به نام زیباترین

مقدمه‌ای بر بیوانفورماتیک

پاییز ۱۴۰۳

علی شریفی زارچی

دانشگاه صنعتی شریف

دانشکده‌ی مهندسی کامپیوتر

۲۴ دی ۱۴۰۳	آزمون پایان‌ترم	زمان: ۱۵۰ دقیقه

لطفا پاسخ هر سوال را در یک برگه‌ی جداگانه بنویسید. بالای هر برگه نام و شماره دانشجویی خود را حتما قید کنید. هر سوال ۲۵ نمره دارد. یک سوال نمره‌ی اضافه است.

سوال ۱:

(الف) در یک باکتری از هر ژن یک پروتئین و در هر گیاه از هر ژن چند نوع پروتئین تولید می‌شود. مکانیزم آن را توضیح دهید.

(ب) کدام جهش اثر مخرب‌تری روی ساختار پروتئین تولیدشده از یک ژن دارد؟ حذف یک حرف از ژن یا حذف سه حرف متوالی از ژن.

(ج) چه چیزی باعث می‌شود یک سلول پوست ما به سادگی نتواند به یک سلول قلب تبدیل شود؟

(د) مزیت توالی‌یابی تک‌مولکولی نسبت به سایر روش‌های توالی‌یابی چیست؟

سوال ۲:

همان‌طور که می‌دانید مساله‌ی یافتن ساختار RNA را می‌توان به صورت پرانتزگذاری متناظر کرد. مثلا برای توالی زیر، پرانتزگذاری داده‌شده مشخص‌کننده‌ی ساختار بهینه است.

```
ACAGGCUAGCUGU
(((.((()))))
```

حال برای هریک از صورت مساله‌های زیر الگوریتمی برای یافتن ساختار بهینه ارائه کنید.

(الف) چون زوج CG سه پیوند هیدروژنی و زوج AU دو پیوند هیدروژنی تشکیل می‌دهند، ساختاری را بهینه در نظر بگیرید که بیشترین تعداد پیوندهای هیدروژنی را داشته باشد.

(ب) اگر در مثال بالا دقت کنید حرف شماره ۷ (U) با حرف شماره ۸ (A) پیوند هیدروژنی داده‌اند. این بدین معنا است که ساختار RNA در یک نوکلئوتید چرخش ۱۸۰ درجه‌ای کرده‌است، که در واقعیت ممکن نیست. فرض کنید در ساختار بهینه در هر چرخش حداقل سه نوکلئوتید آزاد وجود دارند. به عبارت دیگر بین هر پرانتز باز و پرانتز بسته‌ی متناظر آن حداقل سه نقطه قرار می‌گیرد. ساختاری بهینه است که بیشترین تعداد مجموع پرانتز باز و بسته را داشته باشد.

سوال ۳:

(الف) فرض کنید بخواهیم تمامی زیررشته‌های kحرفی یک توالی n حرفی را hash کنیم. این کار در ساده‌ترین پیاده‌سازی به زمان اجرای $O(kn)$ نیاز دارد. الگوریتم و شیوه‌ی hash کردن را به گونه‌ای بهبود دهید که زمان اجرا خطی $O(n)$ شود.

(ب) ثابت کنید احتمال یکسان بودن کمینه‌ی hash کل زیرتوالی‌های k حرفی دو متن، برابر شباهت Jaccard آن‌ها است.

سوال ۴:

(الف) یک شبکه‌ی ژنی رسم کنید که دو وضعیت پایدار [۱] داشته باشد. معادلات دیفرانسیل مربوط به آن را بنویسید.

(ب) یک شبکه‌ی ژنی با سه نقطه‌ی پایدار رسم کنید و معادلات دیفرانسیل آن را بنویسید.

(ج) اگر یک گونه جانور ۱۶ نوع سلول مختلف داشته باشد آیا می‌توانید یک شبکه‌ی ژنی برای آن رسم کنید که تولید این ۱۶ نوع سلول پس از دوران جنینی را توضیح دهد؟ تعداد ژن‌ها دلخواه است و نیازی به نوشتن معادلات دیفرانسیل نیست.

سوال ۵:

دانشجویان درس مقدمه‌ای بر بیوانفورماتیک در حل مسئله‌ی alignment مشکل دارند و هر کدام برای پیدا کردن هم‌ترازی میان دو رشته جدول مقادیر برنامه‌نویسی پویا را با روشی متفاوت پر کرده‌اند. در هر قسمت توضیح دهید که دانشجوی گمراه در واقع چه مسئله‌ی هم‌ترازی ای را حل کرده است.

(الف) مینا مقدار را سطر و ستون اول را برابر ۰ قرار داده، اما به سایر خانه‌های جدول اجازه داده مقدار منفی به خود بگیرند. او در نهایت عدد بیشینه‌ی جدول را به عنوان بیش‌ترین امتیاز ممکن گزارش داده است.

(ب) کیان جدول local alignment را به درستی پر کرده اما امتیازی که گزارش داده بیش‌ترین مقدار ستون آخر می‌باشد.

(ج) سپنتا جدول global alignment را پر کرده ولی بیش‌ترین عدد جدول را به عنوان جواب گزارش می‌دهد.

---

[۱] state steady stable