

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 22(2010)年度

2010 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 21 年 8 月 25 日(火)

Tuesday, August 25, 2009

13:30~15:30

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 14 ページです。落丁、乱丁、印刷不鮮明な箇所などがあつた場合には申し出ること。  
This problem booklet consists of 14 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 8 題出題されます。問題 1~8 から選択した合計4問に解答しなさい。ただし、問題 1~8 は同配点です。  
There are eight problems (Problem 1 to 8). Answer four problems out of the eight problems. Note that Problem 1 to 8 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.

(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

以下の文章を読んで設問に答えよ。

整数の配列  $A$  の  $l$  番要素から  $r$  番要素までを分割統治法によって整列する関数 `quicksort` を以下のように再帰的に定義する。

```
quicksort(int A[], int l, int r) {
    int v, i, j, t;
    if (r > l) {
        v = A[r];
        i = l - 1; j = r;
        for (;;) {
            do { i = i + 1; }
              while (A[i] < v);
            do { j = j - 1; }
              while (l <= j && A[j] > v);
            if (i >= j) break;
            t = A[i]; A[i] = A[j]; A[j] = t;
        }
        t = A[i]; A[i] = A[r]; A[r] = t;
        quicksort(A, l, );
        quicksort(A, , r);
    }
}
```

- (1) 配列  $A$  全体を整列するには関数 `quicksort` をどのように呼び出したらよいか。
- (2) 変数  $v$  の役割を説明せよ。
- (3) 関数定義中の  と  の部分にあてはまる内容とその役割を説明せよ。
- (4) 関数 `quicksort` による整列の最悪時間計算量について説明せよ。
- (5) 関数 `quicksort` による整列の平均時間計算量について説明せよ。

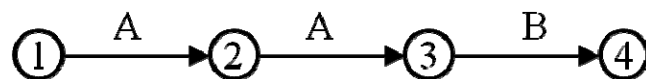
## 問題2

次の条件を満たす有向グラフを考える。

- 頂点数は有限で、閉路（ループ）を含まない。
- 各頂点  $v$  を始点とする辺  $(v, w)$  は高々 1 個しか存在せず、辺が存在しない場合は  $v$  を出口と呼ぶ。出口は唯一つ存在すると仮定する。
- 各辺には文字 A もしくは B がラベルされている。

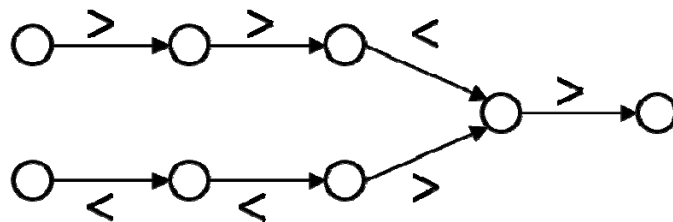
このとき以下の各問に答えよ。

- (1) 各頂点から有向辺に沿って頂点を訪問してゆくと、いつかは唯一の出口に到達することを証明せよ。
- (2) 各頂点  $v$  から有向辺に沿って頂点を訪問し、唯一の出口に到達するまでに辺にラベルされた文字を連結した文字列を  $S(v)$  と置く。 $v$  が出口の場合  $S(v)$  は空の文字列と定義する。以下のグラフでは  $S(4)$  は空の文字列、 $S(3) = B$ 、 $S(2) = AB$ 、 $S(1) = AAB$  である。



任意の辺  $(v, w)$  について記号列  $S(v)$  と  $S(w)$  は等しくないことを示せ。

- (3) 辞書式順序で文字列を比較する。たとえば  $AAB < AB < B$ 。なお、空の文字列を最小と定義する。辺  $(v, w)$  について  $S(v) > S(w)$  のとき辺は  $>$  タイプ、そうでないとき  $<$  タイプと呼ぶ。以下のグラフの各辺のタイプを満たすように、各辺に A もしくは B をラベルせよ。



- (4) グラフの各辺の文字が決まっていないとする。 $w$  が出口のとき辺  $(v, w)$  は  $>$  タイプである。それ以外の辺に  $>$  もしくは  $<$  のどちらのタイプを任意に割り当てても、それを満たす文字を各辺にラベルできることを証明せよ。
- (5) 文字がラベルされたグラフについて、辺の全てにタイプを割り当てる作業を考える。この作業を辺の数に線形比例する時間で実行するアルゴリズムを設計せよ。

### 問題3

DNA 複製の誤りなどにより、ゲノム配列は進化的時間とともにランダムな塩基置換をうける。ゲノムのある位置における変化を、以下の連続時間マルコフモデルで定義しよう。

$$P(b|a,t) = [\exp(tS)]_{b,a}$$

ただし  $t(\geq 0)$  は進化時間、自然数  $a, b \in \{1, 2\}$  はそれぞれ 1=プリン(アデニンまたはグアニン)、2=ピリミジン(シトシンまたはチミン)、 $P(b|a,t)$  は、時刻 0 で塩基  $a$  のとき、時刻  $t$  で塩基  $b$  が観測される確率をあらわす。行列  $S = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}$  ( $\alpha, \beta > 0$ ) は塩基間の置換速度で、

$[A]_{b,a}$  は行列  $A$  の  $b$  行  $a$  列要素を表す。 $\exp(A)$  は行列の指数関数を表し、以下で定義される。

$$\exp(A) = I + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{A^k}{k!} \quad (I: \text{unit matrix})$$

以下の問いに答えよ。

- (1) 方程式  $\det(A - \lambda I) = 0$  の解を行列  $A$  の固有値という。行列  $S$  の固有値  $\lambda_1, \lambda_2$  を求めよ。
- (2) 方程式  $(A - \lambda I)x = 0$  を満たす 0 でないベクトル  $x$  を  $A$  の固有ベクトルという。行列  $S$  の固有ベクトルを求めよ。

- (3) 固有ベクトルを並べた正則行列  $U$  を用いて、 $S = U \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U^{-1}$  と分解できる。この式を満

たす  $U, U^{-1}$  を求めよ。

- (4) 行列  $S$  の固有値分解が  $S = U \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U^{-1}$  のとき、 $\exp(tS)$  の固有値分解は

$$\exp(tS) = U \begin{pmatrix} \exp(t\lambda_1) & 0 \\ 0 & \exp(t\lambda_2) \end{pmatrix} U^{-1} \text{ となる。設問(3)で導いた } U, U^{-1} \text{ を使って、}$$

$\lim_{t \rightarrow \infty} P(b|a,t)$  を求めよ。

- (5) プリンのみで構成された DNA 塩基配列  $L$  を仮定しよう。十分時間の経った後の  $L$  中のプリンとピリミジンの存在比の期待値を答えよ。

#### 問題4

ゲノム情報処理のひとつに、非常に多くの塩基配列を類似配列のグループに分ける作業がある。議論を簡単化するため、等しい長さで A,C,G,T の 4 種からなる文字列のみを扱うと仮定して、以下の設問に答えよ。ただし、ハミング距離とは 2 つの文字列で文字が異なる位置の数とする。以下の設問に答えよ。

- (1) 長さ  $L$  のランダム文字列 2 本を先頭から列毎に比較したとき、ハミング距離が 2 となる確率を、理由とともに示せ。
- (2) 長さ  $L$  のランダム文字列 2 本を先頭から列毎に比較するとき、最後の文字まで比較する前にハミング距離が 2 以下とわかる確率を、理由とともに示せ。
- (3) 長さ  $L$  の文字列  $N$  本をソートする。 $L$  を定数としたとき、この処理を最悪計算量  $O(N)$  でおこなうアルゴリズムを示せ。
- (4)  $L$  が定数のとき、長さ  $L$  の  $N$  本の塩基配列群からハミング距離が 2 以下のペアをすべて求めるアルゴリズムとその時間計算量を、理由とともに示せ。



## 問題5

タンパク質中のアミノ酸残基の中心炭素原子(C)に対して、アミド基との結合(C - N 結合)の周りの二面角をファイ( )、カルボニル基との結合(C - C' 結合)の周りの二面角をプサイ( )であらわす。以下の設問に答えよ。

- (1) および 角の組み合わせのほとんどは、主鎖と側鎖の立体障害のために出現しにくい。この制約を比較的受けずに広範囲のコンホメーションを取れるアミノ酸を一つ上げ、その構造式と理由を示せ。
- (2) 相同タンパク質配列間において、設問(1)で答えたアミノ酸の割合が高くなる位置の例を、理由とともに示せ。
- (3) タンパク質の二次構造には並行シートと逆並行シートがある。それぞれの構造をN末端、C末端を明記して図示し、違いを説明せよ。
- (4) アミノ酸プロリンはヘリックスの中央部よりも末端に含まれる場合が多い。その構造を描いて理由を示せ。
- (5) 膜貫通型のヘリックスを形成しやすいアミノ酸とそうでないアミノ酸の物理化学的違いを述べよ。

## 問題6

以下の設問に答えよ。

- (1) 文章内の[1]~[8]に当てはまる語句または数値を語句リストより選択して答えよ。ただし、同じ語句または数値の複数回の使用は不可とする。

ヒトゲノム計画は1991年に開始され、日米欧の大学等の公的機関からなる国際チームと、米国のセセラ社によってそれぞれ独自に進められた。解読戦略として、国際チームは[1]の物理地図を元にした clone by clone 法を、セセラ社は物理地図を必要としない[2]法をそれぞれ用いた。今日、[2]法はゲノム解読のもっとも一般的な戦略となっている。国際チームは2003年にヒト染色体の[3]領域の約99%を高精度に配列決定し、その解析結果を2004年に *Nature* 誌に発表した。その達成には、サンガー法を原理として開発されたキャピラリー式 DNA シークエンサーを多数装備した大規模解析システムの確立によるところが大きい。ヒトゲノム配列の解析から、タンパク質をコードする遺伝子の数は約[4]個であることがわかった。

ヒト遺伝子の大部分が明らかになったことによって、ハイブリダイゼーションを原理とした[5]技術による定量的かつ網羅的な遺伝子発現解析や、SNPs を用いた相関解析による病気のリスク遺伝子のゲノムワイドな探索が可能になった。さらに最近では、[6]法等の新たなシークエンス技術を原理とした DNA シークエンサーが開発されている。この新型シークエンサーはキャピラリー式シークエンサーよりも少なくとも2桁以上高速の配列決定能力を有しており、これまでその実行が困難であった個人のゲノム解析や環境中の微生物群集を分離しないでそのまま解析する[7]解析等が今日可能となった。このように、近い将来、莫大な量のゲノム情報や関連生物情報を解析する上で、[8]はきわめて重要かつ強力な学問分野となる。

### 語句リスト

5000、12000、25000、100000、250000、バイオインフォマティクス、FISH、GFP、PCR、ノーザン、サザン、ウェスタン、DNA マイクロアレイ、アセンブリ、全ゲノムショットガン、細菌人工染色体、ヘテロクロマチン、ポリメラーゼ、制限酵素、エピゲノム、データベース、電気泳動、SDS-PAGE、パスウェイ、エキソン、マキサムーギルバート、ワトソン、ベンター、エドマン、質量分析計、メタゲノム、プラスミド、パイロシークエンシング、ペプチド、プライマーウォーク、蛍光色素、シンテニー、ユークロマチン、メチル化、プロテオミクス

- (2) サンガー法の原理を簡潔に説明せよ。  
(3) ゲノム配列からタンパク質をコードする部分を情報学的に同定/予測する方法を3つ挙げ、それぞれの方法を簡潔に説明せよ。  
(4) SNPs (または SNP) を簡潔に説明せよ。

\* 問題(2)~(4)については図を用いて説明しても良い。

## 問題7

神経細胞と内分泌細胞は、どちらも遠く離れた別の細胞に情報を伝えることができる。以下の設問に答えよ。

- (1) それぞれの細胞が情報を伝える仕組みを説明せよ。
- (2) 細胞が情報を伝える速さと範囲は両者でどのように異なるか。
- (3) 神経細胞と内分泌細胞が情報伝達に用いる物質の例をそれぞれ2つあげ、それらの機能を説明せよ。
- (4) 細胞表面から核内へ情報が伝わる仕組みの例を2つ挙げ、それぞれ説明せよ。
- (5) 情報伝達物質が細胞外に残留しつづけると新たな情報を受け取る際に支障が生じる。この問題に対して細胞はどう対処しているのか、候補を2つ考えて説明せよ。

## 問題8

以下の設問に答えよ。

：真核生物のゲノム DNA はヌクレオソーム構造を取っている。

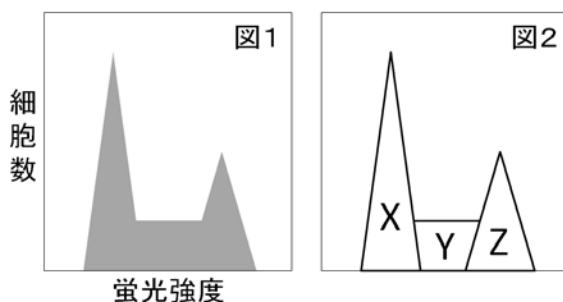
- (1) ヌクレオソーム・コア粒子を形成しているタンパク質を挙げよ。
- (2) ヌクレオソーム・コア粒子を形成しているタンパク質に最も特徴的なアミノ酸を2種類あげて、何故それらが多く含まれているのかを簡潔に説明せよ。
- (3) ヒトゲノム全体をヌクレオソーム・コア粒子で隙間なくカバーするためには、ヌクレオソーム・コア粒子を形成しているタンパク質が何分子必要になるか。ヒトゲノムのサイズを  $3 \times 10^9$  bp であるとして、考え方と計算の筋道を示しながら解答せよ。

：GTP を加水分解する GTP アーゼは細胞内で様々な重要な機能を果たしている。

- (1) タンパク質合成に関わる GTP アーゼを2種類挙げて、その機能を簡潔に説明せよ。
- (2) シグナル伝達に関わる GTP アーゼを2種類挙げて、その機能を簡潔に説明せよ。
- (3) GTP アーゼ活性を持つ細胞骨格タンパク質をひとつ挙げて、その機能を簡潔に説明せよ。

：細胞周期は G1、G2、M、S の4つのステージからなる。

- (1) 4つのステージを細胞周期の進行に沿って並べ替えよ。
- (2) 増殖中の細胞集団を DNA と結合する蛍光色素で染色して、フローサイトメトリーで解析したところ、図1のようなヒストグラムを得た。このヒストグラムを図2のように分解した場合、X、Y、Z にはそれぞれどのステージの細胞が含まれるかを答えよ。



：膜で囲まれた細胞小器官がタンパク質を取り込む輸送機構を3通りに分けて簡潔に説明せよ。ただし、それぞれの機構について、それが使われる細胞小器官名の例を挙げること。また、それぞれの機構について、タンパク質の構造をほどく必要が有るか無いかにも言及すること。

(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)