



## پروژه درس بیوانفورماتیک.

عنوان پروژه: شناسایی ژن های آنزیم زایلاناز از متازنوم شکمبه نشخوارکنندگان

### اهمیت اکتشاف آنزیم های جدید از متازنوم

اکتشاف آنزیم های جدید از متازنومها، دریچه ای به سوی گنجینه ای عظیم از تنوع زیستی میکروبی و پتانسیل های بالقوه آنها می گشاید. متازنوم، به عنوان مخزن ژنتیکی میکروارگانیسم های محیط های مختلف، حاوی اطلاعات ژنتیکی بی شماری هستند که می توانند منجر به کشف آنزیم های جدید با ویژگی های منحصر به فرد شوند. این آنزیم ها می توانند در صنایع مختلف، از جمله صنایع غذایی، دارویی، کشاورزی و بیوتکنولوژی، کاربردهای فراوانی داشته باشند. به عنوان مثال، آنزیم های تجزیه کننده پلیمرهای پیچیده مانند سلولز و همی سلولز، می توانند در تولید سوخت های زیستی و مواد شیمیایی تجدید پذیر نقش مهمی ایفا کنند. همچنین، آنزیم های مقاوم به دما و pH بالا، می توانند در فرآیندهای صنعتی سخت و پرهزینه، کارایی و صرفه جویی قابل توجهی را به ارمغان آورند.

### اهمیت آنزیم های Xylanase میکروبی

آنزیم های زایلاناز میکروبی، به عنوان درشت مولکول های قدرتمند در تجزیه زایلان (دومین پلی ساکارید فراوان در طبیعت)، نقش حیاتی در صنایع مختلف ایفا می کنند. زایلان، به عنوان یکی از اجزای اصلی دیواره سلولی گیاهان، پتانسیل بالایی در تولید سوخت های زیستی، مواد شیمیایی و خوراک دام دارد. آنزیم های زایلاناز، با شکستن زایلان به قندهای ساده تر، امکان استفاده بهینه از این منبع تجدید پذیر را فراهم می کنند. در صنعت خمیر و کاغذ، زایلانازها می توانند با کاهش مصرف مواد شیمیایی و انرژی، فرآیند سفیدسازی را بهبود بخشند. همچنین، در تولید خوراک دام، زایلانازها با افزایش قابلیت هضم خوراک، می توانند عملکرد و سلامت دام را بهبود بخشند. با توجه به اهمیت روزافزون منابع تجدید پذیر و توسعه پایدار، آنزیم های زایلاناز میکروبی به عنوان ابزارهای کلیدی در این مسیر، توجه روزافزونی را به خود جلب کرده اند.

### مقدمه و هدف:

این پروژه با هدف شناسایی ژن های کدکننده آنزیم زایلاناز از متازنوم شکمبه نشخوارکنندگان طراحی شده است. زایلانازها آنزیم های تجزیه کننده زایلان (دومین پلی ساکارید فراوان در زیست کره) هستند و در صنایع مختلف از جمله تولید سوخت های زیستی، خوراک دام و کاغذ کاربرد دارند. متازنوم شکمبه نشخوارکنندگان به عنوان منبعی غنی از تنوع میکروبی، پتانسیل بالایی برای کشف زایلانازهای جدید با ویژگی های منحصر به فرد را داراست.

### داده ها:

یک فایل حاوی توالی های کانتینگ اسمبل شده از متازنوم شکمبه نشخوارکنندگان در گوگل درایو به آدرس زیر قرار داده شده است:

<https://drive.google.com/file/d/14PGwsGuL2ouY-fv0yzzijGnBMSjREU6/view>



### گام اول: شناسایی توالی‌های بالقوه زایلاناز (۴۰ درصد از امتیاز کل)

در این گام، هدف شناسایی توالی‌هایی است که احتمالاً ژن‌های زایلاناز را کد می‌کنند. برای این منظور، از رویکرد مبتنی بر شباهت توالی استفاده خواهد شد. توالی‌های کانتینگ با پایگاه داده‌ای از زایلانازهای شناخته شده (مانند پایگاه داده (CAZy) مقایسه می‌شوند. اگر یک کانتینگ شباهت قابل توجهی با یک یا چند زایلاناز شناخته شده داشته باشد، به عنوان یک توالی بالقوه زایلاناز در نظر گرفته می‌شود. تعیین معیار شباهت بر عهده دانشجویان است.

#### ابزارها و روش‌ها:

برای انجام این گام، می‌توان از ابزارهای بیوانفورماتیکی مختلفی استفاده کرد. به عنوان مثال Blast+ که از طریق لینک زیر در دسترس است:

<https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html#downloadblastdata>

ابزار های دیگری هم به انتخاب دانشجویان قابل استفاده هستند.

مانند BLAST، DIAMOND یا HMMER استفاده کرد. این ابزارها امکان مقایسه سریع و کارآمد توالی‌ها را فراهم می‌کنند. همچنین، می‌توان از روش‌های مبتنی بر یادگیری ماشین برای بهبود دقت شناسایی توالی‌های زایلاناز استفاده کرد.

#### خروجی:

خروجی این گام، فهرستی از توالی‌های کانتینگ است که به عنوان توالی‌های بالقوه زایلاناز شناسایی شده‌اند. برای هر توالی لازم است ترجمه پروتئینی آن نیز به دست آید. این توالی‌ها می‌توانند در گام‌های بعدی برای بررسی‌های بیشتر و تایید عملکرد زایلانازی مورد استفاده قرار گیرند.

### گام دوم: خوشه‌بندی و انتخاب نماینده (۳۰ درصد از امتیاز کل)

پس از شناسایی توالی‌های بالقوه زایلاناز در گام اول، ممکن است تعدادی از این توالی‌ها بسیار شبیه به هم باشند. این شباهت بالا می‌تواند ناشی از افزونگی در داده‌های متاژنومی و یا وجود ایزوفرم‌های مختلف یک ژن باشد. برای کاهش افزونگی و ساده‌سازی تحلیل‌های بعدی، توالی‌های بسیار مشابه در خوشه‌هایی گروه‌بندی می‌شوند.

#### روش خوشه‌بندی:

در این پروژه، از الگوریتم‌های خوشه‌بندی توالی مانند CD-HIT استفاده می‌شود. CD-HIT یک ابزار سریع و کارآمد برای خوشه‌بندی توالی‌های DNA و پروتئین بر اساس شباهت آن‌ها است. این ابزار، توالی‌ها را بر اساس یک آستانه شباهت (در اینجا ۹۷٪) در خوشه‌ها گروه‌بندی می‌کند.

#### انتخاب نماینده:

پس از خوشه‌بندی، از هر خوشه یک توالی به عنوان نماینده انتخاب می‌شود. نماینده معمولاً طولانی‌ترین توالی در خوشه یا توالی با بیشترین شباهت به سایر اعضای خوشه است. انتخاب نماینده به این دلیل انجام می‌شود که تحلیل‌های بعدی را بر روی یک مجموعه کوچک‌تر و غیر افزونه از توالی‌ها متمرکز کنیم.



در گام سوم لازم است که یک مدل برای طولانی ترین ناحیه conserved یک زیر خانواده از زایلانازها ساخته شود. انتخاب این زیرخانواده بر عهده دانشجویان است. به عنوان مثال زایلانازهای ترموفیل، فعال در pH های اکستریم اسیدی یا بازی، زایلانازهای قارچی، ...

این مدل می تواند بر مبنای روش دلخواه دانشجویان ساخته شود. به عنوان مثال PSSM، مدل های مخفی مارکوف، عبارات منظم، ... و یا ترکیبی از آنها. این مدل باید برای فیلتر توالی ها استفاده شود تا به توالی های محدود تری برسیم.

Show drafts

### گام سوم: مدل سازی ناحیه حفاظت شده و فیلتر کردن توالی ها (۳۰ درصد از امتیاز کل)

در این گام، هدف ساخت یک مدل برای ناحیه حفاظت شده (conserved region) طولانی ترین توالی یک زیرخانواده خاص از زایلانازها است. این مدل سپس برای فیلتر کردن توالی های نماینده خوشه ها استفاده می شود تا به مجموعه ای محدودتر از توالی های کاندید زایلاناز برسیم.

#### انتخاب زیرخانواده:

دانشجویان می توانند یک زیرخانواده خاص از زایلانازها را بر اساس ویژگی های موردعلاقه خود انتخاب کنند. برخی از گزینه های ممکن عبارتند از:

- زایلانازهای ترموفیل (مقاوم به دماهای بالا)
- زایلانازهای فعال در pH های اسیدی یا بازی
- زایلانازهای قارچی
- زایلانازهای باکتریایی
- زایلانازهای متعلق به یک خانواده گلیکوزیل هیدرولاز خاص (مثلاً GH10 یا GH11)

#### روش های مدل سازی:

برای مدل سازی ناحیه حفاظت شده، می توان از روش های مختلفی استفاده کرد، از جمله:

- ماتریس امتیازدهی موقعیت-ویژه (PSSM): PSSM یک ماتریس است که احتمال حضور هر اسید آمینه را در هر موقعیت از یک ناحیه حفاظت شده نشان می دهد. این ماتریس ها معمولاً از طریق هم تراز کردن چندگانه توالی های مرتبط ساخته می شوند.
- مدل های مخفی مارکوف (HMM): HMM ها مدل های احتمالاتی هستند که می توانند برای نمایش نواحی حفاظت شده در توالی های پروتئینی استفاده شوند. HMM ها نسبت به PSSM ها پیچیده تر هستند، اما می توانند اطلاعات بیشتری را در مورد ساختار و عملکرد ناحیه حفاظت شده ارائه دهند.
- عبارات منظم (Regular Expressions): عبارات منظم الگوهایی هستند که می توانند برای توصیف توالی های زیستی استفاده شوند. عبارات منظم می توانند برای شناسایی نواحی حفاظت شده ساده و کوتاه استفاده شوند.



- سایر روش ها یا ترکیبی از روش های فوق: می توان از ترکیبی از روش های فوق برای ساخت یک مدل قوی تر و دقیق تر برای ناحیه حفاظت شده استفاده کرد.

#### فیلتر کردن توالی ها:

پس از ساخت مدل، می توان از آن برای فیلتر کردن توالی های نماینده خوشه ها استفاده کرد. توالی هایی که با مدل مطابقت خوبی دارند، به عنوان کاندیدهای قوی تر برای ژن های زایلاناز در نظر گرفته می شوند.

#### خروجی:

خروجی این گام، مجموعه ای محدودتر از توالی های کاندید زایلاناز است که با مدل ناحیه حفاظت شده مطابقت دارند. این توالی ها می توانند در گام های بعدی برای بررسی های دقیق تر و تایید عملکرد زایلانازی مورد استفاده قرار گیرند.

#### گام چهارم: پیش بینی ساختار سه بعدی و هم ترازی ساختاری (امتیاز اضافه حداکثر ۳۰ درصد)

در این مرحله، هدف پیش بینی ساختار سه بعدی سه توالی از توالی های باقی مانده و مقایسه آن ها با ساختار یک زایلاناز صنعتی موجود در پایگاه داده پروتئین (PDB) است. این کار به ما کمک می کند تا شباهت ها و تفاوت های ساختاری بین زایلاناز های پیش بینی شده و زایلاناز صنعتی را بررسی کنیم و اطلاعاتی در مورد عملکرد و ویژگی های احتمالی آن ها به دست آوریم.

#### پیش بینی ساختار سه بعدی:

برای پیش بینی ساختار سه بعدی، می توان از ابزار های مختلفی استفاده کرد، از جمله:

- **AlphaFold:** یک ابزار مبتنی بر یادگیری عمیق که در سال های اخیر پیشرفت های چشمگیری در زمینه پیش بینی ساختار پروتئین داشته است.
- **Phyre2:** یک ابزار آنلاین که از روش های مدل سازی مقایسه ای برای پیش بینی ساختار پروتئین استفاده می کند.
- **SWISS-MODEL:** یک ابزار آنلاین دیگر که از روش های مدل سازی مقایسه ای و همولوژی استفاده می کند.
- **RoseTTAFold:** یک ابزار مبتنی بر یادگیری عمیق که مشابه AlphaFold عمل می کند.
- ...

#### هم ترازی ساختاری:

پس از پیش بینی ساختارها، می توانیم آن ها را با ساختار یک زایلاناز صنعتی موجود در PDB هم تراز کنیم. برای این کار، می توان از ابزار های مختلفی استفاده کرد، از جمله:

- **TM-align:** یک ابزار سریع و دقیق برای هم ترازی ساختاری پروتئین ها
- **PyMOL:** یک نرم افزار تجسم مولکولی که امکان هم تراز و مقایسه ساختارها را فراهم می کند.
- **ChimeraX:** یک نرم افزار تجسم و تحلیل ساختارهای مولکولی که امکانات زیادی برای هم تراز و مقایسه ساختارها دارد.



...

### گزارش نتایج:

نتایج هم‌ترازی ساختاری می‌تواند شامل موارد زیر باشد:

• **RMSD (Root Mean Square Deviation):** یک معیار عددی که میزان انحراف بین دو ساختار را نشان می‌دهد.

• **TM-score:** یک معیار عددی دیگر که شباهت کلی بین دو ساختار را نشان می‌دهد.

...

### نتیجه‌گیری:

با انجام این گام، می‌توانیم اطلاعات ارزشمندی در مورد شباهت‌ها و تفاوت‌های ساختاری بین زایلانازهای پیش‌بینی‌شده و زایلاناز صنعتی به دست آوریم. این اطلاعات می‌تواند به ما در درک بهتر عملکرد و ویژگی‌های این آنزیم‌ها کمک کند و در نهایت منجر به طراحی و تولید زایلانازهای جدید و بهبودیافته برای کاربردهای صنعتی شود.

### اطلاعات تکمیلی پروژه

- **کار گروهی:** پروژه به صورت گروهی انجام می‌شود. هر گروه می‌تواند حداکثر ۳ عضو داشته باشد.
- **سرگروه:** هر گروه باید یک سرگروه داشته باشد.
- **اسامی اعضا:** سرگروه موظف است اسامی اعضای گروه را در گروه اسکایپ قرار دهد.
- **نمره‌دهی:** نمره پروژه به صورت یکسان به همه اعضای گروه تعلق می‌گیرد.
- **زمان تحویل:** زمان تحویل پروژه به احتمال زیاد روز شنبه ۲۴ مرداد ماه خواهد بود. تایید نهایی زمان تحویل متعاقباً اعلام خواهد شد.
- **گزارش:** هر گروه باید یک گزارش مکتوب شامل جزئیات انجام پروژه به همراه فایل‌های تولید شده ارائه کند.

### جزئیات گزارش:

گزارش باید شامل بخش‌های زیر باشد:

۱. **مقدمه:** توضیح مختصری در مورد اهمیت زایلانازها و اهداف پروژه.
۲. **روش‌ها:** شرح دقیق مراحل انجام پروژه (گام‌های ۱ تا ۴).
۳. **نتایج:** ارائه نتایج حاصل از هر گام به همراه جداول، نمودارها و تصاویر مرتبط.
۴. **بحث:** تفسیر نتایج، مقایسه با مطالعات قبلی، و بیان محدودیت‌ها و پیشنهادات برای تحقیقات آینده.
۵. **نتیجه‌گیری:** خلاصه نتایج اصلی و دستاوردهای پروژه.
۶. **منابع:** فهرست منابع مورد استفاده در پروژه.

### فایل‌های تولید شده:



فایل‌های زیر باید به همراه گزارش ارائه شوند:

- فایل‌های حاوی توالی‌های شناسایی شده در هر گام
- فایل‌های حاوی مدل‌های ساخته شده (PSSM)، HMM، یا عبارت منظم
- فایل‌های حاوی ساختارهای پیش‌بینی شده
- فایل‌های حاوی نتایج هم‌ترازی ساختاری

**توجه:**

- رعایت اصول اخلاق پژوهشی (عدم سرقت ادبی، ارجاع صحیح به منابع) در گزارش ضروری است.
- گزارش باید به صورت واضح، مختصر، و علمی نوشته شود.

**موفق باشید!**  
**کاوه کاوسی**