Computational methods in bioinformatics: Lecture 1

Graham J.L. Kemp

4 November 2014

What is biology?

Ecosystem	Rain forest, desert, fresh water lake, digestive tract of an animal
Community	All species in an ecosystem
Population	All individuals of a single species
Organism	One single individual
Organ System	A specialised functional system of an organism, e.g. nervous system or immune system
Organ	A specialised structural system of an organism, e.g. brain or kidney
Tissue	A specialised substructure of an organ, e.g. nervous tissue, smooth muscle
Cell	A single cell, e.g. neuron, skin cell, stem cell, bacteria
Molecule	e.g. protein, DNA, RNA, sugar, fatty acid, metabolites, pharmaceutical drugs

What is bioinformatics?

> *"Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data."*
>
> *"Bioinformatics applies principles of information sciences and technologies to make the vast, diverse, and complex life sciences data more understandable and useful."*

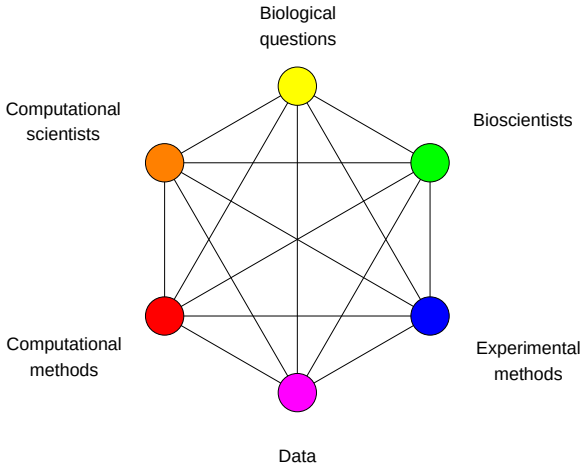
Working definition by the NIH Biomedical Information Science and Technology Initiative Consortium, 2000
http://www.bisti.nih.gov/docs/CompuBioDef.pdf

What is computational biology?

> *"The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.*
>
> *"Computational biology uses mathematical and computational approaches to address theoretical and experimental questions in biology."*

Working definition by the NIH Biomedical Information Science and Technology Initiative Consortium, 2000
http://www.bisti.nih.gov/docs/CompuBioDef.pdf

Addressing biological questions



Sequences, structures and systems

Sequences

- Nucleic acids (DNA and RNA) and proteins are (unbranched) polymers. Their composition can be described by the sequence of units (nucleotides or amino acid residues) in a chain.

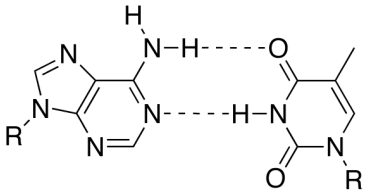
Structures

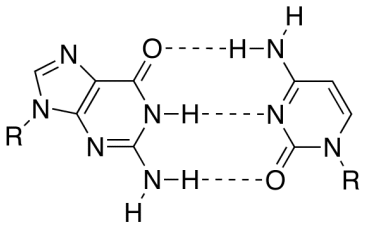
- Three-dimensional structures can give insights into the molecular basis of biological functions.

Systems

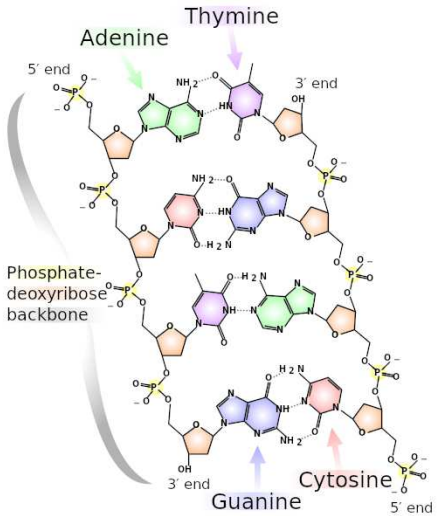
- Biological processes consist of the coordinated actions of molecules.

Base pairing in DNA





Structure of DNA



Protein structure

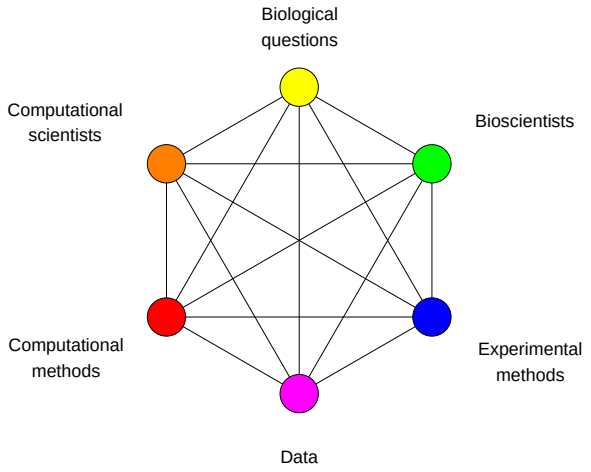
Primary structure
- sequence of amino acid residues linked in a chain

Secondary structure
- locally, the main chain forms helices and strands

Tertiary structure
- the 3-D structure
- assembly and interaction of helices and sheets

Quaternary structure
- assembly of subunits

Addressing biological questions



Biological sequences: some experimental methods

- DNA sequencing
- Protein sequencing
- Next-generation sequencing (NGS)

Biological sequences: some questions

- How similar are a pair of sequences?
- Identify the corresponding units in a pair of homologous molecules that have undergone substitutions and insertions/deletions during their evolutionary history (*pairwise sequence alignment*).
- Given a new sequence, has anything similar (in whole or part) been seen before?
- Reconstruct a phylogenetic tree from the sequences of a set of homologous molecules.
- Given the sequences of many overlapping DNA fragments from a single organism, assemble them to reconstruct a full genome.
- Given the sequences of many DNA fragments from a mixture of organisms, identify the species present in the mixture.

Biological structures: some experimental methods

Find the atomic structure of a macromolecule or complex

- X-ray crystallography
- Nuclear magnetic resonance (NMR) spectroscopy

Identify a low-resolution "envelope" enclosing a large macromolecular complex

- Cryo-electron microscopy
- Small-angle x-ray scattering

Biological structures: some questions

- Can differences in the functions of two similar proteins be explained by differences in their structures?
- Can a drug be designed to fit into the active site of a target protein?
- Can the safety and efficacy of a potential therapeutic protein be predicted from its structure?
- Can the function of a protein be altered by changing its composition, and hence its structure?
- Can a protein's structure be predicted from its sequence?
  - the protein folding problem
- Given the structures of two proteins, will they associate with one another? If so, how will they fit together?
  - the protein docking problem

Biological systems: some experimental methods

Which mRNA molecules are being expressed?

- Microarray gene expression
- RNA-Seq

Which proteins are being expressed?

- (2-D) gel electrophoresis
- Mass spectrometry

In which tissue(s) are particular genes expressed?

- *in situ* hybridization

Biological systems: some questions

- Which genes/proteins are co-expressed (i.e. have similar expression profiles)?
- Which genes are expressed in tumour cells but not in healthy cells?
- If a gene is "knocked out", will an organism survive, and how will the expression of other genes be affected?
- Can protein expression profiles identify proteins that could be targets for drug development?
- Can an individual's expression profile indicate whether they are likely to respond to a particular therapeutic treatment?
- How do biological networks respond to injury or to treatment with a therapeutic drug?