

平成26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。すべ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、試験監督の指示があつ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。**

問 1

反復配列、ゲノムの重複に関する以下の文章の (a) ~ (c) 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。

1 ~ 4 bp の反復単位で 10 ~ 20 回ほど繰り返す縦列型の反復配列を (a) という。例として “ATATATAT...” などがある。Alu はヒトと近縁種のゲノム上に存在する短鎖散在反復配列 (b) の一種であり、ヒトゲノム中では約 100 万コピー存在する。また、基本染色体セットの整数倍の染色体数をもつ現象を (c) という。

- | | | | |
|---|-----------------|----------|-----------|
| 1 | (a) マルチジーンファミリー | (b) LINE | (c) 遺伝子重複 |
| 2 | (a) マイクロサテライト | (b) SINE | (c) 遺伝子重複 |
| 3 | (a) マイクロサテライト | (b) SINE | (c) 倍数性 |
| 4 | (a) マルチジーンファミリー | (b) LINE | (c) 倍数性 |

問 2

生物の形質は量的形質と質的形質とに大別される。ヒトの形質を示した以下の選択肢のうち、質的形質に該当する項目はどれか。もっとも適切なものを選択肢の中から一つ選べ。

- 1 身長
- 2 体重
- 3 血液型
- 4 血糖値

問 3

原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 核が存在しない。
- 2 原核生物は真正細菌と古細菌に分けられる。
- 3 分布範囲は高温、無酸素状態の極限環境までおよぶ。
- 4 周囲の環境に応じてオートファジーを行う。

問 4

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 ゲノム DNA は環状構造をとり、重要な領域の前後にはテロメアが存在している。
- 2 翻訳に寄与しない遺伝子領域であるイントロンを除くスプライシング機構がある。
- 3 様々な細胞内小器官へのタンパク質の輸送経路が発達している。
- 4 独自にゲノム DNA をもつ細胞内小器官が存在する。

問 5

核酸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 遺伝情報を担う分子としての DNA は二本鎖で存在する。
- 2 ヌクレオチドは塩基、五炭糖、リン酸という 3 つの要素で構成される。
- 3 高分子である DNA や RNA は、ヌクレオチドが重合したポリヌクレオチドである。
- 4 DNA を構成する五炭糖はリボースであり、RNA を構成する五炭糖はデオキシリボースである。

問 6

生体膜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 CO_2 などの疎水性の小分子や、 H_2O などの電荷をもたない極性小分子は、単純拡散によって脂質二重層の生体膜を通過できる。
- 2 グルコースなどの大型で電荷をもたない極性分子や、 Na^+ などの水溶性のイオンは、脂質二重層の生体膜をほとんど通過できない。
- 3 分子やイオンを特異的に輸送するために、トランスポーターやチャネルとよばれる膜タンパク質が細胞膜に埋め込まれている。
- 4 ATP などのエネルギーを用いず、生体膜の内側と外側における物質の濃度勾配にしたがう方向の輸送を能動輸送という。

問 7

生体内の膜タンパク質に関する以下の文章の (a) ~ (c) 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。

膜貫通型タンパク質における膜貫通部分の多くは (a) 構造をとるか、(b) 構造が樽のように複数回貫通する (c) 構造をとる。

- | | | | |
|---|--------------------|--------------------|--------------------------|
| 1 | (a) α ヘリックス | (b) β シート | (c) β バレル |
| 2 | (a) α ヘリックス | (b) β ターン | (c) α / β バレル |
| 3 | (a) β シート | (b) α ヘリックス | (c) α / β バレル |
| 4 | (a) β シート | (b) β ターン | (c) β バレル |

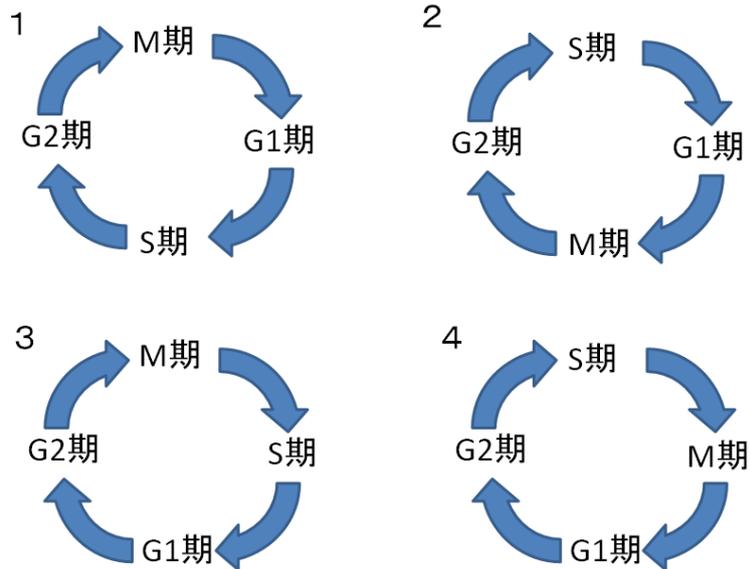
問 8

植物細胞の内部には、脂質二重層で区画された様々な細胞内小器官が見られるが、内膜と外膜の二重膜構造をもたないものとして、もっとも適切なものを選択肢の中から一つ選べ。

- 1 核
- 2 ミトコンドリア
- 3 葉緑体
- 4 液胞

問 9

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。この細胞周期における4つの期の順序として、もっとも適切なものを選択肢の中から一つ選べ。ただし、G0期は考慮しないものとする。



問 10

ヒト細胞の減数分裂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 1個の母細胞から4個の細胞が形成される。
- 2 第一分裂において、対合した相同染色体間での相同組換えが起こる。
- 3 第二分裂において、相同染色体を1組ずつもつ細胞が形成される。
- 4 減数分裂によって生じた細胞は、すべて同じ配列を有する相同染色体を1組ずつもっている。

問 11

細胞の 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 の複製では、DNA ヘリカーゼの働きにより DNA 鎖は複製フォークと呼ばれる形状をとる。
- 2 DNA の複製では、複製起点から相補鎖合成が開始される。
- 3 DNA ポリメラーゼは、5' → 3' 方向に合成鎖を伸長する。
- 4 DNA ポリメラーゼによる DNA 鎖の合成では、プライマーは不要である。

問 12

次の①～⑤の文章は、真核生物における遺伝子発現についての記述である。①～⑤の過程の時系列に沿った並べ方として、もっとも適切なものを選択肢の中から一つ選べ。

- ① mRNA にリボソームが付着する。
- ② アミノ酸がつながり、タンパク質が合成される。
- ③ DNA の遺伝情報が mRNA 前駆体に転写される。
- ④ tRNA 上のアンチコドンが、mRNA 上のコドン情報にしたがったアミノ酸をリボソームに運んでくる。
- ⑤ mRNA 前駆体のイントロンとよばれる領域が取り除かれ、mRNA となって核膜孔から細胞質基質に移る。

- 1 ③ → ② → ① → ④ → ⑤
- 2 ⑤ → ④ → ③ → ① → ②
- 3 ⑤ → ③ → ④ → ① → ②
- 4 ③ → ⑤ → ① → ④ → ②

問 13

以下の表は、真核生物におけるタンパク質への翻訳後修飾の種類、修飾部位のアミノ酸残基、修飾による効果をまとめたものである。(a) ~ (d) 内に入る翻訳後修飾の名称や修飾部位のアミノ酸残基として、もっとも適切な組み合わせを選択肢の中から一つ選べ。

名称	修飾を受ける残基	修飾によるタンパク質への効果
パルミトイル化	(a)	タンパク質の疎水性を高める。
O-結合型グリコシル化	(b)	タンパク質の親水性を高める。
(c)	セリン、スレオニン、チロシン、ヒスチジン	タンパク質に負電荷を導入する。
(d)	システイン	側鎖の-SH 基同士の結合により、タンパク質構造安定性が高まる。

- 1 (a) システイン (b) セリン、スレオニン (c) リン酸化 (d) ジスルフィド化
 2 (a) メチオニン (b) アスパラギン (c) 脱リン酸化 (d) ジエステル化
 3 (a) システイン (b) アスパラギン酸 (c) リン酸化 (d) ユビキチン化
 4 (a) メチオニン (b) セリン、スレオニン (c) アセチル化 (d) ジスルフィド化

問 14

アミノ酸に関する以下の記述について、(a) ~ (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を選択肢の中から一つ選べ。

pH 7 の水溶液中において、(a) と (b) の側鎖のカルボキシ基は負に荷電しており、アルギニン、リジン、(c) の側鎖のアミノ基やアミド基は正に荷電している。ただし、(c) は pH 5 の水溶液中では脱プロトン化によって電荷を失う。荷電性以外の側鎖の構造に着目すると、芳香環をもつ(d) は、20 種類のアミノ酸残基のうちもっとも分子量大きい。

- 1 (a) アスパラギン (b) グルタミン酸 (c) プロリン (d) フェニルアラニン
 2 (a) アスパラギン酸 (b) グルタミン (c) メチオニン (d) トリプトファン
 3 (a) アスパラギン (b) グルタミン酸 (c) ヒスチジン (d) チロシン
 4 (a) アスパラギン酸 (b) グルタミン酸 (c) ヒスチジン (d) トリプトファン

問 15

酵素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 酵素は反応の種類や基質特異性に基づき、EC 番号によって分類されている。
- 2 酵素として機能するリボ核酸 (RNA) のことをリボザイムといい、RNA の切断・合成などの活性をもっている。
- 3 タンパク質からなる酵素は熱、pH、塩濃度、溶媒など置かれた条件が大きく変化しても、その機能が失われることはない。
- 4 活性部位への結合によって酵素の機能を低下させる分子を、酵素阻害剤という。

問 16

グルコースの代謝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 解糖系は嫌気的な代謝である。
- 2 解糖系は細胞質で行われる。
- 3 クエン酸回路および酸化的リン酸化はゴルジ体で行われる。
- 4 解糖系では 1 分子のグルコースから ATP を 2 分子産生することができる。

問 17

細胞間のシグナル伝達のタイプとその特徴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 内分泌型 — 血中に入ったシグナル分子が全身に運ばれ、標的細胞の受容体にシグナル分子が結合することによりシグナルを伝える。
- 2 パラクリン型 — 細胞外液にシグナル分子が拡散して、周囲の標的細胞の受容体にシグナル分子が結合することによりシグナルを伝える。
- 3 神経型 — 神経細胞の軸索を伝ってシグナル分子が運ばれ、標的細胞に電気シグナルとして伝わる。
- 4 接触型 — シグナル分子を分泌せず、細胞表面のタンパク質がシグナル分子として直接結合することで標的細胞にシグナルを伝える。

問 18

生体分子内や分子間に起こる化学結合や相互作用に関する以下の記述について、(a) ~ (d) に対応する化学結合や相互作用の名称として、もっとも適切なものを表中の選択肢の中から一つ選べ。

- (a) 酸性のアミノ酸と、塩基性のアミノ酸の間に働く相互作用
- (b) 2つのシステインの間における酸化反応による結合
- (c) 遊離アミノ酸のアミノ基と、別の遊離アミノ酸のカルボキシル基間の脱水縮合による結合
- (d) DNA や RNA の塩基間における結合

- 1 (a) ファンデルワールス力 (b) グリコシド結合 (c) ジスルフィド結合 (d) 静電相互作用
- 2 (a) ファンデルワールス力 (b) π - π 相互作用 (c) グリコシド結合 (d) ホスホジエステル結合
- 3 (a) 静電相互作用 (b) ジスルフィド結合 (c) ペプチド結合 (d) 水素結合
- 4 (a) 静電相互作用 (b) 疎水性相互作用 (c) 水素結合 (d) ホスホジエステル結合

問 19

体液性免疫で抗体産生を行う細胞として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 T 細胞
- 2 B 細胞
- 3 マクロファージ
- 4 NK 細胞

問 20

1 ng の二本鎖 DNA の全領域を PCR 法で増幅したい。適切なプライマーを用いてこの DNA 全量の全領域を完全に増幅できた場合、10 サイクルで理論上およそ何 μg に増幅されるか、もっとも適切な数値を選択肢の中から一つ選べ。

- 1 0.1
- 2 0.5
- 3 1
- 4 2

問 21

二進数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 整数 n を表現するには少なくとも $\log_2 n$ 桁必要となる。
- 2 負の整数に 2 の補数表現を用いた場合、正の整数 n と負の整数 $-n$ の表現は等しくなる。
- 3 負の整数に 2 の補数表現を用いることで、減算を加算として行なうことができる。
- 4 整数 n をビット列で表現した場合、左シフト・右シフトの演算は n に対して 2 のべき乗を乗算・除算することに対応する。

問 22

計算機性能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 CPU コアの数 n 倍にしても、すべてのプログラムの動作速度が n 倍になるとは限らない。
- 2 CPU の動作周波数が 2 倍になれば、プログラムの動作速度が必ず 2 倍になる。
- 3 FLOPS とは、1 秒間に浮動小数点は何回計算できるかによってあらわしたコンピュータの性能指標の一つである。
- 4 ムーアの法則とは、計算機の処理性能の向上の長期的傾向に関する経験的な法則である。

問 23

TCP/IP でサブネット化されたローカルエリアネットワークにコンピュータを接続し、Web ブラウザのアドレスバーに <http://www.jsbi.org/> と入力して、バイオインフォマティクス学会の Web ページを閲覧した。この時、必ずしも設定の必要が無い項目を選択肢の中から一つ選べ。ただし、このローカルエリアネットワークには DHCP サーバが存在しないとする。

- 1 メールアドレス
- 2 IP アドレス
- 3 ゲートウェイアドレス
- 4 DNS サーバアドレス

問 24

以下のプロトコルのうち、通信経路が暗号化されるため、もっとも秘匿性の高いと考えられるプロトコルを選択肢の中から一つ選べ。

- 1 FTP
- 2 TELNET
- 3 SMTP
- 4 SSH

問 25

データ構造の実装について述べた以下の文の (a) および (b) にそれぞれに入る用語の組み合わせとして、もっとも不適切なものを選択肢の中から一つ選べ。

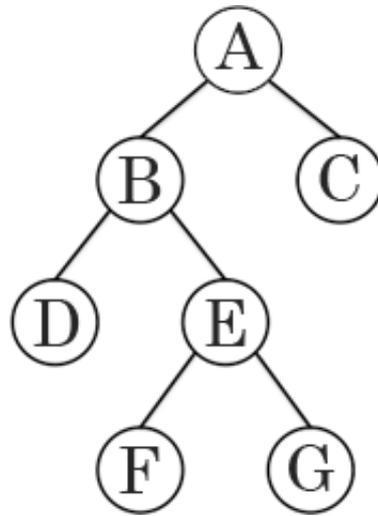
(a) は (b) を用いて効率よく実装することができる。

- | | |
|------------|---------|
| 1 (a) スタック | (b) 配列 |
| 2 (a) キュー | (b) リスト |
| 3 (a) ヒープ | (b) リスト |
| 4 (a) スタック | (b) リスト |

問 26

以下の疑似コードは、木を巡回し、各節点 v を $\text{output}(v)$ で出力するプログラムを表す。

```
traverse(node  $v$ ):  
begin  
  output( $v$ );  
  foreach  $c \in \text{children}(v)$  do:  
    traverse( $c$ );  
  end  
end
```



図の木に対して $\text{traverse}(A)$ を実行した際の出力として、もっとも適切なものを選択肢の中から一つ選べ。ただし、 $\text{children}(v)$ は、節点 v の子の節点を左優先の順序つき集合としてすべて返す関数とする。

- 1 A, B, C, D, E, F, G
- 2 A, B, D, E, F, G, C
- 3 D, F, G, E, B, C, A
- 4 D, B, F, E, G, A, C

問 27

時間計算量・空間計算量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 $O(1)$ 時間アルゴリズムを $O(1)$ 回実行するのに必要な時間は $O(1)$ である。
- 2 ある問題に対し、平均時間計算量が $O(n^2)$ 、空間計算量が $O(1)$ のアルゴリズムを実行したとき、十分大きなサイズ n の入力に対して平均 1 秒を要した。このとき、サイズ $10n$ の入力に対しては平均 100 秒程度の実行時間を要すると予想できる。
- 3 $O(n)$ 時間アルゴリズムを $O(n)$ 回実行するのに必要な時間は $O(n)$ である。
- 4 同じ問題を解く際に $O(n)$ 時間アルゴリズムよりも $O(n^2)$ 時間アルゴリズムの方が高速に動作する場合がある。

問 28

スタック構造にデータ列 A, B, C, D を、この順番でプッシュまたはポップ操作を行った場合に、ポップ操作で取り出されるデータとして、いかなるプッシュおよびポップ操作の組み合わせでも実現しない配列を、選択肢の中から一つ選べ。例えば、A をプッシュした後に、B をプッシュし、その後ポップを 2 回行くと、1 回目のポップで B、2 回目のポップで A が取り出されるので、B, A が取り出されるデータの列となる。

- 1 A, D, C, B
- 2 A, B, C, D
- 3 C, D, A, B
- 4 B, A, D, C

問 29

一定の順序に並べられたデータの列から、目的の値を見つけ出す探索法の一つに二分探索法がある。二分探索法によって下記に示したデータから「7」を見つけ出す例を示す。

データの列 1 3 7 9 13 21 22

- ① 中央の値 9 と目的の値 7 を比較し、範囲を「1, 3, 7」に絞る。
- ② 「1, 3, 7」の中央の値 3 と目的の値 7 を比較し、範囲を「7」に絞る。
- ③ 範囲中に値は 1 つしかないため、その値 7 と目的の値 7 を比較し、目的の値が見つかる。

このように、二分探索法では、探索の範囲を半分に絞り込む操作を繰り返す。上の例では、2 回の二分走査で目的の値を見つけることができた。

同様にして、8 回の二分操作で目的の値を必ず見つけ出すことができるデータ列の最大長として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 256
- 2 511
- 3 6561
- 4 15625

問 30

以下の「表：実験データ一覧」の実験データに対して、以下に示す SQL 文を実行した時に、結果として得られるデータの件数として、もっとも適切なものを選択肢の中から一つ選べ。

```
SELECT 実験番号 FROM 実験データ一覧
WHERE 平均リード長 > 100 AND エラー率(<) < 1.0 ;
```

表：実験データ一覧

実験番号	総リード数	平均リード長	総塩基数	カバー率	エラー率(<%)
HA582	1,094,909,696	114.3	125,033,878,215	41.7	1.1
HX331	595,155,854	147.8	87,964,037,220	29.3	0.13
SP013	7,904,937,577	38.1	301,178,121,731	100.4	0.8
PL061	3,577,460	1,747.7	6,252,328,113	2.1	8.7
RE071	6,509,412	450.6	2,933,141,255	1.0	0.025

- 1 1 件
- 2 2 件
- 3 3 件
- 4 4 件

問 31

コンピュータ上でのデータの記述形式として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。XML 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 XML では、各要素を表で表現し、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 内のデータを検索する言語として XQUERY が存在する。
- 4 XML データベースには、画像や音声を扱うことができるものがある。

問 32

リレーショナルデータベー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 リレーショナルデータベースは、一般に情報をハードディスクに格納する。
- 2 リレーショナルデータベースは、テーブル間の関係を用いてテーブルを結合することが可能である。これは SQL の JOIN 句で実現できる。
- 3 リレーショナルデータベースにおいて、一つのデータベースは常に一つのテーブルに対応する。
- 4 レコード数が増えると、検索に時間を要する場合がある。これを解消するために、索引を作成することができる。これは SQL の CREATE INDEX 文で実現できる。

問 33

X と Y を連続確率変数とする。 $\text{Var}(X)$ および $\text{Var}(Y)$ をそれぞれ X と Y の分散とする。 E を期待値演算子とする。分散および期待値に関する以下の記述のうち、もっとも適切なものを選択肢の中から一つ選べ。

- 1 常に $\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)$ が成り立つ。
- 2 X と Y が負の相関をもつとき $E(X+Y) > E(X)+E(Y)$ となる。
- 3 X と Y が正の相関をもつとき $E(X+Y) < E(X)+E(Y)$ となる。
- 4 X と Y の共分散が正のとき $\text{Var}(X+Y) > \text{Var}(X)+\text{Var}(Y)$ となる。

問 34

以下の (a) ~ (c) にそれぞれに入る語句と式の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

母平均 μ および母分散 σ^2 をもつカイ二乗分布 p 、および p から独立試行によって得た標本観測値の標本平均 m について考える。標本数を増やすほど標本平均 m の分布は (a) に漸近する。標本数を n とすると、 m の平均は μ となり、 m の分散は (b) となる。一方、 p が正規分布の場合、 m は標本数が (c) 正規分布にしたがう。

- 1 (a) 正規分布 (b) $\sigma^2 / (n-1)$ (c) 十分多い場合に限り
- 2 (a) 正規分布 (b) σ^2 / n (c) 少なくとも
- 3 (a) カイ二乗分布 (b) $\sigma^2 / (n-1)$ (c) 十分多い場合に限り
- 4 (a) カイ二乗分布 (b) σ^2 / n (c) 少なくとも

問 35

下の (a) ~ (c) にそれぞれ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。

正規分布にしたがう母集団の平均がある値 μ_0 に等しいか検査する t 検定においては、帰無仮説を「母集団の平均が μ_0 に等しい」として (a) を行う。得られた p 値が 0.0014 であったとする。このとき、有意水準 0.01 において、帰無仮説は (b) され、母集団の平均は μ_0 に (c) と結論付けられる。

- | | | | |
|---|----------|--------|---------------|
| 1 | (a) 両側検定 | (b) 受容 | (c) 等しいといえる |
| 2 | (a) 片側検定 | (b) 受容 | (c) 等しいとはいえない |
| 3 | (a) 片側検定 | (b) 棄却 | (c) 等しいといえる |
| 4 | (a) 両側検定 | (b) 棄却 | (c) 等しいとはいえない |

問 36

ある事象 A に対し、余事象を A^c と書くことにする。以下の式のうち不適切なものを選択肢の中から一つ選べ。

- 1 $1 - P(A \cap B) = P(A^c \cup B^c)$
- 2 $1 - P(A \cup B) = P(A^c \cap B^c)$
- 3 $P(A^c) = 1 - P(A)$
- 4 $P(A \cup B) = P(A) + P(B) + P(A \cap B)$

問 37

スチューデント t 分布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 スチューデント t 分布の自由度パラメータが無限大に近づくにつれ、正規分布に近づく。
- 2 スチューデント t 分布の自由度パラメータが 2 に等しいとき、2 項分布に一致する。
- 3 スチューデント t 分布の場合、累積密度関数の値は 1 を超えることがない。
- 4 スチューデント t 分布の場合、確率密度関数の値は 1 を超えることがある。

問 38

以下の (a) ~ (d) にそれぞれに入る語句と式の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

ある情報検索システムの性能を評価したいとする。TP を真陽性数、FP を偽陽性数、FN を偽陰性数、TN を真陰性数とする。性能評価には精度 (Precision)、再現率 (Recall)、F 値 (F-measure) がよく用いられる。精度は (a) で、再現率は (b) で定義され、F 値は精度と再現率の (c) で与えられる。これらの指標は、(d) の値が正確にわからなくても算出することができる。

- | | | | | |
|---|----------------------|----------------------|----------|----------|
| 1 | (a) $TP / (TP + FP)$ | (b) $TP / (TP + FN)$ | (c) 調和平均 | (d) 真陰性数 |
| 2 | (a) $TP / (TP + FN)$ | (b) $TP / (TP + FP)$ | (c) 幾何平均 | (d) 真陰性数 |
| 3 | (a) $TP / (TP + FP)$ | (b) $TN / (TN + FP)$ | (c) 調和平均 | (d) 偽陰性数 |
| 4 | (a) $TP / (TP + FP)$ | (b) $TN / (TN + FP)$ | (c) 幾何平均 | (d) 偽陰性数 |

問 39

学習アルゴリズムなどにより構築した判別器の性能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 交差検証法の目的は、未知の入力データに対する判別器の汎化誤差を推定して、判別器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 つをテストのために確保し、残りの $n-1$ 個を判別器の学習に用いる方法を n -fold 法という。
- 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のために、残りの半分を学習のために用いることになるが、これを特に分割統治法という。
- 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のために、残りのデータを学習のために用いることになるが、これを特に leave-one-out 法という。

問 40

人工ニューラルネットワーク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 人工ニューラルネットワークとは、人工ニューロンのネットワークが学習によってシナプス結合強度を変えながら、問題解決能力をもつようになるモデルのことである。
- 2 人工ニューラルネットワークの一つである、3層パーセプトロンは、入力層、増幅層、解決層からなる。
- 3 3層パーセプトロンの学習には、誤差逆伝搬法と呼ばれる勾配法がしばしば使われる。
- 4 ディープラーニング法は、人工ニューラルネットワークを用いている。

問 41

RNA-seq 解析をはじめとする、大量の cDNA 配列解析の過程で実施する配列データの品質検査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 品質検査ツールを用いて、クオリティ値の低い配列データが大量に含まれていないか検査する。
- 2 既知の mRNA 配列を網羅したデータベースに対してホモロジー検索を行い、類似配列数を検査する。
- 3 遺伝子ごとの発現量を求め、主要なマーカー遺伝子が検出できているか検査する。
- 4 マッピングできなかった配列について、他の生物種由来の配列とのホモロジー検索を行い、サンプルに混入がないか検査する。

問 42

脊椎動物における DNA メチル化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 遺伝子転写開始点近傍の CpG アイランドにおけるメチル化の程度は、その遺伝子の発現量と負の相関がある。
- 2 遺伝子内部領域の CpG におけるメチル化の程度は、その遺伝子の発現量と正の相関がある。
- 3 バイサルファイトシークエンスによって、メチル化されているシトシン (C) はチミン (T) に置換され、メチル化されていないシトシンはそのまま保存される。
- 4 がん細胞では正常細胞とは著しく異なる DNA メチル化パターンが観測され、遺伝子の発現パターンも異なっている。

問 43

次世代シーケンサを用いたオーミクス解析の実験手法とその目的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 全ゲノムシーケンシングにより、一塩基変異の情報を網羅的に測定する。
- 2 全エキソームシーケンシングにより、エキソンの位置を網羅的に測定する。
- 3 RNA-seq により、遺伝子の発現情報を網羅的に測定する。
- 4 ChIP-seq により、DNA 結合タンパク質の結合位置を網羅的に測定する。

問 44

オーソログ遺伝子データベース COGs (Clusters of Orthologous Groups of proteins) 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 COGs はオーソログ関係に基づき、主に原核生物の完全長ゲノムのタンパク質を分類している。
- 2 COGs のオーソログ関係は、精密な遺伝子系統樹に基づいて、すべて自動的に構築されている。
- 3 COGs はアノテーションの際に、タンパク質の機能分類の指標の一つとして用いられる。
- 4 COGs と同様の方法で、真核生物を対象とした KOGs、プラスミドを対象とした POGs も構築されている。

問 45

メタゲノム解析とは、体内や環境中に生息する生物群集に由来するゲノム DNA の混合物を対象にしたゲノム配列解析のことである。個別の菌を分離・培養してから行う通常のゲノム解析と比べると、得られる情報は不完全であるものの、培養できない菌も含めて環境中にどのような生物が存在するかを網羅的に調べることができる。この目的のために、得られたメタゲノム配列から各配列が由来する生物種やコードする遺伝子の機能などを推定するが、その際に有効な解析手法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 配列相同性検索を行う。
- 2 固定長 k の塩基配列 (k -mer) の頻度を算出する。
- 3 GC 含量を算出する。
- 4 DNA の一方の鎖における G と C の含量の差に基づく値 (GC-skew) を算出する。

問 46

大腸菌ゲノム中のある遺伝子が水平伝播によって他の細菌から獲得されたとする根拠として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 この遺伝子の GC 含量は、大腸菌の他の遺伝子と比較してかなり低い。
- 2 この遺伝子のコドン使用頻度は、他の大腸菌遺伝子と比較して大きく異なる。
- 3 系統樹解析を行うと、この遺伝子は大腸菌が属するガンマプロテオバクテリア綱以外のグループに分類される。
- 4 この遺伝子のオーソログ遺伝子はバクテリアからヒトまで多くの生物種に存在する。

問 47

接尾辞配列 (suffix array) は、長大な DNA 配列の索引付けに効果的なデータ構造の一つで、文字列の接尾辞 (任意の位置から最後まで文字列) の開始位置を要素とする配列であり、その順序は接尾辞を辞書順にソートしたものである。以下の表は、DNA 配列 TCAGATGA に対して作成された接尾辞配列を表している (ただし表の一部は伏せられている)。この表により、部分文字列 AT が位置 5 のみに出現することなどが高速に検索できる。表の (a) と (b) に入る値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。

位置	1	2	3	4	5	6	7	8
配列	T	C	A	G	A	T	G	A

表: DNA 配列「TCAGATGA」の接尾辞配列

ソート後の順番	配列位置	接尾辞
1	8	A
2	3	AGATGA
3	5	ATGA
4	2	CAGATGA
5	7	GA
6	(a)	?
7	1	TCAGATGA
8	?	(b)

- | | | |
|---|-------|-----------|
| 1 | (a) 6 | (b) GATGA |
| 2 | (a) 6 | (b) TGA |
| 3 | (a) 4 | (b) GATGA |
| 4 | (a) 4 | (b) TGA |

問 48

BLAST などの配列類似性検索ツールにおいては、配列上に出現する固定長 k の単語 (k -mer) の出現位置を記録することによって、検索を効率化するための索引づけをおこなうことがよく行われる。これ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 長さ $L (\geq k)$ の配列には $L - k + 1$ 個の k -mer が含まれる。
- 2 4 種類の塩基からなる DNA 配列の場合、 4^k 通りの k -mer がある。
- 3 長さ 5,000 の塩基配列において、各 5-mer の出現頻度は、平均すると 1 未満である。
- 4 検索の索引づけにおいて、 k を大きくするほど検索の効率を上げられるが、不一致を許した検索の精度は下がる。

問 49

アミノ酸置換行列 BLOSUM 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 アミノ酸の物理化学的性質に関わる様々な指標を統計的に処理することによって作成されている。
- 2 一般に、物理化学的性質がよく似たアミノ酸に置換した場合は正のスコア、似ていないアミノ酸に置換した場合は負のスコアをとる。
- 3 アミノ酸が一致した場合のスコアは一定ではなく、アミノ酸の種類によって異なる。
- 4 BLOSUM40 と BLOSUM80 では、BLOSUM80 の方が進化的に近い配列を比較するのに向いている。

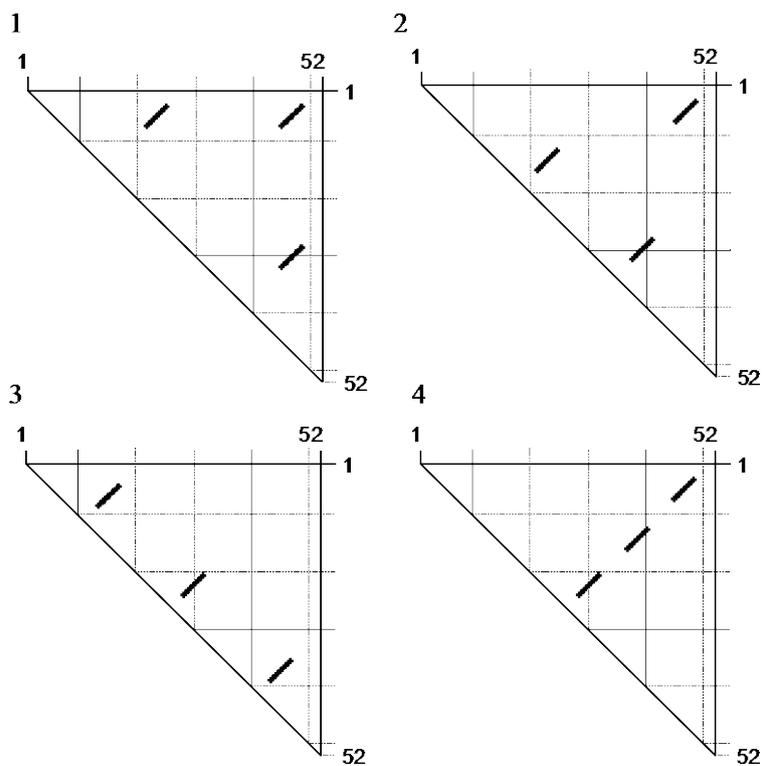
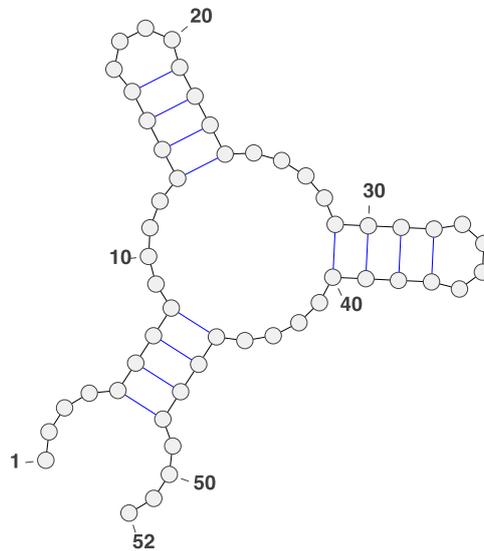
問 50

次世代シーケンサによる解析において得られた短い配列 (リード) を、リファレンス配列と照合して位置づけること (マッピング) は基本的な操作の一つである。マッピングした結果を格納するファイルの形式として SAM 形式が標準的に使われているが、この SAM 形式のファイルの、ヘッダ部分を除く本体の各行が含む情報として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 リードがリファレンス配列上にマッピングされた左端の位置。
- 2 リファレンス配列上にマッピングされたリードについて、その位置がどの程度正確であることを示す、マッピングクオリティ値。
- 3 ギャップを考慮してリードをリファレンス配列にアラインメントしたときの、挿入または欠失の位置。
- 4 リファレンス配列上の各位置に、リードが全体で何回マッピングされたかという頻度。

問 51

以下の図のような二次構造をとる RNA の配列がある。この二次構造における相補塩基対の位置を、この配列同士のドットプロット上に示した図として、もっとも適切なものを選択肢の中から一つ選べ。



問 52

下表に示すように、4本の配列中に長さ5塩基のDNA配列モチーフを見つけた。このモチーフの保存性を評価するための相対エントロピースコアとして、もっとも適切な値を選択肢の中から一つ選べ。ただし長さ N のモチーフに対する相対エントロピースコア S は次式で与えられるものとする。

$$S = \sum_{j=1}^N \sum_{a \in \{A, T, G, C\}} f_j(a) \log_2 \frac{f_j(a)}{p(a)}$$

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 の出現頻度確率を表し、 $p(a)$ は塩基 a がモチーフ以外の場所で観測される背景確率で、 $p(A) = p(G) = p(T) = p(C) = 0.25$ とする。また、 $0 \log 0 = 0$ とする。例えば、1文字目のスコアについては、 $0 \log_2 \frac{0}{0.25} \times 3 + 1 \log_2 \frac{1}{0.25} = 2$ と計算される。

表：モチーフ配列

	塩基位置番号				
	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	T	T
配列 4	T	A	A	C	T

- 1 6.5
- 2 7.0
- 3 7.5
- 4 9.0

問 53

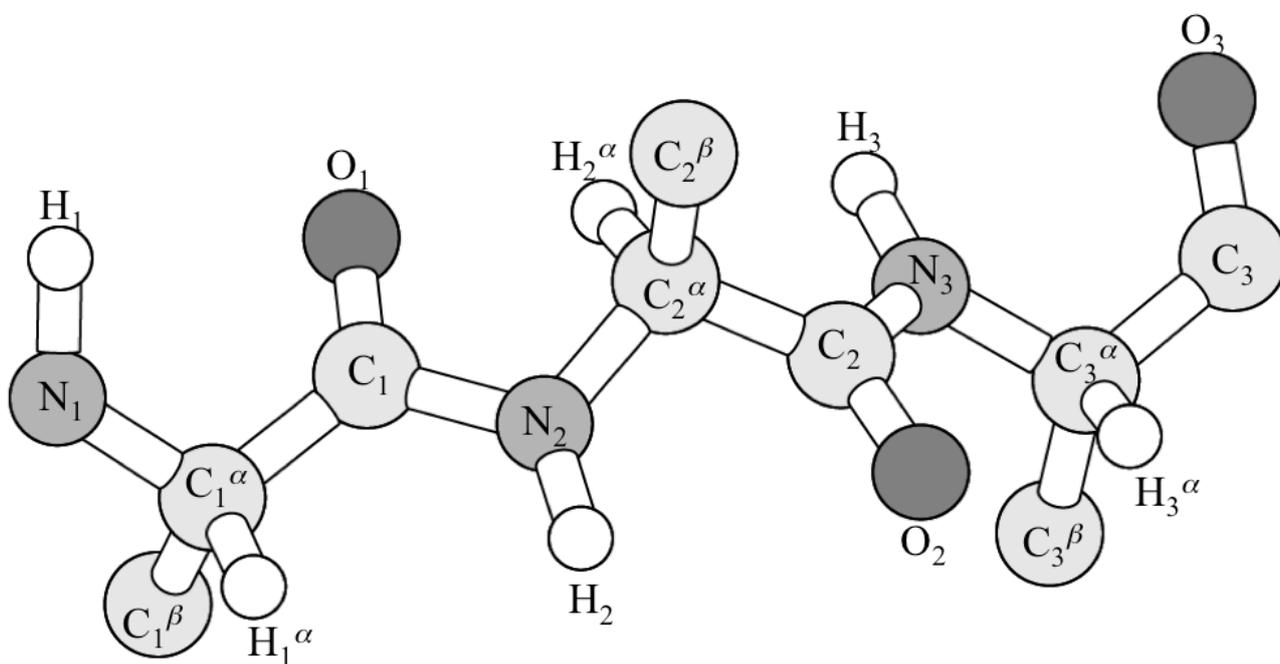
立体構造の類似した 2 つのタンパク質において、対応関係にある原子がわかっている場合には、RMSD (Root Mean Square Deviation) と呼ばれる数値を計算することにより、立体構造の類似性を数値化することができる。RMSD は、すべての対応する原子間の距離を 2 乗した値の平均値 (相加平均) の平方根として定義される。あるタンパク質における 5 つの原子 A, B, C, D, E と別のタンパク質の 5 つの原子 A', B', C', D', E' がこの順序で対応することが分かっており、互いにもっともよく重なるように回転・並進して重ねた時の対応原子間の距離は以下の表のようになった ($1 \text{ \AA} = 0.1 \text{ nm}$ である)。5 つの原子に対する RMSD 値はいくらになるか。もっとも適切な値を選択肢の中から一つ選べ。

対応する原子対	距離
原子 A - 原子 A'	2.0 \AA
原子 B - 原子 B'	1.0 \AA
原子 C - 原子 C'	0.0 \AA
原子 D - 原子 D'	1.0 \AA
原子 E - 原子 E'	3.0 \AA

- 1 1.0 \AA
- 2 1.4 \AA
- 3 1.7 \AA
- 4 3.9 \AA

問 54

タンパク質の主鎖構造は、アミノ酸残基ごとの三つの二面角 φ, ψ, ω によって特徴づけられる。二面角とは四つの結合した原子の並び $a-b-c-d$ で定まる、面 $a-b-c$ と $b-c-d$ のなす角であり、結合 $b-c$ を軸とする回転角である。 φ, ψ は C^α 原子のそれぞれ前後の結合を軸とする二面角であり、 ω はペプチド結合（アミノ酸間の共有結合 $C-N$ ）を軸とする二面角である。以下の図に示された三つの連続したアミノ酸残基において、2 番目（中央）のアミノ酸残基の φ, ψ, ω を定める四つの原子の並びとして、もっとも適切な組み合わせを選択肢の中から一つ選べ。ただし、アミノ酸番号は下付き添え字で示されている（例えば、1 番目のアミノ酸の N 原子は N_1 、2 番目のアミノ酸の C^α 原子は C_2^α である）。



- | | | | |
|---|---|--|--|
| 1 | $\varphi : C_1 - N_2 - C_2^\alpha - C_2$ | $\psi : N_2 - C_2^\alpha - C_2 - N_3$ | $\omega : C_2^\alpha - C_2 - N_3 - C_3^\alpha$ |
| 2 | $\varphi : C_1^\alpha - C_1 - N_2 - C_2^\alpha$ | $\psi : C_1 - N_2 - C_2^\alpha - C_2$ | $\omega : N_2 - C_2^\alpha - C_2 - N_3$ |
| 3 | $\varphi : N_2 - C_2^\alpha - C_2 - N_3$ | $\psi : C_1^\alpha - C_1 - N_2 - C_2^\alpha$ | $\omega : C_1 - N_2 - C_2^\alpha - C_2$ |
| 4 | $\varphi : C_1 - N_2 - C_2^\alpha - C_2$ | $\psi : C_1^\alpha - C_1 - N_2 - C_2^\alpha$ | $\omega : C_2^\alpha - C_2 - N_3 - C_3^\alpha$ |

問 55

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

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

問 56

以下は PDB のタンパク質立体構造データ (PDB コード 5hvp) の一部である。このフォーマットでは、すべてのデータを `_{カテゴリーグループ}_{カテゴリー}. {アイテム}` という階層的なタグで分類したテーブル形式で記述する。データが複数行に及ぶ場合は `loop_` の後にタグを列挙し、その後にデータが記述される。このデータ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。

```
data_5HVP
#
(省略)
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.Cartn_x_esd
_atom_site.Cartn_y_esd
_atom_site.Cartn_z_esd
_atom_site.occupancy_esd
_atom_site.B_iso_or_equiv_esd
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
(省略)
ATOM 89 N N . VAL A 1 11 ? 25.369 30.691 11.795 1.00 17.93 ? ? ? ? ? ? 11 VAL A N 1
ATOM 90 C CA . VAL A 1 11 ? 25.970 31.965 12.332 1.00 17.75 ? ? ? ? ? ? 11 VAL A CA 1
ATOM 91 C C . VAL A 1 11 ? 25.569 32.010 13.808 1.00 17.83 ? ? ? ? ? ? 11 VAL A C 1
ATOM 92 O O . VAL A 1 11 ? 24.735 31.190 14.167 1.00 17.53 ? ? ? ? ? ? 11 VAL A O 1
ATOM 93 C CB . VAL A 1 11 ? 25.379 33.146 11.540 1.00 17.66 ? ? ? ? ? ? 11 VAL A CB 1
ATOM 94 C CG1 . VAL A 1 11 ? 25.584 33.034 10.030 1.00 18.86 ? ? ? ? ? ? 11 VAL A CG1 1
ATOM 95 C CG2 . VAL A 1 11 ? 23.933 33.309 11.872 1.00 17.12 ? ? ? ? ? ? 11 VAL A CG2 1
(省略)
```

- 1 `_atom_site.B_iso_or_equiv` は温度因子であり、図に示された原子の最大値は 17.93 である。
- 2 `_atom_site.label_comp_id` は残基名であり、11 番目の残基の残基名は VAL である。
- 3 `_atom_site.Cartn_y` は原子座標値であり、11 番目の残基の C^β 炭素の y 座標値は 33.146 である。
- 4 `_atom_site.label_asym_id` はタンパク質鎖などの分子の識別子であり、ここに示された残基の値は A である。

問 57

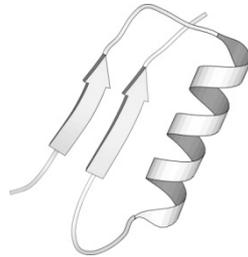
以下の図は、DNA に結合したタンパク質 (PDB コード 4l0z) の分子グラフィックスである。DNA は灰色、タンパク質は白のリボンモデルで示され、タンパク質のアミノ酸残基のうち、アルギニン (Arg) またはリシン (Lys) が黒、アスパラギン酸 (Asp) またはグルタミン酸 (Glu) が白の棒球 (ボール&スティック) モデルで示されている。このグラフィックスに関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。



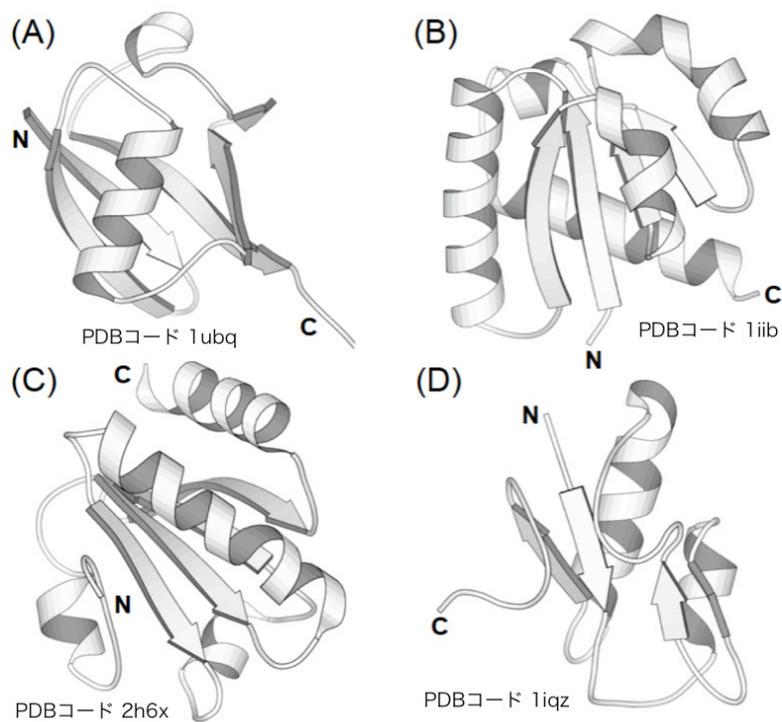
- 1 図右側の β シートを多く含むタンパク質は主に DNA 二重らせんの主溝に入り込む形で相互作用している。
- 2 図左側のタンパク質は α ヘリックスの側面で DNA 二重らせんの主溝と相互作用している。
- 3 DNA 二重らせんに接しているアミノ酸は主に負電荷をもっている。
- 4 タンパク質は B 型の DNA 二重らせんに結合している。

問 58

以下の図は、 β - α - β と呼ばれる超二次構造であり、 α/β クラスのタンパク質に含まれる特徴的な部分構造である。



以下の図 (A) ~ (D) のタンパク質のうち、 β - α - β 超二次構造を含む構造の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

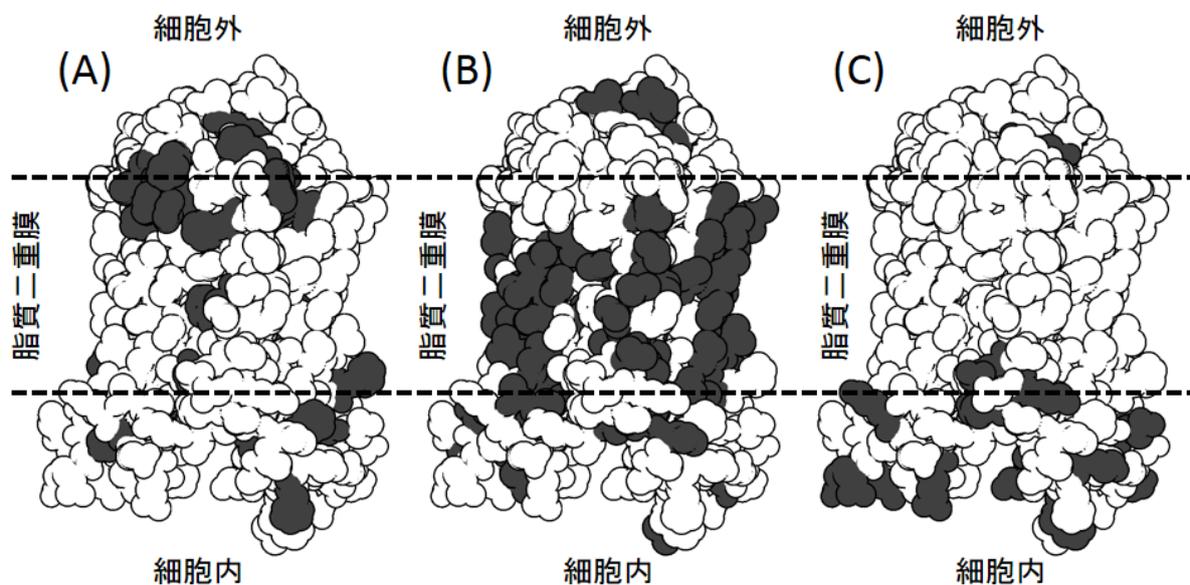


- 1 (A) と (B)
- 2 (C) と (D)
- 3 (A) と (C)
- 4 (B) と (C)

問 59

膜タンパク質の立体構造の一つ (PDB コード 3p0g の A 鎖) を、細胞外が上、細胞内が下になるように向きをそろえ、分子全体が白になるようにして、分子描画ソフトで表示した。その後、アミノ酸残基アラニン (Ala)、ロイシン (Leu)、イソロイシン (Ile)、バリン (Val) を黒に変えて表示した。次に、もう一度分子全体を白に戻し、アミノ酸残基アルギニン (Arg) とリシン (Lys) を黒に変えて表示した。これら二つの操作により表示される図に関する以下の記述のうち、もっとも適切なものを選択肢の中から一つ選べ。

- 1 Ala, Leu, Ile, Val だけを黒にすると図 (A)、Arg, Lys だけを黒にすると図 (B) になった。
- 2 Ala, Leu, Ile, Val だけを黒にすると図 (A)、Arg, Lys だけを黒にすると図 (C) になった。
- 3 Ala, Leu, Ile, Val だけを黒にすると図 (B)、Arg, Lys だけを黒にすると図 (A) になった。
- 4 Ala, Leu, Ile, Val だけを黒にすると図 (B)、Arg, Lys だけを黒にすると図 (C) になった。



問 60

以下は分子シミュレーションにおけるポテンシャルエネルギーに関する記述である。文中の(a) ~ (c) に入るグラフ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。

ファンデルワールス相互作用は以下のレナード・ジョーンズポテンシャル $V(r)$ の式で表される。

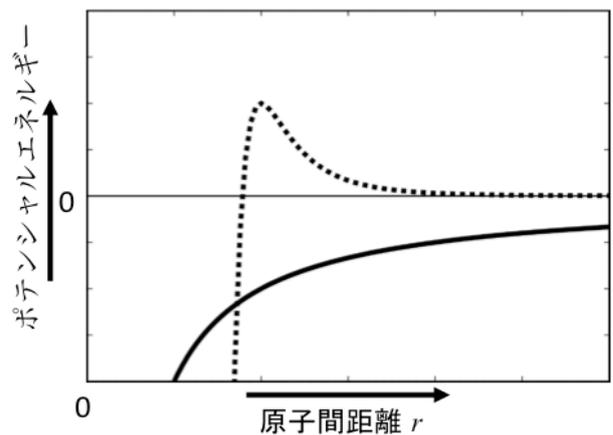
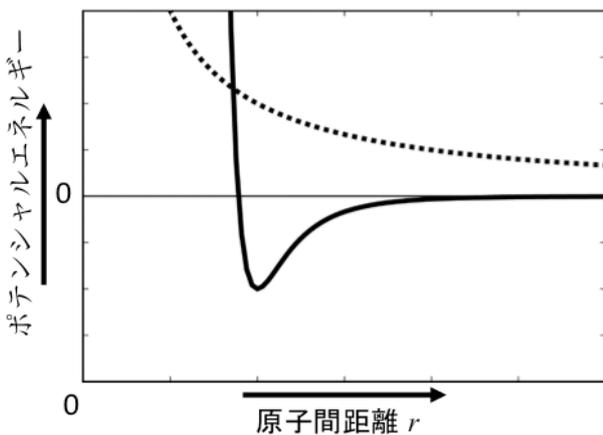
$$V(r) = A \left(\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right)$$

また、静電相互作用は、以下のクーロンポテンシャル $E(r)$ の式で表される。

$$E(r) = B \frac{q_1 q_2}{r}$$

ここで、 r は2つの原子間の距離、 r_0 、 A 、 B は正の値のパラメータである。 q_1 と q_2 は原子の中心に置かれた電荷の値であり、正負どちらの値もとることができる。 $V(r)$ は以下の (a) の形をしている。

$E(r)$ は q_1 と q_2 の符号が同じ場合 (両方がプラス、あるいは両方がマイナス) は反発力となり (b) の形になる。一方、 $E(r)$ は q_1 と q_2 の符号が異なる場合 (片方がプラス、もう片方がマイナス) は引力となり、(c) の形になる。



- | | | | |
|---|-----------|-----------|-----------|
| 1 | (a) 左図の実線 | (b) 左図の点線 | (c) 右図の実線 |
| 2 | (a) 左図の実線 | (b) 右図の点線 | (c) 左図の実線 |
| 3 | (a) 左図の点線 | (b) 左図の実線 | (c) 右図の点線 |
| 4 | (a) 左図の実線 | (b) 右図の実線 | (c) 左図の点線 |

問 61

アミノ酸配列からタンパク質二次構造を予測または同定する方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 二次構造を α ヘリックス、 β ストランド、コイルの 3 状態で予測する際に、一般にもっとも高い精度で予測できるのは α ヘリックスである。
- 2 二次構造予測では、 α ヘリックスと β ストランドは独立に予測できない。
- 3 PSIPRED 法は、ニューラルネットワークの入力に PSI-BLAST の位置特異的スコア行列を利用することで、高い正答率を示した。
- 4 二次構造予測に対して正解となる二次構造は、DSSP プログラムを用いて、立体構造を入力として同定することができる。

問 62

以下のタンパク質の立体構造予測に関する記述の (a) ~ (d) に入る語句と数値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。

タンパク質立体構造の予測法は、(a) と (b) の二つに大きく分類される。(a) では、分子シミュレーション技法により、標的アミノ酸配列をもつタンパク質分子の膨大な配座空間から、もっとも自由エネルギーが低い配座を探索する。(b) では、立体構造データベースの中から、標的アミノ酸配列の立体構造にもっとも近いと予測される鋳型立体構造を選択し、それを標的アミノ酸配列に合わせて修正して予測構造を得る。特に、鋳型立体構造の選択をアミノ酸配列相同性により行う(b) のことを、(c) という。(c) の予測精度は、標的と鋳型立体構造のアミノ酸配列一致度に大きく依存する。主鎖構造のおおまかな予測が成功するための配列一致度の下限は、(d) 程度であるとされる。

- 1 (a) 第一原理的方法 (b) 比較モデリング法 (c) ホモロジー・モデリング法 (d) 30%
- 2 (a) 比較モデリング法 (b) 第一原理的方法 (c) フラグメント・アセンブリ法 (d) 30%
- 3 (a) 第一原理的方法 (b) 比較モデリング法 (c) ホモロジー・モデリング法 (d) 80%
- 4 (a) 比較モデリング法 (b) 第一原理的方法 (c) フラグメント・アセンブリ法 (d) 80%

問 63

以下は、ある生物のゲノム配列が決定されたときに、コードされるタンパク質に網羅的に注釈付け（アノテーション）する方法についての記述である。文中の（ a ）～（ c ）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。ただし、ゲノム上の遺伝子同定は既に行われ、すべての転写産物（タンパク質）のアミノ酸配列が得られているものとする。

- ① まず、BLAST などの標準的な配列相同性検索ツールを使ってゲノム中のすべての配列のデータベース検索を行い、有意な相同性を示すタンパク質のアノテーションを参照する。このとき、（ a ）などの基準を用いて、なるべくオーソロガスなタンパク質を参照することに注意する。
- ② 有意に相同なタンパク質が見つからない場合は、タンパク質ドメインデータベース Pfam などに登録された既知ドメインと、プロフィール法や（ b ）を用いて、弱い相同性がないか確認する。
- ③ それでも情報が得られない場合は、二次構造予測法、翻訳後修飾予測法、（ c ）など、相同性に依存しない方法を適用しアノテーションを行う。

- 1 (a) 双方向ベストヒット (b) 隠れマルコフモデル法 (c) 膜貫通ヘリックス予測法
- 2 (a) 双方向ベストヒット (b) 細胞内局在予測法 (c) 膜貫通ヘリックス予測法
- 3 (a) コドン使用頻度 (b) 隠れマルコフモデル法 (c) 分子ドッキング法
- 4 (a) コドン使用頻度 (b) 細胞内局在予測法 (c) 分子ドッキング法

問 64

以下のタンパク質間相互作用に関する記述の（ a ）～（ c ）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。

タンパク質間相互作用を大規模に計測する代表的な実験手法として（ a ）と（ b ）がある。（ a ）では、ある転写因子を D ドメインと A ドメインに分離し、それぞれ相互作用を計測したいタンパク質 X と Y と融合させて、D-X, A-Y として発現させる。タンパク質 X, Y が結合すると、一緒に D ドメインと A ドメインも結合し、本来の転写因子の活性を取り戻し、レポーター遺伝子の発現量を上げるため、X と Y の相互作用を確認できる。一方、（ b ）では、相互作用を計測したいタンパク質 X に、タグと呼ばれる認識配列を融合させ、測定するサンプルにタグ融合 X を加える。次に、タグと選択的に結合する分子やカラムを用いて、タンパク質をサンプルから抽出する。このとき、タグ融合 X と結合するタンパク質も一緒に抽出されるので、これを同定する。（ b ）のうち、タグの認識に抗原-抗体反応を利用する手法は（ c ）とも呼ばれる。

- 1 (a) ツーハイブリッド法 (b) プルダウン法 (c) 共免疫沈降法
- 2 (a) プルダウン法 (b) ツーハイブリッド法 (c) 共免疫沈降法
- 3 (a) プルダウン法 (b) ツーハイブリッド法 (c) 免疫蛍光法
- 4 (a) ツーハイブリッド法 (b) プルダウン法 (c) 免疫蛍光法

問 65

以下の表は、標準的なコドン表で Arg をコードするコドンの使用頻度を、ヒトのミトコンドリア、大腸菌および高度好熱菌について示したものである。(a) ~ (c) に対応する生物種の説明がもっとも適切なものを選択肢の中から一つ選べ。

	(a)	(b)	(c)
CTG	0.38	0.02	0.11
CGC	0.40	0.34	0.41
CGA	0.06	0.01	0.45
CGG	0.10	0.43	0.03
AGA	0.04	0.01	*
AGG	0.02	0.19	*

*: 終止コドン

- | | | | |
|---|----------------|----------------|----------------|
| 1 | (a) ヒトのミトコンドリア | (b) 大腸菌 | (c) 高度好熱菌 |
| 2 | (a) ヒトのミトコンドリア | (b) 高度好熱菌 | (c) 大腸菌 |
| 3 | (a) 大腸菌 | (b) 高度好熱菌 | (c) ヒトのミトコンドリア |
| 4 | (a) 大腸菌 | (b) ヒトのミトコンドリア | (c) 高度好熱菌 |

問 66

以下の多重遺伝子族の進化過程に関する記述について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 遺伝子が加速進化することで形成される。
- 2 遺伝子が非相同組換えによって重複することで形成される。
- 3 遺伝子から転写された mRNA が、逆転写されることで形成される。
- 4 染色体全体が重複することで形成される。

問 67

塩基置換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 トランスバージョン変異（プリンとピリミジン間の変異）よりも、トランジション変異（プリン間またはピリミジン間の変異）の方が起こりやすい。
- 2 同義置換はすべてコドンの 3 塩基目に起きる塩基置換である。
- 3 非同義置換速度は遺伝子によって数十倍異なることがあるが、同義置換速度は遺伝子によらず比較的似ている。
- 4 遺伝子のサイトあたりの非同義置換数 (dN) とサイトあたりの同義置換数 (dS) の値は、通常 $dN < dS$ である。

問 68

多型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 ハプロタイプは、遺伝子型のみを調べることにより決定できる。
- 2 複数の座位で集団内のアレルの存在が独立であることを連鎖平衡という。
- 3 父親に由来する遺伝子のプロモーターに見られる多型は、おもに父親由来のハプロタイプに存在するコード領域の多型に影響するが、母親由来のハプロタイプに存在するコード領域には影響しない。
- 4 組換えのホットスポットに挟まれた染色体の領域は、そのブロック単位で遺伝するのでそのブロック内の一部の SNP を調べるだけで残りの SNP も推測できる。

問 69

ある受容体の遺伝子について、塩基配列をチンパンジーとヒトとで比較した結果を下に示した。この結果の解釈として、もっとも不適切なものを選択肢の中から一つ選べ。なお、図中の「.」は参照配列と同じ塩基であることを意味する。図示した部分以外に変異は認められず、シークエンスエラーはないものとする。

参照配列	TGCCACCTTGGACTCCGCAGC
チンパンジー 1	.A.....G.A.T...T.
チンパンジー 2	.A.....C...G.G.T...C.
チンパンジー 3	.A.....G.A.T...C.
チンパンジー 4	.A.....G.A.T...C.
チンパンジー 5	.AT.....AG.A.T...C.
チンパンジー 6	GA...T.....G.A.T...C.
チンパンジー 7	GA.....G.A.T...C.
ヒト 1T..A.....T...
ヒト 2T..A..T...AT...
ヒト 3	...G..T..A.....AT...
ヒト 4	...G..T..A....T.AT...
ヒト 5	...GG.T..A.....AT...
ヒト 6	...G..T..A.....T...
ヒト 7	...G..T..A..T...AT...

- 1 チンパンジー1～7の間では、この受容体の遺伝子座に変異が認められるのは7塩基である。
- 2 ヒトとチンパンジーの間で完全に保存されている部位は2塩基である。
- 3 ヒトではこの受容体のハプロタイプが少なくとも7種類ある。
- 4 チンパンジー1とヒト1の配列の間には8塩基の違いがある。

問 70

遺伝子の進化的性質の一つに分子時計がある。分子時計に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 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である。
- 2 分子時計は、その遺伝子をもつ生物の世代時間に比例して遅くなる。
- 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が一組必要である。
- 4 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。

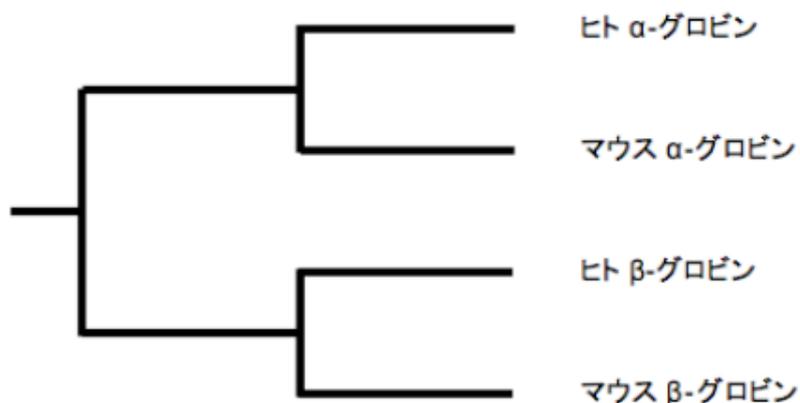
問 71

分子系統樹の作成に使われる方法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 近隣結合法 (neighbor-joining)
- 2 ウォーターシェッド法 (watershed)
- 3 最尤法 (maximum likelihood)
- 4 最大節約法 (maximum parsimony)

問 72

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。この系統樹に関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。



- 1 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。
- 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、オーソログとよばれる。
- 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、パラログとよばれる。
- 4 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、複合系統樹とよばれる。

問 73

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

- 1 DNA のメチル化解析
- 2 選択的スプライシングの解析
- 3 タンパク質の立体構造解析
- 4 微生物集団のメタゲノム解析

問 74

次世代シーケンサの進歩により、個人のゲノム情報を解析することが広く行われるようになりつつある。個人のゲノム情報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 薬の効きやすさや副作用には個人差がある。個人のゲノム情報に基づいて、一人ひとりの体質に合わせた医療を実現することを目指している。
- 2 指の長さなどの量的形質の多くは、環境の影響を受けづらい。そのため、個人のゲノム情報を用いることで、これら量的形質を精度よく推定することができる。
- 3 保険会社が個人のゲノム情報を利用し、保険加入における差別を行うことが懸念されている。一方で、人びとが個人のゲノム情報を利用して保険購入の判断を行うことで、保険会社の経営が悪化することも懸念されている。
- 4 個人のゲノム情報の解析結果に疑問があった場合には、医師や遺伝カウンセラーに相談することが望ましい。

問 75

インタラクトーム解析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 酵母ツーハイブリッド法 (Y2H) はレポーター遺伝子の活性化を測定することで二つのタンパク質間の物理的な相互作用をスクリーニングする手法である。
- 2 アフィニティー精製後の質量分析 (AP-MS) はベイト (釣り餌) タンパク質を用いて同じ複合体に含まれるタンパク質を網羅的に同定する手法である。
- 3 Database of Interacting Proteins (DIP) はヒトを含む複数生物種のタンパク質間相互作用の実験情報を集めたデータベースである。
- 4 タンパク質間相互作用ネットワークにおいては、どのタンパク質もほぼ一定数のタンパク質と相互作用する傾向がみられる。

問 76

以下の (a) ~ (d) 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。

代表的なプロテオーム解析の手法では、まず二次元電気泳動でタンパク質を (a) と分子量により分離する。次に単一成分となったタンパク質を (b) などアミノ酸残基特異的に分解する。生成したペプチドの混合物を (c) で分析し、結果を (d) と照らし合わせることで、そのペプチドを含んでいたタンパク質を同定する。

- 1 (a) 等電点 (b) トリプシン (c) 質量分析装置 (d) ゲノム配列情報
- 2 (a) 活性 (b) アミラーゼ (c) 質量分析装置 (d) 文献データベース
- 3 (a) 等電点 (b) トリプシン (c) アミノ酸シークエンサ (d) タンパク質構造データベース
- 4 (a) 活性 (b) アミラーゼ (c) X線結晶構造解析装置 (d) タンパク質構造データベース

問 77

タンパク質をノード、タンパク質間の相互作用をエッジとした無向グラフを用いて、タンパク質間相互作用ネットワークを表すことができる。タンパク質間相互作用ネットワークに関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 タンパク質間相互作用ネットワークではノード数に対して平均最短経路が短い傾向（スモールワールド性）が観察される。
- 2 タンパク質間相互作用ネットワークのグラフィカルモデルとして、無向グラフに適したマルコフ確率場モデルがある。
- 3 タンパク質間相互作用ネットワークは、連結グラフであれば完全グラフになる。
- 4 分子シャペロンは多数のタンパク質と相互作用をするため、タンパク質間相互作用ネットワークにおけるハブになる。

問 78

質量分析装置を用いた研究手法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ゲノムシークエンス解析
- 2 プロテオーム解析
- 3 メタボローム解析
- 4 有機化学合成

問 79

メタボロミクスにおける代表的な測定システム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 GC-MS (ガスクロマトグラフィー質量分析器) は分解能に優れているが、揮発性化合物以外は直接測定できず、誘導体化しなければならない。
- 2 LC-MS (液体クロマトグラフィー質量分析器) は広範囲の化合物を分析できるが、分離能が比較的低い。
- 3 CE-MS (キャピラリー電気泳動質量分析器) は一次代謝物質に多く見られる荷電分子の分離能に優れるが、中性分子の分離能は低い。
- 4 NMR (核磁気共鳴分光法) は分子の同位体を高感度に検出できるが、質量分析器に比べて網羅性が低い。

問 80

4つの遺伝子 (遺伝子 A, B, C, D) から構成される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もっとも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 と遺伝子 C の発現量が減少した。

遺伝子 B を破壊すると、遺伝子 C の発現量が減少した。

遺伝子 C 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 A, B, C 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 B は遺伝子 C の発現量を増加させる。
- 3 遺伝子 B は遺伝子 D の発現量を増加させる。
- 4 遺伝子 A が直接的に遺伝子 C の発現量を増加させているかは分からない。

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