Protein Structure Prediction

Amino acid sequence $\rightarrow$ 3D conformation

$\alpha$-helix $\rightarrow$ $\beta$-sheet

Testing structure

$P = (p_1, p_2, \ldots, p_N)$

$\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N)$

$\hat{p}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$

$\hat{p}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$

RMSD: Root mean squared distance

$\hat{p}$ is predicted

$\hat{p}$ is predicted
\[ R_{\text{MSD}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \text{dist}(p_i, \hat{p}_i)^2} \]

\[ \text{dist}(p_i, \hat{p}_i) = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2} \]

\[ p_i = (x_i, y_i, z_i) \quad \hat{p}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i) \]

\[ R_{\text{MSD}} \leq 2\AA^0 \]

\[ 2 \times 10^{-10} \text{ m} \]

\( \hat{\lambda} = 1 \text{ Å} \)

\( 1 \text{ Ångstrom} \)

2 ways to determine the true structure:

1. X-ray crystallography
2. NMR Spectroscopy

PDB

170k known structures

1 million putative


- **Lattice model**

  - Simplified molecular dynamics model
  - \( H \): hydrophobic amino acids
  - \( P \): polar amino acids

  - Hydrophobic collapse

  - \( \Sigma = \sum \{ H, P \} \)

  - Each grid cell has

  - 2D lattice (4 neighbors)
  - 3D lattice (6 neighbors)

- Known structures: over 1 million protein sequences.
Energy (Configuration) = - (\# \delta_{\text{H-H bonds that are non-adjacent}})

Sequence: H P P H H P P H P

Energy = -2

Each grid cell has 6 neighbors in 3D lattice.

Self-avoiding walk on the lattice.
Given: He-sequence 3D lattice

Markov Chain Monte Carlo simulations

state to state transitions

All possible conformations

Energy landscape

Local move set

1. End move
   Move one of the ends to an available grid cell

2. Corner move
Crank-shaft move

1, 0, 0 are far less restricted in terms of configuration

Point move: this allows all possible configuration

\[ 2 \]
try very large number of steps/moves
look $\rightarrow$ 1 million or more.
report the lower energy configuration

Even in 2D HT model, finding the lower energy configuration is NP-hard.

2D lattice

6 internal H

1 terminal H

Any internal H can have at most 2 other H contacts

Any terminal H can have at most 3
\[ \left\lfloor \frac{6 \times 2 + 1 \times 3}{2} \right\rfloor = \left\lfloor \frac{12 + 3}{2} \right\rfloor = \left\lfloor \frac{15}{2} \right\rfloor = 8 \]

Upper bound

In an integer lattice \((2D/2D)\) only even \& odd

\(H\) 's can contribute to the energy.

\[ \frac{2 \cdot 10}{2} = 10 \]

\[ -3 \]

1. Find a division point such that there is an equal number of even & odd pairs on either side.
\[ N_L(\varepsilon) = \text{# of even } H's \text{ on left} \]

\[
\begin{array}{c}
\text{even} \\
N_L(\varepsilon)
\end{array} \quad \text{or} \quad \begin{array}{c}
\text{odd} \\
N_R(\text{odd})
\end{array}
\]

\[
\begin{array}{c}
5 \\
N_L(\text{odd})
\end{array} \quad \begin{array}{c}
7 \\
N_R(\varepsilon)
\end{array}
\]

\[
\max \left( \begin{array}{c}
3 \\
\min( N_L(\varepsilon), N_R(\text{odd}) )
\end{array} \right) = 3
\]

\[
\min \left( \begin{array}{c}
5 \\
\min( N_L(\text{odd}), N_R(\varepsilon) )
\end{array} \right) = 5
\]

\[
\max \left( \begin{array}{c}
3 \quad 5
\end{array} \right) = 5
\]

\[
\begin{array}{c}
1 \quad 2 \quad 3 \quad 9 \\
10 \quad 11 \\
12 \quad 13 \quad 14 \quad 15 \quad 16
\end{array}
\]

- \( N_L(\varepsilon) = 3 \)
- \( N_L(\emptyset) = 0 \)
- \( N_R(\varepsilon) = 2 \)
- \( N_R(\emptyset) = 2 \)

\[
2 = \max \left( \begin{array}{c}
\min(3,2) \\
\min(0,1)
\end{array} \right) = 2
\]

\[
\begin{array}{c}
\varepsilon = 3 \\
\emptyset = 2
\end{array}
\]

\[
\begin{array}{c}
N_L(\varepsilon) = 3 \\
N_L(\emptyset) = 0 \\
N_R(\varepsilon) = 2 \\
N_R(\emptyset) = 2
\end{array}
\]