

Random Sequence of Proteins

ie YFRYW
YFFVF
FFRYW
...

Steps

- 1 Calculate freq. of each amino acid
- 2 Global Alignment of pairs of Random proteins with 8 Scoring
Avg alignment is of order $\max(\log) \text{length}$ (time complexity is $m \cdot n$)
Avg length of local alignment optimum is $\max \{\log n, \log m\}$

Phase Transition

$$O(N) \rightarrow O(\log N)$$

Kawlin & Artshul statistic : 5 axioms

- 1 Positive score in matrix (min. 1)
- 2 Expected score is negative

$$S = \begin{vmatrix} s_{1,1} & s_{1,2} & \dots \\ s_{2,1} & & \end{vmatrix} \sum_{i,j=1}^k s_{i,j} \cdot p_i \cdot p_j < 0$$

→ pulling from overall dist. of Amino Acids

- 3 Letters independently, identically distributed (Proteins are Random sequences of Amino Acids)

- 4 Sequences infinitely long

- 5 No gaps

Take Unmapped alignments



→ compute the distribution of amino acid pairs aligned in

the max scoring segments

$$q_{i,j} = p_i p_j e^{-\lambda S_{i,j}}$$

all prob. uniform

Row score

Score of aligning i, j in the max scoring segment

The lady tasting tea : "Design of Experiments"

Sir Fisher introduced the concept of Null hypothesis

"Never proved or established but is possibly disproved in the course of the experimentation"

→ Fisher gives her 8 randomly selected cups of tea (4 milk first, 4 tea first) → $\binom{8}{4}$ assignments = 70

→ H_0 : No ability to distinguish

→ she needs to be first

Of the 70, here is distribution of ^{Correct tea} answers under H_0

0 answers Correct: $1/70$

1 Success: there are 4 correct cups (for tea first) to guess $\binom{4}{1} \neq 1$ places to put the 4 wrong cups

2 Success: $\binom{4}{2} \binom{4}{2} = 36$

3 Success: $\binom{4}{3} \binom{4}{1} = 16$

4 Success: 1

$1/70 \leq 0.05$, at this level, we accept the H_1