

# مقدمه‌ای بر بیوانفورماتیک

نیم سال دوم ۹۷-۹۸



زمان: ۷۵ دقیقه

دانشکده‌ی مهندسی کامپیوتر

۲۳ اردیبهشت ۱۳۹۸

## راهنمایی حل آزمون میان‌ترم

### مسئله‌ی ۱. زیست زیباست

هریک از مفاهیم زیر را در یک سطر توضیح دهید:

- رونویسی (Transcription): فرایند تولید mRNA از یک ژن که توسط RNA Polymerase انجام می‌گیرد. این فرایند مقدمه‌ای برای تولید پروتئین از ژن است.
- تنظیم بیان ژن: مجموعه‌ی سازوکارهایی که باعث افزایش یا کاهش بیان یک ژن (نرخ رونویسی از آن) می‌شود، نظیر اتصال فاکتورهای رونویسی یا تغییرات اپی‌ژنتیک.
- اپی‌ژنتیک (Epigenetics): تغییراتی که توالی DNA را تغییر نمی‌دهند اما با اتصال مولکول‌های کوچکی مثل متیل یا استیل به خود DNA یا هیستون‌ها، باعث تفاوت بیان ژن‌ها در نوع سلول‌های مختلف می‌شوند.
- هلیکاز (Helicase): آنزیم بازکننده‌ی مارپیچ دورشته‌ی DNA در هنگام تکثیر آن و تقسیم سلول.

### مسئله‌ی ۲. پلاناریا

الف) کرم پلاناریا دو گونه‌ی متفاوت دارد که دو سیستم تولید مثل کاملاً متفاوت دارند

- گونه‌ی اول با تولید مثل جنسی، که در آن یک کرم جدید از سلول تخم حاصل از لقاح یک سلول جنسی نر و یک سلول جنسی ماده به وجود می‌آید.
- گونه‌ی دوم با تولید مثل غیرجنسی، که در آن یک کرم از وسط نصف می‌شود و دو کرم جدید به وجود می‌آیند.

به نظر شما سلول‌های یک کرم از گونه‌ی اول ژنوم شبیه‌تری دارند، یا سلول‌های یک کرم از ژنوم دوم؟ علت چیست؟

سلول‌های یک کرم از گونه‌ی اول، چون همه‌ی آن‌ها از یک سلول به وجود آمده‌اند و در نتیجه طی تقسیم‌های سلولی، دچار خطای کپی‌برداری کم‌تری شده‌اند. هر کرم از گونه‌ی دوم در حین تقسیم هزاران سلول را به هریک از دو فرزند خود منتقل می‌کند که ژنوم آن‌ها به خاطر تقسیم‌های متعدد سلولی که در کرم والد و اجداد آن انجام شده از یک‌دیگر تفاوت قابل ملاحظه‌ای پیدا کرده‌اند.

ب) در هریک از این کرم‌ها، سلول‌های مختلفی مثل عصبی، گوارشی، بینایی و غیره وجود دارند. به نظر شما چطور می‌توانند این سلول‌ها از یک‌دیگر متمایز شوند؟

سلول‌های مختلف یک حیوان در عین داشتن ژنوم واحد، اپی‌ژنتیک و پروفایل بیان ژنی و مجموعه‌ی فاکتورهای رونویسی متفاوتی دارند. همین موضوع می‌تواند تنظیم بیان ژن‌ها را در هر نوع سلول از نوع سلول دیگر متفاوت کند و باعث تفاوت پروفایل پروتئینی بین سلول‌ها و در نتیجه تفاوت شکل و عملکرد آن‌ها شود. به این موضوع می‌توانید تمام سازوکاری که وادینگتون در منظره‌ی اپی‌ژنتیک توضیح داده را نیز اضافه کرد.

### مسئله ۳. خطای کپی برداری

در هنگام تکثیر ژنوم یک باکتری، تقریباً یک باز در هر ۱۰۰۰ باز اشتباه کپی می‌شود. نرخ این خطا در ژنوم انسان تقریباً ۱ باز در هر صد میلیون باز است.

• چه تفاوتی بین سلول‌های انسان و باکتری باعث این اختلاف چشم‌گیر می‌شود؟  
 ساختار DNA Polymerase در انسان، برخلاف باکتری‌ها، دارای قابلیت تصحیح خطا است. **proofreading**

• چگونه این دو نرخ خطای متفاوت به بقای باکتری و انسان کمک کرده است؟

باکتری‌ها خیلی بیشتر از انسان تکثیر می‌شوند. بالا رفتن نرخ خطا منجر به افزایش جهش در آن‌ها می‌شود که این موضوع می‌تواند به از بین رفتن برخی باکتری‌ها به خاطر جهش مخرب در ژن‌های حیاتی بیانجامد. در عین حال باکتری‌هایی که زنده می‌مانند از تنوع بیشتری برخوردار بوده و شانس آن‌ها که زیرمجموعه‌ای از آن‌ها در مقابل تغییرات محیط، مثل بروز یک شرایط سخت یا مواجه شدن با آنتی بیوتیک، مقاوم شود افزایش می‌یابد.

در انسان اگرچه یک نرخ پایین‌تر جهش باعث بروز تفاوت‌های بین انسان‌های مختلف می‌شود، اما نرخ زیاد جهش می‌توانست پروتئین‌های حیاتی هر انسان را دستخوش تغییرات مخربی کند که با توجه به نرخ کمتر تولید مثل در انسان اثر مخربی دارد.

### مسئله ۴. برهم‌نهی

ماتریس جریمه‌ی تبدیل هریک از آمینواسیدهای زیر به یکدیگر را در نظر بگیرید (این ماتریس بخشی از PAM250 است):

	G	I	L	M	F	V
G	5	-3	-4	-3	-5	-1
I	-3	5	2	2	1	4
L	-4	2	6	4	2	2
M	-3	2	4	6	0	2
F	-5	1	2	0	9	-1
V	-1	4	2	2	-1	4

دو رشته‌ی MIGV و FFIGL را با روش برنامه‌سازی پویا (Dynamic Programming) به یکدیگر برهم‌نهی (Alignment) کرده‌ایم. حاصل جدول ذیل شده است:

	-	M	I	G	V
-	0	-2	-4	-6	-8
F	-2	0	-1	-3	-5
F	-4	-2	1	-1	-3
I	-6	-2	3	1	3
G	-8	-4	1	A	B
L	-10	-4	-1	C	D

• این برهم‌نهی محلی بوده یا سراسری؟ جریمه‌ی فاصله (Gap) چقدر است؟

سراسری، چرا که هزینه‌ی فاصله در سطر و ستون اول صفر نیست. با نگاه به مقادیر سطر و ستون اول مشخص می‌شود جریمه‌ی هر فاصله ۲- است.

- مقدار خانه‌های خالی را مشخص کنید. کافی است بنویسید هریک از مقادیر  $A, B, C$  و  $D$  چقدر است.

$$A = ۸, B = ۶, C = ۶, D = ۱۰$$

- برهم‌نهی این دورشته را بنویسید. بدین منظور، دو رشته را در دو سطر منطبق بر هم بنویسید به طوری که حرف‌های منطبق بر یکدیگر روبروی هم قرار بگیرند. فاصله‌ها را با خط تیره مشخص کنید.

\_MIGV  
FFIGL

یا

M\_IGV  
FFIGL

- اگر الگوریتم برهم‌نهی سراسری را به محلی تغییر دهید، دقیقاً چه تغییراتی در الگوریتم ایجاد می‌شود؟ چه زمان بهتر است از الگوریتم برهم‌نهی سراسری استفاده کنیم و چه زمان از محلی؟

مقادیر سطر و ستون اول صفر (فاصله‌های شروع هر رشته) می‌شود. همچنین به جای آن که جواب مساله مقدار نوشته‌شده در خانه‌ی پایین سمت راست جدول باشد، بسته به نوع مساله بیشینه مقدار سطر آخر، ستون آخر، یا کل جدول را به عنوان جواب برمی‌گردانیم. مثلاً اگر به دنبال یافتن یک ناحیه مشترک بین دو رشته باشیم، یا به دنبال یافتن شبیه‌ترین ناحیه‌ی یک ژنوم با یک ژن باشیم از برهم‌نهی محلی استفاده می‌کنیم. اگر به تطابق کل دو رشته باشیم، مثلاً مجموعه‌ی جهش‌ها و حذف و اضافه‌هایی که یک ژن در موجود اول را به همان ژن در موجود دوم تبدیل کرده، از برهم‌نهی سراسری روی آن دو ژن استفاده می‌کنیم.

## مسئله‌ی ۵. چکمه

یک رشته DNA را با الگوریتم BWT به رشته‌ی CAGATTCA\$AT تبدیل کرده‌ایم.

- رشته‌ی اصلی را بیابید. فرض کنید \$ در انتهای رشته قرار می‌گیرد و از همه‌ی حروف بزرگ‌تر است. این رشته برابر با TCGATCAAAT\$ است. کافی است ابتدا تمام حروف را مرتب و در یک ستون بنویسیم و بعداً رشته‌ی اولیه را به عنوان ستون سمت چپ آن اضافه کنیم و مجدداً همه‌ی سطرها را مرتب و دوباره رشته‌ی اولیه را به عنوان ستون سمت چپ اضافه کنیم و این کار را به اندازه‌ی طول این رشته تکرار کنیم تا یک جدول  $n \times n$  به دست آید. رشته‌ی مورد نظر در سطری است که در انتهای آن علامت \$ قرار دارد.

- توضیح دهید از تبدیل BWT چگونه می‌توانیم در برهم‌نهی توالی‌ها استفاده کنیم.

ابتدا به ازای هریک از چهار نوکلئوتید، یک بردار می‌سازیم که مقدار  $i$  ام آن برابر تعداد دفعات وقوع آن نوکلئوتید در  $i$  حرف ابتدایی رشته‌ی BWT است. سپس دو اشاره‌گر  $p_1$  و  $p_2$  به سطر اول و آخر جدول می‌گیریم که بازه‌ی جستجو را نشان می‌دهند. با یک حلقه از آخرین به اولین حرف توالی جستار (query) هر بار مقدار این دو اشاره‌گر را به روز می‌کنیم: اگر حرف  $i$  ام از رشته‌ی جستار کاراکتر  $C$  باشد، ابتدا تعداد دفعات وقوع  $C$  در بازه‌ی  $p_1$  تا  $p_2$  را در متغیر  $w$  می‌ریزیم. سپس  $p_1$  پس از این عمل برابر موقعیت شروع حرف  $C$  در ستون اول جدول BWT (معادل تعداد حرف‌های کوچکتر از  $C$  در کل رشته‌ی رفرنس + ۱) به اضافه‌ی تعداد دفعات وقوع  $C$  در حرف‌های اول تا  $p_1 - ۱$  و مقدار  $p_2$  بعد از این عمل برابر با مقدار جدید  $p_1$  به اضافه  $w$  خواهد بود.