

Exercise Sheet 6 for Computational Biology (Part 2), SS 14

Hand In: Until Tuesday, 22.07.2014, 10:00 am, email to `wild@cs...` or in lecture.

Problem 16

1 + 1 + 1 + 1 + 3 points

Consider Zuker's algorithm as given on pages 201f in the lecture script.

- a) Show that the recursion for $E(i, j)$ can be equivalently written as

$$E(i, j) := \min \begin{cases} E(L_{i,j}) \\ \min_{i \leq k < j} E(i, k) + E(k + 1, j) \end{cases} \quad (1)$$

(Note the \leq instead of the $<$ in the inner minimum!)

- b) Show that E fulfills a kind of *triangle inequality*:

$$\forall i \leq k < j : E(i, j) \leq E(i, k) + E(k + 1, j)$$

- c) Consider again recurrence (1) and assume that the value of $E(i, j)$ resulted from the second alternative, i. e. formally $\exists k : E(i, j) = E(i, k) + E(k + 1, j)$. Let k be *minimal* with this property.

Show that $E(i, k) = E(L_{i,k})$, i. e. the minimum for computing $E(i, k)$ was attained by the first alternative in (1).

Note: This means that in the bifurcation alternative, we only need to consider split points k , where the optimal substructure for range $i \dots k$ includes the base pair (i, k) !

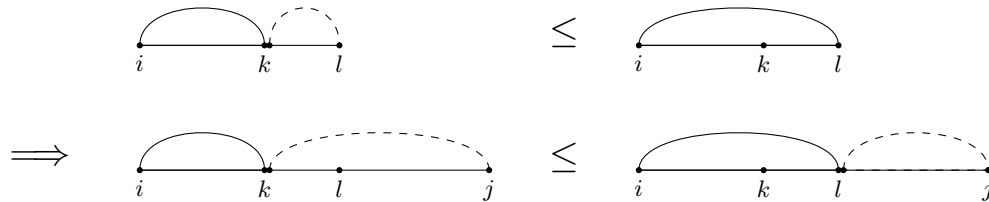
d) Prove the following *domination relation*:

Let $i < k \leq l$ be indices such that $E(L_{i,k}) + E(k + 1, l) \leq E(L_{i,l})$. Then for any $j \geq l$ also

$$E(L_{i,k}) + E(k + 1, j) \leq E(L_{i,l}) + E(l + 1, j)$$

holds.

For the more visually inclined, the claim says



Note: The dominance relation says that if for substructure $i \dots l$, including base pair (i, l) did not improve energy, neither does it when we extend the substructure to the right.

e) Use the results of c) and d) to design a variant of Zuker’s algorithm that does *not* naïvely iterate over all possible values for k in the bifurcation alternative.

Problem 17

2 + 4 points

Consider RNA primary structure $r \in \{A, C, G, U\}^n$ with compatible secondary structure $\mathcal{S} \subset [n]^2$. A pair $(i, j) \in \mathcal{S}$ closes a *multiloop* iff two or more base pairs are reachable from (i, j) (cf. Definition 63, script page 197).

a) How are multiloops handled in Nussinov’s and Zuker’s algorithm, respectively? Give an example structure with a multiloop and show the recursive decomposition implied by the algorithms.

Discuss why this decomposition is adequate for Nussinov’s objective, but might be problematic for Zuker’s algorithm.

b) Extending the energy model from lecture (Definition 64), let $e_m(k, d)$ be the free energy contribution of a multiloop of order d and length k , i.e. the multiloop comprises k unpaired bases and d base pairs are reachable from (i, j) .

Design a dynamic programming recurrence based on Zuker’s algorithm (as given on page 201 of the lecture script) incorporating proper contributions $e_m(k, d)$ for multiloops. Describe how the recurrence can be efficiently computed and analyze time and space complexity of your solution. Can you find an algorithm whose complexity is independent of the maximal multiloop degree?

Problem 18

2 points

Let $s := s_1 s_2 \cdots s_n$ be an RNA sequence of length n , i. e. its primary structure is given. Assume that base pairs can be formed between Watson-Crick base pairs (A–U and C–G) and G–U wobble pairs. Moreover, hair pins must span at least one unpaired base. Give a recurrence relation for the number of possible secondary structures for the given (fixed) primary structure s .