

平成 24 年度

日本バイオインフォマティクス学会 (JSBi)

バイオインフォマティクス技術者認定試験

試験問題

注意事項

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

余白

余白

余白

問 1

原核生物と真核生物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 生物は、明確な核を持たない原核生物と核を持つ真核生物に分類することができる。
- 2 原核生物には細胞膜の外側に細胞壁が存在するが、真核生物には存在しない。
- 3 原核生物は、細胞膜の脂質や遺伝子組成の違い等から真正細菌と古細菌に分類される。
- 4 真核生物の細胞は、原核生物の細胞に比べ 1,000 倍以上の体積をもつこともある。

問 2

真核生物に関する以下の記述のうち、もっとも不適切ものを選択肢の中から一つ選べ。

- 1 鞭毛は、中心に 2 本、周囲に 9 本の二連の微小管が配置する 9+2 構造をとる。
- 2 染色体は、環状 DNA とヒストンなどのタンパク質の複合体である。
- 3 真核生物のゲノムサイズは、原核生物のそれに比べて比較的大きく、タンパク質をコードしていない領域の割合も高い。
- 4 細胞小器官によって細胞内を膜で仕切り、物質や機能を局在化している。

問 3

生物の栄養的分類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 原始地球における大気の酸素濃度の増加には、好気性独立栄養生物が関わったとされる。
- 2 嫌気性従属栄養生物は発酵によってエネルギーを得る。
- 3 嫌気性独立栄養生物にはカビやきのこの仲間が含まれる。
- 4 動物は基本的に好気性従属栄養生物である。

問 4

生体膜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もっとも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 (a) であり、それ以外にステロールや糖脂質が含まれる。(a) は (b) の部分と炭化水素鎖の (c) の部分からなる極性分子であり、(b) の部分を外側に、(c) 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 (d) として球状の構造になる。

- | | | | |
|------------|---------|---------|---------|
| 1 (a) リン脂質 | (b) 親水性 | (c) 疎水性 | (d) ミセル |
| 2 (a) 糖鎖 | (b) 疎水性 | (c) 親水性 | (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 5

分泌タンパク質や膜タンパク質の輸送に関わる細胞内構造として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 粗面小胞体
- 2 ゴルジ体
- 3 ミトコンドリア
- 4 分泌小胞

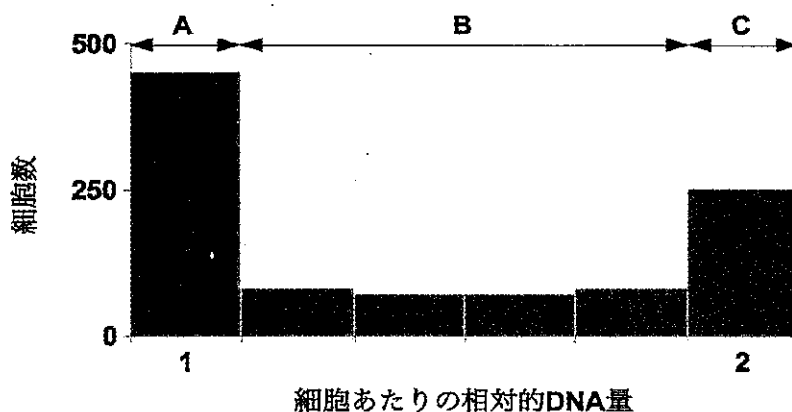
問 6

ミトコンドリアと葉緑体は、始原真核細胞に原核細胞が共生することで成立したオルガネラだと考えられている（細胞内共生説）。ミトコンドリアと葉緑体には、それぞれがかつて独立の生命体であったことを示唆するいくつかの共通した特徴が見られる。細胞内共生説を支持する、両者に共通する特徴についての記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ミトコンドリアと葉緑体は、核内 DNA とは別の独自の DNA を持つ。
- 2 ミトコンドリアと葉緑体は、細胞内で自律的に分裂して自己増殖することができる。
- 3 ミトコンドリアと葉緑体内部での転写や翻訳の仕組みは、真核細胞よりも原核細胞のそれに近い。
- 4 ミトコンドリアも葉緑体も、一重膜に囲まれている。

問 7

活発に増殖しているヒトの培養細胞集団から 1,000 細胞を採取し、各細胞の DNA 量を計測したところ、下のようなヒストグラムが得られた。もとの培養細胞集団を、細胞周期の進行を G1 期で阻害する薬剤で十分な時間処理した後、同じように 1,000 細胞採取してヒストグラムを描いた場合、最も期待されることを選択肢の中から一つ選べ。なお、細胞死などは起こらないものとする。

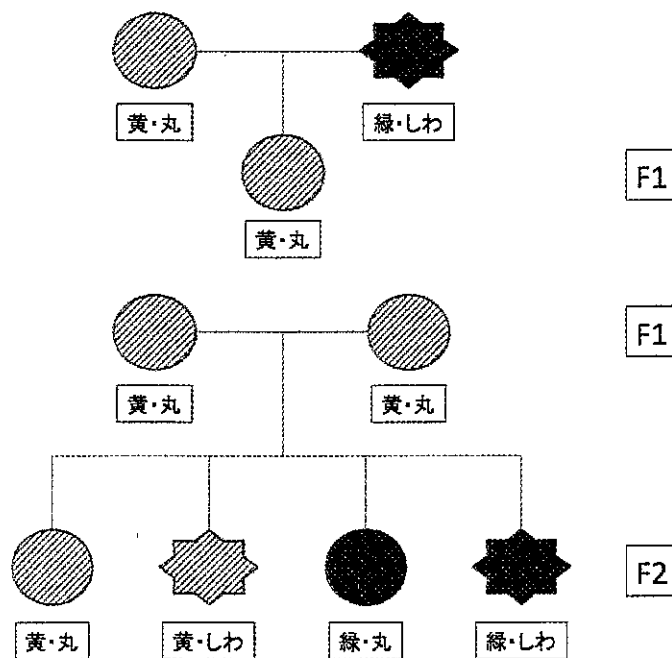


- 1 A・B 区画の細胞数が減り、C 区画の細胞数が増える。
- 2 B・C 区画の細胞数が減り、A 区画の細胞数が増える。
- 3 A 区画の細胞数が減り、B・C 区画の細胞数が増える。
- 4 B 区画の細胞数が減り、A・C 区画の細胞数が増える。

問 8

図はメンデルの二遺伝子雑種の実験を表したものである。胚乳の色が黄色で種子の形が丸いエンドウ（黄・丸）と、緑色でしわのエンドウ（緑・しわ）を交配すると、雑種第一代（F1）はすべて（黄・丸）となった。F1 どうしを交配した雑種第二代（F2）では、（黄・丸）、（黄・しわ）、（緑・丸）、（緑・しわ）の4種類のエンドウが得られた。F2 で得られる種子の数の比（黄・丸）：（黄・しわ）：（緑・丸）：（緑・しわ）として正しいものを選択肢の中から一つ選べ。

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



問 9

細胞のシグナル伝達における二次メッセンジャーとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 サイクリック AMP
- 2 イノシトール 3 リン酸
- 3 チロシン
- 4 カルシウムイオン

問 10

ウイルス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 ウイルスは、増殖する条件が揃うと分裂を繰り返しながら対数的に増加する。
- 2 ウイルスは、核酸のゲノムがカプシドと呼ばれるタンパク質の殻に覆われた基本構造を持つ。
- 3 ウイルスのうち、細菌に感染し増殖するタイプのものを特にバクテリオファージと呼ぶ。
- 4 ウイルスのうち、ゲノムとして RNA を持ち、かつ逆転写酵素を持つタイプのものを特にレトロウイルスと呼ぶ。

問 11

ヒトの免疫系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 異物を取り込み、抗原として提示する細胞にはマクロファージや樹状細胞がある。
- 2 抗体として働く免疫グロブリンは T 細胞から分化した細胞によって産生される。
- 3 自己と非自己の認識に関係しているのは、MHC(主要組織適合遺伝子複合体)分子である。
- 4 抗体分子の多様性は主に V(D)J 組換えと呼ばれる遺伝子再構成によって生じる。

問 12

DNA を構成する塩基 (A: アデニン、T: チミン、G: グアニン、C: シトシン) に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 A と G はピリミジンと呼ばれる。
- 2 核酸中の塩基は分子内に窒素を含む芳香環化合物である。
- 3 二本鎖 DNA の融解温度は塩基の組成によって変化する。
- 4 二本鎖 DNA 構造内において、通常 A は T と、G は C と水素結合する。

問 13

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

- 1 DNA 複製においては、それぞれの鋳型 DNA 鎖と同じ塩基配列の鎖が合成される。
- 2 DNA 複製はデオキシリボヌクレオチドを基質とする重合反応である。
- 3 DNA 複製は常に 2-デオキシリボースの 5' から 3' の方向に合成が進行する。
- 4 DNA 複製の主役となる酵素は DNA ポリメラーゼである。

問 14

真核生物の DNA 複製機構に関する以下の記述について (a) から (d) 内に入る酵素名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。

二重らせん構造をとる二本鎖 DNA 構造を (a) がほどこし、ほどけてできた一本鎖 DNA 上に (b) が短い RNA 断片を合成する。この RNA 断片を足がかりとし、(c) が一本鎖 DNA に相補的な DNA 鎖の合成を行う。なおこの工程は、親鎖がほどけてできる 2 つの一本鎖 DNA のうち、リーディング鎖と呼ばれる側では連続的に行われるが、ラギング鎖と呼ばれる側では不連続的に行われる。不連続的な合成でできる DNA 断片同士の結合は (d) が行う。

- 1 (a) DNA リガーゼ
(b) DNA ヘリカーゼ
(c) DNA ポリメラーゼ
(d) DNA プライマーゼ
- 2 (a) DNA ヘリカーゼ
(b) DNA プライマーゼ
(c) DNA ポリメラーゼ
(d) DNA リガーゼ
- 3 (a) DNA ヘリカーゼ
(b) DNA ポリメラーゼ
(c) DNA リガーゼ
(d) DNA プライマーゼ
- 4 (a) DNA リガーゼ
(b) DNA ポリメラーゼ
(c) DNA プライマーゼ
(d) DNA ヘリカーゼ

問 15

真核生物の mRNA のプロセッシング（修飾）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 前駆体 mRNA の 5'端にはタンパク質合成に必要なキャップ構造が付加される。
- 2 前駆体 mRNA からスプライシングによってイントロンが除去される。
- 3 前駆体 mRNA の 3'端には、数十以上の A(アデニル酸)がつながったポリ A 鎖が付加される。
- 4 プロセッシングの終了した成熟 mRNA は核内でタンパク質合成に使われる。

問 16

種々の RNA 分子に関する以下の記述のうち、もっとも不適切な記述を含むものを選択肢の中から一つ選べ。

- 1 siRNA は二本鎖、miRNA は一本鎖の RNA である。
- 2 tRNA、rRNA、miRNA はノンコーディング RNA に分類される。
- 3 siRNA や shRNA を使って、狙った遺伝子の細胞内の mRNA 量を減らすことができる。
- 4 rRNA は mRNA のコドンに対応するアミノ酸を運ぶ役割をしている。

問 17

次に示した現象の中で、エピジェネティクスと直接関わりがないものを選択肢の中から一つ選べ。

- 1 一卵性双生児でも病気のかかりやすさに違いがある。
- 2 通常、三毛猫の模様をもつのは雌であり、クローンでも同じ模様にならない。
- 3 移植、がんワクチンなどの適応に HLA（ヒト白血球抗原）型を用いる。
- 4 DNA のメチル化は遺伝子発現の抑制に関わる。

問 18

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

- 1 PCR 法には、耐熱性の DNA ポリメラーゼが利用される。
- 2 ごく微量の試料から DNA を増幅することが可能である。
- 3 二本鎖 DNA を PCR 法に使用することはできない。
- 4 プライマーは DNA 合成の開始に必要である。

問 19

以下の方法のなかで、遺伝子の組織特異的発現を確認する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 RNA *in situ* ハイブリダイゼーション法
- 2 RT-PCR 法
- 3 DNA マイクロアレイ法
- 4 サザンブロットティング法

問 20

以下の実験法のなかで、次世代シーケンサの利用法として一般的でないものを選択肢の中から一つ選べ。

- 1 メタボローム解析
- 2 メタゲノム解析
- 3 ChIP-seq 解析
- 4 ヒトゲノムリシーケンシング解析

問 21

コンピュータ上では、数値は二進数で表される。二進数同士の論理和 (OR)、論理積 (AND) は、各桁ごとにそれぞれ論理和、論理積をとったものであり、否定 (NOT) は、各桁ごとに 01 を反転させたものである。たとえば、 $1110_{(2)} \text{ OR } 0101_{(2)} = 1111_{(2)}$ 、 $1110_{(2)} \text{ AND } 0101_{(2)} = 0100_{(2)}$ 、 $\text{NOT } 1001_{(2)} = 0110_{(2)}$ である (数字の右の₍₂₎は、その数字が2進数であることを示している)。

また、左シフト($x \ll i$) は2進数 x の各桁を左に i 桁ずらしたもの (空いたビットは0にし、あふれた最左 i 桁は捨てられる)、右シフト($x \gg i$) は同様に2進数 x の各桁を右に i 桁ずらしたものである。たとえば対象とする二進数を4桁であると仮定すると、 $1010_{(2)} \ll 2 = 1000_{(2)}$ 、 $1010_{(2)} \gg 2 = 0010_{(2)}$ である。

このとき、2つの任意の同じ桁数の二進数 x, y に対して、成り立つとは限らない式を選択肢の中から一つ選べ。

- 1 $x \text{ AND } (\text{NOT } y) = \text{NOT } ((\text{NOT } x) \text{ OR } y)$
- 2 $x \text{ OR } (\text{NOT } y) = \text{NOT } ((\text{NOT } x) \text{ AND } y)$
- 3 $(x \ll 2) \text{ AND } (y \ll 2) = (x \text{ AND } y) \ll 2$
- 4 $(x \ll 2) \text{ AND } (y \gg 2) = x \text{ OR } y$

問 22

浮動小数点数による数値演算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 0.1 は2進数で表現した場合には無限小数(循環小数)となるため、浮動小数点数で表現すると丸め誤差が生じるが、これを避けることは不可能である。
- 2 数値の大きさがほぼ同じ2つの数値の差を計算すると、有効桁数の減少が起こることがある。
- 3 絶対値が大きい値に対し、絶対値が非常に小さい値を加算すると情報落ちが起きることがある。
- 4 無限大を表現するために NaN (Not a Number) と呼ばれる特別な浮動小数点数が用いられる。

問 23

コンピュータに周辺機器を接続するために用いるインターフェース規格として不適切なものを選択肢の中から一つ選べ。

- 1 HiSeq 2000
- 2 USB 3.0
- 3 シリアル ATA
- 4 IEEE 1394

問 24

メモリが 2GB、ハードディスクの容量が 500GB の計算機がある。この計算機内のハードディスクから、プログラムを用いて 5GB のデータを読み込んでいる時、突然データの読み込みが遅くなった。この原因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 他のユーザのプログラムがハードディスクに対して読み込み・書き込みをし始めた。
- 2 ディスプレイに障害が発生した。
- 3 ハードディスクに障害が発生した。
- 4 メモリスワップが発生した。

問 25

(a) から (d) 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。

次世代シーケンサから得られる配列情報は、計算機のメモリより (a) 場合が一般的である。この配列情報をソートするために外部 (エクスターナル) ソートが利用される。外部ソートでは、データを計算機のメモリより (b) 単位に (c) した上で、データをソートした後、これらを (d) することで限られたメモリでもソートできるように工夫する。

- | | | | | |
|---|---------|---------|--------|--------|
| 1 | (a) 小さい | (b) 大きい | (c) 併合 | (d) 分割 |
| 2 | (a) 小さい | (b) 小さい | (c) 分割 | (d) 併合 |
| 3 | (a) 大きい | (b) 小さい | (c) 分割 | (d) 併合 |
| 4 | (a) 大きい | (b) 大きい | (c) 併合 | (d) 分割 |

問 26

スタックは、コンピュータ上でよく用いられるデータ構造の一つで、最後に入ったデータが最初に出る、後入れ先出し(Last In First Out: LIFO)が特徴である。ここで、十分な大きさのスタックがあるとし、PUSH 命令はスタックにデータを格納する操作、POP 命令は、データを取り出す操作とする。下記のように 9 回の命令を実行した時、最後の POP 命令によって出力されるデータは何か。適切なものを選択肢の中から一つ選べ。

- (1) PUSH a
- (2) PUSH b
- (3) PUSH c
- (4) PUSH b
- (5) POP
- (6) POP
- (7) PUSH b
- (8) POP
- (9) POP

- 1 a
- 2 b
- 3 c
- 4 空なので何も得られない。

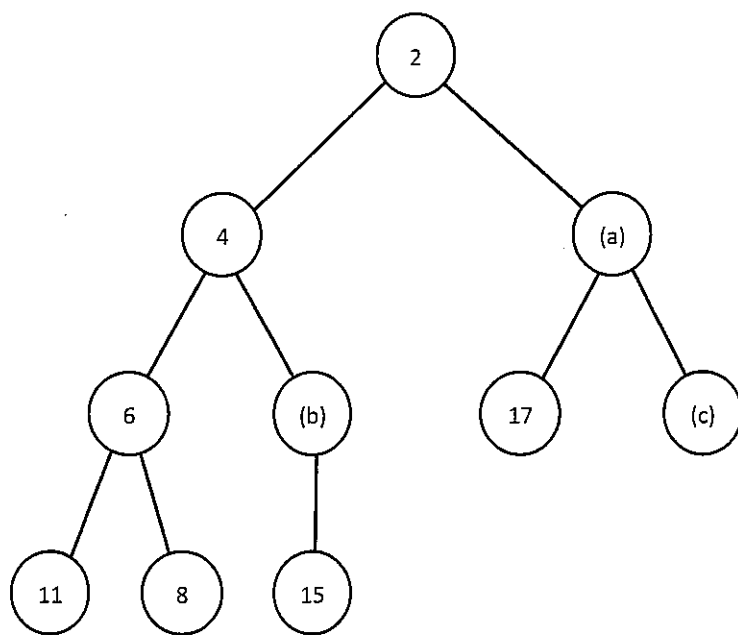
問 27

以下の記述の中から、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 任意の数列を $O(n \log n)$ でソートすることが可能なアルゴリズムで知られているものはクイックソートのみである。
- 2 長さがそれぞれ n と m の 2 つのソート済みの数列が与えられた時、双方の数列に含まれるすべての数をソートした数列は $O(n+m)$ で計算することができる。
- 3 任意の長さ n の数列の中央値（数列中で $\lceil n/2 \rceil$ 番目に大きな値）は $O(n)$ で求めることができる。
- 4 1 から 10 までの整数だけからなる長さ n の数列は、バケットソートを用いて $O(n)$ でソートすることが可能である。

問 28

数列 (16, 11, 6, 14, 2, 9, 15, 17, 4, 18) を入力として適当な方法でヒープを作成したところ、以下の図のようになった。この図の中で (a), (b), (c) に入る数字の組み合わせとして正しいものを選択肢の中から一つ選べ。



- 1 (a) 14 (b) 16 (c) 18
- 2 (a) 16 (b) 14 (c) 18
- 3 (a) 18 (b) 16 (c) 14
- 4 (a) 18 (b) 14 (c) 16

問 29

二分探索法に関する以下の記述について、(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。

一定の順序に並べられたデータ n 個が与えられたとき、二分探索法を用いてそこに含まれる目的のデータを見つけ出すのに必要な二分操作の回数は高々 (a) 回で済む。

例えばデータの列：「1 3 4 6 7 9 10」からデータ (b) の位置を探すには (c) 回の二分操作が行われる。

- | | | | |
|---|-------------------|-------|-------|
| 1 | (a) $O(n \log n)$ | (b) 7 | (c) 2 |
| 2 | (a) $O(\log n)$ | (b) 4 | (c) 2 |
| 3 | (a) $O(\sqrt{n})$ | (b) 9 | (c) 3 |
| 4 | (a) $O(\log n)$ | (b) 6 | (c) 2 |

問 30

次世代シーケンサから得られる配列をゲノム配列と対応付ける手法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 Burrows-Wheeler 変換法
- 2 ハッシュ法
- 3 動的計画法
- 4 Newton-Raphson 法

問 31

ある中学校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として、もっとも適切なものを選択肢の中から一つ選べ。

```
SELECT 学生番号 FROM 学籍簿
WHERE 組 >= 3 AND (評価点 > 50 OR 出席日数 > 150);
```

学籍簿

学生番号	名前	評価点	出席日数	組
2401	伊藤 かおり	85	261	1
2402	大隈 京子	36	132	3
2403	桂 正	54	240	4
2404	原 新一	21	146	1
2405	吉田 太郎	75	125	3

- 1 2403
- 2 2402, 2404
- 3 2403, 2405
- 4 2401, 2402, 2405

問 32

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以下の説明のうち、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データとして扱う。
- 2 データを 2 次元の表(テーブル)として表現する。
- 3 データの操作言語として、SQL が知られている。
- 4 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。

問 33

正規分布に関する説明文として、正しいものを選択肢の中から一つ選べ。

- 1 正規分布の場合、確率密度関数の値も累積密度関数の値も 1 を超えることはない。
- 2 正規分布の場合、確率密度関数の値も累積密度関数の値も 1 を超えることがある。
- 3 正規分布の確率密度関数の値は 1 を超えることがあるが、累積密度関数の値は 1 を超えることはない。
- 4 正規分布の確率密度関数の値は 1 を超えることはないが、累積密度関数の値は 1 を超えることがある。

問 34

次に示した説明文の中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 カルバック・ライブラー情報量は相対エントロピーとも呼ばれる。
- 2 尤度関数を最大にするようにパラメータを決めることを MAP 推定という。
- 3 コイン投げの試行をモデル化するとき、0 か 1 の値を許すベルヌーイ分布がしばしば用いられる。ベルヌーイ分布の確率変数は平均も分散も常に 0 以上 1 以下の値をとる。
- 4 t 分布においても正規分布においても、いかなる実現値でも、確率密度は 0 より大きい。

問 35

3 つの事象 A, B, C に対して、 C の条件下で A と B が同時に生起する確率 $P(A, B | C)$ の満たす等式として適切なものを選択肢の中から一つ選べ。

- 1 $P(A, B | C) = P(A | C) P(B | A, C)$
- 2 $P(A, B | C) = P(A | C) P(B, C)$
- 3 $P(A, B | C) = P(C | A, B) P(A, B)$
- 4 $P(A, B | C) = P(A | C) P(B | C) P(C)$

問 36

統計学的仮説検定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。以下では α をあらかじめ定めた有意水準とする。

- 1 観測したデータ D に対して、 $P(D|H_0)$ を帰無仮説 H_0 の有意確率 (p -値) とすると、 $P(D|H_0) < \alpha$ が成り立つとき、帰無仮説を棄却する。
- 2 帰無仮説が棄却された場合には、対立仮説が採択される。
- 3 帰無仮説が棄却されない場合は、帰無仮説が採択される。
- 4 複数のデータに対して検定を行う場合に、検定が 1 つの場合と同じ有意水準を用いると偽陽性が増大する。

問 37

次世代シーケンサを用いて遺伝子発現量を調べる方法に RNA-seq 法がある。RNA-seq 法では、ある遺伝子の発現がサンプル間で異なるかを調べる際に、各遺伝子の発現を箱からボールを抽出する事象としてモデル化する。まず、細胞内で発現している全遺伝子の数と長さに比例した数のボールを箱の中に入れる。次に、ボールの内、調査対象の遺伝子に相当するものを赤に、それ以外を青に塗る。RNA-seq で 1 本のリードを観測する事は、ボールを 1 個取り出すことに相当する。このため、箱のなかからボールを RNA-seq のリード数と同数取り出し、その中に含まれる赤のボールの個数を表す確率分布が、遺伝子発現を表す。この確率分布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 ポアソン分布
- 2 二項分布
- 3 負の二項分布
- 4 一様分布

問 38

(a) から (d) 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。

(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 39

陽性か陰性かの二値分類の予測を行うとき、その予測精度を表す指標として真陽性(True Positive)、真陰性(True Negative)、偽陽性(False Positive)、偽陰性(False Negative)がある。(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。

真陽性とは陽性と正しく予測された標本、(a) とは陽性と誤って予測された標本、(b) とは陰性と正しく予測された標本、(c) とは陰性と誤って予測された標本のことをいう。感度(Sensitivity)は 1 から (c) 率を引いた値に等しく、特異度(Specificity)は 1 から (a) 率を引いた値に等しい。

- 1 (a) 真陰性 (b) 偽陽性 (c) 偽陰性
- 2 (a) 偽陽性 (b) 真陰性 (c) 偽陰性
- 3 (a) 偽陽性 (b) 偽陰性 (c) 真陰性
- 4 (a) 真陰性 (b) 偽陰性 (c) 偽陽性

問 40

(a) と (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はどれか。選択肢の中から一つ選べ。

典型的なサポートベクトルマシン(SVM)の学習アルゴリズムは (a) で表されており、カーネル法の一つとして知られている。カーネルとして線形カーネルを用いると識別面は常に (b) になる。

- | | |
|------------------|---------|
| 1 (a) 組み合わせ最適化問題 | (b) 超曲面 |
| 2 (a) 整数計画問題 | (b) 超円錐 |
| 3 (a) 二次計画問題 | (b) 超平面 |
| 4 (a) 凸最適化問題 | (b) 超球 |

問 41

2本の塩基配列をニードルマン・ Wunsch (Needleman-Wunsch) アルゴリズムでアラインメントを行ったところ、下図(A)のようなアラインメントが得られた。ただし、アラインメントのスコアは、塩基の一致に対する加点 (m)、不一致に対する減点 (u)、およびギャップペナルティ (g) の3つのパラメータで定義されるものとする。このとき、(m, u, g)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。

(ヒント) 下図(B)の非最適アラインメントと比較せよ。

(A)	(B)
GGAT-GCCACCT	GGATGCCACCT
GGATTGC-ACCT	GGATTGCACCT

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

問 42

生物種 A～E の遺伝子配列から、下表のような距離行列が得られた（対称行列なので上半分に影をつけてある）。この表をもとに非加重結合法（UPGMA）を用いて作成した系統樹のニューウィック形式（Newick format）での表現として適切なものを選択肢の中から一つ選べ。

	A	B	C	D	E
A	0	0.6	0.2	0.6	0.4
B	0.6	0	0.6	0.4	0.6
C	0.2	0.6	0	0.6	0.4
D	0.6	0.4	0.6	0	0.6
E	0.4	0.6	0.4	0.6	0

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

問 43

下の表は、ある微生物ゲノムにおけるダイマー（2連塩基）の出現頻度である。このとき、条件付き確率 $P(G^2|T^1)$ の値として適切なものを選択肢の中から一つ選べ。ここで、 T^1 および G^2 はダイマーの1塩基目、2塩基目がそれぞれ T および G であることを表す。

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

- 1 0.07
- 2 0.20
- 3 0.24
- 4 0.28

問 44

アラインメント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 BLASTX は塩基配列を問い合わせに、アミノ酸配列データベースを検索するプログラムである。
- 2 アラインメントされた配列間には相同性がある。
- 3 スミス・ウォーターマン (Smith-Waterman) 法は動的計画法を用いて二つの配列間で局所的にもっとも類似した領域を見出す方法である。
- 4 ClustalW は累進法を実装した多重アラインメントプログラムである。

問 45

以下は、アミノ酸配列のアラインメントを基に定義されたプロファイルを示している。これは、アラインメント中の各位置でのアミノ酸の出現頻度をそのままスコアとしたもので、出現頻度が 0 のセルは空白で表されている。また、N 末端からの位置が「残基番号」の列に指定されている。このプロファイルに、ギャップなしのアラインメントで、もっともよく合致するアミノ酸配列を選択肢の中から一つ選べ。

残基番号	アミノ酸のスコア																			
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
5	1									2								13		
6			16																	
7					16															
8																7	1		5	3
9	16																			
10			1	4									2			1	7	1		

- 1 VDFSATVDFT
- 2 VDFSDFVDFT
- 3 ADFVDFVSAT
- 4 VDFSVDVSAT

問 46

遺伝子の水平伝搬に関する以下の記述において、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 遺伝子の水平伝搬には、DNA を直接細胞に取り込む形質転換、ファージを介して伝達する形質導入、細胞間で表面の一部を結合して伝達する接合などのメカニズムがある。
- 2 真核生物の核ゲノムには、ミトコンドリアや葉緑体ゲノムから移動した遺伝子が含まれることがあり、それらは一般に真正細菌の遺伝子と比較的高い類似性を示す。
- 3 遺伝子の水平伝搬は GC 含量の近い生物種間でのみ起こるため、GC 含量の違いを手がかりに水平伝搬の痕跡を探すことは極めて難しい。
- 4 遺伝子の水平伝搬がある場合、遺伝子の分子系統が生物種の系統と大きく異なる可能性があることに注意しなければならない。

問 47

以下に示すデータはGenBank形式で記されたRefSeqのエントリーの一部を改変したものである。このデータ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、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 18-JUN-2012
SOURCE     Escherichia coli ATCC 8739
FEATURES             Location/Qualifiers
     source          1..2031
                     /organism="Escherichia coli ATCC 8739"
                     /mol_type="genomic DNA"
     CDS             complement(1..2031)
                     /gene="maltS"
                     /locus_tag="EcolC_0143"
                     /note="periplasmic enzyme that degrades maltodextrins that
                     enter via the outer membrane porin LamB; hydrolyzes
                     alpha-(1,4) glycosidic linkages; contains internal
                     disulfide bonds"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic alpha-amylase"
                     /translation="MKLAACFLTLLPGFAVAASWTSPGFPAFSEQGTGTFVSHAQLPK
                     ***** 略 *****
                     GTRPLTLNFDQQC"
     sig_peptide     complement(1978..2031)
                     /gene="maltS"
                     /locus_tag="EcolC_0143"
                     /note="Signal predicted by SignalP 3.0 HMM (Signal peptide
                     probability 1.000) with cleavage site probability 0.723 at
                     residue 18"

ORIGIN
      1 ttactgttgc cctgccaga cgaccagcac ttatogtgc ccatgctcac gaacaaagcc
     61 gtagccctgc ttcagcaaaa gtgtgtttg ttgcccgcg ccaattggg gatggcgggc
        ***** 略 *****
    1981 ggcaacggcg aagccaggaa ggagtgtcag aaacaggcg gcgagtttca t
//

```

- 1 2031 塩基からなる大腸菌のある遺伝子の全長である。
- 2 遺伝子産物がアミラーゼ活性をもつ。
- 3 cDNA から塩基配列が解読されている。
- 4 シグナルペプチドをもつタンパク質をコードすると予測されている。

問 48

BLAST では類似性スコアが bit を単位として表されており、このとき E-value(E)はスコア (S)とクエリ配列の長さ(m)、データベースの大きさ(n)を用いて $E=mn2^{-S}$ と表される。いま、あるクエリ配列を、あるデータベースに対して検索したときのスコアが 48 で、E-value が 0.0001 であったとする。同じクエリ配列を 8 倍の大きさのデータベースに対して検索したとき、E-value が同じ値 0.0001 になるのに必要なスコアはいくらになるか。選択肢の中から一つ選べ。

- 1 45
- 2 51
- 3 96
- 4 384

問 49

ゲノム中には様々なタイプの繰り返し配列が存在しており、それらを検出する方法もいろいろある。以下の解析のうち、ゲノム配列中の繰り返し配列を検出するための手段として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 GC skew 解析(ゲノム DNA の一方の鎖での G と C の含量の違いに着目した解析)
- 2 ゲノム配列中出现する固定長の単語の出現頻度統計
- 3 自分自身の配列に対する相同性検索
- 4 トランスポゾンなど既知の散在性反復配列を集めたデータベースに対する相同性検索

問 50

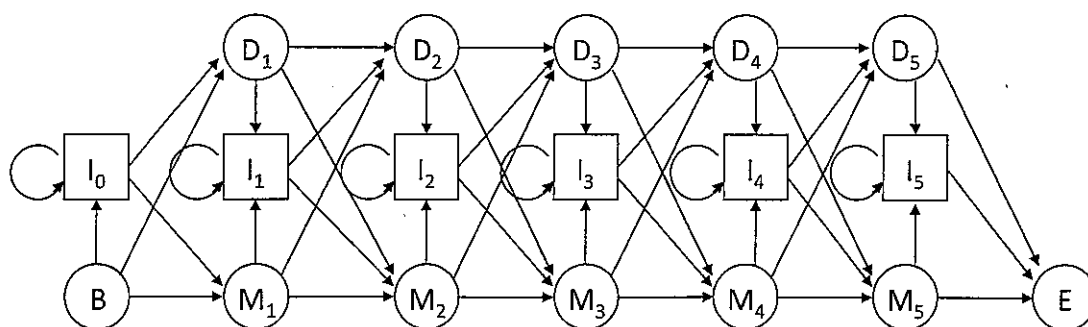
配列アラインメントに関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 比較する配列間で共通の相同ドメインが存在するかどうかを確認したい場合は、ローカルアラインメントを用いる。
- 2 ローカルアラインメントは、一般にグローバルアラインメントのアルゴリズムで、両端のギャップペナルティを 0 に設定することにより計算できる。
- 3 アラインメントのギャップペナルティは、しばしば 2 つのパラメータ a, b を用いて、ギャップの長さが L のとき $aL+b$ となるように計算される。このとき、 a を b に対して絶対値の小さな値に設定すると、ギャップがまとまって入りやすくなる。
- 4 アラインメントのスコアが一致に対する加点、不一致に対する減点、および上記の a, b で定義されるギャップペナルティの 4 種類のパラメータによって計算されるとき、各パラメータを一律に 2 倍にして計算した最適アラインメントは、元のアラインメントと全く同じになる。

問 51

以下の図(A)は、アミノ酸配列のマルチプルアラインメントの計算に用いられる隠れマルコフモデルを表している。ここでBとEはそれぞれ開始と終了、 M_n , I_n , D_n はそれぞれマッチ、挿入、欠失の状態に対応しており、マッチと挿入の各状態からは 20 種類のアミノ酸が一定の確率で出力されている。また、表(B) はアミノ酸配列をこのモデルに当てはめることにより生成されたアラインメントを示しており、大文字と小文字のアミノ酸は、それぞれマッチ状態および挿入状態から出力されたアミノ酸を、「.」はギャップを表している。各配列は、モデルに当てはめたときの最適パスに応じてアラインメントされており、Seq1, Seq2, Seq3 に対応するパスが表(B)右端のカラムに示されている。このとき、Seq4 に対応するパスとしてもっとも適切なものを選択肢の中から一つ選べ。

図(A)



表(B)

	アラインメント	パス
Seq1	MYL...PK	B-M ₁ -M ₂ -M ₃ -M ₄ -M ₅ -E
Seq2	VLV...SK	B-M ₁ -M ₂ -M ₃ -M ₄ -M ₅ -E
Seq3	MYFpadPK	B-M ₁ -M ₂ -M ₃ -I ₃ -I ₃ -I ₃ -M ₄ -M ₅ -E
Seq4	L.Ldk.SK	?

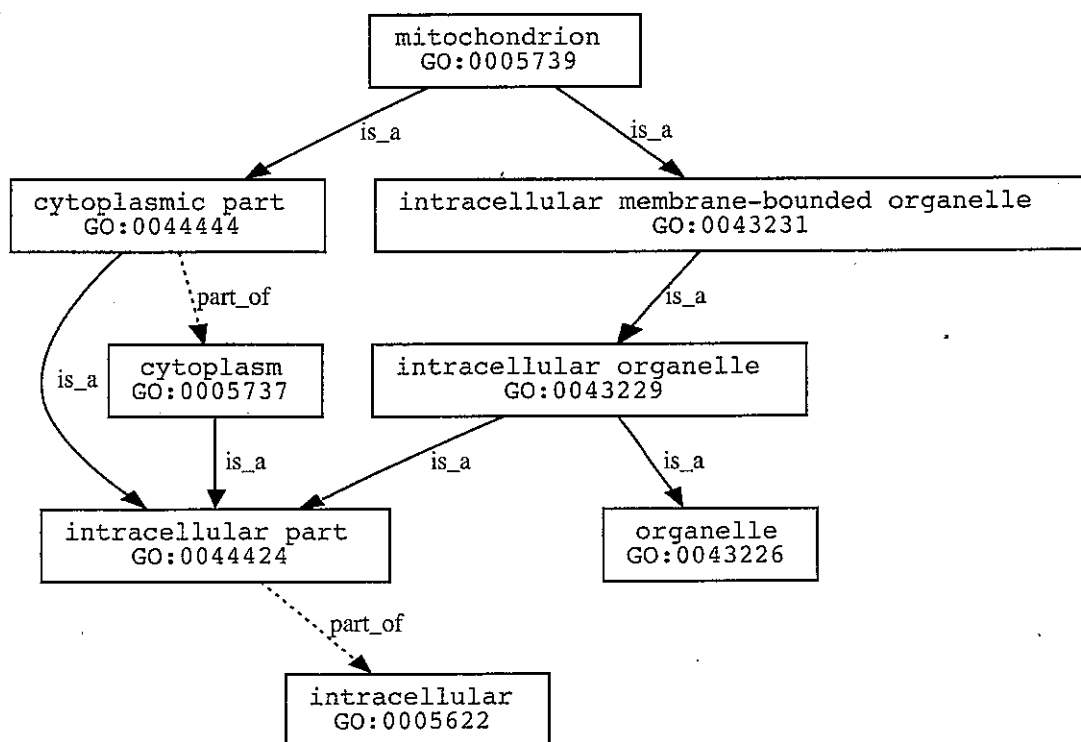
- 1 B-M₁-I₁-M₂-M₃-D₄-M₅-E
- 2 B-M₁-D₂-M₃-I₃-I₃-M₄-M₅-E
- 3 B-M₁-I₁-M₂-M₃-I₃-D₄-I₄-M₅-E
- 4 B-M₁-D₂-M₂-I₂-I₃-D₃-M₄-M₅-E

問 52

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、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 "organelle" is_a "intracellular part"
- 4 "intracellular membrane-bounded organelle" part_of "intracellular"

Copyright © 2012 Japanese Society for Bioinformatics. All Rights Reserved.

問 53

ある触媒機能が保存された、細胞外分泌酵素ファミリーのアミノ酸配列のマルチプルアラインメントの一部を以下の図に示した。このアラインメントをもとに、部位 1、部位 2、...、部位 7 の立体構造上の特徴や機能的な役割について推察を行った(部位はアラインメント最上段に数字で示した)。この推察としてもっとも不適切な記述を選択肢の中から一つ選べ。

1	2	3	4	5	6	7
-YHFCGGSLINSQWVVSAAH	CYKSGIQVRLG-EDNINVV					
-YHFCGGSLINDQWVVSAAH	CYQYHIQVRLG-EYNIDVL					
-YHFCGGSLINDQWVVSAAH	CYSRIQVRLG-EHNINVL					
-SHICGGSLITDQWVLSAAH	CYHPQLQVRLG-EHNIYEI					
TSLRCGGVLIDHRWVLTAAH	CSGSRYWVRLG-EHSLSQL					
LSFHCAGVLVDQSWVLTAAH	CGNKPLWARVG-DDHLLLL					
QQLLCGGVLVGGNWVLTAAH	CKKPKYTVRLG-DHSLQNK					
NQLHCGGVLVNERWVLTAAH	CKMNEYTVHLG-SDTLG--					
RRFLCGGALLSGQWVITA	AHCGRPILQVALG-KHNLRRW					
GFHFCGGSLINENWVVTA	AHCGVTTSDV VVAGEFDQGSS					

- 1 部位 1 と部位 5 では、部位 5 のほうが活性部位である可能性が高い。
- 2 部位 4 と部位 5 では、部位 4 のほうが活性部位である可能性が高い。
- 3 部位 2 と部位 6 は、ジスルフィド結合を形成している可能性がある。
- 4 部位 3 と部位 7 では、部位 7 のほうがタンパク質表面に露出している可能性が高い。

問 54

下図は PDB データベースのあるエントリーの一部である。この PDB エントリーについて、もっとも不適切な記述を選択肢の中から一つ選べ。

```

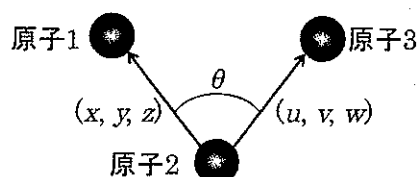
HEADER      PROTO-ONCOGENE                                27-JAN-97   1AA9
TITLE       HUMAN C-HA-RAS(1-171) (DOT)GDP, NMR, MINIMIZED AVERAGE
TITLE       2 STRUCTURE
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: C-HA-RAS;
COMPND      3 CHAIN: A;
COMPND      4 FRAGMENT: RESIDUES 1 - 171;
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE      3 ORGANISM_COMMON: HUMAN;
SOURCE      4 ORGANISM_TAXID: 9606;
SOURCE      5 GENE: HUMAN C-HA-RAS GENE;
EXPDTA      SOLUTION NMR
AUTHOR      Y. ITO, Y. YAMASAKI, Y. MUTO, G. KAWAI, S. NISHIMURA, T. MIYAZAWA,
AUTHOR      2 S. YOKOYAMA, RIKEN STRUCTURAL GENOMICS/PROTEOMICS INITIATIVE
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 O LEU A 53 N LYS A 42
SHEET 3 S1 6 THR A 2 VAL A 9 1 O LEU A 6 N LEU A 56
SHEET 4 S1 6 GLY A 77 ILE A 84 1 O LEU A 79 N VAL A 9
SHEET 5 S1 6 PRO A 110 LYS A 117 1 O VAL A 112 N CYS A 80
SHEET 6 S1 6 PRO A 140 THR A 144 1 O 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 HE1 MET A 1 130.217 16.701 -13.847 1.00 2.88 H
ATOM 18 HE2 MET A 1 129.544 18.142 -13.072 1.00 2.81 H
ATOM 19 HE3 MET A 1 131.290 17.937 -13.192 1.00 2.81 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 これは核磁気共鳴法で構造決定されたタンパク質である。
- 2 このエントリーの ID は 1AA9 である。
- 3 このタンパク質の GLU126 から TYR137 までのアミノ酸残基は、 α ヘリックス構造をとっている。
- 4 このタンパク質の MET1 の C α 原子の z 軸座標値は、THR2 の C α 原子の z 軸座標値に比べて小さい。

問 55

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



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

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

3 $\theta = \cos^{-1} \left(\frac{x \cdot u + y \cdot v + z \cdot w}{\sqrt{(x-u)^2 + (y-v)^2 + (z-w)^2}} \right)$

4 $\theta = \sin^{-1} \left(\frac{x \cdot u + y \cdot v + z \cdot w}{\sqrt{x^2 + y^2 + z^2} \cdot \sqrt{u^2 + v^2 + w^2}} \right)$

問 56

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通りであった。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 1.5 Å

2 1.6 Å

3 2.0 Å

4 5.0 Å

問 57

タンパク質構造解析に関する用語の相互に関連の深い組み合わせとして、もっとも不適切なものを選択肢の中から一つ選べ。

1 立体構造 – PDB

2 超二次構造 – β α β

3 構造分類 – CATH

4 分子動力学計算 – GEO

問 58

分子力学法と分子動力学法に関して、もっとも不適切な記述を選択肢の中から一つ選べ。

- 1 分子力学法は、ポテンシャルエネルギーが極小となる構造を探索するために用いられる。
- 2 分子動力学法を用いることで、分子の運動の軌跡を計算し、熱力学的アンサンブルを生成することができる。
- 3 一つの原子を表現するために、分子力学法では一つの中心点だけを用いるが、分子動力学法ではいくつかの原子軌道の重ね合わせを用いる。
- 4 生体高分子の分子力学法や分子動力学法では、ポテンシャルエネルギー関数として AMBER と CHARMM がよく用いられている。

問 59

タンパク質の分子間結合部位の特徴について、もっとも不適切な記述を選択肢の中から一つ選べ。

- 1 核酸分子は負に帯電しているため、タンパク質の核酸結合部位には正の電荷を持つグルタミン酸、アスパラギン酸が多く見られる。
- 2 低分子との結合部位は、結合分子を覆い囲むようなポケット型の形状となっていることが多い。
- 3 核酸やタンパク質など高分子との結合部位は、ポケット型の形状ではないことも多い。
- 4 他の分子が結合することで、結合部位周辺の立体構造が、開いた形から閉じた形へ変化することがある。

問 60

タンパク質のアミノ酸配列の一致度、立体構造の違い、および相同性についてもっとも不適切な記述を選択肢の中から一つ選べ。

- 1 アミノ酸配列の一致度が 30%以上なら、多くの場合、 $C\alpha$ に関する RMSD（立体構造の違い）は数 Å 以下である。
- 2 アミノ酸配列の一致度が 100%であっても、実験条件の違いなどによって、立体構造が多少異なる場合がある。
- 3 立体構造の違いが数 Å 以下であっても、アミノ酸配列の一致度が 30%未満であれば、相同である可能性はない。
- 4 相同で機能の保存されたタンパク質を立体構造でアラインメントした場合、アミノ酸配列の一致度が 30%以下であっても、機能部位の位置は対応し、そのアミノ酸種は同一であることが多い。

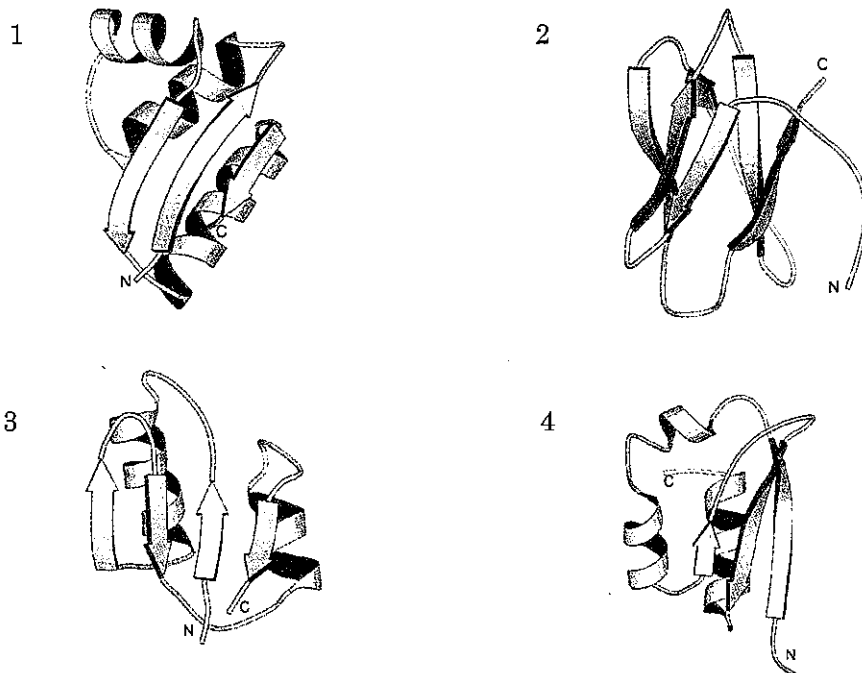
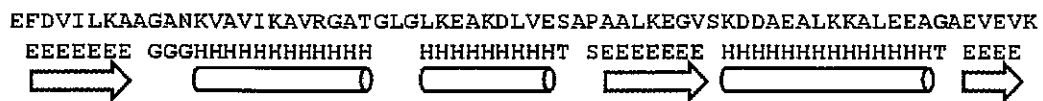
問 61

タンパク質立体構造の SCOP 分類において、全タンパク質をスーパーファミリー、ファミリー、フォールドに分類したとき、それぞれの総数の大小関係をもっとも適切に示す式を選択肢の中から一つ選べ。

- 1 ファミリー数 $\geq$ フォールド数 $\geq$ スーパーファミリー数
- 2 ファミリー数 $\geq$ スーパーファミリー数 $\geq$ フォールド数
- 3 フォールド数 $\geq$ スーパーファミリー数 $\geq$ ファミリー数
- 4 スーパーファミリー数 $\geq$ ファミリー数 $\geq$ フォールド数

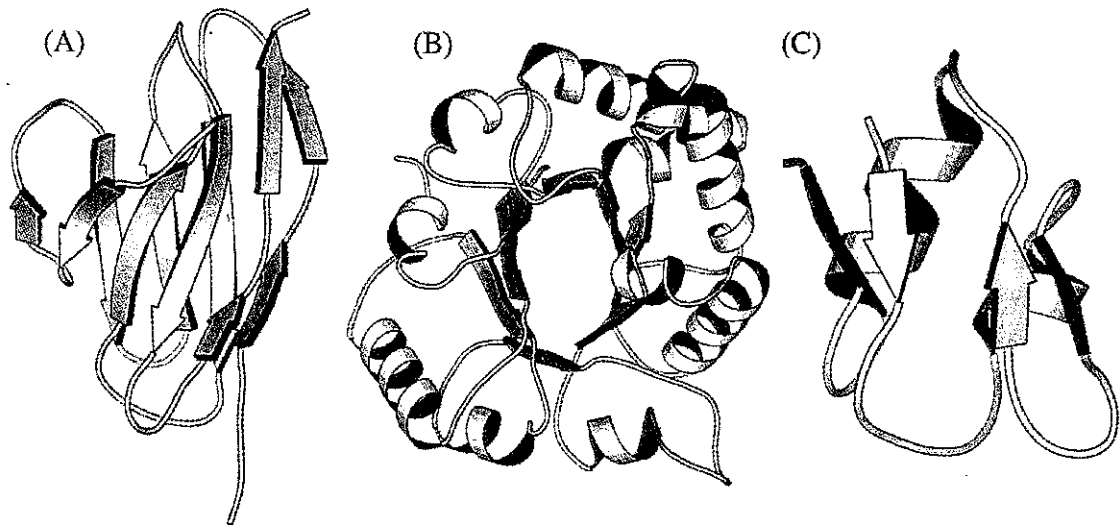
問 62

あるタンパク質のアミノ酸配列とその二次構造を下図に示す。記号 H やシリンダーは α ヘリックス、記号 E や矢印は β ストランドを表している。このタンパク質の立体構造の概形としてもっとも適切なりボン模型の図を、選択肢の中から一つ選べ。ただし、N,C はそれぞれ N 末端、C 末端を示す。



問 63

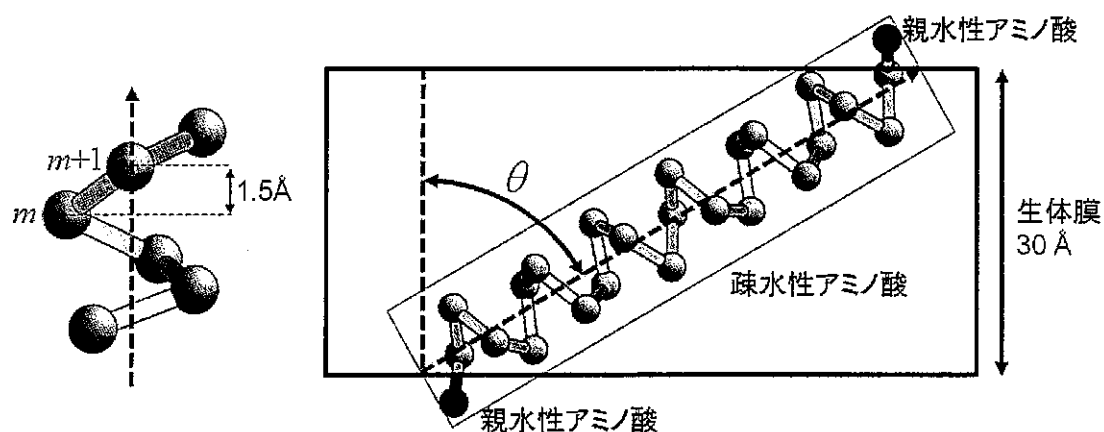
以下の3つのリボン模型で表されたタンパク質 (A), (B), (C) のフォールド名の正しい組み合わせを選択肢の中から一つ選べ。



- 1 (A) 免疫グロブリンフォールド
(B) ロスマンフォールド
(C) フェレドキシンフォールド
- 2 (A) フェレドキシンフォールド
(B) ロスマンフォールド
(C) 免疫グロブリンフォールド
- 3 (A) 免疫グロブリンフォールド
(B) TIM バレルフォールド
(C) フェレドキシンフォールド
- 4 (A) フェレドキシンフォールド
(B) TIM バレルフォールド
(C) 免疫グロブリンフォールド

問 64

膜貫通ヘリックスの長さから、膜に対する傾斜角を求めたい。 $30\text{\AA}$ の一定の厚さを持つ生体膜を、40 残基の長さの膜貫通ヘリックス 1 本が貫通しているとする。この膜貫通ヘリックスは、主に疎水性のアミノ酸から構成されるが、その両端には親水性アミノ酸を持つ。疎水性のヘリックス領域は完全に膜内に埋め込まれ、両端の親水性アミノ酸は絶対に膜内に入れないとしたとき（下図の右）、膜貫通ヘリックスの長軸の、膜に垂直な軸に対する傾斜角 θ として、もっとも適切な値を選択肢の中から一つ選べ。なお、膜貫通ヘリックス内の隣りあう残基（ m 番目の残基と $m+1$ 番目の残基）の位置は、ヘリックスの長軸上で $1.5\text{\AA}$ 異なるとして計算せよ（下図の左）。



- 1 20°
- 2 30°
- 3 45°
- 4 60°

問 65

ゲノムワイド関連解析 (genome-wide association study: GWAS) において、メタ・アナリシスを実施されることが増えている。メタ・アナリシスについての説明としてもっとも不適切なものを選択肢の中から一つ選べ。

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

問 66

連鎖する二つの遺伝子座 A, B を考える。A には a1 と a2 の 2 種類のアレルが、B にも b1 と b2 の 2 種類のアレルがあるとする。連鎖不平衡とは、この二つの遺伝子座のアレルの特定の組み合わせ（ハプロタイプ、例えば a1・b2）のある集団における頻度が有意に高く、その集団の祖先の持っていたハプロタイプが保持されている状態を意味する。

連鎖不平衡の指標として、よく用いられる r^2 は次のように定義される。

$$\frac{(p_{1-1}p_{2-2} - p_{1-2}p_{2-1})^2}{p_{1*}p_{2*}p_{*1}p_{*2}}$$

式中、 p_{ij} は、ハプロタイプ $a_i \cdot b_j$ の頻度、 p_{i*} , p_{*j} はそれぞれ a_i , b_j の頻度を表す。連鎖不平衡と r^2 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 連鎖平衡のとき、 $r^2 = 0$ である。
- 2 連鎖平衡であるならば、ハプロタイプ a1・b1 の頻度 p_{1-1} は、 p_{1*} と p_{*1} との積である。
- 3 a1・b1, a1・b2, a2・b1 の 3 種類のハプロタイプのみが存在する集団を考える。このとき、 r^2 は次の式で表される。

$$\frac{p_{1-2}p_{2-1}}{(p_{1-1} + p_{1-2})(p_{1-1} + p_{2-1})}$$

- 4 a1・b1 のハプロタイプのみを持つ集団と、a2・b2 のハプロタイプのみを持つ集団が混じり合うと、その集団には、a1・b1, a2・b2 の 2 種類のハプロタイプのみが存在する。遺伝子座 A と B の間に組換えが起きなければ、2 種類のハプロタイプのみが存在する状態が維持される。このとき、a1 アレルの頻度と b1 アレルの頻度は等しく、 $r^2 = 1$ である。

問 67

2 つ遺伝子座 X, Y それぞれに一塩基多型(SNP)があり、一方の遺伝子座 X は A または G, 他方の遺伝子座 Y は C または T のアレルを持つとする。

それぞれのアレル頻度を p_A^* , p_G^* , p_C^* , p_T^* とする。また、この 2 つの SNP は、ある染色体上で相互に近い座位にあり、4 つのハプロタイプ A-C, A-T, G-C, G-T を構成する。この 4 ハプロタイプの頻度を p_{A-C} , p_{A-T} , p_{G-C} , p_{G-T} とする。また、連鎖不平衡の指標としてよく用いられる D' は次のように定義される。

$$\frac{|p_{A-C}p_{G-T} - p_{A-T}p_{G-C}|}{\min(p_A^*p_T^*, p_C^*p_G^*)} \quad p_{A-C}p_{G-T} - p_{A-T}p_{G-C} > 0 \quad \text{の場合}$$

$$\frac{|p_{A-C}p_{G-T} - p_{A-T}p_{G-C}|}{\min(p_A^*p_C^*, p_T^*p_G^*)} \quad p_{A-C}p_{G-T} - p_{A-T}p_{G-C} < 0 \quad \text{の場合}$$

この二つの遺伝子座の連鎖不平衡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 $p_A^*=0.3$, $p_C^*=0.6$ で、連鎖平衡を仮定したときの p_{A-T} の値は 0.12 である。
- 2 $p_A^*=0.3$, $p_C^*=0.7$, $p_{A-T}=0.3$ のとき、連鎖不平衡係数 D' は 1 である。
- 3 $p_A^*=0.3$, $p_C^*=0.6$, $p_{A-T}=0.2$ のとき、連鎖不平衡係数 D' は 1 である。
- 4 $p_A^*=0.3$, $p_C^*=0.6$, $p_{A-T}=0.3$ のとき、連鎖不平衡係数 D' は 1 である。

問 68

分子系統樹に関する以下の記述について、(a) と (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はどれか。選択肢の中から一つ選べ。

分子系統樹は、塩基配列やアミノ酸配列の違いをもとに、それぞれの (a) 関係を木の形で表現したものである。もっとも古い分岐の位置が明示されているタイプの系統樹を有根系統樹とよび、そうでないものを無根系統樹とよぶ。無根系統樹を有根系統樹にするには、比較に用いた遺伝子やタンパク質よりも遠い関係のものを、(b) として加えることが一般的である。

- 1 (a) 親子 (b) 外群
- 2 (a) 親子 (b) 祖先型
- 3 (a) 類縁 (b) 外群
- 4 (a) 類縁 (b) 祖先型

問 69

2 種の生物の間でオーソログをベストヒット法で推定する場合を考える。比較している 2 種類の生物が共通の祖先から分枝した後に生じたイベントの中で、遺伝子やアミノ酸配列のベストヒット法によるオーソログの推定に影響を与えるものとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 スプライシングパターンが大きく変化した遺伝子がある。
- 2 その遺伝子を含む領域が、転座して染色体上の位置が変化した。
- 3 分断または融合を起こした遺伝子がある。
- 4 パラログとなる遺伝子が多数できた。

問 70

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選択肢の中から一つ選べ。

- 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 遺伝子やタンパク質の性質の違いによって、塩基置換やアミノ酸置換の速度は異なる。
- 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组は必要である。
- 4 遺伝子の同義置換速度は、バクテリアから、植物、菌類、動物に至るまで、遺伝子の種類によらず、どの系統でもほぼ一定である。

問 71

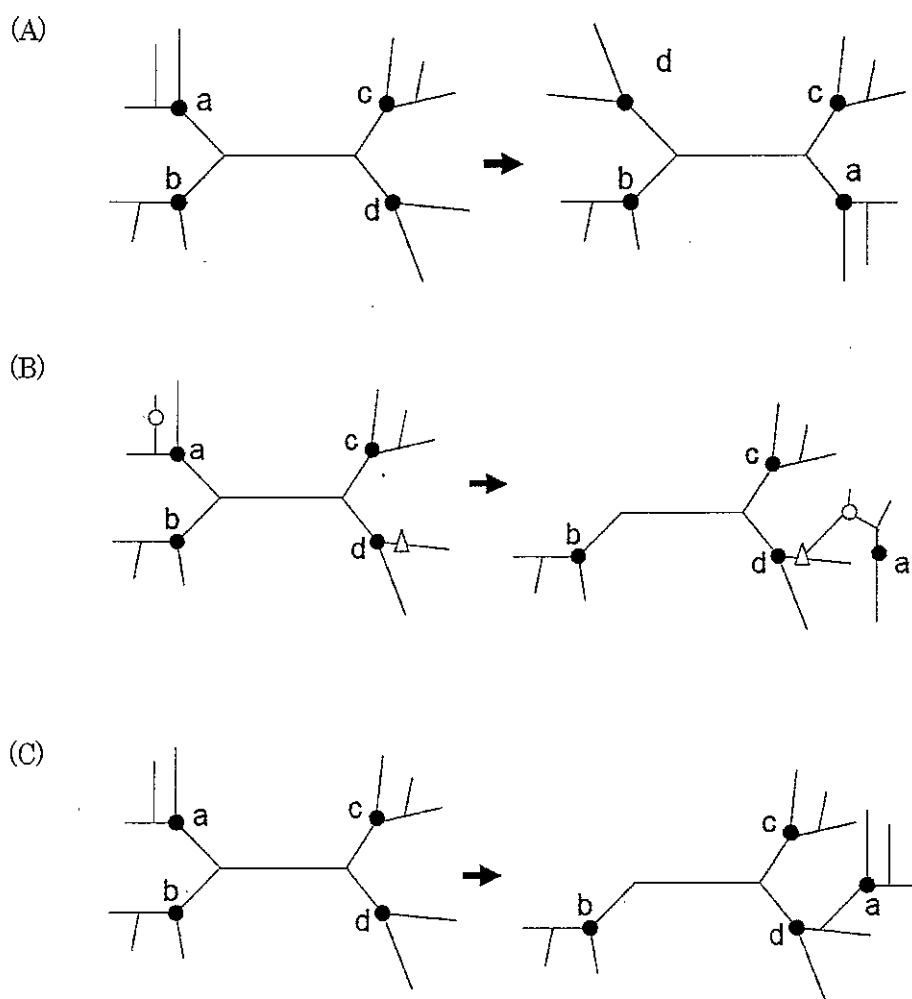
分子系統樹の作成法に関する以下の記述について、(a) から (c) に入る語句の組み合わせとしてもっとも適切なものはどれか。選択肢の中から一つ選べ。

分子系統樹の作成法のうち、UPGMA と (a) は、配列のアラインメントから計算される (b) をもとに系統樹を推定するのに対し、最大節約法と (c) は、配列のアラインメントの各サイトを形質と見なしたとき、形質の変化数が最小になるように、あるいは観測された形質の状態 (アラインメント) の確率を最大化するように系統樹を推定する方法である。

- | | | |
|-------------|----------|-----------|
| 1 (a) 近隣接合法 | (b) 進化距離 | (c) 最尤法 |
| 2 (a) 最尤法 | (b) 進化速度 | (c) 動的計画法 |
| 3 (a) 動的計画法 | (b) 進化速度 | (c) 最尤法 |
| 4 (a) 最尤法 | (b) 進化距離 | (c) 動的計画法 |

問 72

分子系統樹の樹形(topology)を探索する方法として系統樹の再編がある。再編の代表的な手法として、最近隣枝交換(nearest neighbor interchange)、部分木剪定・接木(subtree pruning and regrafting)、系統樹切断・再接続(tree bisection and reconnection)がある。以下の3つの再編処理(左の系統樹を右の形に変換)を表すものとして、適切な組み合わせを選択肢の中から一つ選べ。ただし、図(B)の○と△は、再編前は枝上にあるが、再編後に節となる位置を表している。



- | | | |
|----------------|---------------|---------------|
| 1 (A) 最近隣枝交換 | (B) 部分木剪定・接木 | (C) 系統樹切断・再接続 |
| 2 (A) 最近隣枝交換 | (B) 系統樹切断・再接続 | (C) 部分木剪定・接木 |
| 3 (A) 部分木剪定・接木 | (B) 最近隣枝交換 | (C) 系統樹切断・再接続 |
| 4 (A) 部分木剪定・接木 | (B) 系統樹切断・再接続 | (C) 最近隣枝交換 |

問 73

複雑ネットワークを数学的に表すモデルの一つとしてバラバシ・アルバート (Barabasi-Albert)モデルがあり、以下の手順でグラフネットワークを作成する。

- ① 次数 1 以上の m_0 個 ($m_0 \geq 2$) のノードからなるグラフから開始
- ② 単位時間毎に 1 個のノードを加え、既存の m 個のノードに対し、それぞれのノード (i) の次数 k_i に応じて確率 $p_i = \frac{k_i}{\sum_j k_j}$ (j は全ての既存ノード) でエッジを結ぶ

このモデルで生成されるネットワークの性質を説明する以下の文章に関して、(a) から (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はどれか。選択肢の中から一つ選べ。

1999 年に発表されたバラバシ・アルバートモデルでは、単純なモデルながら、ノードの次数がべき乗則の次数分布をもつ (a) 性を有しつつ、ノード数の増加に従って平均最短経路長が緩やかにしか増加しない (b) 性も示す事が知られている。一方で、複雑ネットワークで重要な 3 つの性質の残り一つである (c) 性は持たないことが知られており、いくつかの拡張モデルも提案されている。

- | | | |
|----------------|--------------|-------------|
| 1 (a) クラスタ | (b) スモールワールド | (c) スケールフリー |
| 2 (a) スモールワールド | (b) スケールフリー | (c) クラスタ |
| 3 (a) スケールフリー | (b) スモールワールド | (c) クラスタ |
| 4 (a) スモールワールド | (b) クラスタ | (c) スケールフリー |

問 74

グルコース濃度の変動を表す常微分方程式

$$\frac{d[\text{Glucose}]}{dt} = -k[\text{Glucose}]$$

をオイラー法で数値積分した時の、 $t=2.0\text{sec}$ での値はどれか。ただし、 $t=0.0$ の時 $[\text{Glucose}]=1.0\text{mM}$ 、反応定数 $k=0.2\text{sec}^{-1}$ 、積分ステップ幅 $h=1.0\text{sec}$ とする。

- 1 0.00 mM
- 2 0.33 mM
- 3 0.64 mM
- 4 0.80 mM

問 75

近年のタンパク質間相互作用に関する実験データの蓄積と共にデータベースの整備が進んできた。タンパク質間相互作用に関するデータベースとしてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 IntAct
- 2 HPRD
- 3 BIOGRID
- 4 INSDC

問 76

ハイスループットなタンパク質間相互作用検出法として酵母ツーハイブリッドや共免疫沈降法などがよく使われる。これらハイスループットな実験法の問題点として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 相互作用するタンパク質ペアの情報は得られるが、複合体の 3 次元構造情報は得られない。
- 2 細胞内局在を考慮しないことによる偽陰性が多い。
- 3 リン酸化のような修飾が起こったときにのみ起こる相互作用を検出できないことがある。
- 4 一時的な相互作用か安定な相互作用かを判別するのが困難である。

問 77

トランスクリプトーム測定やメタボローム測定等の定量的実験では、データの定量性を見積もるために実験を反復することが多い。実験の反復には大きく分けて生物学的反復 (biological replicates) と技術的反復 (technical replicates) がある。これらについて述べた以下の文章のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 生物学的反復とは、生物サンプルの調製から反復することである。
- 2 技術的反復とは、実験により得られたデータを異なるプログラムで反復して解析することである。
- 3 測定に高いコストがかかる場合には、測定サンプルの調製は複数回行うが、測定試料をプールし測定は 1 回ですませ、生物学的反復とすることもある。
- 4 高精度な測定のためには、技術的反復の回数だけでなく生物学的反復の回数を増やすことが重要である。

問 78

遺伝子発現量データの解析に利用される教師なし学習法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 サポートベクトルマシン
- 2 k -平均法
- 3 自己組織化マップ
- 4 階層的クラスタリング

問 79

以下のマイクロアレイのうち、未知の遺伝子を検出できるものを選択肢の中から一つ選べ。

- 1 タイリングアレイ
- 2 エキソンアレイ
- 3 cDNA アレイ
- 4 化合物マイクロアレイ

問 80

時間の関数である2次元のベクトル変数 $\mathbf{y}$ からなるシステムの連立微分方程式が

$$\frac{d\mathbf{y}}{dt} = \mathbf{A}\mathbf{y}$$

のように表される。ただし行列 $\mathbf{A}$ は以下で与えられるとする。

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

このシステムの安定性は行列 $\mathbf{A}$ の固有値を用いて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、システムは不安定である。
- 2 固有値は1, 2であるので、システムは安定である。
- 3 固有値は-1, -2であるので、システムは不安定である。
- 4 固有値は-1, -2であるので、システムは安定である。

平成 24 年度

日本バイオインフォマティクス学会 (JSBi)

バイオインフォマティクス技術者認定試験

【解答】

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

余白

余白

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。

Copyright © 2012 Japanese Society for Bioinformatics. All Rights Reserved.