



STON: روشی جدید برای مقایسه ساختمان سه بعدی

پروتئینها

عرب سید شهریار^{1*}، اصلاحچی چنگیز^{2,3*}، صادقی مهدی^{2 و 4}، پزشکی حمید^{2 و 5}، رحیمی امیر مسعود²، معبودی افخم حیدر³

1. دانشگاه تهران، مرکز تحقیقات بیوشیمی و بیوفیزیک IBB، گروه بیوانفورماتیک
2. پژوهشگاه دانشهای بنیادین (IPM)، پژوهشگاه علوم کامپیوتر، هسته بیوانفورماتیک
3. دانشگاه شهید بهشتی، دانشکده علوم ریاضی
4. پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری
5. دانشگاه تهران، پردیس علوم، دانشکده ریاضی آمار و علوم کامپیوتر، گروه آمار

چکیده

مقایسه ساختمانی پروتئینها یکی از مسائل مهم در بیوانفورماتیک می باشد که کاربردهای زیادی در مطالعه ژنومیک ساختمانی و عملکردی دارد. در طول دهه های گذشته، روشهای مکاشفه ای زیادی برای حل مسئله مقایسه ساختمان پروتئینها ابداع شده اند. بیشتر روشهای مقایسه ساختمان پروتئین، انطباقی براساس کمترین RMSD را می دهند که از خیلی تطابق های محلی که ممکن است برای مطالعات تکاملی یا شناخت فعالیت آن مهم باشد صرف نظر می کنند. ما برای پیدا کردن تطابق باقیمانده های اسید آمینه ای با RMSD مطلوب الگوریتمی جدید را ارائه کردیم. که با تغییر در مقادیر پارامترهای فاصله و گام زاویه چرخش در برنامه این امکان را برآورده می سازد که در بین قطعات دو پروتئین بتوان جستجو را برای قطعات با شباهت بالا و یا با شباهت کمتر نیز انجام داد.

کلمات کلیدی: مقایسه ساختمان پروتئین؛ انطباق؛ انتقال جسم صلب؛ زاویه دو سطحی

مقدمه

یکی از روشهای اساسی که به درک خواص فیزیکی، شیمیایی و زیستی یک پروتئین کمک می کند مقایسه ساختمانی پروتئینها می باشد که براساس تشابهات و اختلافات بین ساختمان سه بعدی پروتئین ها عمل می کند. از آنجایی که ساختمان پروتئینها در طول تکامل نسبت به توالی آنها بیشتر حفظ شده اند، بنابراین مقایسه ساختمانی نسبت به مقایسه ای که بر پایه آنالیز توالی است اطلاعات بیشتری را درباره تاریخچه تکامل می دهد. در حال حاضر در حدود 46000 ساختمان پروتئین در بانک اطلاعات پروتئین PDB موجود می باشد (Sep 2007) و تعداد ساختمانی جدید به سرعت رو به افزایش است. برای طبقه بندی ساختمانی شناخته شده تلاشهای بسیاری انجام شده است که منجر به ایجاد بانکهای اطلاعاتی ساختمانی همچون SCOP (1)، CATH (2) و FSSP (3) گردیده است. مقایسه ساختمانی جدید بدست آمده با ساختمانی طبقه بندی شده موجود، به تشخیص خانواده پروتئین، روند تکاملی و فعالیت پروتئین کمک می کند. مقایسه ساختمانی پروتئین یک مسئله NP-سخت می باشد بنابراین در این سه ده گذشته روشهای مکاشفه ای برای حل این مسئله ابداع گردید (4-6). مسئله انطباق ساختمان پروتئین ها علیرغم وجود روشهای فراوان، هنوز حل نشده است و نتایج روشهای مختلف با یکدیگر تفاوت دارد (4 و 7 و 8). دو مشخصه که برای انتخاب بهترین انطباق بین دو پروتئین استفاده می شود عبارتند از RMSD (جذر میانگین توان دوم اختلاف ها) و تعداد جفت اسید آمینه ها که با یکدیگر منطبق شده اند. در این



مقاله، ما الگوریتمی را برای پیدا کردن اسیدهای آمینه متناظر در انطباق دو پروتئین با یک مقدار مناسب RMSD ارائه دادیم. قابل توجه است که پارامترهای برنامه قابل تغییر می باشد که ما را قادر می سازد قطعات با شباهت بالا و یا اینکه قطعاتی که شباهت کمتری نیز به هم دارند را نیز در پروتئین پیدا کنیم.

روش

در هر پروتئین، برای هر چهار اسید آمینه متوالی i ، $i+1$ ، $i+2$ و $i+3$ ساختمانی متناظر با کربنهای آلفای آن در نظر گرفته میشود. برای اسید آمینه i سه زاویه $(\phi_{i_1}^A, \phi_{i_2}^A, \phi_{i_3}^A)$ تعریف می شود، که $\phi_{i_1}^A$ زاویه بین بردارهای $i(i+1)$ و $(i+1)(i+2)$ ؛ $\phi_{i_2}^A$ زاویه بین بردارهای $(i+1)(i+2)$ و $(i+2)(i+3)$ و $\phi_{i_3}^A$ زاویه دو سطحی است که این چهار کربن آلفا با یکدیگر تشکیل می دهند. چنانچه A و B دو پروتئین که به ترتیب دارای m و n اسید آمینه هستند را در نظر بگیریم جدول ماتریس $M=[m_{ij}]$ چنین تعریف می شود که برای هر i و j ، چنانچه $\frac{|\phi_{i_1}^A - \phi_{j_1}^B| + |\phi_{i_2}^A - \phi_{j_2}^B| + |\phi_{i_3}^A - \phi_{j_3}^B|}{3} < \epsilon$ باشد $m_{ij}=1$ و در غیر اینصورت $m_{ij}=0$. سپس بزرگترین مجموعه $S = \{m_{i_1j_1}, m_{i_2j_2}, \dots, m_{i_kj_k}\}$ از مقادیر ماتریس M را در نظر گرفته می شود. در گام بعد کربن آلفای r' از پروتئین B را روی کربن آلفای r از پروتئین A قرار می دهیم و بردار $\vec{r'(r'+1)}$ را طوری حرکت می دهیم که بر روی بردار $\vec{r(r+1)}$ قرار گیرد. در صورتی که با چرخشی که پروتئین B حول بردار $\vec{r'(r'+1)}$ میکند، فاصله بین مراکز جرم ساختمانیهای S_i از A و S_j از ساختمان B کمتر از δ آنگستروم باشد، فاصله بین مراکز جرم ساختمانیها یک پارامتر قابل تنظیم است که با انتخاب آن می توان تطابق های با فاصله کمتر یا بیشتر را انتخاب کرد. پارامتر دیگر گامهای زوایای چرخش است که با کم کردن مقدار آن، تعداد حالات بیشتری از تطابق دو ساختمان بررسی می شود که حساسیت تطابق را بالا می برد.

نتایج و بحث

1. مقایسه با دیگر روشها برای تشخیص شباهتهای جزئی: برای آزمون عملکرد روش STON به منظور شناسایی تشابهات کم ساختمانی و همچنین مقایسه با روشهای دیگر، آنرا را روی جفت ساختمانی که به عنوان نمونه های سخت شناخته می شوند (9) بکار بردیم. نتایج با برنامه های DALI (6)، CE (5)، FAST (10) و FATCAT (11) مقایسه شد. در STON 3-5، حد آستانه فاصله بین اسید های آمینه متناظر 3 آنگستروم و گام چرخش نیز 5 درجه می باشد. هنگامیکه حد آستانه فاصله کوچک باشد، حساسیت پایین می آید و اختصاصی بودن افزایش می یابد بنابراین قطعاتی که شباهت بیشتری دارند تطابق می یابند. همانطور که در جدول 1 نشان داده شده است روشهای مختلف در RMSD و تعداد اسید آمینه های تطابق یافته با یکدیگر تفاوت دارند. بطور کلی، نتایج بدست آمده از STON 5-5 با روشهای دیگر قابل مقایسه و در برخی موارد بهتر از بقیه می باشد. در STON 3-5 گرچه تعداد اسید آمینه های تطابق یافته کمتر از بقیه روشها است، اما نسبت تعداد



آسید آمینه های تطابق یافته به RMSD در تمام موارد از بقیه روشها بیشتر می باشد.

2. **مقایسه اسیدهای آمینه ای که در فعالیت پروتئین دخالت دارند:** یکی از اهداف مقایسه ساختمانی پروتئینها بررسی مکانهای فعال و تطابق اسیدهای آمینه ای است که در این مکانها درگیر هستند. ساختمان مکان فعال یک پروتئین حتی اگر همولوژی کلی توالی پائین باشد میتواند مدل خوبی برای پروتئینهای مربوطه ارائه دهد. تطابق کلی ساختمان پروتئین برای رسیدن به اسیدهای آمینه متناظر بیشتر و RMSD کمتر ممکن است منجر به تطابق نادرست اسید آمینه ها در دو پروتئین شود. عملکرد روش STON برای تعیین اسید آمینه هایی که در مکان فعال پروتئین هستند توسط مقایسه جفتی ساختمانی از اعضاء سه خانواده مختلف بررسی شد. در هر خانواده، پروتئینها مکان فعال مشابه ولی شباهت توالی کمی با یکدیگر داشتند. اسید آمینه هایی که در مکان فعال شرکت داشتند از پایگاه اطلاعاتی EzCatDB (12) استخراج شدند. در اکثر موارد RMSD اسید های آمینه مکان فعال در روش STON معادل و یا بهتر از روشهای DALI و CE بود. کاهش فاصله آستانه تا حد 1Å منجر به نتایج بهتری برای تطابقهای محلی و همچنین قرار گرفتن اسید های آمینه مکان فعال بروی یکدیگر خواهد شد.

3. در آدرس <http://bioinf.cs.ipm.ac.ir/software/ston> یک سرویس دهنده وب که بر پایه این الگوریتم کار می کند موجود می باشد. ورودی برنامه ساختمان پروتئینها در قالب فایل PDB و یا شماره شناسایی پروتئین در بانک PDB می باشد. خروجی برنامه یک فایل PDB است که حاوی دو زنجیره پروتئین که بر هم منطبق شده اند و یک فایل که اطلاعات انطباق اسیدهای آمینه در آن درج شده است می باشد.

جدول 1: مقایسه تطابق ساختمانی برای 10 جفت ساختمان مختلف که توسط روشهای CE, DALI, FAST, FATCAT و STON بدست آمده اند.

structure pair	CE			DALI			FAST			FATCAT			STONE-3-5 ^a		STON-5-5 ^b	
	res	RMSD	common ^c	res	RMSD	common ^c	res	RMSD	common ^c	res	RMSD	common ^c	res	RMSD	res	RMSD
1bge:B(159) 2gmf:A(121)	107	3.9	0.23	94	3.3	0.63	83	3.10	0.73	100	3.19	0.62	69	1.85	90	2.92
1cew:I(108) 1mol:A(94)	81	2.3	0.71	81	2.3	0.88	74	1.99	0.82	83	2.44	0.88	69	1.52	77	2.31
1cid: (177) 2rhe: (114)	97	2.9	0.87	97	3.2	0.90	72	2.46	0.79	100	3.11	0.89	77	1.60	90	2.84
1crl: (534) 1ede: (310)	219	3.8	0.69	211	3.5	0.75	-	-	-	272	3.57	0.57	154	2.01	180	3.12
1fxi:A(96) 1ubq: (76)	64	3.8	0.78	60	2.6	0.90	52	2.63	0.78	63	3.01	0.68	41	1.62	55	3.08
1ten: (89) 3hrh:B(195)	87	1.9	0.91	86	1.9	0.93	78	1.67	0.91	87	1.90	0.91	80	1.53	85	2.43
1tie: (166) 4fgf: (124)	116	2.9	0.83	114	3.1	0.43	93	3.29	0.75	117	3.05	0.43	87	1.73	108	3.03
2sim: (318) 1nsb:A(390)	275	3.0	0.82	291	3.3	0.85	-	-	-	286	3.07	0.83	194	1.94	264	3.23
2aza:A(129) 1paz: (120)	84	2.9	0.50	81	2.5	0.87	74	2.45	0.84	87	3.01	0.86	65	1.72	79	2.81
3hla:B(99) 2rhe: (114)	84	3.4	0.18	75	3.0	0.86	63	2.87	0.77	83	3.11	0.50	58	1.81	70	3.00

^a STON با پارامترهای 3Å برای آستانه فاصله و گامهای 5 درجه برای چرخش، ^b STON با پارامترهای 5Å برای آستانه فاصله و گامهای 5 درجه برای چرخش، ^c تعداد مکانهای تطابق یافته یکسان با STON 3-5 تقسیم بر تعداد کل تطابقها در STON 3-5



تشکر و قدردانی بخشی از هزینه های این پژوهش با کمک
هزینه پژوهشی پژوهشگاه دانشهای بنیادی IPM (شماره CS1385-1-02) تامین شده است.

منابع:

1. A. Murzin, S.E. Brenner, T. Hubbard, C. Chothia (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536-540.
2. C.A. Orengo, et al (1997) A hierarchic classification of protein domain structures. *Structure* 5:1093-1108.
3. L. Holm, C. Sander (1994) The fssp database of structurally aligned protein fold families. *Nucleic Acids Res.* 22:3600-3609.
4. M. Gerstein, M. Levitt (1998) Comprehensive Assessment of Automatic Structural Alignment Against a Manual Standard, the SCOP Classification of Proteins. *Protein Sci.* 7:445-456.
5. I.N. Shindyalov, P.E. Bourne (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 11:739-747.
6. L. Holm, C. Sander (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233:123-138.
7. J.M. Sauder, et al (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins* 40:6-22.
8. R. Kolodny, P. Koehl, M. Levitt (2005) Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *J. Mol. Biol.* 346:1173-1188.
9. D. Fischer, A. Elofsson, D. Rice, D. Eisenberg (1996) Assessing the performance of fold recognition methods by means of a comprehensive benchmark, *Pacific Symposium on Biocomputing*, (World Scientific Publishers, New Jersey p. 300
10. J. Zhu, Z. Weng (2005) FAST: A Novel Protein Structure Alignment Algorithm, *Proteins* 14:417-423.
11. Y. Ye, A. Godzik (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19:246-255.
12. N. Nagano (2005) EzCatDB: the enzyme catalytic-mechanism database, *Nucleic Acids Res.* 33:D407-D412



STON: A novel method for protein three dimensional structure comparison

Seyed Shahriar Arab^{1,2*}, Changiz Eslahchi^{2,3*}, Mehdi Sadeghi^{2,4}, Hamid Pezeshk^{2,5}, Amir Massoud Rahimi², Heydar Maboudi Afkham³

1. Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
2. School of Computer Science, Institute for Studies in Theoretical Physics and Mathematics (IPM), Tehran, Iran
3. Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran
4. Center of Excellence in Biomathematics, School of Mathematics, Statistics and Computer Sciences, University College of Science, University of Tehran, Tehran, Iran,
5. National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

Abstract

Protein structure comparison is an important problem in bioinformatics and has many applications in the study of structural and functional genomics. During the last decades, various heuristic methods have been developed to solve the protein structure comparison problem. Most of the protein structure comparison methods give the alignment based on the minimum RMSD and ignore many significant local alignments that may be important for evolutionary or functional studies. We have developed a new algorithm to find aligned residues in two proteins with desired RMSD value. The parameterized distance and rotation in this program enable us to search for strongly or weakly similar aligned fragments in two proteins.

Keywords: protein structure comparison; alignment; greedy algorithm; rigid body transformation; dihedral angle