

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

東京大学 大学院新領域創成科学研究科
Graduate School of Frontier Sciences, The University of Tokyo
メディカル情報生命専攻
Department of Computational Biology and Medical Sciences
令和 6 (2024) 年度
2024 School Year
大学院入学試験問題
Graduate School Entrance Examination Question Booklet

専 門 科 目
Specialties

令和 5 年 8 月 1 日 (火)
Tuesday, August 1, 2023

10:00~12:00

注意事項 Instructions

- 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 booklet until the start of examination is announced.
- 本冊子の総ページ数は 33 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 booklet consists of 33 pages. If you find missing, misplaced, and/or unclearly printed pages, notify it to the staff.
- 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。
Only black pencils (or mechanical pencils) are allowed to answer the questions.
- 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 12 exam questions (Question 1 to 12). Answer 4 questions out of the 12 questions. Note that Question 1 to 12 are equally weighted.
- 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 4 answer sheets. You must use one answer sheet for each question. You may continue to write your answer on the back of the answer sheet if you cannot conclude it on the front. But you must not proceed to write on the second sheet.
- 解答は日本語または英語で記入しなさい。
Answers should be given in Japanese or in English.
- 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 question number you are to answer. Fill the designated blanks at the top of this page with your examinee number.
- 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 booklet.
- 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.
- 解答できない場合でも、解答用紙すべてに受験番号を記入して提出しなさい。
Turn in the answer sheet with your examinee number, even if you cannot solve the question.
- 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 this 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)

問題 1

A. 以下の (1) から (10) の問いに答えなさい。

(1) 中性溶液中で正に荷電した側鎖を持つリジン以外の天然アミノ酸を 2 つ答えなさい。

(2) 側鎖に R1、R2 を持つ 2 つのアミノ酸からなるジペプチドの化学構造を記しなさい。

(3) 以下のアミノ酸配列を持つタンパク質をトリプシンで完全分解した時に得られるペプチドのうち、2 番目に長いペプチド配列を記しなさい。

SYSMEHFRWKGKPVGKKRRPVKVYPDAGEAQSAAEAFPLEF

(4) 0.02 M の水酸化ナトリウムの溶液の pH を求めなさい。Log₁₀5 = 0.7 として計算しなさい。

(5) 光路長が 1 cm の吸光度計で 0.015 M のタンパク質溶液の吸光度を測定するとその値は 0.9 だった。このタンパク質のモル吸光係数を求めなさい。単位も記すこと。

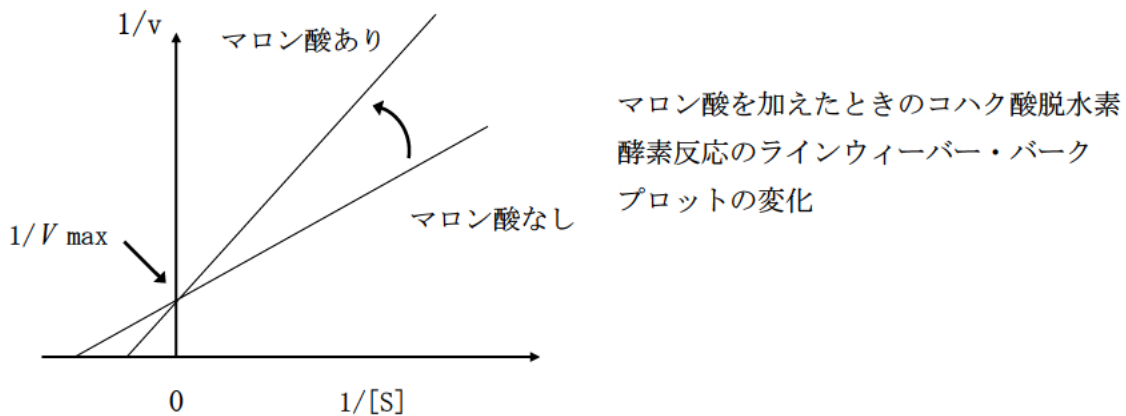
(6) 塩基部分にアデニン (A) を持つヌクレオチドが 2 つ連結した RNA 分子の化学構造を記しなさい。ただし、塩基部分の構造は簡略化して A と記してよい。

(7) DNA と RNA ではどちらがアルカリ性水溶液中で容易に加水分解されるか。100 文字程度で理由も説明しなさい。

(8) DNA 二重鎖は純水中と 0.5 M 食塩水中ではどちらがより安定か。100 字程度で理由とともに答えなさい。

(9) Michaelis-Menten の速度論に従うある酵素反応において、K_m が 1.0×10^{-5} M と算定された。また、基質濃度が 0.1 M のとき、この酵素反応の反応初速度が 1.0×10^{-6} mol/min であった。基質濃度が 1 μM のとき、この酵素反応の反応初速度を求めよ。

(10) コハク酸脱水素酵素はコハク酸をフマル酸へ酸化する。コハク酸の濃度[S]を変えた時のコハク酸脱水素酵素によるフマル酸の合成速度 v をラインウィーバー・バーク プロット (Lineweaver-Burk plot) で解析した。この反応にマロン酸を加えて活性を測定すると次ページの図のようなグラフの変化を得た。マロン酸の構造の特徴を踏まえて、その作用機序を 100 文字程度で説明しなさい。



B. 以下の（１）から（４）の問いに答えなさい。

（１）真正細菌と真核生物の mRNA の違いを以下の４つの語句を全て用いて 100 字程度で説明しなさい。

キャップ構造、シストロニック、スプライシング、ポリ A 鎖

（２）ヒトでは細胞に発現しているタンパク質は 5 ～10 万種類と見積もられている。一方、遺伝子（タンパク質をコードしている領域）の総数はタンパク質の種類数より少ない 2.2 万程度である。1 つの細胞の中の遺伝子とタンパク質の種類数が異なる理由を 100 字程度で説明しなさい。

（３）以下は、あるレトロウイルスの複製に必要なタンパク質 α をコードする遺伝子の open reading frame (ORF)（センス鎖）の一部を記したものである。以下の(i) ～ (iv)の問いに答えなさい。また次ページの遺伝暗号表を参照しなさい。

……CATTAGCCCTATTGAAACTGTACCGTAAAACTAAAGCCAGGAATGGATGGCCCAAA……

(i) レトロウイルスの生活環を次の３つの語句を全て用いて説明しなさい。

宿主細胞、ゲノム DNA、組み込まれる

（問題 1：次ページに続く）

(ii) 下の①から③で示される読み枠のうち、タンパク質 α はどの読み枠でコードされているか。ただしフレームシフトは起こらないこととする。

① ……CAT TAG CCC TAT TGA AAC TGT ACC AGT AAA ACT AAA GCC AGG AAT GGA TGG CCC AAA……

② ……CA TTA GCC CTA TTG AAA CTG TAC CAG TAA AAC TAA AGC CAG GAA TGG ATG GCC CAA A……

③ ……C ATT AGC CCT ATT GAA ACT GTA CCA GTA AAA CTA AAG CCA GGA ATG GAT GGC CCA AA……

(iii) 次に示す変異が起きたとき、ウイルス増殖にどのような影響が出るか 50 字程度で理由をつけて答えなさい。

[T]が欠失した。

……CATTAGCCC [T] ATTGAAACTGTACCAGTAAAACTAAAGCCAGGAATGGATGGCCCAA……
 ↓ [T]が欠失
 ……CATTAGCCC [] ATTGAAACTGTACCAGTAAAACTAAAGCCAGGAATGGATGGCCCAA……

(iv) 次のような変異が起きたとき、ウイルス増殖にどのような影響が出るか 50 字程度で理由をつけて答えなさい。

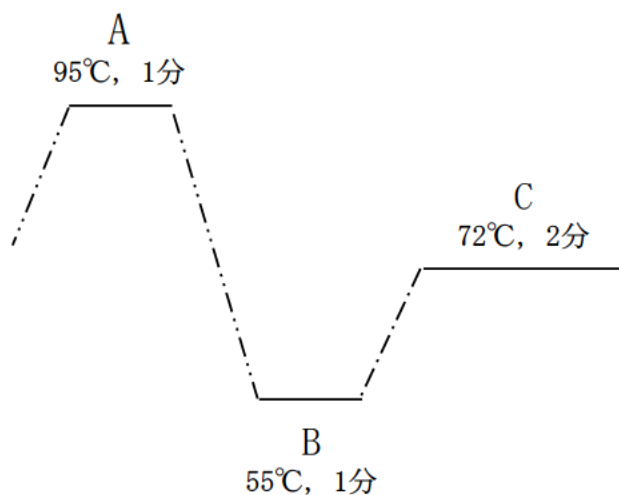
[T]が[A]に置換された。

……CATTAGCCC [T] ATTGAAACTGTACCAGTAAAACTAAAGCCAGGAATGGATGGCCCAA……
 ↓ [T]が[A]へ置換
 ……CATTAGCCC [A] ATTGAAACTGTACCAGTAAAACTAAAGCCAGGAATGGATGGCCCAA……

	Phe	C	Ser	A	yr	G	Cys
C	Leu	CC		AC		GC	
A		CA		AA	S OP	GA	S OP
G		CG	AG	GG	rp		
C		CC	CA	His	CG	Arg	
C C	CCC	CAC	Gln	CGC			
C A	CCA	CAA		CGA			
C G	CCG	CAG		CGG			
A	Ile	AC	hr	AA	Asn	AG	Ser
A C		ACC		AAC	AGC		
A A		ACA		AAA	Lys	AGA	Arg
A G	Met	ACG		AAG		AGG	
G	Val	GC	Ala	GA	Asp	GG	Gly
G C		GCC		GAC		GGC	
G A		GCA		GAA	Glu	GGA	
G G		GCG		GAG		GGG	

遺伝暗号表

(4) 図に示すような温度変化を1サイクルとしたPCRを15サイクル行い、ある特定の2本鎖DNA断片を増幅し、114 ngの試料DNAを得た。以下の(i)、(ii)の問いに答えなさい。



(i) 図のA、B、Cのステップで起きている反応について、以下の語句を全て用いてそれぞれ30字程度で説明しなさい。

dNTP、*Taq DNA* ポリメラーゼ、1本鎖、2本鎖、アニーリング、プライマー

(ii) はじめに用意した特定の2本鎖DNA断片の質量をpg単位で答えなさい。ただし、全ての反応が完全に進行したとする。

(問題1 終わり)

問題 2

A. 以下の(1)から(5)の問いの答えとして適切なものをAからDの4つの選択肢から全て選び記号で答えなさい。ただし、適切なものが存在しない場合には「なし」と答えなさい。

- (1) 真核生物において、核とは異なる独自のゲノムを維持している細胞小器官はどれか。
- A. 葉緑体
 - B. ゴルジ体
 - C. ミトコンドリア
 - D. ペルオキシソーム
- (2) 大腸菌株をプラスミド DNA pUC19 の導入により形質転換して、プラスミドを保有する大腸菌のみを取得したい。そのためにプラスミドに必要な不可欠な要素はどれか。
- A. テロメア
 - B. 選択マーカー遺伝子
 - C. マルチクローニングサイト
 - D. pMB1 複製起点
- (3) ゲノムのある場所から別の場所へ動く遺伝因子はどれか。
- A. 補酵素
 - B. 偽遺伝子
 - C. トランスロコン
 - D. トランスポゾン
- (4) 折り畳みに誤りがあるなど、異常なタンパク質の選択的分解を担う細胞内複合体はどれか。
- A. エンドソーム
 - B. ヌクレオソーム
 - C. プロテアソーム
 - D. リボソーム
- (5) RNA 分子とタンパク質分子から構成される細胞内複合体はどれか。
- A. リボヌクレアーゼ P
 - B. シグナル認識顆粒 (Signal Recognition Particle)
 - C. スプライソソーム
 - D. テロメラーゼ

B. 以下の(1)から(5)の記述は正しいか、誤りか答えなさい。正しい場合には「正」と答えなさい。誤りの場合には下線部のみを修正した正しい文章を作り、(1) ① - DNA、③ - タンパク質のように、修正した部位を全て答えなさい。

- (1) DNA ポリメラーゼが担う DNA 複製反応でも、RNA ポリメラーゼが担う転写反応でも、伸長される核酸の末端にあるヌクレオシドの 5'位のヒドロキシ基 ① と、基質となるヌクレオシド 三リン酸 ② の 3'位のリン酸基 ③ の間に ホスホジエステル結合 ④ が生じる。
- (2) 多くの DNA ポリメラーゼは DNA を末端から分解する エキソヌクレアーゼ ① 活性をもつ。3'から5' ② の方向の分解は合成中の DNA 鎖に誤って共有結合されたヌクレオチドを除去する 校正 ③ 機構を担う。
- (3) 多くの真核細胞では DNA ポリメラーゼにより染色体 DNA の複製が行われる時期は限られており、これを細胞周期の S 期 ① と呼ぶ。一方で、細胞分裂が行われる時期を M 期 ② と呼び、DNA 複製から細胞分裂の間の期間を G1 期 ③、細胞分裂から DNA 複製の間の期間を G2 期 ④ と呼ぶ。
- (4) mRNA を鋳型にしたタンパク質の翻訳反応では、伸長される新生ペプチドの末端にあるアミノ酸残基の アミノ基 ① と、mRNA ② に共有結合したアミノ酸の カルボキシ基 ③ の間に ペプチド結合 ④ が生じる。
- (5) アミノアシル tRNA 合成酵素 ① は tRNA とアミノ酸の結合に ATP ② を消費する。翻訳の伸長因子 EF-G や eEF-2 は ATP ③ を消費することでリボソーム上の翻訳反応を進行させる。

(問題 2 : 次ページに続く)

C. 以下の（１）から（５）の問いに答えなさい。

- (1) 大腸菌 K-12 株から単離された DNA ポリメラーゼ I は DNA 配列から 928 個のアミノ酸残基により構成されると考えられる。この遺伝子の開始コドンから終止コドンまでをコードする DNA の長さは何塩基か。その数を答えなさい。
- (2) この DNA ポリメラーゼ I の分子機能に重要なアミノ酸残基を検証するためにアラニンスキニング法を用いて変異体を作成したい。この遺伝子を発現させる DNA 断片をもとに、このタンパク質中の 1 箇所のアミノ酸残基だけがアラニン残基に置換された点変異体を全部で何種類作ることができるか。その数を答えなさい。なおこの DNA ポリメラーゼ I は 99 個のアラニン残基を含むものとする。
- (3) アミノ酸をコードしているコドンの突然変異によって終止コドンが出現することをナンセンス変異と呼び、なかでも UAG が出現する変異をアンバー変異と呼ぶ。アミノ酸をコードしているコドンのうち、1 塩基のみの突然変異によりアンバー変異となりうるコドンは何種類存在するか。以下の表を参考にその数を答えなさい。

表 遺伝暗号表

	Phe	C	Ser	A	yr	G	Cys	
C		CC		AC				GC
A	Leu	CA		AA		S OP		GA
G		CG	AG		GG		rp	
C		CC	CA	His	CG	Arg		
C C		CCC	CAC		CGC			
C A		CCA	CAA		Gln		CGA	
C G	CCG	CAG	CGG					
A	Ile	AC	hr	AA	Asn		AG	Ser
A C		ACC		AAC		AGC		
A A		ACA		AAA	Lys	AGA	Arg	
A G	Met	ACG	AAG	AGG				
G	Val	GC	Ala	GA	Asp	GG	Gly	
G C		GCC		GAC		GGC		
G A		GCA		GAA	Glu	GGA		
G G		GCG		GAG		GGG		

- (4) 免疫システムでは細胞が積極的に DNA に変異を生じさせている。脊椎動物の B 細胞における変異導入の第一段階では、DNA 鎖上の塩基を修飾する反応によりウラシル残基を生じさせる。どのような酵素活性によりどのような塩基がウラシルに変換されるか、60 文字程度で簡潔に説明しなさい。
- (5) 細胞は生じた変異を修復する機構を備えている。二本鎖 DNA 上のウラシル残基を修復する塩基除去反応ではどのような酵素反応を経て変異が修復されるか、120 文字程度で簡潔に説明しなさい。

- (4) 大腸菌の親株からラクトース非存在下でもラクトースオペロンが抑制されない変異体を多数分離したところ、領域 A および領域 B それぞれに一定数の変異が検出された。このうち、領域 A に変異を持つ変異体は、影響を受ける分子機構上、複数のタイプに分類されることが明らかになった。そのうち、2つのタイプの変異体についてどのようなものが推定できるか、それぞれ 120 文字程度で簡潔に答えなさい。

(問題 2 終わり)

問題 3

A. 以下の (1) ~ (10) までの問いの文章を読み、全て答えなさい。

(1) ヒト常染色体上のある座位に 5 種類のアレルがある時、可能な遺伝型の種類はいくらか。計算とともに示すこと。

(2) ある遺伝子周囲の非コード配列の遺伝的バリエーションの中で、その遺伝子産物であるタンパク質のアミノ酸配列を改変するバリエーションの種類を一つ挙げなさい。

(3) ゲノム上の遺伝的距離と物理的距離が異なる理由はなぜか。50 文字程度で答えなさい。

(4) 両親が同じであるきょうだいのゲノム上のある位置を適当に選んだ時、完全に同じ祖父母由来 (同祖) の染色体の組み合わせである確率はどれくらいか。

(5) 単一遺伝子疾患において、リスクとなる遺伝型を持っているときに疾患を発症する確率をなんと呼ぶか。

(6) 2022 年、長鎖シーケンス技術の活用により、より完全なヒトゲノム配列の解読が行われた。これまでのシーケンス技術では解読が困難であった領域が数多く解読された。これらの新しく解読された領域のうち、染色体の中ほどに位置する反復配列の豊富な領域を何と呼ぶか答えなさい。

(7) エクソン領域の DNA を濃縮して網羅的に解読する手法を全エクソームシーケンス解析という。この手法について次の中から最もあてはまらないものを選びなさい。

- A) 同じ量のシーケンスデータを取得した場合、全ゲノムシーケンスと比べて、標的領域あたりのシーケンス深度 (depth) は大きくなる。なお、「シーケンス深度」とは、標的領域の各塩基が読み取られた回数の平均値を指す。
- B) 転写産物のスプライシングパターンを網羅的に調べることができる。
- C) エクソン領域の DNA を濃縮する手法として、エクソン領域に設計されたプローブとのハイブリダイゼーションがよく使われている。
- D) タンパク質コード領域を解読できるので、アミノ酸置換を伴う多型や変異の検出を効率よく行うことができる。

(問題 3 : 次ページに続く)

(8) ヒト細胞において、ヒストン H3 における 27 番目のリジン残基のアセチル化(H3K27ac)のパターンを解析したところ、ある遺伝子の転写開始点付近にこのヒストン修飾が見られた。この遺伝子の転写は一般に活性化されているか、抑制されているか、答えなさい。

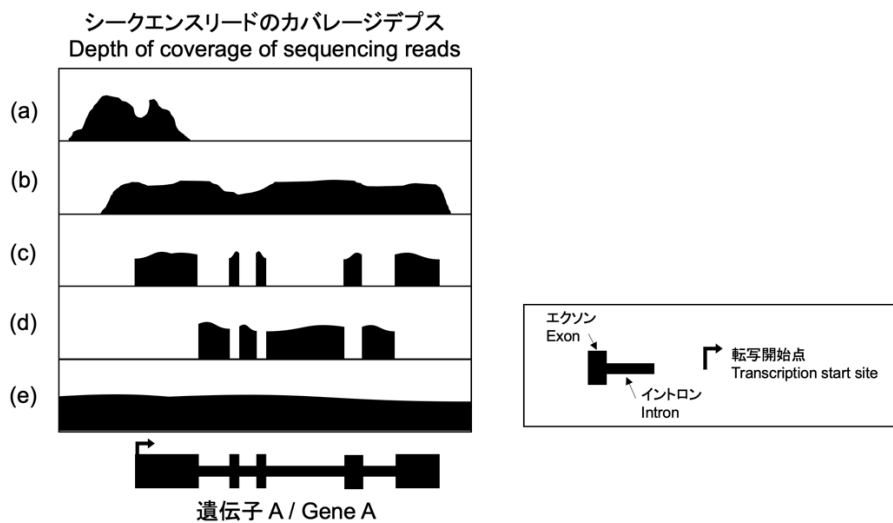
(9) 次の文章を読み、(a)~(e)に最も当てはまる語句を下記の語群から選びなさい。

ジデオキシ法は、(a)法とも呼ばれる塩基配列決定法である。この手法では、(b)が(c)の代わりに取り込まれた場合に、DNA の(d)反応が停止し、(e)の異なる DNA 断片が得られる。その DNA 断片をゲル電気泳動等によって解析して、DNA の配列を決定する。

【語群】

サンガー、ワトソン、マキサム・ギルバート、プライマー、ATP、ADP、dNTP、ddNTP、伸長、増幅、長さ、量、修飾

(10) さまざまな多層オミクスシーケンス解析技術から得られたシーケンスリードをヒトゲノムリファレンス配列にマッピングした。下図はゲノムビューアーを用いて表示した、遺伝子 A 近傍におけるシーケンス深度 (depth) の分布を示している。RNA シーケンスデータは(a)~(e)のいずれか答えなさい。



B. 下記の文章を読み、問 1~4 に答えなさい。

単一遺伝子疾患は主として単一の座位のアレルによって発症が決定される疾患で、一般に (1) の法則に従った遺伝様式を示す。遺伝様式としては常染色体顕性 (優性)、潜性

(劣性)、または X 連鎖潜性 (劣性) を主要なものとして挙げるができるが、他にも様々な遺伝様式が見られる。

レット症候群は発達障害の一つで、そのうち 95% を占める *MECP2* 遺伝子を原因とする型では、X 連鎖顕性 (優性) 遺伝を示す。この遺伝形式は、遺伝子座が X 染色体上にあり、ヘテロ接合体女性で発症 (する / しない) ^(A) ことを意味する。常染色体遺伝との違いは、罹患男性から生まれた男児は罹患しない ⁽¹⁾ ことがヒントとなる。実際には原因遺伝子 *MECP2* の少なくとも 1 つのコピーは初期発生に必要であることから、この疾患については患者の性別はほぼ (男性 / 女性) ^(B) のみである。

レリー・ワイル症候群は低身長を特徴とし、擬 (偽) 常染色体遺伝を示す。この疾患では原因遺伝子 *SHOX* は性染色体上に位置するが、擬常染色体領域にある。この領域は X 染色体と Y 染色体で相同であるため男性の (2) 時にも組換えが起こる。そのため通常の X 連鎖顕性 (優性) 遺伝とは異なり、(女-女 / 女-男 / 男-女 / 男-男) ^(C) 伝達が見られる。

プラダー・ウィリ症候群とアンジェルマン症候群はいずれも 15 番染色体 15q11-q13 領域の微細欠失を主要な原因とし、知的障害、発達遅滞を示すが、そのほかにプラダー・ウィリ症候群では筋弛緩、摂食障害、肥満、低身長が起こり、一方アンジェルマン症候群は特徴的な微笑み、失調性歩行や振戦が見られ、症状から見ると別の疾患である。この領域には父親染色体と母親染色体で DNA メチル化パターンの違いがあることがわかっており、そのため父親由来、または母親由来の染色体からしか発現しない遺伝子がそれぞれ存在する。したがって、この領域の片親由来染色体の欠失は通常とは異なる遺伝様式を取る ⁽²⁾。

単一遺伝子疾患には、核ゲノムの遺伝子変異によるものだけでなく、ミトコンドリアゲノムの変異による疾患が含まれる。ミトコンドリア遺伝病は (父系 / 母系) ^(D) 遺伝を示すことが、核ゲノムによる遺伝性疾患と区別するきっかけとなる。

知的障害や発達遅滞は一般に遺伝性のある疾患とされるが、家系に障害がない病児も多く見られる。その場合に病児に対してマイクロアレイ染色体検査 (CGH 法) を行うと遺伝子コピー数の変化を検出することがある。このような遺伝的変異が両親には見られない場合、このような変異を (3) 変異と呼び、典型的なメンデル遺伝様式を示さないことになる。

問 1 (1) ~ (3) に当てはまる最も適した語句を次の語群から選びなさい。

【語群】メンデル、モルガン、ベイトソン、有糸分裂、減数分裂、コンパウンド、デノボ (*de novo*)、ハプロタイプ

問 2 (A) ~ (D) について、それぞれの括弧の中の語群から適したものを選びなさい。

(問題 3 : 次ページに続く)

問3 下線部(1)について、なぜそうなるか説明しなさい。

問4 下線部(2)について、常染色体顕性（優性）または潜性（劣性）の遺伝様式と比べると遺伝様式の違いはどこにみられるか。

C. 下記の文章を読み、問1～4に答えなさい。

ヒト由来のある培養細胞株 X と Y について、網羅的に遺伝子発現と転写制御パターンを計測するために、RNA シークエンス解析とバイサルファイトシークエンス解析を実施した。RNA シークエンス解析は、転写産物を網羅的にシークエンス解読する手法である。バイサルファイトシークエンス解析は、DNA の非メチルシトシンをウラシルに変えるバイサルファイト変換を活用して、DNA メチル化パターンを解析する手法である。なお、細胞株 X および Y はそれぞれ単一細胞 x または y 由来のモノクローナルな細胞集団であるとする。

問1 あるヒトゲノム領域における DNA 配列 (5'-CGTTTGGACGCGTG-3') のすべてのシトシンはメチル化されていなかった。このゲノム領域をバイサルファイトシークエンスした時の結果を下記1～4から1つ選びなさい。なお、この解析では、ヒト由来の細胞より抽出した DNA サンプルをバイサルファイト変換後に PCR 増幅し、得られた DNA ライブラリを次世代シーケンサーによって解読している。

1. 5'-CGTTTGGACGCGTG-3'
2. 5'-UGTTTGGGAUGUGTG-3'
3. 5'-UUTTTGGAUUUUTG-3'
4. 5'-TGTTTGGATGTGTG-3'

問2 遺伝子発現量と DNA メチル化状態について、ヒト培養細胞株 X と Y の間で比較を行った。その結果、細胞株 X では、遺伝子 α の RNA 発現量が細胞株 Y よりも低かった。また、遺伝子 α について、プロモーターの CpG アイランドにおける DNA メチル化率が X と Y で大きく異なっていた。細胞株 X では、遺伝子 α のプロモーター領域の CpG アイランドにおける DNA メチル化状態が、一般にどのような状態になっていると考えられるか、25 字程度で答えなさい。

問3 遺伝子 α の転写領域内 (gene body) においては、細胞株 X と Y 間で DNA メチル化率に大きな差がなかった。一般にヒトの分化した体細胞において、転写領域内の DNA メチル化状態はどのような状態になっていると知られているか、25 字程度で答えなさい。

問4 細胞株 Y の遺伝子 β のプロモーター領域における、ある CpG サイトの DNA メチル化状態を調べたところ、DNA メチル化率が約 50%であった。

(a) 細胞株 Y におけるすべての細胞が同じ DNA メチル化パターンを示す場合、この CpG サイトの DNA メチル化状態について、ヒト細胞におけるゲノム DNA のコピー数に言及しつつ、100 字程度で簡潔に説明しなさい。

(b) 一部のゲノム領域では DNA メチル化パターンが配偶子形成の過程ですでに決定しており、正常な個体発生のための遺伝子発現制御において重要な機能を果たしていることが明らかとなっている。配偶子形成過程でみられるこのエピジェネティック制御機構を特に何と云うか答えなさい。

(問題3 終わり)

問題 4

A. 以下の (1) ~ (9) までの問いの文章を読み、全て答えなさい。

- (1) B 型肝炎に関する以下の記述のうち、最も適切なものを 1つ 選びなさい。
- A. ウイルス感染予防のために抗体投与は無効である。
 - B. 生ガキの摂食により感染する可能性がある。
 - C. 性交渉によって感染する可能性がある。
 - D. 有効な抗ウイルス薬は開発されていない。
 - E. 感染後最初に抗体価が上昇するのは IgA 型抗体である。
- (2) 下記のワクチンのうち、抗原が体内で作られるものを 1つ 選びなさい。
- A. 合成ペプチドワクチン
 - B. 組換えタンパクワクチン
 - C. 不活化ウイルスワクチン
 - D. mRNA ワクチン
 - E. 弱毒生ワクチン
- (3) 季節性インフルエンザに関する以下の記述のうち、正しいものを 1つ 選びなさい。
- A. インフルエンザウイルスは 1 本鎖マイナス鎖の RNA をゲノムとして持つ。
 - B. インフルエンザウイルスに対する宿主細胞表面の受容体は CCR5 である。
 - C. インフルエンザウイルス粒子の表面糖タンパク質であるヘマグルチニン (HA) とノイラミニダーゼ (NA) のうち、宿主細胞表面受容体に結合するのは NA である。
 - D. インフルエンザ治療薬として用いられる NA 阻害剤の作用機序は、ウイルスタンパク質の合成阻害である。
 - E. 季節性インフルエンザを引き起こすのは A 型と C 型である。
- (4) 下記のうち、エンベロープを持たないウイルスを 1つ 選びなさい。
- A. ヒト免疫不全ウイルス
 - B. アデノウイルス
 - C. インフルエンザウイルス
 - D. 単純ヘルペスウイルス
 - E. 水痘・帯状疱疹ウイルス
- (5) 哺乳類の 1 次リンパ組織のうち、T 細胞の成熟に関わる組織の名称を答えなさい。

(6) 貪食作用(食作用)を持つ免疫担当細胞を 2つ答えなさい。

(7) ヒトの免疫グロブリンの抗原結合部位は (a) の (b) で形成される。下記 A ~ F のうち、組み合わせとして正しいものを 1つ選びなさい。

- A. (a) 2 本の H 鎖 ; (b) V 領域
- B. (a) 2 本の H 鎖 ; (b) C 領域
- C. (a) 2 本の L 鎖 ; (b) V 領域
- D. (a) 2 本の L 鎖 ; (b) C 領域
- E. (a) 1 本の H 鎖と 1 本の L 鎖 ; (b) V 領域
- F. (a) 1 本の H 鎖と 1 本の L 鎖 ; (b) C 領域

(8) 制御性 T 細胞について、マスター転写因子(a)、分泌する主なサイトカイン(b)、主な機能(c)について正しい組み合わせを下記 A ~ H から 1つ選びなさい。

- A. (a) FOXP3; (b) IL-12; (c) マクロファージの活性化
- B. (a) FOXP3; (b) IL-12; (c) 他の T 細胞の抑制
- C. (a) FOXP3; (b) IL-10; (c) マクロファージの活性化
- D. (a) FOXP3; (b) IL-10; (c) 他の T 細胞の抑制
- E. (a) GATA3; (b) IL-12; (c) マクロファージの活性化
- F. (a) GATA3; (b) IL-12; (c) 他の T 細胞の抑制
- G. (a) GATA3; (b) IL-10; (c) マクロファージの活性化
- H. (a) GATA3; (b) IL-10; (c) 他の T 細胞の抑制

(9) NK 細胞は細胞表面に発現する Fc γ RIII によって、IgG1 抗体もしくは IgG3 抗体で覆われた細胞を認識し、標的細胞を殺傷する。このメカニズムの名称を答えなさい。

(問題 4 : 次ページに続く)

B. 以下の文章を読み、問 1～7 に答えなさい。

ヒトに感染症を引き起こす病原体には様々な種類がある。その中にはインフルエンザや単純ヘルペスなどのウイルスに加え、細菌、真菌、原虫、寄生虫などが含まれる。さらに DNA や RNA などの遺伝情報を持たない病原体もある (1)。

ウイルスは遺伝情報として DNA または RNA を持ち、遺伝情報が DNA であるか RNA であるか、一本鎖であるか二本鎖であるか、一本鎖の場合プラス鎖であるか、もしくはマイナス鎖であるかにより分類される。例えば、パンデミックとなった新型コロナウイルス (SARS-CoV-2) は () の () (2) ウイルスに分類される。RNA ウイルスの一種であるレトロウイルスは、() 酵素 (3) によってゲノム RNA を DNA に変換し、宿主ゲノムに組み込む。ウイルス感染症に対して近年では、抗ヘルペスウイルス薬、抗 HIV 薬、抗インフルエンザ薬などの様々な抗ウイルス薬が開発されている。細菌感染症に対しては 抗生物質などの抗菌剤が使用される (4) が、近年、抗生物質耐性菌 (5) の増加が問題になっている。また、寄生性原虫も重篤な疾患を起こす (6) ものがあり、適切な治療が行われないと死に至る疾患もある。

問 1. 下線部 (1) に相当する病原体を挙げなさい。

問 2. 下線部 (1) の病原体によって発症する疾患を 1 つ挙げなさい。

問 3. 下線部 (2) の括弧内に入る言葉の組み合わせで正しいものを下記の A～D から 1 つ選びなさい。

- A. 一本鎖 — DNA
- B. 一本鎖 — RNA
- C. 二本鎖 — DNA
- D. 二本鎖 — RNA

問 4. 下線部 (3) に入る適切な言葉を答えなさい。

問 5. 下線部 (4) について抗生物質が殺菌作用を発揮する作用機序を 1 つ答えなさい。

問 6. 下線部 (5) について、遺伝子変異や耐性遺伝子の伝搬などによって細菌が抗生物質耐性となる時の機序を 3 つあげて 150 字程度で説明しなさい。

問 7. 下線部 (6) について、原虫の感染が原因で起こる感染症を 1 つ答えなさい。

C. 以下の文章を読み、(1)～(7)の問いに答えなさい。

T細胞やB細胞などのリンパ球のクローン選択とクローン増殖は、(a)免疫応答で誘導される。この2次免疫応答は再度侵入してきた病原体に対して、(b)によって誘導される迅速で強力な免疫応答である。

MHCは、2次免疫応答を誘導するために細胞表面に抗原を提示する分子群である。MHCクラスI^①はヘテロ二量体を形成する分子で、ほぼ全ての細胞で発現する。抗原となるペプチド^②を提示したMHCクラスI分子はリンパ球の一種である(c)によって認識され、(d)分子と会合する。一方、抗原提示能力が高い一部の細胞で発現するMHCクラスII^③は、(e)によって認識され、(f)分子と会合する。MHC分子によって抗原提示を受けた細胞は、受容体を介してシグナルを受け取り活性化する。

免疫応答において自己、非自己の区別は極めて重要である。MHCによって抗原提示を受けたエフェクター細胞や自己反応性抗体(自己抗体)が、正常な細胞や組織に異常に反応するようになったために引きこされる疾患を自己免疫疾患^④という。自己免疫疾患は免疫寛容^⑤を維持する機構の総合的な破綻の現れといえる。

(1) 括弧(a)、(b)について、下記A～Dの組み合わせから正しいものを1つ選びなさい。

- A. (a) 自然 ; (b) 免疫記憶
- B. (a) 適応 ; (b) 免疫記憶
- C. (a) 自然 ; (b) T細胞受容体の分泌
- D. (a) 適応 ; (b) T細胞受容体の分泌

(2) MHCクラスI^①とヘテロ二量体を形成する分子名を答えなさい。

(3) 細胞外抗原に由来するペプチド^②を提示するのはMHCクラスIとMHCクラスIIのどちらか答えなさい。

(4) 括弧(c)、(d)、(e)、(f)に当てはまる言葉について、下記語群の中からそれぞれ選び、(c)- ○○、(d)- ○○のように答えなさい。

【語群】

造血幹細胞、NK細胞、ヘルパーT細胞、細胞障害性T細胞、マクロファージ、記憶B細胞、ナイーブB細胞、CD4、CD8、CD16、CD19、PD-1

(問題4 : 次ページに続く)

- (5) MHCクラスII^③ を発現するプロフェッショナル抗原提示細胞を 3つ 答えなさい。
- (6) 以下の A~F の疾患のうち、自己免疫疾患^④ を すべて 選びなさい。
- A. 多発性硬化症
 - B. 後天性免疫不全症候群
 - C. 多発性骨髄腫
 - D. 1型糖尿病
 - E. ハンチントン病
 - F. 関節リウマチ
- (7) 免疫寛容^⑤ について、自己抗体の産生を防ぐためのメカニズムを1つ挙げ、100字程度で説明しなさい。

(問題4 終わり)

問題 5

A. 以下の (1) ~ (10) までの問いの文章を読み、全て答えなさい。

(1) 細胞増殖とアポトーシスの調節を通じて器官のサイズを制御するシグナル伝達経路は何か。答えなさい。

(2) X 染色体不活化の過程に関与する代表的な非コード RNA の名前を 1 つ答えなさい。

(3) 発生期に細胞膜表面に存在し、細胞間接着の形成に重要な役割を果たしている分子の名前を答えなさい。

(4) 下記の文を読み、(a)、(b)に当てはまる語句を答えなさい。

1962 年、ガードン博士は、オタマジャクシの小腸の細胞から取り出した (a) を、元の (a) を壊した (b) に移植し、カエルに成長させた。この実験により、分化した細胞の中にも体を作る全ての遺伝情報があることが証明された。

(5) 幹細胞の自己組織化能を活用して、試験管内で作られた 3 次元構造体(ミニ臓器)のことを何というか答えなさい。

(6) 哺乳類の性染色体のうち、長いのはどちらか答えなさい。

(7) 21 番染色体が 1 本過剰になることによって引き起こされる先天性疾患を何というか答えなさい。

(8) 中枢神経系を構成するニューロン以外の細胞 (グリア細胞) のうち、中胚葉に由来する細胞の名前を答えなさい。

(9) イモリを用いた Spemann-Mangold の移植実験で発見された、周囲の組織に作用し、特定の構造を誘導する働きをもつ胚の領域を何と呼ぶか答えなさい。

(10) BCL2 ファミリータンパク質のうち、アポトーシスを促進するものと抑制するものを 1 つずつ答えなさい。

(問題 5 : 次ページに続く)

B. 次の文章を読み、(1)～(8)の間に答えなさい。

高度に統制の取れた幹細胞₍₁₎の機能の維持と制限は、正常な個体の発生に必要不可欠である。哺乳類のように複雑な生物も、1つの受精卵が分裂して個体を形成する。つまり個体発生において、受精卵こそが最初で最重要な幹細胞である。受精後(2)期まではほぼ未分化のまま幹細胞が増殖し、その後分化する方向が決まっていく。それぞれの細胞の分化の方向性の決定₍₃₎は、転写因子の使い分け、シグナル因子の組み合わせ₍₄₎、そしてエピゲノム変化₍₅₎などによって微調整され、やがて高度に専門化した個々の細胞が正しい組織に配置され、個体の維持に携わる。

近年、人工的に開発した幹細胞を活用した再生医療の進歩が目覚ましい。最初に人類が手に入れた人工幹細胞は ES細胞₍₆₎ で、適切な培養条件下ではほぼ無限に増殖させることが可能。うえ、養母個体に戻すと個体のあらゆる種類の細胞に分化させることができる。一方、iPS細胞₍₇₎ はリプログラミングに必要な複数の因子をマウスの線維芽細胞に導入することで樹立された。当初は24種類あった因子が2006年には4因子にまで絞られ、その後開発者の山中伸弥博士が2012年にノーベル医学・生理学賞を受賞した。これらのES細胞やiPS細胞の開発は、再生医療の発展に大きく貢献している。造血幹細胞移植は、生体にもともと存在する幹細胞を用いた再生医療の1つと言える。₍₈₎ 白血病や再生不良性貧血など、血液疾患を持つ多くの患者の命を救ってきた有効な治療法である。

(1) 下線1について、幹細胞は全能性(totipotent)幹細胞、多能性(pluripotent)幹細胞、多分化能(multipotent)幹細胞に分類することができる。これらの幹細胞のタイプの違いについて、それぞれ例をあげて 250字程度で説明しなさい。

(2) 括弧2に入る発生段階を答えなさい。

(3) 下線3について、内胚葉と外胚葉からできる組織を それぞれ2つずつ答えなさい。

(4) 下線4について、Notch とそのリガンド Delta による「側方抑制」について、200字程度で説明しなさい。

(5) 下線5について、ショウジョウバエでは *Hox* ファミリー遺伝子群の発現量が体の極性や体節の決定に関与している。この *Hox* ファミリー遺伝子群の発現量を調節する上で主要な役割を果たす2種類のヒストン制御因子群を答えなさい。また、それぞれのヒストン制御因子群が、*Hox* ファミリー遺伝子の発現をどのように制御するか答えなさい。

(6) 下線 6 について、ES 細胞の正式名称 (日本語または英語) を答えなさい。また、その作成方法について説明した下記の文章の (A) と (B) に入る適切な言葉をそれぞれ答えなさい。

【文章】ES 細胞は、マウスや人の (A) 期の初期胚の (B) を適切な培養条件下で増殖させたもの。

(7) 下線 7 について、iPS 細胞の英語での正式名称を答えなさい。

(8) 下線 8 について、なぜ造血幹細胞移植は再生医療の 1 つと考えられるのか。機能的に分化した様々な種類の血液細胞の名称と、それらが作られるメカニズムを含めて 200 字程度 で説明しなさい。

(問題 5 終わり)

問題6

A. 以下の(1)～(8)までの問いの文章を読み、全て答えなさい。

(1) 一部のがんはワクチンにより予防可能である。ワクチンと、そのワクチンによって予防可能ながんの組み合わせを1つ挙げなさい。

(2) 下記に合致する学説の名前を答えなさい。

「発がんには、がん抑制遺伝子の対立遺伝子の両方が変異あるいは欠失などにより機能を失う必要があるとする説。Knudsonにより提唱され、後に網膜芽細胞腫の原因遺伝子 *RB* の発見によって証明された。」

(3) *BRC A2* の生殖細胞変異が最も高頻度に見られるがん種を、下記から1つ答えなさい。

卵巣がん、膵がん、大腸がん、食道がん

(4) 以下の要素は細胞老化の指標として用いられている。このうち、老化細胞で一般的に亢進もしくは増加を認めるものをすべて答えなさい。

p16、Ki-67、テロメア長、核の大きさ

(5) 大腸がんの多段階発がんモデルにおいて、以下の3つの現象が発生する順に並べ、(A)(B)(C)の様に答えなさい。

(A) *KRAS* 遺伝子の変異、(B) *TP53* 遺伝子の変異、(C) *APC* 遺伝子の変異

(6) アポトーシスの特徴【語群A】とそれを検出する方法【語群B】の正しい組み合わせを4つ答えなさい。

【語群A】 脂質二重膜の変化、Caspaseの活性化、クロマチン凝縮、DNAの断片化

【語群B】 DAPI染色、TUNEL染色、Annexin V染色、PARPの切断

(7) 血管新生は既存の血管から新たな血管枝を形成して血管網を構築する現象であり、腫瘍の増殖に重要な役割を果たす。血管新生の過程において、以下の4つの現象が発生する順に並べ、(A)(B)(C)(D)の様に答えなさい。

(A) VEGFの分泌

(B) 腫瘍塊における低酸素状態

(C) 血管内皮細胞の増殖

(D) HIF-1タンパク質の安定化

(8) 下記に合致する学説の名前を答えなさい。

「がん細胞がある特定の臓器で増殖できるのは、がん細胞が育つのにその臓器の環境が適しているからであるとする考え方。英国の Paget 博士が提唱した説で、がんの臓器選択的転移メカニズムを説明する概念である。」

B. 以下の (1) ~ (6) までの問いの文章を読み、全て答えなさい。

(1) 紫外線は皮膚がんリスクを高めることが知られている。紫外線はどのような機序で発がんリスクを高めるのか、100 字程度で説明しなさい。

(2) 免疫チェックポイント阻害剤 (抗 PD-1/PD-L1 抗体や抗 CTLA-4 抗体) の臨床効果は、一般的に遺伝子変異を多く認めるがん患者で高い。遺伝子変異頻度の高い患者で治療効果が高い理由を「ネオアンチゲン」という単語を使って 100 字程度で説明しなさい。

(3) 以下の表は、複数のがん抑制遺伝子の体細胞変異パターンを示している。TP53 遺伝子はミスセンス変異が多い。ミスセンス変異とはどのような変異か 50 文字程度で説明しなさい。また TP53 遺伝子でミスセンス変異が多い理由について、「ナンセンス変異依存 mRNA 分解」と「ドミナントネガティブ効果」の 2 つの語句を使って 200 字程度で説明しなさい。

	frame shift	missense	in frame deletions/ insertions	nonsense	silent	others
TP53	9%	74%	2%	8%	4%	3%
APC	51%	4%	0%	32%	9%	4%
ATM	56%	28%	2%	14%	0%	0%
BRCA1	54%	30%	0%	11%	0%	5%

(4) 日本において、全がんの罹患率の年次推移は増加傾向が認められるが、年齢調整を行うと横ばい傾向になる(1)。下線 (1) の理由を 100 字程度で説明しなさい。尚、年齢調整罹患率とは、ある基準年度における人口構成と同じであると仮定した場合の罹患率である。

(問題 6 : 次ページに続く)

(5) 抗がん剤治療の主要な2つの課題は、薬剤耐性と副作用である。当初有効であった薬剤が抗がん剤の治療経過中に効かなくなる獲得耐性が生じる機序を2つ挙げ、合わせて100字程度で説明しなさい。

(6) あるがん患者の腫瘍細胞の染色体分析の結果、Philadelphia 染色体という染色体異常がみつかった。この異常から推定されるがん種を答えなさい。またイマチニブという薬剤は Philadelphia 染色体陽性のがん種に対して有効であるが、その作用機序を 100文字程度で説明しなさい。

(問題6 終わり)

問題7

実数 $n (> 1)$ と非負整数 i に対し、 n に関数 f を i 回適用した関数を $f^{(i)}(n)$ と定義する (すなわち $f^{(0)}(n) = n$ 、 $i > 0$ のとき $f^{(i)}(n) = f(f^{(i-1)}(n))$)。

いま、 $f^{(i)}(n) \leq 1$ となる i が存在するとき、そのような i の最小値を $f^*(n)$ と定義し (すなわち $f^*(n) = \min\{i > 0 \mid f^{(i)}(n) \leq 1\}$)、存在しないとき $f^*(n)$ は未定義とする。下記の各質問をその理由とともに答えよ。

(1) $f(n) = n/2$ のとき、 $f^*(n) = \lceil \log_2 n \rceil$ を示せ ($\lceil x \rceil$ は x 以上の最小の整数値)。

(2) $f(n) = \log_2 n$ のとき、 $f^*(n) = 5$ を満たす n の最小値を答えよ。

(3) $f(n) = \log_2 n$ のとき、 $f(f^*(n))$ と $f^*(f(n))$ の大小関係を答えよ。

問題8

推移行列 $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$, $0 < p < 1, 0 < q < 1, p+q \neq 1$, に関して以下の問に答えよ。

(1) ベクトル $\begin{pmatrix} p \\ -q \end{pmatrix}$ が、 P の固有ベクトルであることを示せ。

(2) n 次推移行列 P^n を求めよ。

(3) $\lim_{n \rightarrow \infty} P^n$ を求めよ。

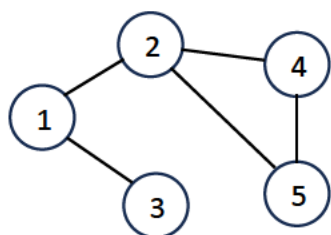
$n \times n$ 実行列 $A = (a_{ij})_{1 \leq i, j \leq n}$ の trace は、 $\text{tr } A = \sum_{i=1}^n a_{ii}$ と定義される。以下の問に答えよ。

(4) A の固有値を $\lambda_1, \dots, \lambda_n$ とするとき、 $\text{tr } A = \sum_{i=1}^n \lambda_i$ であることを示せ。

(5) $A^m = O$ (O はゼロ行列) となる自然数 m が存在するとき、 $\text{tr } A = 0$ であることを示せ。

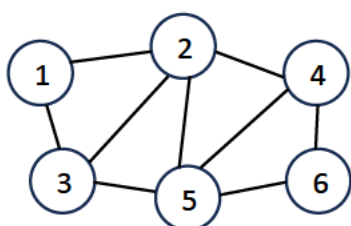
問題9

以下はグラフとそれに対応する隣接リストの一例である。



1: 2, 3
2: 1, 4, 5
3: 1
4: 2, 5
5: 2, 4

(1) 以下のグラフに対応する隣接リストを書け。



(2) 以下の隣接リストに対応するグラフを描け。

1: 2, 3
2: 1
3: 1
4: 5, 6
5: 4, 6, 7
6: 4, 5, 7
7: 5, 6

- (3) このページの初めに示したグラフに対し、深さ優先探索をノード2から開始するとする。このときノードの訪問順序として可能なものをひとつ書け。その最初のノードはノード2で、2番目のノードはノード2に隣接するノードの1つでなければならない。
- (4) このページの初めに示したグラフに対し、幅優先探索をノード2から開始するとする。このときノードの訪問順序として可能なものをひとつ書け。
- (5) あるグラフは n 個のノードを持ち、ノード i は m_i 個のエッジを持つとする ($1 \leq i \leq n$)。対応する隣接リストを A_{ij} とする ($1 \leq i \leq n, 1 \leq j \leq m_i$)。このとき深さ優先探索を用いてグラフの連結成分数を出力するアルゴリズムの擬似コードを書け。アルゴリズムは A_{ij} の各要素を明示的に用いるものでなければならない。

(問題9：次ページに続く)

- (6) (5) において、深さ優先探索の代わりに幅優先探索を用いてグラフの連結成分数を入力するアルゴリズムの擬似コードを書け。アルゴリズムは A_{ij} の各要素を明示的に用いるものでなければならない。

(問題 9 終わり)

問題 10

N 個の要素からなる整数配列 V に対して、以下の手続き `quick_sort` は配列 V の全体を `quick_sort(0, V, 0, N - 1, $-\infty$, $+\infty$)` と呼び出すことで整列する。ただし、`swap(x, y)` は変数 x と y の値を入れ替える特殊な手続きとする。

```
def quick_sort(d, V, li, hi, lv, hv):
    if hi <= li: return
    p = li; l = li + 1; h = hi
    while l < h:
        while l <= hi and V[l] < V[p]: l += 1
        while li <= h and V[p] <= V[h]: h -= 1
        if h <= l: break
        swap(V[l], V[h])
        l += 1; h -= 1
    swap(V[p], V[h])
    if lv < V[h]: quick_sort(d + 1, V, li, l - 1, lv, V[h])
    if V[h] < hv: quick_sort(d + 1, V, h + 1, hi, V[h], hv)
```

(A)

- (1) (A)では配列 V にどのような処理を行っているか答えよ。
- (2) d は何を表しているか答えよ。
- (3) 整列の手続き全体を通して d はさまざまな値をとるが、そのうち最大の d を d_{\max} とする。 N が与えられているとき、 d_{\max} が最も大きくなる配列 V の例を1つ挙げよ。
- (4) lv と hv は何を表しているか答えよ。
- (5) d が偶数のときにはそのままの動作を、 d が奇数のときには下線部(B)と(C)の不等号を交換した動作をさせることとする。このとき、(3)で答えた配列 V を入力として整列を行うと何が起こるか説明せよ。
- (6) (5)の処理を行うメリットを説明せよ。

問題 11

n 個の独立な非負実数値確率変数 X_i ($i = 1, \dots, n; n \geq 2$) はすべて同一の確率密度関数 $f(x) = e^{-x}$ を持つとする。以下の問に数式による導出も含め答えよ。

- (1) X_i の平均と分散を求めよ。
- (2) ある i について、 X_i の値 x_i が定まったとする。このとき $k \neq i$ となる一つの k について $X_i \leq X_k$ となる確率 $\mathbb{P}(X_i \leq X_k | X_i = x_i)$ を求めよ。
- (3) (2) の確率に $f(x_i)$ をかけて x_i で積分することにより $\mathbb{P}(X_i \leq X_k)$ を求めよ。
- (4) 集合 $\{X_i | i = 1, \dots, n\}$ の最小値を $X_{\min} = \min_{k=1, \dots, n} X_k$ とおく。正実数 x が与えられたとき $x \leq X_{\min}$ となる確率 $\mathbb{P}(x \leq X_{\min})$ を求めよ。

n 個の正実数 λ_i ($i = 1, \dots, n$) が与えられたとき、確率変数 Z_i ($i = 1, \dots, n$) を $Z_i = X_i / \lambda_i$ で定義する。

- (5) ある i について、 X_i の値 x_i が定まったとする。このとき $k \neq i$ となる一つの k について $Z_i \leq Z_k$ となる確率 $\mathbb{P}(Z_i \leq Z_k | X_i = x_i)$ を求めよ。
- (6) ある i について、 X_i の値 x_i が定まったとする。このとき $k \neq i$ となる全ての k について $Z_i \leq Z_k$ となる確率 $\mathbb{P}(\bigcap_{k=1, \dots, n; k \neq i} \{Z_i \leq Z_k\} | X_i = x_i)$ を求めよ。
- (7) 集合 $\{Z_i | i = 1, \dots, n\}$ の中で最小の要素の一つを Z_I とする。このとき添字 $I = \operatorname{argmin}_{k=1, \dots, n} Z_k$ が従う確率分布 $\mathbb{P}(I = i)$ ($i = 1, \dots, n$) を答えよ。

問題 12

- (1) 以下は、2本の配列 $x = x_1 \cdots x_m, y = y_1 \cdots y_n$ の大域アラインメントのスコアを計算するアルゴリズムにおける漸化式である。ただし、 $s(a, b)$ は文字 a と b のマッチスコアであり、 $0 < e < d$ を満たす。初期値の設定はここでは記載していない。

for $i = 1, \dots, m$

for $j = 1, \dots, n$

$$M_{i,j} = \max \begin{cases} M_{i-1,j-1} + s(x_i, y_j) \\ X_{i-1,j-1} + s(x_i, y_j) \\ Y_{i-1,j-1} + s(x_i, y_j) \end{cases}$$

$$X_{i,j} = \max \begin{cases} M_{i-1,j} - d \\ X_{i-1,j} - e \\ Y_{i-1,j} - d \end{cases}$$

$$Y_{i,j} = \max \begin{cases} M_{i,j-1} - d \\ X_{i,j-1} - d \\ Y_{i,j-1} - e \end{cases}$$

- (1-1) 長さ k のギャップに対するペナルティを式で示せ。

- (1-2) 初期値として設定すべき値の一部を

$$M_{0,0} = 0, X_{0,0} = -\infty, Y_{0,0} = -\infty,$$

$$\text{for } j = 1, \dots, n: M_{0,j} = X_{0,j} = -\infty,$$

$$\text{for } i = 1, \dots, m: M_{i,0} = Y_{i,0} = -\infty$$

とするとき、

初期値として $X_{i,0}$ ($i = 1, \dots, m$) および $Y_{0,j}$ ($j = 1, \dots, n$) に設定すべき値を示せ。

- (1-3) 同様のギャップペナルティを使用し、局所アラインメントの最大スコアを求めるための漸化式を示せ。

- (1-4) (1-3)の計算を用いて、最大スコアを持つ局所アラインメントを得る方法を説明せよ。

(問題 12 : 次ページに続く)

(2) 4種類の文字、 a, c, g, t からなる配列 $x = x_1 \cdots x_m$ がある。

a, c, g, t の相補文字をそれぞれ

$$\text{comp}(a) = t, \text{comp}(c) = g, \text{comp}(g) = c, \text{comp}(t) = a$$

と定義する。

(2-1) 以下のアルゴリズムは何を報告するアルゴリズムか、説明せよ。

```
for  $i = 1, \dots, m$ :  $H_{i,m+1} = 0$ 
for  $j = 1, \dots, m + 1$ :  $H_{0,j} = 0$ 
for  $i = 1, \dots, m$ 
  for  $j = m, \dots, i$ 
    if  $x_i = \text{comp}(x_j)$  then
       $H_{i,j} = H_{i-1,j+1} + 1$ 
    else
       $H_{i,j} = 0$ 
      if  $H_{i-1,j+1} \geq k$  then
        report a pair of ranges  $[i - H_{i-1,j+1}, i - 1]$  and  $[j + 1, j + H_{i-1,j+1}]$ .
```

(2-2) それぞれの長さが $p + 1, q + 1$ の2本の部分配列 x_i, \dots, x_{i+p} と x_{j-q}, \dots, x_j の「逆相補アラインメントスコア」を2本の配列 x_i, \dots, x_{i+p} と x_j, \dots, x_{j-q} の大域アラインメントの最大スコアと定義する。 x_j, \dots, x_{j-q} は逆順であることに注意せよ。

ただし、アラインメントの置換行列を

$$s(a, b) = \begin{cases} 1 & \text{if } \text{comp}(a) = b \\ -1 & \text{otherwise} \end{cases}$$

とし、ギャップペナルティはギャップの数とする（長さ k のギャップのペナルティは k ）。

$x = x_1 \cdots x_m$ の部分配列（空配列になることもあり得る）の組で、逆相補アラインメントスコアが最大のもものを報告するアルゴリズムを示せ。

(問題 12 終わり)

(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)

