

تعیین پروتئین‌های واسط در شبکه تعامل پروتئین - پروتئین با در نظر گرفتن بیماری‌های شایع در پروتئین‌های مونولایت

فرشید شیرافکن¹، سجاد قرقانی^{2*}

1- دانشجوی دکتری، گروه بیوانفورماتیک، مرکز تحقیقات بیوشیمی بیوفیزیک، دانشگاه تهران، تهران، ایران
2- استادیار، گروه بیوانفورماتیک، مرکز تحقیقات بیوشیمی بیوفیزیک، دانشگاه تهران، تهران، ایران صندوق پستی: 1365-13145

*نویسنده مسئول: s.gharaghani@ut.ac.ir

پذیرش: 1400/12/15

دریافت: 1400/6/30

چکیده

پروتئین‌های مونولایت یک زیرکلاس از پروتئین‌های چندعملکردی است که در آنها بیش از یک عملکرد مستقل یا به‌طور معمول متمایز در یک زنجیره پلی‌پپتیدی منفرد رخ می‌دهد. با تحلیل شبکه‌های تعاملی پروتئین‌ها در سلول، می‌توان درک کرد که چگونه فرایندهای پیچیده باعث بیماری می‌شوند. به کمک زیست‌شناسی سیستم‌ها، می‌توان سیستم‌های بزرگ‌تر و پیچیده‌تر را مطالعه کرد و به اساس مولکولی چند بیماری توجه داشت. پروتئین‌های ارگانیزم انسان که مونولایت هستند، بیشتر در بیماری‌های سرطان، کم‌خونی و بیماری مرتبط با سیستم عصبی درگیر هستند. در این کار با توجه به شبکه PPI انسانی، یک زیرشبکه ایجاد شد که در آن نودها، پروتئین‌های ایجادکننده سه بیماری منتخب و یال‌ها، ارتباط این پروتئین‌ها با یکدیگر می‌باشند. قدرت آثار غیرمستقیم واسط‌های غیربیماری بین سه گروه از بیماری‌ها اندازه‌گیری شد و پروتئین‌های واسط کلیدی متصل‌کننده بیماری‌ها شناسایی شدند. نتایج کار نشان‌دهنده ارتباط بین نقش واسط و مرکزیت و همچنین بین نقش واسط و ویژگی‌های عملکردی این پروتئین‌ها می‌باشد. مشاهده شد که یک پروتئین که نقش واسط کلیدی غیرمستقیم بین دو بیماری را بازی می‌کند، به‌طور لزوم در شبکه PPI، هاب نمی‌باشد. بنابراین همان‌طور که به پروتئین‌های هاب توجه می‌شود گیرند، باید به پروتئین‌های واسط نیز توجه شود. ما مشاهده کردیم که واسط‌های بین بیماری‌های کم‌خونی و اعصاب از نظر عملکردی اهمیت زیادی در سلول دارند. پروتئین‌های واسط‌ای که در اینجا پیشنهاد شده‌اند، باید به‌صورت تجربی به‌عنوان پروتئین‌های فرضی مرتبط با بیماری آزمایش شوند.

کلید واژگان: پروتئین‌های مونولایت، شبکه تعاملی پروتئین - پروتئین، هاب، بیماری.

مقدمه

در رابطه با روش‌های محاسباتی برای ژن‌های بیماری به‌وسیله خان [7] و شاران [8] انجام شده است. تعامل پروتئین - پروتئین⁵ یا همان PPI در اثر تماس فیزیکی دو یا بیش از دو پروتئین در نتیجه نیروهای الکترواستاتیک یا رویدادهای شیمیایی با یکدیگر رخ می‌دهد. اغتشاش در شبکه اندرکنش پروتئین‌ها می‌تواند منشأ بیماری‌های مختلف باشد. به همین علت، تعامل پروتئین - پروتئین برای هر فرایندی در سلول زنده اهمیت زیادی دارد. اطلاعات در مورد این تعاملات درک ما را از بیماری‌ها بهبود می‌بخشد و می‌تواند پایه‌ای برای روش‌های جدید درمانی باشند. در کاربردهای زیادی از مدل‌سازی شبکه‌های PPI با گراف ساده استفاده شده است. تجزیه و تحلیل خصوصیات توپولوژیکی شبکه PPI و بررسی ویژگی‌های عملکرد پروتئین‌ها می‌تواند موجب درک بهتر بیولوژی بیماری و شناسایی جایگاه‌های احتمالی کاندیدهای دارویی باشد. پژوهش‌های خوبی روی شبکه‌های PPI و بیماری‌ها در سال‌های گذشته انجام شده است [9]. پروتئین‌های بیماری ممکن است برای ایجاد بیماری به‌طور مستقیم به یکدیگر اثر نکنند و از راه یک واسط یا گروهی از واسط‌ها به همدیگر اثر کنند. بنابراین ارتباط بین پروتئین‌های ایجادکننده بیماری، بسیار پیچیده است.

پروتئین‌های مون‌لایت زیرمجموعه‌ای از پروتئین‌های چند عملکردی را تشکیل می‌دهند که در آنها یک زنجیره پلی‌پپتیدی بیش از یک عملکرد بیوشیمیایی یا بیوفیزیکی دارد [1]. به عبارت دقیق‌تر، می‌توان کلمه مون‌لایت را برای پروتئین‌هایی با حداقل دو عملکرد غیرمرتبط متفاوت به کاربرد، زیرا این چندعملکردی نتیجه همجوشی ژن¹، دمین‌های متعدد، انواع اتصال چندگانه، قطعه‌های پروتئولیتیک، خانواده‌های همولوگ یا اثر پلی‌تروپی نمی‌باشند [2].

اولین نمونه پروتئین‌های مون‌لایت در اواخر دهه 1980 به‌وسیله پیاتیگورسکی و ویستو گزارش شد [3]. آنها متوجه شدند که کریستالین، یک پروتئین ساختاری در لنز چشم، نقش آنزیمی هم دارد. البته واژه مون‌لایت در سال 1999 به‌وسیله جفری² ابداع شد. این پژوهشگر با جمع‌آوری اطلاعات پروتئین‌های مون‌لایت، پایگاه داده مون پرات³ را طراحی کرده است [4]. تاکنون پروتئین‌های مون‌لایت در پستانداران، مخمرها، کرم‌ها، باکتری‌ها، گیاهان، ویروس‌ها، آرکی‌ها و بسیاری از موجودات دیگر کشف شده‌اند. پایگاه داده مون پرات شامل اطلاعات حدود 400 پروتئین مون‌لایت است که تعداد 69 پروتئین آن مربوط به انسان است. شکل 1 نشان می‌دهد که یک پروتئین چگونه می‌تواند مون‌لایت باشد.

تغییرات در ژنوم انسان می‌تواند منجر به بیماری شود. ژن‌های مربوط به ایجاد بیماری‌ها، ژن‌های بیماری⁴ نامیده می‌شوند و پروتئین آنها محصولات پروتئین بیماری هستند [6]. ژن‌های بیماری را با روش‌های مختلفی مانند استفاده از حاشیه‌نویسی یا آنالیز توالی و یا در شبکه تعاملی پروتئین می‌توان بررسی کرد. بررسی‌های کلیدی

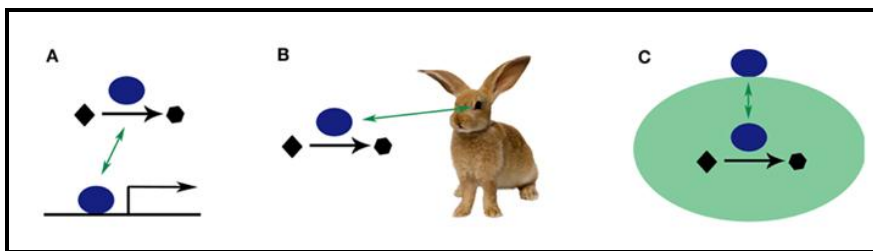
1 Gene fusion

2 Jeffery

3 MoonProt

4 Disease genes

5 Protein-Protein Interaction



شکل 1 چگونه یک پروتئین می‌تواند مون‌لایت باشد. یک پروتئین می‌تواند یک واکنش آنزیمی را کاتالیز کند و همچنین (الف) به DNA متصل شود و به‌عنوان یک عامل رونویسی عمل کند؛ (ب) نقش غیرکاتالیتی در عدسی چشم (کریستالین) داشته باشد یا (ج) در سطح سلول قرار بگیرد تا به‌عنوان گیرنده عمل کند [5].

برای فهم مسیرهای خاص، کافی نباشد و نیاز به شناسایی واسط‌ها باشد. با ایجاد یک زیرشبکه از شبکه PPI انسانی و انتخاب پروتئین‌های ایجادکننده چند بیماری منتخب به‌عنوان گره‌های شبکه و ارتباط این پروتئین‌ها با یکدیگر به‌عنوان یال‌های شبکه، قدرت آثار غیرمستقیم واسط‌های غیربیماری بین گروه‌های بیماری‌ها، اندازه‌گیری می‌شوند و پروتئین‌های واسط کلیدی متصل‌کننده بیماری‌ها شناسایی خواهند شد. ممکن است یک پروتئینی که نقش واسط کلیدی غیرمستقیم بین دو بیماری را بازی می‌کند، به‌طور لزوم در شبکه PPI هاب نباشد و هدف شناسایی آنها است و باید مانند هاب به آن توجه شود.

مواد و روش‌ها

در این پژوهش از بانک اطلاعاتی PPI به‌عنوان پایگاه داده اینتراکشن‌های پروتئین انسان استفاده شد. همچنین پروتئین‌های درگیر در سه بیماری A, N, C برای ارگانیزم انسان از پایگاه داده یونی پرات⁴ استخراج شدند. براساس این اطلاعات، زیرشبکه‌ای به نام DPN⁵ شامل پروتئین‌های بیماری ایجادکننده سه بیماری به همراه شرکای آنها نیز ایجاد شد. اینتراکشن‌ها در این شبکه بین دو پروتئین بیماری یا بین یک پروتئین بیماری⁶ و یک پروتئین غیربیماری است. ما فقط اینتراکشن‌های

می‌توان برای درک تعیین غیرمستقیم فرایندهای بیولوژیکی از تئوری گراف و آنالیزهای شبکه استفاده کرد. تجزیه و تحلیل شبکه، اطلاعاتی در رابطه با ویژگی‌های محلی و عمومی سیستم فراهم می‌کند. در این مقاله از روش تحلیل شبکه برای تعیین کمیت آثار غیرمستقیم از راه پروتئین‌ها در شبکه تعاملی آنها به کمک شاخص‌های شبکه استفاده شده است [10-12]. هدف، تعیین پروتئین‌های واسط آثار غیرمستقیم در ارتباطات پروتئین‌های درگیر در سه بیماری سرطان¹، کم‌خونی² و بیماری مرتبط با سیستم اعصاب³ (که به اختصار با C, A و N نشان می‌دهیم) است. علت انتخاب این بیماری‌ها، درگیر بودن بیشتر پروتئین‌های مون‌لایت انسانی در آنها می‌باشد.

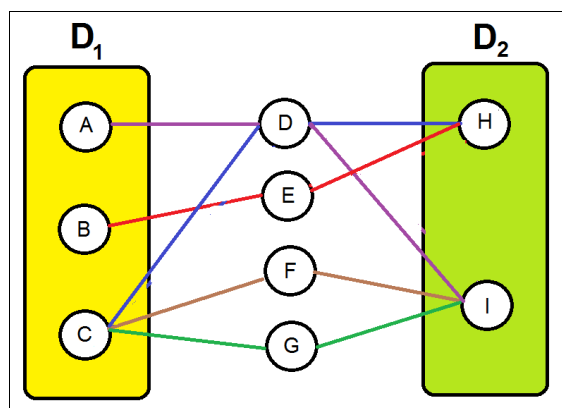
مسئله این پژوهش، بررسی اهمیت عملکردی پروتئین‌های واسط است. در بیشتر بررسی‌های انجام‌شده بر شبکه‌ها، بیشتر این موضوع که هاب‌ها از نظر عملکردی مهم‌تر هستند، مطرح شده است [13-15]. بنابراین در این مطالعه ما علاقه‌مند شدیم تا به بررسی اینکه آیا هاب‌ها بیشتر مسئول اتصال بین پروتئین‌های بیماری هستند؟، بپردازیم و در این کار بر پروتئین‌های واسط کلیدی که شرکای تعاملی پروتئین‌های بیماری در PPI هستند، تمرکز کردیم. ممکن است شناسایی هاب‌ها

4 Uniprot
5 Disease Protein Network
6 Disease Protein

1 Cancer
2 Anemia
3 Neurodegeneration

(شکل 2). فرض کنیم درجه گره‌ها در شبکه شکل 2 برابر باشد با:

$$A=8, B=13, C=15, D=4, E=5, F=9, G=10, H=7, I=20$$



شکل 2 یک شبکه فرضی که نحوه اندازه‌گیری آثار غیرمستقیم را نشان می‌دهد. در این شبکه دو نوع بیماری D1 و D2 نشان داده شده است. پروتئین‌های A، B و C در بیماری D1 و پروتئین‌های H و I در بیماری D2 درگیر هستند. پروتئین‌های D، E، F و G واسطه‌های بین دو نوع بیماری هستند.

مقدار شاخص M برای پروتئین D (زمانی که بین A و I میانجی‌گری می‌کند) برابر است با:

$$M_D = \frac{1}{D_D} \left(\frac{1}{D_A} + \frac{1}{D_I} \right) = \frac{1}{4} \left(\frac{1}{8} + \frac{1}{20} \right) = 0.014$$

مقدار شاخص M برای پروتئین D (زمانی که بین C و H میانجی‌گری می‌کند) برابر است با:

$$M_D = \frac{1}{D_D} \left(\frac{1}{D_C} + \frac{1}{D_H} \right) = \frac{1}{4} \left(\frac{1}{15} + \frac{1}{7} \right) = 0.038$$

مقدار شاخص M برای پروتئین‌های E، F و G برابر است با:

$$M_E = \frac{1}{D_E} \left(\frac{1}{D_B} + \frac{1}{D_H} \right) = \frac{1}{5} \left(\frac{1}{13} + \frac{1}{7} \right) = 0.03$$

$$M_C = \frac{1}{D_C} \left(\frac{1}{D_A} + \frac{1}{D_I} \right) = \frac{1}{15} \left(\frac{1}{8} + \frac{1}{20} \right) = 0.0053$$

$$M_F = \frac{1}{D_F} \left(\frac{1}{D_C} + \frac{1}{D_I} \right) = \frac{1}{9} \left(\frac{1}{15} + \frac{1}{20} \right) = 0.0059$$

غیرمستقیم دومارحله‌ای¹ بین پروتئین‌ها را بررسی کردیم. لینک‌ها در شبکه‌ها، بدون جهت و بدون وزن در نظر گرفته شده‌اند.

شبکه PPI استفاده‌شده در روش ما، شامل 18161 نود و 187096 یال است. همچنین شبکه DPN ایجادشده شامل 153375 یال و تعداد 16697 نود است. تعداد پروتئین‌های درگیر در سه بیماری سرطان، کم‌خونی و عصبی به ترتیب عبارتند از 7143، 274 و 364.

مقدار 55 درصد پروتئین‌های مون‌لایت انسانی، در بیماری درگیر می‌باشند. تعداد 26 پروتئین در سرطان، 11 پروتئین در کم‌خونی و 7 پروتئین در سیستم عصبی درگیر هستند. ارتباطات مستقیمی بین مجموعه پروتئین‌های بیشتر جفت بیماری‌ها وجود دارند. اما جفت‌های A-N و C-N فقط شامل لینک غیرمستقیم هستند.

یک الگوی گراف در مطالعات سیستم بیولوژی، از N گره و M لینک تشکیل می‌شود، به‌طوری‌که گره‌ها مشخص‌کننده پروتئین‌ها و لینک‌ها مشخص‌کننده اینتراکشن بین پروتئین‌ها می‌باشند. به کمک شاخص مرکزیت درجه که تعداد همسایه‌های یک گره را مشخص می‌کند، می‌توان به اهمیت نسبی آن گره پی برد [13]. نودهای با درجه بالا، هاب² نام دارند. درجه گره i را با Di نشان می‌دهیم.

در شبکه‌ای با لینک‌های بدون جهت که آثار آن در هر جهت با احتمال یکسان قابل گسترش باشد، رابطه شماره 1 قدرت یک پروتئین k در میانجی‌گری³ یک مسیر دو مرحله‌ای بین i و j را نشان می‌دهد.

$$M_k = \frac{1}{D_k} \times \frac{1}{D_i} + \frac{1}{D_k} \times \frac{1}{D_j} = \frac{1}{D_k} \left(\frac{1}{D_i} + \frac{1}{D_j} \right)$$

فرض کنید سه پروتئین A، B، C در بیماری D1 و دو پروتئین H، I در بیماری D2 درگیر هستند. همچنین چهار پروتئین D، E، F و G به‌عنوان واسطه عمل می‌کنند

1 two-step-long
2 Hub
3 Mediate

در جدول 1، واسطه‌ها و قدرت آثار غیرمستقیم بین پروتئین‌های موجود در دسته بیماری‌های D1 و D2، نشان داده شده است.

جدول 1 واسطه‌ها و قدرت آثار غیرمستقیم بین پروتئین‌های بیماری‌ها از نوع D1 و D2

D1\D2	H	I
A	-	D=0.014
B	E=0.03	-
C	D=0.038	F=0.0059, G=0.0053

از جمع دو مقدار 0.0059 و 0.0053، مقدار 0.0112 به دست می‌آید.

همان‌طور که مشخص است، قوی‌ترین اثر غیرمستقیم بین دو بیماری به‌وسیله پروتئین واسط D، واسطه‌گری می‌شود. با توجه به مجموع مقادیر در سطرها و ستون‌ها، مشخص است که در میان پروتئین‌های D1، پروتئین C بیشترین تأثیر را از D2 می‌گیرد. همچنین در میان پروتئین‌های D2، پروتئین H بیشترین تأثیر را از D1 می‌گیرد.

نتایج

بعضی پروتئین‌ها همیشه اهمیت کلیدی دارند، مانند Q8IWL3 که یک هاب (nD زیاد در PPI) و یک واسط مهم است (M زیاد در DPN). این پروتئین با توجه به شاخص M، در هر سه لینک C-A، A-N و C-N رتبه بالایی دارد.

بعضی پروتئین‌ها مانند P05412 و Q13573 در PPI هاب بوده ولی یک واسط مهم نمی‌باشند (M کم در DPN)، یعنی آثار غیرمستقیمی بین بیماری‌ها را میانجی‌گری نمی‌کنند.

بعضی پروتئین‌ها در PPI هاب نمی‌باشند (nD کم) و بینایی¹ کمی هم دارند، ولی یک واسط مهم می‌باشند (M زیاد در DPN)، مانند پروتئین O00626 در لینک C-A، پروتئین‌های P45877 و Q9BRQ5 در لینک A-N و پروتئین P45877 در لینک C-N. جدول 2 شامل 10 پروتئین واسط غیربیماری با بیشترین رتبه M است.

تعداد پروتئین‌های واسط در لینک C-A برابر 4005 در لینک A-N برابر 2679 و در لینک C-N برابر 6479 است. واسطه‌های مشترک زیادی در هر سه اتصال A-N، A-C و N-C وجود دارند که به‌طور معمول اهمیت زیادی ندارند. در شبکه DPN، بیشتر پروتئین‌های درگیر در بیماری A به پروتئین‌های N اتصال مستقیمی ندارند ولی پروتئین‌های A یا N وجود دارند که به‌طور مستقیم به پروتئین C متصل باشند. همچنین بیشتر پروتئین‌های واسط بین دو بیماری A و N، پروتئین‌های بیماری نیستند.

میانجی‌گری، قوی‌ترین اثر غیرمستقیم در لینک C-A به‌وسیله P38398 و در لینک A-N و همچنین لینک C-N، توسط Q8IWL3 می‌باشد. در میان پروتئین‌های مون‌لایت انسان، قوی‌ترین اثر غیرمستقیم در هر سه لینک به‌وسیله P13569 میانجی‌گری می‌شود. در میان پروتئین‌های A، پروتئین Q9NRR5 بیشترین تأثیر را از N می‌گیرد.

1 betweenness

جدول 2 ده پروتئین واسط غیربیماری با بیشترین رتبه M در C-A، A-N و C-N

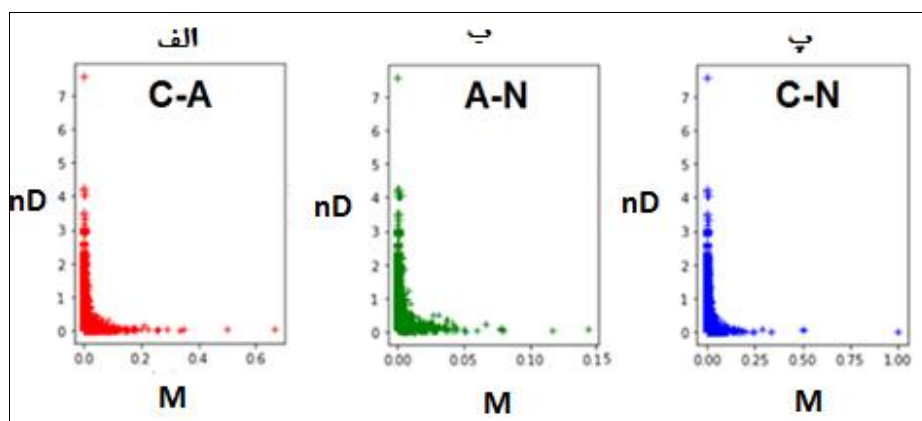
C-A	A-N	C-N
Q13162	Q8IWL3	Q8IWL3
Q86X19	Q13162	Q86X19
Q8IWL3	P45877	Q9NRD5
Q8IUH5	Q86X19	P11279
O00626	Q14197	Q6NUS6
P02768	P02647	Q9P0N5
Q93079	Q9BRQ5	P02768
O14880	O15354	P45877
P02647	Q96RQ3	P49286
Q6NUS6	O14880	P14854

جدول 3 پروتئین‌های مون‌لایت واسط را که در هیچ‌کدام از بیماری‌های C یا A یا N درگیر نمی‌باشند، با توجه به اهمیت آنها (شاخص M) نشان می‌دهد. ارتباط قرار دارند.

جدول 3 پروتئین‌های مون‌لایت واسط را که در هیچ‌کدام از بیماری‌های C یا A یا N درگیر نیستند، با توجه به اهمیت آنها بر اساس شاخص M نشان می‌دهد.

C-A	A-N	C-N
P13569	P13569	P13569
O60568	Q9H211	P27797
P09467	P27797	
P27797		

همان‌طور که در شکل 3 مشخص است، با مقایسه درجه هر گره در شبکه PPI در برابر شاخص M، مشخص می‌شود که واسطه‌های مهم به‌طور لزوم هاب نمی‌باشند، اما می‌توانند باشند.



شکل 3 درجه و نقش واسط بین پروتئین‌های بیماری (الف: C-O، ب: A-N، پ: C-N)، پروتئین‌های واسط مهم بین پروتئین‌های بیماری (اندازه‌گیری شده با شاخص M در شبکه DIP)، در برابر nD در شبکه PPI نشان داده شده است. واسطه‌های مهم در پروتئین‌های بیماری به‌طور لزوم پروتئین‌های هاب نمی‌باشند.

بحث

تحلیل‌های انجام‌شده نشان می‌دهد که بیشتر واسط‌ها، نقش‌های بیماری‌زا دارند. دمین‌های این پروتئین‌ها، دمین‌های مهم در انجام عملکرد پروتئین بوده و در سازوکارهای بیماری‌زا متفاوت درگیر می‌باشند.

در لینک C-A، قوی‌ترین اثر غیرمستقیم به‌وسیله پروتئین Q86X19، با نام پروتئین غشایی¹ میانجی‌گری می‌شود. این یافته هماهنگی خوبی با طیف عملکردهای آن در سلول دارد. این پروتئین در بیماری سندرم دهان- صورت - انگشت² (0FD) درگیر است [16]. پروتئین Q86X19 در پایگاه داده MoonDB به‌عنوان یک پروتئین مونولایت، پیشگویی شده است [17].

در لینک A-N و همچنین در لینک C-N، قوی‌ترین اثر غیرمستقیم به‌وسیله پروتئین Q8IWL3 میانجی‌گری می‌شود. این پروتئین به‌عنوان یک کوچاپرون³ در مونتاژ کلاستر آهن و گوگرد⁴ در میتوکندری و سیتوپلاسم عمل می‌کند [18: 19]. این پروتئین در مسیر⁵ بیوستز کلاستر آهن و گوگرد⁶ درگیر است که قسمتی از بیوستز کوفاکتور⁷ می‌باشد.

بیشتر پروتئین‌های واسط در شبکه DIP، پروتئین بیماری نمی‌باشند، البته واسطه‌هایی با رتبه بالا که در بیماری درگیر باشند، هم وجود دارند، مانند پروتئین P38398 که رتبه یک در لینک C-A دارد و در سرطان درگیر است.

یکی از واسطه‌ها، پروتئین P45877 است که هر سه مسیر، A-N، C-N و C-A را میانجی‌گری می‌کند. نام این پروتئین پپتیدیل- پرولیل سیس ترانس ایزومراز سی⁸ است. این پروتئین، سه نوع از پروتئین‌های بیماری مانند

Q8N9N5(C) ، Q9UMX0(C) ,O43765(C) و Q9NRR5(C,A,N) را میانجی‌گری می‌کند. ممکن است آثار غیرمستقیم بیماری‌ها به‌وسیله این پروتئین انجام شود. مقدار 55 درصد پروتئین‌های مونولایت انسانی، در بیماری درگیر می‌باشند. پروتئین P13569، یکی از پروتئین‌های مونولایت است که در هیچ‌یک از بیماری‌های بررسی‌شده در کار ما درگیر نیست. این پروتئین که به نام تنظیم‌کننده هدایت انتقال غشای فیروز کیستیک⁹ شناخته می‌شود، هر سه لینک A-N، C-N و C-A را میانجی‌گری می‌کند. این پروتئین انتقال یون‌های کلرید از راه غشای سلول را میانجی‌گری می‌کند [20].

یکی دیگر از پروتئین‌های مونولایت که رتبه M بالایی در هر سه لینک دارد، پروتئین P27797 است. این پروتئین کلر تیکولین¹⁰ نام دارد و یک چاپرون متصل به کلیسم است. این لکتین¹¹ با همه گلیکو پروتئین‌های مونوگلوکوزیله¹² که در شبکه آندوپلاسمی سنتز می‌شوند، تعامل دارد [21].

به‌تازگی مطالعات زیادی برای پیش‌بین پروتئین‌های مونولایت [22] و همچنین تحلیل شبکه بیماری‌های مختلف [23] انجام شده است ولی این پژوهش اولین مطالعه برای تعیین پروتئین‌های واسط در شبکه تعامل پروتئین-پروتئین با در نظر گرفتن بیماری‌های شایع در پروتئین‌های مونولایت است.

نتیجه‌گیری

نتایج کار نشان‌دهنده ارتباط بین نقش واسط و ویژگی‌های عملکردی این پروتئین‌ها است. مشخص شد که یک پروتئین که به‌عنوان واسط کلیدی غیرمستقیم بین دو بیماری عمل می‌کند، ممکن است در شبکه PPI، هاب

1 Transmembrane protein 17

2 Oral-facial-digital syndrome

3 Co-chaperone

4 Iron-sulfur cluster assembly

5 Pathway

6 Iron-sulfur cluster biosynthesis

7 Cofactor biosynthesis

8 Peptidyl-prolyl cis-trans isomerase C

9 Cystic fibrosis transmembrane conductance regulator

10 Calreticulin

11 Lectin

12 Monoglucosylated

[4]C. Chen et al. (2018). MoonProt 2.0: An expansion and update of the moonlighting proteins database. *Nucleic Acids Res*, 46(D1):D640{D644.

[5]C. J. Jeffery. (2015). Why study moonlighting proteins? *Frontiers in Genetics*, 6(JUN).

[6]NCBI: Genes and disease. National Library of Medicine (US), NCBI 2007.

[7]Kann MG: Protein interactions and disease: computational approaches to

[8]uncover the etiology of diseases. *Brief Bioinform* 2007, 8:333-346.

[9]Ideker T, Sharan R: Protein networks in disease. *Genome Res* 2008,

[10]18:644-652.

[11]Xu J, Li Y: Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics* 2006, 22:2800-2805.

[12]Thanh-Phuong Nguyen, Ferenc Jordán. A quantitative approach to study indirect effects among disease proteins in the human protein interaction network. *BMC Systems Biology*.

[13]Müller CB, Godfray HCJ: Indirect interactions in aphid-parasitoid communities. *Research on Population Ecology* 1999, 41:93-106.

[14]Jordán F, Liu WC, van Veen FJF: Quantifying the importance of species

[15]and their interactions in a host-parasitoid community. *Community Ecology* 2003, 4:79-88.

[16]Jeong H, Mason SP, Barabási AL, Oltvai ZN: Lethality and centrality in protein networks. *Nature* 2001, 411:41-42.

[17] Batada NN, Hurst LD, Tyers M: Evolutionary and physiological importance of hub proteins. *PLoS Comput Biol* 2006, 2(7):e88.

[18]Zotenko E, Mestre J, O'Leary DP, Przytycka TM: Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between

نباشد. بنابراین به پروتئین‌های واسط به علت مهم‌بودن نقش آنها مانند هاب‌ها باید توجه شود. همچنین مشاهده شد که پروتئین‌های هاب در شبکه PPI، ممکن است اهمیت توپولوژیکی کمی در واسطه‌گری آثار غیرمستقیم بین گروه‌های بیماری پروتئین‌ها در شبکه DPN داشته باشند. همچنین مشخص شد پروتئین‌هایی که اینتراکشن‌های غیرمستقیم را از راه پروتئین‌های بیماری مورد مطالعه، میانجی‌گری می‌کنند، به‌طور معمول در سلول، اهمیت عملکردی بالایی دارند. از مشاهده‌های دیگر می‌توان به این موضوع اشاره کرد که حتی اگر پروتئین‌های A و N به‌طور مستقیم به C متصل نمی‌باشند، پیوندهای غیرمستقیم غنی می‌تواند ایجاد شود.

تشکر و قدردانی: از دانشگاه تهران به دلیل تأمین منابع مالی تشکر می‌شود.

تأییدیه اخلاقی: موردی به‌وسیله نویسندگان گزارش نشده است.

تعارض منافع: هیچ‌گونه تعارض منافی وجود ندارد. سهم نویسندگان: فرشید شیرافکن (نویسنده اول)، نگارنده مقدمه/ پژوهشگر اصلی/نگارنده بحث (50%): سجاد قرقانی نویسنده دوم (روش‌شناس/ پژوهشگر اصلی/ نگارنده بحث 50 درصد)

منابع

[1] Daphne H.E.W. Huberts and Ida J. van der Klei. (2010). Moonlighting proteins: An intriguing mode of multitasking. *Biochimica et Biophysica Acta* 1803, page 520{525.

[2] C. J. Jeffery. (2018). Intracellular/surface moonlighting proteins. *Biochemistry and Molecular Biology*.

[3] J. Piatigorsky and G. J. Wistow. (1989). Enzyme/crystallins: Gene sharing as an evolutionary strategy. *Cell*, 57(2):197{199.

biogenesis with the CIAO1-mediated transfer to recipients. *Hum Mol Genet*.2018.

[24]Ostedgaard L.S., Baldursson O., Vermeer D.W., Welsh M.J., Robertson A.D. A functional R domain from cystic fibrosis transmembrane conductance regulator is predominantly unstructured in solution.

[25]Nauseef W.M., McCormick S.J., Clark R.A. .Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase.

[26]Shirafkan, F., Gharaghani, S., Rahimian, K. *et al.* Moonlighting protein prediction using physico-chemical and evolutionary properties via machine learning methods. *BMC Bioinformatics* **22**, 261 (2021).

[27]do Valle, I.F., Roweth, H.G., Malloy, M.W. *et al.* Network medicine framework shows that proximity of polyphenol targets and disease proteins predicts therapeutic effects of polyphenols. *Nat Food* **2**, 143–155 (2021).

the network topology and essentiality. *PLoS Comput Biol* 2008, 4(8):e1000140.

[19]Erickson, R. P., Bodensteiner, J. B. Oro-facial-digital syndrome IX

[20]with severe microcephaly: a new variant i n a genetically isolated population. *Am. J. Med . Genet.* 143A: 3309-3313, 2007.

[21]Diogo M Ribeiro et al. (2019). Moondb 2.0: an updated database of extreme multifunctional and moonlighting proteins. *Nucleic Acids Research*, 47(D1):D398{D402.

[22]Uhrigshardt H., Singh A., Kovtunovych G., Ghosh M., Rouault T.A.Characterization of the human HSC20, an unusual DnaJ type III protein, involved in iron-sulfur cluster biogenesis. *Hum Mol Genet*.2010.

[23]Ki Soon Kim, Nunziata Maio, Anamika Singh, Tracey A Rouault. Cytosolic HSC20 integrates de novo iron-sulfur cluster

Determination of intermediate proteins in the protein-protein interaction network considering common diseases in Moonlighting proteins

Farshid Shirafkan¹ , Sajjad Gharaghani^{*2}

1- PhD student, Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

2- Assistant Professor, Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

*corresponding author: s.gharaghani@ut.ac.ir

Received: 2021/9/21

Accepted: 2022/3/6

Abstract

Moonlight proteins are a subset of multifunctional proteins in which more than one independent or usually distinct function occurs in a single polypeptide chain. Analyzing the interactive networks of proteins in the cell makes it possible to understand how complex processes cause disease. With the help of systems biology, larger and more complex systems can be studied, and the molecular basis of several diseases can be considered. The proteins of the human organism that are moonlight are mostly involved in cancer, anemia, and neurodegeneration. In this work, we created a subnet according to the human PPI network, in which the nodes, the proteins that cause the three selected diseases, and the edges, are the connection of these proteins with each other. We measured the power of the indirect effects of non-disease mediators between the three disease groups and identified key disease-binding intermediate proteins. The results show the relationship between mediator role and centrality and between mediator role and functional properties of these proteins. We have shown that a protein that plays a key indirect mediator between two diseases is not necessarily a hub in the PPI network. Therefore, as hub proteins are considered, intermediate proteins should be considered. We have observed that the mediators between anemia and neurodegeneration diseases are functionally important in the cell. The mediator proteins suggested herein should be experimentally tested as hypothetical disease-related proteins.