Lecture 15

Protein-Fold Classification

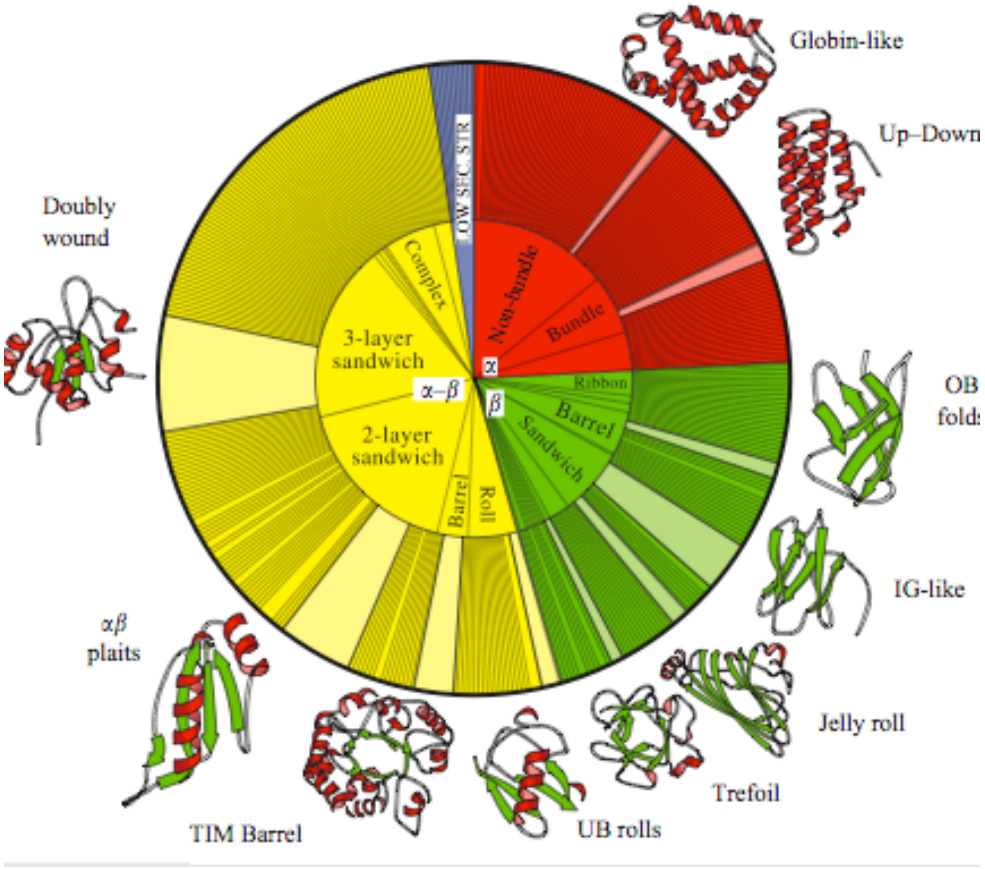
And “The 80% : 20% Law”

Early Thoughts

- Early 1970's
- Late 1970's
  - "Standard Design"
  - Simple Physical Restrictions exist

"The Intermediate Level" is the "folding pattern"

- Alpha and Beta region dependent
- Allows for classification



Classification

- Common Classifications:
  - Dali/FSSP
  - CATH (Class-Architecture-Topology-Homology)
  - SCOP (Structural Classification of Proteins)

- Common Classification Schemes

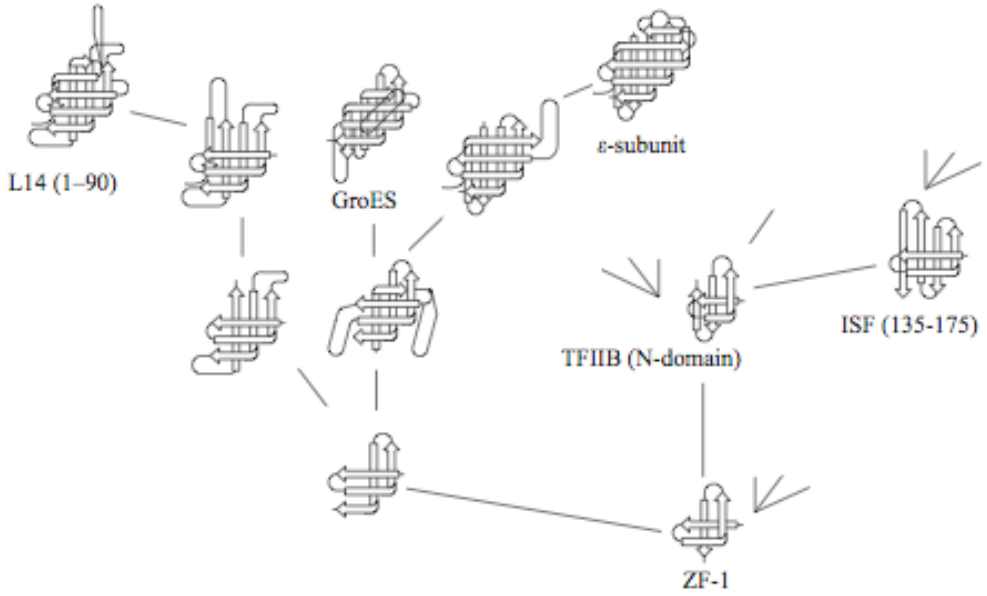
  -Classes

  -Architectures

  -Topologies

Benefits of Classification

- Systematize Structure Studies
- Prediction of Structures
  - Example: β-Prisms



Protein Structure Evolution

- (1) Microscopic evolution of proteins
  - Example: Llama hemoglobin

- (2) Macroscopic evolution of proteins
  - Not observed

80% : 20% Law

- 2 Driving questions
  - What is the physical reason for the simplicity and regularity of typical folding patterns?
  - Why are the same folding patterns shared by completely different proteins?

Regularity of Folding Patterns

- Stability Concerns
  - Simple question: Why the layered structure of globular proteins?
    - Stability due to hiding the Hydrophobic core
    - Irregular regions around the globule
  - Instability concerns
    - The mixed α-β layer instability
    - 4+ Layer instability
      - These states do occur from time to time, but are extremely rare
        - » Example: a chain consisting mostly of hydrophobic amino acids could pack into a large stable globule
        - » This case is very sequence specific

Sequence Specificity

- Sequences of primary structures of proteins
  - Globular protein sequences are “quasi-random”
    - Structural/Energy defects
    - Defect free structures

Folding Patterns

- Volume Covering Problem
  - Secondary Structures and Loops
    - Conditions:
      - » Loops do not intercross
      - » And loops connect anti-parallel
      - » Loop crossing can occur but are very sequence specific
  - The main point: Structures are similar due to constraints that cause “defect” folds to be highly sequence dependent
  - This can be summarized as: occurrence=exp(-E/KTc)

Summary

(1) “Popular” folding patterns look so regular because the framework has physical restrictions (hydrophobic effects, etc..)

(2) The number of “standard” stable folding patterns is not large

(3) “Defects” aren’t not allowed, just rare due to specificity of the sequences that generate them

(4) “Multitude Principle”-> The more sequences fit the given architecture without disturbing its stability, the higher the occurrence of this architectures in native proteins