タンパク質構造の修正処理；水素原子付加の問題

牟田　元, 吉廣　保, 佐藤　文俊
東京大学生産技術研究所　計算科学技術連携研究センター

1. 序論

タンパク質は、生命の構成要素の一つであり、生体内で生命維持のための種々の機能をつかさどる分子である。近年の分析技術の進歩により、タンパク質の巨大で複雑な 3 次元構造が次々と明らかになってきており、Protein Data Bank (PDB)[1]には現時点で 2 万件近い構造データが登録されている。これらの 3 次元構造データは、タンパク質が織り成す複雑な生命現象を解明するために重要である。

我々はこれまでに、密度汎関数理論を用いた量子化学計算により、タンパク質の全電子状態を理論的に求めることが可能な ProteinDF プログラム[2]を開発してきた。従来、このような分子の全電子状態は小さな分子については広範に行われてきたが、タンパク質のような巨大分子には計算時間の爆発的な増大により適用が困難であった。ProteinDF はそのような計算時間の増大を抑えるアルゴリズムを開発・導入し、タンパク質の全電子計算に世界で初めて達成したプログラムである[3]。ProteinDF は、タンパク質解析に新たな概念をもたらすものとして、製薬、バイオ関連など多方面での応用が期待できる。

量子化学計算におけるタンパク質機能解析は、PDB に蓄積されている 3 次元構造データを基に実行するのが一般的である。PDB に登録されているデータの大半は X 線構造解析から得られたものである。X 線構造解析は分子の 3 次元構造を決める方法として古くから用いられ、非常に強力な解析手法である。しかし、水素原子の座標を決定することができないという欠点があるため、これらのタンパク質構造中には、水素原子の座標情報が含まれていない。量子化学計算では水素原子も含めた完全な分子構造を必要とするため、タンパク質の X 線構造解析データを用いるためには、欠損している水素原子を何らかの方法を用いて付加する必要がある。

Fixing of hydrogen-missing protein structure; Problem of hydrogen addition
Hajime Muta, Tamotsu Yoshihiro, Fumitoshi Sato
Collabolative Research Center of Frontier Simulation Software for Industrial Science, Institute of Industrial Science, University of Tokyo

そのような事情から、これまで、タンパク質を扱う種々のソフトウェアには水素付加を行う機能が搭載されている。しかし、タンパク質データを事前にテキストエディタなどで加工する必要があったり、特定の分子に水素原子が付加されないなどの理由により、多くの場合手作業が発生する。数千～数万原子にも及ぶタンパク質では大変な手間である。また、価電基を全て中性（またはイオン型）にしたいなどの特定のニーズに対してはほとんど対応していないのが現状である。

本研究では、容易に様々なタンパク質機能解析のニーズを満たすよう、PDB データへ水素を付加するプログラムを開発した。プログラムは再利用可能な C++言語によるクラスライブラリとし、特定のアプリケーションに依存しない汎用性の高いものを目指した。

2. 水素付加の手法

2.1 水素付加の対象

本研究における水素付加の対象は PDB から提供されている PDB データに限定する。PDB データには各原子のタイプが記載されており、それらの情報を用いて、水素原子が欠けている場所とその数、および結合様式（結合角、二面角）を決定することができる。

PDB データには様々な分子が含まれているが、基本はタンパク質を構成する 20 種類の通常アミノ酸である。また、PDB の Hetero Group Dictionary を使用すれば、PDB に含まれるアミノ酸以外のヘテログループ、および DNA などの核酸分子、溶媒分子への水素付加にも対応できる。一般分子において水素原子の位置を決めることは困難であるが、本研究では、このようなナレッジベースの手法により、その困難を克服した。

2.2. 内部座標を用いた水素原子の追加方法

一般に、既知の分子骨格に、ある原子を追加する場合、その原子の分子骨格に対する相対的な位置を表す情報が必要である。その与え方と

して最も一般的な方法は、内部座標による指定である。それは図 1 のように、3 原子 i, j, k を指定し、新たな原子 l を以下に示す 3 つの内部座標、

・原子 k との結合長 r_{kl}
・原子 j, k との結合角 θ_{jkl}
・原子 i, j, k との二面角 τ_{ijkl}

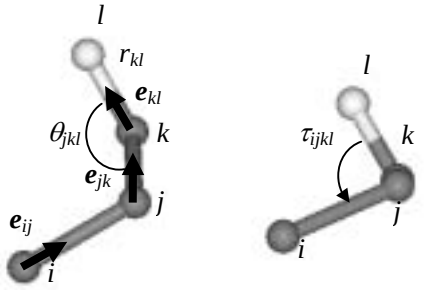

図 1：内部座標による原子の指定

によって指定する方法である。PDB データは内部座標ではなく XYZ 座標で表されているため、内部座標を XYZ 座標に変換する。

図 1 の各原子において、原子 i から j へ、j から k へ、k から l へ向かう単位ベクトル $\boldsymbol{e}_{ij}, \boldsymbol{e}_{jk}, \boldsymbol{e}_{kl}$ を定義する。ここで、$\boldsymbol{e}_{kl}$ を、$\boldsymbol{e}_{ij}$ と $\boldsymbol{e}_{jk}$ で表すことができれば、原子 i, j, k の座標と原子 l の内部座標 $r_{kl}, \theta_{jkl}, \tau_{ijkl}$ から、原子 l の XYZ 座標に相当する $r_{kl} \cdot \boldsymbol{e}_{kl}$ を求めることができる。$\boldsymbol{e}_{kl}$ を求める式は以下のようになる。[4]

$$\boldsymbol{e}_{kl} = a\boldsymbol{e}_{jk} + b(\boldsymbol{e}_{ij} \times \boldsymbol{e}_{jk}) + c\boldsymbol{e}_{ij}$$
$$a = -\cos\theta_{jkl} + p\cos\theta_{ijk}\cos\tau_{ijkl}$$
$$b = p\sin\tau_{ijkl}$$
$$c = p\cos\tau_{ijkl}$$
$$p = \sin\theta_{jkl} / \sin\theta_{ijk}$$

2.3．水素付加の順序

PDB データの中には種々の官能基（原子団）が存在するが、その中には水素原子の位置が一意に決められるものと、そうでないものがある。例えば、メチレン鎖はほとんどの場合 CH_2 であり、2 つの水素原子の位置も、厳密な構造の歪みを無視すれば、ほぼ一意に決められるのに対し、ヒドロキシル基 -OH の水素原子の向きは一般に 3 通り考えられ、任意性がある。この他、タンパク質を構成する官能基で、水素付加に任意性のあるものとしては、カルボキシル基（中性）、チオール基、アミノ基、グアニジノ基（カチオンも含む）、イミダゾール環などがある。しかし、これらは水素原子の向きによっては、他の水素原子の位置と競合する場合がある。そこで当プログラムでは、先に任意性のない基から水素付加を行い、変数を減らした上で、任意性のある基を対象に競合が起こらないような位置に水素を付加する。

2.4．様々な水素付加方法のサポート

上記のように当プログラムの水素付加の方法について説明した。このうち、2.1 および 2.3 節の内容は水素付加のいわば戦略であり、2.2 節の内容は実際に水素を付加する操作を表している。ここで、水素付加の戦略には、2.3 の方法以外にも様々なバリエーションが考えられ、将来もっと良い戦略が導入できる可能性もある。従って、当水素付加プログラムを設計するにあたって、水素付加の戦略に当たる部分を実際の操作と分離し、任意の戦略を、水素付加の別の機能に影響を与えずに自在に定義・変更できる仕組みを導入した。これにより、水素付加のアルゴリズムの改良が容易となる他、様々な水素付加のケースに柔軟に対応することも可能となった。

謝辞

本研究は文部科学省リサーチ・レボリューション計画（RR2002）、IT プログラム「戦略的基盤ソフトウェアの開発」の支援の下におこなわれました。

参考文献

[1] 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/

[2] F. Sato, Y. Shigemitsu, I. Okazaki, S. Yahiro, M. Fukue, S. Kozuru, H. Kashiwagi, Int. J. Quant. Chem., 63 (1997) 245.

[3] F. Sato, T. Yoshihiro, M. Era, H. Kashiwagi, Chem. Phys. Lett., 341 (2001) 645.

[4] 計算化学シリーズ　分子力学法，大澤映二編，町田勝之輔　著，講談社サイエンティフィク，pp. 5-8 (1994)