

## NGHIÊN CỨU PHƯƠNG PHÁP LIÊN KẾT ĐỈNH TRONG DỮ LIỆU PHỔ CỘNG HƯỞNG TỪ HẠT NHÂN NMR

Nguyễn Thị Oanh\*, Phạm Thị Liên, Lương Thị Minh Huế, Đào Thị Hằng  
*Trường Đại học Công nghệ thông tin và Truyền thông – ĐH Thái Nguyên*

### TÓM TẮT

Bài viết này trình bày một phương pháp liên kết đỉnh hiệu quả cho dữ liệu phổ cộng hưởng từ hạt nhân (*Nuclear magnetic resonance - NMR*). Phân tích phổ cộng hưởng từ hạt nhân có nhiều tính năng mạnh mẽ, là một công cụ hữu ích cho cả phân tích định lượng và định tính. NMR và các kỹ thuật nhận dạng mẫu là công cụ kết hợp không thể thiếu được sử dụng thường xuyên trong hệ thống sinh học và trong ngành công nghiệp dược phẩm. Hiện nay đã có khá nhiều bài viết của các nhà khoa học trên thế giới đã đề cập, nghiên cứu các phương pháp liên kết cho các dạng dữ liệu tín hiệu trong đó có dữ liệu NMR.

Trong nghiên cứu này chúng tôi đã phân tích các phương pháp đã có, đưa ra nhận xét và cải tiến một phương pháp đạt hiệu quả tốt hơn. Dữ liệu sử dụng trong nghiên cứu là cơ sở dữ liệu phổ cộng hưởng từ thực tế, là mẫu nước tiểu của một bệnh viện. Chúng tôi tiến hành nghiên cứu, thử nghiệm với các mức, các khoảng chia dữ liệu khác nhau. Phương pháp phân tích phổ tín hiệu này đáp ứng được mục tiêu phân tích phổ được nhanh chóng, hiệu quả hơn so với một số phương pháp đã có.

**Từ khóa:** *NMR, cộng hưởng từ hạt nhân, quang phổ, căn chỉnh đỉnh, phân tích đỉnh*

*Ngày nhận bài: 06/12/2018; Ngày hoàn thiện: 24/01/2019; Ngày duyệt đăng: 28/02/2019*

## RESEARCH PEAK ALIGNMENT METHOD FOR NUCLEAR MAGNETIC RESONANCE (NMR) SPECTRA

Nguyen Thi Oanh\*, Pham Thi Lien, Luong Thi Minh Hue, Dao Thi Hang  
*University of Information and Communication Technology - TNU*

### ABSTRACT

This paper presents an efficient peak alignment method for nuclear magnetic resonance (NMR). It has many powerful features, is a useful tool for both quantitative and qualitative analysis. NMR and template recognition techniques are indispensable tools commonly used in biological systems and in the pharmaceutical industry. Currently, there have been many articles of scientists in the world mentioned, researching the peak alignment methods for the types of signal data including NMR data.

In this research we have analyzed existing methods, made remarks and improved a better method. The data used in the study are the actual magnetic resonance spectral database, a urine sample of a hospital. We conduct research, experiment with different levels, data divisions. This peak alignment method satisfies the goal of rapid spectrum analysis, which is more efficient than some existing methods.

**Keywords:** *NMR, nuclear magnetic resonance, spectral, peaks alignment, peak analysis*

*Received: 06/12/2018; Revised: 24/01/2019; Approved: 28/02/2019*

\* Corresponding author: *Tel: 0981 368 808, Email: ntoanh@ictu.edu.vn*

## GIỚI THIỆU

Trong những năm gần đây, nhiều nhà khoa học có xu hướng quan tâm đến khoa học omics trong đó quang phổ cộng hưởng từ hạt nhân (NMR) đóng một vai trò trung tâm. NMR là một kỹ thuật linh hoạt vì nó cung cấp một số lượng lớn các tín hiệu của các phân tử khác nhau trong một phổ NMR có nhiều tính năng mạnh mẽ, làm cho nó trở thành một công cụ hữu ích cho cả phân tích định lượng và định tính, kỹ thuật nhận dạng mẫu và NMR là những công cụ kết hợp không thể thiếu. thường xuyên được sử dụng trong các hệ thống sinh học và trong ngành công nghiệp dược phẩm.

Mặc dù quang phổ NMR là một công cụ phân tích mạnh mẽ cho định dạng chuyển hóa định lượng, một trong những khía cạnh cản trở phân tích vi phân mạnh mẽ là thực tế là tần số cộng hưởng của các đỉnh có thể trải qua sự thay đổi. Một loạt các yếu tố, thường liên quan đến việc kiểm soát hoàn toàn các điều kiện thí nghiệm, góp phần làm thay đổi đỉnh không đồng nhất, bao gồm tương tác hóa lý và sự khác biệt về pH [1] nhiệt độ, ma trận nền hoặc cường độ ion [2]. Khi phân tích các dữ liệu thu thập được hoặc các mẫu phức tạp sẽ khó khăn và tốn thời gian, đó là thách thức đối với các nhà khoa học máy tính với các chương trình thống kê và tính toán sẽ giúp phân tích tự động nhanh chóng và hiệu quả.

Trong các phần tiếp theo trình bày các nội dung: Phần 2 trình bày các phương pháp liên kết đỉnh trong dữ liệu phổ cộng hưởng từ hạt nhân, đưa ra nhận xét ưu nhược điểm, so sánh các phương pháp để tìm thử nghiệm cải tiến các phương pháp đó. Phần 3 giới thiệu cách xây dựng phương pháp liên kết đỉnh trong dữ liệu NMR. Phần 4 trình bày quá trình thử nghiệm phương pháp mới xây dựng và phương pháp đã có với bộ dữ liệu thực tế và kết quả. Cuối cùng là phần kết luận hướng nghiên cứu tiếp theo được trình bày trong phần 5 của bài báo.

## CÁC PHƯƠNG PHÁP PHÂN TÍCH ĐỈNH TRONG DỮ LIỆU PHỔ CỘNG HƯỞNG TỪ HẠT NHÂN

Hiện tại, đã có nhiều phương pháp xử lý vấn đề này và chúng hoạt động tốt cho các tín hiệu có mức nhiễu thấp. Binning là một phương pháp đơn giản và phổ biến trong phân tích dữ liệu quang phổ. Phương pháp này sẽ chia quang phổ thành các xô nhỏ (small buckets) với kích thước lý tưởng là 0,04 ppm, mỗi phần đó sẽ có chứa các biến thể thay đổi đỉnh [3].

Thuật toán đầu tiên được phát triển liên quan đến việc áp dụng thuật toán di truyền để liên kết các phân đoạn của quang phổ [4]. Thuật toán tiếp theo là ứng dụng của một phương trình tuyến tính phù hợp để sắp xếp một phần quang phổ [5]. Ngoài ra còn có phương pháp tìm kiếm trong vùng phổ bù là phương pháp phân tích thành phần chính - Principle Component Analysis (PCA) [6]. Hầu hết các phương pháp này chưa được áp dụng rộng rãi do thiếu hiệu năng liên kết và / hoặc chi phí tính toán cao. Wong et al. [5] giải quyết vấn đề không hiệu quả tính toán bằng cách sử dụng công cụ tương quan Fast Fourier Transformation (FFT) để tăng tốc các thuật toán liên kết đỉnh bằng FFT - Peak Alignment by FFT (PA FFT) và đồng thời sử dụng các khoảng phổ thông thường để được liên kết riêng. Veskelov et al. [7] kết hợp các tính chất của phương pháp chọn đỉnh với FFT và các tính năng khoảng thời gian của PA FFT.

Một trong những giải pháp hiện đại nhất là phương pháp thay đổi tương quan giữa các khoảng tối ưu - interval correlation optimized shifting (icoshift) [8]. Phương pháp này độc lập sắp xếp từng tín hiệu NMR với một đích chọn trước theo tương quan chéo giữa các tín hiệu trong khoảng thời gian do người dùng xác định.

Nguyên tắc cơ bản của icoshift là khá giống với các phương pháp đã công bố khác cho sự liên kết của các tín hiệu quang phổ và sắc ký:

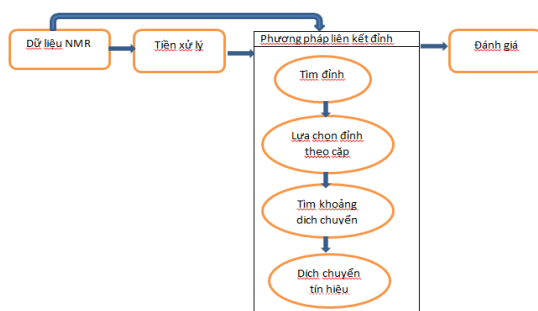
Liên kết đỉnh bằng FFT (PAFFT)[9], Liên kết đỉnh đệ quy theo FFT - Recursive Peak Alignment by FFT (RAFFT)[4] và phân đoạn đỉnh liên kết đệ quy - Recursive Segment-wise Peak Alignment (RSPA)[10].

Thuật toán icoshift dựa trên sự thay đổi tương quan của các khoảng quang phổ và sử dụng thuật toán FFT để căn chỉnh đồng thời tất cả các quang phổ. Thuật toán có thể sử dụng các giá trị thiếu (Not a Number - NaN) như là một thay thế để tránh xuất hiện biến thể của đỉnh tại ranh giới của các phân đoạn tín hiệu. Thuật toán là một công cụ trợ giúp đầy đủ cùng với các thuật toán cũng như là một bản demo có thể làm việc trên một bộ dữ liệu NMR thực [11].

Phương pháp Icoshift là một công cụ linh hoạt cho sự liên kết đỉnh nhanh chóng cho tín hiệu NMR. Tuy nhiên, giống như phần lớn các phương pháp liên kết hiện tại, phương pháp Icoshift không thể sửa đổi thứ tự các đỉnh. Trong khi đó chúng ta đang có một nhu cầu rõ ràng về các phương pháp tính toán để căn chỉnh chính xác các đỉnh tương ứng trên phổ. Nếu các đỉnh được dịch chuyển không đều giữa các phổ khác nhau, chúng sẽ không được kết hợp đúng và phân tích định lượng đơn biến hoặc đa biến về cường độ tín hiệu của chúng có thể bị tổn hại.

### XÂY DỰNG PHƯƠNG PHÁP LIÊN KẾT ĐỈNH MỚI TRONG DỮ LIỆU PHỔ CỘNG HƯỞNG TỬ HẠT NHÂN

Để nghiên cứu một phương pháp mới cho bài toán liên kết đỉnh này, nhóm đề xuất các bước thực hiện như hình 1 dưới đây.



**Hình 1.** Quá trình nghiên cứu phương pháp liên kết đỉnh mới cho phổ NMR

Phương pháp liên kết đỉnh được chia thành 4 bước như sau:

- Xác định các đỉnh,
- Lựa chọn với đỉnh theo cặp,
- Tìm khoảng dịch chuyển
- Chuyển dịch tín hiệu

### Xử lý dữ liệu ban đầu - Preprocessing dataset

Xử lý dữ liệu ban đầu là một bước trung gian giữa dữ liệu quang phổ thô và phân tích dữ liệu. Mục tiêu chính của bước này là chuyển đổi dữ liệu sao cho các mẫu trong tập dữ liệu có thể so sánh được nhiều hơn, dễ dàng hơn và cải thiện việc phân tích dữ liệu.

### Chuẩn hóa dữ liệu – Normalization

Mục đích chính là làm cho tất cả các mẫu có thể so sánh với nhau bằng cách loại bỏ hoặc giảm thiểu tổng lượng vật liệu trên mỗi mẫu hoặc pha loãng chất chuyển hóa. Cách chuẩn hóa điển hình là phép nhân của mỗi hàng (tức là mỗi phổ NMR) theo một hằng số [12]. Hằng số này có thể được tính theo nhiều cách khác nhau. Các phương pháp chuẩn hóa dữ liệu: chuẩn hóa tích phân, chuẩn hóa từng quang phổ riêng lẻ thành cường độ tích hợp tổng không đổi trên toàn bộ cấu hình [13], chuẩn hóa không thể thiếu hay còn gọi là chuẩn hóa tổng không đổi [12]. Trong nghiên cứu này, chúng tôi sẽ giả định một sự chuẩn hóa dựa trên sự thống nhất hoặc dựa trên tính năng mở rộng quy mô. Để chuẩn hóa dữ liệu một cách thống nhất, chúng tôi đưa tất cả các giá trị dữ liệu về giá trị từ 0 đến 1.

Phương trình (1) được sử dụng để thực hiện chuẩn hóa dựa trên sự thống nhất:

$$X_{i_{new}} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Trong đó :

$X_i$  = Điểm dữ liệu thứ  $i$

$X_{min}$  = Giá trị nhỏ nhất trên các điểm dữ liệu

$X_{max}$  = Giá trị lớn nhất trên các điểm dữ liệu

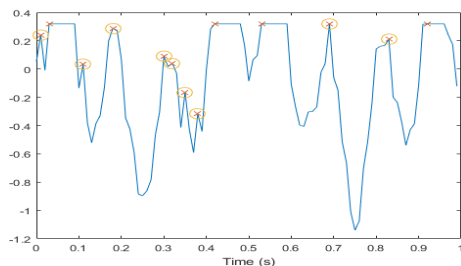
$X_{i_{new}}$  = Điểm dữ liệu thứ  $i$  sau khi chuẩn hóa có giá trị từ 0 đến 1

### Làm mịn dữ liệu - Smoothing dataset

Một số thuật toán làm mịn tín hiệu phổ biến: Bộ lọc trung bình trượt - Moving Average Filter – MAF, thuật toán Savitzky-Golay, Ensemble Average, Applet và các thuật toán khác [14].

### Tìm đỉnh - Find peak

Các giải pháp để xử lý và so sánh quang phổ với ca làm việc bao gồm sự liên kết đỉnh. Trước hết, chúng ta cần phải tìm ra tất cả các đỉnh trong phổ tín hiệu toàn bộ. Một đỉnh được định nghĩa là giá trị cực đại của tín hiệu. Để tránh chọn các artifact (các đỉnh tiếng ồn), chúng ta chỉ xem xét các cực trị có các giá trị trên một ngưỡng nhất định. Đỉnh tìm được có giá trị lớn hơn giá trị đỉnh liền trước và liền sau nó đồng thời giá trị đó phải lớn hơn giá trị ngưỡng đã cho.



**Hình 2.** Các đỉnh của tín hiệu sau khi tìm được

### Sắp xếp đỉnh theo cặp - Match peak

Bước tiếp theo là so khớp các đỉnh sau khi tìm vị trí của chúng. Dữ liệu đầu vào là ma trận vị trí đỉnh được tìm thấy ở bước trước và giá trị ngưỡng là giá trị được sử dụng để so sánh sự khác biệt giữa các vị trí đỉnh khi so khớp chúng theo cặp. Dữ liệu đầu ra của bước này là một ma trận với mỗi cột là vị trí của các đỉnh tương ứng.

### Chọn chuyển dịch cho mỗi tín hiệu – Choose shifting

Trong bước này, hàm sẽ tính toán cách dịch chuyển cho mỗi tín hiệu để căn chỉnh phổ theo mục tiêu đã chọn. Nếu chúng ta có sự lựa chọn tốt của một mục tiêu phù hợp, đó là do đó quan trọng nhất cho một kết quả tốt hơn và không thiên vị. Khi chọn mục tiêu, nó phải

được xem xét rằng cả hình dạng và vị trí của nó (sự thay đổi hóa học) đều quan trọng. Mặc dù trung bình hoặc phổ trung bình thường là lựa chọn tốt, đôi khi (trong trường hợp lệch mạnh), chúng không cung cấp hình dạng tối ưu cho mục tiêu trong mỗi khoảng thời gian và cần lựa chọn cẩn thận một trong những phổ thực nhất.

### Dịch chuyển tín hiệu – Shift signal

Trong bước trước, chúng tôi có các giá trị lưu trữ mảng thay đổi để thay đổi mọi tín hiệu theo tín hiệu đích đã chọn. Để tiếp tục, sử dụng các thuật toán shiftSignal để chuyển tín hiệu bằng các giá trị được lưu trữ trong mảng dịch chuyển. Có tín hiệu sẽ dịch chuyển sang trái với giá trị dịch chuyển là âm và dịch sang phải nếu giá trị là dương. Vì vậy, sau khi thay đổi độ dài của tín hiệu sẽ tăng lên. Theo tính toán, số lượng tín hiệu cột sẽ tăng bằng tổng giá trị tuyệt đối của giá trị dịch chuyển tối thiểu và giá trị dịch chuyển tối đa.

### THỬ NGHIỆM VÀ ĐÁNH GIÁ

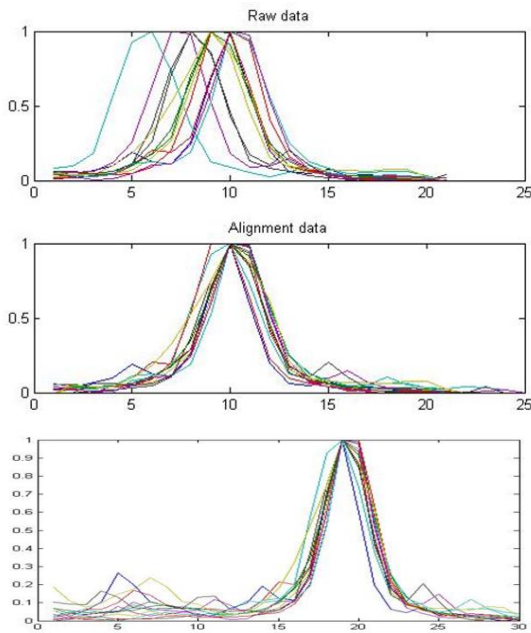
Trong nghiên cứu này, nhóm tác giả sử dụng công cụ MATLAB để minh họa thuật toán. Dữ liệu được chọn thử nghiệm thuật toán là bộ dữ liệu thu thập từ 91 mẫu nước tiểu của bệnh nhân, kích thước 91 mẫu x 4691 biến.

Thử nghiệm sẽ thực hiện trên thuật toán Icoshift và thuật toán mới nhóm nghiên cứu đã xây dựng (Alignment). Mỗi thử nghiệm sẽ chọn nhóm tín hiệu trong một phân đoạn dữ liệu nào đó. Kết quả của các thử nghiệm được trình bày trong hình vẽ, hình đầu tiên là dữ liệu thô, hình thứ 2, 3 là kết quả sau khi lần lượt áp dụng thuật toán Alignment và Icoshift

### Thử nghiệm 1

Chọn dữ liệu sau khi chuẩn hóa từ bộ dữ liệu Urine.mat. Sử dụng một đoạn tín hiệu ngẫu nhiên để kiểm tra.

Áp dụng thuật toán Alignment cho 15 tín hiệu từ đoạn dữ liệu 2320 đến 2350, giá trị ngưỡng là 20 và giá trị nhỏ nhất để tìm đỉnh là 0,7.



**Hình 3.** Kết quả thử nghiệm 1

Kết quả của thử nghiệm 1 được trình bày trong hình vẽ trên, hình đầu tiên là dữ liệu thô, hình thứ 2, 3 là kết quả sau khi lần lượt áp dụng thuật toán Alignment và Icoshift. Có thể thấy trong hình trên từ dữ liệu thô sau khi áp dụng cả hai thuật toán liên kết đỉnh đều cho kết quả khá tốt, các tín hiệu đã hội tụ gần với tín hiệu mẫu và không có sự sai khác quá nhiều giữa hai phương pháp.

**Thử nghiệm 2**

Chọn 50 tín hiệu từ 1 đến 50 trong phân đoạn dữ liệu từ 3330 đến 3350 thể hiện ở hình 4.

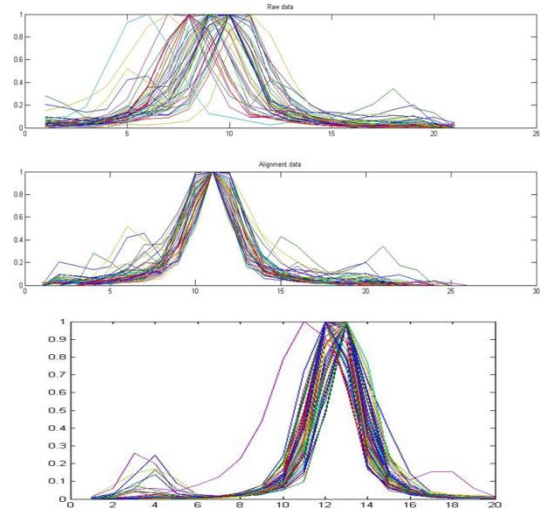
Quan sát kết quả trên ta thấy phương pháp Alignment tốt hơn phương pháp Icoshift trong việc dịch chuyển tín hiệu. Trong kết quả phương pháp Alignment các tín hiệu hội tụ tại đỉnh, còn kết quả phương pháp Icoshift, các tín hiệu mặc dù có sự dịch chuyển nhưng lại chưa hội tụ vào một đỉnh.

**Thử nghiệm 3**

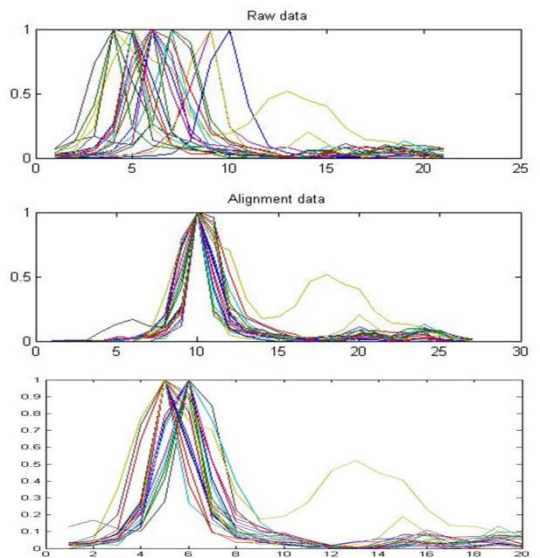
Chọn tập dữ liệu khác cho thử nghiệm này, với 20 tín hiệu từ 1 đến 20 trong phân đoạn từ 3560 đến 3580.

Kết quả thử nghiệm này cho thấy, sự liên kết đỉnh bằng phương pháp Alignment vẫn có kết quả tốt hơn so với Icoshift. Tất cả các đỉnh

trong 20 phổ được căn chỉnh chính xác vị trí. Trong thuật toán icoshift, các tín hiệu đã được dịch chuyển và sắp xếp lại nhưng vị trí đỉnh vẫn không thống nhất, cụ thể được đặt thành hai điểm.



**Hình 4.** Kết quả thử nghiệm 2



**Hình 5.** Kết quả thử nghiệm 3

**\* Đánh giá phương pháp Alignment**

Đánh giá phương pháp dựa trên tỷ lệ đỉnh phù hợp là một phép đo lường từ của phương pháp căn chỉnh đỉnh.

A là ma trận phù hợp với kích thước đỉnh  $n \times m$ , A lưu trữ vị trí tương ứng của các đỉnh, trong đó n là số tín hiệu được sử dụng để khớp, m là các đỉnh số của tín hiệu đích.

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & a_{ij} & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}$$

$$R = \left( \frac{1}{m} \sum_{j=1}^m \frac{r_j}{n} \right) * 100\% \tag{2}$$

*n* là số hàng ma trận

*m* là số cột ma trận

Trong ma trận A,  $a_{ij}$  là phần tử hàng *i*, cột *j* của ma trận và giá trị của nó là vị trí của đỉnh trong tín hiệu *i*. Cột đầu tiên tất cả các phần tử  $a_{11}, \dots, a_{n1}$  có giá trị khác 0 vì chọn tín hiệu đầu tiên làm tín hiệu đích.

$r_j$  là số phần tử khác 0 hoặc số đỉnh khớp với đỉnh *j* trong tín hiệu đầu tiên. Tỷ lệ phù hợp với đỉnh R được tính theo công thức (2):

Bảng dưới đây trình bày tỷ lệ đỉnh phù hợp khi áp dụng phương pháp căn chỉnh đỉnh Alignment. Trong các thí nghiệm chúng tôi sẽ thay đổi các giá trị tham số được sử dụng trong phương pháp căn chỉnh đỉnh: phân đoạn dữ liệu, thứ tự tín hiệu, giá trị ngưỡng, giá trị tối thiểu.

**Bảng 1.** Tỷ lệ căn chỉnh đỉnh khi giá trị ngưỡng thay đổi

Phân đoạn	Số tín hiệu	Giá trị ngưỡng	Giá trị tối thiểu	Tỷ lệ (%)
4100:4200	10:30	<b>20</b>	0,5	94,64
4100:4200	10:30	<b>15</b>	0,5	91,07
4100:4200	10:30	<b>10</b>	0,5	85,12

Từ tỉ lệ thu được ở bảng trên ta thấy thí nghiệm này có thể được coi là thu hẹp giá trị ngưỡng tới khoảng cách tương ứng giữa các đỉnh sẽ giảm tỉ lệ phù hợp đỉnh.

**Bảng 2.** Tỷ lệ căn chỉnh đỉnh khi thay đổi giá trị tối thiểu

Phân đoạn	Số tín hiệu	Giá trị ngưỡng	Giá trị tối thiểu	Tỷ lệ (%)
4100:4200	10:30	20	<b>0,3</b>	98,57%
4100:4200	10:30	20	<b>0,5</b>	94,64%
4100:4200	10:30	20	<b>0,7</b>	52,38%

Kết quả bảng 2, sau khi thay đổi giá trị tối thiểu để tìm đỉnh ta thấy giá trị tối thiểu càng cao thì giá trị tỷ lệ càng thấp.

**Bảng 3.** Tỷ lệ liên kết đỉnh khi tín hiệu phân đoạn thay đổi

Phân đoạn	Số tín hiệu	Giá trị ngưỡng	Giá trị tối thiểu	Tỷ lệ (%)
<b>4100:4200</b>	1:30	20	0,5	90,83
<b>1400:1500</b>	1:30	20	0,5	90
<b>2500:2600</b>	1:30	20	0,5	100
<b>3500:3800</b>	1:30	20	0,5	80

Nhìn vào các bảng trên, ta thấy đánh giá về giá trị của tỷ giá khác nhau tùy thuộc vào giá trị được lựa chọn của các yếu tố tương ứng. Các chỉ số phù hợp với từng phân đoạn tín hiệu sẽ dẫn đến tỷ lệ cao và ngược lại. Việc áp dụng phương pháp căn chỉnh đỉnh này cho tập dữ liệu sẽ đạt được kết quả tốt nếu có các tùy chọn giá trị phù hợp cho mỗi đối số.

**KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO**

Nhóm nghiên cứu đã thực hiện tìm hiểu các phương pháp liên kết đỉnh đã có, đưa ra ưu nhược điểm đồng thời xây dựng một phương pháp liên kết đỉnh mới cho dữ liệu NMR và so sánh với các phương pháp đã có dựa trên việc áp dụng vào cùng một bộ dữ liệu thực tế. Phương pháp mới này khi đánh giá trên các thực nghiệm cho thấy hiệu quả hơn, chính xác hơn các phương pháp đã có khi cùng áp dụng trong bộ dữ liệu thực tế. Chúng tôi dự định sẽ chỉnh sửa và phát triển hơn cho phương pháp của mình đồng thời thử nghiệm nhiều hơn với những bộ dữ liệu thực tế khác để đánh giá và nâng cao chất lượng.

## TÀI LIỆU THAM KHẢO

1. H. Winning, F. H. Larsen, R. Bro, S. B. Engelsen (2008), *Quantitative analysis of NMR spectra with chemometrics*, J. Magn. Reson. 190 (2008), pp. 26–32.
2. C. Z. Gao, R. T. Liu, R. D. Pan and Y. Zheng (2006), *Transactions of the Institute of Metal Finishing*, 84, pp. 5.
3. Beneduci, A.; Chidichimo, G.; Dardo, G.; Pontoni, G. (2011), *Highly routinely reproducible alignment of <sup>1</sup>H NMR spectral peaks of metabolites in huge sets of urines*. *Anal. Chim. Acta*, 685, 186–195.
4. J. Forshed, I. Schuppe-Koistinen, S. P. Jacobsson (2003), *Peak alignment of NMR signals by means of a genetic algorithm*, *Anal. Chim. Acta* 487 (2003), pp. 189–199.
5. J. W. H. Wong, C. Durante, H. M. Cartwright (2005), *Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets*, *Anal. Chem.* 77 (2005), pp. 5655–5661.
6. R. Stoyanova, A. W. Nicholls, J. K. Nicholson, J. C. Lindon, T. R. Brown (2004), *Automatic alignment of individual peaks in large high-resolution spectral data sets*, *J. Magn. Reson.* 170 (2004), pp. 329–335.
7. K. A. Veselkov, J. C. Lindon, T. M. D. Ebbels, D. Crockford, V. V. Volynkin, E. Holmes, D. B. Davies, J. K. Nicholson (2009), *Recursive segment-wise peak alignment of biological H-1 NMR spectra for improved metabolic biomarker recovery*, *Anal. Chem.* 81 (2009), pp. 56–66.
8. F. van den Berg, G. Tomasi, N. Viereck (2005), *Warping: investigation of NMR preprocessing and correction*, in: S.B. Engelsen, P.S. Belton, H.J. Jakobsen (Eds.), *Magnetic Resonance in Food Science: The Multivariate Challenge*, Royal Society of Chemistry, Cambridge, 2005, pp. 131–138.
9. Wong J. W. H., Durante C., Cartwright H. M. (2005), *Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets*. *Anal. Chem.*, 77, pp. 5655–5661.
10. G. C. Lee, D. L. Woodruff (2004), *Beam search for peak alignment of NMR signals*, *Anal. Chim. Acta* 513 (2004), pp. 413–416.
11. F. Savorani, G. Tomasi, S. B. Engelsen (2010), *icoshift: A versatile tool for the rapid alignment of 1D NMR spectra*, *J. Magn. Reson.* (2010) 202, pp. 190–202.
12. Craig A., Cloarec O., Holmes E., Nicholson J. K., Lindon J. C. (2006), *Scaling and normalization effects in NMR spectroscopic metabonomic data sets.*, *Anal Chem* 2, 78(7), pp. 2262–2267.
13. Spraul M., Neidig P., Klauk U., Kessler P., Holmes E., Nicholson J. K., Sweatman B. C., Salman S. R., Farrant R. D., Rahr E., Beddell C. R., Lindon J. C. (1994), *Automatic Reduction of Nmr Spectroscopic Data for Statistical and Pattern-Recognition Classification of Samples*. *Journal of Pharmaceutical and Biomedical Analysis* 1994, 12, (10), pp. 1215–1225.
14. [http://www.chem.uoa.gr/applets/appletsmooth/appl\\_smooth2.html](http://www.chem.uoa.gr/applets/appletsmooth/appl_smooth2.html).