

共同利用(産業利用トライアルユース:先端研究施設共用促進事業『みんなのスパコン』TSUBAMEによるペタスケールへの飛翔) 成果報告書 平成 19 年度 戦略分野利用推進「計算化学手法による創薬技術の開発」 i07qa

利用課題名 巨大生体分子の非経験的分子軌道法による設計指針構築
英文:Molecular Design of Large Biomolecules by ab-initio MO Method

中村 振一郎
Shinichiro Nakamura

株式会社三菱化学科学技術研究センター
Computational Science Laboratory, R&D Division, Mitsubishi Chemical Corp.

邦文抄録 FMO 法を利用し、量子化学計算による生体高分子 NMR 化学シフトを予測する手法を開発した。フラグメント分割により生じるゲージ依存の問題を解消するため GIAO 法、及び CSGT 法との組み合わせを適切に行うプログラムを開発した。10 残基ペプチド、32 残基 α ヘリックス、 β シートとユビキチンたんぱく質に対して計算を行い、生体高分子 NMR 化学シフトを高精度に計算できることを示した。さらに計算精度を高めるためにカットオフ長を導入し、化学シフト計算に適したフラグメント分割をおこなう Cutoff-FMO-NMR 法を開発した。

英文抄録 The new method of quantum chemistry is developed for evaluating NMR chemical shift, being based on FMO method. An introduction of cut off enabled to provide reasonable accuracy for the shift of large bio-molecules.

Keywords: FMO, NMR ケミカルシフト、カットオフ

背景と目的

本プロジェクトでは、(目的と解決手段と、成果を述べてください)を XX によって解決し YY の成果を得た。

NMR(核磁気共鳴)は原子レベルでの物質の分析、同定手段として幅広い研究分野において使用されている。なかでも化学シフトは重要である。それは化学シフトから個々の原子が存在する周囲の環境が特定できるためである。例えば、化学シフトの情報から、原子の結合状態や原子と原子間距離を特定することができる。水素結合などの生体高分子の構造を決めていく上で重要な情報にも敏感に応答する。こうした特徴を生かして、医薬開発においては、標的生体高分子と薬の分子の相互作用をあきらかにすることで、機能性が高い薬の開発に役立てつつある。

一方で、NMR の観測データから分子構造やその動的な情報を得る際には理論計算が必要になってくる。量子化学による化学シフトの計算法、特に第一原理計算法は NMR 計算法のなかでは最も信頼性が高い方法である。しかし、第一原理計算には大きな計算コストをかけるために、これまでに取り扱われてきた系は主に

気相中にある小さな有機分子であった。従って、生体中(特に水中)における大きな生体分子の構造について

の情報を得ようとする場合、溶媒や周囲の生体分子からの効果なども含めた第一原理計算法による化学シフトの解析が要求されることになり、その構造決定への応用には様々な困難が存在する。

本プロジェクトでは生体高分子の化学シフトを高精度で予測することを目的とする。そのため第一原理計算に基づいた新たな NMR 計算法を開発した¹⁻³。その新手法とは、第一原理計算法のもつ高い計算精度を保ちつつ、低い計算コストで化学シフトを計算するものである。以下に簡潔にその成果をまとめた。

- 1) FMO というフラグメント分割に基づく量子化学計算法を活用し、かつゲージ依存の問題を適切に解消することによって、第一原理による化学シフトの計算を実現した。フラグメント分割により計算コストの問題は解消され、これまで大きすぎて計算できなかった分子に対しても、短時間で化学シフトの計算を行うことに成功した¹。
- 2) カットオフ長(L_{cutoff})を導入することによって、より高精度に化学シフトを計算することに成功した³。
- 3) 水分子を 2) で開発した Cutoff-FMO-

図 4: 水分子を発生したユビキチンたんぱく質の構造(12671 原子)

NMR 計算手法に取り込むことによって、たんぱく質の周りの水分子による化学シフトのずれの問題を解決した。

概要

巨大生体分子であるたんぱく質と薬の分子の間の相互作用に伴う化学シフトの変化を精度よく計算するため、第一原理計算手法に基づいた手法構築、およびその精度の検証を行う。最終的には開発した手法が産業利用することができるようにするための研究開発を行う。

研究開発に当たって、計算手法の開発、ソフトの実装、計算資源(ハードウェア)の充実が必要になる。詳細な過程を以下に述べる

1)我々はまず大きな生体高分子を小さなフラグメントに分割して、FMO(フラグメント分子軌道)を用いた量子化学計算を行い、化学シフト予測する手法を開発した。FMO 計算の際には、一つのフラグメントに対して、一つの CPU を割り振るように実装した。FMO 計算には GAMESS を、化学シフトの計算には Gaussian03 をというように、二つの量子化学計算ソフトを利用して、大規模な並列計算システムを作り上げた(図1参照)。最終的に2つのフラグメント(dimer)を一つの単位とする計算を GAMESS でおこなわせ、その結果を GAUSSIAN に受け渡すように実装した¹。(以下、これを FMO-NMR dimer と呼ぶ。)

この計算を実行するために、大量かつ高速に演算処理できる計算機が必要となる。東工大 TSUBAME 計算機システムはまさにそのような用途に適したシステムであった。開発した手法を TSUBAME 上で使うことにより、従来第一原理計算により化学シフトを計算してきた系に比べ、数百倍大きいサイズの系に対して短時間で大規模第一原理計算ができるようになった。

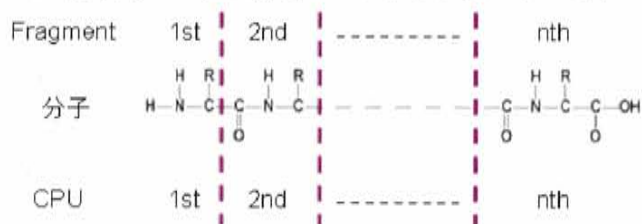


図 1: FMO による高分子量子化学並列計算の模式図

2)次に、1)で開発した手法の計算精度を高めるために、フラグメントの切り分け方を変更した。具体的にはカットオフ長 L_{cutoff} を導入し、新たな NMR 計算用のフラグメント分割法 (Cutoff-FMO-NMR 法) として提案した(図 2 参照)³。その分割法は1)で述べた単純に生体高分子の一次元配列によるフラグメントの分割方と違って、生体高分子の三次元情報も含まれているので、より正確に原子周辺化学環境の記述ができた。

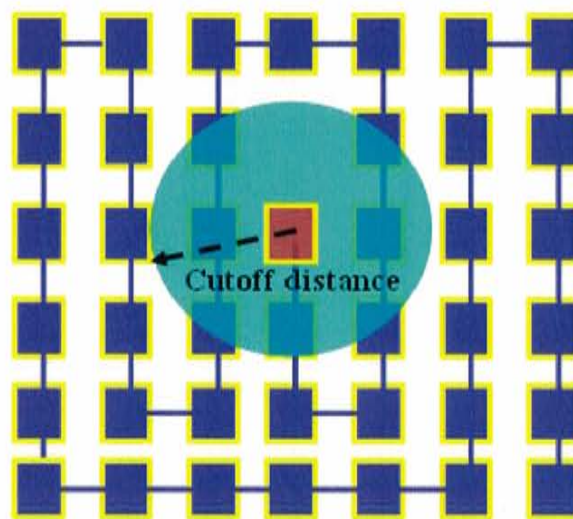


図 2: Cutoff-FMO-NMR 法の概念図: L_{cutoff} で囲んでいる範囲(緑色の円)を新しいフラグメントとする。

一方で、ハードウェアの面において、Cutoff-FMO-NMR の計算を実行するために、大きなメモリ装備(数十 GB)が要求される。その点においても、東工大 TSUBAME 演算サーバで搭載されている 32GB、64GB、128GB メモリを持つ三種類ノードのお陰で、解決された。

現在、我々は Cutoff-FMO-NMR を利用して、生体高分子の様々な構造について溶媒効果などの影響を取り入れながら、NMR 化学シフトの計算を行っている。その産業利用には、更なるソフトの改善とより強力なハードウェアが必要になるだろう。このハード、及びソフトの両面における進展があれば、計算科学は実験をサポートし解釈を助けるとともに、実験的入手が不可能な情報(薬物とたんぱく質の相互作用にともなう遷移状態の構造など)を提供することができる。計算精度の

向上によりライフサイエンスの根源的な中心課題である巨大生体高分子の計算例を蓄積し、大規模計算によって得られた知見をデータベースとして公開したうえで、その知見の共有を図れば、創薬や診断予防技術の飛躍的發展が期待される。

結果および考察

我々は計算手法を検証するために、10 残基ペプチド、及び、32 残基の α ヘリックスと β シートのそれぞれの構造に対して、NMR 化学シフトを計算し、考察を行った。ここでは、32 残基の β シートを例として議論する。他の構造についても同様な結果が得られた。

検証で使われた 32 残基 β シートは Green Fluorescent Protein(GFP) たんぱく質 (PDBID : 1Q4B) の 198~299 番目の配列である。この構造に対してまず AMBER により水素を付加し、その後 MOPAC2007 に組み込まれている MOZYME を使って構造最適化した(図 3)。

最適化した構造に対して2種類のフラグメント分割が行われた。1)一次元配列の順番に 2 残基を 1 フラグメントとする FMO-NMR dimer 計算。2) $L_{\text{cutoff}}=6\text{\AA}, 8\text{\AA}, 10\text{\AA}$ で三次元空間上のフラグメントを決定する。それぞれのフラグメントに対して NMR 化学シフトの計算を行った。これに対し、周囲のフラグメントからの影響を考慮するために、モデル I(他のフラグメントからの影響を点電荷の分布によるものとして近似する)とモデル II(他のフラグメントからの影響のうち静電ポテンシャルの影響を取り入れる)の二つの計算モデルを構築した。最終的にそれぞれのモデルでの計算結果と従来の量子化学の計算結果と比較した。なお NMR のゲージに依存しない計算手法として定評がある GIAO 法と、それとは別に CSGT 法を使った計算手法を開発したが、ここでは CSGT 法での結果に限定して議論する。GIAO 法に関して CSGT 法と同様な結果が得られた。ここで示した結果は HF/6-31G(d)での計算によるものである。(HF/6-311G(d,p)に関しても同様な結果が得られた)

1) FMO-NMR dimer の計算結果について:

1.a) モデル I では $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{15}N と ^1H 原子において、化学シフトの isotropic 値は従来の量子化学計算結果とよく一致した(90%以上)。 ^{15}N と ^1H は ^{13}C より大きなエラーを示した。その原因は ^{15}N と

^1H は ^{13}C より周囲の環境の影響に敏感であるためである。さらに、anisotropic 値 ($\Delta\sigma = \sigma_3 - (\sigma_1 + \sigma_2)/2$) は isotropic 値より大きなエラーを示したことから、これらの原子近傍における電子状態が完全には再現できていない²⁾。

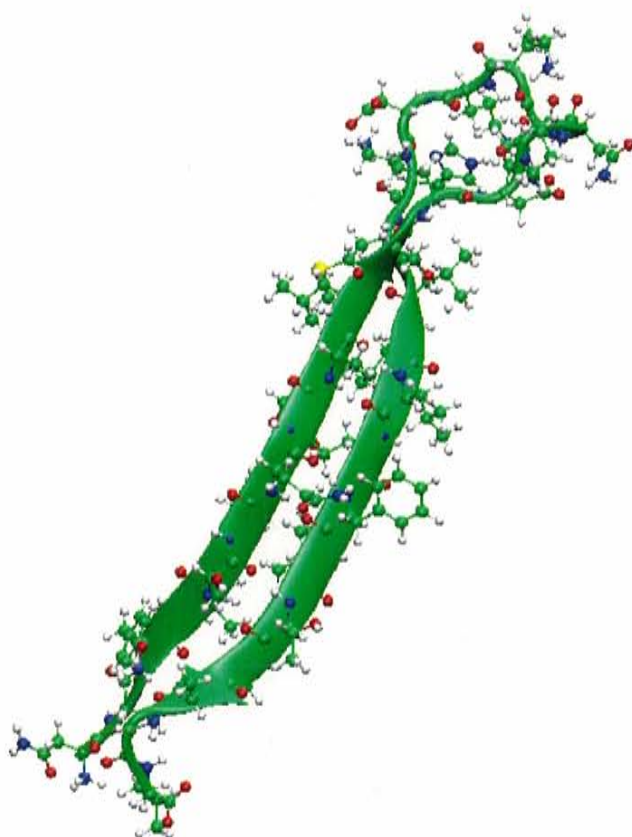


図 3: 検証のために使った β シートの構造

1.b) モデル II では $^{13}\text{C}_\alpha$, ^{15}N と ^1H のエラーは isotropic と anisotropic 値両方において、モデル I より小さいことがわかった²⁾。 $^{13}\text{C}_\beta$ の場合はややモデル I のほうがエラーが小さかった。それらの結果から、静電ポテンシャルを概ね正確に再現してるモデル II は点電荷近似を採用しているモデル I より高い精度で化学シフトを再現できていることがわかった。一方で、時折モデル I のほうが計算結果がよい場合もあったが、その原因はフラグメントの分割が適切ではないため、そのことにもなるエラーが大きすぎることに
よるものと考えられる。

2) $L_{\text{cutoff}}=6\text{\AA}, 8\text{\AA}, 10\text{\AA}$ に対応する Cutoff-FMO-NMR の計算結果について³⁾:

2.a) モデル I では $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{15}N と ^1H 原子に

において、isotropic 値と anisotropic 値ともに 1) で示した結果より大きく改善された。とくに $L_{\text{cutoff}}=8\text{\AA}$ のときに、とても大きな改善が見られた。計算コストと精度のバランスを考えると、 $L_{\text{cutoff}}=8\text{\AA}$ を使うことが適切であると結論付けられた。またテストした原子種別に解析してみると、 $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ と ^1H の改善度は ^{15}N より数倍程度大きいことがわかった。 ^{15}N はほかの原子種と比べ、より周囲の環境に敏感で、電子状態変化しやすいという特徴がある。従って、 ^{15}N の改善度が低い原因は点電荷近似の問題点が現れていると考えられる。

2.b) モデル II では isotropic 値と anisotropic 値ともにすべての原子種類に対して、Cutoff-FMO-NMR モデル I の結果よりさらに再現性がよい結果が得られた。とくに isotropic 値の改善が目立つ。またモデル I で計算したときに相対的に改善度が低い ^{15}N 原子に対しても、ほかの原子種とほぼ同じオーダーの改善度が得られた。従って、点電荷近似をおこなわなかったことにより、静電ポテンシャルがよりよく再現され、非常に高い精度で化学シフトを予測することに成功したと考えられる。

現在我々は Cutoff-FMO-NMR の手法を使って、

たんぱく質の NMR 化学シフトの計算を行っている。

一例として、ユビキチンたんぱく(図 4) 周辺に水分子を発生させた巨大系(原子数 12671)に対して NMR 化学シフトの計算を行った場合について示す。水分子を計算に取り込むことによって、真空中での計算とくらべ、より正確に化学シフトの実験値が再現された。

最後に計算コストについて表 1 にまとめた。32 残基の場合、積算で 128 CPU を使いかつ必要なメモリを割り当てることにより、FMO-NMR dimer, $L_{\text{cutoff}}=6\text{\AA}$, $8\text{\AA}$, $10\text{\AA}$ の計算時間は従来の第一原理計算法によるものに比べ約 35.5%, 51.6%, 60.2%, 88.4% に短縮された。さらに従来の第一原理計算法によると、化学シフト計算の計算コストが分子サイズ大きさの 6 乗に比例するため、巨大たんぱく質分子の計算は非現実になる。それに対して、Cutoff-FMO-NMR 法では大きな分子でも、大きな計算資源を利用すれば、短い時間で計算できる。例えば、304 CPU でユビキチンを計算する場合 $L_{\text{cutoff}}=8\text{\AA}$ を使うと、32 残基の計算よりわずか 1 日しか計算時間が増えていないことがわかる。

まとめ、今後の課題

我々は FMO 法を利用し、量子化学計算による生体高分子 NMR 化学シフトを予測する手法を開発した。このときフラグメント分割により生じるゲージ依存の問題を解消するため GIAO 法、及び CSGT 法との組み合わせを適切に行うためのプログラムを開発した。周囲のフラグメントの影響を取り入れるために、我々は 2 つのモデルを提案した。この 2 つの計算モデルを用いて、東工大 TSUBAME スーパーコンピューティングシステムのサポートにより 10 残基ペプチド、32 残基 α ヘリックス、 β シートとユビキチンたんぱく質に対して計算を行い、短い時間で生体高分子 NMR 化学シフトを高精度に計算できることを示すことができた。

さらに計算精度を高めるためにカットオフ長を導入し、化学シフト計算に適したフラグメント分割をおこなう Cutoff-FMO-NMR 法を開発した。その方法を用いて、同様に TSUMABE 演算サーバで生体分子 (32 残基 α -helix と β -sheet) の NMR 化

表 1: 計算法による計算コストの比較

計算法	全利用 CPU 数	最大必要 メモリ	全計算 時間 (day)
32 残基 β -sheet (501 原子)			
第一原理計算	8	24GB/8CPU	6.53
FMO-NMR dimer	128	2GB/4CPU	2.32
Cutoff-FMO-NMR $L_{\text{cutoff}}=6\text{\AA}$	128	6GB/4CPU	3.37
Cutoff-FMO-NMR $L_{\text{cutoff}}=8\text{\AA}$	128	6GB/4CPU	3.93
Cutoff-FMO-NMR $L_{\text{cutoff}}=10\text{\AA}$	128	6GB/4CPU	5.77
ユビキチン+水分子 (12671 原子)			
Cutoff-FMO-NMR $L_{\text{cutoff}}=8\text{\AA}$	304	8GB/4CPU	4.93

学シフトの計算を行った。そこで、我々は計算精度のカットオフ長依存性を調べるために、異なるカットオフ長 (6 $\text{\AA}$, 8 $\text{\AA}$, 10 $\text{\AA}$) に対応するフラグメントを生成し、計算及び考察を行った。また計算精度と基

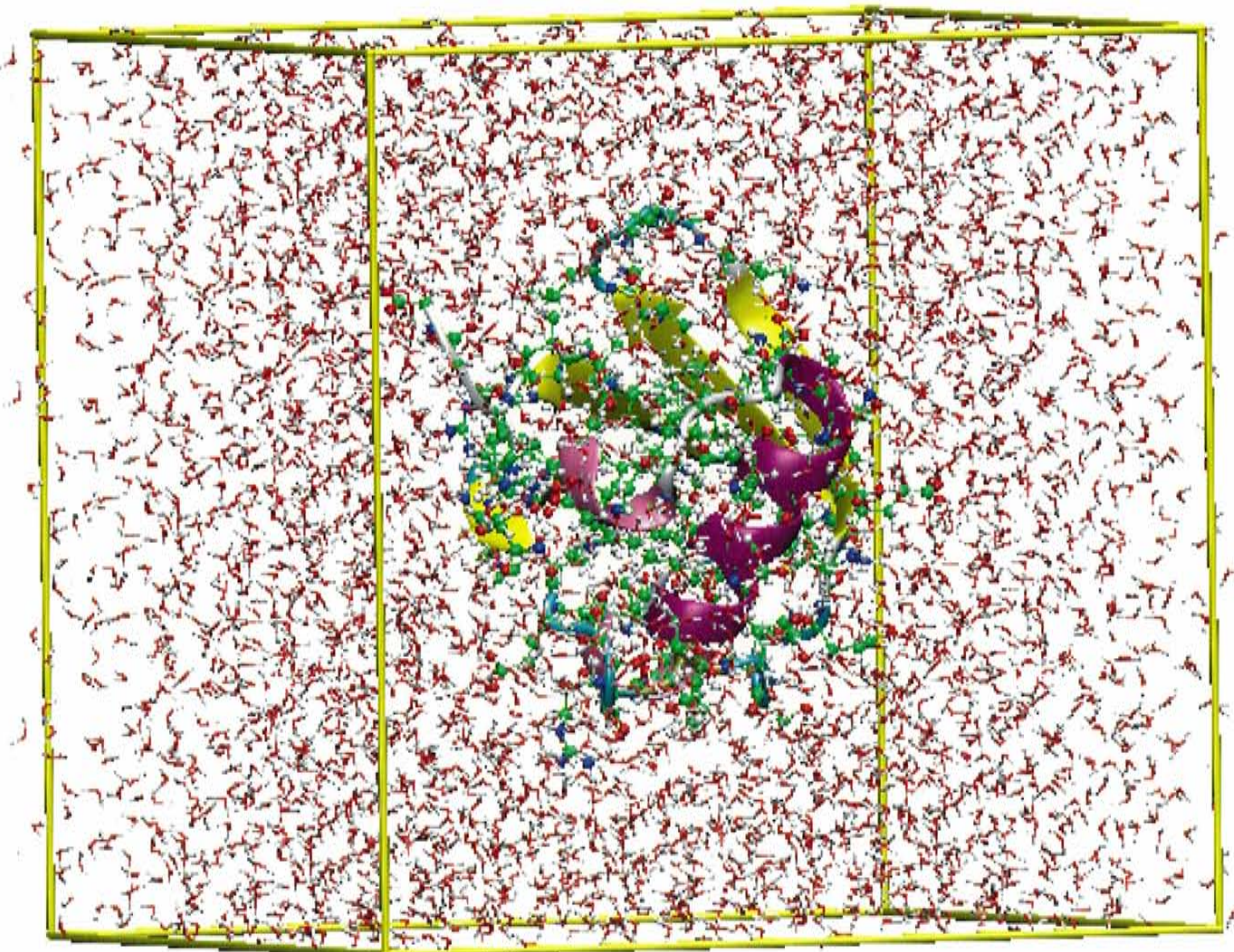


図 4: 水分子を発生したユビキチンたんぱく質の構造 (12671 原子)

底関数 (6-31G(d), 6-311G(d,p)) の関係についても考察した。それらの考察から $L_{\text{cutoff}} \geq 8 \text{ \AA}$ の場合において、Cutoff-FMO-NMR モデル II がほぼ完全に従来の第一原理計算の結果を再現していることを見出した。

現在我々は Cutoff-FMO-NMR 法を利用して、ユビキチンたんぱく質の化学シフトの計算を行っている。溶媒効果も含めたこれらのアンサンブル計算と実験結果との比較により、更なる化学シフトの高精度計算のために必要な他の要素についての知見を蓄えつつある。

これからの課題は、基礎と応用の両面が挙げられよう。基礎面においては電子相関も含めた電子状態を解く方法(例: DFT, Post-HF)での検証が必要である。また Cutoff-FMO-NMR 法で決められたフラグメントはそれなりに大きなサイズになるため、比較的大きな計算コストが必要となる。この問題にたいして、手法を改善し、いまよりさらに一桁程度速い高精度な計算手法を開発していきたい。

一方、応用面においては、引き続き Cutoff-FMO-NMR 法を用いて、様々なたんぱく質に対して、化学シフトの計算を行う。それらの結果からたんぱく質の構造やダイナミクスに関する情報を解析する。さらに、構造ゆらぎ、温度効果に対しても考察を行う。それらの経験を蓄積し、系統的な解析手法を構築していく。その知見は、標的生体分子と薬との相互作用の解析に応用し、現実的な医薬品及び予防診断薬の設計に乗り出すための大きな助けとなる。現実的になればなるほど、構造不明瞭というジレンマが存在するが、それについては既にMD等既存の方法があり、このジレンマを克服は不可能ではない。むしろ今回のわれわれの手法は、手がかりが希薄で、夢であった。

参考文献

¹Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, *Chem. Phys. Lett.* **445**, 331-339 (2007).

²S. Yokojima, Q. Gao, and S. Nakamura, 'Ab Initio NMR Chemical Shift Calculations Using

Fragment Molecular Orbitals', page 164-167, in AIP Conference Proceedings 1102 'Theory and Applications of Computational Chemistry - 2008', ed. by D.-Q. Wei (AIP, 2009).

³Q. Gao, S. Yokojima, and S. Nakamura, in preparation.