

Computational biology

Cell dynamics group

Department of Mathematical Sciences
Middle Tennessee State University
Box 34 Murfreesboro TN 37132

email:....@mtsu.edu

Information flow

- A major task in computational molecular biology is to *decipher* information contained in biological sequences.
- Since the nucleotide sequence of a genome contains all information necessary to produce a functional organism, we should in theory be able to duplicate this decoding using computers.

Review of basic biochemistry

- DNA makes RNA makes protein
- Sequence determines structure determines function

Structure

- macromolecular structure divided into
 - primary structure (1D sequence)
 - secondary structure (local 2D & 3D)
 - tertiary structure (global 3D)
- DNA composed of four nucleotides or bases:
A, C, G, T

Structure

- RNA composed of four also: *A, C, G, U* (*T* transcribed as *U*)
- proteins are composed of amino acids

General form for amino acid

- $R - CH(NH_2) - COOH$
 - $COOH$ carboxy group
 - NH_2 is amino group
 - CH , C is called central carbon
 - R is a chemical group(called the residue chain). There are 20 different chain residues.

Connection of two amino acids

- **Peptide bond** is the reaction between carboxyl group of the first amino acid and amino group of the second amino acid.
- Using the peptide bond, long linear chains of amino acids (i.e. proteins) can be generated.
- The peptide bond itself is usually planar, which means that there is no free rotation around this bond.

DNA properties - base composition

- Some properties of long, naturally-occurring DNA molecules can be predicted accurately given only the base composition, usually expressed as either
 - % GC (the percent of all base pairs that are $G : C$), or
 - χ_{GC} (the mole fraction of all bases that are either G or C)

Melting temperature

- Example of zero order sequence properties
 - T_m , the melting temperature, defined as the temperature at which half of the DNA is single-stranded and half is double-stranded.

DNA structure - restriction maps

- **Restriction enzymes** cut DNA at specific sequences.
- A **restriction map** is a graphical description of the order and lengths of fragments that would be produced by the digestion of a DNA molecule with one or more restriction enzymes.

Types of Protein Structure

- Primary Structure \cong sequence of AAs.
- Secondary Structure \cong folding of parts of the AA chain.
- Tertiary Structure \cong real 3-D conformation.
- Quaternary Structure \cong 3-D arrangement of several domains.

Facts about proteins

- Proteins are chain molecules built from 20 different amino acids (*AA*).
- Linear polymer of amino acids, linked together by peptide bonds. Average size is around 200 amino acids, large proteins can reach over 1000 amino acids.

Facts about proteins

- To a large extent, cells are made of proteins. Proteins determine shape and structure of a cell, and serve as the main instruments of molecular recognition and catalysis.
- Proteins have a complex structure with four hierarchical levels.
- The function of a protein is determined by its 3-dimensional structure.

Facts about proteins

- Antibodies in the human immune system recognizes antigens by having a complementary surface to that of the antigen(**key and lock paradigm**)
- Enzymes are proteins, which, by forming a substrate to lower the reaction energy, facilitate or accelerate chemical reactions which would otherwise might never take place.
- The function of a protein is determined by its 3-dimensional structure.

Type of proteins

- Enzymes
- Collagen
- Membrane

Amino acids

- Each amino acid contains an amine group (NH_3) and a carboxy group (COOH) (shown in black in the diagram).
- nonpolar and hydrophobic. All other are hydrophilic (*water loving*)
- hydrophilic and polar.
- acidic (carboxy group in the side chain).
- basic (amine group in the side chain).
- 100 amino acids occur naturally, but only 20 are commonly used in protein synthesis.

Protein Structure and Function

- Tertiary Structure decides the biochemical function of a protein. → change in 3-D structure usually results in a loss of function.
- Knowing the 3-D structure of a protein is the key to understanding its function and to protein engineering.

Protein Structure and Function

- Finding the primary structure is easy.
- Finding the secondary structure is difficult.
- Finding the tertiary structure is very difficult.
- Finding the function of a protein is difficult.

Secondary Structure

- Secondary structure consists of fold patterns (rough structure). Most important: (i) α -helix (ii) β -strand
- In contrast to the tertiary structure, short range interactions in the AA chain are important for the secondary structure.

Protein folding

- The whole structure (primary to quaternary) is determined by the primary sequence and their physico-chemical interactions in the medium.
- Therefore, the folding of a protein is defined by the genetic material itself, as the three-dimensional structure with the minimal free energy.
- The structure of a protein determines its functionality.

Properties of α -helices

- Not only the AA sequence determines α -helix formation
- α -helices consist of 6-20 AAs in the primary structure
- Only 10% of the AA chain of a protein belong to α -helices. data consists of batches of 6 : : : 20 yes interspersed between longer batches of *no*.