

平成 30 年 6 月 11 日現在

機関番号：82401

研究種目：新学術領域研究(研究領域提案型)

研究期間：2013～2017

課題番号：25120003

研究課題名(和文)スパースモデリングによるNMR計測・解析の高速高精度化

研究課題名(英文) Initiative for faster and more precise NMR data measurement and analysis with sparse modeling

研究代表者

木川 隆則 (Kigawa, Takanori)

国立研究開発法人理化学研究所・生命システム研究センター・チームリーダー

研究者番号：20270598

交付決定額(研究期間全体)：(直接経費) 58,800,000円

研究成果の概要(和文)：本研究では、スパースモデリングの導入により、特に複雑な生命分子系を対象にしたNMR計測やデータ解析の高速高精度化の実現による飛躍的発展を目指した。【課題1】では疎に計測した信号の再構成法に関して、スペクトル再現性の高い方法を開発した。【課題2】では、スペクトル分解とアミノ酸判別を同時におこなう復号法を構築し、「符号化標識法」の復号性能を飛躍的に向上させた。【課題3】ではベイズ推定を応用した立体構造計算手法を構築し、真核細胞中のタンパク質のde novo立体構造決定に成功した。【課題4】では、疎に計測したデータからシグナル同定、構造動態情報の取得までを一体的に実現する全く新規の手法を開発した。

研究成果の概要(英文)：In this project, focusing on the sparseness of NMR data and structure information of biomolecule, we have aimed to achieve faster and more precise NMR measurement and analysis of biomolecules by solving the issues arising from complexity and complicatedness of NMR analysis with the sparse modeling. In the subproject 1, we have successfully developed the spectrum reconstruction method with high reproducibility. In the subproject 2, decoding process of "SiCode" was dramatically improved by concurrently deconvoluting spectra and identifying amino-acid assignment. In the subproject 3, the method for faster and more precise biomolecular structure calculation was established by using Bayesian inference, and was successfully applied to de novo structure determination of proteins in living eukaryotic cells. In later introduced subproject 4, we have successfully developed a novel method enabling concurrent signal identification and structural dynamics analysis.

研究分野：構造生物学

キーワード：スパースモデリング NMR 圧縮センシング スペクトル解析 ベイズ統計 レプリカ交換モンテカルロ法 テンソル分解 安定同位体標識

1. 研究開始当初の背景

核磁気共鳴法(NMR)は、非侵襲的に分子構造動態を探る最も強力で汎用性の高い計測手法の一つである。近年の装置技術の著しい進展により、タンパク質等の複雑な生命分子の計測が可能となったが、計測データ量が多く計測時間が長く、得られるスペクトルは複雑でデータ解析は煩雑である。さらに解析データ数の増大により、立体構造を得るための計算も煩雑で時間を要することから、既存手法の利用が限界に達しつつあった。

2. 研究の目的

本研究では、NMR データや生命分子の構造情報が有するスパース性に着目して、スパースモデリング(SpM)の導入により NMR 解析の複雑さ・煩雑さに起因する問題を解決する。生命分子は不安定で長時間計測に適さない試料が多いため、SpM を用いたデータ処理手法を活用して短時間で高精度の NMR データ計測を可能とする。また、対象分子の複雑度が増すにつれ得られるスペクトルの分離識別は困難となるため、SpM によりスペクトル分離・識別能を向上させスペクトル解析の高精度化と高速化を実現する。さらに、ベイズ統計を応用した新しい NMR 構造決定法を開発し最適構造を導くことを容易にすることにより、立体構造計算を高速・高精度化する。これにより、自然な状態に近い「生きた細胞内」の生命分子を解析する in-cell NMR 法など、複雑な生命分子系を対象にした NMR 計測・解析に利用していくことにより、生命分子の動態と機能の理解に大きく貢献する。

3. 研究の方法

本研究では、研究遂行中に着想した課題も含めて以下課題に関して研究を進めた。

【課題 1】NMR 計測の高速・高精度化

NMR 信号を非線形標本化(NUS)などによりスパースに計測し、圧縮センシング等の SpM 手法によりデータ欠落を補間する信号処理をおこない、計測時間の大幅短縮とスペクトル品質の改善を実現する。

【課題 2】データ解析の高速・高精度化

複雑な多次元 NMR スペクトルのピーク分離精度を向上させることにより煩雑なスペクトル解析を容易化し、人間の判断に依存しない解析アルゴリズムを構築することにより、データ解析のボトルネックの一つを解消して NMR 解析の高速・高精度化を実現する。

【課題 3】立体構造計算の高速・高精度化

レプリカ交換によりタンパク質の構造空間を広域に渡り探索して構造の事前確率を決定するベイズモデルに基づくデータ処理により、客観・定量的なパラメータ設定を可能とする新規のアルゴリズムを開発することにより、立体構造計算を高速・高精度化する。

【課題 4】NMR 計測・解析・立体構造計算を一体化した手法の開発

上記課題を遂行する過程で得られた新規の

着想に基づく本課題では、タンパク質の構成単位であるアミノ酸の情報を符号語で表現して情報科学のモデルとして扱うことを可能とした新規概念を基盤として、シグナル分離の問題を「どのアミノ酸であるか」というモデル選択の問題、すなわちアミノ酸判別の問題として単純化することにより、統一的に取り扱い、これを更に発展させることで、NMR 計測・解析・立体構造計算を一体化した新たな生命分子の構造解析手法を開発して質的量的に著しい向上を実現する。

4. 研究成果

(1) 多次元 NMR 測定において、間接観測軸の全点について測定をおこなう均一サンプリングではなく、一部の時間軸データ点について測定を省略する NUS によって、測定時間の短縮あるいは時間あたりの解像度や感度の向上が得られるが、少ない時間軸観測データから周波数軸スペクトルを求める問題は劣決定問題である。多次元の周波数スペクトルではシグナルがまばらにしか存在しないことを利用し、圧縮センシング(CS)などのスパース推定をおこなう方法は有力である。我々は、大自由度班(C01-3)の竹田・小淵・井上・樺島と共同で、周波数スペクトルに L_p 正則化($p=0$ or 1)を科す CS を活用した再構成アルゴリズムを構築した。この方法を、シグナル強度のばらつきが大きい 3 次元 NOESY スペクトルに適用したところ、広く用いられているプログラム mddNMR に比べ、強度の再現性がよかった(図 1)。特に、強度の弱いシグナルについても正確な強度が得られており、立体構造計算を正確におこなう上で特に有効である。

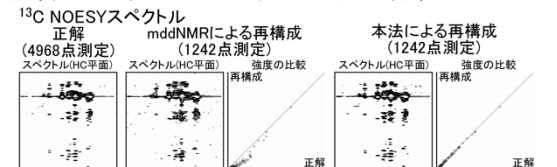


図 1. CS によるスペクトル再構成

(2) 我々は、タンパク質の NMR 解析で必須の安定同位体標識手法に関して、「アミノ酸の情報を安定同位体標識率に符号化しておき、観測したスペクトルから復号する」との概念に基づく「符号化標識法 (Stable isotope encoding; SiCode)」を開発した(図 2)。情報科学に基づく手法であるため、ノイズ耐性を向上させる標識パターンを設計したり、誤り符号機構を導入してシグナル同定の誤りを検出することも可能な、強力な手法である。

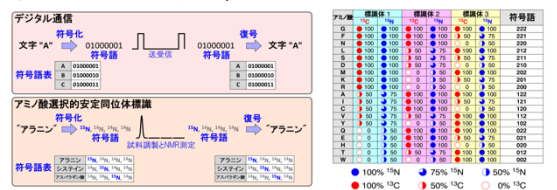


図 2: 符号化標識法の概念と標識パターンの一例
本手法におけるシグナルの同定(安定同位体標識率として測定試料に埋め込まれたアミノ

酸情報の NMR スペクトルからの“復号”)において、シグナルが重複・縮重した場合には同定が困難であることが、本手法の弱点であった。そこで我々は、重複したシグナルの同定に関して、スパースモデリング班(B01-2)の永田・岡田と共同で取り組んだ。

永田らが開発したベイジアンスペクトル分解の手法を活用し、アミノ酸の組み合わせをモデルとし、各モデルについてレプリカ交換モンテカルロ法(REMC)でピークフィットをおこないつつ自由エネルギーを算出し、もっとも自由エネルギーの低いモデルを正解として選択することで、アミノ酸の情報を得ることができた。この方法は、2~3 個程度のシグ

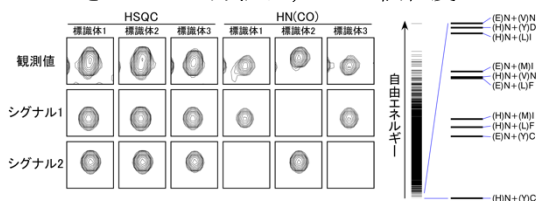


図3: 重複シグナル分離とモデル選択

ナルが重複した場合には有効であったものの、アミノ酸の組み合わせ数は一般的に1シグナルあたり数百通りあり、シグナルが増えるにつれ検討しなければならないモデル数が指数関数的に増加するため、4 個以上のシグナルが重複した場合の解析は現実的でなかった。そこで、解の近傍を重点的に探索できるというマルコフ連鎖モンテカルロ法の特徴を活かすためにも、モデル(アミノ酸の組み合わせ)そのものも同時に探索することとした。説明変数として連続変数(ピーク位置・強度・線幅)と離散変数(アミノ酸)を同時に扱うことになるが、REMCを自然に拡張して実現できる。アミノ酸変数の値域を、アミノ酸配列にあらわれる組み合わせのみに制限し、それ以外の組み合わせへの遷移を無条件に棄却することで、アミノ酸配列の情報も利用した探索とすることができた。この方法により、条件により異なるが10~20シグナル程度の同時解析がおこなえるようになった。

アミノ酸配列と標識パターンを事前知識として活用し、観測されたスペクトルをもっともよく説明するモデルを提示するというこの方法は、シグナルが重複した場合だけでなく、極端に強度の低い雑音に埋もれたシグナルの解析にも有効である。in-cell NMR のデータは、高濃度で存在する周囲の分子の影響で極めて信号強度が低くなるため一般的に解析が困難であるが、生きた細胞内にあるタンパク質も当該手法により解析可能となった(図4)。

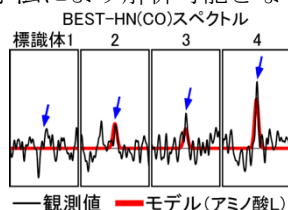


図4: 細胞内タンパク質信号からのアミノ酸情報復号

(3) NMR を用いた蛋白質立体構造決定は、スペクトルから得られるピーク位置と強度を、原子間距離や角度情報に変換し、これを拘束条件として大域最適構造を探索する。一方で、生きた細胞中のタンパク質や不安定試料のスペクトルは、様々な雑音によりS/N比は大きく低下し、得られる構造情報も限られる。我々は、こうした構造情報がスパースな条件においても、正確な立体構造を構築する新たな計算手法として、ベイズ推定を利用した新しい立体構造計算アルゴリズムを開発した。事前確率は、経験的に求められた分布と、分子のポテンシャルエネルギーから求められるボルツマン分布により推定され、尤度は実験データと任意の時点での分子の位置座標から計算される。蛋白質分子の関数空間は、数千次元を超え、膨大な局所最適解を持つため、効率的な探索を行うために、ハイブリッドモンテカルロとレプリカ交換法を組み合わせた探索手法により事後確率分布を推定した。

生きた細胞内のタンパク質を測定した in-cell NMR データに対して、本手法を適用した(図5)。原核細胞中の TTHA1718 と GB1 タンパク質(図5a,b)と真核細胞中の GB1 タンパク質(図5c)の立体構造を示す。これまで真核細胞中の de novo 蛋白質立体構造決定に成功した例はなく、本計算手法により初めて側鎖原子の位置まで正確に構造決定が可能となった。開発したベイズ推定を用いた NMR 立体構造計算手法は、従来法に比べて、情報量が少なく、質の低いデータにおいても正確な構造決定が可能であることを示した。

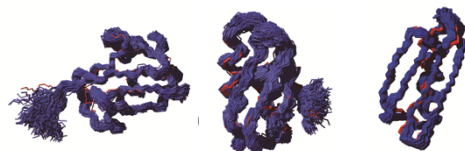


図5: 細胞内タンパク質立体構造

(4) NMR スペクトルはシグナルの集まりであり、多次元シグナルは1次元シグナルの直積と見做せる。SiCode法ではサンプル間のシグナル強度比からアミノ酸情報を得るが、これもベクトルと見做すことができる。そこで、SiCode法のデータを一つのテンソルとして解析することにより、シグナルの分離とアミノ酸情報の取得を同時に実施することが可能となる。この手法は、周波数軸スペクトルだけではなく時間軸観測データを直接扱うことも可能であり、さらに、シグナル緩和測定のための複数の多次元測定を纏めてテンソルと見做せば、分子動態情報を定量的に与える緩和パラメータも同時に抽出することができる。この全く新たな発想に基づき開発した NMR 計測・解析手法を、“Stable-isotope-assisted Parameter extraction (SiPex)”と名付けた。この手法は、NMR 計測・解析一体化した新たな生命分子の構造解析手法である。

5. 主な発表論文等

[雑誌論文] (計 3 件)

- ① Ikeya, T., Hanashima, T., Hosoya, S., Shimazaki, M., Ikeda, S., Mishima, M., Güntert, P. & Ito, Y., Protein NMR Structure Refinement based on Bayesian Inference, Sci. Rep., vol. 6, 2016, 38312 DOI:10.1038/srep38312 (査読有)
- ② Kasai, T., Nagata, K., Okada, M. & Kigawa, T., NMR spectral analysis using prior knowledge, J. Phys.: Conf. Ser., vol. 699, 2016, 012003 DOI:10.1088/1742-6596/699/1/012003 (査読有)
- ③ Kasai, T., Koshihara, S., Yokoyama, J. & Kigawa, T., Stable isotope labeling strategy based on coding theory, J. Biomol. NMR, vol. 63, 2015, 213-221 DOI:10.1007/s10858-015-9978-8 (査読有)

[学会発表] (計 5 件)

- ① 葛西卓磨、タンパク質試料への情報の符号化を用いた NMR 解析法、2018 年電子情報通信学会総合大会、2018 年 3 月 22 日、東京電機大学東京千住キャンパス (東京都・足立区)
- ② Kigawa, T., Protein dynamics in cellular environment analyzed by stable isotope-aided NMR spectroscopy, The 42nd Naito Conference “In the Vanguard of Structural Biology: Revolutionizing Life Sciences”, October 5, 2016, Chateraise Gateaux Kingdom Sapporo (Sapporo, Hokkaido)
- ③ Ikeya, T., Improved in-cell structure determination of proteins at near-physiological concentration, UK/Japan international Symposium, July 4, 2016, Leicester (UK)
- ④ Kigawa, T., NMR Analysis of Protein Structure and Dynamics with Computational and Information Sciences, 29th Annual Symposium of the Protein Society, July 25, 2015, Barcelona (Spain)
- ⑤ Ikeya, T., Three-Dimensional Protein Structure and Dynamics in Living Cells, 7th Korea-Japan Seminars on Biomolecular Sciences, November 11, 2014, Seoul (Korea)

[図書] (計 1 件)

- ① Ikeya, T., Ito, Y., Springer, Experimental Approaches of NMR Spectroscopy, 2017, 63

[産業財産権]

○出願状況 (計 0 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 1 件)

名称：タンパク質解析装置、タンパク質解析方法、および、プログラム
発明者：木川隆則、葛西卓磨
権利者：国立研究開発法人理化学研究所
種類：特許権
番号：6191927
取得年月日：2017 年 8 月 18 日
国内外の別：国内

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

木川 隆則 (Kigawa Takanori)
国立研究開発法人理化学研究所・生命システム研究センター・チームリーダー
研究者番号：20270598

(2) 研究分担者

池谷 鉄兵 (Ikeya Teppei)
首都大学東京・理工学研究科・助教
研究者番号：30457840

(3) 連携研究者

葛西 卓磨 (Kasai Takuma)
国立研究開発法人理化学研究所・生命システム研究センター・研究員
研究者番号：70446516

(4) 研究協力者

()