



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

مقدمه‌ای بر بیوانفورماتیک

پاییز ۱۴۰۳

استاد: علی شریفی زارچی

مسئول تمرین: هانیه ملائی

مهلت ارسال نهایی: ۸ بهمن

تمرین چهارم

مهلت ارسال بدون تاخیر: ۵ بهمن

- 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 کوئرا 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 `HW۴_[STD_ID].pdf` آپلود کنید.
- گردآوردندگان تمرین: هانیه ملائی، ریحانه ابراهیمی، زهرا عسگری، رضا دولتی

سوالات (۱۰۰ نمره)

725

۱. (۲۵ نمره)

- با استفاده از دو رقم آخر شماره دانشجویی خود (XX)، یک پروتئین با شناسه PDB که به این رقم ختم می‌شود، پیدا کنید. اگر چنین پروتئینی پیدا نکردید، مقدار XX را یک واحد افزایش یا کاهش دهید.
- مشاهده سه‌بعدی پروتئین‌ها به شما کمک می‌کند تا ساختارهای عملکردی مانند سایت‌های فعال و لیگاندها را شناسایی کنید. فایل pdb. مربوط به پروتئین انتخابی خود را از پایگاه داده دانلود کنید و از ابزارهایی مانند Chimera، PyMOL یا ... برای مشاهده ساختار سه‌بعدی پروتئین استفاده کنید.
- یک تصویر از نمای کلی ساختار پروتئین (surface یا cartoon) ارائه دهید.
 - یک تصویر حاشیه‌نویسی شده از یک ناحیه خاص مانند سایت فعال یا لیگاند متصل در صورت وجود اضافه کنید.
- ساختارهای PDB همیشه کامل نیستند. ممکن است بخش‌هایی از توالی حذف شده باشند یا برچسب‌های اضافی به آن‌ها اضافه شده باشد. با نوشتن کد، توالی پروتئین از فایل PDB را با توالی UniProt مقایسه کنید. اختلافات (مانند residueهای گم‌شده یا tagهای اضافی) را شناسایی و تحلیل کنید. همچنین به صورت کلی توضیح دهید این اختلافات به چه دلایلی ممکن است ایجاد شده باشند (مانند بی‌نظمی ساختاری یا مهندسی آزمایشگاهی).
- ویژگی‌های ساختار ثانویه نشان‌دهنده نحوه پایداری و عملکرد پروتئین شما را با ارتباط بین ساختار و عملکرد پروتئین آشنا می‌کند. با استفاده از ابزارهایی مانند Stride ساختارهای ثانویه مانند تعداد و محل helices- α و sheets- β را شناسایی کنید.
- در نهایت گزارشی از این روند ارائه دهید.

۲. (۲۵ نمره)

پیش‌بینی ساختار پروتئین از دنباله‌های آمینواسید یکی از چالش‌های بزرگ در زیست‌شناسی ساختاری است. مدل AlphaFold برای پیش‌بینی ساختار پروتئین‌ها با دقت بسیار بالا شناخته شده است. در این سوال، شما باید از مدل AlphaFold برای پیش‌بینی ساختار پروتئین دلخواهتان استفاده کنید و نتایج آن را با ساختار واقعی پروتئین مقایسه کنید.

مراحل انجام کار:

- ابتدا یک پروتئین از پایگاه داده UniProt یا PDB انتخاب کنید. اطمینان حاصل کنید که این پروتئین دارای ساختار سه‌بعدی مشخص در قالب فایل‌های PDB می‌باشد.
- دنباله پروتئینی مربوطه را استخراج کنید. این دنباله باید به فرمت FASTA یا مشابه آن درآید. شما می‌توانید از UniProt برای استخراج دنباله‌های پروتئینی استفاده کنید.
- پس از آماده‌سازی داده‌ها، به کمک API مدل AlphaFold برای پیش‌بینی ساختار پروتئین اجرا کنید.
- پس از مدل AlphaFold شما باید ساختار پیش‌بینی‌شده را با ساختار واقعی پروتئین که از منابعی مانند PDB دریافت کرده‌اید مقایسه کنید. برای مقایسه می‌توانید از معیارهای استاندارد مانند Mean Root Deviation Square (RMSD) و Test Distance Global (GDT-TS) استفاده کنید. این معیارها تفاوت بین ساختار واقعی را اندازه‌گیری می‌کنند و نشان می‌دهند که مدل چقدر دقیق عمل کرده است.
- نتایج را بررسی کرده و تحلیل کنید. مقایسه‌ای بین ساختار پیش‌بینی‌شده و ساختار واقعی انجام دهید. به دقت مدل در نواحی خاص ساختار توجه کنید.

در نهایت گزارشی از این روند ارائه دهید.

۳. (۲۵ نمره) به سوالات زیر پاسخ دهید:

- (آ) توضیح دهید چگونه خطا در جفت شدن آنتی‌کدون tRNA و کدون mRNA می‌تواند بر Reading Frames تأثیر بگذارد و چه پیامدهای احتمالی برای ساختار نهایی پروتئین دارد؟ آیا سیستم ترجمه مکانیسمی برای شناسایی و اصلاح چنین خطاهایی دارد؟
- (ب) اگر یک tRNA مصنوعی طراحی شود که بتواند به بیش از یک نوع آمینواسید متصل شود، این tRNA چگونه می‌تواند فرآیند ترجمه را تحت تأثیر قرار دهد؟ آیا سلول می‌تواند چنین چالش‌هایی را تحمل کند و یا مکانیسمی برای جلوگیری از آن وجود دارد؟
- (ج) آزمایشی را فرض کنید که در آن یک پروتئین با استفاده از یک توالی RNA که فاقد کدون شروع است، ترجمه شده است. آیا این پروتئین قابل تولید و عملکرد (قابل استفاده) خواهد بود؟ آیا غیبت کدون شروع می‌تواند بر کل فرآیند ترجمه تأثیر بگذارد؟
- (د) در یوکاریوت‌ها، فرآیند ترجمه به دلیل وجود ساختارهای پیچیده‌ای مانند اسپلایسینگ (Splicing) و پلی‌آدنیلایسون وابسته به زمان و مکان خاصی است. اگر یک RNA پیام‌رسان با توالی اسپلایس نشده به طور مستقیم وارد سیتوپلاسم شود و توسط ریبوزوم ترجمه شود، چه پیامدهایی بر محصول پروتئینی نهایی و تنظیم ژنتیکی خواهد داشت؟ آیا چنین پدیده‌ای در شرایط خاص زیستی امکان‌پذیر است؟

۴. (۲۵ نمره) فایلی در قالب notebook jupyter در اختیار شما قرار گرفته است که بایست تمام بخش‌های خواسته شده آن را به صورت کامل و بدون خطا اجرا نمایید و در کنار بقیه ی موارد آپلود کنید.