1J0945

残基間平均距離統計に基づくコンタクトマップによる
ドメインの位置の予測— α/β タンパクへの応用—
○菊地 武司 (倉敷芸科大・産業科技・生命化学)

これまで残基間平均距離統計に基づくコンタクトマップ (Average Distance Map, ADM) を用いてドメインの位置の予測を試みてきた。特に最近のゲノム解析の著しい進歩を考慮し、その次の重要なステップはゲノム解析によって得られた ORF の立体構造単位、あるいは機能単位の決定である。本方法を予備的にいくつかのゲノム ORF 配列に応用し、本方法が全般的にタンパク立体構造単位あるいは構造のコアの位置を予測していることがわかっていく。しかしながら、特に α/β タンパクやバレルタンパクのドメインの位置の予測が困難であることも明らかになった。本研究においては、Lang ら (D. Lang et al., Science 289, 1546 (2000)) によって構造の特徴が研究されている His 合成経路の酵素 HisA と HisF をとりあげ、ADM によるドメインの位置の予測と、その構造特性とを比較し、ADM 法によるドメインの位置予測法の改善の方向を探ることを目的とする。Lang らによれば、この 2 つのタンパクは立体構造的によく似ているが、一つのタンパクにおいても非常に似た 2 回繰り返し構造から成っている。このため、この 2 つのタンパクは一つの共通の遺伝子の複製により進化したものと考えられている。このタンパクの ADM を用いて、ドメインの位置を予測すると 4 つの部分の立体構造単位と予測される。それぞれを詳細に Lang らの結果と比較すると、基質結合部位と HisF におけるリン酸結合部位に対応することがわかった。これは、この場合には ADM が機能単位に対応する部分を予測していることを示しているが、周期的な構造単位の位置を拾いあげているようにも見える。現在、この結果に基づき α/β タンパクのドメイン予測可能性について検討中である。

T.Kikuchi : Prediction of location of domains with contact maps based on the interresidue average distance statistics for alpha/beta proteins

1J1015

PC クラスタを用いた高速溶媒和エネルギー計算 Web
アプリケーション

○山下 雄介¹、升屋 正人² (¹ 鹿児島大院・理工学・情報工学、² 鹿児島大・総合情報処理センター)

アミノ酸配列の情報からモンテカルロ法により分子の立体構造を予測する試みや、分子動力学法による立体構造予測シミュレーションなどが行われている。これらにおいては、構造エネルギーに加えて溶媒によるエネルギー及びそのデリバティブの計算が必要となる。ところが、溶媒によるエネルギーとそのデリバティブの計算には多くの時間がかかるため大規模なシミュレーションはこれまで実現できていない。低コストで大規模なシミュレーションを実現するため、PC クラスタによる並列計算システムの研究が進みつつあるが、溶媒によるエネルギーとそのデリバティブの計算に時間を要することは解決されていない。高速に溶媒和自由エネルギーを計算する方法としては、表面積に比例するパラメータを用いた溶媒和自由エネルギーの近似計算法と、分子表面積の数値解法を組み合わせた方法がある。しかし、この方法は並列計算システムに最適化されておらず、十分な速度を達成できていない。そこで本研究では高速溶媒和自由エネルギー計算アルゴリズムを PC クラスタによる並列計算システム上で最適化するとともに、タンパク質分子の溶媒和自由エネルギーの算出をサービスとして提供することのできる Web アプリケーションを開発した。

Y.Yamashita and M.Masuya : Web application for fast solvation free energy calculation using PC cluster

1J1000

Real value prediction of accessible surface area
in proteins

○Ahmad Shandar¹、Gromiha Michael²、皿井 明倫¹ (¹ 理研・筑波研究所、² 産総研・生命情報科学研究センター)

Prediction of solvent accessibility or accessible surface area (ASA) has been a widely studied topic in computational molecular biology, due to its significance in overall structure prediction as well as functional properties such as binding, mutations and stability. However, all the efforts for its prediction have been limited to first classifying residues into buried and exposed states and then making a binary or tertiary prediction. Lot of information about the actual accessible area is therefore lost for building the model as well as making predictions thereafter. We have developed the first algorithm, by which real value of accessible surface area of protein residues can be predicted. This method employs a neural network with floating point outputs, which maps a binary space of sequence information onto a finite real space of ASA. It has been shown that such a prediction carries more information than even a 100% binary prediction of ASA states. Our model could predict real value ASA of protein residues within 18-19% mean absolute error, defined as per residue absolute difference between the experimental and predicted values of relative ASA. Cross-validated data sets of more than 80,000 residues were used for this study. As expected, prediction error is found to strongly correlate with variability of ASA for different residue types.

S.Ahmad, M.M.Gromiha, A.Sarai : Real value prediction of accessible surface area in proteins

1J1030

PC クラスタを用いたホモロジー検索 Web アプリ
ケーション

○上國料 享¹、升屋 正人² (¹ 鹿児島大院・理工学・情報工学、² 鹿児島大・総合情報処理センター)

蛋白質をそのアミノ酸配列の類似度、すなわちホモロジーにより分類する試みがこれまで行われてきた。分類された各グループはファミリーと呼ばれ、同じファミリーに属する蛋白質は類似の構造と機能を持つ場合が多い。このため、配列の類似度の比較が構造・機能が未知の蛋白質の構造予測・機能予測に用いられてきた。厳密な立体構造を決定することができなくても、どのファミリーに属する蛋白質であるかの知見を得ることが、構造と機能が未知の蛋白質を扱う創薬の分野などでは有用である。そこで本研究では、実験室で決定したアミノ酸配列を Web ブラウザの画面に入力するだけで、どのファミリーであるのかを表示するシステムを開発した。本システムはインターネットを経由して全世界からアクセスできるため、いつでもどこでもファミリーの検索が可能である。配列同源性検索アルゴリズムとしては FASTA を用い、PC クラスタによる並列計算システム上で計算を行うことで、高速化を図った。また、Protein Data Bank に登録されている蛋白質のアミノ酸配列データを本システムの入力として計算を実行し、蛋白質ファミリーの数がおおよそ 1000 であるという Chothia の報告の検証を行った。

T.Kamikokuryo and M.Masuya : Web application for homology search of proteins using PC cluster