

戦略的創造研究推進事業
発展研究（SORST）

研究終了報告書

研究課題「立体構造情報を利用した蛋白質間
相互作用予測法の開発」

研究期間：平成16年12月1日～
平成20年3月31日

木下賢吾
（東京大学医科学研究所、准教授）

1. 研究課題名

立体構造情報を利用した蛋白質間相互作用様式の予測法の開発

2. 研究実施の概要

背景

近年、構造ゲノム科学プロジェクトの進展に伴い数多くのタンパク質立体構造が、その機能が同定される前に解明されている。生物を分子レベルで理解するためにはタンパク質の機能を同定する必要があるが、これを実験的に行うには多大な時間と労力を要する。そのため、計算による機能予測が重要視されてきた。タンパク質の機能は、既に知られているタンパク質との配列相同性によって進化的類縁関係をたどり機能を予測する方法が広く用いられているが、既に知られているタンパク質との配列相同性が低い構造も多く決定されており、配列-機能相関を用いた機能予測には限界があることが次第に明らかになってきた。そこで、配列が構造を決め構造が機能を決めるという原理に基づき、タンパク質の立体構造情報から直接、機能を予測する方法を開発する必要が生じてきた。

そこで本研究では、タンパク質の立体構造から機能を予測する方法の確立を目指し、タンパク質分子表面上における物理化学的性質（特に静電ポテンシャル）と表面の形状に着目し、タンパク質の高次機能を決定しているタンパク質間相互作用を予測する方法の開発を行った。

ねらい・着眼点

タンパク質間相互作用を明らかにすることは、生物を分子レベルから理解する基本的なステップである。そのため数多くの実験がなされ、どのタンパク質ペアが相互作用するかに関しては多くの情報が得られつつある。しかし、それらが「どのように」相互作用するかに関しては、近年複合体の結晶構造解析などにより多くのタンパク質で明らかにされつつあるものの、「どのタンパク質ペアが」という情報に比べると圧倒的に少ない。タンパク質分子がどのように相互作用するかを知ることは、タンパク質間相互作用の理解のみならず、その相互作用の制御などの応用を念頭に置くと、生物学的あるいは医学・薬学的にも非常に重要なステップである。

研究内容

タンパク質複合体には大きく分けて、同じサブユニットから構成されるホモオリゴマーと異なる数種のサブユニットから構成されるヘテロ複合体が存在する。これらの異なる種類の複合体は、機能及び構造のどちらの観点でも異なる側面を持っている。近年の構造解析の進展や、2007年3月に成功裏に終了したたんぱく3000プロジェクトやその後のターゲットタンパク質などの構造ゲノム科学プロジェクトにより、数多くのタンパク質立体構造が明らかにされ、また複合体の構造の数も増えては来ている。しかしその数はやはりまだまだ十分とはいえない。特にヘテロ複合体の数は少なく、2007年5月最初の時点で、全PDBのエントリー（約41000エントリー）中、約15%しかない。一方ホモ複合体は、約36%を占めヘテロ複合体に比べて大分構造情報が蓄積していることが見て取れる。このような現状を踏まえて、ホモ複合体に関してはどのように相互作用しているのかを比較分類したDBの構築を行い、ヘテロ複合体に関しては、新しいドッキング手法の開発を行うこととした。なお、これらの手法はどちらも、本研究に先行して行ったPRESTO研究による分子表面の形状と物性（静電ポテンシャル及び疎水性）を集めたデータベースeF-siteを基礎としている。

ホモ複合体

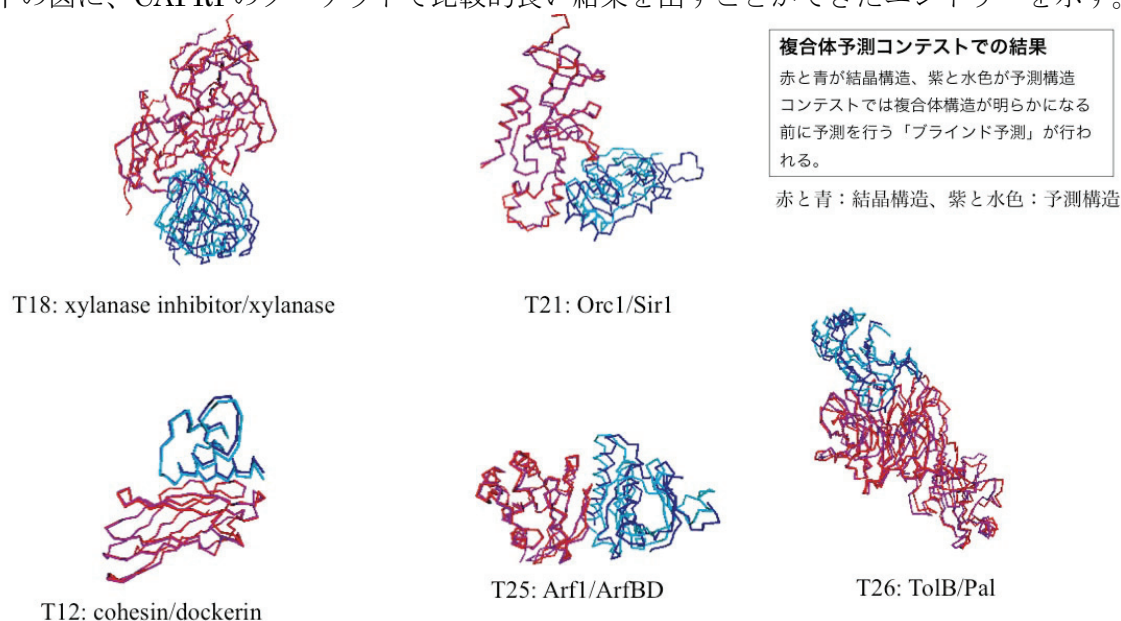
先に述べたように、ホモ複合体に関してはかなりの立体構造情報が利用できるようになってきた。そこで我々はホモ複合体の相互作用面の比較・分類を行い（<http://pre-s.protein.osaka-u.ac.jp/~classppi>）、生物学的な相互作用面と結晶構造解析の

弊害であるクリスタルコンタクトによる相互作用を区別する手法の開発へと発展させた (<http://pre-s.protein.osaka-u.ac.jp/~prebi>)。

相互作用面の分類と物性との相関の解析により、ホモオリゴマーの形成にかかわる 3 つのルールを見いだすことができた。(1) 巻き付き型では複雑な立体構造であるにもかかわらず、すべての物性に関して相補性が高い。さらに、ほとんどの例において活性部位や DNA 結合部位が相互作用面に存在する。以上のことから、**twisted-dimer** 型相互作用面を持つホモ複合体は、機能を発揮するためによくデザインされた相互作用面を持つ複合体を形成していると考えられる。(2) 並行型では、基本的に静電相互作用を用いることができない。これは、このタイプの相互作用面が静電ポテンシャルでの相補性を持つためには、相互作用面の対称性も同時に満たすために、非常に狭い範囲に異なる符号を持った静電ポテンシャル分布を持たなくてはならないことになるが、狭い範囲に異なる電荷分布を持つと、強い電場を持つことになり、安定性などの観点で振りになることが考えられる。(3) 相互作用面に対称性が存在する必要のないオリゴマータイプでは、相互作用面の片側に正または負どちらかの静電ポテンシャルが分布しており、正負両方が同時に分布することはほとんどない。これは、相互作用面に対称性を持たなくてはならない他の相互作用面とそうではないオリゴマータイプの決定的な違いである。

ヘテロ複合体

ヘテロ複合体に関しては、分子表面の形状と物性（静電ポテンシャルと疎水性度）の相補性に着目したドッキング手法を新たに開発した。これらの手法を利用して、複合体の予測コンテスト (CAPRI) に参加を行った。本研究期間を開始して時点で既に CAPRI は数回行われ、問題(ターゲットと呼ぶ)数は既に 20 個になっていた。そこで、最初は過去問から開始し、ターゲット 21 番（以下、T21 などと略記）からリアルタイムで参加した。以下の図に、CAPRI のターゲットで比較的良好な結果を出すことができたエントリーを示す。



図を見ても分かるように、予測と実際の結果の区別が困難なぐらい良い結果を得ることができたターゲットもあったが、ここに示していないターゲットに関しては良い予測結果を得ることができなかった。また、ここに示したターゲットに関しても、候補構造としては上げることができたが、コンテストの際に答えとして選ぶことができる 10 個には選ぶことができなかった。言い換えると、我々の予測のアプローチでは、候補構造の探索は非常に良く機能しているが、構造を選ぶ段階にまだ未成熟な点があることがわかったことを意味する。

まとめ

タンパク質複合体の予測は非常に重要な分野であり、世界的にも数多くのグループが取り

組んでいる問題である。しかし、CAPRI の結果などを見ると、一つのグループがぬきんでて良い予測をできている訳ではなく、あるターゲットに対して良い予測をしたグループも、別のターゲットでは非常に良くない結果を出しているのが現状である。その点では我々の方法も、ぬきんでて良い手法とまでは行かないが、いくつかのターゲットで良い結果を出していること、及び手法の新規性の高さでは注目を集めており、実際 2007 年 4 月に行われた CAPRI の評価会議では数少ない口頭発表に選ばれる結果となった。今後は、タンパク質間相互作用だけでなく、タンパク質 DNA/RNA との複合体予測への発展や、タンパク質構造の柔らかさを考慮した拡張などやるべき課題はまだまだあるが、当初の目標であったタンパク質複合体の構造予測としては良い成果を上げることができたと考えている。

3. 研究構想

当初は、PRESTO で行った研究の延長として、複合体の構造予測を、類似性検索を基礎として展開する事をもくろんでいた。しかし実際に複合体の構造の解析を行うと、ホモ複合体の解析で明らかにしたように、相互作用の様式は非常に多様であった。また、利用できるヘテロ複合体の数も多くはなく、方向転換を余儀なくされた。これに対して、PRESTO 時代からの基盤としていた分子表面形状と物性に着目した手法で、新規の複合体構造予測法を開発し、それを実際に複合体の予測コンテスト CAPRI に応用することで良好な成績を修めることができた。しかしながら、まだまだ課題も多く残っている。特に、タンパク質複合体の天然構造が予想以上に疑似構造と差がない事は、単純に予測を困難にしている理由を明らかにした以上に、タンパク質相互作用の本質について考えられる結果を得たと考えている。本研究課題は複合体の予測法の開発なので、その観点では一定の成果を得たが、今後はタンパク質間相互作用の本質をより明らかにしていきたいと考えている。タンパク質は生体分子で有るとはいえ、物理化学の法則には従っているので、ホモ複合体の解析で明らかにしたような、複合体の形成ルールがヘテロ複合体でも存在するはずであるので追求していきたいと考えている。

4. 研究実施内容

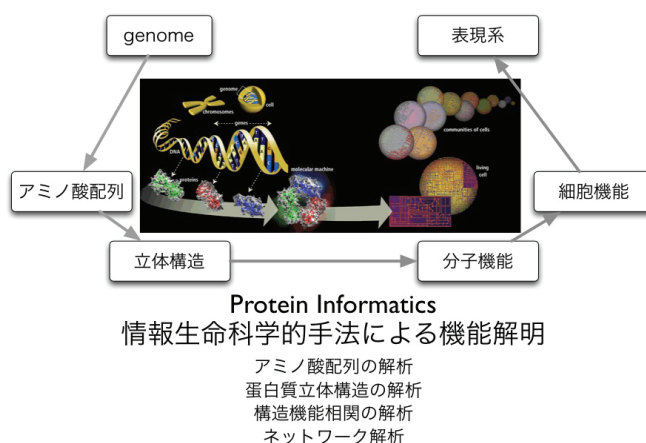
(1) 実施の内容

研究の背景

近年、たんぱく 3000 などの構造ゲノム科学プロジェクトの進展に伴い、数多くのタンパク質立体構造がその機能が同定される前に解明されている。生物を分子レベルで理解するためにはタンパク質の機能を同定する必要があるが、これを実験的に行うには多大な時間と労力を要するため、計算による機能予測が重要となってくる。これまでタンパク質の機能は、既に知られているタンパク質との配列相同性によって進化的類縁関係をたどり機能を予測する方法が広く用いられているが、既に知られているタンパク質との配列相同性が低い構造も多く決定されており、配列-機能相関を用いた機能予測には限界があることが次第に明らかになっている。そこで、配列が構造を決め構造が機能を決めるという原理に基づき、タンパク質の立体構造情報から直接、機能を予測する方法を開発する必要性が生じてきた。

タンパク質の機能には構造から決まる分子機能と、たくさんが相互作用ネットワークを形成することで達成されている生物学的な機能（細胞機能）がある。本研究では生物における情報の流れに従って、PRESTO 時代には構造情報からの低分子の結合部位の予測を行い、本発展研究では生物学的機能の第一歩となるタンパク質間相互作用を対象として、その複合体構造を予測法の開発を行った。

これを受け本研究では、タンパク質の立体構造から機能を予測する方法の確立を



目指し、タンパク質分子表面上における物理化学的性質（特に静電ポテンシャル）と表面の形状に着目し、タンパク質の高次機能を決定しているタンパク質間相互作用を予測する方法の確立を目指す。

研究内容

タンパク質複合体には大きく分けて、同じサブユニットから構成されるホモオリゴマーと異なる数種のサブユニットから構成されるヘテロ複合体が存在する。これらの異なる種類の複合体は、機能及び構造のどちらの観点でも異なる側面を持っている。機能面では、前者は各サブユニットが同じ機能を担っており、複合体を形成するとそれらの機能の間に調整メカニズムが発生するといった関係があることが多いが、後者の場合は、異なる機能を担うサブユニットが一時的に相互作用して相手に変化をもたらす場合が多い。そのため、複合体の構造を考える際には考えるべき問題が異なる。ホモ複合体では同じサブユニットの繰り返しなので、X線結晶構造解析を行った際に、その相互作用が結晶学的な相互作用なのか生物学的に意味のある相互作用なのかを判別することが大きな問題になる一方、ヘテロの場合は、そもそも複合体の安定性が高くないのでX線結晶構造解析などの実験的手法での構造解析が困難である。このように、ホモとヘテロでは全く異なる側面を持つので、その相互作用面の特徴はそれぞれ異なっていると考えられる。本研究ではまずホモ複合体での複合体構造予測法の開発を行い、続いてヘテロ複合体での複合体構造予測法の開発という順序で解析を進めてきた。

近年の構造解析の進展や、2007年3月に成功裏に終了したたんぱく3000プロジェクトやその後のターゲットタンパク質などの構造ゲノム科学プロジェクトにより、数多くのタンパク質立体構造が明らかにされ、また複合体の構造の数も増えては来ている。しかしその数はやはりまだまだ十分とはいえない。特にヘテロ複合体の数は少なく、2007年5月最初の時点で、全PDBのエントリー（約41000エントリー）中、約15%しかない。一方ホモ複合体は、約36%を占めヘテロ複合体に比べて大分構造情報が蓄積していることが見て取れる。このような現状を踏まえて、ホモ複合体に関してはどのように相互作用しているのかを比較分類したDBの構築を行い、ヘテロ複合体に関しては、新しいドッキング手法の開発を行うこととした。なお、これらの手法はどちらも、PRESTO時代に開発を行った、分子表面の形状と物性（静電ポテンシャル及び疎水性）を集めたDB、eF-siteを基礎としている。

ホモ複合体の相互作用面に関する研究

先に述べたように、ホモ複合体に関してはかなりの立体構造情報が利用できるようになってきた。そこで我々はホモ複合体の相互作用面の比較・分類を行い（<http://pre-s.protein.osaka-u.ac.jp/~classppi>）、生物学的な相互作用面と結晶構造解析の弊害であるクリスタルコンタクトによる相互作用を区別する手法の開発へと発展させた。

ホモ複合体の分類は相互作用面の分子表面形状及び静電ポテンシャルと疎水性度の相補性に着目して行った。データセットとしては、2004年2月の時点でのPDBを利用し、結晶構造解析により構造が解かれ、その解像度が $2.5\text{\AA}$ より良いエントリー8,609エントリーに対して行った。まず、それぞれの複合体エントリーに対してすべての相互作用しているタンパク質鎖ペアを同定し、その配列の比較を行う。それらペアの配列一致度が85%以上あるものをホモインターフェース、それ以外をヘテロインターフェースとした。その結果、13,021のホモインターフェースと5,442のヘテロインターフェースを同定できた。これらを最初にSCOPデータベースを用いてファミリー分類を行い、867ファミリーにグルーピングを行った。次に、これらをPDBの記述に従って467の生物学的に意味のある相互作用面を同定した。この中から63個の記述ミスと考えられるエントリーを除き、さらに複数箇所接触しているエントリー11個を除いて、最終的に393インターフェースの構造分類を行った。

ホモオリゴマーの相互作用面は、まず大きく2つに分けることができた。相互作用面がタンパク質の異なる面かできているオリゴマータイプと同じ面からできているダイマータ

イプである。これは、必ずしも全体がオリゴマーかダイマーかという区別とは一致しない。例えば、4量体タンパク質でダイマーが2つ結合してできているようなタイプでは、ダイマータイプインターフェースとオリゴマータイプインターフェースの両方を持つことがあるし、相互作用様式によってはすべてがオリゴマータイプの事もあり得る。一方2量体では必ずダイマータイプであるし、3量体では必ずオリゴマータイプのインターフェースとなる。

次にダイマータイプのインターフェースは対称軸を持つことから、対称軸に対する広がり具合で3種類に分類することができる。相互作用面が特定の方向に広がっていない、つまり円形をしている円形タイプ、2階軸方向に平行に広がっている並行タイプ、垂直に広がっている垂直タイプである。ただし、相互作用面が複雑に絡み合っているために、このような分類に適さない27例は最初に巻き付き型として別に分類を行った。詳細は割愛するが、これらの分類はすべてインターフェースの特徴量を計算し、自動で定量的に行っている。以上の分類の結果は図にまとめてある（括弧内の数字が各分類に属するインターフェースの数）。

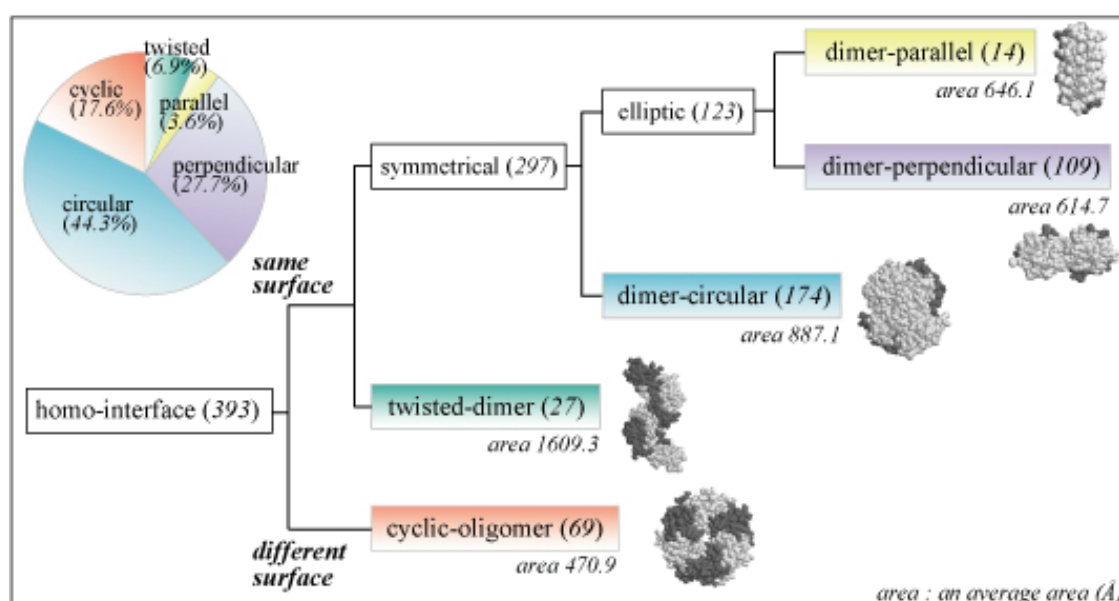
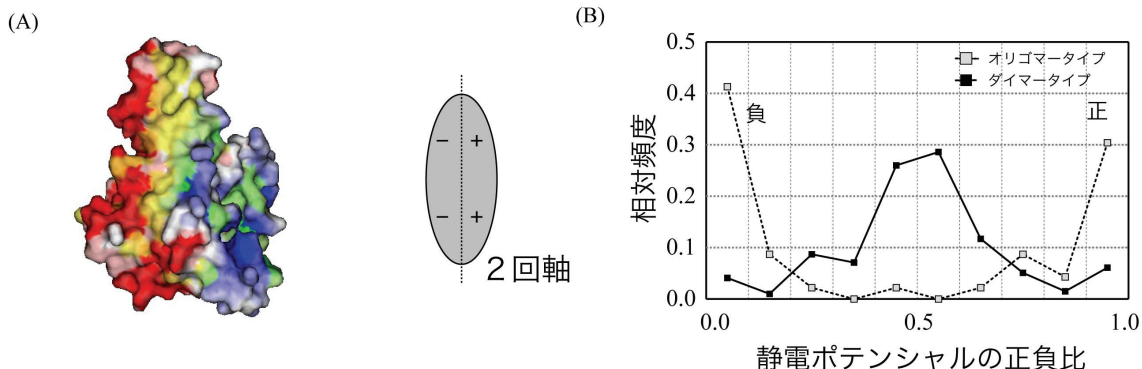


Figure classification of homo protein-protein interfaces

さて以上のような分類により、ホモタンパク質がどのように相互作用するのかというバラエティを整理することができた。現在分かっているホモ複合体では、ダイマータイプの相互作用面が比較的多く、その中では円形タイプ、垂直タイプが主な相互作用タイプであり、並行タイプはまれにしか見られない。それでは、これらの分類とインターフェースの物理化学的な特徴には関係があるのだろうか？この疑問に答えるべく、インターフェースの物理化学的な特徴として、静電ポテンシャルと疎水性度の相補性と構造分類の相関を解析した。その結果、ホモオリゴマーの形成にかかわる3つのルールを見いだすことができた。(1) 巻き付き型では複雑な立体構造であるにもかかわらず、すべての物性に関して相補性が高い。さらに、ほとんどの例において活性部位やDNA結合部位が相互作用面に存在することを確認した。以上のことから、**twisted-dimer**型相互作用面を持つホモ複合体は、機能を発揮するためによくデザインされた相互作用面を持つ複合体を形成していると考えられる。(2) 並行型では、基本的に静電相互作用を用いることができない。これは、このタイプの相互作用面が静電ポテンシャルでの相補性を持つためには、相互作用面の対称性も同時に満たすために、非常に狭い範囲に異なる符号を持った静電ポテンシャル分布を持たなくてはならないことになるが、狭い範囲に異なる電荷分布を持つと、強い電場を持つことになり、安定性などの観点で振りになることが考えられる（下図のA）。(3) 相互作用面に対称性が存在する必要のないオリゴマータイプでは、相互作用面の片側に正または負どちらかの静

電ポテンシャルが分布しており、正負両方が同時に分布することはほとんどない(下図 B)。これは、相互作用面に対称性を持たなくてはならない他の相互作用面とそうではないオリゴマータイプの決定的な違いである。



ヘテロ複合体構造の予測

以上のように、ホモ複合体ぐらい数が増えると体系的に分類し、その結果を解析することで相互作用の構築原理の一端をかいま見ることができた。一方先に述べたように、ヘテロ複合体では構造が解析されているタンパク質はまだ多くはない。そのためヘテロ複合体では「一時的な相互作用をするハブタンパク質」という特殊なケースでの構造的基盤の解析などを行うことはできたが、一般的な分類という観点では明確な方針と複合体形成のルールは見えてこないのが現状である。そのため本研究では、ドッキング法によりタンパク質複合体予測法の開発を行った。

この方法の開発にあたって我々がこれまで開発を行ってきた表面に着目し、表面形状の相補性度と配列上の保存部位情報を利用した。この方法は2段階からなる。第1段階は、計算の速い形状の相補性と表面の重要度からなる評価関数(下図)を用い、モンテカルロ法と遺伝的アルゴリズムを用いて数多くの候補をリストアップする。第2段階では、候補構造の中から静電相互作用や疎水性度と言ったより計算量のかかる方法を利用して評価を行い、その評価と文献情報などを総合して、より尤もらしい複合体を選ぶ。

評価関数 $F = aF_1 + bF_2$

$$F_1 = \sum_{\text{all vertex pairs}} I^2 \times w_1^2 \times w_2^2 \times D$$

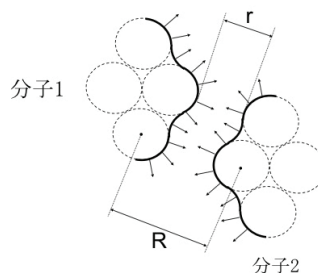
$$I = (\vec{n}_1 \times \vec{n}_2 + 1) / 2$$

$$D = \begin{cases} 0.25 / r^2 & (r > 0.5) \\ 1.0 & (r \leq 0.5) \end{cases}$$

w_1, w_2 : weight of conservation, $\vec{n}_1, \vec{n}_2$: normal vector
 r : distance between vertices

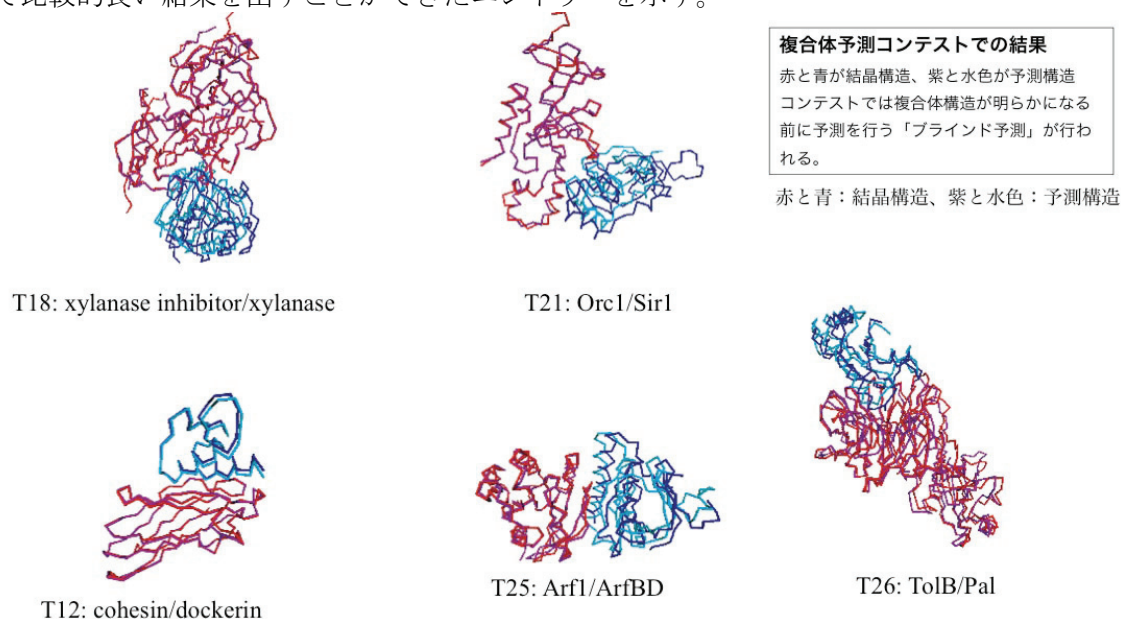
$$F_2 = \sum_{\text{atom pairs}} 4.0((\sigma / R)^6 - (\sigma / R)^{12})$$

R : distance between atoms



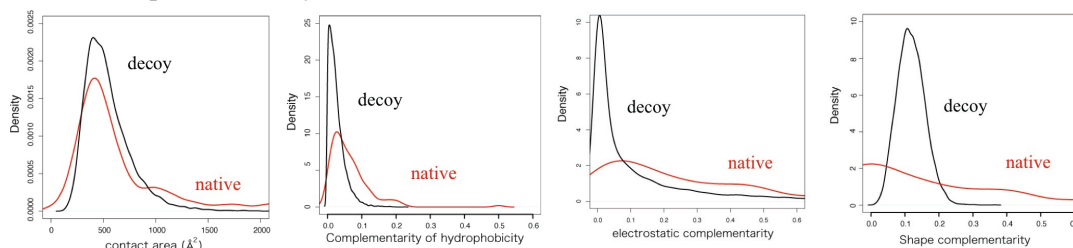
これまで、これらの手法を利用して、複合体の予測コンテスト (CAPRI) に参加を行った。CAPRIは毎年2から3回程度行われる予測のブラインドコンテストである。ブラインドコンテストというのは、予測結果の評価がおこなわれるまで複合体の構造は一切公開されず、参加者がすべて等しい条件で各々の方法の評価することを意味する。そのため、これまでのように方法の開発者が自前の評価で善し悪しを議論していたのとは異なり、非常に客観的に評価を行うことができる。予測は多くの場合、非結合状態のタンパク質の構造が提供され開始されるが、場合によっては、参加者が自分でホモロジーモデリングをする必要があるきわめて困難な例や結合状態から予測をすれば良い比較的容易な例まで多岐にわたる。本研究期間を開始して時点で既にCAPRIは数回行われ、問題(ターゲットと呼ぶ)数は既に20個になっていた。そこで、最初は過去問から開始し、ターゲット21番(以

下、T 2 1 などと略記) からリアルタイムで参加した。以下の図に、CAPRI のターゲットで比較的良好な結果を出すことができたエントリーを示す。



図を見ても分かるように、予測と実際の結果の区別が困難なぐらい良い結果を得ることができたターゲットもあったが、ここに示していないターゲットに関しては良い予測結果を得ることができなかった。また、ここに示したターゲットに関しても、候補構造としては上げることができたが、コンテストの際に答えとして選ぶことができる 10 個には選ぶことができなかった。言い換えると、我々の予測のアプローチである第 1 段階目は多くの場合非常に良く機能しているが、第 2 段階目の構造を選ぶ段階にまだ未成熟な点があることがわかったことを意味する。

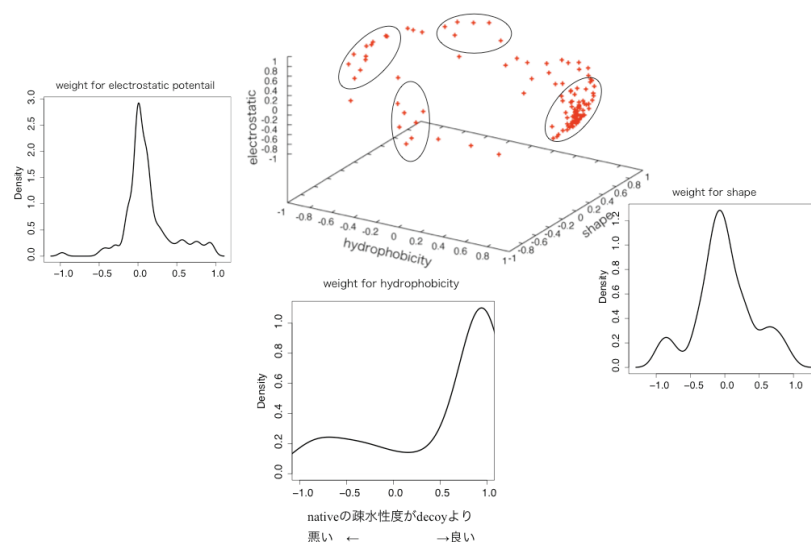
解析の詳細は割愛するが、最大の問題は、配列の保存度の計算にあることがわかってきた。そこで、ホモ複合体の構造予測でうまく機能した「物理化学的な特徴の相補性度」をヘテロ複合体に応用を行った。具体的には、すでに複合体構造がわかっているヘテロ複合体に対して、我々の手法を利用して、偽の複合体を人工的に発生させ、偽の構造を本物の構造を区別できるように評価関数の構築を行った。非冗長なヘテロ 2 量対 125 個に対して、1 複合体体当たり 94.3 個、総数で 11,791 個の偽構造の構築を終わり、中身の解析を行った。その結果、驚くべき事に偽構造が天然構造と比べても非常に良い複合体が多数見つかった(下図、complementarity は各物性の相補性度を定量化した指標 (後述))



これは、ヘテロ複合体では不安定な複合体を作っていることと考え合わせると天然構造が必ずしも良くできていないという点でおもしろい知見であり、次の相補性評価関数(*comp*)の開発(後述)と共に現在論文を準備中である。

次に、物理化学的な物性(本研究では静電ポテンシャルと疎水性度)も考慮するために、ホモ複合体の解析の際に考案した相補性度 ($comp = (W_h H_c + W_e E_c + W_s S_c)$, W_h , W_e , W_s はそれぞれ、疎水性度、静電ポテンシャル、形状の違いを規格化する重みパラメータ)を利用して、天然構造と偽構造を区別する関数の開発を行った。その際重要なポイントは、従来はすべての複合体に利用できる評価関数を構築することがなされていたが、ここではタンパク質の多様性を評価できるようにタンパク質毎に重みパラメータを決めた点である。

理由は、多様な相互作用面では、ホモ複合体の相互作用面の解析で明らかにしたように、ある相互作用面は静電ポテンシャル支配的であり、別の相互作用面では疎水相互作用が支配的であるという事が起こりうるので、一つの評価関数では評価が難しいと考えたからである。そこで、各タンパク質複合体に対して、天然構造と疑似構造のセットを利用して、天然構造の相補性度が尤も良くなるようにウェイトパラメータの最適化を行った（結果は下図）。



偽構造を用いて決定したパラメータの分布

その結果、(1) 31%のタンパク質でしか、天然構造の相補性度が最大にならないこと及び、(2) 最適化したパラメータは非常に様々であることが分かった。1 点目は、先ほど各物性の相補性度の分布で示唆された以上に、天然構造は疑似構造に比べて「良くできた」相互作用部位を持っていないを意味していると考えられる。この事は、我々の用いている相補性度がまだ十分に成熟したモノでない可能性もあるが、天然構造がそれほどしつかりした相互作用面を持っていないことを意味しているように思える。ヘテロ複合体の場合、多くの場合は一時的な相互作用であることを考えると、あまりにしつかりした（安定した）相互作用面を作ることは、その機能上不利益であることを示唆している可能性があり興味深い。一方 2 点目は、ヘテロ複合体の相互作用面が非常に多様であること、言い換えるとヘテロ複合体の数が十分で無い事を示している。立体構造情報は近年急激な勢いで増えてはいるが、複合体、特にヘテロ複合体の構造はまだ十分ではないのではないだろうか？

まとめ

本研究では、タンパク質複合体構造の予測法の開発を行った。複合体の予測は、どのような複合体を作る可能性があるかをリストアップするステップと、その中からより尤もらしい複合体を選ぶステップからなるが、後者の問題を考えるにあたって、形成メカニズムが大きく異なると考えられる、ホモ複合体とヘテロ複合体に分けて研究を行った。ホモ複合体では十分な数の複合体を得ることができたので、統計解析により複合体形成のルールを見いだすことができた、一方、ヘテロ複合体では、構造情報の少なさから多様であるという事実とヘテロ複合体の天然構造が比較的弱い相互作用である可能性を示すこととなった。また、複合体予測コンテストCAPRIへ参加を行った結果、良い結果を出すことができ、我々の開発した複合体の予測手法が世界的に見ても高いレベルに有ることを示すことができた。

(2) 得られた研究成果の状況及び今後期待される効果

今回開発を行ったタンパク質分子表面の相補性に基づく複合体の構造予測法は現在も改良を続けている。直近の課題としては、次回のCAPRIへの参加を行うために、タンパク

質・RNA/DNAの複合体構造予測に対応する事を予定している。タンパク質・DNAに関しては、PRESTOの開発の一環として評価関数は開発を行っているので、候補構造のリストアップの部分でわずかな改良により対応できる予定である。その後は、タンパク質の柔軟さを入れた複合体予測への拡張も予定している。しかしすべての自由度を考えるのは現実的でないので、相互作用面の側鎖のロータマーのみを変化させることから始める予定である。これらの改良と並行して、複合体予測のwebサーバとしても公開の可能性を考えている。PRESTOの際に開発を行った低分子の結合部位の予測をwebサーバeF-seekとして公開した経験を元に、タンパク質複合体予測のサーバ構築も可能であると考えているが、計算量が低分子の比ではないので、現在eF-seekで行っている阪大蛋白研とヒトゲノム解析センター間でのGridをより積極的に利用することが必要になってくると思われる。これら一連の改良とサーバの構築を行うことによって、今後は創薬などのより実用的な応用に寄与することが期待される。

5. 類似研究の国内外の研究動向・状況と本研究課題の位置づけ

タンパク質複合体の予測は非常に重要な分野であり、世界的にも数多くのグループが取り組んでいる問題である。しかし、CAPRIの結果などを見ると、一つのグループがぬきんでて良い予測をできている訳ではなく、あるターゲットに対して良い予測をしたグループも、別のターゲットでは非常に良くない結果を出しているのが現状である。その点では我々の方法も、ぬきんでて良い手法とまでは行かないが、いくつかのターゲットで良い結果を出していること、及び手法の新規性の高さでは注目を集めており、実際2007年4月に行われたCAPRIの評価会議では数少ない口頭発表に選ばれる結果となった。

タンパク質複合体の予測は、学問的に重要なだけでなく創薬の観点でも最近は大きな注目を集めるようになってきた。従来は創薬のターゲットとなるタンパク質の機能部位は低分子との相互作用部位が主なモノであったが、最近ではタンパク質間相互作用を阻害する薬剤の開発がすすんでいる。その観点でも、我々が開発を行った手法は今後ますます重要になっていくと考えている。

今後は、タンパク質間相互作用だけでなく、相互作用ネットワークにどう取り組むかが重要な課題になっていくと思われるが、その際にも、各相互作用の信頼性が重要であり、その観点でも我々の手法が重要になることが期待できる。

6. 研究実施体制

個人型研究なので該当せず。

7. 研究期間中の主な活動

(1) ワークショップ・シンポジウム等

年月日	名称	場所	参加人数	概要
なし				

(2) 招聘した研究者等

氏名（所属、役職）	招聘の目的	滞在先	滞在期間
なし			

8. 発展研究による主な研究成果

(1) 論文発表 (英文論文 15件 邦文論文 4件)

1. S Daron, A Kinjo, K Kinoshita, H Nakamura, Protein Structure Databases with new web services for structural biology and biomedical research, *Brief. Bioinform*, in press
2. H Chiba, R Yamashita, K Kinoshita, and K Nakai, Weak correlation between sequence conservation in promoter regions and protein-coding regions of human-mouse orthologous gene pairs, *BMC Genomics* in press
3. ○T Ikura, K Kinoshita, and N Ito, A cavity with an appropriate size is the basis of the PPIase activity, *PEDS*, in press
4. T Obayashi, S Hayashi, M Shibaoka, M Saeki, H Ohta and K Kinoshita, COXPRESdb: a database of coexpressed gene networks in mammals, *Nucleic Acids Res*, **36**, D77-82, 2008
5. ○M Higurashi, T Ishida and K Kinoshita, Identification of transient hub proteins and gentle possible structural basis for their multiple interactions, *Protein Sci*, **17**, 72-78, 2008
6. ○E Kanamori, Y Murakami, Y Tsuchiya, DM Standley, H Nakamura and K Kinoshita, Docking of protein molecular surfaces with evolutionary trace analysis, *Proteins*, **69**, 832-838, 2007
7. T Obayashi, K Kinoshita, K Nakai, M Shibaoka, S Hayashi, M Saeki, D Shibata, K Saito and H Ohta, ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis, *Nucleic Acids Res*, **35**, D863-869, 2007
8. R Koike, K Kinoshita and A Kidera, Probabilistic alignment detects remote homology in a pair of protein sequences without homologous sequence information, *Proteins*, **66**, 655-663, 2007
9. ○K Kinoshita, Y Murakami and H Nakamura, eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape, *Nucleic Acids Res*, **35**, W398-402, 2007
10. T Ishida and K Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence, *Nucleic Acids Res*, **35**, W460-464, 2007
11. Y Hosaka, M Iwata, N Kamiya, M Yamada, K Kinoshita, Y Fukunishi, K Tsujimae, H Hibino, Y Aizawa, A Inanobe, H Nakamura and Y Kurachi, Mutational analysis of block and facilitation of HERG current by a class III anti-arrhythmic agent, nifekalant, *Channels*, **1**, 198-208, 2007
12. ○Y Tsuchiya, K Kinoshita and H Nakamura, Analyses of homo-oligomer interfaces of proteins from the complementarity of molecular surface, electrostatic potential and hydrophobicity, *Protein Eng Des Sel*, **19**, 421-429, 2006
13. Y Tsuchiya, K Kinoshita, N Ito and H Nakamura, PreBI: prediction of biological interfaces of proteins in crystals, *Nucleic Acids Res*, **34**, W320-324, 2006
14. N Sierro, T Kusakabe, KJ Park, R Yamashita, K Kinoshita and K Nakai, DBTGR: a database of tunicate promoters and their regulatory elements, *Nucleic Acids Res*, **34**, D552-555, 2006
15. K Kinoshita, M Kusunoki and K Miyai, Analysis of the three-dimensional structure of the "CXGXC" motif in the CMGCC and CAGYC regions of alpha- and beta-subunits of human chorionic gonadotropin: Importance of Glycine residue in the motif, *Endocrine J*, **53**, 51-58, 2006
16. 木下賢吾、タンパク質の機能部位の予測、実験医学増刊号 (バイオインフォマティクスツールの開発と生命研究への応用の最前線)、2008 (印刷中)

17. 木下賢吾、タンパク質間相互作用に関するデータベースの構築、*生体の科学*、**58**, 334-337, 2007
18. 木下賢吾、eF-siteを利用した機能部位の予測、*バイオテクノロジージャーナル*、**6**, 444-447, 2006
19. 木下賢吾、プロテインインフォマティクスによるタンパク質の機能推定、*日本結晶学会誌*、**47**, 341-347, 2005

(2) 口頭発表

①学会

国内 4 件, 海外 0 件

1. 木下賢吾、タンパク質の配列・構造・機能関連の博物学、生物物理学会、2007 年
2. 木下賢吾、表面形状と物性から見たタンパク質間相互作用部位の特徴、タンパク質科学会、2005 年
3. 木下賢吾、表面形状と物性から見たタンパク質間相互作用様式の予測法の開発、生理学研究会、2005 年
4. 木下賢吾、タンパク質立体構造からの機能推定、第 252 回 CBI 学会研究講演会、2005 年

②その他

国内 7 件, 海外 3 件

1. 木下賢吾、Ordered and disordered structure of proteins and their functions, タンパク質構造研究の未来像、東海大学、2008 年
2. K. Kinoshita, Four topics on the relation between protein structure and functions, Japan-UK symposium on Protein Interactome, Leeds University, 2008
3. 木下賢吾、タンパク質立体構造情報を利用した機能推定、第 41 回遺伝医学研究会、2007 年
4. E. Kanamori, Y. Murakami, Y. Tsuchiya, D. Standley, K. Kinoshita, H. Nakamura, Prediction of protein-protein interaction using evolutionary trace method and molecular surface complementarity, CAPRI evaluation meeting, Canada, 2007
5. 木下賢吾、タンパク質複合体の予測法の開発、分子脳科学・タンパク 3000 合同セミナー、2007 年
6. 木下賢吾、計算科学的アプローチによるタンパク質立体構造からの機能推定、第 16 回細胞電気薬理研究会、2006 年
7. K. Kinoshita, History of fragments in PDB: New folds need new fragments? GSIC international symposium, Japan, 2006
8. 木下賢吾、eF-site: electrostatic surface of functional-site, 第 3 回オントロジー研究会、2006 年
9. 木下賢吾、表面形状と物性から見たタンパク質間相互作用様式の予測法の開発、生理学研究会、2005 年
10. 木下賢吾、タンパク質立体構造からの機能推定、第 252 回 CBI 学会研究講演会、2005 年

(3)特許出願（SORST 研究の成果に関わる特許（出願人が JST 以外のものを含む））

	件数
国内出願	0
海外出願	0
計	0

(4)その他特記事項

Biotechnology Japan: 2007年3月1日ニュース「タンパク3000の成果であるたんぱく質3040種の構造情報を一般公開、類似性検索Webサービス「eF-seek」も最近公開」(<http://biotech.nikkeibp.co.jp/bionewsn/detail.jsp?id=20042386>)で研究成果が紹介された。

9. 結び

本研究では、タンパク質複合体構造の予測法の開発を行った。複合体の予測は、どのような複合体を作る可能性があるかを探索するステップと、その中から尤もらしい複合体を選ぶステップからなるが、後者の問題を解決するために、形成メカニズムが大きく異なると考えられるホモ複合体とヘテロ複合体に分けて研究を行った。ホモ複合体では十分な数の複合体を得ることができたので、統計解析により複合体形成のルールを見いだすことができた。一方、ヘテロ複合体では、構造情報の少なさから多様であるという事実とヘテロ複合体の天然構造が比較的弱い相互作用である可能性を示す事となった。これら複合体構造の解析と並行して、複合体構造の予測コンテストCAPRIへ参加し良い結果を出すこともできた。この事は、今回開発を行った複合体の予測手法が世界的に見ても高いレベルに有ることを示すと考えられる。その観点では想定していた以上に良い結果を出すことができたと考えているが、すべてのタンパク質の場合に良い結果を出せるわけではなく、良い結果を出せる場合もあるという程度であるという点ではまだまだ課題も残されている。しかしながら、多くの研究者が取り組みながら決定版となる方法が見いだせない現状に一定の成果を上げることができたと思う。

JSTにはPRESTOに引き続いてSORSTでも研究を支援していただいたことに感謝しています。JSTのグラントは科研費と比べて柔軟に利用できること、事務手続が簡素である点など非常に良いグラントだったと思います。また、PRESTOに引き続いて研究を発展させる上でSORSTの役割も重要なものがあったと考えています。その点で、最近のPRESTOでは発展としてのSORSTという道筋が無くなったというのは残念に思っています。

最後になりましたが、今回は個人型の研究ではありますが、特に後半においては自分のグループの立ち上げとも並行して、様々な共同研究者との共同研究の研究になりました。すべての共同研究者への感謝の意味も込めて、研究室のメンバーの写真を添えさせていただきます。

