

Stat 246, Lecture 3

Biological sequence analysis

The objects of our study

DNA, RNA and proteins: macromolecules which are unbranched polymers built up from smaller units.

DNA: units are the nucleotide residues A, C, G and T

RNA: units are the nucleotide residues A, C, G and U

Proteins: units are the amino acid residues A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W and Y.

To a considerable extent, the chemical properties of DNA, RNA and protein molecules are encoded in the **linear sequence** of these basic units: their primary structure.

The use of statistics to study linear sequences of biomolecular units

Can be descriptive, predictive or everything else in between.....**almost** business as usual.

Stochastic mechanisms should never be taken literally, but nevertheless can be amazingly useful.

Care is always needed: a model or method can break down at any time without notice.

Biological confirmation of predictions is almost always necessary.