

1P10

ニューラルネット、アニーリング法を利用した $^{15}\text{N}/^{13}\text{C}$ 標識タンパク質 NMR シグナル帰属の自動化

○小林直宏¹、木川隆則^{1,2}、横山茂之^{1,3}

1 理化学研究所横浜研究所、ゲノム科学総合研究センター

2 東京工業大学大学院、総合理工学研究科

3 東京大学大学院理学系研究科、生化学

【緒言】 $^{15}\text{N}/^{13}\text{C}$ ラベル化タンパク質を利用したタンパク質の物性研究は近年ほぼ一般化されており、タンパク質の立体構造決定、動力学的性質の解析、薬剤分子との相互作用解析などの研究に広く利用されている。特に有用なタンパク質主鎖シグナルである、 H_N , ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ といったシグナルをアミノ酸配列特異的に連鎖帰属するためのスペクトル実験法が既に確立されている。しかしながら、その帰属解析作業は大部分手動で実施されることが多く、スペクトルデータの質や数、アミノ酸配列数によっては人為的ミスも生じやすい。我々は既に一人の研究者が同時に多くのスペクトルデータ、化学シフトデータを把握しながら整理統合された解析環境を与えるプラットフォームソフトウェア、KIJIRA を完成し発表した[1]。主鎖シグナル解析に関与する NMR スペクトルデータは、一般的に測定時あるいはデータ処理時により生じたノイズ、アーティファクトが多く混入しており、化学交換により数%のシグナルは観測不能くらいブロードニングすることが多く、この問題は特に有用なスペクトルである 3D C(CCO)NNH での解析に大きな影響を与える。本研究では解析プラットフォーム KIJIRA 上で動作する、主鎖シグナル完全自動連鎖帰属機能をもつプラグインモジュールを開発し、ニューラルネット、アニーリングによる手法を応用し効率的な帰属作業の自動化できることに成功したので報告する。

【方法】主鎖シグナルの 2D $^1\text{H}-^{15}\text{N}$ HSQC スペクトル上における位置情報を ID 化する。これにより主鎖シグナル帰属確率行列 A を作成すし、どの ID がどのアミノ酸残基番号に帰属されるかの確率を記述を行う。

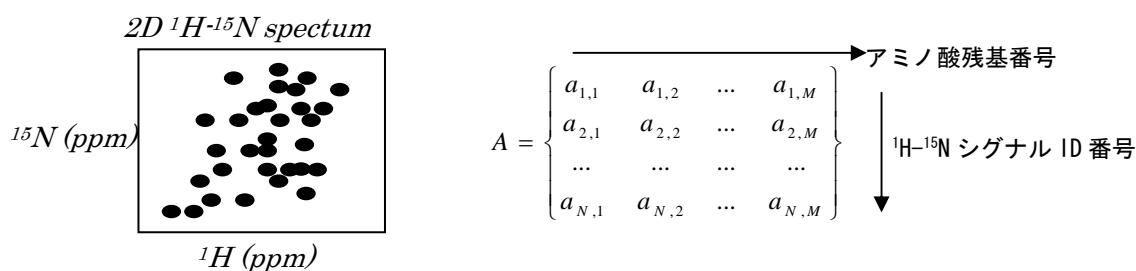


図1 2D $^1\text{H}-^{15}\text{N}$ スペクトルの概念図と主鎖シグナル帰属確率行列

この行列に対して、 (i) および $(i-1)$ 番目のアミノ酸残基タイプ、 $(i-1)-(i)$ あるいは $(i)-(i+1)$ における各シグナルの化学シフト差より計算されるペナルティ関数群 P_x を作成し、アニーリング計算で最も重要な評価関数として用いる。これらの関数に最適化された係数をかけた総和を評価関数とし、以下の式に従って焼きなまし計算(アニーリング)を行う。

$$P_{total} = k_{ID} P_{ID} + k_{RN} P_{RN} + k_{RT} P_{RT} + k_{CN} P_{CN}$$

$$\begin{cases} K = \exp\left\{-\frac{(P_1 - P_0)}{k_B T}\right\} & (P_1 - P_0) > 0.0 \\ K = 1.0 & (P_1 - P_0) \leq 0.0 \end{cases}$$

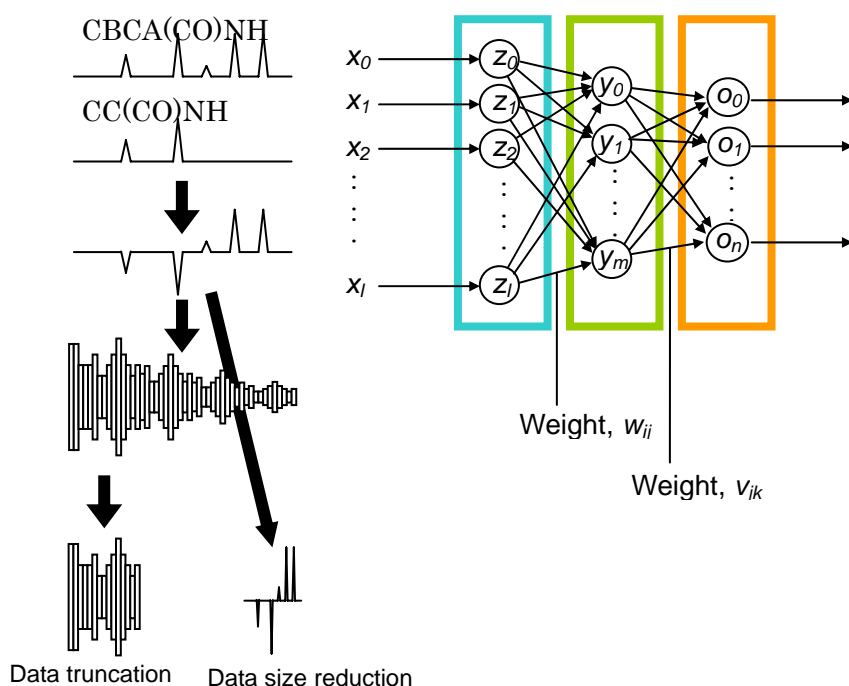


図2 スペクトルデータ加工と使用したニューラルネットの構造

本研究では(i-1)アミノ酸残基タイプ情報を自動的に得る目的として、C(CO)NH, CBCA(CO)NH スペクトルを加工して得られた入力情報を教師データとし、3層ニューラルネットのBPNNにより学習させて得られた重み行列を使ってFNNで認識させる。

【結果】 208 残基のタンパク質についてアニーリングによる自動連鎖帰属計算を行ったところ5%のシグナルの帰属を失敗した。手動で決定した(i-1)残基目のアミノ酸タイプ情報を加えると帰属ミスは0%になり、計算時間も半減した。

トレーニングデータとして6個の独立した100残基程度のタンパク質について540個の主鎖シグナルについて解析した結果を教師データとしてBPNNトレーニングを行った。得られた重み行列を用いて、テストデータとして120残基タンパク質から得られたCCCONNHシグナルの帰属を実行した。帰属すべきシグナル84個の87%が正解として帰属された。また帰属すべきでないノイズ群については約99%の精度で排除することが出来た。また、208残基のタンパク質についてFNNによって自動認識された(i-1)残基目のアミノ酸タイプ情報を加えると帰属のミスは1%程度に留まり、計算時間も半減することが出来た。本研究により、手動帰属にはまだ一歩及ばないが帰属作業の大部分を自動化することに成功したといえる。

参考文献

[1] KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. Kobayashi N, Iwahara J, Koshiba S, Tomizawa T, Tochio N, Guntert P, Kigawa T, Yokoyama S. *J Biomol NMR*. **39**(1):31-52. (2007)