

گزارش تحقیق شبکه تعاملی و فعل و انفعالات پروتئین‌ها

شقایق راجی
گرایش مهندسی نرم‌افزار
shaqayq.raji.۰۲۱@gmail.com

علیرضا سلطانی نشان
گرایش مهندسی نرم‌افزار
a.soltani@iau-tnb.ac.ir

۷ فروردین ۱۴۰۳

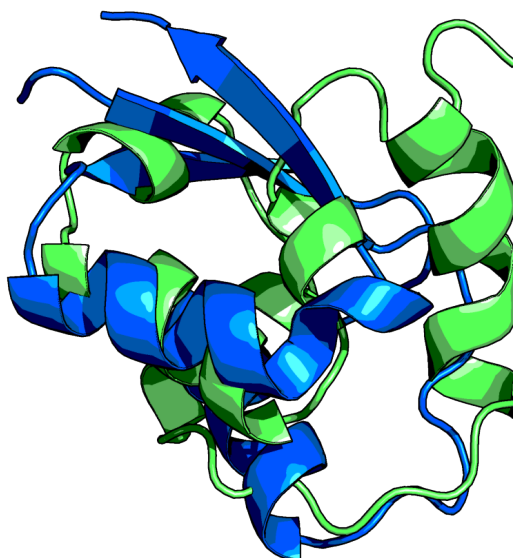
چرایی این گزارش

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 [۲] 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 اورگانسیم 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.

فهرست مطالب

۳	۱	آشنایی قبلی
۳	۱.۱	AlphaFold
۳	۲.۱	پروتئین‌ها
۳	۳.۱	بسیار یا پلیمر
۴	۴.۱	اسیدهای آمینه
۴	۲	روابط و پروتئین‌های بررسی شده
۴	۱.۲	دادهای تحلیل شده
۴	۲.۲	لیست پروتئین‌های مورد بررسی در این تحقیق
۴	۳.۲	توصیف کار انجام شده
۶	۴.۲	میانگین درجه هر پروتئین
۶	۵.۲	نظریه Betweenness centrality [۳]
۸	۶.۲	نظریه Minimum Spanning Tree [۴]



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

۲.۱ 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 Trait‌ها 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

۳.۱ 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.

۴.۱ 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.

۲ روابط و پروتئین‌های بررسی شده

۱.۲ داده‌های تحلیل شده

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

۲.۲ لیست پروتئین‌های مورد بررسی در این تحقیق

نام پروتئین	جدول ۱: پروتئین‌های بخش سروتونین کاربرد
TPH1	راهبرد در سنتز سروتونین
TPH2	کلیدی در تولید سروتونین: (تریپتوفان هیدروکسیلاز ۲)
COMT	تجزیه کاتکولامین‌ها: (کاتکول آمینه ترانسفراز)
SLC18A2	حمل وستیکول‌های سروتونین: (پروتئین حمل‌کننده وستیکولر ۲ بخشی)
SLC6A4	حمل سروتونین از فضای سیناپسی: (پروتئین حمل‌کننده سروتونین)
HTR1B	نقش در تنظیم خواب و اضطراب: (گیرنده ۱ اسروتونین)
HTR2C	تأثیرگذاری بر اشتها و تغذیه: (گیرنده ۲ اسروتونین)
HTR2A	نقش در تنظیم خواب و شناخت: (گیرنده ۲ اسروتونین)
HTR1A	نقش در تنظیم در خلق و هیجان: (گیرنده ۱ اسروتونین)
HTR7	نقش در تأثیر بر حافظه و یادگیری: (گیرنده ۷ اسروتونین)
MAOA	تجزیه نوراپی‌نفرین و سروتونین: (منیزم آمین اکسیداز منوآمین)
GABBR1	تأثیرگذاری بر عملکرد سیستم عصبی: (گیرنده بتا ۱ گاما آمینوبوتیریک اسید)
GABBR2	تأثیرگذاری بر سیستم عصبی مرکزی: (گیرنده بتا ۲ گاما آمینو بوتیریک اسید)
POMC	پیش ماده‌ای برای هرمون‌هایی مانند آدرنوکورتیکوتروپین: (پرواپیوملانوکورتین)
GNAI3	در موازی‌سازی تأثیرات گیرنده‌های پروتئین: (پروتئین باکتریایی ۳)
NPY	نشق در کنترل اندازه مشت غذایی: (پیتاید یو)
ADCY1	تأثیرگذار در سیگنالینگ سلولی: (آنزیم آدنیلات سیکلازی نوع ۱)
PDYN	ماده معرفتی مانند مورفین: (پرودینورفین)
GRM2	یک نوع از گیرنده‌های متابوتروپی گلوتامات: (گروه متابوتروپی ۲)
GRM3	نقش در یادگیری و حافظه: (گروه متابوتروپی ۳)

۳.۲ توصیف کار انجام شده

برای دسترسی به کد [۱] این تحقیق به مراجع این برگه مراجعه کنید. گرافی که ما از داده‌های تحلیل شده، تهیه کرده‌ایم گرافی با وزن و بدون جهت می‌باشد:

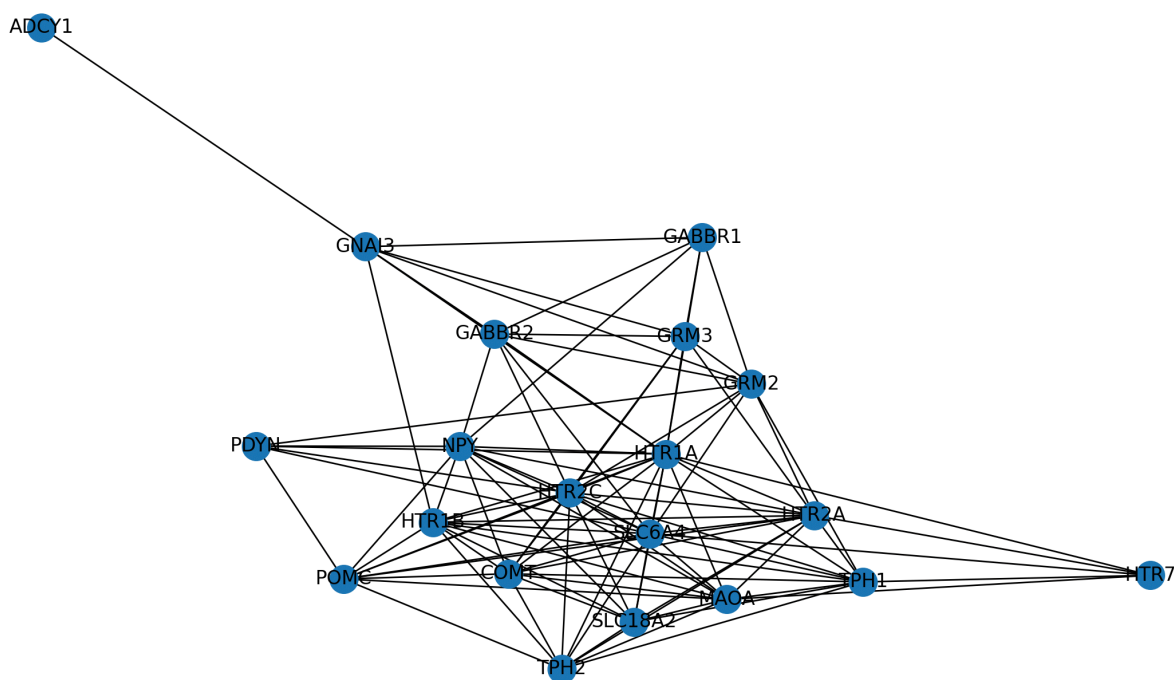
- دلیل اصلی بدون جهت بودن گراف آن است که تعامل بین پروتئین A با پروتئین B هیچ فرقی با رابطه پروتئین B با پروتئین A ندارد.
- گراف وزن دار است زیر را هر یال وزن و هزینه‌اش را بر اساس امتیاز تعامل بین دو پروتئین مشخص می‌کند.

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

پروتئین (آ)	پروتئین (ب)	امتیاز
TPH1	GRM2	۰/۴۳۶
TPH1	HTR7	۰/۶۰۷
TPH1	COMT	۰/۶۳۶
TPH1	SLC18A2	۰/۶۵۳
GABBR2	HTR1A	۰/۴۰۳
GABBR2	HTR2C	۰/۴۳۹
SLC6A4	COMT	۰/۹۴۳
SLC6A4	HTR2A	۰/۹۶۱
TPH2	POMC	۰/۴۰۹
TPH2	COMT	۰/۵۹۹
TPH2	MAOA	۰/۶۸۷
TPH2	SLC18A2	۰/۷۲
TPH2	HTR2A	۰/۷۶۲

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



شکل ۲: خروجی اتصالات پروتئین‌ها

به طور کلی می‌توان گفت که ۲۰ پروتئین در این گراف به عنوان رئوس حاضر هستند که با ۱۰۱ یال آن‌ها با یکدیگر ارتباطی با امتیاز مشخص دارند:

نودها Protiens	تعداد ارتباطات
۲۰	۱۰۱

جدول ۳:

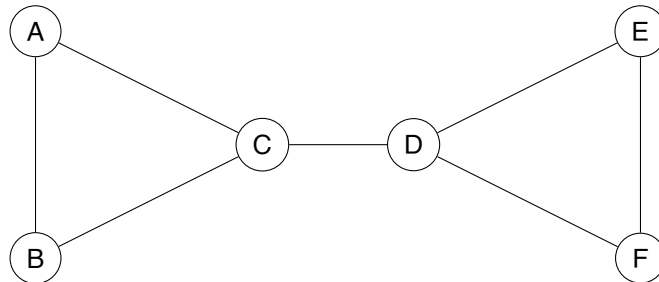
۴.۲ میانگین درجه هر پروتئین

پروتئین	درجه
TPH1	۱۲
COMT	۱۲
SLC18A2	۱۳
HTR1B	۱۲
HTR2C	۱۱
HTR2A	۱۱
MAOA	۱۲
TPH2	۱۳
HTR1A	۱۰
HTR7	۱۴
SLC6A4	۱۱
GABBR2	۱۱
POMC	۱۳
GNAI3	۹
NPY	۱۲
ADCY1	۷
PDYN	۱۳
GRM2	۱۱
GRM3	۱۱
GABBR1	۱۰

۵.۲ نظریه Betweenness centrality [۳]

$$BC(V) = \sum_{u, w \in V} \left(\frac{\sigma_{vw}(V)}{\sigma_{vw}} \right) \quad (۱)$$

- σ_{vw} : مجموع تعداد کوتاه‌ترین مسیرها بین نود u و w
- $\sigma_{vw}(V)$: مجموع تعداد کوتاه‌ترین مسیرهایی که از بین (V) گذر می‌کند.



مقدار BC را برای راس C به شکل زیر مشخص می‌کنیم.

$$(A, B) = \sigma_{A,B} \rightarrow 1, (A, B) = \sigma_{A,B(C)} \rightarrow 0, (\sigma_{A,B(C)} / \sigma_{A,B}) = 0/1 \rightarrow 0 \quad (۲)$$

$$(A, D) = \sigma_{A,D} \rightarrow 1, (A, D) = \sigma_{A,D(C)} \rightarrow 1, (\sigma_{A,D(C)} / \sigma_{A,D}) = 1/1 \rightarrow 1 \quad (۳)$$

$$(A, E) = \sigma_{A,E} \rightarrow 1, (A, E) = \sigma_{A,E(C)} \rightarrow 1, (\sigma_{A,E(C)} / \sigma_{A,E}) = 1/1 \rightarrow 1 \quad (۴)$$

$$(A, F) = \sigma_{A,F} \rightarrow 1, (A, F) = \sigma_{A,F(C)} \rightarrow 1, (\sigma_{A,F(C)} / \sigma_{A,F}) = 1/1 \rightarrow 1 \quad (۵)$$

$$(B, D) = \sigma_{B,D} \rightarrow 1, (B, D) = \sigma_{B,D(C)} \rightarrow 1, (\sigma_{B,D(C)}/\sigma_{B,D}) = 1/1 \rightarrow 1 \quad (۶)$$

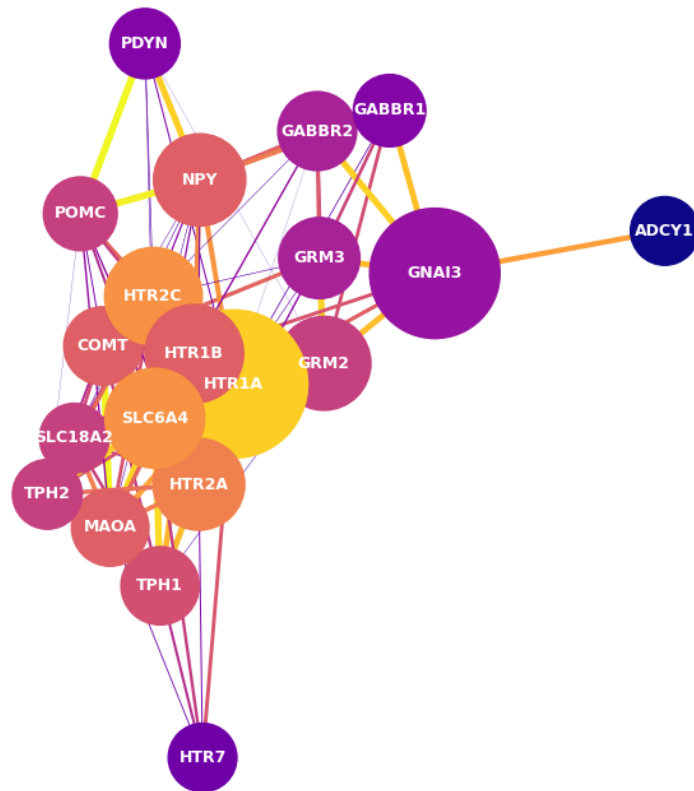
$$(B, E) = \sigma_{B,E} \rightarrow 1, (B, E) = \sigma_{B,E(C)} \rightarrow 1, (\sigma_{B,E(C)}/\sigma_{B,E}) = 1/1 \rightarrow 1 \quad (۷)$$

$$(B, F) = \sigma_{B,F} \rightarrow 1, (B, F) = \sigma_{B,F(C)} \rightarrow 1, (\sigma_{B,F(C)}/\sigma_{B,F}) = 1/1 \rightarrow 1 \quad (۸)$$

$$(E, F) = \sigma_{E,F} \rightarrow 1, (E, F) = \sigma_{E,F(C)} \rightarrow 0, (\sigma_{E,F(C)}/\sigma_{B,F}) = 0/1 \rightarrow 0 \quad (۹)$$

$$(D, E) = \sigma_{D,E} \rightarrow 1, (D, E) = \sigma_{D,E(C)} \rightarrow 0, (\sigma_{D,E(C)}/\sigma_{D,E}) = 0/1 \rightarrow 0 \quad (۱۰)$$

مقدار Betweenness centrality براساس راس C برابر با ۶ است.
به دلیل آنکه C و D با یکدیگر در یک راستا و سطح گراف هستند پس مقدار راس D هم مانند راس C برابر با ۶ می باشد.
تمامی گوشه های گراف مقدار BC آنها برابر با ۰ می باشد.

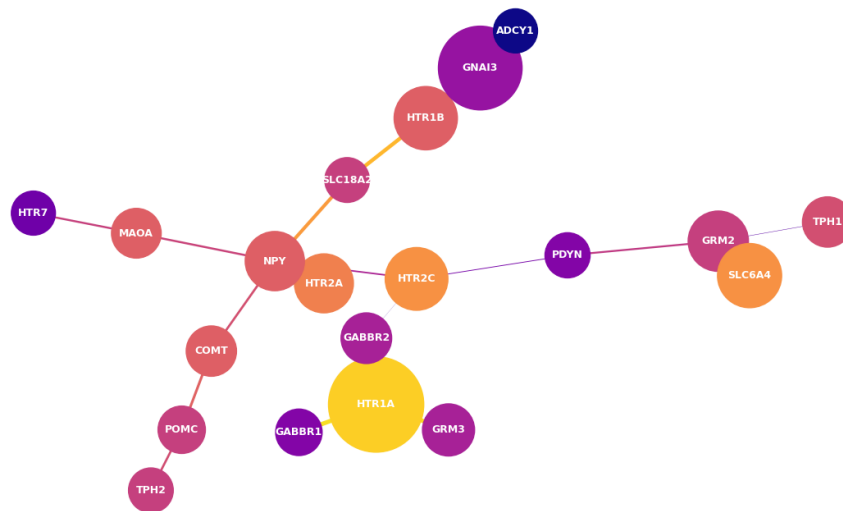


شکل ۳: گراف Betweenness centrality

از Degree و BC هر کدام از پروتئین ها را استفاده می کنیم که دو پارامتر رنگ و اندازه را از طریق وزن یال ها بدست آوریم.
رنگ ها از بنفش پررنگ شروع می شود تا زرد روشن. هر چقدر یک پروتئین زردتر باشد درجه بیشتری را دارا می باشد. بزرگ ترین پروتئین
بیشترین BC را بین پروتئین های دیگر دارد.

۶.۲ نظریه Minimum Spanning Tree [۴]

گرافی است بدون حلقه، زیر مجموعه تمام لبه‌هایی که گره‌ها را با کمترین وزن ممکن به یکدیگر متصل می‌کند.



شکل ۴: بهینه‌سازی گراف با mst

- [1] asn Soltani Neshan, Alireza. Serotonin Pathways Interaction Networks.
- [2] Konc, Janez and Janežič, Dušanka. Protein binding sites for drug design. *Biophysical Reviews*, 14(6):1413–1421, 2022.
- [3] Wikipedia contributors. Betweenness centrality — Wikipedia, the free encyclopedia, 2024. [Online; accessed 20-March-2024].
- [4] Wikipedia contributors. Minimum spanning tree — Wikipedia, the free encyclopedia, 2024. [Online; accessed 20-March-2024].