# Exam II

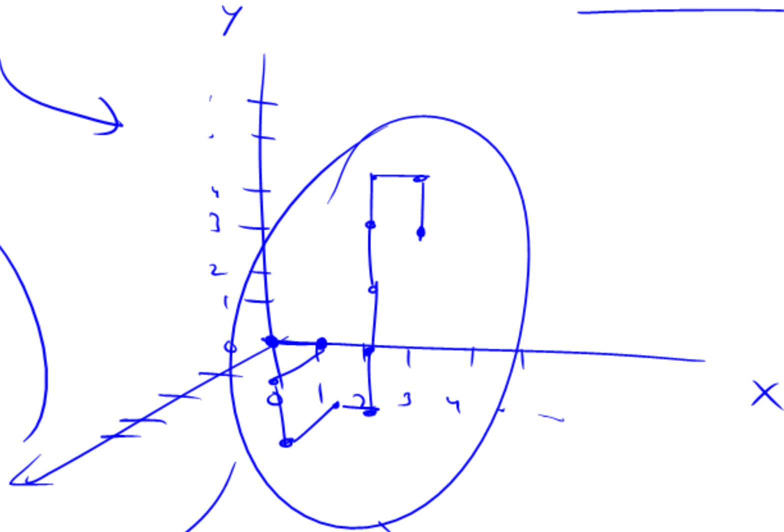
- Suffix Trees
- Suffix Arrays
- Genome Scale Alignment
- Phylogenetic Trees
  - Parsimony → DP : Cost matrix
  - Distance-based → UPGMA / NJ
  - Maximum Likelihood
- Genome Rearrangement → signed ←, unsigned ←
- Protein Structure Alignment
  - DP + Superposition
  - Double DP ←
  - Geometric Hashing
  - Distance Matrix
- Structure Prediction
  - HP model

# Lattice Model : HP : hydrophobic / Polar residue

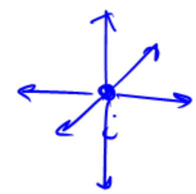
$Z_2$

2D lattice

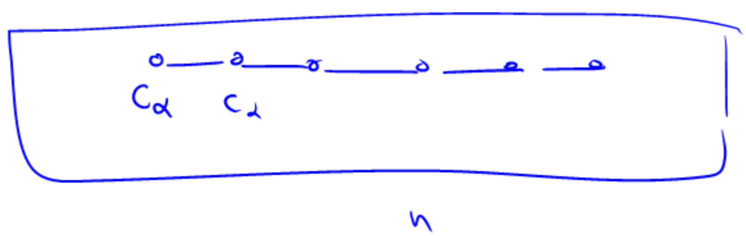
($Z_3$: 3D lattice)




"hydrophobic core"



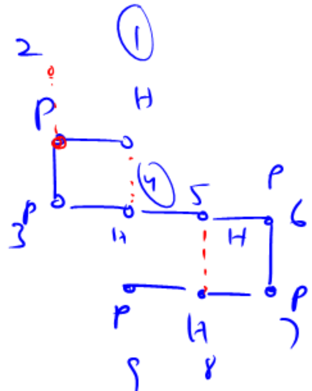
2d : # of choice




unfolded protein

Energy (conformation) = − ("Hydrophobic contact")

Self-avoiding walk



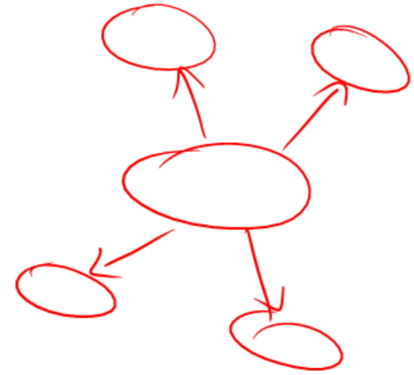

1 2 3 4 5 6 7 8 9
H P P H H P P H P

H − H non consecutive in sequence

Energy = − 2

# 1) Simulations (Monte Carlo)

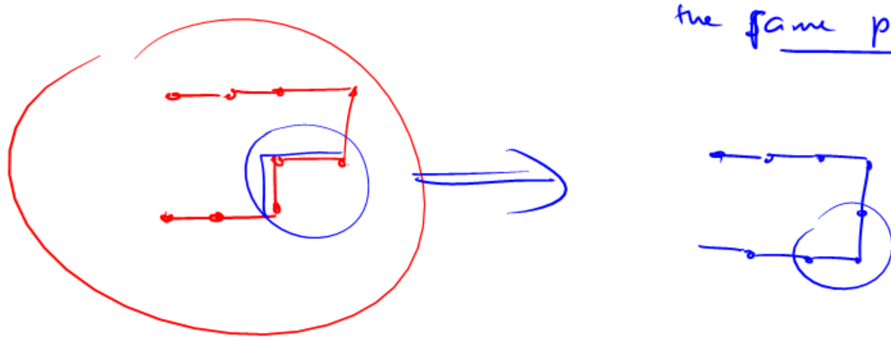
## Set of neighbor move



restricted sampling

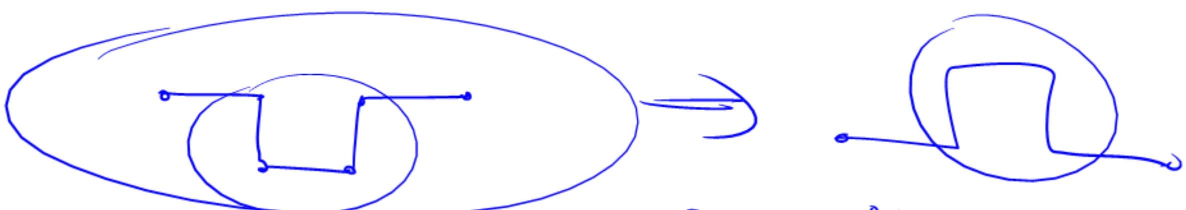
a) End move

b) Corner move : rotation/flip in the same plane



(XY
YZ
XZ
planes
in
3D)

c) Crank-shaft move : rotate $90^\circ$ (3D)

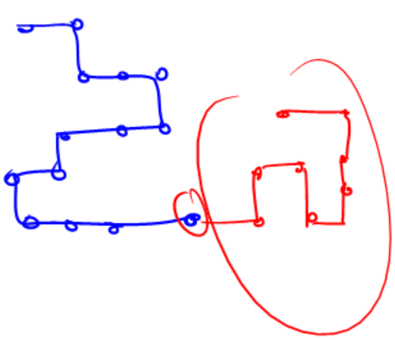


2D ($180^\circ$)

+

pivot
———
complete search

d) pivot move : choose a pivot point $i$
divide the conformation into 2
rotate one side in
any of the
$2d-1$ possible
$d=3 \Rightarrow 5$
neighboring
cells

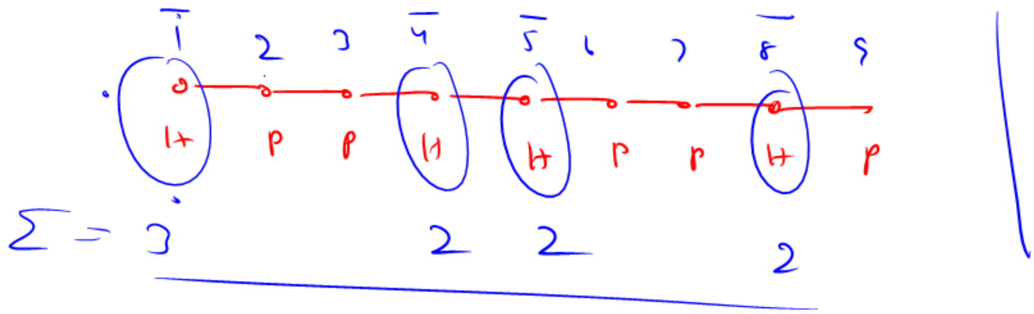


make sure self-avoiding

2) Deterministic approach

1 2 3 4 5 6 7 8 9

H P P H H P P H P



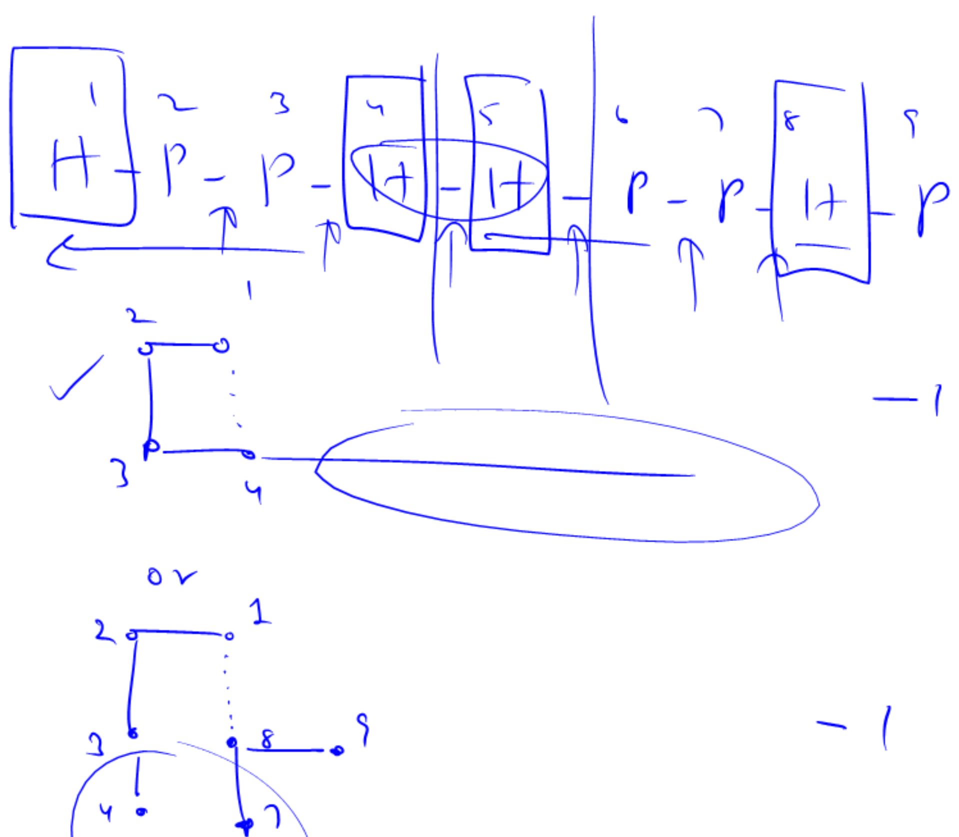

2D

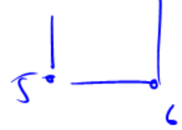
$$\text{optimal lower bound} = \left\lceil \frac{9}{2} \right\rceil = 5 = \left( \#\ \text{internal } H + \#\ \text{terminal } H \right) / 2$$

Theorem: In an integer lattice, only possible H-H non-consecutive contacts are between even & odd polarity cells

Pivot

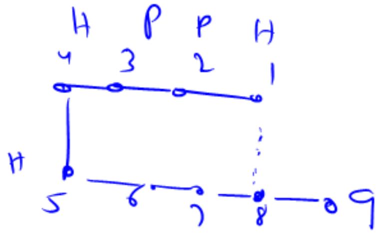
P23 :





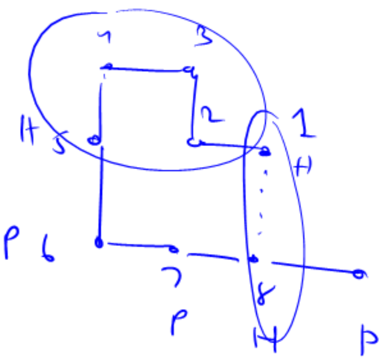
Pivot P34 : same as P23

Pivot P45 :



−1

P52 :



−1

best : −1

approx ratio: 4 (asymptotically $n \to \infty$)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P | H | P | H | P | P | P | H | P | P | H | P | H | H | P | H |

Even | odd | even

max { # of even H ← min $N_E^L$ : 3 | $N_E^R = 2$

# of odd H ← min $N_O^L$ : 0 | $N_O^R = 2$

left | right

1) define the even & odd maximal block

2) try the pivot point that are between blocks

3) balance the # of odds with the # of evens on either side

$$\text{Score}(L, R) = \max \begin{cases} \min(N_E^L, N_O^R) \\ \min(N_O^L, N_E^R) \end{cases}$$

Choose the max score c/r $(L, R)$