

به نام خدا

امیررضا رجبی ۹۸۳۱۱۲۶

تمرین دوم

سوال اول :

	A	G	C	T	-
A	0	5	5	5	3
G	5	0	4	5	3
C	5	4	0	5	3
T	5	5	5	0	3
-	3	3	3	3	∞

$seq_1 = ACTGG$

$seq_2 = ACCTAGA$

(الف)

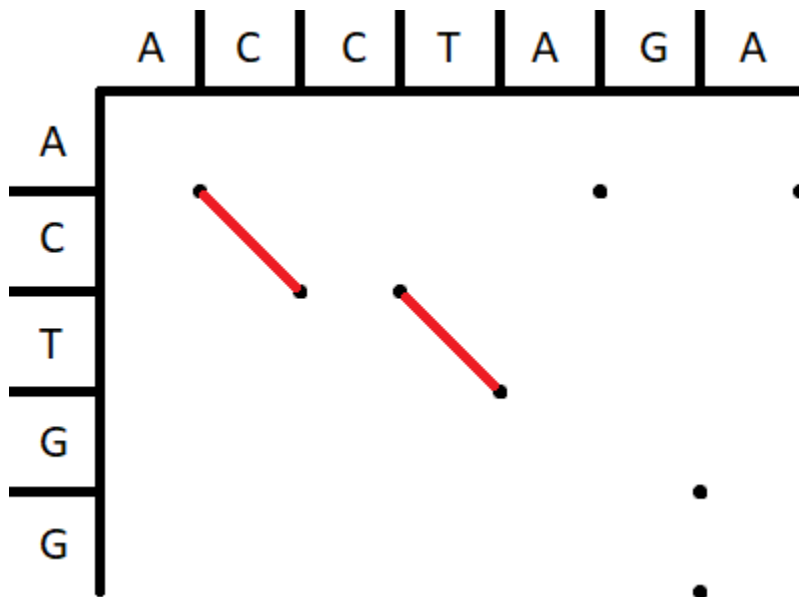
	-	A	C	C	T	A	G	A
-	0	-3	-6	-9	-12	-15	-18	-21
A	-3	0	-3	-6	-9	-12	-15	-18
C	-6	-3	0	-3	-6	-9	-12	-15
T	-9	-6	-3	-5	-3	-6	-9	-12
G	-12	-9	-6	-7	-6	-8	-6	-9
G	-15	-12	-9	-10	-9	-11	-8	-11

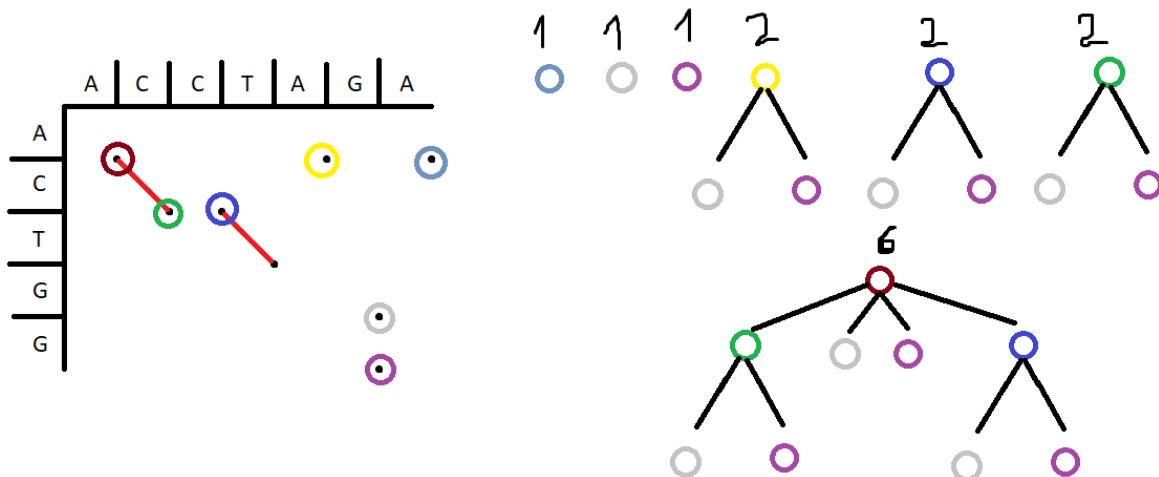
	-	A	C	C	T	A	G	A
-	0	-3	-6	-9	-12	-15	-18	-21
A	-3	0	-3	-6	-9	-12	-15	-18
C	-6	-3	0	-3	-6	-9	-12	-15
T	-9	-6	-3	-5	-3	-6	-9	-12
G	-12	-9	-6	-7	-6	-8	-6	-9
G	-15	-12	-9	-10	-9	-11	-8	-11

۴ دسته جواب وجود دارد

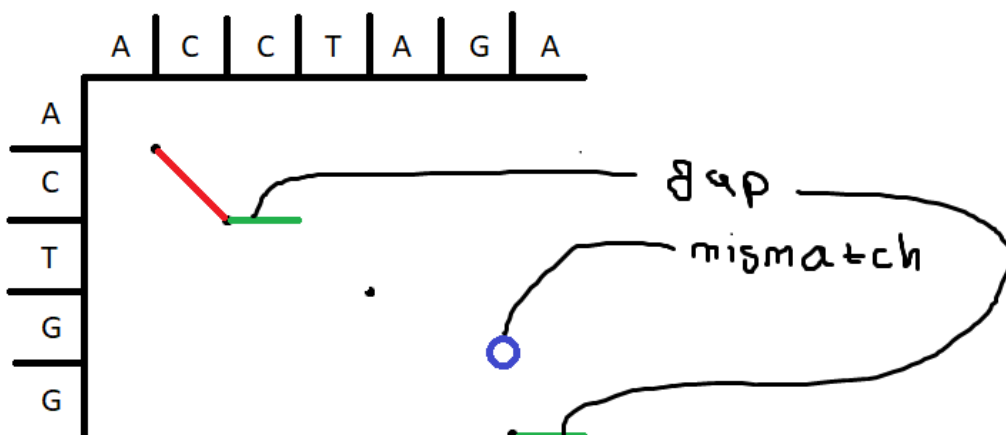
AC-TGG-	A-CTGG-	AC-T-GG	A-CT-GG
ACCTAGA	ACCTAGA	ACCTAGA	ACCTAGA

(ب)






برای بدست آوردن جواب با استفاده از روش دات ماتریس باید روی تمام حالات بی اف اس زد به این شکل داریم $1 + 1 + 1 + 2 + 2 + 2 + 6$ مساوی است با ۱۵ شکل و بعد از انجام این کار بالاترین امتیاز را به عنوان جاب پس می دهیم تفاوت این است اما میتوان با دیدن این که کدام خط ها قطری طول بیشتری دارد این را نیز فهمید با توجه به این که ما در اینجا سیستم نمره دهی داریم نه فقط تعداد **match** و **mismatch** برایمان مهم باشد پس جواب مد نظر این است که **AC-TGG-** و **ACCTAGA**



سوال دوم :

(الف)



	-	T	G	C	A	T	T	A	C	G	G	A
-	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-1	-2	-3	0	-1	-2	-3	-4	-5	-6	-7
G	-2	-2	2	1	0	-1	-2	-3	-4	-1	-2	-3
T	-3	1	1	1	0	3						
T	-4											
C	-5											
G	-6											
A	-7											

از سطر اول با توجه به خطی بودن گپ میتوان تشخیص داد gap برابر با 1- است

از نکته 1 روی شکل که تقاطع A, T است و mismatch رخ داده است میتوان میزان

mismatch را 1- بدست آورد زیرا که اگر گپ بود مقدار خانه برابر 2- بود

از نکته 2 روی شکل که تقاطع A, A است و match رخ داده است میتوان میزان match را 3

بدست آورد زیرا که اگر گپ بود مقدار خانه برابر 4- بود

$$\text{Match} = 3, \text{gap} = -1, \text{mismatch} = -1$$

(ب) حال جدول را پر میکنیم

	-	T	G	C	A	T	T	A	C	G	G	A
-	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-1	-2	-3	0	-1	-2	-3	-4	-5	-6	-7
G	-2	-2	2	1	0	-1	-2	-3	-4	-1	-2	-3
T	-3	1	1	1	0	3	2	1	0	-1	-2	-3
T	-4	0	0	0	0	3	6	5	4	3	2	1
C	-5	-1	-1	3	2	2	5	5	8	7	6	5
G	-6	-2	2	2	2	1	4	4	7	11	10	9
A	-7	-3	1	1	5	4	3	7	6	10	10	13

امتیاز سراسری برابر با ۱۳ میشود

	-	T	G	C	A	T	T	A	C	G	G	A
-	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	-1	-2	-3	0	-1	-2	-3	-4	-5	-6	-7
G	-2	-2	2	-1	0	-1	-2	-3	-4	-1	-2	-3
T	-3	1	1	1	0	3	2	1	0	-1	-2	-3
T	-4	0	0	0	0	3	6	5	4	3	2	1
C	-5	-1	-1	3	2	2	5	5	8	7	6	5
G	-6	-2	2	2	2	1	4	4	7	11	10	9
A	-7	-3	1	1	5	4	3	7	6	10	10	13

TGCATTACGGA

AG - -TT - C - GA

و با توجه به مسیر جواب ما میشود

ج (

اگر مسیر traceback را ببینیم متوجه میشویم که در خانه مشخص شده با سبز که مقدار ده دارد

TGCATTACGGA

AG - -TT - C G - A

مسیر ما دو شاخه میشود و نتیجه میگیریم که دو جواب دارد و جواب دوم :

سوال سوم :

	-	T	A	T	A	G	C
-	0	0	0	0	0	0	0
G	0	0	0	0	0	2	0
T	0	2	0	2	0	0	1
T	0	2	1	2	1	0	0
A	0	0	4	2	4	2	0
T	0	2	2	6	4	3	1
C	0	0	1	4	5	3	5

TAT

TAT

یک جواب وجود دارد و آن هم

است

Seq	Length
A	214
B	181
C	389

سوال چهارم :

(الف)

۱۷۳ تا کاراکتر در A و B مشابه هستند (در ۱۸۱ طول B بدون در نظر

گرفتن gap مشاهده با پیمایش سائِه) برای مقایسه A, B هم تریزی سراسری پیشنهاد میشود

چون طولی تقریبا مشابه دارند و طبق بالا شباهت زیادی نیز دارند اما برای A, C هم تریزی شبه

سراسری یا محلی پیشنهاد میشود چون طول C حدود دو برابر A است و بامقایسه نیمه سراسری

میتوان اورلپ های موجد را بدست آورد یا با محلی میتوان قسمت های مشابه را بدست آورد

(ب)

global of A, B

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 214
# Identity: 174/214 (81.3%)
# Similarity: 175/214 (81.8%)
# Gaps: 33/214 (15.4%)
# Score: 872.0
#
#=====
EMBOS_001      1 MEAWNVEKASADGNLPEVISINIKETLKIVSRTPVNITMAGDSGNHGHSFI 50
|||||
EMBOS_001      1 MEAWNVEKASADGNLPEVISINIKETLKIVSRTPVNITMAGDSGNHGHSFI 50
|||||
EMBOS_001     51 SALRNTGHEGKASPTTGLVKATQRCASYFSSHFNNVLDPGTGSATTT 100
|||||
EMBOS_001     51 SALRNTGHEGKASPTTGLVKATQRCASYFSSHFNNVLDPGTGSATKT 100
|||||
EMBOS_001    101 LEIVLMEHQFNRYDFIMVASAQFSNHHVWLAKTAEDMGKKFYIVNTKLDH 150
|||||
EMBOS_001    101 LEIVLMEHQFNRYDFIMVASAQFSNHHVWLAKTAEDMGKKFYIVNTKLDH 150
|||||
EMBOS_001    151 DLSTGALPEVQLQIRENVLENLQKERLACHEKYLKSTPENSTRPRNIPS 200
|||||
EMBOS_001    151 DLSTGALPEVQLQIRENVLENLQKERV-CEP----- 181
|||||
EMBOS_001    201 RRKLYVNLIRFINS      214
EMBOS_001    182 ----- 181
#-----
#-----
```

global of A, C

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 390
# Identity: 101/390 (25.9%)
# Similarity: 140/390 (35.9%)
# Gaps: 177/390 (45.4%)
# Score: 410.5
#
#=====
EMBOS_001      1 MXPALSLHTPLSTSTFPAAVWNPNGILSKTGAINIENAVGGGKLLVED 50
|||||
EMBOS_001      1 -----MEAWNVEKASADGNLPEVI 19
|||||
EMBOS_001     51 SIVRGTLTESSTPV-XAVTGGSSNMHSSFKALRVTGKKEEATAPTQNV 99
|||||
EMBOS_001     20 SNIKETLKIVSRTPVNITMAGDSGNHGHSFIISALRNTGHEGKASPTTEL 69
|||||
EMBOS_001    100 RTTXIPTCYSSSVIPSVLWHPGTGTATQSPENYLEMWFXYELFIIS 149
|||||
EMBOS_001     70 KATQRCASYFSSHFNNVLWLPDGTGSATTTLEIVLMEHQFNRYD-FIMV 118
|||||
EMBOS_001    150 ISEQFSNLIKLAQITQSLGKRIFYIWNKLDRLSTSAFWEELNQIQE 199
|||||
EMBOS_001    119 ASAQFSNHHVWLAKTAEDMGKKFYIWNKLDRLSTGALPEVQLQ-IRE 167
|||||
EMBOS_001    200 NTQKNLLKEGVCEPIIFLVSSIDPLNHPVPVPROTLHRYHGLNLEPYT 249
|||||
EMBOS_001    168 NVLENLQKERLACHEKYLKSTPE-----NSTRPR-----NTPSR 201
|||||
EMBOS_001    250 HEKVINVEVTSIXVKTASKFFQDTLGFQADDLGELCKAYHLFXVDOEC 299
|||||
EMBOS_001    202 RKLYVN-----LIRFINS----- 214
|||||
EMBOS_001    300 LQQVAQNMKGKPHVEYKTKNSQDLHTVHPRETLALYMNKNTASYISQI 349
|||||
EMBOS_001    215 ----- 214
EMBOS_001    350 PLLDDTIINVTRQKYRQFLGIVTKDTKILKLIQDFII 389
EMBOS_001    215 ----- 214
#-----
#-----
```

local of A, C

```
#####
# Aligned_sequences: 2
# 1: EBBOSS_001
# 2: EBBOSS_001
# Matrix: ELOSOUP62
# Gap_penalty: 10.0
# Extend_penalty: 0.5

# Length: 178
# Identity: 173/178 (97.2%)
# Similarity: 174/178 (97.6%)
# Gaps: 0/178 (0.0%)
# Score: 876.0

#####
EBBOSS_001 1 MEAHNWEKASADGLPEVLT
EBBOSS_001 1 MEAHNWEKASADGLPEVLT
EBBOSS_001 51 SALRTGHEGKASPTPLGV
EBBOSS_001 51 SALRTGHEGKASPTPLGV
EBBOSS_001 101 LENVLHLEQRIYRLQIRNV
EBBOSS_001 101 LENVLHLEQRIYRLQIRNV
EBBOSS_001 151 DLSTGLPEVLLQIRNV
EBBOSS_001 151 DLSTGLPEVLLQIRNV
```

```
#####  
# Aligned_sequences: 2  
# 1: EMBOS5_001  
# 2: EMBOS5_001  
# Matrv: EBLGUSW62  
# Gap_penalty: 10.0  
# Extend_penalty: 0.5  
#  
# Length: 176  
# Identity: 92/176 (52.3%)  
# Similarity: 120/176 (68.2%)  
# Gaps: 3/176 ( 1.7%)  
# Score: 419.0  
#  
#  
#  
  
EMBOS5_001      3   ANNVKEKASADGNLPVTSI  
|.....|..|..  
EMBOS5_001      34  AINZINAVGGSGKLVEVDI  
|.....|..|..  
  
EMBOS5_001      53  LRINTHEGKASPPTLVKA  
|.|||...|..|..  
EMBOS5_001      83  LRVTHGEKETATPGTWS  
|.....|..|..  
  
EMBOS5_001     103  NLYLEMFQIRYD-FIMVA  
|.....|..|..  
EMBOS5_001     133  NLYLEEMFXYELFIISIS  
|.....|..|..  
  
EMBOS5_001     152  LSTGLPEVLQLQ-IREIN  
|.....|..|..  
EMBOS5_001     182  LSTSPAFNEEILNQITEN  
|.....|..|..  
  
#####
```

local of A, C

```
#####
#
# Aligned_sequences: 2
# 1: EMBOS5_001
# 2: EMBOS5_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 214
# Identity: 174/214 (81.3%)
# Similarity: 175/214 (81.8%)
# Gaps: 33/214 (15.4%)
# Score: 872.0
#
#
#####
```

```

# Aligned_sequences: 2
# 1: EMBOS5_001
# 2: EMBOS5_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5

#
# Length: 390
# Identity: 101/390 (25.9%)
# Similarity: 140/390 (35.9%)
# Gaps: 177/390 (45.4%)
# Score: 410.5
#
#

```

```
#####
# Aligned_sequences: 2
# 1: EMBROSS_001
# 2: EMBROSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 178
# Identity: 173/178 (97.2%)
# Similarity: 174/178 (97.6%)
# Gaps: 0/178 ( 0.0%)
# Score: 876.0
#
#
#####
```

```
#####
#
# Aligned_sequences: 2
# 1: EMBOS001
# 2: EMBOS001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 176
# Identity:      92/176 (52.3%)
# Similarity:   120/176 (68.2%)
# Gaps:         3/176 ( 1.7%)
# Score: 419.0
#
#####
```

همانطور که میبینیم نتیجه محلی بر A, C بهتر از سراسری بود.

در A, B نیز همین نتیجه را مشاهده میکنیم اما با توجه به طول برابر آن ها میتوان نتیجه گرفت

که سراسری هم مناسب است البته که درصد تشابه در محلی بسیار بالا است

سوال پنجم :

PAM:

برای محاسبه pam1 این مراحل طی شده است :

محاسبه فرکانس جابه جایی برای جفت آمینو اسید را در خانواده های پروتینی نزدیک به هم

بدست می آوریم و آن را A می نامیم با هم ترازوی سراسری بدست می آید

فرکانس وقوع یک آمینو اسید را در کل مجموع دنباله پروتین ها مینامیم و آن را f می نامیم

جهش نسبی را برای هر آمینو اسید میابیم و m می نامیم

$$m_i = f_i . (number\ of\ times\ i\ is\ observed\ to\ mutate)$$

ماتریس احتمال جهش را برای هر آمینواسید که با دیگری جایگزین شده است محاسبه میکنیم M

$$M_{ij} = m_i \cdot \frac{A_{ij}}{\sum_{k=1}^{20} A_{kj}}$$

محاسبه مقیاس فاصله تکاملی به طوری که تنها ۱ درصد آمینو اسید تغییر می کند

$$M_{ii} = 1 - \lambda \cdot \sum_{j=1}^{20} M_{ij}$$

که لاندا ضریب تکامل است

نحوه محاسبه PAMN به این شکل است که PAM1 را به توان N میرسانیم

شماره های بیستر نشان دهنده فاصله تکاملی بیشتر است و در ساخت درخت فیلوژنی کاربرد دارد

BLOSUM :

مراحل زیر را طی میکنیم

دنباله هایی که بیشتر از یک ترشهلدی یکسان هستند را حذف میکنیم که با دو راه انجام میشود این را می توان با حذف دنباله ها از بلوک یا فقط با یافتن دنباله های مشابه و جایگزینی آنها با دنباله های جدیدی که می تواند خوشه را نشان دهد، انجام داد.

پایگاه داده که ترازهای توالی حفاظت شده ترین مناطق خانواده های پروتئین را ذخیره می کند. این ترازها برای استخراج ماتریس های BLOSUM استفاده می شوند. فقط دنباله هایی با درصد هویت کمتر از ترشهلد استفاده می شود. با استفاده از بلاک جفت آمینو اسید هارا در هر ستون از تراز ها می شماریم

نسبت وقوع آمینو اسید در هر ترکیب در داده های مشاهده شده را به مقدار مورد انتظار وقوع جفت نشان می دهد و در ماتریس جایگزینی استفاده می شود

این ماتریس بر اساس ۲۰۰۰ بلاک که هر کدام شامل ۵۰۰ دنباله است ساخته میشود

در واقع ماتریس بلوسام از مشاهدات آماری مستقیم بدست می آید و در آن از هم تراز می محلی استفاده میشود و کاربرد آن برای پیدا کردن دومین های پروتئینی و مناطق حفاظت شده در پایگاه داده است

و بر خلاف PAM شماره بیشتر نشان دهنده شباهت بیشتر و تکامل کمتر است و عدد N جلوی آن نشاندهنده شباهت در منطقه آماری است که ماتریکس از آن ساخته شده است (همان ترشهلد که در بالا آمده است)