

データベース解析に基づくタンパク質の水和構造予測

(1. 慶應大院・理工、 2. 理研・播磨) ○松岳大輔¹、中迫雅由^{1,2}

【序論】

水は生命にとって欠かせない物質であり、タンパク質がその機能を発揮するためには一水和層程度の水分子が必要であるという報告がこれまで数多くなされてきた。水分子はタンパク質分子と主に水素結合によって相互作用すること、水素結合が方向性を持った相互作用であることから、第1層の水分子はタンパク質表面の極性原子周辺では、相互作用に起因した特徴的分布を有することが期待される。1988年に Thanki らは Protein Data Bank (PDB) に登録されている16個のタンパク質の常温下高分解能構造モデルについて、タンパク質の極性原子周辺の水分子空間分布の傾向を明らかにした[1]。現在では、常温下よりも多くの水分子の位置が観測可能な低温 X 線結晶解析で得られた 50,000 を超えるタンパク質構造モデルが PDB に登録されており、そこに含まれる膨大な水分子の座標データを用いて統計学的に信頼できる水分子の空間分布情報を得ることが可能となっている。我々はまずデータベース解析を行って、極性原子周辺の水分子の詳細な空間分布を得た[2]。さらに得られた分布を用いてタンパク質の極性原子周辺の水和構造予測を開始したところである。

【方法と結果】

まず、PDB より温度 150 K 以下、かつ分解能 2.2 Å 以上で回折強度データ収集がなされ、その結晶学的 R 因子が 0.23 以下のタンパク質立体構造モデル 17,894 個を選択した。以前開発したプログラム "FESTKOP" [3,4] を基にして水和構造解析用プログラム "PD_hydra" [2] を作成し、主鎖および 11 種類の親水性アミノ酸側鎖の極性原子周辺の水分子を得た。解析に使用した水分子の *B*-factor の上限は 45 Å² とした。得られた分布は N-H、O-H 結合の方向ならびに C=O 結合の孤立電子対方向に分布極大を有していた。水素供与基である N-H 基周辺では水分子は N-H 結合方向に局在したのに対し、水素受容基である C=O 基の周辺では、水分子は C=O 結合の周りに幅広く分布した (Fig.1)。

データベース解析で得た統計学的に信頼できる水分子の分布を用い、タンパク質の溶媒露出表面を形成する極性原子周辺の水和構造を確率分布として予測するプログラム "DM_hydra" を開発している。その方法は、得られた分布をタンパク質表面において重ね合わせるだけという簡単なものである。非常に単純なアルゴリズムであるので計算時間も速く、分子量が約 40 万のタンパク質でも計算時間は 50 分程度である。まず、低温 X 線結晶解析で得た分解能 0.95 Å のヒトリゾチーム構造モデルに適用し、タンパク質の親水性表面の水和構造の予測が可能であることが確認した。さらに、タンパク質内部の空隙へと通じるチャンネル

内の水和構造が予測可能であった。

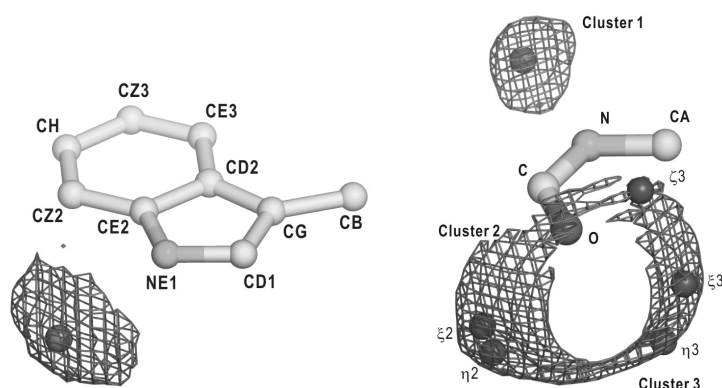


Fig.1.

データベース解析により得られた Trp 側鎖周辺(左)ならびに主鎖周辺(右)の水分子の分布

この結果を受けて、我々は膜タンパク質であるバクテリオロドプシンにこのプログラムを適用し、チャンネル内の水和構造を予測することができることを確認した。バクテリオロドプシンはいくつかの中間状態の結晶構造が解かれているが、その構造変化に伴う水和構造の変化も予測することができた(Fig.2)。

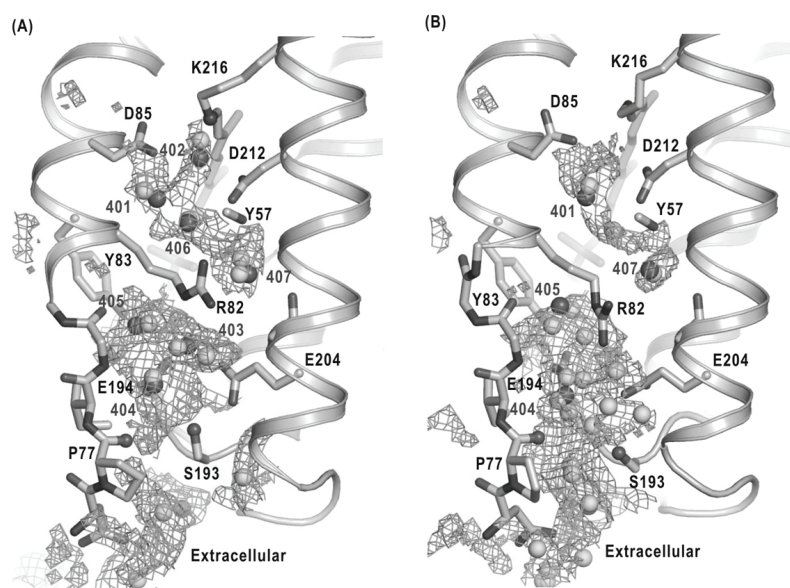


Fig.2.

バクテリオロドプシンに対する予測結果。

(A) trans BR 状態

(B) late M 中間状態。

メッシュはデータベース解析で得られた分布から計算された水分子の存在確率分布を示す。濃い色の球は X 線結晶構造解析で見つかった水分子位置、薄い色の球は確率密度内の局所極大から予測された水和水結合位置を示す。

【今後の課題】

今後、本プログラムを疎水性表面にまで拡大して利用するために、水素結合ネットワークを利用して疎水性表面の水和部位を予測するアルゴリズムを考える必要がある。

[1] N. Thanki, J. M. Thornton & J. M. Goodfellow (1988) *J. Mol. Biol.*, **202**, 637-657

[2] D. Matsuoka & M. Nakasako (2009) *J. Phys. Chem. B* (in press)

[3] M. Nakasako (1999) *J. Mol. Biol.*, **289**, 547-564

[4] M. Nakasako (2001) *Cell. Mol. Biol.*, **47**, 767-790