

特別 WEB
コラム新型コロナウイルスのタンパク質に対する
フラグメント分子軌道計算による解析事例

望月 祐志

立教大学

1. まえがき

新型コロナウイルスの感染が世界的な脅威となっており、対応策に関する研究開発が多様かつ精力的に行われています。こうした中、筆者らの研究グループではフラグメント分子軌道 (FMO: Fragment Molecular Orbital)^{1~3)} 計算を自主開発してきた ABINIT-MP プログラム⁴⁾ を使い、(国研) 理化学研究所計算科学研究センターが管理するスーパーコンピュータ「富岳」の上で行い、新型コロナウイルスの関連タンパク質を解析する試みを続けています⁵⁾。「富岳」は、スーパーコンピュータの世界ランキング⁶⁾ で、2020年6月時点で頂点に立っているのですが、そのうちのごく一部の資源を利用するだけでも、前世代機の「京」に比して100倍程度の計算力のポテンシャルを実感しています。このコラムでは、「富岳」によって可能となった大規模・高精度な FMO 計算による相互作用解析を、コロナウイルスの増殖に関与するメインプロテアーゼ、ならびに細胞への感染に関わるウイルス表面のスパイクタンパク質を例に紹介させていただきます。

2. FMO 法と ABINIT-MP プログラム

FMO 法は、北浦和夫氏によって1999年に創案¹⁾された、いわゆる分割&統治系のアプローチの1つで、並列処理を駆使してタンパク質のような巨大分子系を量子論的かつ非経験的なフレームワークで扱うことが可能です^{2~4)}。タンパク質の場合、FMO 計算からはフラグメントとして分割された各アミノ酸残基の間、あるいは活性領域にはまりこんだ薬物分子(リガンド)と周囲のアミノ酸残基の間の相互作用エネルギーなどの数値指標が得られます。言い換えると、高速の量子化学的手法としてメリットだけでなく、計算対象の解析ツールとしても使えるのが FMO 法の大きな強みで、生物物理学や創薬科学の分野で広く使われています^{2,4)}。なお、応用物理学分野における FMO 計算の適用事例については、筆者らの解説記事が機関誌の10月号⁷⁾に掲載予定ですので、参考にいただければ幸いです。

ここで、FMO 計算の概略をまとめます^{1~4)}。計算対象の分割によって生じた個々のフラグメントをモノマー、その2量体すなわち対をダイマーと呼びます(基本の2体展開の FMO)。最初の段階は、与えられた基底関数の下、基本的な近似であるハートリー・フォック (Hartree-Fock: HF)

計算によって各モノマーの分子軌道・電子密度を互いに静電ポテンシャル (ElectroStatic Potential: ESP) を自己無撞着的に課しつつ決定するモノマー SCC (Self-Consistent Charge) になります。これにより、各モノマーに対して分極が考慮されます。次の段階では、モノマーで決められた静電ポテンシャルを使ってダイマーでの HF 計算を行っていきます (SCC 条件は課しません)。ダイマーの計算によって、モノマー間の電子の非局在化が取り込まれます。こうして得られたモノマーとダイマーの HF エネルギーの和から対象系の全電子エネルギーを求めることができます。2次のメラー・プレセット摂動 (MP2) などの電子相関の計算は、モノマー SCC の完了後に各モノマーの HF 計算に続けて、各ダイマーについても個々の HF 計算の後に行い、エネルギー補正は加法的に加えます。電子相関の導入は、HF での過剰なイオン性の緩和や分散力の記述による定量性の向上に不可欠です。FMO 計算全体としては処理の高速化のため、図1に示すような階層的な並列化をはじめ、さまざまな技法が用いられています。ただ、計算対象のタンパク質が大きくなると、経過時間的に計算ノード数が豊富なスーパーコンピュータの利用が望ましくなります。

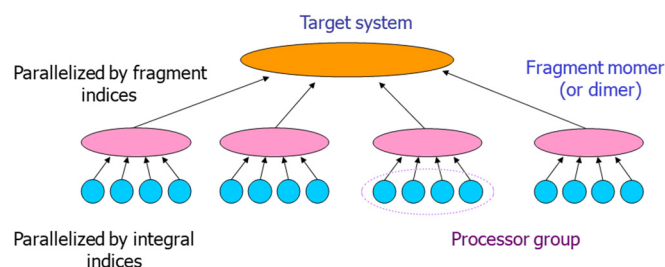


図1 FMO 計算における階層的な並列化の模式図。桃楕円が計算対象を分割したフラグメントのモノマー(ないしダイマー)のリストとなり、各々は青丸で示すプロセッサコア群で積分計算などが並列処理される。スーパーコンピュータでの実行の場合、桃楕円は計算ノードに対応すると考えてもよい。

FMO の電子エネルギーはフラグメント間の相互作用に対応するエネルギー (Inter-Fragment Interaction Energy: IFIE ないし Pair Interaction Energy: PIE と呼ばれます) を使った形^{3,4)}に書き換えられ、対象系の解析に供されます。最近では、相互作用エネルギーの内訳を静電項

や相関補正項などに分割してさらに詳しく解析する (PIE Decomposition Analysis: PIE DA) ことも増えてきています。

FMO 計算が可能な主なプログラムとしては, D. G. Fedorov 氏らによる GAMESS-US⁸⁾, 石川岳志氏による PAICS⁹⁾, そして筆者らの ABINIT-MP⁴⁾があります。ABINIT-MP の特徴は, フラグメント分割の自動化などの使い勝手のよさ, 結果の信頼性を高める電子相関計算の高速性, スレッド・プロセス (OpenMP/MPI) の混成並列化によるスーパーコンピュータへの高い適用性にあります¹⁰⁾。また, 専用のグラフィカルユーザーインターフェースである BioStation Viewer によって, IFIE などの計算結果の可視化が簡単に行えます。ABINIT-MP には 20 年ほどの歴史があり, 東京大学拠点の複数のプロジェクト (直近では, ポスト「京」重点課題 6) などの支援を得て開発が続けられてきました。

3. 新型コロナウイルスのタンパク質の解析

3.1 メインプロテアーゼ

筆者らは, 「富岳」の高い計算能力を2つの方向で使うことにしました。1つは, 同種計算の大量処理 (capacity computing) の文脈で, 生体内環境を想定した分子動力学 (Molecular Dynamics: MD) シミュレーションによって生成される構造揺らぎを含むタンパク質の構造セット多数をまとめて処理し, 相互作用エネルギー解析を統計的に行うことです。新型コロナウイルスに関しては, メインプロテアーゼと阻害剤 N3 との複合系の水和構造のセット 4000 個を, 電子相関の補正を MP2 で行う FMO-MP2/6-31G* レベルで計算しました。サンプル構造 1 個あたりの計算時間は「富岳」の半ラック (192 ノード) を使うと僅か 0.6 時間で, 多ノードを生かした同時並行処理によって全ての FMO 計算を約 5 時間で終わることができました。実は, メインプロテアーゼと N3 の複合体の結晶構造に関する FMO 計算は退任直前の名古屋大学の FX100 で行っていた¹¹⁾ のですが, 構造揺らぎを含めると水素結合などの相互作用に関わるアミノ酸残基の寄与に顕著な差が出てくることが明らかになりました。図 2 は, N3 の中でも最も重要な部位と周囲の残基との位置関係, それに個々の相互作用エネルギー (IFIE) を示したもので, 左側が結晶構造¹¹⁾での結果, 右側が MD サンプル構造での統計的な結果を示します。生体内環境 (水和, 常温) を想定した MD によって動的な構造揺らぎを含めることで, 相互作用描像をリアルに捉えられるようになりました¹²⁾。こうした動的な知見は, メインプロテアーゼの阻害剤の分子設計に関する有意義な情報であると考えています。FMO 創薬コンソーシアム¹³⁾の「京」を用いた研究活動の中で, タンパク質-リガンド複合体の (結晶由来の) 単一構造での FMO 解析をルーチン的に行うことはできていましたが, 数百個単位の構造をベースにする統計的な評価は, 計算コスト的に限定的でした。新型コロナウイルスのタンパク質に限らず, 「富岳」の時代には (生体内環境での) 構造揺らぎを含めた相互作用の統計的解析が一般化することは間違いないと思われ, よりリアルな薬の設計・評価が進むと期待されます。

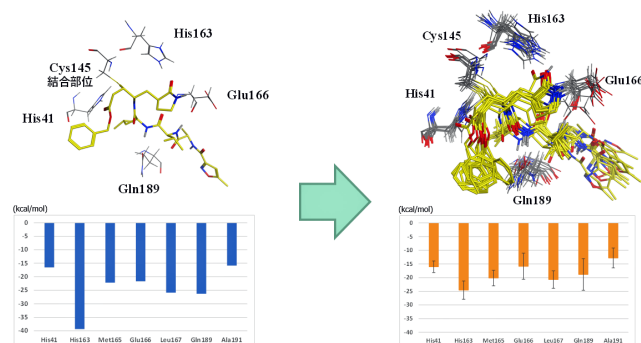


図 2 メインプロテアーゼと阻害剤 N3 の重要部位 (PDB ID: 6LU7) との相互作用。左は結晶構造での結果, 右は MD による構造揺らぎを含めた結果。

3.2 スパイクタンパク質

「富岳」の計算力を生かすもう 1 つの軸線は, 大規模で複雑な問題を扱う処理 (capability computing) です。FMO のフレームワーク内で, 電子相関の摂動補正レベルを 2 次 (MP2) から 3 次 (MP3) とすると, テンソル量の縮約計算の複雑さが増してコストが上がり, さらにメモリ要求も大きくなります¹⁴⁾。もちろん, MP3 よりも摂動の次数を上げた (それゆえコストもさらに大きい) 結合クラスター展開 (CCSD(T) が最も信頼性が高いとされます) が望ましいのですが, MP3 に簡便なスケールリングを施すと実効的に CCSD(T) 並の相互作用エネルギーの算定値が得られることが知られています (MP2.5 法)。したがって, FMO-MP3 計算が行えることは計算結果の信頼性を高めるうえで大きな意味があります。残念ながら, 「京」では計算ノードあたりのコア数が少なく, また利用可能なメモリが限られていたために不可能でした。しかし, 「富岳」ではノードの計算資源が大幅に向上したため, FMO-MP3¹⁴⁾ が実用的に使えるようになります。この立場から, 1.1 千個のアミノ酸残基の鎖が 3 つ結合して形成される巨大なスパイクタンパク質を MP3 レベルで計算しました。基底関数を 6-31G* よりも記述性が高い cc-pVDZ とし, 8 ラック (3072 ノード) でジョブを流したところ, 閉じた (closed) 状態の構造 (PDB ID: 6VXX) では 3.4 時間, 感染に関わるとされる開いた (open) 状態の構造 (同 6VYB) では 3.9 時間で完了させることができました。これは驚異的な速さで, 現時点で世界最大の 2 次摂動相関を超えた (beyond MP2) レベルの FMO 計算になっています (実効的に CCSD(T) の信頼性を期待)。図 3 の左側は, 閉じた構造と開いた構造で各鎖のレセプター結合領域 (RBD: Receptor Binding Domain) の安定化相互作用エネルギー (スケールリングした MP2.5 で評価した IFIE の和) を図示したもので, B 鎖の RBD では開いた構造で安定化が大きく減少していることが明確に示されています¹²⁾。図 3 の右側には, B 鎖の RBD の位置を示しますが, 浮き上がった状態で細胞表面のアンジオテンシン転換酵素 2 (ACE2) との結合が生じやすい形となっています。IFIE の和の取り方を変えてみても, B 鎖-RBD の相対的重要性が確認されました。別途, RBD 部分と ACE2 の結合モデル (PDB ID: 6MOJ) を計算してみると, 安定化エネルギーが上述の

減損分をほぼ補填する量となっていましたので、ACE2 との結合は妥当と思われます（抗体との結合モデルも計算中）。スパイクタンパク質の詳細な解析は特異値分解¹⁵⁾なども駆使して進行中ですが、鎖間の相互作用には荷電性のアミノ酸残基群が重要な役割を担っていることがわかってきています。荷電性残基間の相互作用は静電的ですので、構造変化による安定化（不安定化）の差分エネルギーが大きくなると考えられます。2020年の秋期学術講演会（オンライン開催）でも、これらの解析結果を報告させていただくことになっています。いずれにせよ、FM0計算から得られるこうした知見は、スパイクタンパク質の阻害剤の選別・開発や抗体のデザインにも貢献できると思われます。

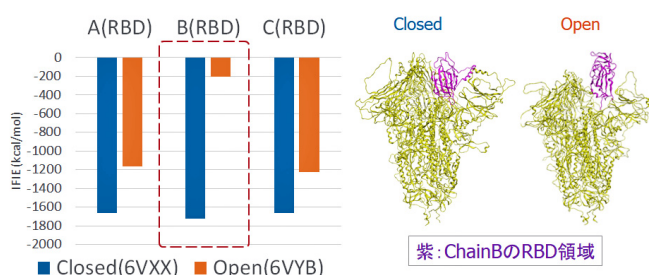


図3 左はスパイクタンパク質のクローズ型（PDB ID: 6VXX）とオープン型（PDB ID: 6VYB）のレセプター結合領域（RBD）の安定化相互作用エネルギーの比較。右は感染に関与するとされるチェーンBのRBD位置の比較。

4. まとめ

本コラムでは、最新鋭のスーパーコンピュータ「富岳」を使ったFM0計算による新型コロナウイルスの関連タンパク質の解析事例をご紹介しました。研究は現在も進行中で、メインプロテアーゼとネルフィナビルとの複合系や構造緩和を含めたスパイクタンパク質の計算も行われつつあります。解析結果を論文化して公知とするとともに、元のFM0計算データのデータベースサイト¹⁶⁾での公開も進めていきたいと考えています。

謝辞

このコラムで記した成果は、文部科学省と（国研）理化学研究所が連携して実施している「新型コロナウイルス対策を目的としたスーパーコンピュータ「富岳」の優先的な試行的利用」⁵⁾の第4課題「新型コロナウイルス関連タンパク質に対するフラグメント分子軌道計算」の研究で得られたもので¹²⁾、田中成典（神戸大学）、福澤薫（星薬科大学）の両先生をはじめ、以下の皆様（敬称略）とのコラボレーションで進められています：奥脇弘次、畑田峻、秋澤和輝（以上、立教大学）、川嶋裕介、半田佑磨（以上、星薬科大学）、古明地勇人（（国研）産業技術総合研究所）、坂倉耕太（（公財）計算科学振興財団）、渡邊啓正（HPCシステムズ（株））。なお、立教大学学術推進特別重点資金（立教SFR）からも活動支援を受けました。

文献

- 1) K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi: Chem. Phys. Lett. **313**, 701 (1999).
- 2) *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, ed. D.G. Fedorov and K. Kitaura (CRC Press, Boca Raton, FL, 2009).
- 3) D.G. Fedorov, T. Nagata, and K. Kitaura: Phys. Chem. Chem. Phys. **14**, 7562 (2012).
- 4) S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, and K. Fukuzawa: Phys. Chem. Chem. Phys. **16**, 10310 (2014).
- 5) <https://www.r-ccs.riken.jp/post-k/corona/>
- 6) <https://top500.org>
- 7) 望月祐志, 奥脇弘次: 応用物理 (2020年10月号, 掲載予定).
- 8) D.G. Fedorov: WIREs Comput. Mol. Sci. e1322 (2017).
- 9) T. Ishikawa, T. Ishikura, and K. Kuwata: J. Comp. Chem. **30**, 2594 (2009).
- 10) http://www.cenav.org/abinit-mp-open_ver-1-rev-22/
- 11) R. Hatada, K. Okuwaki, Y. Mochizuki, Y. Handa, K. Fukuzawa, Y. Komeiji, Y. Okiyama, and S. Tanaka: J. Chem. Inform. Model. **60**, 3593 (2020).
- 12) <https://www.r-ccs.riken.jp/outreach/formedia/200617.html>
- 13) <https://fmodd.jp>
- 14) Y. Mochizuki, K. Yamashita, K. Fukuzawa, K. Takematsu, H. Watanabe, N. Taguchi, Y. Okiyama, M. Tsuboi, T. Nakano, and S. Tanaka: Chem. Phys. Lett. **493**, 342 (2010).
- 15) S. Tanaka, C. Watanabe, T. Honma, K. Fukuzawa, K. Ohishi, and T. Maruyama: J. Mol. Graph. Model. **100**, 107650 (2020).
- 16) <https://drugdesign.riken.jp/FMODB/>

脱稿日 2020年7月18日



望月 祐志 (もちづき ゆうじ)

1990年北海道大学理学研究科博士後期課程修了。理学博士。日本電気株式会社、日本原子力研究所付随研究員、東京大学生産技術研究所プロジェクト研究員などを経て、06年立教大学助教授、10年同大教授。専門分野は量子化学・計算化学、ABINIT-MPプログラムの取りまとめ役。