

科学研究費助成事業 研究成果報告書

令和 3 年 6 月 7 日現在

機関番号：82401

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K06152

研究課題名(和文)人工知能型自動NMRデータ解析システムの開発

研究課題名(英文)Development of fully automated NMR data analysis assisted by AI

研究代表者

小林 直宏 (Kobayashi, Naohiro)

国立研究開発法人理化学研究所・放射光科学研究センター・上級研究員

研究者番号：80272160

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：NMR法は他の方法、X線結晶解析、クライオ電子顕微鏡などと比較し、経験や手動解析に依存した部分が多い。研究代表者は更なる高度化のため人工知能の技術を取り入れる世界初の試みに成功し、経験が無い初学者も高速・高精度に解析可能なシステムとして公開した。研究代表者は世界に先駆けて人工知能によるノイズ除去を高精度に実行可能な機能を搭載した解析ツールを公開し、国際誌へ掲載した。また多次元スペクトルの非線型取り込み法により測定時間を25%まで短縮し、合計18件におよぶ新規NMR構造を自動解析により達成し、うち12件は測定から構造解析、データ検証の完了までわずか数日、さらに2件は完全自動化を実現した。

研究成果の学術的意義や社会的意義

令和2年度までの成果として人工知能により高度化、高速化された解析システムが完成し、公開論文化された。公共データベースであるPDBへの述べ新規登録件数は18件以上にのぼり、そのうち13件は小型のタンパク質ながら新規フォールドであり測定、信号帰属解析、検証、登録には通常半年から1年を要することはごく普通である。今回の開発研究により大規模な測定期間短縮、解析やデータ検証の自動化が高度化されシステムとして統合され、その性能は世界最高レベルに達したと言っても過言ではない。この創薬研究などへの社会的インパクトは極めて大きいと考え、令和3年度内での国際学会発表、国際誌への投稿準備中である。

研究成果の概要(英文)：Compared to other methods such as X-ray crystallography and cryo-electron microscopy, the NMR method relies heavily on experience and manual analysis. We have succeeded in the world's first attempt to incorporate AI technology for further advancement, and has released a system that enables highly quick and accurate analysis even by beginners of NMR study. We published our result to a international journal as the first analysis tool equipped with a function that can perform noise removal with high accuracy using AI. In addition, by using a non-linear sampling method of multidimensional NMR spectra, the measurement time was reduced to 25%, and a total of 18 new NMR structure were determined by automated analysis, 12 of which took only a few days from measurement to completion of structural analysis and data verification, and two of which were fully automated. We are planning to publish all these strategy and automated system.

研究分野：NMR分光学

キーワード：NMR 自動解析 人工知能

1. 研究開始当初の背景

核磁気共鳴法 (NMR) は生体高分子などに対し原子分解能レベルで立体構造、動的状態、薬剤相互作用など多様な情報を得る事ができるその分解能の高さから分光学的にも極めて特異な存在である。さて、この15年を振り返って同じく構造モデル構築が可能であるX線結晶解析、クライオ電子顕微鏡解析の技術が大規模に自動化されている一方で NMR 解析は現在もお経験論や手動解析に依存した部分が多い。NMR 解析における自動化の障壁となっているものはその測定・解析・研究のために求められる知識として有機化学、生体高分子を扱う上での物理化学、解析のための代数論や計算機アルゴリズム、果ては量子力学まで含め全て初等レベルで良いにせよ極めて高度な知識を前提として取り組む必要があり、なおかつ信号帰属や構造モデル構築のための測定法の複雑さ・多様さは他の分光学とは比肩できない次元である。これらが初学者あるいは他分野研究者・非経験者の参入障壁となってきた。

2. 研究の目的

NMR 解析における自動化を長らく阻んできた上記障壁を取り除き、初学者でも容易に参入でき、なおかつその後も研究の多様さや解析・研究レベルがいかに高度になっても対応可能な柔軟かつ統合的な NMR 自動解析システムの高度化を目的とした。

3. 研究の方法

本課題においては自動化のモデルケースとして難易度の低いもの、高いものも含む熱安定性・溶解度共に高い人工タンパク質を14件、およびSS結合を高密度にもつ小型のタンパク質、極めて難易度の高いSS結合5つづく柔軟性に富むタンパク質それぞれ1例を対象とした測定・解析の高度化を行なった。また未発表であるが23kDaの膜タンパク質アクアポリン(AQZ)の固体NMR解析、40kDaに達する大きめのタンパク質(p120-GAP)の解析について特殊同位体ラベルを行いつつ解析効率化のためのシステム開発を行なった。人工タンパク質およびp120-GAP解析のためのほとんどの3次元あるいは4次元スペクトル測定については非線形信号取り込み法(NUS)を用いた。解析システムには当初予定していた深層学習による高度な信号検出・ノイズ除去技術を搭載し、信号シミュレーションを高速実行する特殊同位体ラベルのために有限オートマトン理論を用いたGraph_Robotを改良する。4次元スペクトルにはスペクトル解析を容易にするSync-Jump機能を拡張させたDouble-Sync-Jumpの開発に取り組んだ。また、解析の客観評価のためにP4ファージを用いて微弱に配向させた試料からもとめられる残余双極子を自動解析できるシステムの開発を行なった。

4. 研究成果

申請者は高い自動化性能を求めため深層学習の技術を取り入れる世界初の試みに成功し、令和2年度においてはこれらの技術的基盤をさらに発展させ、経験が無い初学者も高速高精度に解析を達成し、創薬研究に取り組めるシステムとして公開することを達成した。安定同位体ラベル化された蛋白質を用いる場合、多数の多次元スペクトルより得るべきNMR信号数は例えば76残基の蛋白質ですら4千個近い真の信号を検出する必要がある。これは従来の自動ピーク検出法で得ようとするならば多数のノイズを含む2万個以上の信号を検出してしまう。研究代表者は世界に先駆けて人工知能の一種である深層学習によるノイズ除去を高精度に実行可能な機能を搭載した解析ツールを公開し、2018年12月に国際誌への掲載に成功した。入力として検出されたNMR信号を画像化し、畳み込みニューラルネットに加え信号数・強度分布による統計的情報を最終段のニューラルネットへ入力し、ノイズの可否を総合判定する。この機能は単純な信号強度のみならず、信号近傍の状況を視覚的に人工知能(AI)が経験したイメージに基づいた判別が行われ人間の直感に近い処理が実行される。また、多次元スペクトルの非線形型取り込み法(NUS)によるNMR測定を代表者自らが実行することで測定時間を3次元スペクトルでは20~25%、4次元スペクトルで4~6%まで短縮することで貴重なマシンタイムを従来の4分の1程度まで

で高速化に成功し、13 件の構造解析に成功した。これらは全て現存しない新規フォールドであり、残余双極子による検証を経て X 線結晶解析に匹敵する精度での解析事例となった。そのうち 12 件は測定から構造解析、データ検証の完了までわずか数日、さらに 2 件は人間の手をほぼ全く介さない完全自動化を実現した。これらの成果は分子研古賀准教授らとの共同研究として国際誌への投稿準備中である。同グループとの共同研究として、98 残基中 Val を 30 個(メチル基信号 60 個以上)も含むタンパク質の極めて難度の高い解析にも成功し、その成果は 2020 年の国際誌に掲載された。また SS 結合を 32 残基中 3 箇所含む小型のチャンネル阻害タンパク質の解析はほぼ完全自動解析により達成し、慶應大学大澤教授らとの共同研究として国際誌に掲載された。化学交換により柔軟な構造を持ちつつ 168 残基中 5 箇所もの SS 結合を有する発光酵素である GLuc の構造解析も成功し、理研山崎 TL ら・農工大黒田教授との共同研究として国際誌に掲載された。これらに加え、未発表ながら理研山崎・長島らとの共同研究として膜タンパク質であるバクテリア由来アクアポリン(AQZ)の固体 NMR 解析も行い、溶液 NMR 法より更に複雑で難解な多次元スペクトルを改良した Graph_Robot と特殊 Sync-Jump を活用することで、230 残基中ほぼ全残基主鎖・側鎖信号の半自動的帰属解析に成功した。また、産総研竹内らとの共同研究としてがん関連タンパク質である p120-GAP (以下 GAP, 336 残基, 40kDa) の解析にも取り組んだ。その過程においては全重水素化 ^{13}C , ^{15}N ラベル体、メチル基選択的ラベル体など従来法よりはるかに安価に作成でき、かつ多様なラベル体を用いた。GAP の主鎖信号およびメチル基信号の帰属をおよそ 8 割達成し、現在その帰属結果を基にしたリガンド相互作用実験あるいは動力学的実験を竹内らが鋭意取り組んでいる。この規模のタンパク質をわずか 1 ヶ月程度の測定期間で完了し、連鎖帰属・メチル基帰属を達成する事例は極めて少なく、これらを支援するために新規に開発された Graph_Robot は特殊な同位体ラベルによる磁化移動シミュレーションを実現し、新機能である 4 次元スペクトル解析を容易にする double-sync-jump を完成させ、解析のさらなる自動化・高度化に至った。これらの成果は次世代人工知能型統合解析システムと共に 2021 年度内に国際誌あるいは国際学会にて発表を予定している。最終的に令和 2 年までに合計 18 件もの新規構造を自動解析により達成し、本研究課題は人工知能技術として近年注目されている深層学習および有限オートマトン理論による強固な支援により開発された次世代人工知能型統合解析システムの完成と共に、多数の成果を上げたと言える。

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 8件/うち国際共著 4件/うちオープンアクセス 1件）

| | |
|---|-------------------------------|
| 1. 著者名 Yamada T., Hayashi T., Hikiri S., Kobayashi N., Yanagawa H., Ikeguchi M., Katahira M., Nagata T. and Kinoshita M. | 4. 巻 59 |
| 2. 論文標題 How Does the Recently Discovered Peptide MIP Exhibit Much Higher Binding Affinity than an Anticancer Protein p53 for an Oncoprotein MDM2 | 5. 発行年 2019年 |
| 3. 雑誌名 Journal of Chemical Information and Modeling | 6. 最初と最後の頁 3533 ~ 3544 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acs.jcim.9b00226 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 該当する |
| 1. 著者名 Bala S., Shinya S., Srivastava A., Ishikawa M., Shimada A., Kobayashi N., Kojima C., Tama F., Miyashita O. and Kohda D. | 4. 巻 1864 |
| 2. 論文標題 Crystal contact-free conformation of an intrinsically flexible loop in protein crystal: Tim21 as the case study | 5. 発行年 2020年 |
| 3. 雑誌名 Biochimica et Biophysica Acta (BBA) - General Subjects | 6. 最初と最後の頁 129418 ~ 129418 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.bbagen.2019.129418 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 該当する |
| 1. 著者名 Romero R., Kobayashi N., Wedell R.J., Baskaran K., Iwata T., Yokochi M., Maziuk D., Yao H., Fujiwara T., Kurusu G., Ulrich L.E., Hoch C.J. and Markley L.J. | 4. 巻 14 |
| 2. 論文標題 BioMagResBank (BMRB) as a Resource for Structural Biology | 5. 発行年 2020年 |
| 3. 雑誌名 Methods Molecular Biology | 6. 最初と最後の頁 187 ~ 218 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-1-0716-0270-6_14 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 該当する |
| 1. 著者名 Kobayashi N, Hattori Y, Nagata T, Shinya S, Guentert P, Kojima C, Fujiwara T | 4. 巻 34 |
| 2. 論文標題 Noise peak filtering in multi-dimensional NMR spectra using convolutional neural networks | 5. 発行年 2018年 |
| 3. 雑誌名 Bioinformatics | 6. 最初と最後の頁 4300-4301 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/bty581 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 該当する |

| | |
|---|---------------------------|
| 1. 著者名 Koga R., M. Yamamoto, K. Takahiro*, N. Kobayashi*, T. Sugiki, T. Fujiwara and N. Koga | 4. 巻 117 |
| 2. 論文標題 Robust folding of a de novo designed ideal protein even with most of the core mutated to valine. | 5. 発行年 2020年 |
| 3. 雑誌名 Proc. Natl. Acad. Sci. USA | 6. 最初と最後の頁 31149-31156 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1073/pnas.2002120117 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|--|---------------------------|
| 1. 著者名 Wu N., Kobayashi N., Tsuda K., Unzai S., Saotome T., Kuroda Y. and Yamazaki T. | 4. 巻 10 |
| 2. 論文標題 Solution structure of Gaussia Luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear NMR. | 5. 発行年 2020年 |
| 3. 雑誌名 Sci. Rep. | 6. 最初と最後の頁 20069-20074 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-020-76486-4 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている (また、その予定である) | 国際共著 - |

| | |
|--|-----------------------|
| 1. 著者名 Nakayama N., Sakashita G., Nagata T., Kobayashi N., Yoshida H., Park S.Y., Nariai Y., Kato H., Obayashi O., Nakayama K., Kyo S. and Urano T. | 4. 巻 8 |
| 2. 論文標題 Nucleus Accumbens-Associated Protein 1 Binds DNA Directly through the BEN Domain in a Sequence-Specific Manner. | 5. 発行年 2020年 |
| 3. 雑誌名 Biomedicines | 6. 最初と最後の頁 608-612 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/biomedicines8120608 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|---|--------------------|
| 1. 著者名 Matsumura K., Takushi S., Yoshihiro K., Takayuki O., Kobayashi N., Imai S., Yanase N., Akimoto M., Fukuda M., Yokogawa M., Ikeda K., Kurita J., Nishimura Y., Shimada I. and Osawa M. | 4. 巻 22 |
| 2. 論文標題 Mechanism of hERG inhibition by gating-modifier toxin, APETx1, deduced by functional characterization | 5. 発行年 2021年 |
| 3. 雑誌名 BMC Mol. Cell Biol. | 6. 最初と最後の頁 3-11 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12860-020-00337-3 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計8件（うち招待講演 0件 / うち国際学会 3件）

| |
|--|
| 1. 発表者名 小林直宏, 杉木俊彦, 南慎太郎, 佐久間航也, 長島敏雄, 児嶋長次郎, 藤原敏道, 小杉貴洋, 古賀理 恵, 古賀信康, 山崎俊夫 |
| 2. 発表標題 高速・高精度な MagRO による NMR 構造解析の完全自動化 |
| 3. 学会等名 第59回NMR討論会 |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 A. Wickramasinghe, Y. Xiao, N. Kobayashi, T. Yamazaki and Y. Ishii |
| 2. 発表標題 Structural differences and Novel Polymorphs of Synthetic and brain-derived A _β 42 Fibrils by 1H-detected SSNMR |
| 3. 学会等名 第59回NMR討論会 |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 小林直宏, 長嶋敏夫, 山崎俊夫 |
| 2. 発表標題 Nearly complete assignments for 23kDa membrane protein by solid-state NMR using MagRO system assisted by Graph_Robot based on finite automaton theory. |
| 3. 学会等名 第58回NMR討論会 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 A. Wickramasinghe, Y. Xiao, N. Kobayashi and I. Ishii |
| 2. 発表標題 Structural characterization of A _β (1-42) fibrils in nano-mole scale by 1H-detected SSNMR under UFMAS Structural characterization of A _β (1-42) fibrils in nano-mole scale by 1H-detected SSNMR under UFMAS |
| 3. 学会等名 第58回NMR討論会 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 佐久間航也、小林直宏、杉木俊彦、鈴木かの、小林直也、小杉貴洋、村田武士、古賀理恵、古賀信康 |
| 2. 発表標題 Design of Globin-like Complicated Folds |
| 3. 学会等名 64th BioPhysical Society Annual Meeting, SanDiego (国際学会) |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 南慎太郎、小林直宏、杉木俊彦、鈴木かの、千見寺浄慈、古賀理恵、古賀信康 |
| 2. 発表標題 Exploration of novel protein folds by de novo design |
| 3. 学会等名 64th BioPhysical Society Annual Meeting, SanDiego (国際学会) |
| 4. 発表年 2020年 |

| |
|--|
| 1. 発表者名 小林直宏、杉木俊彦、永田崇、Peter Guentert, 児嶋長次郎、藤原敏道 |
| 2. 発表標題 Deep-MagRO: 深層学習を応用した高速かつ高精度な統合的NMR解析システム |
| 3. 学会等名 第57回NMR討論会 |
| 4. 発表年 2018年 |

| |
|--|
| 1. 発表者名 N. Kobayashi, T. Sugiki, Y. Hattori, T. Nagata, S. Shinya, J.M. Wuerz, R. Koga, N. Koga, P. Guentert, C. Kojima and T. Fujiwara |
| 2. 発表標題 Application of Deep Neural Networks for Highly Accurate NMR Structure Analysis by Deep-MagRO |
| 3. 学会等名 ICMRBS2018 (国際学会) |
| 4. 発表年 2018年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

NMR Tool Box
https://bmrbddep.pdbj.org/en/nmr_tool_box.html

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 |
|-------|--|---------------------------------------|----|
| 研究分担者 | 児嶋 長次郎 (Kojima Chojiro) (50333563) | 横浜国立大学・理工学部・教授 (12701) | |

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 |
|---------|---------|
|---------|---------|