

3P106

スパースモデルによる水溶液構造の解析

水上 卓¹, 杉山 歩², 川崎 隆史², 高木 啓行², ホー・ツー・バオ², ダム・ヒョウ・チ²

(北陸先端大院・マテリアル¹, 北陸先端大院・知識科学²)

Structural analysis of protein hydration water by means of sparse model

Taku Mizukami¹, Ayumu Sugiyama², Kawasaki Takafumi²,

Takagi Hiroyuki², Ho Tu Bao², Dam Hieu Chi²

(JAIST Materials Sci. ¹, JAIST Knowledge Sci. ²)

【序】

我々はデータマイニング法を用いた物理モデル設計支援システムの構築を目指している。データから知識を取り出すには、物理的モデルに基づいた解釈をおこなう必要があり、それは従来人間の直観にもとづいて設計されてきた。一方、近年の計測・計算・通信技術の発展により、多次元ビッグデータを取り扱う必要性が増している。しかし、大量のデータを対象にして実際に直観を働かせるにいたるまでには様々な手続きや処理が必要であり、人間の直観のみによるモデル設計には大きな困難がともなう。

我々は分子動力学計算からのトラジェクトリデータを混合分布モデルによって特徴空間に変換し、データマイニング的手法によって、タンパク質に水和する水分子の振る舞いをクラス分けしその特徴を同定してきた。今回、水分子のダイナミクスにおける隠れた属性間の関係を明かにすることを目的に、水和・バルク 2 種類の水分子の振る舞いを定量化し、それをさらにグラフモデルによって解析した。その結果水和水・バルク水の属性の関係性を示す特徴的なグラフ表現を得たので報告する。

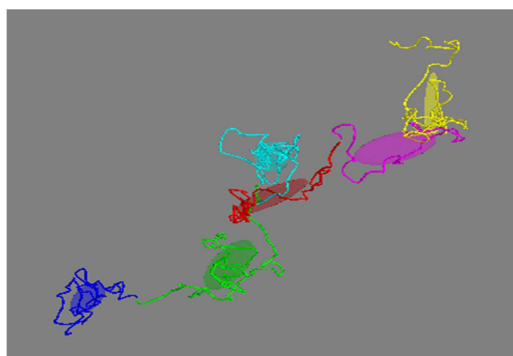


図 1. 混合分布モデルによる水（酸素原子）トラジェクトリのクラスタ化

【方法】

分子動力学シミュレーションにおいては、力場 Amber03、TIP3P モデル水分子を約 3.3×10^4 個、周期的境界条件、温度 300K、NVT アンサンブルの条件で実行した。トータルのランタイムは約 10 ns. また蛋白質(PDBID:1PSV)周辺に TIP3P 水分子を約 6×10^3 個配置し同条件で約 15 ns のトラジェクトリデータを作成した。

混合分布モデルは関数としてガウス分布を用い、EM アルゴリズム (Expectation-maximization algorithm) を応用した Gaussian Mixture Model[1]を採用した。これを水分子の酸素原子トラジェクトリに適用し、実空間上の軌跡をさまざまな大きさ・形状の回転楕円体に対して近似する複数のクラスタに分解した (図 1)。これらのクラスタの形状を示すパラ

メータ、およびクラスに含まれる軌跡点の個数等から特徴空間を構築し、その空間上で主成分解析(PCA)および特徴空間上でのさらなる混合分布プロセスを走らせることによって、水分子のダイナミクスをクラス分けした。バルク水の系と蛋白質+水の系を比較し、水和水のクラスを同定した。

次に Gaussian Mixture Model によってクラス分けされた水分子の振る舞い属性を、複数の時間スケールにおいて再計算し、階層的時間における特徴空間を構築した。空間上において線型モデルによる regression およびスパース性を考慮したグラフマイニング[1]を行い、属性間の関係性を表現するグラフを得た。

【結果と考察】

バルク水の系および水+蛋白質の系をそれぞれデータマイニングすることによって特徴空間上の点の集合が得られる。各要素は水分子のある区間の運動の軌跡を特徴の座標軸上にマッピングされた点である。今回、蛋白質をバルク水に“投入”したことにより、大きくポピュレーションが増加したクラスを蛋白質表面に水和水分子であると予想し、それに含ま

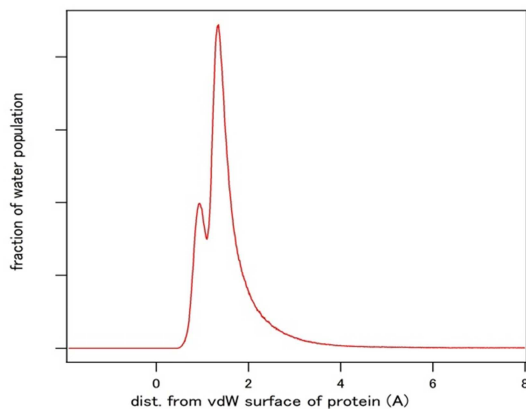


図 2. 蛋白質 vdW 表面からの酸素原子の空間分布

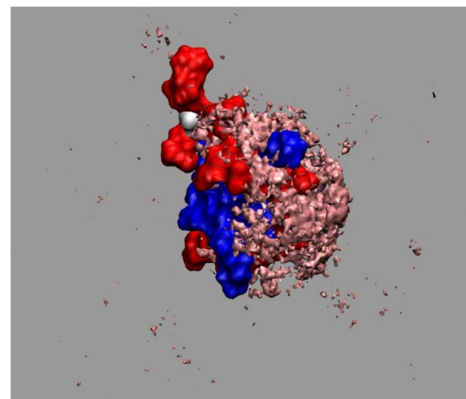


図 3. 第一水和水に対応すると示唆される水分子クラスの空間分布。

れる水分子を可視化した。

蛋白質のファンデルワールス(vdW)表面と水の酸素原子との間の距離分布を図 2 に示す。2 個のピークは平均すると約 1.4Å の距離にあった。このときの水分子の空間分布を示す(図 3)。蛋白質の疎水性、親水性のアミノ酸残基をそれぞれ赤色、青色で示し、水分子の分布はピンク色で示す。親水基の近傍に偏って水分子が分布しているのが認められ、このことから親水基に強く水素結合している第一水和水に相当するクラスであること判断される。

次に水和水、およびバルク水における属性間の関係性に関してグラフマイニングを行い、それぞれ特徴のあるグラフを得ることに成功した。これは、2 つの状態にある水分子の振る舞いの時空構造が異なることを示唆している。詳しい考察は講演でおこなう予定である。

今回おこなったグラフマイニングによる属性間の関係性の描出は、水分子のクラス分けの次のステップに位置づけられる。この方法によって、データマイニングによってモデルの構築を逐次的・再帰的に行い、新たな知識を獲得していくことを目指す。

【参考文献】

[1] Data Mining 2nd ed. I.H. Witten E. Frank, Elsevier (2005)