

Genome 540 Discussion

Conor Camplisson

February 2nd, 2023

Outline

- Homework 5 overview
- Homework 4 & 5 questions

Outline

- Homework 5 overview
- Homework 4 & 5 questions

Homework 5 Overview

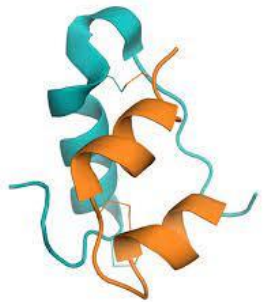
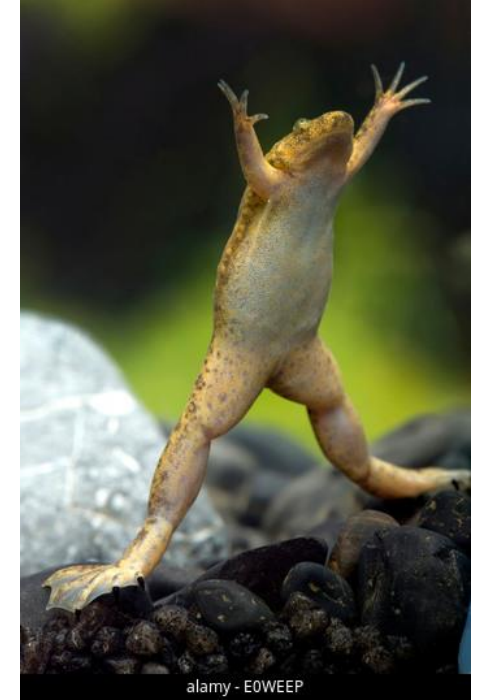
Homo sapiens



Bubalus bubalis



Xenopus tropicalis



Insulin

```
>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens GN=INS PE=1 SV=1  
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGQVELGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

```
>tr|Q25C78|Q25C78_BUBBU Insulin (Fragment) OS=Bubalus bubalis OX=89462 GN=bpi PE=2 SV=1  
FVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEGPQVGALELAGGPGAGGLEGPPQKRGIVEQC CASVCSLYQLENYCN
```

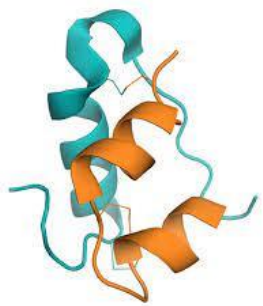
```
>tr|F6QRS1|F6QRS1_XENTR Insulin OS=Xenopus tropicalis OX=8364 GN=fbxw8 PE=3 SV=3  
MLPLSSIMALWMQCLPLVLLVLLFSTPNTEALANQHLCGSHLVEALYLVCGRGFFYYPKIKRDIEQAMVNGPQDNELDGMQLQPQEYQKMKRGIVEQCCHSTCSLFQLESYCN
```

Homework 5 Overview

Goal: find the best (local) multiple sequence alignment of three insulin proteins using the BLOSUM62 score matrix

Approach:

- Write a program that builds a weighted, 3D edit graph for the three sequences and exports it to .txt representation
- Determine the max weight path using your HW4 program



Insulin

```
>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens GN=INS PE=1 SV=1  
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

```
>tr|Q25C78|Q25C78_BUBBU Insulin (Fragment) OS=Bubalus bubalis OX=89462 GN=bpi PE=2 SV=1  
FVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEGPQVGALELAGGPGAGGLEGPPQKRGIVEQC CASVCSLYQLENYCN
```

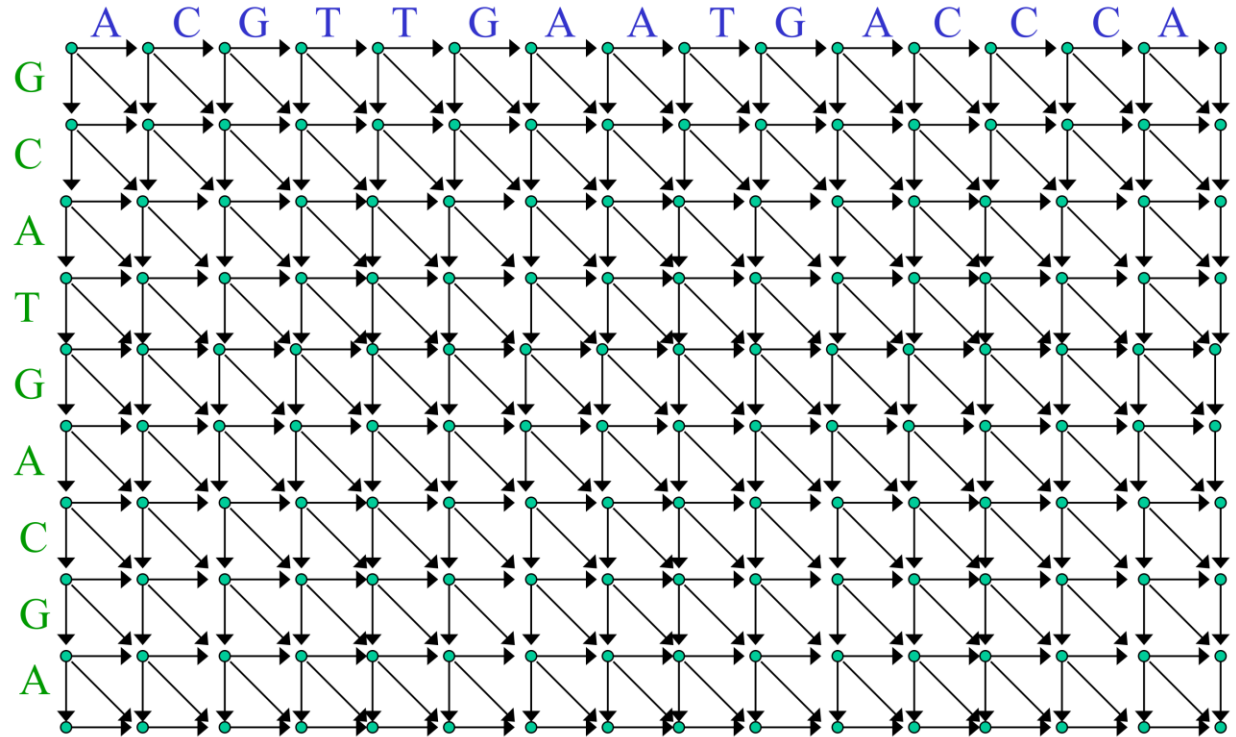
```
>tr|F6QRS1|F6QRS1_XENTR Insulin OS=Xenopus tropicalis OX=8364 GN=fbxw8 PE=3 SV=3  
MLPLSSIMALWMQCLPLVLVLLFSTPNTTEALANQHLCGSHLVEALYLVCGRGFFYYPKIKRDIEQAMVNGPQDNELDGMQLQPQEYQKMKRGIVEQCCHSTCSLFLQLESYCN
```

Homework 5 Overview

1 seq \rightarrow 1-D sequence graph



2 seqs \rightarrow 2-D Edit graph(pairwise alignment)



Homework 5 Overview

Sequence 1: from 1 to N1
 Sequence 2: from 1 to N2

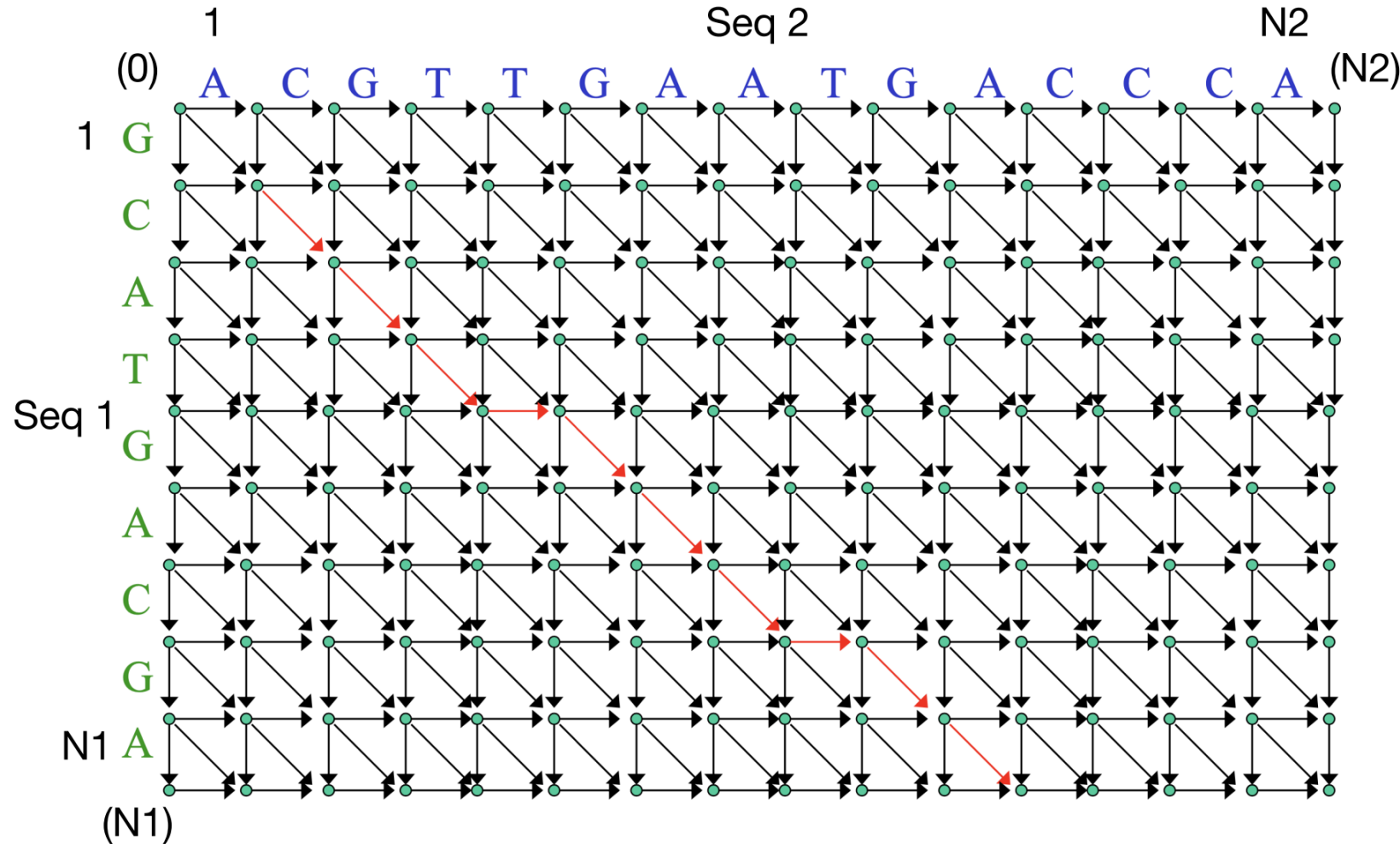
Pairwise alignment

Vertice: (0,0) (0,1) (0,2) ... (0,N2)
 (1,0) (1,1)
 (2,0) ...
 ...
 (N1,0) (N1,N2)

Vertice: two for loops

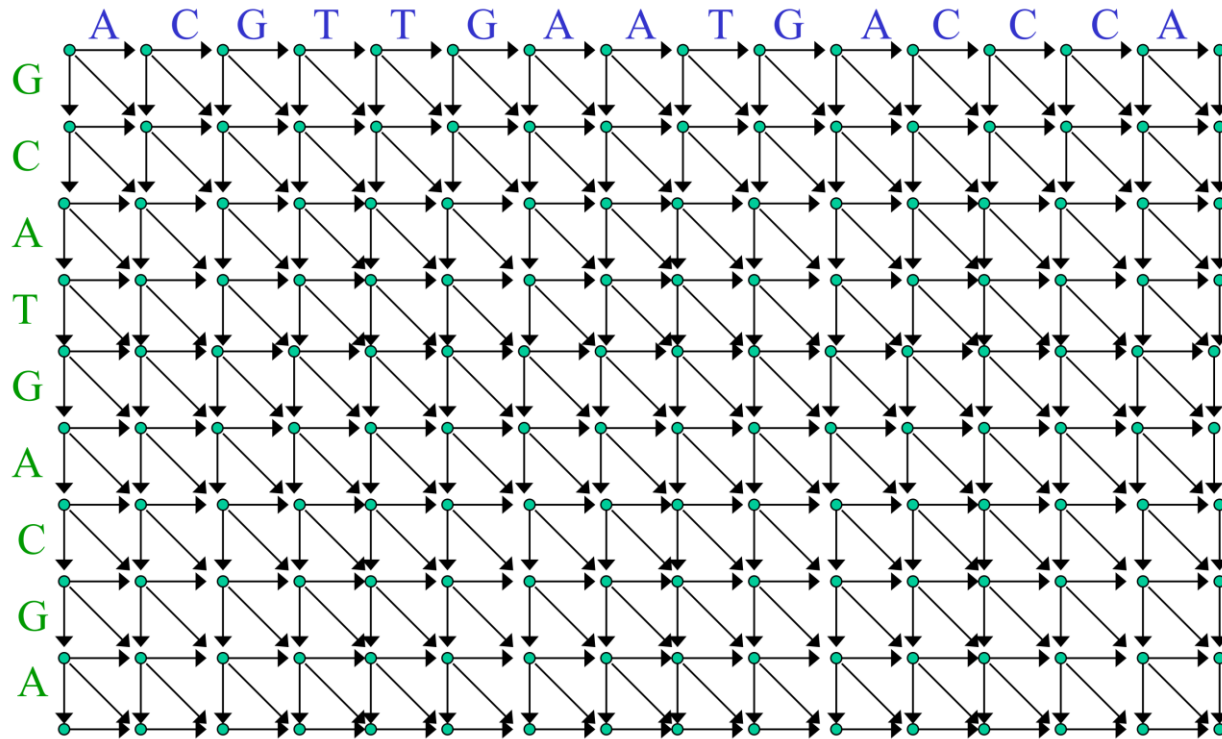
Edges: (0,0) (0,1) weight ($_A$)
 (0,0) (1,0) weight ($G_$)
 (0,0) (1,1) weight (GA)
 ...

Edges: for any node (i, j)
 (i, j) -> (i+1, j)
 (i, j) -> (i, j+1)
 (i, j) -> (i+1, j+1)

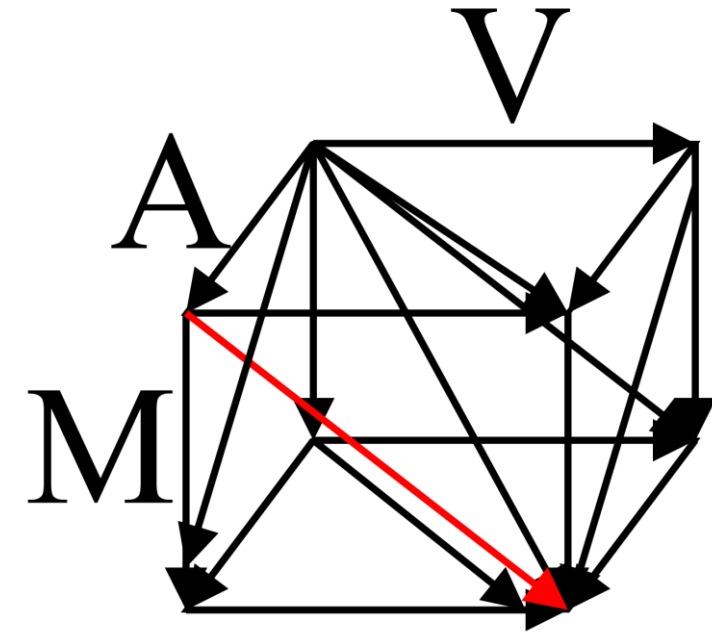


Homework 5 Overview

2 sequences → 2-D (pairwise alignment)

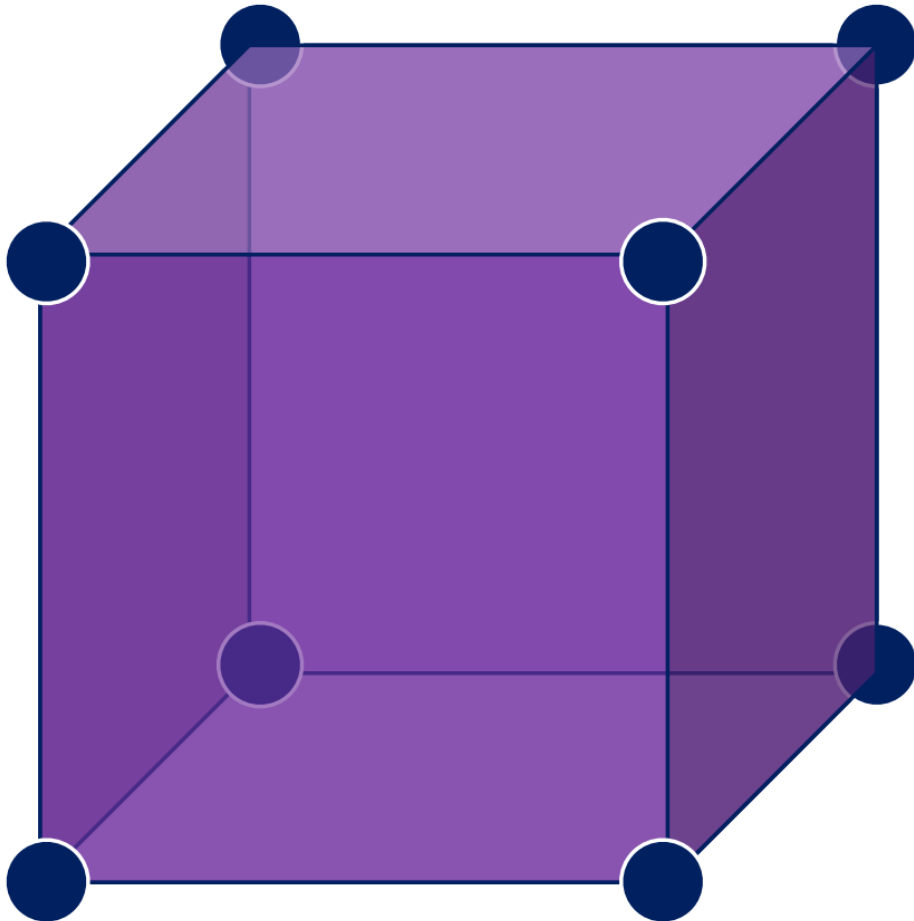
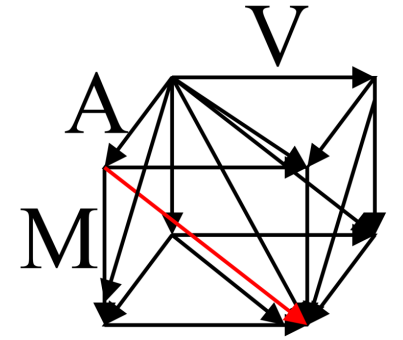
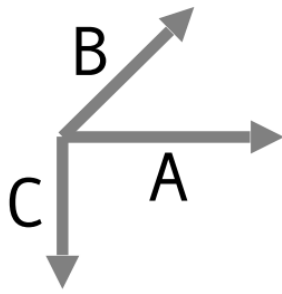


3 sequences → 3-D (multiple alignment)



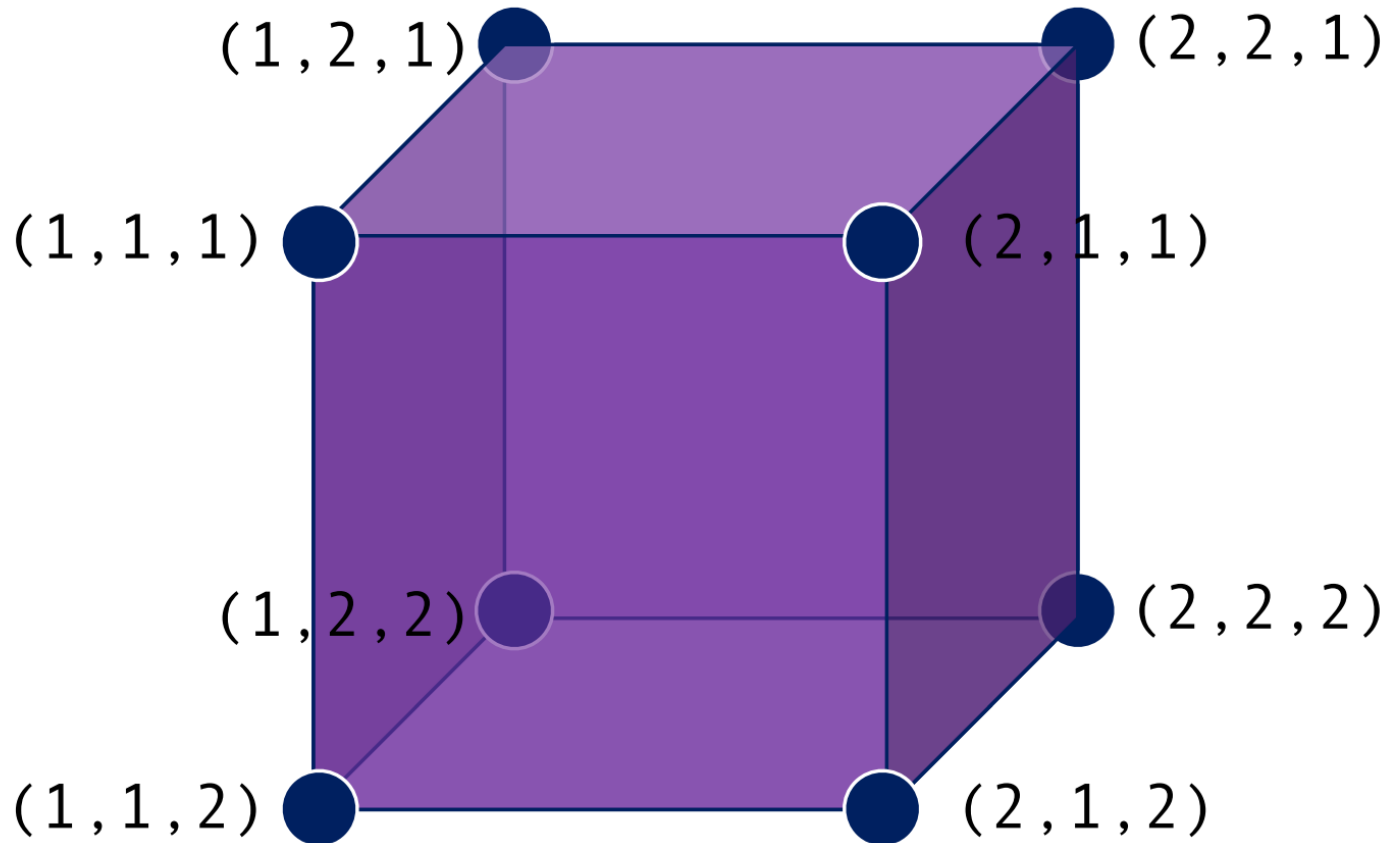
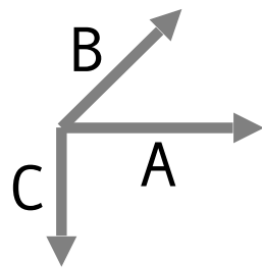
(each cell ^)

Homework 5 Overview



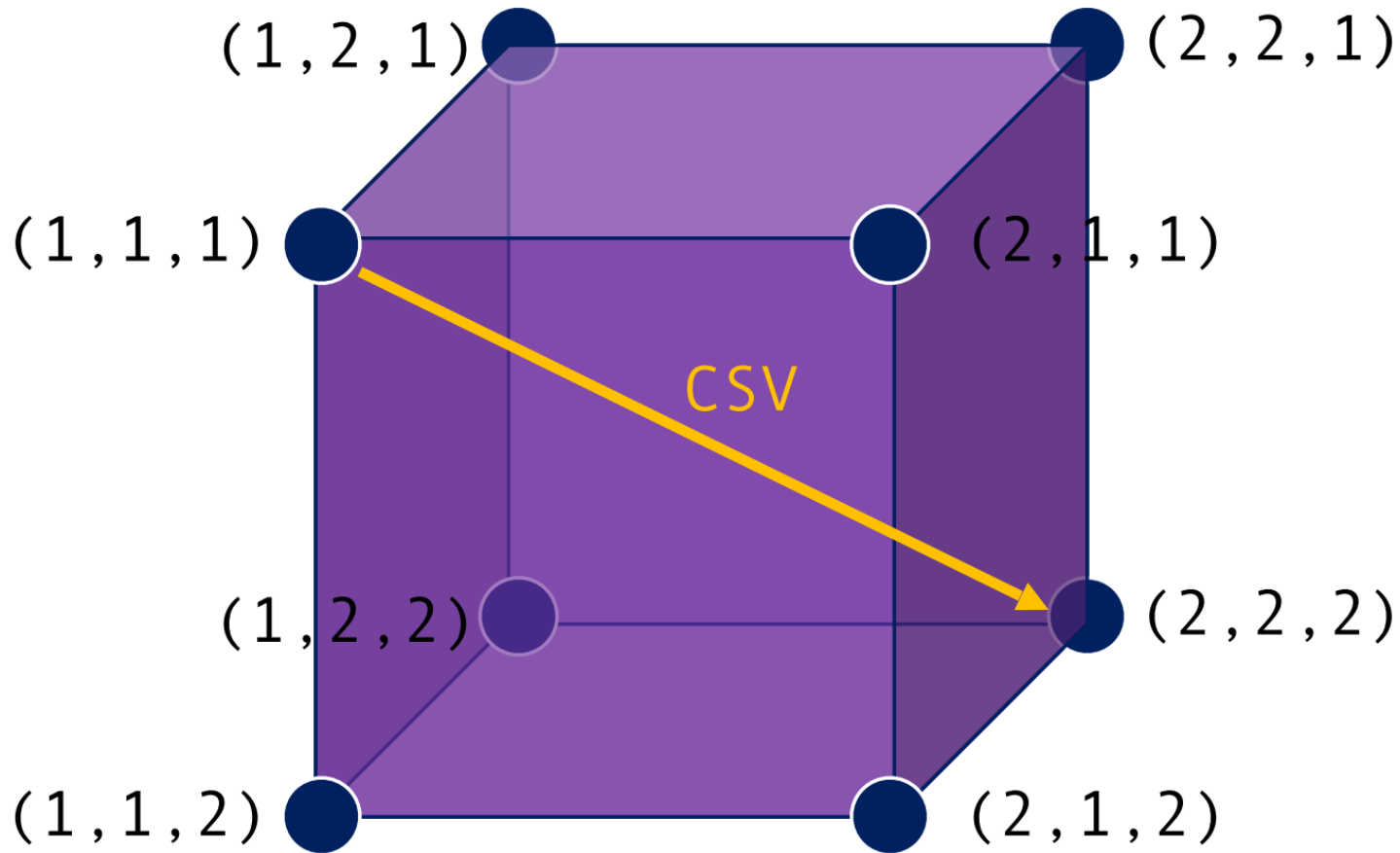
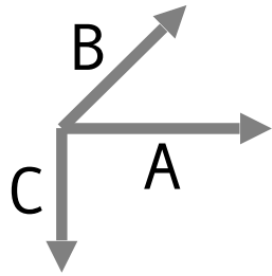
A = •M•C•D•R•...
B = •M•S•D•E•...
C = •M•V•D•R•...

Homework 5 Overview



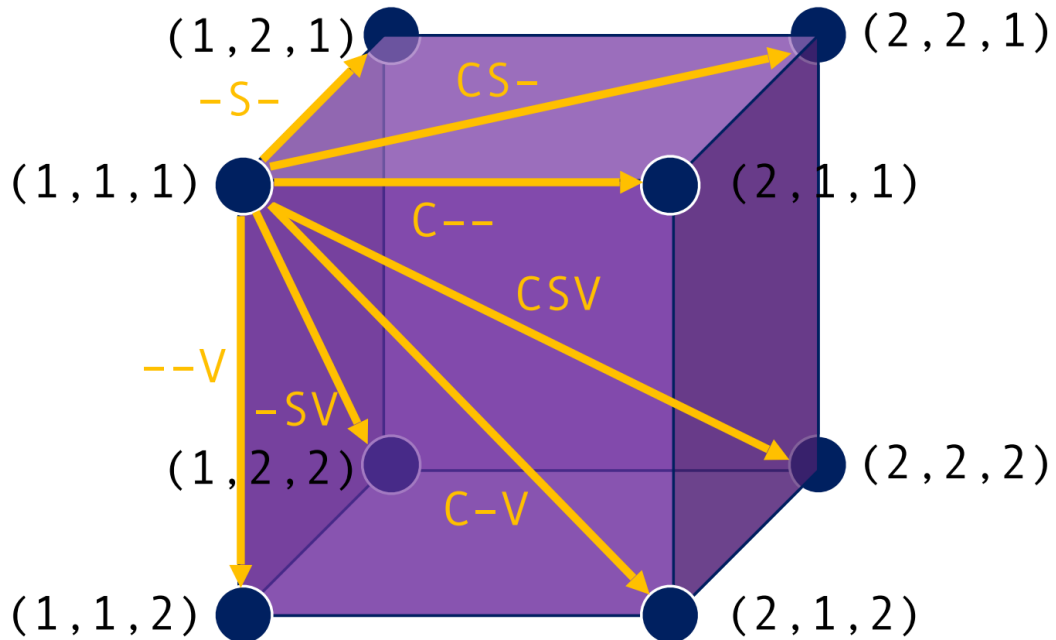
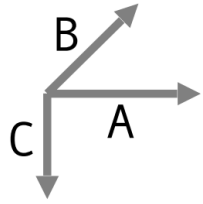
A = •M•C•D•R•...
B = •M•S•D•E•...
C = •M•V•D•R•...

Homework 5 Overview



$\dot{A} = \bullet M \bullet C \bullet D \bullet R \bullet \dots$
 $B = \bullet M \bullet S \bullet D \bullet E \bullet \dots$
 $C = \bullet M \bullet V \bullet D \bullet R \bullet \dots$

Homework 5 Overview



\cdot
 $A = \cdot M \cdot C \cdot D \cdot R \cdot \dots$
 $B = \cdot M \cdot S \cdot D \cdot E \cdot \dots$
 $C = \cdot M \cdot V \cdot D \cdot R \cdot \dots$

$$\text{weight}(\text{CSV}) = \text{score}(\text{CS}) + \text{score}(\text{CV}) + \text{score}(\text{SV})$$

$$\text{weight}(\text{CS-}) = \text{score}(\text{CS}) + \text{gap_penalty} + \text{gap_penalty}$$

Homework 5 Overview

the BLOSUM62 score matrix for the pairwise scores:

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	-4
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	-4
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	-4
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-4
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	-4
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	-4
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

Gap penalty: -6

Sum of pairs scoring

Edge:
x1, x2, x3

Edge weight:
sum((x1, x2), (x1, x3), (x2, x3))

- the corresponding score matrix entry if x_i and x_j are both residues
- the gap penalty if one of x_i and x_j is a residue, and the other is a gap character
- 0 if both x_i and x_j are gap characters

Homework 5 Overview

Your program should output the following:

1. The maximum path score
2. A list of all edge weights (sorted alphabetically by edge name)
3. A histogram of edge counts (again, sorted alphabetically by edge name)
4. The highest-scoring alignment, formatted vertically (as described above)

```
Assignment: GS540 HW5
Name: {YOURNAME}
Email: {YOUREMAIL}
Language: {YOURLANGUAGE}
Runtime: {YOURRUNTIME}
```

```
Score: 82.0
```

```
Edge weights:
```

```
--A = -12
--C = -12
--D = -12
--E = -12
--F = -12
```

```
.
.
.
list all edge weights in alphabetical order
(only first/last 5 shown here)
```

```
.
.
.
 YYS = 3
  YYT = 3
  YYV = 5
  YYW = 11
  YYY = 21
```

```
Edge counts:
```

```
--A = 8832
--C = 17664
--D = 52992
--E = 70656
--F = 44160
```

```
.
.
.
list all the edge counts in alphabetical order
(only first/last 5 shown here)
```

```
.
.
.
 YYS = 48
  YYT = 24
  YYV = 72
  YYW = 24
  YYY = 60
```

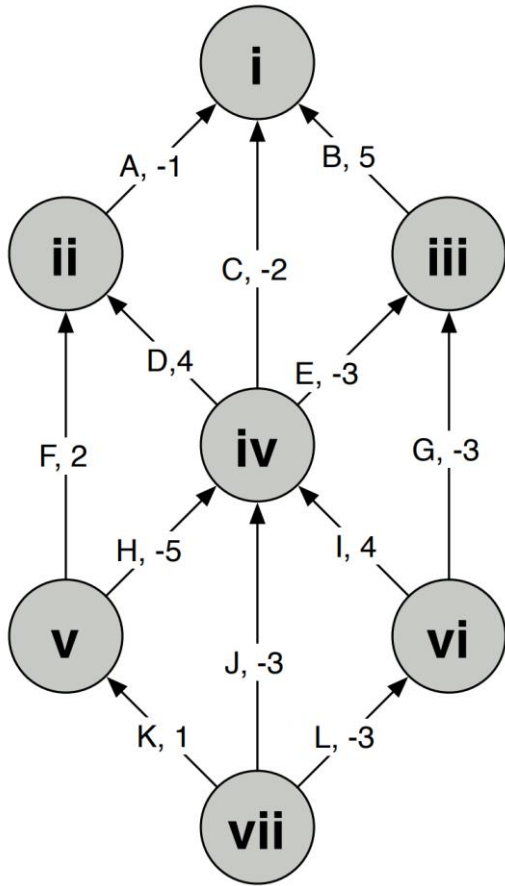
```
Local alignment:
```

```
KKK
DLK
YWY
G--
LFL
KVN
REH
IPI
QRQ
KPN
SNS
AVV
FFF
MVE
GSG
SGE
LVI
KKS
KDE
HSP
AVQ
LAI
NQL
STN
LVL
LVK
TDY
DYE
LLL
RLN
RQK
MYV
VTI
--R
--K
NPD
VVI
DED
SSS
-G-
VLV
IMI
VLL
FFF
```

Outline

- Homework 5 overview
- Homework 4 & 5 questions

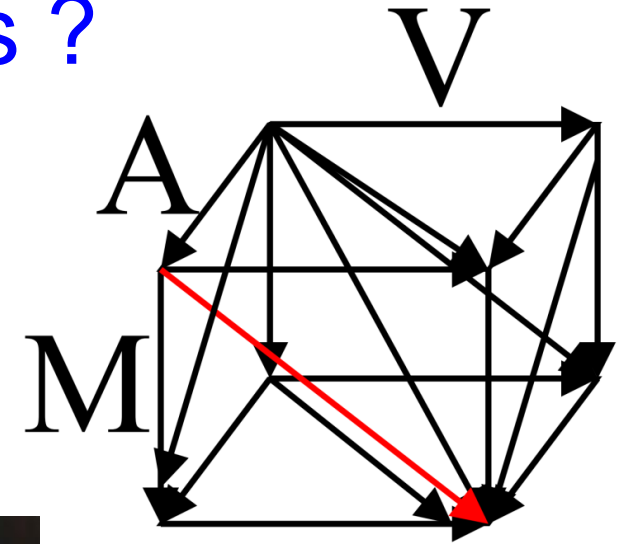
Homework 4 & 5 Questions ?



Assignment: GS 540 HW4
 Name: Conor Camplisson
 Email: concamp@uw.edu
 Language: C++/Python
 Runtime: 0m17.545s

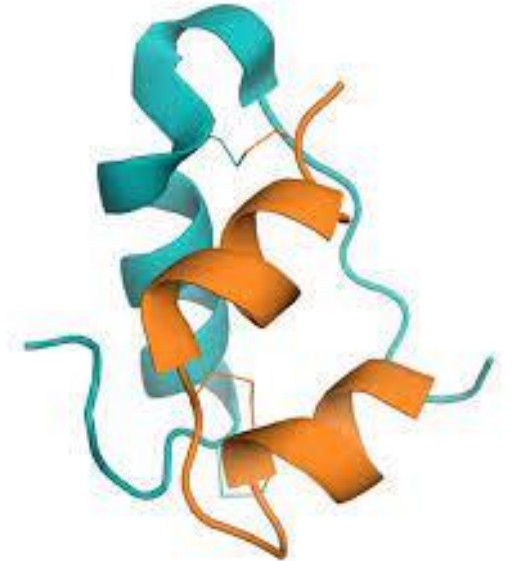
Part 1
 Score: 8
 Begin: vi
 End: ii
 Path: ID

Part 2
 Score: 4
 Begin: vii
 End: i
 Path: LIDA



Local alignment:

KKK
 DLK
 YWY
 G--
 LFL
 KVN
 REH
 IPI
 QRQ
 KPN
 SNS
 AVV
 FFF
 MVE
 GSG
 SGE
 LVI
 KKS
 KDE
 HSP



Insulin



Applications & Products

Home > Genome Editing > Products > Cas9 Nuclease, *S. pyogenes*

Cas9 Nuclease, *S. pyogenes* NEB r3.1 RR 37° 65°

Reminders

- Homework 4 due this Sunday Feb. 5, 11:59 pm
 - name in the file: `camp_lissson_hw4.txt.gz`
- Homework 5 due next Sunday Feb. 12, 11:59 pm

