

平成 20 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

※無断転載を禁じます。

Copyright © 2008 Japanese Society for Bioinformatics. All Rights Reserved.

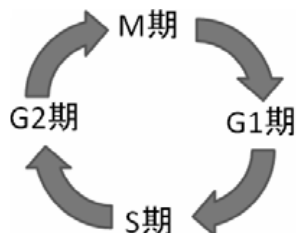
問 1

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

- 1 真核生物の細胞小器官であり、外膜と内膜の二つの脂質二重膜に包まれている。
- 2 好気性細菌が真核生物の祖先に共生することによって獲得されたと考えられている。
- 3 ミトコンドリア DNA は、父親から子に受け継がれる特性を利用して、家系を追跡するための研究に利用される。
- 4 脂肪酸の β 酸化や、電子伝達系による酸化リン酸化によるエネルギー生産を行う細胞小器官である。

問 2

細胞分裂における細胞周期では、細胞の状態は4つの「期」に分けられて解釈されている。G1期において細胞内部で起こっている現象としてもっとも適切なものを選択肢の中から一つ選べ。



- 1 細胞分裂期
- 2 細胞分裂の準備期
- 3 DNA 合成期
- 4 DNA 合成の準備期

問 3

転写調節に関連する記述として適切でないものを選択肢の中から一つ選べ。

- 1 シス調節領域とは、転写活性を調節するタンパク質に認識される DNA 上の塩基配列をさす。
- 2 シス調節領域は、それによって発現が調整される構造遺伝子の上流に常に存在する。
- 3 DNA 上の特定の配列を認識して、転写活性を調節するタンパク質をトランス調節因子と呼ぶ。
- 4 プロモーターとエンハンサーはシス調節領域である。

問 4

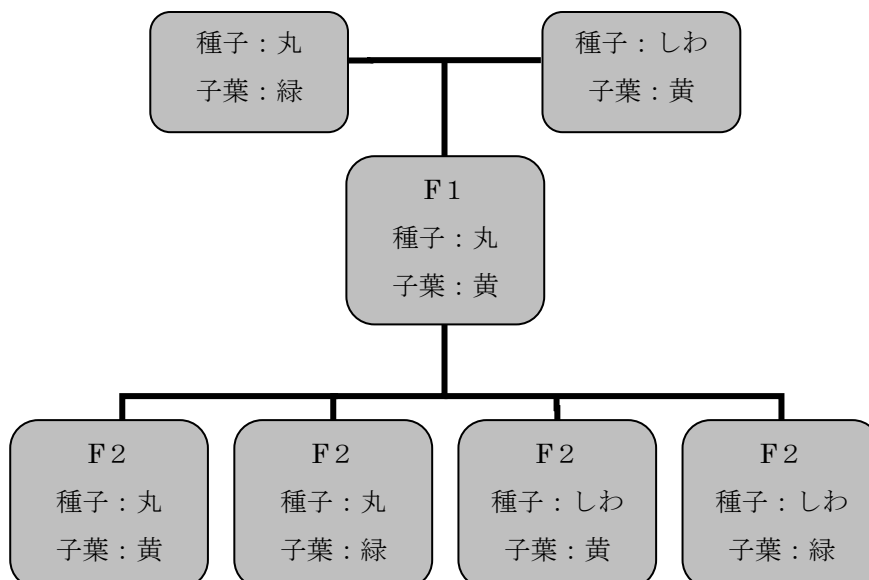
ヒトの血液型の遺伝に関して、B 型の父親と AB 型の母親から産まれる子供の血液型は何か。可能性をもつ全ての血液型の組合せとしてもっとも適切なものを選択肢の中から一つ選べ。

- 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 5

エンドウの種子の形には「丸」と「しわ」の対立形質があり、子葉の色には「黄」と「緑」の対立形質がある。純系の「丸・緑」と「しわ・黄」の F1 は「丸・黄」となるが、その F2 表現型の分離比（丸・黄：丸・緑：しわ・黄：しわ・緑）は次のうちのどれになるか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 9 : 3 : 3 : 1
- 2 4 : 2 : 2 : 1
- 3 1 : 1 : 1 : 1
- 4 2 : 1 : 2 : 1



問 6

ヒトの生命活動に必要な ATP を産生するしくみ（内呼吸）ともっとも関係の低いものを選択肢の中から一つ選べ。

- 1 カルビン回路
- 2 解糖系
- 3 電子伝達系
- 4 クエン酸回路

問 7

生体内で、一般的にエネルギー通貨（貨幣）と呼ばれている物質はどれか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 ATP
- 2 CTP
- 3 GTP
- 4 TTP

問 8

次に示したアミノ酸配列のうち、もっとも疎水性が高いと考えられるものを選択肢の中から一つ選べ。

- 1 Asp (D) - Thr (T) - Glu (E)
- 2 Val (V) - Leu (L) - Ile (I)
- 3 Gln (Q) - Asn (N) - Ser (S)
- 4 His (H) - Arg (R) - Lys (K)

問 9

タンパク質の翻訳後修飾に関する以下の記述について不適切なものを選択肢の中から一つ選べ。

- 1 リン酸化、糖鎖付加（グリコシル化）、脂質付加、水酸化、ユビキチン化などがある。
- 2 翻訳後修飾によって、機能の発現と調節、安定化、細胞内移行や局在化、細胞外への分泌などが制御される。
- 3 GPI アンカーはプロテインフォスファターゼにより切断される。
- 4 *O*-グリコシル化はセリン残基やスレオニン残基への糖鎖付加によって起こる。

問 10

生体膜に関する以下の記述について から 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。

生体膜の基本をなすのは であり、それ以外にステロールや糖脂質が含まれる。

は の部分と炭化水素鎖の の部分からなる極性分子であり、 の部分を外側に、 の部分を内側に向けて互いに集まり、流動性の脂質二重層を作るか、または として球状の構造になる。

- 1 (a) 糖鎖 (b) 疎水性 (c) 親水性 (d) ミセル
- 2 (a) リン脂質 (b) 親水性 (c) 疎水性 (d) ミセル
- 3 (a) 糖鎖 (b) 親水性 (c) 疎水性 (d) ラフト
- 4 (a) リン脂質 (b) 疎水性 (c) 親水性 (d) ラフト

問 11

神経伝達ともっとも関係の低いものを選択肢の中から一つ選べ。

- 1 GABA
- 2 シャペロニン
- 3 アセチルコリン
- 4 セロトニン

問 12

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

- 1 4,000 塩基
- 2 6,000 塩基
- 3 8,000 塩基
- 4 12,000 塩基

問 13

コドンとアンチコドンの塩基対形成に関連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 コドンの三番目とアンチコドンの一番目の間で形成される非 Watson-Crick 型の塩基対のことをシュードノット (pseudoknot) と呼ぶ。
- 2 コドンの三番目のシトシンやウラシルは、アンチコドンの一番目のイノシンと対合できる。
- 3 アンチコドンの一番目にあるウラシルは修飾されている場合がある。
- 4 コドンの三番目のアデニンと対合するアンチコドンの一番目の塩基は、原核生物の場合ウラシルのみだが、真核生物の場合ウラシルとイノシンがある。

問 14

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。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 15

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不適切なものを選択肢の中から一つ選べ。

- 1 RB や p53 はがん抑制遺伝子として知られている。
- 2 ras や myc はがん遺伝子として知られている。
- 3 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。
- 4 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。

問 16

クローニングに関する以下の記述について不適切なものを選択肢の中から一つ選べ。

- 1 ショットガンクローニングの一つの方法として、制限酵素を用いてゲノム DNA を断片化し、その断片を制限酵素で切断したプラスミドに DNA リガーゼを用いて組み込むことが行われる。
- 2 多数の DNA クローンの塩基配列の重なりを見つけながら、それらのクローンをアセンブルすることによってゲノム全体の配列を知ることができる。
- 3 プラスミドやファージは、細菌内で増殖するので、それらにクローニングしたい遺伝子を組み込み、細菌内に送り届けるベクターに使うことができる。
- 4 大腸菌と酵母のような異なる生物種の両方で増殖するプラスミドベクターは、両生物種で共通に機能する一つの複製開始点 (ori) をもつ。

問 17

PCR (ポリメラーゼ連鎖反応) について、不適切なものを選択肢の中から一つ選べ。

- 1 プライマーが必要である。
- 2 少量のサンプル (DNA または RNA) があれば増幅することができる。
- 3 タンパク質を増幅することができる。
- 4 制限酵素認識部位を作成することができる。

問 18

次に示した分子生物学の実験手法名の中で、通常使われないものはどれか。選択肢の中から一つ選べ。

- 1 イースタンブロットティング
- 2 サザンブロットティング
- 3 ウェスタンブロットティング
- 4 ノーザンブロットティング

問 19

プロテオームに関する以下の記述について不適切なものを選択肢の中から一つ選べ。

- 1 2次元ゲル電気泳動はタンパク質の等電点と分子量の2つのパラメーターの違いに基づいて、タンパク質を分離する。
- 2 タンパク質を構成するアミノ酸の配列を知るために質量分析装置が利用できる。
- 3 タンパク質相互作用を実験的に同定するために酵母ツーハイブリッド法を用いることができる。
- 4 タンパク質はしばしばリン酸基で修飾されているが、リン酸基の分子量はタンパク質全体からみれば小さいので、質量分析装置ではそのようなタンパク質修飾を検知できない。

問 20

タンパク質など生体高分子の立体構造を決定する実験的手法として不適切なものを選択肢の中から一つ選べ。

- 1 赤外分光法
- 2 X線回折法
- 3 NMR法
- 4 電子顕微鏡法

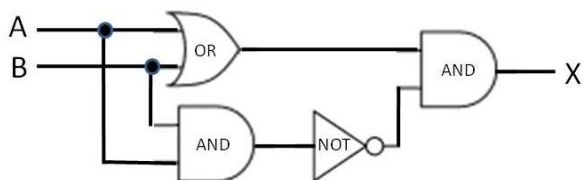
問 21

コンピュータ上では、数値は二進数で扱われる。十進数の 2008 を二進数で表現すると「11111011000」となる。ここで、2008 を 4 で割った商である十進数の 502 は、二進数ではどのように表現されるか。適切なものを選択肢の中から一つ選べ。

- 1 111110110
- 2 111011000
- 3 111110101
- 4 111010111

問 22

下記の図（ア）は、いくつか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 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（イ）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適切なものを選択肢の中から一つ選べ。



A	B	X
0	0	(a)
0	1	1
1	0	1
1	1	(b)

図（ア） 論理回路図

表（イ） 真理値表

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

問 23

近年のコンピュータシステムでは、ある技術により、システム上に物理的に実装されているメモリ量を超えるような大規模なプログラムや複数のプログラムを実行することが可能となっている。この技術の名称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 レジスタ
- 2 仮想記憶
- 3 キャッシュ機構
- 4 半導体ディスク

問 24

コンピュータをローカルエリアネットワーク（DHCP で無い場合）に接続する際の設定で必要でないものを選択肢の中から一つ選べ。

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

問 25

与えられたデータの列を一定の順序に並べ替えるソートアルゴリズムの一つに、バブルソートがある。バブルソートでは、データ列の最初の要素から最後の要素に向かって、隣接する2つの要素を順に比較し、2つの要素が正しい順番に並んでいない時には交換を行うという処理を必要なだけ繰り返していく。データ間の交換がもう必要ないと判断できた場合には、そこで処理を終了するものとする。このバブルソート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 バブルソートは、各種のソーティングアルゴリズムのうち交換法的一种に分類される。
- 2 最初の要素から最後の要素までの比較が済むと、最後の要素については最小値または最大値として確定するので、次回からは比較の範囲を1つ狭めても良い。
- 3 データが初めから正順に並んでいたときは計算が最も早く終了し、逆順に並んでいたときは計算が最も遅くなる。
- 4 理解しやすい簡便な方法ではあるが、選択法に比べて必要なメモリ容量が多くなるという欠点がある。

問 26

ソートのアルゴリズムに関する以下の記述について間違っているものを選択肢の中から一つ選べ。

- 1 与えられたデータ数が n のとき、バブルソートの計算量 $O(n^2)$ である。
- 2 与えられたデータ数が n のとき、クイックソートの計算量は $O(n \log n)$ である。
- 3 左から右に数値データを昇順に並べるとき、クイックソートではまずピボット M より小さいデータを M の左に、 M より大きいデータを M の右にあつめる。そのあと、 M の左側および右側に対してクイックソートを再帰的に適用することによりソートが完了する。
- 4 左から右に数値データを昇順に並べるとき、直接挿入法は、左側から昇順に整列させながら、その整列済みの範囲を左側にむかって徐々に拡大させながら整列する方法である。

問 27

データ格納方法としてキューを用いる。アルファベットの文字データをキューに出し入れする。データの操作を

A enqueue

B enqueue

C enqueue

dequeue

dequeue

D enqueue

A enqueue

dequeue

dequeue

のように行ったとき、最後の dequeue で取り出されるデータとして正しいものを選択肢の中から一つ選べ。ただし、enqueue はデータを追加する操作、dequeue はデータを取り出す操作である。

1 A

2 B

3 C

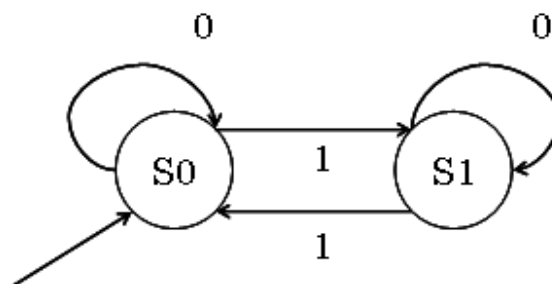
4 D

問 28

下図の決定性有限オートマトンは 2 つの状態 S0 と S1 をもつ。S0 は初期状態である。入力数列は 1 あるいは 0 である。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。

表 状態遷移表

入力状態	1	0
S0	S1	S0
S1	S0	S1



- 1 入力数列が 00111 のとき、状態は S1 となる。
- 2 入力数列中に 1 が偶数個あるときは、状態は S0 となる。
- 3 入力数列が 01010010101 のとき、状態は S1 となる。
- 4 入力数列中に 0 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、インデックス（索引）がある。インデックス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 件数の大きなデータほど、インデックスをあらかじめ作成しておく効果大きい。
- 2 検索されるデータの 1 件ごとのサイズが大きいときは、あらかじめインデックスを作成しておく効果があまり得られない。
- 3 一般的には、データが頻繁に挿入または削除されるケースよりは、データが静的な場合のほうがこの手法を導入しやすい。
- 4 データの挿入または削除があっても、インデックスの大幅な作り直しは保留して、データを特殊な領域に書き留めておくという方法もある。

問 30

次に示した説明文のなかで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2次元のテーブルとして表現することができる。
- 2 結合演算を用いて、リレーション間の情報を結びつけることができる。
- 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 Ruby がある。

問 31

大学4年生の学籍簿に対して下記のSQL文を実行する。結果として得られるデータでもっとも適切なものを選択肢の中から一つ選べ。

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

表：学籍簿

学籍番号	名前	評価点	研究室番号
1000	伊藤かおり	71	302
1001	坂本京子	83	301
1002	田中正	77	202
1003	藤堂新一	82	301
1004	山田太郎	85	101

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

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 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。
- 4 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。

問 33

コインを投げ、表が出たか、裏が出たかを調べるという試行をくり返す。このコインの表、裏が出る確率はそれぞれ $1/2$ であり、各試行は独立であるとする。この試行に関連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 この試行を繰り返した時、10 回続けて表が出た。この時、その次の試行では裏が出る確率が高い。
- 2 このコイン投げを 100 回行った時、表が出る回数の期待値は 50 回である。
- 3 5 回の試行で、表が 3 回、裏が 2 回出る確率は、 $\frac{5!}{3!2!} \left(\frac{1}{2}\right)^5$ と計算される。
- 4 この問題では、コインの表あるいは裏が出ること以外の事象（コインが立つ等）の確率は 0 と仮定されている。

問 34

二つの確率変数 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。

$P(A, B)$ A 、 B の同時確率

$E(A)$ A の平均

$V(A)$ A の分散

- 1 X 、 Y が独立の時、 X 、 Y の同時確率 $P(X, Y)$ は、それぞれの確率の積 $P(X)P(Y)$ に等しい。
- 2 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 $P(X|Y)$ は、次の式で表される。 $P(X, Y) / P(Y)$
- 3 X の分散 $V(X)$ は、 X^2 の平均と X の平均の二乗を用いて次のように表される。 $E(X^2) - [E(X)]^2$
- 4 X 、 Y が独立でない時、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。

問 35

それぞれ、平均 0、分散 1 の標準正規分布に従う独立な確率変数 X と Y 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 X と X^2 の相関係数は 0 である。
- 2 $X + Y$ の分散は 1 である。
- 3 $X + Y$ は正規分布に従う。
- 4 X と Y の相関係数は 0 である。

問 36

次のクラスタ解析の手法の中で非階層的クラスタリングではないものを選択肢の中から一つ選べ。

- 1 K-平均法
- 2 自己組織化マップ
- 3 Fuzzy c-means
- 4 UPGMA

問 37

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

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

問 38

1,000 人のサンプルに関してある性質の陽性・陰性の予測を試みたところ、50 人が陽性、950 人が陰性との予測結果になった。その後、全てのサンプルに対して厳密な実験を行ったところ、実際には下記の表に示すような確定結果を得た。この手法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。

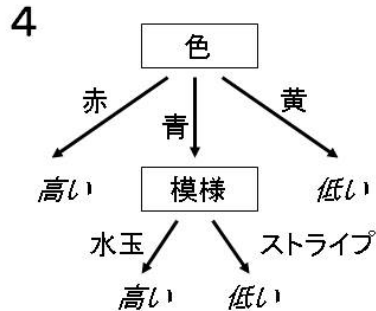
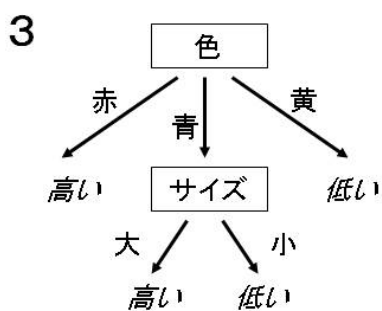
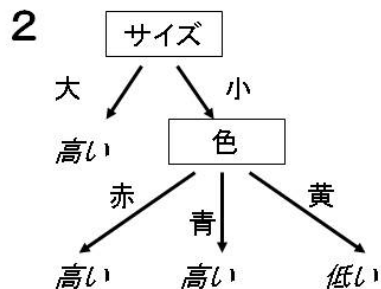
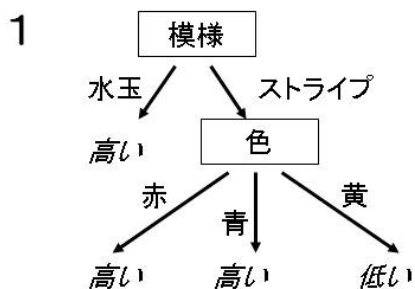
		確定結果	
		陽性	陰性
予測結果	陽性	45 人	5 人
	陰性	5 人	945 人

- 1 確定結果が陽性である人について、90%の確率で正しく陽性と予測できた。
- 2 確定結果が陰性である人について、99%以上の確率で正しく陰性と予測できた。
- 3 陽性と予測された人のうちおよそ 10%は、実際には陰性であった。
- 4 陰性と予測された人のうちおよそ 5%は、実際には陽性であった。

問 39

意志の決定プロセスなどをグラフを用いて表現する方法に決定木がある。ある商店において、売られている 6 種類のマグカップの売れ行きを調べた結果、次のような結果が得られた。この表を元にして売れ行きを予測するための決定木を作成した。もっとも適切なものを選択肢の中から一つ選べ。

マグカップの特徴			マグカップの 売れ行き
サイズ	色	模様	
大	赤	ストライプ	高い
小	赤	水玉	高い
大	青	ストライプ	低い
大	青	水玉	高い
小	黄色	ストライプ	低い
大	黄色	水玉	低い



問 40

下図のようなプロファイル HMM がある。状態間の遷移確率は図中に示されている。
各状態での文字の出力確率は、M1、M2、M3 の各状態では以下のようにになっている。

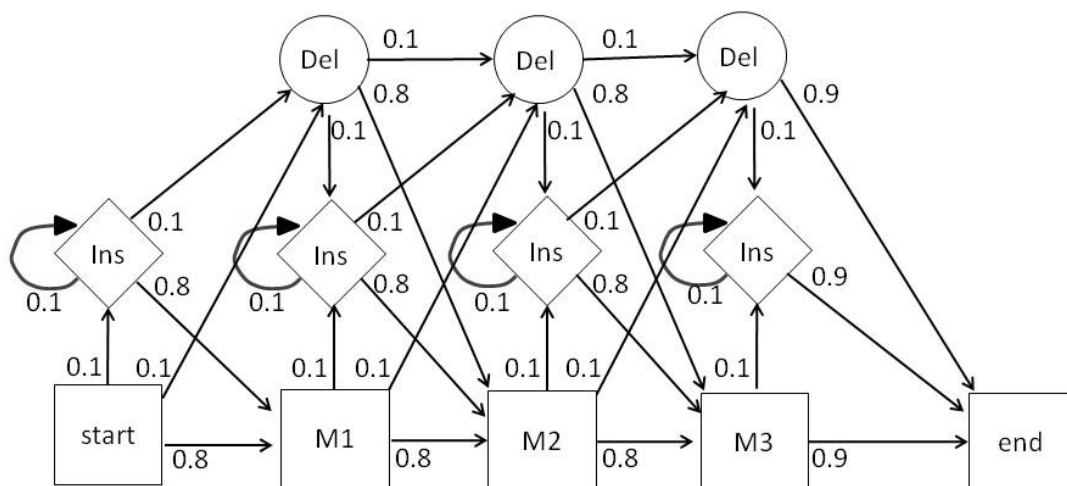
M1 : $P(A)=0.0$ 、 $P(T)=0.5$ 、 $P(G)=0.5$ 、 $P(C)=0.0$

M2 : $P(A)=1.0$ 、 $P(T)=0.0$ 、 $P(G)=0.0$ 、 $P(C)=0.0$

M3 : $P(A)=0.0$ 、 $P(T)=0.0$ 、 $P(G)=0.9$ 、 $P(C)=0.1$

また Ins 状態では、位置によらず $P(A)=P(T)=P(G)=P(C)=0.25$ である。

この HMM から、「start → T → A → G → A → end」という出力文字列が観測されたとする。以下の選択肢に示された状態遷移パスのうち、この出力文字列を発生させる確率をもっとも大きいものを一つ選べ。



T A G A

- 1 start → Ins → M1 → M2 → M3 → end
- 2 start → M1 → Ins → M2 → M3 → end
- 3 start → M1 → M2 → Ins → M3 → end
- 4 start → M1 → M2 → M3 → Ins → end

問 41

次に示したデータベースに登録されているデータの内容とデータベース名の組み合わせの中で不適切なものを選択肢の中から一つ選べ。

- 1 疾患データベース – CDD、CATH、FSSP
- 2 塩基配列データベース – GenBank、EMBL、DDBJ
- 3 遺伝子多型データベース – dbSNP、HGVB、HapMap
- 4 パスウェイデータベース – KEGG、TransPath、AfCS

問 42

タンパク質モチーフ統合データベースである InterPro に組み込まれているデータベースとして不適切なものを選択肢の中から一つ選べ。

- 1 SMART
- 2 COG
- 3 PROSITE
- 4 Pfam

問 43

遺伝子発現のデータベースとして不適切なものを選択肢の中から一つ選べ。

- 1 NCBI GEO
- 2 ArrayExpress
- 3 MIAME
- 4 Stanford Microarray Database

問 44

以下に示すデータはフラットファイル形式で記された GenBank の遺伝子エントリーの一部である。このデータ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。なお、出題と関係していない項目については、適宜省略してある。

```

LOCUS       NC_010468                2031 bp    DNA    linear    BCT 30-JUL-2008
SOURCE      Escherichia coli ATCC 8739
FEATURES             Location/Qualifiers
     source          1..2031
                     /organism="Escherichia coli ATCC 8739"
                     /mol_type="genomic DNA"
     CDS             1..2031
                     /locus_tag="EcolC_0143"
                     /note="PFAM: alpha amylase catalytic region
                     SMART: alpha amylase catalytic sub domain
                     KEGG: sfx:S4154 alpha-amylase
                     /codon_start=1
                     /transl_table=11
                     /product="alpha amylase catalytic region"
                     /translation=" MKLAACFLTLLPGFAVAASWTSPGFPAFSEQGTGTFVSHAQLPK
                     ***** 略 *****
                     GTRPLTLNFDQQC"
     sig_peptide     1..54
ORIGIN
1 atgaaactcg cgcctgttt tctgacactc cttcctggct tcgccgttgc cgccagctgg
61 acttctccgg ggtttcccg ctttagcgaa caggggacag gaacatttgt cagccacgcg
***** 略 *****
1981 cgtgagcatg gcgacgataa agtgctggtc gtctgggcag ggcaacagta a
//

```

- 1 2031 塩基配列からなる大腸菌の遺伝子である。
- 2 cDNA から塩基配列が解読されていることがわかる。
- 3 遺伝子産物がアミラーゼ活性をもつことがわかる。
- 4 シグナルペプチドをもつタンパク質をコードしていることがわかる。

問 45

下図は PDB データベースで、あるがん遺伝子を検索して得られたデータの一部である。その記述について不適切なものを選択肢の中から一つ選べ。

なお、出題と関係していない項目については、適宜省略してある。

```

HEADER      PROTO-ONCOGENE                      27-JAN-97  1AA9
TITLE       HUMAN C-HA-RAS (1-171) (DOT) GDP, NMR, MINIMIZED AVERAGE
TITLE       2 STRUCTURE
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: C-HA-RAS;
COMPND      3 CHAIN: A;
COMPND      4 FRAGMENT: RESIDUES 1 - 171;
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM SCIENTIFIC: HOMO SAPIENS;
SOURCE      3 ORGANISM COMMON: HUMAN;
SOURCE      4 GENE: HUMAN C-HA-RAS GENE;
EXPDTA      NMR
AUTHOR      Y. ITO, Y. YAMASAKI, Y. MUTO, G. KAWAI, S. NISHIMURA, T. MIYAZAWA,
AUTHOR      2 S. YOKOYAMA, RIKEN STRUCTURAL GENOMICS/PROTEOMICS INITIATIVE
AUTHOR      3 (RSGI)
REVDAT      1 29-JUL-97 1AA9 0
JRNL        AUTH  Y. ITO, K. YAMASAKI, J. IWAHARA, T. TERADA, A. KAMIYA,
    
```

中略

```

DBREF 1AA9 A 1 171 UNP P01112 RASH_HUMAN 1 171
SEQRES 1 A 171 MET THR GLU TYR LYS LEU VAL VAL VAL GLY ALA GLY GLY
SEQRES 2 A 171 VAL GLY LYS SER ALA LEU THR ILE GLN LEU ILE GLN ASN
    
```

中略

```

HELIX 1 H1 LYS A 16 ILE A 24 1 9
HELIX 2 H2 ASP A 69 ARG A 73 1 5
HELIX 3 H3 GLU A 91 LYS A 104 1 14
HELIX 4 H4 GLU A 126 TYR A 137 1 12
HELIX 5 H5 VAL A 152 GLN A 165 1 14
SHEET 1 S1 6 SER A 39 ILE A 46 0
SHEET 2 S1 6 GLU A 49 ASP A 57 -1 0 LEU A 53 N LYS A 42
SHEET 3 S1 6 THR A 2 VAL A 9 1 0 LEU A 6 N LEU A 56
SHEET 4 S1 6 GLY A 77 ILE A 84 1 0 LEU A 79 N VAL A 9
SHEET 5 S1 6 PRO A 110 LYS A 117 1 0 VAL A 112 N CYS A 80
SHEET 6 S1 6 PRO A 140 THR A 144 1 0 ILE A 142 N GLY A 115
    
```

中略

```

ATOM 1 N MET A 1 129.640 14.669 -14.847 1.00 2.13 N
ATOM 2 CA MET A 1 130.624 13.884 -14.049 1.00 1.42 C
ATOM 3 C MET A 1 130.508 12.401 -14.404 1.00 1.16 C
ATOM 4 O MET A 1 129.625 11.993 -15.134 1.00 1.52 O
ATOM 5 CB MET A 1 130.340 14.070 -12.561 1.00 0.96 C
ATOM 6 CG MET A 1 131.298 15.114 -11.985 1.00 1.06 C
    
```

中略

```

ATOM 17 H1 MET A 1 129.044 14.019 -15.399 1.00 2.33 H
ATOM 18 H2 MET A 1 129.043 15.230 -14.208 1.00 2.50 H
ATOM 19 H3 MET A 1 130.148 15.307 -15.492 1.00 2.68 H
ATOM 20 N THR A 2 131.392 11.591 -13.891 1.00 0.74 N
ATOM 21 CA THR A 2 131.334 10.134 -14.195 1.00 0.55 C
ATOM 22 C THR A 2 130.993 9.362 -12.927 1.00 0.46 C
ATOM 23 O THR A 2 131.108 9.865 -11.830 1.00 0.48 O
    
```

後略

- 1 この情報は、ヒト Ha-Ras タンパク質の立体構造に関するものである。
- 2 このエントリーの ID は 1AA9 である。
- 3 このタンパク質の 16 (Lys) から 24 (Ile) までは、 α ヘリックス構造をとっている。
- 4 このタンパク質の 1 (Met) の C α の座標は (130.508, 12.401, -14.404) である。

問 46

生物学分野での代表的なオントロジーデータベースである **Gene Ontology** では、遺伝子産物を 3つの意味概念によって階層的に分類している。そのうち、二つは、**molecular function** と **cellular component** である。残り一つは、何と呼ばれるか。選択肢の中から一つ選べ。

- 1 molecular structure
- 2 biological process
- 3 metabolic pathway
- 4 sequence homology

問 47

4 塩基よりなる塩基配列のモチーフを、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、AGAGGTC を検索した時に、最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 AGAG
- 2 GAGG
- 3 AGGT
- 4 GGTC

問 48

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

```

D D D G W
| . . | |
D E E G W
    
```

- 1 35
- 2 27
- 3 23
- 4 22

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

問 49

長さ 5 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、適切な値の一つを選べ。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、 $p(A) = p(G) = p(T) = p(C) = 0.25$ 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 50

以下の 2 本の塩基配列において、両配列間で対応する塩基数が最大となるように、ギャップの挿入を許すアラインメントを行う。塩基が対応するとは、A-A、T-T、G-G、C-C というように塩基が完全に一致することである。簡単のために、ギャップペナルティ、塩基配列の不一致については考慮しない。アラインメントした両配列の塩基が一致する最大数で、もっとも適切なものを選択肢の中から一つ選べ。

ATGCATGC

AATCAACG

- 1 3
- 2 4
- 3 5
- 4 6

問 51

DNA 塩基配列 2 本のグローバルアラインメントを動的計画法を用いて作成する。動的計画法の漸化式は、

$$D(i, j) = \text{Max} \begin{cases} D(i-1, j-1) + s(i, j) \\ D(i-1, j) - p \\ D(i, j-1) - p \end{cases}$$

とする。ここで、 $s(i, j)$ は、第一の配列の i 番目の塩基と第二の配列の j 番目の塩基が一致していれば 1、不一致であれば 0 の値をとる。 p は、ギャップペナルティであり、正の値 2 をとる。漸化式を 5'側から解き、 $D(i-1, j-1)$ 、 $D(i-1, j)$ 、 $D(i, j-1)$ は図のように既に求まっているとする。一方の配列の i 番目の塩基は G、他方の配列の j 番目の塩基は T とする。この時、 $D(i, j)$ の値を選択肢の中から一つ選べ。

		...	
...		$D(i-1, j-1)=9$	$D(i-1, j)=10$
...		$D(i, j-1)=8$	$D(i, j)$
		...	

- 1 7
- 2 8
- 3 9
- 4 10

問 52

FASTA に関する記述について不適切なものを選択肢の中から一つ選べ。

- 1 FASTA は、DNA の塩基配列やタンパク質のアミノ酸配列のデータベース検索を行うためのソフトウェアである。
- 2 塩基配列検索のとき、タプル (tuple) のサイズを k から $k+2$ にすると、検索速度は 32 倍速くなる。
- 3 FASTA は、部分一致文字列の検出にハッシュ表を用いている。
- 4 タプルのサイズが大きくなるとホモロジー検索速度は向上するが、検索の感度は低下する傾向がある。

問 53

相同性検索に用いられる BLAST には、クエリ配列と対象データベースのデータの種類によって使い分けられるいくつかの異なるバージョンがある。BLAST に含まれる blastp プログラムでの、クエリ配列と対象データベースの組み合わせとして適しているものを選択肢の中から一つ選べ。

	クエリ配列	対象データベース
1	タンパク質 (アミノ酸配列)	タンパク質 (アミノ酸配列)
2	タンパク質 (アミノ酸配列)	DNA 配列
3	DNA 配列	タンパク質 (アミノ酸配列)
4	DNA 配列	DNA 配列

問 54

配列データベースに対して相同性検索を行ったとき、あるしきい値 X よりも高いスコアを持つヒットが何個くらい得られるかについては、Karlin-Altschul の理論がある。すなわち、ギャップ無しの局所アラインメントに関しては、得られるヒット数の期待値 E は下式で与えられる。

$$E = m n 2^{-S}$$

ただし、 m は入力した問い合わせ配列の長さ、 n はデータベース側の配列の全長、 S はしきい値 X をビットスコアと呼ばれるスコアに換算した値である。

ここで、長さ 400 残基の配列を、全長 25 億残基 (2.5×10^9) のデータベースに対して検索したとき、ビットスコア $S=30$ のヒットはおおよそ何個得られるか。もっとも適切なものを選択肢の中から一つ選べ。ただし $\log_{10} 2 = 0.3010\dots$ である。

- 1 およそ 10 個
- 2 およそ 100 個
- 3 およそ 1,000 個
- 4 およそ 10,000 個

問 55

2 つのゲノム配列をハープロットによって解析してわかるものとして不適切なものを選択肢の中から一つ選べ。

- 1 相同性
- 2 反復部位
- 3 欠損部位
- 4 置換行列

問 56

真核生物のゲノム配列の中からアミノ酸をコードする部分を検出する際に利用する情報として不適切なものを選択肢の中から一つ選べ。

- 1 オープンリーディングフレーム (ORF)
- 2 GC 含量
- 3 スプライシングのドナーとアクセプター
- 4 一塩基変異 (SNP)

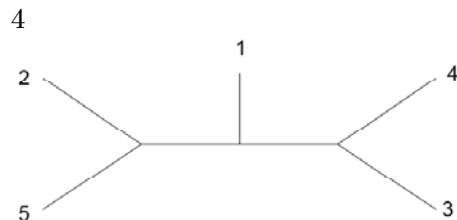
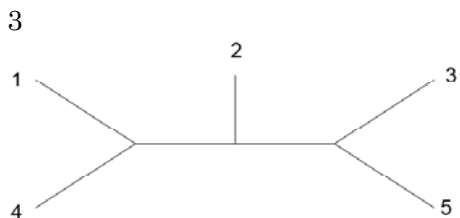
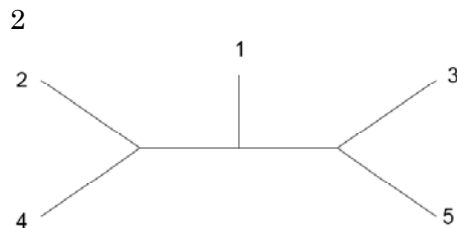
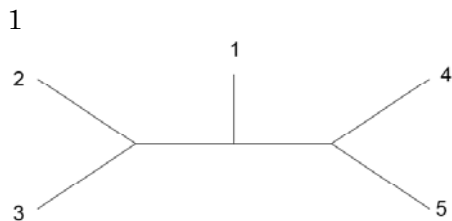
問 57

タンパク質をコードしている遺伝子の進化の過程における塩基置換は同義置換と非同義置換に分類される。次の同義置換と非同義置換に関する記述の中で不適切なものを選択肢の中から一つ選べ。

- 1 同義置換はアミノ酸を変化させない置換で、非同義置換はアミノ酸を変化させる置換である。
- 2 タンパク質レベルでの選択圧を受けないので、一般的には同義置換率（同義置換数/同義サイト数）は、非同義置換率（非同義置換数/非同義サイト数）よりも小さい。
- 3 ウイルスの表面抗原の抗体によって認識される部位では、抗体による認識を避けるためアミノ酸置換が加速されることがあり、その部位では、非同義置換率は同義置換率よりも大きくなる。
- 4 マウスとヒトのオーソログ遺伝子を比較する場合、同義置換率は遺伝子によらずほぼ一定の値になるが、非同義置換率はコードされる遺伝子によって大きく異なる。

問 58

系統樹を表記する方法として Newick format とよばれる括弧を利用した入れ子表現がある。5つの生物種の進化的関係についての Newick format による表現が $((1, 4), 2, (3, 5))$ であるとする。この表現で表わされる系統樹を選択肢の中から一つ選べ。



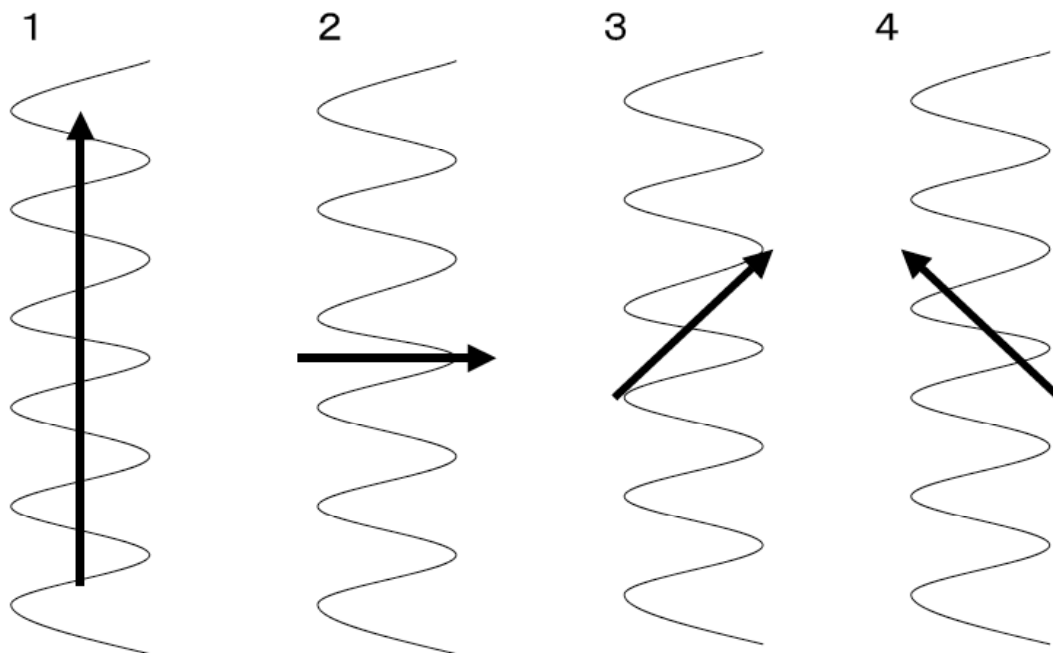
問 59

タンパク質の立体構造のグラフィクス表示プログラムではないものを選択肢の中から一つ選べ。

- 1 Chime
- 2 PyMol
- 3 MEGA
- 4 MolScript

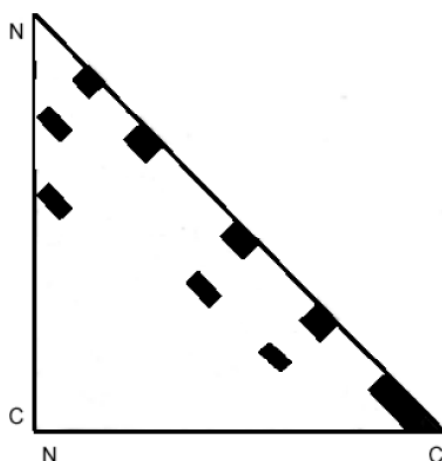
問 60

α ヘリックスが 15 個のアミノ酸より構成されているとする。一つのアミノ酸を、その $C\alpha$ のデカルト座標 (x, y, z) を 3 つの変数として持つデータと考え、先の α ヘリックスを構成する 15 個のアミノ酸に対して主成分分析を行う。第一主成分に対応するベクトルの、 α ヘリックスの構造に対しての向きを選択肢の中から一つ選べ。

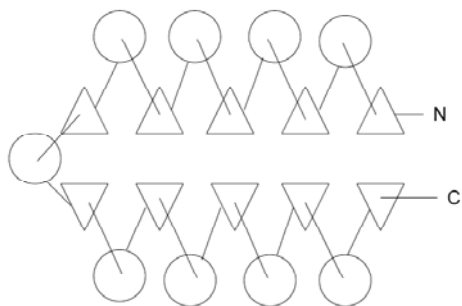


問 61

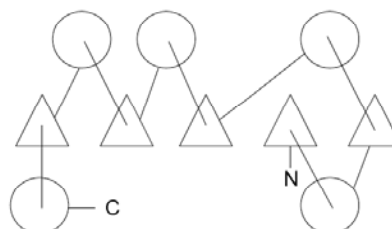
次の図は、あるタンパク質の立体構造のコンタクトマップの模式図である。図中斜めの太い線のそれぞれは、10Å以内にある残基対の集合を表す。このタンパク質は、N 末側から $\beta \alpha \beta \alpha \beta \alpha \beta \alpha \beta \alpha$ という二次構造の構成をとっている。βストランドの配向に着目した時、このタンパク質の構造として正しいものを選択肢の中から一つ選べ。図中、○はヘリックス、△はストランドであり、同じ向きの△の隣接は平行シート、逆向きの△の隣接は逆平行シートの形成を表しているとする。



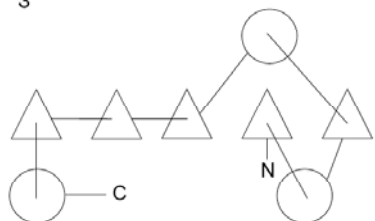
1



2



3



4



問 62

タンパク質の二次構造予測法として不適切なものを選択肢の中から一つ選べ。

- 1 GOR 法
- 2 NJ 法
- 3 Chou-Fasman 法
- 4 PHD 法

問 63

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つ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 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 64

タンパク質の立体構造予測に関連した下記の記述のうち、不適切なものを選択肢の中から一つ選べ。

- 1 Sippl らは、タンパク質配列を正しい立体構造と偽の立体構造とに当てはめたとき、この両者を残基間の安定性の差として区別できることを見だし、**Differential Display**法と名付けた。
- 2 Sali らが開発した **MODELLER** プログラムは、ホモロジー法に基づくタンパク質立体構造予測を行う。
- 3 Dunbrack らが開発した **SCWRL** は、側鎖ロータマーのライブラリである。
- 4 Eisenberg らは、与えられた立体構造から、それに最もふさわしいアミノ酸配列を求める逆フォールディング問題に取り組んだ。

問 65

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 66

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、不適切なものを選択肢の中から一つ選べ。

- 1 ゲノムアラインメントとは、オーソログ遺伝子のゲノム上の位置に基づき、それら遺伝子が最も対応する位置にくるように、全ゲノム構造を配置することである。
- 2 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。
- 3 ゲノムリアレンジメントに基づく生物間の進化的距離の一つに、編集距離 (**edit distance**) がある。
- 4 生物の種分化に伴い分岐した二つのゲノム中の相同遺伝子の関係をパラログスと呼ぶ。

問 67

5種類の生物について、6つの遺伝子 (A, B, C, D, E, F) のゲノム中の有無を調べ、各遺伝子の系統プロファイルを作成した。系統プロファイルは、生物種の数 (この例では 5 つ) の要素を持つベクトルとして表現され、遺伝子が有る場合は 1、無い場合は 0 が要素となる。以下の系統プロファイルから、その産物が相互作用すると予測される遺伝子の組み合わせを選択肢の中から一つ選べ。ただし、ベクトル中の要素 (生物種) の並びは、同じになっている。

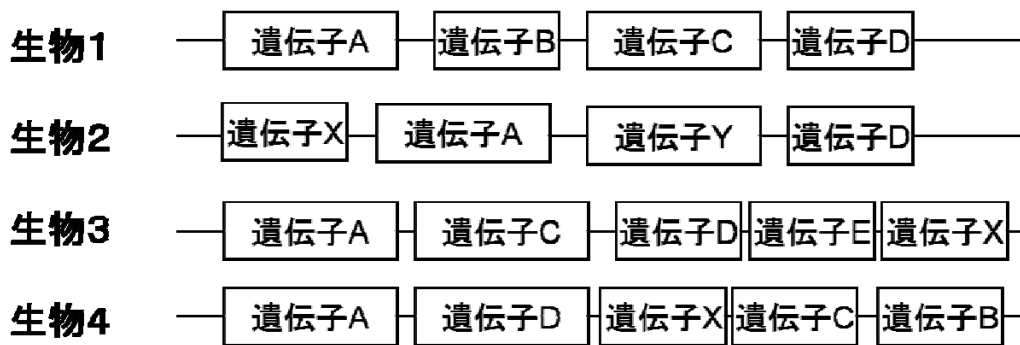
<系統プロファイル>

遺伝子 A	(1, 0, 0, 1, 1)
遺伝子 B	(1, 0, 1, 0, 0)
遺伝子 C	(0, 1, 0, 1, 0)
遺伝子 D	(1, 0, 0, 1, 1)
遺伝子 E	(1, 1, 1, 1, 1)
遺伝子 F	(0, 1, 1, 0, 0)

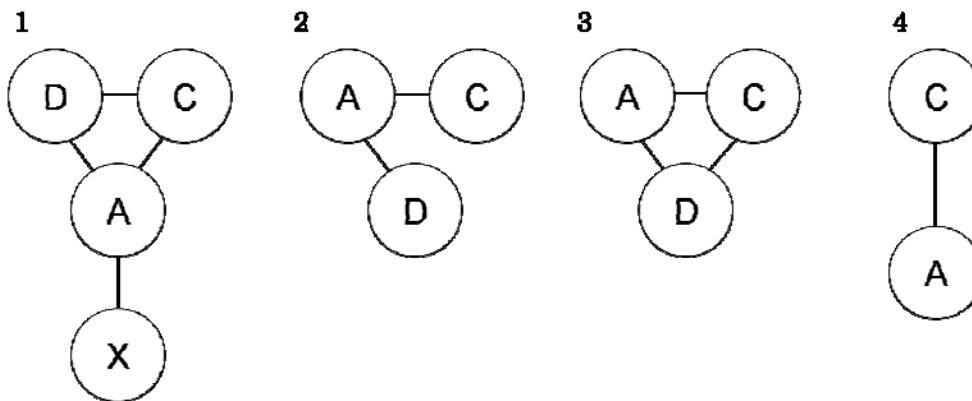
- 1 遺伝子 A、遺伝子 D
- 2 遺伝子 B、遺伝子 C
- 3 遺伝子 A、遺伝子 B、遺伝子 E
- 4 遺伝子 B、遺伝子 D、遺伝子 F

問 68

たがいに遠い進化的関係にある4種の原核生物のあるオペロン中の遺伝子の並びを示す。4種の原核生物中、3種以上の間で、オペロンの中で遺伝子の順番が保存されている場合、それらの遺伝子にコードされているタンパク質は相互作用すると予測する。遺伝子 α と β の順番が保存するとは、例えば、オペロン中で遺伝子 α が上流で、遺伝子 β が下流にあるという関係が複数のゲノムで観察されることをさす。この時、他の遺伝子はその二つの遺伝子 α と β の間に存在しても良いものとする。次のオペロン構造の図から予測されるタンパク質間相互作用としてもっとも適切なものを選択肢の中から一つ選べ。



4つの生物種のアペロン中での遺伝子の並び



○はその中の文字で示される遺伝子の産物、線は遺伝子産物間の相互作用を示す。

問 69

1 塩基多型 (SNP) の種類 (a) ~ (d) と、その意味 (イ) ~ (ニ) の正しい組み合わせを選択肢の中から一つ選べ。

- (a) cSNP
- (b) iSNP
- (c) rSNP
- (d) sSNP

(イ) 遺伝子の発現制御領域の SNP

(ロ) イントロン領域の SNP

(ハ) タンパク質をコードしている領域のアミノ酸置換を生じない SNP

(ニ) タンパク質をコードしている領域の SNP (狭義にはアミノ酸置換を生じる SNP)

- 1 (a) - (イ)、(b) - (ロ)、(c) - (ハ)、(d) - (ニ)
- 2 (a) - (イ)、(b) - (ハ)、(c) - (ロ)、(d) - (ニ)
- 3 (a) - (ニ)、(b) - (ロ)、(c) - (イ)、(d) - (ハ)
- 4 (a) - (ニ)、(b) - (イ)、(c) - (ロ)、(d) - (ハ)

問 70

ゲノムワイド関連解析 (GWAS) 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。

- 1 GWAS を行うことによって、がん細胞の突然変異部位が見つかる。
- 2 GWAS は、遺伝的に影響のある疾患関連遺伝子を探す手法である。
- 3 GWAS を行うためには、より多くの検体 (症例数) が望まれる。
- 4 GWAS を行うためには、より多くの SNP についての解析を行うことが望まれる。

問 71

疾患における環境因子と遺伝因子の影響を調べるために、ハワイの日系人を用いた疫学調査が有名である。□ a □ から □ c □ 内に入る語句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。

ハワイの日系人は日系人同士の結婚が多いので、□ a □ は日本人と似ていると考えられる。そのため、日本人と似た疾患パターンが得られた場合には □ b □ の影響が強いと考えられる。一方、世代が進むにつれて現地人と似た疾患パターンになる場合には、□ c □ の影響が強いと考えられる。

- | | | | |
|---|---------|----------|----------|
| 1 | (a) 食生活 | (b) 環境因子 | (c) 遺伝因子 |
| 2 | (a) 遺伝子 | (b) 環境因子 | (c) 遺伝因子 |
| 3 | (a) 遺伝子 | (b) 遺伝因子 | (c) 環境因子 |
| 4 | (a) 食生活 | (b) 遺伝因子 | (c) 環境因子 |

問 72

マイクロアレイにおいて二色法を用いて測定された生データの傾向を分析するために、MA-plot と呼ばれる図法が用いられる場合がある。M と A の二つの軸が描かれた二次元平面上に、各遺伝子に関する蛍光強度のデータがプロットされる。この MA-plot 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 縦軸は M (Minus) と呼ばれ、二色の強度の比 (対数の差) を表す軸である。
- 2 横軸は A (Add) と呼ばれ、二色の強度の積 (対数の和) を表す軸である。
- 3 誤差のない実験が行われていれば、プロットは完全に直線上に並ぶはずである。
- 4 もしも全体として右上がり (または右下がり) の傾向があれば、蛍光強度の強い領域で二つの色の片方が常に強めに (弱めに) 出ていることになり、系統的誤差が発見できる。

問 73

DNA マイクロアレイ実験において、複数のチップのデータを解析する際に、チップ間でのばらつきを補正する必要がある。この補正は一般に何と呼ばれるか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 正規化
- 2 log 変換
- 3 ブートストラップ
- 4 交差検証法

問 74

遺伝子発現データのクラスタリング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。

- 1 遺伝子発現データから、類似した発現パターンを示す遺伝子群をグループ化することをクラスタリングという。
- 2 遺伝子発現量の時間変化が類似している場合、それらの遺伝子が細胞内で協調して機能することが推定でき、遺伝子の未知機能や遺伝子間の依存関係を発見できる可能性がある。
- 3 遺伝子発現パターンのクラスタ数を決めるためには、発現データから推定した遺伝子間相互作用ネットワークの最大次数の値を用いるとよい。
- 4 遺伝子発現クラスタリングの精度を上げるために、関連の遺伝子のゲノム上にある転写因子のシス調節領域の比較が行われる。

問 75

5つの遺伝子 (A, B, C, D, E) から構成される人工的な遺伝子ネットワークの構造を推定するために、それぞれの遺伝子を破壊する実験を行った。実験と結果は以下の通りである。この実験の評価として不適切なものを選択肢の中から一つ選べ。

<実験と結果>

遺伝子 A を破壊すると、遺伝子 B、C の発現量が抑制された。

遺伝子 C を破壊すると、遺伝子 D の発現量が増加した。

遺伝子 B を破壊しても他の遺伝子の発現量には変化が見られなかった。

遺伝子 D を破壊すると、遺伝子 E の発現量が増加した。

- 1 遺伝子 A は、遺伝子 B の発現量を増加させる。
- 2 遺伝子 C は、遺伝子 D の発現量を抑制する。
- 3 遺伝子 A の発現量が増加すると、遺伝子 D の発現量は増加する。
- 4 遺伝子 C の発現量が増加すると、遺伝子 E の発現は増加する。

問 76

トランスクリプトーム解析で通常に比べて高い発現が見られた遺伝子の産物が、プロテオーム解析においては発現量に変化が無い場合、どの段階で制御されていたと考えられるか。もっとも不適切なものを選択肢の中から一つ選べ。

- 1 mRNA の分解による制御
- 2 タンパク質の分解による制御
- 3 翻訳開始における制御
- 4 翻訳速度による制御

問 77

タンパク質濃度 y が時間 t に関する微分方程式

$$\frac{dy}{dt} = k - y^2$$

で表される。 k は正の定数である。この方程式の記述に関して不適切なものを選択肢の中から一つ選べ。

- 1 定常状態($\frac{dy}{dt} = 0$) では、 $y = \sqrt{k}$ である。
- 2 この方程式はタンパク質合成のためのミカエリス・メンテン式とよばれる。
- 3 k をタンパク質 y の合成速度と考えることは妥当である。
- 4 y の分解は自己促進分解 (y が増加すればするほど、 y の分解速度が大きくなる) と考えることは妥当である。

問 78

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。

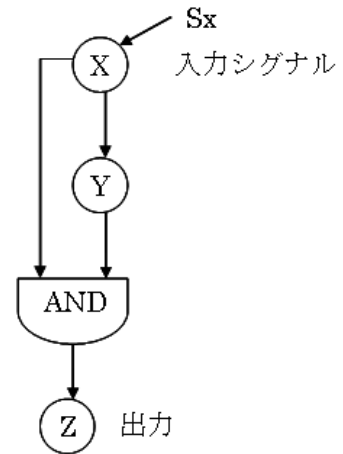
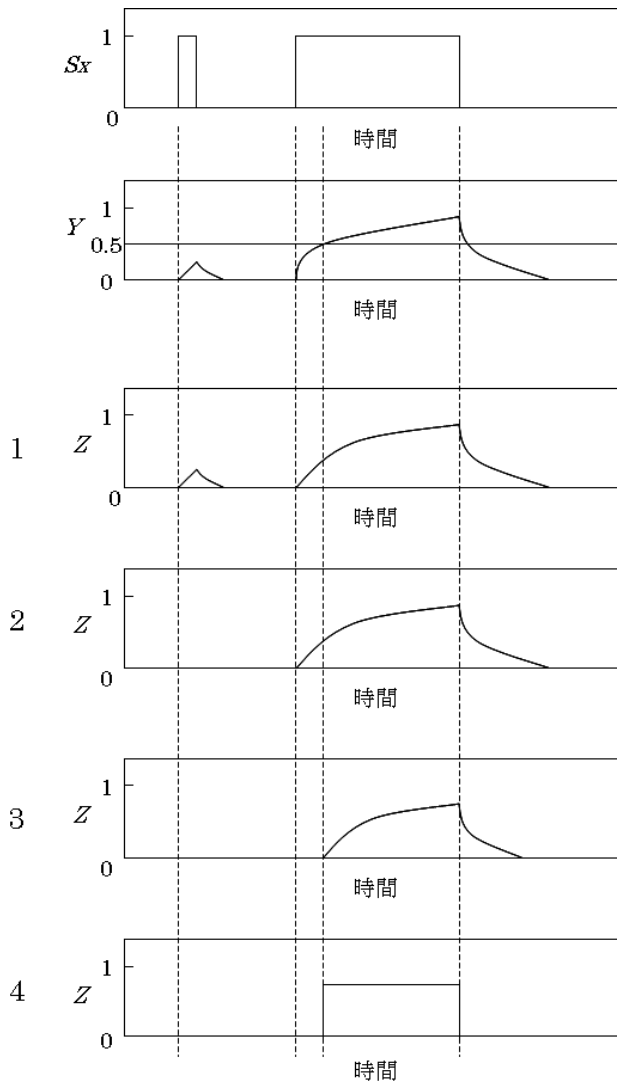
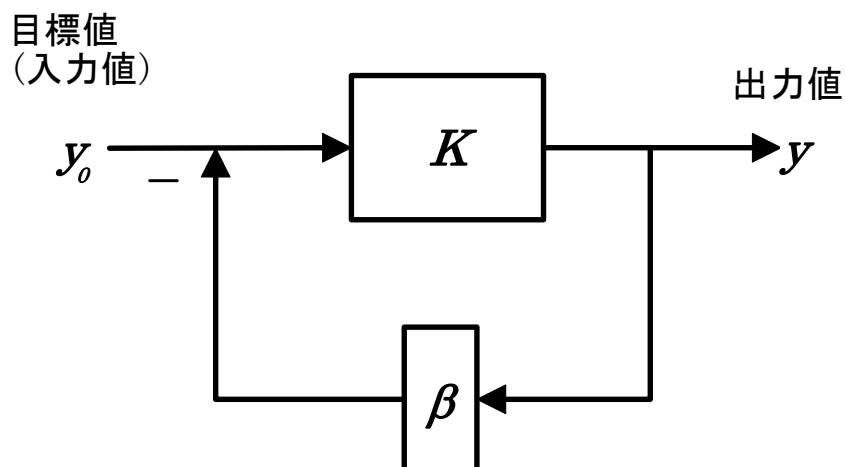


図 タンパク質 X、Y、Z のフィードフォワードループ



問 79

下図は出力値 y を目標値（入力値） y_0 に制御するフィードバックシステムである。出力値 y は β 倍されて入力値にフィードバックされる。 K はゲインである。このとき、 $K(y_0 - \beta y) = y$ の関係が成立するとする。このフィードバックシステムの記述に関して不適切なものを選択肢の中から一つ選べ。



- 1 $y = \frac{K}{1 + \beta K} y_0$ なので、 $K \gg 1$ のとき、出力値 $y \rightarrow \frac{y_0}{\beta}$ となり、出力値 y は K の値の変化に対してロバストになる。
- 2 $K \gg 1$ のとき、 β の変化に対して、出力値 y はロバストになる。
- 3 $K \gg 1$ 、 $\beta = 1$ のとき、出力値 $y \rightarrow y_0$ となる。
- 4 β が大きくなると、出力値 y は小さくなる。

問 80

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。

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