Constitutional and Stereochemical Features of Chemical Systems in the Computer Assisted Design of Syntheses. Tetrahedron 1974; 30: 1845-59.
70. Hendrickson JB. A Systematic Characterization of Structures and Reactions for Use in Organic Synthesis. J Am Chem Soc 1971; 93:6847-54.
71. Gelernter HL, Sridharan NS, Hart AJ, Yen SC. The Discovery of Organic Synthetic Routes by Computer. Top Curr Chem 1973; 41:113-50.
72. Funatsu K, Sasaki S. Computer-assisted organic synthesis design and reaction prediction system, "AIPHOS", Tetrahedron Computer Methodology 1988; 1:1:27-37.