タンパク質量子化学計算システム統合環境 ProteinEditor の開発

○西村康幸[1]、吉廣　保[1]、佐藤文俊[1, 2]
1 東京大学 生産技術研究所
2 東京大学 情報基盤センター

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]。本システムは「計算エンジン ProteinDF」を中心に、「自動計算法」、「構造最適化・*ab initio* MD 計算」、「励起状態計算」、「タンパク質波動関数データベース」で構成されており、これらを統括するグラフィカル･ユーザー･インターフェース（GUI）が「ProteinEditor」である。これまで ProteinEditor は量子化学計算に関わる研究者のみならず、生物物理学や生化学の研究者を想定して開発され、現在では複雑なタンパク質の基底状態全電子計算達成をサポートする機能をほぼ全て装備している[3]。

本発表では、ProteinEditor の各機能、特にタンパク質構造の編集機能、更には Virtual Reality（VR）技術を取り入れたタンパク質モデリング支援機能について報告する。

タンパク質量子化学計算システム： 今でこそコスト高なタンパク質量子化学計算も、今後の計算機の発展を考慮に入れれば、まもなく誰でも簡単に実行できる時代が到来することは明らかである。これまで古典論による解析が主であったタンパク質の研究に一大革命をもたらすであろう。このような時代に先駆けて、ソフトウエアを整備することが急務の課題となっている。当研究グループでは ProteinDF をベースに、1,000 残基規模の超大規模密度汎関数計算が可能なソフトウエアとタンパク質の精密な分子動力学計算や励起状態計算が可能なソフトウエアを開発することを目標としている。また、実際に 100 種類程度のタンパク質の全電子計算を実施し、タンパク質の高精度の電子状態データベースを構築する。これらの目標をカップルさせることによって、ProteinDF システムとして公開し、本ソフトウエア、データベースともに世界標準（デファクトスタンダード）とすることを目指している。

ProteinEditor： ProteinEditor は、システムを統括する機能に加えて、水素付加、アミノ酸置換、化学修飾などのグラフィカルな分子編集機能、古典論による構造最適化および構造妥当性評価機能、更には各種の機能表現に必要なタンパク質のための大規模分子グラフィックス、およびその GUI で構成されている。ProteinEditor は全てスクラッチから開発されており、Windows に対応し、分子グラフィックスは OpenGL を用いて描画している。グラフィックスデータの一つは AVS の field data に準拠している。タンパク質をターゲットにした GUI では、分子グラフィックスの性能が特に重要で、大量の原子と複雑な物理量を如何にわかりやすく、しかも高速に表示するかが鍵となる。ProteinEditor では総原子数が 12 万原子のタンパク質を Pentium4 3.06GHz+ RADEON9700PRO を使用して、全原子モデル構築に 2 秒で対応することができる。一度モデルが構築されれば、回転、拡大縮小などの操作はそれほど違和感のないスピードで行うことができ、十分な性能を達成している。また、タンパク質の全電子計算を安全に達成するには高品位な分子構造が必要となる（後述）。図 1 に、ProteinEditor を用いたタンパク質全電子計算の一連の作業の流れを模式的に示した。分子編集機能による計算構造作成、ProteinDF による量子化学計算ジョブ制御ならびに結果解析に至る各種操作が対話的に行なえることが特徴である。

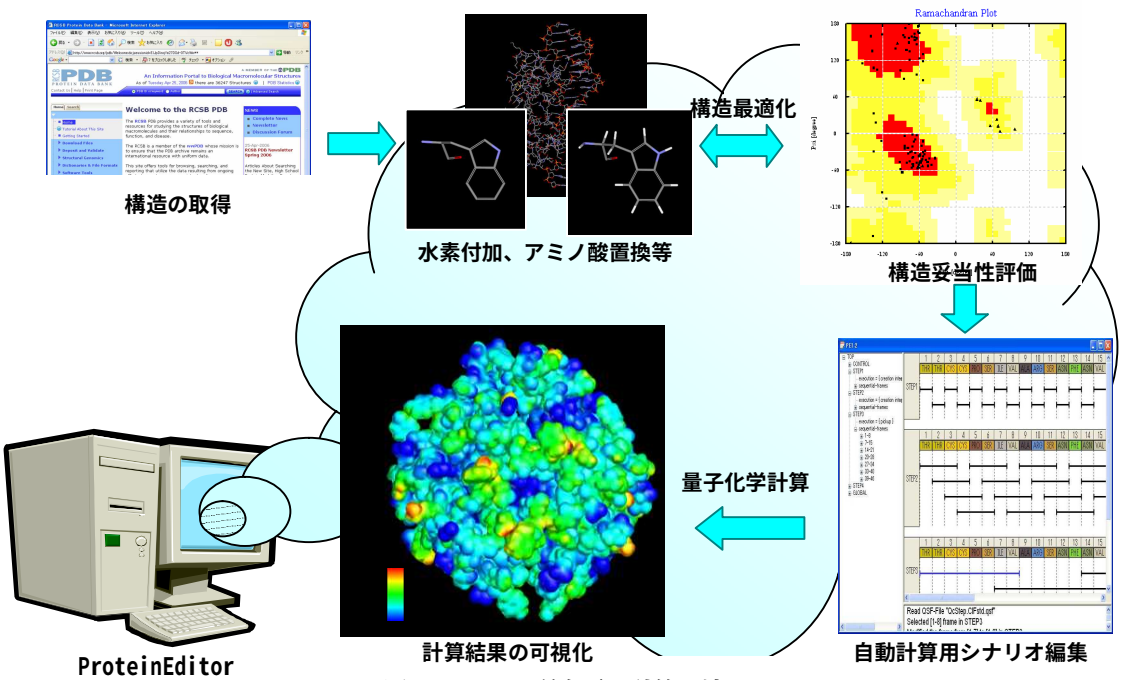

図 1 タンパク質全電子計算の流れ

量子化学計算に最適なタンパク質立体構造の作成支援機能： ProteinDF によるタンパク質の量子化学計算では、計算の初期構造に Protein Data Bank (PDB)[4]で配布されている三次元構造データを用いることが多い。PDB に登録されている立体構造は X 線構造解析や多次元 NMR などで求められたものである。PDB データの 8 割以上が X 線構造解析によって求められており、これらは水素原子のデータを持っていない。さらに多くのデータに共通して、化学的な見地から問題と思われるほどの構造歪みや原子間衝突があり、そのままの構造を用いた量子化学計算結果は判断を誤らせる危険がある。そのため、タンパク質の量子化学計算を実行するためには、PDB データに水素付加などの処理を行うだけでは不十分で、分子構造に問題が無いか、その妥当性をチェックし、場合によっては補正する必要がある。そこで ProteinEditor には、PDB データへの水素付加機能や古典分子動力学(MD)と連動させて構造歪みを緩和する機能、および構造に問題がないか構造妥当性のチェックを支援する機能を実装した。また、アミノ酸置換や化学修飾を行うことも可能である。このような構造の精密化が GUI で容易に行うことができるようになっている。

VR 技術を取り入れたタンパク質モデリング支援機能： 当グループではさらに、ProteinEditor のタンパク質モデリング支援機能に VR 技術を取り入れる試みを始めている。この VR 技術を取り入れたタンパク質モデリング支援機能(タンパク質モデリング VR システム[5])は、複雑なタンパク質立体構造の 3 次元空間における原子位置関係とそれらが感じる場を正確に把握し、インタラクティブな構造（原子、アミノ酸残基、基質、配位子）の編集を支援する機能である。従来の分子モデリングでは、奥行き感があまり得られない 2 次元的な表示と平面的な動きしか表すことの出来ないマウスを用いることが多かったため、編集する構造の位置を直感的に決定することは困難であった。また数十原子以下の一般分子ではその必要性も低かった。数千原子からなる巨大で複雑なタンパク質の密構造を編集する場合には、構造の 3 次元空間的位置関係や力を直感的に把握し、最適な状態へと構造を変形させる操作の機能開発は有効である。

図 2 にタンパク質モデリング VR システムの構成を示す。本システムは反力デバイス PHANTOM による触感インターフェース、及び裸眼立体視ディスプレイによる視覚インターフェースを取り入れた。モデリングは反力デバイスを使用してタンパク質の異常構造を変異・修正させるが、その時の力を MD で計算し、

デバイスにフィードバックさせることによって触感を通じてタンパク質が作る場を認識することができる。また、分子グラフィックスの立体視表示とこの触感によって３次元構造を正確に認識できるため、直感的かつ精密に構造を修正することができる。なお、本趣旨によりタンパク質の力計算はリアルタイム性が重要となるため、MD シミュレーション専用計算機 MDGRAPE-3 の組み込みも予定している。

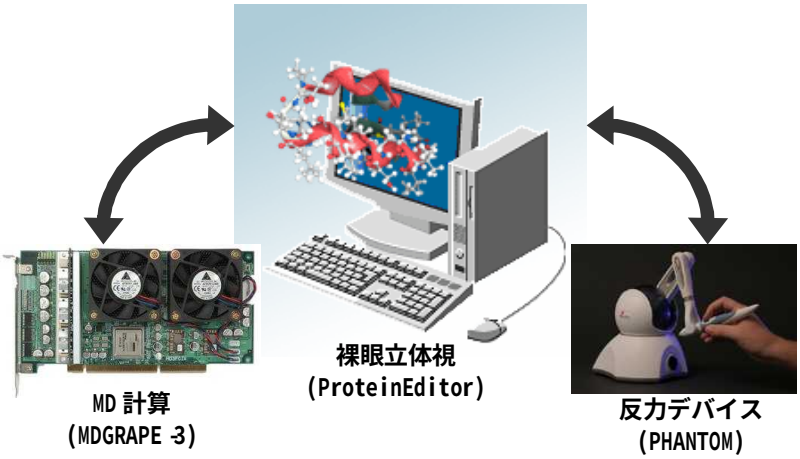

図 2 タンパク質モデリング VR システムの構成

従来、構造歪み除去は MD による構造最適化を行ってきたが、側鎖や官能基が自由に回転できない局所解に落ち込み、緩和できない構造歪みが残る問題があった。VR 技術を取り込んだ本 ProteinEditor を使用することによって、視覚・触覚に訴えた直観的な編集作業ができるため、このような構造から強制的に局所壁を乗り越えさせ、安全に適切な構造にピンセット修正できるものと考えられる。

タンパク質活性部位の認識： タンパク質の重要な反応の一つに酵素反応がある。タンパク質（酵素）の表面や内部にはポケットと呼ばれるくぼみや空洞（鍵穴）があり、作用を受ける基質分子（鍵）はポケットに結合して、適切な化学変化を受ける。一般に、この活性部位の立体構造は、鍵穴と鍵の関係のように特定の基質とぴったり合うようになっており、酵素は基質の立体構造を認識することができる分子といえる。このドッキングのシミュレーション（Virtual Screening）を行う上で、ポケットの認識は非常に重要なステップである。本研究のタンパク質モデリング VR システムを用いることにより、この酵素と基質の詳細なドッキング解析が可能となる。しかし、ポケットの構造は複雑な上にタンパク質内部に隠れた状態になっていることがあり、その場合タンパク質構造を粗視化するような単純な立体表示ではポケットと基質の結合状況、さらにはタンパク質立体構造との関係を正確に認識することが難しい。このような問題は、例えば没入型の CAVE システムを利用すれば、ポケット内部から全方向を見渡すことができるため、タンパク質活性部位のドッキング状況をより分かり易く認識できると考えられる。

本研究は、文部科学省 IT 基盤構築のための研究開発プログラム「革新的シミュレーションソフトウェアの研究開発」、および東京大学生産技術研究所の「選定研究」において実施された。

参考文献：

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

[2] 革新的シミュレーションソフトウェアの研究開発; http://www.rss21.iis.u-tokyo.ac.jp/.

[3] 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).

[4] 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** (2000) 235. http://www.rcsb.org/pdb/.

[5] 西村康幸, 吉廣保, 石川寛人, 佐藤文俊; 分子構造総合討論会, 4P067 (2006).