B741

立体構造帰属データベースシステム「GENIUS II」の開発

○矢葦 幸光^{1,2}、向井 有理¹、諏訪 牧子¹（¹産総研・生命情報研、²（株）情数研）

GENIUS IIは、多重媒介配列検索（Multiple Intermediate Sequence Search:MISS）法を用いて、構造既知のタンパク質を各生物種の持つ全ORFに帰属した立体構造帰属データベースシステムである。MISS法とは、構造既知のタンパク質を問い合わせ配列、ターゲットをnon-redundantなアミノ酸配列データベースとしてPSI-BLASTを行い、その結果集められた多重媒介配列を介することによって立体構造をORFに帰属させる手法のことを言う。具体的には、一度多重媒介配列データベースが作成できれば、各ゲノムのORFを問い合わせ配列、多重媒介配列データベースをターゲットとしたローカルアライメント検索での帰属が可能であり、毎回PSI-BLASTを行う手間が省けることからより高速な帰属を行うことができる。立体構造の帰属を行う際には、CATHの立体構造分類情報に基づいて算出した閾値が用いられ、偽陽性を1%以下に抑えた高感度な帰属を実現している。また、帰属率が立体構造データ、アミノ酸配列データの増加に強く依存していることから、立体構造の情報及びアミノ酸配列の増加によって将来的にも帰属率の上昇が期待される。現在、GENIUS IIで利用できる機能は以下の通りである。1) 全ゲノムが決定された130種の生物種に対する立体構造帰属情報の閲覧 2) 多重媒介配列データベースをターゲットとしたWebからの配列検索 3) PDBコードを入力することによる特定の立体構造が帰属したORFの検索 4) 全ORFのアノテーションを対象としたキーワード検索 GENIUS IIサーバは産業技術総合研究所 生命情報科学研究センター内に置かれ、誰でもWeb上からアクセスすることができる。URLは<http://genius.cbrc.jp>。

Y.Yabuki, Y.Mukai and M.Suwa: GENIUS II: A database system for assigning ORFs in complete genomes to three-dimensional structures

B744

膜貫通ヘリックス間3to3コンタクトペアの解析

○孫 鵬飛¹、小西 祐伸¹、西尾 卓広²、清水 俊夫¹（¹弘前大・理工・電子情報システム工、²浜松医大・物理）

複数の膜貫通ヘリックス（TMセグメント）を含む膜貫通タンパク質においては、異なるTMセグメント中のアミノ酸側鎖間のコンタクトが立体構造を決める因子の一つであると考えられる。セグメント同士のコンタクトでは、隣接した2つのヘリックス中の3残基ずつが互いに入り込んで接触しているペアが数多く見出される。我々はこのペア（3to3コンタクトペア）を、アミノ酸の位置関係によって下表のように4タイプに分類・定義し、立体構造が知られているTMタンパク質から探し出して解析した。

TMPDBに含まれるTMタンパク質の15エントリーから937組の3to3コンタクトペアを取り出した。その内約8割は、逆平行のTMセグメント間に存在するType IおよびType IIであった。また、側鎖の長さによってアミノ酸を3クラスに分類し、ヘリックス内での各クラスのアミノ酸の存在確率と、クラスの組合せとしてのコンタクトペアの出現数から3to3 contact likelihood matrixを求めた。

本発表では、このマトリックスの意味と3to3コンタクトペアの特徴について報告する。更にこのマトリックスに基づいて、構造未知のTMタンパク質中のコンタクトペアを予測する方法を議論する。

セグメント配置	タイプ	接触する残基の組合せ
逆平行	Type I	(n, m+3), (n, m+4), (m, m+3), (m, m+4)
	Type II	(n, m+4), (n+1, m+4), (m, m+4), (m+1, m+4)
平行	Type III	(n, m), (n, m+1), (m+4, m+3), (m+4, m+4)
	Type IV	(n, m), (n+1, m), (m+3, m+4), (m+4, m+4)

n, m は各TMセグメント中の残基番号である。タイプごとに、各残基ペアの組合せにおいて、側鎖間最近距離（水素原子を除く）が4 Å以下のものを3to3コンタクトペアとして定義した。

P-F.Sun, S.Konishi, T.Nishio and T.Shimizu: Analysis of 3to3 contact pairs between transmembrane helices

B742

高信頼性膜貫通トポロジー予測データを得るための一つのアプローチ

○夏 俊雄¹、池田 修己^{1,2}、清水 俊夫¹（¹弘前大・理工・電子情報システム工、²岩手大院・連合農・生物資源）

膜貫通（TM）タンパク質の機能とTMトポロジーとの間には密接な対応関係があり、正確なTMトポロジー情報があれば、TMタンパク質配列の機能分類・同定が可能である。プロテオーム解析が進む中、TMタンパク質の機能分類・同定のための信頼性の高い予測TMトポロジーデータが切実に求められている。

本発表では、信頼性の高い予測TMトポロジーデータを獲得するための、新たなコンセンサス予測法（ConPred.elite）を提案する。このアプローチでは、次の二つの場合にだけトポロジー予測を成立させる：5つの予測法（TMHMM 2.0、MEMSAT 1.8、HMMTOP 2.0、TMPred、TMAP）によるトポロジーモデルが、(i) 全て一致している、(ii) セグメント本数・位置は全て一致しているが、貫通方向が2つに割れる（4対1あるいは3対2で）。予測配列数を犠牲にする代わりに、予測精度を大幅に上げようという戦略である。この方法によって原核生物由来配列の30%、真核配列では22%を予測することができ、正しい予測の割合はそれぞれ98%、95%であった。

このConPred.eliteを原核ゲノム29種、真核ゲノム10種から予測されたTMタンパク質（それぞれ19,513、54,572）配列に対して適用し、それぞれ3,871配列（19.8%）、6,980配列（13.3%）のTMトポロジーモデルを得ることができた。ConPred.eliteの具体的な方法、得られたTMトポロジーデータの内容などの詳細については当日報告する。

J-X. Xia, M. Ikeda and T. Shimizu: A prediction approach to obtaining highly reliable transmembrane topology models

B745

膜貫通ヘリックスバンドル内側のアミノ酸残基は保存される傾向がある

○小西 祐伸、清水 俊夫（弘前大・理工・電子情報システム工）

膜貫通タンパク質は実験による立体構造の決定が困難な上、立体構造予測もまだまだ精度は不十分である。そこで、膜貫通タンパク質の立体構造と関係のある傾向を見つけ立体構造予測につなげたい。水溶性タンパク質の場合、内側のアミノ酸残基が良く保存されている事が知られている。これをもとに今回は膜貫通タンパク質の膜貫通ヘリックス部位において、膜貫通ヘリックスバンドルの内側又は外側で保存されている残基の割合について研究した。

本研究では膜貫通ヘリックスバンドルの内側、外側を分かりやすくとらえる為C α （膜貫通ヘリックス部位のみ）の位置を平面に投影した。この時、膜貫通ヘリックスバンドルを的確に投影する為、各膜貫通ヘリックスの傾きから平均の傾きを計算し、そのベクトルに沿って投影する方法をとった。さらに各膜貫通ヘリックスの平面上における平均C α 位置を計算し、これを各膜貫通ヘリックス中心とした。これらの中心からなる多角形内を膜貫通ヘリックスバンドル内側、多角形外を膜貫通ヘリックスバンドル外側とした。

残基の保存度を、CLUSTALWを用いMultiple Alignmentを行い、その結果から完全に保存されている、強く保存されている、弱く保存されている、全く保存されていない、の4種類に分けた。

Bacteriorhodopsin (BACR.HALN1)との配列類似度が30%以上である配列をSwiss-Prot Release 41から11配列を選択した場合、保存性と位置関係を調べた結果、完全に保存されている残基は67.7%が内側に、全く保存されていない残基は19.6%が内側に存在した。

S.Konishi and T.Shimizu: Inner positions in the transmembrane-helix-bundle are more conservative