

پیش‌گویی انعطاف‌پذیری قطعات پروتئینی، با استفاده از پروفایل انرژی

سیدشهریار عرب^{۱*}، مهدی صادقی^۲، چنگیز اصلاح‌چی^۳، حمید پزشکی^۴

۱- استاد یار بیوانفورماتیک، پژوهشگاه دانش‌های بنیادی (IPM)، پژوهشکده علوم کامپیوتر، گروه بیوانفورماتیک

۲- استاد یار بیوفیزیک، پژوهشگاه ملی تحقیقات ژنتیک و بیوتکنولوژی

۳- دانشیار ریاضی، دانشگاه شهید بهشتی، دانشکده علوم ریاضی

۴- استاد آمار، دانشگاه تهران، دانشکده ریاضی آمار و علوم کامپیوتر، قطب بیومت

*تهران صندوق پستی ۹۳۹۵-۵۷۴۶

shahriar.arab@gmail.com

(دریافت مقاله: ۸۹/۵/۱۲، پذیرش مقاله: ۸۹/۸/۴)

چکیده - در این مقاله، پیشگویی انعطاف‌پذیری قطعات پروتئینی، بر پایه توالی اسیدهای آمینه آن ارائه خواهد شد. برای تعریف فاکتور تشخیص میزان انعطاف‌پذیری، از انرژی دو اسید آمینه متوالی، استفاده شده است. با استفاده از تحلیل‌های آماری روی این انرژی، عددی بین صفر و صد به هر جفت اسید آمینه تعلق می‌گیرد که آن را عدد انعطاف‌پذیری می‌نامیم. در نظر گرفتن آثار همسایه‌ها تا چهار اسید آمینه، باعث بهبود پیش‌گویی انعطاف‌پذیری قطعات پروتئین می‌شود. برای ارزیابی پیش‌گویی انعطاف‌پذیری قطعات پروتئینی، از پروتئین‌هایی که با NMR تعیین ساختار شده بودند، استفاده شد. در روش NMR، عدم قطعیت در تعیین ساختار مناطقی که انعطاف‌پذیری بالایی دارند، مثل ابتدا و انتهای زنجیره و حلقه‌ها، باعث ایجاد مدل‌های مختلف ساختمانی می‌شود. مقایسه بین پیش‌گویی انعطاف‌پذیری و پراکندگی زوایا در مدل‌های مختلف، همبستگی بالایی را نشان می‌دهد که حاکی از دقت قابل اعتماد روش پیش‌گویی است.

کلیدواژه‌ها: انعطاف‌پذیری ساختمانی، پیشگویی عملکرد پروتئین، پیش‌گویی انعطاف‌پذیری قطعات پروتئین، پیشگویی ساختمان پروتئین

۱- مقدمه

ساختمان‌های شناخته شده پروتئین‌ها، محققین دریافته‌اند که برخی از پروتئین‌ها می‌توانند تغییر ساختار دهند. پروتئین‌های کروی، دارای حرکات داخلی زیادی هستند. این حرکات، از تغییر شکل نامنظم الاستیک کل پروتئین، به واسطه برخورد با مولکول‌های حلال، به وجود می‌آیند

عملکرد یک پروتئین غالباً توسط جزئیات خاص ساختمان سه بعدی شکل طبیعی پروتئین به دست می‌آید. در گذشته این باور وجود داشت که ساختمان طبیعی پروتئین، از نظر سطح انرژی، در پایین‌ترین سطح قرار دارد. با رشد

نظری انعطاف‌پذیری پروتئین مثل B-factors از کریستالوگرافی [۶]، پارامترهای نظم به‌دست آمده از اندازه‌گیری فروپاشی فلورسانس ناهمسان‌گردی^۵ [۷، ۸] و هم‌چنین پارامترهای نظم به‌دست آمده از دینامیک مولکولی (MD) [۹، ۱۰]، مقایسه شده است. مطالعات انجام شده، همبستگی بین پارامترهای نظم و ویژگی‌های مولکولی، مانند عناصر ساختمان دوم، حجم زنجیره‌های جانبی اسیدهای آمینه [۱۱، ۱۲] و توالی‌های حفظ‌شده اسیدهای آمینه [۱۳] را نشان می‌دهند. همبستگی بین تغییرات ساختاری، بین مدل‌های NMR و حرکت حرارتی تأیید شده است، این تغییرات با تعیین ساختار پیوسته، قابل شناسایی است [۱۲].

پیش‌گویی انعطاف‌پذیری، از روی توالی اسیدهای آمینه، بسیار مورد توجه است و در دهه اخیر، روش‌های مختلفی برای این پیش‌گویی ارائه شده است [۱۴-۱۹]. در این جا روشی برای پیش‌گویی انعطاف‌پذیری قطعات پروتئینی، براساس پروفایل انرژی جفت اسیدآمینه‌ها ارائه کرده‌ایم. روش موردنظر، برپایه قواعد ساده‌ای استوار است. پراکندگی انرژی آزاد دو اسید آمینه، به‌عنوان عاملی برای تشخیص میزان انعطاف‌پذیری به‌کار رفته است. پایگاه داده‌ای، از زوایای دایهدرال دو اسید آمینه پشت سر هم تهیه و انرژی این ساختارها با میدان نیروی Amber99 محاسبه شد.

با استفاده از تحلیل‌های آماری، سطح انرژی دی‌پپتیدهای موجود در پایگاه داده، عددی نرمال بین صفر و صد، به هر جفت اسیدآمینه، اختصاص یافت که آن‌را عدد انعطاف‌پذیری می‌نامیم. هرچه این عدد به صد نزدیک‌تر باشد، احتمال این‌که دی‌پپتید در قطعه‌ای با انعطاف بالا باشد، بیش‌تر است؛ و برعکس هرچه این عدد

[۱]. بررسی حرکات داخلی، به وسیله‌ی روش‌های تجربی مختلفی، مثل تحلیل تفرق اشعه‌ی ایکس [۲]، NMR و شبیه‌سازی کامپیوتری [۱، ۳، ۴] انجام می‌شود. تحرکات در مقیاس بزرگ‌تر، مثل حرکات لولایی، تغییرات آلوستریک، تغییرات ساختاری پروتئین‌های موتور و غیره، به‌صورت گسترده شناخته شده‌اند که این حرکات نقش مهمی را در عملکرد پروتئین بازی می‌کنند.

گرچه پروتئین‌ها می‌توانند ساختمان‌های متفاوتی داشته باشند ولی برخی قطعات که به‌عنوان قطعات حفظ شده می‌توان از آن‌ها یاد کرد، در هیچ کدام از آرایش‌های فضایی تغییری نمی‌کنند. قطعات پروتئین را براساس تغییرات آرایش فضایی، می‌توان به دو دسته تقسیم کرد: مناطق سخت^۱ و مناطق با انعطاف‌پذیری بالا^۲. از جمله مناطقی که از انعطاف‌پذیری بالا برخوردار است مناطق پیونددهنده انعطاف‌پذیر^۳ است که وظیفه اتصال مناطق سخت و حفظ شده را برعهده دارد. بنابراین شناخت مناطق با انعطاف‌پذیری کم و یا زیاد، می‌تواند جزئیات بیش‌تری از ساختمان سه بعدی پروتئین‌ها را آشکار سازد و هم‌چنین می‌تواند اطلاعاتی درمورد مراحل تا خوردگی و چگونگی آن ارائه کند. دومین‌ها یا مناطقی که تا خوردگی مستقل دارند، عموماً با مناطقی با انعطاف‌پذیری بالا، به یک‌دیگر متصل می‌شوند. شناسایی مناطق انعطاف‌پذیر پروتئین‌ها، می‌تواند اطلاعاتی چند، در خصوص مناطقی که به‌صورت مستقل تا می‌خورند در اختیار قرار دهد.

برخی از محققین، برای مطالعه بیش‌تر انعطاف‌پذیری آرایش ساختمانی پروتئین‌ها، از داده‌هایی که از نتایج NMR در دسترس است استفاده می‌کنند [۵]. پارامترهای نظم^۴ به‌دست آمده از NMR، با دیگر معیارهای تجربی و

1. Rigid
2. Flexible regions
3. Linkers
4. Order Parameter

شد و سپس زنجیره جانبی به آن اضافه شد. برای یافتن بهترین آرایش شاخه جانبی، شاخه اصلی ثابت نگاه داشته شاخه جانبی از نظر انرژی بهینه شد.

۲-۳- ساختن پروفایل انرژی

انرژی دو اسید آمینه متوالی، به عنوان عاملی برای تشخیص انعطاف پذیری در قطعات پروتئین استفاده شد. انرژی تمام دی پپتیدهای بازسازی شده، در پایگاه داده، با استفاده از میدان نیروی Amber99 محاسبه شد و این داده‌ها به دیتابیس اضافه شد. پراکندگی انرژی دی پپتید، ما را برای تشخیص مناطق با انعطاف پذیری بالا راهنمایی می‌کند. این مقادیر پراکندگی انرژی در مناطقی که انعطاف پذیری بیشتری دارند، بیش تر است و در مناطق حفظ شده کم تر است. عدد انعطاف پذیری F^2 ، برای هر دی پپتید چنین تعریف می‌شود:

$$F = \text{StDev}(E) = \sqrt{\frac{\sum (E - \bar{E})^2}{N - 1}}$$

E ، انرژی هر آرایش فضایی، \bar{E} ، میانگین انرژی و N ، تعداد رخداد دی پپتید در دیتابیس است. هرچه این عدد به ۱۰۰ نزدیک تر باشد، احتمال حضور دی پپتید در قطعه‌ای با انعطاف پذیری بالا بیش تر است. از طرف دیگر هرچه این عدد به صفر نزدیک تر باشد، احتمال این که قطعه‌ای که این دی پپتید در آن قرار دارد از تحرک کم تری برخوردار باشد بیش تر می‌شود.

۲-۴- پیش گویی انعطاف پذیری

برای پیش گویی انعطاف پذیری، به هر اسید آمینه، با توجه به اسید آمینه بعدی، یک عدد اولیه انعطاف پذیری نسبت

کوچک تر باشد، احتمال این که این قطعه، عضوی از یک منطقه حفظ شده باشد، بالاتر است. در نظر گرفتن اثر همسایه‌ها تا چهار اسید آمینه، سبب بهبود پیش گویی انعطاف پذیری قطعات پروتئینی شد.

۲- مواد و روش‌ها

۲-۱- مجموعه داده پروتئینی

مجموعه داده‌ای شامل ۱۰۸۱۲ پروتئین از سایت بانک اطلاعات پروتئین RCSB- [۲۰]، انتخاب شد. در این مجموعه، پروتئین‌هایی انتخاب شدند که تعیین ساختار آنها به وسیله کرسنالوگرافی اشعه‌ی ایکس، با درجه تفکیک کم تر از ۲/۵ آنگسترم ممکن بوده هم چنین برای جلوگیری از حضور پروتئین‌های مشابه در بانک، پروتئین‌هایی که بیش از نود درصد مطابقت در توالی داشتند نیز از بانک کنار گذاشته شدند.

۲-۲- پایگاه داده‌ای از آرایش‌های فضایی

پایگاه داده‌ای از زوایای دیهدرال، برای هر جفت اسید آمینه پشت سرهم ساخته شد. در این پایگاه، داده زوایای دیهدرال همه پروتئین‌های انتخاب شده، از بانک اطلاعات پروتئینی استخراج شد. به این منظور، از برنامه DSSP استفاده شد. در این پایگاه داده، ما اطلاعات آرایش‌های فضایی برای هر جفت اسید آمینه، شامل آرایش‌های فضایی مختلف و هم چنین تعداد رخداد این آرایش‌ها را داریم. در هر جفت اسید آمینه، سه زاویه دیهدرال در رشته اصلی وجود دارد که آرایش فضایی این دی پپتید را تعیین می‌کند. زاویه ψ در اسید آمینه اول و زوایای ϕ و ψ از اسید آمینه دوم. پس از آن، به ازای همه زوایای موجود در بانک، رشته اصلی^۱ آرایش‌های فضایی مربوطه بازسازی

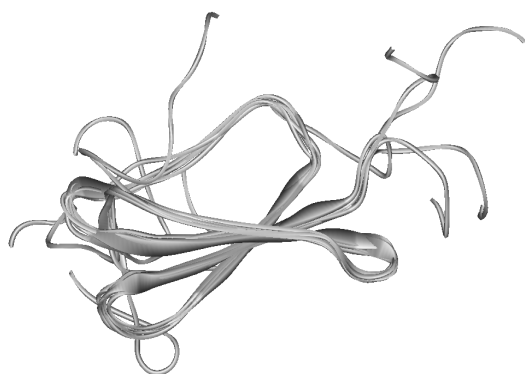
1. Backbone

2. Flexibility number

پراکندگی زوایای رشته اصلی در مجموعه پروتئین‌ها، بین ۰/۲۸ تا ۰/۷۱ است که معرف همبستگی بالایی این دو پارامتر است. بر همین اساس می‌توان از این معیار برای پیش‌گویی انعطاف‌پذیری استفاده کرد.

شکل ۱ پنج مدل اول ساختمان سه بعدی پروتئین Anastellin (IQ38) را نشان می‌دهد. شکل ۲ نمودار پراکندگی زوایا را به ازای هر اسیدآمینه و هم‌چنین پیش‌گویی این قطعات را در همان پروتئین، نشان می‌دهد. همبستگی بالایی که بین نمودار پیش‌گویی و نمودار پراکندگی زوایا وجود دارد، شاهدی بر دقت بالای پیش‌گویی به حساب می‌رود. این نمودار، مناطق سخت و هم‌چنین مناطقی که از انعطاف‌پذیری بالایی برخوردارند را به خوبی نشان می‌دهد.

مناطق پیونددهنده انعطاف‌پذیر^۳، قسمت‌های اساسی هستند که نقش حفظ انعطاف‌پذیری و فعالیت پروتئین را بر عهده دارند. همان‌طور که در شکل‌های ۱ و ۲ نشان داده شده است، نقطه‌اوج منحنی انعطاف‌پذیری با این مناطق پیونددهنده تا حد قابل‌قبولی منطبق است.



شکل ۱ ساختمان سه بعدی پروتئین IQ38 که به‌روشنی NMR تعیین ساختار شده است. برای نمایش بهتر، ۵ مدل اول پروتئین نشان داده شده است.

3. Linker

می‌دهیم. این عدد از پروفایل انرژی ۴۰۰ تایی جفت اسیدآمینه‌ها، که در بخش پیشین، نحوه تهیه آن بیان شد، استخراج می‌شود. برای اعمال آثار همسایه‌های هر اسید آمینه در طول توالی، پنجره‌های ۵ تایی اسیدآمینه در نظر گرفته می‌شود و میانگین اعداد انعطاف‌پذیری در این پنجره را به اسیدآمینه وسطی نسبت می‌دهیم.

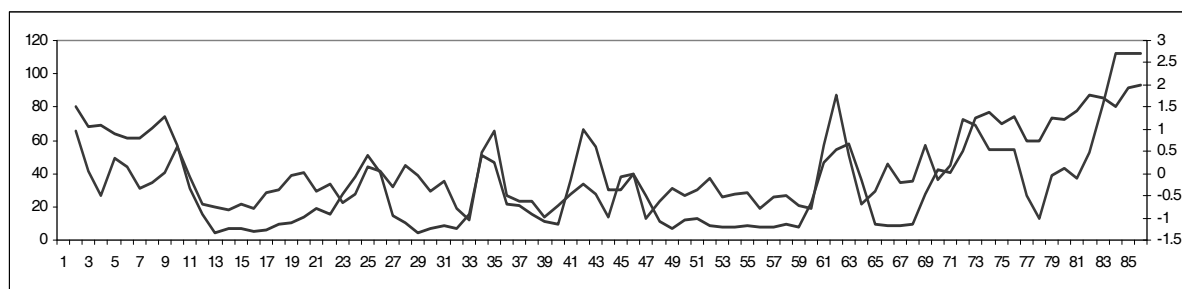
۲-۵- ساخت دیتابیس پراکندگی زوایای رشته اصلی

برای بررسی این‌که آیا پیش‌گویی انجام شده، با پراکندگی زوایای سازنده رشته اصلی همبستگی دارد یا خیر، پروتئین‌هایی که ساختمان آن‌ها توسط طیف‌سنجی NMR تعیین شده بودند و بیش از ۲۰ مدل ساختمانی داشتند، انتخاب شد برنامه‌ای نوشته شد که مدل‌های تعیین ساختار شده با NMR خوانده شود و زوایای رشته اصلی در آن مدل محاسبه و در دیتابیس ذخیره شود. با اجرای این برنامه روی کل مجموعه پروتئین‌های انتخاب شده دیتابیس پراکندگی، زوایای رشته اصلی ساخته شد. این دیتابیس شامل تمام زوایای هر اسید آمینه در مدل‌های متفاوت است. زوایای رشته اصلی، در مناطق با انعطاف‌پذیری پایین^۱، بسیار حفظ شده است و برعکس در مناطقی که انعطاف‌پذیر است این زوایا مقادیر متفاوتی دارند. انحراف معیار^۲ مقادیر زوایا، معیاری مناسب برای تعیین انعطاف‌پذیری رشته اصلی است.

۳- نتایج و بحث

همبستگی پیش‌گویی انعطاف‌پذیری با پراکندگی زوایای رشته اصلی: همبستگی بین مقادیر انعطاف‌پذیری و

1. Rigid Regions
2. Standard deviation



شکل ۲ نمودار پیش‌گویی انعطاف‌پذیری و انحراف معیار زوایای دیهدرال در پروتئین 1Q38. نمودار آبی، پیش‌گویی انعطاف‌پذیری و نمودار قرمز، انحراف معیار زوایای دیهدرال در مدل‌های مختلف است.

measurements and dynamic simulations. *Methods Enzymol* 1986, 131:362-389.

- [2] Ringe D, Petsko GA: Study of protein dynamics by X-ray diffraction. *Methods Enzymol* 1986, 131:389-433.
- [3] Opella SJ: Protein dynamics by solid state nuclear magnetic resonance. *Methods Enzymol* 1986, 131:327-361.
- [4] Wagner G, Wuthrich K: Observation of internal motility of proteins by nuclear magnetic resonance in solution. *Methods Enzymol* 1986, 131:307-326.
- [5] G.Nodet, D.Abergel: An overview of recent developments in the interpretation and prediction of fast internal protein dynamics. *European Biophysics Journal* 2007, 36:985-993.
- [6] R.Powers, G.M.Clore, D.S.Garrett, A.M.Gronenborn: Relationships between the precision of high-resolution protein NMR structures, solution-order parameters and crystallographic B factors.

دانستن انعطاف‌پذیری قطعات پروتئینی، کمک بزرگی به بهبود بخشیدن به کیفیت پیش‌گویی ساختمان دوم، ساختمان سوم و میزان در دسترس بودن حلال می‌کند. این اطلاعات نیازمند دانستن جزئیات زیادی از ساختمان پروتئین است و به همین دلیل برای تعداد کمی از پروتئین‌ها این اطلاعات در دسترس است.

به این منظور، ما روشی را برای پیش‌گویی مناطق انعطاف‌پذیر از روی توالی پروتئین ارائه کردیم. همان‌طور که نشان داده شد اسیدهای آمینه با انعطاف‌پذیری بالا توسط این روش، قابل پیش‌گویی و شناسایی است و همبستگی قابل قبولی هم با میزان انعطاف‌پذیری مشاهده شده در پروتئین‌ها از دارد. این روش نسبت به اسیدهای آمینه موجود در مناطق پیونددهنده که نقش پیوند و تنظیم موقعیت نسبی دومین‌ها را برعهده دارند، بسیار حساس است.

۴- تشکر و قدردانی

این مطالعه توسط گرانت CS1385-1-02، از پژوهشگاه دانش‌های بنیادی (IPM) مورد حمایت قرار گرفته است.

۵- مراجع

- [1] Dobson CM, Karplus M: Internal motion of proteins: nuclear magnetic resonance

- derived backbone order parameters. *J Mol Biol* 2000, 295:963-978.
- [12] Anderson CAF, Palmer AG, Brunak S, Rost B: Continuum secondary structure captures protein flexibility. *Structure* 2002, 10:175-184.
- [13] Mittermaier A, Davidson AR, Kay LE: Correlation between ^2H NMR side-chain order parameters and sequence conservation in globular proteins. *J Am Chem Soc* 2003, 125:9004-9005.
- [14] Iakoucheva LM, Dunker AK: Order, disorder, and flexibility: Prediction from protein sequence. *Structure* 2003, 11:1316-1317.
- [15] Yuan Z, Bailey TL, Teasdale RD: Prediction of protein B-factor profiles. *Proteins-Structure Function and Bioinformatics* 2005, 58:905-912.
- [16] Chen K, Kurgan LA, Ruan JS: Prediction of flexible/rigid regions from protein sequences using k-spaced amino acid pairs. *Bmc Structural Biology* 2007, 7.
- [17] Dosztanyi Z, Sandor M, Tompa P, Simon I: Prediction of protein disorder at the domain level. *Current Protein & Peptide Science* 2007, 8:161-171.
- [18] Trott O, Siggers K, Rost B, Palmer AG: Protein conformational flexibility prediction using machine learning. *Journal of Magnetic Resonance* 1993, 101:325-327.
- [7] A.G.Palmer, R.Hochstrasser, D.P.Millar, M.Rance, P.E.Wright: Side chain dynamics of a zinc finger peptide characterized by ^{13}C NMR relaxation measurements and fluorescence anisotropy decay. *Journal of the American Chemical Society* 1992, 115:6333-6345.
- [8] Moncrieffe MC, Juranic N, Kemple MD, Potter JD, Macura S, Prendergast FG: Structure-fluorescence correlations in a single tryptophan mutant of carp parvalbumin: solution structure, backbone and side-chain dynamics. *J Mol Biol* 2000, 297:147-163.
- [9] Philippopoulos M, Mandel AM, Palmer AG, III, Lim C: Accuracy and precision of NMR relaxation experiments and MD simulations for characterizing protein dynamics. *Proteins* 1997, 28:481-493.
- [10] D.C.Chatfield, A.Szabo, B.R.Brooks: Molecular dynamics of staphylococcal nuclease: comparison of simulation with ^{15}N and ^{13}C NMR relaxation data. *Journal of the American Chemical Society* 1998, 120:5301-5311.
- [11] Goodman JL, Pagel MD, Stone MJ: Relationships between protein structure and dynamics from a database of NMR-

- [20] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: The Protein Data Bank. *Nucleic Acids Res* 2000, 28:235-242.
- Journal of Magnetic Resonance 2008, 192:37-47.
- [19] Schlessinger A, Yachdav G, Rost B: PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics* 2006, 22:891-893.