

گزارش تحقیق شبکه تعاملی و فعل و انفعالات پروتئین‌ها

علیرضا سلطانی نشان

۱ فروردین ۱۴۰۳

۱ چرایی این گزارش

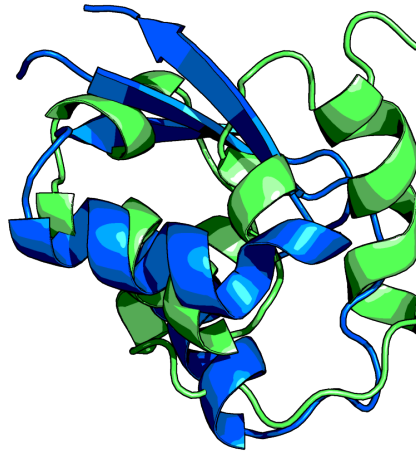
داده‌هایی که مربوط به روابط پروتئین‌ها در سطح اینترنت وجود دارد، داده‌هایی بسیار ارزشمندی هستند که به ما در درمان بیماری‌های مختلف کمک خواهند کرد. وجود این داده‌ها به ما این کمک را می‌کند که بتوانیم پیچیدگی بین مولکولی را کشف کنیم و عملکرد آنها را بیاموزیم که بعد از آن بتوانیم با ترکیب آنها به پروتئین‌های جدیدی برسیم که کاربردهای بسیار مهمی را در زندگی بشر ایفا می‌کنند.

منبع‌های بسیاری در این رابطه وجود دارد مانند ProBIS که در مقاله Protein binding sites for drug design که در نشریه Springer می‌توان آن را یافت، بسیار صحبت می‌کند. اما ما می‌توانیم از Api‌های آزاد منبع String^۱ استفاده کنیم که داده‌های آن تا به امروز به بیشتر از ۲۰ میلیون روابط پروتئین‌ها را دارا می‌باشد. در این بین بایستی اشاره کرد که در دیتابیس‌های آن ۶۷،۵۹۲،۴۶۴ پروتئین را از ۱۴،۰۹۴ اورگانیزم پوشش می‌دهد. در این گزارش سعی بر شناخت برخی پروتئین‌ها را داریم و علاوه بر آن سعی کردیم که روابط بین پروتئین‌هایی که بعداً به آن‌ها اشاره می‌کنیم را به وسیله ابزاری مناسب به سریع‌ترین حالت ممکن در گراف نمایش دهیم تا شناخت و تشخیص آن‌ها را برای ما ملموس کند.

^۱ STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

۲ آشنایی قبلی

۱.۲ AlphaFold



شکل ۱: نمونه‌ای از آلفافولد

یک ابزار هوش مصنوعی معرفی شده توسط دیپ‌ماینند گوگل است که به وسیله داده‌های فراوانی که دارد می‌تواند ساختار سه بعدی پروتئین‌ها و آمینواسیدها را با دقت بالا پیش‌بینی کند. استفاده از آن می‌تواند در موارد زیر بسیار مفید باشد:

- درک بهتر عملکرد پروتئین‌ها و توابع آنها
- طراحی داروهای جدید
- با استفاده از پیش‌بینی‌هایی که این ابزار انجام می‌دهد، تحلیل و مقایسه ساختاری پروتئین‌ها را برای متخصصین آسان می‌کند که از طریق آنها می‌توانند به دنبال الگوها و تفاوت‌های مهم در ساختار باشند.
- سایر کاربردهای پزشکی و زیست‌شناسی

۲.۲ پروتئین‌ها

اجزای اساسی مهم زندگی هستند. در تشکیل بسیاری از سلول‌ها، بافت‌ها و اعضای بدن انسان و سایر موجودات زنده نقش بسیار مهمی دارد. ساختار و ویژگی هر پروتئین را آمینواسیدهایی که به آن وصل هستند مشخص می‌کند. (آمینواسیدها به صورت پیوسته متصل هستند). هر آمینواسید در یک زنجیره پروتئینی در یک نقطه از دایره می‌باشد. وقتی نقاط را به هم وصل می‌کنیم، یک زنجیره خطی به وجود می‌آید که امکانات و قابلیت‌های مختلفی دارد. این امکانات و قابلیت‌ها می‌توانند به عنوان ویژگی‌های پروتئین‌ها تعریف شوند که رفتار آنها را مشخص می‌کند. مانند Struct هایی هستند که درون آنها Trait ها و ویژگی‌های مختلفی نوشته شده است. طبق نوع و محل حضور پروتئین‌ها در عملکرد متفاوت هستند:

۱. فعالیت آنزیمی به منظور کاتالیز فرآیند
۲. شناسایی میکروب‌ها و سلول‌های سرطانی
۳. انتقال موادی مانند گازهای تنفسی و سیگنال‌دهی

۳.۲ بسیار یا پلیمر

یک درشت مولکول است که از تعداد انبوهی از اجزای کوچک‌تری به نام مونومر تشکیل شده است، به گونه‌ای که زنجیره‌ای به هم متصل هستند (بس: بسیار، پار: پاره، قطعه). پروتئین‌ها مانند زنجیره‌ای از یک کلافی سه بعدی از بسپارهایی هستند که از ترکیب اسیدهای آمینه حاصل می‌شوند.

۴.۲ اسیدهای آمینه

اسیدهای آمینه یا آمینواسیدها ترکیبات آلی متشکل از گروه‌های عاملی آمینو و کربوکسیلیک اسید هستند، بیشتر از ۵۰۰ آمینواسید در طبیعت وجود دارد. مهم‌ترین آمینواسید از نوع آلفا می‌باشد. نکته: ۲۲ اسید آمینه آلفا واحدهای تشکیل دهنده پروتئین هستند.

۳ روابط و پروتئین‌های بررسی شده

۱.۳ داده‌های تحلیل شده

مجموعه داده‌های ما شامل فعل و انفعالات جفتی برای تعداد انگشت شماری از پروتئین‌هایی است که در مسیرهای سروتونین دخیل هستند.

۲.۳ توصیف کار انجام شده

گرافی که ما از داده‌های تحلیل شده، تهیه کرده‌ایم گرافی با وزن و بدون جهت می‌باشد:

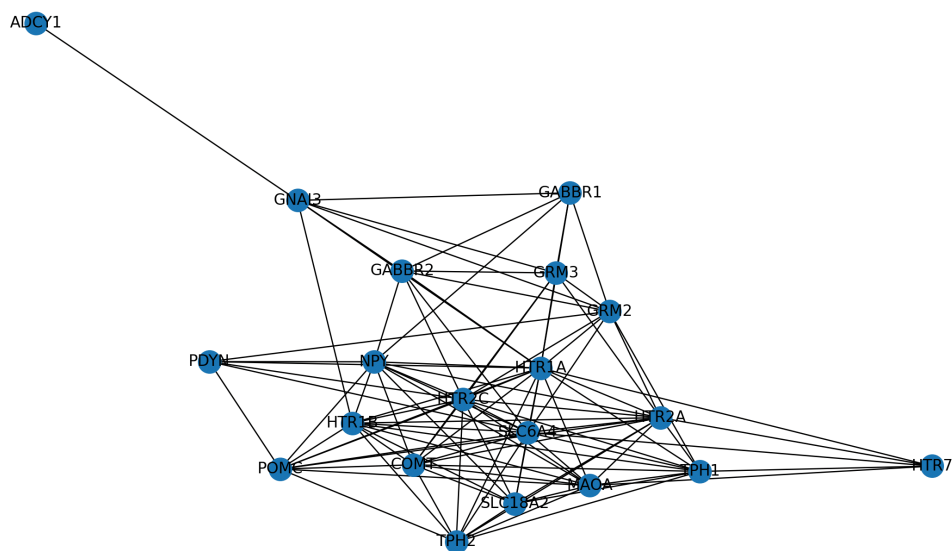
- دلیل اصلی بدون جهت بودن گراف آن است که تعامل بین پروتئین A با پروتئین B هیچ فرقی با رابطه پروتئین B با پروتئین A ندارد.
- گراف وزن دار است زیرا هر یال وزن و هزینه‌اش را بر اساس امتیاز تعامل بین دو پروتئین مشخص می‌کند.

در این برنامه ورودی، ۲۰ عدد از پروتئین‌هایی که بالاتر توضیح داده شد می‌باشد، که به وسیله Api که از دیتابیس String بدست آوردیم، آن لیست را وارد کردیم تا بتوانیم رابطه دو پروتئین را با یک وزن مشخص کنیم. سپس بعد از مشخص کردن پروتئین اول و دوم و هزینه اتصال بین دو راس، اقدام به ترسیم گراف آن کردیم که بتوانیم متوجه شوم هر پروتئین به چه پروتئین‌های دیگر می‌تواند متصل شود.

جدول ۱: نمونه‌ای از ارتباط پروتئین‌ها با وزن مشخص از داده‌های خام دیتابیس String

امتیاز	(ب) پروتئین	(آ) پروتئین
۴۳۶.۰	GRM۲	TPH۱
۶۰۷.۰	HTR۷	TPH۱
۶۳۶.۰	COMT	TPH۱
۶۵۳.۰	SLC۱۸A۲	TPH۱
۴۰۳.۰	HTR۱A	GABBR۲
۴۳۹.۰	HTR۲C	GABBR۲
۹۴۳.۰	COMT	SLC۶A۴
۹۶۱.۰	HTR۲A	SLC۶A۴
۴۰۹.۰	POMC	TPH۲
۵۹۹.۰	COMT	TPH۲
۶۸۷.۰	MAOA	TPH۲
۷۲.۰	SLC۱۸A۲	TPH۲
۷۶۲.۰	HTR۲A	TPH۲

نکته: درجه یک راس (پروتئین) در حقیقت قدرت اتصال آن به پروتئین‌های دیگر است.



شکل ۲: خروجی اتصالات پروتئین‌ها