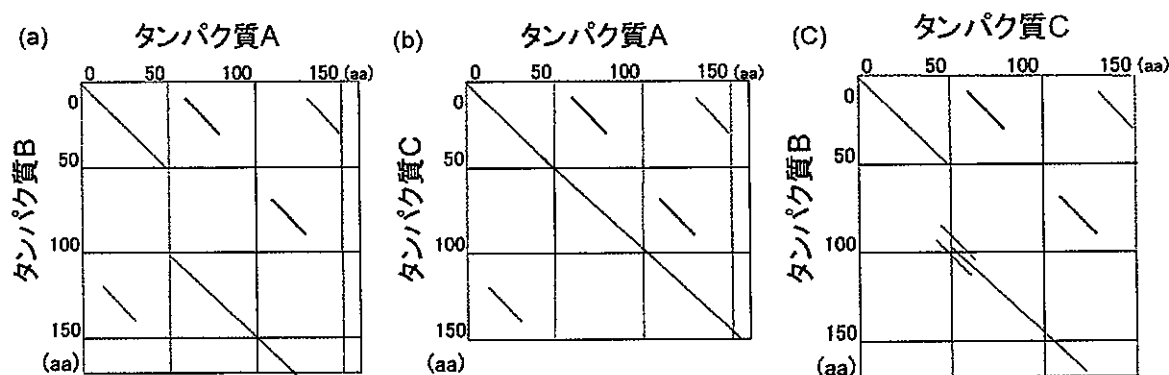
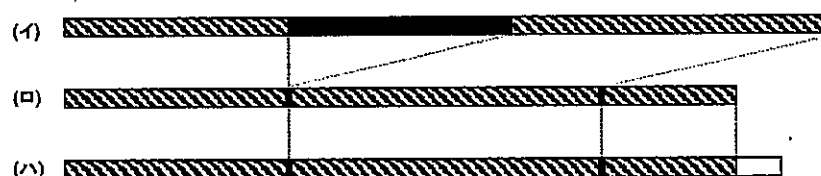


情報生物学（1 / 2）

〔1〕 タンパク質のアミノ酸配列を文字列として見なし、その二本の配列の類似領域を次のように図示化する。二つの配列の一方が左端から右方向に N 末端から、他方は上方から下方に N 末端から並べられている。20 アミノ酸残基のサイズのウィンドウで部分配列を比較し、20 残基ペア中的一致するアミノ酸の数が 13 個以上の時に、その残基ペアに対応する場所に斜めの線を引いて、類似している領域を表している。以下の設問に答えなさい。（計 50 点）



（1）図から予測される三つのタンパク質 A, B, C の一次構造の模式図を次に示す。模式図は左が N 末端、右が C 末端である。模式図中のパターンが同じ部分は、配列が類似している領域を示す。次の模式図中の(イ)、(ロ)、(ハ)と A, B, C の対応関係の中で正しいものはどれかを選択肢から選びなさい。（25 点）。

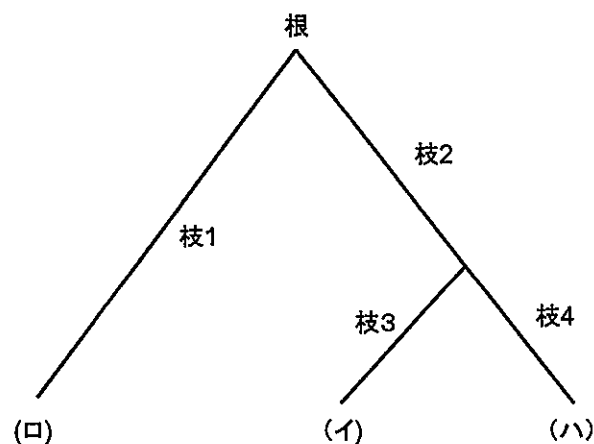


- | | |
|--------------------------|--------------------------|
| (1) (イ)–A, (ロ)–B, (ハ)–C, | (2) (イ)–B, (ロ)–C, (ハ)–A, |
| (3) (イ)–C, (ロ)–A, (ハ)–B, | (4) (イ)–A, (ロ)–C, (ハ)–B, |
| (5) (イ)–C, (ロ)–B, (ハ)–A, | (6) (イ)–B, (ロ)–A, (ハ)–C |

（2） 3つのタンパク質(イ), (ロ), (ハ)の配列アライメントから構築された分子系統樹が図に示すように構築された。図中の根は外群によって決定された。配列(イ)中の黒く塗りつぶされた領域は挿入か欠失か、また系統樹中どの位置で生じたか、正しい組み合わせを選択肢から選びなさい（25 点）。（最小イベント数で説明できるものを選択すること）

- (1) 枝 1、挿入、(2) 枝 1、欠失、(3) 枝 2、挿入、(4) 枝 2、欠失、
(5) 枝 3、挿入、(6) 枝 3、欠失、(7) 枝 4、挿入、(8) 枝 4、欠失、

情報生物学（2 / 2）



〔2〕 次の漸化式に従う動的計画法で塩基配列 A と B のグローバルアライメントを行なう。

$$D(0,0)=0, \quad D(0,n)=-n, \quad D(m,0)=-m$$

$$D(i,j)=\max \begin{cases} D(i-1,j-1)+s(A(i),B(j)) \\ D(i-1,j)-1 \\ D(i,j-1)-1 \end{cases}$$

$D(i,j)$ は、配列 A の 5'側から i 番目までの部分配列と配列 B の 5'側から j 番目までの部分配列のアライメントスコアを表す。 $s(a,b)$ は塩基 a と b が一致の場合は 1、不一致の場合は -1 をとる。 $A(i)$ と $B(j)$ はそれぞれ配列 A、B の 5'側から i 番目、 j 番目の位置の塩基を表す。この漸化式を解く過程で現れる A の部分配列（長さ m ）と B の部分配列（長さ n ）のアライメントを考える。

$A(m)=T$, $B(n)=C$, $D(m-1,n-1)=9$, $D(m,n-1)=11$, $D(m-1,n)=12$ であるとする。以下の問いに答えなさい。（計 50 点）

(1) 上記の漸化式の場合、長さ L のギャップに対するペナルティはどのように表されるか？選択肢から正しいものを選びなさい。（20 点）

(イ) $-2+(L-1) \times (-1)$, (ロ) $-2+L \times (-1)$, (ハ) $(L-1) \times (-1)$, (ニ) $L \times (-1)$

(2) A の長さ m の部分配列と B の長さ n の部分配列のアライメントスコア $D(m,n)$ はいくつになるか？ (15 点)

(3) この部分配列のアライメントの 3'末端は、次のどの状態となっているか？選択肢から正しいものを選びなさい。（15 点）

(イ)配列 A	T	(ロ)配列 A	T	(ハ) 配列 A	—
配列 B	C	配列 B	—	配列 B	C

注：選択肢中のハイフンはギャップを表す。