



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

مقدمه‌ای بر بیوانفورماتیک

پاییز ۱۴۰۳

استاد: علی شریفی زارچی

مسئول تمرین: امیررضا آرانپور

مهلت ارسال نهایی: ۱۸ آبان

تمرین اول

مهلت ارسال بدون تاخیر: ۱۴ آبان

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روزهای مشخص شده است.
- در طول ترم، برای هر تمرین می‌توانید تا ۴ روز تأخیر داشته باشید و در مجموع حداکثر ۸ روز تأخیر مجاز خواهید داشت. توجه داشته باشید که تأخیر در تمرین‌های عملی و تئوری به صورت مشترک محاسبه می‌شود. پس از اتمام تأخیرهای مجاز، می‌توانید با تاخیری ساعتی ۱ درصد تمرین خود را ارسال کنید.
- حتماً تمرین‌ها را بر اساس موارد ذکر شده در صورت سوالات حل کنید. در صورت وجود هرگونه ابهام، آن را در صفحه تمرین در سایت کوثر مطرح کنید و به پاسخ‌هایی که از سوی دستیار آموزشی مربوطه ارائه می‌شود، توجه کنید.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- فایل پاسخ‌های سوالات نظری را در قالب یک فایل pdf به فرمت `[STD_ID]_HW1.pdf` آپلود کنید.
- گردآوردندگان تمرین: امیررضا آرانپور، رها قره‌داغی، عرفان محمدی، زهرا رجالی

سوالات نظری (۱۰۰ نمره)

۱. (۱۵ نمره) فرض کنید ژنی که مسئول تولید پروتئینی خاص در سلول‌های عضلانی است، دچار جهش شده است. این جهش باعث شده است که عضلات توانایی بازسازی پس از آسیب را از دست بدهند.

الف) توضیح دهید که چگونه جهش در یک ژن می‌تواند باعث کاهش یا توقف بازسازی سلول‌های عضلانی شود.

ب) فرض کنید که این جهش ژنی در یک کروموزوم X قرار دارد و بیماری حاصل از این جهش به صورت وابسته به جنسیت منتقل می‌شود. اگر یک دختر و یک پسر این جهش را از یکی از والدینشان به ارث ببرند، توضیح دهید که هر کدام از آن‌ها در چه وضعیتی خواهند بود.

دختر XX
پسر XY
حل.

الف) جهش در یک ژن مسئول تولید پروتئین خاص در سلول‌های عضلانی می‌تواند منجر به تغییر ساختار یا عملکرد آن پروتئین شود. این پروتئین ممکن است نقش حیاتی در فرایندهای بازسازی و ترمیم عضلات پس از آسیب داشته باشد، مانند تنظیم فعالیت سلول‌های بنیادی عضلانی (satellite cells)، هماهنگی سیگنال‌های بین‌سلولی، یا فراهم کردن ساختار لازم برای تشکیل فیبرهای عضلانی جدید. اگر پروتئین دچار جهش شود و عملکردش مختل گردد، این فرایندهای بازسازی نمی‌توانند به درستی انجام شوند. به عنوان مثال، سلول‌های بنیادی ممکن است نتوانند به درستی فعال شده، تمایز یابند یا فیوز شوند تا بافت عضلانی آسیب‌دیده را جایگزین کنند. در نتیجه، توانایی عضلات برای بازسازی و ترمیم پس از آسیب کاهش یافته یا کاملاً متوقف می‌شود.

ب) از آنجایی که دختر دارای دو کروموزوم X است، اگر یک نسخه جهش‌یافته و یک نسخه سالم از ژن را داشته باشد، نسخه سالم ژن می‌تواند نقش محافظتی ایفا کند و اجازه دهد که ترمیم عضلات تا حدی به صورت طبیعی انجام شود. بنابراین، دختر ممکن است علائم بسیار خفیفی را تجربه کند یا هیچ علائمی نداشته باشد. از آنجا که پسران تنها یک کروموزوم X دارند، اگر ژن جهش‌یافته را به ارث ببرند، نمی‌توانند از نسخه سالم دیگری برای جبران نقص استفاده کنند. بنابراین، این پسر احتمالاً علائم شدیدی از بیماری را نشان خواهد داد.

۲. (۱۵ نمره) در فرآیند تکثیر DNA، آنزیم‌های متعددی به طور هماهنگ کار می‌کنند تا تکثیر دقیق و کامل انجام شود. حال فرض کنید که جهشی در ژن کدکننده DNA Polymerase رخ داده که دقت این آنزیم را در جایگزینی نوکلئوتیدهای صحیح کاهش داده است. علاوه بر این، DNA Helicase نیز به دلیل جهشی دیگر قادر به باز کردن دو رشته DNA با سرعت معمول نیست.

الف) جهش در DNA Polymerase ممکن است منجر به افزایش تعداد خطاهای تکثیری شود. توضیح دهید که این خطاها چگونه می‌توانند بر توالی ژنی و در نهایت بر عملکرد پروتئین‌های تولیدی تأثیر بگذارند.

ب) اگر جهش در DNA Helicase باعث شود که این آنزیم نتواند به موقع دو رشته DNA را باز کند، توضیح دهید که چگونه این امر باعث ایجاد اختلال در تکثیر DNA و مشکلات ساختاری در رشته‌های جدید DNA می‌شود.

حل.

الف) DNA Polymerase مسئول اضافه کردن نوکلئوتیدهای صحیح به رشته جدید در فرآیند تکثیر DNA است. اگر جهشی در این آنزیم رخ دهد و دقت آن کاهش یابد، احتمال افزایش خطاهای تکثیری وجود دارد. این خطاها می‌توانند منجر به جایگزینی نوکلئوتیدهای نادرست در توالی ژنی شوند که به جهش‌های نقطه‌ای منجر می‌شود. اگر این جهش‌ها در نواحی کدکننده ژن‌های مهم رخ دهند، ممکن است توالی‌های پروتئینی به صورت اشتباه سنتز شوند. این امر می‌تواند منجر به تولید پروتئین‌های معیوب با عملکرد نادرست شود که در نهایت بر روی عملکرد سلول و فرایندهای حیاتی آن تأثیر منفی می‌گذارد. در برخی موارد، جهش‌ها می‌توانند باعث تولید پروتئین‌های ناپایدار یا غیرقابل عملکرد شوند که به بیماری‌های ژنتیکی یا سرطان منجر می‌شود.

ب) DNA Helicase وظیفه باز کردن دو رشته DNA را دارد تا تکثیر بتواند به درستی آغاز شود. اگر این آنزیم به دلیل جهش نتواند به موقع یا به درستی عمل کند، سرعت باز شدن دو رشته DNA کاهش یافته یا به طور نامنظم انجام می‌شود. این امر باعث ایجاد تأخیر در تکثیر DNA خواهد شد و ممکن است بخش‌های بزرگی از DNA ناتمام بمانند. در نتیجه، ساختار نهایی DNA ناقص و ناپیوسته خواهد بود و ممکن است مشکلاتی در طول کروموزوم‌ها ایجاد شود که می‌تواند در چرخه سلولی مشکل‌ساز باشد. چنین نقص‌هایی ممکن است منجر به متوقف شدن تقسیم سلولی، مرگ سلولی یا تولید سلول‌هایی با مواد ژنتیکی ناقص شود که در نهایت به بروز ناهنجاری‌های ژنتیکی و اختلالات سلولی منجر می‌شود.

۳. (۱۰ نمره) یک زیست‌شناس گونه‌ای از باکتری‌ها را کشف کرده که در محیط‌های با دمای بسیار بالا مانند چشمه‌های آب‌گرم زندگی می‌کنند. این باکتری‌ها توانایی تولید آنزیم‌هایی دارند که در دماهای بالا نیز فعالیت خود را حفظ می‌کنند. او به دنبال شناسایی ژن‌های مسئول تولید این آنزیم‌های مقاوم به حرارت است و از روش RT-PCR برای بررسی بیان ژن‌ها استفاده می‌کند. به نظر شما چه چالش‌هایی ممکن است در فرآیند جداسازی پروتئین‌های این باکتری وجود داشته باشد و چگونه می‌توان از پایداری آنزیم‌ها در طی مراحل جداسازی و خالص‌سازی اطمینان حاصل کرد؟

حل.

در فرآیند جداسازی پروتئین‌های باکتری‌های مقاوم به حرارت، چندین چالش عمده ممکن است پیش آید. یکی از این چالش‌ها، احتمال تغییر شکل یا دناتوره شدن پروتئین‌ها در اثر دماهای بالا است که می‌تواند فعالیت آنزیم‌ها را کاهش یا کاملاً متوقف کند. علاوه بر این، پروتئین‌های مقاوم به حرارت ممکن است تمایل به تشکیل تجمعات داشته باشند یا دارای ویژگی‌های حل‌شدگی متفاوتی باشند که جداسازی آن‌ها را دشوار می‌کند. حضور آنزیم‌های تجزیه‌کننده مانند پروتئازها نیز می‌تواند منجر به تخریب پروتئین‌های هدف شود. برای اطمینان

از پایداری آنزیم‌ها در طی مراحل جداسازی و خالص‌سازی، می‌توان از بافرهای مناسب حاوی عوامل تثبیت‌کننده مانند گلیسرول یا نمک‌ها استفاده کرد تا ساختار پروتئین حفظ شود. همچنین افزودن مهارکننده‌های پروتئاز به نمونه‌ها می‌تواند از تخریب پروتئین‌ها جلوگیری کند. انجام مراحل فرآیند در دماهایی که حداقل تغییرات در ساختار پروتئین ایجاد می‌کنند و استفاده از روش‌های خالص‌سازی بدون استفاده از مواد دناتوره‌کننده نیز از دیگر راهکارهای مؤثر برای حفظ فعالیت آنزیم‌ها می‌باشد. به علاوه، بهینه‌سازی شرایط کروماتوگرافی و سایر تکنیک‌های جداسازی به گونه‌ای که محیطی ملایم و سازگار با پایداری پروتئین‌ها فراهم شود، نقش مهمی در حفظ فعالیت آنزیم‌ها ایفا می‌کند.

۴. (۱۵ نمره) شما دو توالی DNA از دو گونه مختلف دارید.

$x = \text{AGCTGAC}$
 $y = \text{AGCAGTC}$

الف) Hamming Distance را برای این دو توالی محاسبه کنید.
 ب) Edit Distance را برای این دو توالی محاسبه کنید. توضیح دهید که چگونه این فاصله می‌تواند اطلاعات بیشتری نسبت به Hamming Distance درباره تغییرات ژنتیکی بین توالی‌ها ارائه دهد.

حل.

(آ) **فاصله‌ی همینگ:** فاصله‌ی همینگ نشان‌دهنده‌ی تعداد جایگاه‌های متفاوت بین دو توالی با طول برابر است. به عنوان مثال:

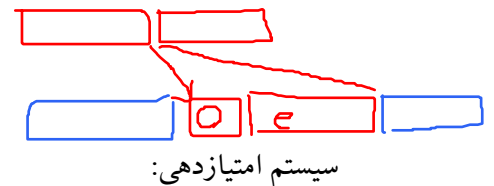
$x = \text{AGCTGAC}, y = \text{AGCAGTC}$

فاصله‌ی همینگ بین x و y برابر با ۲ است، زیرا دو کاراکتر متفاوت دارند.

(ب) **فاصله‌ی ویرایش Distance Levenshtein:** حداقل تعداد عملیات (درج، حذف، جایگزینی) برای تبدیل یک توالی به توالی دیگر است. فاصله‌ی ویرایش برای توالی‌های داده شده نیز برابر با ۲ است.

۵. (۱۵ نمره) در بیولوژی و توالی‌های ژنتیکی، فاصله‌ها (gaps) به بخش‌هایی از توالی اشاره دارند که در یک توالی وجود دارند اما در توالی دیگر حذف شده‌اند یا وجود ندارند. این فاصله‌ها به دلیل تغییرات طبیعی مانند حذف (deletion)، درج (insertion) یا جهش (mutation) در فرآیندهای تکاملی ایجاد می‌شوند. در هم‌ترازی توالی‌ها، دو نوع جریمه به نام‌های open gap و extension gap برای کنترل فاصله‌ها استفاده می‌شود. جریمه open gap بالاتر است زیرا ایجاد یک فاصله جدید تغییر چشمگیری در ساختار توالی به وجود می‌آورد و اغلب به تغییرات تکاملی مهم‌تری مانند حذف یا درج اشاره دارد. از سوی دیگر، extension gap جریمه کمتری دارد چون صرفاً به گسترش یک فاصله موجود می‌پردازد و تغییری جدید ایجاد نمی‌کند. در نتیجه، برای جلوگیری از ایجاد فاصله‌های زیاد و بی‌رویه، جریمه open gap بیشتر است تا هم‌ترازی دقیق‌تری بین توالی‌ها برقرار شود. با استفاده از سیستم امتیازدهی داده شده، دو توالی زیر را هم‌تراز کنید.

$x = \text{GCAC}$
 $y = \text{GCC}$



match: ۲

mismatch: -۲

gap open: -۵

gap extend: -۱

matrix

حل. برای حل این سوال از ۳ ماتریس استفاده میکنیم: ماتریس (F) نشان میدهد که در کاراکتر آخر با هم تراز شده باشند. ماتریس Ix آخرین کاراکتر x با گپ تراز شده باشد ماتریس Iy آخرین کاراکتر y با گپ تراز شده باشد ماتریس ها از روی هم به شکل زیر پر میشوند:

$$F(i, j) = \sigma(x_i, y_j) + \max \begin{cases} F(i-1, j-1) & \leftarrow \text{Continuing alignment} \\ I_x(i-1, j-1) & \leftarrow \text{Closing gaps in y} \\ I_y(i-1, j-1) & \leftarrow \text{Closing gaps in x} \end{cases}$$

$$I_y(i, j) = \max \begin{cases} F(i, j-1) + d & \leftarrow \text{Opening a gap in x} \\ I_y(i, j-1) + e & \leftarrow \text{Gap extension in x} \end{cases}$$

$$I_x(i, j) = \max \begin{cases} F(i-1, j) + d & \leftarrow \text{Opening a gap in y} \\ I_x(i-1, j) + e & \leftarrow \text{Gap extension in y} \end{cases}$$

	y = G C C			
x =	0	-∞	-∞	-∞
G	-∞	2		
C	-∞			
A	-∞			
C	-∞			

F

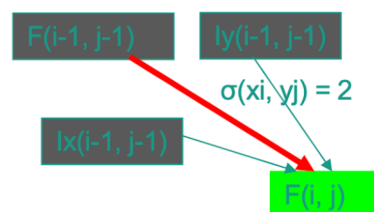
	y = G C C			
x =	-∞	-∞	-∞	-∞
G	-5			
C	-6			
A	-7			
C	-8			

Ix

m = 2
s = -2
d = -5
e = -1

	y = G C C			
x =	-∞	-5	-6	-7
G	-∞			
C	-∞			
A	-∞			
C	-∞			

Iy



حالا با ادامه این مراحل داریم:

	y = G C C			
x =	0	-∞	-∞	-∞
G	-∞	2	-7	-8
C	-∞	-7	4	-1
A	-∞	-8	-5	2
C	-∞	-9	-6	1

F

	y = G C C			
x =		-5	-6	-7
G	-∞	-∞	-3	-4
C	-∞	-∞	-12	-1
A	-∞	-∞	-13	-10
C	-∞	-∞	-14	-11

l₁

	y = G C C			
x =		-∞	-∞	-∞
G	-5	-∞	-∞	-∞
C	-6	-3	-12	-13
A	-7	-8	-1	-6
C	-8	-13	-2	-3

l_x

	y = G C C			
x =				
G				
C				
A				
C				

m = 2
s = -2
d = -5
e = -1

x GCA
C
y |||
GC-C

در نتیجه همترازی نهایی به دست می آید:

GCAC

GC-C