بردار کشیدگی به عنوان پارامتری برای تشخیص آبدوستی پروتئین

قنبری ممان لیلا*[۱]، دیده ور فرزاد[۲]، مینوچهر زرّین**[۳]، مددکارسبحانی آرمین[۴]، نماینده راحله[۵]، نجفی عرب هاجر[۶]**

[۱]دانشجوی کارشناسی ارشد دانشگاه صنعتی امیرکبیر، l_ghanbari_62@aut.ac.ir
[۲]عضو هیئت علمی دانشگاه صنعتی امیرکبیر، didehvar@aut.ac.ir
[۳]عضو هیئت علمی پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری،.minuchehr@nigeb.ac.ir
[۴]عضو هیئت علمی انستیتو بیوشیمی-بیوفیزیک دانشگاه تهران armin@ibb.ut.ac.ir
[۵]دانشجوی کارشناسی دانشگاه صنعتی امیرکبیر namayande@aut.ac.ir
[۶]دانشجوی کارشناسی دانشگاه صنعتی امیرکبیر hajar.najafi@aut.ac.ir

چکیده

در سال ۲۰۰۵ میلادی مجله Science پیش بینی ساختار پروتئین را به عنوان یکی از ۱۲۵ مسئله سخت حل نشده در علم مطرح کرد(۱). در سال های اخیر برای حل چنین مسئله ای، راهکارها و پیشنهادات بسیار زیاد و متنوعی ارائه شده و از دیدگاه های متفاوتی مورد بررسی قرار گرفته است. در زمینه هندسه نیز روش های متفاوتی برای حل مسئله بیان شده که از آن ها در مسئله پیش بینی ساختار پروتئین ها و مقایسه آن ها با یکدیگر استفاده شده است. تقریب پروتئین با بیضی گون یکی از این موارد می باشد(۲). در این مقاله، ابتدا برداری به نام "بردار کشیدگی"[۱] معرفی شده است که توسط آن می توان شکل پروتئین را تقریب زده و به مقایسه آن ها با یکدیگر بر اساس این بردار پرداخت. سپس اثر خصوصیت آبدوستی پروتئین در یکی از پارامترهای این بردار مورد بررسی قرار گرفته است. در واقع به کمک روش پیشنهادی تنفورد[۲] (۳) و بر اساس استاندارد کایت[۳]-دولیتل[۴] kd (۴) برای پروتئین ها خصوصیت آبدوستی تعریف کرده و اثر آن را بر بردار کشیدگی بررسی کردیم. نتایج بدست آمده و تحلیل های انجام شده در این زمینه بیانگر آن است که هرچه پروتئین آبدوست تر باشد، پارامتر مربوطه در بردار کشیدگی، مقدار بیشتری داشته و هرچه آبگریزتر باشد، این پارامتر مقدار کمتری خواهد داشت. از ویژگی های این بردار می توان به جامعیت آن نسبت به مدل بیضی گون اشاره کرد. در ضمن انعطاف بیشتری در نوع تقریب های غیر متقارن نیز دارد.

واژه‌های کلیدی: ساختار پروتئین، آبدوستی پروتئین، بردار کشیدگی.

---

[1] Elongation Vector
[2] Tanford
[3] Kyte
[4] Doolittle

مقدمه

از سال ۱۷۴۷ که پروتئین ها و نقش اساسی آن ها در سیستم های زنده کشف شد، بسیاری از محققان به تلاش بر روی عملکرد آن ها پرداختند. در بین موضوعات مطرح شده در این زمینه، مسئله مقایسه ساختار پروتئین ها به عنوان یکی از مهم ترین و اساسی ترین مسائل شناخته شد؛ چراکه عملکرد یک پروتئین به ساختار و شکل سه بعدی آن مربوط می شود و شاید بتوان گفت که پروتئین های با ساختار مشابه، عملکردهای مشابه نیز دارند. در نتیجه ارائه راه کارهای دقیق تر برای مقایسه این ساختارها در زمینه بیوانفورماتیک، به عنوان یکی از اهداف مهم برشمرده می شود. روش های بسیار متنوعی برای این کار پیشنهاد شده است که تکنیک تقریب پروتئین با بیضی گون یکی از متداول ترین آن ها است و در آن به جای پروتئین، از شکل تقریبی برای مقایسه استفاده می شود(۲). روشی که در این مقاله معرفی شده است، جامع تر از تخمین با بیضی گون بوده و در ضمن انعطاف بیشتری نیز دارد چراکه توانایی آن را دارد که ساختارهای نامتقارن را با شکل های نامتقارن تقریب بزند. در این مقاله به معرفی "بردار کشیدگی" پرداخته و اثر خاصیت آبدوستی و آبگریزی پروتئین بر روی پارامتر اول آن، بررسی می شود.

مواد و روش‌ها

بردار کشیدگی

این بردار درواقع یک بردار سه تایی یکه به صورت $EV = (EP_1, EP_2, EP_3)$ می باشد که از آن برای معرفی شکل سه بعدی یک پروتئین و تقریب شکل محدب آن با در نظر گرفتن پراکندگی اسیدآمینه های پروتئین، می توان استفاده کرد. در ادامه نحوه محاسبه این بردار در۴ مرحله بیان می شود:

۱. بزرگترین راستای کشیدگی را به کمک مختصات اتم های پروتئین بدست آورده (این راستا، راستای خط رگرسیون سه بعدی اتم های پروتئین می باشد) و آن را R_1 می نامیم. سپس بزرگترین فاصله بین سایه عمود تمام اتم ها بر روی آن را l_1 می نامیم.

۲. راستای R_1 (که در مرحله قبل محاسبه شد) را به عنوان نرمال صفحه ای مثلP در نظر می گیریم. سپس تصویر تمامی اتم ها را بر رویP بدست آورده و راستای خط رگرسیون سه بعدی نقاط تصویر شده را R_2 می نامیم (در نظر داشته باشید که برای مستقل بودن سه درایه بردار کشیدگی باید راستای آن ها بر هم عمود باشد؛ در اینجا نیز هر راستایی در صفحه P بر R_1 عمود است). با محاسبه بزرگترین فاصله بین سایه عمود تمام اتم ها بر روی آن، l_2 بدست می آید.

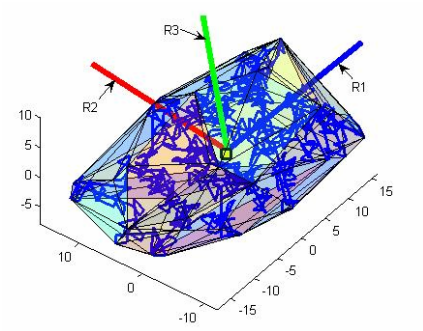
۳. برای بدست آوردن راستای درایه سوم فقط کافی است که از حاصلضرب خارجی R_1 و R_2 استفاده شود و به روش مشابه l_3 نیز محاسبه گردد.

۴. برای آن که تأثیر اندازه پروتئین ها بر محاسبات از بین برود به طوری که مقیاس خوبی برای مقایسه درست آن ها در دست داشته باشیم، بردار $L = (l_1, l_2, l_3)$ را یکه کرده و آن را "*بردارکشیدگی*" (EV) می نامیم.

با توجه به نوع تعریف و استفاده از مفهوم رگرسیون در آن به راحتی می توان خصوصیات ذیل را برای بردارکشیدگی بیان کرد:

- $EP_1 \geq EP_2 \geq EP_3 \geq 0$
- $\sqrt{EP_1^2, EP_2^2, EP_3^2} = 1$

در شکل۱ نمونه ای از سه راستای کشیدگی برای پروتئین 1abz را ملاحظه می کنید.



شکل۱: سه راستای کشیدگی (R_1, R_2, R_3) برای پروتئین 1abz با بردار کشیدگی (۰,۳۸و۰,۵۳و۰,۷۵).

درجه آبدوستی پروتئین

برای بررسی این بردار و تغییرات مربوط به آن در انواع پروتئین ها، از مفهومی به نام درجه آبدوستی پروتئین استفاده می کنیم. همان طور که می دانیم، برای اسیدآمینه ها خصوصیتی تحت عنوان "آبدوستی" مطرح می شود که بیان کننده میزان تمایل هرکدام برای ایجاد پیوند با آب می باشد. استانداردهای متفاوتی نیز برای اندازه گیری آن وجود دارد از جمله روش های کایت-دولیتل (kd) (۴)، ولفندن[5] و همکاران(۵). در این مقاله از روشkd برای محاسبه درجه آبدوستی اسیدآمینه ها استفاده می شود. در این روش، هرچه درجه نسبت داده شده به اسیدآمینه ای منفی تر باشد، نشان دهنده آن است که آن اسیدآمینه آبدوست تر است و هرچه این مقدار مثبت تر باشد، اسیدآمینه آبگریزتر است. برای پروتئین ها نیز روش های متنوعی جهت کمینه کردن این خصوصیت ارائه شده است. به عنوان مثال تنفورد (۳) از میانگین درجه آبدوستی برای تقریب زدن این مقدار استفاده می کند. در انجام محاسبات، ما نیز از همین روش برای محاسبه آبدوستی پروتئین ها (که آن راHP می نامیم) استفاده می کنیم. به دلیل استفاده از روش kd برای محاسبهHP، درمورد پروتئین ها نیز هرچه HPکمتر باشد، آبدوست تر است و هرچه HP بیشتر آبگریز تر.

نتایج و بحث

بررسی اثر HP بر EP_1

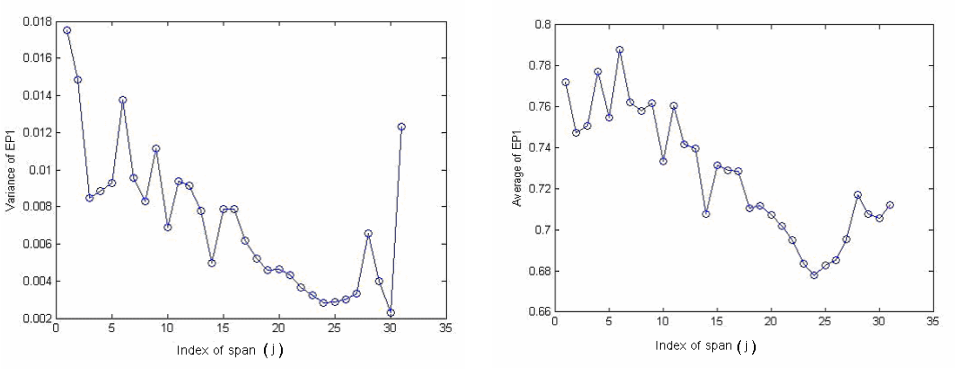
در این مقاله، به جای بررسی خود پروتئین ها، از دامنه ها[6] استفاده می کنیم. دامنه های پروتئینی در واقع سازنده ساختار پروتئین و عملکرد آن می باشند(۶،۷) این واحدها که مستقل از هم تا می خورند، معمولا منجر به عملکردهای بیولوژیکی متفاوتی می شوند. یکی از مجموعه دامنه هایی که بر اساس ساختار جمع آوری شده است، مجموعه داده CATH (۸،۹) می باشد که در این مقاله از آخرین نسخه آن (نسخه ۳,۲) استفاده شده است.

طبق روش ارائه شده، بردار کشیدگی و درجه آبدوستی برای تمام ۸۸۷۱ نمونه دامنه پروتئینی موجود در CATH محاسبه شد. کمترین و بیشترین میزان HP در بین پروتئین ها را بدست آورده و بازه $[\min(HP_i), \max(HP_i)]$را در نظر گرفتیم و آن را به زیربازه های کوچک تقسیم کردیم. پروتئین هایی که درجه آبدوستی آن ها در هر زیربازه مثل زیر بازه $[a_j, a_{j+1}]$ قرار می گیرد را یافته و میانگین EP_1 را برای هر زیربازه محاسبه می کنیم. به عنوان مثال شکل۲ نمودار مربوط به مقدار میانگین و مقدار واریانس EP_1 در ۳۱ زیربازه را نشان می دهد. همانطور که در این شکل نیز مشاهده می شود، به طورکلی از چپ به راست (ازآبدوست به آبگریز)، مقدار میانگین EP_1 کاهش یافته و مقدار واریانس نیز کاهش می یابد.

---

[5] Wolfenden

[6] Domains



ب **الف**

شکل۲: (الف) نمودار مقدار میانگین EP_1 در هر زیربازه. **(ب)** نمودار مقدار واریانس EP_1 در هر زیربازه.

مقدار کم میانگین و واریانس EP_1 درپروتئین های آبگریز نشان دهنده آن است که تنوع ساختاری از نظر بردار کشیدگی در پروتئین های آبگریزتر کمتر است. از طرفی به دلیل بیشتر بودن میانگین و واریانس EP_1 در پروتئین های آبدوست، می توان نتیجه گرفت که این پروتئین ها در کل تنوع ساختاری بیشتری داشته و کشیدگی اصلی آن ها بیشتراست.

کاربردهای بردارکشیدگی و کارهای آتی

بردارکشیدگی درواقع اطلاعاتی راجع به نوع ساختار و کلیات شکل محدب سه بعدی آن در اختیار ما قرار می دهد، که به نوعی جامع تر از تخمین با بیضی گون بوده و درضمن انعطاف بیشتری نسبت به آن دارد؛ چراکه با در نظرگرفتن مرکز ثقل اسیدآمینه ها، می توان برای پروتئین هایی که ساختار متقارن ندارند، توسط این بردار، شکل نامتقارنی تقریب زد. در حالی که این نوع تقریب از عهده مدل های مبتنی بر بیضی گون خارج است. از این بردار می توان در مدل هایی که بر اساس بیضی گون عمل می کنند(۱۰) استفاده کرد و اطلاعات با دقت بیشتری را بدست آورد.

تشکر و قدردانی

با تشکر و سپاس فراوان از دانشگاه صنعتی امیرکبیر و پژوهشگاه ملی و مهندسی ژنتیک و زیست فناوری، کـه ایـن تحقیـق را حمایت کرده و امکانات و شرایط لازم برای انجام آن را در اختیار قرار دادند.

منابع

1. 'So Much More to Know …', *Science*, Vol. 309, no. 5731, July 1, 2005 2005, pp. 78b-102.
2. S. E. Harding, J. C. Horton and H. Colfen, 'The ELLIPS suite of macromolecular conformation algorithms', *Eur Biophys J*, Vol. 25, no. 5-6, 1997, pp. 347-59.
3. Charles Tanford, 'Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins', *Journal of the American Chemical Society*, Vol. 84, no. 22, 1962, pp. 4240-7.
4. J. Kyte and R. F. Doolittle, 'A simple method for displaying the hydropathic character of a protein', *J Mol Biol*, Vol. 157, no. 1, May 5 1982, pp. 105-32.
5. R. Wolfenden, L. Andersson, P. M. Cullis and C. C. Southgate, 'Affinities of amino acid side chains for solvent water', *Biochemistry*, Vol. 20, no. 4, Feb 17 1981 ,pp. 849-55.
6. Christine Vogel, Matthew Bashton, Nicola D. Kerrison, Cyrus Chothia and Sarah A. Teichmann, 'Structure, function and evolution of multidomain proteins', *Current Opinion in Structural Biology*, Vol. 14, no. 2, 2004, pp. 208-16.
7. C. P. Ponting and R. R. Russell, 'The natural history of protein domains', *Annu Rev Biophys Biomol Struct*, Vol. 31, 2002, pp. 45-71.
8. C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells and J. M. Thornton, 'CATH - a hierarchic classification of protein domain structures', *Structure*, Vol. 5, no. 8, 1997, pp. 1093-108.
9. C. A. Orengo, F. M. Pearl and J. M. Thornton, 'The CATH domain structure database', *Methods Biochem Anal*, Vol. 44, 2003, pp. 249-71.

10. P. Deschavanne and P. Tuffery, 'Exploring an alignment free approach for protein classification and structural class prediction', *Biochimie*, Vol. 90, no. 4, Apr 2008, pp. 615-25.

Abstract

In 2005, Science named the protein folding problem one of the biggest unsolved problems in science (1). In recent years, there are many approaches to this problem and researchers proposed many ideas for it. In Geometry there are different methods for simplifying the problem. One of these methods is approximating the protein structure with ellipsoid and sphere. We introduce a new method for approximating this feature with "Elongation Vector", which would be used for protein classification. In this paper we analyzed the relation between one of the parameters of this vector and the hydrophobicity of proteins (according to Tanford (3) definition and Kyte-Doolittle method (4)). Results show that the parameter of the vector in hydrophobe proteins is smaller than the hydrophil proteins in average. Therefore we can use this vector in order to determine the hydrophobicity degree of proteins. One of the advantages of this vector over ellipsoid model is the Elongation Vector which is more general and more flexible than ellipsoid model.