Protein evolution and sequence space
Proteins are diverse
Protein evolution and sequence space

- Evolution — can we measure it?
- Forms of evolution
- Protein Evolution — mechanism
  - Sequence space
  - Protein design
Evolution
Evolution

Natural selection

Artificial selection

Genetic drift
Drug resistance

(A) wt protein

(B) wt protein

(C) mutant protein

ligand A' binding

B-B' (drug) binding

A-A' binding

Normal biological action

No biological action

Normal biological action
Evolutionary distance: phylogenetic trees

Tree structure

Nodes are some classification of biological species

Edges represent relation and ancestry

Branched or unbranched
Phylogenetic tree of life

All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct.
Protein evolution can also be traced.

The human kinome
Protein evolutionary relationships

- Phylogenetic trees describe similarities among biological entities
- Implications: Studies of one biological system can be used to infer the behavior of another

Hamming distance for sequence identity

- $D(1,2) = 3$
- $D(2,3) = 7$
- $D(1,3) = 10$

**Sequence**

1. LVRKVAEENGRSVNS
2. LVAKVAEENKRSTNS
3. MVAYVGEDHKRCTIS
Molecular clock

Mutation rate is linear in time
Mechanisms of evolution

DNA → RNA → Protein
Mechanisms of evolution

- Nucleotide insertion
Protein evolutionary relationships

Homologs
- Orthologs
- Xenologs
- Paralogs

Ancestor
- Orthologs
- Paralogs
- New function
Protein structures are tolerant of some sequence changes.

Sequences

Sequences that fold

Unique folds

Sequence Space

Seq(N) = 20^N

Conformational Space

Conf(N) = 3^{2N}
Simplified models help understand the conformational/sequence space

HP lattice model:
- Polar residues (P)
- Hydrophobic (H)
- H-H interactions stabilize the fold
Protein structures are robust to single mutations

- Single mutations
  - 30-50% deleterious
  - 50-70% neutral
  - 0.01-0.5% beneficial

- Folding stability is intricately related to **designability**
Knowledge-based design in the HP lattice model
Protein Design

Directed evolution

Knowledge based design
Design in real systems

Principles for designing ideal protein structures

Nobuyasu Koga, Rie Tatsumi-Koga, Gaohua Liu, Rong Xiao, Thomas B. Acton, Gaetano T. Montelione & David Baker


Knowledge + Directed evolution
Other foldameric polymers

- Protein mimetics, foldamers, peptoids

Zuckermann and Dill