

2P047

タンパク質のための波動関数データベースの開発
(東大生研) ○平野敏行, 佐藤文俊

【はじめに】

本研究グループはタンパク質のための密度汎関数(DF)法プログラムProteinDF^[1]を基に、タンパク質の統合的な量子化学計算シミュレーションシステムの開発を行っている^[2]。ProteinDFを用いた全電子計算から、タンパク質の精密な波動関数や電子密度が得られ、これらから様々な物理量が計算できる。従って、タンパク質全電子計算の結果は、タンパク質の機能の解明といった基礎研究のみならず、産業界においても医薬品、触媒、遺伝子治療、遺伝子改変、環境有害物質の解析等に役立つデータである。

一方、タンパク質の全電子計算は達成が難しい上に、現在の計算環境では多くの時間と大容量の記憶領域を必要とする。例えば、インスリン単量体のDZ基底関数計算(残基数 51, 原子数 790, 電子数 3078, 軌道数 4439, 補助基底関数 8060)^[3]では、本研究グループが開発した自動計算法^[4]により効率的に計算できるようになったものの、41.6 GFlops の計算機を使って、CPU 時間として約 1 日半を要し、得られたデータを全て残すと記憶容量は約 27 GBにも及ぶ。100 残基になると、CPU 時間として約 4 日、記憶容量として約 132 GB必要になる。今後、計算されるタンパク質の数が増えるにしたがって、ますます膨大なデータ量となることは必至である。現在のところ、個人のレベルでは多くのタンパク質の波動関数を気軽に計算することは難しい。

このように、タンパク質の波動関数はデータベース(DB)として提供するに値するデータであり、その第一弾として 2 次元の画像データによるタンパク質波動関数計算事例集を出版する^[5]。本研究グループではさらに、誰もが自由にタンパク質の解析結果を利用できるよう公開したいと考えている。このような特徴を持ったデータに特化した計算結果検索サービス、取得サービス、登録サービス、データ再利用サービスを行う、タンパク質波動関数DBを構築する方法を検討した。

【設計】

ProteinDF の計算結果から、ユーザに提供するタンパク質波動関数の基本データとして、以下の情報を抽出した。

- ・ 計算に使用した分子座標と原子、アミノ酸、ヘテロ分子の情報
- ・ 計算に使用した基底関数、補助基底関数の情報、その他計算条件
- ・ 残基数、原子数、電子数、軌道数、補助基底関数の数、計算機と計算時間、ディスク容量
- ・ 全 LCAO 係数、電子密度と KS 軌道エネルギー分布
- ・ 最高被占有軌道(HOMO)と最低非占有軌道(LUMO)近傍の軌道の三次元グラフィックスデータ
- ・ 等電子密度面、静電ポテンシャル分布の三次元グラフィックスデータ

具体例として図 1、図 2 にそれぞれインスリンの分子軌道(HOMO)図、KS軌道エネルギー図を示した。KS軌道エネルギー分布(図 2)から明らかのように、主に反応に関与する軌道として注目されるHOMO、LUMOの周辺 (-10 ~ 0 eV) には多くの軌道が密集しており、これらも反応に関与する可能性が高い。そこで、HOMO、LUMO だけでなく、その近傍の軌道データも登録する必要があるだろう。タンパク質の会合や基質との相互作用を解析する上で、タンパク質表面の静電ポテンシャル分布といった物理量も欠くことができない。こうしたグラフィックスデータの作成にも、多くの計算リソースを要する。そこで、予めグラフィックスデータをサーバ側で作成・蓄積

することにした。これらのデータはローカルにダウンロードすることで、ProteinDFシステムのGUIプログラム(プロテイン・エディタ)^[6]などで表示することができる。

ProteinDF 固有の入出力データの他、参照した PDB(<http://www.rcsb.org/pdb/>)のエントリ番号やアミノ酸シーケンス、構造緩和の手法等、アノテーション情報も登録される。こうした情報を基に本 DB を検索することが可能である。PDB をはじめとする公共分子生物学 DB との連携を図り、相補的に利用することで、タンパク質に関する新しい情報や知見を容易かつ詳細に得ることが期待できる。

【本 DB へのアクセス】

タンパク質波動関数データベースを一般に公開する手段として、多くの分子生物学DB同様、インターネットを利用した。webシステムは、LAMP(Linux, Apache, MySQL, PHP)により構築した。サーバマシンはIBM-PC/AT 互換機(CPU: Pentium4 2.8 GHz, RAM: 2 GB)を使用した。webページはユーザの指示に従い、動的に作成される。出力されるページは標準的なHTMLで記述されているため、ユーザはInternet ExplorerやNetscape, Mozilla等で検索・表示することができる。今後webブラウザに限らず、プロテイン・エディタ^[5]においても、SOAP等を用いて、本DBをシームレスに利用できる環境を用意する予定である。

【データの標準化】

XML は機種に依存しないデータ形式であり、情報の管理と交換に適したフォーマットである。近年、公共分子生物学 DB においても XML が積極的に利用されている。本 DB においても XML 文書の入出力をサポートしている。XML データを利用することで、ユーザは自由にデータの処理・加工を行うことができる。現在、本 DB が用いる XML のスキーマは独自のものを利用している。今後、タンパク質の量子化学計算に適した XML スキーマの標準化が望まれる。

本研究は、文部科学省 IT プログラム「戦略的基盤ソフトウェアの開発」において実施された。

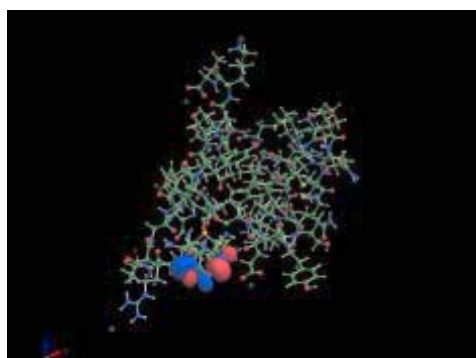


図1 インスリンの分子軌道(HOMO)図

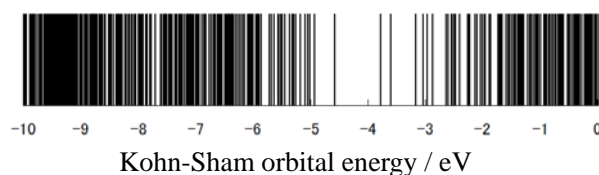


図2 インスリンのKS軌道エネルギー分布図

【参考文献】

- [1] F.Sato, T. Yoshihiro, M. Era, H. Kashiwagi, *Chem. Phys. Lett.*, **341** (2001), 645.
- [2] 佐藤文俊 他, 分子構造総合討論会, 3D05 (2005).
- [3] T. Inaba, S. Tahara, N. Nisiwaka, H. Kashiwagi, F. Sato, *J. Comput. Chem.*, **26** (2005), 987.
- [4] H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro, F. Sato, *Mol. Phys.*, **101** (2003), 81.
- [5] “ProteinDFによるタンパク質量子化学計算事例集”, アドバンスソフト (2005) in press.
- [6] 西村康幸, 吉廣保, 西野典子, 上野哲哉, 佐藤文俊, 分子構造総合討論会, 4P125 (2004).