Why Study Bioinformatics?

PROF. (DR.) AMAR P. GARG
Vice Chancellor
Shobhit University, Meerut – 250110
Email: vicechancellor@shobhituniversity.ac.in

RESEARCH SHOULD FOCUS ON

- ***It is not enough to identify a problem***
- ***Not enough to solve a problem***
- ***Need to identify and fix the root cause-otherwise problem is likely recur***

IDENTIFICATION OF

ROOT CAUSE IS NECESSARY

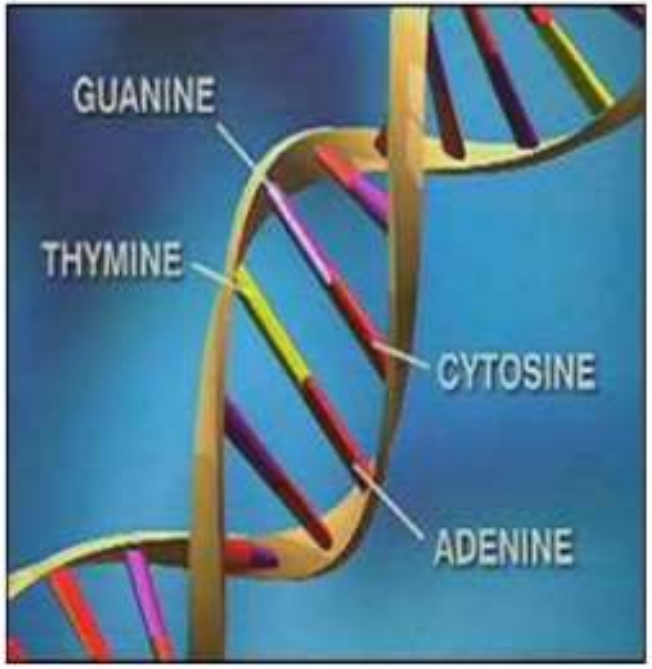
Research in 21st Century

" ***The new paradigm, now emerging is that all the 'genes' will be known (in the sense of being resident in databases available electronically), and that the starting "point of a biological investigation will be theoretical."***

- Walter Gilbert

What is Bioinformatics?

- "The mathematical, statistical and computing methods that aim to solve biological problems using DNA and amino acid sequences and related information"



Bioinformatics is being used in following fields:

1. Microbial genome applications
2. Molecular medicine
3. Personalized medicine
4. Preventative medicine
5. Gene therapy
6. Drug development
7. Antibiotic resistance
8. Evolutionary studies
9. Waste cleanup Biotechnology
10. Climate change Studies
11. Alternative energy sources
12. Crop improvement
13. Forensic analysis
14. Bio-weapon creation
15. Insect resistance
16. Improve nutritional quality
17. Development of Drought resistant varieties
18. Veterinary Science

Microbial genome applications

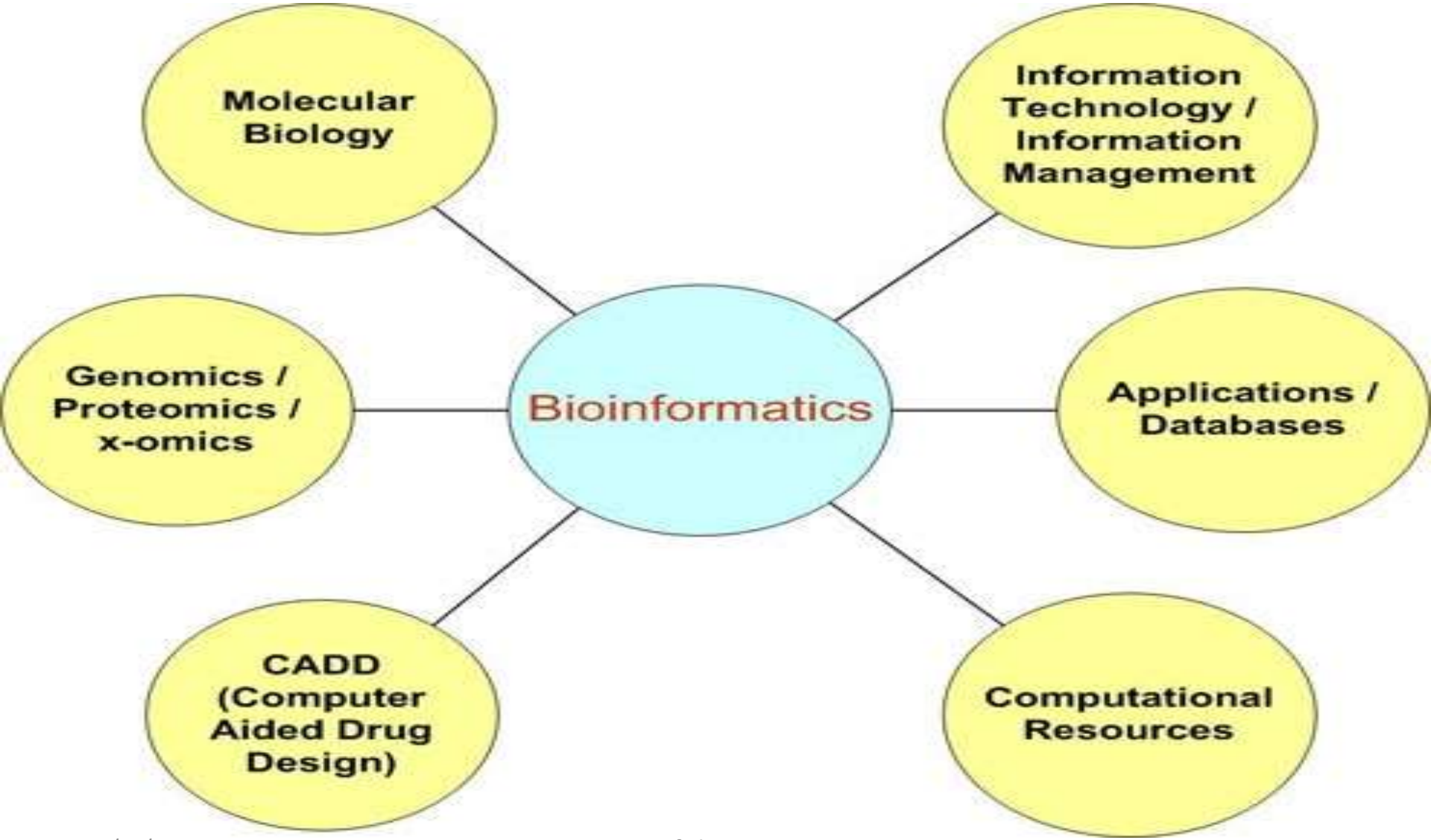
1. *Genome assembly*
2. *Re-sequencing*
3. *Comparative analysis*
4. *Evolutionary studies*
5. *Antibiotic resistance*
6. *Waste cleanup*
7. *Biotechnology*
8. *Genome Assembly*

• Genome assembly is a very complex computational problem due to enormous amount of data to put together and some other reasons reason. •

• Ideally an assembly program should produce one contig for every chromosome of the genome being sequenced. But because of the complex nature of the genomes, the ideal conditions just never possible, thus leading to gaps in the genome. 7 De Novo assembly - puzzle without

Applications of Bioinformatics in Drug Discovery

Bioinformatics



Bioinformatics Tools

The processes of designing a new drug using Bioinformatics tools have open a new area of research. In order to design a new drug one need to follow the following path.

1. ***Identify target disease***
2. ***Study interesting compounds***
3. ***Detection of molecular bases for disease***
4. ***Rational Drug designing technique***
5. ***Refinements of compounds***
6. ***Quantitative Structure Activity Relationships***
7. ***Solubility of molecule***
8. ***Drug Testing***

Identify Target Disease

1. First know all about the disease and its existing or traditional remedies.

2. Target identification alone is not sufficient in order to achieve a successful treatment of a disease. A real drug needs to be developed.

3. This drug must influence the target protein in such a way that is does not interfere with normal metabolism.

4. Bioinformatics methods have been developed to virtually screen the targets for compounds that bind and inhibit the protein.

Study Interesting Compounds

1. ***Identify and study the Lead compounds that have some activity against a disease.***

2. These may be only marginally useful and may have severe side effects.

3. These compounds provide a starting point for refinement of the chemical structures.

Detect the molecular bases for disease

1. ***If it is known that a drug must bind to a particular spot on a particular protein or nucleotide then a drug can be tailor made to bind at that site.***

2. This is often modeled computationally using any of several different techniques.

3. Traditionally, the primary way of determining what compounds would be tested computationally was provided by the researchers understanding of molecular interactions.

4. A second method is the brute force testing of large numbers of compounds from a database of available structures.

Refinements of compounds

1. Once you get a number of lead compounds, computational and laboratory techniques are applied in refining the molecular structures to give a greater drug efficacy with lesser side effects.

2. Based on laboratory and computational analysis, determine those aspects that are responsible for both the drug efficacy and its side effects.

Computer Aided Drug Design (CADD)

.

Computer Aided Drug Design (CADD) is a specialized discipline that uses computational methods to simulate drug-receptor interaction.
CADD methods heavily dependent on bioinformatics tools, applications and databases. As such, there is considerable overlap in CADD research and Bioinformatics.

Bioinformatics supports CADD Research
Virtual High-Throughput Screening (vHTS)

1. Pharmaceutical companies are always searching for new leads to develop new drug compounds.

2.One search method is virtual High-throughput screening. In vHTS protein targets are screened against database of small molecule compounds to see which molecules bind strongly to the target.

3. If there is a hit with a particular compound, it can be extracted from the database for further testing.

Sequence Analysis

1. In CADD research, one often knows the genetic sequence of multiple organisms or the amino acids of proteins from several species.

2. It is very useful to determine how similar and dissimilar the organisms are based on gene or protein sequences.

3.With this information one can infer the evolutionary relationships of the organisms, search for similar sequences in bioinformatics databases and find related species to those under investigation.

4.There are many bioinformatics sequence analysis tools that can be used to determine the level of sequence similarity.

Bioinformatics support CADD Research

1.Another common challenge CADD research in determining the 3-D structure of proteins.
2. Most Drug targets are proteins, so it is important to know their 3-D structure in detail. It is estimated that the human body has 500,000 to million proteins.
3. However, the 3-D structure is known for only a small fraction of these. Homology modeling is one method used to predict 3-D structure.
4. In homology modeling, the amino acid sequence of a specific protein (target) is known, and the 3-D structure of proteins related to target (templates) are known.
5. Modeller is a well known tool in Homology Modeling, and the SWISS-MODEL Repository is a database of protein structures created with homology modeling.

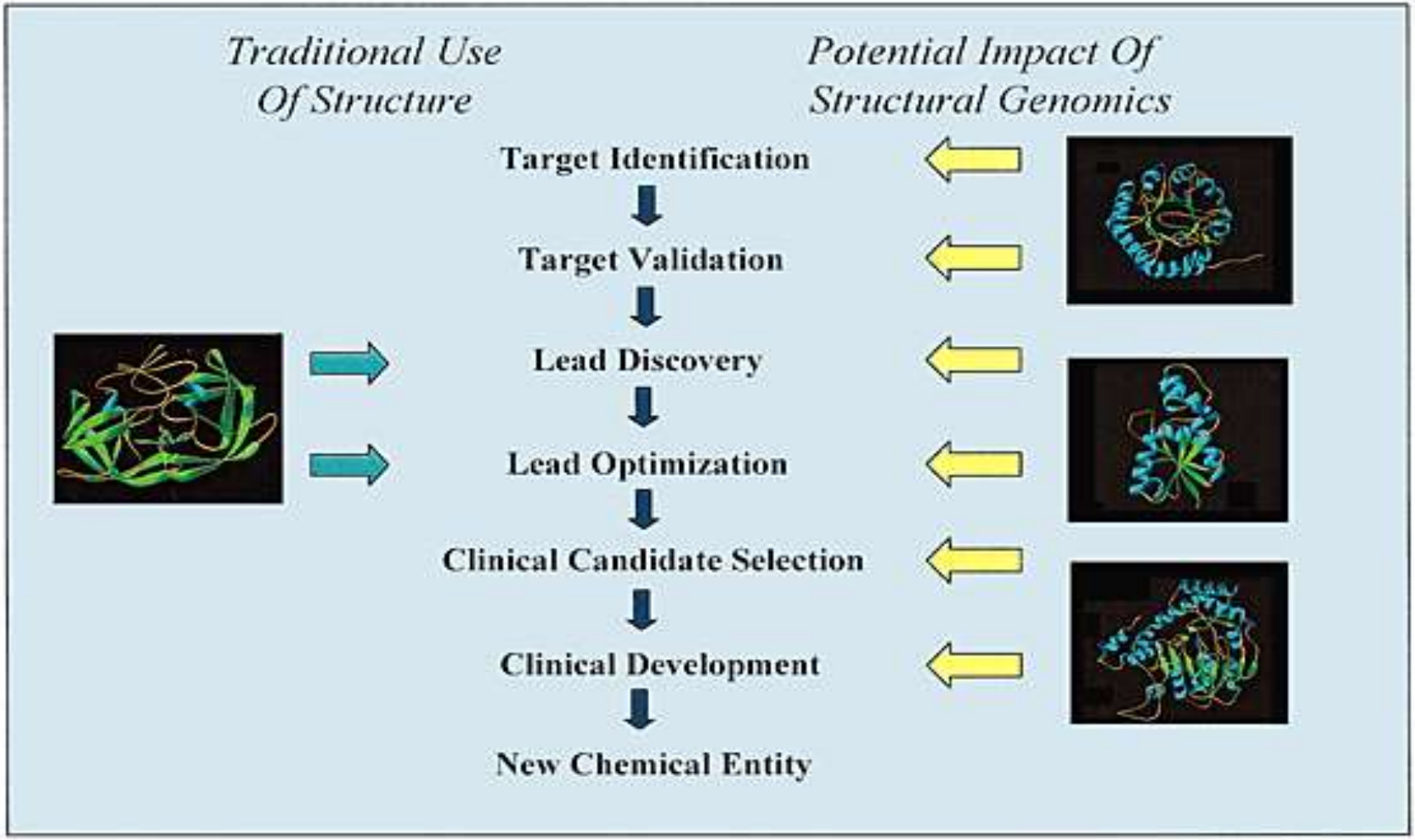
Structural Genomics

- ***Organise all known proteins into families.***
- ***Determine structures of at least one member of every family.***
- ***Solve structures of more than 10,000 protein in next 10 years.***
- ***Generate knowledge and rules from known protein structures.***
- ***Apply this knowledge to predict the structure of each and every protein of known organisms.***

Objectives of Structural Genomics

- ***Selection of Targets for structure determination to obtain maximum information return on total efforts***
- ***Develop mechanism that facilitates cooperation and prevent work duplication***

Impact of Structural Genomics on Drug Discovery

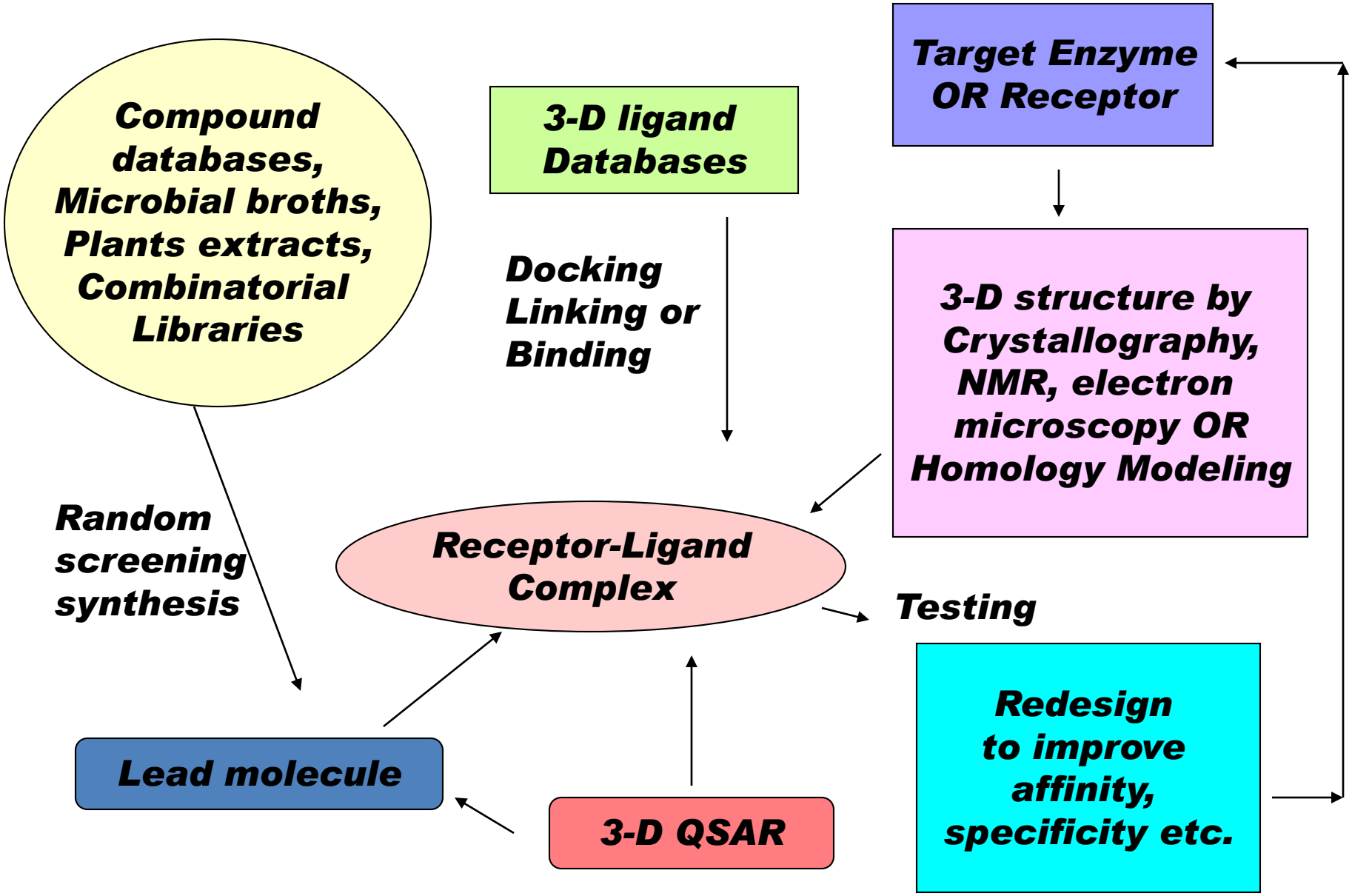


Dry, S. et. al. (2000) Nat. Struc.Biol. 7:976-949.

Drug Development Flowchart

- ***Check if structure is known***
- ***If unknown, model it using KNOWLEDGE-BASED HOMOLOGY MODELING APPROACH.***
- ***Search for small molecules/ inhibitors***
- ***Structure-based Drug Design***
- ***Drug-Protein Interactions***
- ***Docking***

STRUCTURE-BASED DRUG DESIGN



Docking Methods

- ***Docking of ligands to proteins is a formidable problem since it entails optimization of the 6 positional degrees of freedom.***
- ***Rigid vs Flexible***
- ***Speed vs Reliability***
- ***Manual Interactive Docking***

GRID Based Docking Methods

- Grid Based methods
  - GRID (Goodford, 1985, J. Med. Chem. 28:849)
  - GREEN (Tomioka & Itai, 1994, J. Comp. Aided. Mol. Des. 8:347)
  - MCSS (Mirankar & Karplus, 1991, Proteins, 11:29).
- Functional groups are placed at regularly spaced (0.3-0.5A) lattice points in the active site and their interaction energies are evaluated.

Automated Docking Methods

- ***Basic Idea is to fill the active site of the Target protein with a set of spheres.***
- ***Match the centre of these spheres as good as possible with the atoms in the database of small molecules with known 3-D structures.***
- ***Examples:***
  - ***DOCK, CAVEAT, AUTODOCK, LEGEND, ADAM, LINKOR, LUDI.***

Prediction & Design of New Drugs

- ***Prediction of 3-D PfDHFR using bacterial DHFR and homology modeling approach.***
- ***Search for the compounds using bifunctional basic groups that could form stable H-bonds in a plane with carboxyl group.***
- ***Optimize the structure of small molecules and then dock them on PfDHFR model.***
- ***Toyoda et. al. (1997). BBRC 235:515-519 could identify two compounds.***

How molecular modeling could be used in identifying new leads

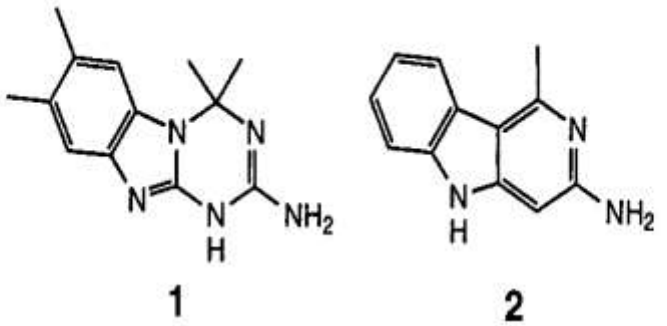


Figure 3. *Inhibitors* ***1****, a triazinobenzimidazole, and* ***2****, a pyridoindole, were found to be active with* K_i *values of 0.54 and 8.7 μM, respectively, against recombinant* P. falciparum *dihydrofolate reductase*[20].

- ***These two compounds***
  - ***a triazinobenzimidazole***
  - ***&***
  - ***a pyridoindole were found to be active with high*** K_i ***against recombinant wild type DHFR.***

- ***Thus demonstrate use of molecular modeling in malarial drug design.***

Genome Update: Public domain

- ***Published Complete Genomes:*** ***59***
  - ***Archaeal*** ***9***
  - ***Bacterial*** ***36***
  - ***Eukaryal*** ***14***

- ***Ongoing Genomes:*** ***335***
  - ***Prokaryotic*** ***203***
  - ***Eukaryotic*** ***132***

Private sector holds data of more than 100 finished & unfinished genomes.

Challenges in Post-Genomic era: Unlocking Secretes of quantitative variation

- ***For even after genomes have been sequenced and the functions of most genes revealed, we will have no better understanding of the naturally occurring variation that determines why one person is more disease prone than another, or why one variety of tomato yields more fruit than the next.***
- ***Identifying genes like fw2.2 is a critical first step toward attaining this understanding.***

Value of Genome Sequence Data

- ***Genome sequence data provides, in a rapid and cost effective manner, the primary information used by each organism to carry on all of its life functions.***

- ***This data set constitutes a stable, primary resource for both basic and applied research.***

- ***This resource is the essential link required to efficiently utilize the vast amounts of potentially applicable data and expertise available in other segments of the biomedical research community.***

Challenges

- ***Genome databases have individual genes with relatively limited functional annotation (enzymatic reaction, structural role)***

- ***Molecular reactions need to be placed in the context of higher level cellular functions***

Data Mining: Finding the Needle in the Haystack

- Data mining refers to the new genre of BI tools used to sift through the mass of raw data.
- DM applications should be able to process -
  - **TEMPORAL** (Time studied) and
  - **SPATIAL** (Organism, organ, cell type etc)
  - The gained 'knowledge' to reprocess data.
  - Data using techniques beyond Bayesian (similarity search) methods.
- An extension of DM is the concept of **'KNOWLEDGE DISCOVERY'**, which open up new avenues of research with new questions and different perspectives.

Other Areas of Research leading to Drug Discovery

- ***Life Science – Biosensors, Genome Sequence and genetic disorders, Microbiome, Stem Cell and organ culture, Biotechnology, Genetically Modified Organisms, Developmental Biology, Bio-energy, Astrobiology, Neuroscience, Medical Technology***
- ***Artificial Intelligence***
- ***Machine Learning***