

タンパク質分子構造を例とする 高性能計算結果の可視化システムの試作

平尾智也[◇], 水原隆道[◇], 齊藤哲哉[◇],
安藤 誠[†], 秋山 泰[‡], 城 和貴^{*}, 國枝義敏[◇]

◇ 和歌山大学 システム工学部
† 日本鋼管(株) 情報システム部
‡ 技術研究組合 新情報処理開発機構
* 奈良女子大学 理学部

高性能計算を通じて生成される膨大な計算結果の可視化手法の研究の一環として、我々はタンパク質分子構造の高度な可視化を行なう“ProteinView”システムを作成した。ProteinViewはWindows 95/98/NT4.0上で動作し、タンパク質分子の立体構造を三次元の仮想空間内に効率よく表示するとともに、ジョイスティックを用いて操作者の視点位置を滑らかに移動させることができる。従来のシステムの多くが、分子構造全体を外から眺めることに重点を置いてきたのに対し、本手法では巨大分子の内部空間における分子の局所構造の詳細な検討が容易に行なえる。本稿では、ProteinViewの設計上の特徴や、実装手法などを述べる。

ProteinView: a Highly Interactive Visualization System for Protein Structure Analysis

Tomoya Hirao[◇], Takamichi Mizuhara[◇], Tetsuya Saito[◇],
Makoto Ando[†], Yutaka Akiyama[‡], Kazuki Joe^{*}, Yoshitoshi Kunieda[◇]

◇ Faculty of System Engineering, Wakayama University
† NKK Corporation
‡ Real World Computing Partnership
* Faculty of Science, Nara Women's University

We develop a highly interactive protein conformation visualising system “ProteinView.” ProteinView visualises a protein molecular structure in a 3D virtual space and be able to move user's viewpoint smoothly using a joystick on Windows 95/98/NT4.0. Many conventional systems respect to view the molecular structure from their outside. But this method provides the way to check up the local structure within a huge molecule in detail. In this paper, we describe design features and an implementation of ProteinView.

1. はじめに

多くの高性能計算は膨大な量の計算結果を生成する。そのすべてを把握し、計算結果の比較や検討を行うのは難しい。そういった問題を解決するために、計算結果の可視化手法の研究が盛んに行われている。特に高性能計算を通じて

膨大な量の計算結果が得られる分野としては、流体力学、構造解析などがある。このような分野で高性能計算の結果を視覚化を行う場合、AVS [1] のような汎用の視覚化ツールを用いることが多い。しかし、表示速度を要求されたり、表示の対象となるものの形状が複雑であるような場合には独自のツールが開発されてきた。

今回、我々がタンパク質構造の描画をターゲットとした理由として、この分野では高速な描画と高度な操作性を要求される研究課題が多いこと、将来的に多くの利用者が期待できることが挙げられる。

近年、ゲノム(全遺伝子)解析プロジェクトの進展とともに、タンパク質に関する膨大なデータがデータベースに登録されつつある。タンパク質の配列データでは約8万件(SWISS-PROTデータベース)、立体構造データでは約9千件(PDBデータベース [2])のデータが公開されており、急速なペースでデータ数も増加している。これらの既知の立体構造の特徴を調査したり、あるいは配列情報から計算機によって立体構造を推定する研究においては、多数の原子から構成されるタンパク質の立体構造を高速に描画することが極めて重要である。タンパク質分子の場合、数百原子から数万原子までの膨大なオブジェクトを描画しなければならず、描画の効率のよさが求められる。

さらに、前述したような他の高性能計算分野と比較して、観察者の意図に応じた回転・平行移動・拡大・縮小などの処理がきわめて頻繁に行なわれる。これは、タンパク質分子が巨大な階層的な系であり、観察者は、その全体だけではなく、詳細な部分構造をさまざまなレベルで吟味する必要があるからである。このために、きわめて操作性に優れたシステムである事が求められる。

次の2節では我々が新たなタンパク質構造可視化システムの開発に踏み切った理由について触れる。3節ではその設計方針、4節では現在の ProteinView の実装の状況を述べる。5節では導入を検討している機能や改善、改良のポイントについて議論する。6節では関連のある研究に触れ、7節で結論を述べる。

2 背景

タンパク質は通常、溶媒中ではある決まった構造に折りたたまれていることが知られている。折りたたまれることで原子の相互位置が変化して生じる種々の形態のことをタンパク質の立体配座といい、結合の回転(ねじれ)によって生成される異性体のことを配座異性体という。タンパク質の立体配座を解析するシステムとして、安藤らによる“ESCAPE” [3]がある。このESCAPEによる解析結果を視覚化するこ

とを主目的として ProteinView は開発が開始された。このため、ProteinView は次のような目的で利用することができる。

- PDB データベースに登録されているタンパク質の構造を静的に表示する。
- 立体構造予測システムから出力されたタンパク質構造を表示する。
- 立体配座解析システムからの出力を表示する。

3 ProteinView の設計

3.1 オブジェクトの表現

ProteinView では、原子を表す原子オブジェクトと、原子間の結合を表す結合オブジェクトの2つを組み合わせることで分子構造を三次元仮想空間内に描画する。原子オブジェクトは球で表現される。原子間の結合には共有結合、イオン性結合、水素結合、疎水性結合などがあるが、配座解析の用途では、共有結合を円柱として表現している。

前述のESCAPEは存在可能な立体配座を解析する際に原子を球、共有結合を円柱とみならず、解析の際のモデル化と計算結果の可視化の際のオブジェクトの表現を一致させることで、ユーザにわかりやすいビューを提供できると考え、このような仕様とした。多くのタンパク質構造を可視化するシステムには、分子をリボンモデルなど、さまざまな形式で表示する機能が実装されている。しかし、本システムでは、開発目的から考えて必ずしも必要でないこと、現在のところ試作段階であることなどの理由で、現段階では、リボンモデルでの表示は実装していない。

3.2 分子構造の表示

分子構造の3次元空間への描画を実現するために、OpenGLを用いた。OpenGLを用いることで、容易に3次元空間への描画を実現できるだけでなく、OpenGLのハードウェアサポートを受けられる環境であれば、かなり高速に描画できるものと期待できる。また、OpenGLによる描画に時間がかかる環境でも利用が可能にするために、通常の描画のほかに、ワイヤフレームによる表示もサポートする。

3.3 入力ファイル

ProteinView は HTML に似せた独自のタンパク質構造記述仕様が記述されたテキストファイルを入力ファイルとしている。この記述仕様は基本的に原子オブジェクト、結合オブジェクトの座標と色を羅列しているだけなので、タンパク質に限らず、あらゆる分子構造を記述することができる。

3.4 出力ファイル

出力ファイルとして、2種類の形式を準備している。

一つは VRML ソースファイルである。ProteinView は三次元仮想空間内に再現したタンパク質構造をそのまま VRML ソースファイルにコンバート・出力することができる。これによって、タンパク質構造を WWW 経由で公開することが容易に実現できるだけでなく、VRML ブラウザが動作する環境であれば、OS やその他の環境に依存すること無く、どこでもタンパク質構造を確認することができ、プレゼンテーションにも威力を発揮するものと期待できる。なお、出力されるソースファイルは、VRML 2.0 の仕様に基づいている。

もうひとつの出力形式は BMP ファイルである。ProteinView は Windows 9x/NT4 上で動作するため、その標準的な画像形式である BMP をサポートする。

3.5 入力デバイス

視点の移動のために、キーボード、マウス、ジョイスティックの3つのデバイスによる入力を可能にしている。表示されている原子オブジェクトにカーソルをあわせてクリックもしくはジョイスティックのトリガを引くと、その原子オブジェクトに関連する情報が表示される。

3.6 変形操作

ProteinView は表示されている分子構造を変形する機能を提供する。配座解析を行うために必要な機能である。タンパク質内の共有結合は、結合角、結合長はほぼ不変であるが、結合部を軸とした回転は大きな自由度を持つ。そこで、

表示された分子構造の中から任意の共有結合を選び出し、そのいずれかの端を共有結合を軸に回転させる。回転させたい共有結合を指定する際に、結合オブジェクトのどの位置をマウスなどでポイントしたかによって、共有結合のどちらの端を回転させるかを決定する。回転量はマウスの移動量によって決定する。

4 実装

4.1 データ構造

ファイルから入力されたデータは原子オブジェクト、結合オブジェクトそれぞれの線形リストとして内部に保持される。これらの線形リストは各オブジェクトの座標、大きさ、色等のデータを保持している。描画の際は、これらの線形リストをそれぞれ1回ずつ走査する。データ構造が単純なので、データの読み込みから描画までに無駄に時間がかかることが無い。

4.2 動作環境

ProteinView は OpenGL を利用することができる Windows 95/98/NT4.0 上で動作する。ジョイスティックは現在のところ Sidewinder force feedback pro にのみ対応しており、その利用には DirectX 5.0 が必要になる。

なお、開発は主に Windows NT4.0 workstation 上で Visual C++ を用いて行った。

4.3 現状

本稿執筆時点で実装済みの機能、未実装の機能はそれぞれ以下のとおりである。

- 実装済みの機能
 - 分子構造の3次元表示
 - ワイヤフレーム表示のサポート
 - ジョイスティックによる操作
 - VRML および BMP によるファイル出力
- 未実装の機能
 - 変形操作
 - PDB データベース形式への対応

図1は ProteinView で分子構造を表示しているところである。画面内の x, y, z の3軸については表示/非表示の切り替えが可能になっている。また、画面下部のメッセージボックスには入力ファイルのエラー情報や、マウスなどによって指定した元素名など、原子オブジェクトの固有情報が表示される。

図2は図1で描画されている分子構造を VRML ファイルとして出力した結果を VRML ブラウザで表示している様子である。

5 議論

この ProteinView は現段階では、入力ファイルから与えられた原子の座標と原子間の結合をそのまま表示するだけのものである。しかし、将来的に高性能計算によって得られた予測構造を表示するためのツールにするために次のような改善を現在検討中である。

1. 共有結合部の回転および、原子の衝突判定
高性能計算によって予測された配座を可視化した結果にユーザが修正を加える必要が出てくるのが考えられる。このような場合に、簡単に配座を変更できるように、共有結合部を回転することを可能にする必要がある。共有結合部を回転させた結果、原子同士が干渉する可能性もあるので、衝突判定の機能も実装すべきであると考えている。なお、VRML のソースファイルを出力した場合にも、共有結合部が回転することができるように、Java や JavaScript との連携も検討している。
2. PDB データベース形式への対応
タンパク質を扱う以上、タンパク質立体構造データベース PDB (Protein Data Bank) から得られる構造を直接入力として扱えるようにする必要があると考えられる。このためには、PDB 形式のパーザを ProteinView に実装する必要がある。PDB のパーザはすでに PAPIA (PARallel Protein Information Analysis) library [4] に準備されているので、これを利用することも検討中である。PDB データベース形式に対応することで、大幅に実用性が増すものと期待している。
3. アミノ酸残基レベルでの表示/非表示の切り替え

観察の対象でないアミノ酸残基が表示されているとタンパク質構造の観察を邪魔することがある。アミノ酸の配列情報が表示される別のウィンドウ内でアミノ酸残基ごとに表示・非表示の設定をすれば、不要なアミノ酸に邪魔されることなくタンパク質構造を観察できるように拡張する予定である。

4. コンパス機能の導入

視界からタンパク質が外れるような位置に視点が移動してしまう構造物との相対的な位置関係がつかめなくなり、どこに構造物があるのかわからなくなってしまう。これを防ぐためには、タンパク質構造を見失ったときにタンパク質構造はどの方向にあるのか、自分はどちらを向けばよいのかを示すコンパスのようなものを導入すればよい。また、視点がタンパク質内部に入り込んでいるようなときには、常にユーザが指定したオブジェクトを指しつづけるようなコンパスがあれば便利である。タンパク質構造の可視化という目的から考えると、あまり本質的な問題ではないので現在のところは実装していないが、操作性を考えると、将来的には導入すべき機能である。

6 関連研究

- RasMol [5]
Roger Sayle が作成したフリーウェアである。球と円柱だけでなく、さまざまな描画方式がサポートされている。ただし、離れた視点からの二次元的な画像しか得られない。回転処理は非常に高速なので、全体像を眺めるのには向いている。
- MOSBY [6]
電総研の上野が開発したシステムである。MOSBY 単体では PDB データベースのタンパク質の立体構造を表示するプログラムとして利用できるが、これに解析・計算用のモジュールをプラグインとしてユーザが自由に加えることができる。これに対し、我々の ProteinView は、高性能計算が表示系の外部にあると仮定して設計と実装を進めている。原子の衝突チェックのように、計算機能が描画機能の中に密接に入ったほうが良い場合も多いが、一方で、並列計算システムによる計算と、表示系を分け

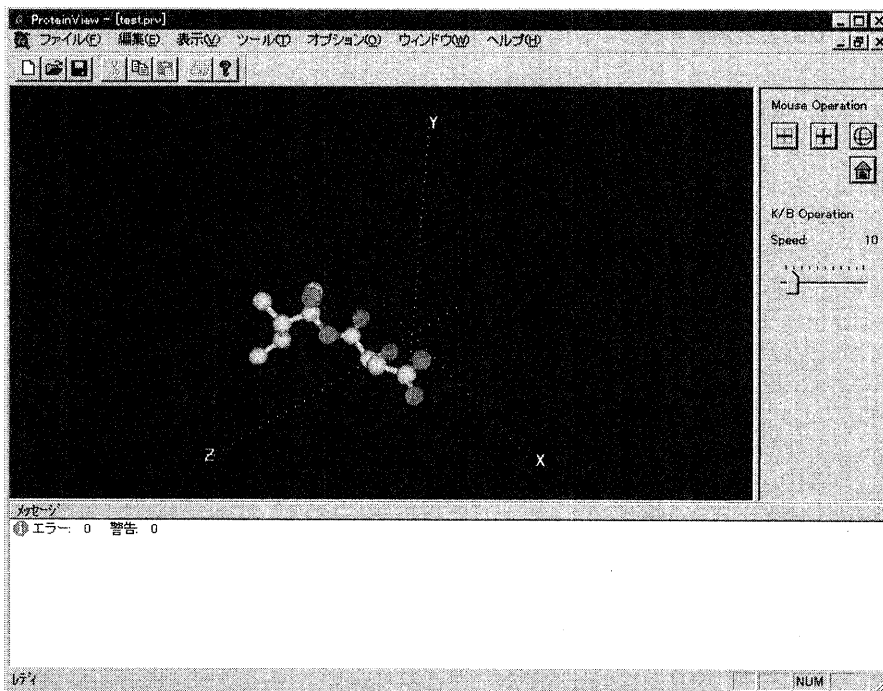


図 1: ProteinView による分子構造の表示

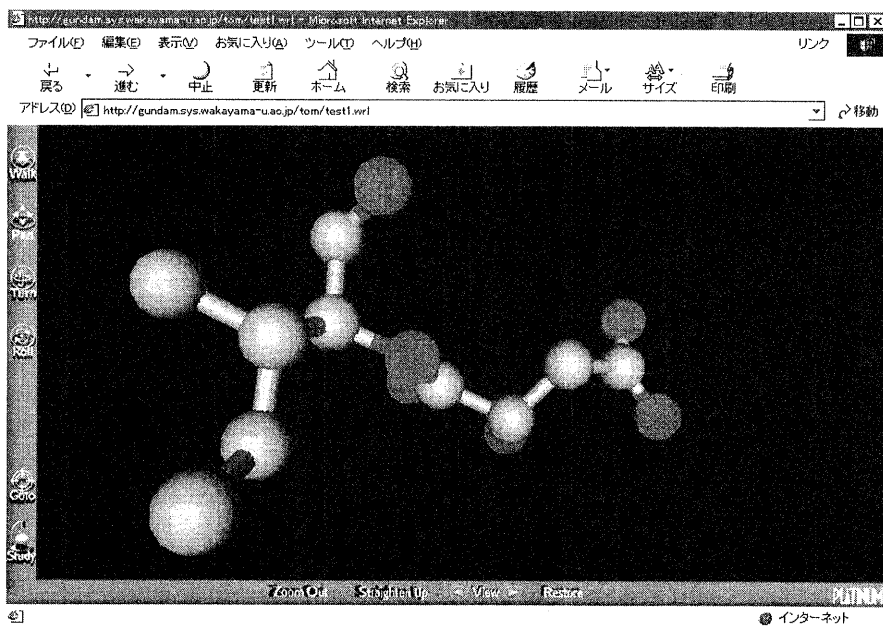


図 2: VRML ブラウザでの表示

たい場合などには、我々のアプローチの方が適切であると考えられる。

- **insightII** [7] 米国 MSI 社の高額の商用ソフトウェア。種々の描画モードがあり、別売の計算モジュールなどと連携して、計算結果を表示させられる。回転や拡大縮小などはマウス、または画面上のボタンのクリック等で行なう。基本的に視点はタンパク質の外部に置かれている。この点が、内部を自由に視点移動できる ProteinView との大きな差異である。同様のソフトに QUANTA があったが、会社の合併により現在は insightII に機能が吸収されている。同様のものとして、富士通の BioResearch シリーズや、Protein Adviser などの多数のパッケージソフトがある。

7 結論

ProteinView では、仮想空間内を視点が移動するという、この種のシステムとしては他に無い視覚化手法を実装した。視点が移動するという手法は、ジョイスティックでの操作との親和性も高く、この結果、優れた操作性を得ることに成功した。一方で、ProteinView はまだ開発は始まったばかりであり、他の類似のシステムと比較すると、実装している機能は不十分なものとなっている。今後は、ProteinView の機能の充実を図り、タンパク質構造可視化システムとしてより実用的なものにしていく予定である。

ProteinView はタンパク質構造の視覚化に特化したシステムであるが、これを応用することで、各種の CAD システムなどさまざまな分野での可視化システムにとりして利用可能になると考えられる。

なお、本システムはプログラム並列化支援ツール “NaraView” [8] を Windows NT 上に移植した “NaraView/NT” の 3 次元表示システムを応用し、タンパク質の分子構造の描画用に発展させたものである。

参考文献

- [1] AVS, Advanced Visual Systems Inc., <http://www.avs.com/>
- [2] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi: The Protein Data Bank: a computer-based archival file for macromolecular structures, *J. Mol. Biol.*, **112**, pp. 535-542 (1977). (<http://www.pdb.bnl.gov/>)
- [3] 安藤 誠, 秋山 泰, 鬼塚 健太郎, 野口 保: 木探索アプローチによるタンパク質立体配座解析と大規模並列計算機上での高速解析システムの構築, 情報処理学会論文誌, Vol.40, No.SIG2(TOM1), pp.91-104 (1999).
- [4] Y. Akiyama, K. Onizuka, T. Noguchi, M. Ando: Parallel Protein Information Analysis (PAPIA) system running on a 64-node PC Cluster, Proceedings of the 9th Genome Informatics Workshop (GIW'98), *Universal Academy Press*, pp.131-140 (1998).
- [5] RasMol ホームページ <http://www.umass.edu/microbio/rasmol/>
- [6] Ueno, Y. and Asai, K.: A High-Throughput Graphics Library Designed for a Portable Molecular Structure viewer, Proceedings of the Pacific Symposium on Biocomputing *Altman et al., eds. World Scientific Press*, pp.201-212 (1998).
- [7] insightII, Molecular Simulations Inc., <http://www.msi.com/>
- [8] Sasakura, M., Joe, K., Kunieda, Y., Araki, K.: NaraView: an interactive 3D visualization system for parallelization of programs, *International Journal of Parallel Programming*, *Plenum Press*, Vol.27, No.2, pp.111-129 (1999).