# 4 Protein structure prediction

## Proteins

Translated from DNA to mRNA to sequance of amino acids

| DNA | | mRNA |
|---|---|---|
| A | | A |
| C | → | C |
| G | | G |
| T | | U |

translated → sequence of amino acids

↓ folds to its tertiary structure ⇒ function

it is beliened that the tertiary structure is encoded in the primary structure.

Finding structure

- X-ray crystallography
  - expensive
  - Some Proteins impossible to crystalize
- Simulations
- Modelling
  - Free energy

# HP model

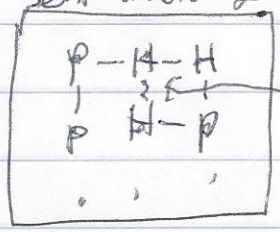
2D grid model.

Minimize ~~free~~ energy

$S \in \{H, P\}^*$ where H are hydrophobic (water hating) and P are polar (water loving).

Each aminoacid is assigned a group H or P.

Self-avoiding walk




HPHHPP

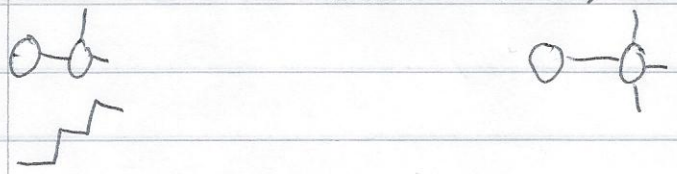
point scored

Idea is to have water-hating protected inside water loving amino acids.

Is it a hard problem?

$2^{n-2} < 2,638^n < 3^{n-2}$

↑ when n is large

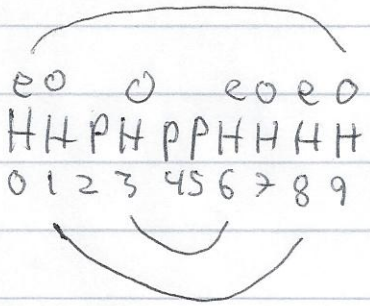


The optimal solution problem is NPC

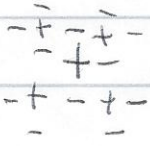
## 1/4 Approximation algorithm

How it works:

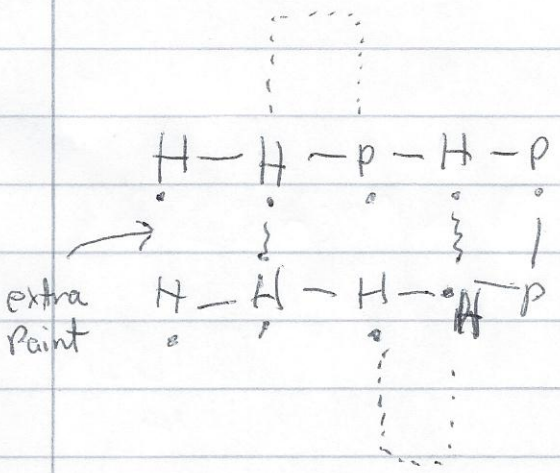
1) Match evens from left with odds from right
2) Match odds from left with evens from right
3) Pick the largest matching




why evens and odds?



Evens can only match odds.




HP score $\geq$ ~~size~~ size of matching

why $\frac{1}{4}$ ?

$\exists S_1, S_2 \quad S = S_1 \cup S_2$ and

Find $S[1, \ldots, i]$ such $i$ is minimum and holds half evens or odds.

1) $\text{Evens}(S_1) \geq \frac{1}{2}\text{Evens}(S)$ and $\text{Odds}(S_2) \geq \frac{1}{2}\text{Odds}(S)$
2) or $\text{odds}(S_1) \geq \frac{1}{2}\text{odds}(S)$ and $\text{evens}(S_2) \geq \frac{1}{2}\text{Evens}(S)$

Since we match in 1) all evens in $S_1$ against odds in $S_2$, one of them will limit the matching, namely the smallest one. Same for 2)

$\Rightarrow$ Size of matching $\geq \frac{1}{2}\min(\text{evens}(S), \text{odds}(S))$

Upper bound on OPT(S)

An $h$ can form at most two bonds. odd $h$ with even $h$.

$\text{OPT}(S) \leq 2 \cdot [\min(\text{evens}(S), \text{odds}(S))$

Put together

$$\begin{aligned} \text{Hp-score} &\geq \frac{1}{2}\min(\text{evens}(S), \text{odds}(S)) \\ &\geq \frac{1}{2}\left(\frac{1}{2}\text{OPT}(S)\right) \\ &\geq \frac{1}{4}\text{OPT}(S) \end{aligned}$$