タンパク質分子表面形状と 物性のデータベース eF-site による 分子機能類似性検索

理化学研究所ゲノム科学総合研究センター

木下 賢吾*, 中村 春木**

兼: * 横浜市立大学総合理学研究科

兼: ** 大阪大学蛋白質研究所

1. タンパク質の立体構造情報の活用

近年の構造生物学の進展は目覚ましく、すでに15,000個を超えるタンパク質の立体構造が明らかにされている。その結果、数多くのタンパク質において、その立体構造と機能がどのように関係しているかが調べられている。しかし、これらの知見はまだ十分なものではない。生物学的機能がわかっているタンパク質に関しては、立体構造情報を利用して、その機能の化学的なメカニズムのより深い理解を得られるようにはなっている。しかし、機能がわからないタンパク質の立体構造を与えられたときに、その生物学的機能および生化学的機能を推定できるほどには立体構造情報を十分活用できていない。

一方、ゲノムプロジェクトにより、すでに50種を超える生物種において全塩基配列が明らかにされ、さらに多くの種のゲノム配列が明らかにされようとしている。しかしながら、すでに明らかにされた塩基配列の中にコードされている遺伝子のうち約半分の遺伝子の機能は未知のままである¹⁾。生物を分子レベルで理解することが、ゲノムプロジェクトの目的だとするならば、これら機能未知の遺伝子あるいはその生産物であるタンパク質の機能を同定することが不可欠になる。

そこで我々は、ゲノムプロジェクトの一環として進展しつつある構造ゲノム科学プロジェクトを念頭に、機

能未知の遺伝子産物としてのタンパク質の立体構造が明らかにされたときに、その機能を立体構造情報から推定する方法の開発を試みた。

2. タンパク質分子表面の形状と物性

タンパク質の立体構造情報から機能を予測するには、さまざまなアプローチが考えられる。ここではすでに多くの立体構造が明らかにされている現状を受けて、立体構造と機能の関係を利用したアプローチをとる。具体的には、(1) 機能部位の構造上の特徴を収めたデータベースを構築し、(2) 機能未知・立体構造既知のタンパク質を、その機能データベースに対して類似性検索することにより機能を推定する。このようなアプローチで最も問題になるのは、どのような構造的特徴に着目すればいいのかである。一口に立体構造といっても、表現の仕方により得られる情報は変わってくる。例えば、よく使われる立体構造の表現としては、フォールド(主鎖原子の折りたたみ構造)がある。しかし、フォールドの類似性と機能の関係は、構造・機能相関というよりむしろ、進化的な類縁関係がフォールドの類似性をもたらす、結果として機能が似ているという間接的なものなので、フォールドが似ているからといって機能が似ているとは限らない²⁾。一方、原子位置の空間配置の類似性は機能に密接にかかわることが予想される。しかし、タンパク質の柔軟性により同じ機能のタンパク質でも原子の空間配置は変化するので、原子空間配置の類似性を検出するのが困難になる。

そこで我々は構造生物学の分野で、構造を決定し、機能を議論する際によく用いられ、原子位置よりは荒っぽいフォールドよりは機能に直接関係していると考えられる、分子表面の形状と物性(静電ポテンシャル分布および疎水性度)に着目して、構造からの機能予測法の開発を試みた。

3. eF-site データベース

機能部位の構造上の特徴を収めたデータベースとして我々は、eF-site (electrostatic surface of Functional site) データベースを構築した³⁾。このデータベースでは、タンパク質の立体構造を分子表面⁴⁾で表現し、そこにタンパク質が作り出す静電ポテンシャル⁵⁾と、分子表面

Identification of Protein Molecular Functions from eF-site, a Molecular Surface Database

Kengo KINOSHITA* and Haruki NAKAMURA**

* Graduate School of Integrated Science, Yokohama City Univ., RIKEN Genomic Sciences Center

** Institute of Protein Research, Osaka Univ., RIKEN Genomic Sciences Center

図1 分子表面の例

静電ポテンシャルのタンパク質表面での値を、+0.1V 以上は青、-0.1V 以下は赤で、+0.1 ~ -0.1 の間は右下の2段のカラーバーの上側の色で色づけし、表面近傍に疎水性の側鎖をもつアミノ酸残基があればカラーバーの下側の色で色付けして表現している。

近傍の疎水性アミノ酸残基のあるなしを色付けして表現した(図1) また、4種類の機能部位についてそれぞれ、eF-site/Antibody (抗体の抗原結合部位)、eF-site/Prosite (Prosite 配列モチーフ⁶⁾)、eF-site/P-site (モノヌクレオチド結合タンパク質のリン酸結合部位)、eF-site/ActiveSite (SwissProt データベースの活性部位) を構築した。

4. 類似性検索法

分子表面は本来滑らかな表面であるが、形状の比較にあたっては、計算機が取り扱いやすい形である3角メッシュでの表現を用いる。ただし、もとの表面の情報をできるだけ残すように、各3角メッシュの頂点での表面の外向き単位法線ベクトルの空間配置を用いて表面を記述する。また物性としては静電ポテンシャルと疎水性度を考える。静電ポテンシャルは、タンパク質の各原子に AMBER パラメータの部分電荷を置き、タンパク質内部は誘電率が2の誘電体、外部は誘電率が80の誘電体として Poisson-Boltzmann 方程式を用いて計算し、この3角メッシュの各頂点にその値をマップする⁵⁾。疎水性度は、各頂点から最も近い位置に存在するアミノ酸側鎖原子の Kyte & Doolittle の疎水性インデックスを各頂点にマップした⁷⁾。以上のような表現を用いると、表面の形状と物性は、物性つきベクトルの空間配置として表現でき、表面の物性と形状の類似性検索は、物性つきベクトルの空間配置の類似性の検索問題になる。

このように問題を定式化すると、問題は2次構造要素

図2 表面の比較のためのグラフ

グラフは点と線からなる。表面の比較では、点が「似た物性をもつ法線ベクトルの対応付け」を意味し、線が「対応付けられた法線ベクトル間の相対配置が似ていること」を意味する(図3) 例えば、図中の1-2という点は、表面Aの1番目の法線ベクトルと表面Bの2番目の法線ベクトルが対応する(似た物性をもつ)ことを表現しており、1-2という点と2-3という点をつなぐ線は、表面Aの1番目と2番目の法線ベクトルの相対配置(図3)と2番目と3番目の相対配置が似ていることを意味している。このような意味づけの結果、グラフの中で互いに線で結ばれた部分グラフに着目すると、その部分グラフに対応する法線ベクトルの対応付けは、その相対配置が互いに似ていることを意味していることになる。

図4 構造変化の例

(A) 50.1 抗体、(B) BV04-01 抗体。それぞれ結合状態と非結合状態の結合部位の様子を示している。

図3 類似性検索: 抗体タンパク質への応用結果

抗体タンパク質での類似性検索の応用結果を典型的な4つのケースに関して示す。4つのケースはそれぞれ、(A) D2.3抗体、(B) N1G9抗体、(C) 50.1抗体、および(D) BV04-01抗体の抗原結合型((+)をつけて示す)をクエリー構造として、eF-site/Antibodyに対して検索を行った結果の類似度の分布。類似度はZ値(平均値からのずれを標準偏差の倍数で表現した値)で規格化して示してある。なお(-)は非結合状態の構造である。

の空間配置の比較とほぼ等価となる。違いとしては、(1) 2次構造要素の空間配置の比較の場合は、物性が2次構造要素の種類(ヘリックスかストランド)という2種類であるのに対して、表面の場合は静電ポテンシャルと疎水性度という連続値であること、(2) 2次構造要素の場合は少ない数の対応付けでかなりのバリエーションがあるが、表面の場合は多くの要素を対応付けてはじめて違いが出てくること、および(3) 2次構造要素の場合はその要素数がたかだか100程度であるが表面の場合は平均的な大きさのタンパク質で1万数千程度であることである。(1)の違いはベクトルを対応付けるときに物性の類似度に応じて閾値を導入することで対応できる。(2)と(3)の違いはクリーク検索法を応用することで対応した⁸⁾⁻¹¹⁾。

クリークは、数学でいうところのグラフ(点と線の集合)の一種で、すべての点がお互いに線で結ばれた特殊なグラフである。この方法では、図2のように、ベクトルの対応付けと相対配置の類似性をグラフで表現してあるので、そのグラフの中から最も大きなクリークを探すことが、できるだけ多くのベクトルを対応付けるこ

とに対応している。つまりグラフの表現を工夫することにより、類似性検索問題をクリーク検索問題に置き換えてやるわけである。クリーク検索問題は、数学的には古くから研究され、1979年には効率的なアルゴリズムが提案されている¹⁰⁾。

5. 応用例～抗原認識のデータベース検索～

今回開発した方法の有効性を検証するために抗体タンパク質において抗原結合部位が正しく認識されるかどうかを調べた。この計算ではeF-site/Antibodyの中にある103個の抗体に対して、抗体タンパク質の抗原結合部位を切り出してきたものをクエリー構造として検索を行った。図3に4つの典型的な検索結果を示す。

D2.3抗体(図3A)やN1G9抗体(図3B)の例では、非結合状態のみならず、似た抗原に対する抗体もうまく検索することができた。50.1抗体(図3C)は図4Aに見られるように構造変化していることが知られているが、その構造変化は側鎖の配向のみが変化するごく局所的な構造変化であるため、本手法で類似性が検出できている。しかし、BV04-01(図3D)はうまく検出でき

ていない。この場合は、図4Bに見られるように、構造変化が結合部位全体に及んでいるため、本手法では類似性が検出できていない。このように今回開発した方法は、主鎖の大きな構造変化がない場合には、ある程度うまく類似機能表面を検出できていることがわかる。

6. 今後の展開

今回我々は、分子表面の構造と物性に着目して、機能部位のデータベース eF-site を構築し、そのデータベースに対する検索法の開発を行った。これにより原理的には、機能未知のタンパク質の立体構造が明らかにされたときにその機能部位の推定を行う枠組みができたことになる。今後は eF-site をさらに拡張していくとともに検索法の改善を進めていきたい。検索法の改善として、構造変化にいかに取り組むかが大きな課題である。抗体タンパク質の場合には構造変化する例はあまり多くなく、今の方法がうまく機能しているが、いろいろな機能を考える際には、構造変化の問題に正面から取り組む必要がある。現在考えられる方法として、非結

合状態の結晶構造からの分子動力学シミュレーションを利用した構造サンプリングを行い、構造変化した場合でも類似性を検索できる方法の開発を進めている。

文 献

- 1) Yura, K. <http://www.bio.nagoya-u.ac.jp:8001/~yura/wholegenome.html>
- 2) 中村春木 (1999) 蛋白質 核酸 酵素 **44**, 590-597.
- 3) Kinoshita, K. *et al.* <http://pi.protein.osaka-u.ac.jp/eF-site/>
- 4) Connolly, M. L. (1983) *Science* **221**, 709-713.
- 5) Nakamura, H. *et al.* (1987) *J. Phys. Soc. Jpn.* **56**, 1609-1622.
- 6) Bairoch, A. *et al.* (1997) *Nucleic Acids Res.* **25**, 217-221.
- 7) Kyte, J. *et al.* (1982) *J. Mol. Biol.* **157**, 105-132.
- 8) Nussinov, R. *et al.* (1991) *Proc. Natl. Acad. Sci. U S A* **88**, 10495-10499.
- 9) Artymiuk, P. J. *et al.* (1989) *J. Inform. Sci.* **15**, 287-298.
- 10) Bron, C. *et al.* (1973) *Communications of the ACM* **16**, 575-577.
- 11) Kinoshita, K. *et al.* (2001) *J. Struct. Funct. Genomics* **2**, 9-22.

木下賢吾 (きのした けんご)
 横浜市立大学総合理学研究科 助手
 連絡先: 〒230-0045 横浜市鶴見区末広町1-7-29
 E-mail: kinoshita@tsurumi.yokohama-cu.ac.jp
 中村春木 (なかむら はるき)
 大阪大学蛋白質研究所生体分子解析研究センター 教授
 連絡先: 〒565-0871 大阪府吹田市山田丘3-2
 E-mail: harukin@protein.osaka-u.ac.jp