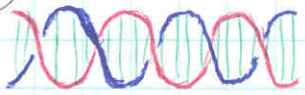
Protein folding: biochemistry background

6.885
L15.1
Nov. 3, 2004

DNA (Deoxyribonucleic acid)

- genetic information in every cell & many viruses
- essentially a string over alphabet {A, C, G, T}
- in fact, two such strings woven together into a double helix,
  always meeting in complimentary pairs:
  A-T, T-A, C-G, or G-C.



[Watson & Crick 1953]

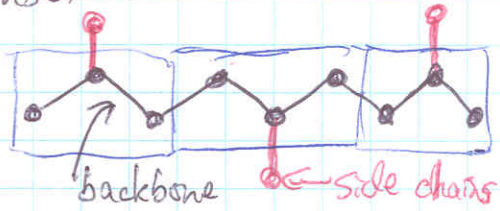
- A=adenine, C=cytosine, G=guanine, T=thymine
- replicate by unzipping the two strands and generating second copies from soup of nucleotides

RNA (Ribonucleic acid)

- similar but (usually) one strand & use U=uracil instead of T=thymine
- essentially short-term DNA for transfer out of cell

Protein

- fundamental building block of life:
  - form enzymes (speedup, chemical reactions) (catalyze)
  - form the cytoskeleton (scaffolding of a cell)
  - form antibodies (immune response)
  - etc.
- essentially a chain of amino acids: (20 kinds)
  (side-chains which hang off backbone)



- translated from (portions of) RNA/DNA by genetic code:
  - triplets of letters from {A, C, G, U} decode to 20 amino acids
  - AUG & GUG → "start"
  - UAG, UGA, & UAA → "stop"
  - AUG → Methionine (also) } uniques
  - UGG → Tryptophan } uniques
  - ...

[redacted] [Nirenberg, Matthaei, and Leder by 1965]

  - CGU, CGC, CGA, CGG, AGA, AGG → Arginine } most popular
  - UUA, UUG, CUU, CUC, CUA, CUG → Leucine } most popular
  - UCU, UCC, UCA, UCG, AGU, AGC → Serine } most popular