

平成 23 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

問 1

水は生体内で最も多く存在する分子として知られる。水の性質に関する以下の記述について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 水分子を形成する酸素原子と 2 つの水素原子の間の共有結合角は約 105 度で、立体構造は点対称性をもたない。
- 2 水分子を形成する酸素原子と水素原子の電気陰性度の差により、共有結合電子の分布が酸素原子側に偏っている。
- 3 水分子の酸素原子は、他の水分子の水素原子と静電的に結合する。これをファンデルワールス結合という。
- 4 水が状態変化を起こす際、水分子間の結合を切断するエネルギーを必要とする。そのため、メタンなど分子量の近い非極性分子に比べて融点や沸点が高い。

問 2

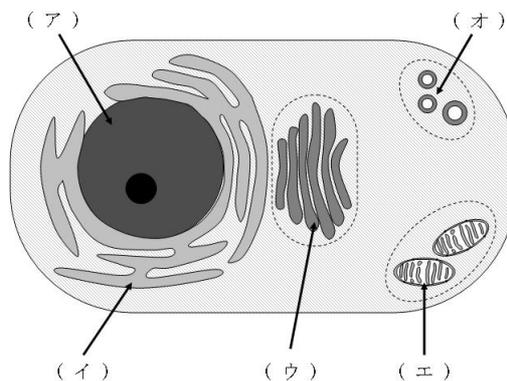
生体内のエネルギーの通貨と言われる ATP (アデノシン 3 リン酸) に関する以下の記述の中でもっとも不適切なものを選択肢の中から一つ選べ。

- 1 生体内に取り込まれた糖や脂肪は細胞質やミトコンドリアで分解され、エネルギーが ATP に貯蔵される。
- 2 ATP を ADP とリン酸に加水分解すると約 11-13 kcal/mol のエネルギーとなる。
- 3 細胞内外における物質の能動輸送は ATP の加水分解のエネルギーで駆動されている。
- 4 ATP から合成されるサイクリック AMP もエネルギーの貯蔵に使われる。

問 3

次に示す図は、真核生物の細胞を模式的に描いたものである。

(ア) から (オ) で示している細胞内小器官のうち、ミトコンドリア、ゴルジ体 (ゴルジ装置)、核に相当するものはどれか。もっとも適切な組み合わせを選択肢の中から一つ選べ。



選択肢	ミトコンドリア	ゴルジ体 (ゴルジ装置)	核
1	(イ)	(オ)	(ア)
2	(ウ)	(ア)	(エ)
3	(エ)	(ウ)	(ア)
4	(オ)	(ウ)	(イ)

問 4

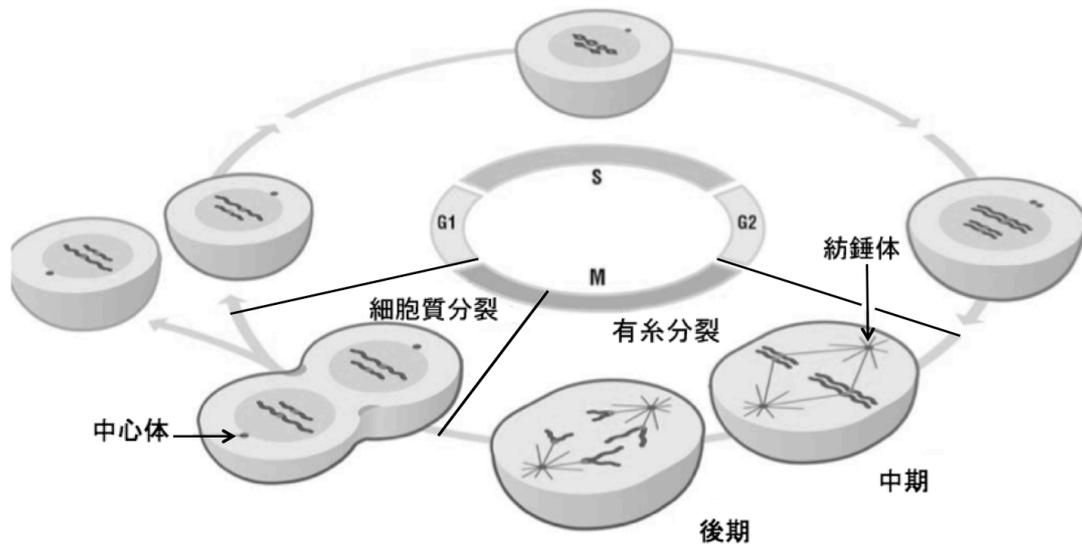
ミトコンドリアに関する以下の記述について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ミトコンドリアは地球上の全ての生物の細胞内に存在する。
- 2 ミトコンドリアは独自のゲノムを持ち、母系遺伝する。
- 3 ミトコンドリアには核の染色体ゲノム由来のタンパク質も存在する。
- 4 ミトコンドリアは内膜と外膜の二重膜で形成されている。

問 5

真核細胞の細胞周期に関する以下の記述について図を参考に以下の(a)から(d)内に入る語句の組み合わせとして適切なものを選択肢の中から選べ。

細胞周期のうち (a) が起こる時期を M 期、(b) が起こる時期を S 期と呼び、M 期から S 期の間を G1 期、S 期から M 期の間を G2 期と呼ぶ。M 期は有糸分裂期ともいわれ、核膜が消失して紡錘体が出現し、染色体中の (c) と呼ばれる部分に紡錘糸が付着することにより染色体が分離する。最後に細胞質が分裂し再び G1 期へと戻るが、S 期へと進まず (d) へと進む場合もある。



(出典 : http://commons.wikimedia.org/wiki/File:General_cell_cycle.jpg 著者:David O Mrgan, 1 January 2007, より改変)

- | | | | | |
|---|------------|------------|----------|---------|
| 1 | (a) 細胞分裂 | (b) DNA 複製 | (c) 動原体 | (d) 休止期 |
| 2 | (a) DNA 複製 | (b) 細胞分裂 | (c) 動原体 | (d) 休止期 |
| 3 | (a) 細胞分裂 | (b) DNA 複製 | (c) テロメア | (d) 分裂期 |
| 4 | (a) DNA 複製 | (b) RNA 合成 | (c) テロメア | (d) 分裂期 |

問 6

減数分裂についての以下の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 二倍体の細胞から一倍体の細胞をつくる過程は減数分裂と呼ばれる。
- 2 減数分裂は有性生殖に特有の細胞分裂様式である。
- 3 減数分裂では、父親由来と母親由来の相同染色体が対合し、交叉を行う。
- 4 交叉が起こらない限り、配偶子に父母由来の遺伝子が入りまじることはない。

問 7

ホメオティック遺伝子についての以下の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 ショウジョウバエのホメオティック遺伝子は、触覚になるべき部分が肢になるなどの、体の構造に異常を生じた突然変異体の解析から同定された。
- 2 ホメオティック遺伝子は、通常各生物種に 1 個存在する。
- 3 ホメオティック遺伝子は転写因子をコードしている。
- 4 ホメオティック遺伝子は大まかな胚の領域の将来の運命の決定に関与する。

問 8

ヒト細胞におけるシグナル伝達についての以下の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 一般的な医薬品の多くは、G タンパク質共役型受容体か核内受容体を標的としている。
- 2 二次メッセンジャーとは、メッセンジャーRNA(mRNA)に続いて情報を伝達するのに使われる低分子のことを指す。
- 3 シグナル伝達において、タンパク質のリン酸化に用いられるアミノ酸残基は、セリン、スレオニン、チロシンの 3 種類である。
- 4 核内受容体は、男性ホルモンなどの脂溶性ホルモンをリガンドとし、直接 DNA と相互作用する。

問 9

細胞接着との関連性がもっとも低いタンパク質はどれか。選択肢の中から一つ選べ。

- 1 フェリチン
- 2 インテグリン
- 3 カドヘリン
- 4 セレクチン

問 10

神経伝達との関連性がもっとも低い分子を選択肢の中から一つ選べ。

- 1 GABA
- 2 セロトニン
- 3 ユビキチン
- 4 アセチルコリン

問 11

DNA を構成する塩基（A：アデニン、T：チミン、G：グアニン、C：シトシン）に関する以下の記述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 核酸中の塩基は分子内に窒素を含む芳香環化合物である。
- 2 アデニン、グアニンはピリミジンとよばれる。
- 3 二本鎖 DNA の融解温度は塩基の組成によって変化する。
- 4 二本鎖 DNA 構造内において A は T と、G は C と水素結合する。

問 12

DNA 複製に関する以下の記述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 DNA 複製では二本鎖 DNA のうちアンチセンス鎖のみが鋳型として利用される。
- 2 DNA ポリメラーゼの反応開始には、プライマーとよばれる短い RNA 断片が必要である。
- 3 DNA 複製では重合をつかさどるポリメラーゼの他にも複数の酵素が関与している。
- 4 複製フォークの進行方向と逆方向に合成される鎖をラギング鎖と呼ぶ。

問 13

真核生物の mRNA について、以下の記述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 mRNA は RNA ポリメラーゼにより 3' から 5' 方向にむかって合成される。
- 2 mRNA の 5' 末端には翻訳に必要なキャップ構造が付加される。
- 3 mRNA は合成後に様々なプロセッシング反応を受け、核から細胞質へと輸送される。
- 4 mRNA の 5' 末端には非翻訳領域があり、AUG という配列が翻訳の開始コドンとなる。

問 14

RNA スプライシングに関する以下の記述について (a) から (d) 内に入る語句の組合せとして適切なものはどれか。選択肢の中から一つ選べ。

(a) において、DNA から転写された (b) には (c) と呼ばれる直接タンパク質のアミノ酸配列をコードしていない部分が存在する。RNA スプライシングは、この (c) を除き、残りの部分を結合して (d) を作ることをいう。

- | | | | | |
|---|----------|--------------|------------|-------------|
| 1 | (a) 原核生物 | (b) tRNA | (c) インtron | (d) タンパク質 |
| 2 | (a) 真核生物 | (b) mRNA 前駆体 | (c) エキソン | (d) 成熟 mRNA |
| 3 | (a) 真核生物 | (b) mRNA 前駆体 | (c) エキソン | (d) タンパク質 |
| 4 | (a) 真核生物 | (b) mRNA 前駆体 | (c) インtron | (d) 成熟 mRNA |

問 15

次に示した DNA あるいはタンパク質の修飾のうち、真核生物の転写制御に深く関わるものとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ヒストンのメチル化
- 2 核内受容体の N-グリコシル化
- 3 ゲノム DNA のメチル化
- 4 RNA ポリメラーゼのリン酸化

問 16

遺伝子工学の技術に関する以下の記述のなかでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 遺伝子組み換え生物による生物多様性への悪影響を防ぐため、カルタヘナ議定書では P1～P4 の 4 段階の物理的封じ込めレベルが定められている。
- 2 逆転写酵素は RNA を鋳型に DNA を合成する働きを持つため、逆転写酵素を用いて mRNA から cDNA を人工的に合成することができる。
- 3 PCR 法では、DNA を合成する酵素である DNA ポリメラーゼを用いることなく目的の DNA を増幅することができる。
- 4 サザンハイブリダイゼーション法は、標識されたプローブ DNA と相補的な配列を持つ DNA を検出する方法である。

問 17

遺伝子組み換えで利用されるベクターに関する以下の記述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 宿主と異なる生物種の DNA を組み込んで運搬することができる。
- 2 宿主の中でコピー数を増やしたりタンパク質を発現することができる。
- 3 ベクター固有の薬剤耐性等のマーカにより組み換え体を選別することができる。
- 4 ベクターの塩基配列はすべて人工的に設計されたものである。

問 18

次世代シーケンサの用途として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 *de novo* シーケンシング
- 2 プロテオーム解析
- 3 エキソーム解析
- 4 メタゲノム解析

問 19

タンパク質分子間の相互作用を調べるための実験手法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 共免疫沈降法
- 2 プロテインチップ法
- 3 酵母ツーハイブリッド法
- 4 ノーザンハイブリダイゼーション法

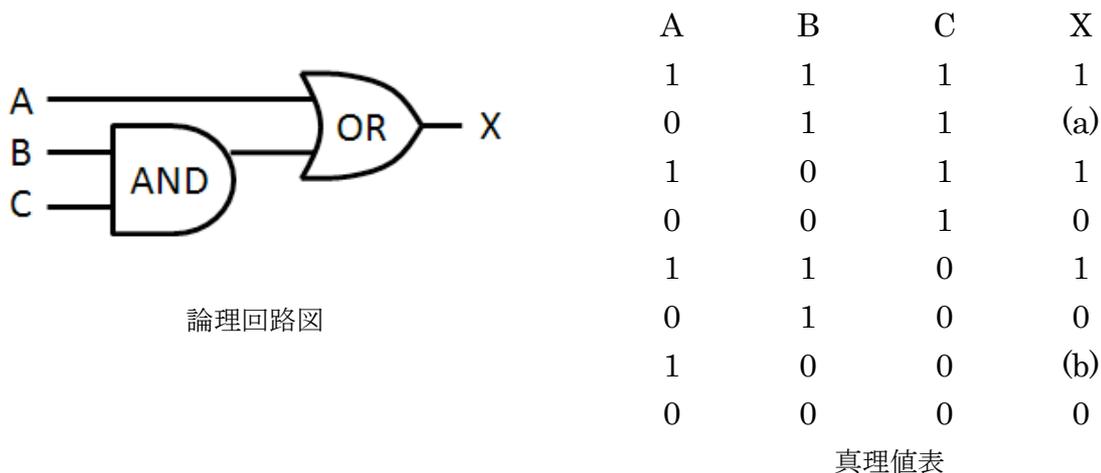
問 20

クロマチン免疫沈降シーケンス法（ChIP-Seq 法）について、以下の記述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 クロマチン免疫沈降（ChIP）によりゲノム DNA の中で転写調節因子が結合する部位を選択的に抽出することができる。
- 2 ChIP-Seq 法の配列解析には主として次世代シーケンサが用いられる。
- 3 ChIP-Seq 法により得られる解析データは DNA 結合タンパク質をコードする塩基配列である。
- 4 修飾ヒストンに対する抗体を用いることによって、ゲノム上で転写が活性化している部位と不活性化している部位を特定することができる。

問 21

下記の図は、2つの論理素子を接続した論理回路を表現している。この回路では、A、B、Cは入力でありXが出力である。この回路に対する入出力の結果によって真理値表を作成した。ここで、真理値表の(a)、(b)に入る値の組み合わせとして正しいものを、選択肢の中から一つ選べ。



- 1 (a) 1 (b) 1
- 2 (a) 1 (b) 0
- 3 (a) 0 (b) 1
- 4 (a) 0 (b) 0

問 22

以下のネットワークに関する記述のうち、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 23

TCP/IP でサブネット化されたローカルエリアネットワークにコンピュータを接続するとき、必要のない設定を選択肢の中から一つ選べ。ただし、DHCP サーバは存在しないものとする。

- 1 メールアドレス
- 2 IP アドレス
- 3 ネットマスク
- 4 ゲートウェイアドレス

問 24

メモリが 2GB、HDD が 500GB の計算機がある。この計算機内のハードディスクから、プログラムを用いて 5GB のデータを読み込んでいる時、突然データの読み込みが遅くなった。何が起こったと考えるのがもっとも適切か、選択肢の中から一つ選べ。

- 1 CPU が故障した。
- 2 ネットワークの帯域を使い尽くした。
- 3 メモリスワップが発生した。
- 4 キーボードが故障した。

問 25

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)と(b)に入る用語の組み合わせとしてもっとも適切なものはどれか。選択肢の中から一つ選べ。

バイオインフォマティクスでは Perl、Python、Ruby などのスクリプト言語が広く利用されている。これらの言語では (a) が、ソースコードを逐次解釈しながら実行する。C や Java に比べ (b) が必要ないため、プログラムを記述後、即座に実行できる特徴がある。

- | | | |
|---|-----------|----------|
| 1 | (a)コンパイラ | (b)コンパイル |
| 2 | (a)コンパイラ | (b)デバッグ |
| 3 | (a)インタプリタ | (b)コンパイル |
| 4 | (a)インタプリタ | (b)デバッグ |

問 26

以下の(a)~(d)に入れる語の組み合わせとして適切なものを選択肢の中から一つ選べ。

n 個の浮動小数点数からなる数列をソートするには、(a) を用いれば平均的には (b) 時間でソートすることができるが、最悪の場合 (c) 時間かかってしまう。一方、(d) を用いれば最悪でも (b) 時間でソートすることが可能である。

- | | | | | |
|---|-------------|-------------------|-------------------|-------------|
| 1 | (a) バケットソート | (b) $O(n)$ | (c) $O(n \log n)$ | (d) クイックソート |
| 2 | (a) クイックソート | (b) $O(n)$ | (c) $O(n \log n)$ | (d) バケットソート |
| 3 | (a) マージソート | (b) $O(n \log n)$ | (c) $O(n^2)$ | (d) クイックソート |
| 4 | (a) クイックソート | (b) $O(n \log n)$ | (c) $O(n^2)$ | (d) マージソート |

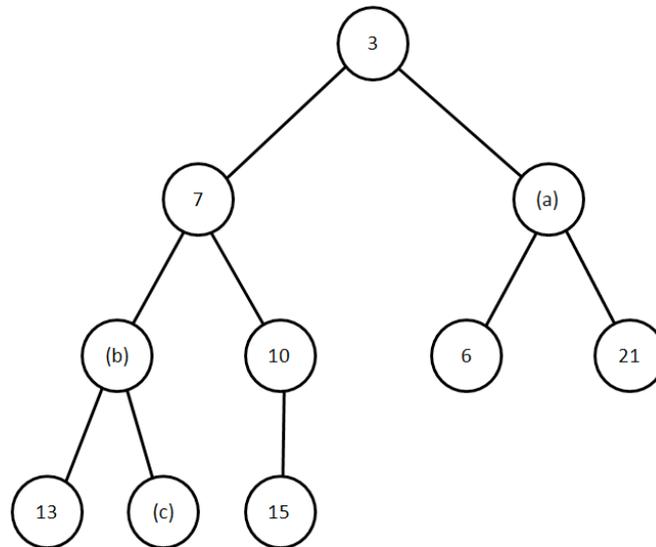
問 27

文字列処理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 文字列中の途中の位置から始まり、その文字列の最後の位置で終わる部分文字列のことを接尾辞という。
- 2 長さ n の 2 つの文字列中で、異なる文字が格納されている位置の数を数えたものをハミング距離といい、これは $O(n)$ 時間で計算することが可能である。
- 3 n 個の固定長の単語からなる集合をソートするバケットソートでは、その集合に含まれる単語を $O(n)$ 時間で辞書順にソートすることが可能である。
- 4 与えられた文字列 P が長さ n の文字列 S に部分文字列として含まれるかどうかの判定は必ず $O(n \log n)$ 時間が必要である。

問 28

数列 (21, 4, 7, 10, 12, 6, 15, 9, 3, 13) を入力としてヒープを作成したところ、以下の図のようになった。この図の中で(a)、(b)、(c)に入る数字の組み合わせとして正しいものを選択肢の中から一つ選べ。



- 1 (a) 4 (b) 9 (c) 12
- 2 (a) 9 (b) 12 (c) 4
- 3 (a) 9 (b) 4 (c) 12
- 4 (a) 12 (b) 9 (c) 4

問 29

スタックおよびキューそれぞれに対して行う操作列のうち、最後に取り出される文字が異なるような操作列を選択肢の中から一つ選べ。

ただし、スタックに対する操作の場合、英字はその文字のプッシュ操作、*はポップ操作を表す。キューに対する操作の場合、英字はその文字を入れる操作、*は取り出す操作を表す。たとえばスタックに対して「AB**」を行うと最後のポップ操作では「A」が出力されるが、キューに対して操作列「AB**」を行うと最後の取り出し操作では「B」が出力されるため、操作列「AB**」においては最後に取り出される文字はスタックとキューで異なる。

- 1 A*A*BB*
- 2 AABB***
- 3 B*BAA**
- 4 A*ABB**

問 30

一列に並んだデータを格納するデータ構造として連結リストを用いた場合と配列を用いた場合とを考える。次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。ただし、以下で n はデータの要素数を表すとする。

- 1 先頭の要素を削除する操作では、連結リストは $O(1)$ 時間で可能であるが、配列は一般に $O(n)$ 時間必要となる。
- 2 連結リスト、配列どちらを用いても、任意の $k (1 \leq k \leq n)$ について k 番目の要素を参照する操作は一般に $O(1)$ 時間で可能である。
- 3 先頭から順番にすべての要素を参照する場合、一般に配列の方が高速である。
- 4 格納する対象のデータが同じでも、連結リスト、配列どちらを用いるかによって使用する総メモリ量が異なる。

問 31

以下の分子特徴テーブルに対し、1~4 の選択肢にある 4 つの SQL 文をそれぞれ実行したが、1 つだけ他の 3 つと異なる結果を返した。他と異なる結果を返した SQL 文を選択肢の中から一つ選べ。

テーブル名：分子特徴

分子名	分子量	溶解度	沸点
メタン	16	3.5	-161.5
エタン	30	4.7	-88.6
エチレン	28	10.7	-103.9
塩化水素	36	67.0	-85.1
硫化水素	34	2.1	-60.2
オゾン	48	0.1	-111.9

- 1 SELECT 分子名 FROM 分子特徴 WHERE 分子量 < 20 ;
- 2 SELECT 分子名 FROM 分子特徴 WHERE 沸点 < -150.0 ;
- 3 SELECT 分子名 FROM 分子特徴 WHERE 溶解度 > 10.0 AND 沸点 < -100.0 ;
- 4 SELECT 分子名 FROM 分子特徴 WHERE 分子量 < 30 AND 溶解度 < 10.0 ;

問 32

データの検索を高速化するための基本的な技法として索引(index)がある。索引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 件数の大きなデータほど、索引をあらかじめ作成すると検索効率が良くなる。
- 2 1 つのデータに、複数の索引を作成することがある。
- 3 索引を作成すると、データの保存に必要な領域を減らすことが出来る。
- 4 一般的に、データが頻繁に挿入または削除される場合よりは、データが静的な場合の方が索引を導入しやすい。

問 33

以下の (a)、(b) 内に入る語句・数値の組み合わせとしてもっとも適切なものはどれか。選択肢の中から一つ選べ。

表・裏の出る確率が未知のコインを 3 回投げたところ、表が 1 回、裏が 2 回出た。表が出る確率を p とすると $p=0.3$ の尤度は $p=0.5$ の尤度 (a)。尤度は $p=(b)$ のとき最大となり、 p の最尤推定値である。

- | | | |
|---|------------|-------------|
| 1 | (a) よりも大きい | (b) 0.66... |
| 2 | (a) よりも小さい | (b) 0.66... |
| 3 | (a) よりも大きい | (b) 0.33... |
| 4 | (a) よりも小さい | (b) 0.33... |

問 34

2 つのコインを 1 回ずつ同時に投げるとする。1 個目のコインが表になる事象を A、2 個目のコインが表になる事象を B とし、それぞれ生起する確率を $P(A)$ 、 $P(B)$ とする。この 2 つの試行は統計的に独立なベルヌーイ試行とする。 $0 < P(A) < 1$ および $0 < P(B) < 1$ を仮定するが、表と裏が等確率で生起するとは限らないとする。次に示した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 $P(A) < P(A|B)$ が成り立つ場合がある。
- 2 事象 A および事象 B が同時に生起する確率 $P(A,B)$ は、 $P(A)$ と $P(B)$ の積に等しい。
- 3 $P(A,B) < P(A)$ かつ $P(A,B) < P(B)$ が成り立つ。
- 4 1 個目のコインを n 回投げたときにすべて表が出る確率は、 n をどんなに大きくしても必ず 0 より大きい確率になる。

問 35

次に示した説明文の中で、不適切なものを選択肢の中から一つ選べ。

- 1 母分散の方が標本分散より小さくなる場合がある。
- 2 標準偏差の値は 0 未満になる場合がある。
- 3 中央値と平均値は異なる場合がある。
- 4 母平均と標本平均は異なる場合がある。

問 36

次に示した説明文の中で、不適切なものを選択肢の中から一つ選べ。

- 1 事後確率を最大にするようにパラメータの値を決めることを MAP 推定と言う。
- 2 1次元の正規分布では、いかなる実現値でも、確率密度は 0 より大きい。
- 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 ベイズ推定は確率モデルのパラメータを点推定するための方法である。

問 37

以下の (a)~(d) に入れる語の組み合わせとして適切なものを選択肢より一つ選べ。

0 から 1 までの区間の中で任意の実数が同じ確率で現れる分布の事を (a) 分布とよぶ。これは、身長や体重など自然界に多く見られるばらつきを表す際に用いられる (b) 分布とは異なる。特に (b) 分布のうち、平均 (c)、分散 (d) のものを、標準 (b) 分布と呼ぶ。

- | | | | | |
|---|--------|--------|---------|--------|
| 1 | (a) 一様 | (b) 正規 | (c) 1 | (d) 1 |
| 2 | (a) 正規 | (b) 幾何 | (c) 0.5 | (d) 10 |
| 3 | (a) 正規 | (b) 幾何 | (c) 0.5 | (d) 0 |
| 4 | (a) 一様 | (b) 正規 | (c) 0 | (d) 1 |

問 38

k 平均法 (k -means method)は次のように与えられる。

入力 : x_1, \dots, x_n

出力 : クラスタリング結果 C_1, \dots, C_K

初期化 : n 個の例題を K 個のクラスタのどれかに振り分けることで C_1, \dots, C_K を初期化する。

Do

各クラスタの平均 m_k を計算する。

もっとも近い m_k に基づいてクラスタ C_1, \dots, C_K を割り付け直す。

until クラスタ C_1, \dots, C_K が変化しなくなるまで

すなわち、与えられたクラスタの割り当てで平均を計算した後、クラスタを割り当てなおしても割り当てが変わらなければ、 k 平均法の出力になりうる。いま、5 個のデータ 1, 1.5, 2, 3, 6 をクラスタ数 $K=2$ として、 k 平均法を適用したとする。 k 平均法の出力としてありうるものを一つ選べ。

- 1 $C_1=\{1\}, \quad C_2=\{1.5, 2, 3, 6\}$
- 2 $C_1=\{1, 1.5\}, \quad C_2=\{2, 3, 6\}$
- 3 $C_1=\{1, 2, 3\}, \quad C_2=\{1.5, 6\}$
- 4 $C_1=\{1, 1.5, 2, 3\}, \quad C_2=\{6\}$

問 39

k 最近傍法とは、訓練用データから距離が最も近い 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 X_1, X_2, \dots, X_{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 のクラスを予測するために k 最近傍法を用いるために、各データとの距離 $d(X, X_i)$ 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 $d(X, X_i)$ 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正しい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 $d(X, X_i)$	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

問題 40

次に示した説明文の中で、予測手法の性能評価の際に行われる交差検証法 (cross-validation) 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 予測手法が未知データにも対応できるかを検査する目的で行われる。
- 2 一部のデータを学習に使わず残しておき、テスト用に用いて予測性能を測定する。
- 3 leave-one-out 法は、データのうち 2 個のみをテスト用に残しておく方法である。
- 4 n -fold 法は、データのうち $1/n$ をテスト用に残しておく方法である。

問 41

下図は、動的計画法による大域的アラインメントスコアの計算過程を表している。各セルの値は、一番左上のセルからそのセルへ至る最適経路（スコアが最大となる経路）のスコアを表しており、最終的な最適経路上のセルを太枠で示している。図の*印（4行4列目）に入る値として、もっとも適切なものを選択肢の中から一つ選べ。ただし、スコア計算では塩基の一致、不一致、ギャップ（開始、延長、末端の区別はしない）の3状態のみを考慮するものとする。

		A	C	G	T	G
	0	-3	-6	-9	-12	-15
A	-3	2	-1	-4	-7	-10
C	-6	-1	4	1	-2	-5
T	-9	-4	1	*	3	0
G	-12	-7	-2	3	2	5

- 1 1
- 2 2
- 3 3
- 4 4

問 42

下の表は、ある微生物ゲノムにおけるダイマー（2 連塩基）の出現頻度である。このとき、条件付き確率 $P(G^2|A^1)$ の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。ここで、 A^1 および G^2 はダイマーの 1 塩基目、2 塩基目がそれぞれ A および G であることを表す。

		2 塩基目			
		A	C	G	T
1 塩 基 目	A	0.07	0.05	0.05	0.07
	C	0.07	0.06	0.08	0.05
	G	0.06	0.08	0.06	0.05
	T	0.05	0.06	0.07	0.07

- 1 0.05
- 2 0.19
- 3 0.21
- 4 0.24

問 43

マルチプルアラインメントに関する説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 あらかじめ計算した近似的な系統樹（案内木）に従って、類似性の高い配列のグループから順にペアワイズアラインメントを組み上げていく手法を、累進法（ツリーベース法）という。
- 2 累進法を用いる事で、アラインメントの初期に発生したエラーを修正する事ができる。
- 3 SP（sum of pairs）スコアは、マルチプルアラインメントのスコアを全配列ペアのスコアの和として表したものである。
- 4 逐次改善法は、計算済みのマルチプルアラインメントを 2 つのグループに分割し、グループ間で再アラインメントを行うことを反復することにより、アラインメントを改善していく手法である。

問 44

日米欧で協調して運用されている国際塩基配列データベース(INSDC)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 ArrayExpress
- 2 GenBank
- 3 ENA (EMBL-Bank)
- 4 DDBJ

問 45

PSI-BLAST に関する以下の記述において、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 PSI-BLAST はアミノ酸配列を入力とした類似配列検索プログラムである。
- 2 PSI-BLAST は問い合わせ配列や検索結果に基づいた位置特異的スコア行列を導出し、それを利用してデータベースの検索を行う。
- 3 PSI-BLAST は対話的な利用により検索結果を取捨選択するオプションが用意されており、偽陽性の類似配列の混入を防ぐことに役立つ。
- 4 PSI-BLAST は問い合わせ配列からのブートストラップリサンプリングを行い、検索結果の確からしさを評価するプログラムである。

問 46

ゲノム DNA 上にタンパク質コード領域を予測する目的で利用が可能なソフトウェアとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 AMBER
- 2 BLAST
- 3 GenScan
- 4 Glimmer

問 47

異なる塩基配列やアミノ酸配列間で頻繁に観察される配列のパターンをモチーフ配列と呼ぶ。モチーフ配列の表現方法のひとつとして、PROSITE で用いられている正規表現がある。以下はそのような正規表現の例と記述のルールである。

正規表現： $C-x-H-x-[LIVMFY]-C-x(2,4)-C-[LIVMYA]$

大文字アルファベットはアミノ酸残基、 x は任意のアミノ酸、 $-$ は文字の連結を表す。 $[]$ は括弧内のいずれか1つのアミノ酸が現れることを示す。 $x(m, n)$ は任意のアミノ酸が m 残基以上、 n 残基以下現れることを示す。

上記の正規表現に全長が一致するアミノ酸配列を選択肢の中から一つ選べ。

- 1 CLHLLCECY
- 2 CIHIICCCGECY
- 3 CLHLLCGECLA
- 4 CEEHLVCGCLECL

問 48

以下は、BLAST 検索を実行した結果の一部である（途中を適宜省略してある）。これについて述べた文章のうち、もっとも不適切なものを選択肢の中から一つ選べ。

```

Query= clone sle-57c20 map10
      (58,800 letters)

Database: SWISS-PROT protein sequence database
          446,376 sequences; 166,217,564 total letters

Searching.....done

Sequences producing significant alignments:

                                Score   E
                                (bits) Value
sp|Q9FNX5.1|DRP1E_ARATH Dynamin-related protein 1E...      168   2e-78
sp|Q8LF21.2|DRP1C_ARATH Dynamin-related protein 1C...      172   3e-77
sp|Q8S3C9.2|DRP1D_ARATH Dynamin-related protein 1D...      151   7e-67

>sp|Q9FNX5.1|DRP1E_ARATH Dynamin-related protein 1E
      Length = 624

      Score = 168 bits (426), Expect(2) = 2e-78
      Identities = 83/132 (62%), Positives = 108/132 (81%), Gaps = 7/132 (5%)
      Frame = -2

Query: 4688 QELKRFP SLQSAIAAASYEALEKYREEGRKTVLRLVDMEANYLTVEFFRKLPQEVK--- 4518
            +ELKRFP SLQ +AAA+ +LEK+REE +K+V+RLVDME+ YLT EFRKLPQE+E+
Sbjct: 451  EELKRFP SLQVELAAAANSLEKFRRESKKSIVIRLVDMESAYLTAEFFRKLPQETERPVT 510

Query: 4517 -GGNPAATPA---VDRYAE GHFRRIGLVN VSSYINMVTDLRNSIPKAVVYCQVKEAKQSL 4350
            N A+P+ +D+Y +GHFRR I NVS+Y+NMV+DTLRN+IPKA VYCQV++AK +L
Sbjct: 511  NSKNQTSAPSSATLDQYGDGHFRR IASNVSA YVNMVSDTLRNTIPKACVYCQVRQAKLAL 570

Query: 4349 LNYFYTQIGKKE 4314
            LNYFY+QI K+E
Sbjct: 571  LNYFYQSISKRE 582

      Score = 152 bits (384), Expect(2) = 2e-78
      Identities = 81/137 (59%), Positives = 93/137 (67%)
      Frame = -3

Query: 5176 FHELYITYYFFSRPGGDRIYGVFDNQLPAALRKLPFDRLPIQNVRKIVSEADGYQPHLI 4997
            F +++ + R PGGDRIYGVFDNQLPAAL+KLPFDR+L +Q+V+KIVSEADGYQPHLI
Sbjct: 348  FDKIFKEHLDGGRPGGDRIYGVFDNQLPAALKKLPFDRHLSLQSVKIVSEADGYQPHLI 407

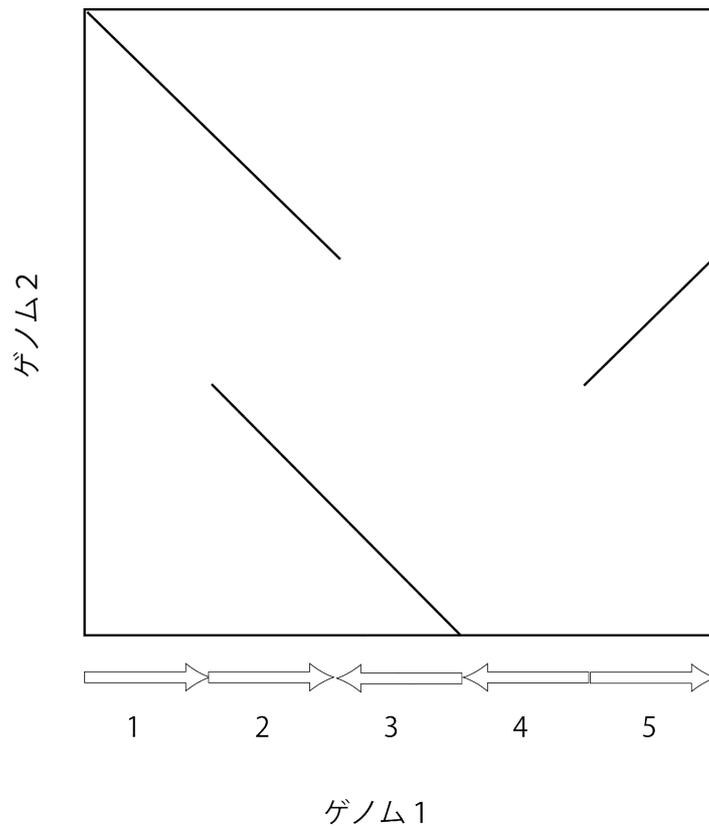
Query: 4996 APEQGYRRLIEGALN YFRGPAEASVDAVCTYFEVY*SQ LTYMFFP*QLT*CI TKQKSIY* 4817
            APEQGYRRLIEGAL YFRGPAEASVDA
Sbjct: 408  APEQGYRRLIEGALGYFRGPAEASVDA----- 434

Query: 4816 VHFVLKELVRKSVGECQ 4766
            VH+VLKELVRKS+ E +
Sbjct: 435  VHYVLKELVRKSISETE 451
    
```

- 1 ここで用いたプログラムは、核酸配列を翻訳して比較する TBLASTN である。
- 2 最初のヒット(DRP1E_ARATH)は問い合わせ配列の相補鎖側とヒットした。
- 3 最初のヒットは、複数の高スコアセグメント対(HSP)からなっているが、最初の HSP は 2 番目の HSP より、ヒットしたアミノ酸配列上で C 末端側に位置している。
- 4 2 番目の HSP には大きなギャップが見られるが、ここには終止コドンが多く含まれるので、問い合わせ配列のその領域はイントロンである可能性がある。

問 49

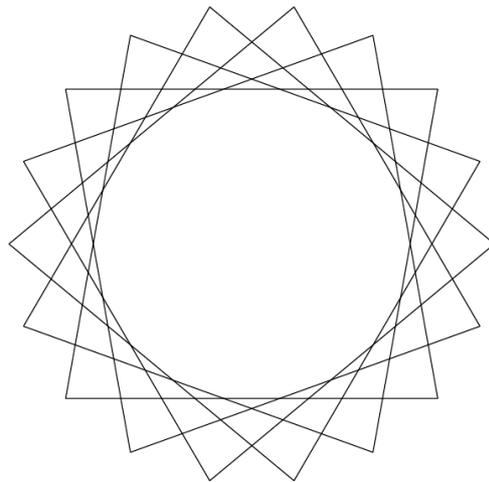
以下の図は、2本のゲノム配列の一部をドットプロットで比較した図で、相同な領域が斜線で示されている。ゲノム1上には、5つの遺伝子が図の下辺に示すように並んでおり、この配置を遺伝子の向きも含めて(1, 2, -3, -4, 5)、もしくは(-5, 4, 3, -2, -1)のように表すこととする(両者は、互いに相補鎖上での同じ配置を表している)。ここで相同な遺伝子を同じ数字で表すこととして、ゲノム2上の遺伝子の配置を同様の方法で表すと、どのようになるか。もっとも適切なものを選択肢の中から一つ選べ。



- 1 (1, 2, 2, 3, -5)
- 2 (1, 2, 5, 2, 3)
- 3 (3, -2, 5, -2, -1)
- 4 (5, -4, 3, -2, -1)

問 50

タンパク質の α ヘリックス構造において、一方の側に親水性、他方の側に疎水性のアミノ酸が偏って存在する場合、その構造は両親媒性を持つという。以下の選択肢に示すのは、アミノ酸配列を、親水性のアミノ酸を **x**、疎水性のアミノ酸を **o** として表したものであるが、このうち、もっとも顕著な両親媒性の α ヘリックス構造を取り得ると考えられる配列を選択肢の中から一つ選べ。ただし以下の **helical wheel** 図 (α ヘリックス構造を上から見た図) を参考にせよ。



- 1 OOOOOOOOXXXXXXXXX
- 2 OOXXXXOOOOXXXXOOOO
- 3 XXOOXXOOXOOXOOXXO
- 4 XOXOXOXOXOXOXOXO

問 51

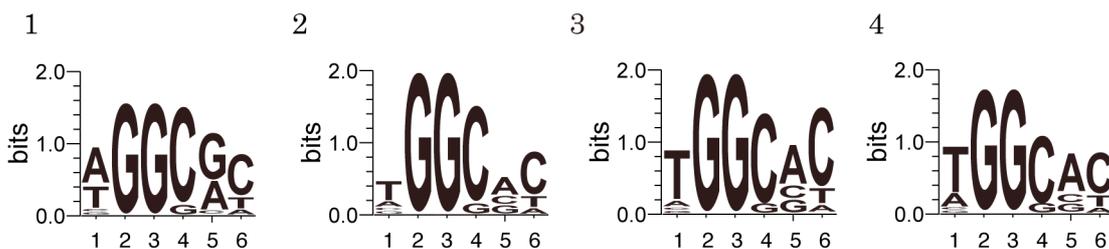
配列ロゴは、配列モチーフを視覚的に表現する方法としてよく用いられている。これは、モチーフの各位置 i における全体の高さを情報量 $R(i)$ と一致させ、その中で各文字 $a \in \{A, T, G, C\}$ の高さを、その文字の出現頻度割合 $f(i, a)$ に比例するように (すなわち $R(i)f(i, a)$ となるように) 書くことにより作成される。ただし、ここで各位置 i の情報量 $R(i)$ は、エントロピー $H(i)$ の最大エントロピー (2 bit) との差、すなわち、

$$R(i) = 2 - H(i) = 2 - \left\{ - \sum_{a \in \{A, T, G, C\}} f(i, a) \log_2 f(i, a) \right\}$$

で定義される量である (一般にはこれに補正項がつくが、本問では考慮しない)。

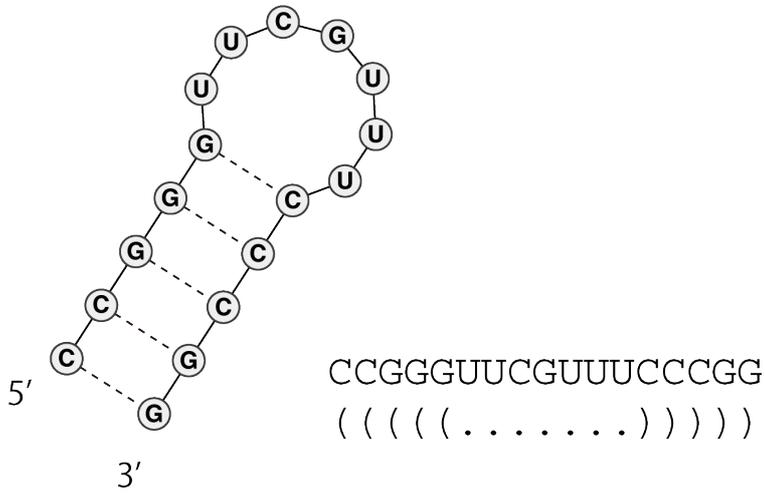
いま、各位置における文字の出現数が以下のような行列で与えられる DNA 配列のモチーフについて、配列ロゴを作成するとどのようになるか。もっとも適切なものを選択肢の中から一つ選べ。

	位置					
	1	2	3	4	5	6
A	9	0	0	0	24	5
T	29	0	0	0	0	9
G	5	48	48	5	12	0
C	5	0	0	43	12	34

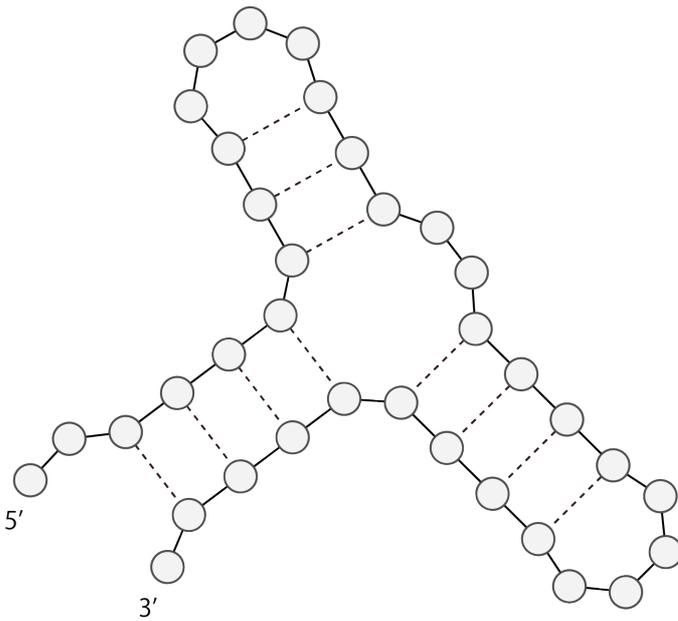


問 52

RNA の二次構造の表現方法のひとつに、塩基対を対応する括弧を用いて表現する方法がある。
 たとえば、以下の構造は((((.....))))と表現される。



これを用いた以下の二次構造の表現として、もっとも適切なものを選択肢の中から一つ選べ。



- 1 ..((((..)).....((((.....))))). .
- 2 ..(((((((.....(((.....)))))).....))))). .
- 3 ..((((..(((.....)))..))).....((((.....))). .
- 4 ..(((((((.....)))..((((.....)))))). .

問 53

タンパク質における水素結合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 主鎖原子間の水素結合は、主として二次構造形成に寄与する。
- 2 タンパク質内で水素結合は最適化されており、タンパク質が水などの溶媒分子と水素結合を作ることはまれである。
- 3 1つの原子が、複数の水素結合を作ることもよく見られる。
- 4 水素結合は、タンパク質の構造維持だけではなく、基質の分子認識などにも用いられる。

問 54

タンパク質の触媒残基として使用される頻度のもっとも低いアミノ酸を、選択肢の中から一つ選べ。

- 1 イオウ原子を含むシステイン
- 2 芳香環を持つヒスチジン
- 3 電荷を有するグルタミン酸
- 4 疎水性のイソロイシン

問 55

PDB (Protein Data Bank) は様々なファイルフォーマット (書式) で立体構造情報を配布しているが、立体構造の記述にはもっとも不適切なフォーマットを選択肢の中から一つ選べ。

- 1 PDB フォーマット
- 2 mmCIF フォーマット
- 3 PDBML フォーマット
- 4 FASTA フォーマット

問 56

あるタンパク質の PDB エントリーの抜粋を以下に示す。このエントリー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。

```

HEADER      PRION PROTEIN                               24-JUN-02   1H0L
TITLE      HUMAN PRION PROTEIN 121-230 M166C/E221C
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: MAJOR PRION PROTEIN;
COMPND     3 CHAIN: A;
COMPND     4 FRAGMENT: RESIDUES 121-230;
COMPND     5 SYNONYM: PRP, PRION PROTEIN, PRP27-30, PRP33-35C;
COMPND     6 ENGINEERED: YES;
COMPND     7 MUTATION: YES
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE     3 ORGANISM_COMMON: HUMAN;
(中略)
EXPDTA     SOLUTION NMR
(中略)
REMARK 999 2 RESIDUES (GLY SER) INSERTED AT THE N-TERMINUS
REMARK 999 MET166 AND GLU221 REPLACED AGAINST CYS
(中略)
HELIX      1  H1  ASP  A  144  MET  A  154  1                      11
HELIX      2  H2  ASN  A  173  LYS  A  194  1                      22
HELIX      3  H3  GLU  A  200  ARG  A  228  1                      29
SHEET      1  S1  2  TYR  A  128  GLY  A  131  0
SHEET      2  S1  2  VAL  A  161  ARG  A  164  -1  N  TYR  A  128  O  VAL  A  161
SSBOND     1  CYS  A  166      CYS  A  221                      1555  1555  2.04
SSBOND     2  CYS  A  179      CYS  A  214                      1555  1555  2.03
(中略)
ATOM       1  N   GLY  A  119      12.606 -28.775  2.220  1.00  0.00      N
ATOM       2  CA  GLY  A  119      11.169 -28.656  2.505  1.00  0.00      C
ATOM       3  C   GLY  A  119      10.509 -28.149  1.245  1.00  0.00      C
ATOM       4  O   GLY  A  119      10.344 -28.945  0.322  1.00  0.00      O
(後略)

```

- 1 2本の β ストランドからなる1枚の平行 β シートが存在する。
- 2 ヒトのタンパク質であるが、人工的改変が施されている。
- 3 核磁気共鳴法で決定された構造である。
- 4 ジスルフィド結合による分子内架橋が存在する。

問 57

生体高分子の動的立体構造解析に用いられる計算手法に関する以下の記述において、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 分子力学法では、分子を構成する各原子の運動はシュレーディンガーの波動方程式で計算される。
- 2 分子動力学法では、分子を構成する各原子の運動は、ニュートンの運動方程式で計算される。
- 3 基準振動解析では、分子の動的構造はラグランジュの運動方程式で計算される。
- 4 分子動力学法は、英語では **Molecular Dynamics** であり、**MD** という略称も使われる。

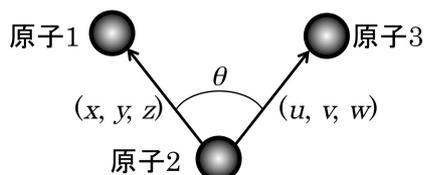
問 58

タンパク質の立体構造ドメイン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 立体構造ドメインは、長い分子進化を経てもあまり変化しない。
- 2 同じ機能を持つタンパク質でも、立体構造ドメインが著しく異なる場合がある。
- 3 1つの立体構造ドメインは、配列上連続した部位からなる。
- 4 立体構造ドメインのシャフリングで、新しいタンパク質ができる場合がある。

問 59

図のように原子 1 と原子 3 が、それぞれ原子 2 に結合した分子がある。原子 2 から原子 1 へのベクトルを (x, y, z) 、原子 2 から原子 3 へのベクトルを (u, v, w) としたとき、原子 1・原子 2・原子 3 の結合角 θ を求める式として適切なものを選択肢の中から一つ選べ。



1

$$\theta = \cos^{-1} \left(\frac{x \cdot u + y \cdot v + z \cdot w}{\sqrt{x^2 + y^2 + z^2} \cdot \sqrt{u^2 + v^2 + w^2}} \right)$$

2

$$\theta = \sin^{-1} \left(\frac{x \cdot u + y \cdot v + z \cdot w}{\sqrt{x^2 + y^2 + z^2} \cdot \sqrt{u^2 + v^2 + w^2}} \right)$$

3

$$\theta = \cos^{-1} \left(\frac{x \cdot u + y \cdot v + z \cdot w}{\sqrt{(x+u)^2 + (y+v)^2 + (z+w)^2}} \right)$$

4

$$\theta = \cos^{-1} \left(\frac{x \cdot u + y \cdot v + z \cdot w}{\sqrt{(x-u)^2 + (y-v)^2 + (z-w)^2}} \right)$$

問 60

立体構造の類似した 2 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、全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った(1 Å = 0.1 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'	1.0 Å
原子 B—原子 B'	0.0 Å
原子 C—原子 C'	1.0 Å
原子 D—原子 D'	3.0 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 61

タンパク質の二次構造予測に用いられるチョウ・ファスマンパラメータは、アミノ酸残基 A の二次構造 S への傾向値を、以下の式で評価するものである。

$$\text{傾向値} = \{(S \text{ 内に存在する } A \text{ の総数}) / (A \text{ の総数})\} / \{(S \text{ 内のアミノ酸の総数}) / (\text{アミノ酸の総数})\}$$

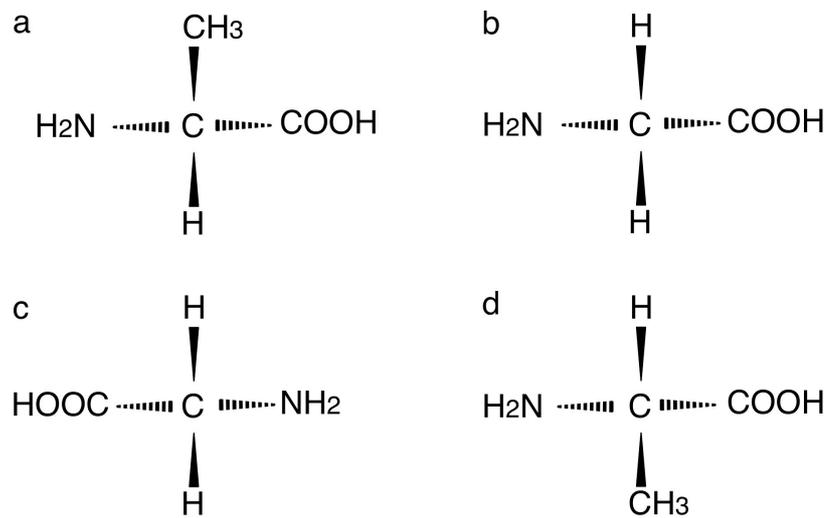
表に示した統計をもとに、ロイシン残基の α ヘリックス傾向性をあらわすチョウ・ファスマンパラメータを求めたとき、もっとも適切な値を選択肢の中から一つ選べ。

α ヘリックス中のロイシンの総数	β ストランド中のロイシンの総数	コイル中のロイシンの総数	α ヘリックス中のアミノ酸の総数	β ストランド中のアミノ酸の総数	コイル中のアミノ酸の総数
300	100	400	3000	1500	4500

- 1 3/8
- 2 3/1
- 3 9/10
- 4 9/8

問 62

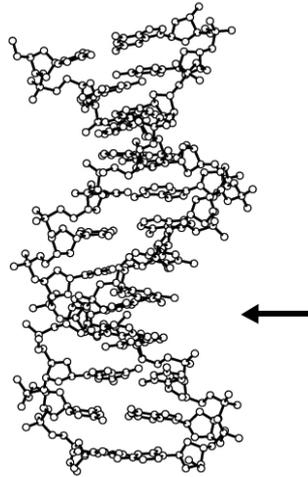
以下の投影図法で描かれた a から d のアミノ酸のうち、互いにエナンチオマー(鏡像異性体)の関係にある分子の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。



- 1 a-b
- 2 b-c
- 3 b-d
- 4 a-d

問 63

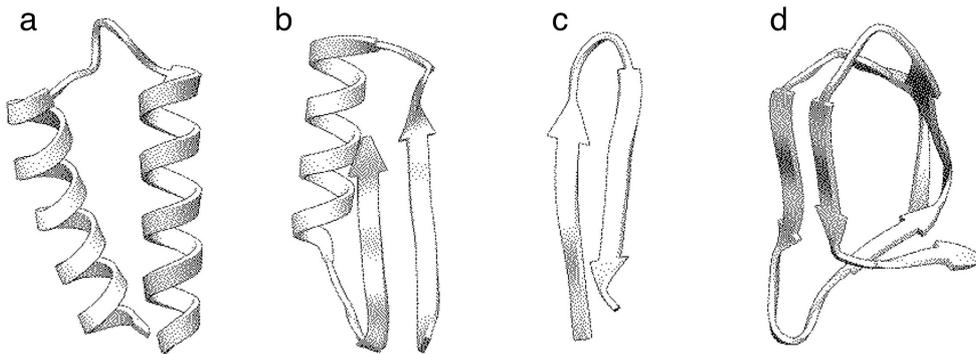
下図は DNA の立体構造を表したものであるが、この DNA についてもっとも不適切な記述を選択肢の中から一つ選べ。



- 1 これは B-form とよばれる構造である。
- 2 矢印のくぼみは副溝(minor groove)とよばれる。
- 3 この構造は 2 重らせんである。
- 4 この DNA は 12 塩基対からなる。

問 64

タンパク質の立体構造に高頻度で現れる二次構造の空間配置を超二次構造(super secondary structure)とよぶ。図に示す a から d の 4 種の超二次構造を、 α ヘアピン - β ヘアピン - $\beta\alpha\beta$ - β ヘリックスの順番に並べたとき、もっとも適切な記号列を選択肢の中から一つ選べ。



- 1 a-b-c-d
- 2 a-c-b-d
- 3 b-c-a-d
- 4 b-c-d-a

問 65

HapMap プロジェクトは、SNP を中心としたヒトの遺伝子多型に関するデータを公開している。この HapMap プロジェクト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 1 つ選べ。

- 1 複数の民族の遺伝子多型データが公開されている。
- 2 日本人サンプルのデータが含まれている。
- 3 SNP タイピングの実験プラットフォームは 1 種類ではない。
- 4 親子トリオのデータを取得したのでいずれの民族のデータもハプロタイプ推定の精度が高い。

問 66

ゲノムワイド関連解析(genome-wide association study, 以後 GWAS と略す)において、メタアナリシスが実施されることが増えている。メタアナリシスについて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 ある遺伝的多型と疾患の関連性についての複数のグループによるスタディの結果をメタアナリシスで統合して解析することで、多型の持つ疾患への影響を正確に推定できる。
- 2 小規模の個別のスタディの統合においても、GWAS のような大規模なスタディの統合においても、メタアナリシスによる統合の手順は同様である。
- 3 出版された研究のみを対象とすることで、データソースに関するバイアスを排除できる。
- 4 メタアナリシスでは、対象候補となるスタディのうち、どのスタディを対象とするか否かの取捨選択を行う。この取捨選択が中立でなければ、恣意的な結論を導きうる。

問 67

ABO 式血液型を決める遺伝子多型は 3 つのアレル A、B、O の組み合わせによって決まり、その遺伝子型は AA、AO、BB、BO、AB、OO の 6 通りある。それぞれの遺伝子型を持つとき、血液型の表現型は A、A、B、B、AB、O となる。

ハーディ・ワインバーグ平衡にある集団において、血液型の表現型の頻度が A、B、AB、O それぞれ、0.09、0.65、0.10、0.16 のとき、この集団の 3 つのアレル A、B、O のアレル頻度として正しい組み合わせを選択肢の中から一つ選べ。

- | | | | | | | |
|---|---|-----|---|-----|---|-----|
| 1 | A | 0.3 | B | 0.2 | O | 0.5 |
| 2 | A | 0.3 | B | 0.3 | O | 0.4 |
| 3 | A | 0.1 | B | 0.4 | O | 0.5 |
| 4 | A | 0.1 | B | 0.5 | O | 0.4 |

問 68

オーソログに関する以下の記述のうち、もっとも不適切なものを一つ選べ。

- 1 一般に偽遺伝子にオーソログは存在しない。
- 2 1つのヒト遺伝子に対して、マウスのオーソログが複数見つかることがある。
- 3 オーソログの遺伝子配列で分子系統樹を作成することにより、生物種間の系統関係を推定できる。
- 4 ヒト遺伝子のオーソログがニワトリに存在しても、マウスには存在しない場合がある。

問 69

分子系統樹の作成方法として不適切なものを選択肢の中から一つ選べ。

- 1 最大節約法 (maximum parsimony)
- 2 近隣結合法 (neighbor-joining)
- 3 動的計画法 (dynamic programming)
- 4 最尤法 (maximum likelihood)

問 70

以下の文章の(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。

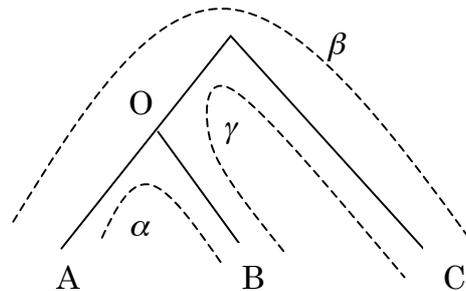
無根系統樹において可能な樹形の数、OTU (Operational Taxonomic Unit; 遺伝子や種など、系統樹を推定する際の操作単位) が4つのときは3通り、OTUが5つのときは(a)通り、6つのときは105通りというように急速に増加する。

- 1 15
- 2 18
- 3 21
- 4 24

問 71

3本の配列 A、B、C について、配列間距離が $d(A-B) = \alpha$ 、 $d(A-C) = \beta$ 、 $d(B-C) = \gamma$ と計算されたとする。これら 3本の配列の系統関係は図のように、A、B が近く、C がそれらに対して遠い関係になる。この図で O は、A、B の共通祖先を表す節である。相対速度テストで用いられる方法によって、O から A の距離を計算する。O-A 間の距離として正しいものを選択肢の中から一つ選べ。

- 1 $d(O-A) = (\alpha + \beta + \gamma)/2$
- 2 $d(O-A) = (\alpha + \beta - \gamma)/2$
- 3 $d(O-A) = (\alpha - \beta + \gamma)/2$
- 4 $d(O-A) = (\alpha - \beta - \gamma)/2$



問 72

2つの遺伝子 A、B が図のようにフレームをずらして長い領域がオーバーラップしていた。近縁なオーソログでも同じオーバーラップが観察されたとする。この時、オーバーラップしている領域の同義置換率について、もっとも適切なものを選択肢の中から一つ選べ。

遺伝子 A

遺伝子 B

- 1 オーバーラップ領域の遺伝子 A の同義置換率は、遺伝子 A のオーバーラップしていない領域の同義置換率よりも高い。
- 2 オーバーラップ領域の遺伝子 A の同義置換率は、遺伝子 A のオーバーラップしていない領域の同義置換率よりも低い。
- 3 オーバーラップ領域の遺伝子 A の同義置換率は、近似的に遺伝子 A と遺伝子 B のオーバーラップしていない領域の同義置換率の平均値になる。
- 4 オーバーラップ領域の遺伝子 A の同義置換率は、近似的に遺伝子 A と遺伝子 B のオーバーラップしていない領域の非同義置換率の平均値になる。

問 73

あるタンパク質 A を培養細胞から、A に特異的な抗体を用いて精製した。精製自体には問題がないように思われたが、SDS-PAGE で確認したところ、移動度から算出した分子量とデータベースに登録されている分子量の間に約 10kDa の差が認められた。この差の原因として確認すべき可能性として、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 A が、総電荷が偏った特殊なアミノ酸組成を持っている可能性。
- 2 タンパク質 A が、他の低分子量タンパク質と複合体を形成する可能性。
- 3 タンパク質 A が、組織特異的な選択的スプライシングを受けている可能性。
- 4 タンパク質 A が、糖鎖修飾などの翻訳後修飾をうけている可能性。

問 74

マイクロアレイ法による遺伝子発現データを用いる際には、測定後の生データに補正を加える。以下の補正手順のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 非特異的なハイブリダイゼーションの影響を取り除くためにスポット強度からバックグラウンドの値を引く。
- 2 遺伝子発現量の変化を比でとらえるためにスポット強度を対数値に変換する。
- 3 RNA の合成速度を見積もるためにスポット強度をエキソンの数で割る。
- 4 サンプル間でのスポット強度を比較するために、スポット強度の中央値をそろえる。

問 75

RNA はタンパク質や代謝物と比較して網羅的解析に適している。その理由となる RNA の性質として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 測定サンプルが少量であっても、PCR 法により量を増やすことが容易である。
- 2 塩基の相補性を利用することで、各配列に特異的なプローブの作成が容易である。
- 3 物理化学的性質が均質であるため、他の生体物質との分離が容易である。
- 4 水溶液中で安定した立体構造を形成し、取り扱いが容易である。

問 76

オーミクス解析に利用される実験機器とその使用目的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。

選択肢番号	実験機器	目的
1	DNA マイクロアレイ	遺伝子発現量の解析
2	次世代シーケンサ	SNP 解析
3	キャピラリーシーケンサ	選択的スプライシングの解析
4	質量分析装置	発現配列タグ (EST) の解析

問 77

あるタンパク質 X に対するモノクローナル抗体とポリクローナル抗体を購入し、ウエスタンブロッティング法によるサンプル間の比較解析を行った。その結果、ポリクローナル抗体を用いた解析では、サンプル A とサンプル B のシグナルに有意な差は見られなかったが、モノクローナル抗体を用いた解析では、サンプル A に対してサンプル B のシグナルの著しい減少が見られた。

この実験結果の考え得る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 2つのサンプルでは、タンパク質 X をコードする mRNA の翻訳の効率に差がある。
- 2 2つのサンプルで、タンパク質 X の一部のアミノ酸が異なる化学修飾状態にある。
- 3 2つのサンプルには、選択的スプライシングの影響で、異なるアミノ酸配列を有するタンパク質 X が異なる比で存在する。
- 4 2つのサンプルでは、タンパク質 X が翻訳後にプロテアーゼによる切断を介した活性化を受ける効率が異なる。

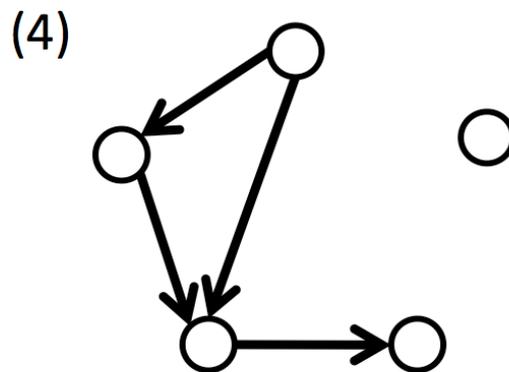
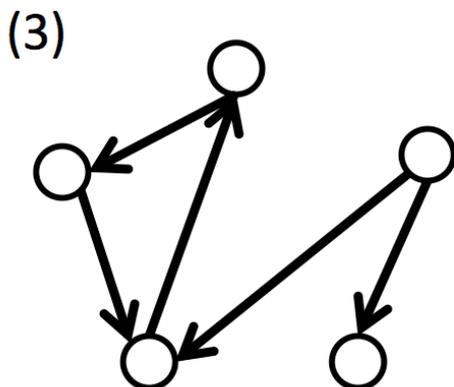
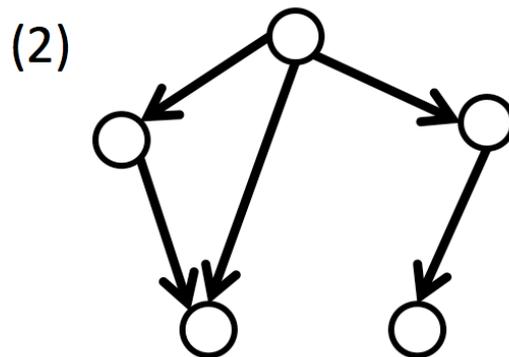
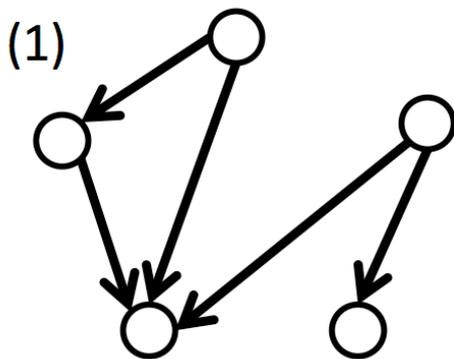
問 78

生体内パスウェイに対する時系列シミュレーションモデルに関する説明のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 システムノイズの影響に敏感なロバストネスが高いモデルがある。
- 2 代謝系の反応、シグナル伝達系の反応、遺伝子の転写反応を扱えるモデルがある。
- 3 確率的な制御反応を扱えるモデルがある。
- 4 細胞内のコンパートメントを考慮してシミュレーションできるモデルがある。

問 79

5つの遺伝子に対して、発現量の測定を行った。次に、各遺伝子の発現量を確率変数と考え、統計モデルによって発現制御の因果関係を推定し、ベイジアンネットワークを用いて表現した。以下の有向グラフのうち、ベイジアンネットワークの構造としてもっとも不適切なものを一つ選べ。



問 80

ある mRNA の時刻 $t (\geq 0)$ での発現量を $X(t)$ とした時、 $X(t)$ は正の定数 a と b を含む以下の方程式に従うとする。

$$\frac{dX(t)}{dt} = a - bX(t)$$

このとき、この mRNA の定常状態での発現量に関する以下の文章の (ア) (イ) に入れるもっとも適切な語句の組み合わせを、選択肢の中から一つ選べ。

$X(0)$ の値が大きくなった場合、定常状態でのこの mRNA の発現量は (ア)。また、mRNA の定常状態での発現量は、 b の値が一定のときに、 a の値が大きくなると (イ)。

- 1 (ア) 変わらない (イ) 小さくなる。
- 2 (ア) 変わらない (イ) 大きくなる。
- 3 (ア) 大きくなる (イ) 小さくなる。
- 4 (ア) 大きくなる (イ) 大きくなる。

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

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。