

Algorithms in Structural Bioinformatics

Assignment 1 (I.Emiris)

announced 21/2/13. Deadline: Tue. 26/3/13

Solutions: either by hand (hand-in at start of class) or upload. If you develop an implementation, upload. All uploads at <http://eclass.uoa.gr/modules/work/work.php?course=D464>.

1. Alignment [under/grads]

Match protein sequences $x = \text{ASRFALFF}$, $y = \text{ASIRVVFALF}$ using dynamic programming. Show the filled-in table and the backtrack path for the optimal solution. Also, write the optimal match after the table, including any inserted gaps.

(a) Apply global alignment [Needleman-Wunch] using +2 for matches, 0 for mutations, -1 for gap insertion.

(b) Apply local alignment [Smith-Waterman] using +2 for matches, -1 for mutations and gaps.

2. Multiple y -motifs [grads]

(a) For the proteins $x = \text{ASRFALFF}$, $y = \text{ASIRFL}$, with lengths $n = 8$, $m = 6$, find a maximum number of local matches scoring beyond threshold T , maximizing the following function (cf 1.dynpr.pdf, p.17), with BLOSUM50 scores (p.6), gap cost $d = 8$, threshold $T = 15$.

Initialize: $F(0, j) = 0$, $j = 1, \dots, m$; for $i = 1, \dots, n$: $F(i, 0) = \max \begin{cases} F(i-1, 0), \\ \max_{j=1, \dots, m} \{F(i-1, j) - T\}. \end{cases}$

Recurse for $i = 1, \dots, n$, $j = 1, \dots, m$:

$$F(i, j) = \max \begin{cases} F(i, 0), \\ F(i-1, j-1) + d(x_i, y_j), \\ F(i-1, j) - d, \\ F(i, j-1) - d. \end{cases}$$

Show the filled-in table and the backtrack paths of the matches scoring $> T$.

(b) If the number of matches is different than 2, can you change T so that there are exactly 2 matches?