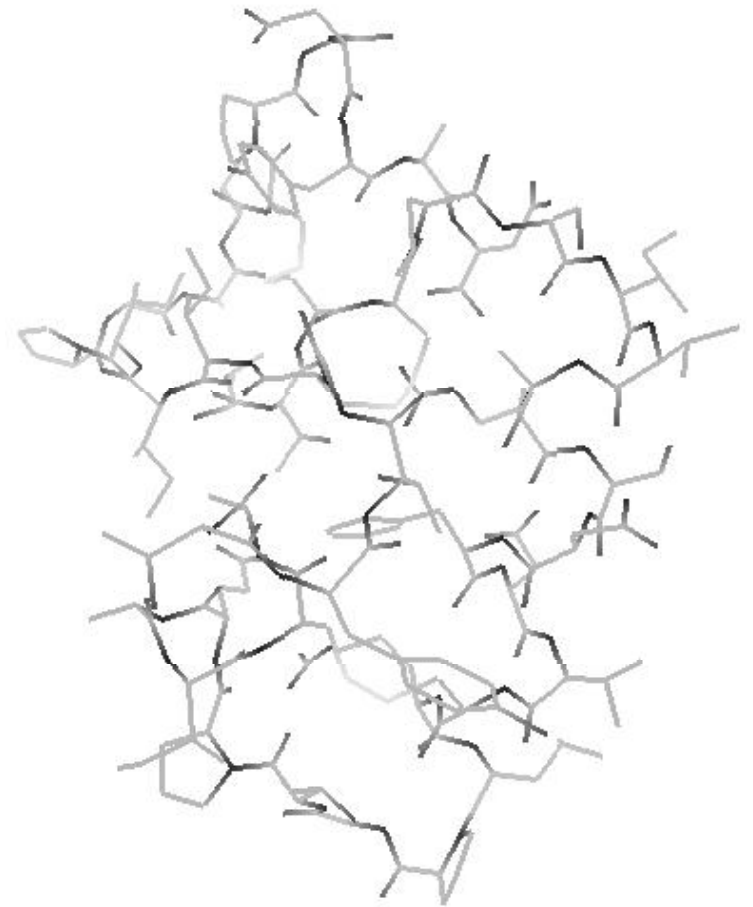
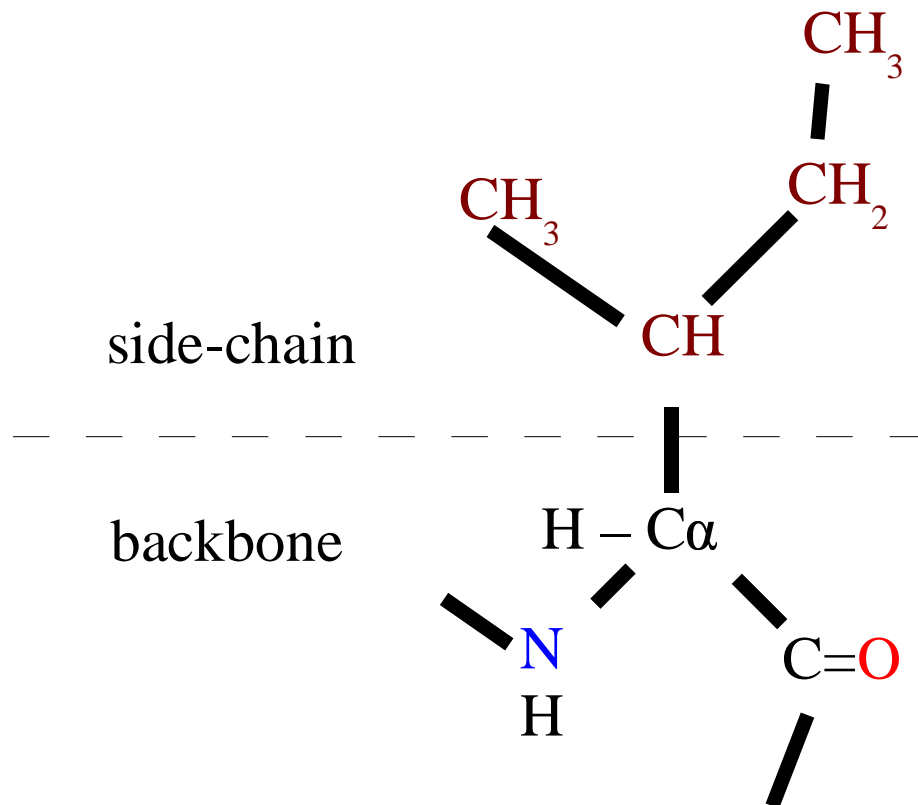


Parallel Self-Consistent Mean Field

Karl Gutwin
18.337 Spring 2005

90-second Protein Structure

- Residues (alanine, proline, tryptophan, **isoleucine**)
- Backbone vs. side-chain



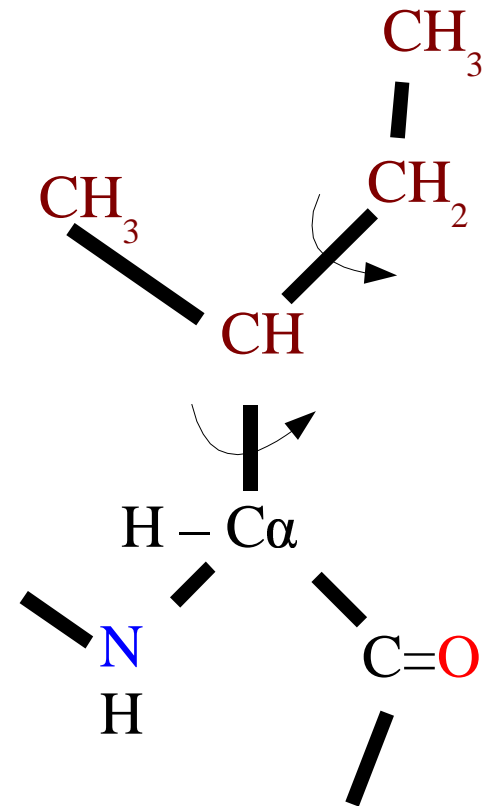
The problem

- Protein repacking
 - implications in protein design
 - Method:
 - Take a known backbone (N,C α ,C,O) and C α – C β vector
 - Find the best conformation of a residue at a given position
 - best == lowest energy, least amount of 'clashes'

Continuous space along >300 dimensions!

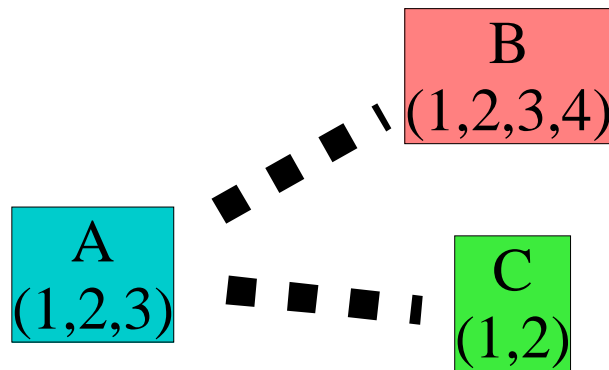
The Rotamer Approximation

- Certain conformations are more likely than the rest
- From infinite combinations to 1-25 (average 10)
- Still, to enumerate all in a 150-residue protein:
 $\sim 10^{150}$ possibilities!



Self-consistent Mean Field

- When calculating the total cost of a given rotamer, weight the contribution of the interacting rotamer's energetic cost by the probability that it will be in the final structure



$$E(i, k) = \sum_{j=1, j \neq i}^N \sum_{l=1}^{K_j} CM(j, l) U(x_{ikC}, x_{jlC})$$

Probabilities:

	1	2	3	4
A	0.1	0.4	0.5	
B	0.2	0.6	0.1	0.1
C	0.8	0.2		

Interaction energies for A1:

	1	2	3	4
B	3.5	1.6	7.2	7.8
C	1.4	4.4		

Self-consistent Mean Field

- Why do we care about energetic cost?
It allows us to calculate probabilities!

$$CM(i, k) = \frac{e^{-E \frac{(i, k)}{RT}}}{\sum_{l=1}^{K_i} e^{-E \frac{(i, l)}{RT}}}$$

Boltzmann Distribution

- Probabilities  Energies

My problem

- Assumption: Residues are independent
 - any issues with joint probabilities must be resolved implicitly
- Approach: Explicit dependence
 - combine residues into residue groups
 - rotamers into rotamer sets
 - Multiplies possibilities! 100 residues, 10 rots/res:
 $d = 1 \quad \text{sets} = 100 * 10 = 1000$
 $d = 2 \quad \text{sets} = 50 * 100 = 5000$
 $d = 3 \quad \text{sets} = 33 * 1000 = 33000$
 $d = 4 \quad \text{sets} = 25 * 10000 = 250000$

Parallel perspective

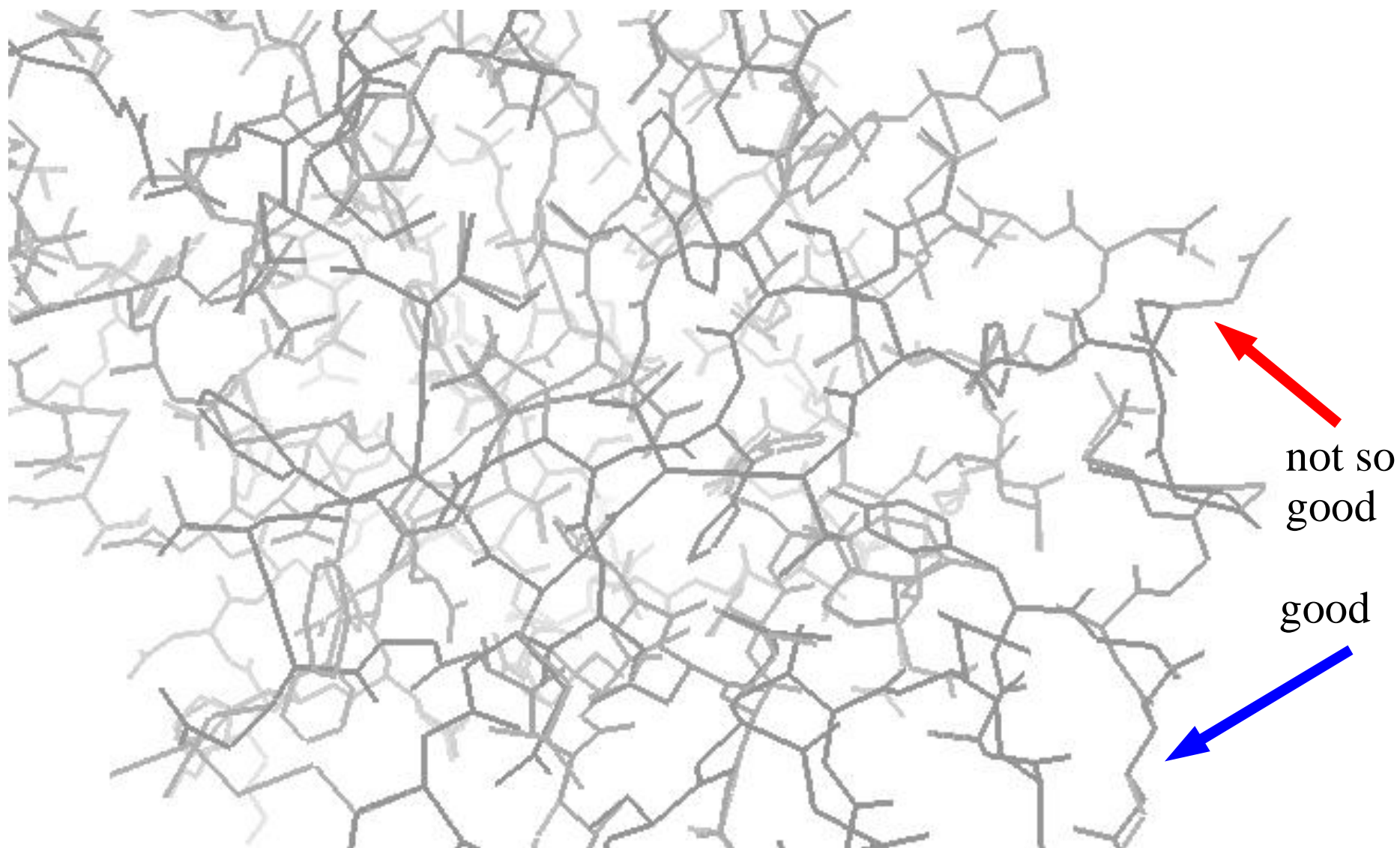
- Basic computation itself is trivial -- matrix-vector multiply (on a very large matrix!)
- Memory issues
 - Normally precalculate and save pairwise potentials
 - 250000^2 is too big to fit in RAM, even spread out
- Load balancing issues
 - want to evenly distribute rotamer sets – basic unit
 - grouped by residue groups, varying number of sets per group!

Results

- Very slow! Bottlenecks abound, much more optimization is possible
- Was unable to determine quantitative differences between structures
 - normal method: RMSD (root mean square deviation) between observed and calculated structures

blue: calculated
green: predicted

Results



green: observed
blue: calculated

Results

