BIOINFORMATICS

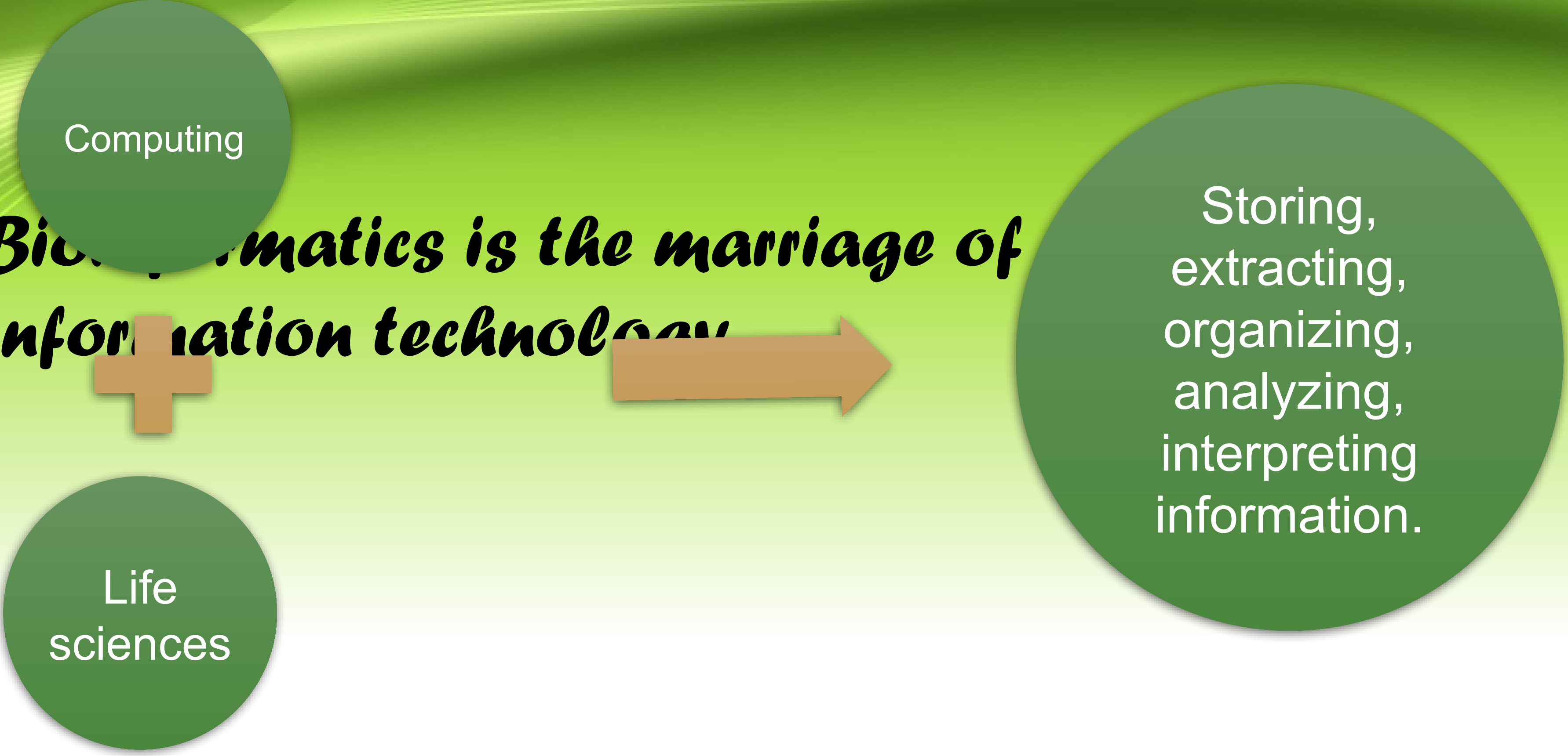
Synopsis

- Introduction
- Definition
- History
- Principle
- Components of bioinformatics
- Bioinformatics databases
- Tools of bioinformatics
- Applications of bioinformatics

✓ Molecular medicine

- ✓ Microbial genomics
- ✓ Plant genomics
- ✓ Animal genomics
- ✓ Human genomics
- ✓ Drug and vaccine designing
- ✓ Proteomics
- ✓ For studying biomolecular structures
- ✓ In- silico testing
- ➡ Conclusion
- ➡ References

- ***Bioinformatics is the marriage of information technology***



Computing

Life sciences

Storing, extracting, organizing, analyzing, interpreting information.

Definition

- the science of using information to understand biology.

- "a subset of larger field of computational biology, the application of quantitative analytical techniques in modelling biological systems".

A short historical overview (I):

- 1954: first protein sequence (insulin by Sanger)
- 1958: first X-ray 3D structure of a protein (myoglobin by Kendrew)
- 1972: first DNA sequencing
- 1977: rapid sequencing techniques (Gilbert and Sanger!)
- 1986: PCR (the photocopying machine of the biologist)
- 1992: sequence of yeast chromosome III ($3*10^5$ bp)
- 1995: sequence of the genome of the bacteria *Haemophilus influenzae* ($2*10^6$ bp)
- 1999: sequence of the genome of a multi-cellular organism (*Caenorhabditis elegans*) (10^8 bp)
- 2000: blue draft of the human genome ($3*10^9$ bp)
- 2002: genome of *Ashbya gossypii*
- Today: 16 archeal, 77 bacterial & 9 eukaryotic genomes completed.

A short historical overview (II):

- 1965: Atlas of protein sequence and structure (Dayhoff)
- 1967: Fitch (phylogenetic trees)
- 1970: Needleman / Wunsch (1st similarity search algorithm)
- 1971: PDB (3D structure database)
- 1982: EMBL nucleotide sequence database and Gene Bank
- 1985: CABIOS (1st scientific journal for bioinformatics)
- 1985: FASTP (ancestor of FASTA, Blast, etc.)
- 1986: Swiss-Prot (protein sequence database)
- 1988: Creation of the NCBI in the USA
- 1993: ExPASy (1st WWW server for the life sciences)

Bioinformatics:
A snapshot 10 years ago

- *Pharmaceutical companies were not interested;*

- *Life scientists believed that it was an outlet for failed biologists that want to play around with computers;*

- *Computer scientists did not even consider it important, they confused it with bio-inspired "computer sciences".*

Bioinformatics in 2002

- *Pharmaceutical companies believe that it is the most efficient way to streamline the process of drug discovery;*

- *Some life scientists believe it is the solution to all problems in life sciences and that it will allow them to avoid doing some experiments;*

- *Computer scientists are very interested.The scope and complexity of the domain makes it the ideal field of application of new software techniques and specialized hardware developments.*

Bioinformatics in 2010

- ***Pharmaceutical companies use it routinely, but have realized that it complements rather than replaces experimental work;***

- ***Life scientists use it efficiently every day and therefore forget that it exists;***

- ***Computer scientists may have jumped on another fancy subject.(Spiritual machines ?)***

COMPONENTS OF BIOINFORMATICS

- Creation of databases
- Development of algorithms & statistics
- Analysis of data & interpretation

BIOINFORMATICS DATABASE

- Biodiversity
- DNA sequencing
- RNA sequencing
- Protein Sequencing
- Metabolomes
- Macromolecular structures
- Chemical diversity

NCBI

EMBL
NUCLEOTIDE
SEQUENCE
DATABASE

DDBJ

RCSB PDB
PROTEIN DATA BANK

Bioinformatics tools

FASTA

-for comparing nucleotide or peptide sequence to sequence database.

BLAST-(Basic local alignment search tool)

-for comparing gene & protein sequences against others in public databases.

ClustalW

-multiple sequence alignment program for DNA or proteins.

RasMol

-powerful research tool to display structure of DNA, proteins & smaller molecules.

Applications

- ***MOLECULAR MEDICINE*** – completion of human genome means that we can search for genes directly associated with different diseases & begin to understand molecular basis of diseases. This will enable better treatments, cures etc.
  - a) More drug targets
  - b) Personalized medicine
  - c) Preventive medicine
  - d) Gene therapy

- **MICROBIAL GENOMICS** – arrival of complete genome sequences & their potential to provide greater insight in microbial world.Their capacities can help in environment, health, energy & industrial applications.
- ✓ MGP in 1994, to sequence genomes of bacteria, useful energy production, environmental cleanup, industrial processing, waste cleanup etc.
  - a) Waste cleanup
  - b) Climate change
  - c) Alternative energy sources
  - d) Antibiotic resistance
  - e) Evolutionary studies
  - f) Bioweapons

- PLANT GENOMICS -The sequencing of genomes of plants has enormous benefits for agriculture. Bioinformatics tools can be used for genes with these genomes & to elucidate their functions.This specific genetic knowledge could then be used to produce stronger ,more drought resistant & insect resistant crops & improve crop quality.

  a) Crops
  b) Insect resistance
  c) Improve nutritional quality
  d) Drought resistance

- ***ANIMAL GENOMICS** – sequencing projects of many farm animals including cows, pigs etc. for better understanding of their biology & have huge impacts on improving production & health of livestock & ultimately benefits for human nutrition.*

HUMAN GENOMICS

✓ to study human genome which has been carried out world wide as HGP started in 1990 in U.S. dept. of energy & National institute of health.

✓ Approx. 30,000 genes & 3 billion bp in humans.

✓ 99.9% of bases of all humans are same.

✓ Highest no. of genes present in chromosome 1 (2968 genes).

✓ Fewest no. of genes present in chromosome Y (231).

✓ Out of 30,000 genes, less than 2% of genome codes protein & 50% of genome are non-coding

DRUG & VACC E DESIGNING

✓ drug oach
Of find n,
base
target ed
drug a
computational emistry to
discover , enhance or study
drugs & related
biologically
active molecules.

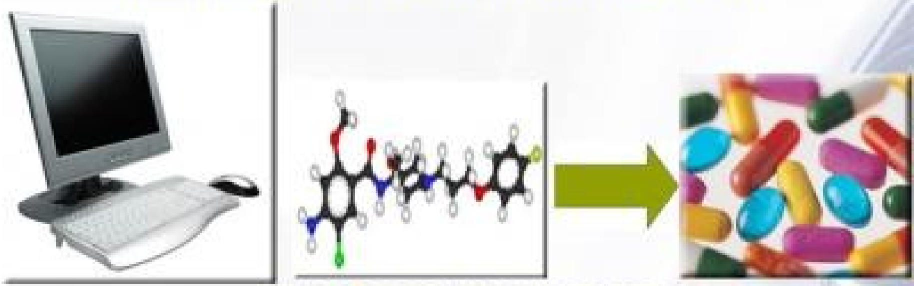
Molecular Biology

Biochemistry

Computer Science

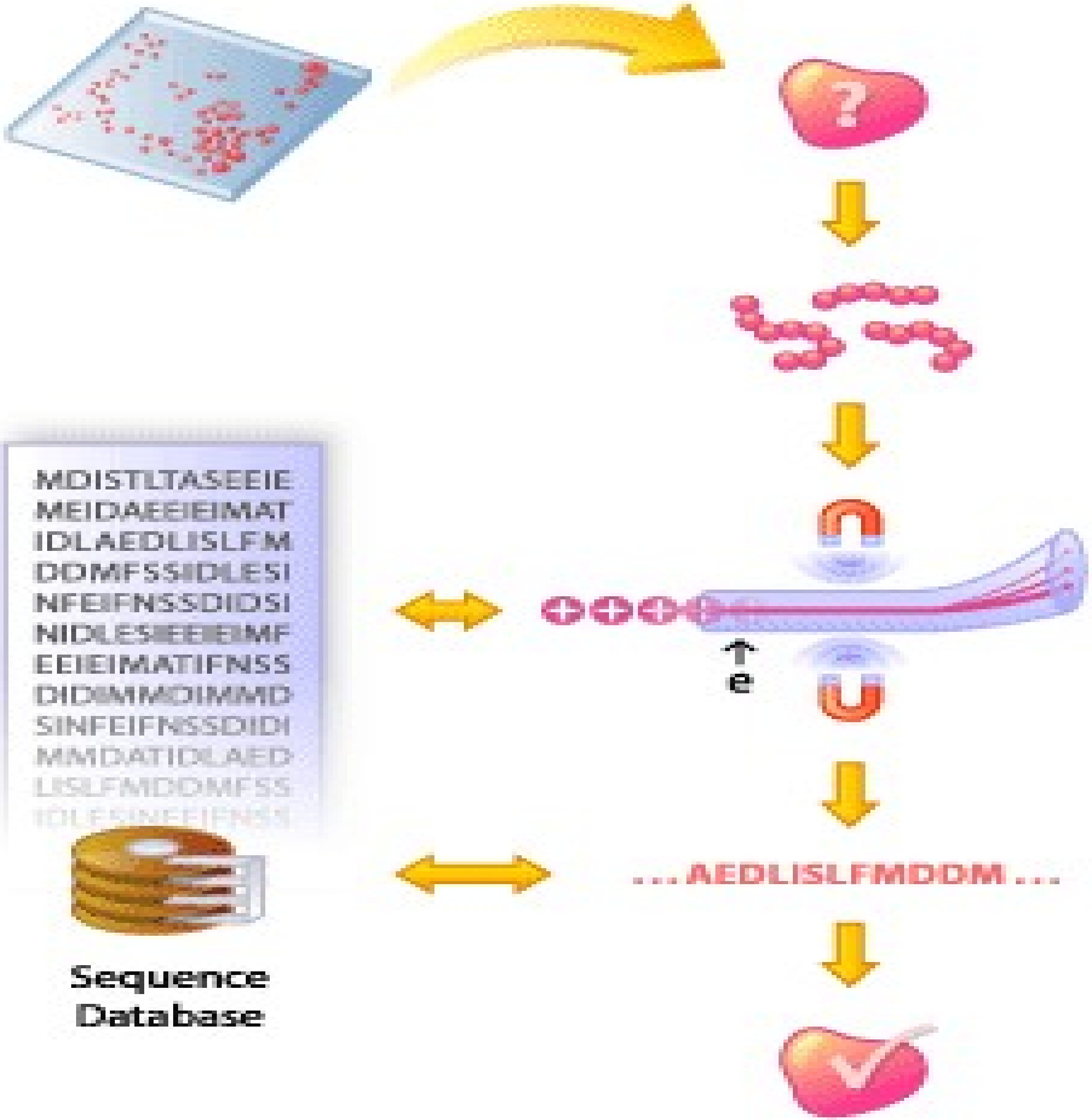
Genetics

DRUG DISCOVERY IN THE TRADITIONAL WAY.



DRUG DISCOVERY IN SILICO

PROTEOMICS



Unidentified protein extracted from gel

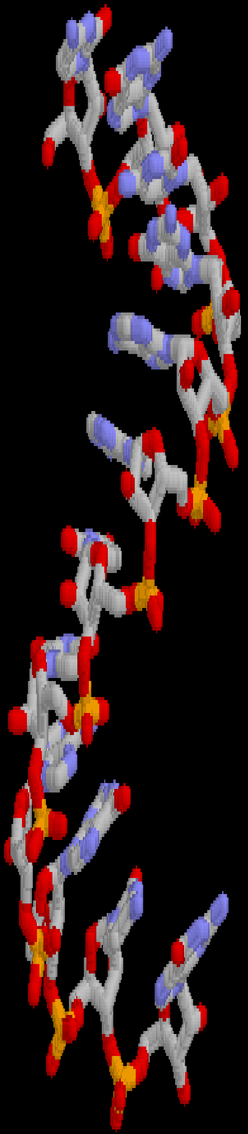
Split into fragments of 5-10 amino acids

Determine mass using MS (Mass Spectrometry)

Determine amino acid sequence and compare with sequence database

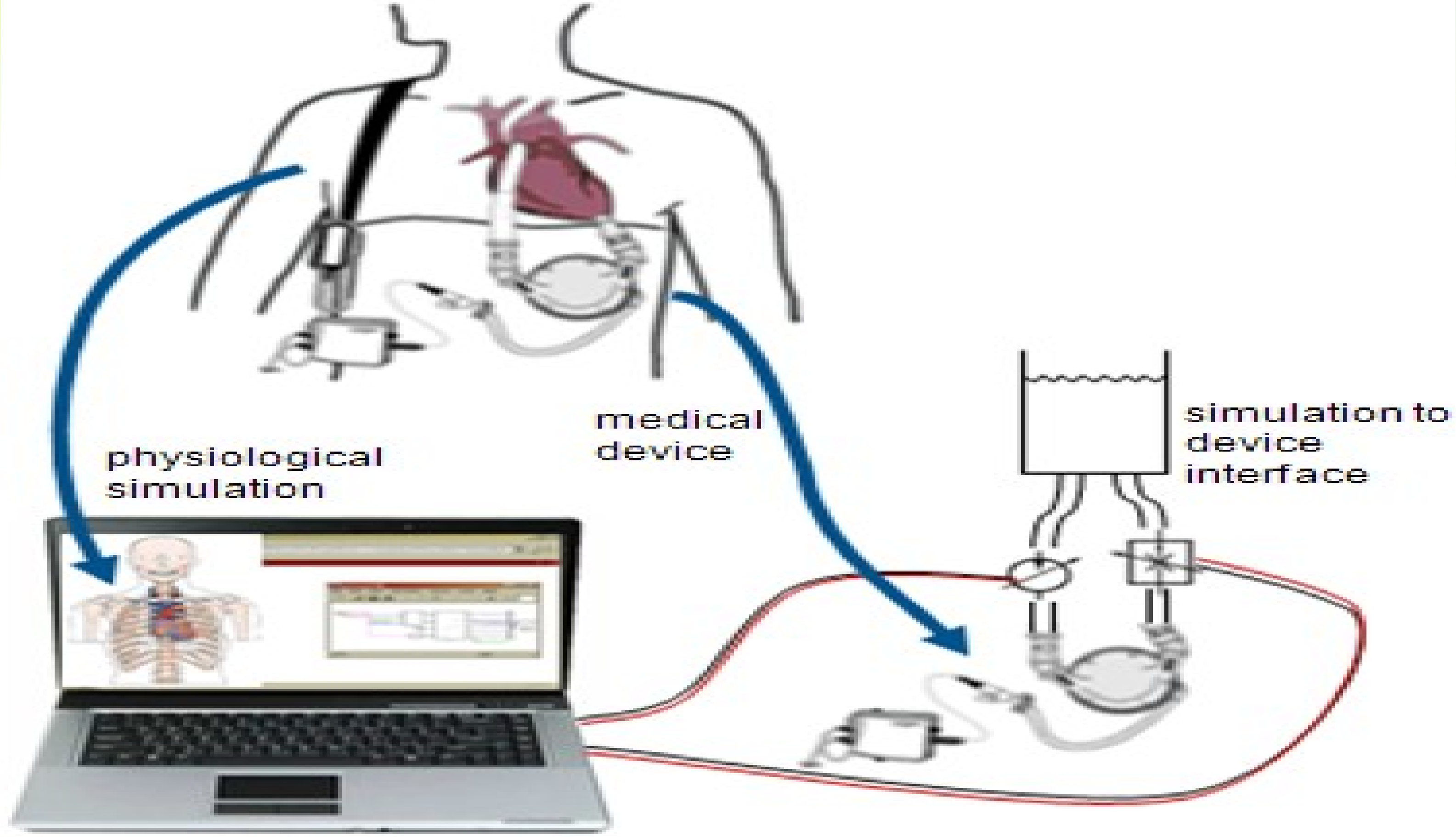
Protein Identified

FOR STUDYING BIOMOLECULAR STRUCTURES



- ✓ Importance of biomolecules is found in various processes like metabolism, differentiation, energetics, signaling etc.
- ✓ Systematic studies of physiochemical properties of biomolecules can permit analysis & prediction of such regulated events.
- ✓ Prediction of protein structure is another important application.
- ✓ In the genomic branch, homology is use to predict the function of a gene.
- ✓ MODELLER is one of the best software for homology modelling.

- ***IN- SILICOTESTING –*** ***provides a greater degree of prediction in animal& human clinical trials which is very fast & cost- effective.***



CONCLUSION