

# 研究報告書

## 「試料への情報の符号化を活用する NMR 計測・解析法」

研究タイプ: 通常型

研究期間: 2016 年 10 月～2020 年 3 月

研究者: 葛西 卓磨

### 1. 研究のねらい

タンパク質は生命にとって重要な機能を担う分子であり、一定の立体構造をとったり、一部や全体が柔軟に動いたりすることで、その活性や相互作用を巧みに制御している。核磁気共鳴(NMR)法は、高磁場下においた試料の原子核を観測する方法であり、タンパク質を含む生体高分子の立体構造や動態を解析する主要な手法である。特に、溶液状態の試料を測定できる特徴から、細胞内のタンパク質や、一定の立体構造をとらない天然変性タンパク質の解析が期待される。しかし、信号が広幅化して感度が低下すること、信号が重なりやすいこと、さらにこれらの理由により信号の帰属が困難であることから、これらのタンパク質は NMR にとっても難度の高い解析対象である。こうした問題を解決することでより多くのタンパク質を NMR によって解析可能とすることが、本研究の目的である。研究者らは以前、タンパク質主鎖アミド信号の帰属に標準的に用いられる連鎖帰属法が困難な場合の代替あるいは補助法として用いられるアミノ酸選択的安定同位体標識法を、「タンパク質試料にアミノ酸情報を符号化し NMR スペクトルから復号する」との新たなとらえ方のもと改良した「符号化標識法」を開発した。本研究では、符号化標識法を、単にアミノ酸の情報を得るために使うだけではなく、信号ごとの安定同位体標識の違いを信号分離に活用することで、重なった信号の解析を可能とする。NMR スペクトルをテンソルと見なしてテンソル分解の手法を適用することで、信号分離をおこなう。これにより、重なった信号の分離、分離された信号からアミノ酸情報を得るによる信号帰属、分離された信号から構造や動態の情報を得ること、が同時に達成される。

### 2. 研究成果

#### (1) 概要

本法は、符号化標識法と、 $^{15}\text{N}$  緩和測定などのタンパク質の立体構造・動態を解析法に必要な2次元スペクトルを集めて4階のテンソルとして扱い、これをテンソル分解して信号を分離し、各信号からアミノ酸の情報と構造・動態情報を抽出するものである。当初は交互最小二乗法を利用したテンソル分解プログラムを作成し、低分子量タンパク質でその性能を検証した。その結果、人工的に作成した信号重複データや、時間軸データにおいて、想定通り重複信号を分離し、情報抽出できることを確認した。続いて、天然変性タンパク質の  $^{15}\text{N}$  緩和測定に応用した。その過程で、別途の方法で帰属をおこなった場合にはその事前知識を用いてより正しいテンソル分解をおこなう方法を開発した。高難度の解析対象において制約条件や正則化項を付加するため、これらを可能にしつつかつ高速なテンソル分解法を、領域内の共同研究で開発した。

#### (2) 詳細

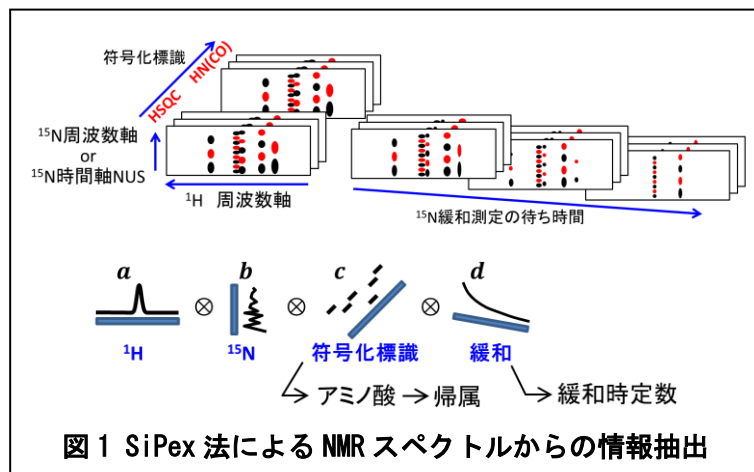
## 研究テーマ A「テンソル分解 NMR 法の基盤技術の確立」

タンパク質の NMR 解析においては、原子核間の結合や距離などの関係を知ることのできる多次元 NMR スペクトルが多用される。中でも、 $^1\text{H}$ - $^{15}\text{N}$  HSQC に代表される  $^1\text{H}$ - $^{15}\text{N}$  2 次元 NMR スペクトルは、各アミノ酸残基のアミド基を観測する、基礎的な測定である。この 2 次元 NMR スペクトルをもとにさらに多次元に展開した複数のスペクトルを組み合わせる解析をおこなうのが通常である。多次元 NMR においては、磁化の時間発展を直接観測するひとつの直接観測軸以外は、パルスシーケンス中の待ち時間を変えて測定する間接観測軸であり、観測時間は間接観測軸の観測点数の積に比例するため、高次元となるほど観測時間が長くなる。アミド基の信号を帰属するために標準的に用いられる連鎖帰属法は、3 次元スペクトルを組み合わせるおこなうが、高難度のタンパク質においては、一部の測定の感度が低く、困難である。

連鎖帰属法によるアミド基帰属が困難である場合には、アミノ酸選択的安定同位体標識法（以下「選択標識法」という）が補助的に、あるいは代替として用いられる。一部のアミノ酸のみを安定同位体標識することによって、各アミド基の信号がどのアミノ酸に由来するかを知ることができ、帰属に役立つ。デュアル選択標識法は、比較的感度の良い  $^1\text{H}$ - $^{15}\text{N}$  HSQC と HN(CO) という 2 種類の 2 次元 NMR スペクトルを用い、 $^{13}\text{C}$  と  $^{15}\text{N}$  の選択標識により、アミド基を含むアミノ酸残基 ( $i$  位) とそのひとつ N 末端側のアミノ酸残基 ( $i-1$  位) の情報を得ることができる、強力な方法であるが、単純なデュアル選択標識ですべてのアミノ酸残基を帰属するには、アミノ酸配列にあらわれるアミノ酸ペアの種類の数だけ標識タンパク質を作らなければならないが、その数が膨大となり現実的でなかった。必要な標識タンパク質数を減らす組み合わせ選択標識法が多くの研究グループから提案されているが、研究者らは、定量的な選択標識を用いて最も少ない標識タンパク質数で網羅的に情報を得ることができる「符号化標識法」を開発した。

本研究で開発した、符号化標識法とテンソル分解を組み合わせたタンパク質解析法 (Stable-isotope-assisted Parameter extraction; SiPex) は、符号化標識法で用いる  $^1\text{H}$ - $^{15}\text{N}$  2 次元 NMR スペクトルを利用する。例えば 3 種類の選択標識タンパク質を用いる場合、HSQC、HN(CO) の 2 種類の NMR スペクトルをそれぞれ 3 種類の選択標識タンパク質ごとに測定しているので、6 枚の 2 次元 NMR スペクトルが存在する。この 2 次元スペクトルの集合は、3 階のテンソルと見なすことができる。本法では、さらに、立体構造や動態を知ることのできる別の NMR スペクトルを組み合わせ

て解析する。例えば、 $^{15}\text{N}$  緩和解析は、 $^1\text{H}$ - $^{15}\text{N}$  2 次元 NMR スペクトルを複数測定し信号の強度変化を観測することで  $^{15}\text{N}$  原子核の磁化緩和の速さを測定するもので、この速さがタンパク質の内部運動を反映することから、タンパク質の動態解析でよく用いられる手



法である。各選択標識タンパク質で緩和測定をおこなうことで、これらの2次元スペクトルの集合を4階のテンソルとみなすことができる。2次元スペクトルの各信号は<sup>1</sup>H軸、<sup>15</sup>N軸のベクトルの直積となっており、符号化標識法の各スペクトル、<sup>15</sup>N緩和測定各スペクトルは信号の強度が変化するのみであるので、4階テンソルはベクトルの直積の和で表現できるはずである。ベクトルの直積の和に分解できれば、符号化標識法の軸はアミノ酸の情報を、<sup>15</sup>N緩和の軸は<sup>15</sup>N原子核の緩和の情報をもつので、信号を分離し、かつ、帰属のための情報と構造・動態の情報がリンクした状態で同時に得られることになる(図1)。

本研究の初期には、交互最小二乗法(ALS)を用いてテンソル分解をおこなうこととし、プログラムを作成した。テンソルをベクトルの直積の和に分解する問題は多峰性の問題であり、局所解を避けるため、乱数で発生させた異なる初期解を用いて最適化をおこない、観測スペクトルとの残差が最も少ないものを選ぶマルチスタート法をとることとした。実NMRデータを用いた検証によって、おおむね数十程度の異なる初期解を用いることで安定的に大域的最適解が得られることがわかった。

低分子量タンパク質であるユビキチン(変異体、77アミノ酸残基、発現用の付加配列を含めて84アミノ酸残基)で本法の検証をおこなった。3種類の選択標識タンパク質を用いてHSQC、HN(CO)、および<sup>15</sup>N緩和測定をおこなった。異なる領域のスペクトルを足し合わせて人工的に信号重複を作成した場合もテンソル分解によって信号を分離でき、アミノ酸の情報を正しく得ることができ、かつ、同時に<sup>15</sup>N緩和情報を得ることができた。<sup>15</sup>N緩和の値は従来法によって得られた値と一致していた(学会発表リスト1~5、論文投稿中)。

本法のもうひとつの特徴は、テンソルの線形にならんの仮定もおいていないため、通常NMRで解析に用いられる周波数軸のスペクトルではなく、生データである時間軸データをそのまま解析できることである。多次元NMRにおける間接観測軸は、測定時間が観測点数の積に比例するため、観測点数を減らすことは測定時間の短縮につながるが、一方でスペクトルの分解能を下げたくない場合には、観測点の一部を省略して(飛び飛びに)観測する方法がとられることがある(不均一サンプリング; non-uniform sampling; NUS)。一部を欠測したデータはそのままでは周波数軸スペクトルにフーリエ変換できないので、圧縮センシング法や最大エントロピー法などでスペクトルを推定することがおこなわれている。しかし一般的には、信号強度を正しく推定することは困難である。本法では、欠測点を推定することなく、時間軸データをそのまま解析できるので、スペクトル推定法の品質に影響されることなく正しい信号強度を用いて解析がおこなえる

のが利点である。ユビキチンを用いて時間軸データの直接解析が可能か検証したところ、もともと64点の<sup>15</sup>N軸観測点があるところ、そのうち8点のみを用いて、5シグナルを正しく分

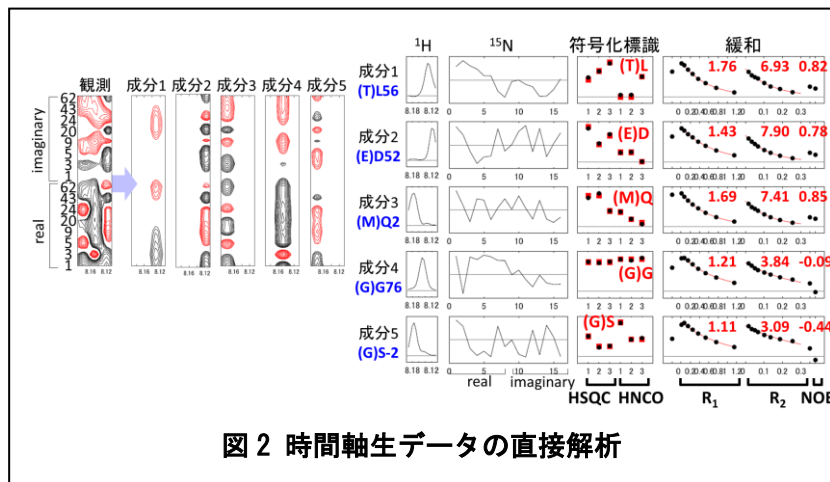


図2 時間軸生データの直接解析

離し、アミノ酸情報が正しく得られ、かつ、従来法と一致した<sup>15</sup>N緩和情報が得られた(図2、論文1、学会発表リスト1~5)。標準的な手法である、均一標識体を用いてすべての間接観測軸観測点を観測して<sup>15</sup>N緩和測定をおこなう方法に比べ、観測点数を1/8に減少させることができた。3種類の標識タンパク質を用いているので、全測定時間を3/8に減少させることができたことになる。単に測定時間を減らしただけでなく、アミノ酸情報という、従来法にはない情報を付加的に得られている点も特筆すべきである。

#### 研究テーマB「天然変性タンパク質への応用」

信号の重複がより問題になる難度の高いタンパク質として、一定の構造をとらずにはたらく天然変性タンパク質X(98アミノ酸残基、発現用の付加配列を含めて99アミノ酸残基)を解析した。天然変性タンパク質は、一定の構造をとらずに溶液中で運動しているので、周囲の化学的環境の影響を受ける共鳴周波数が似通った値となり、信号が重なりやすい。狙い通り、本法ではいくつかの信号重複について、信号を分離し<sup>15</sup>N緩和情報を取り出すことができた(図3)。

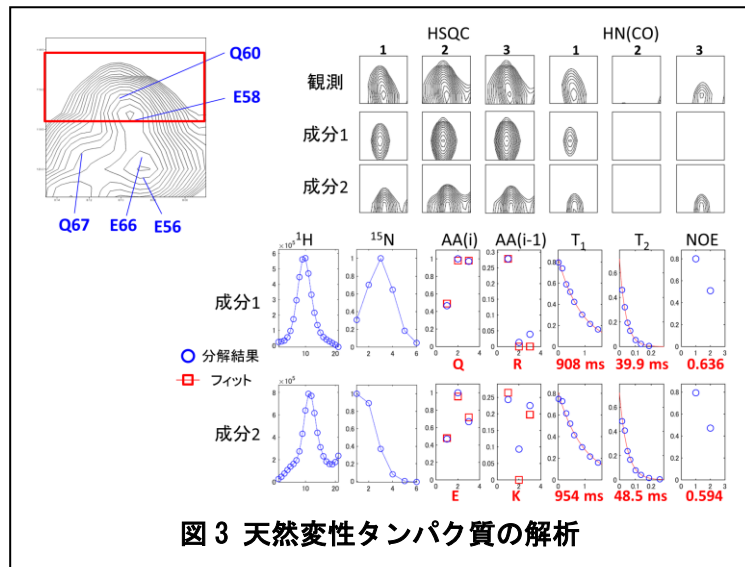


図3 天然変性タンパク質の解析

この天然変性タンパク質Xの解析にあたっては、符号化標識法と連鎖帰属法を組み合わせでアミド基の信号帰属をおこなった。帰属がついている場合には、アミノ酸と安定同位体標識率を既知としてテンソル分解をおこなうほうが、より低感度の場合など難度の高い場合でも正しく信号分離をおこなうことができる。6階のテンソル分解の問題とし、そのうち<sup>15</sup>N標識率、<sup>13</sup>C標識率に対応する2つの軸については既知として固定してテンソル分解をおこなう方法を開発することでこれを実現した。これにより、より多くの重複信号について信号分離<sup>15</sup>N緩和情報を取り出すことができた。

<sup>15</sup>N緩和測定において2次元スペクトル上で重複してしまう信号を分離する方法として、HNCOなどの3次元スペクトルを用いて測定する従来法があるが、時間がかかるのが難点である。2次元測定においては<sup>15</sup>N方向に64点、3次元測定においては<sup>15</sup>N方向に28点、<sup>13</sup>C方向に20点測定すると仮定すると、本法は3次元法に比べ、8倍程度の時間短縮がおこなえることになる。しかし重要なことは、天然変性タンパク質ではそもそもカルボニル炭素の<sup>13</sup>C化学シフトのばらつきが非常に小さく、3次元法ではそもそも信号分離が期待できないことであり、本法のような選択標識を用いた信号分離に頼らざるを得ない。

#### 研究テーマC「高難度タンパク質に適用可能なテンソル分解法の開発」

本研究の初期に用いたALSは、収束が遅いこと、また、非負制約やL<sub>1</sub>正則化などを課そう

とすると多重ループとなり現実的な計算時間で処理できないことが問題であった。そこで、領域の制約や正則化を課さずにテンソル分解していたが、高難度タンパク質を解析対象とした時の低感度の問題など、より厳しい条件下では、適切な制約条件や正則化を課すことが正しいテンソル分解に重要と考えられる。当領域のさきがけ研究者である東京工業大学小野准教授との共同研究で、ALSよりも高速に、かつ制約や正則化を課してテンソル分解をおこなうプログラムを開発した。主双対近接分離法により制約条件・正則化項と観測値一致項を交互に最適化し、その速い収束を活かして内部ループの回数を(経験的に)減らし、外部ループは標準的な交互最適化法とする交互最適化-主双対近接分離法(AO-PDS)によるテンソル分解アルゴリズム部分を小野准教授が作成し、NMR データに対する処理部分を葛西が実装した。これにより、高難度タンパク質のテンソル分解においてさまざまな制約・正則化を試すことができるようになった。

### 3. 今後の展開

本法は、もともと従来法で解析可能なタンパク質をより効率的に解析することと、従来法では困難な解析対象を解析可能にすることの双方に用いることができる。前者については、創薬研究などスピーディーな解析が求められる場面で特に有効と考えられる。符号化標識法については製薬企業などからの問い合わせがあり共同研究に発展した例があるなど興味を持たれているので、その発展形である本法についても訴求していきたい。後者については、試料調製・測定・解析の各段階でのさらなる工夫などより困難な対象を解析できるような技術革新をさらにすすめて、天然変性タンパク質や、細胞内タンパク質に応用していく。

### 4. 自己評価

研究目的に関しては、符号化標識法とテンソル分解を組み合わせたNMR解析法(SiPex)を確立することができた点、低分子量タンパク質でじゅうぶんな性能が得られることを検証できた点、実際の困難な解析対象として天然変性タンパク質を選び、本法の効果が従来法では分離できないような信号を分離して情報が得られた点は評価できると考えている。当初は細胞内タンパク質を最初の対象とすることを想定していたが、天然変性タンパク質も近年注目されていてかつ信号重複が主な解析阻害要因となるタンパク質であるので、本法の解析・デモンストレーション対象としては優れており、対象を切り替えた判断は正しかったと考えている。本課題の研究期間中、スムーズな論文発表ができなかった点が反省点である。はじめに投稿した論文は、重複信号の分離よりもデュアル選択標識による帰属の省略に重きをおいて新しい帰属戦略を提供するものとして書いており、従来法からは飛躍している点が多すぎて本法のメリットが理解されにくかった。新しい帰属戦略を提供することは主張しつつも重複信号の分離という、わかりやすく、かつ分野において具体的に問題になっていることを主題に書き直し、revisionにおいて、天然変性タンパク質のデータを追加することで、アクセプトとなった。

研究実施体制としては、京都大学田中利幸教授、統計数理研究所池田思朗教授のアドバイスを必要に応じてうけつつ、東京工業大学小野准教授と共同で進めた。開始当初はお互いのテクニカルチームがわからずとまどうことがあったが、段々とスムーズに意思疎通できるようになった。この関係が得られたことが、本領域に参加したことの大きな資産であり、今後を活かしていきたい。研究費は、主にタンパク質試料調製用の機器と、安定同位体・試料調製用試薬・

NMR 測定用消耗品に用いたほか、学会発表・論文の投稿料・校閲費などの発表経費、共同研究者との打ち合わせのための旅費に用いた。

本研究は、さきがけ研究期間中には主に手法開発をおこなっており、まだブラッシュアップしなければならない点があるものの、応用としてはこれからが「収穫期」である。学会発表での反応などにおいても、信号重複は多くの NMR 研究者にとって身近でやっかいな問題であり、うまく解消することへのニーズは大きいと感じる。今後、本手法を用いて、天然変性タンパク質や細胞内タンパク質の構造・動態解析を進めていく。

当初、自分一人で計測・情報処理をおこなう計画で当領域に参加したが、小野准教授との共同研究により、専門家の意見として、何が実現可能で、何が(数理的な処理としては)実現困難なのか、はっきりと伺うことができ、それをもとに NMR 研究者側として何ができそうか考えることができた。研究を効率的に進めるうえで、これは非常に貴重な情報であった。戦略目標との関係では、本課題は「データからの特徴量や情報抽出」「生体分子の解析」に該当する研究課題である。戦略目標には必ずしも明記されていないものの、情報と計測の融合には、単に従来の計測データを情報処理するだけでなく、後段での情報処理を前提とした試料調製法・計測法の革新も重要であるというのが、本領域の隠れたキャッチコピーとして「情報×計測」の「掛け算」が掲げられている主旨であると思う。本研究はまさにその点を意識して進めたものであり、情報×計測の融合の一例として見ていただければありがたい。

## 5. 主な研究成果リスト

### (1) 論文(原著論文)発表

1. T. Kasai, S. Ono, S. Koshihara, M. Yamamoto, T. Tanaka, S. Ikeda, T. Kigawa, Amino-acid selective isotope labeling enables simultaneous overlapping signal decomposition and information extraction from NMR spectra, *Journal of Biomolecular NMR*, 74:125-137, 2020.

### (2) 特許出願

研究期間累積件数:0 件

### (3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

主要な学会発表:

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、東京都足立区
2. Shunsuke Ono, Takuma Kasai, “Efficient Constrained Tensor Factorization by Alternating Optimization with Primal-dual Splitting”, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2018), 2018, Calgary, Canada
3. Takuma Kasai, Shunsuke Ono, Toshiyuki Tanaka, Shiro Ikeda, Takanori Kigawa, “Simultaneous assignment and characterization of proteins by integrated analysis of time-domain NMR data”, 28th International Conference on Magnetic Resonance in Biological Systems (ICMRBS), 2018, Dublin, Ireland
4. 葛西卓磨、「タンパク質試料に符号化したアミノ酸情報の NMR スペクトルからの抽出」、日本分析化学会第 67 年会、2018 年、宮城県仙台市

5. 葛西卓磨、小野峻佑、田中利幸、池田思朗、木川隆則、「化学シフト帰属に依存せず時間軸データから直接情報を取り出すタンパク質 NMR 解析法」、第 19 回日本蛋白質科学会年会 第 71 回日本細胞生物学会年会合同年次大会、2019 年、兵庫県神戸市

受賞:

第 56 回 NMR 討論会若手ポスター賞、「in-cell NMR など困難な実験系を対象としたアミノ酸選択的安定同位体標識法」、2017 年

プレスリリース:

葛西卓磨、小野峻佑、木川隆則、「タンパク質の構造や動きを解析する新技術を開発—情報・数理科学の応用による NMR 法の革新—」、理化学研究所・東京工業大学・科学技術振興機構、2020 年 1 月 31 日