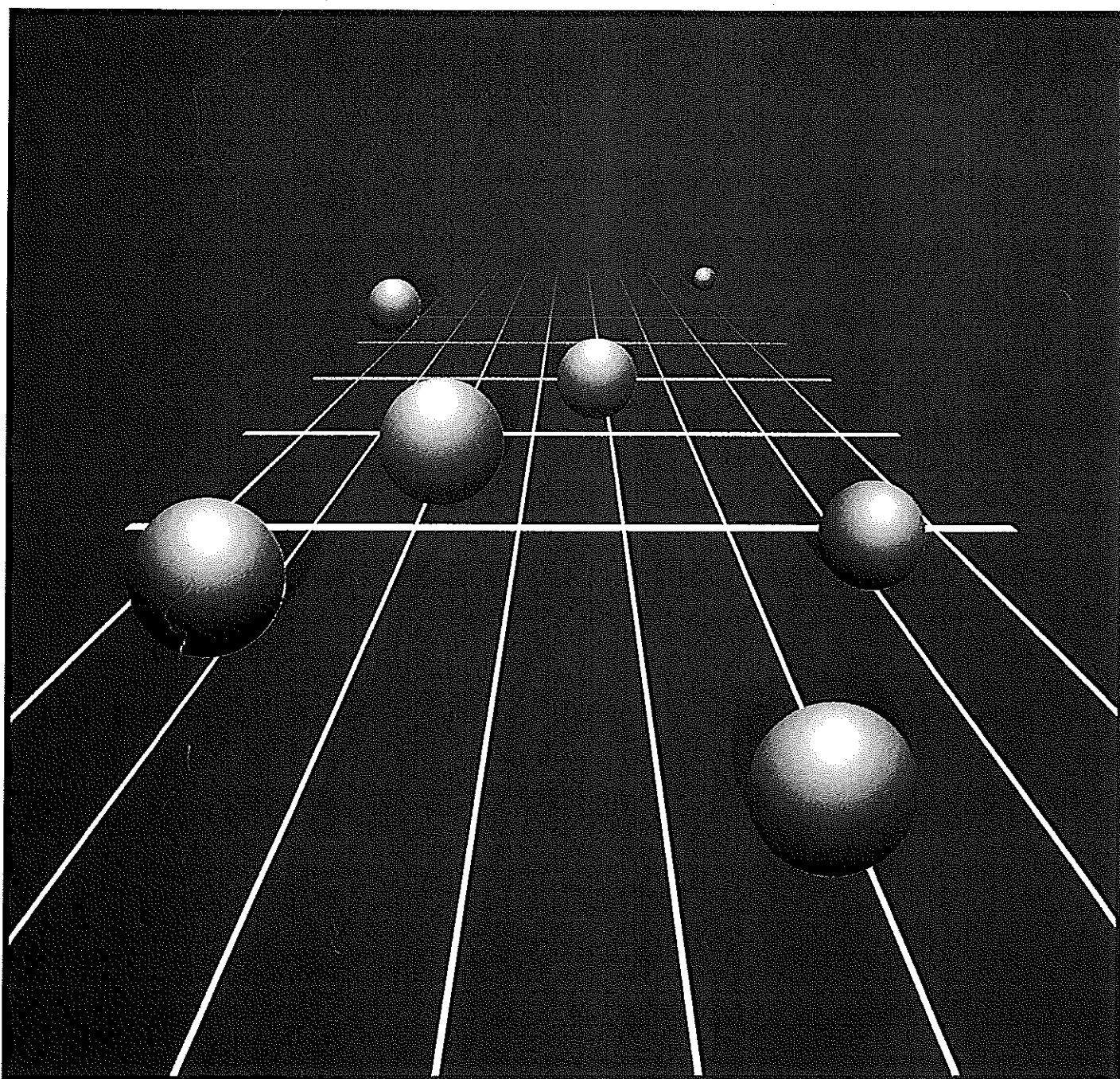


生体の科学

SEITAI  
NO  
KAGAKU

Vol.58 No.5  
2007 Sep.-Oct.

# タンパク質間相互作用



## 1. 総論

## 計算科学によるタンパク質間相互作用解析

竹田-志鷹 真由子 寺師 玄記 梅山 秀明

タンパク質立体構造情報データベースである Protein Data Bank (PDB ; <http://www.rcsb.org/pdb>) に登録されているタンパク質立体構造の数は急速に増加しており (2004 年 4,857 件, 2005 年 5,058 件, 2006 年 6,120 件), 現在 41,000 件を超えている。一方, タンパク質の機能を理解する上で重要なタンパク質複合体の立体構造情報は, 実験そのものの困難さや, 複合体を形成する相互作用の組み合わせが膨大になることもあり, 網羅的に実験構造を得ることは困難である。そこで, タンパク質間相互作用の研究において, コンピューターを用いてタンパク質複合体の立体構造を予測するドッキング解析 (protein-protein docking) は非常に有効である。本稿では, まず筆者らが開発したドッキング解析プログラム SKE-DOCK<sup>1)</sup> の基本的なアルゴリズムを紹介し, ドッキング解析の基本的な方法と精度について説明する。次に, タンパク質相互作用予測の国際コンテスト CAPRI (Critical Assessment of PRediction of Interactions)<sup>2)</sup> について紹介し, ドッキング解析の現状と展望を述べる。

## ● SKE-DOCK

二つのタンパク質立体構造を入力とする SKE-DOCK ([http://www.pharm.kitasato-u.ac.jp/bmd/files/SKE\\_DOCK.html](http://www.pharm.kitasato-u.ac.jp/bmd/files/SKE_DOCK.html)) は, 大きく分けて (1) geometric docking, (2) クラスタリング, (3) 構造の評価, (4) 側鎖の re-packing の四つの処理で構成されている (図 1)。通常 72 時間以内に計算を終え, 候補となる複合体の予測構造を 10 個提示する。各処理は次の通りである。

(1) geometric docking : SKE-DOCK サーバーで使用されている geometric docking は, 局所的

なタンパク質分子の表面を四角錐の五つの座標点 (パッチ) で表現し, 2 分子のパッチ同士のマッチングを高速に行い, 相補性の高いドッキング構造を検索する方法である。この方法を使用することで, 局所的なタンパク質分子の表面で凸部と凹部がうまく組み合わさったドッキング構造のみを効率的に検索することが可能となっている。当然, タンパク質間での衝突が多い構造は取り除かれる。その結果, 形の相補性順に上位 5000 個の候補構造が選択され, 次のステップへ進む。

(2) クラスタリング : geometric docking によって得られた 5000 個の予測構造を, 冗長性を削減するために相互の類似度によるクラスター解析を行う。その結果, 各クラスターの代表構造 3000-4000 個の候補構造に絞られる。

(3) 構造の評価 : 3000-4000 個の候補構造において, 最も安定な構造を選び出す評価を行う。PDB データセットから, 20 種のアミノ酸それぞれがどのような環境 (アミノ酸側鎖の埋没度, 極性原子の存在割合, 二次構造) に存在しやすいかを表した統計スコアを用い, 候補構造それぞれを評価することで順位付けを行う。

(4) 側鎖の re-packing : (3) によって選ばれた上位 10 個の構造に対して, アミノ酸残基の側鎖の再構築を行う。これは geometric docking の際に生じたアミノ酸残基の側鎖同士の衝突を解消するために行われる。

非常に単純な方法ではあるが, SKE-DOCK は入力構造から実験構造が大きな構造変化が生じていない複合体構造の予測では優れた予測を行うことができる。図 2 は, 後述する CAPRI で SKE-DOCK が予測した複合体構造と, 実験で解析され

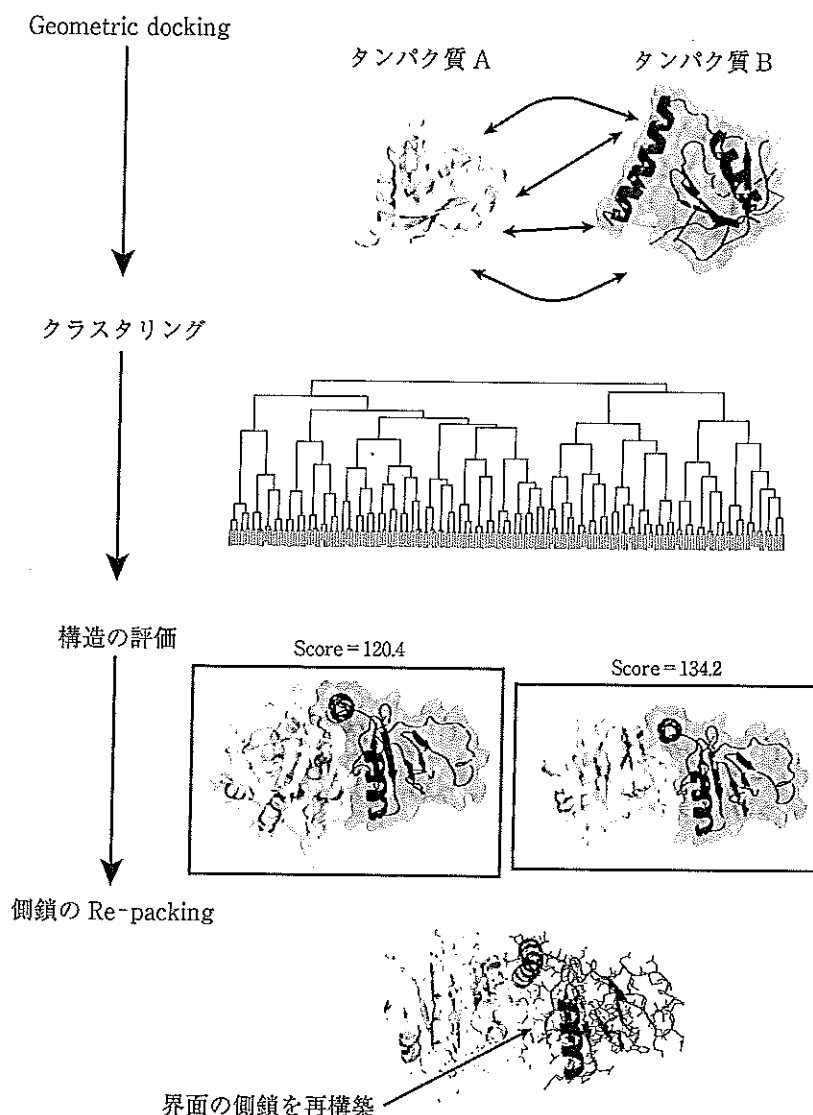


図 1 SKE-DOCK のフローチャート図

た複合体構造とを比較したものである。レセプター分子同士で重ねあわせを行った後の、タンパク質リガンド間の rmsd 値は 3.3Åであった。タンパク質複合体の立体構造予測を精度よく行うためには、上記の四つのステップのうち(1)と(3)に対応する二つの要素で精度の高い計算が行われる必要がある。一つはサンプリング法、もう一つはスコア関数である。次にそれぞれの要素について説明する。

●サンプリング方法

第1の要素、サンプリング方法に求められるのは、複合体の結合状態を密に精度よく探索することである。このサンプリング方法の網羅性が不十分であるならば、そもそも真実のタンパク質複合体の構造に近い候補を得ることは不可能となる。



図 2 SKE-DOCK server が予測したドッキング構造と実験構造(Arf1/ARHGap10, PDB entry 2j59)との比較  
予測構造：黒 実験構造：白

しかし網羅性を重視するあまり、サンプリング方法が非効率的で現在のコンピューター資源では計算することが不可能なほどあまりにも膨大な計算

コストになるならば、たとえ網羅性が高いサンプリング法であったとしても実用には適さない。したがってサンプリング方法には網羅性と効率性の両立が求められる。おもなサンプリング方法としては、Geometric Hashing 法、エネルギー最小化、モンテカルロ法、分子動力学法、高速フーリエ変換、主成分分析、遺伝的アルゴリズムなどがある。ZDOCK<sup>3)</sup>(<http://zlab.bu.edu/zdock/index.shtml>)は、タンパク質構造を3次元グリッドで表現し、高速フーリエ変換(FFT: Fast Fourier Transform)によって高速なドッキングを行っている。FFTを使用したドッキングプログラムは現在の主流の一つとなっており、FTDock<sup>4)</sup>、GRAMM<sup>5)</sup>、ClusPro<sup>6)</sup>(<http://nrc.bu.edu/cluster/>)でも使用されている。前述したSKE-DOCK、PATCHDOCK<sup>7)</sup>(<http://bioinfo3d.cs.tau.ac.il/PatchDock/>)は局所的なタンパク質表面を単純なパッチで表現し、それらをマッチングすることで高速なドッキングを行っている。ランダムな相対配置から、モンテカルロ法を用いて6次元(併進・回転)の検索を行っているものでは、Rosetta-Dock<sup>8)</sup>が挙げられる。

#### ●スコア関数

第2の要素、スコア関数とは、上記のサンプリングによって得られた多数の候補構造の中から実験構造、もしくはそれに最も近い構造を選び出すために使われるものである。もちろんスコア関数には、正しいタンパク質複合体を選び出す「精度」が求められるが、同時にnear native構造と呼ばれる実験構造に近いものをよい構造であると判定できる能力も求められる。特に、多数の候補構造からある程度の精度が期待されるものをいくつか選択し、選ばれたものだけについてさらにその周辺において密なサンプリングを行う場合には、near native構造を選択できる能力が求められるのである。したがってスコア関数はサンプリング方法と非常に密接な関係があるといえる。

スコア関数の代表例は、形の相補性、ファンデルワールスエネルギー、静電ポテンシャル、疎水エネルギー、幾何学的一致度、水素結合、アミノ酸残基ペアポテンシャル、クラスターサイズ、統計スコアなどである。これらのスコア関数は単独

に使われることはまれであり、多くの場合は組み合わせられて用いられる。例えば、タンパク質の性質に関連した疎水エネルギーや水素結合などを使わずに、幾何的なスコア関数(相補性など)のみ評価を行った場合、上位にくる構造の精度を期待することはできない。スコア関数を選ぶ際には、そのスコア関数をサンプリングでの目的関数として使うのか、最終的な順位付けを行うのかといった、どの用途に使用するかについても注意が必要である。SKE-DOCKを例にとるならば、相補性スコア関数がサンプリングでの目的関数と同時にフィルタリングとして用いられ、データベースに基づく(knowledge based)スコア関数が最終的な順位付けに用いられている。相補性スコアで一定のスコア以下の構造を除くというフィルタリングに用いることにより、相互作用領域が非常に少ないといった明らかに相互作用している複合体ではない予測構造を取り除くことができる。

#### ●タンパク質相互作用予測の国際コンテスト

##### CAPRI

CAPRI(<http://capri.ebi.ac.uk/capri.html>)とは、Joël Janinらにより2001年から行われているタンパク質-タンパク質相互作用予測の国際コンテストである。実験により解析されたタンパク質複合体構造情報をCAPRI参加者には隠した状態で、参加者各自のドッキング解析のアルゴリズムが正しく予測をすることができるかを比較するものである。参加チーム数は初年のRound1-2では19チームであったが、2007年のRound6-12では38チームに増え、ドッキング解析が国際的に関心の高い分野であることがわかる。

CAPRIでは通常、ターゲットとなる複合体を形成しているそれぞれの分子の座標情報がPDBフォーマットで参加者に与えられる。座標情報は複合体を形成するそれぞれの分子が単体の状態で構造解析されたもの(unbound構造)と、複合体を形成しているそのままの座標(bound構造)のどちらかが与えられる。現在までのところ、unbound構造同士のドッキングまたはbound構造とunbound構造のドッキングが出題されている。さらには、複合体を形成している分子の立体構造自体をホモロジーモデリング法などにより予測し

た後でドッキング予測を行わなければならないターゲットも出題される。また、1対1の2量体構造の予測に限らず、3量体構造の予測や、ホモダイマー構造の予測などもまれに出題される。これらのターゲットも比較的難易度が高いものに分類される。予測構造の精度の評価は、High, Medium, Acceptable, Incorrectの四つの段階に分けられる。

#### ●今後の展望

ドッキング解析における最大の問題の一つは、「タンパク質立体構造は、周りの環境に応じて柔軟に変化する」ことである。これはすなわちタンパク質の立体構造は単体で存在するときと、複合体としてほかのタンパク質と相互作用して存在するときとは異なるということの意味している。構造の変化の度合いは様々であり、相互作用部位に存在するアミノ酸側鎖部分のパッキングだけが変化するものや、ループ構造部分が増えるもの、ドメイン構造の配置が増えるものまで様々ある。前述したSKE-DOCKにおいては、側鎖の変化やループ構造の若干の変化程度であれば、geometric dockingに大きな影響を与えないことがわかっている。しかし、それ以上の大きな構造変化が相互作用部位で起こる場合には対応ができていないのが現状である。多くのドッキングプログラムにとって、タンパク質の大きな構造変化を扱う場合、計算量の増大が問題となってくる。大きな構造変化が起こるターゲットにおいて、構造変

化を考慮しないドッキング方法(rigid body docking)は、たとえスコア関数の精度が高くても精度の高い予測をすることができない。

2007年現在のCAPRIの結果から、相互作用部位で大きな構造変化が起こっていないターゲットでは、参加グループの半数近くがAcceptable以上の精度の予測をすることが明らかとなっている。それらのグループは、様々なサンプリング法とスコア関数を用いていることから、比較的やさしいターゲットにおいて、その手法に絶対的な優劣はもはやないと思われる。逆にいえば、ドッキング解析の精度は今後、大きな構造変化を含めた予測にいかに対応するかにかかっているといえる。多くの研究グループはすでにその試みを始めており、その成果が期待される。

#### 文献

- 1) Terashi G et al : *Proteins* (in press)
- 2) Janin J et al : *Proteins* 52 : 2-9, 2003
- 3) Pierce B et al : Structure Prediction of Protein Complexes, In Xu Y, Xu D, Liang J (Eds.), *Computational Methods for Protein Structure Prediction and Modeling Volume 2 : Structure Prediction*, pp 109-134, Springer, New York, 2007
- 4) Gabb HA et al : *J Mol Biol* 272 : 106-120, 1997
- 5) Katchalski-Katzir E et al : *Proc Natl Acad Sci USA* 89 : 2195-2199, 1992
- 6) Comeau SR et al : *Bioinformatics* 20 : 45-50, 2004
- 7) Schneidman-Duhovny D et al : *Proteins* 52 : 107-112, 2003
- 8) Gray JJ et al : *J Mol Biol* 331 : 281-299, 2003