

科学研究費助成事業 研究成果報告書

令和 6 年 5 月 31 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2020～2023

課題番号：20K06503

研究課題名(和文) 水素・プロトンまでを含む蛋白質原子モデルの量子化学計算による構築法の開発

研究課題名(英文) Development of a method to build a protein atom model including hydrogen and proton by quantum chemical calculations

研究代表者

恒川 直樹 (Tsunekawa, Naoki)

東京大学・定量生命科学研究所・特任研究員

研究者番号：90638800

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：蛋白質分子のX線結晶構造解析では一般に水素・プロトンは観測されず、それらの位置は決定できない。そこで、観測された非水素原子の位置から推測される水素結合ネットワークモデルを網羅的に生成し、それぞれのモデルに対する量子化学計算によるエネルギーの比較を行うことで、より適切と思われる水素結合ネットワークを絞り込む方法を開発した。

この手法の戦略は正攻法とも言えるが、膨大なモデルに対する量子化学計算が計算リソースを非常に必要とするため、実効性の向上のために、必要な計算精度や効率的な計算の段取りを明らかにすることが重要となる。カルボキシル基が集中する系に適用し、手法の向上化と限界を探った。

研究成果の学術的意義や社会的意義

蛋白質分子構造の結晶解析において、その分子内にある水素結合ネットワークの情報は重要な判断材料となる。そのため、経験的スコア関数を用いた水素原子位置やプロトンの有無を推測するツールは多くある。しかし、それらの信頼性は限定的である。そこで、本手法では水素結合相互作用を適切に再現する量子化学計算によるエネルギーをスコア関数として採用することで、適切な水素・プロトン予測を実現する。そして、この手法は並列計算による高速化が容易である。いずれ構造解析における簡易的なツールとなると考えられる。また、この手法で得られた水素結合ネットワークとエネルギーのデータベースは、新手法の基盤となることが予想される。

研究成果の概要(英文)：In X-ray crystallography of protein molecules, hydrogen atoms and protons are generally not observed, and their positions cannot be determined. Therefore, a method was developed to find candidates from exhaustively generated hydrogen-bonding network models with positions of observed non-hydrogen atoms adopted by comparing their energies using quantum chemical calculations.

The strategy of this method is a straightforward one, but since quantum chemical calculations for a large number of models are computationally very resource-intensive, it is important to clarify the necessary computational accuracy and efficient computational procedures in order to improve the effectiveness of the method. We applied the method to systems including several carboxyl groups and explored the method's improvements and limitations.

研究分野：理論物理学

キーワード：量子化学計算 水素結合 蛋白質 結晶構造解析 密度汎関数法 カルシウムポンプ ナトリウムポンプ 大規模探索

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

水素結合相互作用は、蛋白質分子の構造や構造変化を決定づけている相互作用の一つである。よって、X 線結晶解析などで原子モデルを決定する時に水素結合の有無を考慮することは多々ある。また、構造を変化させて機能する蛋白質分子の場合、水素結合の生成や消滅の可能性を考慮しながら構造遷移を推測する。

一般に、水素結合は電気陰性度が高い原子(陰性原子)間に水素原子が介在する静電的な引力相互作用と理解されているが、多くの陰性原子を含む蛋白質内では水素結合の位置は一意に決定できない。多数ある水素結合は生体分子内でネットワークを形成し、各々の水素結合は周辺から大きな影響を受ける。また、反応過程においては比較的弱い水素結合一つでさえも重要な役割を果たすことがあり、一般にどの水素結合も無視できない。従って、水素結合の同定は避けられない課題である。

しかしながら、膜蛋白質分子のような大きな蛋白質の X 線結晶解析では水素原子の同定は困難である。また、水素原子が見えないが故に、Gln や Asn および His の側鎖の向きも間違えて決定してしまうこともある。よって、理論的に水素原子の位置を決定する必然性がある。

水素原子やプロトンが欠落している非水素原子のみの原子モデルに水素原子を付加するツールは既にいくつかある。また、荷電アミノ酸のプロトン化の有無を占うツールもある。これら中には、経験的なスコア関数を採用して、最適と判断される水素原子の位置を生成してくれる。しかし、その信頼性は限定的である。複数の水素結合がお互いに影響し合う場合があり、分子動力学法などで採用されている一般的な分子力場も含めて、既存の経験的スコア関数はその様な場合に対応できない。個々の局所的な水素結合の適切な再現の精度だけでなく、複数の水素結合の再現、つまり大域的な水素結合ネットワークの適切な再現(エネルギー値)が見込める評価関数(力場)での探索が必要となる。

量子化学計算は様々な低分子の構造を高精度に再現してきた。しかし、膨大な計算リソースを必要とするため、複数の水素結合を含む蛋白質ような系の計算は非現実的であった。しかし、量子化学計算の一種である密度汎関数法(DFT)は 300 残基もある系でも計算実行例があり、想定している水素結合ネットワークを含む蛋白質分子系(50 残基ほどまで)ならば、実行可能であると思われる。つまり、水素結合ネットワークの DFT によるエネルギー比較は可能であると判断できる。

水素結合ネットワークのエネルギー比較を実行し、適切なモデルを導き出すには、可能性ある水素結合ネットワークモデルすべてについて DFT 計算を実行し、エネルギー比較を行う必要がある。水素結合は水素原子を供給するドナー原子と水素原子を受け入れるアクセプター原子で構成されるが、例えば、計算対象系内に 10 個のドナー原子があった場合、それぞれのドナー原子からの水素原子の方向を 20 パターンあるとして水素結合ネットワークモデルを生成すると、その数は 20^{10} もの膨大な数になってしまい、実質計算不可能となる。ただし、各ドナー原子からの水素原子の方向を数パターンまで限定したり、1 パターンに確定してしまえば、その数は一気に減少させることが可能である。

蛋白質系の水素結合は陰性原子間にあると想定されるので、ドナー原子からの水素原子の方向は、その近傍の陰性原子への方向に限定され、数パターンに抑えることが可能であると思われる。その結果、DFT 計算対象となる水素結合ネットワークモデルは多くて数千ほどの数まで抑え込めると考えられた。

2. 研究の目的

水素原子が欠落している X 線結晶解析で得られた原子モデルに対して、適切な水素結合ネットワークの探索、つまり、水素・プロトンの位置の決定をする方法の開発、およびそのツールの作成を目的とした。この目的のために、3つのサブ目標を立てた。まず最初に、DFT 計算の対象となる水素結合ネットワークモデルの網羅的生成方法の確立。第2に、DFT 計算における段階的な計算レベルの確定。最後に、最低エネルギーとして選ばれた水素結合ネットワーク候補の正当性の確認、である。

3. 研究の方法

(1) 網羅的水素結合ネットワークモデルの生成は次のような段階を踏んでいく。与えられた非水素原子の原子モデルに対してドナー原子に成りうる原子とアクセプター原子に成りうる原子を探索する。ドナー原子候補の近傍にあるアクセプター原子候補を探索し、各ドナー原子候補にアクセプター原子候補のリストを作成する。各ドナー原子候補のアクセプター原子候補との距離についてソートし、ドナー原子候補からのアクセプター原子候補の方向が重なるような場合、より遠い距離のアクセプター原子候補をリストから排除する。水素結合を形成しない場合も含ませるために、リスト内に残ったアクセプター原子候補の方向と重ならない方向がある場合のみ、その方向にダミーアクセプター原子候補をリストに加える。ドナー原子候補のリストにあるアクセプター原子候補、つまり、水素原子の方向の組み合わせた水素原子ネットワークモデルを

生成する。1つの水素結合に水素が2つ存在するモデルは排除する。また、陰性原子間が短距離 (< 3Å) でありながら水素結合が形成されず(その間に水素原子がない場合)、かつ、どちらかのドナー原子の水素がアクセプター原子候補に向いていない場合(ダミーアクセプター原子候補に向いている場合)は、そのモデルは排除する。この様な方法で DFT 計算対象となる水素結合ネットワークモデルを網羅的に生成しつつも、その数を減らすことができた。この手法は Bash と Python で実装した。

(2) DFT によるエネルギー比較は3段階で行う。まず最初は、B3LYP/6-31G(d,p)のレベルで、非水素原子の位置と2面角は固定し、最適化終了閾値を Loose にした構造最適化を実行する。得られた最適化構造におけるエネルギーでのソートを実行する。最低エネルギーのモデルから数十 kcal/mol 高いエネルギーのモデルまでを次の段階のモデルとして残す。2段階目は、同レベルで、水素原子のみの構造最適化を一般的な閾値で実行した。3段階目は、B3LYP/6-31+G(d,p)のレベルで、かつ、Grimme's empirical dispersion correction version3(D3)を採用して水素原子のみの構造最適化を一般的な閾値で実行した。エネルギーでのソートを実行し、最小エネルギーのモデルをベスト候補として選ぶ。

(3) ベストモデルの正当性の検証は、非水素原子も含む構造最適化をベストモデルに実行し、結晶構造を保てるか、否か、で判断する。B3LYP/6-31+G(d,p) with D3 のレベルで、結晶構造との差を見る原子(結晶解析で十分に正確であると思われる原子または原子群)から十分に離れた、系の表面近くの C、もしくは、主鎖の原子を固定して実行する。検証対象の原子の位置が結晶構造からズレなければ、水素結合ネットワークのベストモデルの正当性があると判断できる。

4. 研究成果

ナトリウムポンプの膜領域イオン結合近傍の水素・プロトンの位置決定に本手法を適用して、手法の有効性や必要な計算リソースや時間を確認した。

膜領域イオン結合近傍には6つのキーとなるカルボキシル基があり、それらから距離 6Å 以内にある原子を含むアミノ酸残基と水分子およびイオンを探索する対象系とした。非水素原子の位置は PDB-ID 2ZXE から取り出した。その結果、59 アミノ酸残基、7つの水分子、そして2つの K⁺ イオンで構成された系となった。探索する水素結合ネットワークに関わるドナー原子候補は34個あり、探索対象となる水素結合の最大値は35個となった。この時点で、単純に水素結合の組み合わせを行うと $\sim 2.9 \times 10^{17}$ もの数となる。しかし、方法のセクションで述べたように、禁則となる条件に合ったモデルを排除していくと、1260 モデルまで下げることができる。これらのモデルはプロトンの数が異なり、その数は2-6個であった。プロトン数が2のモデルと6のモデルは、最適モデルを含まないであろうと考えたので、排除し、プロトン数が3-5個の866モデルが残った。禁則によるモデルの減少は大いに有効であることが分かった。この処理には大規模な計算リソースは不要で、パソコンですぐに終わるほどの計算量であった。

DFT 計算による段階的エネルギー解析は Xeon CPU が2台載った計算ノード34台ほどを断続的に1ヶ月強使用して、完了させた。第一段階の DFT 構造最適化では、そもそも電子状態計算が失敗してしまうモデルがあったが、それらはエネルギーが高い構造と思われるモデルであった。それら以外は想定内の計算が進み、完了している。DFT のレベルをより下げた B3LYP/STO-3G や semi-empirical な計算も実行したが、失敗してしまうモデルがより多かった。おそらく、計算対象の系に含まれる相互作用、特に水素結合相互作用を的確に再現できない場合に失敗をしているのではないかと推測する。

水素結合ネットワークモデルはプロトン数についてグループ分けをして、それぞれのグループ内でエネルギー比較を実行し、それぞれのベストモデルを決定した。プロトン数が異なるベストモデル間の比較にはプロトンの水和エネルギーを考慮した比較が必要となる。しかし、そもそも正確なプロトン水和エネルギーが分からない。DFT のエネルギーの精度を鑑みても、プロトン数が異なるモデル間の比較は有効ではないと判断している。

しかし、正当性の検証の結果で、プロトン数が異なるベストモデルの優劣が決定できた。更に、ベストモデルと思われるプロトン数のグループでは、最低エネルギーのモデルとエネルギー値がほぼ変わらない ($E = \sim 0.1$ kcal/mol) モデルと 6.0 kcal/mol ほど高い3つのモデルが得られたが、これらは非水素原子の構造最適化で結晶構造をベストモデルと同じように保った。それらのモデルは水素結合ネットワークの違いは僅かで、共存しているのかもしれない。

ナトリウムポンプへの検証は想定内の結果が得られ、実用性があったと判断できるものであるが、分解能が高く、非水素原子の位置が正確であったからであるとも言える。実際、分解能 3.0Å 近くの X 線結晶解析の部分的な水素結合ネットワーク探索を実行したが、うまくいかなかった。そもそも非水素原子の位置が不正確である可能性が高く、非水素原子の位置のより正確な refinement が必要なのであると推測している。しかしそれを実行するに当たり、水素結合に関わる情報が切望されているのも実際である。卵が先か鶏が先か、という状況に陥る。

また、分解能が高く、そして、原子間距離が 2.3 Å という酸素原子対を持つ結晶構造の検証を行う機会があったが、B3LYP/6-311++G(d,p)のレベルではその距離は標準的な距離(~ 2.7 Å)になってしまった。結晶解析にイレギュラーな問題があるのか、DFT 計算では low barrier HB の的確な再現をする能力が備わっていないのか、判断できかねぬ状況である。PDB で調査すると、意外に

も、low barrier HB を示唆するほどの短距離な陰性原子間が見受けられる。このような low barrier HB の存在は今回の手法では想定していなかった。よって、low barrier HB が蛋白質系にある程度存在しうるならば、この手法はもう一度検証し直す必要があるかもしれない。

幾つかの懸念点があるが、本手法は並列計算に適した方法であり、並列度の高い計算リソースが十分にあるならば、先に説明したナトリウムポンプの水素結合ネットワークの探索の掛かる時間は2、3日まで圧縮できると考えている。また、今日では GPU の計算スピードが上がり、かつ、量子化学計算のツールもそれらに対応しつつあるので、更に容易に実行できるものになっていくであろう。

また、semi-empirical ではあるが、DFTB+のような精度を保ちつつ必要な計算リソースを抑制したパッケージの開発も進んでいる。これらの可能性も調査していく必要があると思われる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------