6.581J / 20.482J

LECTURE 3: DISCRETE CONFORMATIONAL SEARCH

REVIEW:

A.) COORDINATE REPRESENTATION – PROTEIN STRUCTURE
- 1.) CARTESIAN – ABSOLUTE
- 2.) INTERNAL – RELATIVE
  - -bond lengths } "hard"
  - -bond angles }
  - -torsional – "soft"

$$\vec{X}^{3N} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{3N-2} \\ X_{3N-1} \\ X_{3N} \end{bmatrix} \begin{matrix} \} \text{1st atom} \\ \\ \\ \\ \} N^{\text{th}} \text{ atom} \\ \\ \end{matrix}$$

B) ENERGY FUNCTION – PHYSICS BASED
- level of "atom"
- $\vec{X}^{3N} \rightarrow E(\vec{X}^{3N}) = U_{bonds} + U_{bond\ angles} + U_{improper\ dihedrals} + U_{torsions} + U_{vdw} + U_{elec}$

$U_{covalent}$ $O(N)$; $U_{non\text{-}covalent}$ $O(N^2)$

WHAT CAN WE DO WITH AN ENERGY FUNCTION?

* Fold a protein
- Find the final, folded shape
- Understand how a protein finds its final shape
- Average folding time for a protein: ~1ms – 1s typical
- Most detailed simulations: 1 cpu day ⇔ 1ns
  1ms ⇒ 10^6 CPU days ⇒ 2,740 years

Two Studies of Historical Note:

1) Levitt & Warshel Nature 253 694-698 (1975).
- Simplified model of a small protein
- BPTI 58 residues
- Every center represents an amino acid
  - ran detailed atom-level simulations on pairs of amino acids to produce higher level model
- Accelerated simulation methodology

RESULT
Authors claim close to "native" structure
Subsequent analysis question this

2) Duan & Kollman Science 282: 740-744 (1998).
1 μs simulation of 36-residue protein (villin headpiece subdomain)
- model: all-atoms plus ~3,000 water molecules
- took ~4 months on a 256-processor Cray T3E

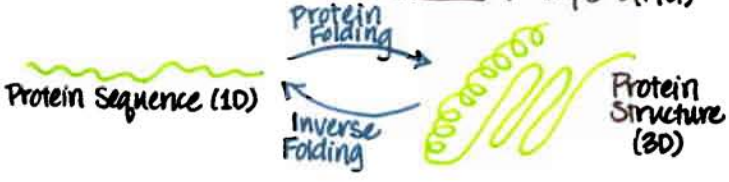
RESULT
Marginally stable state "close" to native state adopted at ~150ns, persisted briefly, then dissolved

FUNDAMENTAL PROBLEM

1. What result does the model produce?
   Interplay: Simulation Methodology ⇔ Model
   Pathway ⇔ Final Structure
2. Can't relate problems to individual terms in the energy function

Two Visionary Statements:

C.O. Pabo Nature 301: 200 (1983)
E. Drexler Proc Natl Acad Sci USA 78: 5275-8 (1981)



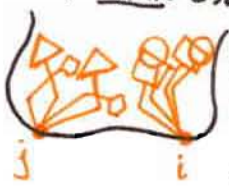
PROBLEM STATEMENT:

① Fixed backbone topology
② For each side chain attachment site
- which of the 20 amino acids belongs there?
- what is its conformation?

Our energy function
- adopts relaxed local (covalent) conformations
- vdW potential: efficient packing of space
- elec: complementary charge patterns

2 simplifications
1. fix all bond lengths & bond angles
2. discretize side chain torsions

The Dead-End Elimination Algorithm (DEE)
Desmet et.al. Nature 356: 539-42 (1992).



At each of N positions select from m_i choices for the amino acid and its conformation

Solve for $\vec{M} = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_N \end{bmatrix}$ that minimizes $E(\vec{X}^{3N})$ over all possible $\{\vec{M}\}$

Imagine $m_i = m \Rightarrow$ space is m^N

NOTE: OUR ENERGY FUNCTION HAS SIMPLE FORM

$$E_{\vec{M}} = \sum_{i=1}^{N} E^{self}_{M_i} + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} E^{pair}_{M_i,M_j} + \text{const.}$$

terms between sc & bb ; sc-sc

Looking at just one position

$E^{self}_{M_i^s} + \sum_{\substack{j=1 \\ j \neq i}}^{N} E^{pair}_{M_i^s,M_j}$ is the contribution if I choose m_i^s at position i

But I can bound the contribution

Lower Bound: $E^{self}_{M_i^s} + \sum_{\substack{j=1 \\ j \neq i}}^{N} \min_u E^{pair}_{M_i^s,m_j^u}$

Upper Bound: $\max_u$

DEE Singles Criterion

If $E^{self}_{m_i^t} + \sum_{\substack{j=1 \\ j \neq i}}^{N} \min_u \left(E^{pair}_{m_i^t,m_j^u}\right) > E^{self}_{m_i^s} + \sum_{\substack{j=1 \\ j \neq i}}^{N} \max_u \left(E^{pair}_{m_i^s,m_j^u}\right)$

For some $t \neq s$, then m_i^s can not be in the global optimum

This method is provably correct, but can't prove running time

Approach:

1. Eliminate iteratively singles until no more possible
2. Develop higher-order eliminations (pairs) ↑
3. When all eliminations done, enumerate remaining space

Application:

Insulin: 76 positions: 2.7×10^{76} conformations

Iterative approach (9 iterations)

$2.7 \times 10^{76} \rightarrow 7200 \rightarrow$ search by enumeration

93% of buried positions "correct"