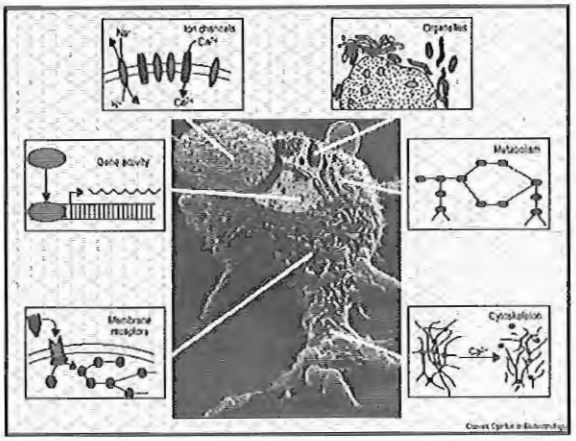
SY11 USE OF BIOINFORMATICS IN DRUG DISCOVERY

รศ.ดร.สุรศักดิ์ วงศ์รัตนชีวิน

ภาควิชาจุลชีววิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

Use of Bioinformatics in Drug Discovery

Surasak Wongratanacheewin, Ph.D.
Department of Microbiology,
Faculty of Medicine, KKU
http://web.kku.ac.th/~sura_wng/lecture.html
E-mail:sura_wng.kku.ac.th



Bacteria or Hunman
⇩
Tissue or Cells
⇩
DNA
⇩
DNA Seguencing
⇩
Gene Anotation
⇩
Functional analysis by Bioinfomatics or Proteomics

In vitro
In vivo
In silico

What is Bioinformatics

Bioinformatics

- is a science of recent creation that uses biological data and knowledge stored in computer databases,complemented by computational methods, to derive new biological knowledge.

Bioinformatics

- is making a key contribution to the organization and analysis of the massive amount of biological data from genome sequencing projects and, increasingly, from other areas of 'high-throughput','massively parallel', robotized and miniaturized methods of biological experimentation.

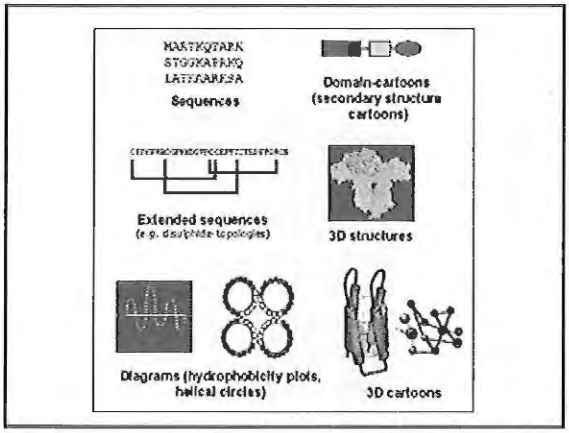
Bioinformatics Goals

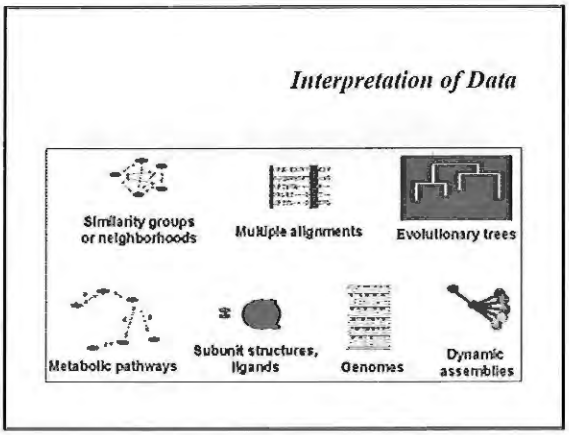
The ultimate goal of the field is

- to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned.
- There are three important sub-disciplines within bioinformatics:

Sub-disciplines within bioinformatics

- The development of new algorithms and statistics with which to assess relationships among members of large data sets.
- The analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and
- The development and implementation of tools that enable efficient access and management of different types of information.





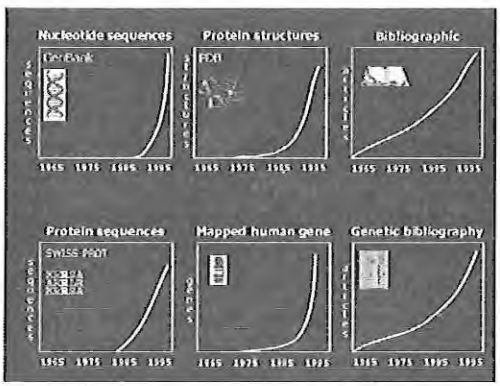
Why use Bioinformatics

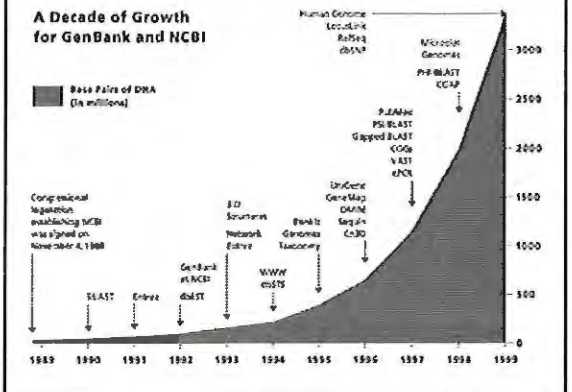
The rationale for applying computational approaches to problems in biology include:

- an explosive growth in the amount of biological information which necessitates the use of computers for information cataloging and retrieval.
- a more global perspective in experimental design. As we move from the one scientist-one gene/protein/disease paradigm of the past to a consideration of whole organisms, we gain opportunities for new insights into health and disease.

Why use Bioinformatics

- Data-mining - the process by which testable hypotheses are generated regarding the function or structure of a gene or protein of interest by identifying similar sequences in better characterized organisms. For example, new insight into the molecular basis of a disease may come from investigating the function of homologs of the disease gene in model organisms. Equally exciting is the potential for uncovering phylogenetic relationships and evolutionary patterns

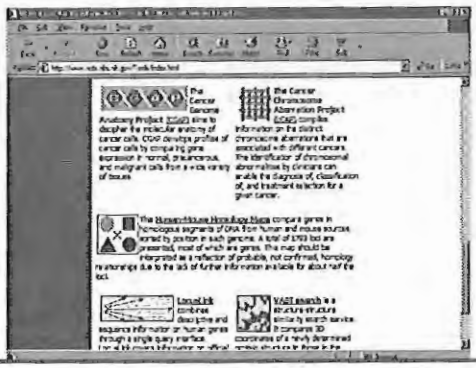




Data Mining

Data Mining

Data Mining



The Bioinformatics Gold Rush

Genomic Projects

- generating gigantic databases containing the details of
- When and in which tissue of the body various genes are turned on.
- The shapes of the proteins the genes encode
- How the proteins interact with another and the role those interactions play in disease.

The Bioinformatics Gold Rush

- Gene Myers, Jr., Vice president of informatics research at Celera Genomics ::: The new discipline of bioinformatics- a marriage between computer science and biology. It is destined to change the face of biomedicine.
- The race and competition will be who can mine it best. There will be such a wealth of riches.

Bioinformatics and Drug Companies

The reason drug companies are so willing to line up and pay for bioinformatics are.

- It offer the prospect of finding better drug targets earlier in the drug development process. It could mean you could grab maybe $500 million in sales.
- It significantly decresing overall costs.
- It lengthening the time a drug is on the market before its patent expires.

Bioinformatics and Drug

- Cathepsin K an enzyme that might turn out to be an important target for treating osteoporosis.

1993 Researchers, at Smith Kline, asked
Human Genome Scientists ⟹ analyze genetic material from Osteoclasts cell

⇩

Sequence and search for homology

⇩

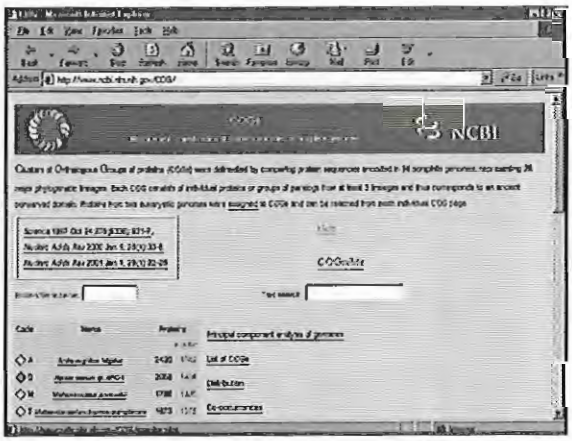
Look for some over expressed proteins: cathepsins

Molecular and Genomic Databases

- **Primary sequence databases**
  - **DDBJ EBI GenBank GSDB**
- **Specialized databases**
  - **Blocks GDS PDB PIR**
  - **PROSITE REBASE SWISS-PROT**

Bioinformatics In gene and drug discovery

- Bioinformatics
- has become a key aspect of drug discovery in the genomic revolution, contributing to both target discovery and target validation.
- Discussion
- on genome-wide data sources that have become available to the industry, including expressed sequence tags, microbial genome sequences, model organism sequences, polymorphisms, gene expression data and proteomics. However, these knowledge sources must be intelligently integrated.



Bioinformatics In target validation

The challenge to bioinformatics is evolving from that of

1. Creating long lists of genes.
2. Creating short lists of the targets most likely to be crucial in disease and least likely to fail for 'developability' reasons.
3. Target selection between the desire to study targets that already have a well understood role in disease and the desire to study those that might offer a completely novel mode of action.
4. Bioinformatics should provide the bridge that reconciles these goals, primarily by providing as many clues as possible to function and role.

Bioinformatics In target selection

1. The characterization of targets, such as the classification and subclassification of protein families.
2. The understanding of targets, such as their behavior in a larger biochemical and/or cellular context.
3. The development of targets, such as making predictions about uptake or reuptake, detoxification, the stratification of patient populations and other gene-based variations.

Bioinformatics In data integration

Bioinformatics will established "wet-dry" cycles.

Refinements of models

Computational models- Biological models

Prediction

Bioinformatics In data integration

Data Integration

1. Expressed sequence tags (ESTs)
2. Microbial sequences
3. Human genome sequences
4. Single-nucleotide polymorphisms (SNPs)

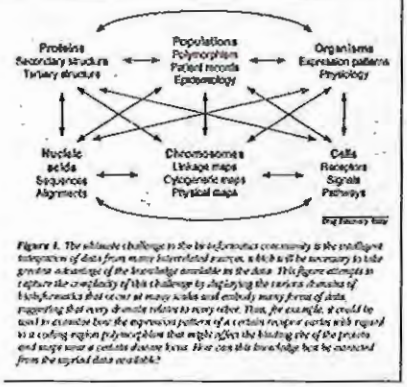


Figure 1. The ultimate challenge in the bioinformatics community is the intelligent integration of data from many interrelated sources, which will be necessary to take greatest advantage of the knowledge available in the data. This figure attempts to capture the complexity of this challenge by depicting the various domains of bioinformatics that occur at many scales and embody many forms of data, suggesting that every domain relates to many others. Thus, for example, it could be useful to examine how the expression pattern of a certain receptor varies with regard to a coding region polymorphism that might affect the binding site of the protein and maps near a certain disease locus. How can this knowledge best be extracted from the myriad data available?

Expressed sequence tags (ESTs)

The immediate challenge that faced the industry in making use of these data was that of data management.

build a computational infrastructure

⇩

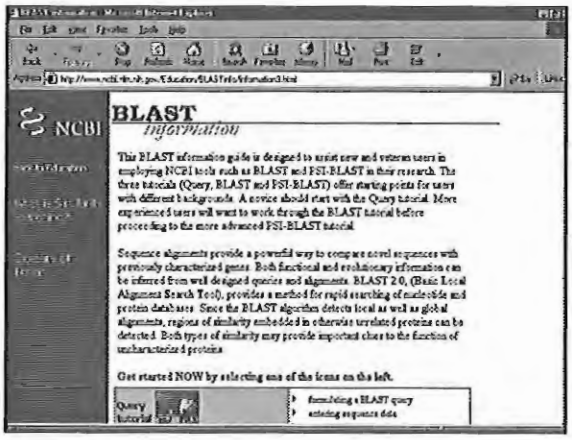
rapidly identify interesting homologues
BLAST (Basic Local Alignment Search Tool) database searches
and then present the results to biologists for evaluation.

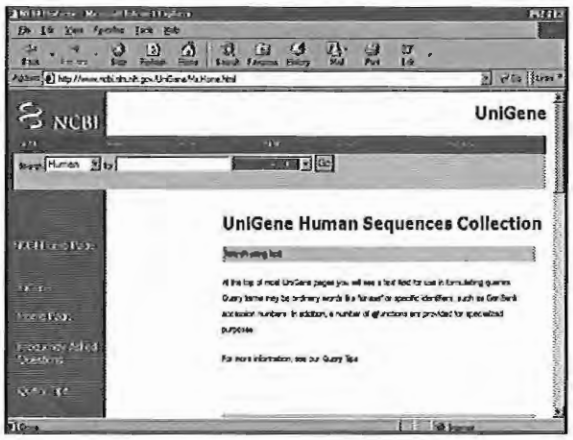
⇩

ESTs

⇩

clustering or (even better) by assembling
overlapping fragments to create a dataset that more closely
represented the actual underlying genes (public Unigene
resource)





Bacterial genome sequencing and drug discovery

- 1928, Alexander Fleming first demonstrated that a fungal extract could inhibit the growth of the bacterial pathogen.
- The development of Penicillin required 13 years.
- 1995, *Haemophilus influenzae* genome complete.
- 1996, *Saccharomyces cerevisae.* 1997, *Escherichai coli*
- The drug discovery paradiam has shift away from finding compound active against whole cells to identifying compound that are active against selected protein targets.

Factors influencing the need for new antibiotics.

•**Drug resistance** multiple resistance systems are particularly damaging, often severely limiting options for effective treatment. Many resistance mechanisms are also mobile and have spread rapidly through bacterial species.

•**More immunocompromised patients.** Numbers of patients more susceptible to infection are growing, partly due to an ageing population but also as a result of advances in other areas of medicine.

•**Advances in surgery.** Patients with any sort of indwelling device, from hip joints to mechanical assisted hearts are more prone to infection.

Factors influencing the need for new antibiotics

•**Greater awareness.** There are now numerous of the role numerous examples of bacterial of infectious agents may bacterial infection being play in other examples implicated in various diseases conditions, the best documented being *Helicobacter pylori* and peptic ulcers and probably stomach cancer .

•**Advances in diagnostics.** More powerful bacterial diagnostics are just on the horizon, however, the ability to detect more'unusual' pathogens, such as *Mycoplasma* and *Chlamydia,* have already demonstrated a high association with community acquired pneumonia cases in the US.

Factors influencing the need for new antibiotics.

Advances in diagnostics. More powerful bacterial diagnostics are just on the horizon, however, the ability to detect more'unusual' pathogens, such as *Mycoplasma* and *Chlamydia,* have already demonstrated a high association with community acquired pneumonia cases in the US.

Comparison of the relevant properties of potential antibacterial targets.

Decision matrix	
Essential	Known, likely to be, unknown?
Novel	At least not associated with Known resistance mechanisms
Spectrum	Conservation of target sequence and function in RTI, UTI, STD pathogens etc.
Selective	Comparison against nearest human homologue and function

Comparison of the relevant properties of potential antibacterial targets.

Function	Known, likely to be, unknown?
Chemically tractable	Enzyme or macromolecular binding site
Assay	Know, likely to be, unknown?
Accessible	Inside or outside cell membrane
Patient position	
Structural information	

Microbial genomes

- Data management
- Large scale anotation of genome identifying gene
- Phylogenetic distribution of genes
- High degree of similarity across bacteria.
- Putative targets must also be essential for the survival of the pathogens etc citric acid cycles in 19 complete genomes.
- Conservation of the gene, lateral gene transfer.

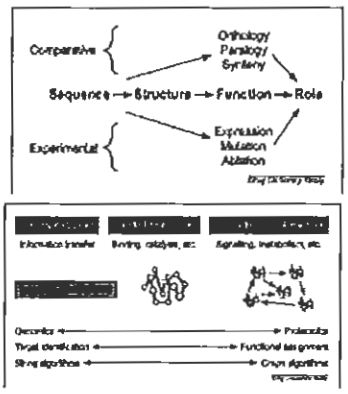
Genome Sequence

- 10-20% are missing from EST.
- Low abundance when expressed or a highly specific pattern of expression.
- Best drug targets precisely.
- The potential to predict more full-length genes. This will be useful for screening mutations, analysis of promoters and regulatory regions, cluster of related genes, relationship to model-organism genome and candidate genes from mapping studies.

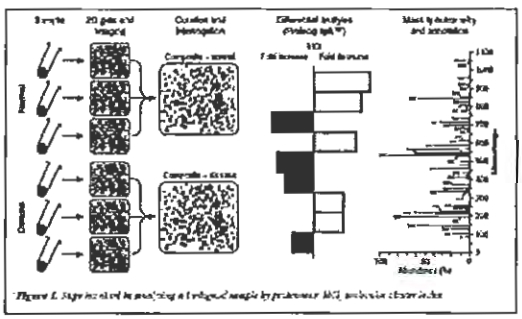
Polymorphisms

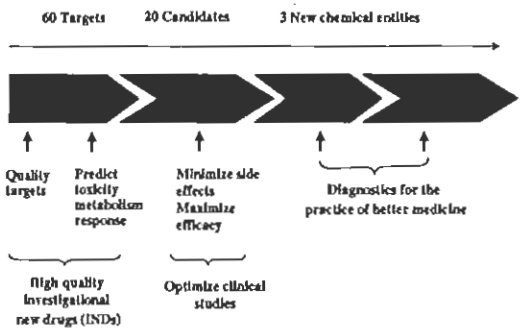
- Genetic markers in the human genome.
- Microsatellite markers
- SNPs and the mapping of disease associations
- underlying genes

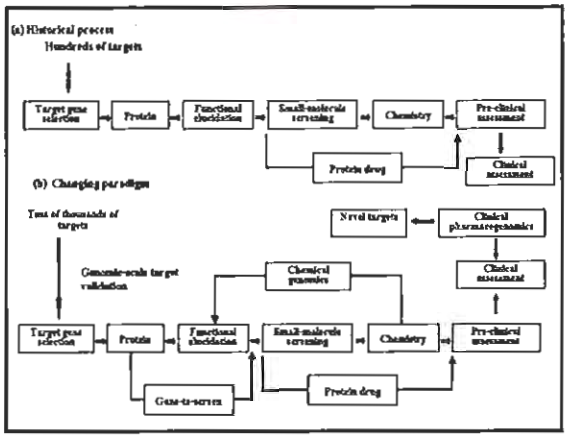
This all information will need bioinformatics to manage and intrepretation leading to candidate genes for drug development



Proteomics







Genomic for target identification and validation

- Tenet of genomics-based drug discovery is
- Sequence database searching that can separate organism's genes into several broad classifications
  1. Genes that are conserved amongst all/many living organisms.
  2. Genes conserved within a particular phylogenetic Kingdom.
  3. Genes conserved within a particular Order or Geuns.
  4. Organism-specific gene.

Genomic for target identification and validation

The information will be useful for target identification. Valid drug targets are

Those gene that encode proteins (sturtural/catalytic RNA molecules) required for the cell to grow. These can be done by screening for temperature-sensitive mutation, transposon mutagenesis. Extensive amalysis of bacterial and the *S. cerevicae* genome have found that roughly 25% of the genes are required for normal growth.

Genomic for target identification and validation

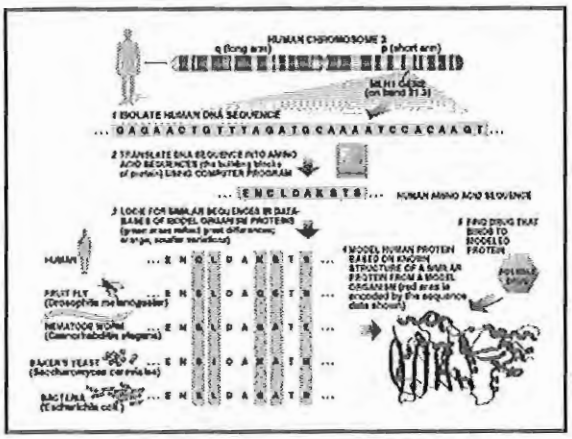
- After comparing with mammalian, from these genes, less than 10% of microbe's genes is a potential selective agents for antimicrobials.
- For *E. coli*, this translates into 400 potential target genes.
- Genetic-based "generic " assay used to screen the targets.
- Cell-free assay formats for genomic targets.
- Genomic target prioritization.

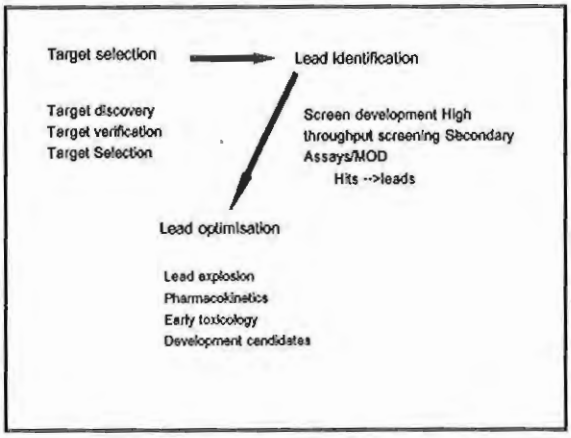
Searching for anti-microbial drug targets

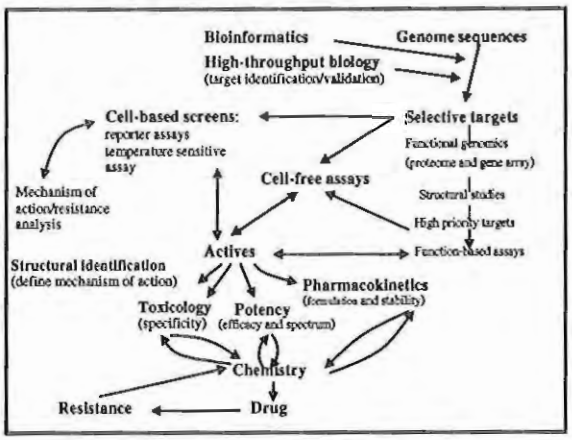
- Identify conserved genes.
- Select the ones, essential for survival of the bacteria.
- Find out the cellular function(s) of it.
- Design or select an inhibitor.
- Determine the active dosage *in vitro*
- Check for toxicity
- Check for activity *in vivo*

Clinical properties translated to molecular target characteristics

Clinical profile	*Molecular target*
Essential	
• Broad spectrum	→ **Present in all target species
• Required function	→ *Essential or growth modifier
• Not associated with known resistance mechanisms	→ **Novel target
• Selective (no inherent toxicity	→ **Absent from host or substantially different
Desirable	
• Lethal	→ Not saved by transient expression
Practical	→ **Accessible
	→ **Function easily established
	→ *Chemically tractable







Acknowledgements

•Faculty of Medicine, Khon Kaen University

•Bruno A Gaeta (Australian National Genomic Information System, ANGIS)

•International Center for Genetic Engineering and Biotechnology (ICGEB) Triste, Italy

•Associate Prof. Dr. Rasana Wongratanacheewin, KKU