

平成 22 年度

日本バイオインフォマティクス学会 (JSBi)

バイオインフォマティクス技術者認定試験

試験問題

問 1

タンパク質の構造安定性に寄与する化学結合について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ホスホジエステル結合
- 2 ジスルフィド結合
- 3 イオン結合
- 4 水素結合

問 2

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 生体膜の基本をなすのはリン脂質であり、それ以外にコレステロールや糖脂質が含まれる。
- 2 リン脂質は疎水性の頭部と、親水性の炭化水素鎖からなる。このように同一分子内に両方の性質をもつことを両親媒性とよぶ。
- 3 脂質分子の運動には、側方拡散、屈曲、回転、フリップ・フロップなどがある。
- 4 生体膜中のタンパク質は、「流動モザイクモデル」が示すように膜内を二次元的に流動する。

問 3

ゴルジ体について、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 扁平な円板状の小嚢が数層重積した層板構造をもつ。
- 2 細胞外へ分泌されるタンパク質の糖鎖修飾を行う。
- 3 脂質をタンパク質に付加し、リポタンパク質に変換する。
- 4 電子伝達系によるエネルギー生産を行う。

問 4

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。
- 3 減数分裂に対し、神経細胞や上皮細胞など一般の細胞分裂を体細胞分裂と呼ぶ。
- 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 5

ヒトの血液型の遺伝に関して、AB 型の父親と AB 型の母親から産まれる子供の血液型は何か。可能性をもつ全ての血液型の組合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 型、B 型
- 2 B 型、AB 型
- 3 A 型、B 型、AB 型
- 4 O 型、A 型、B 型、AB 型

問 6

ある生物種の遺伝に関する観察結果について書かれた次の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。

観察結果

この生物は二倍体で有性生殖する。また、この生物の常染色体上には、体色に関する対立遺伝子 A と a がコードされていることがわかっている。体色が赤で AA の遺伝子型を有する雄と、体色が白で aa の遺伝子型を有する雌を交配させたところ、その第一世代の子孫の体色は全て赤であった。

- 1 A は優性遺伝子で、 a は劣性遺伝子である。
- 2 第一世代の子孫の遺伝子型は、 Aa である。
- 3 第一世代の子孫同士を交配させた時、第二世代の子孫にもっとも多く見られる遺伝子型は AA である。
- 4 第一世代の子孫同士を交配させた時、第二世代の子孫では、体色が赤の個体と白の個体の比率は、 $3 : 1$ となる。

問 7

真核細胞のタンパク質をコードする遺伝子の転写について、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 RNA ポリメラーゼが転写を開始するために最初に結合する部位をプロモーターという。
- 2 転写活性化を制御するエンハンサー配列は遺伝子上流下流どちらに位置しても良い。
- 3 転写開始点上流にあるプロモーター領域の代表的な配列は TATA ボックスである。
- 4 RNA ポリメラーゼによってゲノムから mRNA に転写されない領域を UTR という。

問 8

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

問 9

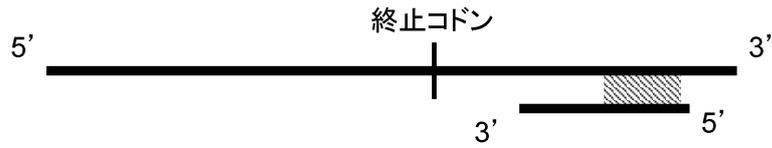
タンパク質の翻訳とそこで用いられる標準型のコドンについて、以下の記述の中でもっとも不適切なものを選択肢の中から一つ選べ。

- 1 翻訳される塩基配列の方向は 5'から 3'の一方のみである。
- 2 翻訳の開始コドンは 1 種類あるが終止コドンは 3 種類存在する。
- 3 対応するアミノ酸が決まっているコドンは全部で 20 種類である。
- 4 コドンは相補的なアンチコドンを持つ tRNA によって読みとられる。

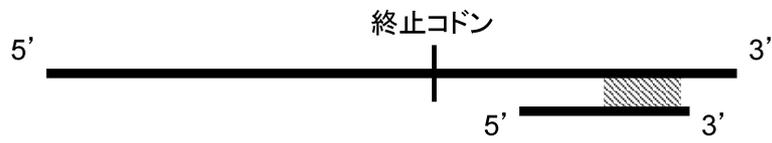
問 10

マイクロ RNA とその標的となる mRNA が相補的に結合した様子として典型的なものを、以下の選択肢から 1 つ選べ。なお、標的 mRNA を上段に、マイクロ RNA を下段に示している。また、図中の斜線部はマイクロ RNA のシード領域における塩基対を示している。

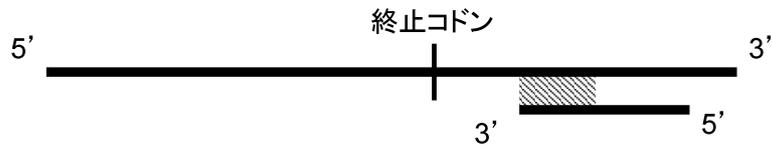
1



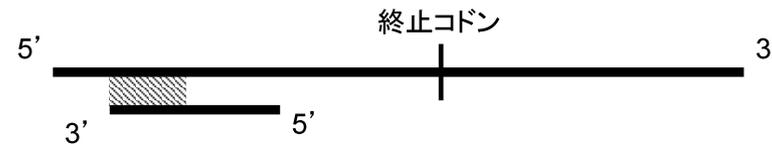
2



3



4



問 11

翻訳後修飾の 1 つであるタンパク質への糖鎖の付加（糖鎖修飾、グリコシル化）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 アスパラギン側鎖のアミドの N 原子への糖鎖付加は、*N*-結合型グリコシル化という。
- 2 タンパク質への糖鎖付加は、細胞内小器官の滑面小胞体に多数局在する糖転移酵素の働きによって行われる。
- 3 セリン側鎖、あるいはトレオニン側鎖のヒドロキシ基の O 原子への糖鎖付加は、*O*-結合型グリコシル化という。
- 4 タンパク質に付加された糖鎖は、細胞の形態維持や細胞間コミュニケーション、免疫反応の自己認識機能などに関連している。

問 12

細胞の情報伝達について、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 神経細胞と神経細胞の間で情報伝達が行われる部位をシナプスという。
- 2 G タンパク質は細胞質側に存在し GTP が結合することにより活性型となる。
- 3 シグナル分子には細胞から分泌されるものと細胞表面に提示されるものがある。
- 4 イオンチャネル型受容体に結合するシグナル分子は Na⁺や K⁺などのイオンである。

問 13

がん遺伝子やがん抑制遺伝子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 *ras* や *myc* はがん遺伝子として知られている。
- 2 *Rb* や *p53* はがん抑制遺伝子として知られている。
- 3 がん抑制遺伝子は、がんにならないようにする機能を持っている。
- 4 ゲノムのがん遺伝子に変異が起こることで、がん化が抑えられる。

問 14

ヒトのゲノムについて、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 ヒトゲノムは核ゲノムとミトコンドリアゲノムからなる。
- 2 ヒトのゲノムには SNP と呼ばれる 1 塩基多型を示す部位が存在する。
- 3 ヒトのゲノムにあるタンパク質をコードする遺伝子の数は約 10 万である。
- 4 ヒトの核ゲノムは通常 46 本の染色体 DNA からなり、そのうち 2 本は性染色体である。

問 15

下の図はさまざまな生物のもつ一倍体ゲノム DNA に含まれる塩基対の数を示している。フグとヒトはそれぞれ硬骨魚類と哺乳類の仲間であるが、以下の図ではどの位置にあたるか。適切な組合せを選択肢の中から一つ選べ。

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

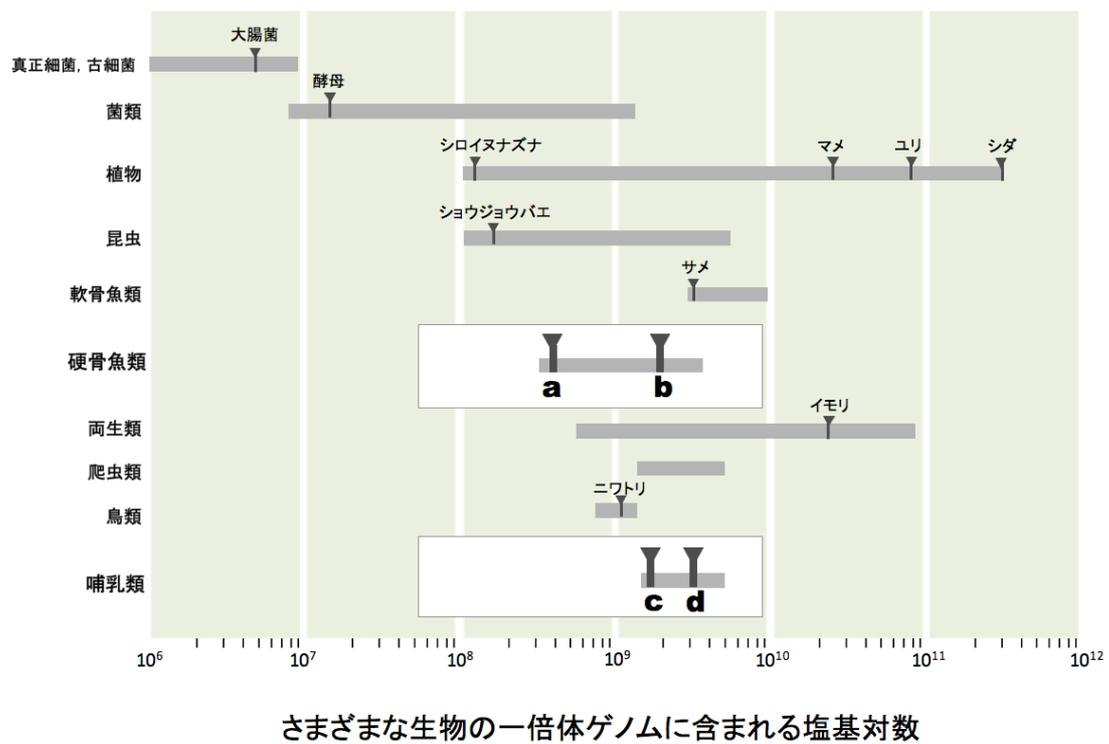


図 細胞の分子生物学第 5 版（邦訳ニュートンプレス刊）の図を改変

問 18

タンパク質の分子量を測定する目的で、SDS ポリアクリルアミドゲル電気泳動法 (SDS-PAGE) がよく用いられる。これについて以下の中から正しい組合せを選択肢の中から一つ選べ。

SDS-PAGE では、試料を (a) 側におく。泳動が進むと、分子量が (b) タンパク質ほど速く泳動する。

- 1 (a) 陽極 (b) 大きい
- 2 (a) 陽極 (b) 小さい
- 3 (a) 陰極 (b) 大きい
- 4 (a) 陰極 (b) 小さい

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

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

問 21

コンピュータ上では、数値は二進数で表される。二進数同士の論理積(AND)とは、2 つの二進数の各桁に対してそれぞれ論理積をとったものであり、たとえば $110_{(2)} \text{ AND } 101_{(2)} = 100_{(2)}$ である。なお、ここで数値の最後に付加した(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

$$f(x) = x \text{ AND } 101_{(2)}$$

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 $y = 110_{(2)}$ $z = 100_{(2)}$
- 2 $y = 101_{(2)}$ $z = 100_{(2)}$
- 3 $y = 011_{(2)}$ $z = 101_{(2)}$
- 4 $y = 000_{(2)}$ $z = 101_{(2)}$

問 22

コンピュータ言語に関する用語の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 インタープリタは、ソースコードあるいは中間言語を解釈しながら計算機上でソフトウェアの実行を行う。例として Perl などがあげられる。
- 2 コンパイラは、ソースコードを計算機が解釈可能な機械語に直接変換を行う。この変換作業をコンパイルという。例として Ruby や MySQL などがあげられる。
- 3 ジャスト・イン・タイム・コンパイラ (JIT コンパイラ) は、ソフトウェア実行時にコンパイルすることによってインタープリタと比べ実行速度の向上を図ったものである。
- 4 アセンブリ言語は、機械語を人間にわかりやすい形で記述するための言語である。

問 23

コンピュータアーキテクチャに関する用語の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 CISC (Complex Instruction Set Computer)に対して、RISC(Reduced Instruction Set Computer)は、少数の単純な命令セットしか持たないことでハードウェアを単純化したアーキテクチャである。
- 2 どのようなアーキテクチャにおいても、機械語の1命令は、1クロック以内で動作するように設計されている。
- 3 メモリインターリーブは、メモリを複数個の単位に分割し、それらへ並列にアクセスすることでメモリアクセスを高速化したものである。
- 4 キャッシュとは、CPU と主記憶の間に配置する小容量の高速な記憶装置のことである。

問 24

以下の(a)～(d)に入る語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。

(a)は、計算機間で遠隔ログインを行うための通信プロトコルの一種であるが、認証を含めすべての情報が暗号化されない平文のまま送受信されるというセキュリティ上の問題がある。そのため、近年では、その通信・認証が公開鍵暗号によって暗号化されている(b)が用いられるようになった。また、ファイルを交換するプロトコルである(c)も同様のセキュリティ上の問題があり、近年では通信・認証が暗号化される(d)などがよく用いられている。

- | | | | | |
|---|------------|----------|------------|----------|
| 1 | (a) scp | (b) ssh | (c) telnet | (d) ftp |
| 2 | (a) ftp | (b) sftp | (c) telnet | (d) ssh |
| 3 | (a) telnet | (b) ftp | (c) scp | (d) ssh |
| 4 | (a) telnet | (b) ssh | (c) ftp | (d) sftp |

問 25

ある問題に対するあるアルゴリズムの計算量が $O(n^2)$ であるとはどのようなことを表しているか。以下の記述のうち、もっとも適切なものを選択肢の中から一つ選べ。

- 1 そのアルゴリズムを用いて最大 n^2 回の計算機上の演算でその問題を解く（計算する）ことができることを表す。
- 2 そのアルゴリズムを用いてその問題を解く（計算する）のに最小でも n^2 回の計算機上の演算が必要であることを表す。
- 3 $n > c$ であるどのような n に対しても、そのアルゴリズムを用いて最大 $a \cdot n^2$ 回の計算機上の演算でその問題を解く（計算する）ことができる、というような定数 a および c が存在することを表す。
- 4 $n > c$ であるどのような n に対しても、そのアルゴリズムを用いてその問題を解く（計算する）のに最小でも $a \cdot n^2$ 回の計算機上の演算が必要である、というような定数 a および c が存在することを表す。

問 26

以下の(a)~(c)に入る語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。

ソートアルゴリズムには様々なアルゴリズムが存在するが、(a)は、最悪計算時間は $O(n^2)$ であるが平均計算時間は $O(n \log n)$ である。一方、(b)のように最悪計算時間も $O(n \log n)$ であるソートアルゴリズムも存在する。また、(c)のように、ある定数以下の正の整数のみからなる数列を線形時間でソートすることができる、といったアルゴリズムも存在する。

- | | | |
|---------------|-------------|-------------|
| 1 (a) クイックソート | (b) マージソート | (c) バケットソート |
| 2 (a) クイックソート | (b) バケットソート | (c) マージソート |
| 3 (a) マージソート | (b) クイックソート | (c) バケットソート |
| 4 (a) マージソート | (b) バケットソート | (c) クイックソート |

問 27

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、最後に取り出された文字が同じになるような操作列を選択肢の中から一つ選べ。

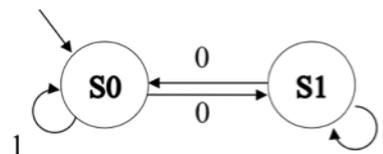
ただし、スタックに対する場合、英字はその文字のプッシュ操作、*はポップ操作を表す。キューに対する場合は英字はその文字をキューに挿入する操作、*は取り出し操作を表す。たとえば、スタックに対して「AB**」を行うと最後のポップ操作では「A」が出力されるが、キューに対して同じ操作列「AB**」を行うと最後の取り出し操作では「B」が出力されるため、この操作列「AB**」においては最後に取り出される文字は異なる。

- 1 AA* B**
- 2 AB* A**
- 3 AB* B**
- 4 AB* C**

問 28

下図の決定性有限オートマトンは 2 つの状態 S0, S1 を持つ。S0 は初期状態である。入力数列は 0 または 1 からなる。このオートマトンに関する記述で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
入力	0	S1	S0
	1	S0	S1



- 1 入力数列が 11000 のとき、状態は S1 となる。
- 2 入力数列が 1010110 のとき、状態は S1 となる。
- 3 入力数列中に 0 が偶数個あるときは、状態は S0 となる。
- 4 入力数列中に 1 が偶数個あるときは、状態は S1 となる。

問 29

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。データ全体の中から検索のキーとなりうる語を列挙して、あらかじめ整理しておくことにより、その語が検索される時には高速に語の存在する位置へのアクセスが可能となる。索引等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 索引上の語をソートしておくことにより、二分探索法などを用いて高速な検索が可能となる。
- 2 データベースでは、索引の検索に二分探索木ではなく、B-木と呼ばれる多分木を用いることが多い。
- 3 B-木はバランス木であり、根から葉までの長さがほぼ揃っている。
- 4 B-木では、各節点での分岐数は一般に 16 程度が限度であり、これより多い分岐数になるとハードディスクへのアクセスの回数が増えてしまうので効率が低下する。

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。ハッシュ表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 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呼ぶ。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ある。
- 3 全てのハッシュ関数では、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 に対して必ず $h(a)+h(b) = h(a+b)$ 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より小さくない場合には、衝突の起こらない完全ハッシュ関数を求めることが可能である。

問 31

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の件数として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 塩基配列一覧  
WHERE サブタイプ = 'H1N1' AND 年号 > 1976 ;
```

表： 塩基配列一覧

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 3 件
- 2 4 件
- 3 5 件
- 4 6 件

問 32

コンピュータ上でのデータの記述形式には、様々なものが開発されているが、その中でも XML (eXtensible Markup Language) はもっとも普及した形式の一つである。次に示した記述のうち、不適切なものを選択肢の中から一つ選べ。

- 1 XML では、各要素をタグで囲み、階層的なデータ構造を入れ子で表現する。
- 2 XML では、ユーザが新たなタグを定義して、言語を拡張していくことができる。
- 3 XML データベースでは、画像や音声などのデータは扱えない。
- 4 XML を効率的に扱える XML データベースが商品化されている。リレーショナルデータベースの中にも XML を格納可能なものが開発されている。

問 33

以下に示した 6 個の数値のデータが与えられたとする。このデータにおける 3 個の基本統計量（平均値、中央値、最頻値）の値として正しい組み合わせを選択肢の中から一つ選べ。ここで中央値はメジアン、最頻値はモードとも呼ばれる。

3, 3, 3, 8, 9, 10

選択肢	平均値	中央値	最頻値
1	5.5	6	3
2	5.5	3	6
3	6	5.5	3
4	6	3	5.5

問 34

2つの確率変数 X, Y を考える。不適切な記述を選択肢の中から一つ選べ。

ここで、それぞれの記号は次の意味で用いられる。

$\Pr(A, B)$ A と B の同時確率

$E(A)$ A の平均

$\text{Var}(A)$ A の分散

- 1 X, Y が統計的独立ではないとき、積 XY の平均 $E(XY)$ は、それぞれの平均の積 $E(X)E(Y)$ に等しい。
- 2 X, Y が統計的独立のとき、積 X, Y の同時確率 $\Pr(X, Y)$ は、それぞれの確率の積 $\Pr(X)\Pr(Y)$ に等しい。
- 3 X の分散 $\text{Var}(X)$ は、 X の平均の二乗と X^2 の平均との間に次の関係がある。
$$\text{Var}(X) = E(X^2) - (E(X))^2$$
- 4 Y が生じたという条件のもとで X が生じる条件付き確率 $\Pr(X|Y)$ は、 $\Pr(X, Y)/\Pr(Y)$ であらわされる。

問 35

X と Y は独立に標準正規分布 $N(0,1)$ に従うとする。標準正規分布 $N(0,1)$ とは、平均 0、標準偏差 1 の正規分布のことを指す。この 2 変数 X, Y に関して、不適切なものを選択肢の中から一つ選べ。

- 1 X と Y は同じ値を生起するとは限らない。
- 2 $X + Y$ はカイ二乗分布に従う。
- 3 $X + Y$ の分散は 2 である。
- 4 $2X$ の分散は 4 である。

問 36

統計解析の説明として、不適切なものを選択肢の中から一つ選べ。

- 1 尤度関数を最大にする推定値を最尤推定値という。
- 2 母数を推定する際、1点で推定することを点推定と言い、母数の信頼区間を求めることを区間推定という。
- 3 正規分布の確率密度関数 $p(x)$ において、 $p(x) < 0$ となる実現値 x が存在する。
- 4 分散が小さい正規分布では、平均値付近で確率密度が1以上になる場合がある。

問 37

次のクラスタ解析の手法の中で階層的クラスタリングによるものを選択肢の中から一つ選べ。

- 1 K-平均法 (K-means method)
- 2 UPGMA (Unweighted Pair Group Method with Arithmetic mean)
- 3 自己組織化マップ (Self-Organizing Map)
- 4 Fuzzy C-means

問 38

10 個のデータ X_1, X_2, \dots, X_{10} に対して、各データには 2 クラス (+1 : 陽性、-1 : 陰性) 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 -最近傍法を用いることとし、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 $d(X, X_i)$ が次の表のようであるとき、 k -最近傍法の予測結果として正しい選択肢を選べ。ただし、 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識別法である。

標本番号 i	1	2	3	4	5	6	7	8	9	10
クラスラベル y_i	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1
距離 $d(X, X_i)$	4.2	5.7	2.7	6.3	7.9	5.2	0.2	3.1	6.1	8.7

- 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 39

予測手法の性能評価の際に行われるクロスバリデーション法（交差検証法）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 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。
- 2 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。これを特に二分探索法と呼ぶ。
- 3 LOO (leave-one-out)法は、データのうち 1 個のみをテスト用に残しておく方法である。
- 4 クロスバリデーション法は、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ために用いることができる。

問 40

陽性か陰性かの二値分類の予測を行うとき、その予測性能の指標には様々なものがある。

下記の記述のうちで、もっとも不適切なものを選択肢の中から一つ選べ。

ただし以下では、TP = True Positive (真陽性)、TN = True Negative (真陰性)、FP = False Positive (偽陽性)、FN = False Negative (偽陰性)を表すものとする。

- 1 再現率(recall)は、 $TP / (TP + FN)$ で表される。
- 2 適合率(precision)は、 $TP / (TP + FP)$ で表される。
- 3 F 値(F-measure)は再現率と適合率の幾何平均であり、 $2TP / (TP+FN+TP+FP)$ で表される。
- 4 一般的に再現率と適合率はトレードオフの関係があり、一方を高めようとすると、もう一方は低下する傾向がある。

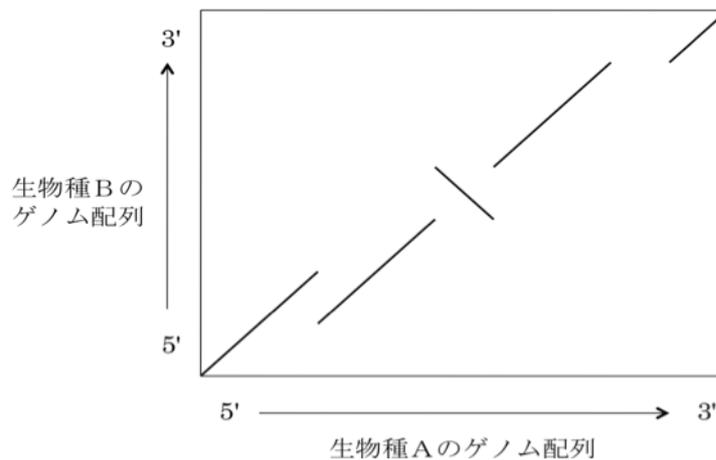
問 41

BLAST によるホモロジー検索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 DNA 配列をクエリとしてアミノ酸配列データベースに対して検索することはできるがアミノ酸配列をクエリとして DNA 配列データベースに対して検索することはできない。
- 2 アミノ酸配列での比較と DNA 配列での比較では、一般にアミノ酸配列の方がより遠縁の相同配列を検出しやすい。
- 3 低複雑性(low-complexity)領域をマスクすることで、マスクしていない場合に比べ、生物学的に意味のある相同配列を得ることができるが、配列一致度(% identity)は別途計算する必要がある。
- 4 PSI-BLAST は、位置特異的スコア行列(position-specific scoring matrix; PSSM)を反復的に作りながら、より遠縁の相同配列を検出するプログラムである。

問 42

下に示す図は、共通祖先から分岐した生物種 A, B のゲノム配列間の相同領域を表している。横軸、縦軸はそれぞれ生物種 A、生物種 B のゲノム配列に相当し、相同性のある領域を斜線で示してある。図の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。



- 1 生物種 B との分岐後、生物種 A でゲノム領域の反転が起きた。
- 2 分岐後、生物種 A または生物種 B でゲノム領域の反転が起きたが、この図からだけではどちらの種で反転が起きたのかは特定できない。
- 3 分岐後、生物種 B にゲノム領域の重複が起きた。
- 4 分岐後、生物種 A にゲノム領域の欠失が起きた。

問 43

ClustalW などで行われている、多重アラインメントの典型的な手法のひとつである累進法 (progressive method) の説明として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 あらかじめ作成した近似的な系統樹に従って配列を順に加えていって、全体のアラインメントを完成させていく方法である。
- 2 ギャップの入れ方を少しずつ変更しながら、より高いスコアを持つアラインメントになるように改善していく方法である。
- 3 まず局所的に保存されたブロックを抽出し、それらを少しずつ組み合わせながら全体のアラインメントを完成させていく方法である。
- 4 データベース検索を繰り返して、より弱い類似性を持つ配列を加えることによって、アラインメントを徐々に改善していく方法である。

問 44

BLAST では、単位を標準化したビットスコア(bit score)が類似性の評価に用いられている。これについて述べた文のうち、もっとも適切なものを次の中から一つ選べ。

- 1 スコア行列を BLOSUM45 から BLOSUM80 に変更しても、ビットスコアの値は変わらない。
- 2 スコア行列とギャップペナルティの各値を一律 5 倍にしたスコア体系を用いても、ビットスコアの値は変わらない。
- 3 検索対象とするデータベースの大きさを 5 倍にすると、ビットスコアの値は 5 分の 1 になる。
- 4 ビットスコアの値を s とすると、 10^s の値が、探索空間の大きさ（データベースの配列全長×クエリの配列長）より十分小さいことが有意性の目安となる。

問 45

BLAST などの相同性検索プログラムでは、局所的に最大のスコアを持つアラインメントを考える。以下に示す、ギャップなしでアラインメントされた 2 本の配列において、一致を+3 点、不一致を -2 点として、その部分領域における局所的なスコアを考える（アラインメントはこのままで固定し、ギャップの挿入などにより変形することは考えない）。たとえば、2 塩基目から 6 塩基目までの領域をとると、一致が 2 つ、不一致が 3 つあるので、局所的なスコアは $3 \times 2 - 2 \times 3 = 0$ 点である。ここで、あらゆる領域の中で局所的なスコアが最大となるような領域をとったとき、そのスコアの値はいくらになるか。次の中から選べ。

```
Seq1  GAGCAGCTAGATCGGATTGTTAGA
      | | | | | | | | | |
Seq2  CATGTGATAGACCGTATTAATGCA
```

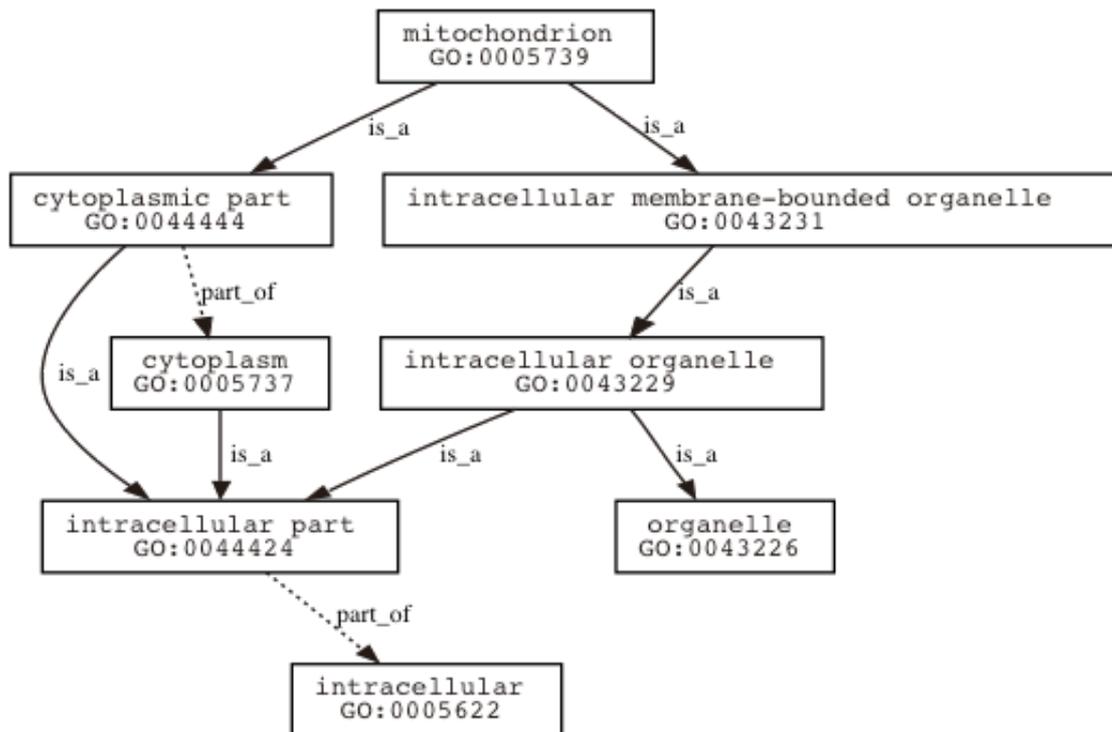
- 1 21
- 2 22
- 3 23
- 4 24

問 46

Gene Ontology では、遺伝子の機能に関わる様々な用語(GO term)の間に、A **is_a** B (AはBの一種である) や A **part_of** B (AはBの部分である；より正確には、すべてのAは、あるBの部分である) といった関係が定義されている。以下に示す図は、一部の GO term 間の関係を表した図で、A **is_a** B および A **part_of** B の関係が、それぞれ A から B への実線および破線の矢印で表されている (例えば、mitochondrion **is_a** cytoplasmic part が成立している)。この図から読み取れることとして、適切でないものを次の選択肢から選べ。ただし、以下の推論規則を用いて良い。

推論規則

- A **is_a** B かつ B **is_a** C ならば、A **is_a** C である。
- A **part_of** B かつ B **part_of** C ならば、A **part_of** C である。
- A **is_a** B かつ B **part_of** C ならば、A **part_of** C である。
- A **part_of** B かつ B **is_a** C ならば、A **part_of** C である。

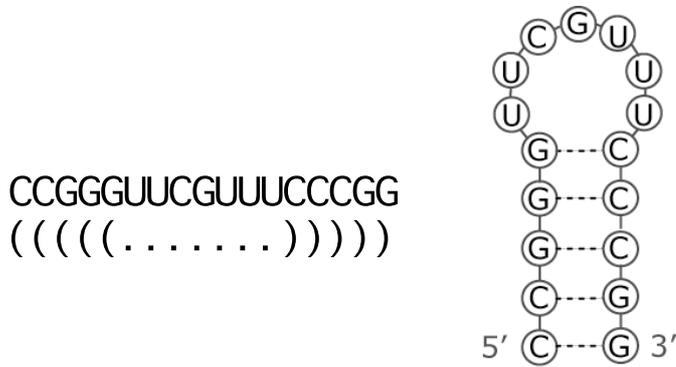


- 1 cytoplasmic part **is_a** intracellular part
- 2 cytoplasmic part **part_of** intracellular part
- 3 mitochondrion **is_a** organelle
- 4 intracellular organelle **is_a** intracellular

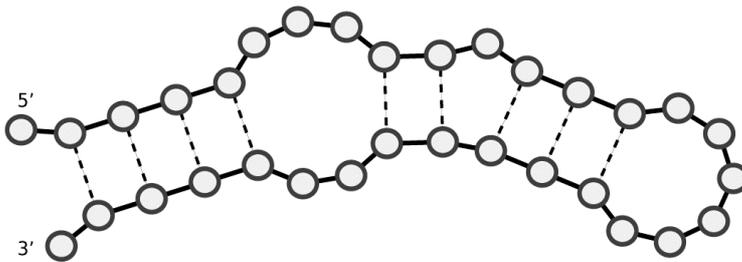
問 47

RNA の二次構造表現方法のひとつに、塩基対を括弧の対応で表現する方法がある。

例 :



下図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 48

以下の A~E はアラインメントの評価についての記述である。正しい記述の組み合わせを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、挿入や欠失などの発生頻度を計数化したものである。
- C PAM や BLOSUM は核酸配列のための置換行列である。
- D 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も重く設定される。

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

問 49

4 残基長のある DNA 配列モチーフを、次のような重み行列で表現した。

	position 1	position 2	position 3	position 4
A	10	-21	-11	-10
T	1	-22	-15	23
G	-20	13	12	-21
C	-20	-22	3	-15

この重み行列を用いて、7 塩基の長さの配列、GAGGTCT を検索した時に、もっとも高いスコアを示す部分配列はどれか。選択肢の中から一つ選べ。

- 1 AGGT
- 2 GAGG
- 3 GTCT
- 4 GGTC

問 50

5 残基長のある DNA 配列モチーフを考える。いま下記の 4 つの配列が実際に観測されたとき、このモチーフに対する相対エントロピースコアを求めよ。選択肢の中から、適切な値を一つ選べ。ただし相対エントロピースコアは次式で与えられるものとする。

ここで $f_j(a)$ は、配列モチーフの j 文字目における塩基 a の出現頻度を表す。 $p(a)$ は塩基 a が一般に観測される背景確率で、 $p(A) = p(T) = p(G) = p(C) = 0.25$ とする。

$$\text{相対エントロピースコア} = \sum_{j=1}^5 \sum_{a \in \{A, T, G, C\}} f_j(a) \cdot \log_2 \frac{f_j(a)}{p(a)}$$

モチーフ配列

	1	2	3	4	5
配列 1 :	T	A	T	A	T
配列 2 :	T	A	T	T	T
配列 3 :	T	A	A	G	T
配列 4 :	T	A	A	C	T

- 1 3.5
- 2 5.0
- 3 6.5
- 4 7.0

問 51

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、ハイフン(-)は文字の連結を表す。角括弧([])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 CTICGGENNRTCFCV
- 3 CLIHGGENDRTCWVC
- 4 CTMCGGEHNNRTCFCV

問 52

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 5'-ATGGGCCGCCCAT-3'
3'-TACCCGGCCGGTA-5'
- 2 5'-ATGGGCCTTTAAAG-3'
3'-TACCCGAAATTTTC-5'
- 3 5'-ATGGGCCCCGGTA-3'
3'-TACCCGGGCCCAT-5'
- 4 5'-ATGGGCCATGGGCC-3'
3'-TACCCGGTACCCGG-5'

問 53

タンパク質はアミノ酸がペプチド結合を介して重合したものであるが、そのペプチド結合は自由回転できないため、シス(*cis*)とトランス(*trans*)の 2 種類の異性体 (シス-トランス異性体) をとりうる。この異性体についてもっとも適切な記述を選択肢の中から一つ選べ。

- 1 トランスが圧倒的に多いが、プロリンのアミノ末端側ではシスも比較的多く見られる。
- 2 トランスが圧倒的に多いが、アラニンのアミノ末端側ではシスも比較的多く見られる。
- 3 シスが圧倒的に多いが、プロリンのアミノ末端側ではトランスも比較的多く見られる。
- 4 シスが圧倒的に多いが、アラニンのアミノ末端側ではトランスも比較的多く見られる。

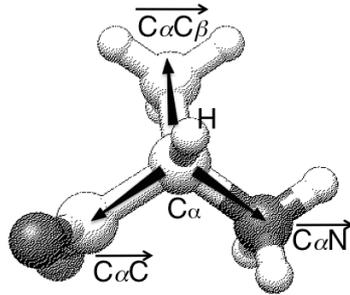
問 54

天然タンパク質を構成する 20 種類のアミノ酸は、中心となる C α 原子に、アミノ基・カルボキシル基・水素原子・側鎖の 4 つの異なる化学団が結合しているため、互いに鏡像の関係にある光学異性体が存在するが、あるアミノ酸だけは例外である。そのアミノ酸を選択肢の中から一つ選べ。

- 1 メチオニン
- 2 グリシン
- 3 プロリン
- 4 システイン

問 55

天然タンパク質を構成するアミノ酸は、2つの可能な光学異性体（L-、D-アミノ酸）のうち、下の図に示す L-アミノ酸が主に使用されている。あるアミノ酸の $C\alpha$ 原子から N、C、 $C\beta$ 原子へのベクトルを、それぞれ、 $\overrightarrow{C\alpha N}$ 、 $\overrightarrow{C\alpha C}$ 、 $\overrightarrow{C\alpha C\beta}$ としたとき、そのアミノ酸が L-アミノ酸であることを判別する式として正しいものを選択肢の中から一つ選べ。ただし、 \cdot と \times はそれぞれベクトルの内積、外積を表す。



L-アミノ酸

- 1 $(\overrightarrow{C\alpha N} \times \overrightarrow{C\alpha C}) \cdot \overrightarrow{C\alpha C\beta} > 0$
- 2 $(\overrightarrow{C\alpha N} \times \overrightarrow{C\alpha C}) \cdot \overrightarrow{C\alpha C\beta} = 0$
- 3 $(\overrightarrow{C\alpha N} \times \overrightarrow{C\alpha C}) \cdot \overrightarrow{C\alpha C\beta} < 0$
- 4 $(\overrightarrow{C\alpha N} \times \overrightarrow{C\alpha C}) \cdot \overrightarrow{C\alpha C\beta} = 1$

問 56

生体高分子の立体構造解析に用いられる計算手法のうち、分子動力学法と分子力学法に関する以下の記述において、不適切なものを選択肢の中から一つ選べ。

- 1 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。
- 2 分子力学法は、英語では **Molecular Mechanics** であり、**MM** という略称も使われる。
- 3 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記述される。
- 4 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記述される。

問 57

あるタンパク質の PDB(Protein Data Bank)エントリー内容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 (次ページ) 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)
KEYWDS   PRION PROTEIN, PRION, BRAIN, GLYCOPROTEIN, GPI-ANCHOR,
KEYWDS   2 REPEAT, POLYMORPHISM, DISEASE MUTATION, MAJOR PRION
KEYWDS   3 PROTEIN
EXPDTA   SOLUTION NMR
(中略)
REMARK   999
REMARK   999 SEQUENCE
REMARK   999 2 RESIDUES (GLY SER) INSERTED AT THE N-TERMINUS
REMARK   999 MET166 AND GLU221 REPLACED AGAINST CYS
DBREF    1H0L A 119 120 PDB 1H0L 1H0L 119 120
(中略)
SEQRES   9 A 112 GLN ALA TYR TYR GLN ARG GLY SER
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
(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
ATOM     5 H GLY A 119 12.835 -29.359 1.441 1.00 0.00 H
ATOM     6 HA2 GLY A 119 11.022 -27.952 3.323 1.00 0.00 H
ATOM     7 HA3 GLY A 119 10.757 -29.629 2.772 1.00 0.00 H
(後略)

```

- 1 ヒトのタンパク質であるが、天然配列に対して2残基の挿入と2残基のアミノ酸置換が施されている。
- 2 3本の α ヘリックスと、2本の β ストランドからなる1枚の β シートが存在する。
- 3 X線結晶解析法で決定された構造である。
- 4 各原子の温度因子が0.00と記載されているが、構造解析手法から考えて、この値は信頼すべきではない。

問 58

主に α ヘリックスや β シートなどの二次構造から構成される、タンパク質主鎖の折りたたみパターンをフォールドと呼ぶ。天然タンパク質でよく見られる以下のフォールドのうち、 β ストランドのみで構成されるものを選択肢の中から一つ選べ。

- 1 TIM バレルフォールド
- 2 免疫グロブリンフォールド
- 3 グロビンフォールド
- 4 ロスマン(Rossmann)フォールド

問 59

次の立体構造分類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 スーパーフォールドとは、進化的類縁関係が考えられないようなタンパク質間であっても共通して現れる類似フォールドの集合である。
- 2 SCOP、CATH などの構造分類データベースは非階層的に構築されている。
- 3 SCOP の構造分類における立体構造のクラスの代表的なものとして、all α 、all β 、 α/β 、 $\alpha+\beta$ がある。
- 4 タンパク質の立体構造は、通常ドメインを単位として分類される。

問 60

以下に示したホモロジーモデリングによるタンパク質の立体構造モデルの構築についての記述から、もっとも不適切なものを一つ選べ。

- 1 信頼性の高いモデルを得るには、立体構造既知のアミノ酸配列と立体構造未知のアミノ酸配列の正確なアラインメントが要求される。
- 2 一般に、立体構造既知のアミノ酸配列と立体構造未知のアミノ酸配列の一致度が低かったり、一方に大きな挿入または欠失が含まれていると、高精度のモデル構造は構築できない。
- 3 基本的には、ホモログの既知立体構造を鋳型として、モデル構造が構築される。
- 4 一般に、保存度の高い活性中心よりも保存度の低いループ領域の方が、モデル中で正確に構築されている。

問 61

立体構造を推定したいタンパク質のアミノ酸配列があるが、PDB(Protein Data Bank)に登録された分子に対して BLAST による配列検索をしたところ、有意な類似性を示すタンパク質は見つからなかった。この後に試みる予測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 フォールド認識法
- 2 *ab initio* 構造予測法
- 3 二次構造予測法
- 4 ホモロジーモデリング法

問 62

それぞれ 5 アミノ酸残基（残基 A1~A5 と残基 B1~B5）からなる 2 つのペプチドの立体構造を重ね合わせたとき、対応する残基の C α 原子間距離が以下のとおりであった。この構造重ね合わせの C α 原子間 RMSD(Root Mean Square Deviation)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。

A1 – B1 間 3.0 Å

A2 – B2 間 1.0 Å

A3 – B3 間 0.0 Å

A4 – B4 間 1.0 Å

A5 – B5 間 3.0 Å

1 1.6 Å

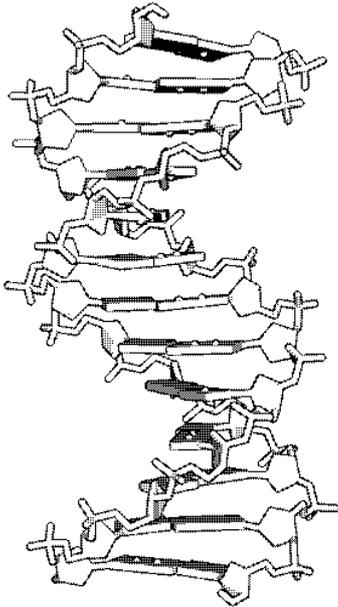
2 2.0 Å

3 3.0 Å

4 8.0 Å

問 63

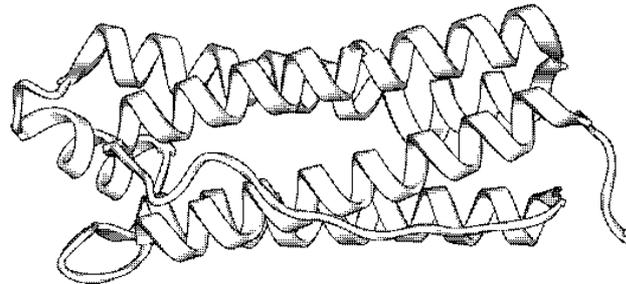
以下の図は DNA の立体構造を模式的に示したものである。この図の説明としてもっとも適切なものを選択肢の中から一つ選べ。



- 1 B型構造である。
- 2 上から主溝(major groove)-副溝(minor groove)-主溝を正面に向けている。
- 3 全部で26塩基対の構造が示されている。
- 4 左巻き二重らせん構造である。

問 64

以下の図はあるタンパク質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適切なものを選択肢の中から一つ選べ。



- 1 SCOP の立体構造分類で $\alpha + \beta$ クラスに分類される構造である。
- 2 マルチドメインタンパク質である。
- 3 空間充填模型(space filling model)表示された分子グラフィックスである。
- 4 4-ヘリックスバンドルに分類される構造である。

問 65

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。次に示した用語のうちヒトゲノムにおける遺伝子多型の呼び名ではないものを一つ選べ。

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

問 66

ABO 式血液型を決める遺伝子多型は 3 つのアレル A、B、O の持ち方によって決まり、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。

今、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 0.1 B 0.5 O 0.4
- 3 A 0.1 B 0.4 O 0.5
- 4 A 0.3 B 0.3 O 0.4

問 67

以下の遺伝因子解析に関連する記述のうち、説明としてもっとも不適切なものを一つ選べ。

- 1 複数の遺伝要因と複数の環境要因の影響による疾患を複合遺伝性疾患とよぶ。
- 2 同一染色体上にある多型の連鎖不平衡は、距離が遠いものほど強くなる傾向がある。
- 3 複合遺伝性疾患における遺伝因子の寄与の指標である遺伝率が大きい場合、遺伝因子を同定できる可能性が高い。
- 4 集団がいくつかの分集団より構成されている場合には、異なる染色体上にある多型のアレルであっても、連鎖不平衡のように見える関連が認められることがある。

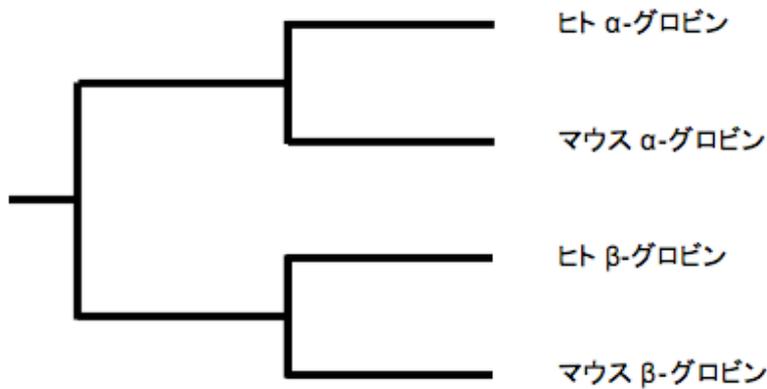
問 68

遺伝子の進化的性質の一つに分子時計がある。分子時計について述べた次の文章のうちもっとも不適切なものを一つ選べ。

- 1 分子時計とは、アミノ酸置換や塩基置換がほぼ一定の速度で起こる現象を指す。
- 2 分子時計では、放射性同位体が一定の割合で崩壊することを利用して年代決定を行う。
- 3 分子時計によって分岐年代を推定するには、化石などの分子以外の証拠から分岐時間がわかっている生物種が一組は必要である。
- 4 遺伝子やタンパク質によっては分子時計が成り立たないこともあるので、分子時計を利用して生物の分岐時間を推定する場合には使う遺伝子やタンパク質の進化速度が近似的に一定になっていることを確認しておくことが必要である。

問 69

次に示す系統樹は、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。この系統樹について述べた次の選択肢のうち、もっとも不適切なものを一つ選べ。



- 1 このような遺伝子重複によってできた2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 ヒト α -グロビンとマウス 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（オーソログスな関係にある遺伝子）と呼ばれる。
- 3 ヒト α -グロビンとマウス 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（パラログスな関係にある遺伝子）と呼ばれる。
- 4 α -グロビンと β -グロビンの遺伝子重複はヒトとマウスの種分岐よりも後に起こっている。

問 70

次の文章はいくつかの系統樹推定法についてその特徴を述べたものである。文章中の(a)、(b)内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。

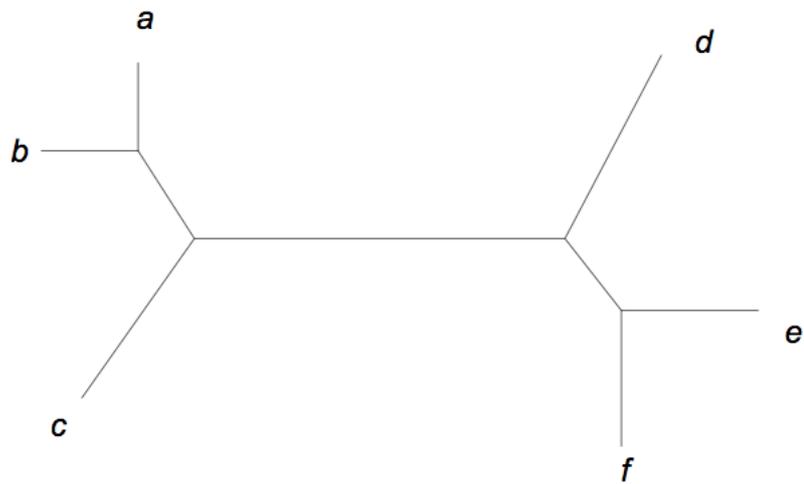
系統樹推定法には大きく分けて、平均距離法や (a) のように距離行列を用いる方法と、最尤法 (ML 法) のように配列データを直接用いる方法がある。

最尤法では、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) が増えるに従って考慮すべき樹形の数が爆発的に増えていくことに注意する必要がある。無根系統樹において可能な樹形の数は、OTU が 4 つのときは 3 通り、OTU が 5 つのときは (b) 通り、OTU が 6 つのときは 105 通りというように急速に増加する。

- 1 (a) 近隣結合法 (NJ 法) (b) 15
- 2 (a) 近隣結合法 (NJ 法) (b) 21
- 3 (a) 最大節約法 (MP 法) (b) 15
- 4 (a) 最大節約法 (MP 法) (b) 21

問 71

系統樹の表記法としてニューウィック形式 (Newick format) とよばれる方法がある。ニューウィック形式では、一つの系統樹が複数の方法で表現できる。次に示す無根系統樹の記述法として不適切なものを選択肢から一つ選べ。



- 1 $((a,b), c, (d, (e, f)))$
- 2 $((((a,b),c), d), (e,f))$
- 3 $((a, b), (c, d), (e, f))$
- 4 $(c, ((e, f), d), (b, a))$

問 72

次の文章の空白部分(a)、(b)、(c)に入る式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。

今、二本の塩基配列のアラインメントが与えられ、総サイト数は n 、差異のあるサイト数は x であったとする。また、挿入/欠失はないものとする。この二本の配列の進化的距離 d を推定したい。そこで、二本の相同な塩基配列の進化的距離を d とした時、任意のサイトで塩基が異なる確率は $p = (3/4) - (3/4)e^{-(4d/3)}$ であるというモデルを考え、最尤推定を行なう。尤度 $L(d; x)$ は、二項確率として次のように与えられる。

$$L(d; x) = \frac{n!}{x!(n-x)!} \quad \boxed{\text{(a)}}$$

ここで、 $\frac{n!}{x!(n-x)!}$ は d に依存しないので除いて、 $\boxed{\text{(a)}}$ の部分の対数をとったものを

対数尤度 $\ell(d; x)$ とする。ここで、 $\boxed{\text{(b)}}$ という式を d について解くと、 d の最尤推定値

$\boxed{\text{(c)}}$ を得る。

- | | | | |
|---|--------------------------------|--------------------------------------|--|
| 1 | (a) $p^x(1-p)^{(n-x)}$ | (b) $\frac{d\ell(d; x)}{dd} = 0$ | (c) $-\frac{3}{4} \log\left(1 - \frac{4}{3} \times \frac{x}{n}\right)$ |
| 2 | (a) $p^x(1-p)^{(n-x)}$ | (b) $\frac{d^2\ell(d; x)}{dd^2} = 0$ | (c) $-\frac{4}{3} \log\left(1 - \frac{3}{4} \times \frac{x}{n}\right)$ |
| 3 | (a) $(3/4) - (3/4)e^{-(4d/3)}$ | (b) $\frac{d\ell(d; x)}{dd} = 0$ | (c) $-\frac{3}{4} \log\left(1 - \frac{4}{3} \times \frac{x}{n}\right)$ |
| 4 | (a) $(3/4) - (3/4)e^{-(4d/3)}$ | (b) $\frac{d^2\ell(d; x)}{dd^2} = 0$ | (c) $-\frac{4}{3} \log\left(1 - \frac{3}{4} \times \frac{x}{n}\right)$ |

問 73

マイクロアレイには 1 色法と 2 色法があるが、それらの特徴について述べた以下の記述のうち、もっとも不適切なものを一つ選べ。

- 1 1 色法では、複数のアレイスライドを比較するので、それらのスライド間でスポットの品質が安定していることが必要である。
- 2 2 色法では、2 種類の色素の特性の違いによる偽陽性を排除するため、色素を入れ替えて繰り返し実験を行うことがある。
- 3 2 色法では、コントロールサンプルに加えて 2 サンプルの測定が可能であるため、10 点からなる時系列サンプルの測定には 5 枚のスライドを用いればよい。
- 4 1 色法、2 色法に関わらず、ハイブリダイゼーションの効率はプローブごとに異なるため、異なる遺伝子の発現レベルの比較は精度が低い。

問 74

次世代シーケンサを用いたトランスクリプトーム解析について述べた以下の記述のうち、もっとも不適切なものを一つ選べ

- 1 リード数に応じて mRNA の検出感度が変化する。
- 2 キャピラリーシーケンサと比べるとリード長が短い。
- 3 未知の遺伝子の mRNA は検出できない。
- 4 出力されるデータ量が多いため、膨大な記録媒体が必要になる。

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 A においてサンプル B の 2 倍量発現している。
- 2 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。
- 3 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。
- 4 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。

問 76

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 1 つ選べ。

- 1 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。
- 2 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。
- 3 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。
- 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 77

遺伝子の転写速度 v が転写因子 Y の濃度 $[Y]$ を用いて、

$$v = k \frac{K}{K + [Y]}$$

と表すことができる場合を考える。ただし、 k は転写速度定数、 K は結合定数である。この式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 最大転写速度 v_{\max} (v の最大値) は k に等しい。
- 2 最大転写速度の半分の速度 ($\frac{v_{\max}}{2}$) を与えるときの $[Y]$ は K と等しい。
- 3 $[Y]$ が一定のとき、 K の値が大きくなると転写速度 v は速くなる。
- 4 転写因子 Y は転写を活性化する転写因子である。

問 78

細胞濃度 X の時間変化が比例定数 k (>0) を用いて、

$$\frac{dX}{dt} = kX$$

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 ダブリングタイムは $\frac{\ln 2}{k}$ で与えられる。
- 2 ダブリングタイムは $\frac{k}{\ln 2}$ で与えられる。
- 3 細胞濃度 X が高くなるにつれてダブリングタイムは長くなる。
- 4 細胞濃度 X が高くなるにつれてダブリングタイムは短くなる。

問 79

時間の関数である変数 y_1, y_2 からなるシステムの連立微分方程式が

$$\frac{dy_1}{dt} = 2y_1 - 3y_2$$

$$\frac{dy_2}{dt} = 4y_1 - 5y_2$$

のように表される。行列

$$\begin{pmatrix} 2 & -3 \\ 4 & -5 \end{pmatrix}$$

の固有値を用いて、システムの安定性を論じることができるが、平衡状態（原点）におけるシステムの安定性に関する記述でもっとも適切なものを選択肢の中から一つ選べ。安定とは、システムに摂動が与えられた場合、すみやかに元の平衡状態にもどれることをいう。

- 1 固有値は1, 2であるので、システムは安定である。
- 2 固有値は1, 2であるので、システムは不安定である。
- 3 固有値は -1, -2であるので、システムは安定である。
- 4 固有値は -1, -2 であるので、システムは不安定である。

問 80

右図のようなタンパク質 X、Y、Z からなるフィードフォワードループがある。シグナル S_x を感知して、タンパク質 Z が合成される。Z の合成式は

$$\frac{dZ}{dt} = \theta_x \cdot \theta_y - Z$$

$$\theta_x = \begin{cases} 1 & S_x > 0 \\ 0 & S_x = 0 \end{cases} \quad \theta_y = \begin{cases} 1 & Y > 0.5 \\ 0 & Y \leq 0.5 \end{cases}$$

である。斜体の文字は対応する物質の濃度を表すものとする。このとき、下図に与えられた S_x と Y の時間変化に対する Z の時間変化曲線でもっとも適切なものを選択肢の中から一つ選べ。

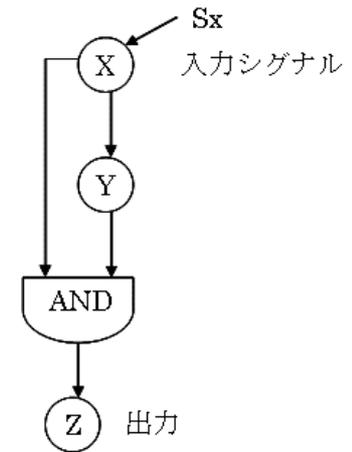
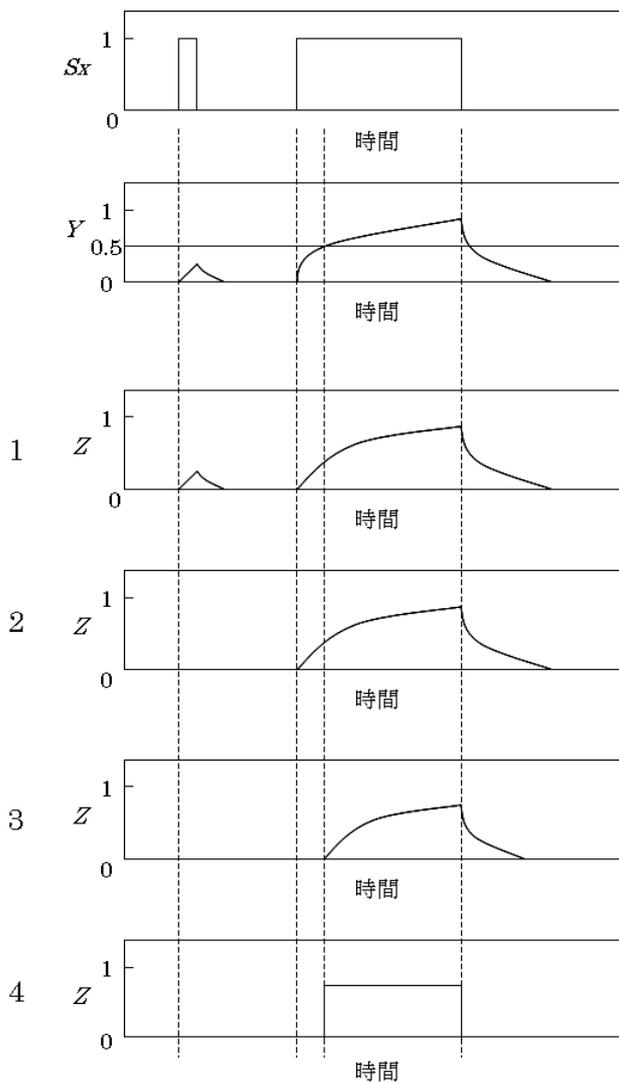


図 タンパク質 X、Y、Z のフィードフォワードループ



平成 22 年度

日本バイオインフォマティクス学会 (JSBi)

バイオインフォマティクス技術者認定試験

【解答】

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