

分類	量子化学計算
キーワード	Protein Data Bank、水素付加、構造評価
開発者	牟田 元、吉廣 保、佐藤文俊
作成年月日	2004年 3月
コード名	ProteinDF-PDBDataTool
使用言語	C++

◇ Protein Data Bank データツール

本研究グループはこれまでに、タンパク質をターゲットとした量子化学計算ソフトウェア ProteinDF [1,2]を開発してきた。ProteinDF によるタンパク質の分子軌道計算には、主に Protein Data Bank (PDB) [3] で配布されている三次元構造データを用いる。しかし、PDB データは約85%が水素原子を持たない X 線構造解析によるデータであり、さらに構造ひずみや原子間衝突などの構造エラーを抱えていることも多く、そのまま分子軌道計算に用いることは不可能である。従って、タンパク質の分子軌道計算を実行するためには、PDB データに水素付加などの修正処理を行い、さらに分子構造に問題が無いか、その妥当性をチェックする必要がある。タンパク質は巨大で複雑な構造を持つため、このような機能はコンピュータによる支援および自動化が必須である。本研究では、タンパク質の量子化学計算を支援するための PDB データ修正・編集を行う分子モデリングプログラムを開発した。本プログラムは PDB データへの水素付加および構造妥当性評価機能を搭載し、タンパク質の構造補正処理を支援する。

◇ プログラムの概要

水素付加：PDB データには、標準のアミノ酸や核酸以外に様々な種類のヘテログループが存在する。ヘテログループはアミノ酸や核酸に比べてはるかにバラエティに富んだ分子構造を有する。従って、XYZ 座標データと簡単な結合情報のみから成る PDB データを用いてヘテログループに確実に水素付加を行うことは困難な場合がある。そこで本研究では、PDB で配布されている PDB Hetero Group Dictionary ファイルを用い、この困難を克服することを試みた。PDB Hetero Group Dictionary には、PDB に登録されているデータに含まれる全てのアミノ酸・核酸・ヘテログループの、水素原子も含めた原子および結合情報（結合原子のペア、および結合の種類）が含まれており、PDB ファイルと原子名を照合することにより結合情報を容易に補完することが可能である。さらに、この方法では比較のみで済むためアルゴリズムも非常に単純となる利点がある。以上の照合結果を元に、各原子について結合の種類（単結合、二重結合、三重結合、共鳴（1.5重）結合など）がどれだけあるかから、その原子の混成（立体構造）お

よび不足原子価を判定することにより、各原子上に水素原子をどのように付加するかが決定される。例えば、単結合を2つ持つ炭素原子は sp^3 混成 (四面体構造) で原子価があと2つ不足しているの、空いている四面体の方向に2つ水素原子を付加すれば良いことが分かる。なお、水素原子の位置は、原子種や混成などの情報からそれぞれ典型的な結合長、結合角、二面角を指定し、決定する。

構造妥当性評価：本研究では、タンパク質の量子化学計算のための構造妥当性評価を行うことを目的とするため、評価対象は水素原子や溶媒原子、ヘテログループを含むすべての原子に対して行えるようになっている。タンパク質部分については、Engh と Huber の提案した構造パラメータ [4] に加えて、水素原子も含んだ構造のリストを作成することを可能とした。また、 ϕ 、 ψ 、 ω 、 χ 、ジスルフィド結合、水素結合、原子間衝突、原子価、その他種々のタイプの結合長、結合角、二面角のリストを得ることも可能である。これらの情報を整理することで、異常な構造を素早く見つけることが可能なインターフェースも提供する。

◇計算例

ヘテログループを含むタンパク質への水素付加例を Fig. 1 に示した。さらに、100種類の PDB データに対する実行テストの結果、本プログラムは他の同等機能を有するソフトウェアに比べてほぼ確実に水素付加処理を行えることを確認した。これは、PDB Hetero Group Dictionary という結合情報を用いて水素付加を行うことにより、バラエティに富んだ分子構造を有するヘテログループへの水素付加がより確実になったためと考えられる。

構造妥当性評価の例として、Ramachandran プロット、および sp^3 - sp^3 周りの二面角の分布図を作成したものを Fig. 2 に示した。 sp^3 - sp^3 周りの二面角は $\pm 60^\circ$ および $\pm 180^\circ$ 付近にほぼ局在化するが、一部これらから大きく外れる箇所がある。この部分は重なり型のコンホメーションを採っているためエネルギー的に不利であり、計算結果に少なからぬ影響を与えると考えられる。

本ソフトウェアは文科省 IT プログラム「戦略的基盤ソフトウェアの開発」プロジェクト (<http://www.fsis.iis.u-tokyo.ac.jp/>) より2004年6月に公開を予定している。

◇参考文献

- (1) F. Sato, Y. Shigemitsu, I. Okazaki, S. Yahiro, M. Fukue, S. Kozuru, H. Kashiwagi, *Int. J. Quant. Chem.*, **63** (1997) 245. (2) F. Sato, T. Yoshihiro, M. Era, H. Kashiwagi, *Chem. Phys. Lett.*, **341** (2001) 645.
- (3) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *The Protein Data Bank, Nucleic Acids Research*, **28**, pp. 235-242 (2000) <http://www.rcsb.org/pdb/> (4) R. A. Engh, R. Huber, *Acta. Cryst.*, **A47** (1991) 392-400.

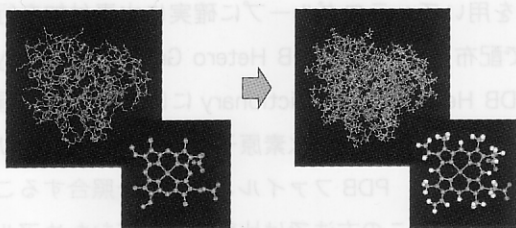


Fig. 1 水素付加の例 (myoglobin, 1A6N)

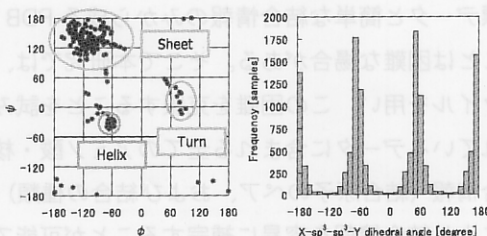


Fig. 2 Ramachandran プロット (左、Hiv-1 protease, 1A30) および sp^3 - sp^3 周りの二面角分析 (右、Enterotoxin, 1TII)