The Protein Folding Problem

Matt Holford

Department of Biology and Biomedical Science
Yale University

CBB 752 Discussion Section 4b
February 18, 2010

Introduction

- Three questions
  - What is the folding code?
  - Can we predict folding structure from amino acid sequence?
  - What is the folding mechanism?
- This once seemed a daunting task...
- But today...
  - Foldable proteins are designed and used
  - Structures of small proteins can be predicted by computers
  - Provable theory: Large global optimization solved by being broken into smaller local optimizations

Anfinsen's hypothesis

- What balance of forces encodes native protein structures?
- Anfinsen's Thermodynamic Hypothesis
  - Native structure = thermodynamically stable structure
  - Depends only upon amino acid sequence and conditions of solution
  - **NOT** on folding route
  - This makes *in vitro* folding a viable research approach
  - The problem is reduced to one of physical chemistry

Old and New Views of Folding

- The old view:
  - Folding is the sum of many small interactions
    - e.g. H-bonds, ion pairs, Van der Waals forces, hydrophobic interactions
  - Primary sequence $\rightarrow$ secondary sequence $\rightarrow$ tertiary sequence
- The new view:
  - Informed by statistical mechanical modeling
  - Hydrophobic interactions dominate
  - Secondary structure $\leftrightarrow$ tertiary structure

Predicting Protein Structure From Sequence

- Rise of protein databases, eg PDB
- CASP (Critical Assessment of Techniques for Protein Structure Prediction)
  - Biennial, community-wide blind test of unsolved structures
- Small proteins can now be accurately predicted to 2-6 $\mathring{A}$
- Best methods combine bioinformatics techniques with physics

Folding Mechanism I

- How do proteins fold so fast?
- The search has led to many important advances
  - Experimental techniques
  - Computational techniques
  - Protein theory
- Plaxco, Simons and Baker (PSB) discover a universal feature of protein folding kinetics
  - Folding speed correlated with topology of native structure
    - Fast folding $\rightarrow$ mostly local, e.g. helices, turns
    - Slow folding $\rightarrow$ mostly non-local, e.g. β-sheets

Folding Mechanism II

- Proteins have funnel-shaped energy landscapes
  - Many high-energy states, few low-energy states
  - “Cartoonization” which shows folding as a pathway from disorder to order
    - Chain entropy increases as structures become less stable
    - There are multiple folding routes
    - Different routes dominate under different folding conditions
  - ZA (Zipping and Assembly) hypothesis
    - Global optimization found by divide-and-conquer
    - Avoids search of all of conformational space
    - Process takes only small-conformational-entropy-loss steps
    - Many short, parallel steps in the beginning
    - Fewer, more sequential steps at the end

Pure Physics-Based Modeling

- Better than mixed-bioinformatics methods for:
  - Predicting conformational changes, such as induced fit
  - Understanding action mechanisms, such as motion, folding and catalysis
  - Understanding how proteins respond to external factors: e.g. solvents, ions, etc.
  - Designing synthetic proteins with noncanonical amino acids or non-biological backbones
- **BUT...**
  - Very expensive computationally
  - Semiemperical atomic physical force-fields not entirely reliable