1. Scientific Computing and Applied Statistics in Life Sciences:
Scientific Computing and Life Sciences have gained the most important place in the field of science and technology in the last decade. They are positioned at the intersection of the high performance computing, quantitative modeling and modern biology. The trio - Scientific computing, Life Sciences and Applied Statistics altogether help provide a fundamental understanding of complex biological systems and offers the potential to significantly impact a wide variety of technologies, including drug discovery, novel therapies for human, animal and plant diseases, metabolic engineering and efficient production of traditional and high-value foodstuffs.

1.1. Scientific Computing:
Computational science (also scientific computing or scientific computation) is concerned with constructing mathematical models and quantitative analysis techniques and using computers to analyze and solve scientific problems. In practical use, it is typically the application of computer simulation and other forms of computation from numerical analysis and theoretical computer science to problems in various scientific disciplines.[1]

The essence of computational science is numerical algorithm[2] and/or computational mathematics. In fact, substantial effort in computational sciences has been devoted to the development of algorithms, the efficient implementation in programming languages, and validation of computational results. Scientific computing is the cross-disciplinary field at the intersection of modeling scientific processes, and the use of computers to produce quantitative results from
these models. It is what takes a domain science and turns it into a computational activity. As a definition, we may describe it: “The efficient computation of constructive methods in applied mathematics.”[3]

This clearly indicates the three branches of science that scientific computing touches on:

- **Applied mathematics:** The mathematical modeling represents the models of the real-world phenomena. Such modeling lead to the implicit descriptions. Complex problems can be concluded with certain mathematical calculations that finally take the shape of models. In order to obtain actual tangible results we need a constructive approach, which can be arrived at by applying such mathematical models.

- **Numerical analysis:** Numerical analysis is the area of mathematics and computer science that creates, analyzes, and implements algorithms for solving numerically the problems of continuous mathematics. Such problems originate generally from real-world applications of algebra, geometry and calculus, and they involve variables which vary continuously; these problems occur throughout the natural sciences, social sciences, engineering, medicine, and business[4]. Numerical analysis varies from quite theoretical mathematical studies to computer science issues.

- **Computer Science** that takes numerical algorithms and program them to automate the process of implementing them. It analyzes the efficacy of implementing them on actually existing, rather than hypothetical, computing engines.[5].

**1.1.1 Methods and Algorithms:**
There are numerous methods which prove vital to the application of Computational Science in various fields. Such applications include pattern recognition, machine learning, multistage modeling of biological data, metrology, geometric modeling, tomography, signal and image processing etc.
Various Methods/Mathematical Models used in Scientific Computing:

The methods used in various applications of Scientific Computing are varied. The methods which are commonly applied include:

- Numerical analysis
- Application of Taylor series as convergent and asymptotic series
- Computing derivatives by Automatic differentiation (AD)
- Computing derivatives by finite differences
- Graph theoretic suites High order difference approximations via Taylor series and Richardson extrapolation
- Methods of integration on a uniform mesh: rectangle rule (also called midpoint rule), trapezoid rule, Simpson’s rule
- Runge Kutta method for solving ordinary differential equations

Use of Various Algorithms in Scientific Computing:

Real-world changing conditions are modeled and implemented by the Scientific Computing application programs. Such application include weather forecasting, flow and pattern of air around a plane, automobile body distortions in a crash, the motion of stars in a galaxy, an explosive device, etc. Each specific condition/problem contains a set of vague information, which are identified, refined, defined and recorded in the form of various absolute as well as discrete data sets. Each data set corresponds to an area in space and contains information about that space relevant to the model. Consider for example, the information pertaining to the weather models that include discrete flows of details viz areas under consideration in square kilometer; land elevation, current wind direction, humidity, temperature, pressure, etc. The algorithm based on the data fetched from such information would calculate the likely next state based on the current state, in simulated time steps, solving equations that describe how the system operates. The entire process is then repeated to calculate the next state.
1.1.1.2 Computer Modeling and Simulation:

A model is a physical, mathematical, or logical representation of a system entity, phenomenon, or process. A simulation is the implementation of a model over time. A simulation brings a model to life and shows how a particular object or phenomenon will behave. It is useful for testing, analysis or training where real-world systems or concepts can be represented by a model [7]

![Figure 1.1.1.2: Emergence of computer based Modeling and Simulation [24]](image)

- Various classes of computer based modeling and simulation:

The three classes of models and simulations are: virtual, constructive, and live:

1. Virtual simulations represent systems both physically and electronically. Examples are, the, Close Combat Tactical Trainer, certain modules for the delivery of built-in training, various scientific models which function on computational powers and appear physically for the further gesture.

2. Constructive simulations represent a system and its employment. They include computer models, analytic tools, mockups, IDEF, Flow Diagrams, and some engineering software tools like Computer-Aided Design/ Manufacturing (CAD/CAM).

3. Live simulations are simulated operations with real operators and real equipment. Examples are fire drills, operational tests, and the systems that work at laboratories to aid to the medical functionalities.
• Goals/benefits of computer based modeling and simulation:

The following are the main goals or benefits derived from the computer based modeling and simulation:

• Environment structured to solve the specific problem allows the customization of the process resulting in ease and convenience in handling data.
• Fast data entry as well as retrieval save the time and assures efficiency.
• Expert system technology generates details automatically while windows and pop-up menus guide the user through the modeling process.
• Changes can be made quickly and easily with far less chances of errors.
• Forms built and embedded to capture the data make the user more productive, so that the time wasted in programming efforts becomes less.
• Verification and testing of designs become fast. More available alternatives make the system more user friendly.
• Most of the Simulation systems can be embedded with 3-D graphics that make more précised pictorial representation of the generated data.
• Real time simulations allow the user to obtain the accurate results based on the real time reflections.

1.1.2 Computational Life Sciences:

Rapid advances in computer science, biology, chemistry, and other disciplines are enabling powerful new computational tools and models for simulating the applications of various faculties of applied sciences into the field. These computational tools hold tremendous promise for advancing applied and basic science. They provide the environment for simplifying the process from streamlining drug efficacy and safety testing to increasing the efficiency and effectiveness of risk assessment for environmental chemicals.
The Computational approaches to Life Sciences also offer the potential to improve experimental design and reduce the overall number of experimental trials needed. Computational approaches are ideally suited to organize, process, and analyze the vast libraries and databases of scientific information and to simulate complex scientific phenomena.

1.1.2.1 Computational life sciences task categories:

1. In silico drug discovery:

In silico drug discovery involves identifying potentially biologically active substances by modeling large numbers of complex virtual molecular interactions; it is generally considered to be a extremely computationally intensive task. Screening the large libraries of candidate molecules for activity, and employing more sophisticated mathematical models of binding and interaction and such other tasks which are currently limited by computing power is solved by supplying the Peak Computing Facility. In the next few years in silico drug discovery will benefit from advances in computational biomodelling of physiological processes, allowing prediction not just of inter-molecular interactions, but subsequent physiological effects.
2. **Clinical Research Informatics:**
   The development of new technologies for generating large amounts of 'state' data on biological samples [including genomics, proteomics, metabolomics, high-throughput sequencing] allows a far more complex picture to be recorded for disease progression in individual patients. Clinical Research Informatics - using statistical and data mining techniques to identify predictors of patient outcome from clinical datasets including patient histories, pedigrees, physiological measurements and treatment regimes - will fundamentally change, moving beyond relatively small numbers of parameters and predictors, towards large scale genomic, metabolomic and proteomic datasets which will enable far more detailed associations between patient and disease progression, for improved outcomes.

3. **Biomedical Image Analysis:**
   Rapid increases in the resolution and usage patterns of, and data generated by, modern imaging technologies (Functional MRI, PET, multislice CT), both clinical and research, necessitate an corresponding increase in computational resources to process and analyse the image data. Many imaging projects are currently restricted by limited available computing power.

4. **Biomolecular Structure Prediction:**
   HPC advantage: Structure prediction from physical principles is very computationally demanding and currently the available power is limiting in both the sophistication of the modelling approaches and the size of molecule that can be modelled; more computing power will immediately allow the simulation of more complex biomolecules.

5. **Computational biomodelling** envelops a large amount of simulation technologies spanning cell, tissue, organ and organ system modelling, physiological state modelling, through to social interactions and population dynamics. These areas have common necessities in computational methods for solving various simultaneous differential equations, parameter estimation, probabilistic interaction modeling, population dynamics and other methods involving many interacting units.
6. Bioinformatics (or Computational 'omics'): Bioinformatics refers to the computational analysis of the information encoded in the biological data and their expression patterns. The field is very vast and implements various high-throughput quantification methods such as microarray binding assays.

1.1.2.2 Major computational methods for life sciences computation:
There are a number of computational methods which support the life sciences computations. Depending on the computational requirements, single or the combination of methods can be put to use for varied tasks. They are:

1. Data Mining:
It is an interdisciplinary subfield of computer science[9]. It can be regarded as the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The main aim of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

2. Mathematical Modelling:
It involves describing the system using mathematical concepts and language. The mathematical modeling in life sciences generally includes the problems that thrive upon the mathematical solutions for the given problems. Eg. Differential equations, fitting the curves to lines etc.. They are designed for almost all the scientific disciplines including natural and applied sciences as well as the field of engineering.

3. Image analysis and visualization:
Image analysis generally deals with the extraction of important information from the images. Digital image processing techniques can be used for the purpose of segregating vital information from mainly digital images. These techniques are inspired by human visualization perception models.

Visualization is an effective way to communicate the concrete and abstract ideas through visual imageries. It has an ever expanding applications in all the sciences to
facilitate the acknowledgement of simple to complex information through the visual graphics.
Both the image analysis and visualization are used together in the field of life sciences to register, segment, reconstruct and render various images.

4. Sequence Analysis:
Various Bioinformatics methodologies including sequence alignment, searching biological databases and other such techniques are used to understand the features, functions, structures or evolution of DNA, RNA or such other peptide sequences[10]. They are used for the pattern matching, gene identification application etc.

5. Data Indexing:
Uncontrollable flow of large amount of data has generated huge databases, which can be possibly arranged for the quick search by the technique of Data Indexing. It can be defined as quite large datasets which are likely to be commonplace soon because of the next-gene sequencing and image technologies. Data indexing would include scalable techniques and related indexing technologies for these datasets.

1.1.3 Scientific computing and applied statistics:
Statistics is fundamentally concerned with the understanding of structure in data. One of the effects of the information-technology era has been to make it much easier to collect extensive datasets with minimal human intervention. Fortunately, the same technological advances allow the users of statistics access to much more powerful ‘calculators’ to manipulate and display data. Statistics are basically the carefully obtained observations that are recorded in numerical form. It is essential to understand that statistics are not facts and therefore incontrovertible, but observations about facts and therefore fallible.

Both - the accuracy with which the facts are observed as well as the proper interpretation of processed data can define the extent to which the reliability in the statistical calculations are assured. These can also affect the extent to which they truly represent the subject matter of that information. Hence, Statistics provides basis for
producing trustworthy data; analyzing them to make their meaning clear; and, drawing practical conclusions from so collected and interpreted data [11]. Computational statistics is somewhat recent discipline which has emerged out of the increasing applications of computers to sort and simulate the statistical problems. It basically aims at the design of algorithm for implementing statistical methods on computers, especially those, which otherwise could not have been implemented without computers, per say bootstrap, simulations as well as some other analytically intractable problems[12]

1.1.3.1 Various application areas of statistics:
Statistical techniques are used in a wide range various scientific and social researches, including:
- biostatistics
- computational biology
- computational sociology
- network biology
- Social science, sociology and social research.

Some fields of inquiry use applied statistics so extensively that they have specialized terminology.

1.1.3.2 Computational statistics:
Computational statistics is somewhat recent discipline which has emerged out of the increasing applications of computers to sort and simulate the statistical problems. It basically aims at the design of algorithm for implementing statistical methods on computers, especially those, which otherwise could not have been implemented without computers, per say bootstrap, simulations as well as some other analytically intractable problems[12].

Though, the advances and new developments in expanding field of statistics have not changed the foundation methodologies and paradigm, it has certainly improved the methods and specific ways of solving the problem. Even, the advances in computational power have enabled newer and more complicated statistical methods. The slight shift in the basic statistical paradigm has occurred due to the exponentially
Increasing computational power that enables the perfect execution of the newer and more complicated statistical methods. Using computational power, many alternative views of the same data set can be obtained. Even more models can be explored to solve the same problem from various angles. Generation of massive amounts of data has made it possible to study and simulate various studies/models to make more efficient and better solution available.

**Emergence and development of the field:**
Early developments in the field occurred due to the increasing usage of computer in solving numeric problems. In early 50’s few statisticians started using the computational power to sort out the numeric equations in the primary formats. Gradually, the usage started increasing during two following decades. Discussion forums in various journals and information sharing during various conferences and meetings put the life of computational statistics on pace. In early 80’s, the introduction and spread of PC brought about even more strong platform to build and run various mathematical/statistical applications on the user friendly software. The democratization of computing resulted in the rapid growth in the field and also in the rapid growth of the specialized software for the statistical computing. It also contributed to the changing role of the data sciences.

The ability of the computer generated algorithm to perform large number of computations almost instantaneously and to generate the graphical representations of the results immediately has opened many new frontiers to simulate those data sets and models, which otherwise could not have been executed as efficiently.

The hardware and software to perform these operations are readily available and hence, it has become possible to establish a two way communication between the statistical theory and statistical computing. The advances in statistical computing suggest new methods and developments of the supporting theory; at the same time, the advances in the theory and methods require new computational methods.
Introduction

- **The Concept:**
The term “statistical computing” refers to the computational methods that enable statistical methods i.e. allows the execution of the statistical methods. Statistical computing includes numeric analysis, database methodologies, computer graphics, software engineering and the computer-human interface i.e. popularly known as user friendly forms/ screens. The term is used more broadly to include not only the methods of statistical computing, but also the statistical methods which are computer intensive. Hence, more broadly, it can be regarded as a large set of modern statistical methods with the computational execution abilities. Computational statistics is grounded in mathematical statistics, statistical computing and applied statistics.

- **The cross currents of the Computational Statistics:**
The Computational Statistics is regarded now as a separate discipline in itself. In fact, it is a crux of various fundamental disciplines, and in a true sense actually, it is a multidisciplinary paradigm. Among any other inclusions, it is more closely related to statistics, though. It directly relates itself to the applied statistics, where computationally intensive methods are becoming commonly used in various application areas. Developments in other areas like numeric science and statistical analysis are also related to this area in a relevant manner.

- **Statistical software packages:**
Now it has become possible to use more sophisticated and complicated models. It is also possible to use the larger datasets which can either be structured/unstructured, homogeneous/heterogeneous containing more observations as well as more variables. Computer aids heavy use of visualization which permits better representation of various data generated after computational analysis. Extensive simulations can exactly reproduce the virtual image of the physical models. All these together make the hallmark software packages which are easy to use and efficient in calculations.

1.1.3.3 **Statistical Bioinformatics:**
Technological advances in all the fields including applied statistics as well as Bioinformatics are providing levels and magnitudes of life science related data that were unimaginable even five years ago. This leads every component of what scientists
do to be put forward by stretching, changing, and projecting forward in anticipation of further more development, both in the research area as well as in the practical implementation. The largest shift has been in the way we do science, it is no longer single laboratory science, it is now multidisciplinary efforts that bridge many disciplines and many species.

![Figure 1.1.3.3: Role and stance of Statistical Bioinformatics in Life and Information Sciences](image)

Statistical Bioinformatics is involved at every level of accumulation, organization, and analysis of biological data. Hypotheses that are initiated and tested can be refined, and new experiments formulated for the purpose of supplying more information can be shaped based on the inferences provided by the Statistical Bioinformatics. It acknowledges the inherent variation found in data that are generated as part of the Bioinformatics investigation and attempts to utilize experimental structure and design to partition variation into biological and technical components. The ultimate goal of statistical bioinformatics is to statistically identify significant changes in biological processes (e.g., changes in DNA sequence, quantitative trait locus identification, differential expression of genes, or changes in protein abundance) for the purpose of answering biological questions.
1.2 Concepts of Bioinformatics and its applications in various life sciences:
Modern scientific research depends on computer technology to organize and analyze large data sets. In fact, nowadays it is difficult to imagine an area of knowledge that can continue developing without the use of computers and informatics. Tremendous data flow produced with new innovations has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. At the same time, the growth of the computational power for processing and storage has also increased the necessity for deeper knowledge in the field. New interdisciplinary approaches have opened the arena of entirely naval innovations which otherwise could not have been done, had each of the discipline been working on its individual account. Researchers are addressing the challenge of dealing with exploding data by adopting mathematical and statistical software, computer modeling, and other computational and engineering methods. As a result, bioinformatics has become the latest engineering discipline. With its ability to organize, sort, reproduce and analyze the biological data with high throughput simulations, Bioinformatics has become an inevitable discipline to solve and sort the life.

1.2.1 The Definition and the concept:
The classical definition of the Bioinformatics is “the mathematical, statistical and computing methods that aim to solve the biological problems using DNA and amino acid sequences and related information”.

In a broader sense, the term can be defined as the study of information content and information flow in biological systems and processes. It has evolved to serve as the bridge between observations (data) in diverse biologically-related disciplines and the derivations of understanding (information) about how the systems or processes function, and subsequently the application (knowledge).
1.2.1.1 Aims and Objectives:

The aims of bioinformatics are basically three-fold. They are

1. Organization of data in such a way that it allows researchers to access existing information & to enter new information as they are produced.

2. To develop tools and resources that help in the analysis of data. For example, having sequenced a particular process, it has to simulate the process to arrive at the meaningful outcome. This requires more than just a straightforward database search. As such, many bioinformatics applications such as FASTA and PSI-BLAST pay attention to what constitutes a biologically significant resemblance.

3. Use of these tools to make analysis of the individual process in detail, check and produce the resemblances, as say for an example, done in case of making comparisons between the outcome on the quality check on two different drug profiles. A Bioinformatics application generates computational tools, databases, and methods to support the biological information.

1.2.1.2 Scope of Bioinformatics:

Bioinformatics implements its applications in the area of computer science, information science, and computer and information technology, communication technology to solve complex problems in life sciences and particularly in biotechnology. Due to rapid growth in quantitative biological data, such as complete genomes of biological species, protein 3-D structures, metabolic pathways databases, biodiversity related information, innovation in computer aided Drug design and discovery etc., data capture, data warehousing and data mining have become major issues for biotechnologists and biological scientists.

Big repositories of data are accessed and used by various scientists from different fields of life sciences due to advancements in information technology, particularly the Internet. Bioinformatics has become the central point to introduce platform for the Research & Development activities in the areas of functional genomics, proteomics, discovery of new drugs and vaccines, molecular diagnostic kits and pharmacogenomics. Bioinformatics makes it possible to create multimedia databases,
provides tools to carry out data analysis and facilitates modeling of molecules and biological systems on computer workstations as well as in a network environment.

Bioinformatics has emerged not only a supporting branch of science, but it also has become directive of the future course of research in biotechnology and life sciences. The importance and usefulness of Bioinformatics is realized in last few years by many industries. Therefore, large Bioinformatics R & D divisions are being established in many pharmaceutical companies, biotechnology companies and even in other conventional industry dealing with biological.

![Figure 1.2.1.2: Scope of Bioinformatics][15]

Bioinformatics can be thought of as a central hub that unites several disciplines and methodologies - molecular biology; information technology/information management; applications/databases; computational resources; CADD (Computer Aided Drug Design); and Genomics/Proteomics/x-omics. Bioinformatics brings together these activities and hence constitutes a large application area.
1.2.2 Bioinformatics Software and Tools

Bioinformatics basically deals with database creation, data analysis and modeling. Databases in biology are generally constituted with the data entries sourced from either printed material or through simulators. They are in the multimedia format organized as per the relational database model. Modeling is done not only on single biological molecule but also on multiple systems thus requiring a use of high performance computing systems.

Tools for bioinformatics range from simple command-line tools, to more complex graphical programs and standalone web-services available from various bioinformatics companies or public institutions.

1.2.2.1 Open-source bioinformatics software

Many free and open-source software tools have existed and continued to grow since the 1980s[16]. There are a number of freely available open code bases that help to create opportunities for all research groups. The open source tools often act as incubators of ideas, or community-supported plug-ins in commercial applications. They may also provide de facto standards and shared object models for assisting with the challenge of bioinformation integration.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Software</th>
<th>Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Bioconductor</td>
<td>This is the suit of applications that provides tools for the analysis and comprehension of high throughput data[17]</td>
</tr>
<tr>
<td>2.</td>
<td>BioPerl</td>
<td>It is basically organized community efforts that produce Perl code in the applicable format in the field of biology/life sciences[18]</td>
</tr>
<tr>
<td>3.</td>
<td>BioPython</td>
<td>It is a set of freely available tools for biological computation written in Python by an international team</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4.</td>
<td><strong>BioJava</strong></td>
<td>It is an open-source project dedicated to providing a Java framework for processing biological data.</td>
</tr>
<tr>
<td>5.</td>
<td><strong>BioRuby</strong></td>
<td>It comes with a comprehensive set of free development tools and libraries for bioinformatics and molecular biology, for the Ruby programming language.</td>
</tr>
<tr>
<td>6.</td>
<td><strong>EMBOSS</strong></td>
<td>It includes “sequence alignment,” “rapid database searching with sequence patterns,” “protein motif identification, including domain analysis,” “nucleotide sequence pattern analysis” etc.</td>
</tr>
<tr>
<td>7.</td>
<td><strong>FASTA and SEARCH</strong></td>
<td>This tool provides sequence similarity searching against protein databases using the FASTA suite of programs. FASTA provides a heuristic search with a protein query. FASTX and FASTY translate a DNA query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).</td>
</tr>
<tr>
<td>8.</td>
<td><strong>Mathematica</strong></td>
<td>It is renowned as the world's ultimate application for computations.</td>
</tr>
<tr>
<td>9.</td>
<td><strong>BLAST</strong></td>
<td>The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences.</td>
</tr>
<tr>
<td>10.</td>
<td><strong>MATLAB</strong></td>
<td>It is a high-level language and interactive environment for numerical computation, visualization, and programming.</td>
</tr>
</tbody>
</table>

Table 1.2.2.1 various open source Bioinformatics software

1.2.2.2 Open source bioinformatics Web services:

For extensive usability, the Bioinformatics applications are designed to allow an application to run on one computer in one part of the world and to use algorithms, data and computing resources on servers in other parts of the world. A number of
application are designed to configure themselves on SOAP (Simple Object Access Protocol) and REST (Representational State Transfer) based interfaces.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Software</th>
<th>Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>SSS (Sequence Search Services), MSA (Multiple Sequence Alignment) and BSA (Biological Sequence Analysis)</td>
<td>These services range from a collection of standalone tools with a common data format under a single, standalone or web-based interface, to integrative, distributed and extensible bioinformatics workflow management systems[27]</td>
</tr>
<tr>
<td>2.</td>
<td>Cytospace</td>
<td>It is an open source software platform for visualizing complex networks and integrating these with any type of attribute data[28]</td>
</tr>
<tr>
<td>3.</td>
<td>Galaxy</td>
<td>Galaxy is an open, web-based platform for data intensive biomedical research. Whether on the free public server or your own instance, you can perform, reproduce, and share complete analyses[29]</td>
</tr>
</tbody>
</table>

Table 1.2.2.2 various open source Bioinformatics web services

1.2.2.3 Bioinformatics Workflow management systems:

Any workflow management system should be designed specifically to make up and perform a set of computational or data manipulation steps or a workflow. A Bioinformatics application is the specialized work flow management system that allows integrative data capture, efficient and pertinent work flow management processes and composite reports in the prescribed format.
Various platforms for running the workflow management systems:

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Software</th>
<th>Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Galaxy</td>
<td>It is a scientific workflow, data integration and data and analysis persistence and publishing platform. The basic objective is to make the computational biology accessible to those which don’t have computer related experience. Originally, it was developed for genomics research, but now is being used as a general bioinformatics workflow management system[29]</td>
</tr>
<tr>
<td>2.</td>
<td>Kepler</td>
<td>Kepler facilitates various application areas in process and data monitoring, provenance information, and high-speed data movement solutions. The Workflows are represented as directed graphs, in which the nodes represent discrete computational components, and the edges represent paths, along which data and results can flow between components[30] In all, Kepler is a free software system for designing, executing, reusing, evolving, archiving, and sharing scientific workflows.</td>
</tr>
<tr>
<td>3.</td>
<td>Anduril</td>
<td>It is used for translating fragmented large-scale data into testable predictions. The Anduril framework allows rapid integration of heterogeneous data with state-of-the-art computational methods and existing knowledge in bio-databases. Anduril automatically generates thorough summary reports and a website that shows the most relevant features of each gene at a glance, allows sorting of data based on different parameters, and provides direct links to more detailed data on genes, transcripts or genomic regions. Anduril is open-source; all methods and documentation are freely available[31]</td>
</tr>
</tbody>
</table>

Table 1.2.2.3 various open source Bioinformatics workflow management systems
1.2.3 Bioinformatics Applications in Pharmaceutical Sciences:
The research activities in the field of life sciences have led to the development of ever-increasing amounts of databases, containing masses of biological information accessible to people. The information technology based tools of Bioinformatics act as an pooling interface between computers and biological sciences.

The field of bioinformatics deals with organizing, analyzing and distributing the biological information in specially designed databases powered by smart mathematical/statistical algorithms and the latest database techniques.

1.2.3.1 Pharmaceutical science in the modern era:
The pharmaceutical science today has evolved into a wide spectrum covering a number of disciplines. It’s ever expanding boundaries now envelope a wide range of interdisciplinary areas that work together to cater to requirements of the discovery and development of new drugs and therapies. Pharmaceutical sciences can be broadly classified into a number of main categories, with many specialized fields within each category.

1.2.3.2 Various branches of Pharmaceutical Science:
There are a number of areas recently covered under the big roof of the Pharmaceutical Science. As the new discoveries and inventions advance and extend their scope, subspecialties continue to be added to the list. Basically, the underlying concepts that form the foundation for the subjects remain more or less the same. Even, when the knowledge advances, the boundaries between these specialization areas begin to blur. The below listed are the main branches, within which another categories or specific disciplines fall:

1. Pharmacology
2. Pharmaceutical Chemistry
3. Pharmaceutics
4. Pharmacognosy
1.2.3.3 In-silico applications in various Pharmaceutical Sciences:
In less than a generation or so, every area including those from science has experienced the catalytic role of computers in almost all the paradigms of development world-wide. It has now become possible to draw different concepts and ideas from various fields and merge them into the specific domain of innovation, as the computer has come up with the strengths of facilitating the crux of multidisciplinary approaches altogether. Particularly, in the area of Pharmaceutical Sciences, it has changed the scenario from the roots. From the limited usage in form of spread sheets in laboratories, it has spanned across almost all the stages from manufacturing through delivery of drugs. Modern information management and computer applications can bridge the hiatus theoreticus, the gap between the scientific and theoretical knowledge in drugs on one hand and its application on the other[32]. It is since past two and half decades that computational methods have facilitated the drug discovery process and smoothened up the financial and experimental efforts of scientists as well as enterprises.

Research and development in the pharmaceutical industry is a time-consuming and expensive process. It takes a lot of efforts on the part of the investigating scientists to design and test new drug while still maintaining its compatibility and viability. At the same time, the newly developed drugs must be formulated in such a way that it should successfully run on commercial basis also. Hence, thorough consideration of every aspect of the formulation and process development must be optimized to create an overall satisfying drug. The wide spread use of computer applications have proved to be the facilitating pool between the concepts of design and their implementation. By the means of statistically organized experiments, artificial intelligence and other computational methods, it is possible now to work on every aspect of drug discovery and design process. Simultaneous development and investigation of pharmaceutical products and processes enables application of quality by design concept that is being promoted by the regulatory authorities worldwide.

The in silico methods have provided support in terms of databases, data mining of large genomes, network analysis, systems biology on the bioinformatics front and structure–activity relationship, similarity analysis, docking, and pharmacophore methods for lead design and optimization. This review highlights some of the
applications of bioinformatics and chemoinformatics methods that have enriched the field of drug discovery. In addition, the review also provided insights into the use of free energy perturbation methods for efficiently computing binding energy. These in silico methods are complementary and can be easily integrated into the traditional in vitro and in vivo methods to test pharmacological hypothesis.

Computer applications for the pharmaceutical sciences are applicable to almost all the areas of pharmaceutical sciences that include

- Pharmacodynamics
- Pharmacokinetics
- Computational Chemistry
- Combinatorial Chemistry
- ADME Informatics
- Chemoinformatics
- Toxicology
- Metabolic Modeling

They combine the fundamentals of experimental design application and interpretation in pharmaceutical technology and devise certain methods, which can be broadly categorized into various heads, all of which are essential tools for successful building of quality into pharmaceutical products and processes from the early stage of their development to selection of the optimal ones.

1.3 In-silico approach to Drug Discovery and Development:

1.3.1 Drug Discovery Pipeline:
In the last four decades, computational technologies for drug Research and Development have advanced very quickly, particularly in last few years with the unprecedented development of biology, biomedicine, and bioinformatics applications. In the post genomic era, the dramatic increase in small-molecule and biomacromolecule information has induced the development and application of computational tools to almost every stage of drug Research and Development. This has greatly changed the strategy and pipeline for drug discovery[33]

Drug research and development (R & D) is a comprehensive, expensive, and time-consuming practice, and it is full of risk throughout the process[34]. Numerous new technologies have been developed and applied in drug R & D to shorten the research
cycle and to reduce the expenses. Among them, computational approaches have revolutionized the pipeline of discovery and development[33].

1.3.1.1 Various stages of the pipeline and applications of in-silico techniques on each of them:

The drug discovery process contains a number of stages, each having its own importance and implications on the successful formation of any drug. Drug discovery involves Different stages, which include: basic exploratory biology on target identification and validation; assay development; lead identification, which usually requires access to high-throughput screening; medicinal chemistry and pharmaceutical lead optimization; and drug candidate selection [35]. Continuous development in the field induces new and faster techniques to gear up the entire set of processes.

![Figure 1.3.1.1 various stages and related software tools of Drug Discovery Pipeline](image)

1.3.1.2 Various computational techniques applied on different stages:

Computational approaches are gaining considerable importance in the process of Drug design and discovery. The available range of software spans almost all stages in
discovery and development pipeline, from target identification to lead discovery, from lead optimization to preclinical or clinical trials.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Software</th>
<th>Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>High Throughput Screening (HTS)</td>
<td>It is the method of scientific experimentation, used particularly in the drug discovery process. With the help of specialized software, robotics and sensitive detectors, it allows conducting millions of chemical/genetic/pharmacological tests with the high speed and accuracy. The outcomes of the HTS experimentations facilitate modulating a biomolecular pathway that guides the drug designing process.</td>
</tr>
<tr>
<td>2.</td>
<td>Virtual screening:</td>
<td>Virtual Screening basically focuses on designing and optimizing targeted combinatorial libraries and enriching libraries of available compounds from in-house compound repositories or vendor offerings. As the accuracy of the method has increased, virtual screening has become an integral part of the drug discovery process [36].</td>
</tr>
<tr>
<td>3.</td>
<td>Molecular modeling:</td>
<td>The behavior of molecules can be modeled or simulated using computers under molecular modeling techniques. Complex calculations in the field of computational chemistry/biology, material science or drug discovery require atomistic level description of the molecular system.</td>
</tr>
<tr>
<td>4.</td>
<td>Sequence similarity searching:</td>
<td>This is the method for searching sequence databases by using the alignment to a query sequence. Once the statistical assessment of finding the similarity between the databases and query sequence is done, the homology can be inferred and the information can be transferred to the query sequence.</td>
</tr>
</tbody>
</table>
Drug lead optimization:

This stage of drug discovery concentrates on adding new or more functionality to the compounds, which may thus result in an increase in the values of the physico chemical properties like molecular weight and number of hydrogen bonding groups. The recent focus hence, during this stage, is on selecting such screen sets, which have more lead-likeness rather than drug-likeness [78-79].

| Table 1.3.1.2 various computational techniques applied on different stages of drug discovery pipeline |
|---|---|
| **5. Drug lead optimization:** | This stage of drug discovery concentrates on adding new or more functionality to the compounds, which may thus result in an increase in the values of the physico chemical properties like molecular weight and number of hydrogen bonding groups. The recent focus hence, during this stage, is on selecting such screen sets, which have more lead-likeness rather than drug-likeness [78-79]. |

1.3.2 Computer Aided Drug Design (CADD):

Pharmaceutical and biotech industries have witnessed increasing growth of CADD application software since the late 1980s. In the 1990s, the rapid development of small molecule combinatorial chemistry/parallel synthesis and the high-throughput screening (HTS) technologies spurred renewed interests in the quantitative structure-activity relationship (QSAR) technique.

As the process of drug discovery is quite lengthy and time consuming, various efforts to apply computational power are continuously employed to facilitate various stages of the process in order to streamline the successive steps of drug discovery, design, development and optimization. In biomedical arena, computer-aided or in silico design is adopted to accelerate and facilitate various processes like hit identification, hit-to-lead selection, optimize the absorption, distribution, metabolism, excretion as well as toxicity profiles.

The development of any potential drug begins with the determination of specific receptors (targets), which can be efficiently done using the multidisciplinary solution approach which combines the computational and automatic analytical power of computer and the underlying concepts of pharmaceutical sciences. Starting from this first stage, the computer aided applications extend to all the processes through the final drug outcome. Hence, CADD applications have spanned to almost all the stages of drug discovery pipeline.
There are a lot many software which aid in drug design and development. Various in-silico models and applications can be subdivided into a number of categories as mentioned below:

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Application Category</th>
<th>Available software</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Web enabled datasets and databases</td>
<td>ZincDatabase, ChEMBL, Chemspider, Bingo, JChemforExcel, ProteinDataBank(PDB), ChemDiff, TTD, STITCH, SMPDB, BindingMOAD(MotherOfAllDatabase), LigandProteinDataBase(LPDB) etc.</td>
</tr>
<tr>
<td>3.</td>
<td>Software for molecular modeling</td>
<td>CHARMM, GROMACS, Amber, SwissParam, CHARMM-GUI, CHARMMing.org, SwissSideChain</td>
</tr>
<tr>
<td>4.</td>
<td>Software for homology modeling</td>
<td>Modeller, I-TASSER, LOMETS, SWISS-MODEL, SWISS-MODELRepository, Robetta</td>
</tr>
<tr>
<td>5.</td>
<td>Various free ware for biding site prediction</td>
<td>MED-SuMo, CAVER, FINDSITE, sc-PDB, CASTp, Pocketome, 3DLigandSite, metaPocket, PocketAnnotate</td>
</tr>
<tr>
<td>6.</td>
<td>Docking software</td>
<td>Autodock, DOCK, GOLD, SwissDock, DockingServer, 1-ClickDocking</td>
</tr>
</tbody>
</table>
Table 1.3.2 Various CADD Applications category wise

Computer-aided drug design uses computational chemistry as a base line concept to study, discern and enhance drugs and related biologically active molecules with the most elementary goal to predict whether a given molecule will bind to a target and if so how strongly. To predict the conformation of the small molecule and to model conformational changes in the biological target, the basic theories of molecular mechanics or molecular dynamics are most often used. Semi-empirical, ab initio quantum chemistry methods, or density functional theory are often used to provide optimized parameters for the molecular mechanics calculations and also provide an estimate of the electronic properties (electrostatic potential, polarizability, etc.) of the drug candidate that will influence binding affinity.
Semi-quantitative prediction of the binding affinity might also be provided using the molecular mechanics methods. Also, knowledge-based scoring function may be used to provide binding affinity estimates. These methods use linear regression, machine learning, neural nets or other statistical techniques to derive predictive binding affinity equations by fitting experimental affinities to computationally derived interaction energies between the small molecule and the target[37][38].

1.3.3 Contemporary drug discovery and development:

1.3.3.1 The shift in drug discovery paradigm:
The process of drug discovery and development has undergone tremendous changes over past couple of decades. It isn’t an old history, when the process of drug discovery was carried out in the sequential steps, wasting too much a time on the repetitive processes during lead selection and structure optimization.

In earlier days, various compounds were synthesized in the processes against a battery of in-vivo biological screens. These processes were very time consuming and lengthy.
The output of the processes would be those promising compounds, which were studied further for the pharmacokinetic properties, its probable metabolism and the potential toxicity. Adverse findings were often made at this stage. They led to the decision regarding continuing or halting or restarting the project to find another more promising drug candidate. This would lead to the unacceptable burden on the research and development activities of the unit.

Nevertheless, this paradigm has shifted today in several ways. The drug or compound is tested for its pharmacokinetic, metabolism or toxicity much earlier, even before the decision is taken to evaluate the compound in the clinic as shown below:

![Figure 1.3.3.1(2): the contemporary process of drug discovery (parallel drug discovery)](image)

Today, the rate at which the biological screening data are obtained has increased enormously. The new High Throughput Screening (HTS) has blown the drug discovery and development regime to the new and greater extent. In response to these developments, a new discipline of chemistry – the Combinational Chemistry has emerged to feed this new thirst of research activities. The new and rapidly developing
era of Combinational Chemistry, Chemoinformatics and other related disciplines form the interactive pool of knowledge; where from naval inputs to the research and discovery process are derived. Together with the increasing capacity for biological screening and chemical synthesis, more demands for large quantities of early information on Absorption, Distribution, Metabolism and Excretion are rising. The highly increasing need for the medium and high throughput in-vitro screens is the outcome of this changing requirement. In addition, there is an increasing need of good tools for predicting these properties to serve two main purposes:

1. Predicting the ADME as well as physicochemical properties at the design stage of compounds and compound libraries to reduce the risk of attrition at the later stage. Today, the 40% of the attritions rate of CDs is directly related to the poor PK profiles.
2. Optimizing the screening and testing processes by making available only potential and relevant candidates. These reliable screening factors are highly desirable to avoid the risk of late attrition.

1.4 Prediction of Physicochemical properties in early stage of drug discovery:
Understanding the physicochemical properties of molecules in early stages of the drug discovery process can help identify and eliminate candidate molecules that are unlikely to survive later stages of development. In this respect, the solubility issues are fundamental for drug discovery research and robust in silico methods for solubility assessment can contribute significantly to the reduction of the overall cost of drug development[41].

The physicochemical properties of a drug have an important impact on its pharmacokinetic and metabolic fate in the body. Basic physicochemical properties are explained in the following table.
<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Property name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lipophilicity</td>
<td>It refers to the ability of a chemical compound to dissolve in fats, oils, lipids, and non-polar solvents. Lipophilicity is the key physicochemical parameter linking membrane permeability and hence drug absorption and distribution with the route of clearance[42].</td>
</tr>
<tr>
<td>2</td>
<td>Solubility</td>
<td>It is the property of a solid, liquid, or gaseous chemical substance called solute, to dissolve in a solid, liquid, or gaseous solvent to form a homogeneous solution of the solute in the solvent. The solubility of a substance fundamentally depends on the physical and chemical properties of the solute and solvent as well as on temperature, pressure and the pH of the solution. Ideally, only soluble compounds would be synthesized in a drug-discovery program.</td>
</tr>
<tr>
<td>3</td>
<td>pK&lt;sub&gt;a&lt;/sub&gt;</td>
<td>pKa indicates a compound’s ionizability. It is a function of the acidity or basicity of group(s) in the molecule.</td>
</tr>
<tr>
<td>4</td>
<td>Hydrogen bonding</td>
<td>The hydrogen-bonding capacity of a drug solute is now recognized as an important determinant of permeability. In order to cross a membrane, a drug molecule needs to break hydrogen bonds with its aqueous environment[42].</td>
</tr>
<tr>
<td>5</td>
<td>permeability</td>
<td>It is an important property, which is a big determinant of intestinal absorption and oral bioavailability.</td>
</tr>
</tbody>
</table>

Table 1.4 various Physicochemical properties
1.5 Drug dissolution:
Drug dissolution in chemistry is the process of dissolving a solid substance into a solvent to yield a solution. Over recent years, drug release / dissolution from solid pharmaceutical dosage forms has been the subject of intense and profitable scientific developments. Whenever a new solid dosage form is developed or produced, it is necessary to ensure that drug dissolution occurs in an appropriate manner[43].

1.6 Drug dissolution rate:
Drug dissolution rate is an important parameter that affects oral drug absorption. A drug is defined as being poorly soluble when its dissolution rate is so slow that dissolution takes longer than the transit time past its absorptive sites, resulting in incomplete oral absorption.[44].

1.7 Solubility of substance:
Water is integral to the structure and function of all living material. The human body is approximately 60% water by mass[45] and biology might be reasonably dependent on water content. The behavior of a drug in water governs many uptake, movement and elimination issues within the body, which affect the later stages of the drug development process. The ‘solubility’ of drug substance is one of the most pertinent properties of drug molecules in this context.

Solubility is the property that makes considerable impact on the bioavailability, as it has the direct relation with the dissolution and hence, in-vivo absorption of the substance. It is one of two factors defining the biopharmaceutics classification system (BCS). According to an IUPAC definition, solubility is the analytical composition of a saturated solution expressed as a proportion of a designated solute in a designated solvent. Solubility may be stated in units of concentration, molality, mole fraction, mole ratio, and other units.
1.8 Drug dissolution and drug solubility:

The term ‘dissolution’ involves disintegration of substance into smaller particles from which the medicinal drug is released more and more rapidly. This process of dissolution, followed by absorption determines, in part, the bioavailability of the drug. To be absorbed across cell membranes or pass through intercellular pores a medicinal drug must have been properly dissolved in the solution. The rate at which a drug gets into solution is important, but if its solubility is low its pharmacokinetic and pharmacodynamic properties might be adversely affected[46]

The dissolution of a chemical into water is a process fundamental to both chemistry and biology. The persistence of a chemical within the environment and the effects of a chemical within the body are dependent primarily upon aqueous solubility.

Solubility, the phenomenon of dissolution of solute in solvent to give a homogenous system, is one of the important parameters to achieve desired concentration of drug in systemic circulation for desired (anticipated) pharmacological response.[47] The term “good solubility” often implies the tendency for fast and complete dissolution[48]. In practice, high drug solubility is associated with the high dissolution rate[46].

Solubility is a critical factor as drug substances have to be dissolved before they can be absorbed. Solubility and rate of dissolution are the crucial players in the famous Biopharmaceutical Classification System (BCS)[49], as absorption of passively transported drugs across the gastrointestinal (GI) tract is the combined product of both permeability and solubility according to Fick’s first law[50].

1.9 Importance of solubility:

The aqueous solubility of a drug is an important factor that influences its absorption, distribution and elimination in the body. Poor aqueous solubility often causes a drug to appear inactive and may cause other biological problems[41]. The major challenge with the design of oral dosage forms lies with their poor bioavailability. The oral bioavailability depends on several factors including aqueous solubility, drug permeability, dissolution rate, first-pass metabolism, pre-systemic metabolism, and susceptibility to efflux mechanisms. The most frequent causes of low oral bioavailability are attributed to poor solubility and low permeability[47].
The drug solubility manages dissolution of compounds and maximum concentration reached to the gastrointestinal fluid. Solubility is one of the important parameters to achieve desired concentration of drug in systemic circulation for achieving required pharmacological response[51]. Poorly water soluble drugs often require high doses in order to reach therapeutic plasma concentrations after oral administration. Low aqueous solubility is the major problem encountered with formulation development of new chemical entities as well as generic development. Any drug to be absorbed must be present in the form of an aqueous solution at the site of absorption[47]. Generally, the compounds with high solubility and membrane permeability easily reach to site of infection in required amount. So, a drug with high solubility and membrane permeability is considered exempt from bioavailability problems. Otherwise, it is a problematic candidate or needs careful formulation work[52]. The compounds with less solubility should further be introduced to formulation step and lead optimization steps during drug discovery.

The solubility of drugs and drug-like compounds has been the subject of extensive studies aimed at finding a way to predict solubility from molecular structure. Because of its important role in the determination of drugs’ ADME-Tox property, aqueous solubility is central in vitro screening assays. Various technologies like combinatorial chemistry and High Throughput Screening have given rise to the high speed discovery of millions of compounds through synthesis and screening during the activity assays. However, the solubility measurement, particularly the intrinsic solubility, is a low-throughput process. The fact that the prediction of the solubility of the relative compound must be predicted accurately intimates the need for the development of high quality solubility \textit{in-silico} models.

In the last 20 years, a numerous solubility models have been developed. However, as the statistical result of the Solubility Challenge indicated, the high accurate and reliable solubility has not emerged[53]
1.10 Why to perform dissolution/solubility studies at early stage of drug discovery?

The oral absorption of a drug is the tandem process of the dissolution and the intestinal membrane permeation of a drug in the gastrointestinal (GI) tract. Therefore, low solubility, a low dissolution rate and low permeability, can all result in incomplete and variable oral absorption. Intestinal membrane permeability is mostly governed by the chemical structure of a drug. In the current drug discovery and development paradigm, modifications of a chemical structure are performed only during drug discovery. Therefore, a candidate compound must achieve an acceptable intestinal membrane permeability at some point during the drug discovery stage. In contrast, the dissolution profile can be improved by salt/solid form selection and formulation, as well as through chemical structure modification. The salt/solid form selection and the formulation studies start at the final stage of drug discovery, or in the early development stage. Since the discovery-development transition is practically irreversible, it is necessary to map out a successful strategy to achieve an acceptable dissolution profile during the discovery stage. Therefore, a comprehensive assessment of solubility and dissolution is required in drug discovery whether or not it can be improved in the later stages[48]

1.11 In silico approach to prediction of aqueous solubility:

Recent analytical studies have shown that in spite of increased investment in preclinical pharmaceutical research, the overall number of new drugs registered by regulatory agencies remained approximately unchanged over the last decade. It is estimated that the total preapproval cost of production of a new drug is in the range from US $800 million[34] to more than $1.7 billion dollars. The cost of drug development is currently 55% higher than the average cost from 1995 to 2000 and it is rising largely as a result of an increasing failure rate for prospective drugs in clinical trials. The poor pharmacokinetic and toxicological properties of compounds constitute one of the most important issues of the failure of candidate drugs in the later phases of clinical testing[34].

It is assumed that the chemical space exceeds $10^{60}$ molecules and it is impossible for mankind to make all those molecules. So far, only about 27 million compounds have been registered[54]. However, the number of different small molecules within our
own bodies is about a few thousands. As a consequence, it is tough to discover small molecules that actively interact with protein targets since the biological related chemical entities only represent an amazingly small fraction of the entire chemical space.

The importance of aqueous solubility for drug design can be recognized at all steps of drug development. The solubility is extremely important as it determines uptake, movement and elimination of the substance from the body. Aqueous solubility is one of the major physicochemical properties to be optimized in drug discovery. It is related to “A” and “D” in the ADME-Tox Aqueous solubility and membrane permeability are the two key factors that affect a drug’s oral bioavailability[55]. Aqueous solubility governs both the rate of dissolution of the compound and the maximum concentration reached in the gastrointestinal fluid.

The first step in the drug absorption process is the disintegration of the tablet or capsule, followed by the dissolution of the active drug. Obviously, low solubility is detrimental to good and complete oral absorption, and so the early measurement of this property is of great importance in drug discovery[42] Reflecting this need, rapid, robust methods have been developed to efficiently measure the solubility of large numbers of compounds.

With advent of computational biology and combinatorial chemistry, numerous solubility prediction methods and models based on these methods have been developed. However, most of these models were developed based on statistical analysis. Such models are found to be less accurate and non-reliable in predicting diverse values. The variations lying in different chemicals belonging to diverse chemical spaces have greatly reduced the efficiency of models. However, the Empirical Methods for building predictive models of the relationships between molecular structure and useful properties are becoming increasingly important.

Still the basic limitation for academic researcher remains with the non/limited availability of related Software applications. ADME intestinal Absorption and ADME Solubility are two important parameters, which are possible to predict only using Discovery studio. This software is commercial software; hence its free usage is
prevented. The DSA calculator created in present study was designed to address this issue and it was found capable of satisfactorily predicting both of these parameters depending upon use of freely available online Open source tools.

1.12 The Objectives of Present study:

The current study is carried out with the following objectives:

- Data mining for available computational tools for drug Solubility
- on line Workbench creation for drug Solubility
- in silico study for novel heterocyclic compounds for solubility
- Virtual screening and lead optimization for better solubility
References:


[Accessed: 08-Jul-2014].

C. Lauro, “Computational statistics or statistical computing, is that the

“Statistics | Research | Statistical Bioinformatics Center.” [Online]. Available:

L. M. C. Heitor Silvério Lopes, Computational Biology and Applied
Bioinformatics. 2011.

http://swift.cmbi.ru.nl/gv/students/mtom/how_bioinf_suport_drug.htm.

“Open Bioinformatics Foundation.” [Online]. Available: http://www.open-

[Accessed: 08-Jul-2014].


2014].

Introduction


Introduction


Introduction


