

平成 22 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(2014 年 7 月 10 日改訂)

問 1 正解【1】

ホスホジエステル結合は核酸の骨格構造形成において重要。ジスルフィド結合はタンパク質のシステイン残基間で、イオン結合は酸性・塩基性アミノ酸残基間で、水素結合は二次構造形成などで、それぞれタンパク質の構造安定性に寄与する。

問 2 正解【2】

リン脂質は両親媒性であるが、頭部は親水性で、尾部の炭化水素は疎水性である。

問 3 正解【4】

電子伝達系によるエネルギー生産を行うのは、ゴルジ体ではなく、ミトコンドリアである。

問 4 正解【2】

減数分裂の結果、ヒトなどの生殖細胞の染色体数は、体細胞の $1/2$ になる。

問 5 正解【3】

AB 型の親は A と B の両方の遺伝子型をもっているため、その両親から生まれた子の遺伝子型は、AA, AB, BB のいずれかになる。

問 6 正解【3】

第一世代の体色が赤なので、A は優性である。第一世代の遺伝子型は Aa なので、そのかけあわせから生まれる遺伝子型は、AA : Aa : aa = 1 : 2 : 1 になる。

問 7 正解【4】

UTR は untranslated region (非翻訳領域) のことで、mRNA の 5'側と 3'側にそれぞれ存在する。

問 8 正解【1】

RNA スプライシングは真核生物に見られ、mRNA 前駆体からイントロンを取り除くなどのプロセッシング反応の結果、成熟 mRNA が形成される。成熟 mRNA の情報をもとにタンパク質が作られる。

問 9 正解【3】

コドンは $4 \times 4 \times 4 = 64$ 通りあるが、このうち 3 種類の終止コドンを除いた 61 通りはすべて 20 種類のアミノ酸のどれかに対応している。

問 10 正解【1】

マイクロ RNA(miRNA)は、通常その 5'端近くにあるシード領域が、標的となる mRNA の 3'非翻訳領域と相補的に結合することにより、タンパク質合成を阻害したり、mRNA の分解を促進したりする。相補的な結合の場合、核酸の向きはそれぞれ逆になることに注意。

問 11 正解【2】

糖転移酵素は主にゴルジ体に存在する。小胞体でも糖鎖付加反応は行われるが、それは翻訳が活発に行われる粗面小胞体においてである。

問 12 正解【4】

イオンチャネル型受容体においては、適当なりガンド(アセチルコリンなどのシグナル分子)が結合することによって、チャネルが開閉され、そのイオンの通過が調節される。イオン自身はシグナル分子ではないことに注意。

問 13 正解【4】

がん遺伝子の産物は、通常はシグナル伝達系などにおいて細胞機能を受け持っているが、遺伝子上の突然変異などによって正常な機能を果たせなくなると、細胞増殖のシグナル伝達経路に制御がきかなくなって、がん化に至るものと考えられている。

問 14 正解【3】

ヒトゲノム計画が始まる前は、ヒトのもつ遺伝子数(タンパク質をコードするもの)は 10 万個程度であると推定されていたが、ヒトゲノムを解読してみた現在では、2 万数千個程度であるとされている。選択的スプライシングなどによって、実際に形成されるタンパク質の種類は 10 万種類程度になる可能性がある。

問 15 正解【2】

フグのゲノムサイズは約 3 億 5 千万から 4 億塩基対で、脊椎動物の中でも特にゲノムが小さいことで有名である。ヒトゲノムのサイズは約 32 億塩基対である。

問 16 正解【3】

ヒトゲノムにおいて、反復配列の示す割合は約 50%である。イントロンを含む遺伝子領域はゲノムの 20%以上を占めるが、タンパク質をコードするエクソンは全体の 1.5%程度を占めるに過ぎない。

問 17 正解【4】

配列(a)は領域 A の下側の配列に相補的なので、もう一方のプライマーは、領域 B の上側の配列に相補的である必要がある。配列は 5'側から読むことに注意。

問 18 正解【4】

SDS-PAGE では、試料タンパク質を負電荷をもつ SDS で覆い尽くしてしまう。この分子は陽極の方に引っ張られて行く。ゲルは分子ふるいとして働くため、分子量の大きな分子ほど、ゲルの編み目に引っかかって泳動速度が遅くなる。

問 19 正解【1】

サザンハイブリダイゼーション法は、ゲル中に存在する特定の DNA 分子の存在を検出する方法。共免疫沈降法は、抗体によって沈殿させたタンパク質と相互作用しているために一緒に沈殿するタンパク質を調べる方法。

問 20 正解【2】

X線結晶構造解析で得られるのは、タンパク質などを構成する原子の電子密度分布である。これに対して、NMR法では主として水素(^1H)や炭素(^{13}C)などの原子間距離情報が得られる。(2011年6月7日改訂)

問 21 正解【1】

101₍₂₎でビットの値が 1 となっているのは1桁目と3桁目である。このため y と z の間で1桁目と3桁目のビットがそれぞれ同じであれば $f(y) = f(z)$ となるが、そのような組み合わせは選択肢1だけである。

問 22 正解【2】

Ruby はインタープリタ、MySQL はリレーショナルデータベースのひとつである。

問 23 正解【2】

多くのアーキテクチャでは機械語の1命令は1クロック以内では完了しない。そのためパイプライン処理などを行って、全体の高速化をはかることが多い。

問 24 正解【4】

telnet, ftp はネットワーク上を平文で情報が流れるため、インターネット上で用いるにはセキュリティ上の問題がある。一方、ssh、sftp は暗号化されている。

問 25 正解【3】

選択肢 3 が計算量の定義である。最悪計算量ともいう。最悪の場合ではなく、平均の演算数で議論する場合は平均計算量とって区別することに注意。

問 26 正解【1】

クイックソートは平均的には高速でよく用いられるが、最悪計算時間は決して速くなく、マージソート、ヒープソート等の方が最悪計算時間は良い。アルファベットや整数などをソートする場合には、バケットソート(バケツソートともいう)を用いると速いことも多い。また、バケットソートはラディックスソート等のより複雑なソート手法のベースとなっている。

問 27 正解【2】

スタックあるいはキューが空にならない限りは、スタックでは最後にプッシュした文字、キューにおいては最初に挿入した文字が最後に出力される。

問 28 正解【4】

たとえば、入力数列「11」の後の状態は S0 である。

問 29 正解【4】

B-木では、各節点において数百、数千の分岐数を持つことが少なくない。データベースの索引に使う場合において、むしろ分岐数が少ないほうが、段数が増えるためにハードディスクへのアクセスの回数が増えて効率が低下する可能性が高い。

問 30 正解【3】

ハッシュ関数には様々なものを使うことができるが、広い定義域を、比較的狭い値域に射影することが目的であるので、一般に線形性は成り立たない。大きな除数で割った剰余を返すハッシュ関数が良く用いられるが、この場合に線形性が成り立たないことは明らかである。

問 31 正解【1】

サブタイプが H1N1 で、年号が 1977 年以降のものが選ばれる。

問 32 正解【3】

XML データベースでも画像や音声データを扱うことは可能である。

問 33 正解【3】

N をデータ数とすると、平均値はデータの合計を N で割ったもの、中央値はデータを昇順に並べて $N/2$ 番目のデータと $N/2+1$ 番目のデータの平均、最頻値は頻度が最も多かった値、と定義されている。

問 34 正解【1】

X, Y が統計的独立のとき $E(XY) = E(X)E(Y)$ を満たす。

問 35 正解【2】

正規分布に従う2変量の和は正規分布に従う。

問 36 正解【3】

いかなる確率密度関数においても $p(x) < 0$ にはならない。選択肢4については、確率関数の全積分値が1を上回ることはないが、確率密度が1を越える事は可能である。(2011年6月7日改訂)

問 37 正解【2】

UPGMA 以外は非階層的クラスタリングを行う手法である。

問 38 正解【2】

まず $k=1$ の場合から考える。最も近いデータは第7標本である。第7標本は陰性なので、よって、予測結果は陰性となる。 $k=3$ の場合は次のようになる。最も近い3個のデータは、第7標本、第3標本、第8標本となる。陰性が2個、陽性が1個なので、多数決により予測結果は陰性となる。

問 39 正解【2】

二分探索法は検索のアルゴリズムの一種で、クロスバリデーション法とは関係ない。

問 40 正解【3】

F値は再現率と適合率の調和平均である。すなわちF値の逆数が、再現率の逆数と適合率の逆数の算術平均と等しい。選択肢3の式はF値の式として正しいが、説明文が誤っている。

問 41 正解【1】

BLAST はよく使われるプログラム群であり、クエリやデータベースがアミノ酸配列でも DNA 配列でもどのような組み合わせでも比較することが可能である。アミノ酸配列は DNA 配列に比べ、残基の種類が多く、同義置換もあることなどから、より遠縁の相同配列を検出しやすい。

問 42 正解【2】

相同性が切れるところ(ブレイクポイント)から、X軸・Y軸に垂線を引くとわかりやすい。

問 43 正解【1】

系統樹に従って近い配列から順にアラインメントを作製する。この方法は比較的早い時期に成功した多重アラインメントの代表的方法で、他の選択肢も実際にあり得る方法なので少し紛らわしいが、累進法と言えばこの方法を指すのが普通である。また、選択肢にはないが、他の多重アラインメントの手法として、反復改善法と呼ばれるものなどがある。

問 44 正解【2】

ビットスコアはスケールを調整して標準化した値なので、選択肢 2 のとおりスケールの変更には依存せず、スコアの値がそのまま有意性の目安になる。選択肢 4 はその目安を説明しているが、二重に間違いがある。ビットスコアの単位ビット(bit)は、底を 2 とした対数で定義される情報量の単位なので、 $10^A S$ ではなく $2^A S$ の値に意味がある。また、スコアが大きいほど類似性が高いのだから「十分小さい」は間違い。

問 45 正解【4】

「左端からスコアを順に合計していき、和が 0 以下になったら 0 にリセットする」という Smith-Waterman 法を簡略化した方法で系統的に解ける。別法として、中央にある一致が集中したところに注目して、両端にどこまで伸ばせるかを考えてもよい。その場合は、一致 1 つにつき不一致 1 つはプラスだが、不一致 2 つはマイナスである点に注目すると早い。

問 46 正解【4】

推論規則を見ると、途中で part_of の関係が入れば、全体としての関係も part_of となることがわかる。したがって選択肢 4 は間違い。cytoplasmic part と intracellular part の間には 2 つの経路があるので、1 と 2 はどちらも正しい。

問 47 正解【1】

塩基対の数を正しく追っていくとともに、バルジループやヘアピンループなど、塩基対を形成していない部分にも着目すること。

問 48 正解【4】

挿入や欠失はギャップペナルティとして扱われるので、B の記述は誤り。PAM や BLOSUM はアミノ酸の置換行列であるので、C の記述は誤り。

問 49 正解【1】

選択肢にある部分配列それぞれについて、順にスコアを計算していくことで、容易にもっとも高いスコアを示す部分配列がわかる。

問 50 正解【4】

完全一致している部位でのスコアは 2 になる(1、2、5 番目の部位がこれに相当)。2 種の塩基が等量現れる部位でのスコアは 1 になる(3 番目の部位がこれに相当)。出現残基の傾向が全く見られない部位ではスコアは 0 になる(4 番目の部位がこれに相当)。

問 51 正解【2】

モチーフに現れる C の距離制約に注目すると解きやすい。ここで扱った正規表現は、プログラミング等に利用される正規表現と区別する際には、プロサイトモチーフ表現と呼ばれる場合がある。(2011年6月7日改訂)

問 52 正解【1】

DNA の回文配列とは、もとの鎖に対して相補鎖が同じ配列になるものである。DNA 鎖に方向性があることに注意。選択肢 3 は単純に左右対称な文字列にすぎないので誤り。

問 53 正解【1】

ペプチド結合 $-C(=O)-N(-H)-$ は共鳴により平面性(C-N 結合のねじれ角が 0° か 180° 付近に制限される)を示し、 $=O$ と $-H$ が C-N に対して同一方向にある場合をシス型、逆方向の場合をトランス型という。通常はトランス型がエネルギー的に安定であるが、側鎖が N に結合したプロリンではシス型も比較的多くみられる。

問 54 正解【2】

アミノ酸の光学異性は C_α 原子に結合する 4 つの化学団がすべて異なることに起因する。グリシンだけは、側鎖($-H$)と水素原子($-H$)が同一化学団になるので、光学異性体が存在しない。

問 55 正解【1】

原子座標系は通常右手系で表現される。その場合、 \overrightarrow{CaN} 、 \overrightarrow{CaC} 、 $\overrightarrow{CaN} \times \overrightarrow{CaC}$ は右手系をなす。すなわち、外積 $\overrightarrow{CaN} \times \overrightarrow{CaC}$ は \overrightarrow{CaN} と \overrightarrow{CaC} に垂直で紙面奥に向かうベクトルである。 $\overrightarrow{CaC\beta}$ は同じく紙面奥に向かっているので、外積ベクトルとのなす角が鋭角(内積が >0)であることにより、L-アミノ酸の判別が可能である。この正の値がちょうど 1 になることは期待できないので、選択肢 4 は不適切である。(2011年6月7日改訂)

問 56 正解【4】

現在の水準の分子力学法では、分子軌道を考慮する場合はフラグメント分子軌道(FMO)法を用い、シュレーディンガー波動方程式を分子全体について解くことは行われない。

問 57 正解【3】

アミノ酸置換と欠失は REMARK 行に、2 次構造は HELIX 行と SHEET 行に記述されている。解析手法は EXPDTA 行にあるが、溶液 NMR で測定されたとあるので選択肢 3 が最も不適切である。NMR 構造では温度因子(ATOM 行)が定義されず 0.00 が記載されるが、信頼すべき値ではない。

問 58 正解【2】

TIM バレルフォールドとロスマン(Rossmann)フォールドは α ヘリックスと β ストランドが交互に現れる典型的な α/β フォールドである。またグロビンフォールドは α ヘリックスのみからなるフォールドの典型である。免疫グロブリンフォールドは β シートのみからなるフォールドの典型である。

問 59 正解【2】

タンパク質構造分類データベースは、SCOP (Class – Fold – Superfamily – Family の 4 階層)、CATH (Class - Architecture - Topology – Homologous superfamily の 4 階層)の様に階層分類されていることが特徴である。

問 60 正解【4】

一般に、ホモロジーモデルの精度は鋳型構造との配列一致度(保存度)が高いほど向上する。

問 61 正解【4】

これらの手法の中で、ホモロジーモデリング法だけが、相同性のある既知構造の存在を前提とする方法であるので、有意な類似性を示すタンパク質が見つからない場合は適用不能である。

問 62 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、 $\{(3.0^2 + 1.0^2 + 0.0^2 + 1.0^2 + 3.0^2)/5\}^{1/2} = 2.0$ が正解である。

問 63 正解【1】

B 型 DNA は、図の様に塩基対の面がらせん軸とはほぼ直角をなす。溝の幅から、図中央で主溝が正面を向いていることがわかる。塩基を数えると、図には 13 塩基対(26 塩基)が示されている。リン酸骨格をたどると、この DNA が右巻き(どちらに骨格をたどっても右回りする)であることがわかる。

問 64 正解【4】

この図には α ヘリックスしか示されていない。また、ドメイン(ペプチド鎖のかたまり)は 1 つしか認められない。空間充填模型は個々の原子を球で表現するものであり、この図は、あきらかにリボンモデルである。この構造には長い α ヘリックスが 4 本あり、束を作っている(左端の短い α ヘリックスは余分であり無視される)ので 4-ヘリックスバンドルタンパク質である。

問 65 正解【2】

CNV はコピー数多型(copy number variation)、SNP は一塩基多型(single nucleotide polymorphism)、VNTR はタンDEM反復配列の反復数の変異(variable numbers of tandem repeats)をそれぞれ表し、遺伝子多型に関連する。QTL は量的形質座位(quantitative trait locus)のことで、多型には関係しない。

問 66 正解【2】

O 型の遺伝子型 OO しかありえないので、0.16 の平方根をとって O の頻度は 0.4 であり選択肢 2 か 4 しかない。A 型の遺伝子型は、AA と AO なので、選択肢 2 の A の頻度から A 型の頻度を計算すると、 $0.1 \times 0.1 + 2 \times 0.1 \times 0.4 = 0.01 + 0.08 = 0.09$ となる。選択肢 4 の場合、同様に計算すると A 型の頻度は $0.3 \times 0.3 + 2 \times 0.3 \times 0.1 = 0.09 + 0.06 = 0.15$ となる。問題の A 型の頻度は、0.09 なので 2 が正解。

問 67 正解【2】

染色体上近接する 2 点の間では組み換えが距離の生じにくいと、その 2 点の間で連鎖不平衡を生じやすい。したがって、不適切なのは選択肢 2 である。

問 68 正解【2】

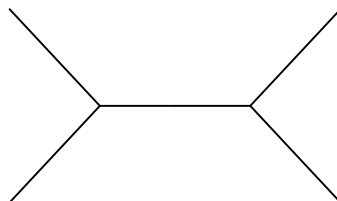
分子時計は、オーソログな遺伝子の中の塩基配列あるいはアミノ酸配列の間の差異が、その遺伝子を持つ生物の分岐年代にほぼ比例することを意味しており、放射性同位体の崩壊とは直接関係しないので、不適切なのは選択肢 2 である。

問 69 正解【4】

オーソログな α グロブリン、 β グロブリンそれぞれにおけるヒトとマウスでの分岐以前に、 α グロブリンと β グロブリンは遺伝子重複で分岐したパラログな関係にある。したがって、選択肢 4 が不適切。(2011 年 6 月 7 日改訂)

問 70 正解【1】

[イ]に対応する選択肢の中で、距離行列が必要となる方法は、NJ 法だけであるので、選択肢は 1 か 2 しかありえない。OUT が 4 つの時のトポロジーは、次のようになる。



新たに枝を付加できる辺が 5 つある。問題文にあるように、葉の部分にラベルがついているとすると、4 生物種の系統樹は 3 通りなので、 $3 \times 5 = 15$ で、選択肢 1 が正解となる。

問 71 正解【3】

ニューウィック形式では、系統樹上隣接するペアが括弧内にくられる。選択肢 3 では、隣接していない c と d が括弧でくられているので、これが不適切なものである。

問 72 正解【1】

問題文に二項確率とあるので、(a)は選択肢 1 か 2 しかありえない。(b)は最尤推定についての問いであり、対数尤度を 1 階微分したものを 0 とおけば良いので選択肢 1 と 2 では 1 が選ばれる。この時点で正解にたどりつくが、(b)の式に問題文の p を代入して解けば、(c)が導かれる。

問 73 正解【3】

二色法マイクロアレイは、コントロールを含めて 2 サンプルの測定である。

問 74 正解【3】

次世代シーケンサを用いた解析では、未知の遺伝子も検出できる。

問 75 正解【4】

リアルタイム PCR における Ct 値は PCR で増幅された標的配列の量が一定値に到達したサイクル数のことで、もともとの標的濃度の対数值に対して傾きが負の直線関係を示す。標的配列の濃度が高いほど PCR に要するサイクル数は少なくて済むことに注意。

問 76 正解【3】

モノクローナル抗体は特定のエピトープを認識する。エピトープが化学的に修飾されたり、スプライシングやプロテアーゼ消化によって失われるとモノクローナル抗体は標的タンパク質を認識できなくなる。

問 77 正解【4】

遺伝子発現調節を表現する典型的数式で、分子レベルのメカニズムから導出できる。最大転写速度やリプレッサーが転写速度に与える影響を簡潔に表現している。アクチベータによる転写速度式が類似の形式で表現できる。

問 78 正解【1】

細胞増殖速度を表現する基礎的微分方程式に関する問題である。2 倍増期間(ダブリングタイム)は、2 分の 1 減衰期間(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る。

問 79 正解【3】

ダイナミックモデルにおける固有値は、システムに摂動を与えたとき、どれだけすばやく元の状態にもどれるかを示す尺度であり、システムのロバストネスを評価する指標のひとつである。固有値から、そのシステムの挙動(安定／不安定)を判定することができる。

問 80 正解【3】

フィードフォワードループは、生物回路中のいたるところに出現するネットワークモチーフである。フィードフォワードループの数学モデルを見て、その動的挙動の特徴を理解することが大切である。この問題では、応答遅延による短パルス信号の除去機能を示している。

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。