3H1045

蛋白質分子活性部位の表面形状と物性に関するデータベース (eF-site/ActiveSite) の構築

○古井 淳一、木下 賢吾、中村 春木 (阪大・蛋白研)

従来、蛋白質の立体構造の比較・分類やその機能を理解するために、蛋白質の主鎖構造に主眼をおいた議論がなされてきた。しかし、結合基質は同一でありながら、基質結合部位周辺の主鎖構造が大きく異なる場合も報告されており、主鎖構造のみでは蛋白質の構造・機能相関の議論には不十分であることが示唆されている。そのため主鎖構造のみでなく、蛋白質分子表面の形状およびその表面物性も用いて蛋白質構造の比較・分類を行う事が重要であると言える。ところで、蛋白質立体構造とその機能の関連を考察するためには、その蛋白質の形状・物性と基質結合部位・活性部位の情報を結びつけて考慮する事が不可欠である。しかしながら、蛋白質の活性部位についての情報のあるデータベースはその表記が多様で、多数の蛋白質を比較検討するのには不向きである。また PDB との連携も簡便な方法では提供されておらず、蛋白質立体構造と活性部位の情報を画一的に扱うことの出来るデータベースは現状では存在していない。そこで、蛋白質分子表面を用いた比較・分類および機能との関連を研究する前段階として、蛋白質表面形状と活性部位の情報をリンクさせ、画一的に扱い表示させることの出来るデータベースの構築を行っており、<http://pi.protein.osaka-u.ac.jp/eF-site/ActiveSite> にて公開を行う予定である。

J.Furui, K.Kinoshita and H.Nakamura : eF-site/ActiveSite : Database for molecular surfaces of protein active sites

3H1115

ヒトゲノムに G タンパク質共役型受容体 (GPCR) はどれだけあるか？

○高江洲 宏智¹、美宅 成樹¹、武田 茂樹²、門脇 嗣郎²、芳賀 達也² (¹東京農工大・工、²東大医)

本研究の目的は、ヒトゲノム全体にどのくらいの GPCR があるかを評価することであるが、ヒトゲノムに関しては ORF の予測が難しいので、GPCR の予測を次の手順で行った。(1) ヒトの DNA 配列の Phase0 から Phase3 のすべてを用い、ストップコドンからさかのぼって、適当な長さ (メチオニンまでに 250~1000 残基) を持った配列を解析対象の配列とした。(2) この配列について、膜タンパク質予測システム SOSUI で解析を行い、6~8 本の膜貫通ヘリックスを持つものを GPCR の候補とした。(3) ヒトのアミノ酸配列で、データベースとして発表されている 262 個の GPCR、嗅覚受容体として報告されている 28 個を基本として、SOSUI で候補となった配列のホモロジー解析を行った。そして、既知の GPCR ないし嗅覚受容体にホモロジーのあるものと、ホモロジーのないものに分類した。この情報処理の結果、SOSUI によって候補として約 1150 個の配列が得られた。そのうち、新規の受容体で、既知の GPCR と何らかのホモロジーがあるものとして約 100 個の配列が見つかった。さらに既知の嗅覚受容体と高いホモロジーを持つ配列が見つかり、嗅覚受容体として総計 157 個が同定された。これによって、かなり確実な GPCR が約 520 個列挙された。残りの約 630 個の中には、GPCR だけではなく、他の機能のタンパク質もかなり含まれているはずであるが、配列の特徴から GPCR と見られるもの 30~80 程度が推定された。最終的に解析したヒトゲノムの中には 600 個以上の GPCR があるという結果となった。ただ解析したものは DNA 配列の 82 % を含んでいると考えられること、SOSUI の予測精度、イントロンを含んでいるものは解析されていないことなどを考慮すると、この 2 倍くらい (約 1200 個) の GPCR があると考えられる。しかし、これは 2000 個以上の GPCR がヒトゲノムには含まれているという従来の推定値よりかなり少ない。

H.Takaesu, S.Mitaku, S.Takeda, S.Kadowaki, T.Haga : How many G-Protein Coupled Receptor (GPCR) genes are contained in human genome ?

3H1100

糖結合蛋白質の基質結合部位の比較

○定浪 圭史¹、木下 賢吾^{2,3}、木寺 詔紀¹ (¹京大院理、²阪大蛋白研、³理研 GSC)

構造ゲノム科学におけるバイオインフォマティクスの課題のひとつとして、大量に生産される機能未知の立体構造から、その分子機能を推定することが挙げられる。我々はこの作業をコンピュータを用いて自動的に行うため、機能に密接に関わる構造モチーフを、PDB の立体構造の比較を通して作り上げることを考えた。具体例として生物学的に重要な機能を担う糖結合蛋白質に注目し、糖結合構造モチーフはどのようなものがあるかを調べることにした。まず、PDB から糖結合部位 194 個を選びだし、類似する原子の空間配置を検索する cliquedetection のアルゴリズムを用いて all-against-all の比較を行った。その結果、いくつかの進化的に類縁関係のない蛋白質に共通して存在する糖結合モチーフを同定することが出来た。しかし、我々が現在考えている極めて限定された部位の立体構造では、糖結合部位は、一通りの原子の空間配置では特徴づけられないことが判明した。多くの場合スーパーファミリー内では非常に強く保存されているが、スーパーファミリーを越える構造モチーフは稀にしか見付からない。そこで、基質結合部位を、原子の空間配置以外の物理化学的パラメータ (水素結合のパターン、静電ポテンシャル) で符号化し、これを機能予測に用いることにした。2 つの基質結合部位が与えられたときに、それぞれの記述子について類似度を定義し、all-against-all の比較を行ったあと、同様にクラスタ分析を行った。二つの異なる記述子でどのようにグループ分割が変化するかをみることににより、機能予測にふさわしい記述子を見つけ出すことが出来る。本発表では、この発見的な手法と、どのような性質が糖鎖結合蛋白質の基質結合部位に一般に保存されているかを紹介する。

K.Sadanami, K.Kinoshita, A.Kidera : A structural analysis of carbohydrate-binding sites in proteins

3H1130

膜貫通ヘリックス予測の精度向上を目的とした親水性膜貫通ヘリックスの解析

○林部 敏之¹、広川 貴次²、美宅 成樹¹ (¹東京農工大学・工、²変化システム)

当研究室で開発された膜タンパク質判別システム SOSUI は非常に高精度で膜内在予測をし、同時に膜貫通ヘリックスの領域予測を行う。膜タンパク質は生体内で受容体や輸送体として重要な機能を有しているものが多く、また膜貫通回数と機能との相関が示唆されている。このため膜貫通ヘリックスの領域予測に関して、より高い精度が要求されている。また膜タンパク質の機能を担っていると考えられるセグメントには膜中に存在するにも関わらず親水性が高いものがあり、予測することが困難である。

本研究では膜貫通領域予測の精度向上を目的として、膜貫通領域内におけるアミノ酸分布の解析を行った。解析に用いたデータセットは SWISSPROT rel.38 から膜貫通部位の記述のある 7 本膜貫通型タンパク質を相同性 25 % の条件で集めた。解析にあたっては膜内部のアミノ酸残基分布可視化ツールを開発した。このツールは膜貫通ヘリックスの長さを規格化し、抽象化することでターゲットとするアミノ酸を膜の厚さに対する相対位置に描画することができる。

このツールを用いて各タンパク質の膜貫通セグメントを 1 本目から 7 本目を分け、SOSUI の結果を考慮して描画・観察した。結果、平均疎水性値と膜内でのアミノ酸残基の相対分布に各貫通領域で個別の傾向があることが解った。平均疎水性値は 1.5, 6 本目のセグメントで高く、これに対して 3, 7 本目では低かった。アミノ酸分布については D, E, N, Q, R, K, H に特異的な分布傾向がみられた。さらに 7 本目の膜貫通領域では親水的なアミノ酸残基が数残基の間隔を置いてペアを組むように分布する傾向が見出された。このような残基の分布と親水的なセグメントの膜への組み込みの相関を考察し、膜貫通領域予測に反映させたいと考えている。

T.Hayashibe, H.Hirokawa and S.Mitaku : Amino Acid Sequence Analysis for High Performance Transmembrane Helix Prediction.