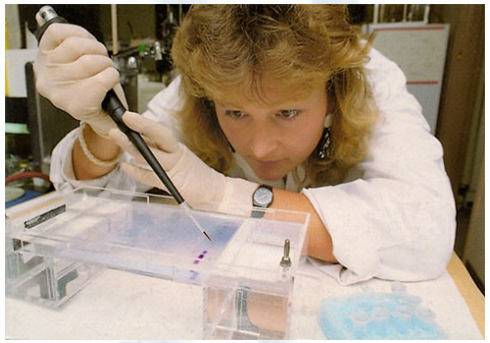
BIOINFORMATICS

Bioinformatics application in Drug Discovery

P. Paulsharma Chakravarthy

BIOINFORMATICS

The “old” biology



The most challenging task for a scientist is to get good data

P. Paulsharma Chakravarthy

The “new” biology

The most challenging task for a scientist is to make sense of lots of data

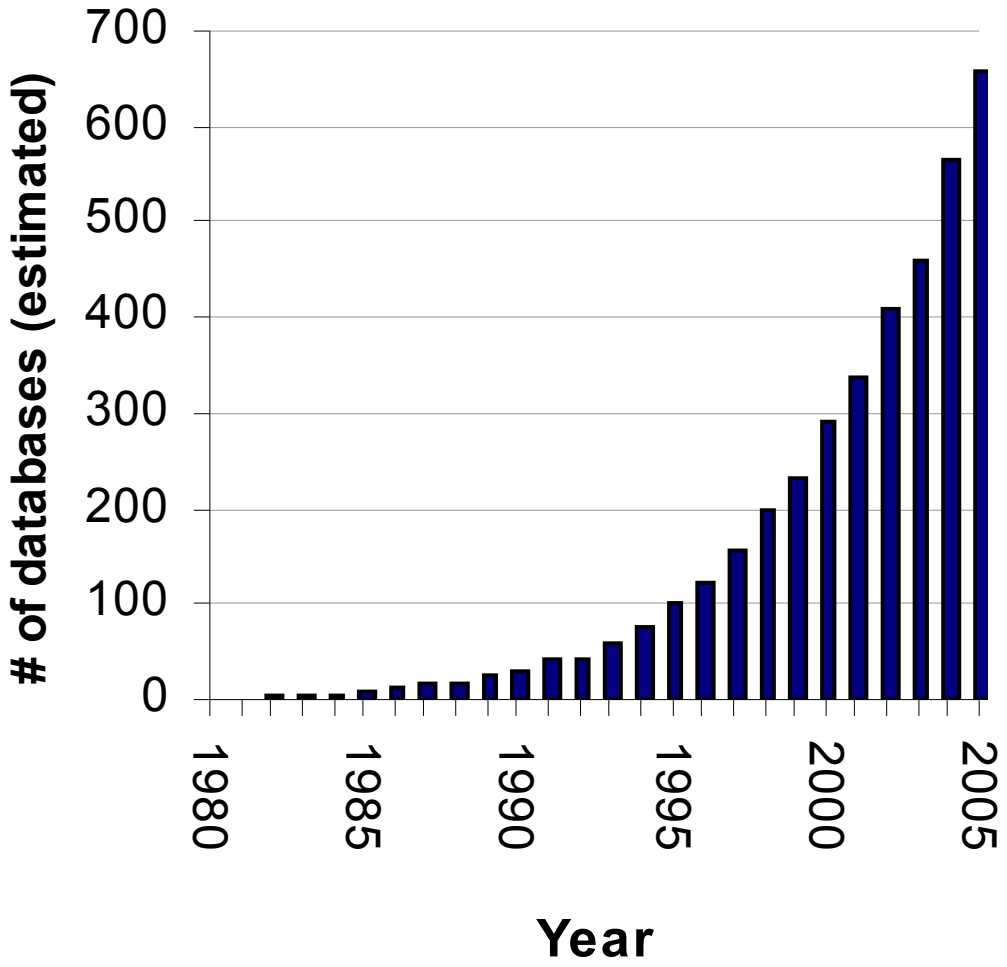
P. Paulsharma Chakravarthy

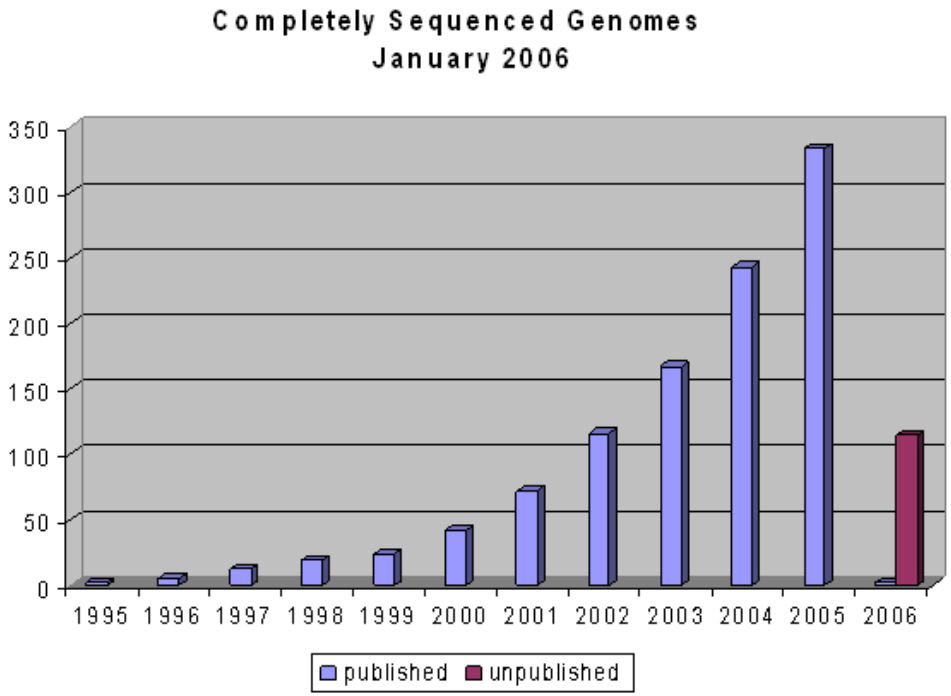
Old vs New - What's the difference?

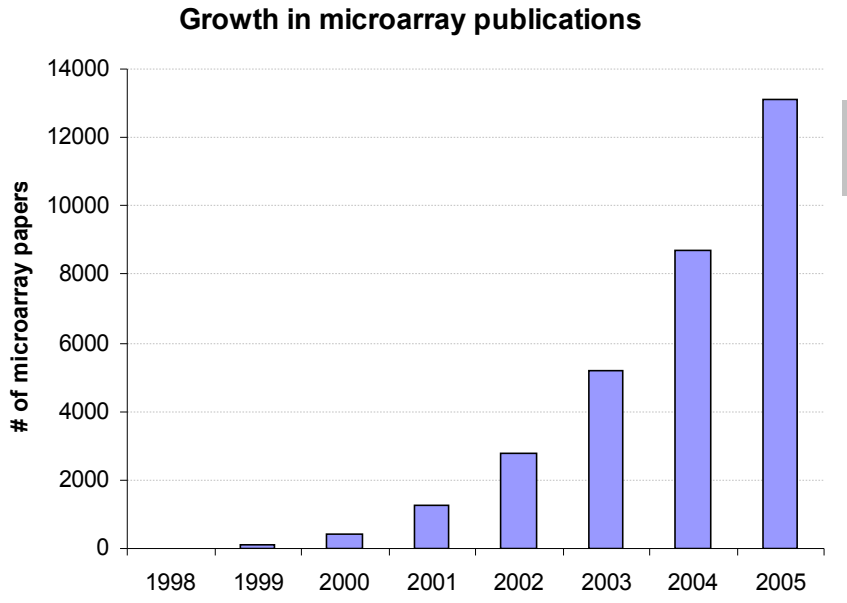
- Miniaturize – less cost
- Multiplex – more data
- Parallelize – save time
- Automate – minimize human intervention
- Thus, you must be able to deal with large amounts of data and trust the process that generated it

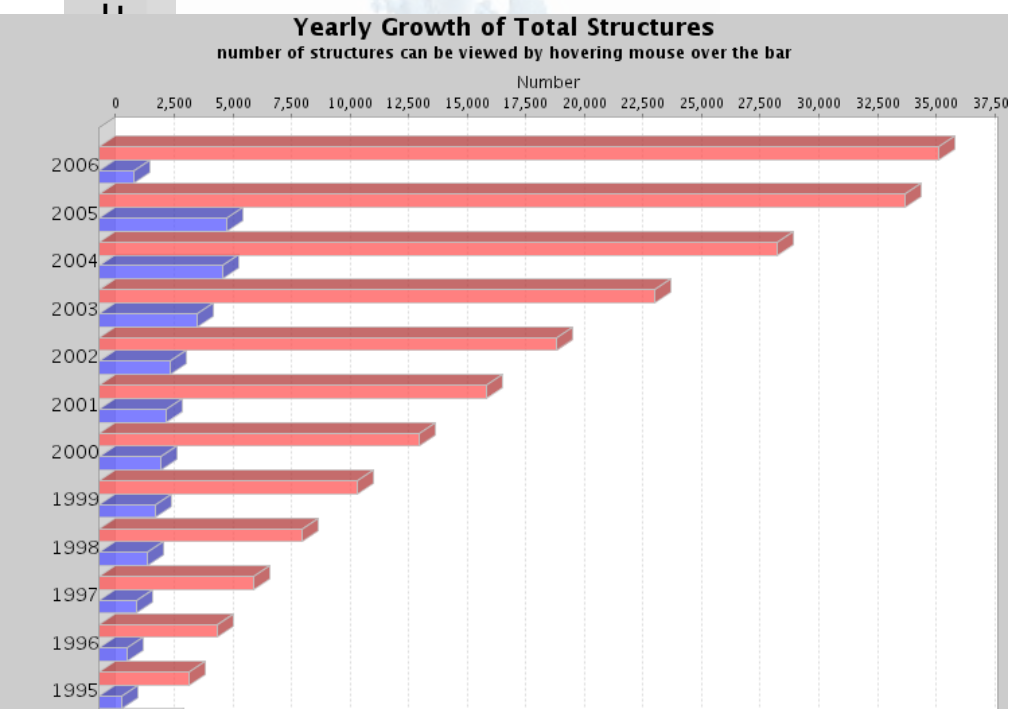
BIOINFORMATICS

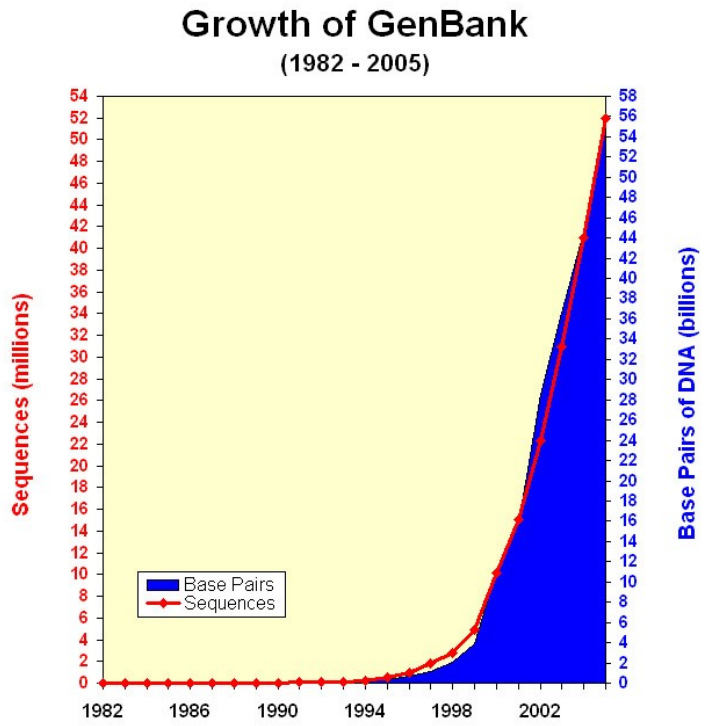
Data is being collected faster and in greater amounts




Completely Sequenced Genomes
January 2006
350
300
250
200
150
100
50
0
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006
published
unpublished

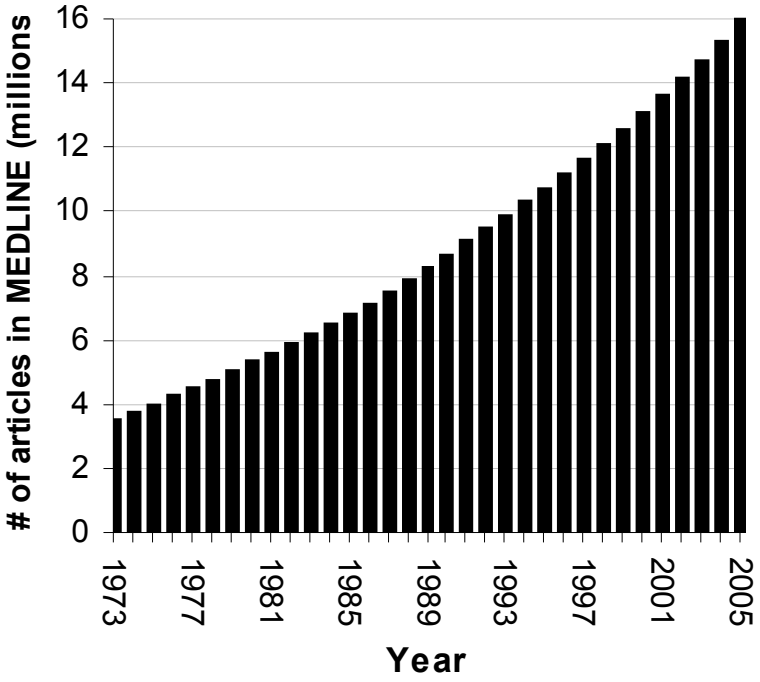

Growth in microarray publications
# of microarray papers
14000
12000
10000
8000
6000
4000
2000
0
1998 1999 2000 2001 2002 2003 2004 2005


Yearly Growth of Total Structures
number of structures can be viewed by hovering mouse over the bar
Number
0 2,500 5,000 7,500 10,000 12,500 15,000 17,500 20,000 22,500 25,000 27,500 30,000 32,500 35,000 37,50
2006
2005
2004
2003
2002
2001
2000
1999
1998
1997
1996
1995


Growth of GenBank
(1982 - 2005)
Sequences (millions)
Base Pairs of DNA (billions)
Base Pairs
Sequences
1982 1986 1990 1994 1998 2002

BIOINFORMATICS

Growth in information & knowledge



>4,800 Journals

>16,000,000 records

672,000 new papers in 2005
(~1,840 per day)

P. Paulsharma Chakravarthy

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.

6. Identify target disease
7. Study Interesting Compounds
8. Detection the Molecular Bases for Disease
9. Rational Drug Design Techniques
10. Refinement of Compounds
11. Quantitative Structure Activity Relationships (QSAR)
12. Solubility of Molecule
13. Drug Testing

Bioinformatics Tools

Identify Target Disease:-

1. One needs to know all about the disease and existing or traditional remedies. It is also important to look at very similar afflictions and their known treatments.

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

Bioinformatics Tools

Identify Target Disease:-

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

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

Bioinformatics Tools

Study Interesting Compounds:

1. One needs to 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.

Bioinformatics Tools

Detect the Molecular Bases for Disease:-

3. 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.

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

Bioinformatics Tools

Detect the Molecular Bases for Disease:-

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

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

Bioinformatics Tools

Refinement of compounds:-

- Once you got a number of lead compounds have been found, computational and laboratory techniques have been very successful in refining the molecular structures to give a greater drug activity and fewer side effects.
- Done both in the laboratory and computationally by examining the molecular structures to determine which aspects are responsible for both the drug activity and the side effects.

Computer-Aided Drug Design (CADD)

- Computer-Aided Drug Design (CADD) is a specialized discipline that uses computational methods to simulate drug-receptor interactions.
- CADD methods are 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 into drug compounds.

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

Bioinformatics Supports CADD Research

Virtual High-Throughput Screening (vHTS):-

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

4. With today’s computational resources, several million compounds can be screened in a few days on sufficiently large clustered computers.

5. Pursuing a handful of promising leads for further development can save researchers considerable time and expense.

e.g.. ZINC is a good example of a vHTS compound library.

Bioinformatics Supports CADD Research

Sequence Analysis:-

3. In CADD research, one often knows the genetic sequence of multiple organisms or the amino acid sequence of proteins from several species.
4. It is very useful to determine how similar or dissimilar the organisms are based on gene or protein sequences.
5. With this information one can infer the evolutionary relationships of the organisms, search for similar sequences in bioinformatic databases and find related species to those under investigation.
6. There are many bioinformatic sequence analysis tools that can be used to determine the level of sequence similarity.

Bioinformatics Supports CADD Research

Homology Modeling:-

3. Another common challenge in CADD research is determining the 3-D structure of proteins.

2. Most drug targets are proteins, so it's important to know their 3-D structure in detail. It's 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.

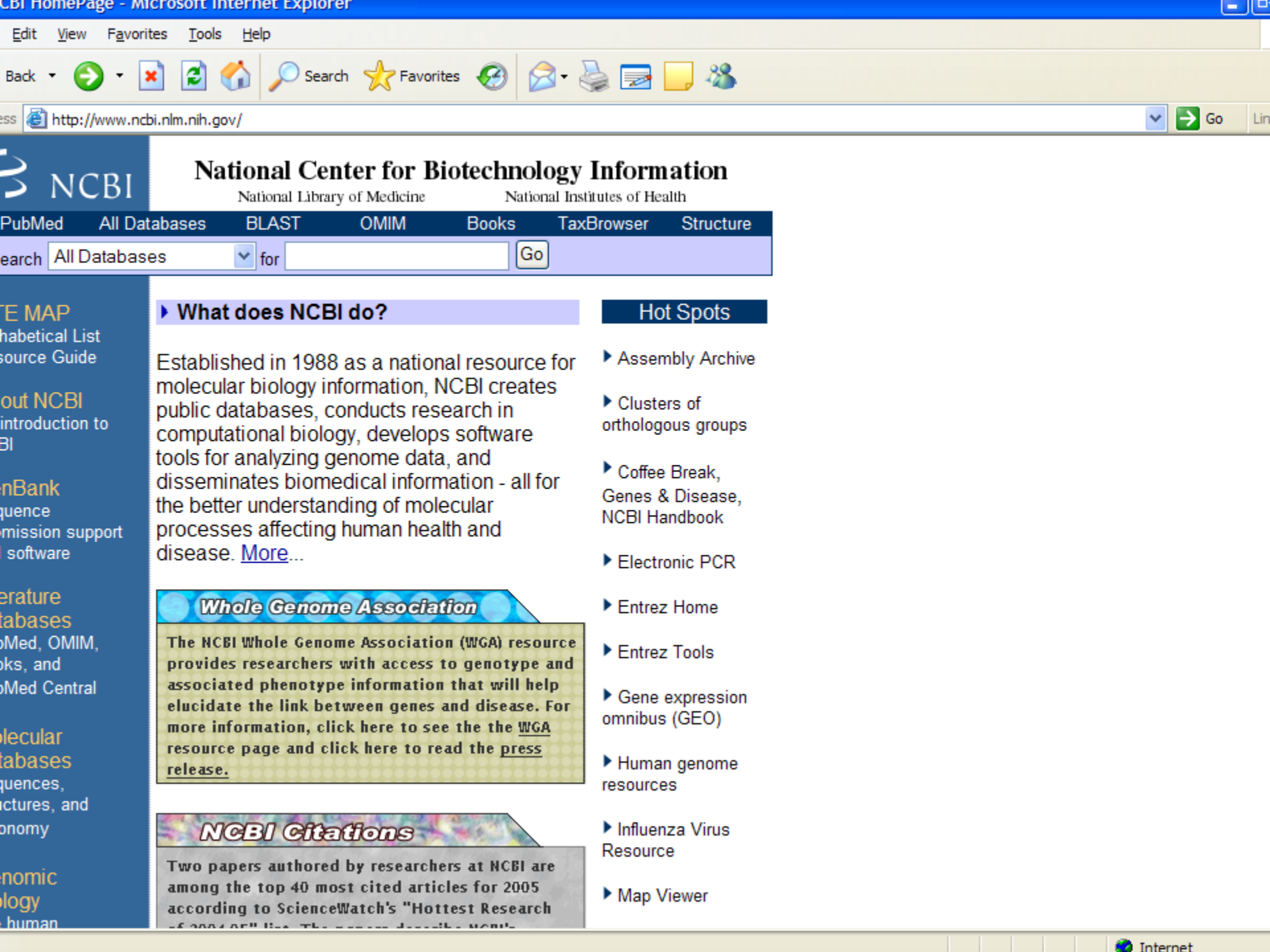
Bioinformatics Supports CADD Research

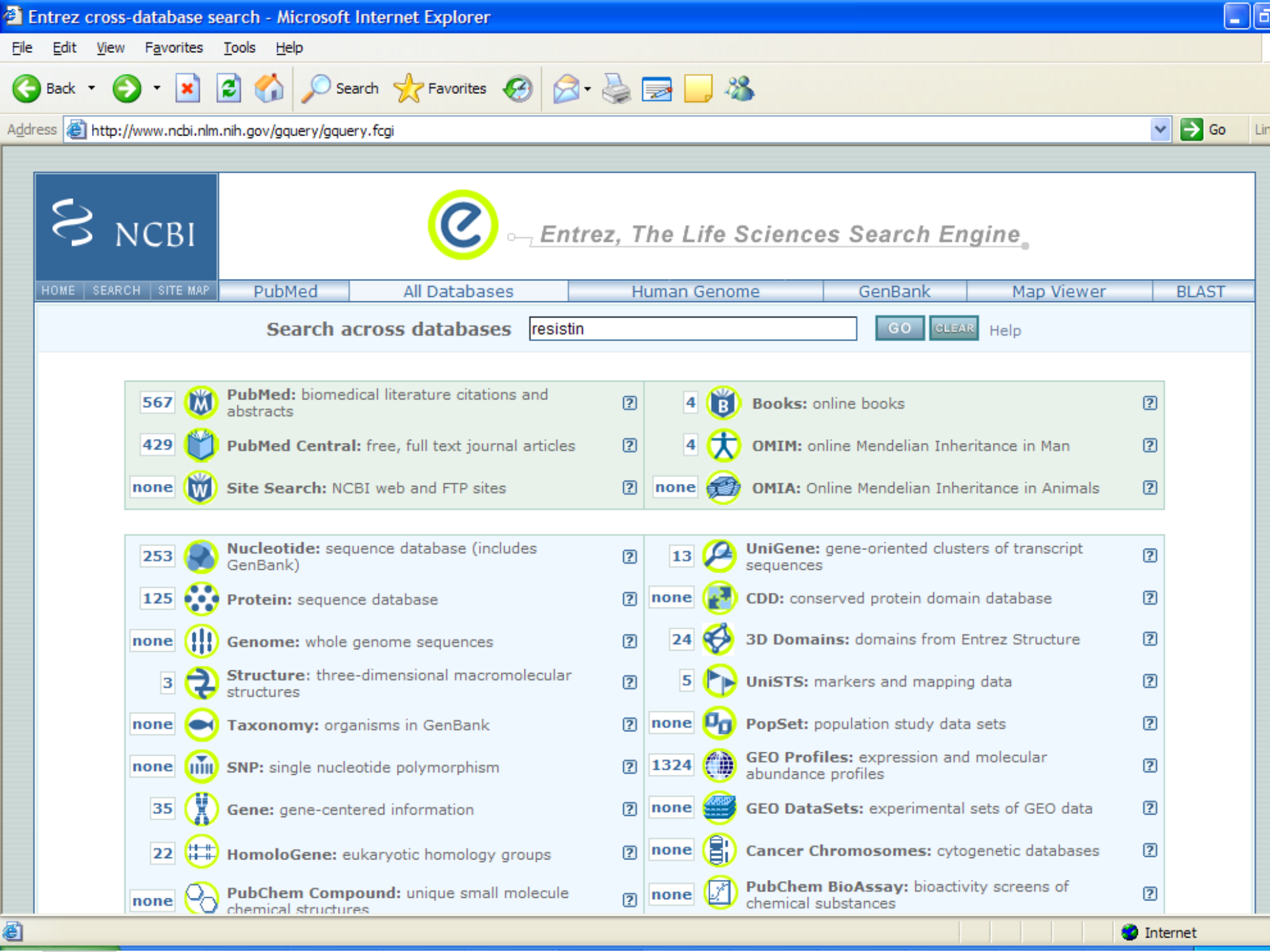
Homology Modeling:-

4. In homology modeling, the amino acid sequence of a specific protein (target) is known, and the 3-D structures of proteins related to the target (templates) are known.

5. Bioinformatics software tools are then used to predict the 3-D structure of the target based on the known 3-D structures of the templates.

6. MODELLER is a well-known tool in homology modeling, and the SWISS-MODEL Repository is a database of protein structures created with homology modeling.


CBI HomePage - Microsoft Internet Explorer
Edit View Favorites Tools Help
Back Search Favorites
http://www.ncbi.nlm.nih.gov/
Go
NCBI
National Center for Biotechnology Information
National Library of Medicine National Institutes of Health
PubMed All Databases BLAST OMIM Books TaxBrowser Structure
earch All Databases for Go
TE MAP
habetical List
source Guide
out NCBI
introduction to
BI
nBank
quence
mission support
software
erature
tabases
oMed, OMIM,
oks, and
oMed Central
olecular
tabases
quences,
uctures, and
onomy
nomic
ology
human
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
Whole Genome Association
The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the the WGA resource page and click here to read the press release.
NCBI Citations
Two papers authored by researchers at NCBI are among the top 40 most cited articles for 2005 according to ScienceWatch's "Hottest Research
Hot Spots
Assembly Archive
Clusters of orthologous groups
Coffee Break, Genes & Disease, NCBI Handbook
Electronic PCR
Entrez Home
Entrez Tools
Gene expression omnibus (GEO)
Human genome resources
Influenza Virus Resource
Map Viewer
Internet


Entrez cross-database search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi
Go
NCBI
Entrez, The Life Sciences Search Engine
HOME SEARCH SITE MAP
PubMed All Databases Human Genome GenBank Map Viewer BLAST
Search across databases
resistin
GO CLEAR Help
567 PubMed: biomedical literature citations and abstracts
429 PubMed Central: free, full text journal articles
none Site Search: NCBI web and FTP sites
4 Books: online books
4 OMIM: online Mendelian Inheritance in Man
none OMIA: Online Mendelian Inheritance in Animals
253 Nucleotide: sequence database (includes GenBank)
125 Protein: sequence database
none Genome: whole genome sequences
3 Structure: three-dimensional macromolecular structures
none Taxonomy: organisms in GenBank
none SNP: single nucleotide polymorphism
35 Gene: gene-centered information
22 HomoloGene: eukaryotic homology groups
none PubChem Compound: unique small molecule chemical structures
13 UniGene: gene-oriented clusters of transcript sequences
none CDD: conserved protein domain database
24 3D Domains: domains from Entrez Structure
5 UniSTS: markers and mapping data
none PopSet: population study data sets
1324 GEO Profiles: expression and molecular abundance profiles
none GEO DataSets: experimental sets of GEO data
none Cancer Chromosomes: cytogenetic databases
none PubChem BioAssay: bioactivity screens of chemical substances
Internet

BIOINFORMATICS

Thanks You

Mail: bioinfopaul@gmail.com
Tel: +91-9884042119