McCaskill: Efficient Base Pair Probabilities

Idea: Compute $p_{kl} := \Pr[(k,l)|S]$ recursively (DP!),
recurse from long base pairs (outside) to small ones (inside)

1) simple case (external base pair)

Definition (Probability of external base pair)

$p^E_{kl} := \Pr[\mathcal{P}^E_{kl}]$,
where $\mathcal{P}^E_{kl} := \{P \mid P \in \mathcal{P}, (k,l) \text{ is external base pair in } P\}$,
where (k,l) is *external base pair in P* iff
$(k,l) \in P$ and $\nexists (i,j) \in P : i < k < l < j$.

$$Z_{\mathcal{P}^E_{kl}} = Q_{1k-1} Q^b_{kl} Q_{l+1n}$$

$$p^E_{kl} = \frac{Z_{\mathcal{P}^E_{kl}}}{Z} = \frac{Q_{1k-1} Q^b_{kl} Q_{l+1n}}{Q_{1n}}$$

McCaskill: Efficient Base Pair Probabilities

2) general case

a) (k, l) is external base pair

b) (k, l) limits stacking, bulge, or interior loop closed by (i, j)

c) (k, l) inner base pair of multiloop closed by (i, j)

McCaskill: Efficient Base Pair Probabilities

2) general case

a) (k,l) is external base pair ✓

b) (k,l) limits stacking, bulge, or interior loop closed by (i,j)

$$\begin{aligned} p^{SBI}_{kl}(i,j) &:= p_{ij} \Pr[\text{loop } i,j,k,l|(i,j)] \\ &= p_{ij} \frac{Z_{\{P \in \mathcal{P}_{ij} | P \text{ has loop } i,j,k,l\}}}{Z_{\mathcal{P}^b_{ij}}} \\ &= p_{ij} \frac{\exp(-\beta\, \text{eSBI}(i,j,k,l)) Q^b_{kl}}{Q^b_{ij}}. \end{aligned}$$

c) (k,l) inner base pair of multiloop closed by (i,j)

McC: Base Pair Probabilities — Multiloop Case

2) general case

c) (k, l) inner base pair of multiloop closed by (i, j)

$$p^M_{kl}(i, j) :=$$
$$p_{ij} \Pr[\text{multiloop with inner base pair } (k, l) \text{ closed by } (i, j) \mid (i, j)]$$

Three cases: position of (k, l) in the multiloop

(i) (k,l) leftmost base pair

(ii) (k,l) middle base pair

(iii) (k,l) rightmost base pair

McC: Base Pair Probabilities — Multiloop Case

2) general case

c) (k, l) inner base pair of multiloop closed by (i, j)

$$p^M_{kl}(i,j) :=$$
$$p_{ij} \Pr[\text{multiloop with inner base pair } (k,l) \text{ closed by } (i,j) \mid (i,j)]$$

Three cases: position of (k, l) in the multiloop

(i) (k,l) leftmost base pair
$Q^b_{kl} Q^m_{l+1j-1} \exp(-\beta(a + b + (k - i - 1)c))$

(ii) (k,l) middle base pair
$Q^m_{i+1k-1} Q^b_{kl} Q^m_{l+1j-1} \exp(-\beta(a + b))$

(iii) (k,l) rightmost base pair
$Q^m_{i+1k-1} Q^b_{kl} \exp(-\beta(a + b + (j - l - 1)c))$

McC — Multiloop Case (Ctd.)

Recall

$p^M_{kl}(i,j) :=$

$p_{ij}\,\Pr[\text{multiloop with inner base pair } (k,l) \text{ closed by } (i,j) \mid (i,j)]$

putting the three cases of (k,l) position together

$$
\begin{aligned}
p^M_{kl}(i,j) \;=\; & p_{ij}[Q^b_{kl} Q^m_{l+1j-1} \exp(-\beta(a+b+(k-i-1)c)) \\
& + Q^m_{i+1k-1} Q^b_{kl} Q^m_{l+1j-1} \exp(-\beta(a+b)) \\
& + Q^m_{i+1k-1} Q^b_{kl} \exp(-\beta(a+b+(j-l-1)c))] Q^{-1}_{ij}
\end{aligned}
$$

McCaskill — Base Pair Probabilities — Summary

$$p_{kl} = p^{E}_{kl} + \sum_{i<k,l<j} p^{SBI}_{kl}(i,j) + \sum_{i<k,l<j} p^{M}_{kl}(i,j)$$

Remarks

- Recursive formula for p_{kl} furnishes DP
- Efficient calculation of all p_{kl} in $O(n^4)$ time/$O(n^2)$ space
- Time reduction to $O(n^3)$ possible (not shown, but you learned the "trick")
- The algorithm by the p_{kl} recursion is an outside algorithm; in contrast the algo for computing Z and the Q_{ij} is inside. For getting the probabilities, we combined inside and outside.

Summary Part I

Algorithms

- Nussinov
- Zuker
- McCaskill

Common

- $O(n^3)$ time, $O(n^2)$ space
- non-crossing structure (= "no pseudoknots")

Differences

- realism: base pairs $\leftrightarrow$ free energy (loop-based)
- mfe $\leftrightarrow$ ensemble

Next?

Comparing RNAs