

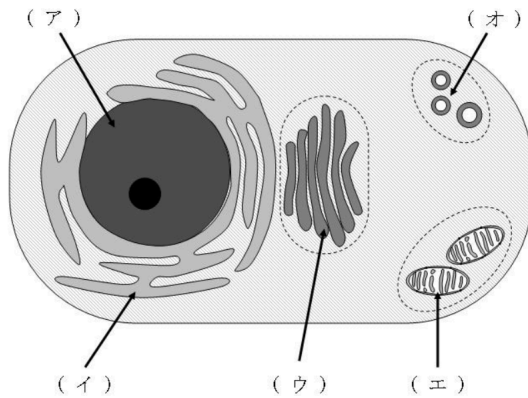
2019年度
日本バイオインフォマティクス学会 (JSBi)
バイオインフォマティクス技術者認定試験
試験問題

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。

問 1

次に示す図は、真核生物の細胞を模式的に描いたものである。（ア）から（オ）で示している細胞内小器官のうち、ミトコンドリア、小胞体、核に相当するものはどれか。適切な組み合わせを選択肢の中から一つ選べ。



選択肢	ミトコンドリア	小胞体	核
1	ウ	ア	エ
2	イ	オ	ア
3	オ	ウ	イ
4	エ	イ	ア

問 2

核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 真核生物の細胞内小器官であり、外膜と内膜の 2 層の脂質二重膜をもっている。
- 2 核膜に存在する核膜孔を通してタンパク質や mRNA など高分子が選択的に輸送される。
- 3 核内のクロマチンは、ヒストンタンパク質に DNA が巻き取られた構造をとる。
- 4 球状の核小体では RNA ポリメラーゼによって mRNA が合成される。

問 3

細胞分裂、細胞周期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 体細胞分裂における細胞周期は G0 期、G1 期、S 期、G2 期、M 期からなる。
- 2 卵や精子などの生殖細胞は 2 回の細胞分裂を経て形成される。
- 3 体細胞分裂の G1 期において細胞に含まれる DNA の量は G2 期の倍である。
- 4 減数分裂時に起こる相同染色体間の組換えは遺伝的多様性を生み出す元となっている。

問 4

RNA の転写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 RNA ポリメラーゼは 4 種類のリボヌクレオシド三リン酸を基質として RNA を合成する。
- 2 真核生物の RNA ポリメラーゼは 3 種類あり、その中の 1 種である RNA pol I は rRNA の合成に関与している。
- 3 原核生物の mRNA の 3' 末端にはアデニンの連続した配列であるポリ A 鎖が付加される。
- 4 ステロイドホルモン受容体の一種であるエストロゲン受容体は転写因子としての機能をもつ。

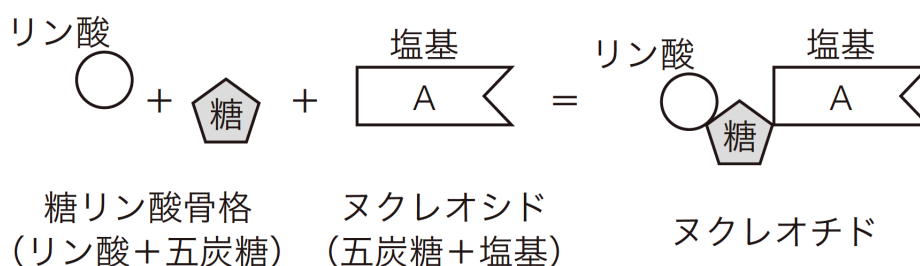
問 5

タンパク質の生合成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 成熟 mRNA にはタンパク質のアミノ酸配列をコードする翻訳領域とコードしない非翻訳領域が存在する。
- 2 リボソームは 1 種類の rRNA とタンパク質の複合体で大小 2 つのサブユニットからなる。
- 3 ミトコンドリア DNA 上の遺伝子にコードされたタンパク質はミトコンドリア内のリボソームで合成される。
- 4 コドンは相補的なアンチコドンを持つ tRNA によって読みとられる。

問 6

ヌクレオチドは五炭糖の骨格に塩基とリン酸が結合した分子である。ヌクレオチドの構造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 ヌクレオチドは、五炭糖の 1 位の炭素に塩基、5 位の炭素にリン酸基が共有結合した分子である。
- 2 ヌクレオチドの五炭糖のヒドロキシ基(-OH)がリン酸基とジスルフィド結合することにより重合したものを核酸とよぶ。
- 3 リボヌクレオチドに含まれる五炭糖の 2 位の炭素にはヒドロキシ基(-OH)が共有結合している。
- 4 デオキシリボヌクレオチドに含まれる五炭糖の 2 位の炭素には水素原子(-H)が共有結合している。

問 7

タンパク質を構成する 20 種類の主なアミノ酸に関する次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 メチオニンとシステインは側鎖に硫黄原子を含む。
- 2 セリン、スレオニンおよびチロシンは側鎖にヒドロキシ基(-OH)を含む。
- 3 アスパラギン、アルギニン、グルタミン、トリプトファン、ヒスチジンおよびリシンは側鎖に窒素原子を含む。
- 4 生体内に含まれる 20 種類のアミノ酸は全て側鎖に炭素原子を含む。

問 8

膜タンパク質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 膜タンパク質のうちイオン輸送の機能を持つものは、一般に一回膜貫通型タンパク質である。
- 2 膜タンパク質の中には、翻訳後修飾で付加された糖脂質によって生体膜に付着しているものもある。
- 3 膜タンパク質には、膜貫通部分の二次構造に α ヘリックスを有するものも β シートを有するものもある。
- 4 膜タンパク質には、アミノ末端が細胞内に露出しているものも細胞外に露出しているものもある。

問 9

タンパク質の翻訳後修飾について述べた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 ペプチドホルモンの中には翻訳後に酵素で切断されて活性型となるものがある。
- 2 タンパク質への糖鎖修飾には N -グリコシル化とユビキチン化がある。
- 3 タンパク質へのリン酸化はキナーゼ酵素による反応である。
- 4 ヒストンのメチル化は、遺伝子発現の制御にかかわっている。

問 10

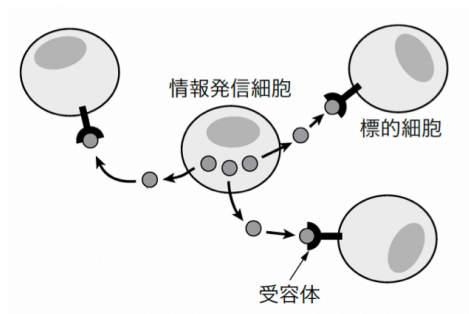
解糖系に関する以下の記述について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 解糖系によって 1 分子のグルコースは 2 分子のピルビン酸に代謝される。
- 2 解糖系によって 1 分子のグルコースが代謝される際に 2 分子の酸素が消費される。
- 3 解糖系によって 1 分子のグルコースが代謝される際に 2 分子の ATP が生成される。
- 4 解糖系によって 1 分子のグルコースが代謝される際に 2 分子の NADH が生成される。

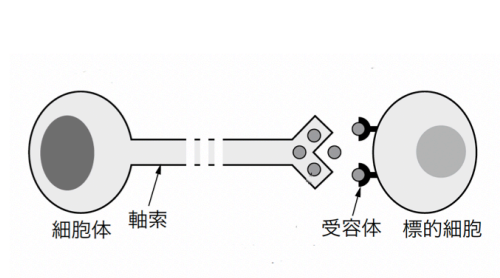
問 11

シグナル伝達は速さと距離に応じて、図（ A ）から（ D ）の 4 つに分類できる。それぞれの名称の組み合わせとして適切なものはどれか。選択肢の中から一つ選べ。

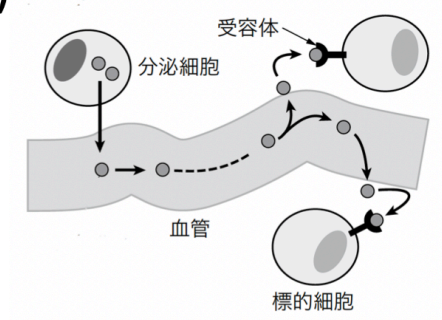
(A)



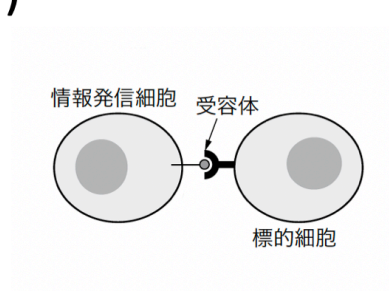
(B)



(C)



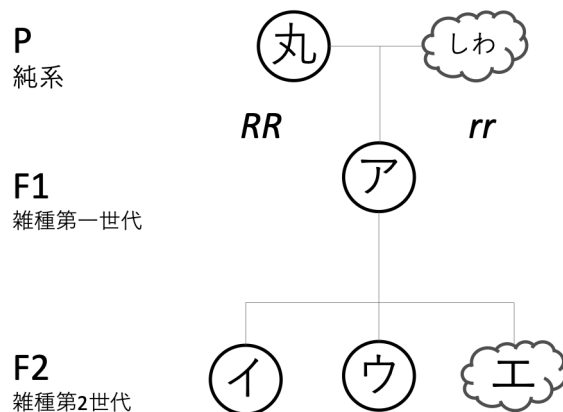
(D)



- | | | | | |
|---|-----------|--------|-----------|-----------|
| 1 | (A)内分泌型 | (B)接触型 | (C)神経型 | (D)パラクリン型 |
| 2 | (A)パラクリン型 | (B)神経型 | (C)内分泌型 | (D)接触型 |
| 3 | (A)神経型 | (B)接触型 | (C)内分泌型 | (D)パラクリン型 |
| 4 | (A)内分泌型 | (B)神経型 | (C)パラクリン型 | (D)接触型 |

問 12

エンドウの種子の形には「丸」と「しわ」の形質があり、丸い形質を示す遺伝子を R 、しわを r とすると、 R と r は対立遺伝子で R が r に対して顕性（優性）である。丸としわの純系をかけ合わせたときの F_1 の遺伝子型(ア)、 F_1 どうしをかけ合わせた F_2 の遺伝子型(イ)~(エ)に当てはまる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 | (ア) Rr | (イ) RR | (ウ) Rr | (エ) rr |
| 2 | (ア) RR | (イ) rr | (ウ) Rr | (エ) Rr |
| 3 | (ア) Rr | (イ) rr | (ウ) Rr | (エ) RR |
| 4 | (ア) RR | (イ) RR | (ウ) Rr | (エ) rr |

問 13

ゲノム塩基配列に関する次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- ヒトの細胞核のゲノムサイズ（約 30 億塩基対）は大腸菌のゲノムサイズ（約 460 万塩基対）の約 650 倍なのに対して、タンパク質をコードする遺伝子の推定数は 5~6 倍程度しか変わらない。
- ヒトの細胞核のゲノムには *Alu* とよばれる短鎖散在反復配列(SINE、short interspersed nuclear element)が約 100 万コピー存在する。
- ヒトの細胞の中には数百のミトコンドリアがあり、各ミトコンドリアにミトコンドリアゲノムが複数コピー存在するが、全て父親由来であることが分かっている。
- ヒトの Y 染色体は X 染色体の約 3 分の 1 のサイズであり、Y 染色体には偽常染色体領域(pseudoautosomal region)という X 染色体と相同の領域が存在する。

問 14

DNA の塩基の略号およびタンパク質のアミノ酸の 1 文字表記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 は塩基だとアデニンの略号、アミノ酸だとアラニンの 1 文字表記である。
- 2 C は塩基だとシトシンの略号、アミノ酸だとシステインの 1 文字表記である。
- 3 G は塩基だとグアニンの略号、アミノ酸だとグリセリンの 1 文字表記である。
- 4 T は塩基だとチミンの略号、アミノ酸だとスレオニンの 1 文字表記である。

問 15

ヒトのゲノムに関する、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 ヒトのゲノムには CNV と呼ばれるコピー数多型を示す部位が存在する。
- 2 ヒトのゲノムには SNP と呼ばれる 1 塩基多型を示す部位が存在する。
- 3 ヒトのゲノムには VNTR と呼ばれる繰り返し配列の多型を示す部位が存在する。
- 4 ヒトのゲノムには LINE と呼ばれる GC 含量の高い部位が遺伝子のプロモータ領域に存在する。

問 16

ポリメラーゼ連鎖反応 (PCR) 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 ごく微量の試料から DNA を増幅することが可能である。
- 2 反応液の組成を変化させることで変性、アニーリング、伸長の 3 つのステップを繰り返し、増幅を行う。
- 3 増幅反応過程の DNA 量を逐次検出することで、サンプル中の標的 DNA 量を定量することが出来る。
- 4 高熱環境に生息するバクテリアから単離された耐熱性の DNA ポリメラーゼがよく利用される。

問 17

次世代シーケンサを用いた解析手法の説明としてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 RNA-Seq は mRNA の塩基配列を大量に決定することでサンプル中に含まれる転写物を計測するトランスクリプトーム解析手法である。
- 2 ChIP-Seq はタンパク質を免疫沈降した際に同時に回収された DNA の塩基配列を決定することで、当該タンパク質と相互作用しているゲノム領域を同定するエピゲノム解析に使うことができる。
- 3 細菌の集団（細菌叢）から個々のクローンを単離したりせずに直接 DNA を精製し、その配列決定を行うことでその集団中の微生物の組成を明らかにする解析手法をメタゲノム解析という。
- 4 キャプチャプローブを用いてゲノム DNA から主にエキソン領域を濃縮・回収し、配列決定を行う手法をメタボローム解析という。

問 18

ゲノム編集技術の CRISPR/Cas9 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 CRISPR は原核生物における免疫機構に関係する反復配列である。
- 2 Cas9 は DNA の二本鎖切断(DSB:double strand break)を誘導するヌクレアーゼである。
- 3 CRISPR/Cas9 法は哺乳類ゲノムにしか利用できない手法である。
- 4 オフターゲット効果とは想定していない部位に変異が生じることである。

問 19

タンパク質の立体構造を決定する際に広く用いられている X 線結晶解析と NMR 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 NMR ではタンパク質の結晶化が不要なため、一般に NMR は X 線結晶解析よりも高分子量のタンパク質の立体構造決定に適している。
- 2 NMR では原子間の距離情報を収集し、原子間の距離情報を満足するように分子モデルを構築して立体構造を決定する。
- 3 NMR では溶液中のタンパク質を解析するため、一般に NMR は X 線結晶解析よりも動的に変化する立体構造の観察に適している。
- 4 X 線結晶解析では電子密度を求め、電子密度にアミノ酸残基モデルを当てはめて立体構造を決定する。

問 20

ヒトゲノム・遺伝子解析研究に関する倫理指針（平成 29 年 2 月 28 日一部改正）において定められている基本方針としてあげられている以下の事項（抜粋）のうち、不適切なものを選択肢の中から一つ選べ。

- 1 事前の十分な説明と自由意思による同意（インフォームド・コンセント）を基本とすること。
- 2 人類の知的基盤、健康及び福祉に貢献する社会的に有益な研究を基本とすること。
- 3 倫理審査委員会による事後の承認による研究の迅速性を確保すること。
- 4 研究の実施状況の第三者による実地調査及び研究結果の公表を通じた研究の透明性を確保すること。

問 21

整数 $w \in \{0, 1\}$ に対し、 $t(w)$ を、 $w=1$ のときに真理値「真」、 $w=0$ の時に真理値「偽」を返す関数とする。また、真理値 $v \in \{\text{「真」}, \text{「偽」}\}$ に対し、 $\text{int}(v)$ は、 $v=\text{「真」}$ のときに整数 1、 $v=\text{「偽」}$ のときに整数 0 を返す関数とする。また、以下の式で AND は論理積、OR は論理和、XOR は排他的論理和を表すものとする。

このとき、0 または 1 の値をとる 3 つの整数 x, y, z に対し、 $(x+y+z)$ を 2 で割った余りを表す式として正しいものを選択肢の中から一つ選べ。

- 1 $\text{int}((t(x) \text{ AND } t(y)) \text{ AND } t(z))$
- 2 $\text{int}((t(x) \text{ OR } t(y)) \text{ OR } t(z))$
- 3 $\text{int}((t(x) \text{ XOR } t(y)) \text{ XOR } t(z))$
- 4 $\text{int}((t(x) \text{ OR } t(y)) \text{ AND } t(z))$

問 22

以下の選択肢の中からもっとも不適切な記述を選択肢の中から一つ選べ。

- 1 投機的実行とは、プロセッサ等において、条件分岐の前に分岐を予測し分岐後の一部の計算処理を行うことによって、そのプロセッサ等の平均的性能を上げることを目指す技法のことをいう。
- 2 プロセッサにおけるパイプライン処理とは、1つの命令を複数要素に分割し、流れ作業的に行うことによって、複数の命令を平行して処理し、プロセッサの資源を有効に使うことを目指す技法のことを言う。
- 3 FPGA (Field Programmable Gate Array) は、製造後に任意の論理機能を（回路サイズが許す範囲内で）実装・変更できるようにした集積回路で、一般的にはハードウェア記述言語で構成を指定する。
- 4 GPGPU (General-Purpose computing on Graphics Processing Units) とは、クロックを用いずすべての計算を非同期で行うことによって高速化したプロセッサの総称である。

問 23

セキュリティに関する以下の説明文において、(a)～(c)に入る語の組み合わせとしてもっとも適切なものを選択肢の中から一つ選べ。

RSA 暗号や ElGamal 暗号に代表される(a)においては、(b)を用いて暗号化された文書は(c)を用いて復号することが可能であるが、(b)のみを用いて復号することは極めて困難であるとされる。

- | | | | |
|---|-----------|---------|---------|
| 1 | (a) 共通鍵暗号 | (b) 公開鍵 | (c) 秘密鍵 |
| 2 | (a) 公開鍵暗号 | (b) 公開鍵 | (c) 秘密鍵 |
| 3 | (a) 共通鍵暗号 | (b) 秘密鍵 | (c) 公開鍵 |
| 4 | (a) 公開鍵暗号 | (b) 秘密鍵 | (c) 公開鍵 |

問 24

自然数 x をキーとするデータを、ハッシュ関数 $h(x) = x \bmod n$ を用いたハッシュ表によって管理する。 n はハッシュ表の大きさであり、 $x \bmod n$ は x を n で割ったときの剰余である。このとき、キー x_1 とキー x_2 が衝突する条件を選択肢の中から一つ選べ。

- 1 $x_1 + x_2$ が n の倍数
- 2 $x_1 - x_2$ が n の倍数
- 3 n が $x_1 + x_2$ の倍数
- 4 n が $x_1 - x_2$ の倍数

問 25

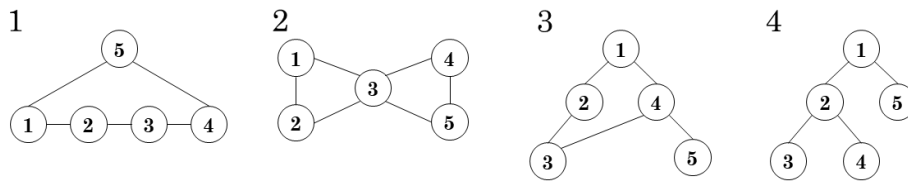
以下の隣接行列で表現されるグラフを以下の 4 つのグラフより選べ。

なお、隣接行列の i 行 j 列目の成分は、ノード i とノード j を結ぶエッジがある場合に 1、ない場合に 0 をとる。

隣接行列：

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(選択肢)



問 26

以下に示した擬似コード INSERTSORT(A) は、長さ n の相異なる数字からなる数列 A を昇順にソートする挿入ソートとよばれるアルゴリズムを示したものである。

INSERTSORT(A)

```

for each  $i \leftarrow 2$  to  $n$  {
   $x \leftarrow A[i]$ 
   $j \leftarrow i - 1$ 
  while  $j \geq 1$  and  $x < A[j]$  {
     $A[j + 1] \leftarrow A[j]$ 
     $j \leftarrow j - 1$ 
  }
   $A[j + 1] \leftarrow x$ 
}
```

このとき、すでに昇順にソートされている長さ n の相異なる数字からなる数列 A に対し、このアルゴリズム INSERTSORT(A) を再度適用することを考える。その際の計算量についての以下の記述のうち、もっとも適切なものを選択肢の中から一つ選べ。

- 1 常に $\Theta(n)$ 時間で終了する。
- 2 常に $\Theta(n \log n)$ 時間で終了する。
- 3 常に $\Theta(n^2)$ 時間で終了する。
- 4 アルゴリズムが停止するとは限らない。

問 27

入力サイズ n がある定数以下ならば定数時間の処理を行い、それ以外の場合はサイズ $n/2$ の問題二つを再帰的に解いてから $\Theta(n)$ 時間の後処理を行う (n が奇数の場合はサイズ $\lfloor n/2 \rfloor$ と $\lceil n/2 \rceil$ の問題を再帰的に解く) 枠組によって問題を解くアルゴリズムを考える。そのようなアルゴリズムの例としては、マージソートが挙げられる。そのようなアルゴリズムの実行時間としてもっとも適当なものを選択肢の中から一つ選べ。

- 1 $\Theta(n)$
- 2 $\Theta(n \log n)$
- 3 $\Theta(n^2)$
- 4 $\Theta(2^n)$

問 28

相異なる $20 \cdot n$ 個 (n は正の整数) の整数からなるリスト L から互いに独立かつ一様にランダムに 1 つ値を選ぶことを m 回繰り返し、得られた m 個の値のうち最大の値を出力することを考える。整数は毎回ランダムに選ぶため、得られた m 個の値には重複がある可能性がある。すると、出力する値がリスト L の整数のうち 上位 10 個の値に含まれない 確率は

$$\left(1 - \frac{10}{20 \cdot n}\right)^m$$

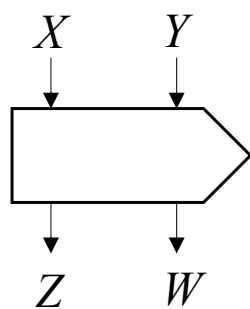
と計算できる。

このとき、出力する値がリスト L の整数のうち 上位 n 個の値に含まれる 確率が 99%以上となるように m の値を設定したい。ただし、 m は可能な限り小さい値を設定するものとする。このとき、 m と n の関係について、もっとも適切な説明を選択肢の中から一つ選べ。

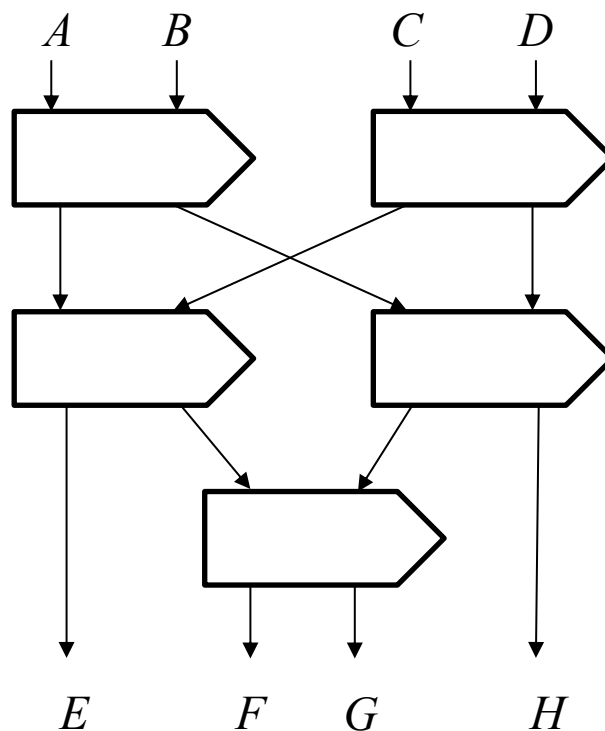
- 1 n を大きくしても、設定すべき m の値は変わらない
- 2 n を大きくすると、設定すべき m の値も大きくする必要がある
- 3 n を大きくすると、設定すべき m の値は小さくする必要がある
- 4 n を大きくしたとき、設定すべき m の値は大きくなることも小さくなることもある

問 29

図 1 は、相異なる 2 つの実数値 X , Y を入力として受け取り、そのうち小さい方の数値を Z 、大きい方の数値を W として出力する機械を表している。この機械を 5 つ用いて図 2 のような、4 つの相異なる実数値 A, B, C, D を入力とし、実数値 E, F, G, H を出力とする機械を作成した。図 2 の機械の出力の、 E, F, G, H の間の大小関係で間違っているものを選択肢の中から一つ選べ。



(図 1)

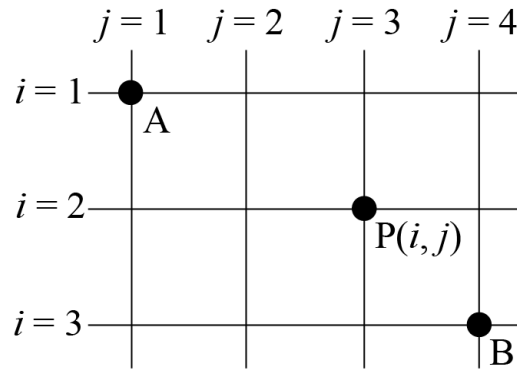


(図 2)

- 1 $E < F$
- 2 $F < G$
- 3 $G < H$
- 4 $H < E$

(問 30・31)

以下の図は、ある都市の道路の様子を模式的に表したものである。ここで、交差点 A から交差点 B への最短経路の数を数えることを考える。



問 30

以下の文章で、(a) に入る数字として正しいものを選択肢の中から一つ選べ。

組み合わせ論を使うと、どの最短経路も 2 つの垂直方向の部分経路と 3 つの水平方向の部分経路から成り立つことから、最短経路の数は (a) と計算できる。

- 1 60
- 2 12
- 3 10
- 4 5

問 31

以下の文章で、(b) に入る数式として正しいものを選択肢の中から一つ選べ。

交差点 A から水平方向の道 i ($1 \leq i \leq 3$) と垂直方向の道 j ($1 \leq j \leq 4$) が交わる交差点へ行く最短経路の数を $P(i, j)$ とする。 $P(1, j) = 1$ ($1 \leq j \leq 4$), $P(i, 1) = 1$ ($1 \leq i \leq 3$) を初期条件とすると、 $1 < i \leq 3$, $1 < j \leq 4$ のとき (b) が成り立つ。つまり、道 $i = 1$ から始めて、各道 i を左から右に計算していくと、最終的に最短経路の数は $P(3, 4)$ で与えられる。このように、部分問題の解から元の問題の解を求める手法を動的計画法と呼ぶ。

- 1 $P(i, j) = P(i - 1, j) + P(i, j - 1)$
- 2 $P(i, j) = P(i - 1, j) + P(i + 1, j)$
- 3 $P(i, j) = P(i, j - 1) + P(i, j + 1)$
- 4 $P(i, j) = P(i - 1, j - 1) + 1$

問 32

XML (Extensible Markup Language) 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 XML データベースを検索するための標準言語として SQL が知られている。
- 2 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。
- 3 HTML ではあらかじめ定義済みのタグを用いるのに対し、XML では、ユーザが新たなタグを定義して、文書構造を記述できる。
- 4 PDBML は XML 形式による PDB (Protein Data Bank) データフォーマットである。

問 33

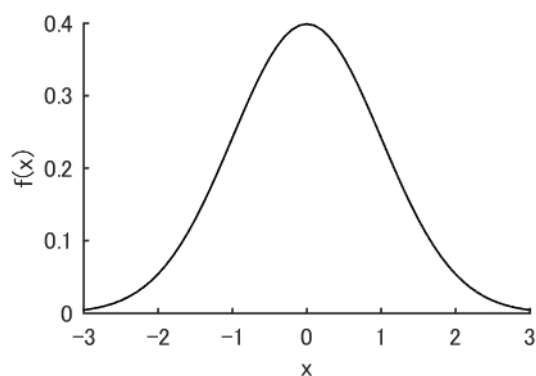
地域別に分かれている同じ構造の 3 つの商品表、「東京商品」、「札幌商品」、「福岡商品」がある。次の SQL 文と同等の結果が得られる関係代数式を選択肢の中から一つ選べ。ここで、3 つの商品表の主キーは「商品番号」である。

```
SELECT * FROM 札幌商品
WHERE 商品番号 NOT IN (SELECT 商品番号 FROM 東京商品)
UNION
SELECT * FROM 福岡商品
WHERE 商品番号 NOT IN (SELECT 商品番号 FROM 東京商品)
```

- 1 (札幌商品 $\cup$ 福岡商品) $-$ 東京商品
- 2 (札幌商品 $\cap$ 福岡商品) $-$ 東京商品
- 3 東京商品 $-$ (札幌商品 $\cup$ 福岡商品)
- 4 東京商品 $-$ (札幌商品 $\cap$ 福岡商品)

問 34

確率変数 X が平均 0、分散 v を持つ正規分布に従うとする。次の図は、平均 0、分散 1 を持つ標準正規分布 $N(0,1)$ の密度関数 $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ をプロットしたものである。



確率変数 X の実現値が非負になる確率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 $P(0 \leq X) = 0$
- 2 $P(0 \leq X) = 0.5$
- 3 $P(0 \leq X) = 1$
- 4 $P(0 \leq X) = 2$

問 35

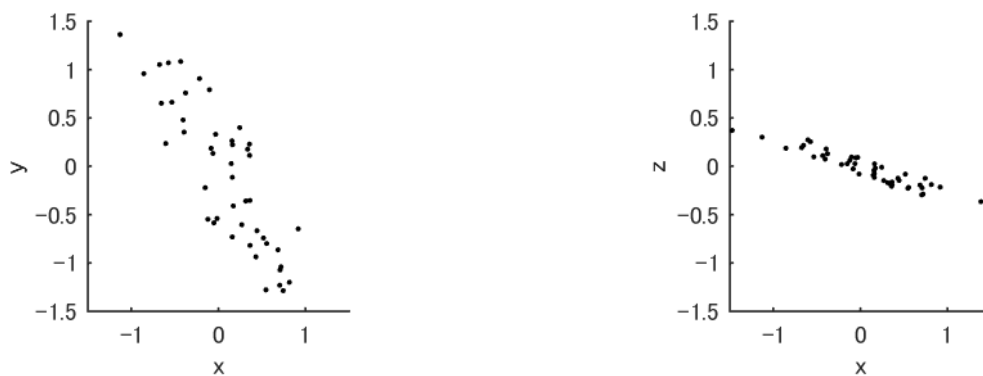
標準正規分布 $N(0,1)$ の密度関数 $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ の累積密度関数を $F(x) = \int_{-\infty}^x f(t)dt$ とする。

標準正規分布の実現値が区間 $[-1, +2]$ に落ちる確率を表す数式として適切なものを選択肢の中から一つ選べ。

- 1 $F(+2) - F(-1)$
- 2 $F(-1) - F(+2)$
- 3 $F(-1) / F(+2)$
- 4 $F(+2) / F(-1)$

問 36

ある大きさ 50 の標本 $(x_1, y_1, z_1), \dots, (x_{50}, y_{50}, z_{50})$ に対して、左図は (x_i, y_i) の散布図をプロットしたもの、右図は (x_i, z_i) の散布図をプロットしたものである。 (x_i, y_i) のピアソン相関係数 r_{xy} 、および (x_i, z_i) のピアソン相関係数 r_{xz} の大小関係を表す不等式としてもっとも適切なものを選択肢の中から一つ選べ。



なお、 (x_i, y_i) のピアソン相関係数 r_{xy} は次のように定義される。

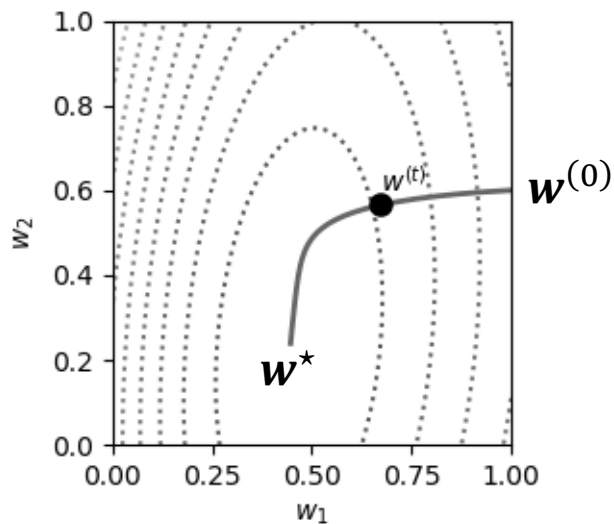
$$r_{xy} = \frac{\sum_{i=1}^{50} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{50} (x_i - \bar{x})^2 \sum_{i=1}^{50} (y_i - \bar{y})^2}}$$

ただし、 $\bar{x}$ は $x_1, \dots, x_{50}$ の標本平均、 $\bar{y}$ は $y_1, \dots, y_{50}$ の標本平均である。

- 1 $-1 < r_{xz} < 0 < r_{xy} < +1$
- 2 $-1 < r_{xz} < r_{xy} < 0 < +1$
- 3 $-1 < 0 < r_{xy} < r_{xz} < +1$
- 4 $-1 < r_{xy} < 0 < r_{xz} < +1$

問 37

誤差関数 $E(\mathbf{w})$ を最小にするパラメータ $\mathbf{w} = (w_1, w_2)$ の値を最急降下法によって求める機械学習問題を考える。図中の点線は等高線を示している。実線は、初期値 $\mathbf{w}^{(0)} = (1.00, 0.60)$ から最急降下法を開始し、最小解 $\mathbf{w}^* = (0.45, 0.24)$ に到達したときの軌跡を表している。



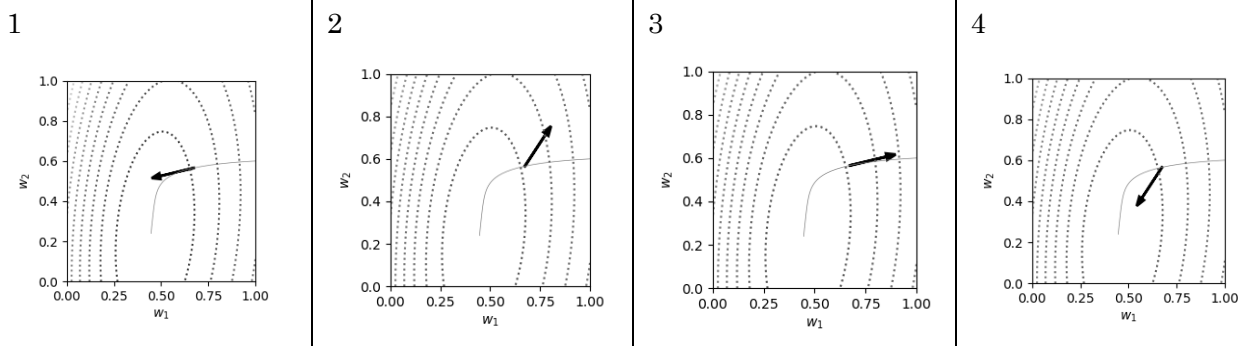
最急降下法では、時刻 t において、

$$\mathbf{w}^{(t)} = \mathbf{w}^{(t-1)} - \eta \nabla E(\mathbf{w}^{(t-1)})$$

によって、パラメータの値を更新する。ただし、 $\mathbf{w}^{(t-1)}$ は時刻 $t-1$ におけるパラメータの値、 η は正の定数、 $\nabla E(\mathbf{w}^{(t-1)})$ は誤差関数の勾配を表す。

ある時刻 t におけるパラメータ値が $\mathbf{w}^{(t)} = (0.67, 0.57)$ であった。誤差関数の勾配 $\nabla E(\mathbf{w}^{(t-1)})$ としてもっとも適切な方向を選択肢の中から一つ選べ。

(選択肢)



問 38

線形識別器に関する以下の説明文の中で (a) および (b) に入れる語としてもっとも適切なものを選択肢の中から一つ選べ。なお、説明文中のベクトルは縦ベクトルであるものとする。

次のような線形識別器を考える： d 次元ベクトル $\mathbf{w}$ とスカラー b をパラメータに持ち、入力ベクトル $\mathbf{x} \in R^d$ に対して、 $\mathbf{w}^T \mathbf{x} + b > 0$ なら陽性、 $\mathbf{w}^T \mathbf{x} + b < 0$ なら陰性、と予測する。この識別器の識別境界 $\{\mathbf{x} \in R^d | \mathbf{w}^T \mathbf{x} + b = 0\}$ は d 次元空間における (a) をなす。2 個の入力ベクトル $\mathbf{x}_a, \mathbf{x}_b \in R^d$ がいずれも陰性と予測されるとき、それらの中点に位置する入力ベクトル $\mathbf{x}_c$ (すなわち, $\mathbf{x}_c = (\mathbf{x}_a + \mathbf{x}_b)/2$) は (b) と予測される。

- | | | |
|---|---------|--------|
| 1 | (a) 超球 | (b) 陽性 |
| 2 | (a) 超球 | (b) 陰性 |
| 3 | (a) 超平面 | (b) 陽性 |
| 4 | (a) 超平面 | (b) 陰性 |

問 39

次の選択肢の中からクラスタリングを行う手法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 隠れマルコフモデル (Hidden Markov Model)
- 2 k -平均法 (k -means method)
- 3 自己組織化マップ (Self-Organizing Map)
- 4 UPGMA (Unweighted Pair Group Method with Arithmetic mean)

問 40

次に示した説明文の中で、予測手法の性能評価の際に行われるクロスバリデーション法の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 予測手法が未知データにも対応できるかを検査する目的で行われる。
- 2 クロスバリデーションにおいては、学習器に対し訓練を行いその結果を用いてテストを行うことを訓練データ、テストデータを変えながら複数回の検証が行われるが、ある回の検証において訓練データとして用いられたデータが、別の回の検証においてはテストデータとして用いられることがある。
- 3 leave-one-out 法は、データのうち 1 個を除いたデータを用いて訓練を行い、残る 1 個を用いてテストを行うことを、すべてのそのような訓練データ・テストデータの組み合わせについて行う。
- 4 n -fold 法は、データの $1/n$ を訓練に用い、残りをテストに用いる方法である。

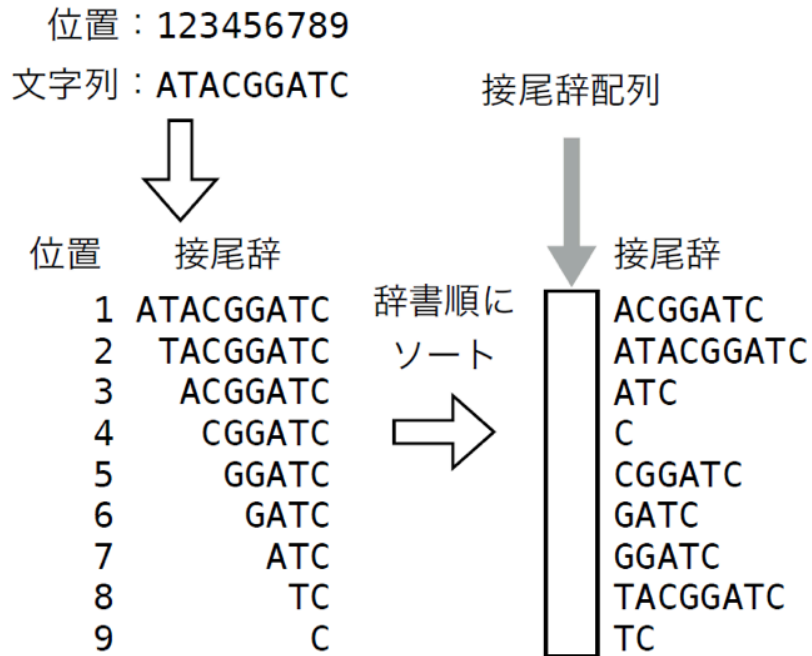
問 41

ショットガンメタゲノム解析では、環境から抽出した様々な細菌ゲノム由来の DNA 配列全体をまとめてシークエンスし、その全体像を得ることを目的とする。しかし、ショートリードの次世代シークエンスを用いて、複数の環境サンプルより得られた大量のリードを混ぜてアセンブルしても、完全長の細菌ゲノム配列を得ることは難しく、たいていの場合、短いコンティグ配列が多く得られるにとどまる。そこで、このコンティグ配列を更にクラスタリングし、同一菌種ごとにコンティグを分類する Binning という解析がよく行われる。Binning の際に考慮される方策として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 $k=4$ 程度の短い k -mer の頻度が似たコンティグを同一のクラスターに分類する。
- 2 同じ単一コピー遺伝子のオーソログを含む異なるコンティグが、同一のクラスターとならないように分類する。ただし「単一コピー遺伝子」とは、ほとんどの細菌ゲノムに 1 コピー存在する遺伝子のことである。
- 3 Clusters of Orthologous Groups (COGs) データベースなどを用いてコンティグに含まれる遺伝子に機能カテゴリを割り当て、機能カテゴリの割合が似たコンティグを、同一のクラスターに分類する。
- 4 各コンティグに、環境サンプル($s_1, s_2, \dots$)ごとに得られたリードをマップしてその被覆度(coverage) ($c_1, c_2, \dots$)を算出し、($c_1, c_2, \dots$)の値の組が似たコンティグを同一のクラスターに分類する。

問 42

下図では、配列 ATACGGATC を接尾辞配列へ変換する過程を示している。空欄の接尾辞配列としてもっとも適したものを選択肢の中から一つ選べ。



- 1 7, 1, 3, 9, 6, 4, 2, 5, 8
- 2 3, 1, 7, 9, 6, 4, 5, 2, 8
- 3 7, 1, 3, 9, 4, 6, 2, 5, 8
- 4 3, 1, 7, 9, 4, 6, 5, 2, 8

問 43

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 のエントリーである。ここから読み取れる情報として不適切なものを選択肢の中から一つ選べ。

- 1 DNA 塩基配列の長さは 191 bp である。
- 2 DEFINITION に「恐竜ゲノム DNA と推定される部分配列」と記載されている。
- 3 「恐竜の卵の化石から増幅された DNA 配列の再解析」に関する学術論文が出版されており、文献 DB (PUBMED) に登録されている。
- 4 ORIGIN に、この配列の先頭が複製開始点であることが記載されている。

```

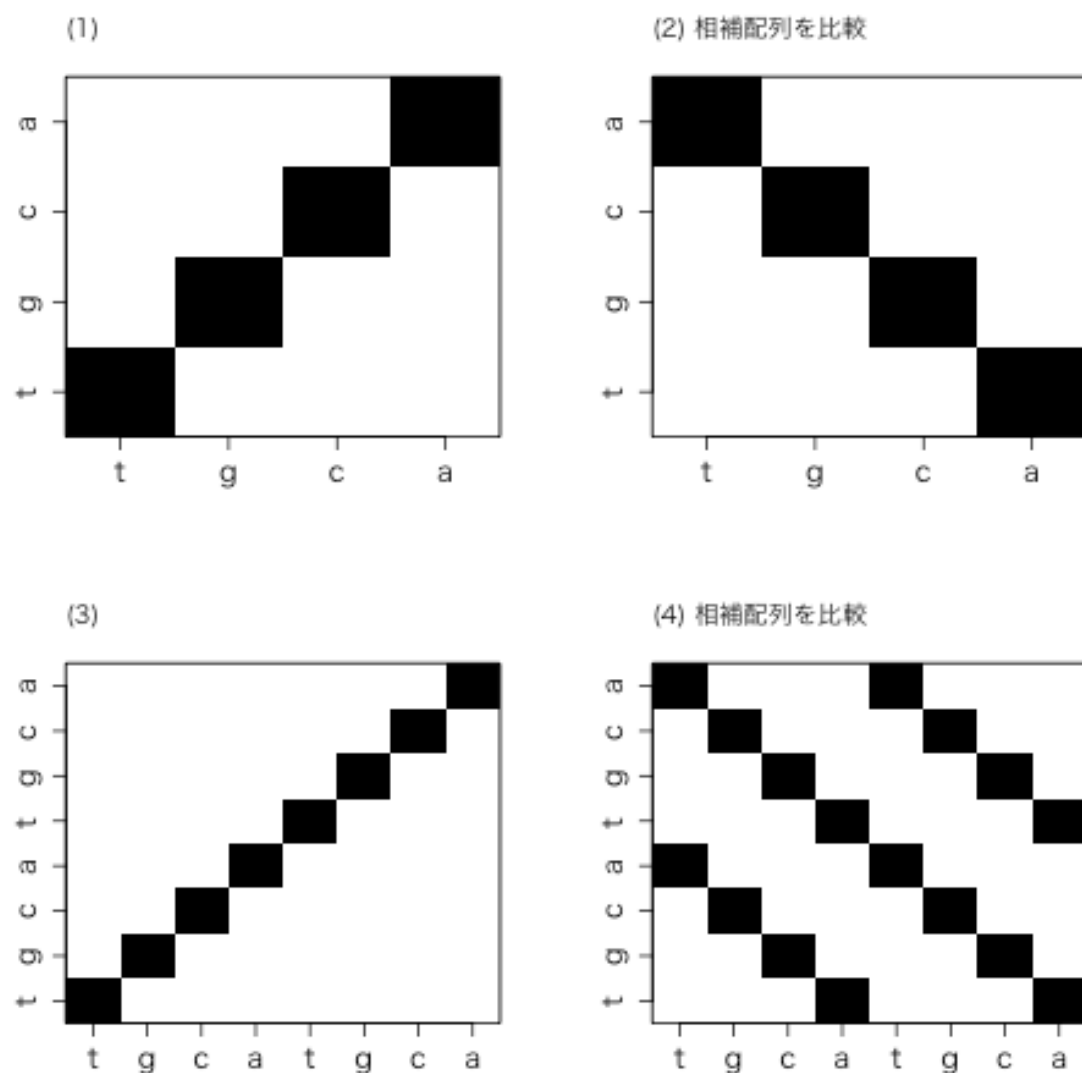
LOCUS      XXU41319                      191 bp    DNA        linear    UNA 11-JUN-1997
DEFINITION Putative dinosaur genomic DNA, partial sequence.
ACCESSION  U41319
VERSION    U41319.1
KEYWORDS   .
SOURCE     unidentified
  ORGANISM unidentified
            unclassified sequences.
REFERENCE  1 (bases 1 to 191)
  AUTHORS  Li,Y., An,C.-C., Zhu,Y.-X., Zhang,Y., Liu,Y.-F., Qu,L., You,L.-T.,
            Liang,X.-W., Li,X.-H., Qu,L.-J., Zhou,Z.-Q. and Chen,Z.-L.
  TITLE    DNA isolation and sequence analysis of dinosaur DNA from Cretaceous
            dinosaur egg in Xixia Henan, China
  JOURNAL  Acta Scientiarum Naturalium Universitatis Pekinensis 31, 148-152
            (1995)
REFERENCE  2 (bases 1 to 191)
  AUTHORS  Wang,H.L., Yan,Z.Y. and Jin,D.Y.
  TITLE    Reanalysis of published DNA sequence amplified from cretaceous
            dinosaur egg fossil
  JOURNAL  Mol. Biol. Evol. 14 (5), 589-591 (1997)
  PUBMED   9159936
REFERENCE  3 (bases 1 to 191)
  AUTHORS  Jin,D.-Y.
  TITLE    Direct Submission
  JOURNAL  Submitted (26-NOV-1995) Dong-Yan Jin, LMM/Molecular Virology
            Section, NIAID/NIH, 9000 Rockville Pike, Bethesda, MD 20892-0460,
            USA
FEATURES   Location/Qualifiers
  source    1..191
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
            /note="The sequence was originally described as dinosaur
            DNA but was found to share no significant homology with
            any sequence in the databases. There is no strong evidence
            to support the authenticity of its ancient origin."
ORIGIN
  1 ctatttcctta attaatgtct acatggctat ttttaatggtt attactgttt gtcactataa
  61 aaaaacgctc atttgagaca atactgacat taactgcttc aacttctacg cacggaactt
  121 ttaattaaat tagcacagga atgttaaatt taatanacaa aaggttattt cgctgtatga
  181 taaaaaaaac c
//

```

問 44

ドットプロットは、比較したい2つのDNA塩基配列をそれぞれ横軸（左から右へ）と縦軸（下から上へ）に配置し、一致した点にドットをプロットしていくというもので、同じ配列がある領域は対角線上にドットが並ぶ領域として観察される。以下に示すDNA塩基配列間のドットプロット解析結果のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 配列"tgca"（横軸）と配列"tgca"（縦軸）のドットプロット。
- 2 配列"tgca"（横軸）と配列"tgca"（縦軸）の相補配列についてのドットプロット。
- 3 配列"tgcatagca"（横軸）と配列"tgcatagca"（縦軸）のドットプロット。
- 4 配列"tgcatagca"（横軸）と配列"tgcatagca"（縦軸）の相補配列についてのドットプロット。



問 45

2本のDNA配列間の最適なアラインメントを決めるために、アラインメントの類似性スコアを定義し、これを最大化することを考える。簡単なスコアとしては、一致(match)、不一致(mismatch)、ギャップ(gap)の3状態で定義するもので、例えば塩基の一致を+3、不一致を-1、ギャップペナルティを2とする。この場合、以下の配列アラインメントの類似性スコアの値を選択肢の中から一つ選べ。

GAATTC

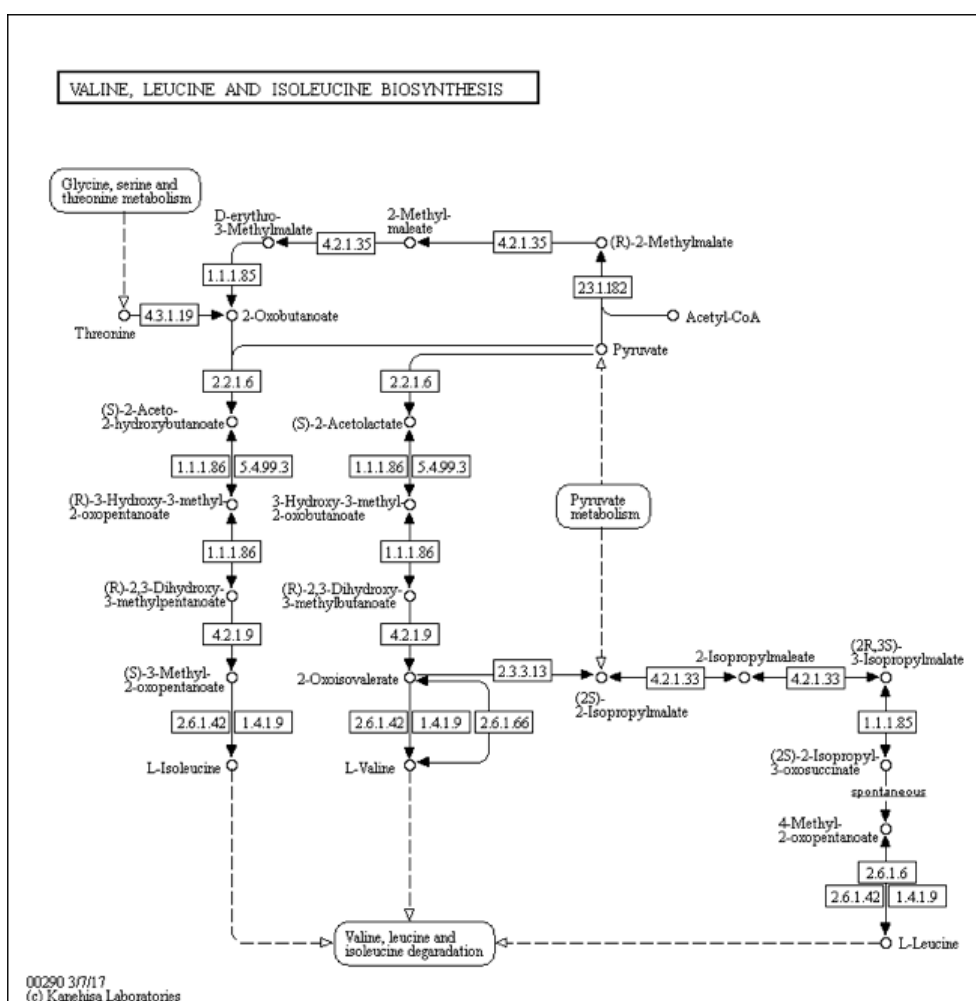
GA-TTA

- 1 7
- 2 9
- 3 11
- 4 12

問 46

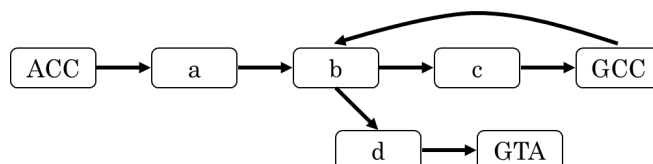
下記の図は KEGG のパスウェイマップ(Valine, leucine and isoleucine biosynthesis)である。このパスウェイマップに関して述べた次の記述のうち、不適切なものを選択肢の中から一つ選べ。

- 1 図中の○は、代謝物質を意味している。
- 2 図中の四角は、代謝の過程の中間生産物を意味し、中の数字は対応する酵素番号(EC number)である。
- 3 化合物によっては、酵素タンパク質の反応を経ずに別の化合物になることもある。
- 4 ゲノムアノテーションにより同定された酵素タンパク質をこのマップ上に表示することにより、その生物ではたらいっている代謝経路を推定することができる。



問 47

ACCCGCCGTA という塩基配列に対して 3-mer をノードとするド・ブラン(de Bruijn)グラフを得た。グラフ中の空白内の各番号に入る 3-mer の正しい組み合わせを選択肢の中から一つ選べ。



- 1 (a) CGC (b) CTG (c) CGC (d) CGC
- 2 (a) CCC (b) CCG (c) CGT (d) CGC
- 3 (a) CCC (b) CCG (c) CGC (d) CGT
- 4 (a) CGC (b) CTG (c) CGC (d) CGT

問 48

次世代シーケンサから得られた RNA-Seq のデータを用いた解析として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ゲノム配列中の繰り返し配列の検出
- 2 遺伝子配列に対するマッピングによる遺伝子発現量解析
- 3 ゲノム配列からの新規遺伝子の発見
- 4 *de novo* アセンブリに基づく CDS 配列の構築

問 49

長さ 30 塩基からなる RNA の配列があり、各塩基に順に 1 から 30 までの番号がつけられている。その二次構造は、以下の 9 つの塩基対によって形成されているとする。

2-29	9-22
3-28	10-21
4-27	13-20
5-26	14-19
8-23	

このとき、この二次構造について述べた以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 塩基 2 から 5 まではステムを形成している。
- 2 塩基 11 から 12 まではバルジループを形成している。
- 3 塩基 15 から 18 まではヘアピンループを形成している。
- 4 塩基 20 から 21 まではシュードノットを形成している。

問 50

アミノ酸置換行列 PAM 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 様々なアミノ酸の物理化学的指標を組み合わせることで評価することにより類似度が定義されており、その組合せ方によって複数の行列が存在する。
- 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傾向がある。
- 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 PAM120 と PAM250 では、PAM250 の方が進化的に遠い配列を比較するのに向いている。

(問 51・問 52)

以下の表は、ある生物のゲノムにおけるダイマー（2 連塩基）の出現頻度を表している（たとえば、ダイマーAC の出現頻度は 0.05、CA の出現頻度は 0.07 である）。これについて、以下の問いに答えよ。

		2 塩基目			
		A	C	G	T
1 塩 基 目	A	0.1	0.05	0.07	0.08
	C	0.07	0.05	0.01	0.07
	G	0.06	0.04	0.05	0.05
	T	0.07	0.06	0.07	0.1

問 51

このゲノム配列における G の出現頻度 $P(G)$ 、およびこのゲノム配列中に G が出現したときに次の塩基が T である条件付き確率 $P(T|G)$ の値の組合せとして、もっとも適切なものを選択肢の中から一つ選べ。

選択肢	$P(G)$	$P(T G)$
1	0.2	0.05
2	0.2	0.25
3	0.3	0.07
4	0.3	0.23

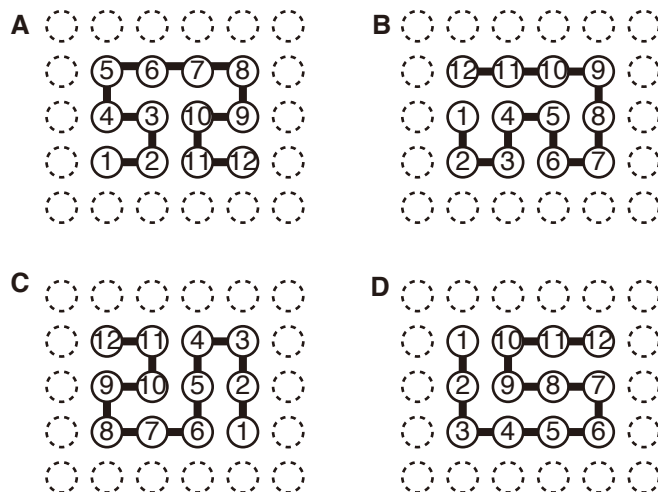
問 52

このダイマーの出現頻度に基づいて、ランダムな塩基配列を生成する確率モデルを考えよう。これは、ある部位における文字の出現確率が、その一つ前の部位の文字に依存して決まるという 1 次のマルコフ連鎖によって表現される。このモデルによると、長さ 3 の文字列 ABC が出現する確率は、 $P(ABC) = P(A)P(B|A)P(C|B)$ のようにして計算される。これを用いて計算すると、配列 GTC が出現する確率はいくらになるか。選択肢の中から一つ選べ。

- 1 0.003
- 2 0.005
- 3 0.01
- 4 0.015

(問 53・問 54)

タンパク質を下図のような二次元格子モデルで表現するとする。アミノ酸残基ひとつがひとつの格子点を占め、配列上隣り合うアミノ酸は必ず上下左右のいずれかに配置される。図中の番号は N 末端を 1 としたときのアミノ酸残基番号を示している。アミノ酸残基で占められていない格子点（破線の円）はタンパク質を取り囲む溶媒（水）に占められている。このモデルに基づいて、問 53 および問 54 に答えよ。



問 53

このモデルでは、あるアミノ酸の上下左右が別のアミノ酸で占められている場合、そのアミノ酸は溶媒に露出せず埋もれているものと見なせる。アミノ酸残基の疎水性の観点から、下記のアミノ酸配列を表すフォールド構造としてもっとも適切なものを上図 A~D より一つ選べ。

配列：(N 末端) -RSEKIDTQNVED- (C 末端)

- 1 図 A
- 2 図 B
- 3 図 C
- 4 図 D

問 54

タンパク質立体構造の予測において、マルチプルアラインメントのアミノ酸共進化情報を用いる手法が近年よく用いられるようになっている。ここで述べるアミノ酸の共進化とは、あるタンパク質のマルチプルアラインメントにおいて異なる座位でのアミノ酸変異が相関していることを意味する。共進化する座位同士は、構造的ないし機能的に関連性が高いことが期待される。配列上で4残基以上離れたアミノ酸残基同士での接触について共進化情報からの予測が有効であると仮定した場合、下図のマルチプルアラインメントから推定されるフォールドとしてもっとも適切なものを前ページの図A～Dより選べ。

```
QKLTGSTNAVDE
NKLTGSTNAVDE
NDLTGSTNAVKE
PDLTGSTNAVKE
PDVTGSTNALKE
PDVTGSTNALKD
...*****...
```

なお、*および.はアミノ酸が完全一致する座位と、そうでない座位を示している。また、各座位における主要なアミノ酸を太字で示した。

- 1 図 A
- 2 図 B
- 3 図 C
- 4 図 D

問 55

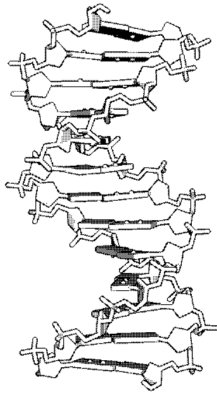
下図は、固定幅テキスト形式で記述された蛋白質立体構造データベース(Protein Data Bank; PDB)のエントリを表している。この形式 (PDB フォーマット) は、立体構造データのフォーマットとして従来広く用いられてきたが、様々な問題点があることから廃止され、別の形式 (mmCIF または PDBML フォーマット) へ移行することが決まっている (特に立体構造データの登録に関しては、PDB フォーマットによる提出は 2019 年 7 月より廃止された)。PDB フォーマットの問題点に関する記述としてもっとも不適切なものを選択肢の中から一つ選べ。

HEADER	HYDROLASE										23-SEP-02		1MUI		
TITLE	CRYSTAL STRUCTURE OF HIV-1 PROTEASE COMPLEXED WITH														
TITLE	2 LOPINAVIR.														
COMPND	MOL ID: 1;														
COMPND	2 MOLECULE: PROTEASE;														
COMPND	3 CHAIN: B, A;														
(中略)															
REMARK	3	AUTHORS		:	BRUNGER, ADAMS, CLORE, DELANO, GROS, GROSSE-										
REMARK	3			:	KUNSTLEVE, JIANG, KUSZEWSKI, NILGES, PANNU,										
REMARK	3			:	READ, RICE, SIMONSON, WARREN										
REMARK	3														
(中略)															
ATOM	1	N	PRO	B	1	-13.147	39.120	31.398	1.00	32.19					
ATOM	2	CA	PRO	B	1	-13.097	39.509	29.972	1.00	31.99					
ATOM	3	C	PRO	B	1	-13.789	38.466	29.102	1.00	31.07					
(中略)															
HETATM	1541	O1	AAB1	B	100	-13.116	16.186	33.516	0.50	14.18					
HETATM	1542	O1	BAB1	B	100	-7.426	19.577	22.183	0.50	14.18					
HETATM	1543	C10AAB1	B	100		-10.526	15.812	32.587	0.50	14.18					
HETATM	1544	C10BAB1	B	100		-8.270	17.106	23.134	0.50	14.18					
HETATM	1545	C11AAB1	B	100		-10.356	15.313	31.126	0.50	14.18					
(後略)															

- 1 メタデータ記述 (REMARK 行など) が十分に構造化されておらず、プログラムによる判読が難しい。
- 2 原子や残基などの数について上限値が低く、扱える構造のサイズが限られている。
- 3 図中矢印のように異なる意味の文字、数字同士が隣接してしまうため、プログラムによる判読が難しい。
- 4 リガンドの名称が数字およびアルファベット 3 文字に限定されているため、近い将来に全ての可能な名称が使用されてしまい、新たな名前が付けられなくなる。

問 56

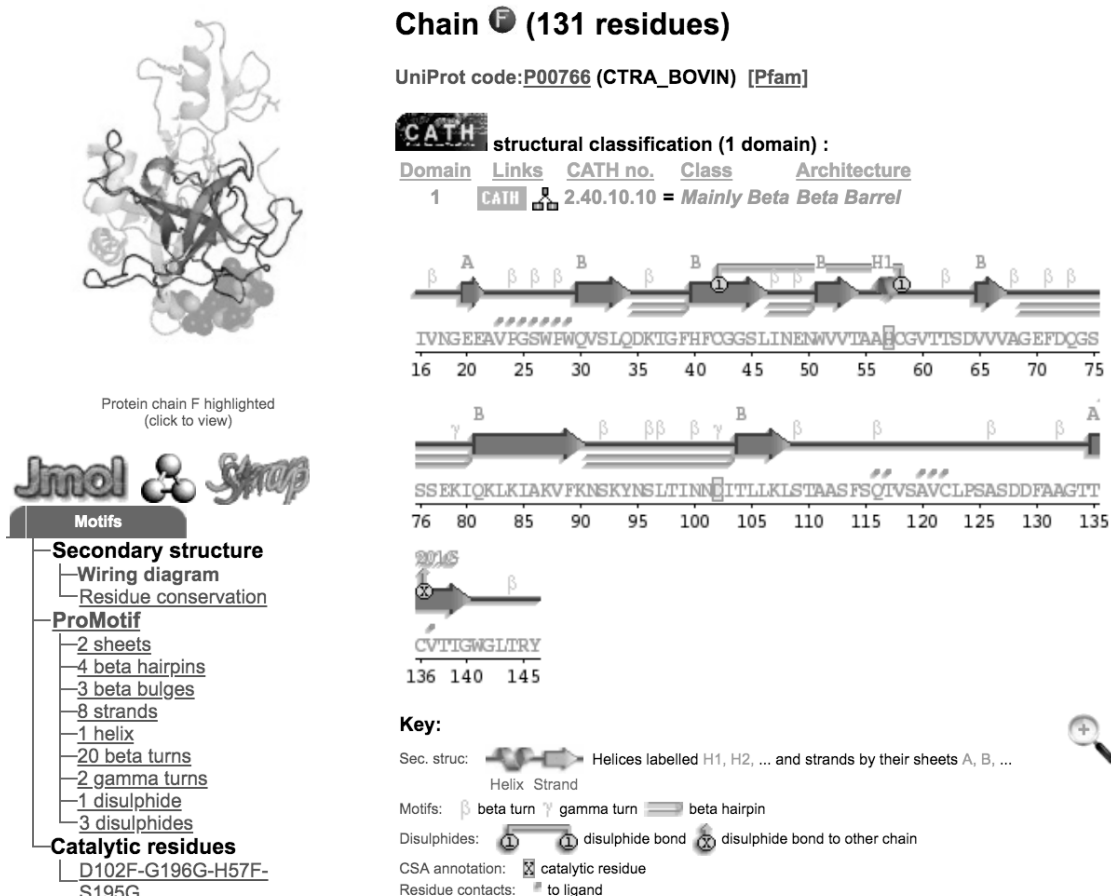
以下の図は DNA の立体構造を模式的に示したものである。この図の説明としてもっとも不適切なものを選択肢の中から一つ選べ。



- 1 B 型構造である。
- 2 上から副溝(minor groove)- 主溝(major groove)-副溝を正面に向けている。
- 3 全部で 13 塩基対の構造が示されている。
- 4 左巻き二重らせん構造である。

問 57

セリンプロテアーゼであるウシのキモトリプシン(PDB ID: 1cho)の F 鎖についての立体構造関連情報を、PDBsum データベースで閲覧した際に表示される図の一部を下に示した。左下図には、二次構造、モチーフ情報、触媒に重要な役割を果たすアミノ酸 (Catalytic residues、触媒残基) についての情報が記載されている。右図には、F 鎖の一次構造に沿って二次構造情報が示され、リガンドとの結合に関与するサイトの上にはドットが示されている。下図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 触媒残基のうち His 57 と Asp 102 は F 鎖上に、Ser 195 と Gly 196 は G 鎖上にある。
- 2 F 鎖に対応するアミノ酸配列に対し、UniProt データベースでは P00766 というアクセッション番号が割り当てられている。
- 3 リガンドと結合するタンパク質のアミノ酸は、一次構造上 25 番目のサイト近傍、120 番目のサイト近傍、そして 137 番目のサイト近傍に位置している。
- 4 CATH データベースにおける階層分類の Architecture では、Alpha-Beta Barrel に分類されている。

問 58

タンパク質の立体構造、機能についての次の記述のうち、収斂進化の記述としてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 進化的に独立起源の配列同士であっても、共通の機能的制約のために立体構造や機能が類似してることがある。
- 2 相同なタンパク質同士の進化的に対応しないサイトに共通の機能的制約がはたらくことで、共通の機能に関連する同じアミノ酸が生じることがある。例えば、シトクロム **c551** とミトコンドリアのシトクロム **c** では、一次構造上対応しない **Trp** とヘムプロピオン酸が水素結合を形成している。
- 3 種系統樹上遠縁な小コウモリ類と歯クジラ類という 2 系統で、エコロケーション（発した超音波の反射音を聞き取り周囲の様子を捉える）能力が独立に進化する際、聴覚に関与するプレスチン分子に複数のアミノ酸置換が共通して起こったと考えられている。
- 4 解析対象とする配列間の類似性の確認が困難な場合に、立体構造が類似していても機能が異なる場合は独立起源のタンパク質がたまたま同じフォールドをとっていると考え、相似とよぶ。

問 59

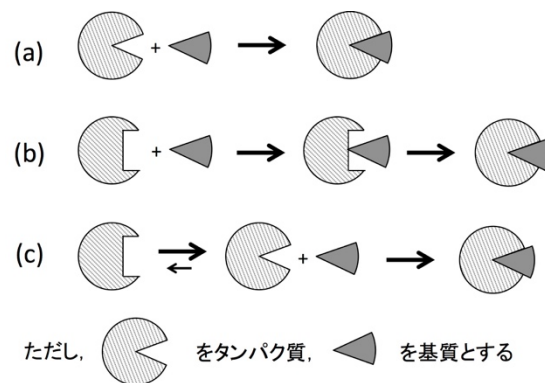
タンパク質間相互作用ネットワークを観察すると複数の相手と相互作用するハブタンパク質がみられる。ハブタンパク質についての記述としてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 天然変性領域を持つものが多いと考えられている。
- 2 膜タンパク質に分類されるものが多い。
- 3 ハブタンパク質の相互作用部位が同じでも、相互作用するタンパク質が同じであるとは限らない。
- 4 複数の相手と同時に相互作用することがある。

問 60

タンパク質と基質の結合を考える時に代表的な結合モデルとして「鍵と鍵穴 lock-and-key モデル」「induced-fit モデル」「pre-existing (selected-fit) モデル」が提唱されている。モデルの概念図(a)～(c)に対応するモデルの名称として、正しい組み合わせを選択肢の中から一つ選べ。

- | | | | |
|---|----------------------|----------------------|----------------------|
| 1 | (a) 鍵と鍵穴モデル | (b) pre-existing モデル | (c) induced-fit モデル |
| 2 | (a) induced-fit モデル | (b) pre-existing モデル | (c) 鍵と鍵穴モデル |
| 3 | (a) 鍵と鍵穴モデル | (b) induced-fit モデル | (c) pre-existing モデル |
| 4 | (a) pre-existing モデル | (b) 鍵と鍵穴モデル | (c) induced-fit モデル |



問 61

ショートリードシーケンサを活用してある個人の全ゲノムショットガンシーケンスを行い、標準ヒトゲノム配列にマッピングした。ある領域のマッピングを見てみると、以下の図に示すように、標準的なベースコールエラー率から期待されるよりも有意に多くの塩基置換や欠失が示唆された。更に、この領域にマップされたリード数は、そのほかの領域にマップされたリード数の平均よりも大きかった。このとき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。

標準ヒトゲノム配列	ATACGATTTGACGAATGACCATGACCATTGGCAATATACGA
	ATAC T ATT
	ATACGATT C GACGA
	ATAC C ATTTGACGAATGAC
	TACGATTTGACGAATGACCATGA
	TAC C ATTTGACGAA AGATC CTGA
	AC T ATTTGACGAATGACCATGAC
	C TATTTGACGA AGGA -CATGAC
	TT-GACGA AGGA -CATGACCA
	TTGACGAATGACCATGACCAT
	GACGA AGGA -CATGAC TAT -G
	GACGAATGACCATGACCATTG
	CGAATGACCATGACCATTGGC
	CGAA AGATC CTGACCATTGGC
	GAATGACCATGACCATT TGCA
	ATGACCATGACCATT TGCAAC
	ATGACCATGACCATTGGCAAT
	GGA -CATGAC TAT -GGCAATA
	CCATGACCATT AGCAATATAC
	CCATG CCT ATTGGCAATACAC
	CATGACCAT-GGCAATATACT T
	ATGACCATTGGCAATATACT A
	ATGACCATT AGCAATAGACTA
	CCT ATTGGCAATACACT A
	TTGGCAATATACT A
	CAATAGACT A
	ATAGACT A

ショートリード
(不一致は太字で示した)

- 1 この個人のゲノムでは、本領域に超優性の変異があるため多型が保持されていた。
- 2 ヒトゲノムの本領域にはコピー数多型があり、この個人のゲノムでは、コピー数の増加が生じていた。
- 3 標準ヒトゲノム配列領域では本領域と似た配列の領域が未決定であり、それら未決定領域に由来するリードが誤ってマップした。
- 4 サンプル調整中に他の生物由来の配列が混入し、その配列がこの領域と偶然一致した。

問 62

逆転写酵素の遺伝子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 レトロウイルスのゲノムに含まれる。
- 2 ヒトゲノム中の反復配列 LINE-1 に含まれる。
- 3 ヒトゲノム中の反復配列 Alu に含まれる。
- 4 ヒトゲノム中の LTR トランスポゾンに含まれる。

問 63

大腸菌ゲノムのある遺伝子が、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 この遺伝子の塩基の k -mer 頻度は、大腸菌の他の遺伝子と比較して大きく異なる。
- 2 この遺伝子がコードするタンパク質の構造モチーフは、大腸菌の他の遺伝子産物と比較して大きく異なる。
- 3 この遺伝子のコドン使用頻度は、大腸菌の他の遺伝子と比較して大きく異なる。
- 4 分子系統樹解析を行うと、この遺伝子は大腸菌の属するガンマプロテオバクテリア綱以外のグループの遺伝子群に分類される。

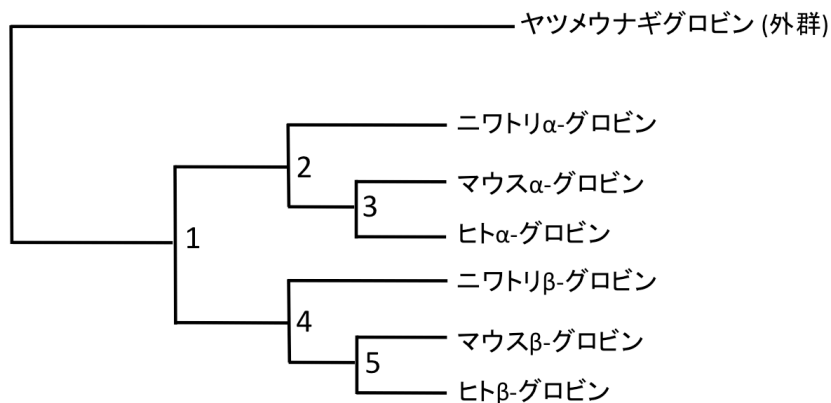
問 64

連鎖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 連鎖不平衡は、同一染色体上に存在する遺伝子座の間にのみ成立する。
- 2 連鎖不平衡とは、複数の遺伝子座の対立遺伝子の特定の組み合わせ頻度が有意に高くなる現象である。
- 3 対立遺伝子の特定の組み合わせをハプロタイプとよび、相同染色体間でのハプロタイプの組み合わせをディプロタイプとよぶ。
- 4 LOD スコアを用いた家系分析を行うことで、遺伝的疾患の原因遺伝子が連鎖地図上でマーカー遺伝子からどの程度離れた位置にあるかを推定できる。

問 65

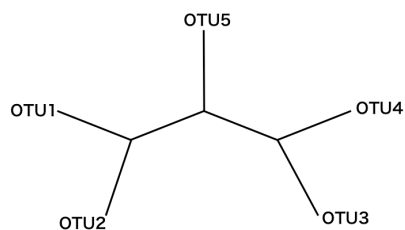
以下の図はグロビン族の遺伝子を用いて作成した分子系統樹であり、ノードに示した数字はそのノードから右に連なる遺伝子の最終共通祖先遺伝子を示す。この系統樹に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。



- 1 ニワトリ β -グロビンとヒト β -グロビンはパラログの関係にある。
- 2 マウス α -グロビンとヒト α -グロビンはオーソログの関係にある。
- 3 ノード 2 とノード 4 に位置する祖先配列はパラログの関係にある。
- 4 ノード 2 とノード 4 それぞれから右に連なる部分系統樹は、いずれも種系統樹と見なせる。

問 66

無根系統樹において可能な樹形の数、OTU (Operational Taxonomic Unit; 遺伝子や種など、系統樹を推定する際の操作単位) が 4 つのときは 3 通り、OTU が 5 つのときは 15 通りである(下図は OTU が 5 つの場合に可能な樹形の 1 例である)。OTU が 6 つのときに可能な樹形の数として適切なものを選択肢の中から一つ選べ。



- 1 27
- 2 75
- 3 105
- 4 210

問 67

分子進化の中立説では、集団に固定される遺伝子変異の大部分は、自然選択に対して有利でも不利でもない中立なものであると考えられる。以下のタンパク質をコードする遺伝子の進化速度についての観察事実の中から、中立説で説明できる観察事実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 何らかの原因で機能を失ってしまった、いわゆる偽遺伝子は非常に速い進化速度で配列が変化する。
- 2 タンパク質をコードする領域の塩基配列では、同義置換は非同義置換に較べて頻繁に起こっている。
- 3 タンパク質をコードする領域の塩基配列では、コドンの 3 番目の位置の進化速度がもっとも速い。
- 4 ウイルスの表面抗原において、抗体によって認識される部位ではアミノ酸置換が加速されることがある。

問 68

全ゲノム相関解析(Genome-Wide Association Study; GWAS)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 GWAS は、集団をある疾患に罹患した疾患群と、その疾患に罹患していない対照群に分けてそれぞれ遺伝子配列を解析し比較する研究である。
- 2 通常の GWAS の統計解析では、SNP などの遺伝子変異について帰無仮説「この SNP は対象疾患と関連しない」を棄却する場合の危険率(p 値)を評価する。
- 3 GWAS 統計解析では通常、 p 値の閾値 (有意水準) は検定した SNP などの遺伝子変異数や遺伝子数に応じて補正される。
- 4 $-\log_{10} p$ 値はオッズ比とよばれ、この値が低いことは、当該 SNP の存在により対象疾患が発症するリスクが高いことを表す。

問69

保存血液中の乳酸は保存条件によって、濃度が増加することがある。乳酸の濃度変動について、簡単のために以下のように記述するとする。

$$\frac{dL(t)}{dt} = kL(t)$$

初期状態($t=0.0$ sec)の乳酸濃度を 1.0 mM、反応定数 $k = 0.10 \text{ sec}^{-1}$ とした場合、時間積分ステップ幅 $h = 1.0 \text{ sec}$ でのオイラー法で数値積分した場合の、 $t = 2.0 \text{ sec}$ における乳酸濃度を選択肢の中から一つ選べ。なお、 $L(t)$ は乳酸の濃度を示す。

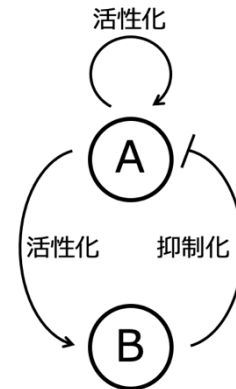
- 1 1.20 mM
- 2 1.21 mM
- 3 2.10 mM
- 4 2.21 mM

(問 70・問 71)

問 70

A, B の遺伝子からなる遺伝子制御ネットワークとして、図で示したような構造を想定する。A は自分自身と B を活性化し、B は A を抑制化する。このネットワークにおける A, B の逐次的な時間変化への、A, B それぞれの寄与について、定性的な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 B が増えることにより B はさらに増える
- 2 B が存在しないとき、A が増えることにより A はさらに増える
- 3 B が増えることにより A は減る
- 4 A が増えることにより B が増える



問 71

次に A, B の発現量の時間変化の時間的変化を議論する。A, B の発現量が、以下の連立方程式

$$\begin{cases} \frac{d[A]}{dt} = k_1[A] + k_2[B] \\ \frac{d[B]}{dt} = k_3[A] \end{cases}$$

で表されるとする。ここで係数 $k_1 = 3, k_2 = -2, k_3 = 1$ とする。このとき、行列

$$\begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}$$

の固有値を用いて、システムの平衡点まわりの安定性について、議論することができるが、もっとも適切なものを選択肢の中から一つ選べ。なお、システムが平衡点まわりで安定であるとは、平衡点にある状態に摂動が加えられた場合に、元の平衡点に状態にもどる挙動を示すことをいう。

- 1 固有値は1, 2であるので、システムの平衡点まわりは不安定である。
- 2 固有値は1, 2であるので、システムの平衡点まわりは安定である。
- 3 固有値は-1, -2であるので、システムの平衡点まわりは安定である。
- 4 固有値は-1, -2であるので、システムの平衡点まわりは不安定である。

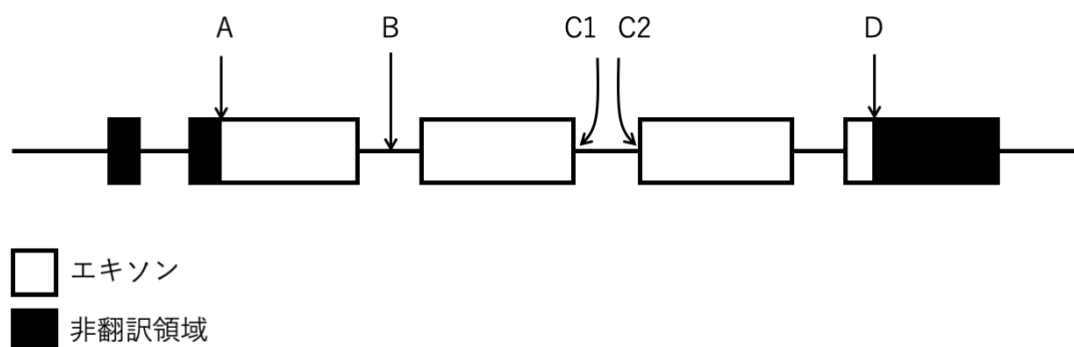
問 72

DNA メチル化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 哺乳類の CpG 部位の多くは DNA メチル化を受けており、DNA メチル化シトシンは変異を受けやすいためヒトゲノム中の CpG の出現頻度は他のジヌクレオチドよりも低い。
- 2 通常 DNA メチル化により遺伝子発現は促進され、脱メチル化により抑制される。
- 3 DNA メチル化を測定する方法に 1 つにバイサルファイト(bisulfite)法がある。バイサルファイト処理により、シトシンは脱アミノ化してウラシルへ変換されるが、DNA メチル化を受けたシトシンは脱アミノ化を受けないため、シーケンシング等により区別することができる。
- 4 近年、バイサルファイト法を使用せず、シトシンとメチル化シトシンの分子構造や電気抵抗の違いなどを利用した DNA メチル化解析が可能となった。

問 73

下図は真核生物のゲノム構造の模式図である。転写・スプライシング関わる記述について図を参照しながら、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 A は開始コドンである。一部の例外を除きゲノム上の配列としては ATG である。
- 2 B はイントロンである。転写後にスプライシングにより取り除かれる。
- 3 C1 と C2 はスプライスジャンクションである。ゲノム上の配列としては C1 が GT、C2 が AG の配列を持つことが多い。
- 4 D は終止コドンである。一部の例外を除きゲノム上の配列としては TAA、TAT、TAG の 3 種である。

問 74

オーミクス解析に使われる実験機器とその使用目的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。

【実験機器】

【使用目的】

- | | |
|---------------|--------------|
| 1 次世代シーケンサ | 環境微生物叢組成の推定 |
| 2 核磁気共鳴装置 | 代謝産物の定量 |
| 3 DNA マイクロアレイ | 遺伝子多型の検出 |
| 4 質量分析装置 | タンパク質立体構造の決定 |

問 75

A および B の 2 つの試料の細菌叢プロファイルを同定するために、メタ 16S 解析(16S Amplicon sequencing)ならびにフローサイトメトリー法による細菌数計測を実施した。各試料において、細菌 X および細菌 Y に帰属されたりードの数、得られたりードの総数、試料単位質量あたりの細菌数はそれぞれ表の通りであった。細菌の存在量に関する推定として、もっとも不適切なものを選択肢の中から一つ選べ。

表：メタ 16S 解析および細菌数計測の結果

試料	細菌 X のリード数	細菌 Y のリード数	総リード数	細菌数 [cell/g]
A	200	400	1000	5.0×10^5
B	600	400	2000	2.5×10^5

- 1 試料 A において、細菌 Y は細菌 X よりも多く存在している。
- 2 細菌 X の菌密度は、試料 A よりも試料 B の方が大きい。
- 3 試料 B において、細菌 X は細菌 Y よりも多く存在している。
- 4 細菌 Y の菌密度は、試料 B よりも試料 A の方が大きい。

問 76

細菌叢データの多様性解析に関する以下の説明文において、（ A ）～（ D ）内に入る語句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。

メタゲノム解析やメタ 16S 解析で取得した細菌叢データの分析において、（ A ）を表す α 多様性や、（ B ）を表す β 多様性などの生態学的な多様性指標が使用される。代表的な α 多様性指標としては、情報エントロピーの概念を用いて種の豊富さ(richness)と均等度(evenness)を複合的に評価する（ C ）がある。また、 β 多様性については、検出された細菌群の系統樹の比較から算出される（ D ）を用いた解析が近年は主流となっている。

選択肢	（ A ）	（ B ）	（ C ）	（ D ）
1	特定環境での種多様性	環境間の種多様性の違い	UniFrac Distance	Shannon Index
2	環境間の種多様性の違い	特定環境での種多様性	UniFrac Distance	Shannon Index
3	特定環境での種多様性	環境間の種多様性の違い	Shannon Index	UniFrac Distance
4	環境間の種多様性の違い	特定環境での種多様性	Shannon Index	UniFrac Distance

問 77

遺伝子ネットワークとは、遺伝子をノードとし、相互作用など様々な関係性をエッジとして抽象化したデータ構造である。遺伝子ネットワークの特徴についての記述として、もっとも不適切なものを選択肢の中から一つ選べ。

ただし、選択肢の中で、スケールフリー性とは、次数（各ノードが持つエッジの数）の分布がべき乗則に従う性質を、ランダムネットワークとは、エッジがノード間に一定の確率でランダムに張られているネットワークを、さらに、完全グラフは、どの二つのノードも必ずエッジで連結されている部分ネットワークをそれぞれ表す。

- 1 遺伝子ネットワークには、スケールフリー性が見られる例がある。
- 2 遺伝子ネットワークの性質を解析する上で、ランダムネットワークにおける性質と比較することは効果的である。
- 3 遺伝子ネットワーク中で、多数の遺伝子と相互作用をするハブの役割をもつ遺伝子がみられる場合がある。
- 4 遺伝子ネットワークの内、タンパク質間相互作用している遺伝子で出来る部分ネットワークは、完全グラフとなる。

問 78

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

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

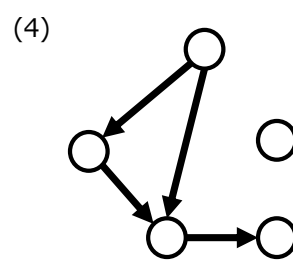
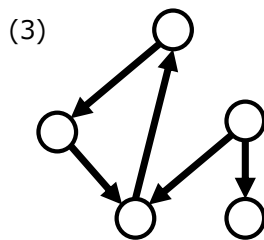
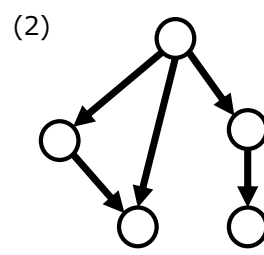
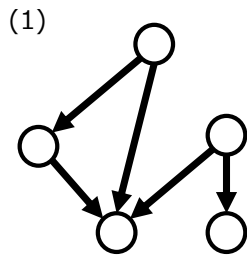
問 79

あるタンパク質 X に対するモノクローナル抗体とポリクローナル抗体を購入し、ウエスタンブロッティング法によるサンプル間の比較解析を行った。その結果、ポリクローナル抗体を用いた解析では、サンプル A とサンプル B のシグナルに有意な差は見られなかったが、モノクローナル抗体を用いた解析では、サンプル A に対してサンプル B のシグナルの著しい減少が見られた。この実験結果の考え得る解釈として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 2 つのサンプルで、タンパク質 X の一部のアミノ酸が異なる化学修飾状態にある。
- 2 2 つのサンプルには、選択的スプライシングの影響で、異なるアミノ酸配列を有するタンパク質 X が異なる比で存在する。
- 3 2 つのサンプルでは、タンパク質 X をコードする mRNA の翻訳の効率に差がある。
- 4 2 つのサンプルでは、タンパク質 X が翻訳後にプロテアーゼによる切断を介した活性化を受ける効率が異なる可能性がある。

問 80

5つの遺伝子に対して、発現量の測定を行った。次に、各遺伝子の発現量を確率変数と考え、統計モデルによって発現制御の因果関係を推定し、ベイジアンネットワークを用いて表現した。以下の有向グラフのうち、ベイジアンネットワークの推定結果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。



2019 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
4	4	3	3	2	2	4	1	2	2
問 11	問 12	問 13	問 14	問 15	問 16	問 17	問 18	問 19	問 20
2	1	3	3	4	2	4	3	1	3
問 21	問 22	問 23	問 24	問 25	問 26	問 27	問 28	問 29	問 30
3	4	2	2	4	1	2	1	4	3
問 31	問 32	問 33	問 34	問 35	問 36	問 37	問 38	問 39	問 40
1	1	1	2	1	2	3	4	1	4
問 41	問 42	問 43	問 44	問 45	問 46	問 47	問 48	問 49	問 50
3	4	4	3	2	2	3	1	4	1
問 51	問 52	問 53	問 54	問 55	問 56	問 57	問 58	問 59	問 60
2	3	3	1	3	4	4	4	2	3
問 61	問 62	問 63	問 64	問 65	問 66	問 67	問 68	問 69	問 70
1	3	2	1	1	3	4	4	2	1
問 71	問 72	問 73	問 74	問 75	問 76	問 77	問 78	問 79	問 80
1	2	4	4	2	3	4	4	3	3

試験問題に記載されている会社名または製品名は、それぞれ各社の商標または登録商標です。
なお、試験問題では、®および™を明記していません。

Copyright © 2019 Japanese Society for Bioinformatics. All Rights Reserved