

# 構造生物学と IT

有井 康博

(要旨) 生命科学研究における IT の利用の一つを紹介する。著者が携わってきた構造生物学という分野では、様々な IT の恩恵を授かっている。近年、クラウド技術を利用した研究ツールやデータベースの統合化が始まったことから、この分野における科学情報の利用しやすさが格段と上がった。そのクラウドシステムの中にある、タンパク質を可視化するソフトである PyMOL を選んで、パソコン上でタンパク質のモデル図を描き、レンダリングした画像を紹介する。これまで研究レベルで用いられてきたツールやデータベースが学部教育の現場で利用される時代は直ぐそこまで訪れている。

**キーワード** : 構造生物学, 創薬等支援技術基盤プラットフォーム, PyMOL

## 1 はじめに

著者は広義な意味で生命科学に携わっている。様々な科学という営みの中で、生命科学という分野は IT の多大な恩恵を受けている分野の一つと言える。今や、インターネット上のツールを使用することなく、生命科学の研究を進めるのは困難になってきている。生命科学の分野の潮流として、1930年代後半から生命現象を理解するために分子を観察しようという試みが頻繁に行われてきた。この行為は、'森を知るために森の中の木を観よう'、'森に住む生き物の動きを観察しよう'と例えられる。この試みを進める原動力となったのが分子生物学である。分子生物学では、'分子が分かれば、生命が解ける'という思想のもとに研究が進められてきた。しかしながら、ヒトゲノムの全配列がほぼ解読され、テクノロジーの進歩で組換え型タンパク質を生産することが格段に容易くなり、様々なタンパク質の機能が明らかとなってきているが、生命現象の最大の謎である '生きていく' という点に関しては未だ分からないままである。そして、多くの研究者が分子生物学に限界を感じ始めており、著者は分子生物学の全盛期は終焉したと感じている。勿論、分子生物学が生み出した技術や知見は未来永劫に残って行くであろうし、新しい研究分野に引き継がれていくだろう。そのような流れの中で、分子生物学の発展の後半を支えてきた情報生物学という分野における高度化された IT は、研究のみならず教育にも利用できるほどに成熟している。本稿では、著者が携わってきた構造生物学研究で用いられる IT の一端を紹介する。

## 2 創薬等支援技術基盤プラットフォーム

構造生物学とは、タンパク質や核酸などの巨大な生体分子について分子あるいは原子レベルで構造・機能を研究する分野である。この研究分野はタンパク質の構造と機能には相関性があるという不文律のもとに成り立っている。タンパク質は 20 種類のアミノ酸がペプチド結合により繋がったポリペプチドである。ポリペプチドには方向性があり、生体がもつ遺伝子情報により順序と配列が決めている。これをタンパク質の一次構造と呼ぶ。1973年に Anfinsen が行った変性タンパク質の巻き戻し実験の結果から、タンパク質の立体的な天然構造は熱力学的に安定な状態であり、一次構造によって決まると考えられている<sup>1)</sup>。ちなみに変性状態はタンパク質の立体的な形(天然構造)が崩れた状態を意味する。これらのことを基に、構造生物学では一次構造から立体構造を予測し、機能も推測できると旗を上げ、タンパク 3000 プロジェクトやターゲットタンパク研究プログラムという国家プロジェクトが立ち上げられた<sup>2)</sup>。この動きは国際的な動きで、世界中で激しい研究競争が繰り広げられてきた。その競争は多彩な研究成果、様々な新たなデータベース、様々な研究ツールを生み出した。近年、その情報基盤を様々な研究開発に活用できるように、使い勝手のよい情報基盤「構造生命科学データクラウド」として提供する試みが始まった。文部科学省創薬等支援技術基盤プラットフォーム情報拠点である<sup>3)</sup>。この試みの計画目的には次の3つが上げられている。

- ① これまで構築されたデータベースと利用ツールを継承・更新・運用し、さらに高度化して、創



図1 創業等支援技術基盤プラットフォーム  
情報拠点のホームページ  
[http://p4d-info.nig.ac.jp/mediawiki/index.php/情報拠点\\_ツールとデータベース](http://p4d-info.nig.ac.jp/mediawiki/index.php/情報拠点_ツールとデータベース)

業等支援技術基盤プラットフォームの解析拠点と制御拠点が必要とする情報の蓄積・提供と両拠点の成果普及を行うことによって、関連分野の研究を支援すること

- ② 構造生物学にゲノム情報・遺伝子発現情報・遺伝子ネットワーク等の他分野の生命情報を取り込みつつ、多種多様なデータベースを利用形態の差異を意識することなく活用できるようにすること
- ③ 「構造生命科学データクラウド」を構築して、広範な生命科学の研究分野で構造生物学の先端研究成果を活用できる情報基盤を提供すること

目的からは研究色が強く感じられるが、この試みは、これまでに得た知的財産を、大学における専門教育にも活用できるレベルまで整備するという点だと著者は理解している。故に、生命現象を学ぶ学部学生たちには積極的に利用して欲しいと願う。

さて、そのクラウドの中には、タンパク質の配列・構造・相互作用・解析に関するデータベースとツールが集積された便利なサイトが用意されている(図1)。実に250種類以上である。全サイト項目に簡単な説明が付記されており、調べたいことに応じた情報を初心者にも見つけ易いように工夫されている。是非とも参照していただきたい。

### 3 PyMOL

ここでは、上述のクラウドシステムの中からダウンロードし使用できる、タンパク質の立体構造を可視化

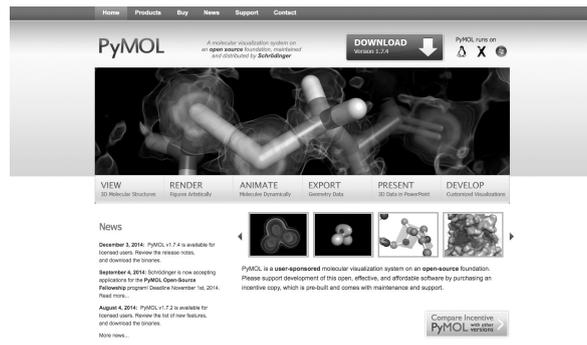


図2 PyMOLのTOPページ

するソフトの一つであるPyMOLEduを紹介する。なお、PyMOLEduは教育を目的に使用するソフトである。PyMOL(図2)はタンパク質立体構造をグラフィックス表示し、レンダリングによる美しい画像、アニメーションを制作できるオープンソースソフトである<sup>4)</sup>。教育目的には無料ソフトが配布されており、簡単な手続きを経て、誰でも利用することができる。ちなみに、タンパク質の立体構造をPyMOLで表示するには、タンパク質の原子を座標化したデータファイルが必要である。そのデータファイルはResearch Collaboratory for Structural Bioinformatics(RCSB)が管理するアーカイブサイトのProtein Data Bank(PDB)<sup>5)</sup>に保存されており、PDBファイルと呼ばれる。このファイルにはX線結晶構造解析などの技術によって明らかにされた立体構造の構造データが書き込まれており、厳正な審査を経て登録されている。利用は無料である。X線結晶構造解析とは、高純度に精製した高濃度のタンパク質を結晶化し、その結晶にX線を照射することで得られる回折像をもとにコンピューティングによりタンパク質中の原子の位置を決定する技術である。波長の短いX線を長時間照射するほど高解像度のデータが得られるが、結晶が溶けるリスクを伴うために、良質な結晶を得ることが不可欠となる。1.0 Å以下の高解像度の回折データを得ると水素原子を観ることが可能。多くの化学反応では、化学反応の詳細を知るために水素原子の観察が重要な情報となる。そのような高解像度のデータを得るために、研究者たちは筑波、播磨、佐賀にある大型放射光施設を利用する。PDBで目的のタンパク質名あるいは学術論文に記されているPDBIDを入力し、検索にかけると、目的のデータが表示される。表示された構造データの中から、自分の観察したい構造を含む構造データをダウンロードして保存する。ファイルには構造データ以外の有用な様々な情報が含まれる。それらの情報、例えば生物種、基質の結合の有無、分解能などが使用するPDBファイルの選択肢となる。

実際に PyMOL が映し出すグラフィックスを観てみよう。ここでは、著者が X 線結晶構造解析を用いて明らかにし、PDB に登録した、ナトリウム結合型アネキシン IV の立体構造 (PDBID: 2zhi)<sup>6)</sup> を利用して、スティックモデル (図 3)、カートゥーンモデル (図 4) を表示し、レンダリングした画像を示した。さらに、ナトリウムが結合する様子をクローズアップして表示した (図 5)。このような画像の提示は、研究の世界においては説得力のある強力な武器となり、思考の手助けとなる。本ソフト上では、図のようなタンパク質分子をカラーでクルクルと回転させることができ、クローズアップすることも可能である。まさに手に取るようにタンパク質を原子レベルでバーチャルに観察することができる。是非とも教育の現場で利用し、学生の想像力をかき立ててもらいたい。

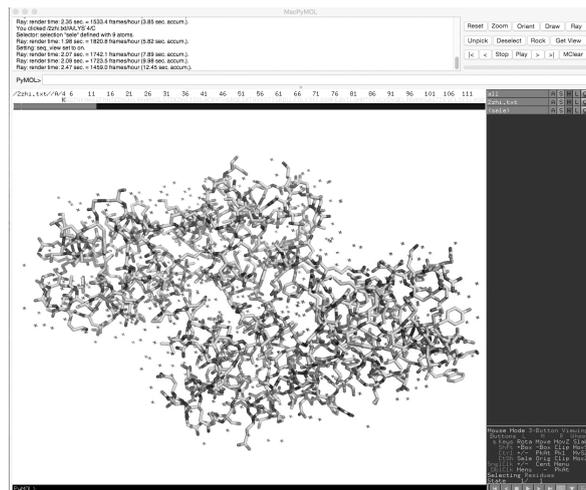


図 3 PyMOL の画面とスティックモデル表示

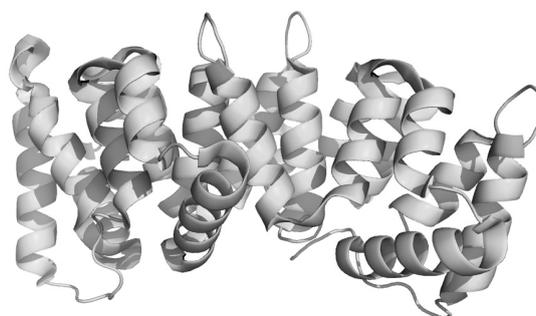


図 4 カートゥーンモデル表示

#### 4 参考文献リスト

- (1) Anfinsen, CB (1973) Principles that govern the folding of protein chains. *Science*, 181, 223-230.
- (2) ターゲットタンパク研究プログラム  
[http://www.tanpaku.org/about/old\\_project01.php](http://www.tanpaku.org/about/old_project01.php)
- (3) 文部科学省 創薬等支援技術基盤プラットフォーム 情報拠点  
<http://p4d-info.nig.ac.jp/mediawiki/index.php/情報拠点 TOP>
- (4) PyMOL  
<http://www.pymol.org>
- (5) RCSB Protein Data Bank  
<http://www.rcsb.org/pdb/home/home.do>
- (6) Butsushita, K, Fukuoka, S-I, Ida, K, Arii, Y (2009) Crystal structures of sodium-bound annexin A4. *Biosci. Biotechnol. Biochem.*, 73, 2274-2280.

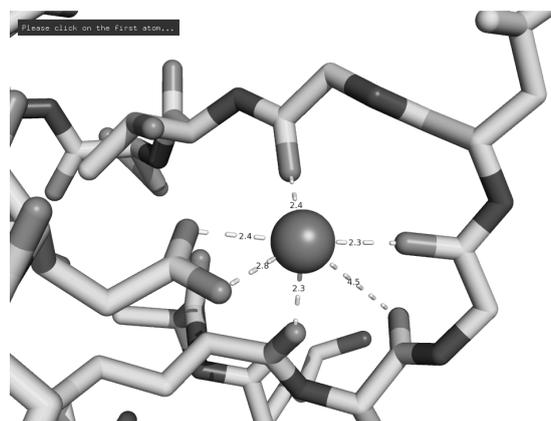


図 5 ナトリウムの結合様式  
中央のボールがナトリウムである。