

SEGMENTATION DE PICS D'IMAGES TRIDIMENSIONNELLES
DE RESONANCE MAGNETIQUE NUCLEAIRE DE PROTEINES

Alain Rouh, Gilles Bertrand* , et Jean-Yves Lallemand

Département de Chimie et de Synthèse Organique, Ecole Polytechnique,
91 128 Palaiseau CEDEX, France

* Laboratoire I.A.A.I., E.S.I.E.E. Noisy le Grand, France

RÉSUMÉ

ABSTRACT

La spectroscopie par Résonance Magnétique Nucléaire permet de déterminer la structure spatiale de macro-molécules, afin d'étudier les éventuelles relations entre l'activité biologique de la molécule et sa structure. L'étape décisive de l'étude consiste à identifier et analyser les raies des spectres de la molécule. Cet article présente une méthode de segmentation en deux étapes pour l'extraction automatique du support des pics dans les spectres bi et tri-dimensionnels. Dans un premier temps, le signal est localisé par détection Bayésienne, puis le support exact des pics est déterminé en utilisant un opérateur morphologique d'épaississement, prenant en compte la superposition partielle des pics.

Nuclear Magnetic Resonance spectroscopy is a powerful tool for the determination of the spatial structure of macro-molecules, in order to study the relationship between the biological activity of the molecule and its structure. The decisive step of the study is the identification and the analysis of the lines in the spectra of the molecule. This article presents a two-step segmentation method that enables the automatic extraction of the lines from two and three dimensional spectra. First the signal is localized by a Bayesian detection, then the precise pic extensions are constructed, using a new morphological thickening operator, which takes into account the partially overlap of the lines.

1 INTRODUCTION

L'étude des relations entre la structure spatiale des macro-molécules biologiques et leur activité est l'un des enjeux majeurs de la chimie moderne. La Résonance Magnétique Nucléaire (RMN) consiste à enregistrer des spectres à deux dimensions [1] ou plus [2], dont l'interprétation, basée sur la mesure des pics, permet d'obtenir des contraintes de distance entre les noyaux de la molécule. Ces contraintes sont injectées dans des programmes de modélisation moléculaire qui permettent de déterminer la structure spatiale de la molécule étudiée.

Devant la complexité de l'interprétation des expériences, de nombreux efforts ont porté sur l'extraction automatique de l'information des spectres bi-dimensionnels [3]. Depuis peu les techniques de RMN tri-dimensionnelle se généralisent [2], cependant il n'existe encore que peu d'outils d'analyse automatique.

Nous proposons un algorithme original qui permet d'extraire automatiquement les pics des spectres tri-dimensionnels. Une pré-sélection des régions contenant l'information utile est réalisée par détection statistique du signal. Au sein de ces régions, le sup-

port d'un pic est reconstruit à partir de son sommet par un algorithme de croissance qui garantit la présence d'un pic unique par composante connexe du support, y compris lors de superpositions partielles des pics. Cette propriété est obtenue par l'utilisation d'un nouvel opérateur morphologique d'épaississement, qui préserve la topologie tri-dimensionnelle [4]. La mesure de la position des pics et de leur intégrale est alors réalisée par un algorithme d'extraction de particules [5].

L'algorithme est appliqué en séquence sur des portions de spectre avec gestion des effets de bord, afin d'exploiter une anté-mémoire de blocs qui permet de traiter les expériences tri-dimensionnelles volumineuses.

2 PRINCIPE DE SEGMENTATION

L'information utile dans un spectre est l'ensemble des paramètres caractéristiques des raies du spectre, c'est-à-dire leurs coordonnées et leurs intégrales. Après analyse spectrale par transformée de Fourier, des méthodes de restauration telles que la correction



des distortions de ligne de base [6] sont utilisées afin d'obtenir des données prêtes à l'analyse. Cette analyse peut être abordée par une approche mathématique globale, la prédiction linéaire [7], mais des problèmes numériques et son manque de robustesse dans le cas de la RMN limitent son utilisation.

La formulation de cette analyse en termes de segmentation consiste à identifier le support des différents pics afin d'en estimer les caractéristiques. Il s'agit alors d'obtenir une image binaire qui indique la fonction d'appartenance des points du spectre au support d'un pic. De plus, afin de distinguer les pics partiellement superposés, il est nécessaire d'ajouter une contrainte supplémentaire, basée sur la notion de composante connexe :

Il doit exister une bijection entre les pics du spectre et les composantes connexes de l'image segmentée.

Il est possible de définir trois critères quantifiant la qualité d'une segmentation en RMN. Le premier critère concerne la qualité de l'extraction du signal. En effet le rapport signal à bruit d'un spectre RMN est souvent critique ce qui peut entraîner soit le manque de certaines raies peu intenses, d'où une sous-segmentation, soit la détection intempestive d'un faux pic, d'où une sur-segmentation. Le second critère est la précision des frontières entre les pics et le fond : le bruit et les distortions qui affectent les spectres peuvent occasionner des déformations géométriques des supports déterminés par la segmentation. Enfin le troisième critère est la bonne séparation des supports des pics : l'information des spectres de protéines est très dense, et comporte des pics partiellement ou totalement superposés.

Il est particulièrement difficile de concevoir une méthode de segmentation satisfaisante pour ces trois critères, ce qui conduit à l'élaboration d'une segmentation en deux étapes. Tout d'abord une méthode de détection statistique du signal permet d'obtenir une extraction du signal utile de bonne qualité. Ensuite un traitement morphologique permet de déterminer avec précision les supports des pics, et de résoudre le problème de la superposition partielle.

2.1 DETECTION BAYESIENNE DU SIGNAL

Un modèle déterministe des raies est utilisé (Figure 1) : les raies sont modélisées par un jeu de voisinages élémentaires caractéristiques des différents types de points composant un pic, c'est-à-dire le sommet ou l'un des côtés.

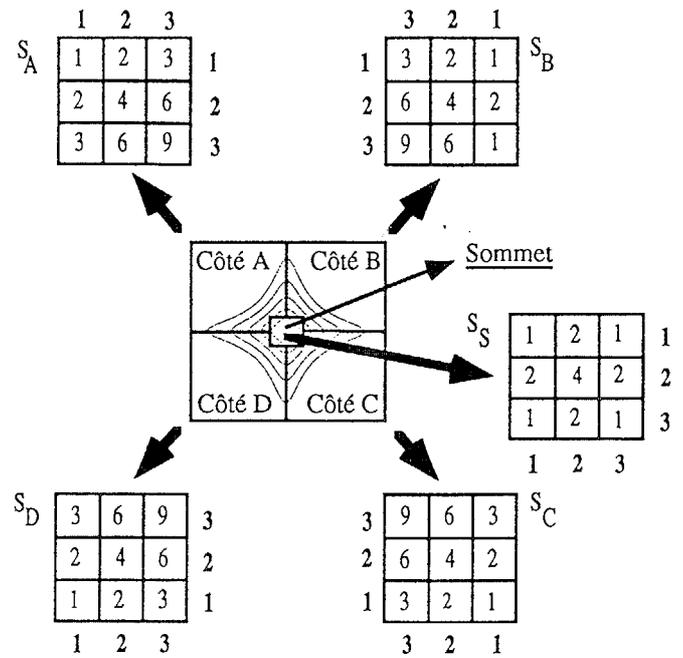


Figure 1: Modèle des configurations locales pour les différents types de points d'une raie lorentzienne bidimensionnelle : S_S le sommet, S_A , S_B , S_C et S_D les quatre types de côtés.

Par ailleurs le bruit peut être supposé stationnaire, blanc, et ayant une distribution de probabilité gaussienne (Voir la Figure 2).

La détection consiste à décider pour chaque point du spectre si l'on est en présence d'une des configurations caractéristiques, ou s'il y a absence de signal. La formulation Bayésienne de ce test [8] combine l'information a priori obtenue par l'interprétation statistique de l'histogramme du spectre et l'information a posteriori obtenue par l'examen du spectre en utilisant les voisinages de référence et le modèle de distribution du bruit. Cette méthode permet d'obtenir une extraction du signal d'excellente qualité, alors que la précision est dégradée par la taille du voisinage utilisé pour tester les configurations caractéristiques. De plus le cas des pics superposés n'est pas pris en compte, et une seconde étape de traitement morphologique est nécessaire afin d'obtenir les supports recherchés.

2.2 TRAITEMENT MORPHOLOGIQUE DU SIGNAL UTILE

Le traitement morphologique itératif consiste à construire la solution X_i à l'étape i par un mécanisme de croissance de X_{i-1} , au sein de la région D pré-sélectionnée par la détection Bayésienne. Le point de départ de la construction est le sommet des pics. La bonne correspondance entre les composantes ainsi

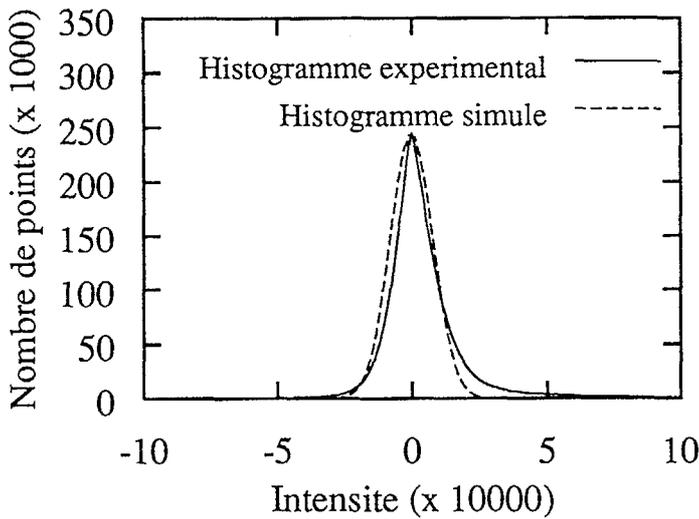


Figure 2: Comparaison entre l'histogramme expérimental d'un spectre tri-dimensionnel de la Capsicéine, et l'histogramme simulé en supposant une distribution gaussienne du bruit.

obtenues et les supports des pics repose sur l'hypothèse d'unicité du sommet pour chaque pic, et l'homotopie du mécanisme de croissance.

Un nouvel opérateur morphologique d'épaississement binaire homotopique en deux et trois dimensions est utilisé [4]. L'opérateur repose sur une caractérisation nécessaire et suffisante des points qui peuvent être ajoutés à une image sans modifier son homotopie.

L'image initiale des supports, notée X_0 , est formée par les sommets des pics, identifiés en recherchant les maxima locaux M_0 du spectre sur un voisinage élémentaire : $X_0 = M_0 \cap D$. L'intensité de ces points est mise à zéro dans le spectre. A chaque itération i , l'ensemble des points candidats à la croissance des supports, noté C_i , est formé par la dilatation homotopique de X_i , ce qui permet de garantir que deux composantes identifiées par deux sommets différents resteront distinctes au cours du processus.

L'image tri-dimensionnelle est définie sur une grille à maille cubique, et la dilatation considérée concerne le voisinage élémentaire formé de 26 points. Ce voisinage est formé des 6 points les plus proches appelés les 6-voisins, puis des 12 points les plus proches appelés les 18-voisins, et enfin des 8 angles du cube élémentaire qui sont appelés les 26-voisins. L'opérateur d'épaississement homotopique concerne les 6-voisins, et trois itérations sont nécessaires afin de générer les chemins éventuels vers les 26-voisins. Après le second et le troisième épaississement, l'intersection avec la dilatation de X permet de créer une transformation qui n'affecte que le voisinage élémentaire.

Soient B l'élément structurant unité, $X_i \oplus B$ le dilaté de X_i par B , et $X_i \odot B$ l'épaississement homotopique de X_i par B , la dilatation homotopique de X_i par B est donc définie par :

$$C_i = \{[(X_i \odot B) \odot B] \cap (X_i \oplus B)\} \cap (X_i \oplus B)$$

Soit M_i l'ensemble des maxima locaux du spectre sur un voisinage élémentaire, l'intensité des points de $M_i \cap [(X_i \oplus B)/C_i]$, où $/$ représente la différence ensembliste, est mise à zéro dans le spectre afin de ne pas bloquer le mécanisme de reconstruction. En effet ces points sont des points du support d'un pic, mais ne sont pas ajoutés à X_i pour préserver son homotopie. Par contre les points de $M_i \cap C_i$ sont ajoutés à X_i pour former X_{i+1} , et leur intensité est également mise à zéro dans le spectre.

La recherche des maxima locaux successifs M_i garantit la précision du support final en utilisant le critère de décroissance de l'intensité des points d'un pic à partir de son sommet, alors que la construction de l'ensemble des points candidats C_i en partant du sommet et en utilisant un épaississement homotopique binaire assure la séparation des raies superposées.

3 DISCUSSION

Un algorithme morphologique d'extraction de particules [5] est ensuite utilisé pour identifier les composantes connexes de la segmentation afin d'estimer les paramètres des pics associés à chaque composante.

Cet algorithme a été appliqué au spectre tri-dimensionnel de la Capsicéine. La Figure 3.A est une représentation en courbes de niveaux d'une section. Les contours des supports des pics extraits sont présentés Figure 3.B où les croix représentent les sommets trouvés dans la section, et les lettres numérotées les noyaux correspondants aux pics obtenus par interprétation manuelle du spectre. Les supports des pics extraits sont corrects, en particulier les deux pics étiquetés Y_{12-A} et Y_{12-B} sont dissociés. Les sommets trouvés dans ce plan sont ceux identifiés lors de l'analyse complète de la Capsicéine, les supports sans sommet tels que ceux du noyau T_{94} correspondent à un pic tri-dimensionnel dont le sommet se trouve dans un plan voisin. Ces résultats ont été obtenus en 4:28':46" sur une station de travail Silicon Graphics 4D30 pour un spectre de $64*256*256$ points avec 256 niveaux de gris.

4 CONCLUSION

Un algorithme original de segmentation a été présenté pour l'analyse automatique des spectres bi et tri-

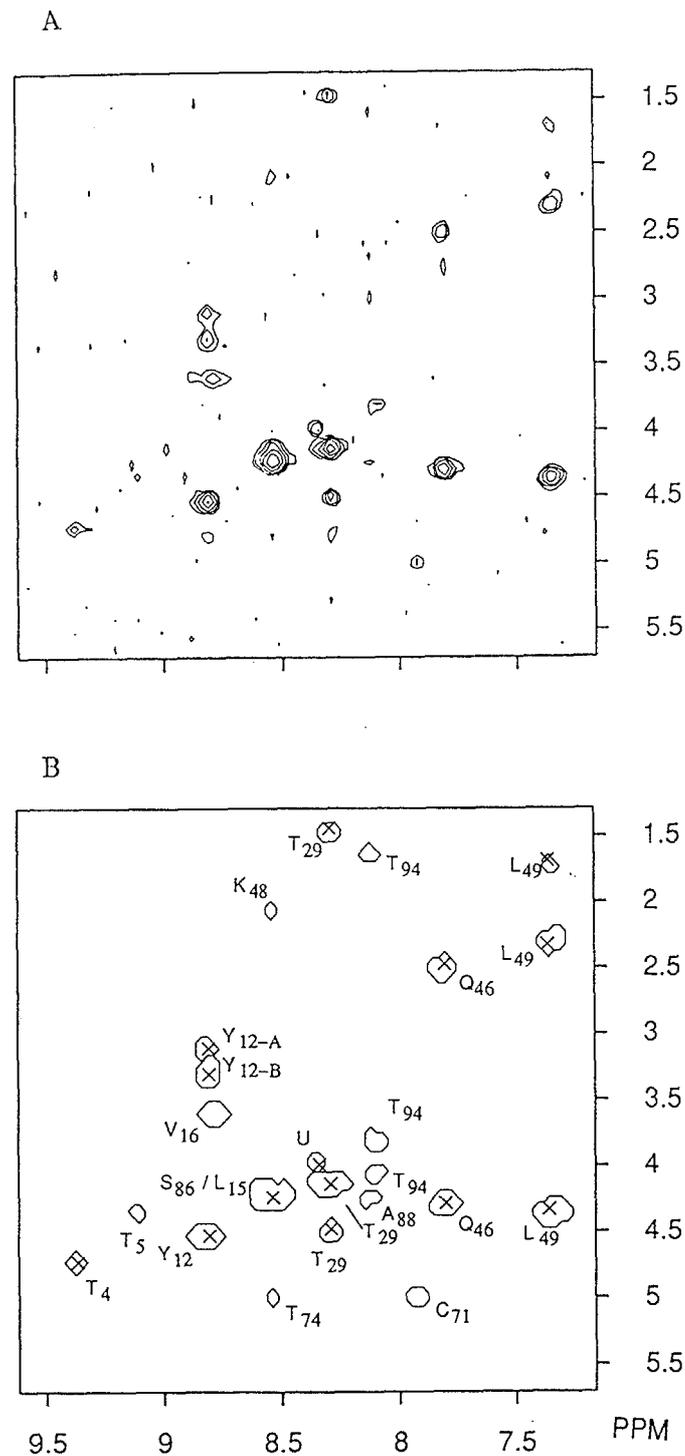


Figure 3: A) Représentation en courbes de niveaux d'une section du spectre tri-dimensionnel de la capscicine, B) Contours des supports des pics extraits ; les croix représentent les sommets des pics trouvés dans la section.

dimensionnels de RMN. Cet algorithme combine la qualité de l'extraction du signal par détection Bayésienne avec une bonne précision dans la localisation des frontières des supports des pics, y compris en cas de superposition partielle. Cette qualité indispensable a été obtenue par une approche morphologique utilisant un nouvel opérateur d'épaississement homotopique binaire en trois dimensions.

Références

- [1] K. WÜTHRICH, "NMR of Proteins and Nucleic Acids", Wiley, New York (1986)
- [2] E. GUITTET, "Application de la R.M.N. tridimensionnelle à l'étude des protéines : quelques développements récents.", GERM XII, La Colle sur Loup, 10-15 Mars 1991.
- [3] G. J. KLEYWEGT, R. BOELEN, M. COX, M. LEINAS AND R. KAPTEIN, "Computer-assisted assignment of 2D 1H NMR Spectra of Proteins : Basic Algorithms and Application to Pharatoxin B.", *Journal of Biomolecular NMR* 1, 23-47 (1991)
- [4] G. BERTRAND, G. MALANDAIN, "A new topological classification of points in 3D images", *2nd European Conf. Computer Vision*, pp 710-714, (1992)
- [5] J. SERRA, "Image Analysis and Mathematical Morphologie", Volume I and II, Academic Press, New-York (1982-1988).
- [6] A. ROUH, G. BERTRAND, AND J.Y. LALLEMAND, "Baseline Distortion Correction of Nuclear Magnetic Resonance Spectra Using Mathematical Morphology", Workshop "Mathematical Morphology and its Applications to Signal Processing", Barcelona, (Mai 1993).
- [7] T. E. MALLIAVIN, "Problèmes inverses en RMN : application à la détermination de la structure des protéines", Thèse, 1992
- [8] R.O. DUDA, P.E. HART, "Pattern Classification and Scene Analysis", Wiley, New York (1973)
- [9] A. ROSENFELD AND A.C. KAK, "Digital Picture Processing", Academic Press, Inc., Orlando, (1982)