H-P model of protein folding energetics

6.885
L17.6
Nov. 10, 2004

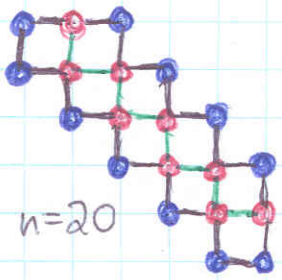
- simple model capturing an often prominant feature of energy: hydrophobic bonds
- protein = chain of H or P nodes
  - H → hydrophobic
  - P → hydrophilic/polar
- folding = embedding on square grid in 2D or 3D
  (also several variations — equilateral △ grid in 2D, cannonball packing in 3D)
- score of folding = # pairs of H nodes adjacent on lattice
  (usually ignoring adjacencies from chain)
- optimal folding = maximum score (avoid surrounding water)
- introduced by Ken Dill (theoretical biophysist) in 1990

Finding optimal folding is NP-hard

- in 3D [Berger & Leighton 1998]
- in 2D [Crescenzi, Goldman, Papadimitriou, Piccolboni, Yannakakis 1998]
- various O(1)-approximation algorithms
  [Hart & Istrail 1996; Mauri et al. 1999; Newman 2002; Agarwala, Batzoglou, Dancik, Decatur, Farach, Hannenhalli, Muthukrishnan, & Skiena 1997]

Unique optimal foldings [Aichholzer, Bremner, Demaine, Meijer, Sacristán, Soss 2003]

- exist in square grid in 2D
  - for all $n \equiv 0 \pmod 2$ for closed chains
  - for all $n \equiv 0 \pmod 4$ for open chains


n=20

OPEN: design protein with unique optimal folding with approximately a desired shape

OPEN: (1-ε)-approximation for optimal folding?

OPEN: optimal folding NP-hard among uniquely optimally foldable? other grids than square?