Computational Biology

Midterm Examination

CSIE, GIBBT 210071

National Chi Nan University

Nov 12, 2007

You should give all *details* in your answers in order to get the points.

Problem 1 (10 points) Given any two strings $X = x_1 \cdots x_n$ and $Y = y_1 \cdots y_m$ as the input, we define the following recurrence relation

$$F(i,j) = \max \begin{cases} F(i-1,j-1) + s(x_i, y_i) \\ F(i-1,j) - 1 \\ F(i,j-1) - 1 \end{cases}$$

where $s(x_i, y_j) = 1$ if $x_i = y_j$, and -1 otherwise. The initial condition sets $F(0,0) = F(0,j) = F(i,0) = 0$ for $1 \leq i \leq n$ and $1 \leq j \leq m$. Please give an intuitive interpretation for the meaning of entries which take the maximum value in the last row (i.e., $i = n$) and the last column (i.e., $j = m$).

Problem 2 (10 points) Please list all possible DNA fragments that can map to the the following polypeptide chain by the standard genetic code:

Phe Phe Leu Ile Lys Arg Arg Gly Tyr.

Problem 3 (10 points) In the Partial Digest Problem, ΔA is defined as the multiset $\{|x-y| : x, y \in A\}$ where A is any set of integers. Show that there exist sets B and C such that $\Delta B = \Delta C$ (in multiset) but $B \neq C$ (in set).

Problem 4 (10 points) Suppose we have the following alignment

```
1:  A T C G A T A A A
2:  A T C A C A A T A
3:  A T G G A T A A A
4:  T C G A A A G G G
5:  A C C G C A A A A
```

Determine their *consensus* string.

Problem 5 (10 points) Let the input be a finite sequence of integers $x_1, x_2, \ldots, x_n$. Please design a linear-time algorithm to identify indices i and j such that $\sum_{k=i}^{j} x_k$ is maximum over $1 \leq i \leq j \leq n$. (Hint: Let $S(t) = \sum_{k=1}^{t} x_k$ for $1 \leq t \leq n$. The problem becomes to locate i and j such that $S(j) - S(i)$ is maximum. Define TWO arrays $A(t)$ and $B(t)$ where $A(t)$ records the maximum interval within $[1, t]$ and $B(t)$ records the maximum interval ending at t. Then apply the dynamic programming to evaluate A and B in a mixed way.)

Problem 6 (10 points) Given any string X, its *prefix reversal* is $y^r z$ where $yz = x$ and y^r is the *reverse* string of y. For example, if $X = abcde$, then $\underline{cba}de$ is a prefix reversal of X since we reverse the order of the first three characters in X. Give a derivation to transform $X =$ `ATCGTAAA` into $Y =$ `AAAATTCG` by reversal operations.

Problem 7 (10 points) Let permutation $\pi = 1\ 3\ 2\ 5\ 6\ 7\ 4$. Evaluate the number of break points in π after appending 0 and 8 to the front and end of π, respectively.

Problem 8 (10 points) Suppose we have two restriction enzymes A and B, and we want to determine the restriction map of A and B on a DNA sequence. When we add only A into this sequence, we get fully-cleaved fragments of lengths $3, 6, 9, 12$. When we add only B, we get $10, 20$. However, if we add both A and B, we get $3, 5, 6, 7, 9$. Please reconstruct the restriction sites (i.e., the locations) for A and B on this sequence.

Problem 9 (10 points) Show that there exists a function from natural numbers to natural numbers that it is neither $O(n^3)$ nor $\Omega(n^2)$.

Problem 10 (10 points) Give an example to show that there are strings A, B, C such that the longest-common subsequence of A, B, and C is shorter than the longest-common subsequence for any two strings out of A, B, and C.