

受験番号 Examinee number

--	--	--	--	--	--

東京大学 大学院新領域創成科学研究科 情報生命科学専攻

Department of Computational Biology, Graduate School of Frontier Sciences  
the University of Tokyo

平成19(2007)年度

2007 School Year

大学院入学試験問題 修士・博士後期課程

Graduate School Entrance Examination Problem Booklet, Master's and Doctoral Course

専 門 科 目

Specialties

平成18年8月10日(木)

Thursday, August 10, 2006

13:00~15:00

注意事項 Instructions

1. 試験開始の合図があるまで、この冊子を開いてはいけません。  
Do not open this problem booklet until the start of examination is announced.
2. 本冊子の総ページ数は 35 ページです。落丁、乱丁、印刷不鮮明な箇所などがあった場合には申し出ること。  
This problem booklet consists of 35 pages. If you find missing, misplaced, and/or unclearly printed pages, ask the examiner.
3. 解答には必ず黒色鉛筆(または黒色シャープペンシル)を使用しなさい。  
Use black pencils (or mechanical pencils) to answer the problems.
4. 問題は 12 題出題されます。問題 1~12 から選択した合計4問に解答しなさい。ただし、問題 1~12 は同配点です。  
There are twelve problems (Problem 1 to 12). Answer four problems out of the twelve problems. Note that Problem 1 to 12 are equally weighted.
5. 解答用紙は計4枚配られます。各問題に必ず1枚の解答用紙を使用しなさい。解答用紙に書ききれない場合は、裏面にわたってもよい。  
You are given four answer sheets. You must use a separate answer sheet for each problem. You may continue to write your answer on the back of the answer sheet if you cannot conclude it on the front.
6. 解答は日本語または英語で記入しなさい。  
Answers should be written in Japanese or English.
7. 解答用紙の指定された箇所に、受験番号と選択した問題番号を記入しなさい。問題冊子にも受験番号を記入しなさい。  
Fill the designated blanks at the top of each answer sheet with your examinee number and the problem number you are to answer. Fill the designated blanks at the top of this page with your examinee number.
8. 草稿用紙は本冊子から切り離さないこと。  
The blank pages are provided for making draft. Do not detach them from this problem booklet.
9. 解答に関係ない記号、符号などを記入した答案は無効とします。  
An answer sheet is regarded as invalid if you write marks and/or symbols unrelated to the answer on it.
10. 解答用紙・問題冊子は持ち帰ってはいけません。  
Do not take the answer sheets and the problem booklet out of the examination room.
11. 受験生の便宜のために問題の英訳が 20 ページ以降に掲載されていますが、正式な問題文は日本語によるものです。  
For the convenience of applicants, problems are translated into English (pp.20-33). Note that Japanese version is the formal one.

(このページは草稿用紙として使用してよい)  
(Blank page for draft)

(このページは草稿用紙として使用してよい)  
(Blank page for draft)

(このページは草稿用紙として使用してよい)  
(Blank page for draft)

### 問題 1

以下の行列式を各々計算せよ。

$$(1) \begin{vmatrix} a+1 & a \\ b & b+1 \end{vmatrix}$$

$$(2) \begin{vmatrix} 1 & a \\ -1 & b+1 \end{vmatrix}$$

$$(3) \begin{vmatrix} a+1 & a & a \\ b & b+1 & b \\ c & c & c+1 \end{vmatrix}$$

## 問題2

以下の(1)–(5)の用語について説明せよ。

- (1) 記憶階層
- (2) 抽象データ型
- (3) ガーベジコレクション(ゴミ集め)
- (4) 番兵
- (5) Newton 法

### 問題3

$N$ 個の英単語からなる辞書を引く問題を考える。これについて以下の5つの設問に答えよ。

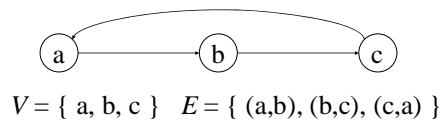
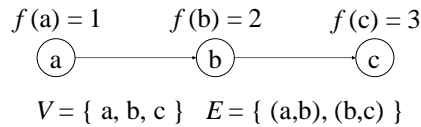
- (1)  $N$  個の英単語をアルファベット順に並べたとき、頭から順番に見ていった場合、望みの英単語を見つけるまで平均何回の比較が必要か？
- (2) 最悪の場合は何回の比較が必要か？
- (3) 比較回数を減らすにはどのような方法が考えられるか？
- (4) その場合の平均比較回数、最悪の比較回数は？
- (5) 上記の辞書引きを例にして、二分探索法、ハッシュ法をそれぞれ説明せよ。

#### 問題4

頂点の集合を  $V$ 、有向辺の集合  $\{(v, w) \mid \text{頂点 } v \text{ から頂点 } w \text{ への有向辺が存在}\}$  を  $E$  とする有向グラフ  $(V, E)$  を考える。 $V$  の各頂点  $v$  に自然数  $f(v)$  を割当て関数  $f$  が  $E$  の全ての辺  $(v, w)$  について  $f(v) < f(w)$  を満たすとき無矛盾と呼ぶ。また有向辺の列

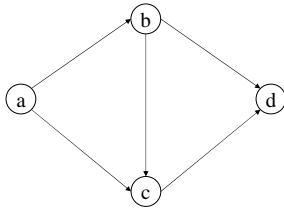
$$(v_1, v_2), (v_2, v_3), \dots, (v_{k-1}, v_k) \quad (k \geq 2)$$

が  $v_1 = v_k$  を満たすときサイクルと呼ぶ。下図では、上に無矛盾な関数の例を、下にサイクルの例を示す。

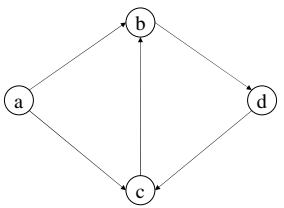


(1) 以下の各々の場合について、無矛盾な関数  $f$  が存在する場合には例を示し、存在しない場合には証明せよ。

A)  $V = \{ a, b, c, d \} \quad E = \{ (a,b), (a,c), (b,d), (c,d), (b,c) \}$



B)  $V = \{ a, b, c, d \} \quad E = \{ (a,b), (a,c), (b,d), (d,c), (c,b) \}$



C)  $E$  がサイクルを含むとき

D)  $E$  がサイクルを含まないとき

(2) 無矛盾な関数が存在するか否かを判定するアルゴリズムを示せ。



## 問題5

以下の語句の組それぞれについて解説し、情報生命科学の観点からみた両者の関係を簡単に記述しなさい。

- 1) EST — ハッシング
- 2) 遺伝子コード領域 — 隠れマルコフモデル
- 3) タンパク質のモチーフ — アライメント
- 4) 逆遺伝学 — DNA マーカー

## 問題6

タンパク質を構成する 20 種類のアミノ酸は、4 種類の核酸塩基の 3 つ組によって指定される。以下の問いに答えよ。解答するのに以下の値を用いてもよい。

$$\log_2 3 = 1.6, \log_2 5 = 2.3$$

- 1) 塩基配列を 4 種の文字からなる文字列と考えたとき、長さ  $N$  の塩基配列に含まれる情報量は何ビットか。
- 2) 長さ  $N$  の塩基配列と同じ情報量を持つアミノ酸配列の長さを  $N$  で表せ。
- 3) アミノ酸配列を指定するのに、塩基の重なりあわない 3 つ組ではなく、1 塩基だけ重なり合う 3 つ組で指定することは可能か、理由とともに述べよ。
- 4) アミノ酸配列と、それをコードする塩基配列との、生物学的な情報量の違いを簡潔に述べよ。

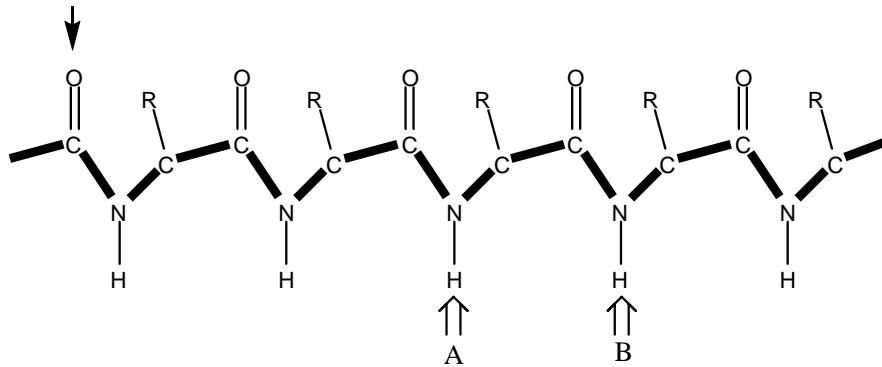
### 問題7

1) 下記の A と B はタンパク質立体構造の性質に関する文である。それぞれに最も関係する用語を1から4の中から重複がないように一つ選び、どのような思考実験または実際の実験を行えばこれを確かめることができるか説明せよ。

- A. タンパク質天然構造は自由エネルギー最小の状態である。
- B. タンパク質の折れたたみにおいて構造空間全探索はおこらない。

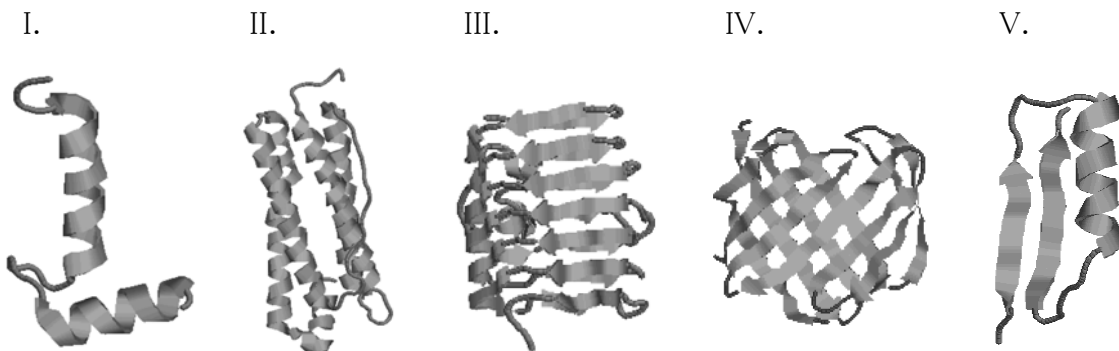
1. Epimenides paradox 2. Anfinsen's dogma 3. Levinthal paradox 4. Central dogma

2) 以下の図はポリペプチド主鎖の化学構造を模式的に示したものである。左上の矢印で示した酸素原子(O)が、AとBで示した水素(H)のどちらかと水素結合を作るとき、それぞれある2次構造の1ターン分が形成することができる。AとBそれぞれについて最適な2次構造を1から5から重複がないように選んで答えよ。



- 1.  $\pi$ ヘリックス
- 2.  $\alpha$ ヘリックス
- 3. 平行 $\beta$ シート
- 4. 反平行 $\beta$ シート
- 5.  $3_{10}$ ヘリックス

3) IからVの図は典型的なタンパク質のモチーフまたはフォールドを示したものである。それぞれの特徴をもっともよくあらわす言葉を1-6から重複がないように選べ。



- 1.  $\beta$ バレル
- 2.  $\beta - \alpha - \beta$ モチーフ
- 3. ヘリックスバンドル
- 4.  $\alpha / \beta$ バレル
- 5. ヘリックスループ-ヘリックス
- 6.  $\beta$ ヘリックス

### 問題8

タンパク質の立体構造を計算機により予測する方法にホモロジーモデリング法がある。これは立体構造既知のタンパク質(鋳型タンパク質)との配列一致率を利用して構造未知のタンパク質の立体構造を予測する方法である。以下の問いに答えよ。

- 1) ホモロジーモデリング法における基本的な3ステップを述べよ。
- 2) ホモロジーモデリング法に必要な配列一致率の閾値を計算機により見積もる方法を説明せよ。
- 3) 鋳型タンパク質を構造データベースから選定する際、配列一致率以外に重要な因子について述べよ。

## 問題9

以下の問に答えなさい。

**問1** メンデルの法則ではエンドウ豆の形(丸い、またはしわ)、色(黄色、または緑)等の形質は独立して遺伝することが知られている。これらの形質の中で、丸はしわに対して優性、黄色は緑に対して優性であることが知られている。今、丸くて黄色の種としわで緑の種から育てたエンドウを掛け合わせたところ、黄色で丸いエンドウと緑で丸いエンドウが同数得られた。このとき、掛け合わせた元のエンドウ豆の遺伝子型を書きなさい。ただし、丸い形質を規定する遺伝子を R、しわを規定する遺伝子を r、黄色を Y、緑を y で表すものとする。

**問2** 形質が独立に遺伝するのはどうしてか、述べなさい。また、独立に遺伝しない場合も知られている。これはどの様な場合か。

**問3** ヒトの場合、ある種の形質は男性だけで表現型が見られ、女性には表現型が見られないことが知られている。これはどの様なメカニズムによると考えられるか。

**問4** ある遺伝病は劣性で常染色体上の単一遺伝子によって支配されていることが知られている。発症率が1万人に一人であるような病気の場合、保因者は何人に一人と考えられるか。

**問5** 生殖方法には上記のような有性生殖と、性を伴わない無性生殖がある。これら各々の特徴をメカニズム、および生物学的意義の観点から 10 行以内で説明しなさい。

## 問題10

RNA 研究に関する以下の文章を読んで、問に答えなさい。

1950 年代に、セントラルドグマという分子生物学の中心原理とも言える概念が提唱された。遺伝情報(DNA)はRNAに写しとられ(転写)、その情報に基づいてタンパク質が生合成(翻訳)されるというものである。この過程において、(Ⅰ)遺伝情報を伝えるRNA、(Ⅱ)翻訳に必要なリボソームの構成成分であるRNA、(Ⅲ)アミノ酸を遺伝子の塩基配列に基づいて運ぶRNA(a)が重要な働きを担うRNA分子であることが認識された。ここでは、RNAの役割は、あくまでもDNAとタンパク質の仲介役にすぎなかった。ところが、セントラルドグマは必ずしも絶対的なものではなかった。1970 年代には、真核生物では、転写されたRNAはそのままタンパク質に翻訳されるのではない(b)こと、また、レトロウイルスとよばれる、RNAをゲノムとして持つウイルスでは、DNAとRNAの間に逆向きの流れがある(c)ことが発見された。さらに、1980 年代になると、RNAには酵素活性をもつものがある(d)こともわかってきた。これらの予想外の発見から、RNAは遺伝情報と酵素活性の両方を持つことが証明され、原始地球上における遺伝子の主体がRNAであるとする、( e ) 仮説が提唱された。これは現在、最も支持されている「生命の起源」説である。この説が一般的に受け入れられるにつれ、RNAはDNAとタンパク質の性質を兼ね備える、生物を構成する最も基本的な高分子であるとみなされるようになってきた。そして、RNAにさまざまな機能が付与され得るとすると、もしかしたら、細胞内には、他にも未知の機能をもつRNAが存在するのではないかと考えられた。このような状況下で、1990 年代後半には、RNAの新しい機能として、低分子の二本鎖RNAが、相補的な配列をもったRNAを分解することによって、遺伝子の機能を抑える、( f ) という現象が見いだされた。さらに、マイクロRNAという低分子のRNAが、ゲノムから多数転写され、( f ) と類似の機構で遺伝子の機能を制御していることもわかってきた。DNAとRNAは、極めて類似した分子であり、化学構造はほんの少ししか変わらない(g)。しかし、RNAは反応性が高く分解しやすいが、DNAは安定であるため、RNAは化学的性質からも柔軟性を持つ分子といえる。今まさにRNAの多様で重要な役割が次々と明らかになりつつある時代といえる。

- 問1 下線部(a)について。真核生物における(Ⅰ)から(Ⅲ)の RNA の種類と、それぞれの RNA の転写に関わる RNA ポリメラーゼの種類を記せ。
- 問2 下線部(b)について。真核生物における RNA のプロセシングの過程を、3つのステップにわけて、概説せよ。
- 問3 下線部(c)について。この流れを、関連する酵素名を用いて2行程度で説明せよ。
- 問4 下線部(d)について。酵素活性をもつ RNA は何と呼ばれるか。また、酵素活性の具体的な例を1つ挙げよ。
- 問5 空欄 (e)に適切な語を入れよ。
- 問6 空欄 (f) に適切な語を入れよ。
- 問7 下線部(g)について。DNA と RNA の違いを化学構造式を用いて説明せよ。ただし、塩基は”R”として良いとする。

## 問題11

下記の問に答えなさい。

コレステロールは水に溶けにくく、血液中ではタンパク質に結合した複合体(A)の形で運搬される。(A)は細胞表面にある受容体に結合すると、クラスリン被覆小胞と呼ばれる小胞に取り込まれる。細胞内でクラスリン被覆小胞は(B)と融合し、そこで(A)と受容体は選別される。受容体は細胞膜に運ばれてリサイクルされるが、(A)はリソソームで分解される。

上皮増殖因子(EGF)などの細胞増殖因子が細胞表面に存在する受容体に結合すると、EGF/EGF受容体の複合体は同様な機構で細胞内に取り込まれる。しかしEGF受容体は、多くの場合EGFとともにリソソームに輸送されて分解される。受容体がリサイクルされるか、リソソームで分解されるかは受容体の種類によって異なる。

真核細胞の細胞質でタンパク質分解の主要な役割を果たしているのは、プロテアソームと呼ばれる巨大なタンパク質分解酵素の複合体である。プロテアソームは少なくとも30種類以上のサブユニットからなるタンパク質分解酵素複合体で、(C)により標識されたタンパク質を分解する。細胞周期の進行に伴ってその発現量が周期的に変化する(D)は、細胞周期の特定の時期にプロテアソームによって分解される事が知られている。

問1 (A)から(D)に適切な語を入れなさい。

問2 リソソームに存在する酵素は、もし細胞質に漏れてきてもその活性を示さない。これは何故か。リソソーム内と細胞質の環境の違いを説明し、リソソーム内の環境を維持するのに重要なタンパク質を挙げて説明しなさい。

問3 EGF が受容体に結合すると、上記のような機構で細胞表面の受容体数が一時的に減少する。このように受容体数が一時的に減少することは、どのような利点があると考えられるか。

問4 増殖中の細胞の核を propidium iodide で染色し、フローサイトメーターで解析して図1のようなヒストグラムを得た。このヒストグラムの各ピークを図2のように分解した場合、XYZそれぞれのピークがどの細胞周期の細胞群に相当するかを示しなさい。



図1

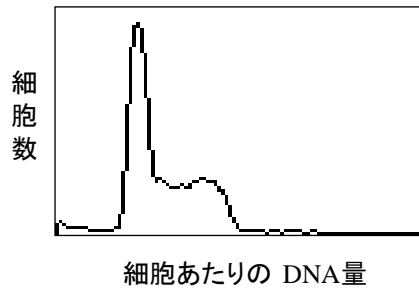
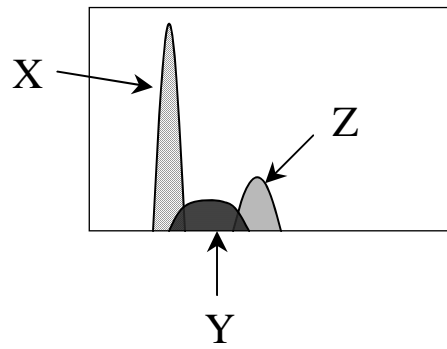


図2



問5 プロテアソームは、変性したタンパク質や折りたたみの異常なタンパク質なども分解するが、これは(C)による修飾が正常なタンパク質に比べこれらのタンパク質で起こりやすいためであると考えられる。変性したタンパク質や折りたたみの異常なタンパク質が、(C)による修飾を受けやすくなる機構について考えを述べなさい。

## 問題12

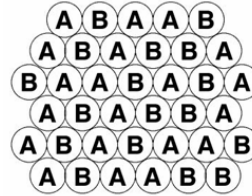
下記の文章を読んで問1～5に答えなさい。

A, Bの2種類の細胞から成る体の組織を考える。カシワインとオダイビンというタンパク質をコードする遺伝子の機能欠失突然変異体では、どちらもこの組織にA細胞がほとんどなくなり、B細胞ばかりになってしまう(図1)。これらの遺伝子の機能を調べるために、次のような実験を行った。

**【実験A】**これらの遺伝子が、正常な組織においてどの細胞で発現しているかを調べるために、カシワイン遺伝子とオダイビン遺伝子の塩基配列からプローブを作成してin situハイブリダイゼーションを行った。その結果、カシワイン遺伝子はA細胞だけで、またオダイビン遺伝子はB細胞だけで発現していることがわかった(図2、染色された細胞は部位、濃淡に関わらず一様にグレーで示す)。

**【実験B】**次に、カシワインとオダイビンのタンパク質に対する抗体を作成して、正常な組織の標本を染色した。カシワインの抗体ではAの細胞だけが染色されたが、オダイビンの抗体ではBの細胞だけでなく組織全体が染色された(図3、染色された細胞は部位、濃淡に関わらず一様にグレーで示す)。

図1. 組織の模式図  
正常な組織



突然変異体の組織

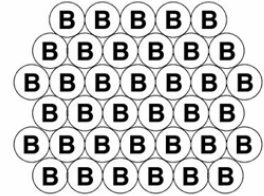


図2. in situハイブリダイゼーション  
カシワインプローブ オダイビンプローブ

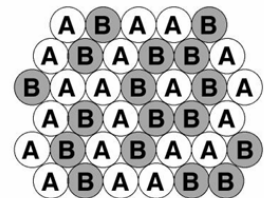
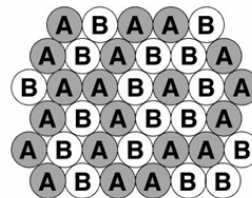
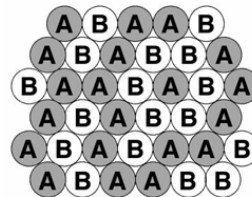
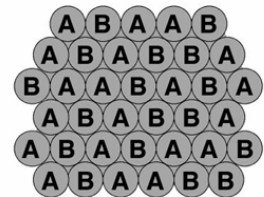


図3. 免疫染色  
カシワイン抗体



オダイビン抗体



問1 以下の文章のうち、これらの実験の結果と符合するものを選び、記号で答えよ。

- カシワインのタンパク質はA細胞に、オダイビンのタンパク質はB細胞にそれぞれ局在している。
- カシワインのタンパク質はA細胞に局在し、オダイビンのタンパク質はB細胞から分泌されている。
- カシワインのタンパク質はA細胞から分泌されており、オダイビンのタンパク質はB細胞に局在している。
- カシワインのタンパク質はA細胞から、オダイビンのタンパク質はB細胞から分泌されている。

**【実験C】**カシワインとオダイビンの遺伝子は胚発生の中期に発現が始まる。正常個体でもどちらの突然変異体でも、すでにこの時点で、この組織には発生完了時と同じ数の細胞がある。しかしそのほとんどはA細胞で、B細胞はわずかな数しかない。その後の発生過程でA細胞が

減少し B 細胞が増えてゆくが、正常個体の組織では B 細胞がある程度の数に達するとそれ以上増えないのに対し、突然変異体では B 細胞がどんどん増え、発生完了時には組織全体が B 細胞になっている。

オダイビン遺伝子の発現が始まる時期に、オダイビン突然変異体の組織から一部分を切り出し、同じ発生時期の正常個体の組織に移植した。発生完了時にこの組織を観察すると、正常な個体に由来する部分で A 細胞と B 細胞が観察されただけでなく、オダイビン変異体に由来する部分でも、B 細胞だけでなく多数の A 細胞が観察された。

**問2** これはなぜだろうか？2～3行程度で述べよ。

**【実験 D】**同様に、同じ時期のカシワイン突然変異体から組織の一部を切り出し、正常な個体の組織に移植して発生を続けさせた。

**問3** 発生完了時において、カシワイン変異体から移植された部分には、A 細胞と B 細胞が両方観察されるだろうか？A 細胞ばかりが観察されるだろうか？それとも B 細胞ばかりが観察されるだろうか？理由とともに2～3行程度で答えよ。

**問4** 正常な個体では、なぜ B 細胞はある程度の数に達するとそれ以上増えないのか、2～3行程度で自分の考えを述べよ。

**【実験 E】**次に、成体からこの組織を取り出してすりつぶし、定量 PCR 法によってオダイビン遺伝子の発現量を調べた。その結果、オダイビン突然変異体の組織では、正常な個体の組織よりもオダイビン遺伝子の発現量が3倍に増加していることが分かった。

**問5** 以上の結果から、「オダイビン突然変異体の組織では、1細胞あたりのオダイビン遺伝子の発現量が通常よりも増加している」と結論してよいか？理由とともに2～3行程度で答えよ。

**Problem 1**

Calculate the determinant of each matrix.

$$(1) \begin{vmatrix} a+1 & a \\ b & b+1 \end{vmatrix}$$

$$(2) \begin{vmatrix} 1 & a \\ -1 & b+1 \end{vmatrix}$$

$$(3) \begin{vmatrix} a+1 & a & a \\ b & b+1 & b \\ c & c & c+1 \end{vmatrix}$$

**Problem 2**

Explain the words from (1) to (5).

- (1) Memory hierarchy
- (2) Abstract data type
- (3) Garbage collection
- (4) Sentinel
- (5) Newton method

### **Problem 3**

Consider the problem of consulting a dictionary that consists of  $N$  English words, and answer the following five questions.

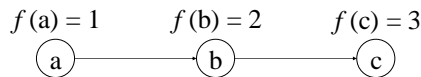
- (1) Suppose that  $N$  English words are sorted in alphabetical order in the dictionary. When one sequentially searches for a word from the beginning of the dictionary, give the expected number of word comparisons required.
- (2) When one searches for a word in the same way as (1), give the number of word comparisons required in the worst case.
- (3) Show a method that can reduce the number of word comparisons.
- (4) In the method of (3), how many word comparisons are required in the average and worst cases, respectively?
- (5) Using the dictionary as an example, explain the two search methods, binary search and hash method.

**Problem 4**

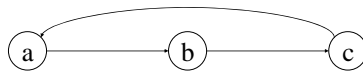
Consider a directed graph  $(V, E)$ , where  $V$  denotes the set of vertexes and  $E$  is the set of directed edges  $\{(v,w) \mid \text{there is a directed edge from } v \text{ to } w\}$ . A function  $f$  that sets each vertex  $v$  to a natural number  $f(v)$  is *consistent* with graph  $(V, E)$  if  $f(v) < f(w)$  for any edge  $(v, w)$ . A series of edges

$$(v_1, v_2), (v_2, v_3), \dots, (v_{k-1}, v_k) \quad (k \geq 2)$$

is called a *cycle* if  $v_1 = v_k$ . In the following figure, the upper portion shows an example of consistent function, and the lower part displays a cycle.



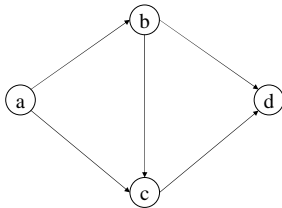
$$V = \{ a, b, c \} \quad E = \{ (a,b), (b,c) \}$$



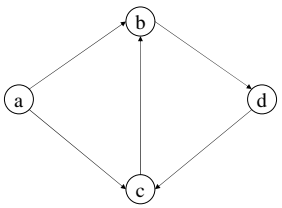
$$V = \{ a, b, c \} \quad E = \{ (a,b), (b,c), (c,a) \}$$

(1) In each of the following cases, if a consistent function exists, provide an example. Otherwise, prove the absence of consistent functions.

A)  $V = \{ a,b,c,d \} \quad E = \{ (a,b), (a,c), (b,d), (c,d), (b,c) \}$



B)  $V = \{ a,b,c,d \} \quad E = \{ (a,b), (a,c), (b,d), (d,c), (c,b) \}$



C)  $E$  contains a cycle.

D)  $E$  has no cycle.

(2) Design an algorithm for checking the existence of a function consistent with the given graph.

**Problem 5**

Explain each of the following two terms, and describe their relationship in terms of computational biology.

- 1) EST – hashing
- 2) protein-coding region – hidden markov model
- 3) motif in protein sequence – alignment
- 4) reverse genetics – DNA marker



**Problem 6**

The 20 amino acids constituting proteins are specified by triplets of nucleic acid bases. Answer the following questions. You may use the following values.

$$\log_2 3 = 1.6, \quad \log_2 5 = 2.3$$

- 1) When a base sequence is considered a string of 4 letters, what is the information content of a base sequence of length  $N$ ?
- 2) Write the length of an amino acid sequence that has the same information content as a base sequence of length  $N$ .
- 3) Is it possible to specify an amino acid sequence using base triplets that have a single-base overlap, instead of non-overlapping triplets? Explain briefly.
- 4) Briefly explain the biological difference of information content between an amino acid sequence and a base sequence that encodes it.

**Problem 7**

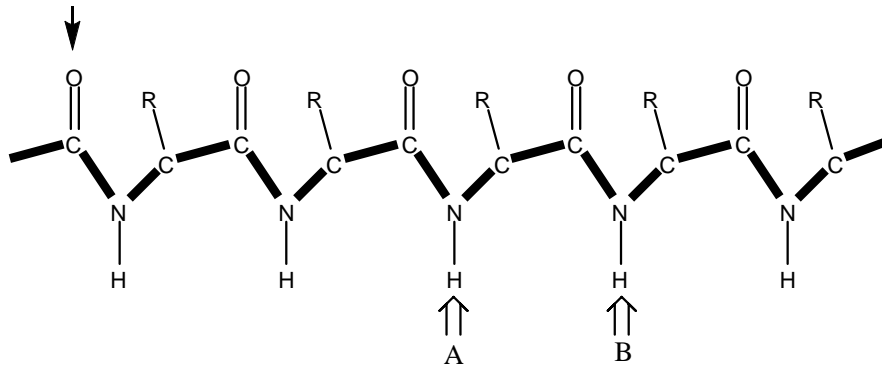
1) The following statements A and B describe the characteristics of protein structure. For each statement, select the most related term from the following list 1 – 4 without repeating any choice, and describe the thought experiment or practical experiment that can support it.

A. The native structure of protein is the free-energy minimum state.

B. Entire search of conformational space does not occur in protein folding.

1. Epimenides paradox    2. Anfinsen's dogma    3. Levinthal paradox    4. Central dogma

2) The following figure schematically shows the main chain chemical structure of a polypeptide. When the oxygen atom (O) indicated by an arrow in the upper left makes a hydrogen bond with one of the hydrogen atoms labeled A or B, one turn of a certain secondary structure can be formed. Select the most appropriate secondary structure from the list 1 – 5 without repeating any choice.



1.  $\pi$  helix    2.  $\alpha$  helix    3. Parallel  $\beta$  sheet    4. Anti-parallel  $\beta$  sheet    5.  $3_{10}$  helix

3) The following figures I – V represent typical protein structural motifs and folds. For each one, select the most appropriate term from the list 1 – 6 without repeating any choice.

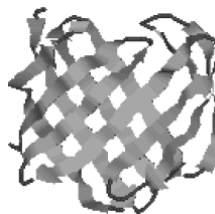
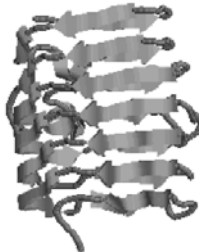
I.

II.

III.

IV.

V.



1.  $\beta$  barrel    2.  $\beta$ - $\alpha$ - $\beta$  motif    3. helix bundle  
 4.  $\alpha/\beta$  barrel    5. helix-loop-helix motif    6.  $\beta$  helix

**Problem 8**

Homology modeling is a computational approach of predicting the 3D structure of a protein based on sequence identity between the target protein and the template of known structure.

- 1) Write 3 general steps in homology modeling.
- 2) Explain the computational method to estimate the threshold of pairwise sequence identity in homology modeling.
- 3) Describe important factors, in addition to sequence identity, for choosing template proteins from structural databases.

## **Problem 9**

Answer the following questions.

**Question 1:** According to the Mendelian principles, two different characteristics of green peas, such as the shape (smooth or wrinkled) and color (yellow or green), are inherited independently. It is known that smooth is genetically dominant over wrinkled and yellow is dominant over green. If you obtained the same number of smooth yellow peas and smooth green peas when you mated plants from smooth yellow seeds and those from wrinkled green seeds, give the genotypes of the parental peas. Use the gene symbol “R” to denote smooth phenotype, use “r” for wrinkled, “Y” for yellow, and “y” for green.

**Question 2:** Explain why different characteristics are inherited independently. It is known, however, that some characters are not inherited independently. Explain why this happens.

**Question 3:** In the case of humans, some phenotypes are observed only in males but not in females. Explain the mechanism involved.

**Question 4:** Suppose that a certain genetic disease is caused by a recessive mutation in a single autosomal gene. If the frequency of the patients is one in 10,000, what is the frequency of the carriers in the population?

**Question 5:** Two modes are known for producing offspring, namely, asexual and sexual reproduction. Describe the characteristics of each reproduction mode particularly from a viewpoint of mechanism and biological significance within 10 lines.

## Problem 10

Read the following sentence on RNA research, and answer the questions.

The central dogma was first enunciated by Francis Crick in the 1950s. This dogma is represented simply as “DNA makes RNA (transcription) makes proteins (translation)” and forms the backbone of molecular biology. In this process, it was recognized that the following three types of RNAs have critical roles as a mediator between DNA and protein: ( I ) an RNA that is transcribed and carries information from DNA, ( II ) an RNA that is one of the components of ribosome, ( III ) an RNA that transfers a specific amino acid to a growing polypeptide chain at the ribosomal site during protein synthesis (a). However, the central dogma has been revised, as new discoveries against the dogma were made later. In the 1970s, it was found that protein was not simply translated from primary RNA in eukaryotes (b), and a reverse flow from RNA to DNA was also observed in retroviruses containing RNA genomes (c). In the 1980s, an RNA molecule that catalyzes a chemical reaction (d) was discovered. Based on these unexpected discoveries, RNA has been proven to carry both genetic information and enzymatic activity. The hypothesis of ( e ) has been proposed, in which RNA is the central molecule in the origin of life. Currently, this is the most widely supported theory of the origin of life. As this theory became accepted, RNA came to be regarded as the most fundamental biopolymer, possessing the properties of both DNA and proteins. Under these circumstances, a small double-stranded RNA was found to inhibit gene expression in a sequence-specific manner by triggering degradation of complementary RNA. This phenomenon was termed as ( f ). Furthermore, a number of small RNAs known as microRNAs were discovered to be transcribed and negatively regulate gene expression in cells. RNA is similar to DNA, but there is a little difference in the chemical structure (g). Since RNA is less stable than DNA and shows a high chemical reactivity, RNA is regarded as a fairly flexible molecule. We are now in the age in which various important functions of RNA are just emerging.

### <Question>

- (1) Underline (a). Specify the type of RNA, ( I ) to ( III ), and the kind of RNA polymerase which transcribes each of them.
- (2) Underline (b). Explain the RNA processing classified into 3 steps.
- (3) Underline (c). Explain this flow using the name of the enzyme involved within about 2 lines.
- (4) Underline (d). State the name of RNA with catalytic activity, and give an example of such an RNA.
- (5) Fill the blank (e) with the most appropriate term.
- (6) Fill the blank (f) with the most appropriate term.
- (7) Underline (g). Explain the difference between DNA and RNA using the structural formula. You can use “R” for the bases.

## Problem 11

Answer the following questions.

Cholesterol is a hydrophobic molecule, and is transported in the blood in the form of lipid-protein particles known as (A). When (A) particles bind to specific receptors on the cell surface, they are incorporated into vesicles called clathrin-coated vesicles. Inside the cell, the clathrin-coated vesicles fuse with (B), where the (A) and the receptors are sorted. The receptors are returned to the plasma membrane, while the (A) is degraded in lysosomes.

When growth factors, such as epidermal growth factor (EGF), bind to the specific receptors on the cell surface, the EGF/EGF-receptor complexes are internalized into the cells by a similar mechanism. However, many of the EGF-receptors end up in lysosomes, where they are degraded along with the ingested EGF. The fate of the receptors, whether returned to cell surface or degraded in lysosomes, depends on the nature of the receptors.

In eukaryotes, a protease complex called the proteasome plays a major role in protein degradation in the cytoplasm. Proteasomes are huge protease complexes composed of more than 30 subunits, and degrade proteins tagged with (C). Proteasomes degrade (D) at a particular cell cycle phase, at which the level of (D) oscillates during cell cycle progression.

### Question 1.

Fill the blank (A)-(D) with appropriate words.

### Question 2.

Lysosomal enzymes do not work in the cytoplasm, even if they are released. Explain why not. Describe the differences in conditions between lysosomes and the cytoplasm, include a key protein to maintain the lysosomal condition.

### Question 3.

The cell-surface EGF receptors are transiently down-regulated in a mechanism described in paragraph 2. What is the advantage of this down-regulation of the cell-surface receptors?

### Question 4.

The nuclei of growing cells are stained with propidium iodide, and analyzed in a flow-cytometer to obtain a histogram as shown in Fig. 1. The peaks of the histogram are designated X, Y and Z, as shown in Fig. 2. Describe which cell cycle phase corresponds to each of the peaks X, Y and Z.

Fig. 1

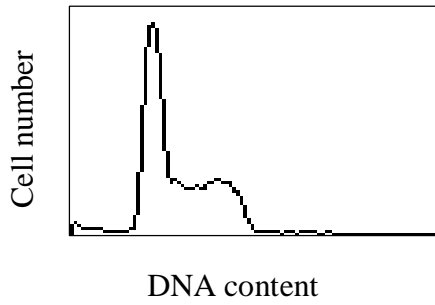
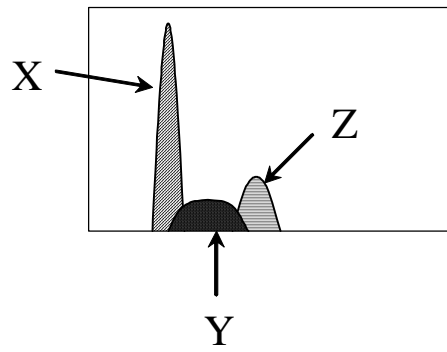


Fig. 2



**Question 5.**

Proteasome degrades denatured proteins and proteins with abnormal folding, to which the modification with (C) preferentially occurs as compared to normal proteins. Describe your opinion on the mechanism by which denatured proteins and proteins with abnormal folding are preferentially modified with (C).

## Problem 12

Read the following sentence and answer the questions.

Think of a body tissue that consists of two types of cells: A-cells and B-cells. A loss-of-function mutant for a gene encoding either Kashiwain or Odaibin lacks almost any A-cells in this tissue; almost all of the cells become B-cells (Figure 1). To analyze the function of these genes, the following experiments are performed.

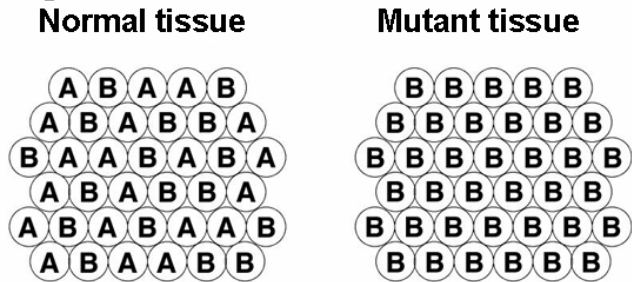
**【Experiment A】** To reveal the cells that express these genes in the tissue of a normal (wild-type) animal, we made probes from the DNA sequence of the *kashiwain* and *odaibin* genes, and performed *in situ* RNA hybridization. We found that *kashiwain* is expressed only in the A-cells, whereas *odaibin* is expressed only in the B-cells (Figure 2). (Note that the cells expressing the genes are uniformly colored in gray, irrespective of the region and intensity of hybridization.)

**【Experiment B】** Next, we generated antibodies against Kashiwain and Odaibin proteins and used them to stain the normal tissue. The anti-Kashiwain antibody labeled only A-cells, but the anti-Odaibin antibody labeled stained all the tissues including the B-cells (Figure 3). (Note that the cells stained by the antibodies are uniformly colored in gray, irrespective of the region and intensity of staining.)

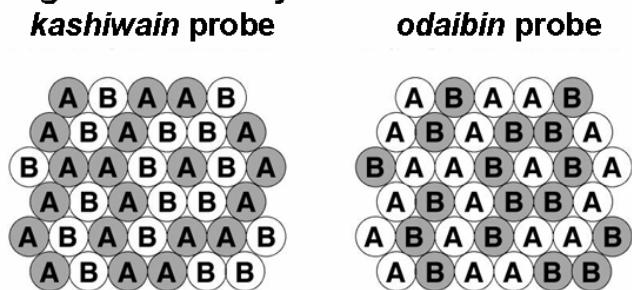
**Question 1:** Select a sentence consistent with the results of these experiments.

- A. Kashiwain proteins are localized in the A-cells, whereas Odaibin proteins are localized in the B-cells.
- B. Kashiwain proteins are localized in the A-cells, whereas Odaibin proteins are secreted from the B-cells.
- C. Kashiwain proteins are secreted from the A-cells, whereas Odaibin proteins are localized in the B-cells.
- D. Kashiwain proteins are secreted from the A-cells, whereas Odaibin proteins are secreted from the B-cells.

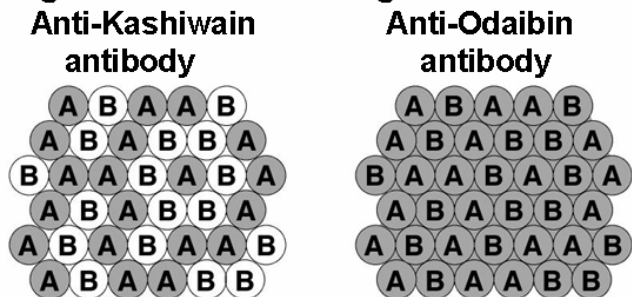
**Figure 1: Schema of the tissue**



**Figure 2: in situ hybridization**



**Figure 3: Immunostaining**





**【Experiment C】** The expression of *kashiwain* and *odaibin* genes start at the middle of the embryonic development. In both normal and mutant animals, the number of the cells in this tissue at this stage is already the same as those in the end of development. However, most of the cells at this stage are A-cells; there are only a few B-cells. During late development, the number of the A-cells decreases whereas that of the B-cells increases. In the normal tissue, the number of the B-cells becomes constant when it reaches a plateau. In the mutant tissue, however, the number of the B-cells continues increasing, and all the cells in the tissue become B-cells at the end of development.

We cut a small part of the tissue from the *odaibin* mutant animal at the middle embryonic stage and transplanted it into the tissue of the normal animal at the same developmental stage. At the end of development, both A- and B-cells were observed in the part of the tissue that is derived from the normal animal. Even in the part of the tissue that is derived from the mutant animal, not only B-cells but also A-cells were observed.

**Question 2:** Explain the reason why A-cells are observed even in the part derived from the mutant animal within 2~3 lines.

**【Experiment D】** Similarly, a part of the tissue was cut off from the *kashiwain* mutant animal and transplanted into the normal tissue.

**Question 3:** At the end of development, will the transplanted tissue have both A- and B-cells, only A-cells or only B-cells? Answer and explain the reasons why within 2~3 lines.

**Question 4:** Why is the number of the B-cells kept at a certain level in the normal animal? Explain the reason why within 2~3 lines.

**【Experiment E】** Next, we examined the expression level of *odaibin* gene in this tissue at the adult stage using a quantitative PCR method. The results demonstrated that the expression level of *odaibin* gene in the *odaibin* mutant is three-fold higher than that in the normal animal.

**Question 5:** Can we conclude that the expression level of *odaibin* gene per cell is increased in the *odaibin* mutant tissue based on these experiments? Answer and explain the reason why within 2~3 lines.

(このページは草稿用紙として使用してよい)  
(Blank page for draft)

(このページは草稿用紙として使用してよい)  
(Blank page for draft)