



بسمه تعالی

تمرین سوم درس مقدمه ای بر بیوانفورماتیک

نیم سال دوم سال ۱۳۹۸-۱۳۹۹

مهلت تحویل ساعت ۲۳:۵۵ روز سه شنبه ۱۳۹۹/۲/۳۰

۱. در کلاس یاد گرفتیم که چگونه یک Position Specific Scoring Matrix بسازیم. دیدیم که یک PSSM می تواند برای امتیازدهی به یک زیررشته در یک رشته بدون برچسب مورد استفاده قرار گیرد. این امتیاز را می توان به صورت $\log \frac{P(s|H_A)}{P(s|H_O)}$ محاسبه کرد که در آن H_A نمونه ای از الگوی حفظ شده است و H_O نمونه ای از توزیع پس زمینه است. ما می توانیم از یک چارچوب مشابه برای مدل سازی یک جفت زیررشته که نمایانگر مکان های اتصال دو پروتئین که با هم در تعامل هستند استفاده کنیم. این تعامل باید خاص پروتئین های مذکور باشد؛ یعنی هر پروتئین باید به شریک خود متصل شود، و نه دیگر پروتئین هایی که لایه های مشابه دارند (ساختار سوم پروتئین). ویژگی تعامل با تعداد کمی از جفت آمینواسیدها در مکان های خاص در هر یک از پروتئین ها کنترل می شود. جایگزینی در یک محل مشخص تعامل با نیازمندی های تعامل پروتئین-پروتئین محدود می شود و تنها در صورتی حفظ خواهد شد که جایگزینی جبران کننده ای در محل مربوطه در پروتئین دیگر اتفاق بیافتد. با مقایسه هم ترازی های چندگانه جفت های در تعامل، می توان دریافت که این مکان ها همبسته هستند. این همبستگی می تواند برای پیش بینی اینکه دو توالی جدید کشف شده با هم تعامل دارند یا نه، مورد استفاده قرار گیرد.

همترازی‌های زیر مکان‌های اتصال چهار جفت در تعامل را نشان می‌دهد. مکان‌های تعیین‌کننده ویژگی اتصال توسط اعداد در بالای ستون‌ها مشخص شده‌اند (اعداد ۱ تا ۴). ستون‌های دارای برچسب مشابه با هم نشان‌دهنده تعامل جفت باقیمانده (آمینواسید) هستند.

	12	3	4			2	1	3	4																				
EnvK:	R	T	L	L	M	A	G	V	S	H	D	L	R	T	P	L	T	R	I	R	L	A	T	E	M	M	S	E	Q
CssK:	R	T	L	L	Q	N	-	I	S	H	D	L	K	T	P	V	T	R	I	R	L	Y	T	Q	S	I	K	D	G
LiaK:	E	R	Q	R	L	A	R	D	L	H	D	A	V	S	Q	Q	L	E	A	I	V	M	M	T	S	A	V	L	E
DegK:	E	R	K	R	V	S	R	E	I	H	D	G	P	A	Q	M	L	E	N	V	V	M	R	S	M	L	I	E	R
RstK:	K	K	Q	L	I	D	G	I	A	H	E	L	R	T	P	L	V	R	L	R	F	R	L	E	M	S	D	N	-
EnvR:	K	I	L	V	V	D	D	D	M	R	L	R	A	L	L	E	R	Y	L	T	E	Q	G	F	Q	V	R		
CssR:	T	I	Y	L	V	E	D	D	D	N	L	N	E	L	L	T	K	Y	L	E	N	E	G	W	N	I	T		
LiaR:	R	V	L	L	I	D	D	H	E	M	V	R	M	V	L	A	K	F	L	E	A	Q	P	D	I	E	V		
DegR:	N	I	V	I	I	D	D	H	Q	L	F	R	E	V	V	K	R	I	L	D	F	E	P	T	F	E	V		
RstR:	T	I	V	F	V	E	D	D	A	E	V	G	A	L	I	A	A	Y	L	A	K	H	D	M	Q	V	T		
Protein 1													Protein 2																

حال با توجه به توضیحات بالا به سوال‌های زیر پاسخ دهید.

الف) یک ماتریس بسازید که فرکانس جفت آمینواسیدهای مشاهده شده در چهار جفت محل مربوط به ویژگی تعامل را نشان دهد. ماتریس فرکانس شما باید چهار ستون با برچسب‌های ۱، ۲، ۳ و ۴ و یک ردیف برای هر جفت آمینواسید مشاهده شده در داده‌ها داشته باشد. (یک ماتریس کامل دارای ۲۰۲ ردیف خواهد بود. برای صرفه جویی در فضا، فقط ردیف‌های مشاهده شده را که برای آنها فرکانس جفت آمینواسیدها غیر صفر است درج کنید.) نیازی نیست از pseudocountها استفاده کنید. راهنمایی: برای مثال فرکانس جفت آمینواسید TL در محل ۱ برابر ۰,۴ است.

ب) برای هر یک از جفت آمینواسیدهای مشاهده شده در ماتریس قسمت (الف)، احتمال مشاهده‌ی تصادفی آن جفت آمینواسید را محاسبه کنید. از فرکانس‌های آمینواسیدهای پس زمینه (Background amino acid frequencies) داده شده در جدول زیر استفاده کنید.

راهنمایی: احتمال مشاهده جفت آمینواسید TL به‌طور تصادفی برابر $0.06 \times 0.091 = 0.0055$ است.

A	C	D	E	F	G	H	I	K	L
0.075	0.018	0.052	0.063	0.039	0.071	0.022	0.055	0.058	0.091
M	N	P	Q	R	S	T	V	W	Y
0.028	0.046	0.051	0.041	0.052	0.074	0.060	0.065	0.013	0.033

ج) از ماتریس های فرکانس و پس زمینه ای که در (الف) و (ب) ساخته اید استفاده کنید تا یک ماتریس نسبت احتمال $\left(\frac{P(s|H_A)}{P(s|H_O)}\right)$ ایجاد کنید که نشان دهنده نسبت احتمال برای این مکان های همبسته است.

د) ماتریس امتیازدهی log-odds را برای این مکان های همبسته ایجاد کنید.

ه) با استفاده از ماتریس log-odds بدست آمده از بخش (د)، امتیاز جفت آمینواسیدهای رشته های زیر را بدست آورید.

```

          12  3  4
YvrK: REEWIAGLSHDLKTPLTEIYVYSMMLESK
          2  1  3  4
YvrR: SILIVDDHKALVDVIKAVLEKEGYRNILDAASAEAAIPVV

```

ر) امتیاز جفت آمینواسیدهای رشته های زیر را نیز محاسبه کنید.

```

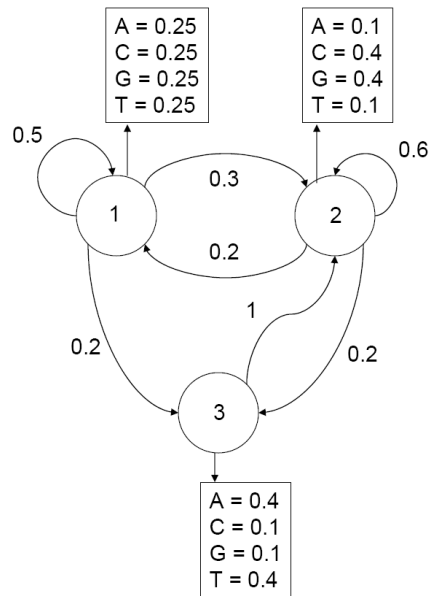
          12  3  4
WalK: RREFVANVSHELRTPLLEMRFYLESLEAG
          2  1  3  4
WalR: KILVVDDHKPFADLLEKNLRKEGYEVHCAHDGNEAVEMVE

```

ز) براساس نتایج PSSM، کدام یک از دو جفت پروتئینی از قسمت های (ه) و (ر) به احتمال زیاد در تعامل با یکدیگر هستند؟

۲. چرا روش پروفایل که در آن از هم ترازی با gap استفاده می شود نسبت به PSSM برای تشخیص توالی motifها در پروتئین ها حساس تر و انعطاف پذیر است؟

۳. با استفاده از مدل مخفی مارکوف زیر و با فرض شروع کار از حالت ۱، محتمل ترین دنباله حالت (مسیر) برای رشته ای AGAT را ضمن پرکردن جدول زیر محاسبه کرده و بنویسید.



	A	G	A	T
1	0.25			
2	0			
3	0			

لطفا به نکات زیر توجه کنید.

۱. تمرینات را به صورت انفرادی انجام دهید.
۲. pdf پاسخ های خود را با کیفیت مناسب و خوانا اسکن کرده و یا تایپ شده به صورت یک فایل درآورید و با نام "HW3_StudentNumber.pdf" روی quera آپلود کنید.
۳. تمیزی و خوانایی پاسخ تمرینات از اهمیت بالایی برخوردار است.
۴. در صورتی پیداشدن هرگونه کپی چه در سوال های تشریحی و چه برنامه نویسی نمره ی هر دو نفر ۱۰۰- در نظر گرفته خواهد شد.
۵. اشکالات خود را می توانید از طریق ایمیل بپرسید.
Mahdis.sfri98@gmail.com
Najmeh.mohammadbagheri77@gmail.com
۶. قاعده ی اعمال جریمه برای ارسال های با تأخیر به صورت زیر است.

$1d\ 100 - ((25 * delay) / 86400)$

2d 75

3d 50

