

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

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 24(2012)年度

2012 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 23 年 8 月 23 日(火)

Tuesday, August 23, 2011

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 12 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 12 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、裏面にわたつてもよい。  
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)

## 問題 1

以下の C 言語で記述された `heap_sort` 関数は、`{key,value}` のペアからなる配列 `a` をヒープソートアルゴリズムによりソートする。

```
typedef struct { int key; int value; } T;
void swap(T a[], int i, int j) { T t = a[i]; a[i] = a[j]; a[j] = t; }
int less(T a[], int i, int j) { return a[i].key < a[j].key; }
void sift_down(T a[], int i, int j) {
    int k;
    for (k = (2 * i + 1); k < j; k = (2 * i + 1)) {
        if ((k + 1 < j) && less(a, k, k + 1)) { ++k; }
        if (less(a, i, k)) { swap(a, i, k); i = k; } else { break; }
    } /* end for */
} /* end sift_down */
void heap_sort(T a[], int n) {
    int i, j;
    for (i = (n - 1); i >= 0; --i) { sift_down(a, i, n); }
    for (j = (n - 1); j > 0; --j) { swap(a, 0, j); sift_down(a, 0, j); }
} /* end heap_sort */
```

但し  $n > 0$  は配列 `a` のサイズである。以下の問に答えよ。

- (1) ヒープデータ構造を説明せよ。
- (2)  $a = \{\{4, 1\}, \{3, 2\}, \{3, 3\}, \{2, 4\}\}$ ,  $n = 4$  に対し `heap_sort(a, n)` を実行した。関数呼び出し後の配列 `a` の内容を求めよ。
- (3) `sift_down` 関数の役割を説明せよ。
- (4) ヒープソートの最悪計算量は、 $O(n \log n)$ であることを説明せよ。

問題 2

2つの異なる生物種のゲノム配列を比較して、進化的に保存された領域を高速に抽出するチェイニング法について考える。文字列  $P$  において、開始位置  $sx$  から終了位置  $ex$  ( $sx < ex$ ) まで連続した部分文字列を  $P[sx, ex]$  と表記する。文字列  $P, Q$  の部分文字列  $P[sx, ex], Q[sy, ey]$  が類似しているときアラインメントとよび、類似度をしめす実数スコア ( $score$  と表記) とともにアラインメント名  $A$  をつかって  $A = (P[sx, ex], Q[sy, ey], score)$  と表記する。各要素  $sx, ex, sy, ey, score$  は  $A.sx, A.ex, A.sy, A.ey, A.score$  と表記する。アラインメント  $A, B$  において  $A.ex < B.sx$  かつ  $A.ey < B.sy$  ならば  $A < B$  と定義する。アラインメント  $A_k$  に対して昇順列  $A_1 < A_2 < \dots < A_k$  ( $k \geq 1$ ) を1つ決めて  $A_k$  の鎖とよぶ。このとき  $\sum_{i=1, \dots, k} A_i.score$  を鎖のスコア ( $A_k.chain\_score$  と表記)、 $k$  を鎖の長さとしよぶ。大きなスコアをもつ鎖は進化的に保存された領域を示唆する。

以下の設問に答えよ。

- (1) 図1の例で、長さ3の鎖を列举せよ。
- (2) 図1の例で、各アラインメントの鎖で最大スコアを与える鎖を示せ。

チェイニング法: 各アラインメント  $A$  について、 $A$  の鎖を  $A$  自身として初期化し2つの組  $(A.sx, A)$ ,  $(A.ex, A)$  をリスト  $X$  に入れる。組の第1引数で  $X$  を昇順に並べる。 $(C.ey, C)$  の形の組をいれるリスト  $Y$  を用意し空にする。 $X$  が空になるまで先頭の要素  $(x, B)$  を削除して以下のステップを繰り返す。

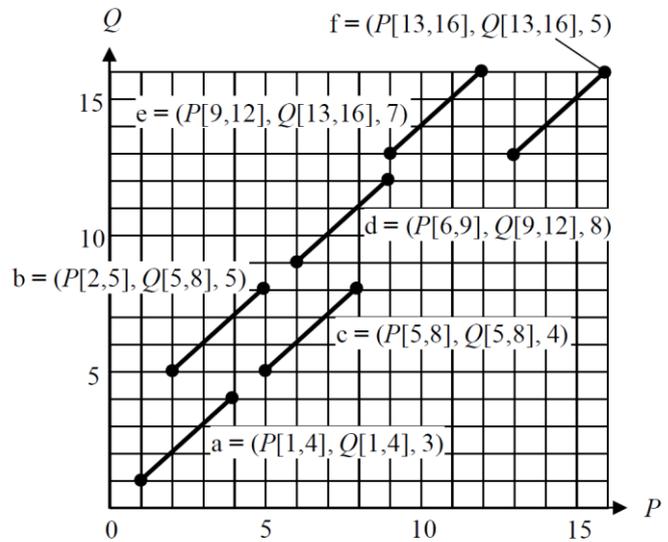


図 1

- I.  $x$  が開始位置  $B.sx$  の場合、 $C < B$  となる組  $(C.ey, C)$  を  $Y$  が含むとき、 $C.ey$  は  $B.sy$  に最も近い値と仮定し、 $C$  の鎖の最後に  $B$  を連結して、あらたに  $B$  の鎖とする。
  - II.  $x$  が終了位置  $B.ex$  の場合、 $C.ey \leq B.ey$  かつ  $C.chain\_score \geq B.chain\_score$  となる組  $(C.ey, C)$  が  $Y$  に存在しなければ、 $(B.ey, B)$  を  $Y$  へ追加し、 $Y$  を組の第1引数で昇順に並べかえ、 $B.ey \leq D.ey$  かつ  $B.chain\_score > D.chain\_score$  となる組  $(D.ey, D)$  はすべて  $Y$  から除く。
- (3) 図1の例にチェイニング法を適用したときのアルゴリズムの動作を説明し、このアルゴリズムが各アラインメントの鎖で最大スコアを与える鎖を計算することを示せ。
  - (4) ステップ II の処理後、リスト  $Y$  の中の組  $(C.ey, C)$  は、第1引数で昇順に並んでおり、かつ  $C$  の鎖のスコアでも昇順に並んでいることを説明せよ。アラインメント数が  $n$  のとき、どのようなデータ構造をどのように用いれば、チェイニング法の最悪計算量を  $O(n \log n)$  にできるか述べよ。

### 問題 3

DNA 配列の 4 種類の塩基をシーケンサーが読み取る場合には、読み取り誤りが起こることがある。本問題で考えるシーケンサーでは、塩基を読み飛ばしたり、余分な読み取り結果を出力したりすることはないが、塩基が読み間違えられる確率は、次のようになることが分かっている。

A を C、C を A、G を T、T を G に読み間違える確率は、それぞれ等しく  $p$ 。

それ以外の組み合わせの読み間違いの確率は、それぞれ等しく  $q$ 。

たとえば、A は確率  $p$  で C に、確率  $q$  で G に、確率  $q$  で T に読み間違えられる。このとき、以下の問いに答えよ。

- (1) 真の配列が ATGGCT であるときに、誤って ATATCT が読み取られる確率を求めよ。
- (2) A, T, G, C が等確率 (=25%) で出現すると仮定する。2 塩基を読んだ時、どちらも C と読み取られた。真の 2 塩基が同一 (必ずしも C でなくてよい) である確率を求めよ。
- (3) A, T, G, C が等確率 (=25%) で出現すると仮定する。長さ 3 の配列 2 種類が、どちらも ATG と読み取られた。真の 2 種類の配列が完全には同一でない確率を求めよ。
- (4) A, T, G, C である確率 (事前確率) がそれぞれ 20%、20%、30%、30% であると仮定する。1 塩基を読んだところ C が得られた。これが正しい結果である確率を求めよ。

問題 4

Operating System (OS) のプロセススケジューリングアルゴリズムについて考える。中央演算装置 (CPU) コアは1つであると仮定する。また、以下の表に示す順序と時刻でプロセスの実行要求が OS に到着し、各プロセスの CPU バースト時間は正確に見積もられている。ただし、あるプロセスの CPU バースト時間とは、そのプロセスを完了するために CPU が動作する合計の時間であると定義する。OS は単一のプロセススケジューリングキュー (プロセス待ち行列) を持つものとする。

プロセス	到着時刻 (ms)	CPU バースト時間 (ms)
$P_1$	0	135
$P_2$	20	35
$P_3$	100	90

- (1) First-Come, First-Served (FCFS) スケジューリングとは、先に到着したプロセスが後に到着したプロセスより優先されるスケジューリングアルゴリズムである。FCFS スケジューリングを行った場合に、プロセス  $P_2$  が終了する時刻を答えよ。
- (2) Shortest-Remaining-Time-First (SRTF) スケジューリングとは、見積もられた残り CPU バースト時間が最も短いプロセスが先に実行されるスケジューリングアルゴリズムである。また、新しく到着したプロセスの CPU バースト時間が最も短い場合には実行中のプロセスを保留して新しく到着したプロセスを代わりに実行する。上記の例において SRTF スケジューリングを行った場合に、FCFS スケジューリングと比べてプロセスが投入されてから実行が終了するまでの平均時間 (以降、平均待ち時間と表記する。) がどれくらい短くなるかを答えよ。
- (3) 一般に SRTF スケジューリングにおいては、プロセスの実行が終了するまでの平均待ち時間が最小化されるという利点がある。しかし、現実世界で実行されているプロセスの性質を考えると SRTF スケジューリングで平均待ち時間を改善することは難しいことが多い。その理由を述べよ。
- (4) Round-Robin (RR) スケジューリングとは、FCFS スケジューリングを拡張したスケジューリングアルゴリズムである。あるプロセスが予め定められた一定の連続した時間 (タイムスライスと呼ばれる) 実行された場合、プロセススケジューリングキューで次に待っているプロセスが取って代わり、取って代わられたプロセスはプロセススケジューリングキューの末尾に配置される。上記の例を用いた場合に、タイムスライスが 50 ms の RR スケジューリングを行うと、時刻と CPU 上で実行中のプロセスの関係がどのようになるか経時的な関係を図示せよ。
- (5) cache-coherent Non-Uniform Memory Access (cc-NUMA) アーキテクチャでコア数が多いコンピュータにおいては、OS が待機中のプロセスを割り付けることのできる空き CPU があっても、プロセススケジューリングキュー中で待っているプロセスを待機させることがある。cc-NUMA の一般的なアーキテクチャを説明したうえで、なぜこのような戦略でプロセスの実行を早く終えることができる場合もあるのか、その理由を述べよ。

## 問題5

原核生物について以下の問に答えよ。

- (1) 原核生物の遺伝子構造、mRNA 構造、転写機構に関して真核生物との相違点を3つ挙げよ。
- (2) 大部分の原核生物のゲノムサイズは百数十 kb から 10Mb の範囲にある。この原核生物のゲノムサイズの違いを生じる要因（進化様式）について説明せよ。
- (3) A の語句（1～3）に最も関係の深い語句を B に挙げた 9 つの語句からそれぞれ1つ選べ。

A

1. Operational taxonomic unit (OTU)
2. Pribnow Box
3. 志賀毒素

B

シャイン - ダルガノ配列、16S リボソーム RNA 遺伝子、孢子、RNA ポリメラーゼ認識部位、GC skew、回文配列、遺伝子の水平伝播、複製開始点、共生

- (4) C の記述（1～3）にもっとも関係の深い細菌を D にリストした 9 つの細菌からそれぞれ1つ選べ。

C

1. この細菌のゲノムは 1995 年に世界で最初に全解読された。
2. この土壌細菌は直鎖状の染色体を持ち、そこには抗生物質などの多くの二次代謝物の生合成遺伝子をコードしている。
3. この細菌はヒト腸内の常在細菌の一種であり、分類学的にアクチノバクテリア門に属する。

D

*Escherichia coli* (大腸菌), *Bacillus subtilis* (枯草菌),  
*Bifidobacterium adolescentis* (ビフィズス菌), *Clostridium tetani* (破傷風菌),  
*Buchnera aphidicola* (ブフネラ菌), *Haemophilus influenzae* (インフルエンザ菌),  
*Streptomyces griseus* (放線菌), *Helicobacter pylori* (ピロリ菌),  
*Staphylococcus aureus* (黄色ブドウ球菌)

## 問題 6

RNA は様々な機能を有する生体高分子である。ほ乳類の細胞には 10-30pg の RNA が存在し、この量は細胞成分の全重量の約 1 % に相当する。RNA は大きくタンパク質をコードする RNA (mRNA) とタンパク質をコードしていない非コード RNA の 2 つに分類できる。非コード RNA は機能性 RNA とも呼ばれ、主に (a)、(b)、(c)、マイクロ RNA などに分類される。マイクロ RNA は、フォールドバック RNA という分子内の塩基対形成により生じたいくつかのヘアピン構造を持つ前駆体として RNA ポリメラーゼ II によって転写される。生成したフォールドバック RNA は核内で (d) という酵素によって個々の pre-miRNA であるヘアピンへと切断され、細胞質に輸送される。ついで、細胞質において pre-miRNA は (e) という酵素によって切断されて 21-25 塩基のマイクロ RNA が生合成される。マイクロ RNA は (f) という生物学的機能を有する。

下記の問題について答えよ。

- (1) 特定の mRNA を定量する方法を 1 つ挙げ説明せよ。
- (2) 機能性 RNA を 3 つ (a)、(b)、(c) を挙げてその機能を説明せよ。
- (3) マイクロ RNA の生合成経路における (d)、(e) の酵素名を答えよ。
- (4) (f) のマイクロ RNA の機能を数行以内で説明せよ。

## 問題 7

DNA 損傷と修復に関する以下の問いに答えよ。

- (1) I に挙げた DNA 損傷にそれぞれ最も関係する塩基を II から選び、その結果として最も典型的なものを III から選べ。

I. 脱アミノ反応, チミン二量体, 脱プリン反応

II. A, T, C

III. 鎌状赤血球貧血症, G-C 塩基対が欠失, G が A に変化, A-T 塩基対が欠失, 色素性乾皮症, C が T に変化

- (2) 1 塩基が損傷した際の DNA 修復に最も関係する酵素を IV から 3 つ選び、それらが機能する順番を答えよ。

IV. テロメラーゼ, DNA ポリメラーゼ, DNA リガーゼ, RNA ポリメラーゼ, DNA フォトリアーゼ, ヌクレアーゼ

- (3) DNA 二本鎖切断は非相同末端結合および相同組み換えによって修復できる。これら 2 つのメカニズムと期待される結果を比較し、それらを簡潔に説明せよ。

## 問題 8

(1) ヨウ素について以下の問いに答えよ。

- (A) 放射性ヨウ素を吸い込むと甲状腺ガンになりやすくなる。なぜ甲状腺に特に影響が出やすいのか、その理由を数行程度で説明せよ。
- (B) 甲状腺ガンの治療法のひとつに、放射性ヨウ素を服用する方法がある。なぜこれによって甲状腺ガンが治療しうるのかを数行程度で説明せよ。

(2) 慢性骨髄性白血病 (CML, chronic myelogenous leukemia) は、10 万人に 1~2 人の割合で発症する、全白血病の約 2 割を占める疾患である。この疾患では、骨髄の造血幹細胞が異常に増殖して血液中の白血球数が増大する。ある女性患者が慢性骨髄性白血病と診断された。この患者の細胞の染色体を核型分析すると、白血病の細胞では 9 番と 22 番染色体の一部が互いに入れ替わった相互転座を有する染色体が、ヘテロ接合の状態を観察された。一方、同じ患者のそれ以外の細胞では、このような転座を持つ染色体は見られなかった。

- (A) この転座によって生じた突然変異は、優性と表現すべきか？劣性と表現すべきか？
- (B) この患者の白血病細胞を調べると、通常は別々の分子として観察される 2 つのタンパク質が連結したような形の特殊なタンパク質が観察された。このタンパク質が作られる分子機構を 1~2 行で説明せよ。
- (C) この患者がその後妊娠した。生まれる子どもが慢性骨髄性白血病を発症する確率は、この病気を持たない女性と比べてどれほどだと考えられるか、またその理由を 1~2 行で説明せよ。

(3) 酸素には、原子量 16 の一般の酸素  $^{16}\text{O}$  の他に原子量 18 の同位体  $^{18}\text{O}$  があり、その微小な質量の違いを利用してこれらを区別できる。通常の大気の酸素や二酸化炭素分子には、 $^{16}\text{O}$  だけが含まれている。

- (A) 通常の大気に  $^{18}\text{O}$  の酸素分子を混合した空気を密閉ケースの中に入れて、その中にネズミを入れて 10 分間呼吸させた。このケースから空気を回収し、質量分析器で分析して、どのような二酸化炭素分子が含まれているかを調べた。 $^{16}\text{O}$  の二酸化炭素分子と  $^{18}\text{O}$  の二酸化炭素分子が、どのような割合で存在するだろうか？呼吸の分子機構から理由を数行程度で説明して答えよ。
- (B) 通常の大気に  $^{18}\text{O}$  の二酸化炭素を混合した空気を密閉ケースの中に入れて、その中にシロイヌナズナの植物を入れ、光を当てて 6 時間光合成をさせた。このケースから空気を回収し、質量分析器で分析して、どのような酸素分子が含まれているかを調べた。 $^{16}\text{O}$  の酸素分子と  $^{18}\text{O}$  の酸素分子が、どのような割合で存在するだろうか？光合成の分子機構から理由を数行程度で説明して答えよ。

(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)