3P005

タンパク質の48残基セグメントの native contact パターンの解析

○伊東 純一¹、三友 大輔¹、池田 和由^{1,2}、肥後 順一¹ (¹ 東薬大院・生命科学科・生物情報、² 生命情報科学研究センター (CBRC))

50 残基程度のセグメントであれば、独自に folding することが可能だと言われている。また、タンパク質の native-contact はその folding のプロセスと深い相関があると言われている。そこで我々は SCOP から all alpha, all beta, alpha/beta, alpha+beta の4つの class の中にある各 fold から代表的なタンパク質を一つずつ抜き出し、計 510 本のタンパク質を用意した。そして、個々のタンパク質を48 残基セグメント単位で1 残基スライドしながら分割し、計 54699 本のセグメントをサンプルとした。次に、N 末端から数えて連続する3 残基を1unit とし、各セグメントを16 unit として新たに定義し、各 unit 間の最短距離(2 つの unit 間の全原子間の最短距離)を計算した。セグメントの各 unit 間の最短距離が 6.5 Å より長い場合は0、短い場合は1 とし、この二値データをそのセグメントの持つ unit 間の native-contact(U-native-contact) パターンとした。次に、各セグメントの U-native-contact パターンをその類似度によりクラスタリングし、およそ 1000 個のクラスターを得た。各クラスターからそのクラスターを代表するパターンを一つずつ取り出した。このおよそ 1000 種類の U-native-contact パターンと、それを持つタンパク質の立体構造や機能との相関関係を調べることににより、タンパク質 folding の新たな特徴を発見することを目的としている。

J.Ito, D.Mitomo, K.Ikeda, and J.Higo : Analysis of native contact pattern of 48-residue segments taken from proteins.

3P007

タンパク質立体構造予測のためのプロファイルベースの構造ポテンシャルの構築

○白倉 弘太、石田 貴士、中村 周吾、清水 謙多郎 (東大・院農・応生工)

既知構造をテンプレートとして用いない蛋白質の立体構造予測は Anfinsen の仮説に基づき、自由エネルギーが最小となるような構造を探索することで立体構造を予測するが、多くの手法では計算量の問題から疎視化されたモデルを利用しており、それに合わせて粗い解像度でも蛋白質の天然構造を他の構造と区別可能なポテンシャルが利用される。我々はこのようなポテンシャルとして、疎水性相互作用を表す埋もれポテンシャルとアミノ酸2体間の関係を表すペアポテンシャルを用いている。従来は、各アミノ酸種の存在確率から統計ポテンシャルとして埋もれポテンシャルやペアポテンシャルが計算されてきた。しかし、近年では配列上連続する残基のアミノ酸種情報や、アミノ酸配列プロファイル情報を利用することで高い精度でポテンシャルを予想することが可能になっている。そこで本研究では埋もれポテンシャル、ペアポテンシャル両方に配列上連続する残基のアミノ酸種情報と、アミノ酸配列プロファイル情報を用いることで精度を向上させた統計ポテンシャルを開発した。本発表ではこの改良されたポテンシャルの詳細と共に、デコイ構造から天然構造を区別する能力と、このポテンシャルを利用した際の予測構造の精度の向上について議論する。

K.Shirakura, T.Ishida, S.Nakamura, K.Shimizu : Development of profile-based potential for protein structure prediction

3P006

混合正規分布モデルを用いた低解像度の蛋白質立体構造の高速重ね合わせ計算

○川端 猛^{1,2}、由良 敬^{2,3}、郷 信広^{1,3} (¹ 奈良先端大・情報、² JST・CREST、³ 原研・計算科学セ・量子生命)

近年、電子顕微鏡による巨大複合体の低解像度の3次元画像に、X線結晶解析の高解像度の部分構造を当てはめて重ねあわせることで、擬似高解像度モデルを得る試みが盛んである。これには巨大分子を並進・回転させ最適な位置を探索する計算が必要であり、その計算コストは小さくはない。本研究では、精度を失わず高速な計算を行うため、混合正規分布モデルを用いた蛋白質立体構造の簡易表現法を提案する。混合正規分布モデルとは、複数の正規分布の線形和で確率分布を近似表現する方法であり、観測点群を出力する尤度の最大化により必要なパラメータを推定する。本研究では、まず、原子モデルと3次元画像のそれぞれを混合正規分布モデルで表現する。原子モデルにおいては、原子の位置を観察点だとみなし、電子顕微鏡による3次元画像では、各画素の位置に濃度に比例した観測数があるとみなすことで、EM アルゴリズムでパラメータを最尤推定を行う。これにより数千原子あるいは、数百万画素が数個から数十個の正規分布で簡潔に表現され、3次元画像に原子モデルを重ねる計算が、複数の正規分布の重なりを最大化する計算に単純化される。混合正規分布モデルの利点の一つは、目的関数となる正規分布どうしの重なり空間積分を解析的に得ることができ、さらに、並進・回転による目的関数の導関数も同様に解析解が得られるため、微分を用いた最適化法が容易に適用できることである。原子モデルとそれをぼかした3次元画像の人工的なテストデータを用意し、それらを混合正規分布モデルに変換した後、重なりを最大化するような並進・回転を最急降下法で計算したところ、良好な結果が高速に得られた。今後はより現実的なデータに対して適用し、本手法の有効性を検証していきたい。

T.Kawabata, K.Yura and N.Go : Fast Fitting Calculation of Low Resolution Protein Structures using Gaussian Mixture Model

3P008

残基間平均距離統計ポテンシャルによる残基対相関関数の特性の解析

○菊地 武司 (立命館大・情報理工・生命情報)

我々はこれまで、 β バレルタンパク、グロビンタンパク、リゾチームなどの残基間平均距離統計に基づく予測コンタクトマップ (ADM) の形状とフォールディング過程との比較を行ってきた。その結果、ADM の局所的なパターンはタンパクのフォールディング過程を反映していることがわかった。特に、ADM により予測されるサブドメイン領域は高い ϕ 値を示す残基を含んでいることがいくつかのタンパクにおいて示された。このことは、残基間の平均距離統計がフォールディングの構造形成に関する情報を有効に含んでいることを示唆している。本研究では、ADM により得られた結果を、平均距離統計を残基間ポテンシャルに変換したものをを用いてシミュレーションを行うことにより再検討することを目的とする。ここではランダム構造を初期構造として構造サンプリングを行い、その過程における残基対相関関数を計算した。残基対相関関数 $g(i,j)$ は残基 C α 原子間距離 R_{ij} が $R_{ij} < R_{cut}$ となる確率として定義される。本研究では R_{cut} を 10 Å とした。まず、 $g(i,j)$ が高い値を示す残基対に対し、それをマップ上にプロットすると天然コンタクト部分に高い $g(i,j)$ を持つ残基対が出現する傾向が見られた。これらは ADM による結果と一致する。さらに、 $g(m) = \sum |i-j|=m g(i,j) / \sum |i-j|=m$ として、 $Q(\mu, \nu) = (g(\mu, \nu) - (\sum |i-j|=m g(i,j) / \sum |i-j|=m)) / D(m)$ という量を計算した。ここで、 $D(m) = (\sum |i-j|=m (g(m) - g(i,j) |i-j|=m) / \sum |i-j|=m) / 2$ である。 $Q(\mu, \nu)$ は特に相関の高い残基対を表す。 $p(\mu) = \sum \nu Q(\mu, \nu)$ として、いくつかのタンパクについて $p(\mu)$ と ϕ 値データと比較すると、それぞれのプロファイルの傾向が一致することが観察された。詳細については現在検討中であり、発表時に報告する予定である。

T.Kikuchi : Properties of Residue Pair Correlation Functions Based on the Potential Derived from the Interresidue Average Distance Statistics