

1P-089

混合分布モデルをもちいた水分子のマイニング-ダイナミクス
水上 卓¹, 杉山 歩², 山口 和宏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)

Mining dynamics of water behavior using a mixture model

Taku Mizukami¹, Ayumu Sugiyama², Kazuhiro Yamaguchi², Ho Tu Bao², Dam Hieu Chi²
(JAIST Materials Sci.¹, JAIST Knowledge Sci.²)

【序】

データから意味のある情報を取り出すには、物理的モデルに基づく必要がある。従来それらは人間の直観にもとづいて設計されてきた。一方、近年の計算・通信技術の発展により、より大量のデータが得られ、それを取り扱う必要性が増している。しかしながら多次元のビッグデータに対しては、人間の直観だけをよりどころにモデルを設計するには大きな困難がともない、実際には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る。

我々はデータマイニング法を用い物理モデルの設計支援システムの構築を目指してきた。今回、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まいをクラス分けすることに成功したので報告する。

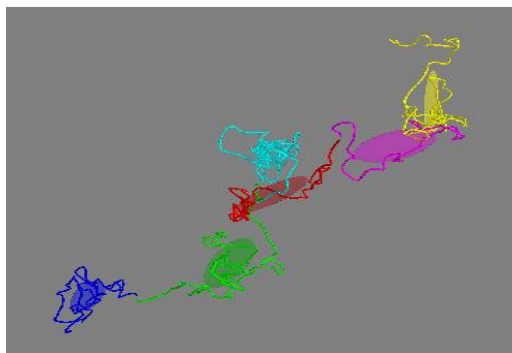


図 1. 混合分布モデルによる水（酸素原子）の軌跡のクラスタ化

【方法】

分子動力学シミュレーションに用いた力場は Amber03、TIP3P 水分子を約 3.3×10^4 個配置し、周期境界条件にて 300K にて NVT アンサンブルで分子動力学シミュレーションを行った。トータルのランタイムは約 10 ns である。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発生させた。

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すると、酸素原子の実空間上の軌跡が、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される (図 1)。これらのクラスタのもつパラメータ、およびクラスタ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 Dirichlet 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。

【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を行った点である。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した水分子であると予想し、そのクラ

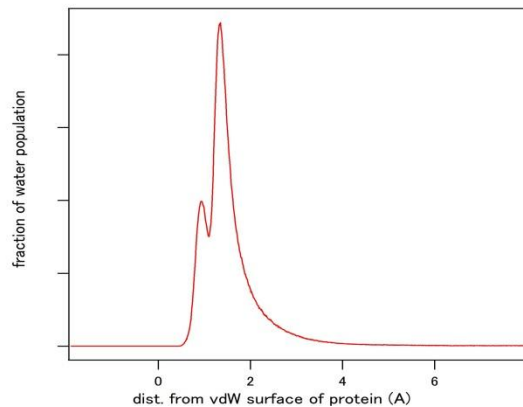


図 2. 蛋白質 vdW 表面からの酸素原子の空間分布

スに含まれる水分子を可視化した。

まず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 個のピークが認められるが平均すると約 1.4Å の距離にあった。次にこのときの水分子の空間分布を示す(図 3)。赤色、青色で示したのは蛋白質のそれぞれ疎水性、親水性のアミノ酸残基であり、水分子の分布はピンク色で示した。全体的には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、これらの点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が推察される。

以上データマイニングにより蛋白質との相対距離の情報を使うことなく、水分子の動的な振る舞いという情報だけを用いて水和水の分類を行うことができた。解析結果にはより微細

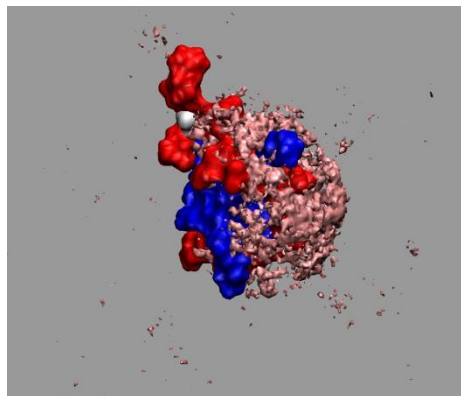


図 3. 第一水和水に相当すると示唆される水分子クラスの空間分布。

な構造が観察されており、特徴空間設計の多様性と併せて考えると、この手法によってさらに高感度/詳細な解析が可能になることが期待できる。

【参考文献】

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