

## 科学研究費助成事業 研究成果報告書

平成30年 8月27日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K06970

研究課題名(和文) データベースを利用したNMR創薬支援パイプラインの開発

研究課題名(英文) Development of NMR drug discovery pipeline using database

研究代表者

小林 直宏 (Kobayashi, Naohiro)

大阪大学・たんぱく質研究所・特任准教授(常勤)

研究者番号：80272160

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：ゲノム配列にコードされ実験的に解明されている立体構造は全体の約10%程度であり、近年では構造予測モデルを使った創薬研究事例が増加する傾向がある。本研究課題では核磁気共鳴(NMR)研究のターゲットとなりうるモデル構造データベースをNMR実験データのデータベースであるBMRBなどの複数の外部生命科学系データベースとリンクをさせる創薬支援データベースを構築した。平成28～29年度においてはデータベースを活用し、深層学習を基盤とした自動解析パイプラインを設計し、創薬研究に必要となるNMRシグナル帰属と溶液立体構造決定を従来の4分の1程度のNMRデータを用いた自動解析により半日程度まで短縮可能にした。

研究成果の概要(英文)：The three-dimensional structure encoded in the genome sequence experimentally determined is about 10% of the total, and recently there are tendencies for cases of drug discovery research using a structure prediction model. In this research project, we have constructed a drug discovery support database that links a model structure database that can be a target of nuclear magnetic resonance (NMR) research, which is lined to several external life science databases such as BMRB, a database of NMR experiment data. In the period from FY2016 to FY2017, utilizing a database, we designed an automatic analysis pipeline based on deep learning, and determined the NMR signal assignments and the solution tertiary structure required for drug discovery research using about 25% of NMR data sets. The new technology has successfully shortened automatic NMR analysis to about half a day.

研究分野：核磁気共鳴分光学

キーワード：核磁気共鳴法 データベース 自動解析

### 1. 研究開始当初の背景

研究開始当時、解読されているヒトゲノムの遺伝子は約 26,000 と言われ、コードされているタンパク質はおよそ 30 万と推定されていた。その一方、実験的に決定された立体構造数は生体高分子の構造データベースである PDB において 2014 年 10 月の時点で 10 万件を超えていたが、ヒトあるいは高等哺乳類に限定し、配列の重複も考慮すると実際に原子分解能で解明されている立体構造は 1~2 万件程度という事実がある。したがってゲノム配列と立体構造との橋渡しには大きなギャップが存在するが、ホモロジーモデリングをはじめとする立体構造予測法がそのギャップを埋める役割を担っている。本研究における連携研究者である本野千恵主任研究員(産業総合研究所)らは JST バイオインフォマティクス推進センター(BIRD)主導によるプロジェクトにより、高精度なモデル構造のデータベースである SAHG (Structural Atlas of Human Genome) を開発し公開している。このデータベースではヒトゲノム配列を網羅的に解析し、約 43,000 件のタンパク質構造について極めて高精度な立体構造予測を実現した。SAHG の特徴として、登録されている構造の約半数がリガンドを有する創薬研究にとって大変興味深いターゲットを多数保有している。標準的な構造モデリング技術によれば、テンプレートとなる立体構造に対して 40%ほどの配列相同性をもつ構造未知のタンパク質については主鎖 RMSD にして 1 程度の精度でモデリングすることが可能である。しかしながら、ループ領域など機能発現に重要な領域に関してはインシリコスクリーニングによるリガンド候補の探索などの創薬研究において、より高精度の構造情報が求められる。この精度の構造情報を得る解析手法としては X 線結晶解析法が主流であるが、NMR 法は立体構造の解明のみならず、得られる化学シフトを利用して、NMR リガンドスクリーニング実験など創薬に直結する研究を展開できる特色がある。NMR 法の問題点としては解析の自動化の完成度が十分であるとはいえず、既存のモデル構造や化学シフトが得られるにもかかわらずそれらを有効に利用して高精度な解析を自動実行する、いわゆる分子置換法が確立されているとはいいがたい。近年の製薬企業を含めた創薬研究業界においては膨大な資金による大規模スクリーニングを行う研究スタイルからよりコストパフォーマンスの高いものが指向されている一方、iPS 細胞などを用いた創薬スクリーニング法などを代表とした新しい技術を利用した新薬開発の機運が高まっている。このような背景において、データベースより得られる有効な創薬候補のモデル構造や NMR データを用いて立体構造解析を迅速に全自動的に実行できるラボスケールのパイプライン化が強く望まれる。

### 2. 研究の目的

本研究においては初期の段階で SAHG に登録されているモデル構造に対して外部データベースとして配列情報を RefSeq、構造座標について PDB、化学シフトについて BMRB、機能情報を UniProt、一塩基置換情報と疾患関連情報を OMIM、相互作用情報を IntAct をセマンティックウェブ技術によりリンクさせる。これにより、創薬ターゲットとしての重要度をより客観的な基準として絞り込み検索が可能ないように構築した創薬支援データベースを完成させる。このデータベースより得られる 40%以上の配列相同性を持つテンプレートより構築されたモデル構造と 60%以上の配列相同性を持つ既知化学シフト情報に加え、NMR 実験データを入力とし、シグナル帰属および立体構造計算の解析を完全自動的に実行させるシステムを開発し、NMR 分子置換法として確立させる。このシステムに解析結果を評価するルーチンと組み合わせることで従来法より優れた精度と実行速度を実現させ、強力なモデル構造データベースとシームレスに連携させたタンパク質立体構造解析のハイスループットな創薬支援パイプラインとして完成させる。

研究代表者および分担研究者より実際のサンプルデータを用いた NMR 解析を行い、創薬支援パイプラインを使った解析を行うことで創薬研究に利用可能な精度で化学シフト帰属と溶液構造とを決定し、パイプラインの実用性について実証する。これらの解析を通じて、パイプラインに設定されているプログラム群の実行パラメータを最適化させ、最終年度までにラボスケールで創薬研究に運用可能なレベルにまで最適化を実施し、全システムの完成を目指す。

### 3. 研究の方法

#### 1) 創薬支援データベースの構築 (小林、本野が担当)

NMR 創薬研究のターゲットとなりうるモデル構造の 2 次データベースをセマンティックウェブ技術により構築する。SAHG データベースに登録されているモデル構造のうち、テンプレート構造の配列に対して 40~90%配列相同性を持つものを選別する。これらは高精度に予測された構造であり、NMR シグナルの自動帰属と立体構造計算を高精度に実現するために利用され NMR 解析における分子置換法を実現することができる。また、これらのモデル構造について NMR 実験データのデータベースである BMRB に登録されている化学シフトデータのうち配列相同性が 60%以上のデータを持つものをリンクさせる。また Uniprot、OMIM、IntAct などの外部のデータベースとのリンクを作成し、機能情報、一塩基置換情報、疾患情報、相互作用情報などをリンクさせた創薬支援データベースを構

築し、創薬研究のための有用なタンパク質を客観的な評価基準により優先順位をつけ効果的に選定を行うことができるよう開発する。

2) NMR 分子置換法による化学シフトと立体構造決定の全自動化 (小林が担当)

化学シフトと溶液構造の解析を高精度に自動実行するための NMR 分子置換システムの開発と自動化パイプラインに用いるパラメータ群の最適化を行う。パイプラインではモデル構造と既知の化学シフトデータを用い、実験的に測定された NMR スペクトルデータを入力させる。研究代表者が開発した MagRO はアニリング法により NMR シグナル自動帰属を実行する `anneal_robot`、実験的に得られた NMR スペクトルと帰属結果とのずれを評価関数として計算する `graph_robot`、構造座標と化学シフトの整合性を元に評価するプログラム `validation_robot` から構成される。これらを外部プログラムである FLYA と SPARTA+ を連携させ、NMR シグナルの自動自動解析を実行するシステムを構築する。MagRO/FLYA/SPARTA+ の計算により、評価関数によりランク付けされた化学シフトテーブルを多数発生させ、外部プログラムである CYANA により溶液構造を計算させる。CYANA により計算された構造は Amber による水分子を考慮した Refinement 計算により精密化される。これらの計算工程を溶液構造と化学シフトを評価するパラメータを持つ NMR 分子置換システムとして開発する。このシステムは実験的に決定されたデータをフィードバックすることにより最適化が可能なように設計し、自動実行が可能なパイプラインとして開発し、平成 27 年度初期においてパイプラインバージョン 1 として完成させる。

3) タンパク質の NMR 解析と分子置換システムへのフィードバック (児嶋、廣明が担当)

本研究においてはモデル構造データベースにおける複数のサンプルについて安定同位体を用いた大腸菌による大量発現、精製、サンプル調整を行う。選ばれたターゲットのタンパク質サンプルについて溶液構造を決定するために必要とされる NMR スペクトル群を測定する。MagRO/FLYA/SPARTA+ による自動的な解析を実行し、平成 27 年度においてはパイプラインバージョン 1 を用いた解析を行う。特に帰属が困難なシグナルの解析、構造精密化解析については手動による解析を部分的に実行し、全自動解析と比較する。この作業は構造解析、シグナル帰属のダブルチェックを行うことになり、より信頼性の高い結果を得ることができると同時にパイプラインのパラメータ最適化における重要なフィードバックとなる。NMR 分子置換法により得られる溶液構造と化学シフトの整合性を `validation_robot` による正解構造と正解化学シフトとのずれを評価関数として最適化するルーチンを分子置換法システムに搭載させる。クオリティーの異なる NMR 実験デー

タ群を教師データとするパイプラインの最適化を行い、パイプラインバージョン 2 として完成させる。自動化が特に難しい NMR シグナルの検出機能や芳香族アミノ酸シグナルの帰属についての決定精度を向上させるため、`graph_robot` と `validation_robot` とを統合したルーチンを開発し、分子置換法システムに組み込む。ここまでのシステムは多くのプログラムモジュールが介在し、大変複雑になるため全システムのプログラムコードをリファクタリングすることで簡素化し、パラメータ群の最適化を完了させ公開用の最終バージョンとしてパイプラインバージョン 3 を完成させる。

平成 28 年度以降も創薬支援データベースより選定されたターゲットのタンパク質の大量発現、精製および NMR スペクトル測定を継続して行い、パイプラインバージョン 2 および 3 を用いた解析を行う。これにより多くの創薬研究に十分利用できる高精度な溶液構造と化学シフトデータを得ることが期待できる。

#### 4. 研究成果

平成 27 年度においては予定していたモデル構造をベースとした NMR 創薬支援データベースは完成し、実際に検索サイトとして稼働させる事が実現した：

[http://bmrdep.pdbj.org/en/mpm\\_search.html](http://bmrdep.pdbj.org/en/mpm_search.html)

また、本研究課題によって開発されたシステムは統合的研究科遺跡環境である MagRO に組み込まれ、化学シフトの統計的なデータや予測をするための機能を獲得できた。もっとも興味深いのは BMRB に登録されている NMR スペクトルデータを訓練データとする深層学習システムの完成である。このシステムはついに計算機が人間の持つ視覚と直感を得た状態に等しく、従来手作業を強いられてきた NMR シグナルのノイズ判定性能を著しく向上させた。平成 28 年度において第 55 回 NMR 討論会の口頭発表に採択され、高速かつ高精度な主鎖帰属性能について発表した。さらに平成 29 年度においては Gordon Research Conference において異例の口頭発表依頼となり当該機能を国際学会の場で発表する事となった。更に深層学習システム、`Graph_robot` と `validation_robot` を組み合わせ高度化された最新のシステムによりほぼ完全なシグナル帰属、立体構造解析に成功した。従来法より 3 分の 1 ないし 4 分の 1 程度の NMR スペクトルデータに対して最も速い場合は数時間での解析完了を可能にする画期的な成果へと結実した。この成果は平成 29 年度第 56 回 NMR 討論会に口頭発表採択され世界に先駆けて発表する事となった。現在、有名国際誌に投稿中であり、ほぼ受理に近い状態である。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 11 件)

1. K. Furuita, S. Kataoka, T. Sugiki, Y. Hattori, N. Kobayashi, T. Ikegami, K. Shiozaki T. Fujiwara and C. Kojima "Utilization of paramagnetic relaxation enhancements for high-resolution NMR structure determination of a soluble loop-rich protein with sparse NOE distance restraints" (2015) *J. Biomol. NMR.* **61**(1), 55-64
2. A. Gutmanas, P.D. Adams, B. Bardiaux, H.M. Berman, D.A. Case, R.H. Fogh, P. Güntert, P.M. Hendrickx, T. Herrmann, G.J. Kleywegt, N. Kobayashi, O.F. Lange, J.L. Markley, G.T. Montelione, M. Nilges, T.J. Ragan, C.D. Schwieters, R. Tejero, E.L. Ulrich, S. Velankar, W.F. Vranken, J.R. Wedell, J. Westbrook, D.S. Wishart, and G.M. Vuister "NMR Exchange Format: a unified and open standard for representation of NMR restraint data" (2015) *Nat. Struct. Mol. Biol.* **22**(6), 433-434
3. M. Yokochi M, N. Kobayashi, E.L. Ulrich, A.R. Kinjo, T. Iwata, Y.E. Ioannidis, M. Livny, J.L. Markley, H. Nakamura, C. Kojima and T. Fujiwara "Publication of nuclear magnetic resonance experimental data with semantic web technology and the application thereof to biomedical research of proteins" (2016) *J. Biomed. Semantics.* **7**(1), 16
4. K. Kuwasako, N. Nameki, K. Tsuda, M. Takahashi, A. Sato, N. Tochio, M. Inoue, T. Terada, T. Kigawa, N. Kobayashi, M. Shirouzu, T. Ito, T. Sakamoto, K. Wakamatsu, P. Güntert, S. Takahashi, S. Yokoyama and Y. Muto "Solution structure of the first RNA recognition motif domain of human spliceosomal protein SF3b49 and its mode of interaction with a SF3b145 fragment." (2017) *Protein Sci.* **26**(2), 281-291
5. J. Yong, J.D. Westbrook, Z. Feng, S. Raul, E. Peisach, T.J. Oldfield, S. Sen, A. Gutmanas, D.R. Armstrong, J.M. Berrisford, L. Chen, M. Chen, L.D. Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P.M.S. Hendrickx, B.P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C.L. Lawson, Y. Liang, S. Mading, L. Mak, M.S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M.R. Sekharan, C. Shao, G.J. Swaminathan, L. Tan, E.L. Ulrich, G. van Ginkel, R. Yamashita, H. Yang, M.A. Zhuravleva, M. Quesada, G.J. Kleywegt, H.M. Berman, J.L. Markley, H. Nakamura, S. Velankar and S.K. Burley, "OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive" (2017) *Structure*, **25**(3),536–545
6. T. Sugiki, N. Kobayashi and T. Fujiwara "Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists." (2017) *Comput. Struct. Biotechnol. J.* **15**, 328-339
7. R. Iwaoka, T. Nagata, K. Tsuda, T. Imai, H. Okano, N. Kobayashi and M. Katahira "Structural Insight into the Recognition of r(UAG) by

- Musashi-1 RBD2, and Construction of a Model of Musashi-1 RBD1-2 Bound to the Minimum Target RNA.” (2017) *Molecules*, **22**(7), E1207
8. R. Iwaoka, T. Nagata, K. Tsuda, T. Imai, H. Okano, **N. Kobayashi** and Katahira M. “Backbone and side chain assignments of the second RNA-binding domain of Musashi-1 in its free form and in complex with 5-mer RNA.” (2017) *Biomol NMR Assign.* **11**(2), 265-268
  9. Y. Nomura, K. Yamazaki, R. Amano, K. Takada, T. Nagata, **N. Kobayashi**, Y. Tanaka, J. Fukunaga, M. Katahira, T. Kozu, Y. Nakamura, Y. Haishima, H. Torigoe and T. Sakamoto. “Conjugation of two RNA aptamers improves binding affinity to AML1 Runt domain.” (2017) *J. Biochem.* **162**(6), 431-436.
  10. A.R. Kinjo, G.J. Bekker, H. Wako, S. Endo, Y. Tsuchiya, H. Sato, H. Nishi, K. Kinoshita, H. Suzuki, T. Kawabata, M. Yokochi, T. Iwata, **N. Kobayashi**, T. Fujiwara, G. Kurisu and H. Nakamura. “New tools and functions in data-out activities at Protein Data Bank Japan (PDBj).” (2017) *Protein Sci.* **27**(1), 95-102
  11. K. Kondo, T. Mashima, T. Oyoshi, R. Yagi, R. Kurokawa, **N. Kobayashi**, T. Nagata, M. Katahira “Plastic roles of phenylalanine and tyrosine residues of TLS/FUS in complex formation with the G-quadruplexes of telomeric DNA and TERRA.” (2017) *Sci. Rep.* **8**(1), 2864.
1. 栢尾尚哉、星川美穂、**小林直宏**、楯真一 “HSP70 タンパク質のアロステリック構造変化の NMR 解析” 第 15 回蛋白質科学会年会、ポスター、2015 年 6 月、徳島
  2. **N. Kobayashi**, M. Yokochi, T. Iwata, B.R. Sahoo, T. Nagata, J.L. Markley, E.L. Ulrich, E. Schmidt, P. Güntert, C. Kojima and T. Fujiwara “New strategy for high-throughput NMR analysis using the NMR database BMRB and tools for automated NMR analysis, MagRO, FLYA and CYANA” ISMAR 2015、口頭、2015 年 8 月、上海
  3. **小林直宏**、横地政史、岩田武史、児嶋長次郎、藤原敏道 “生体高分子 NMR データベース(BioMagResBank)の統合化とその応用” トーゴの日シンポジウム 2015、ポスター、2015 年 10 月、東京
  4. **小林直宏**、横地政史、岩田武史、本野千恵、廣明秀一、児嶋長次郎、藤原敏道 “NMR データベース BioMagResBank の統合的拡張と公開” 第 54 回 NMR 討論会、ポスター、2015 年 11 月、千葉
  5. Y. Hattori, S. Jukab, V. Sychrovský, K. Furuita, T. Sugiki, I. Ohki, T. Ikegami, **N. Kobayashi**, Y. Tanaka, T. Fujiwara and C. Kojima “The Salt Bridge on Protein Surface: Characterization Using the <sup>13</sup>CH<sub>3</sub> NMR Probe” ICMRBS 2016、ポスター、2016 年 8 月、京都
  6. M. Yokochi, **N. Kobayashi**, T. Iwata, E.L. Ulrich, J.L. Markley, Ak.R. Kinjo, H. Nakamura, C. Kojima and T. Fujiwara, “Publication of NMR Data Archived at BMRB with Web Standard Technologies for

Federated and Integrated Search Services" ICMRBS 2016、ポスター、2016年8月、京都

7. **N. Kobayashi**, M. Yokochi, T. Iwata, J.L. Markley, E.L. Ulrich, E. Schmit, P. Güntert, Y. Hattori, C. Kojima and T. Fujiwara "New library system in MagRO for automated analysis of NMR data on biomacromolecules", ICMRBS 2016、ポスター、2016年8月、京都
8. M. Yokochi, **N. Kobayashi**, T. Iwata, E.L. Ulrich, J.L. Markley, A.R. Kinjo, H. Nakamura, C. Kojima and T. Fujiwara "PDBj-BMRB: a web search service for biological NMR data and beyond" 生命医薬情報学連合大会 2015、ポスター、2016年10月、東京
9. **小林直宏**、横地政史、岩田武史、児嶋長次郎、藤原敏道 "高度化された生体高分子 NMR データベース (PDBj-BMRB) による統合的利用" トーゴの日シンポジウム 2016、ポスター、2016年10月、東京
10. **小林直宏**、服部良一、永田崇、児嶋長次郎、藤原敏道, "深層学習に支援された NMR シグナルの自動解析", 第 55 回 NMR 討論会、口頭、2016 年 11 月、広島
11. **N. Kobayashi**, T. Nagata, C. Kojima and T. Fujiwara, "Highly automated NMR signal assignments using peak-noise recognition assisted by Deep Neural Networks", Gordon Research Conference、口頭、2017 年 6 月、米 Sunday River
12. **小林直宏**、永田崇、服部良一、新家粧子、J.M. Würz, P. Güntert, 児嶋長次郎、藤原敏道, "深層学習により強化された NMR 立体構造解析", 第 56 回 NMR 討論会、口頭、2017 年 11 月、東京

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等  
MultipleProteinModelSearch  
[http://bmrdep.pdbj.org/en/mpm\\_search.html](http://bmrdep.pdbj.org/en/mpm_search.html)  
産総研本野らによって開発された SAHG サーバに保存されたエントリに基づいてモデル化された構造を使用して NMR ベースの創薬を支援するために開発され、キーワードでタンパク質モデル構造検索、BMRB、UniProt、OMIM、IntAct などの他のライフサイエンスデータベースとリンクされ NMR ベースの薬物スクリーニングに有用と言える。

#### 6. 研究組織

##### (1) 研究代表者

小林直宏 (KOBAYASHI, Naohiro)  
大阪大学・たんばく質研究所・特任准教授  
研究者番号：80272160

##### (2) 研究分担者

児嶋長次郎 (KOJIMA, Chojiro)  
横浜国立大学・大学院工学研究院・教授  
研究者番号：50333563

##### (3) 連携研究者

廣明秀一 (HIROAKI, Hidekazu)  
名古屋大学・創薬科学研究科・教授  
研究者番号：10336589

本野千恵 (MOTONO, Chie)  
産業技術総合研究所・創薬分子プロファイリング研究センター・主任研究員

研究者番号：80415752