مدلی برای به دست آوردن حفره‌های پروتئین[۱]

سید مجید صابری فتحی

گروه فیزیک دانشگاه فردوسی مشهد

چکیده

در دهۀ اخیر، با توجه به پیشرفت علوم و فنآوری محاسباتی، استفاده از روشهای محاسباتی در تولید داروهای ضد سرطان متداول شده است. با توجه به فرایند تأثیر دارو بر سلول، محققان در "کشف دارو" به دنبال اسید آمینه‌های (لیگاند) تأثیرگذار بر پروتئین‌های سلول‌ها هستند. شرط اولیه ایجاد هرگونه پیوند بین پروتئین و اسید آمینه وجود حفره‌ای مناسب در سطح پروتئین است که از لحاظ هندسی اسید آمینه بتواند در آن جای گیرد. لذا در تحقیقات نظری و محاسباتی کشفِ دارو، بدست آوردن حفره های موجود در سطح پروتئین دارای اهمیت زیادی است، از این جهت که می تواند بسیاری از اسیدهای آمینۀ ناسازگار را از چرخه بررسی‌ها حذف کند. بدین‌گونه تعداداسیدهای آمینه‌ای که در آزمایش‌های تجربی بعدی لازم است امکان پیوند شیمیایی آن‌ها با پروتئین مورد بررسی قرار گیرد کاهش می‌یابد و بدین ترتیب هزینه زمانی و مالی کشف دارو کم می شود. در این مقاله، الگوریتمی جدید و سریع معرفی می‌شود که به وسیلۀ آن می‌توان حفره‌های سطح یک پروتئین را با داشتن موقعیت اتم‌های آن تعیین کرد و خصوصیات بیوشیمیایی و فیزیکی آن‌ها را مشخص نمود.

A model for finding proteins pockets

S. M. Saberi Fathi
Department of Physics, Ferdowsi University of Mashhad
saberifathi@um.ac.ir

Abstract

This paper provides a simple method for a protein-clustering strategy. The basic idea implemented here is to use computational geometry methods to predict and characterize ligand-binding pockets of a given protein structure. In addition to geometrical characteristics of the protein structure, we consider some simple biochemical properties that help recognize the best candidates for pockets in a protein's active site.

مقدمه

بررسی اطلاعات اصلی پروتئین‌ها از طریق ساختار سوم آن‌ها امکان‌پذیر است. کارکرد آنزیمی پروتئین، و هم‌چنین اطلاعاتِ مربوط به پیوند پروتئین‌ها به لیگاندها (مولکول‌های بازدارندۀ[۱] کوچک) [۱] نمونه‌هایی از این اطلاعات هستند. در مطالعۀ پیوند بین پروتئین و لیگاند، برخی از روش‌ها بر پایۀ پیش‌بینی کارکرد آنزیمی پروتئین‌ها بنا نهاده شده‌اند که در آن‌ها اسیدهای آمینه مشخص می شوند، همانند: به دست آوردن خصوصیات موضعی عملکرد اسیدهای آمینه [۲و۳]، کاربرد مدل‌های عمومی[۲] شناخته شده از سایت‌های فعال [۴و۵] یا مشخص کردن شکل سطحی سایت فعال [۶-۱۰].

لیگاندها مولکول‌های کوچکی (اسیدهای آمینه) هستند که می‌توانند با اتم‌های سطحی پروتئین‌ها وارد اندرکنش شده و ترکیب شوند. لازمۀ این ترکیب در مرحلۀ اول وجود حفره‌هایی در پروتئین می‌باشد که از لحاظ سایز و شکل هندسی لیگاندها بتوانند در آن‌ها قرار بگیرند، همانند قرار گرفتن کلید در قفل. از لیگاندها (داروها) در شیمی‌درمانی و از بین بردن سلول‌های سرطانی استفاده می‌شود. برای تعیین لیگاند مناسب برای یک پروتئین بررسی ساختار هندسی آن پروتئین، و تعیین اتم‌های سطحی آن حائز اهمیت است. در مرحلۀ بعد تعیین سایت فعال در پروتئین است که امکان تشکیل پیوند با لیگاند را می‌دهد.

[۱] این مقاله ارائه‌ای است از پژوهش انجام شده توسط مؤلف و همکاران در مرجع شمارۀ [۱۲].

علاوه بر روش‌های تجربی، تحقیق در روش‌های محاسباتی به خاطر قابلیت سرعت بخشیدن به "کشف دارو" در دهۀ اخیر با پیشرفت رایانه‌ها مورد اقبال زیادی بوده است. معمولاً روش‌های مورد استفاده برای تعیین لیگاند مناسب نیمه اتوماتیک هستند به طوری‌که اپراتوری متخصص با کمک برخی نرم‌افزارهای رایانه‌ای این مهم را انجام می‌دهد. یکی از نیازهای موجود در این زمینه به دست آوردن روش‌های تمام اتوماتیک است.

روش‌های محاسباتی معمولاً به دو دسته تقسیم می‌شوند. یکی از آن‌ها، روش‌های هندسی است که حفره‌های ممکن را پیدا می‌کنند همانند SURFNET[۱۱]، PASS [۱۲]و VISGRID [۱۳]. روش‌های هندسی معمولاً از نظر محاسباتی دارای هزینۀ پایین و سرعت بالا هستند. اما از این جهت که قادر به تعیین سایت فعال نیستند لازم است در کنار آن‌ها از نرم‌افزارهای دیگری که توانایی تعیین سایت‌های فعال را دارند استفاده شود. روش دوم روش‌هایی است که سایت فعال را تعیین می‌کنند و بر مبنای دینامیک مولکولی و تعیین خواص فیزیکوشیمیایی لیگاند و حفره پروتئین عمل می‌کنند. در این روش‌ها محاسبۀانرژی پیوندی، بررسی امکانِ پیوندهای واندروالس، یونی وغیره را میسر می‌سازد. هزینۀ محاسباتی این گونه روش‌ها زیاد است و بسیار زمانبر. معمولاً در این روش‌ها به وجود یک اپراتور نیاز است. برخی از این روش‌ها عبارتند از: DEPTH TRAVEL [۱۴]، CAVITY VOIDOO [۱۵]،QSITEFINDER [۱۶]، CAST [۱۷]و غیره. در این مقاله ما روش جدیدی ارائه می‌دهیم که در آن هم خواص بیوشیمیایی و هم خواص هندسی و فیزیکی در نظر گرفته می‌شود و هزینۀ محاسباتی کمی دارد.

روش و الگوریتم

در این روش، اطلاعات پروتئین از پایگاه داده‌های پروتئین، PDB، گرفته می‌شود. سپس پروتئین را در جعبه‌ای مکعب مستطیل شکل که آن را مشبک[۳] کرده‌ایم محصور می‌کنیم. کوچک‌ترین پوستۀ محدب[۴] که پروتئین در آن محصور بشود را تعیین می‌کنیم (شکل۱). سطح یک پوستۀ محدب، در سه بعد، متشکل از مثلث‌هایی است. حجم زیرِ هر یک از این مثلث‌ها ممکن است با حجم برخی مثلث‌های دیگر مقداری هم‌پوشانی داشته باشد. بدین ترتیب می‌توان امکان وجود حفره را با حجم زیرِ یک یا چند مثلث تعریف کرد (شکل۲). در ادامه، ما فاصلۀ هر یک از اتم‌هایی را که در حجم زیرِ یک مثلث قرار گرفته‌اند تا آن مثلث حساب می‌کنیم و اتم‌هایی را سطحی می‌نامیم که کم‌ترین فاصله را تا مثلث داده شده داشته باشند (شکل۲). در این مرحله، برنامه برخی خواص فیزیکی حفره را استخراج می‌کند مانند سطح، عمق، تعداد اتم‌ها در آن و تعداد وُکسِل‌های[۵] خالی که امکان دارداتم‌های لیگاند در آن‌ها قرار گیرند. هم‌چنین خواص بیوشیمیایی اتم‌های سطحی حفره را با استفاده از اطلاعات موجود در فایل‌های PDB در مورد پروتئین استخراج می‌کنیم و سپس با اطلاعات لیگاند (که به طور مشابه استخراج می‌شود) مقایسه می‌کنیم. این اطلاعات برای هر اتم سطحی شامل: پروتون‌دِهی یا پروتون‌گیری اتم، امکان ایجاد پیوندهای واندروالس، یونی و هیدروژنی، امکان قرار گرفتن اتم در C-رینگ و غیره است. در صورتی که اطلاعات مستخرج در مورد حفره‌ای خاص در پروتئین و لیگاند قابلیت ترکیب آن دو را بدهد، حفرۀ مذکور "قابلیت" سایت فعال شدن را دارد. برای ارائه مقایسۀ دقیقی از برنامه از پروتئین‌های انتخاب شده در دو دستۀ متناظر استفاده می‌کنیم: پروتئین بدون لیگاند (به عنوان مثال ۱A۶U) و پروتئینِ متناظری که لیگاندی به آن متصل شده است (برای مثال فوق ۱A۶W). ما برنامه را برای پروتئین بدون لیگاند اجرا می‌کنیم و نتیجۀ به دست آمده را با اطلاعات پروتئین با لیگاند متناظر مقایسه می‌کنیم. سپس معیار کمّی "نسبت صحیح" (cf) را به صورت تعداد اسیدهای آمینه پیش‌بینی شده در یک حفرۀ پروتئین بدون لیگاند توسط برنامه به تعداد اسیدهای آمینۀ موجود در سایت فعال داده شده در فایل PDB پروتیئن متناظر همراه لیگاند تعریف می‌کنیم[۱۸]. روشی که معرفی شد تعداد حفره‌هایی را که لازم است در مورد آن‌ها تحقیقات بعدی انجام شود بسیار کم می‌کند که می‌تواند موجب تسریع کشف دارو شود. زمان محاسبه به وسیلۀ لپ تاپ با CPU intel i۷ بین چند ثانیه تا یک دقیقه برای پروتئین‌هایی با سایز بزرگ است.

جدول۱. حفره‌ها و خصوصیات آن‌ها برای ۱A۶U که توسط روش ما به دست آمده است [۱۸].

Pocket Num.*	Num. of Atoms	Num. of Empty voxels	Surface of Pocket	Depth of Pocket	NoA** HA[a] Bonds	NoA HD[b] Bonds	NoA vdW*** Bonds	NoA Ionic Bonds	NoA Sulfur Bonds	NoA C-Ring	*cf* of the ۱st AS, HAP[c]	*cf* of the ۲nd AS, AC۱[c]
۱	۶۳	۴۰۱	۱۱۶,۲۵	۲۸,۴۰	۵	۸	۰	۱	۰	۲۰	۰,۳۱	۰,۳۳
۵	۸۰	۴۸۱	۲۱,۸۳	۳۸,۶۶	۲	۳	۱۰	۲	۰	۲	۰	۰
۱۸	۱۰۱	۶۴۸	۱۸۷,۲۷	۲۵,۸۳	۵	۷	۶	۲	۰	۱۴	۰,۱۲	۰,۱۱
۱۹	۶۷	۴۱۱	۸۴,۳۶	۱۹,۳۵	۱	۲	۵	۰	۰	۲	۰	۰
۳۸	۴۴	۲۶۶	۱۳۸,۹۰	۲۰,۶۳	۱	۴	۱	۰	۰	۶	۰	۰
۳۹	۸۵	۴۹۹	۸۲,۵۸	۲۸,۲۶	۳	۵	۲	۰	۰	۱۴	۰,۳۱	۰,۲۲
۴۰	۲۱	۱۲۷	۷۷,۹۷	۱۴,۵۳	۲	۳	۰	۰	۰	۴	۰,۰۶	۰
۵۸	۱۱۸	۷۶۵	۳۴۰,۹۰	۲۹,۸۳	۵	۴	۷	۳	۰	۳	۰	۰
۵۹	۸۶	۵۲۹	۲۵۳,۲۰	۲۶,۷۲	۴	۴	۴	۲	۰	۶	۰,۰۶	۰
۸۵	۲۲۶	۱۳۶۰	۳۷۰,۱۴	۳۶,۱۸	۷	۷	۲۶	۳	۱	۲۷	۰	۰
۸۹	۲۱	۱۴۱	۲۱۲,۳۵	۲۱,۴۷	۰	۱	۴	۱	۰	۴	۰	۰
۹۰	۹۲	۵۷۳	۲۹۳,۲۸	۲۸,۵۴	۴	۲	۱۵	۲	۰	۱۱	۰	۰
۱۱۲	۴۴	۲۴۱	۳۶,۳۳	۲۷,۳۹	۱	۲	۱	۰	۰	۶	۰,۰۶	۰
۱۱۷	۳۸	۲۱۵	۷۶,۶۶	۱۷,۴۲	۱	۳	۰	۰	۰	۸	۰	۰
۱۳۷	۱۵	۹۹	۱۲۷,۵۷	۱۷,۵۳	۲	۴	۰	۰	۰	۳	۰,۲۵	۰,۳۳
۱۴۳	۵۵	۳۵۴	۲۵۹,۱۰	۲۴,۲۴	۴	۸	۰	۱	۰	۲۰	۰,۴۳	۰,۵۵

* Pocket number indicates the number in the protein's atomic positions convex hull surface rows, and they correspond to three vertices of triangles.
** NoA means the number of atoms
*** vdW means van der Waals
[a] HA means hydrogen bond acceptor
[b] HD means hydrogen bond donor
[c] These are the *cf*-values (ratio of the number of correct residues to the total number of residues in the active site). For ۱A۶W in PDB two active sites (AS) are reported as HAP and AC۱.

جدول۲. مقایسهٔ کارایی روش ما با روش‌های دیگرهمانند CASTp، LIGSITE، PASS، SURFNET، و VISGRID[۱۸] .

	۴۸ Unbound structures (Top ۱)	۸۶ Unbound structures (Top ۱)
CAST	۳۱ (۶۴,۶٪)	۶۶ (۷۶,۷٪)
LIGSITE	۳۶ (۷۵,۰٪)	۶۹ (۸۰,۲٪)
PASS	۲۷ (۵۶,۳٪)	۵۴ (۶۲,۸٪)
SURFNET	۱۹ (۳۹,۶٪)	۶۳ (۷۳,۳٪)
VISGRID: Top ۰,۸٪ voxels	۳۴ (۷۰,۸٪)	۵۵ (۶۴,۰٪)
Our Method: Overlap ۰,۸	۲۴ (۵۰٪)	۶۶ (۷۸٪)

بحث و نتایج

جدول۱ حفره‌های ممکن برای ۱A۶U را می‌دهد. این پروتئین در مجموع دارای ۱۷۳۷ اتم و جعبه‌ای با ۴۱x۴۹x۴۳ وُکسِل است. پوستهٔ محدب آن دارای ۱۴۸ مثلث سطحی است و فقط ۱۶ حفره با هم‌پوشانی ۸۰٪ با اِعمال شرایط فیزیکی و بیوشیمیایی باقی می‌مانند که قابلیت سایت فعال شدن را دارند. فقط ۴ حفره دارای مقدارِ cf بیشتر از ۰٫۲۵ می‌باشند. لیگاندی که در این مقایسه مورد استفاده قرار گرفته است NIP است که ۱۷ اتم دارد و دارای طول ۸٫۹۷Å طول و سطحی معادل با ۲۰٫۸۷Å^2 است. بنابراین حفره‌هایی که دارای مشخصاتی بزرگ‌تر از این مقادیر باشند باید در نظر گرفته شوند. در جدول ۲ نتایج روش ما با دیگر روش‌ها مقایسه شده است که نشان دهندهٔ هم‌ارز بودن روش ما از لحاظ پیش‌بینی با سایر روش‌هاست.

شکل۱. پوستهٔ محدب مربوط به ABT۱ [۱۸]

شـکل۲. یک مثلث داده شـده برای ABT۱. در شـکل، نحوهٔ تعریف یک حفره مشخص شده است، هم‌چنین فاصلهٔ اتم تا مثلث نیز نشان داده شده است[۱۸] .

مرجع‌ها

[۱] Rupp B, Wang J. Predictive models for protein crystallization. Methods (San Diego, Calif.). ۲۰۰۴ November;۳۴(۳):۳۹۰–۴۰۷.

[۲] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics (Oxford, England). ۲۰۰۶ January ۱۵;۲۲(۲):۱۹۵–۲۰۱.

[۳] Huang B, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. BMC structural biology. ۲۰۰۶;۶:۱۹.

[۴] Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. Journal of molecular graphics. ۱۹۹۵ October;۱۳(۵):۳۲۳–۳۳۰, ۳۰۷–۳۰۸.

[۵] Laurie ATR, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics (Oxford, England). ۲۰۰۵ May ۱;۲۱(۹):۱۹۰۸–۱۹۱۶.

[۶] Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein science: a publication of the Protein Society. ۱۹۹۸ September;۷(۹):۱۸۸۴–۱۸۹۷.

[۷] Peters KP, Fauck J, Frömmel C. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. Journal of molecular biology. ۱۹۹۶ February ۱۶;۲۵۶(۱):۲۰۱–۲۱۳.

[۸] Brady GP Jr, Stouten PF. Fast prediction and visualization of protein binding pockets with PASS. Journal of computer-aided molecular design. ۲۰۰۰ May;۱۴(۴):۳۸۳–۴۰۱.

[۹] Li B, Turuvekere S, Agrawal M, La D, Ramani K, Kihara D. Characterization of local geometry of protein surfaces with the visibility criterion. Proteins. ۲۰۰۸ May ۱;۷۱(۲):۶۷۰–۶۸۳.

[۱۰]. Andricopulo AD, Salum LB, Abraham DJ. Structure-based drug design strategies in medicinal chemistry. Current topics in medicinal chemistry. ۲۰۰۹;۹(۹):۷۷۱–۷۹۰.

[۱۱] Kontoyianni M, Madhav P, Suchanek E, Seibel W. Theoretical and practical considerations in virtual screening: a beaten field? Current medicinal chemistry. ۲۰۰۸;۱۵(۲):۱۰۷–۱۱۶.

[۱۲] Fetrow JS, Godzik A, Skolnick J. Functional analysis of the Escherichia Coli genome using the sequence-to-structure-to-function paradigm: identification of proteins exhibiting the glutaredoxin/thioredoxin disulfide oxidoreductase activity. JMol Biol ۱۹۹۸;۲۸۲:۷۰۳–۷۱۱.

[۱۳] Wallace AC, Borkakoti N, Thornton JM. TESS: a geometric hashing algorithm for deriving ۳D coordinate templates for searching structural databases. Application to enzyme active sites. Protein Sci ۱۹۹۷;۶:۲۳۰۸–۲۳۲۳.

[۱۴] Goldman BB, Wipke WT. QSD quadratic shape descriptors. ۲. Molecular docking using quadratic shape descriptors (Qsdock). Proteins ۲۰۰۰;۳۸:۷۹–۹۴.

[۱۵] Exner TE, Keil M, Brickmann J. Pattern recognition strategies for molecular surfaces. I. Pattern generation using fuzzy set theory. J Comput Chem ۲۰۰۲;۲۳:۱۱۷۶–۱۱۸۷.

[۱۶] Tuccinardi T. Docking-based virtual screening: recent developments. Combinatorial chemistry & high throughput screening. ۲۰۰۹ March;۱۲(۳):۳۰۳–۳۱۴.

[۱۷] Basus VJ, Song G, Hawrot E. NMR solution structure of an alpha-bungarotoxin/nicotinic receptor peptide complex. Biochemistry. ۱۹۹۳ November ۲۳;۳۲(۴۶):۱۲۲۹۰–۱۲۲۹۸.

[۱۸] S. M. Saberi Fathi, Jack A. Tuszynski, A simple method for finding a protein’s ligand-binding pockets , EN , BMC Structural Biology , Volume (۱۴) , ۲۰۱۴-۷, Pages ۱-۹

۱ Inhibitors
۲ Template
۳ Voxelized
۴ Convex Hull
۵ Voxels