

平成25年度

日本バイオインフォマティクス学会 (JSBi)

バイオインフォマティクス技術者認定試験

解説集

問1 正解【2】

大腸菌  $4.6 \times 10^6$  bp、ショウジョウバエ  $1.8 \times 10^8$  bp、ヒト  $3.0 \times 10^9$  bp、  
パンコムギ  $1.7 \times 10^{10}$  bp

問2 正解【4】

遺伝子地図には組換え率から得られる連鎖地図と物理的距離に基づく物理地図とがあり、その作成においては SNP、RFLP、マイクロサテライト等が DNA マーカーとして用いられる。ゲノム解読にはホールゲノムショットガン法が用いられる。

問3 正解【1】

異なる生物のゲノム上にあつて共通祖先に由来するホモログをオーソログといい、同じ生物のゲノム上で遺伝子重複によって生じたホモログをパラログという。

問4 正解【2】

真正細菌のゲノムでもトランスポゾンが見られる。

問5 正解【3】

転写では DNA のアンチセンス鎖を鋳型として mRNA が合成される。

問6 正解【3】

20,000 塩基対 (40,000 塩基) のうち、C と G はそれぞれ 20% で 8,000 塩基ずつ、A と T はそれぞれ 30% で 12,000 塩基ずつとなる。

問7 正解【3】

塩基性アミノ酸では側鎖が正に帯電しており、酸性アミノ酸では側鎖が負に帯電している。

問8 正解【3】

クロマチンはヒストンタンパク質と DNA の複合体である。

問9 正解【1】

細胞膜では、受動輸送 (物質の濃度勾配に逆らわない) と能動輸送 (ATP を消費する輸送体による、物質の濃度勾配に逆らう輸送) とが起こっている。

問10 正解【4】

核・ミトコンドリア・葉緑体は、独自のゲノムを持つ細胞内小器官である。

問 11 正解【3】

ミトコンドリアは母性遺伝する。

問 12 正解【4】

アルコール発酵や乳酸発酵などの発酵の過程においても ATP は生成される。

問 13 正解【1】

N-グリコシル化はアスパラギン (N, Asn) 残基の N 原子への糖鎖付加によって起こる。

問 14 正解【2】

ウェスタンブロッティングは、目的のタンパク質に対する抗体の結合（抗原抗体反応）を利用した技術である。

問 15 正解【1】

DNA フットプリント法は、転写制御因子などのタンパク質が結合している DNA 領域を検出することを目的としている。

問 16 正解【3】

PCR 法では、反応液の温度を変化させ、二本鎖 DNA の変性、プライマーのアニーリング、相補鎖 DNA の合成を繰り返すことによって DNA 断片が増幅される。

問 17 正解【1】

kinase はタンパク質のチロシン、セリン、スレオニン残基へリン酸基を付加する酵素である。

問 18 正解【4】

神経細胞の興奮は  $\text{Na}^+$  と  $\text{K}^+$  のイオン分布による膜電位の変化によって起こる。また筋細胞では、トロポニンと  $\text{Ca}^{2+}$  の結合が筋収縮の引き金となっている。

問 19 正解【2】

二次元電気泳動法では、一次元目に等電点電気泳動、二次元目に SDS-PAGE による分子量分離を行う。

問 20 正解【2】

シャペロンは他のタンパク質分子のフォールディングを助ける。

問 21 正解【4】

Perl, Python, Ruby はインタプリタ型と呼ばれる。

問 22 正解【3】

例えば  $x$  を 0.5 にすると、3 のみ 1 になる。

問 23 正解【2】

通信速度は、機器に依存している。

問 24 正解【1】

ping は IP ネットワーク上で、接続を確認するためのツール。

問 25 正解【1】

物理的なメモリ以上の容量は、仮想記憶を使う。

問 26 正解【4】

$2^3(=8)$  までは割れるが、 $2^4=16$  では割れない。

問 27 正解【3】

OR と AND で NOT は表現できない。

問 28 正解【1】

各回の最小値の選択が最大  $N$ 。これを  $N$  回繰り返すので  $N^2$ 。

問 29 正解【4】

深さ優先では、全てのパスを調べないと最短であることが保証できない。

問 30 正解【3】

ハッシュ関数は、線形性を満たす必要はない。

問 31 正解【3】

擬似コードに従って、順に辿る。

問 32 正解【1】

研究室番号が 300 より大きくて、評価点が 82 以下なのは、1000 番と 1003 番。

問 33 正解【1】

中心極限定理で $(x - \mu)/\sigma^2$  が漸近的に標準正規分布に近づく。

問 34 正解【4】

箱ひげ図の上下に長いひげがある時、値が中央にまとまっているのではなく、裾が長い分布である可能性が高い。

問 35 正解【2】

共分散が 0 ならば、統計的に独立。

問 36 正解【4】

ベルヌーイ分布の確率変数の平均値は 0 以上 1 以下。

問 37 正解【2】

少なくとも 1 回出る確率は  $1 - (5/6)^3$ 。

問 38 正解【2】

ランダムと同等は、AUC=0.5 の時。

問 39 正解【2】

最近傍 3 点は標本 1、3、4。これらのクラスが A, B, B なので、B。

問 40 正解【3】

2-fold 法に特別な名称は無い。Red-Black tree (赤黒木) はデータ構造の名前。

問 41 正解【3】

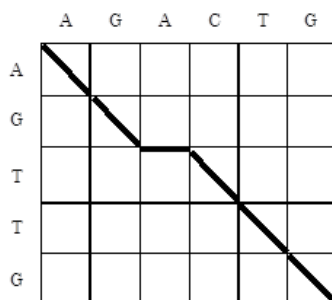
接尾辞配列は、接尾辞木（問 49 参照）と同様の、テキストの任意の部分文字列検索を可能とする索引技術だが、情報を木構造でなく配列として格納することにより記憶容量を削減したもの。バローズ・ホイーラー変換は文字列の並べかえに基づく可逆変換で、変換後の文字列は圧縮しやすく、かつ FM-index という索引を介して接尾辞配列と同様の索引として使用できるため、さらなる記憶域の節約が可能となり、長大なゲノム配列の索引づけに適している。なお、B 木はハードディスク上でのアクセス効率に優れた木構造で、データベースの索引づけなどに用いられる。

問 42 正解【1】

通常、順鎖が一致する領域の間に逆鎖が一致する領域がある場合は、そこで逆位が起きたことが想定されるが、ここでは逆鎖が一致する領域は順鎖を見ても一致している。これは逆位が起きたのではなく、配列中に順鎖と逆鎖が一致するパルンドローム構造が存在することを意味している。

問 43 正解【4】

最適なアラインメントパスが図に示されたもの以外になるような配列を探す。選択肢で異なっているのは 2 文字目と 3 文字目だけで、いずれの場合も相手の配列とは一文字しか一致しない。よって不一致の文字対を転換型から転位型に変えることができれば、スコアを改善できる。選択肢 1 は 2 文字目の C-G が転換型だが、アラインメントの変更によって改善することはできない。選択肢 2 と 3 は、不一致の文字対が転位型の置換であり、これ以上の改善はできない。選択肢 4 は現状で 3 文字目の T-A が転換型だが、ギャップの位置をずらすことにより T-C の転位型に変更できるため（下図参照）、こちらが最適アラインメントとなる。



問 44 正解【4】

非加重結合法では、距離が最小のクラスター対をとって併合し、併合されたクラスターと他のクラスターとの距離を元の距離の平均値で置き換えるという操作を繰り返すことにより階層的なクラスタリングを行う。ここではまず A と B が距離 0.2 で併合され、次に C と E が距離 0.3 で併合され、次にクラスター(A,B)と(C,E)が距離 0.4 で併合され、最後にこのクラスター((A,B),(C,E))と D とが距離 0.6 で併合される。

問 45 正解【4】

RefSeq は GenBank を元に NCBI で再構成して作成されたデータベースであり、元が同じデータであってもアクセッション番号は別のものに付け替えられている。

問 46 正解【3】

ビットスコアは、スコア行列ごとに決まるパラメータを用いて元のスコアを規格化したもので、 $E$ -value は、このビットスコア  $S$  とクエリ配列のサイズ  $m$ 、データベースサイズ  $n$  を用いて、 $E = mn2^{-S}$  で計算される。従って、ビットスコアは、使用したスコア行列の種類に依らない標準的な評価尺度とみることができる。一方、同じ配列アラインメントのビットスコアは、使用するスコア行列によって変化しうるが、これはアラインメントの評価にはスコア行列の選択が影響することを意味している。また、ビットスコアはデータベースサイズに依存しないので、異なるデータベース検索の結果を比較することができる ( $E$ -value は依存するのでできない)。

問 47 正解【2】

式 1 に従って計算する。式 1 の分子はコドン GCG の使用度数 46524、分母は、GCG がコードするアラニンにおけるコドン 1 つあたりの平均頻度  $(20813+35252+27567+46524)/4$  である。前者を後者で割れば答えが得られる。計算は有効数字 2 桁とれば十分。

問 48 正解【3】

スコア行列から D-D のスコアが 6、V-I が 3、K-R が 2 のように読み取れるので、これを順次足していけばよい。スコア行列は対称行列なので、行と列の区別は気にしなくて良い。

問 49 正解【2】

図の接尾辞木から位置 7 における接尾辞は [X]-TA であり、実際の文字列で 7 番目以降の文字列は GCTA なので、[X]には GC が対応する。また、[Y]は A で始まる 4 つの接尾辞の中の一つの位置に対応するが、文字列中で A が出現するのは 1, 4, 6, 10 番目なので、[Y]は 6 となる。

問 50 正解【4】

クオリティ値 20 は、アスキーコードでは 33 を加えた 53 で表される。従って数字の 5 以下、およびそれより小さなアスキー値を持つ記号がクオリティ値 20 以下となる。

問 51 正解【2】

MEGABLAST は核酸配列の類似性検索ツールだが、特に長大で高い類似性を持つ DNA 配列の照合に向けて最適化されたもの、MUMmer は近縁ゲノム配列のアラインメントに適したアラインメントツールである。なお、BLAST シリーズの中で位置特異的行列を使った検索を行うプログラムとしては PSIBLAST や RPSBLAST が、隠れマルコフモデルを用いた検索を行うツールとしては HMMER がある。

問 52 正解【3】

R<sub>4</sub> は CTA, TAC, ACA, CAT, ATC の 5 つのワードがこの順につながったグラフとして表現され、このうち CTA と ATC のノードが、それぞれ図のグラフの 3 番目と 7 番目のノードと一致する。これらを合わせると、3 番目のノードで分岐し 7 番目のノードで合流する 3 のグラフとなる。



問 53 正解【1】

カルシウムなどの金属イオンは疎水性のアミノ酸の側鎖には結合することはできない。

L (ロイシン) 疎水性のアミノ酸であるので、選択肢 1 は不適切である。また、カルシウムイオン( $\text{Ca}^{2+}$ )はプラスの電荷を持つので、同じプラスの電荷を持つアミノ酸 (K (リジン) や R (アルギニン) ) の側鎖は結合できない。よって選択肢 2 は適切である。疎水性のアミノ酸 (I (イソロイシン) 、V (バリン) ) だけが並ぶ部位 3 より、E (グルタミン酸) 、K などの親水性のアミノ酸が混在する部位 1 のほうが、タンパク質表面に露出していると思われるので、選択肢 3 は適切である。また、G (グリシン) や D (アスパラギン酸) は  $\alpha$  ヘリックス構造を形成しにくいので、選択肢 4 は適切である。

問 54 正解【2】

RMSD は対応する原子間距離の二乗平均の平方根であるので、 $\{(2.0^2 + 1.0^2 + 0.0^2 + 1.0^2 + 2.0^2)/5\}^{1/2} = 1.4\dots\text{\AA}$  (選択肢 2 にもっとも近い) が正解である。

問 55 正解【2】

二つの原子間には、ポテンシャルエネルギー関数がより小さくなる方向に力が働く。

$(r_0/r)^{12}$  と  $(r_0/r)^6$  はどちらも常に正の値であり  $r$  が大きくなるとゼロに近づく。よって、 $-$ の符号の 6 乗項が引力、 $+$ の符号の 12 乗項が斥力を表すことがわかる。問題文の式(上)の  $r$  に  $\sigma$  を代入すると、 $v(\sigma) = 4\epsilon[(\sigma/\sigma)^{12} - (\sigma/\sigma)^6] = 0$  すなわち  $r = \sigma$  でポテンシャルがゼロになることがわかる。また、問題文の式(下)を  $r$  について微分すると

$$dv/dr = \epsilon[-12(r_0/r)^{12}/r + 12(r_0/r)^6/r] = 12\epsilon(r_0/r)^6/r[-(r_0/r)^6 + 1]$$

よって、 $r = r_0$  で  $dv/dr = 0$  となり極小値をとることがわかる。

問 56 正解【1】

溶媒露出表面 (溶媒接触表面とも呼ばれる) は、溶質分子に接触した球の中心の軌跡で定義され、(A)図に相当する。(B)図に示した、球が侵入できない領域との境界面は「分子表面」あるいは「コノリー表面」と呼ばれる。ドッキングやフォールディングによって、結合していた水分子が押しのけられて、溶媒露出面積は減少する。以上を考えると、選択肢 1 が正解である。

問 57 正解【4】

SCOP は、配列上明らかな相同性があると考えられるグループをファミリー、配列上はそれほど似ていないが立体構造や機能部位の類似性から相同であると考えられるグループをスーパーファミリー、配列や機能は類似していないが立体構造の概形が類似しているグループをフォールドとして階層的に分類する。ファミリーとして十分な配列の一致度はだいたい 30%程度であり、(A)と(C)は同じファミリーであると考えられる。また、このタンパク質の場合  $\text{Fe}_2\text{S}_4$  が結合する場所が機能部位である。その結合位置の類似性と構造全体の類似性から、配列の類似性が低くても(A)と(B)は同じスーパーファミリーに属すると考えられる。一方(A)と(D)は、立体構造（特に主要な二次構造の配置）は似ているが、配列の一致度は低く、かつ(D)は、 $\text{Fe}_2\text{S}_4$  が結合していない。よって、(A)と(D)はフォールドのみが同じであり、ファミリー・スーパーファミリーは異なると考えられる。よって選択肢 4 が正解である。

問 58 正解【1】

多くのアミノ酸では負の  $\varphi$  角の二つの領域 ( $\alpha$  ヘリックス,  $\beta$  シートに相当) に集中する。一方、グリシンは、側鎖が水素しかないため、側鎖と主鎖の衝突を起こさずに、正の  $\varphi$  角の構造もとることができる。逆に、プロリンは N-C $\delta$  原子間が環状に結合しているため、 $\varphi$  角はせまい領域に固定されてしまう。以上を考えると選択肢 1 が正解であることがわかる。

問 59 正解【3】

グリシン、バリン、グルタミン酸の三つのアミノ酸の中で、最も疎水的なのはバリン、最も親水的なのはグルタミン酸である。また、グルタミン酸やロイシンは  $\alpha$  ヘリックスになりやすく、バリンやイソロイシンは  $\beta$  シートになりやすい。グリシンはフレキシブルなので、どちらの二次構造もとりにくい。以上を考えると選択肢 3 が正解である。

問 60 正解【4】

表により、親水性アミノ酸は膜中よりも水中の方が、ポテンシャルエネルギーが低い一方、疎水性アミノ酸は水中よりも膜中の方がポテンシャルエネルギーが低い。従って、親水性アミノ酸がすべて水環境にあり、疎水性アミノ酸がすべて膜環境にあるペプチド(A)が最低のポテンシャルエネルギーを持つことが判る。(A)と親水性アミノ酸数、疎水性アミノ酸数の構成が同じである残りのペプチドについては(A)を基準としてここから存在環境が変化したアミノ酸についてのエネルギーの差分だけ考慮すれば良い。差分は  $10 \times$  (膜環境にある親水性アミノ酸の数)  $+5 \times$  (水環境にある疎水性アミノ酸数) kcal/mol。よって、(A)に比べて(B)、(C)、(D)はそれぞれ、30、15、5kcal ポテンシャルエネルギーが高くなるので、最もエネルギー差が小さい選択肢 4 が正解である。

#### 問 61 正解【4】

タンパク質は自然な環境で自発的に折り畳まれる。これは、折りたたまれた状態が系の自由エネルギーが最小の状態であると言い換えられる。この時、主に非共有結合（疎水性相互作用、水素結合、イオン結合、ファンデルワールス力）が働く。このうち、疎水性相互作用は、水に溶けにくい領域が水に接する面積を最小化しようとする見かけの相互作用であり、球状タンパク質では、構造の内側に疎水性アミノ酸を配し、親水性アミノ酸を外側に配す駆動力になる。タンパク質内部に埋め込まれた主鎖のアミノ基、カルボニル基は、殆どが水素結合を形成し、これが二次構造を形成する原因となる。球状タンパク質の外側で水と相互作用する親水性アミノ酸は、+電荷を持つ塩基性アミノ酸（アルギニンカリジン）と-電荷を持つ酸性アミノ酸（アスパラギン酸かグルタミン酸）の場合が多い。よって選択肢 4 が正解である。

#### 問 62 正解【1】

二次構造要素のほとんどが  $\alpha$  ヘリックスである構造を all- $\alpha$  型、ほとんどが  $\beta$  ストランドからなる構造を all- $\beta$  型と呼ぶ。 $\alpha$  ヘリックスと  $\beta$  ストランドの両方が混在する構造は、 $\alpha/\beta$  型か  $\alpha+\beta$  型に分類される。 $\alpha/\beta$  型は、配列上  $\alpha$  ヘリックスと  $\beta$  ストランドが交互に現れ、かつ全体に平行の  $\beta$  シートが多い。TIM バレルやロスマン・フォールドが有名である。一方、 $\alpha+\beta$  型は、配列上交互に現れるという規則性はなく、かつ全体に逆平行の  $\beta$  シートが多い。以上を考えると、選択肢 1 が正解であることがわかる。

#### 問 63 正解【4】

ホモロジーモデリング法は、「アミノ酸配列が類似しているほど立体構造も類似している」という経験則に基づき、配列が類似している既知の立体構造を鋳型として予測する。よって、配列類似性が低いほど誤差も大きく、また、必然的に、その予測構造は鋳型とした構造と類似する。一方、フラグメントアセンブリ法は、既知構造群を 9 残基ほどの短い断片（フラグメント）に分割、それらを組合せることで全体の立体構造を構築し、フラグメントの組合せを様々に変えることでポテンシャルエネルギーが低い構造を探索する。よって、組み上げた全体の形は、必ずしも既知のタンパク質の構造全体と似ているわけではない。また、標的アミノ酸配列が長くなると、探索すべきフラグメントの組合せが増大し探索が不十分に終わるため、予測誤差が増大する傾向がある。逆に、ホモロジーモデリング法の誤差は、標的アミノ酸配列の長さあまり依存しない。

問 64 正解【2】

図では、棒状構造がリン酸骨格、五角形の板状構造がデオキシリボース、水平の板状構造が塩基を表す。リン酸骨格をたどると右巻きの二重らせんであることがわかり、塩基を数えると 26 塩基 (=13 塩基対) あることがわかる。B 型構造の特徴は塩基がらせん軸に垂直になることである。片側から見ると、主溝と副溝は交互にあらわれる。よって選択肢 2 が正解である。

問 65 正解【4】

一塩基多型は遺伝的に継承されるものであり、ある個体で後天的に生じる変異は含まない。この変異が子孫に継承される場合は、将来この変異が一塩基多型となる可能性はある。

問 66 正解【1】

コピー数多型について、繰り返し配列の占める領域の上限は定義されていないので、選択肢 1 は不適切である。

問 67 正解【2】

遺伝子が水平伝搬しているかどうかは、種の系統樹と遺伝子の系統樹とが大幅に食い違うことから見いだされる。

問 68 正解【3】

孫花は子花  $RrQq$  と子花  $RrQq$  の掛け合わせであるから、赤花・丸型 ( $RRQq, RRQq, RrQQ, RrQq$ ) : 赤花・しわ型 ( $RRqq, Rrqq$ ) : 白花・丸型 ( $rrQQ, rrQq$ ) : 白花・しわ型 ( $rrqq$ ) = 9 : 3 : 3 : 1 となり、白花・丸型になる確率は  $3/16$  である。

問 69 正解【3】

チンパンジーA~G 間の変異は 7 塩基である (2, 12, 16 番目はチンパンジーでは変異していない事に注意)。ヒト A~G の配列はそれぞれ最低 1 箇所塩基が異なるので、ハプロタイプは 7 種類である。チンパンジーA とヒト A では 2, 7, 10, 12, 14, 16, 18, 20 番目の 8 塩基が異なっている。ヒトとチンパンジーで完全に保存されているのは、9, 19, 21 の 3 塩基部位であるので、選択肢 3 が不適切である。

問 70 正解【2】

CNV(Copy Number Variation)、QTL(Quantitative Trait Locus)、SNP(Single Nucleotide Polymorphism)、VNTR(Variable Number of Tandem Repeat)のうち、QTL は身長などの連続した数値で示される形質を決定する量的形質座位を意味するので、選択肢 2 が不適切である。

問 71 正解【2】

たとえば、系統樹を左(過去)からたどると、まず  $\alpha$ -、 $\beta$ -グロビンが分岐し、次にヒトとウマが分岐する事がわかる。すなわち、 $\alpha$ -、 $\beta$ -グロビン分岐の方が年代的には先に起こっているため、選択肢 2 が不適切である。

問 72 正解【2】

ゲノムの GC 含量は、ゲノムが情報の蓄積物であることと、その物理化学特性、およびその生物が成育する環境とのせめぎ合いによって決定されるとされている。コドンの構造と、GC 対は AT 対よりも水素結合の数が多いことを知っていれば、この問題を解くことができる。

問 73 正解【3】

デ・ノボ・アッセンブルは、得られたリードのみからゲノム配列を構築する技術である。一方、近縁種のゲノム配列が利用できる場合には、その近縁種配列にリードをマッピングすることで効率よくアッセンブルを行うことができる。

問 74 正解【4】

PubMed は文献検索サービスであり、任意の遺伝子群に対してパスウェイ情報を提供する機能はない。

問 75 正解【2】

二次元電気泳動は 2 種類の物理化学的性質を順に利用することで混合物を分離する手法であり、通常、等電点と分子量を用いた大量のタンパク質の分離に用いる。

問 76 正解【1】

プリン塩基(A,G)、ピリミジン塩基(T,C)に関してはグループ内とグループ間で置換頻度に違いがあることが知られている（それぞれ、トランジションとトランスバージョン）。一方、2 本の水素結合を作る塩基対(A,T)、3 本の水素結合を作る塩基対(G,C)にそのような現象は知られていない。

問 77 正解【1】

通常、網羅的解析においては、特定の分離技術に質量分析器(MS)を用いた同定システムを組み合わせる。液体クロマトグラフィー(LC)を分離に用いたシステム LC-MS や、ガスクロマトグラフィー(GC)を分離に用いたシステム GC-MS、キャピラリー電気泳動(CE)を分離に用いたシステム CE-MS などがある。

問 78 正解【2】

オイラー路は複数あり、例えば  $a \rightarrow b \rightarrow e \rightarrow d \rightarrow c$  と通る経路。全ての頂点の次数が偶数の場合オイラー閉路となるが、本題では奇数次数の頂点が 2 つあるので準オイラーグラフ。ハミルトン路は例えば  $G1 \rightarrow G2 \rightarrow G4 \rightarrow G3$  と通る経路で、閉路であるためハミルトングラフとなる。

問 79 正解【2】

代謝流束解析にはネットワークと化学量論関係がわかっているだけで、個別の酵素反応速度（基質濃度依存性）は不要。

問 80 正解【4】

ダブリングタイム（倍化時間）は、半減期と同様に物理的事象の変化速度の目安となる典型的尺度であり、比例定数  $k$  が大きくなればダブリングタイムは小さくなる。

試験問題に記載されている会社名または製品名は、それぞれ各社の商標または登録商標です。

なお、試験問題では、®および™を明記していません。