

蛋白質 - 化合物複合体の相互作用部位プロファイルを用いた 類似相互作用蛋白質検索方式

吉野 公一 (大阪大学)* , 大川 剛直 (神戸大学)

概要

蛋白質 - 化合物複合体の相互作用部位には、化合物との結合に関して特徴的な原子配置が存在する。そこで、同一の化合物に結合する複数の相互作用部位から、化合物と結合する相互作用原子の特徴を抽出し、類似構造を検索することを考える。このとき、化合物と直接結合する原子の特徴だけで相互作用部位全体の特徴を表現することは困難である。

そこで本研究では相互作用部位全体の特徴として相互作用原子の特徴、結合化合物原子の特徴、相互作用原子の配置に影響を与える原子の特徴の3つをプロファイルとして抽出し、プロファイルをクエリーとした類似構造探索を行うことで、類似相互作用蛋白質を検索する方式を提案する。

A method for retrieving similar interaction protein using interaction site profile of protein-compound complex

Koichi Yoshino*(Osaka University), Takenao Ohkawa(Kobe University)

Abstract

In the interaction site of the protein-compound complex, there are features for binding to the compound. However, it is difficult to express the feature of the entire interaction site only by the one of the atoms that binds directly to the compound.

We propose a method for retrieving a protein with a local structure that is similar to the known interaction site in terms of the binding ability to a specific compound by using a interaction site profile. The profile is extracted from the features of some interaction sites that bind to the same compound.

1 はじめに

蛋白質は、通常、表面のごく一部において、他の蛋白質や化合物と結合することにより機能を発現する。この部位のことを、相互作用部位とよぶ [1]。相互作用部位は蛋白質の機能に密接に関わることから、機能未知の蛋白質の相互作用部位を特定することや、結合する化合物を特定することは蛋白質の機能解析の際に重要な役割を果たす。

蛋白質立体構造データベース PDB (Protein Data Bank) ¹には、蛋白質単体の立体構造のみではなく、蛋白質と化合物が結合した状態の立体構造も登録されている。そのようなデータを蛋白質 - 化合物複合体データとよぶ。複合体データを解析することで、相互作用に参与している原子を特定し、相互作用部位の立体構造を得ることができる [2]。

¹<http://www.rcsb.org/pdb/>

類似した相互作用部位を持つ蛋白質は類似した相互作用性を有するため、相互作用部位をクエリーとした類似蛋白質の検索は、その機能解明に有効である。これまでも蛋白質の分子表面に注目した検索方式が多数提案されている [3, 4, 5, 6, 7, 8, 9, 10]。本研究では、特に蛋白質 - 化合物複合体の立体構造情報から得られる相互作用部位の原子構造に着目し、類似する蛋白質を立体構造データベースから検索する手法を提案する。

同一の化合物に結合する複数の相互作用部位は、完全には一致していないが共通した特徴がある [11, 12]。そこで、この特徴を相互作用部位プロファイルと呼び、これを検索のクエリーとして利用する。

化合物との結合において重要な原子として、化合物原子と直接相互作用している相互作用原子がある。しかし、相互作用原子は相互作用部位の表面に位置

する数個～10個程度である。そのため、相互作用原子の特徴のみを用いて相互作用部位全体の特徴を表現するのは困難である。また、相互作用原子が表面に配置するにはその周りの原子の結合力の影響が必要であり、化合物が結合するためには化合物原子が入り込めるような相補的な構造を相互作用部位がとる必要がある。

そこで相互作用部位プロファイルは、相互作用原子の特徴を表現するプロファイル原子と、相互作用原子の配置を決定するのに影響をもつ近接原子の性質、そして化合物原子が結合するために必要な空間を表現する存在不可領域により定義する。各プロファイル原子は、その性質や属するアミノ酸残基によって分類される原子の種類と、原子が存在可能なエリアの属性を持つ。近接原子は、相互作用原子を中心とした時に疎水性をもつ原子のみが存在する領域内の原子であり、プロファイルにはその近接原子の化学的性質が含まれる。存在不可領域は、蛋白質原子が配置できない領域の位置と大きさにより定義する。

2 相互作用部位立体構造からのプロファイル抽出

2.1 相互作用部位の比較

化合物と蛋白質の結合は、相互作用部位において蛋白質原子と化合物原子の相互作用により起こる。この時、化合物原子と相互作用している原子のことを相互作用原子とよぶ。図1は化合物 GUANOSINE-5'-DIPHOSPHATE (GDP) に結合する6種類の蛋白質²の相互作用部位である。中心部で球体が連なっている物体が化合物 GDP であり、その周辺に位置する球体が相互作用原子である。相互作用原子の濃淡は原子の種類を表す。

同じ化合物 GDP に結合する相互作用部位においても、その原子の配置がいくつかのパターンに分かれることがわかる。化合物 GDP に結合する50の相互作用部位を対象として、図1のようなグループ分けを行ったところ、グループ1に属する蛋白質が27種類、グループ2に属する蛋白質が11種類、グループ3に属する蛋白質が6種類、3つのどのグル

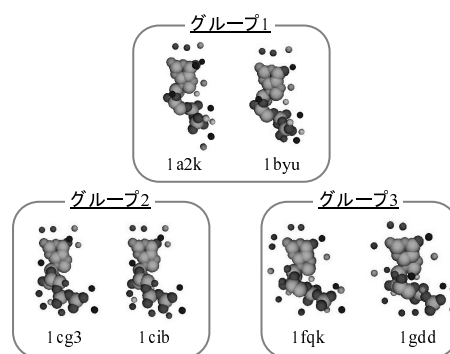


図1: 化合物 GDP に結合する6つの蛋白質のグループ分け

ープにも属さない蛋白質が6種類であった。このグループ分けは、相互作用原子の種類と配置によるおおよその形状をもとに手動で行われたものである。このように、同じ化合物に結合する相互作用部位は幾つかのパターンに分かれるものの、それぞれのパターンにおいては共通した特徴が見られることがわかる。そこでこのような共通した特徴をプロファイルとして抽出する。

2.2 相互作用原子の特徴

2.2.1 原子の分類

図2に化合物 GDP に結合する4種類の蛋白質の相互作用部位を示す。蛋白質原子と化合物とが非共有結合しているとき、それらの原子が空間上で一定距離以下の近接した位置に存在し、両者の化学的性質が相補的になっているものとして抽出される。その結果、抽出される蛋白質原子は表1の10種類のいずれかとなる。

図2からわかるとおり、4つの相互作用部位において同一の性質を持つ原子が同一の配置を取ることが多い。このことから、プロファイルとして抽出した原子をその性質によって分類することで、相互作用原子の特徴を表現する。

また、各原子は属するアミノ酸残基により、さらに詳細に分類でき、同一種類のアミノ酸残基に属する原子が複数の相互作用部位において同一配置を取る場合がある。図3に示すように、図の下部に存在する塩基性原子はヒスチジンやリシンであり、属す

²図の下部に書かれた4文字コードはPDB-IDである。以後、本論文では蛋白質をPDB-IDで表す。

るアミノ酸残基は同一ではないが、図の上部に存在する酸性原子はいずれの相互作用部位においてもアスパラギン酸である。このことから、プロファイルとして抽出する原子は、化学的性質のみでなく属するアミノ酸残基によって分類することも相互作用部位の特徴を表現する。今後、相互作用原子の特徴として抽出した原子をプロファイル原子と呼ぶ。

表 1: 抽出される蛋白質原子

性質	残記名	原子
塩基性 (正に荷電)	リシン	N
	アルギニン	N
	ヒスチジン	N
酸性 (負に荷電)	アスパラギン酸	O
	グルタミン酸	O
非荷電 極性	アスパラギン	N
	グルタミン	N
	セリン	O
	トレオニン	O
	チロシン	O

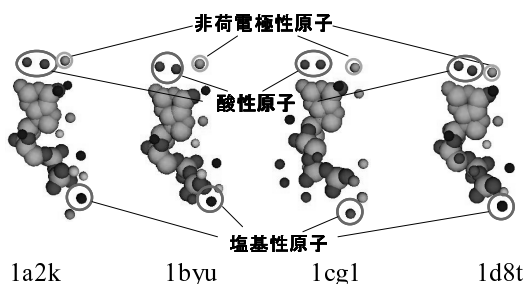


図 2: 化学的性質による相互作用原子の分類

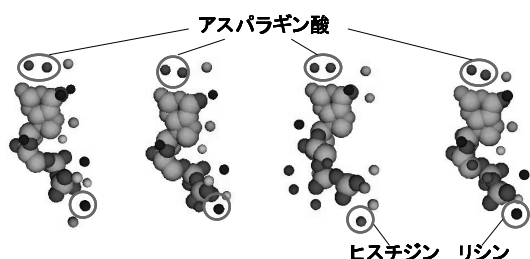


図 3: 属するアミノ酸残基による相互作用原子の分類

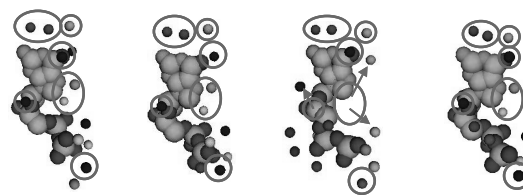


図 4: 相互作用原子の配置

2.2.2 原子の配置

次にプロファイル原子の配置について考える。図 4 を見ると、複数の相互作用部位において共通な原子の相対的な位置が、ほぼ同一な原子もあれば、変化している原子もある。このように、複数の相互作用部位における原子の座標上の変化の大きさが、原子ごとに異なる。このことから、プロファイル原子ごとにその原子の存在可能領域を定義する。この存在可能領域は、プロファイルと検索対象蛋白質との構造比較を行う際、許容可能な座標のずれの誤差範囲として利用する。

また、複数の相互作用部位において頻出しない原子は相互作用に重要な役割を果たさない原子であるとして、特徴として用いない。

これまで述べた、原子の分類と存在可能領域をプロファイル原子のプロパティとする。

2.3 近接原子の特徴

相互作用原子の配置は、その周りの蛋白質原子からの結合力により成立する。様々な原子間結合があるが、とりわけ疎水性結合は、結合力は水素結合よりも弱い、蛋白質の立体構造の形成や化合物との結合に大きな役割を持つことが知られている。また、相互作用部位と疎水性領域の間には深い関係があり、疎水性領域を判定することで蛋白質の相互作用部位を推定する例もある [13]。

そこで、相互作用原子に対して疎水性結合の影響をもつ原子を考える。相互作用原子から、非疎水性原子が含まれない領域内で、一番遠くの疎水性原子までの距離を近接距離とし、その距離内の空間を近接空間とする。近接空間内にある原子を近接原子とする。近接距離の最大値は、図 5 に示すように基本的な水素結合のほとんどが 2.5 Å 以上で起こることから、2.5 Å とする。

例	結合距離(Å)
OH...O	2.72 (O-O間)
OH...N	2.79 (O-N間)
OH...S	3.31 (O-S間)
OH...F	2.72 (O-F間)
OH...Cl	3.12 (O-Cl間)
OH...Br	3.28 (O-Br間)
NH...O	2.89 (N-O間)
NH...N	2.98 (N-N間)
NH...S	3.42 (N-S間)
NH...F	2.92 (N-F間)
NH...Cl	3.23 (N-Cl間)
NH...Br	3.37 (N-Br間)
FH...F	2.44 (F-F間)

図 5: 基本的な水素結合と結合距離

近接原子は全て疎水性原子であるが、属する残基や化学的性質は異なることがある。事前の調査では多くの場合、近接原子の化学的性質はすべて近接するプロファイル原子の化学的性質と同様のものが多かったが、相互作用原子によっては化学的性質が異なる近接原子が混在することもあった。そこで近接原子の化学的性質と、その割合を近接原子のプロパティとし、プロファイルの構成要素にする。

2.4 化合物原子との結合に必要な空間

図 6 に示すようにプロファイル原子の抽出元となる相互作用原子と直接結合する化合物原子の配置を見ると、相互作用原子とおおよそ一定距離に配置している。その距離は化合物原子ごとに異なる。この化合物原子の存在する領域は蛋白質と化合物が結合するために、相互作用部位が相補的な構造を形成する上で重要な領域である。

そこで図 7 に示すように、プロファイル原子の抽

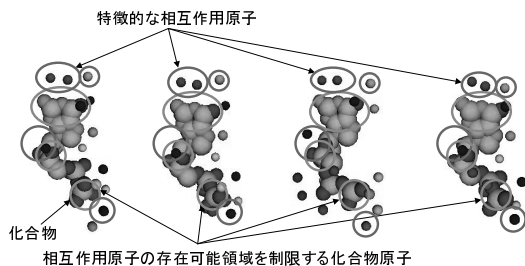


図 6: 相互作用部位の化合物原子と蛋白質原子

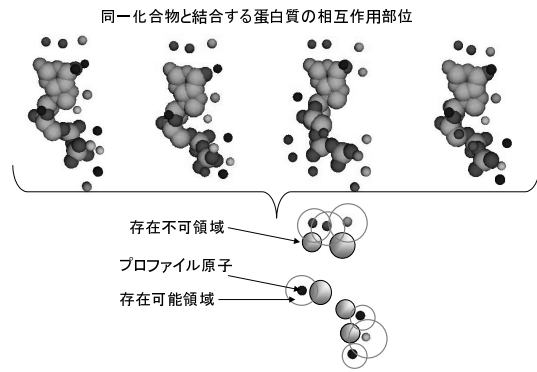


図 7: プロファイル原子と存在不可領域

出とともに、存在不可領域として化合物原子の位置を抽出する。抽出する化合物原子は、プロファイル原子の抽出元の相互作用原子と相互作用している化合物原子である。プロファイル原子の存在可能領域内であっても、存在不可領域には蛋白質原子が存在することはない。

2.5 プロファイルの定義

以上のことを踏まえ、プロファイル P を次のように定義する。

$$P = \langle \{A_i\}, \{S_j\} \rangle \quad (0 \leq i \leq n, 0 \leq j \leq m)$$

$$A_i = \langle xa, ya, za, ra, p, \{C_l\}, kr \rangle \quad (0 < l < q)$$

$$S_j = \langle xs, ys, zs, rs \rangle$$

$$p \in (H \cup K)$$

$$C_l = \langle cp, ra \rangle$$

n : プロファイル原子数

m : 存在不可領域の数

A_i : プロファイル原子 i

S_j : 存在不可領域 j

xa, ya, za : A_i の中心座標

ra : 存在可能領域の半径

p : A_i のプロパティ

q : A_i がもつ近接原子の性質の数

C_l : 近接原子の l 個目のプロパティ

kr : 近接距離

xs, ys, zs : S_j の中心座標

rs : S_j の半径

H : A_i の化学的性質

K : A_i のアミノ酸残基の種類

cp : 近接原子の化学的性質

ra : 全ての近接原子の中で cp が占める割合

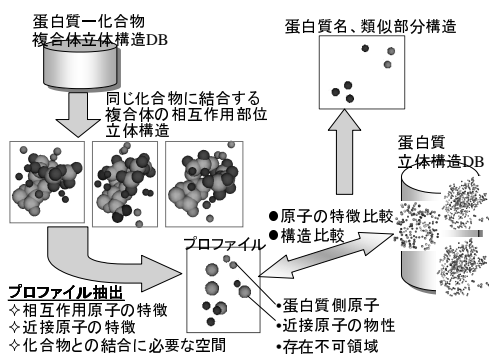


図 8: 検索の概要

3 類似相互作用部位検索方式

図 8 に検索の概要を示す。入力は同一化合物に結合する複数の相互作用部位である。これらはその形状によりグループ分けすることが可能であるが、ここでは、あらかじめグループ分けされ、1つのグループに属するものを想定する。これらの相互作用部位群から検索のクエリーとなるプロファイルを抽出する。このプロファイルと対象となる蛋白質データベース中の蛋白質群と原子レベルでの構造比較を行う。

なお、相互作用は蛋白質表面で起こるものであり、蛋白質内部に埋もれている原子が相互作用部位を構成することはない。そこで、蛋白質の内部に位置する原子をあらかじめ特定し、探索の対象から除外することで、不要な出力の減少による検索精度の向上と検索時間の短縮を図る。

蛋白質の内部原子特定手法として、[14]による原子の周囲 43 個の観測点を用いた手法を用いる。この手法により、相互作用部位の原子を残したまま探索対象原子を 13% ~ 23% 削減することが可能となる。

クエリーであるプロファイルと類似の構造を検索

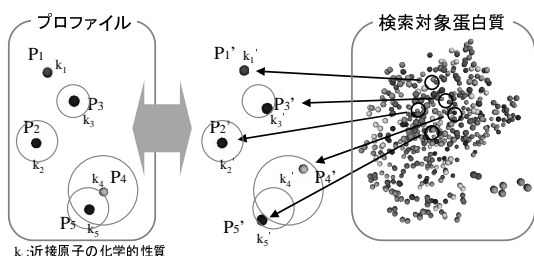


図 9: 類似構造探索の概要

対象蛋白質から探索する。このとき、プロファイルの原子 1 つに注目し、対象蛋白質から条件に合う原子を順に探索していく。その概要を図 9 に示す。

まずプロファイルの 1 つの原子 P_1 に注目する。 P_1 と同じ条件の原子 P_1' を検索対象蛋白質中から 1 つ抽出する。ここでの条件とは原子の種類と近接原子の性質である。次にプロファイルの 2 つ目の原子 P_2 に注目する。 P_2 と同じ条件の原子 P_2' を検索対象蛋白質中から抽出する。この作業を全てのプロファイル原子に関して行う。2 つ目以降の原子に関する条件とは原子の種類と近接原子の性質、そして原子の位置が存在可能領域内にあり、かつ存在不可領域外であるかどうかである。全てのプロファイル原子に関して、条件を満たす原子群を検索対象蛋白質から発見できれば、その原子群を類似構造として決定する。

4 実験

4.1 実験概要

化合物 ADENOSINE-5'-TRIPHOSPHATE (ATP) に結合する 5 種類の蛋白質からプロファイル抽出した。図 10 にプロファイルを示す。図の小さな球体がプロファイル原子、大きな球体が存在不可領域を示す。存在可能領域は近接距離は 2.5 である。近接原子は全て、近接しているプロファイル原子と同じ化学的性質をもつ。また、全てのプロファイル原子に近接原子は 1 つ以上存在した。

検索対象は、化合物 ATP に結合する蛋白質でプロファイル抽出に用いたものとは異なる 8 つの蛋白質である。以上の条件で、提案手法による検索と、プロファイル原子のみによる検索の結果を比較した。

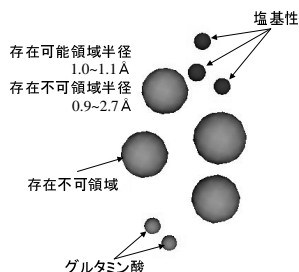


図 10: 作成したプロファイル

表 2: 提案手法による検索結果

蛋白質 PDBID	部位数	比較 手法	提案 手法	R(%)
1dy3	正解	1	1	50
	不正解	2	1	
1gn8	正解	1	1	86
	不正解	7	1	
1h4q	正解	1	1	69
	不正解	84	26	
1jag	正解	1	1	71
	不正解	84	24	
1miw	正解	1	1	86
	不正解	69	10	
1nyr	正解	2	2	75
	不正解	114	29	
1ojl	正解	1	1	68
	不正解	155	49	
1xdn	正解	3	3	33
	不正解	7	4	

R:不正解部位削減率

4.2 実験結果

実験結果を表 1 に示す。比較手法はプロファイル原子のみをクエリとした検索である。不正解部位削減率は以下の式で計算した。

$$\text{不正解部位削減率 (\%)} = 100 \times \frac{\text{比較手法不正解部位} - \text{提案手法不正解部位数}}{\text{比較手法不正解部位数}}$$

結果からわかるように、提案手法を用いることにより、プロファイル原子のみによる検索よりも、検出する不正解部位を少なくしながら相互作用部位を検出できている。このことから存在不可領域と、近接原子の性質を含めた相互作用部位プロファイルを用いることで、効果的に相互作用部位が検索できていることが分かる。

5 まとめ

本稿では、蛋白質 - 化合物複合体における相互作用部位立体構造情報データを元に、相互作用部位の特徴をクエリとして、相互作用が未知の蛋白質立体構造データベース中の蛋白質から類似の立体構造を持つ蛋白質の検索方式について述べた。本手法では相互作用原子の特徴に加え、化合物原子が結合するために必要な空間の特徴と、相互作用原子に近接する原子の特徴を含めた相互作用部位プロファイルを利用する。提案手法を PDB の蛋白質データに対し

て適用した結果、相互作用原子の特徴をクエリーとした検索より、精度が向上することを確認した。

参考文献

- [1] 伊藤隆司, 谷口寿章: プロテオミクス タンパク質の系統的・網羅的解析, 中山書店 (2000).
- [2] G. Kawamura, G. Nagakawa, and T. Ohkawa: "Development of Protein-Compound Interaction Database on Grid Data Service Using the Three-dimensional Structure Data of Complex," *Abstracts of Pacific Symposium on Biocomputing 2004 (PSB2004)*, pp. 87 (2004).
- [3] 兼田佳和, 庄治範匡, 大川剛直, 中村春木: "属性付き法線ベクトルを用いた蛋白質分子表面比較方式," *情報処理学会論文誌*, Vol. 43, No. 1, pp.196-203 (2002).
- [4] Y. Kaneta, N. Shoji, T. Ohkawa, and H. Nakamura: "A Method of Comparing Protein Molecular Surface Based on Normal Vectors with Attributes and its Application to Function Identification," *Information Sciences*, Vol. 146, Issue 1-4, pp. 41-54 (2002).
- [5] 中川忠輔, 兼田佳和, 大川剛直: "分子表面データ間の距離を利用した効率的類似蛋白質検索方式," *情報処理学会論文誌*, Vol. 44, No. 3, pp. 1023-1031 (2003).
- [6] D. W. Ritchie: "Parametric Protein Shape Recognition," A Thesis presented for the degree of Doctor of Philosophy at the University of Aberdeen (1998).
- [7] D. W. Ritchie and G. J. L. Kemp: "Fast Computation, Rotation and Comparison of Low Resolution Spherical Harmonic Molecular Surfaces," *J. Comp. Chem.*, Vol.20, pp. 383-395 (1999).
- [8] M. Ankerst, G. Kastenmuller, H. P. Kriegel, and T. Seidl: "3D Shape Histograms for Similarity Search and Classification in Spatial Databases," in *Proc. of the 6th Int. Symposium on Spatial Databases (SSD)*, Hong Kong, China. *Lecture Notes in Computer Science (LNCS)*, Vol. 1651, Springer Verlag, pp. 207-226 (1999).
- [9] M. Ankerst, G. Kastenmuller, H. P. Kriegel, and T. Seidl: "Nearest Neighbor Classification in 3D Protein Databases," in *Proc. of the 7th Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*, Heidelberg, Germany, AAAI Press, pp. 34-43 (1999).
- [10] H. P. Kriegel, T. Schmidt, and T. Seidl: "3D Similarity Search by Shape Approximation," in *Proc. 5th Int. Symposium on Large Spatial Databases (SSD)*, Berlin, Germany, pp. 11-28 (1997).
- [11] M. Ishiguro and S. Imajo: "Modeling Study on Hydrolytic Mechanism of Class A β -Lactamases," *J Med Chem*, Vol. 39, pp. 2207-2218 (1996).
- [12] R. C. Wilmouth, K. Edman, R. Neutze et al: "X-ray Snapshots of Serine Protease Catalysis Reveal A Tetrahedral Intermediate," *Nat Struct Biol*, Vol. 8, pp. 689-694 (2001).
- [13] 山乙 教之, 広野 修一: "疎水性ポテンシャルを利用したタンパク質上のリガンド分子結合部位の同定," 日本薬学会 構造活性相関部会, K07(2002)
- [14] 野々村祐介, 吉野公一, 中江達哉, 大川剛直: "蛋白質 - 化合物複合体立体構造データに基づく類似相互作用蛋白質の検索方式," *情報処理学会論文誌*, 数理モデル化と応用 (2006 掲載予定)