

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

試験問題

注意事項

1. 試験開始の合図があるまで、問題冊子を開いてはいけません。
2. 問題数は、80問です。すべての問題に解答してください。
3. 答案用紙への受験番号などの記入は、試験監督の指示があつてから始めてください。
4. 試験時間は、13：30～15：30（2時間）です。
5. 試験中にトイレに行きたくなったり気分が悪くなったりした場合は、手を挙げて試験監督に合図してください。
6. 退出可能時間は、14：30～15：20です。
途中で退出する場合には、手を挙げて試験監督に合図をし、解答用紙が回収されてから静かに退出してください。この場合、試験終了まで再入場はできません。
7. 試験問題の内容に関する質問にはお答えできません。文意通りに解釈してください。
8. 解答用紙（マークシート）の記入に当たっては、次の指示に従ってください。
 - (1) 問80までの解答欄を利用してください。問81以降への記入は、無効です。
 - (2) HBかBの黒鉛筆もしくはシャープペンシルを使用してください。
 - (3) 受験番号は、受験票に印字されている通りに記入し、マークしてください。
 - (4) 氏名とフリガナは、受験票に印字されている通りに記入してください。
 - (5) 年月日欄には、生年月日を記入してください（年は、西暦で記入）。
 - (6) 解答は、各解答記入欄に一つずつマークしてください。
 - (7) 訂正の場合は、あとが残らないように消しゴムできれいに消し、消しきずを残さないでください。
9. 試験終了後、この問題用紙は持ち帰ることができます。
10. 解答用紙（マークシート）は、白紙であっても必ず提出してください。

問 1

ゲノムサイズが小さい生物から大きい生物への並びとして、もっとも適切なものを選択肢の中から一つ選べ。

- 1 大腸菌 < ショウジョウバエ < コムギ < ヒト
- 2 大腸菌 < ショウジョウバエ < ヒト < コムギ
- 3 大腸菌 < コムギ < ショウジョウバエ < ヒト
- 4 ヒト < ショウジョウバエ < コムギ < 大腸菌

問 2

遺伝子地図に関する以下の文章の（ a ）～（ d ）に入る語句の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

遺伝子地図は、大きく分けて 2 種類ある。1 つは、組換え率から得られた遺伝子/DNA マーカー間の距離を基に作成した連鎖（遺伝）地図である。もう 1 つは制限酵素切断部位や遺伝子/DNA マーカー間の距離を塩基対数などで表した（ a ）地図である。連鎖地図のマーカー間の距離は cM (センチモルガン) という単位で表され、1 cM は（ b ）% の確率で組換えが起こるマーカー間の距離である。

DNA マーカーの種類には、一塩基多型 (SNP)、制限酵素断片長多型 (RFLP) のほか、数塩基の単位の繰り返し配列である（ c ）などが含まれる。

近年は、次世代シークエンサの圧倒的な塩基配列解読量を活かしたホールゲノム（ d ）法を用いることで、遺伝子地図作成を行わずゲノム解析を行う例が見られるようになった。

- | | | | |
|----------------|---------|---------------|-----------------|
| 1 (a) DNA マーカー | (b) 1 | (c) マイクロアレイ | (d) PCR |
| 2 (a) 物理 | (b) 0.1 | (c) マイクロサテライト | (d) ハイブリダイゼーション |
| 3 (a) DNA マーカー | (b) 100 | (c) マイクロアレイ | (d) ショットガン |
| 4 (a) 物理 | (b) 1 | (c) マイクロサテライト | (d) ショットガン |

問 3

遺伝子重複に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ホモロガスな遺伝子（ホモログ）のうち、異なる生物のゲノム上にあって、共通祖先に由来する遺伝子をパラログといい、同じ生物のゲノム上で遺伝子重複によって生じたものをオーソログという。
- 2 β -グロビンや Hox 遺伝子ファミリーは、ゲノム上で縦列に重複したクラスタを構成していることが、さまざまな生物で報告されている。
- 3 遺伝子重複が起こる要因の一つに、レトロトランスポゾンの転移がある。
- 4 ゲノムの倍数化により、個体の大きさや個々の器官の大きさが増大する現象が知られており、この性質を利用して植物や魚類を大型化する育種法を倍数性育種という。

問 4

真核生物と真正細菌に見られる特徴として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 選択的スプライシングは真核生物で起こる現象であり、真正細菌では起こらない。
- 2 トランスポゾンは真核生物のゲノムには存在するが、真正細菌のゲノムには見られない。
- 3 真核生物のゲノムサイズと遺伝子の数には明確な比例関係はないが、真正細菌のゲノムサイズは遺伝子の数に比例する傾向にある。
- 4 真核生物の転写直後の mRNA の 3' 末端にはアデニンの長い繰り返し配列が付加されるが、真正細菌の転写直後の mRNA の 3' 末端には付加されない。

問 5

DNA の複製と転写に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 DNA の複製にはプライマーが必要だが、転写には必要ない。
- 2 DNA の複製において不連続に合成される短い DNA 鎖を岡崎フラグメントといい、複数の岡崎フラグメントは DNA リガーゼによって結合される。
- 3 DNA のセンス鎖がアミノ酸配列情報をもっているため、センス鎖を鋳型として mRNA が合成（転写）される。
- 4 DNA ポリメラーゼと RNA ポリメラーゼは、いずれもヌクレオチドを $5' \rightarrow 3'$ の方向に重合する酵素である。

問 6

20,000 塩基対で構成された完全に相補的な二本鎖 DNAにおいて、シトシンの含有量が 20% だとすると、この DNA 中にアデニンは何塩基存在するか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 4,000 塩基
- 2 6,000 塩基
- 3 12,000 塩基
- 4 16,000 塩基

問 7

1 文字コードで表したアミノ酸 D, E, I, K, N, Q, R, V を、中性 pH 条件下の側鎖の性質により塩基性・酸性・疎水性・(電荷をもたない) 親水性の 4 種に分類したい。もっとも適切なものを選択肢の中から一つ選べ。

- | | | | |
|--------------|-----------|------------|------------|
| 1 塩基性 [K, R] | 酸性 [N, Q] | 疎水性 [I, V] | 親水性 [D, E] |
| 2 塩基性 [D, E] | 酸性 [K, R] | 疎水性 [N, Q] | 親水性 [I, V] |
| 3 塩基性 [K, R] | 酸性 [D, E] | 疎水性 [I, V] | 親水性 [N, Q] |
| 4 塩基性 [D, E] | 酸性 [N, R] | 疎水性 [K, Q] | 親水性 [I, V] |

問 8

ヌクレオソームに関する以下の文章中の (a) および (b) に当てはまる語句の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

真核細胞の核内に存在するクロマチンは、(a) というタンパク質と (b)との複合体で、糸状の構造を形成している。(a) に (b) が巻きついた 1 単位をヌクレオソームという。

- 1 (a) ミオシン (b) DNA
- 2 (a) ミオシン (b) RNA
- 3 (a) ヒストン (b) DNA
- 4 (a) ヒストン (b) RNA

問 9

ヒトの生体膜に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 細胞膜における物質の透過はすべて受動輸送である。
- 2 生体膜を構成する脂質には多価不飽和脂肪酸が含まれる。
- 3 生体膜の基本構造は、リン脂質、糖脂質、コレステロール等で構成される脂質二重層である。
- 4 ミトコンドリア内膜には、水素イオンの濃度勾配を利用して ATP を合成する酵素が存在する。

問 10

以下に示す細胞内小器官の組み合わせのうち、どちらも独自のゲノムをもたない組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

- 1 リソソーム、葉緑体
- 2 核、液胞
- 3 ミトコンドリア、エンドソーム
- 4 ゴルジ体、小胞体

問 11

ミトコンドリアに関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 真核生物の細胞内小器官であり、外膜と内膜の二つの脂質二重層膜から構成される。
- 2 好気性細菌が真核細胞に共生することによって獲得されたと考えられている。
- 3 ミトコンドリアは父親から子に受け継がれる。
- 4 脂肪酸の β 酸化や、酸化的リン酸化によるエネルギー生産を行う細胞小器官である。

問 12

ATP に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 生物体で用いられるエネルギー保存および利用に関与するヌクレオチドである。
- 2 ATP を ADP とリン酸に加水分解すると、約 11~13 kcal/mol のエネルギーを放出する。
- 3 酸化的リン酸化や光リン酸化の過程で合成される。
- 4 発酵の過程では生成されない。

問 13

タンパク質の翻訳後修飾に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 *N*—グリコシル化はアスパラギン酸残基への糖鎖付加によって起こる。
- 2 *O*—グリコシル化はセリン残基やスレオニン残基への糖鎖付加によって起こる。
- 3 リン酸化はセリン残基、スレオニン残基、チロシン残基、およびヒスチジン残基へのリン酸の付加によって起こる。
- 4 メチル化はアルギニン残基やリジン残基へのメチル基の付加によって起こる。

問 14

核酸が相補的に結合することを利用した技術として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 RNA 干渉
- 2 ウエスタンブロッティング
- 3 サザンブロッティング
- 4 ノーザンブロッティング

問 15

マウスの肝臓に特異的に発現している遺伝子の情報を得る目的に適した方法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 DNA フットプリント法
- 2 マイクロアレイ法
- 3 RT-PCR 法
- 4 ノーザンブロッティング

問 16

PCR 法に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 PCR とは、Polymerase Chain Reaction の略である。
- 2 DNA シークエンシングの際にも利用される。
- 3 反応液の温度を変化させながら、変性、アニーリング、翻訳を繰り返すことによって DNA 断片は増幅される。
- 4 PCR 法で用いられる酵素 Taq ポリメラーゼは、熱水噴出孔に生息している好熱細菌 (*Thermus aquaticus*) から同定された。

問 17

酵素の種類の英語名と、その酵素に直接関連する生化学反応の組み合わせとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 kinase — 糖鎖修飾
- 2 phosphatase — 脱リン酸化
- 3 endonuclease — 核酸の切断
- 4 protease — ペプチドの分解

問 18

ヒトにおける神経細胞の興奮と筋細胞の収縮に密接に関わるイオンの組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

- 1 H⁺ Na⁺ Cl⁻
- 2 Mg²⁺ K⁺ Ca²⁺
- 3 Na⁺ K⁺ Cl⁻
- 4 Na⁺ K⁺ Ca²⁺

問 19

プロテオーム解析に用いられる二次元電気泳動法では、タンパク質を二つのパラメータで分析することができる。二つのパラメータの組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

- 1 Tm 値、等電点
- 2 分子量、等電点
- 3 分子量、疎水性
- 4 Tm 値、疎水性

問 20

タンパク質の折りたたみ（フォールディング）を助ける分子として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 オペロン
- 2 シャペロン
- 3 インターフェロン
- 4 イントロン

問 21

プログラミング言語の実行形式による分類に関する以下の説明において、(a)に入る用語として、もっとも適切なものを選択肢の中から一つ選べ。

バイオインフォマティクスで頻繁に利用される Perl, Python, Ruby などの言語は、作成したプログラム（ソースコード）をコンピュータが理解できる機械語にあらかじめ変換せず、(a)というプログラムがソースコード（または、中間言語）を逐次的に解釈しながら実行を行う。そのため、このようなプログラミング言語は (a) 型言語ともいう。

- 1 CPU
- 2 メモリ
- 3 コンパイラ
- 4 インタプリタ

問 22

0.0 以上 1.0 以下の浮動小数 x に対し以下の 4 つの計算を行う。この中で計算結果が他の 3 つと異なる可能性があるものを選択肢の中から一つ選べ。ただし、 $<< 1$ は左に 1 ビットシフトを行う演算とする。また、int 型への型変換は、C 言語の仕様に即して行われるものとする。

- 1 $(\text{int})x + (\text{int})x$
- 2 $((\text{int})x) << 1$
- 3 $(\text{int})(x * 2.0)$
- 4 $2 * (\text{int})x$

問 23

コンピュータネットワークに関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 TCP/IP はリンク、インターネット、トランスポート、アプリケーションの 4 つのレイヤーから構成される。
- 2 TCP/IP においては通信速度が厳格に規定されているため、いかなる通信機器を用いたとしても任意の 2 点間の通信速度はすべて同じである。
- 3 DNS は IP アドレスとホスト名の対応付けを行うためのシステムであり、そのデータはインターネット上に多く存在する DNS サーバに分散して格納されている。
- 4 FTP はインターネット初期の頃より存在するファイル転送のためのプロトコルであるが、通常はデータが平文で流れるため、セキュリティ上問題がある。

問 24

コンピュータネットワークの保守ツールに関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ping は、属するネットワーク上で DHCP が機能しているかどうかをチェックするツールである。
- 2 traceroute（もしくは tracert）は、IP ネットワーク上で目的のホストまでのゲートウェイを一覧するためのツールである。
- 3 nslookup（もしくは dig）は、ドメイン名から IP アドレス、または IP アドレスからドメイン名を DNS サーバに問い合わせて調べるツールである。
- 4 ifconfig（もしくは ipconfig）は、自分の機器の MAC アドレスや割り当てられた IP アドレスなどの TCP/IP 関連の情報を調べるツールである。

問 25

近年のコンピュータシステムでは、システム上に物理的に実装されているメモリ量を超えるような大規模なプログラムが実行可能である。これを実現する技術として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 仮想記憶
- 2 分岐予測
- 3 アウト・オブ・オーダー実行
- 4 GPGPU

問 26

以下の基数に関する記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ある自然数を 2 進数で表した場合に必要な桁数は、10 進数で表した場合に必要な桁数以上である。
- 2 自然数 x の 2 乗 x^2 を 2 進数で表現するのに必要な桁数は、 x を 2 進数で表現するのに必要な桁数の 2 倍以下である。
- 3 N 桁の 2 進数で表現される自然数 x に対し、 $x + y = 2^N$ となる y を x に関する 2 の補数という。ここで、 2^N の下位 N 桁はすべて 0 であることから、 y を $-x$ の表現とすることで、減算を加算で表現できる。
- 4 2 進数 100110101000_2 は 10 進数 16_{10} で割り切ることができる。

問 27

以下の論理演算に関する記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 論理和 (OR) 演算は、論理積 (AND) 演算と論理否定 (NOT) 演算の組み合わせで表現することができる。
- 2 論理積 (AND) 演算は、論理和 (OR) 演算と論理否定 (NOT) 演算の組み合わせで表現することができる。
- 3 論理否定 (NOT) 演算は、論理和 (OR) 演算と論理積 (AND) 演算の組み合わせで表現することができる。
- 4 任意の論理演算は、否定論理積 (NAND) 演算のみを用いて表現することができる。

問 28

与えられた数値列を並べ替えるソートアルゴリズムの一つに選択法（セレクションソート）がある。選択法によって数値を昇順にソートするには、「与えられたデータの中から最小値を取り出す。次に、それを除いたデータの中から最小値を取り出す」という操作をデータが無くなるまで繰り返す。この選択法に必要な比較演算の回数に関する説明として、もっとも適切なものを選択肢の中から一つ選べ。ただし、与えられた数値列の長さを N とする。

- 1 データの内容に依存せず、およそ N の 2 乗に比例する。
- 2 データの内容に依存せず、およそ N の 3 乗に比例する。
- 3 データの内容によって異なるが、平均的にはおよそ N の 2 乗に比例する。
- 4 データの内容によって異なるが、平均的にはおよそ N の 3 乗に比例する。

問 29

以下のグラフに関する記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 一般に深さ優先探索は、幅優先探索よりも少ないメモリでグラフを探索できる。
- 2 グラフ上のある 1 点から別の 1 点へ到達可能であるかどうかは、深さ優先探索、幅優先探索のいずれの探索法を用いても調べることができる。
- 3 すべての頂点の次数が偶数であるグラフにおいて、グラフ上のすべての辺をただ 1 度通るようなパス（オイラー閉路）は、線形時間アルゴリズムで発見できる。
- 4 一般にグラフ上の 2 点間の最短パスは、幅優先探索に比べて、深さ優先探索の方が高速に求められる。

問 30

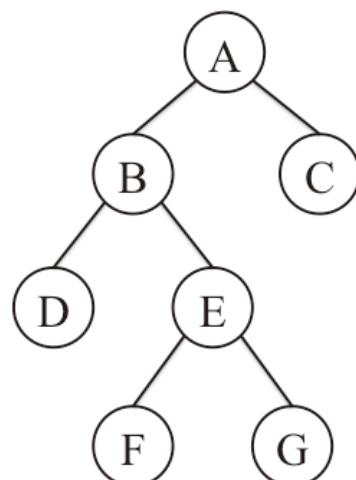
ハッシュ表は、キーと値の組を多数格納したデータ構造であり、与えられたキーと対応づけられた値を高速に検索することを目的としている。以下のハッシュ表に関する記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ハッシュ表は、多くの場合は配列として実装され、その配列の添え字（ハッシュ値）がキーから計算される。ハッシュ値の計算にはハッシュ関数が用いられる。
- 2 異なるキーが同じハッシュ値となることをハッシュの衝突という。衝突が起きた場合には後から格納しようとするキーと値は、隣接する別の格納場所に格納したり、配列中の別の格納場所に格納したりする必要が生じる。
- 3 ハッシュ関数では、線形性が成り立つことが必要である。すなわちハッシュ関数 $h(x)$ において、任意のキー a, b に対して $h(a)+h(b) = h(a+b)$ の関係が成り立つ。
- 4 起こりうるキーの集合を予め知り尽くしており、さらに配列のサイズがその集合のサイズ以上の場合には、衝突の起こらない完全ハッシュ関数を求めることができる。

問 31

以下の疑似コードは、木を巡回し、各節点 v を $\text{output}(v)$ で出力するプログラムを表す。右下の図に示した木に対して $\text{traverse}(A)$ を実行した際の出力として、もっとも適切なものを選択肢の中から一つ選べ。ただし $\text{children}(v)$ は、節点 v の子の節点を左から順番に返す関数とする。

```
traverse(node v):
begin
  foreach  $c \in \text{children}(v)$  do:
    traverse( $c$ );
  end
  output( $v$ );
end
```



- 1 A, B, C, D, E, F, G
- 2 A, B, D, E, F, G, C
- 3 D, F, G, E, B, C, A
- 4 D, B, F, E, G, A, C

問 32

以下の表:学籍簿に対して下記の SQL 文を実行する。結果として得られるデータとして、もっとも適切なものを選択肢の中から一つ選べ。

```
SELECT 学籍番号 FROM 学籍簿 WHERE 研究室番号 > 300 AND 評価点 <= 82 ;
```

表:学籍簿

学籍番号	名前	評価点	研究室番号
1000	伊藤かおり	71	302
1001	坂本京子	83	301
1002	田中正	77	202
1003	藤堂新一	82	301
1004	山田太郎	85	401

- 1 1000, 1003
- 2 1001, 1003
- 3 1000
- 4 1003

問 33

母平均 μ および母分散 σ^2 を持つ確率分布 p から、独立試行によって得られた標本観測値の標本平均 x についての記述として、もっとも不適切なものを選択肢の中から一つ選べ。

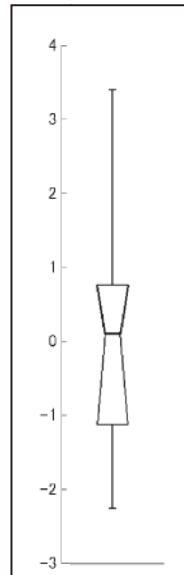
- 1 中心極限定理により、 $x-\mu$ は漸近的に標準正規分布 $N(0,1)$ に近づく。
- 2 大数の法則により、標本観測値の個数が多くなると、標本平均 x は漸近的に μ に近づく。
- 3 中心極限定理は確率分布 p がカイ二乗分布のような左右非対称な分布でも成立する。
- 4 中心極限定理は確率分布 p が 2 項分布のような離散的な分布でも成立する。

問 34

以下の（ a ）～（ c ）に入る語句の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

右図に表わされる箱ひげ図（boxplot）を考える。この箱ひげ図から、このデータは（ a ）が 0.124 であり、下から 25% の値を示す（ b ）が -1.128 であることがわかる。また、ひげが長く伸びていることから、このデータの分布は（ c ）と推定される。

- 1 (a) 平均値 (b) 第 3 四分位点 (c) 補が広い
- 2 (a) 平均値 (b) 第 1 四分位点 (c) くぼみがある
- 3 (a) 中央値 (b) 第 3 四分位点 (c) くぼみがある
- 4 (a) 中央値 (b) 第 1 四分位点 (c) 補が広い



問 35

正規分布に従う二つの確率変数 X および Y を考える。 X と Y の共分散に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 X と Y の共分散の符号は、ピアソン相関係数の符号に一致する。
- 2 X と Y の共分散が負ならば統計的に独立である。
- 3 期待値演算子を E で表すと、共分散は $E(XY) - E(X)E(Y)$ で表される。
- 4 分散共分散行列の対角成分は分散の値、非対角成分は共分散の値である。

問 36

以下の統計に関する説明文の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 尤度関数を最大にするようにパラメータを決めることを最尤推定という。
- 2 F 分布においても t 分布においても、累積密度関数は常に 1 以下の値をとる。
- 3 カルバック・ライブラー情報量は相対エントロピーとも呼ばれる。
- 4 コイン投げの試行をモデル化するとき、0 か 1 の値をとるベルヌーイ分布がしばしば用いられる。ベルヌーイ分布の確率変数の平均値は 0 未満の値をとることがある。

問 37

立方体のサイコロを投げる試行を繰り返す。どの目も $1/6$ の確率で出るとし、各試行は独立であるとする。この試行に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 3 回サイコロを振ってサイコロの目の和を計測したとき、その和の期待値は 10.5 である。
- 2 3 回サイコロを振って少なくとも 1 回 1 が出る確率は $\left(\frac{5}{6}\right)^3$ である。
- 3 6 回サイコロを振って 6 回とも 1 が出る確率と、すべての目が昇順で出る（1, 2, 3, 4, 5, 6 が出る）確率は等しい。
- 4 3 回サイコロを振って同じ目が 2 回以上出る確率は、 $\frac{4}{9}$ である。

問 38

陽性か陰性かの二値分類の予測を行うとき、その予測性能の評価として ROC (Receiver Operating Characteristic) 曲線を使う方法がある。この方法に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ROC 曲線とは、予測器のパラメータを媒介変数とし、False Positive (偽陽性) と True Negative (真陰性) の割合の変化をプロットしたものである。
- 2 ROC 曲線下の面積は AUC (Area Under the Curve) と呼ばれ、AUC = 1 であれば分類器がデータを完全に正しく分類ができる事を示し、AUC = 0 であれば分類器はランダムな仮説と同等であると考えることができる。
- 3 True Negative の割合とは、予測が陰性のもののうち、正解であるものの割合を示す。
- 4 False Positive の割合とは、予測が陽性のもののうち、不正解であるものの割合を示す。

問 39

k -最近傍法とは、訓練用データから距離がもっとも近い k 個を選び、多数決で予測結果を決めるパターン認識法である。10 個のデータ $i=1, 2, \dots, 10$ に対して、各データには 4 クラス (A, B, C, D) のうちいずれかのクラスラベル Y_1, Y_2, \dots, Y_{10} が与えられているとする。今クラス未知のデータ X のクラスの予測に k -最近傍法を用いるために、各データとの距離 $d(X, i)$ を計算した。10 個のデータのクラスラベル Y_i と、10 個のデータとの距離 $d(X, i)$ が下の表で与えられ、近傍数を $k=3$ としたとき、 k -最近傍法の予測結果として、適切なものを選択肢の中から一つ選べ。

標本番号 i	1	2	3	4	5	6	7	8	9	10
クラスラベル Y_i	A	A	B	B	B	C	C	D	D	D
距離 $d(X, i)$	1.2	3.7	0.7	1.3	2.9	5.2	9.2	3.1	6.1	8.7

- 1 A
- 2 B
- 3 C
- 4 D

問 40

回帰分析や学習アルゴリズムを用いてサンプルデータからモデルを構築したとき、そのモデルの良さを確認するために交差検証法 (cross-validation) が行われる。交差検証法に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 交差検証法の目的は、未知の入力データに対するモデルの汎化誤差を推定して、モデルの頑健さを検査することである。
- 2 サンプルデータを n 個の部分に均等に分け、そこから順に 1 つをテスト用に確保し、残り $n-1$ 個をモデルの作成に用いる方法を n -fold 法という。
- 3 n -fold 法における n の値が $n=2$ であるときは、毎回半分のデータをテスト用に、残りの半分を学習用に用いることになるが、これを特に Red-Black tree 法という。
- 4 n -fold 法における n の値がサンプルデータの個数と等しいときは、毎回 1 つのデータだけをテスト用に、残りを学習用に用いることになるが、これを特に leave-one-out 法という。

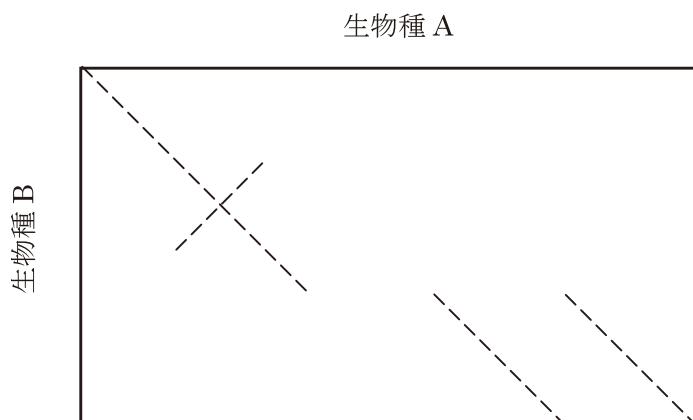
問 41

以下の用語のうち、三つは相互に関連しており、次世代シークエンサから得られる短い配列を参照ゲノム配列上にマップする Bowtie や BWA などのツールにおいて、効率的な索引付けの技術を構成している。この技術において、他との関連性がもっとも低いものを選択肢の中から一つ選べ。

- 1 接尾辞配列 (suffix array)
- 2 バローズ・ホイーラー (Burrows-Wheeler) 変換
- 3 B木 (B-tree)
- 4 FM-index

問 42

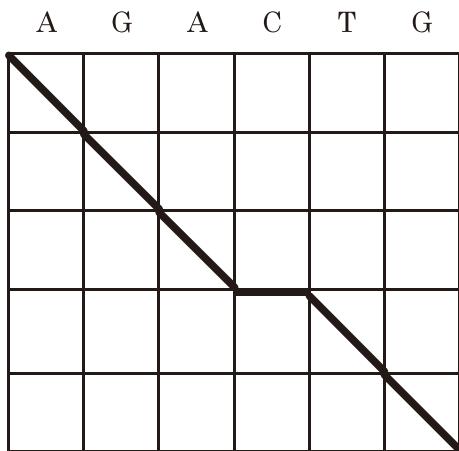
生物種 A, B のゲノムの相同的な領域についてドットプロット解析を行ったところ、以下の図が得られた。図の解釈として、もっとも不適切なものを選択肢の中から一つ選べ。



- 1 いずれかの生物種で逆位が起こっている。
- 2 両生物種にパリンドロームが存在する。
- 3 生物種 A に挿入が起こっているか、または生物種 B に欠失が起こっている。
- 4 生物種 A に重複が起こっている。

問 43

ある配列を配列 AGACTG とペアワイズアラインメントしたところ、最適パスが下図の太線となるような結果が得られた。ただし、一般にプリン塩基間 (A, G) 、およびピリミジン塩基間 (C, T) の置換である転位 (transition) 型の置換は、プリン塩基とピリミジン塩基間の置換である転換 (transversion) 型の置換と比べて起こりやすいため、ここでは後者の方により大きなペナルティが与えられており、さらにギャップにはそれよりも大きなペナルティが与えられているものとする。このとき、アラインメントされた配列として、もっとも不適切なものを選択肢の中から一つ選べ。



- 1 ACATG
- 2 AAATG
- 3 AGGTG
- 4 AGTTG

問 44

生物種 A~E の遺伝子配列から、下表のような距離行列が得られた（対称行列なので上半分に影をつけてある）。この表とともに非加重結合法 (UPGMA) を用いて作成した系統樹 (ニューウィック形式) として、もっとも適切なものを選択肢から一つ選べ。

	A	B	C	D	E
A	0.0	0.2	0.4	0.6	0.4
B	0.2	0.0	0.4	0.6	0.4
C	0.4	0.4	0.0	0.6	0.3
D	0.6	0.6	0.6	0.0	0.6
E	0.4	0.4	0.3	0.6	0.0

- 1 (((A, B), D), (C, E))
- 2 ((A, B), ((C, E), D))
- 3 (((A, B), C), E), D)
- 4 (((A, B), (C, E)), D)

問 45

以下の二つのデータベース RefSeq と GenBank に関する記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 GenBank は、個々の研究室およびゲノムプロジェクトなどから投稿されたデータに基づいて作成されている。
- 2 RefSeq は、GenBank から抽出された配列セットである。
- 3 RefSeq は、精査された重複がない配列セットを提供している。
- 4 RefSeq と GenBank で、対応する配列のアクセッション番号は同じである。

問 46

以下の BLAST の検索結果で得られるビットスコア (bit score) に関する記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ビットスコアが大きくなると、E-value は小さくなる。
- 2 E-value はデータベースとクエリ配列のサイズ、およびビットスコアに依存し、使用したスコア行列の種類には依存しない。
- 3 スコア行列を BLOSUM62 から BLOSUM45 に変更しても、ビットスコアは変わらない。
- 4 複数の異なるデータベースを対象に BLAST 検索を行う場合、同じスコア行列を使用していれば、ビットスコアを使って異なる検索結果を直接比較することができる。

問 47

以下の表は、コドン表に大腸菌 K-12 W3110 株の全遺伝子のコドン使用度数を記述したものである。ここで、相対コドン使用頻度の計算方法を式 (1) とした場合、コドン GCG の相対コドン使用頻度として、もっとも適切な値を選択肢の中から一つ選べ。ただし、計算は四捨五入して小数点第 1 位まで求めるものとする。

2文字目													
		T			C			A			G		
T	Phe	TTT	30462	Ser	TCT	11512	Tyr	TAT	22037	Cys	TGT	7016	T
		TTC	22705		TCC	11802		TAC	16795		TGC	8797	C
C	Leu	TTA	18894	Pro	TCA	9620	終止	TAA	2765	終止	TGA	1249	A
		TTG	18664		TCG	12210	終止	TAG	321	Trp	TGG	20889	G
A	Leu	CTT	15082	Thr	CCT	9540	His	CAT	17791	Arg	CGT	28866	T
		CTC	15272		CCC	7490		CAC	13399		CGC	30530	C
G	Val	CTA	5266	Ala	CCA	11569	Gln	CAA	21121	Arg	CGA	4810	A
		CTG	72898		CCG	32080		CAG	39835		CGG	7401	G
A	Ile	ATT	41644	Asn	ACT	12119	Asn	AAT	24106	Ser	AGT	11924	T
		ATC	34568		ACC	32265		AAC	29581		AGC	22067	C
G	Met	ATA	5733	Lys	ACA	9452		AAA	46116	Arg	AGA	2771	A
		ATG	38167		ACG	19820		AAG	14174		AGG	1496	G
G	Val	GTT	24991	Asp	GCT	20813	Asp	GAT	44217	Gly	GGT	33875	T
		GTC	21050		GCC	35252		GAC	26270		GGC	40849	C
G	Val	GTA	14901	Glu	GCA	27567		GAA	54431		GGA	10774	A
		GTG	36108		GCG	46524		GAG	24629		GGG	15115	G

アミノ酸 a をコードするコドン c ($c \in C(a)$) に対する相対コドン使用頻度 x_c

$$x_c = \frac{f_c}{\frac{1}{|C(a)|} \sum_{i \in C(a)} f_i} \quad (1)$$

ただし、 f_i はコドン i の使用度数、 $C(a)$ はアミノ酸 a をコードするコドンの集合、 $|C(a)|$ は集合 $C(a)$ の要素数を表す。

- 1 0.9
- 2 1.4
- 3 1.1
- 4 0.6

問 48

下記の 2 本のアミノ酸配列のアライメントについて、BLOSUM62 スコア行列（下表の対称行列）を用いてスコアを計算したい。スコアとしてもっとも適切な値を選択肢の中から一つ選べ。

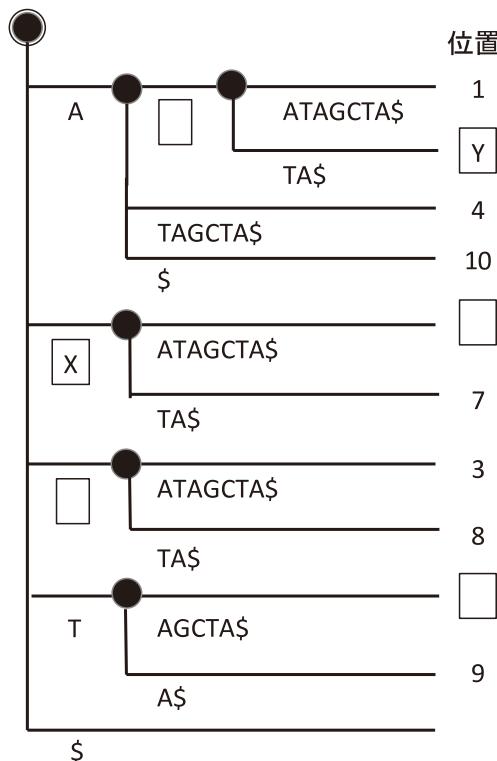
D	V	K	D	G	W
	+	+	+		
D	I	R	E	G	W

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

- 1 23
- 2 27
- 3 30
- 4 35

問 49

接尾辞木 (suffix tree) は、長大な DNA 配列の索引付けに効果的なデータ構造の一つで、文字列上の各位置 i における接尾辞 (i 番目以降の文字列) について、文字列と位置との対応を木構造として格納することにより、高速な検索を可能にする。以下の図は、DNA 配列 AGCATAAGCTA に対して作成された接尾辞木を表している。ただし、枝および葉に割り付けられた情報の一部を伏せてある（図中の四角）。また、\$は文字列の終端を表す記号である。この図から、例えば ATAGCTA が元の配列の位置 4 から始まる接尾辞であることや、文字 T で始まる接尾辞が二つあることなどが木構造をたどることにより検索できる。図の X と Y に入る値の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。



選択肢	X	Y
1	C	6
2	GC	6
3	C	5
4	GC	5

問 50

FASTQ 形式の配列ファイルでは、塩基配列とともに各塩基のクオリティ値が、印字可能なアスキー文字コード（下表参照）を用いて記載されている。以下の FASTQ 形式ファイルから、クオリティ値が 20 以下である塩基を小文字に変換することによって得られる文字列として、もっとも適切なものを選択肢の中から一つ選べ。ただし、クオリティ値はアスキー文字コードから 33 を引くことによって得られるものとする。例えば、! はアスキー文字コードが 33（下表によると、行の数 3 と列の数 30 の和を取って 33）であるため、クオリティ値としては 0 を表す。

FASTQ 形式ファイル

```
@Seq1
AGATCAAGCGAT
+
#09?AFHIJJ+I+
```

印字可能アスキー文字コード表

	30	40	50	60	70	80	90	100	110	120
0		(2	<	F	P	Z	d	n	x
1)	3	=	G	Q	[e	o	y
2		*	4	>	H	R	\	f	p	z
3	!	+	5	?	I	S]	g	q	{
4	"	,	6	@	J	T	^	h	r	
5	#	-	7	A	K	U	_	i	s	}
6	\$.	8	B	L	V	`	j	t	~
7	%	/	9	C	M	W	a	k	u	
8	&	0	:	D	N	X	b	l	v	
9	'	1	;	E	O	Y	c	m	w	

- 1 AGatCAAgcGAT
- 2 agaTCAAGCGAT
- 3 agatcAAGCGAT
- 4 agATCAAGCGAT

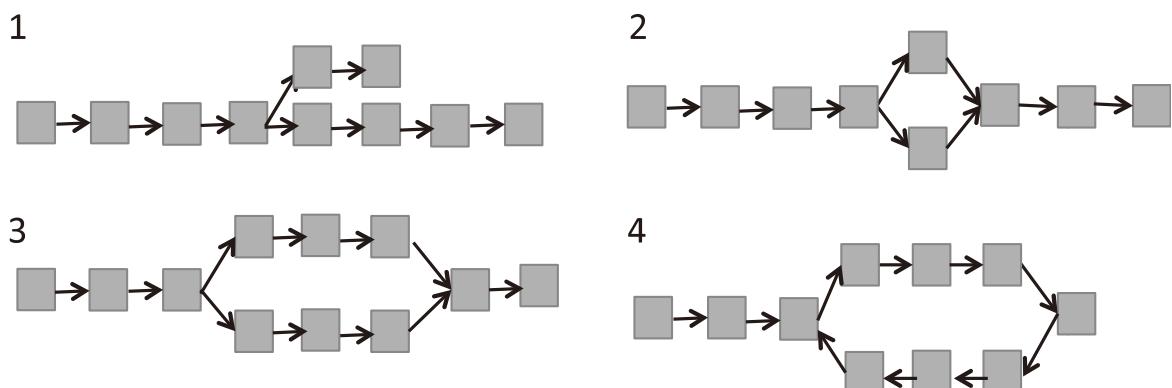
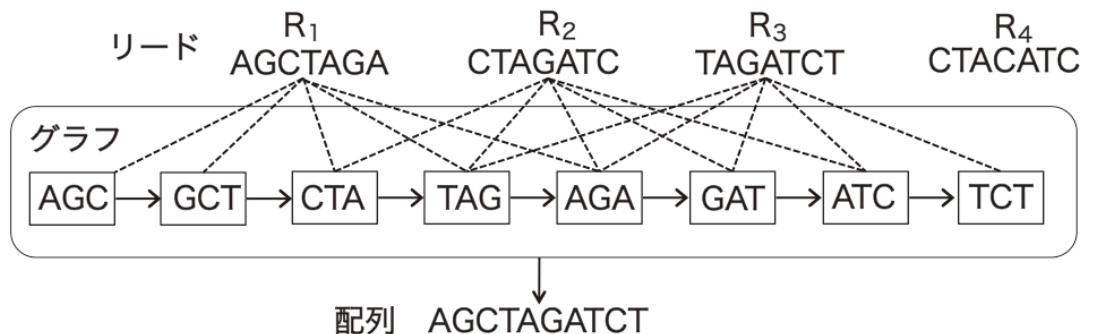
問 51

配列モチーフに関する以下の説明のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 配列モチーフの表現方法としては、照合結果が Yes / No で得られるものとして正規表現や決定木、またスコアや確率値として得られるものとしてプロファイルや位置特異的スコア行列 (PSSM) 、隠れマルコフモデル (HMM) などがある。
- 2 配列モチーフに対する検索プログラムとしては、位置特異的スコア行列を用いた検索を行う MEGABLAST や、隠れマルコフモデルを用いた検索を行う MUMmer などがある。
- 3 配列モチーフを集めたデータベースとしては、隠れマルコフモデルで表現された Pfam や、正規表現やプロファイルで表現された PROSITE などが有名であり、検索の際はそれらを統合したデータベースである InterPro もよく用いられる。
- 4 与えられた配列中に共通に存在する配列モチーフを探索するプログラムとしては、EM アルゴリズムに基づく MEME や、ギブスサンプリング法に基づく Gibbs などがある。

問 52

次世代シークエンサから得られる短いリード配列をアセンブルする際には、ド・ブラン (de Bruijn) グラフを使う方法がよく用いられている。この方法では、各リード配列に含まれる長さ k のワードをすべて抽出し、 $k-1$ の重複をもち、リード中で前後関係にあるワード対をつないだグラフを作成し、そのグラフ上のパスとして配列を再構築する。以下の図は、配列 AGCTAGATCT から得られた 3 つのリード配列、 R_1 : AGCTAGA, R_2 : CTAGATC, R_3 : TAGATCT について、 $k=3$ のド・ブラン・グラフを作成したときの様子を示している。このように直鎖状のグラフが得られる場合は容易にもとの配列を再構築できるが、常にそうなるとは限らない。いま、これら 3 つのリードに加えて、リード R_4 として、 R_2 の中央の塩基にエラーが入った配列 CTACATCA (下線部がエラー) が加わったとする。この場合のグラフの形として、もっとも適切なものを選択肢の中から一つ選べ。



問 53

あるファミリーに属するタンパク質のアミノ酸配列のマルチプルアラインメントの一部を以下の図に示す。これらのタンパク質はすべてカルシウムイオンと結合することができる。このアラインメントに基づいた、部位1、部位2、…、部位6、および図中に枠で囲った領域Aと領域Bの立体構造上の特徴や機能的な役割についての推察として、もっとも不適切なものを選択肢の中から一つ選べ。部位はアラインメント最上段に数字で示し、領域は最下段に示した。

	1	2	3	45	6
CALM_HUMAN :- QLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEA					
CALL3_HUMAN :- QLTEEQVTEFKEAFSLFDKDGDGCI TTRELGTVMRSLGQNPTEA					
CALL5_HUMAN :- ELTPEEEAQYKKAFSAVDTDGNGTINAQELGAALKATGKNLSEA					
CALL6_HUMAN : ERLSAEQUIKEYKGVFEMFDEEGNGEVKTGELEWLMSLLGINPTKS					
TNNC2_HUMAN : SYLSEEMIAEFKAADFMDFADGGGDI SVKELGTVMRMLGQTPTKE					
CETN1_HUMAN : PELTEDQKQEVRREAFDLFDVDGSGTI DAKELKVAMRALGFEPRKE					
CETN2_HUMAN : PELTEEQKQEIRREAFDLFDADGTGTIDVKELKVAMRALGFEPKKE					
領域A		領域B			

- 1 部位 2 と部位 5 では、部位 5 のほうがカルシウムイオン結合部位である可能性が高い。
 - 2 部位 4 と部位 6 では、部位 4 のほうがカルシウムイオン結合部位である可能性が高い。
 - 3 部位 1 と部位 3 では、部位 1 のほうがタンパク質表面に露出している可能性が高い。
 - 4 領域 A と領域 B では、領域 A のほうが α -ヘリックスを形成している可能性が高い。

問 54

立体構造の類似した二つのタンパク質において、対応関係にある原子対がわかっている場合には、RMSD (Root Mean Square Deviation) と呼ばれる数値を計算することにより、立体構造の類似性を数値化することができる。RMSD は、すべての対応する原子間の距離を 2 乗した値の平均値（相加平均）の平方根として定義される。あるタンパク質における 5 つの原子 A,B,C,D,E と別のタンパク質の 5 つの原子 A',B',C',D',E' がこの順序で対応することが分かっており、互いにもっともよく重なるように回転・並進して重ねた時の対応原子間の距離は以下の表のようになつた ($1 \text{ \AA} = 0.1 \text{ nm}$ である)。5 つの原子に対する RMSD 値はいくらになるか。もっとも適切な値を選択肢の中から一つ選べ。

対応する原子対	距離
原子 A – 原子 A'	2.0 Å
原子 B – 原子 B'	1.0 Å
原子 C – 原子 C'	0.0 Å
原子 D – 原子 D'	1.0 Å
原子 E – 原子 E'	2.0 Å

- 1 1.0 Å
- 2 1.4 Å
- 3 1.7 Å
- 4 3.0 Å

問 55

以下の記述について、(a) ~ (d) に入る語句や記号の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

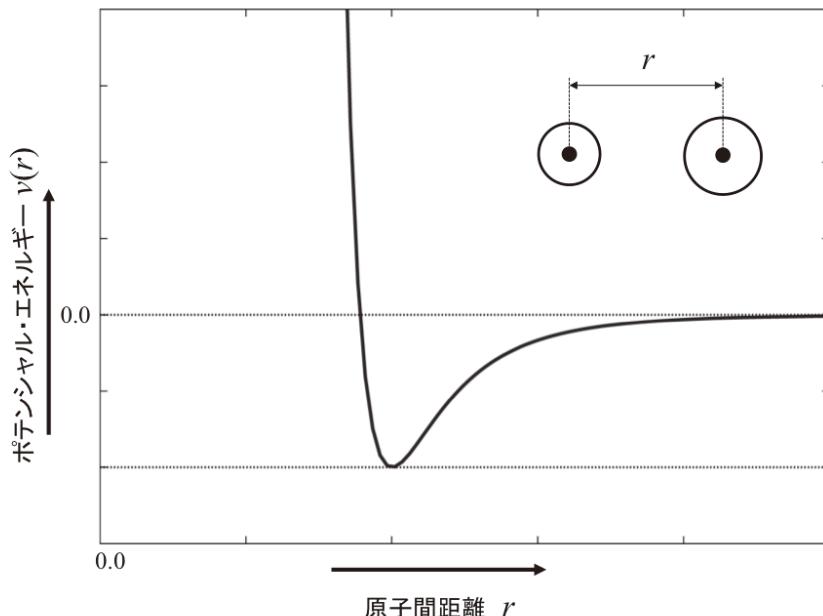
分子動力学法やドッキング計算において、二つの原子間の相互作用を表すために、次式のようなレナード・ジョーンズポテンシャル関数 $v(r)$ が広く使われている。

$$v(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right]$$

r は二つの原子の中心間の距離、 ϵ と σ は二つの原子の種類によって決まる正値のパラメータである。またこの関数は $r_0=2^{1/6}\sigma$ とすると以下のように書き換えられる。 $2^{1/6}$ とは 6 乗すると 2 となる数である。

$$v(r) = \epsilon \left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^6 \right]$$

以下の図にこのポテンシャルの概形を示す。式の中の (a) の項は二つの原子間のゆるやかに引き合う力（引力）を表わし、(b) の項は原子が近づきすぎたときに急激に働く反発力（斥力）を表す。二つの原子の中心間の距離 r が (c) のとき、ポテンシャルは極小値となり、この値より離れれば引力が、近づけば斥力が働く。また、距離 r が (d) のとき、ポテンシャルの値はちょうどゼロとなり、この値より距離が小さい場合にポテンシャルの値は正になる。

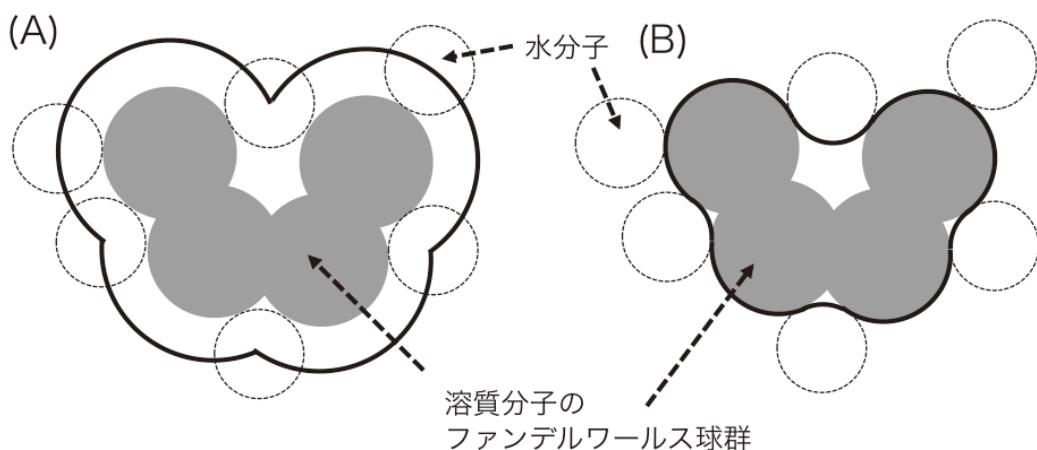


- | | | | | |
|---|----------|----------|--------------|--------------|
| 1 | (a) 12 乗 | (b) 6 乗 | (c) r_0 | (d) σ |
| 2 | (a) 6 乗 | (b) 12 乗 | (c) r_0 | (d) σ |
| 3 | (a) 12 乗 | (b) 6 乗 | (c) σ | (d) r_0 |
| 4 | (a) 6 乗 | (b) 12 乗 | (c) σ | (d) r_0 |

問 56

以下の記述の（ a ）～（ d ）に入る語句の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

溶媒露出表面は、溶質分子のファンデルワールス球群に接するように、水分子の大きさの球をくまなく転がしたときの（ a ）として定義され、下に示した二次元の模式図（ b ）の太線に相当する。その面積は溶媒露出表面積と呼ばれる。溶媒露出表面積に水分子の直径を乗じた値は、溶質分子に接触（水和）できる水分子の総体積の近似値になる。タンパク質が伸びた状態からコンパクトに折りたたまれるフォールディングの過程や、解離した二つの分子が結合するドッキングの過程においては、溶媒露出表面積は（ c ）する。これは、これらの過程の前後において、結合する水分子の数より解離する水分子の数のほうが（ d ）ことを示す。立体構造予測や複合体構造予測では、原子ごとの溶媒露出表面積の変化から水和自由エネルギーを見積もる計算法が広く採用されている。

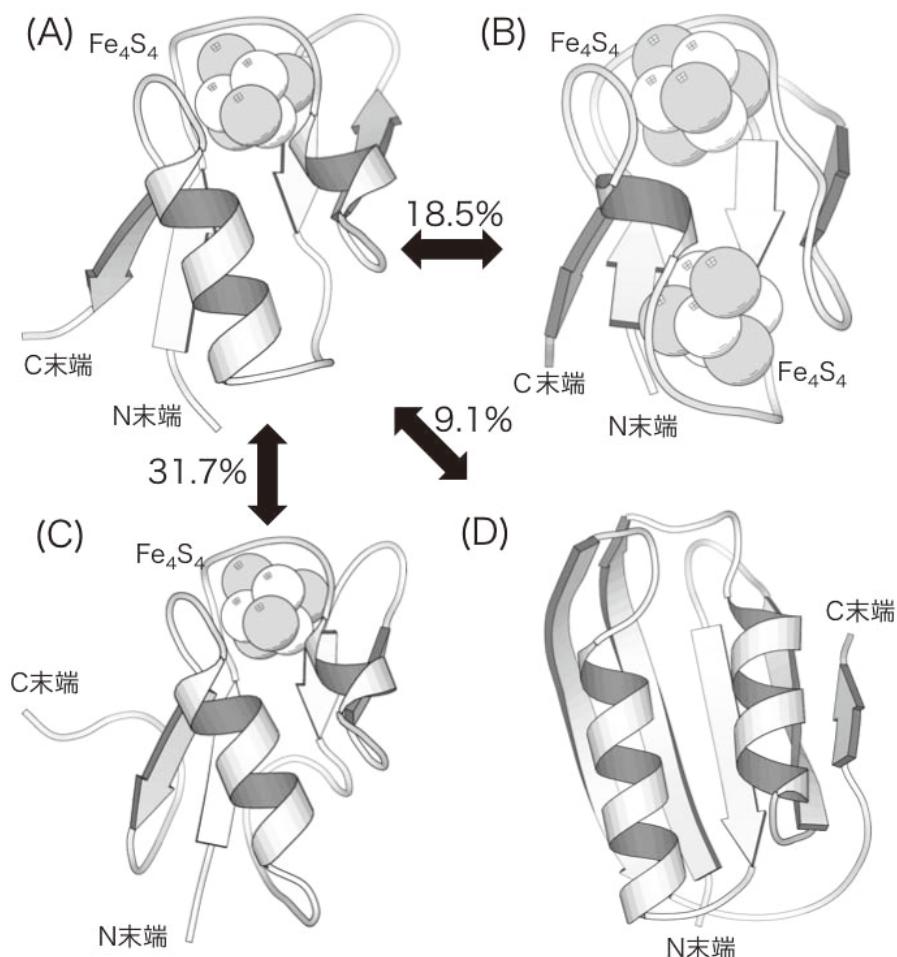


- | | | | |
|-----------------------|-------|--------|---------|
| 1 (a) 球の中心の描く軌跡の面 | (b) A | (c) 減少 | (d) 多い |
| 2 (a) 球の中心の描く軌跡の面 | (b) A | (c) 増加 | (d) 少ない |
| 3 (a) 球が侵入できない領域との境界面 | (b) B | (c) 減少 | (d) 多い |
| 4 (a) 球が侵入できない領域との境界面 | (b) B | (c) 増加 | (d) 少ない |

問 57

以下に立体構造が類似した 4 つのタンパク質 A~D のリボン図を示す。それぞれの立体構造は、立体構造比較プログラムによって、形状が対応するよう向きがそろえである。空間充填モデルで表示された分子はタンパク質に結合している鉄硫黄クラスタ (Fe_4S_4) である。タンパク質 D には Fe_4S_4 は結合できない。矢印の上の数字はアミノ酸配列の一致度を示す。これら 4 つの立体構造の SCOP による立体構造分類について説明した以下の文章の (a) ~ (c) に入る語句の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

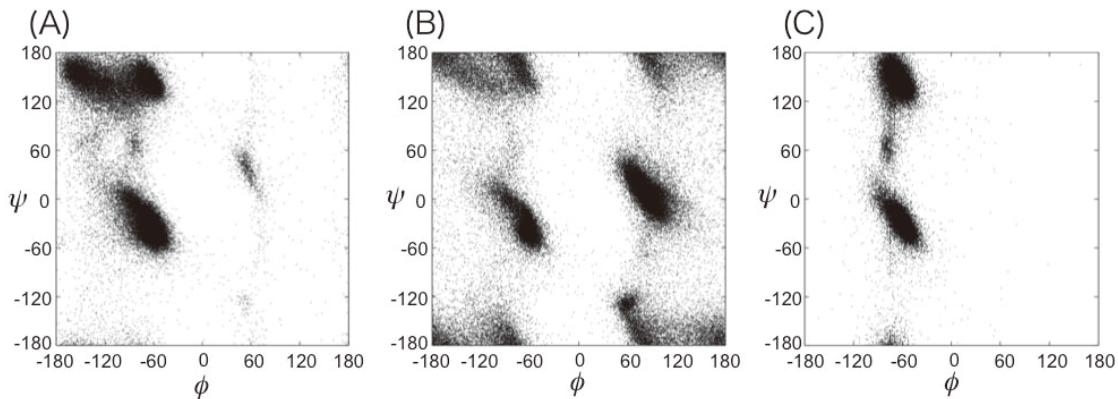
タンパク質 A とタンパク質 C は、同一の (a) に分類される。タンパク質 A とタンパク質 B は (a) は異なるかもしれないが、少なくとも同一の (b) に分類するのが妥当と考えられる。タンパク質 A とタンパク質 D は、同一の (c) であるが、異なる (b) に分類するのが妥当であると考えられる。



- | | | |
|-----------------|---------------|---------------|
| 1 (a) スーパーファミリー | (b) ファミリー | (c) フォールド |
| 2 (a) フォールド | (b) ファミリー | (c) スーパーファミリー |
| 3 (a) フォールド | (b) スーパーファミリー | (c) ファミリー |
| 4 (a) ファミリー | (b) スーパーファミリー | (c) フォールド |

問 58

タンパク質の主鎖の二面角 ϕ と ψ の二次元分布図はラマチャンドランマップと呼ばれる。ラマチャンドランマップは二次構造やアミノ酸ごとに分布に特徴があることから、タンパク質のモデル構造の質の評価などに用いられる。以下の三つのラマチャンドランマップ A, B, C は、アラニン、グリシン、プロリン（順不同）について、タンパク質の代表構造のデータセットからそれぞれのアミノ酸を抽出して作成したものである。アミノ酸とラマチャンドランマップの組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。



- 1 (A) アラニン (B) グリシン (C) プロリン
- 2 (A) グリシン (B) プロリン (C) アラニン
- 3 (A) プロリン (B) グリシン (C) アラニン
- 4 (A) グリシン (B) アラニン (C) プロリン

問 59

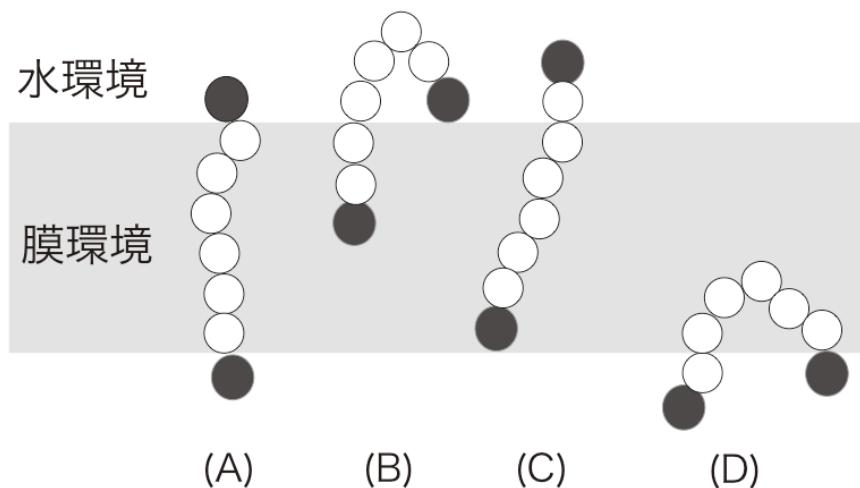
以下の表に、20種のアミノ酸のカイト・ドゥーリトルの疎水性指標、チョウ・ファスマンの α ヘリックス形成能と β シート形成能の値を示した。三つのアミノ酸（a）、（b）、（c）だけ名前が伏せられている。これらは、グリシン、バリン、グルタミン酸のいずれかである。（a）、（b）、（c）に入るアミノ酸の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

	疎水性指標	α ヘリックス形成能	β シート形成能
アラニン	+1.80	+1.45	+0.97
システイン	+2.50	+0.77	+1.30
アスパラギン酸	-3.50	+0.98	+0.80
フェニルアラニン	+2.80	+1.12	+1.28
ヒスチジン	-3.20	+1.24	+0.71
イソロイシン	+4.50	+1.00	+1.60
リジン	-3.90	+1.07	+0.74
ロイシン	+3.80	+1.34	+1.22
メチオニン	+1.90	+1.20	+1.67
アスパラギン	-3.50	+0.73	+0.65
プロリン	-1.60	+0.59	+0.62
グルタミン	-3.50	+1.17	+1.23
アルギニン	-4.50	+0.79	+0.90
セリン	-0.80	+0.79	+0.72
スレオニン	-0.70	+0.82	+1.20
トリプトファン	-0.90	+1.14	+1.19
チロシン	-1.30	+0.61	+1.29
(a)	+4.20	+1.14	+1.65
(b)	-0.40	+0.53	+0.81
(c)	-3.50	+1.53	+0.26

- 1 (a) グリシン (b) バリン (c) グルタミン酸
 2 (a) グルタミン酸 (b) バリン (c) グリシン
 3 (a) バリン (b) グリシン (c) グルタミン酸
 4 (a) グリシン (b) グルタミン酸 (c) バリン

問 60

親水性アミノ酸（下図の黒丸）と疎水性アミノ酸（下図の白丸）で構成された仮想ペプチド A~D が、膜に結合した状態の構造を考える。あるアミノ酸が水環境に存在するときのポテンシャルエネルギー値および、水環境と膜環境に存在する際のポテンシャルエネルギーの差分を下表に示した。他のエネルギーはすべて無視できるとしたとき、A~D の中でポテンシャルエネルギーが 2 番目に低いペプチドを、選択肢の中から一つ選べ。



	親水性アミノ酸	疎水性アミノ酸
水環境中のポテンシャルエネルギー	0 kcal/mol	5 kcal/mol
$\Delta G = G_{\text{水環境}} - G_{\text{膜環境}}$	-10 kcal/mol	5 kcal/mol

- 1 A
- 2 B
- 3 C
- 4 D

問 61

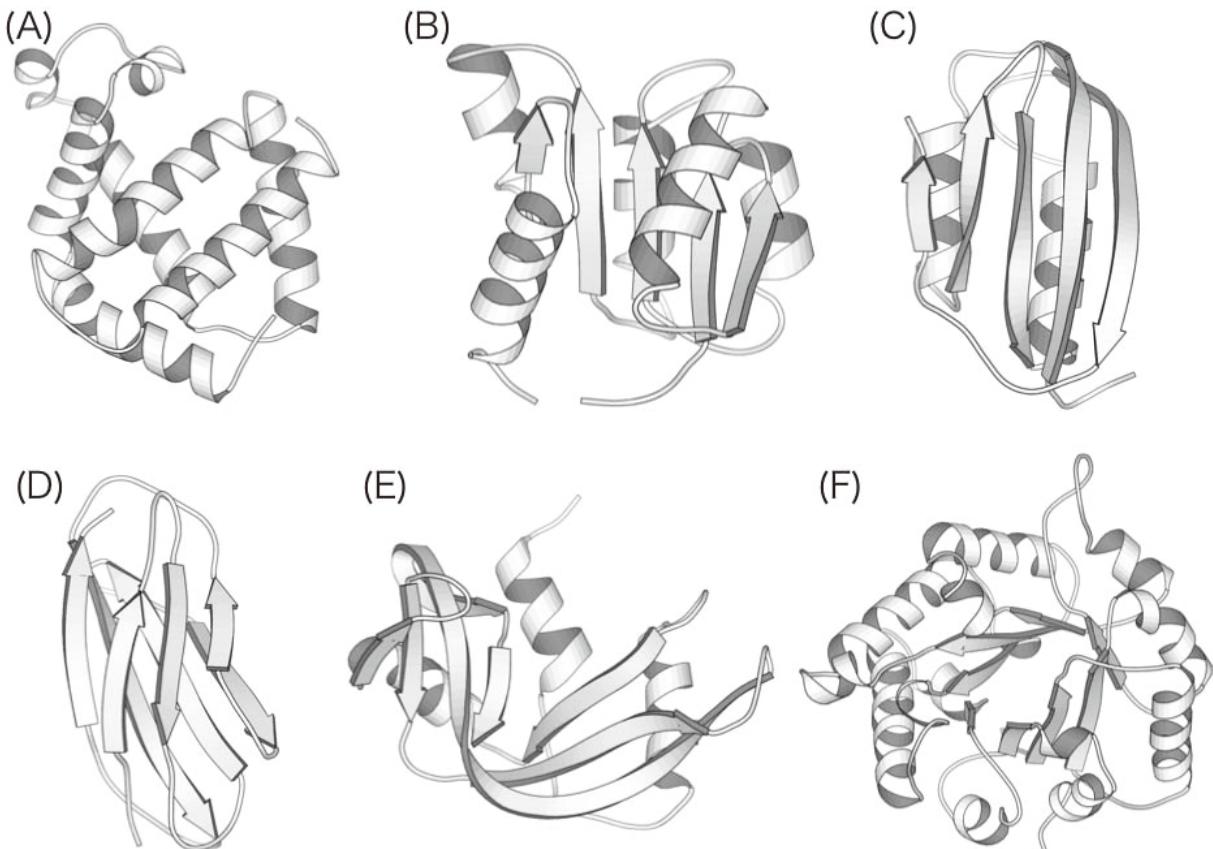
以下のペプチド鎖の性質に関する記述の（ a ）～（ d ）に入る語句の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

タンパク質は、アミノ酸側鎖の種類によってペプチド鎖の構造が制限され、そのペプチド鎖と溶媒水分子の系全体の自由エネルギーが（ a ）なるように折り畳まれる。自由エネルギーは様々な相互作用の総和であり、特に、水に溶けにくい領域と水との接触表面積を最小化させようとする（ b ）相互作用、水素結合、イオン結合、ファンデルワールス力などの（ c ）結合が協同して折りたためた天然構造を安定化させている。この構造内部では、主鎖のアミノ基（NH）とカルボニル基（C=O）のほとんどは（ d ）を作り、 α ヘリックスや β シート構造を形成している。

- | | | | |
|-----------|---------------|---------|-----------|
| 1 (a) 小さく | (b) 疎水性 | (c) 共有 | (d) イオン結合 |
| 2 (a) 小さく | (b) 静電 | (c) 共有 | (d) 水素結合 |
| 3 (a) 大きく | (b) ファンデルワールス | (c) 非共有 | (d) イオン結合 |
| 4 (a) 小さく | (b) 疎水性 | (c) 非共有 | (d) 水素結合 |

問 62

球状タンパク質の立体構造は、all- α 型、all- β 型、 $\alpha+\beta$ 型、 α/β 型の4種類に大きく分類される。以下に立体構造を示した球状タンパク質A~Fが、もっとも適切に分類されている組み合わせを選択肢の中から一つ選べ。



- | | | | | |
|---|---------------------|-----------------------|-------------------------|-------------------------|
| 1 | all- α 型 : A | all- β 型 : D | $\alpha+\beta$ 型 : C, E | α/β 型 : B, F |
| 2 | all- α 型 : A | all- β 型 : D | $\alpha+\beta$ 型 : B, F | α/β 型 : C, E |
| 3 | all- α 型 : A | all- β 型 : C, D | $\alpha+\beta$ 型 : E | α/β 型 : B, F |
| 4 | all- α 型 : C | all- β 型 : A | $\alpha+\beta$ 型 : B, D | α/β 型 : E, F |

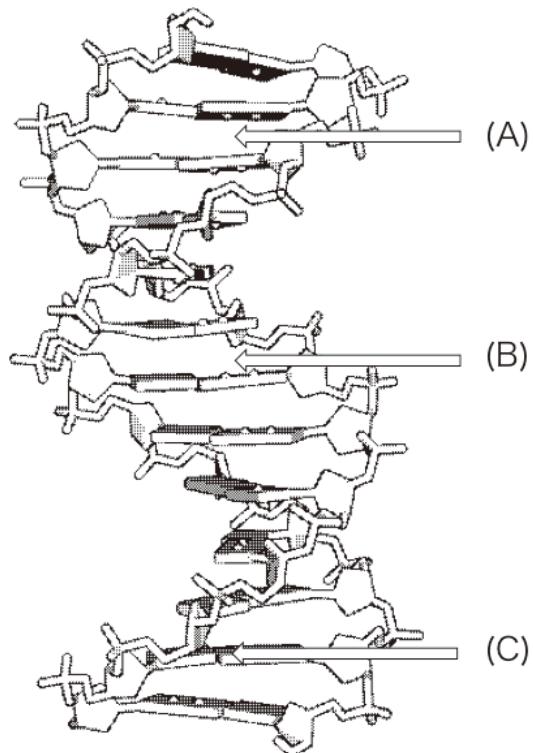
問 63

タンパク質の代表的な立体構造予測法である、ホモロジーモデリング法とフラグメントアセンブリ法の性質に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。標的アミノ酸配列とは、予測対象となるタンパク質のアミノ酸配列のことであり、予測誤差とは正解構造からのずれ(RMSD)を指すものとする。

- 1 ホモロジーモデリング法で予測された立体構造は、既知の立体構造のどれかと全体的に形が似ている。
- 2 ホモロジーモデリング法では、標的アミノ酸配列と鋳型立体構造のアミノ酸配列の類似性が低いほど、予測誤差が増加する傾向がある。
- 3 フラグメントアセンブリ法で予測された立体構造は、全体的な形が既知の立体構造のどれとも似ていないことがある。
- 4 フラグメントアセンブリ法の予測誤差は、標的アミノ酸配列の長さには依存しない。

問 64

以下の図は DNA の立体構造を模式的に示したものである。この図の説明として、もっとも不適切なものを選択肢の中から一つ選べ。



- 1 全部で 13 塩基対の構造が示されている。
- 2 図中 A~C で示された溝は、A と B が主溝 (major groove) 、C が副溝 (minor groove) である。
- 3 B 型構造である。
- 4 右巻き二重らせん構造である。

問 65

一塩基多型に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 一塩基多型の集団内での出現頻度は、通常 1%よりも高い。
- 2 一塩基多型には、表現型に影響を与えるものと与えないものがある。
- 3 一塩基多型の出現頻度は、人種によって異なる。
- 4 一塩基多型には後天的に生じたものも含まれる。

問 66

コピー数多型に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ゲノムの構造を顕微鏡で比較することでは観測できないような、1000 塩基以下の短いゲノム領域での多型である。
- 2 ヒトの集団内に 1%よりも高い頻度で現れる場合、コピー数多型 (Copy Number Polymorphism) という。
- 3 ヒトゲノムのおよそ 1 割以上の領域を占めることが知られている。
- 4 疾患感受性、薬剤感受性などを含めヒトの形質差に関与している。

問 67

遺伝子の水平伝搬に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ヒトゲノムには、RNA ウィルスが約 4000 万年前に取り込まれた形跡がある。
- 2 水平伝搬した遺伝子の分子系統は、種の系統と一致することが多い。
- 3 シロアリのセルロース分解能力の獲得は遺伝子水平伝搬によるものと推測される。
- 4 真核生物より原核生物で頻繁に起こっていると考えられる。

問 68

以下の遺伝に関する文章の（ a ）に入る数値として、もっとも適切なものを選択肢の中から一つ選べ。

ある植物の花の色には赤花と白花があり、赤花は白花に対して優性である。花の色に関する優性遺伝子を R、劣性遺伝子を r とする。また、この植物の種の形には丸型としわ型があり、丸型はしわ型に対して優性である。種の形に関する優性遺伝子を Q、劣性遺伝子を q とする。いま RRQQ と rrqq を両親にもつ子花同士を交配したときに孫花が「白花・丸型」になる確率は（ a ）である。ただし、この植物は二倍体で有性生殖し、花の色と種の形はメンデルの分離・独立・優性の法則にしたがうものとする。

- 1 0
- 2 1/16
- 3 3/16
- 4 1/4

問 69

ある受容体の遺伝子について塩基配列をチンパンジーとヒトで比較した結果を下図に示した。この結果の解釈として、もっとも不適切なものを選択肢の中から一つ選べ。なお図中の「.」は、参照配列と同じ塩基であること意味する。図示した部位以外に変異は認められず、シークエンスエラーはないものとする。

参照配列	TGCCACCTTGGACTCCGCAGC
チンパンジー A	.A.....G.A.T...T.
チンパンジー B	.A.....C...G.G.T...C.
チンパンジー C	.A.....G.A.T...C.
チンパンジー D	.A.....G.A.T...C.
チンパンジー E	.AT.....AG.A.T...C.
チンパンジー F	GA...T.....G.A.T...C.
チンパンジー G	GA.....G.A.T...C.
ヒト AT..A.....T...
ヒト BT..A..T...AT...
ヒト C	...G..T..A.....AT...
ヒト D	...G..T..A....T.AT...
ヒト E	...GG.T..A.....AT...
ヒト F	...G..T..A.....T...
ヒト G	...G..T..A..T...AT...

- 1 チンパンジー A～G の間では、この受容体の遺伝子座で変異が認められるのは 7 塩基である。
- 2 ヒトではこの受容体のハプロタイプが少なくとも 7 種類ある。
- 3 ヒトとチンパンジーの間で完全に保存されている部位は 2 塩基である。
- 4 チンパンジー A とヒト A の配列の間には 8 塩基の違いがある。

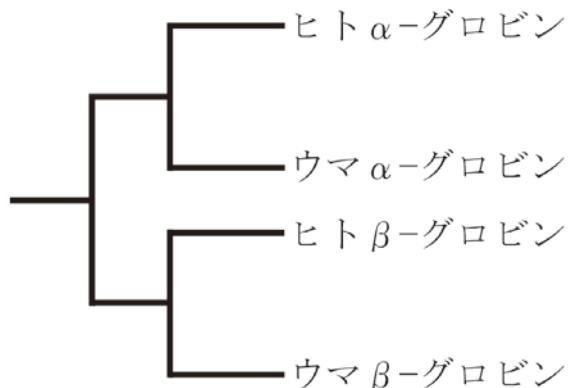
問 70

ヒトのゲノム配列中には、様々な遺伝子多型が存在することが知られている。遺伝子多型の用語として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 CNV
- 2 QTL
- 3 SNP
- 4 VNTR

問 71

下図に示す系統樹は、遺伝子重複によってできた α -グロビン遺伝子と β -グロビン遺伝子の両遺伝子を含む有根系統樹である。この系統樹に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。



- 1 このような遺伝子重複によってできた二つ以上の遺伝子を含む系統樹は、複合系統樹と呼ばれる。
- 2 α -グロビンと β -グロビンの遺伝子重複は、ヒトとウマの種分岐よりも後に起こっている。
- 3 ヒト α -グロビンとウマ α -グロビンは種分岐にともなってできた遺伝子であり、オーソログ（オーソロガスな関係にある遺伝子）と呼ばれる。
- 4 ヒト α -グロビンとウマ β -グロビンは遺伝子重複によってできた遺伝子であり、パラログ（パラロガスな関係にある遺伝子）と呼ばれる。

問 72

さまざまな生物種のゲノム配列が決定された結果、生物種間でゲノムの GC 含量に違いがあることがわかつってきた。この事実に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 タンパク質をコードする遺伝子において、ゲノム全体の GC 含量ともっとも強く相関するのは、コドンの 3 文字目の GC 含量である。
- 2 GC 含量が高い DNA 二重らせんは、GC 含量が低い DNA 二重らせんよりも熱安定性が低く、生物種の成育環境と関連があると考えられる。
- 3 ゲノムの GC 含量が 10%以下の生物は、今のところ見つかっていない。
- 4 ほ乳類ゲノムにおいて、タンパク質をコードする遺伝子は、それ以外の領域と比較して、GC 含量が高い。

問 73

次世代シークエンサを用いた解析に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 RNA-seq 解析では、得られた配列をゲノムにアラインメントすることでスプライシングバリエントを検出できる。
- 2 ChIP-seq 解析では、ゲノムにアラインメントされた配列がバックグラウンドと比較して多数存在する部位を、転写因子結合部位やヒストン修飾部位として推定する。
- 3 ゲノム配列が決まっていない生物種では、近縁種のゲノム配列を用いたデ・ノボ・アッセンブル (*de novo assemble*) によりゲノム配列を決定する。
- 4 エキソーム解析は、全ゲノム解析と比較して特定の領域を低コストで調べることができる。

問 74

注目している遺伝子群に共通する機能的特徴を、データベースを用いて解析したい。解析方法とデータベースの組み合わせとして、もっとも不適切なものを選択肢の中から一つ選べ。

解析方法	データベース
1 遺伝子の機能ドメイン推定	InterPro
2 遺伝子アノテーション解析	Gene Ontology
3 転写因子結合部位推定	TRANSFAC
4 パスウェイ解析	PubMed

問 75

オーミクス解析の目的と実験手法の組み合わせとして、もっとも不適切なものを以下の選択肢の中から一つ選べ。

目的	実験手法
1 遺伝子の発現量推定	DNA マイクロアレイ
2 DNA メチル化部位の検出	二次元電気泳動
3 タンパク質間相互作用の検出	酵母ツーハイブリッド法
4 タンパク質や生理活性物質の検出	質量分析法

問 76

以下の表は、別民族に属し親戚関係のない 2 名の男性、M さんと N さんに対して、次世代シークエンサを用いた全ゲノム解析で検出された 1 塩基変異の部位数をパターン別に示したものである。例えば、M さんが A (アデニン) で、N さんが C (シトシン) であった部位は、ゲノム中に 182,812 ヶ所ある。ただし、変異のホモとヘテロの区別はしないものとする。この表の解釈として、もっとも不適切なものを選択肢の中から一つ選べ。

		Nさん				合計
		A	C	G	T	
M さん	A	182,812	678,902	166,847	1,028,561	
	C	192,920		188,415	711,194	1,092,529
	G	708,884	189,091		190,307	1,088,282
	T	167,494	680,465	183,204		1,031,163
	合計	1,069,298	1,052,368	1,050,521	1,068,348	4,240,535

- 1 二本の水素結合を作る塩基対は、三本の水素結合を作る塩基対より、塩基対間の交換が起こりやすい。
- 2 ピリミジン塩基間、あるいはプリン塩基間の置換は、ピリミジン-プリン塩基間の置換より起こりやすい。
- 3 検出された約 424 万ヶ所の 1 塩基変異は、M さんと N さんの個人差に繋がっている。
- 4 M さんと、M さんと同じ民族の男性 O さんを同様に比較した場合、1 塩基変異の部位数は 424 万ヶ所より少なくなると予想される。

問 77

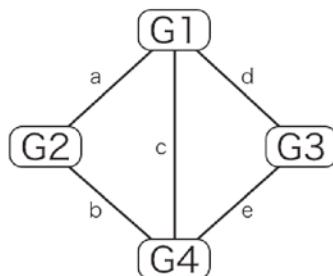
ゲノム・トランスクリプトーム・プロテオームなどの生体内高分子の解析に加えて、近年、糖・アミノ酸・有機酸・脂質などの生体内低分子化合物を網羅的に解析するメタボロミクスが、より表現型に直結する解析手法として注目されている。質量分析機によるメタボロミクスに関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 液体クロマトグラフィーなどの分離技術と X 線による構造解析を組み合わせた分離同定システムが、網羅的解析の基本システムである。
- 2 化合物の同定は分離の特性と推定質量をもとに、標準化合物ライブラリーを検索することで行う。
- 3 単一の分離技術で全代謝物質を精度よく分離する方法は存在しない。
- 4 同位体標識された化合物量の変化を時系列測定することにより、代謝流束（フラックス）解析に応用できる。

問 78

あるグラフのすべての「辺」を 1 度ずつ通る閉路（始点と終点が同じ路）をオイラーーグラフと呼び、すべての辺を 1 度ずつ通る閉路ではない路を持つものを準オイラーーグラフという。同様に、すべての「頂点」を 1 度ずつ通る閉路を持つグラフをハミルトングラフと呼び、すべての頂点を 1 度ずつ通る閉路ではない路を持つものを準ハミルトングラフという。

4 つの遺伝子 G1~G4 を頂点とし、それらの間の 5 つの関係性 a~e を辺とした下図の遺伝子ネットワークのグラフの分類として、もっとも適切なものを選択肢の中から一つ選べ。



- 1 オイラーーグラフであり、かつハミルトングラフ
- 2 準オイラーーグラフであり、かつハミルトングラフ
- 3 オイラーーグラフであり、かつ準ハミルトングラフ
- 4 準オイラーーグラフであり、かつ準ハミルトングラフ

問 79

ある有用物質を生産する微生物の物質生産能を向上させるために代謝流速解析を行いたい。代謝流速解析では細胞が定常状態であることを仮定し、代謝系におけるすべての代謝流速が満たす収支式を構築する。代謝流束モデルを構築するために使用するデータベースとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 化学量論ゲノムスケールモデルのデータベース BiGG (Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions)
- 2 酵素反応モデルのデータベース SABIO-RK (System for the Analysis of Biochemical Pathways - Reaction Kinetics)
- 3 パスウェイのデータベース KEGG (Kyoto Encyclopedia of Genes and Genomes)
- 4 パスウェイ・ゲノムデータベース BioCyc

問 80

細胞濃度 X の時間変化が比例定数 $k(>0)$ を用いて、

$$dX/dt = kX$$

と与えられるとき、細胞濃度が 2 倍になるために要する時間（ダブルリングタイム）の記述として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 細胞濃度 X が高くなるにつれて、ダブルリングタイムは長くなる。
- 2 細胞濃度 X が低くなるにつれて、ダブルリングタイムは長くなる。
- 3 ダブルリングタイムは $k / \ln 2$ で与えられる。
- 4 ダブルリングタイムは $\ln 2 / k$ で与えられる。

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
【解答】

問 1	問 2	問 3	問 4	問 5	問 6	問 7	問 8	問 9	問 10
2	4	1	2	3	3	3	3	1	4
問 11	問 12	問 13	問 14	問 15	問 16	問 17	問 18	問 19	問 20
3	4	1	2	1	3	1	4	2	2
問 21	問 22	問 23	問 24	問 25	問 26	問 27	問 28	問 29	問 30
4	3	2	1	1	4	3	1	4	3
問 31	問 32	問 33	問 34	問 35	問 36	問 37	問 38	問 39	問 40
3	1	1	4	2	4	2	2	2	3
問 41	問 42	問 43	問 44	問 45	問 46	問 47	問 48	問 49	問 50
3	1	4	4	4	3	2	3	2	4
問 51	問 52	問 53	問 54	問 55	問 56	問 57	問 58	問 59	問 60
2	3	1	2	2	1	4	1	3	4
問 61	問 62	問 63	問 64	問 65	問 66	問 67	問 68	問 69	問 70
4	1	4	2	4	1	2	3	3	2
問 71	問 72	問 73	問 74	問 75	問 76	問 77	問 78	問 79	問 80
2	2	3	4	2	1	1	2	2	4

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。
※ P.2-4 は削除されています。ご了承ください。