

平成 21 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

次に示した説明文の中で、遺伝子組換え実験で用いられるプラスミドベクター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 プラスミドベクターは、コスミドベクターに比べて長い DNA 断片を挿入することができる。
- 2 プラスミドベクターは、組み換えに利用しやすいように人工的な配列が組み込まれていることが多い。
- 3 プラスミドベクターは、一般に環状二本鎖 DNA である。
- 4 プラスミドは、大腸菌内で染色体 DNA とは独立して複製される DNA として発見された。

問 2

代謝について、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 代謝においては、物質の合成とエネルギーの移動は表裏一体のものといえる。
- 2 生物がエネルギーを得るための代謝には、発酵、呼吸、光合成の三種類がある。
- 3 ATP のように、電子伝達体 NADPH は生合成反応に関わり、酸化型 NADP⁺となる。
- 4 同化とは、外部の物質を分解して、ATP を合成する代謝である。

問 3

生体膜について、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 脂質二重膜では、生体膜の主成分であるリン脂質が表裏に多数並んでいる。
- 2 コレステロールはリン脂質の一種であり、膜の流動性に影響する。
- 3 リン脂質のうちの疎水性の部分は脂肪酸と呼ばれる。
- 4 糖脂質は、生体膜の成分の一種で、糖と脂質が結合したものである。

問 4

免疫について、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 自然免疫とは、T 細胞を介した抗原抗体反応を伴わない免疫で、原始的な生物のみに備わった仕組みであり、ヒトなど高度な免疫システムを備えた生物には存在しない。
- 2 獲得免疫においては、各細胞表面に提示された自己抗原を認識することで、いわゆる自己免疫反応を防いでいる。
- 3 エイズウイルス HIV は、宿主の免疫システムを破壊する。
- 4 抗体とは、免疫グロブリンとよばれるタンパク質のことである。

問 5

10,000 塩基対で構成された完全に相補的な二本鎖 DNA において、グアニンの含有率が 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 2,000 塩基
- 2 3,000 塩基
- 3 4,000 塩基
- 4 6,000 塩基

問 6

真核生物の細胞内小器官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 小胞体はタンパク質の生合成だけでなく、ATP の合成にも関与している。
- 2 ゴルジ体で糖が付加された膜タンパク質は小胞となってゴルジ体のトランス面より細胞膜に輸送される。
- 3 ミトコンドリアは独自のゲノム DNA を持ち、分裂によって増加する。
- 4 リソソームは加水分解酵素を含み生体内の不要物を消化する役割を担っている。

問 7

多細胞生物の減数分裂に関する次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 減数分裂は生殖細胞が形成される時の細胞の分裂様式である。
- 2 減数分裂に対し、一般の細胞分裂を体細胞分裂と呼ぶ。
- 3 減数分裂の結果生じる生殖細胞の染色体数は、体細胞の 1/4 である。
- 4 減数分裂において相同染色体間で交差が起こり、その結果、遺伝子の組換えが生じる。

問 8

核酸の化学構造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 DNA も RNA も塩基、糖、リン酸という 3 つの成分で構成されている。
- 2 RNA の糖部分の 2' 位は水酸基であるが、DNA では水素である。
- 3 スクレオチドはホスホジエステル結合によって連結され生体高分子となる。
- 4 DNA の二重らせん構造は塩基間の共有結合により形成される。

問 9

DNA 複製に関する次の記述のうち、もっとも適切なものを選択肢の中から一つ選べ。

- 1 DNA 複製は細胞周期の分裂期 (M 期) に起こる。
- 2 大腸菌のゲノム DNA の複製は Ori 領域からスタートする。
- 3 DNA 複製に関与する DNA ポリメラーゼは鋳型を必要としない。
- 4 合成されたリーディング鎖のことを岡崎フラグメントと呼ぶ。

問 10

真核生物においてタンパク質をコードする遺伝子の典型的な構造に関する次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 エキソンは主にアミノ酸配列をコードする部分である。
- 2 イントロンは転写されない部分である。
- 3 遺伝子の 5'側には RNA ポリメラーゼが結合するプロモーターが存在する。
- 4 遺伝子の 3'側には RNA ポリメラーゼの終結を指示するポリ A シグナルが存在する。

問 11

遺伝暗号ではコドンと呼ばれる 3 個の塩基の配列によって対応するアミノ酸が決まっているが、次に挙げるアミノ酸とコドンの組合せのうち、典型的な開始コドンとアミノ酸の組合せはどれか、正しいものを選択肢の中から一つ選べ。

- 1 メチオニン-ATG
- 2 システイン-AGC
- 3 メチオニン-GCC
- 4 システイン-AAG

問 12

タンパク質の翻訳に関する次の記述の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 RNA を鋳型にタンパク質が合成されることを翻訳と呼ぶ。
- 2 どの生物においてもタンパク質の翻訳はリボソームで行われる。
- 3 リボソームは巨大な 1 種類のタンパク質である。
- 4 tRNA は鋳型となる mRNA のコドンに結合するアンチコドン部分を持つ。

問 13

細胞の情報伝達に関する次の記述の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 細胞から分泌されたシグナル分子はリガンドとして標的細胞の受容体に結合する。
- 2 G タンパク質共役型受容体はその構造の特徴から 7 回膜貫通型受容体に分類される。
- 3 イオンチャネル型受容体では、リガンドの結合によりチャネルの開閉が起こり、その結果生じたイオン濃度の変化によって刺激が伝えられる。
- 4 チロシンリン酸化酵素活性を持つタイプの受容体は核内受容体の代表的なものである。

問 14

ゲノム解析に関する次の記述の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ゲノムの解読装置として自動シーケンサが用いられる。
- 2 サンガー法によるゲノム配列の決定には主にショットガン法が用いられる。
- 3 ゲノム解読ではタンパク質に翻訳される部分のみが決定されている。
- 4 次世代シーケンサはサンガー法によらない新しいシーケンシング技術を利用している。

問 15

トランスクリプトーム解析に関する次の記述の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 細胞や組織の転写物を網羅的に解析することをトランスクリプトーム解析と呼ぶ。
- 2 トランスクリプトーム解析技術の代表的なものは、DNA マイクロアレイである。
- 3 トランスクリプトーム解析には精製したタンパク質が試料として用いられる。
- 4 次世代シーケンサがトランスクリプトーム解析に利用されている。

問 16

プロテオーム解析に関する次の記述の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 二次元電気泳動法は、タンパク質間相互作用を一度に検出できる方法である。
- 2 質量分析装置の解析結果とアミノ酸配列データベースとの照合により、タンパク質の同定が可能である。
- 3 細胞内でのタンパク質の局在を観察するためには、蛍光分子等で標識された抗体が用いられることが多い。
- 4 酵母ツーハイブリッド法により、ヒト由来のタンパク質間相互作用も検出することができる。

問 17

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 ゲノム DNA のユビキチン化
- 2 ヒストンのアセチル化
- 3 ヒストンのメチル化
- 4 RNA ポリメラーゼのリン酸化

問 18

次に示した実験手法に関する文章の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 免疫沈降法を用いて、タンパク質間の相互作用を検出することができる。
- 2 DNA マイクロアレイ法を用いて、大規模な遺伝子発現変化を検出することができる。
- 3 FACS 法を用いて、特定の表面抗原を発現している細胞を分画することができる。
- 4 ウェスタンブロッティング法を用いて、特定の細胞での mRNA の発現コピー数を定量的に解析することができる。

問 19

次の真核生物の転写因子に関する文章の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 転写因子はゲノム DNA に結合して、遺伝子発現を正または負に制御する。
- 2 転写因子の中にはゲノム DNA を切断し、アポトーシスを誘導するものがある。
- 3 転写因子の中には同時にいくつもの遺伝子発現を制御するものがある。
- 4 転写因子の中にはホルモン等の受容体としての機能をあわせ持つものもある。

問 20

生体高分子の立体構造を決定する際に広く用いられている X 線結晶解析で直接的に観察されるものは次のどれか。正しいものを選択肢の中から一つ選べ。

- 1 原子（核）の位置
- 2 原子間の結合
- 3 電子の分布（電子密度）
- 4 水素原子（プロトン）間の距離

問 21

DNA シークエンサに関する以下の文章中の (a)、(b) 内に入る語句の組合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。

国際コンソーシアムによるヒトゲノムプロジェクトでは、主に (a) による DNA の塩基配列決定が行われ、その完了には 13 年の歳月と 27 億ドルの費用を要した。しかし、2005 年に発表された新しい DNA シークエンサは、DNA ポリメラーゼが DNA を合成する際に放出するピロリン酸を光に変換してとらえる (b) と呼ばれる技術と微細加工技術を組み合わせることにより、(a) よりも解読塩基長は劣るものの、一回の解析でそれまでとは桁違いの 10 万分子の DNA 配列の同時決定を可能にした。この DNA シークエンサを駆使することにより、ワトソン博士のゲノム配列をわずか 4.5 ヶ月の期間と 150 万ドル以下のコストで決定できたという。(b) とは異なる原理に基づき、解読塩基長はさらに短い、一回の解析で数百万分子の DNA 配列を同時に決定できる機種も発表されており、これらはしばしば次世代シークエンサと総称されている。さらに、数年後にはヒトゲノム配列をわずか 15 分で決定できる DNA シークエンサも登場すると言われている。

(M. Wadman, *Nature* **452**, 788, 2008 の記事をもとに作成)

- 1 (a) サンガー法 (b) マイクロシークエンシング法
- 2 (a) エドマン法 (b) マイクロシークエンシング法
- 3 (a) サンガー法 (b) パイロシークエンシング法
- 4 (a) エドマン法 (b) パイロシークエンシング法

問 22

計算機の CPU (Central Processing Unit) において、連続する命令群の一つ一つの命令における処理を細分化し、それらの細分化された処理を流れ作業的に同時実行し、CPU の計算効率を向上させる技術を何とよぶか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 パイプライン
- 2 タイムシェアリング
- 3 マルチスレッド
- 4 マルチコア

問 23

以下の (a) ～ (d) 内に入る語句の組合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。

(a) などの言語では、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、(b) とよばれるプログラムが、ソースコードあるいは中間言語を解釈しながら実行を行う。一方、(c) などの言語では、(d) とよばれるプログラムが直接機械語に変換を行うため、(b) に基づく言語より実際の計算は高速であることが多い。

- | | | | | |
|---|----------|-------------|----------|-------------|
| 1 | (a) Perl | (b) コンパイラ | (c) C | (d) インタープリタ |
| 2 | (a) Perl | (b) インタープリタ | (c) C | (d) コンパイラ |
| 3 | (a) C | (b) コンパイラ | (c) Perl | (d) インタープリタ |
| 4 | (a) C | (b) インタープリタ | (c) Perl | (d) コンパイラ |

問 24

コンピュータ上では、数値は二進数で表される。二進数同士の論理和(OR)とは、2つの二進数の各桁に対してそれぞれ論理和をとったものであり、たとえば、 $110_{(2)} \text{ OR } 101_{(2)} = 111_{(2)}$ である。なお、ここで数字の最後に付加した (n) は、その直前の数字が n 進数による表現であることを示す。すなわち、たとえば $6_{(10)} = 110_{(2)}$ である。このとき、二進数 x に対する関数 $f(x)$ を

$$f(x) = x \text{ OR } 100_{(2)}$$

と定義する。このとき、以下の選択肢に掲げる y 、 z の組み合わせのうち、 $f(y) = f(z)$ となるものを一つ選べ。

- | | |
|-------------------|-----------------|
| 1 $y = 101_{(2)}$ | $z = 110_{(2)}$ |
| 2 $y = 110_{(2)}$ | $z = 010_{(2)}$ |
| 3 $y = 010_{(2)}$ | $z = 100_{(2)}$ |
| 4 $y = 000_{(2)}$ | $z = 111_{(2)}$ |

問 25

次に示した説明文の中で、コンピュータネットワークに関する用語の説明として不適切なものを選択肢の中から一つ選べ。

- 1 DHCP は、接続するコンピュータに対して、IP アドレス等の必要な情報を自動的に割り当てるプロトコルである。
- 2 MAC アドレスは、ハードウェアに書き込まれており、世界中でユニークな ID である。
- 3 IP アドレスは無限に割り当てることができるため、将来ネットワークを利用する機器がどんなに世界中で増えたとしても、IP アドレスが枯渇することはない。
- 4 ソケット通信では、TCP/IP ネットワーク上で異なるコンピュータ上のプログラム同士がデータを交換するが、その際に通信相手のプログラムを特定するためにポート番号を用いる。

問 26

節点（ノード）数が n である有根の木構造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 枝の数は $n-1$ である。
- 2 葉の数は $(n+1)/2$ 以下である。
- 3 木構造のすべての節点の次数が 3 次以下ならば、根から葉までのパス上の枝の数は $\log_2 n$ 以下である。
- 4 根が子節点を二つ以上持っているならば、少なくとも一つの根の子節点は子孫節点の数が $n/2$ 以下である。

問 27

以下の (a) ~ (d) 内に入る語句の組合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。

「 n 個の実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) を用いればいかなる数列でも (b) 時間でソートすることが可能である。」

- | | | | |
|---------------|-------------------|-------------------|-------------|
| 1 (a) ヒープソート | (b) $O(n)$ | (c) $O(n \log n)$ | (d) クイックソート |
| 2 (a) マージソート | (b) $O(n \log n)$ | (c) $O(n^2)$ | (d) クイックソート |
| 3 (a) クイックソート | (b) $O(n)$ | (c) $O(n \log n)$ | (d) マージソート |
| 4 (a) クイックソート | (b) $O(n \log n)$ | (c) $O(n^2)$ | (d) ヒープソート |

問 28

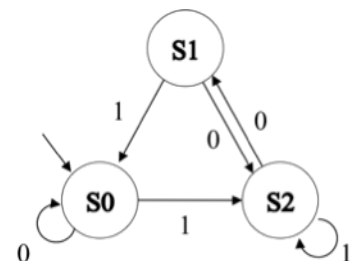
ハッシュに関する以下の記述のうち、不適切なものを選択肢の中から一つ選べ。

- 1 ハッシュ法を用いれば、要素（ハッシュ・キー）を探索することが最悪でも定数時間で可能となる。
- 2 ハッシュ法で要素として管理できるデータとしては、整数、文字列などがあげられる。
- 3 連鎖（chaining）法では、2 つ以上の異なる要素のハッシュ関数の値が同じ値をとった場合に、それらの要素をリストで管理する。
- 4 開番地（open address）法では、2 つ以上の異なる要素のハッシュ関数の値が同じ値をとった場合に、何らかの別の計算によってハッシュ上でまだ用いられていない別の番地を探し、2 つ目以上の要素はその新しく計算した番地で管理することで、一つ一つの要素に必ず一つの番地を割り当てる。

問 29

下図の決定性有限オートマトンは 3 つの状態 S0、S1、S2 を持つ。S0 は初期状態である。入力数列は 0 または 1 である。このオートマトン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。

状態遷移表		状態		
		S0	S1	S2
入力	0	S0	S2	S1
	1	S2	S0	S2



- 1 入力数列が 000111 のとき、状態は S2 となる。
- 2 入力数列 01 の後、さらに 0 が偶数回続くと、状態は S2 となる。
- 3 入力数列 10 の後、さらに 1 が奇数回続くと、状態は S2 となる。
- 4 数列中に 1 が 1 回のみ現れる入力数列の入力後の状態は、S0 ではない。

問 30

データベースのモデルの一つであるリレーショナルデータベースに関する説明として、不適切なものを選択肢の中から一つ選べ。

- 1 データをリレーション（関係）と呼ばれる単位で管理する。リレーションは二次元のテーブルとして表現することができ、データベース内には多数のリレーションを格納できる。
- 2 リレーション間の情報を結びつけるには、結合（join）演算を用いる。近年の CPU の進化により、結合演算はハードウェア上でほぼ 1 命令で実行でき、高速性に寄与している。
- 3 リレーショナルデータベースを管理するソフトウェアは、RDBMS と呼ばれる。
- 4 RDBMS 上で採用される主要なデータベース言語（問い合わせ言語）としては SQL がある。

問 31

データベースのモデルの一つであるオブジェクトデータベースに関する説明として、不適切なものを選択肢の中から一つ選べ。

- 1 データをオブジェクトと呼ばれる単位で管理する。オブジェクトの内部のデータはカプセル化（隠蔽）されており、外部からはあらかじめ提供されたメソッドを通じてしかアクセスできない。
- 2 オブジェクトの特性は、インヘリタンス（継承）と呼ばれる機構で別の種類のオブジェクトの定義に引き継ぐことができ、複雑なデータ構造を取り扱う場合などに、過去に開発したものを有効に利用しやすい。
- 3 オブジェクトデータベースを管理するソフトウェアは、ODBMS と呼ばれる。
- 4 リレーショナルデータベースとオブジェクトデータベースでは、データの基本単位の取扱が異なるため、これらを混用することができない。

問 32

下記の表に対して、次に示す **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 アクセション **FROM** インフルエンザウイルス塩基配列
WHERE サブタイプ = 'H1N1' **AND** 年号 >=1976 **AND** 年号 < 2005 ;

表： インフルエンザウイルス塩基配列

アクセション	宿主	タンパク質	サブタイプ	国・地域	年号
AAD17229	Human	HA	H1N1	USA	1918
ABD95350	Human	HA	H1N1	Russia	1977
ACF54400	Human	HA	H3N2	Hong Kong	1968
ACP41105	Human	HA	H1N1	USA	2009
ACU79959	Human	HA	H2N2	Japan	1957
AAF75994	Swine	HA	H1N2	USA	1999
AAB39851	Swine	HA	H1N1	USA	1976
AAD25304	Avian	HA	H1N1	Canada	1976
AAT65329	Avian	HA	H1N1	Canada	1998

- 1 AAD17229, ABD95350, ACP41105, AAB39851, AAD25304, AAT65329
- 2 ABD95350, ACP41105, AAB39851, AAD25304, AAT65329
- 3 ABD95350, ACP41105, AAF75994, AAB39851, AAD25304, AAT65329
- 4 ABD95350, AAB39851, AAD25304, AAT65329

問 33

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、不適切なものを選択肢の中から一つ選べ。

- 1 5回サイコロをふって1が4回、2が1回出る確率は $\frac{5}{6^5}$ である。
- 2 5回サイコロをふ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17.5である。
- 3 4回サイコロをふって4回とも奇数が出たとき、5回目は偶数が出る確率のほうが高い。
- 4 2回サイコロをふるとする。1回目のサイコロの目の値と、2回目のサイコロの目の値が等しくなる確率は $1/6$ である。

問 34

重さ 20 g（グラム）の物体を天秤で測定する場合を考える。その天秤での測定には誤差が生じ、その誤差は平均 0 g、分散 0.04 g^2 の正規分布 $N(0, 0.04)$ にしたがうものとする。このとき誤差の絶対値が 0.4 g 以上になる確率はいくらになるか。もっとも適切なものを選択肢の中から一つ選べ。必要であれば下記の標準正規分布表を利用してよい。標準正規分布表の数字は、 z で示された位置よりも上側の確率を示している。

z	0	0.2	0.4	0.6	0.8
0.0	0.5000	0.4207	0.3446	0.2743	0.2119
1.0	0.1587	0.1151	0.0808	0.0548	0.0359
2.0	0.0228	0.0139	0.0082	0.0047	0.0026
3.0	0.0013	0.0007	0.0003	0.0002	0.0001

- 1 0.0456
- 2 0.0082
- 3 0.2119
- 4 0.3446

問 35

2つの確率変数 X 、 Y に関する記述において、不適切なものを選択肢の中から一つ選べ。

それぞれの記号の意味は次のとおり。

$V(A)$ A の分散

$\text{Cov}(A, B)$ A と B の共分散

$E(A)$ A の期待値

- 1 X^2 の期待値 $E(X^2)$ は 5 で、 X の平均 $E(X)$ は 1 のとき、 X の標準偏差は 4 である。
- 2 X と Y が独立のときは必ず、 X と Y の共分散 $\text{Cov}(X, Y)$ は 0 である。
- 3 X と Y が独立のときは必ず、 X と Y の和の期待値 $E(X + Y)$ は X の期待値 $E(X)$ と Y の期待値 $E(Y)$ の和に等しい。すなわち、 $E(X + Y) = E(X) + E(Y)$ が成り立つ。
- 4 X と Y が独立のときは必ず、 X と Y の和の分散 $V(X + Y)$ は X の分散 $V(X)$ と Y の分散 $V(Y)$ の和に等しい。すなわち、 $V(X + Y) = V(X) + V(Y)$ が成り立つ。

問 36

統計的検定の説明として不適切なものを選択肢の中から一つ選べ。

- 1 有意水準 5% で P -値が 3% のとき、帰無仮説は棄却される。
- 2 ある確率分布に従う母集団の標本平均は、つねに母平均に等しく母分散の値によらない。
- 3 母平均 10 の正規分布に従う母集団の標本平均の期待値は、母分散の値によらず必ず 10 である。
- 4 2 群の平均値の差の検定において、「2 群の母平均値に差がない」という帰無仮説が棄却されたとき、2 群の平均値は統計的に有意に差があるといえる。

問 37

サポートベクトルマシン (SVM) に関する説明として、不適切なものを選択肢の中から一つ選べ。

- 1 SVM で線形カーネルを使う場合、正則化パラメータ C が十分に大きければ識別面は曲面になる。
- 2 正則化パラメータ C が十分大きいとき、SVM はマージンが最大となる識別面を見つける。
- 3 多項式カーネルは、 $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + c)^p$ とあらわされる。 $p=1$ 、 $c=0$ とすると、識別面は必ず線形となる。
- 4 SVM の学習アルゴリズムは、凸最適化問題で表される。

問 38

3 層パーセプトロンに関する説明として、不適切なものを選択肢の中から一つ選べ。

- 1 3 層パーセプトロンは、二分探索法的一种である。
- 2 3 層パーセプトロンは、入力層、中間層、出力層からなる。
- 3 3 層パーセプトロンの学習には、誤差逆伝搬法と呼ばれる勾配法がしばしば使われる。
- 4 3 層パーセプトロンは、人工ニューラルネットワークの一つである。

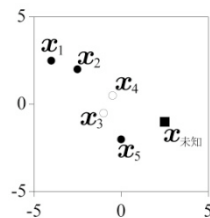
問 39

2次元の特徴ベクトルで表された5個のデータ $\{\mathbf{x}_i \mid i = 1, \dots, 5\}$ がある。各データには、 $y_i \in \{\pm 1\}$ として2クラスのラベル（+1=陽性、-1=陰性）が付けられている。5個のデータとラベルの具体的な値は、以下のとおりである。

$$(\mathbf{x}_1, y_1) = \left(\begin{bmatrix} -4 \\ +2.5 \end{bmatrix}, +1 \right), \quad (\mathbf{x}_2, y_2) = \left(\begin{bmatrix} -2.5 \\ +2 \end{bmatrix}, +1 \right), \quad (\mathbf{x}_3, y_3) = \left(\begin{bmatrix} -1 \\ -0.5 \end{bmatrix}, -1 \right),$$

$$(\mathbf{x}_4, y_4) = \left(\begin{bmatrix} -0.5 \\ +0.5 \end{bmatrix}, -1 \right), \quad (\mathbf{x}_5, y_5) = \left(\begin{bmatrix} 0 \\ -2 \end{bmatrix}, +1 \right)$$

これをプロットすると下図のようになる。（黒丸が陽性、白丸が陰性を表している）



ここで、 k -近傍法 (k -nearest neighbor) を用いて、未知データ

$$\mathbf{x}_{\text{未知}} = \begin{bmatrix} +2.5 \\ -1 \end{bmatrix}$$

のクラスラベル（陽性、陰性）を予測するために、各訓練用データと未知データとのユークリッド距離を計算したところ、次のようになった。

$$\|\mathbf{x}_1 - \mathbf{x}_{\text{未知}}\| = 7.38, \quad \|\mathbf{x}_2 - \mathbf{x}_{\text{未知}}\| = 5.38, \quad \|\mathbf{x}_3 - \mathbf{x}_{\text{未知}}\| = 3.54,$$

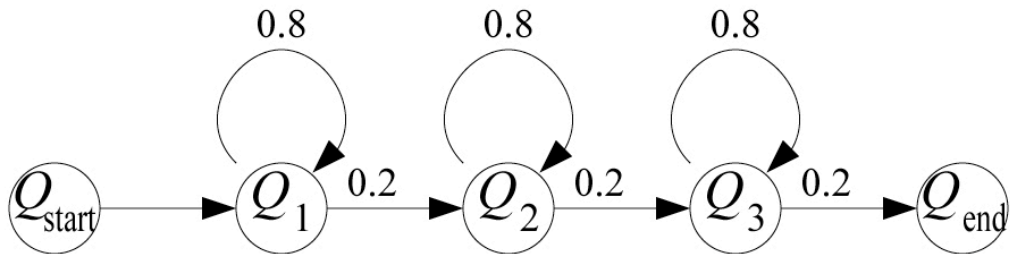
$$\|\mathbf{x}_4 - \mathbf{x}_{\text{未知}}\| = 3.35, \quad \|\mathbf{x}_5 - \mathbf{x}_{\text{未知}}\| = 2.69$$

k -近傍法を用いた予測結果として、もっとも適切なものを選択肢の中から一つ選べ。
ただし k -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識別法である。

- 1 近傍数 $k=1$ の場合でも、近傍数 $k=3$ の場合でも、予測結果は陽性となる。
- 2 近傍数 $k=1$ の場合でも、近傍数 $k=3$ の場合でも、予測結果は陰性となる。
- 3 近傍数 $k=1$ の場合の予測結果は陽性だが、近傍数 $k=3$ の場合の予測結果は陰性となる。
- 4 近傍数 $k=1$ の場合の予測結果は陰性だが、近傍数 $k=3$ の場合の予測結果は陽性となる。

問 40

図のような Left-to-right 型の隠れマルコフモデルを考える。



隠れマルコフモデルの遷移確率を行列 $\mathbf{A} \in R^{3 \times 4}$ を用いて、

$$\forall i, \forall j \in \{1, 2, 3\}: \quad A_{i,j} = P(Q_j | Q_i),$$

$$\forall i \in \{1, 2, 3\}: \quad A_{i,4} = P(Q_{\text{end}} | Q_i)$$

と表し、出力確率を行列 $\mathbf{B} \in R^{3 \times 4}$ を用いて、

$$\forall i \in \{1, 2, 3\}: \quad$$

$$B_{i,1} = P(A | Q_i), \quad B_{i,2} = P(T | Q_i), \quad B_{i,3} = P(G | Q_i), \quad B_{i,4} = P(C | Q_i)$$

と表すとする。

遷移確率と出力確率の値が、

$$\mathbf{A} = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0.8 & 0.2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 \\ 0 & 0.1 & 0.3 & 0.6 \end{bmatrix}$$

と与えられるとする。

このとき観測系列 GCAT を生成する確率が最も大きいパスはどれか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 $Q_{\text{start}} \rightarrow Q_1 \rightarrow Q_1 \rightarrow Q_2 \rightarrow Q_3 \rightarrow Q_{\text{end}}$
- 2 $Q_{\text{start}} \rightarrow Q_1 \rightarrow Q_2 \rightarrow Q_2 \rightarrow Q_3 \rightarrow Q_{\text{end}}$
- 3 $Q_{\text{start}} \rightarrow Q_1 \rightarrow Q_2 \rightarrow Q_3 \rightarrow Q_3 \rightarrow Q_{\text{end}}$
- 4 観測系列 GCAT を出力できるパスは存在しない。

問 41

以下に示すデータは、フラットファイル形式で記された UniProt エントリーの抜粋である。
ここから読み取れる情報として不適切なものを選択肢の中から一つ選べ。

```
ID   PRIO_HUMAN              Reviewed;           253 AA.
AC   P04156; O60489; P78446; Q15216; Q15221; Q27H91; Q8TBG0; Q96E70;
DT   01-NOV-1986, integrated into UniProtKB/Swiss-Prot.
DE   RecName: Full=Major prion protein;
OS   Homo sapiens (Human).
RC   TISSUE=Brain;
CC   -!- FUNCTION: The physiological function of PrP is not known.
CC   -!- SUBUNIT: PrP has a tendency to aggregate yielding polymers
CC   -!- SUBCELLULAR LOCATION: Cell membrane; Lipid-anchor, GPI-anchor.
CC       Golgi apparatus (By similarity).
CC   -!- DISEASE: PrP is found in high quantity in the brain of humans and
CC       animals infected with neurodegenerative diseases known as
CC       transmissible spongiform encephalopathies or prion diseases
DR   EMBL; M13899; AAA60182.1; -; mRNA.
DR   RefSeq; NP_000302.1; -.
DR   UniGene; Hs.472010; -.
DR   PDB; 1E1G; NMR; -; A=125-228.
DR   HGNC; HGNC:9449; PRNP.
DR   MIM; 123400; phenotype.
KW   3D-structure; Cell membrane; Complete proteome;
KW   Direct protein sequencing; Disease mutation; Disulfide bond;
KW   Polymorphism; Prion; Repeat; Signal.
FT   SIGNAL      1         22
SQ   SEQUENCE    253 AA; 27661 MW; 43DB596BAAA66484 CRC64;
      MANLGCWMLV L FVATWSDLG LCKKRPKPGG WNTGGSRYPG QGSPGGNRYP PQGGGGWGQP
      HGGGWGQPHG GGWGQPHGGG WGQPHGGGWG QGGGTHSQWN KPSKPKTNMK HMAGAAAAGA
      VVGGLGGYML GSAMSRPIIH FGSDYEDRYR RENMHRYPNQ VYYRPMDEYS NQNNFVHDCV
      NITIKQHTVT TTTKGENFTE TDVKMMERVV EQMCITQYER ESQAYYQGRS SMVLFSPPV
      ILLISFLIFL IVG
//
```

- 1 ヒト由来のタンパク質で、主に脳で発現する。細胞内では細胞膜に局在する。
- 2 部分的に立体構造が決定されており、PDB に登録されている。
- 3 伝達性海綿状脳症などの神経障害性の病気に関わるタンパク質である。
- 4 未成熟タンパク質全長は 253 アミノ酸残基であるが、少なくとも C 末端の 22 残基が切り取られる。

問 42

相同性検索に用いられる BLAST には、クエリ配列と検索対象配列データベースの組み合わせによって使い分けるいくつかのバージョンがある。それらのうち、blastx プログラムにおけるクエリ配列と検索対象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。

	クエリ配列	検索対象データベース
1	塩基配列	塩基配列
2	塩基配列	アミノ酸配列
3	アミノ酸配列	塩基配列
4	アミノ酸配列	アミノ酸配列

問 43

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。一般的なモチーフ配列の表現方法に正規表現がある。以下は正規表現の例と記述のルールである。

正規表現： $C-X-X-C-x(5,7)-C-[FWY]-[LIV]-C$

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）を、- 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。

上記の正規表現に一致するアミノ酸配列を選択肢の中から一つ選べ。

- 1 CTGCGGCFVC
- 2 CLIHGGENDRTCWVC
- 3 CTMCGGEHNNRTCFVC
- 4 CTICGGENNRTCFVC

問 44

下に示す行列は塩基置換行列である。この塩基置換行列を適用した時、以下の塩基配列のアラインメントのスコアとして正しいものを選択肢の中から一つ選べ。ただしギャップ開始ペナルティを-2 点、ギャップ伸長ペナルティを 0 点とする。

	A	T	G	C
A	2	0	1	0
T	0	2	0	1
G	1	0	2	0
C	0	1	0	2

AATGTCCGT-TGC

AAAAT--GTATGC

- 1 11
- 2 13
- 3 14
- 4 15

問 45

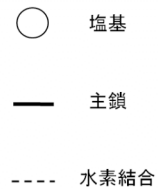
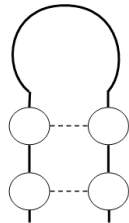
DNA には回文 (palindrome) 配列をもつ部分があり、制限酵素の認識部位によくみられる。以下の二本鎖 DNA 配列のうち、回文配列であるものを選択肢の中から一つ選べ。

- 1 ATGGGCCCCGGGTA
TACCCGGGGCCCAT
- 2 ATGGGCCTTTAAAG
TACCCGGAATTTT
- 3 ATGGGCCGGCCCAT
TACCCGGCCGGGTA
- 4 ATGGGCCATGGGCC
TACCCGGTACCCGG

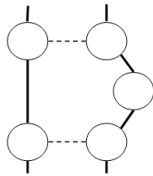
問 46

以下に示す RNA 二次構造のモチーフ A～C にもっとも適した用語の組合せを選択肢の中から一つ選べ。

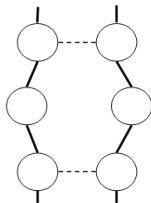
A:



B:



C:



- | | | |
|---------------|-------------|------------|
| 1 A : バルジループ | B : ヘアピンループ | C : 内部ループ |
| 2 A : ヘアピンループ | B : バルジループ | C : 内部ループ |
| 3 A : 内部ループ | B : ヘアピンループ | C : バルジループ |
| 4 A : ヘアピンループ | B : 内部ループ | C : バルジループ |

問 47

以下の選択肢は、RNA 二次構造における塩基対結合の強さの関係を示したものである。正しい関係を示す選択肢を一つ選べ。ただし、結合がより強い（安定な）ものは不等号＞の左辺に書くこととする。

- 1 A·U > G·U > C·G
- 2 G·U > A·U > C·G
- 3 C·G > A·U > G·U
- 4 C·G > G·U > A·U

問 48

マイクロ RNA に関する以下の記述で、(a) ～ (c) 内に入る語句の組合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。

マイクロ RNA は、(a) 塩基長の低分子 RNA である。マイクロ RNA の典型的な性質は、シード領域 (b) 塩基が、標的となるメッセンジャーRNA の (c) に相補的に結合し、標的メッセンジャーRNA の翻訳を阻害することである。

- 1 (a) 20～22 (b) 7 (c) 3'UTR
- 2 (a) 22～25 (b) 18 (c) CDS
- 3 (a) 20～22 (b) 7 (c) CDS
- 4 (a) 22～25 (b) 18 (c) 3'UTR

問 49

以下の A～E は、配列比較に関する記述である。正しい記述を示す組み合わせはどれか。選択肢の中から一つ選べ。

- A アラインメントによって、配列間の類似性を定量的に評価することができる。
- B ホモロジー検索は、大域アラインメントによって行われる。
- C 比較する配列が短いものを局所アラインメント、長いものを大域アラインメントとよぶ。
- D 比較する配列が 1 対 1 のアラインメントをペアワイズアラインメントという。
- E ペアワイズアラインメントの計算量は $O(L)$ である。(L は配列長)

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

問 50

以下の A～E は配列アラインメントの評価についての記述である。正しい記述を示す組み合わせはどれか。選択肢の中から一つ選べ。

- A 最適アラインメントは、比較する配列間の進化的距離が最大になるようなアラインメントである。
- B 置換スコアとは、点突然変異を得点化したものである。
- C BLOSUM は核酸配列のための置換行列である。
- D アミノ酸置換行列 PAM20 と PAM70 では、PAM70 の方が進化的に遠い配列のアラインメントの評価に適している。
- E 一般にギャップ開始ペナルティはギャップ伸長ペナルティより大きく設定される。

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

問 51

動的計画法 (dynamic programming) が配列のアラインメント計算に適用できる理由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 アミノ酸の種類は高々20種類であるから。
- 2 アラインメントスコアは列ごとのスコアの和で表現されているから。
- 3 整数演算は浮動小数点演算より速いから。
- 4 配列の進化速度は時代によって変化するから。

問 52

以下の A～E は配列のマルチプルアラインメントに関する記述である。正しい記述を示す組み合わせはどれか。選択肢の中から一つ選べ。

- A マルチプルアラインメントは 2 本の配列をアラインメントすることである。
- B マルチプルアラインメントの評価指標には、最小エントロピーと SP (sum of pairs) スコアなどがある。
- C マルチプルアラインメントは、配列中の生物学的に重要と思われる部位を特定し、配列間の進化的な関連を明らかにすることに利用される。
- D 3 次元動的計画法による 3 本のマルチプルアラインメントに要する計算量は、 $O(L^2)$ である。(L は配列長)
- E 4 本以上の配列のマルチプルアラインメントを求める手法としては、多次元動的計画法がもっとも高速なアルゴリズムである。

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

問 53

2000 年にある研究者が問い合わせ配列 q を当時のデータベースに対して類似配列検索を行い、ある配列 h がスコア(s_{00}) と e -値 (e_{00}) のヒットとして見つかった。2010 年のデータベース(d_{10}) は 2000 年のデータベースに新しい配列を多数加えたものとする。 d_{10} に対して以前と同じパラメータ設定で q の相同性検索を行った場合、同じ配列 h のスコアと e -値を s_{10} 、 e_{10} で表すことにする。予想される結果について適切なものを選択肢の中から一つ選べ。

- 1 $s_{10} > s_{00}$, $e_{10} > e_{00}$
- 2 $s_{10} = s_{00}$, $e_{10} < e_{00}$
- 3 $s_{10} = s_{00}$, $e_{10} > e_{00}$
- 4 $s_{10} > s_{00}$, $e_{10} < e_{00}$

問 54

真核生物の選択的スプライシングなどによりできるタンパク質アイソフォームが、異なる細胞小器官に局在することがある。配列の違いとして、異なる局在化をもっとも強く示唆するものを選択肢の中から一つ選べ。

- 1 配列の途中で 20 残基程度の欠損がある。
- 2 配列の三個目のグリシン付近に 20 残基程度の欠損がある。
- 3 配列の N 末端付近に 20 残基程度の欠損がある。
- 4 配列の C 末端付近に 20 残基程度の欠損がある。

問 55

ドットマトリックス図（ドットプロットとも呼ばれる）は配列の繰り返し部分や配列間の類似部分を可視化する方法である。以下の図表は 4 塩基長のウィンドウで完全一致を条件として作成したドットマトリックス図である。どの配列の組み合わせを表したかを選択肢の中から一つ選べ。なお、このドットマトリックス図では相補鎖も考慮されている。

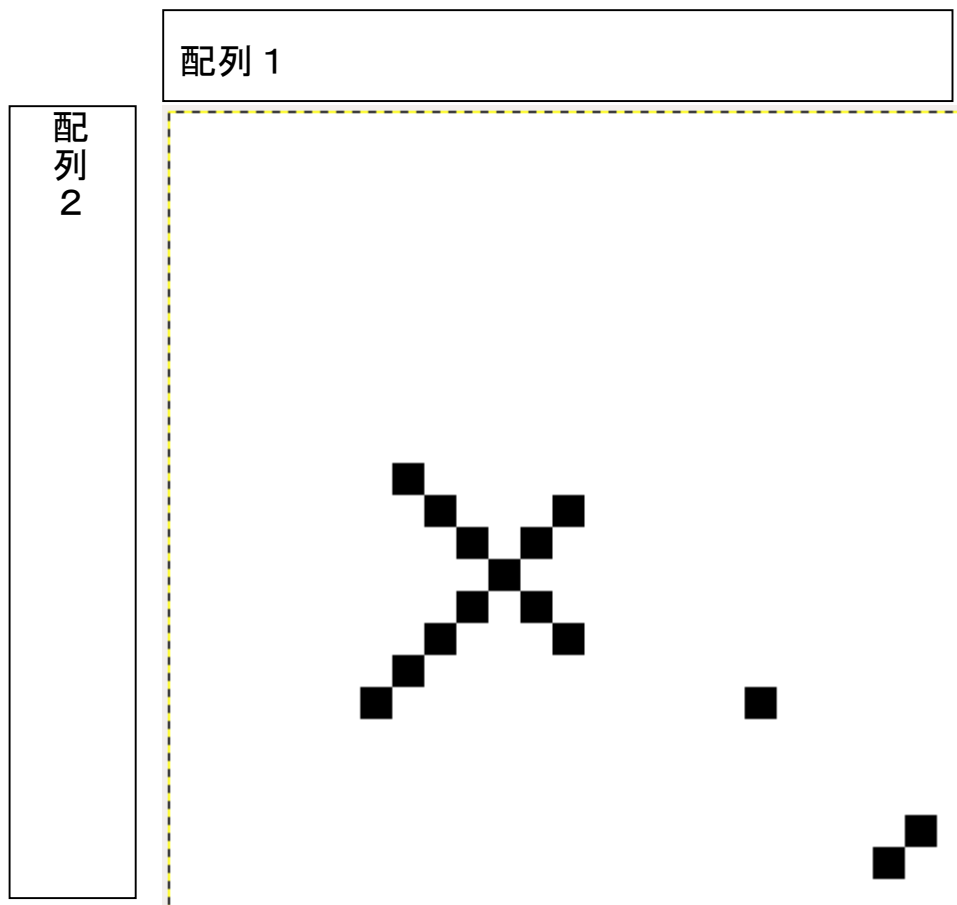
ア TTTTGCTGTATATACTCACAGCATAA

イ TGGAGTACTAGTGTGTGCAGCTGCCC

ウ TTCCTGCCCTCGCCCGCATTCCGATT

エ GCCCCCTGGGTGTATATACAGTTATG

	配列 1	配列 2
1	ウ	ア
2	イ	ウ
3	エ	ア
4	ウ	エ



問 56

タンパク質の天然変性領域を予測するソフトウェアとして適切なものを選択肢の中から一つ選べ。

- 1 AMBER
- 2 DISOPRED
- 3 BLAST
- 4 HMMPRED

問 57

タンパク質の立体構造を図示するソフトウェアとして不適切なものを選択肢の中から一つ選べ。

- 1 Rasmol
- 2 3DCoffee
- 3 Molscript
- 4 PyMOL

問 58

次のタンパク質の二次構造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 α ヘリックスも β シートも主鎖原子間の水素結合が安定化に寄与している。
- 2 β シートでは、平行 (parallel) ストランドと反平行 (anti-parallel) ストランドの出現頻度に大きな差はない。
- 3 α ヘリックスでは、右巻き (right-handed) と左巻き (left-handed) の出現頻度に大きな差はない。
- 4 一次構造 (アミノ酸配列) が保存されている場合、その部分の二次構造も保存されていることが多い。

問 59

以下の図は抗インフルエンザウイルス薬オセルタミビルのターゲットタンパク質であるノイラミニダーゼ (neuraminidase) の立体構造の情報を収めた PDB エントリーの一部抜粋である。この図の説明として不適切なものを選択肢の中から一つ選べ。

```

HEADER      HYDROLASE                               26-JUL-06   2HU4
TITLE       N1 NEURAMINIDASE IN COMPLEX WITH OSELTAMIVIR 2
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: NEURAMINIDASE;
COMPND      3 CHAIN: A, B, C, D, E, F, G, H;
COMPND      4 SYNONYM: NEURAMINIDASE NA;
COMPND      5 MUTATION: YES
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: INFLUENZA A VIRUS;
SOURCE      3 ORGANISM_TAXID: 11320
KEYWDS      N1, NEURAMINIDASE, OSELTAMIVIR, HYDROLASE
EXPDTA      X-RAY DIFFRACTION
AUTHOR      R.J.RUSSELL,L.F.HAIRE,D.J.STEVENS,P.J.COLLINS,Y.P.LIN,
AUTHOR      2 G.M.BLACKBURN,A.J.HAY,S.J.GAMBLIN,J.J.SKEHEL
REVSTAT     3 24-FEB-09 2HU4 1 VERSN
REVSTAT     2 17-OCT-06 2HU4 1 JRNL
REVSTAT     1 05-SEP-06 2HU4 0
JRNL        AUTH R.J.RUSSELL,L.F.HAIRE,D.J.STEVENS,P.J.COLLINS,
JRNL        AUTH 2 Y.P.LIN,G.M.BLACKBURN,A.J.HAY,S.J.GAMBLIN,
JRNL        AUTH 3 J.J.SKEHEL
JRNL        TITL THE STRUCTURE OF H5N1 AVIAN INFLUENZA
JRNL        TITL 2 NEURAMINIDASE SUGGESTS NEW OPPORTUNITIES FOR DRUG
JRNL        TITL 3 DESIGN.
JRNL        REF NATURE V. 443 45 2006
JRNL        REFN ISSN 0028-0836
JRNL        PMID 16915238
JRNL        DOI 10.1038/NATURE05114
REMARK      1
REMARK      2
REMARK      2 RESOLUTION. 2.50 ANGSTROMS.
REMARK      3

中略

HETNAM      G39 5-N-ACETYL-3-(1-ETHYLPROPYL)-1-CYCLOHEXENE-1-
HETNAM      2 G39 CARBOXYLIC ACID
HETSYN      G39 OSELTAMIVIR
FORMUL      9 G39 8(C14 H24 N2 O4)

中略

SITE        1 AC1 13 ARG A 118 GLU A 119 ASP A 151 ARG A 152
SITE        2 AC1 13 TRP A 178 SER A 246 GLU A 276 GLU A 277
SITE        3 AC1 13 ARG A 292 ASN A 294 TYR A 347 ARG A 371
SITE        4 AC1 13 TYR A 406

中略

CRYST1      201.060 201.240 211.740 90.00 90.00 90.00 C 2 2 21 64
ORIGX1      1.000000 0.000000 0.000000 0.000000
ORIGX2      0.000000 1.000000 0.000000 0.000000
ORIGX3      0.000000 0.000000 1.000000 0.000000
SCALE1      0.004974 0.000000 0.000000 0.000000
SCALE2      0.000000 0.004969 0.000000 0.000000
SCALE3      0.000000 0.000000 0.004723 0.000000
ATOM        1 N VAL A 83 -0.603 65.642 77.183 1.00 23.27 N
ATOM        2 CA VAL A 83 -0.883 67.005 76.630 1.00 24.70 C
ATOM        3 C VAL A 83 -1.221 68.015 77.732 1.00 24.24 C
ATOM        4 O VAL A 83 -1.851 67.665 78.731 1.00 23.78 O
ATOM        5 CB VAL A 83 -2.050 66.965 75.617 1.00 25.58 C
ATOM        6 CG1 VAL A 83 -2.332 68.365 75.094 1.00 27.08 C
ATOM        7 CG2 VAL A 83 -1.694 66.047 74.443 1.00 26.86 C
ATOM        8 N LYS A 84 -0.803 69.266 77.541 1.00 23.15 N
ATOM        9 CA LYS A 84 -1.047 70.312 78.526 1.00 23.28 C
ATOM        10 C LYS A 84 -2.520 70.643 78.728 1.00 22.34 C
ATOM        11 O LYS A 84 -3.329 70.575 77.803 1.00 21.41 O
ATOM        12 CB LYS A 84 -0.293 71.609 78.170 1.00 25.35 C
ATOM        13 CG LYS A 84 1.234 71.495 78.128 1.00 27.98 C
  
```

後略

- 1 この PDB のエントリーID は 2HU4 である。
- 2 このエントリーにはポリペプチド鎖が 8 本含まれている。
- 3 このエントリーは X 線の結晶構造解析の結果による。
- 4 83 番のバリンの N 原子の Z 座標は 23.27 である。

問 60

1994 年に開始された、タンパク質の立体構造予測法の国際的な評価会の略称は次のうちどれか。適切なものを選択肢の中から一つ選べ。

- 1 CoMFA
- 2 CAMDA
- 3 CASP
- 4 CAPRI

問 61

タンパク質の主鎖に添って定義される二面角のうち、 ϕ （窒素原子と α 炭素原子の間の二面角）と、 ψ （ α 炭素原子とカルボニル基の炭素原子の間の二面角）の値をそれぞれのアミノ酸残基ごとに、二次元でプロットしたものをラマチャンドラン（Ramachandran）プロットと呼ぶ。天然のタンパク質では二次構造に関連した独特の分布を示すが、この分布が大きく異なるアミノ酸残基が二つある。その一つはグリシンであるが、もう一つは以下のどれか。適切なものを選択肢の中から一つ選べ。

- 1 ヒスチジン
- 2 グルタミン酸
- 3 アルギニン
- 4 プロリン

問 62

立体構造の分類データベース SCOP と CATH に関する次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 SCOP データベースは Alexey G. Murzin を中心とするエキスパートの構造観察により分類されている。
- 2 CATH データベースは、エキスパートによる観察と構造比較プログラム SSAP を併用して分類がなされている。
- 3 SCOP と CATH は階層的な分類を採用しており、最上位階層は共に二次構造の含量に着目した Class の分類である。
- 4 SCOP での分類では、多くのフォールドが 10 以上のスーパーファミリーを含んでいる。

問 63

バイオインフォマティクスの素材として生体高分子の立体構造情報を考えたとき、その信頼性は大きな問題である。X 線結晶構造解析で得られた原子座標の精度に関して、他の条件は同一だとした時に、以下の記述で不適切なものを選択肢の中から一つ選べ。

- 1 精密化に用いられた回折データの解像度が高い方が座標の精度は高い。
- 2 各原子の温度因子が高い方が座標の精度は高い。
- 3 R_{free} -値が低い方が座標の精度は高い。
- 4 類似タンパク質の構造を用いて分子置換法で決定された構造でも、重原子同形置換法など実験的に位相を決めた構造でも、精密化後の座標の精度に大きな差はない。

問 64

ハーディ・ワインバーグ平衡 (HWE) とは、2 倍体生物の遺伝子型に観測される状態のことである。ある集団のある 2 アレル型の座位のアレル頻度と 2 倍体遺伝子型の関係が HWE を満足しているかどうかを調べている。この調査に関連する記述に関し、もっとも不適切なものを選択肢から一つ選べ。

- 1 2 つのアレルの頻度が p_1 と p_2 であるとする、HWE のときこの 2 つのアレルが作るヘテロ接合体の HWE における頻度は $2p_1 p_2$ である。
- 2 近親交配があるとき、ヘテロ接合体の割合が多くなる。
- 3 集団が HWE を満足する複数の亜集団からなり、亜集団間の移動がないとき、ホモ接合体の割合が多くなる。
- 4 HWE を満足しない場合、その理由の一つとして、集団が小さい可能性が挙げられる。

問 65

HAPMAP プロジェクトとは、ヒトゲノム上の SNP の頻度分布、その連鎖不平衡構造についてのデータの収集・公開を目的とした国際プロジェクトである。その HAPMAP プロジェクトに関する記述に関し、もっとも不適切なものを選択肢から一つ選べ。

- 1 3 つの異なる民族に関するゲノムワイドな遺伝子型データが公開されている。
- 2 アジア系サンプルは東アジアの 2 カ国からのサンプルである。
- 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サンプルもハプロタイプの推定精度が高い。

問 66

有病者の数が比較的多い疾患の多くは、複数の遺伝要因と複数の環境要因との影響を受けた複合遺伝性疾患であると考えられている。この複合遺伝性疾患の遺伝子解析に用いられる手法に連鎖不平衡マッピング解析がある。以下の連鎖不平衡と連鎖不平衡マッピング解析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 連鎖不平衡は複数の多型もしくはそのアレルの間の非独立な関係であって、家系に対して定義されるものである。
- 2 同一染色体上にある多型の連鎖不平衡は、距離が遠くなるにしたがって弱くなる傾向がある。
- 3 連鎖不平衡は異なる染色体上の多型の間にも観測されるが、それは、同一染色体上の連鎖不平衡とは異なる理由によるものである。
- 4 連鎖不平衡マッピング解析においては、ある多型の代用として別の多型を用いることができる。

問 67

ヒトゲノム DNA 配列中には、様々な多型が存在することが知られている。次に示した用語のうちヒトゲノムにおける多型の呼び名として適切でないものを一つ選べ。

- 1 マイクロサテライト
- 2 テロメア
- 3 SNP
- 4 VNTR

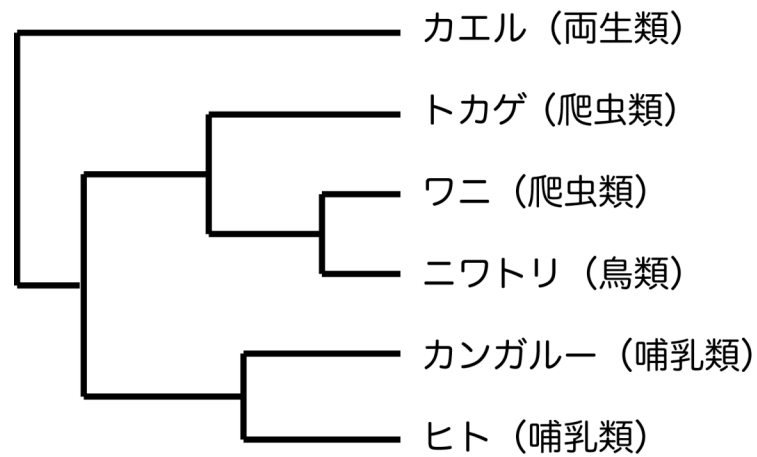
問 68

分子進化の中立説によれば、進化に関与する変異の多くは自然選択に対して有利でも不利でもない、中立なものであると考えられる。以下のタンパク質をコードする遺伝子の進化速度についての観察事実の中から、中立説で説明できるものとしてもっとも不適切なものを選択肢から一つ選べ。

- 1 何らかの原因で機能を失ってしまった、いわゆる偽遺伝子は非常に速い進化速度で配列が変化する。
- 2 タンパク質のアミノ酸配列をコードする領域の塩基配列では一般に、同義置換は非同義置換に比べて頻繁に起っている。
- 3 生物間でタンパク質のアミノ酸配列を比較すると、電荷や大きさといった性質が似たアミノ酸の間での置換は、これらの性質が大きく異なる置換よりも頻繁に起っている。
- 4 ウイルスの表面抗原において、抗体によって認識される部位ではその認識を避けるためにアミノ酸置換が加速されることがある。

問 69

下に示す系統樹は、カエルをアウトグループ（外群）とした無根系統樹である。この系統樹から読み取れる情報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、この系統樹の枝の長さは意味を持たない。



- 1 トカゲとニワトリの分岐はワニとニワトリの分岐よりも前に起こっている。
- 2 哺乳類は単系統（他の分類群を含まない、共通の祖先に由来する生物群）である。
- 3 爬虫類は側系統（共通祖先に由来する子孫を構成要素とする生物群から、特定の単一系統を除いた残りをまとめた生物群）である。
- 4 ワニよりもニワトリの方がヒトに近縁である。

問 70

次の文章はアミノ酸配列の相違度から座位当たりの置換数を推定する方法の 1 つであるポアソン補正法について述べたものである。文章中の[イ]、[ロ] 内に入る式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。

二つのアミノ酸配列の間の相違度 K (異なるアミノ酸座位数/比較した座位数) から、進化の過程で起こった座位当たりの置換率 k を推定するには、同じ座位に 2 回以上置換が起こること (多重置換) を考慮する必要がある。なお、ここでは復帰突然変異は無視するものとする。

座位当たりの置換率が k であるとき、アミノ酸置換は平均が k のポアソン分布にしたがって起こると仮定する。このとき、ある座位で n 回の置換が起こる確率 $P(n)$ は、

$$P(n) = \frac{e^{-k} \cdot k^n}{n!}$$

と表すことができる。ここで、置換が起こらない座位 ($n=0$) は、

$$P(0) = e^{-k}$$

の割合だけ観察されることになる。これは、[イ] として測定できるので、

$$[\text{イ}] = e^{-k}$$

したがって、

$$k = [\text{ロ}]$$

となる。

- | | | | | |
|---|-----|-------|-----|-------------|
| 1 | [イ] | $1-K$ | [ロ] | $-\ln(1-K)$ |
| 2 | [イ] | $1-K$ | [ロ] | $-\ln K$ |
| 3 | [イ] | K | [ロ] | $-\ln(1-K)$ |
| 4 | [イ] | K | [ロ] | $-\ln K$ |

問 71

系統樹を表現する方法としてニューウィック形式 (Newick format) と呼ばれる括弧を利用した入れ子表現がある。これは括弧で閉じられた OTU (数字で表された遺伝子、種など) が優先的に連結されることを示す表現である。以下の選択肢の中で、ニューウィック形式 ((1,2),(3,4),5) で表される系統樹と樹形が異なるものを選択肢の中から一つ選べ

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

問 72

異なる生物のゲノムの比較に関連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 種分化に伴って祖先遺伝子から分岐して生じた相同な遺伝子の間の関係をオーソログスな関係とよぶ。
- 2 ゲノムアラインメントとは、オーソログスな遺伝子のゲノム上の位置に基づいて、それらの遺伝子をもっとも対応する位置にくるように、二つの全ゲノム構造を配置することである。
- 3 多数のパラログス遺伝子が縦列に反復している領域をシンテニック領域とよぶ。
- 4 ゲノムリアレンジメントとは、逆位、転座などのように、染色体のある範囲の断片の向きや位置が変化することさす。

問 73

複数のマイクロアレイデータを比較するために行われる補正には、幾つかの方法が考えられる。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 グローバル法とは、プローブ全体から計算される値を用いて発現量を補正する方法である。
- 2 スパイクインコントロール法とは、ラベリング後の RNA 溶液に濃度が既知の RNA を加え、その RNA の蛍光強度を用いて発現量を補正する方法である。
- 3 インターナルコントロール法とは、アクチン等の発現変動のない遺伝子のデータ値を用いて発現量を補正する方法である。
- 4 分位点法 (Quantile 法) とは、プローブ全体の蛍光強度分布を用いて発現量を補正する方法である。

問 74

タイリングアレイ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 タイリングアレイでは遺伝子の構造とは無関係にプローブを設計することができる。
- 2 タイリングアレイはゲノムが読まれていない生物種に対しても設計することができる。
- 3 タイリングアレイを新規の遺伝子の探索に利用することができる。
- 4 タイリングアレイをクロマチン免疫沈降法と組み合わせた実験に利用することができる。

問 75

サンプル A および B をマイクロアレイに競合的にハイブリダイゼーションしたところ、遺伝子 X に関してはサンプル B 由来のシグナルがサンプル A 由来のシグナルに比べて高い値を示した。この結果を確認するため、リアルタイム PCR 法で遺伝子 X の発現量を調べたところ、サンプル B の Ct 値（増幅曲線がある一定のシグナル値を超えるサイクル数）がサンプル A の Ct 値より 2 だけ低かった。これから導かれる結論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 遺伝子 X はサンプル B においてサンプル A の 4 倍量発現している。
- 2 遺伝子 X はサンプル B においてサンプル A の 2 倍量発現している。
- 3 遺伝子 X はサンプル A においてサンプル B の 4 倍量発現している。
- 4 遺伝子 X はサンプル A においてサンプル B の 2 倍量発現している。

問 76

質量分析計を用いたプロテオーム解析に関連が深いプログラム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 BLASTP と DALI
- 2 DALI と MASCOT
- 3 MASCOT と SEQUEST
- 4 SEQUEST と BLASTP

問 77

ある真核細胞中のすべてのタンパク質の量とすべての mRNA の量を測定することに成功した。それぞれのタンパク質とそれをコードする mRNA の量比は必ずしも一定ではなかった。その理由の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 タンパク質ごとに分解速度が異なる。
- 2 使用されるコドンの組み合わせにより翻訳の効率が異なる。
- 3 翻訳されない mRNA が存在する。
- 4 mRNA の転写速度によるタンパク質発現の調整が行われている。

問 78

酵素反応速度式であるミカエリス・メンテン式は、

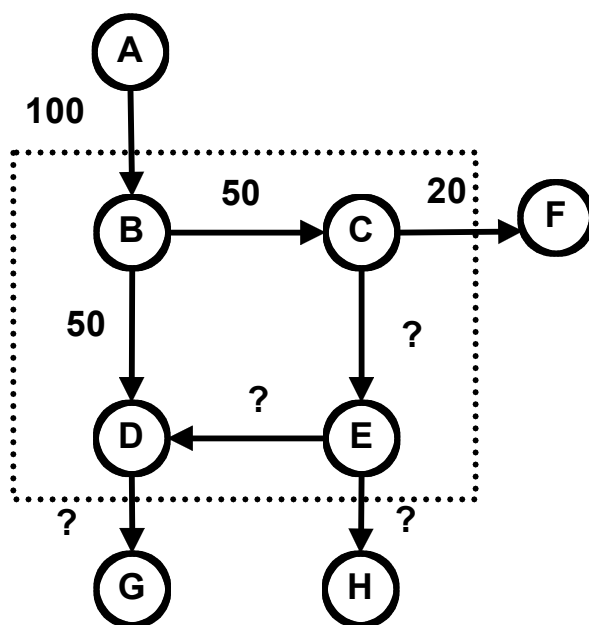
$$\frac{dP}{dt} = \frac{V \cdot S}{K + S}$$

のように示される。 t は時間、 P は生成物濃度、 S は基質濃度、 V は最大反応速度、 K はミカエリス定数である。ミカエリス・メンテン式の記述として不適切なものを選択肢の中から一つ選べ。

- 1 ミカエリス定数が大きいほど、基質と酵素の結合力は弱いと考えてよい。
- 2 基質濃度の逆数と反応速度の逆数をそれぞれ X 軸と Y 軸にプロットして得た直線の X 軸切片の逆数の絶対値をミカエリス定数とみなしてよい。
- 3 基質濃度がミカエリス定数より十分に小さいとき、反応速度は基質濃度に比例するとみなしてよい。
- 4 基質濃度がミカエリス定数の 2 倍のとき、反応速度は最大になると考えてよい。

問 79

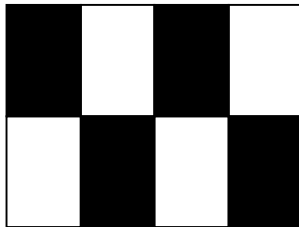
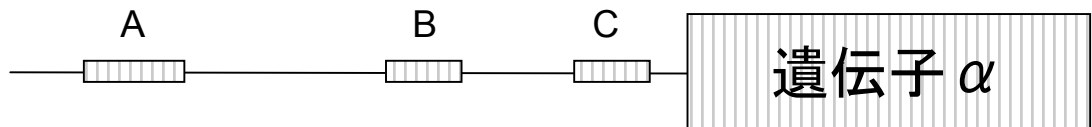
下図は、定常状態における細胞内の代謝物の流束分布を示している。点線内は細胞を表し、測定された流束値を図中に記述して、測定値のない反応には「？」をつけた。A は基質（原料）、F、G、H は生成物である。システム全体の化学反応式は、 $A \rightarrow B$ （流束 100）、 $B \rightarrow C$ （流束 50）、 $B \rightarrow D$ （流束 50）、 $C \rightarrow F$ （流束 20）、 $C \rightarrow E$ 、 $E \rightarrow D$ 、 $D \rightarrow G$ 、 $E \rightarrow H$ である。D→G の流束の下限値と上限値として、もっとも適切なものを選択肢の中から一つ選べ。



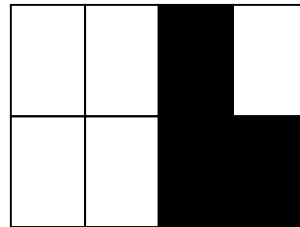
- 1 下限 50、上限 80
- 2 下限 60、上限 70
- 3 下限 0、上限 100
- 4 下限 50、上限 100

問 80

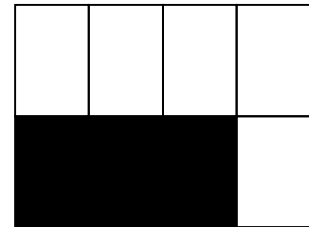
ある節足動物の胚における遺伝子 α の発現を考える。この遺伝子は 5'末端に 3 つのシス調節領域 A、B、C を持つ。A にはアクチベータータンパク質 X が、また B、C にはそれぞれリプレッサータンパク質 Y、Z が結合する。この遺伝子は、アクチベーターが結合していないと発現しない。しかし、アクチベーターが結合していても、リプレッサーが一個でも結合していると、この遺伝子は発現しない。胚は 8 つの区画に分割され、黒く塗りつぶして示してある区画で遺伝子 X、Y、Z が発現しているとする。この次の発生段階における遺伝子 α の発現のパターンとして正しいものを選択肢の中から一つ選べ。



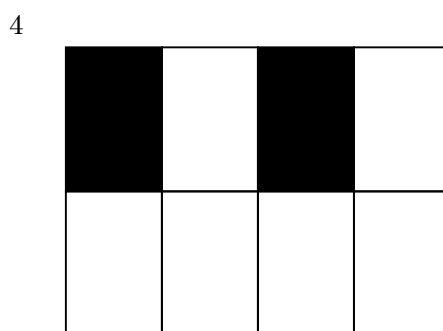
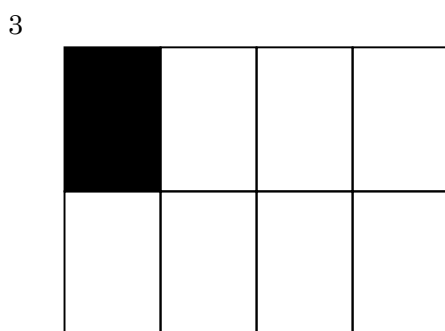
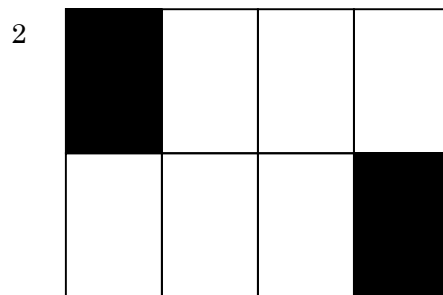
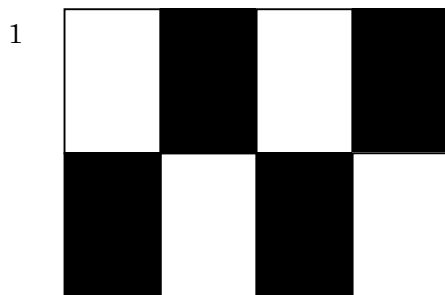
遺伝子Xの発現



遺伝子Yの発現



遺伝子Zの発現



平成 21 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】

(平成 22 年 1 月 21 日修正版)

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

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。