

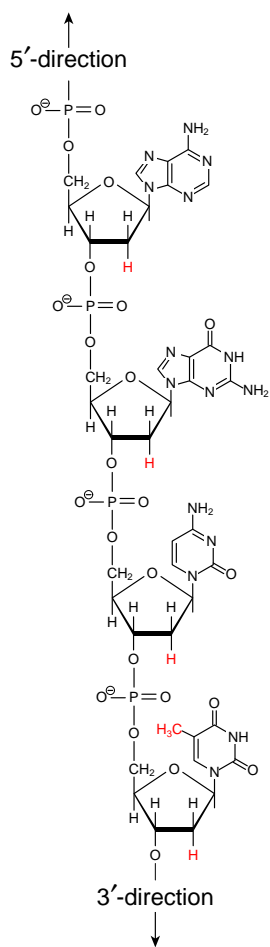
Chapter 1.

Exercise 1.1

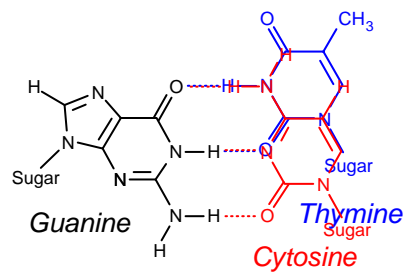
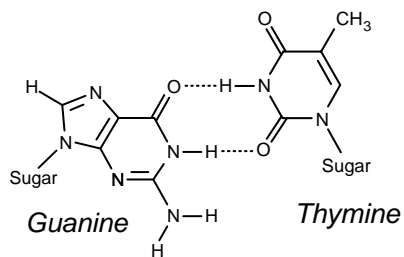
5'-ATATATATATATATATATAT  
3'-TATATATATATATATATATA

Exercise 1.2  $3.2 \times 10^9 \times 0.005 = 1.6 \times 10^7$

Exercise 1.3 The 2'H on the sugars would have to be changed to 2'OH and the 5-methyl group of thymine would have to be changed to a hydrogen to give uracil (shown in red in accompanying figure).



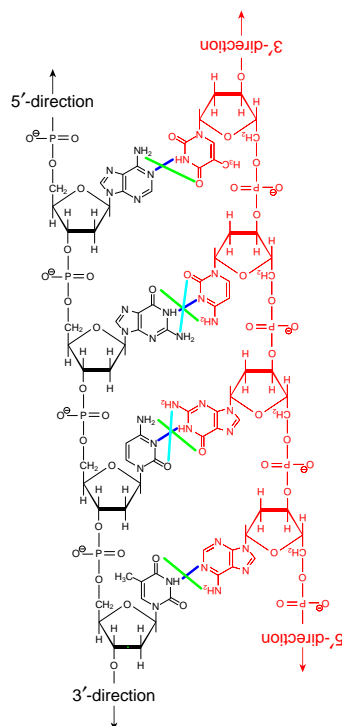
Exercise 1.4



## Introduction to Bioinformatics — Answers to exercises

Exercise 1.5 (a) A self-complementary polynucleotide is one that can form a double helix with another copy of itself.

(b)

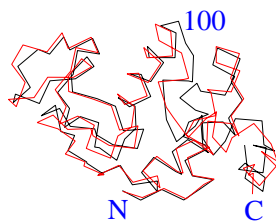


The hydrogen bonds are drawn in blue, green and cyan. Note that they cross each other because the bases are not in the right relative orientation. This is a casualty of the translation from three dimensions to two. Readers can find diagrams similar to this on the web, in which the hydrogen bonds are not crossed. A careful examination of these pictures will reveal that some of the nucleosides are shown as the wrong anomer. This is simply wrong: all nucleosides in natural DNA are the  $\beta$ -anomer, as shown in the figure appearing here.

Exercise 1.6 Length of DNA in one cell:  $3.2 \times 10^9$  bases  $\times 3.4 \text{ \AA} / \text{base} \times 10^{-10} \text{ m/\AA} \times 2$  (because most cells are diploid) = 217.6 m

Total length of DNA in  $10^{13}$  cells =  $2.176 \times 10^{15} \text{ m} \sim 2 \times 10^{12} \text{ km}$ ; this is longer than the semi-major axis of Pluto's orbit.

Exercise 1.7



Exercise 1.8 Same or similar: 37. Opposite charges: 113 R/K, 120 D/K, 130 R/E. Residues at these positions would be expected to point out into the solvent. (They do.)

## Introduction to Bioinformatics — Answers to exercises

Exercise 1.9 Possible ideas: neurodegenerative disorder, New England colonial period, age of onset, discovery of gene for Huntington disease, polyglutamine block, huntingtin, trinucleotide repeat, anticipation, counselling.

Exercise 1.10 (a) Leu-Ala-His-Lys-Tyr-His-STOP (b) change initial c  $\rightarrow$  t (c) change initial c  $\rightarrow$  a (d) change fourth codon aag  $\rightarrow$  tag (e) change final a  $\rightarrow$  t

Exercise 1.11 (a) about 10000 human generations (b) not quite a year

Exercise 1.12 (a) Ile (b) Glu (c) Ser (d) Arg (e) Gly (f) Arg

Exercise 1.13 (a) *nodes*: people working in the building. *edges*: connect people working in same room (b) *nodes*: people working at same installation. *edges*: connect people with immediate superiors in rank; for instance a major with captains under his command, but not a general with a captain. (c) *nodes*: airports served by BA. *edges*: airports linked by a direct flight. (d) *nodes*: papers on the selected topic. *edges*: pairs of articles such that one of them is cited in the other.

Exercise 1.14 (a) 36250 (b) 62.5 (c) 20 million

Exercise 1.15 (a) 20 (b) 1250 (c)  $3.52 \times 10^{16}$ .

Exercise 1.16 about  $\frac{1}{10}$  (omitting the genetic code table, about  $\frac{1}{2}$ ).

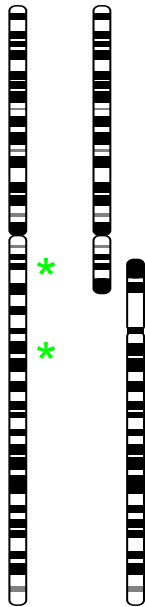
Exercise 1.17 about 3 hours 20 minutes

Exercise 1.18 decrease latency.

## Chapter 2.

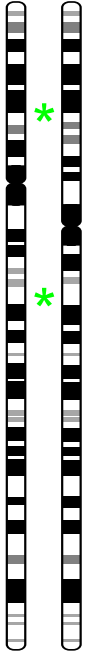
Exercise 2.1 No.

Exercise 2.2 See green asterisks in figure:

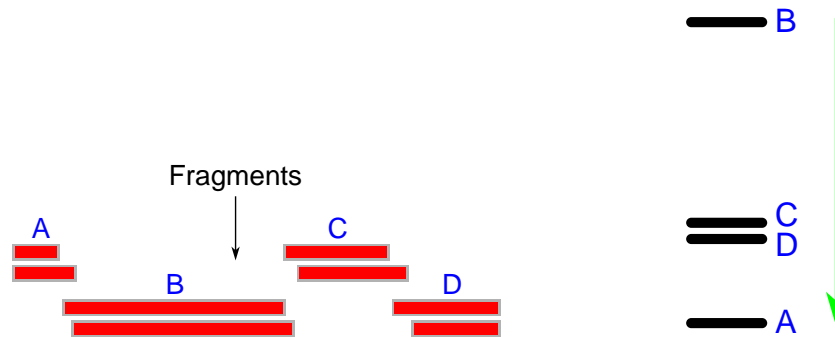


Introduction to Bioinformatics — Answers to exercises

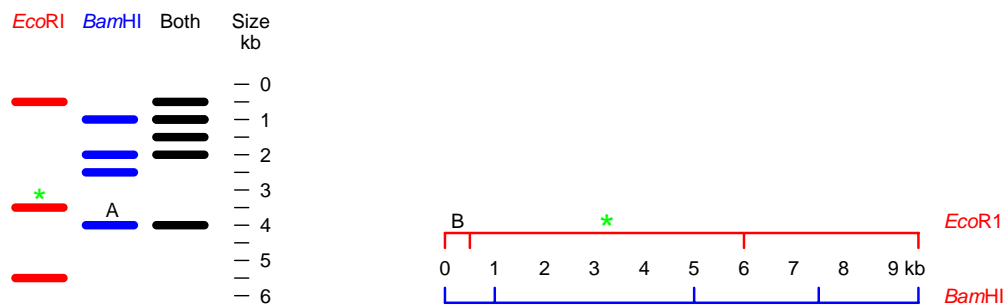
Exercise 2.3 See green asterisks in figure:



Exercise 2.4 Green arrow on right indicates direction of migration. Smaller fragments migrate faster.

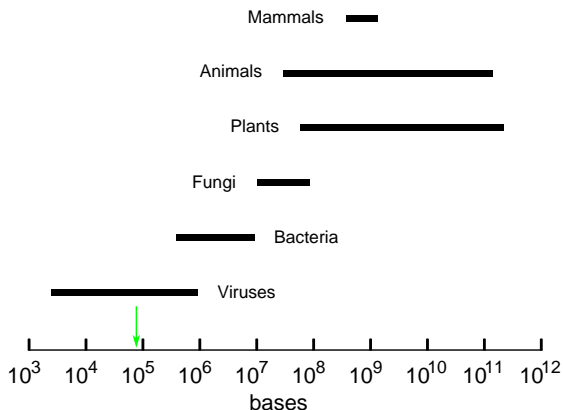


Exercise 2.5



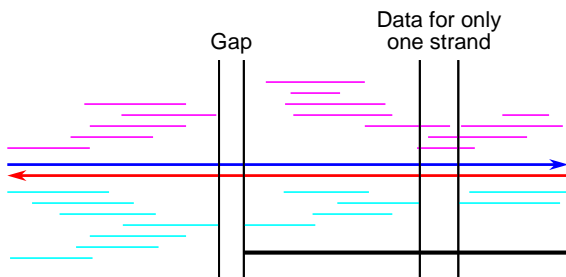
## Introduction to Bioinformatics — Answers to exercises

Exercise 2.6 See green arrow:



Exercise 2.7 In the absence of a complementary second strand, there is no reason to expect a region of single-stranded DNA to maintain the helical structure.

Exercise 2.8 Longest contig shown as thick black line at bottom right.



Exercise 2.9 In Figure 2.13: (a) Obviously: 0. (b) 9. (c) Estimate the average coverage of the entire region? (Hint: measure the total lengths of the fragments and divide by the length of the region.)

Exercise 2.10 Because DNA polymerase can only synthesize in the 5'→3' direction.

Exercise 2.11 (a) gttaatgtagcttaactaaagcaa

(b) gttaatgtagcttaactaaagcaa + gctgtatcccgtggggcgtggttt

Exercise 2.12 The mammoth mitochondrial genome is ~17 kb long. Assuming 300 kb of numts in the mammoth nuclear genome, and a copy-number ratio of 1000 mt:1 nuclear:  $1.7 \times 10^3 \times 1000/3 \times 10^5$  bp of numt sequence = 0.018 = 18%.

Exercise 2.13  $47 \times 10^9$  UK pounds. This is slightly less than half the annual budget for the NHS. Dividing the NHS budget by the population, the country spends on an average approximately 1760 UK pounds per person; this is slightly more than the cost of two human genomes.

Current plans are for the NHS to offer genome sequencing, free to patients with 'serious genetic conditions'. The policy is being developed.

## Introduction to Bioinformatics — Answers to exercises

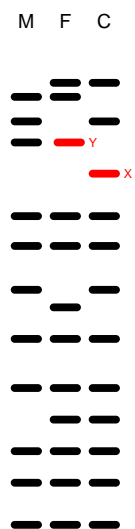
Exercise 2.14

Suffix	Path
inning	root→1→3→8
nning	root→2→5
ning	root→2→6
ing	root→3→9
ng	root→2→7
g	root→4

Exercise 2.15 (a)  $A \rightarrow B \rightarrow D \rightarrow C \rightarrow F \rightarrow G \rightarrow G$  in reverse  $\rightarrow E$  (b) No – this is impossible.

Exercise 2.16 If you build another bridge linking the right island with the upper mainland (call this H), then the path: (a)  $A \rightarrow B \rightarrow D \rightarrow C \rightarrow F \rightarrow G \rightarrow H \rightarrow E$  will traverse each bridge once. This is an Eulerian path. Note however that it is still not possible to construct a path that traverses each bridge once and returns to the starting point, which would be an Eulerian circuit. To achieve this a second new bridge is necessary, for instance linking the left island to the lower mainland.

Exercise 2.17 (a) see figure. (b) impossible. (c) see figure.



Exercise 2.18 After relabeling the lanes, F cannot be the father of C – because the third band down in the (relabelled) C lane is not present in (relabelled) F (or in M, implying that this band must be present in the true father of C).

Exercise 2.19 If M is the mother, neither of the two sailors could be the father. However it is noteworthy that neither of M’s bands appears in the baby. If, upon studying a larger number of M’s bands, should they all, or all but very few, fail to appear in the baby, there would be grounds for suspicion that M is not the baby’s mother. In that case it would not be possible to rule out the paternity of S1 or S2 on purely genetic grounds; however, if there were no other women on the island at the time of conception, that would eliminate them. (Why?)

Exercise 2.20 The gene prevalent in the Sykes family is on the Y chromosome.

Exercise 2.21 (a) The absence of any amylogenin band would make it difficult to identify the individual’s gender. (b) Complete sequencing of DNA derived from the blood.

Exercise 2.22 Incorrect attribution of paternity somewhere in Richard III’s ancestry.

## Introduction to Bioinformatics — Answers to exercises

Exercise 2.23 If D were the father of E, E would have inherited an X chromosome from A, which sequencing would reveal. If E had been male, this would be impossible.

Exercise 2.24 Sequencing of Tom and Jenny would prove that Jenny is not Tom's mother. Screening of people living in the area would reveal that Tom is the son of Bridget Allworthy. The father has died, but Bridget Allworthy knows who he was.

### Chapter 3.

Exercise 3.1 Almost 30000.

Exercise 3.2 If another enzyme provided the function of the product of the gene knocked out, there might be little or no effect on the phenotype.

Exercise 3.3 All but development, and cell-type differentiation, would be suitable for classifying proteins from prokaryotes.

Exercise 3.4 Moon landing.

Exercise 3.5 autosomal dominant.

Exercise 3.6 phenotype correlates with allele of retinoblastoma gene (RB1), not with allele of esterase D gene.

Exercise 3.7 to compare expression patterns in different tissues.

Exercise 3.8 On average  $7.5 \times 10^5$ bp, 6 genes.

Exercise 3.9 risk no higher than if father normal (but see Schwartz, M. & Vissing, J. (2002). Paternal inheritance of mitochondrial DNA. *New Eng. J. Med.* 347, 576–580.)

Exercise 3.10 search for the amino acid sequence in all six-frame translations of the human genome.

Exercise 3.11 exceptions: trpE and trpD, and trpB and trpA, are two pairs of genes contributing to one enzyme; trpD encodes [parts of] two enzymes.

Exercise 3.12 two million.

Exercise 3.13 average transfer rate:  $5.24 \times 10^{-5}$  ORFs/year,  $5 \times 10^{-5}$  kbp/year

Exercise 3.14 (a) both, (b) living genome, (c) both, (d) both, (e) both, (f) computer databases.

### Chapter 4.

Exercise 4.1 4.

Exercise 4.2 6.

Exercise 4.3 agtcc → cgtcc → cgtca → cgctca

Exercise 4.4 If the alignment is not correct, then a conserved residue might be split, in the alignment table, among several different positions.

Exercise 4.5 off main diagonal, expect run of matches on words time and waste, appearing parallel to main diagonal. (b) use PERL program in Box 4.1; use window = 4, thresh = 2.

Exercise 4.6 window = 2, threshold = 2.

Exercise 4.7 PAM250: H↔R more probable, BLOSUM62: W↔F more probable.

## Introduction to Bioinformatics — Answers to exercises

Exercise 4.8

```
THE.RETORT.COURTEOUS-  
THE.REPLY-.CHURLI--SH
```

Exercise 4.9 Set the weights of all route segments into and out of Uppsala to very large negative values.

Exercise 4.10 Do a dotplot of one sequence against the reverse complement of the other.

Exercise 4.11 Change lines 6-10 of the program to:

```
$/ = ""; # read entire paragraph  
$_ = <DATA>; # read input, assume no blank lines  
$_ =~ s/#([\n]*)\n\n/g; # kill comments after # on each line  
s/^(([\n]+)\n\s*(\d+)\s+(\d+)/; # extract job title,  
$title = $1; $nwind = $2; $thresh # window, threshold  
  
# What remains is two sequences, break at FASTA signal >  
  
@seqs = split(>/, $_);  
  
$seqs[1] =~ s/^(([\n]+)\n//; # split out title of first sequence  
$seqt1 = $1; # record title of first sequence  
  
$seq1 = $seqs[1]; # record first sequence  
$seq1 =~ s/\n//g; # kill end-of-line characters  
  
$seqs[2] =~ s/^(([\n]+)\n//; # split out title of second sequence  
$seqt2 = $1; # record title of second sequence  
  
$seq2 = $seqs[2]; # record second sequence  
$seq2 =~ s/\n//g; # kill end-of-line characters
```

Exercise 4.12 Program OK, change data to:

```
Palindrome  
1 1 #WINDOW, THRESHOLD  
MaxIstayawayatsixam  
MAXISTAYAWAYATSIXAM*  
MaxIstayawayatsixam  
MAXISTAYAWAYATSIXAM*
```

Exercise 4.13 change <DATA> to <>, and delete the line `__END__` and all that follows it. Then run the program using the command: `dotplot.pl < input_file`

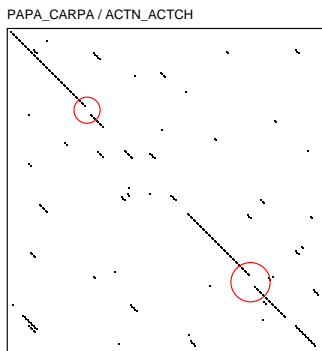
Exercise 4.14 0.16.

Exercise 4.15 (a) more similar, (b) more similar, (c) just more similar, (d) less similar.



Introduction to Bioinformatics — Answers to exercises

Exercise 4.16

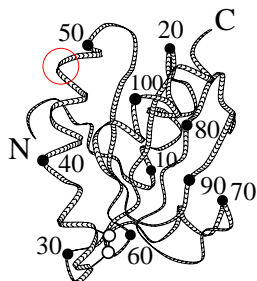


Exercise 4.17 Break the sequence into dipeptide or tripeptide sequences, and reassemble them in random order.

Exercise 4.18 4000000

Exercise 4.19  $65 = \text{Min}(40 + 25, 45 + 20, 50 + 25)$  where  $40 + 25$  corresponds to the vertical move,  $45 + 20$  to the diagonal move (mismatch) and  $50 + 25$  corresponds to the horizontal move. The two arrows appear because two possible predecessors give the same value.

Exercise 4.20 (a)



(b) about residue 48

Exercise 4.21 (a) VDFSAT. (b) score of VDFSAT = -1533, score of VDFSAT = -1503.

Exercise 4.22 (a) Scores computed by adding up six terms, the first of which is: inventory score of  $V \times \sum_{i=1}^{20} \text{BLOSUM62}(V, \text{amino acid } i)$

Residue number	number of																			
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
90					8			2	2	1								3		
91	1			1				2						1		7		4		
92						16														
93	15	1																		
94			2						2	2		6				1		3		
95				4					5				3			2		1		

(b) 32. (c) 49.

## Introduction to Bioinformatics — Answers to exercises

### Exercise 4.23

```

PAX6_human -----MQNSHSGVNQLGGVFNGRPLFDSTRQ 27
eyeless      MFTLQPTTALGTIVVPPWSAGTLIERLPSLEDMAHKGHSGVNQLGGVFNGRPLFDSTRQ 60

PAX6_human   KIVELAHSGARPCDISRIILQVSNQCVSKILGRYYETQSIKRAIGGSKPRVATRVVYSKI 87
eyeless      KIVELAHSGARPCDISRIILQVSNQCVSKILGRYYETQSIKRAIGGSKPRVATRVVYSKI 120

PAX6_human   AQYKRECPISIFAWEIRDRLLEGVCTNDNIPSVSSINRVLNRLASEKQQ----- 136
eyeless      SQYKRECPISIFAWEIRDRLLEGVCTNDNIPSVSSINRVLNRLAAKQEQSTGSGSSSTS 180

PAX6_human   -----MG-----ADG 141
eyeless      AGNSISAKVSVSIGNVSMVASGRGTLSSSTDLMQTATPLNSSESGGATNSGEGSEQEA 240

PAX6_human   MYDELRLMNGQTGS-----MGRTP----- 160
eyeless      IYEKRLLLNTQHAAGPGPLEPARAAPLVQQSPNHLGTRSSHQQLVRGNHQLQQRQQSW 300

PAX6_human   -----GWYFG-----TSVP-----GQP----- 172
eyeless      PPRHYSGSWYPTLSEIPISSAPNIASVTAYASGPLAHLSLSPNDIKSLASIGHQKPC 360

PAX6_human   -----TQDQCQQQGG---GENTNSISSNGBDEACMRLQLKRLQRNRTSFTQ 219
eyeless      VATEDIHLKKELDGHQSDTSGSGSENSNOGASNIQWTEDDQARLILKRLQRNRTSFTN 420

PAX6_human   EQIEALEKEFERTHYDFVAFERLAAKIDLPEARIQWFNFRRAKWRREKLRNQRQAS 279
eyeless      DQIDSLAKEFERTHYDFVAFERLAGKIGLPEARIQWFNFRRAKWRREKLRNQRRTFN 480

PAX6_human   NTFSHIPISSPSTSVYQPIQPTTFVSSPTSGSMLG----- 316
eyeless      STGASATSSSTASATSLTDFNLSLACSSLLSGSAGPVSVTINGLSFSTLSTNVAFT 540

PAX6_human   -----
eyeless      LGAGIDSSSEPTPIPHIRPSCSTSDNDWGRQSEDCRVCSPCLGVGGHQNTHHIQSNQHA 600

PAX6_human   -----RTDTALINTYSALFPMPFTMANNLPMQFPVP 348
eyeless      QGHALVPAISRLNFNSSGFGAMYSNMHTALMSDSYGAVTIPISFNHSVAGLAPPSP 660

PAX6_human   S-----QTSYSQMLFTSP-----SYNGRS 368
eyeless      IFQQGDLTFSSELYPCWMLRPPFMAPAHHHIVFGDGRPAGVGLSGQSANLGAESCSSG 720

PAX6_human   YDTYTP-----PHMQTHMSQP-----MGTS 389
eyeless      YEVLAYALFPFPMASSSAADSSPSAASASANVTPHHTLAQESCSPSCSASHPGVAHS 780

PAX6_human   GTTSTGLISFGVS-----VPVQVPS---EPDMSQYWRILQ----- 422
eyeless      SGPSSDPIISFVAVSYAHMSYNYASSANTMTFSSASGTSAAHVAPKQQFFASCFYSPWV 838
    
```

Exercise 4.24 (a) yes. (b) no.

Exercise 4.25 3: a(bc), b(ac), c(ab).

Exercise 4.26 The reduced distance matrix after combining ATCC and ATGC is shown in the text. The next step is to combine TTCG and TCGG. In the further-reduced distance matrix following that combination, the distance from {ATCC, ATGC} to {TTCG, TCGG} = 3. Therefore the distances from the root to {ATCC, ATGC} or to {TTCG, TCGG} =  $\frac{1}{2} \times 3 = 1.5$ .

Exercise 4.27 distance between ATCC and ATGC = 1 in both original and as sum of edges along shortest path through tree.

distance between ATCC and TTCG = 2 in original and 4.5 as sum of edges along shortest path through tree.

distance between ATCC and TCGG = 4 in original and 4.5 as sum of edges along shortest path through tree.

distance between ATGC and TTCG = 3 in original and 4.5 as sum of edges along shortest path through tree.

distance between ATGC and TCGG = 3 in original and 4.5 as sum of edges along shortest path through tree.

distance between TTCG and TCGG = 2 in both original and as sum of edges along shortest path through tree.

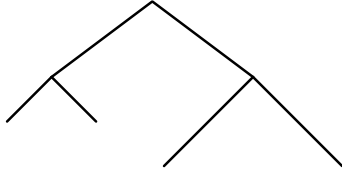
Exercise 4.28 a pentagon with nodes at each vertex, edges between pairs of adjacent vertices (many other possibilities).

Exercise 4.29 2

Exercise 4.30 sheep.

Introduction to Bioinformatics — Answers to exercises

Exercise 4.31



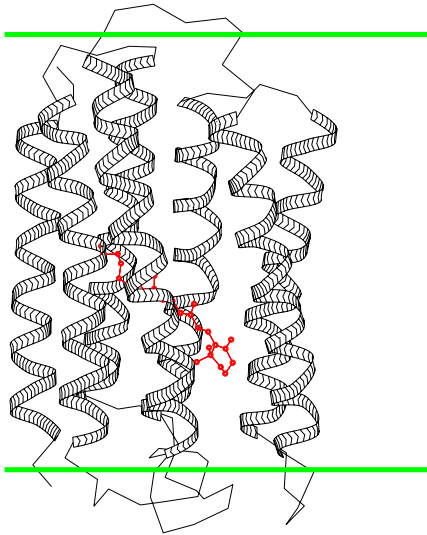
Exercise 4.32 (a) as  $n^2$ , (b) as  $n$ .

**Chapter 5.**

Exercise 5.1 25.5 kJ/hydrogen bond

Exercise 5.3 (a) paralogues, (b) orthologues, (c) paralogues, (d) paralogues, (e) paralogues, (f) neither orthologues nor paralogues.

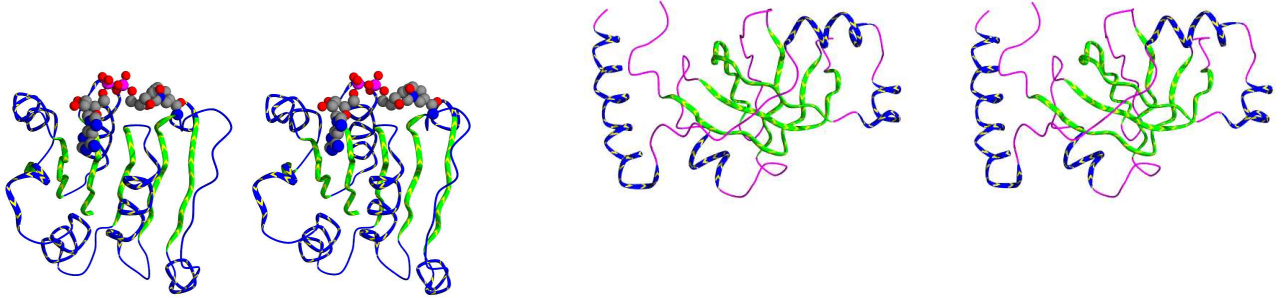
Exercise 5.4



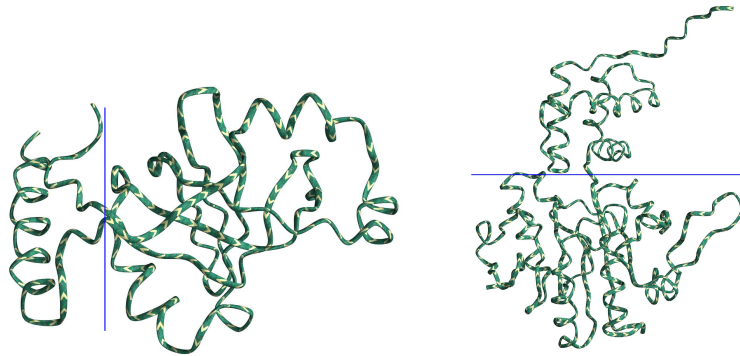
Exercise 5.5 (a) Only helices: a, b, c. (b) Only sheets: e, f. (c) Both helices and sheets: d, g, h, i, j, k, l, m, n. (d) Neither: o

Exercise 5.6  $\alpha$ -hairpin: a, c, l.  $\beta$ -hairpin: e, f, g, h, i, j.  $\beta - \alpha - \beta$  unit: k, l, m. TIM-barrel: n.

Exercise 5.7

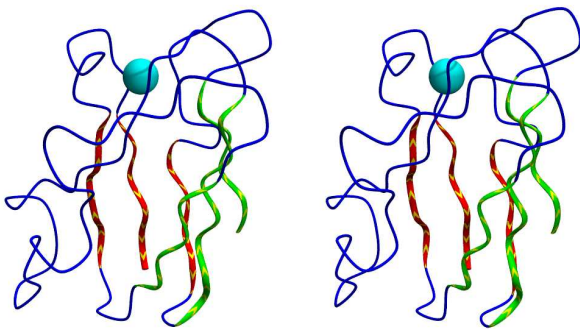


(a)



(b)

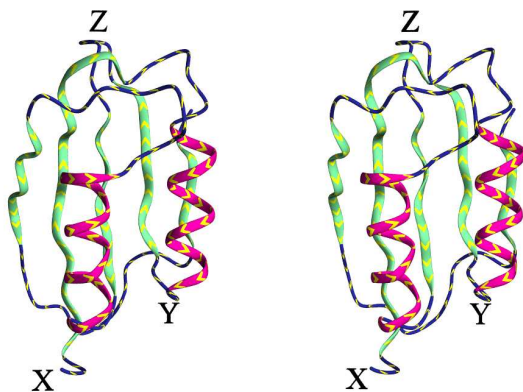
Exercise 5.8



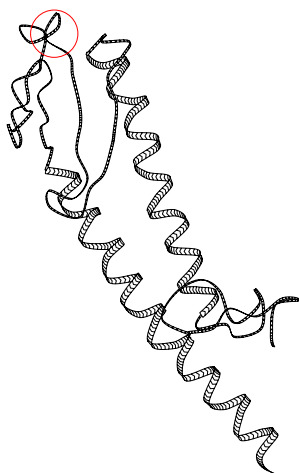
Exercise 5.9 m

Introduction to Bioinformatics — Answers to exercises

Exercise 5.10

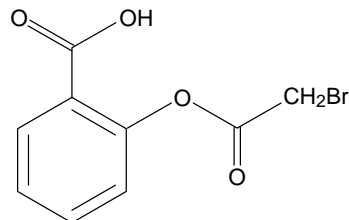
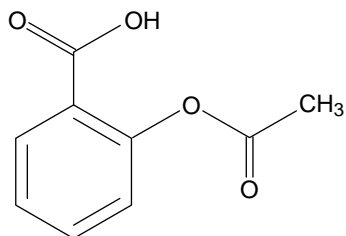


Exercise 5.11



Exercise 5.12 b

Exercise 5.13 (a) aspirin = acetylsalicylic acid (b) 2-bromoacetoxybenzoic acid



Exercise 5.14 You would search for homologous proteins that had different (and ideally separable) modes of determining specificity. Even better you would look for proteins carrying out functions essential to the pathogen that had NO human homologue.

Exercise 5.15 a

## Introduction to Bioinformatics — Answers to exercises

Exercise 5.16 For B, there is an additional entropy change involving the fixing of the conformation of the loop. This involves a negative entropy change. Therefore this would contribute to lowering the entropy change for  $B + X \rightarrow BX$  relative to  $A + X \rightarrow AX$ .

### Chapter 6.

Exercise 6.1 (a) the librarian could allocate resources according to likelihood of use. (b) probably not.

Exercise 6.2 (a) convert both the speech database and the query to text and use standard text-searching algorithms. (b) characterize each successive instant of the speeches and the query by a frequency spectrum and adapt text-searching algorithms to search for patterns in the time-ordered set of frequency spectra. (1/8000 seconds is an appropriate time interval).

Exercise 6.3 Cys.

Exercise 6.4 (a) virtual reality (computer graphics). (b) chemicals. (c) bioinformatics sequences. (d) log reports of web servers

Exercise 6.5

```
<mammals>
  <genus>Homo
    <species>sapiens</species>
      <common_name>human</common_name>
    <species>neanderthalis</species>
      <common_name>neanderthal man</common_name>
  </genus>
</mammals>
```

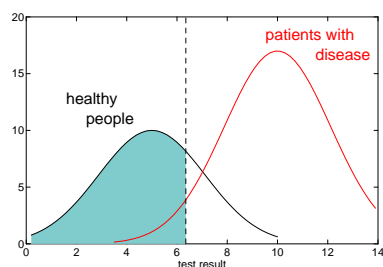
Exercise 6.6 (a) (1) Time appears to pass very quickly. (2) Measure the speed of flies the same way that you would measure the speed of an arrow. (3) A certain type of flies (time flies) are fond of arrows. (b) (3) because there are no time flies, probably (2) because it doesn't make much sense.

Exercise 6.7 <association of> ... <protein name> ... <with *OR* and> ... <protein name>

Exercise 6.8 Examples: base, briefcase, chase, erase, phrase, purchase, staircase; many geological terms such as plagioclase

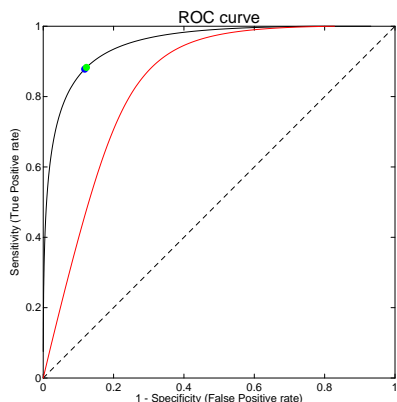
### Chapter 7.

Exercise 7.1 For example, the area shaded in cyan corresponds to healthy people correctly diagnosed as healthy = True Negatives



## Introduction to Bioinformatics — Answers to exercises

Exercise 7.2 (a) Blue circle. (b) Green circle. (Note that the circles are at almost the same position.)



Exercise 7.3 A is the best. Area under curve A = 1.0. Area under curve D = 0.5 Area under curves B and C are intermediate, with larger area under B.

Exercise 7.4 One might take a training set restricted to gram-positive bacteria, and check how well the results function for non-overlapping test sets from Gram-positive bacteria

Exercise 7.5 The PCA may suppress the relevant dimension.

Exercise 7.6 In principle, nucleic acid sequences contain more information.

Exercise 7.7 01100100101000111

Exercise 7.8

Smokers/nonsmokers:

There are 7 smokers – 4 high risk and 3 low risk. There are 3 non-smokers – 1 high risk and 2 low risk.

$$\text{Entropy}(\text{smokers}) = -\frac{4}{7} \log_2 \frac{4}{7} - \frac{3}{7} \log_2 \frac{3}{7} = 0.985$$

$$\text{Entropy}(\text{non-smokers}) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$\text{Total entropy} = 0.7 \times 0.985 + 0.3 \times 0.918 = 0.965$$

$$\text{Reduction in entropy} = 1 - 0.965 = 0.035$$

Exercise habits:

There are 3 people who undertake regular strenuous exercise – 1 high risk and 2 low risk. There is one person who undertakes moderate regular exercise – 0 high risk and 1 low risk. There are 6 people who undertake no regular exercise – 4 high risk and 2 low risk.

$$\text{Entropy}(\text{regular strenuous exercise}) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$\text{Entropy}(\text{moderate regular exercise}) = 0$$

$$\text{Entropy}(\text{no regular exercise}) = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} = 0.918$$

$$\text{Total entropy} = 0.3 \times 0.918 + 0.1 \times 0 + 0.6 \times 0.918 = 0.826$$

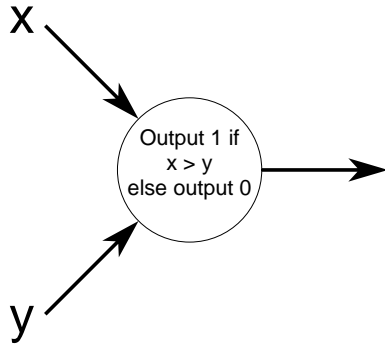
$$\text{Reduction in entropy} = 1 - 0.826 = 0.174$$

Exercise 7.9 14 weights and 14 thresholds

Introduction to Bioinformatics — Answers to exercises

Exercise 7.10 It selects points above and to the right of the line  $x + 2y = 2$ . (This line intersects the x-axis at  $x = 2$  and intersects the y-axis at  $y = 1$ .)

Exercise 7.11



Geometric interpretation: selects points below and to the right of the line  $y = x$ .

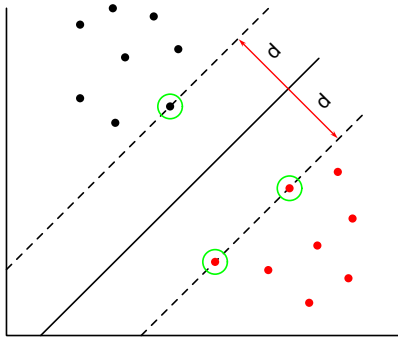
Exercise 7.12 yes.

Exercise 7.13 Because for all points on the left, either  $x$  or  $y$  is 0, therefore  $\sqrt{2}xy = 0$  for all four points.

Exercise 7.14

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

Exercise 7.15



Exercise 7.16 (a) 1. (b) 0. (c) 0.375

**Chapter 8.**

Exercise 8.1 (a)  $V_3-V_4$  (b)  $V_3-V_1$  (c)  $V_5-V_6$  (no matter how you answered (b) (d)  $V_2-V_6$  (e)  $V_3-V_4$



## Introduction to Bioinformatics — Answers to exercises

Exercise 8.2 (a) phylogenetic trees, parts of metabolic pathways, citation patterns, the World Wide Web. (b) metabolic pathways, chemical bonding patterns (if single / double / triple bonds are distinguished).

Graph	Nodes	Edges
Sets of people who have met each other	people	relationships
Electricity distribution systems	power stations	cables
Phylogenetic trees	species (or taxa)	lines of descent
Metabolic pathways	metabolites	enzyme-catalyzed reactions
Chemical bonding patterns in molecules	atoms	bonds
Citation patterns in the scientific literature	articles	references
The World Wide Web	sites	references

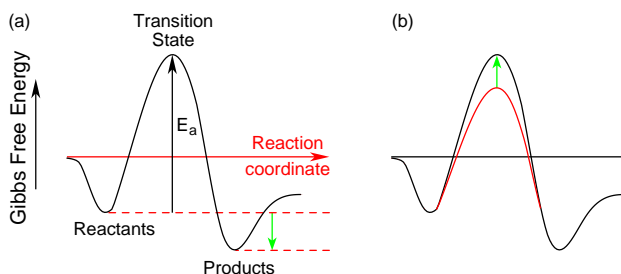
Exercise 8.3 (a) Moorgate → Bank → Waterloo → Embankment (b) King's Cross → Russell Square → Holborn → Tottenham Court Road → Oxford Circus → Euston Square → King's Cross (c) 2/15

Exercise 8.4 (a) Amersham – Upminster (b) Stations between Upney and Upminster, stations between West Brompton and Wimbledon, Chiswick Park (allowing access by National Rail on Richmond spur). (c) Stations between Stanmore and Baker Street (except for Wembley Park, West Hampstead – accessible by National Rail – and Finchley Road, Southwark, Bermondsey

Exercise 8.5 upper graph: 1 vertex has 4 neighbours, 4 vertices have 1 neighbour.

### Chapter 9.

Exercise 9.1 See green arrows in figure:



Exercise 9.2 Ex. 8.2. (a) If  $[S] = K_M$ ,  $v_0 = V_{\max} \times \frac{[S]}{[S]+[S]} = V_{\max} \times \frac{[S]}{2[S]} = \frac{1}{2}V_{\max}$

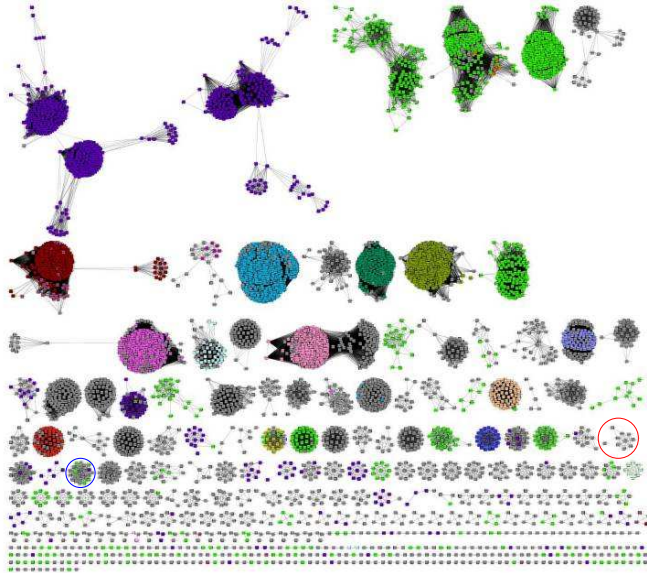
(b) If  $[S] \gg K_M$ ,  $v_0 \simeq V_{\max} \times \frac{[S]}{[S]} = V_{\max}$

(c) If  $[S] = 2K_M$ ,  $K_M = \frac{1}{2}[S]$ , and  $v_0 = V_{\max} \times \frac{[S]}{\frac{1}{2}[S]+[S]} = V_{\max} \times \frac{[S]}{1.5[S]} = \frac{2}{3}V_{\max}$

Exercise 9.3  $V_{\max}$  is the same, because at sufficiently high substrate concentrations the substrate swamps out the inhibitor.  $K_M$  will increase in the presence of inhibitor. (The lower availability of binding sites manifests itself as weaker substrate binding, corresponding to increased  $K_M$ .)

Introduction to Bioinformatics — Answers to exercises

Exercise 9.4 (a) red circle. (b) blue circle. (Not the only possible answers.)



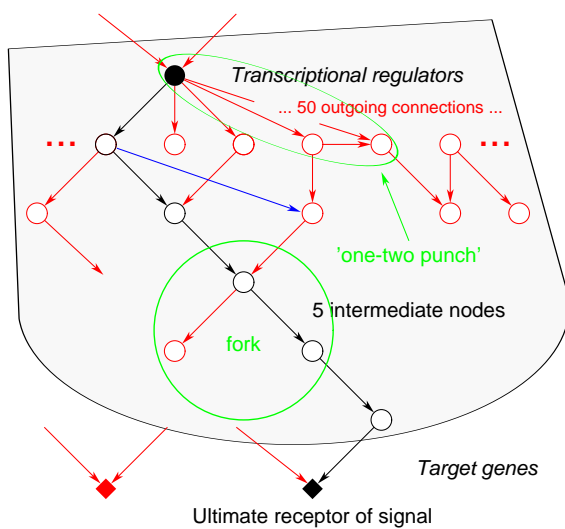
Chapter 10.

Exercise 10.1 (a) phage display. (b) X-ray and NMR structure determinations, Two-hybrid screening systems,

Exercise 10.2 (a) 10. (b) 17. (c) 178.5 kJ

Exercise 10.3 1/21.

Exercise 10.4



Exercise 10.5 (a) closed. (b)  $\beta$ -sheet

Exercise 10.6 0.201