HAL
open science

Etude structurale des protéines et des acides nucléiques par RMN. Etude de la répression du gène de la beta-lactamase chez B. licheniformis 749/I. Augmentation de la résolution des spectres RMN multidimensionnels par filtrage Hadamard.

Hélène van Melckebeke

To cite this version:

Hélène van Melckebeke. Etude structurale des protéines et des acides nucléiques par RMN. Etude de la répression du gène de la beta-lactamase chez B. licheniformis 749/I. Augmentation de la résolution des spectres RMN multidimensionnels par filtrage Hadamard.. Sciences pharmaceutiques. Université Joseph-Fourier - Grenoble I, 2005. Français. NNT : . tel-00011563

HAL Id: tel-00011563

https://theses.hal.science/tel-00011563

Submitted on 8 Feb 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

Présentée pour obtenir le titre de

Docteur de l'Université Joseph Fourier – Grenoble 1

Spécialité : PHYSIQUE

par

Hélène VAN MELCKEBEKE

Etude structurale des protéines et des acides nucléiques par RMN :

Etude de la répression du gène de la β-lactamase chez *B. licheniformis 749/I.*

Augmentation de la résolution des spectres RMN multidimensionnels par filtrage Hadamard.

Soutenue le 29 septembre 2005, devant le jury:

Rapporteur	Eric GUITTET
Rapporteur	Michel KOCHOYAN
Examinateur	Bernard JORIS
Examinateur	Beat MEIER
Présidente	Eva PEBAY-PEYROULA
Directeur de thèse	Jean-Pierre SIMORRE

Préparée au Laboratoire de Résonance Magnétique Nucléaire

de l'Institut de Biologie Structurale Jean-Pierre Ebel (CNRS-CEA-UJF)

Remerciements

Je voudrais tout d'abord remercier Michel Kochoyan et Eric Guittet d'avoir accepté de porter leur jugement sur ce travail de thèse. C'est également un grand honneur et un grand plaisir de compter Eva Pebay-Peyroula, Bernard Joris, et Beat Meier parmi les membres de mon jury. Pour la relecture de ce document, je remercie vivement Jean-Pierre et Bernhard, Christelle et Paul, Dominique, Julien, Célia, Nicolas et Jean.

Ce manuscrit est l'occasion pour moi de faire le bilan des quatre année que j'ai passées à l'Institut de Biologie Structurale. Elles ont été pour moi une véritable source d'épanouissement professionnel et personnel. J'ai pu apprécier la solide formation scientifique qu'on m'a offerte, la qualité et la variété du travail de recherche auquel j'ai participé. C'est également avec beaucoup de plaisir que j'ai pu côtoyer tous les jours l'équipe du LRMN, et je quitterai avec regret l'ambiance motivée et détendue du laboratoire. Je tiens donc à remercier toutes les personnes qui ont contribué à ma formation scientifique durant ces trois années, et qui ont rendu possible ce travail de thèse dans des conditions très agréables:

Je remercie Jean-Pierre Simorre et l'Institut de Biologie Structurale de Grenoble de m'avoir accueillie au sein du laboratoire de RMN, ainsi que le CEA de m'avoir financée pendant trois ans. Ces trois années ont comporté une partie d'enseignement qui m'a également beaucoup apporté au niveau scientifique et pédagogique. Je remercie donc les étudiants de l'UJF et les collègues de l'UFR de physique et de chimie avec qui j'ai travaillé.

Pour la partie de ma thèse concernant la résistance aux antibiotiques, j'aimerais souligner le grand plaisir que j'ai eu à collaborer avec l'équipe de Bernard Joris de Liège, en particulier Christelle Vreuls. Ce travail fait en commun restera pour moi un modèle de collaboration interactive, où j'ai beaucoup appris en matière de biochimie. Bernard, merci pour ton accueil et tes conseils. Christelle, ce fut une grande joie de pouvoir régulièrement partager avec toi des expériences de biochimie et de RMN, des mails pleins de questions et de réponses, des petites soirées chez l'une ou l'autre. Merci pour ton amitié et ta disponibilité ! Du côté grenoblois, un grand merci à Jean-Pierre Simorre et à Pierre Gans de m'avoir soutenue et aidée tout au long de ce projet.

Une deuxième partie du travail que j'ai effectué durant ces trois années concerne l'étude d'ARN par RMN. Même si tous ces projets ne sont pas décrits en détail dans ce manuscrit, ils ont été importants dans la maturité que j'ai acquise pour l'étude des ARN par RMN. Pour les projets concernant le virus de l'hépatite C, je remercie vivement François Penin de l'IBCP de Lyon pour la confiance qu'il m'a accordée. Je remercie également Jean-Pierre Simorre, Bernhard Brutscher et Jérôme Boisbouvier pour leurs conseils avisés sur la manipulation biochimique et spectroscopique des ARN. Enfin, je remercie notre collaborateur Arthur Pardi pour nous avoir fourni les échantillons dont nous avions besoin pour l'étude méthodologique.

Ensuite, je voudrais remercier tous les membres de l'IBS avec qui j'ai eu le plaisir de discuter de sujets scientifiques divers. En RMN, je regretterai les discussions philosophiques sur les principes fondamentaux de physique que j'ai pu avoir avec Dominique Marion, Bernhard Brutscher et Christophe Deprez, les conseils de biochimie de Pierre Gans, Julien Boudet, Denis Bouvier et Monika Nalezkova, les heures passées au spectromètre aux côtés de Jean-Pierre Simorre et Bernhard Brutscher. Merci à Paul Schanda, Pierre Gans et Catherine Bougault pour leurs conseils nombreux et avisés, à Beat Meier, Klaartje Houben et Jérôme Boisbouvier pour leur aide lors des demandes de financement. Un grand merci à tous les autres pour les petits détails qui rendant la vie plus facile et plus agréable : Florence, Beate, Adrien, Nicolas, Guillaume, Phineus, Martin, Pau, Rémy, Laurence, Isabel, et Anne. Un grand merci aussi aux autres personnes de l'institut qui ont partagé avec moi un peu de cristallographie : Mickael Cherrier, Meike Stelter, et Jacques Colletier ; de microscopie électronique : Grégory Effantin, et merci aux collègues de spectrométrie de masse : David Lemaire et David Lascoux.

Enfin, j'aimerais ajouter une attention toute particulière pour mes deux guides scientifiques : Jean-Pierre Simorre et à Bernhard Brutscher. Merci pour votre encadrement et votre confiance, je n'aurais pas rêvé mieux pour pouvoir apprendre en profondeur des aspects aussi bien appliqués que méthodologiques en spectroscopie RMN. Jean-Pierre, un grand merci de m'avoir laissé la liberté de choisir ma voie au sein de la RMN, pour tout ce que tu m'as appris, pour ton écoute, et pour l'équilibre que tu fais régner autour de toi. Bernhard, merci pour tes conseils scientifiques et personnels, je suis sure qu'ils me suivront au-delà de ces trois années.

Abréviations

1D, 2D, 3D, nD	Spectre à 1, 2, 3, n dimensions
ADN	Acide désoxyribo nucléique
ANS	Acide 8-anilino-1-naphtalène-sulfonique
ARN	Acide ribonucléique
B. licheniformis	Bacillus licheniformis
Bis-ANS	Acide 4,4'-dianilino-1,1'-binaphtyl-5,5'-disulfonique
BMRB	BioMagResBank (http://www.bmrb.wisc.edu/)
NTD	Domaine N-terminal
CTD	Domaine C-terminal
CD	Dichroïsme circulaire
CHAPS	3-((3-cholamidopropyl)diméthylammonio)-1-propane-sulfonate
COSY	Correlated spectroscopy
CSA	Anisotropie du déplacement chimique
CSI	Index des déplacements chimiques
CT	Expérience à temps constant
DNP	Polarisation nucléaire dynamique
DSP	Dithiobis(succinimidylpropionate)
DSS	2,2-diméthyl-2-silapentane-5-sulfonate de sodium
EDTA	Ethylène diamine tétra-acétate
FDM	Filter diagonalization method
GdnHCl	Chlorure de guanidinium
GFT	G-matrix Fourier transformation
NAG	N-acétylglucosamine
NAM	Acide N-acétylmuramique
Hepes	Acide N-2-hydroxyéthyl)pipérazine-N'-(2-éthanesulfonique)
HSQC	Heteronuclear single quantum coherence
IBS	Institut de biologie structurale Jean-Pierre Ebel, Grenoble
INEPT	Insensitive nuclei polarization enhancement
IR	Infra-rouge
IRM	Imagerie par Résonance Magnétique
MALDI-TOF	Matrix assisted laser desorption ionization – Time of flight
NOE	Effet Overhauser nucléaire
NOESY	Nuclear Overhauser spectrocopy
PAGE	Polyacrylamide gel electrophoresis
pb	Paire de base
PBP	Penicillin binding protein (http://www.rcsb.org/pdb)
PDB	Protein data bank
pI	Point isoélectrique
rf	Radio fréquence
RMN	Résonance Magnétique Nucléaire
rmsd	Root mean square deviation
S. aureus	Staphylococcus aureus
SDS	Sodium dodécyl sulfate
SW	Largeur spectrale
TF	Transformée de Fourier
TOCSY	Total correlation spectroscopy
TROSY	Transverse relaxation optimized spectroscopy
UV	Ultra-violet
WHP	Winged helix protein

Sommaire

Remerciements, abréviations

Sommaire

Introduction générale ***p. 1-5***

Etude structurale de la répression du gène de la β-lactamase chez B. licheniformis 749/I ***p. 6-61***

Introduction: Résistance des bactéries aux antibiotiques de type β-lactame

I. Mode d'action des antibiotiques à noyau β–lactame 6

I. 1. La chimiothérapie anti-bactérienne 6

I. 2. Le peptidoglycane 8

I. 3. Mode d'action des antibiotiques de type β–lactame 10

II. Résistance des bactéries aux antibiotiques à noyau β–lactame 12

II. 1. Un phénomène inéluctable et de grande ampleur 12

II. 2. Mécanismes de résistance aux antibiobiotiques à noyau β-lactame 14

III. Régulation des gènes de résistance chez B. licheniformis et S. aureus 17

III. 1. Les systèmes de régulation 17

III. 2. Acteurs protéiques 20

III. 3. Mécanismes de répression et d'induction 23

IV. Objectifs 26

Résultats : Etude structurale du répresseur BlaI du gène de la β-lactamase chez *B. licheniformis* 749/I

I. Production et caractérisation des échantillons protéiques.....27

I. 1. La protéine BlaI sauvage 27

I. 2. Les deux domaines fonctionnels de BlaI..... 28

I. 3. Mutagenèse dirigée 30

II. Etude structurale du répresseur BlaI.....31

II. 1. Structure de BlaI-NTD 31

II. 2. Les mystères du domaine de dimérisation C-terminal de BlaI 39

III. Caractérisation par RMN de l'interaction de BlaI avec l'ADN.....44

III. 1. Matériel et méthode..... 44

III. 2. Caractérisation structurale de l'interaction..... 47

III. 3. Détermination de la constante de dissociation BlaI-NTD / demi-opérateur 52

Discussion et perspectives: Répression et induction des gènes de résistance chez *B. licheniformis* 749/I et *S. aureus*

I. Structure des répresseurs BlaI et MecI _____ 54

I. 1. Domaine de liaison à l'ADN 54

I. 2. Domaine de dimérisation 55

I. 3. Structure quaternaire : la pince à ADN 56

II. Interaction des répresseurs avec l'ADN _____ 57

II. 1. Interaction du domaine de liaison avec l'ADN 57

II. 2. Modification de la pince..... 58

II. 3. Organisation des régions intergéniques 59

III. Bilan et conclusion sur l'induction_____ 60

III. 1. Bilan structural 60

III. 2. Conclusion et perspectives..... 60

Augmentation de la résolution des spectres RMN multidimensionnels des protéines et des acides nucléiques par filtrage Hadamard *p. 62-128*

La spectroscopie Hadamard dans le contexte de l'acquisition rapide des spectres mutidimensionnels en RMN

I. Spectroscopie RMN des macromolécules biologiques.....62

I. 1. La RMN: une technique spectroscopique 62

I. 2. Sensibilité des spectres..... 63

I. 3. Résolution spectrale 65

II. Principe des techniques d'acquisition des spectres RMN 69

II. 1. Principe général 69

II. 2. RMN par onde continue 71

II. 3. Spectroscopie Hadamard dans le domaine des fréquences 73

II. 4. La révolution de la RMN impulsionnelle par TF 76

III. Spectroscopie par TF et spectroscopie Hadamard : concurrentes ou complémentaires? 78

III. 1. Fondements théoriques..... 78

III. 2. Spectroscopie par TF et problème du temps expérimental..... 80

III. 3. Implémentation pratique de la spectroscopie Hadamard 86

III. 4. Bilan..... 94

Applications du filtrage Hadamard
pour les protéines et les acides nucléiques

I. Filtrage des corrélations ^{1}H-^{13}C des méthyles selon le type d'acide aminé 96

I. 1. Attribution des corrélations ^{1}H-^{13}C des méthyles 96

I. 2. Application à la recherche de contraintes structurales de type NOE 106

II. Filtrage des corrélations selon le type de base dans l'expérience HCN.....116

II. 1. Importance de l'expérience HCN pour l'attribution des ARN. 116

II. 2. Résultats. 118

III. Bilan et perspectives.....128

Annexes ***p. 129-156***

Annexe 1 …….. 129
Acides aminés et nucléotides : biochimie et données spectroscopiques

Annexe 2 …….. 133
Tableau des antibiotiques couramment utilisés.

Annexe 3 …….. 135
Article : Van Melckebeke *et al.*, 2003

Annexe 4 …….. 145
Article : Vreuls *et al.*, 2004

Bibliographie ***p. 157-168***

Résumé ***Couverture***

Introduction générale

Les protéines et les acides nucléiques sont des macromolécules essentielles à la vie et à la reproduction de tous les êtres vivants. Leurs propriétés et leurs fonctions au sein des cellules sont liées à leur composition chimique (Annexe 1). Les acides nucléiques sont des chaînes linéaires formées de la répétition de quatre motifs de base appelés nucléotides, qui sont légèrement différents dans le cas des ADN (Acides DésoxyriboNucléiques), ou des ARN (Acides RiboNucléiques). Les protéines sont également des polymères, constitués d'une chaîne linéaire faite à partir de vingt acides aminés.

Malgré le faible nombre de motifs de base, les protéines et les acides nucléiques peuvent avoir des fonctions biologiques très variées. Ces fonctions dépendent en grande partie de l'enchaînement des motifs de base au sein de la molécule, qui conditionne une structuration dans une ou plusieurs conformations privilégiées. La stabilisation énergétique induite par les interactions qui existent entre les groupements chimiques des motifs de base est en général assez faible, ce qui permet aux macromolécules biologiques de répondre rapidement aux changements d'environnement (pH, température, force ionique), et d'interagir entre elles ou avec des ligands. L'étude de la structure tridimensionnelle des macromolécules biologiques, de leur dynamique et des interactions qu'elles peuvent avoir entre elles ou avec des ligands est indispensable pour avoir une meilleure compréhension du fonctionnement des êtres vivants.

Les macromolécules que nous étudions ont une taille de l'ordre de 10 nm (Figure 1), ce qui fait d'elles des objets trop gros pour être étudiés actuellement par des calculs de mécanique quantique. De plus, la stabilisation thermodynamique très faible de leur structuration rend une approche *ab initio* délicate. Ils sont par ailleurs trop petits pour être étudiés avec de simples microscopes optiques.

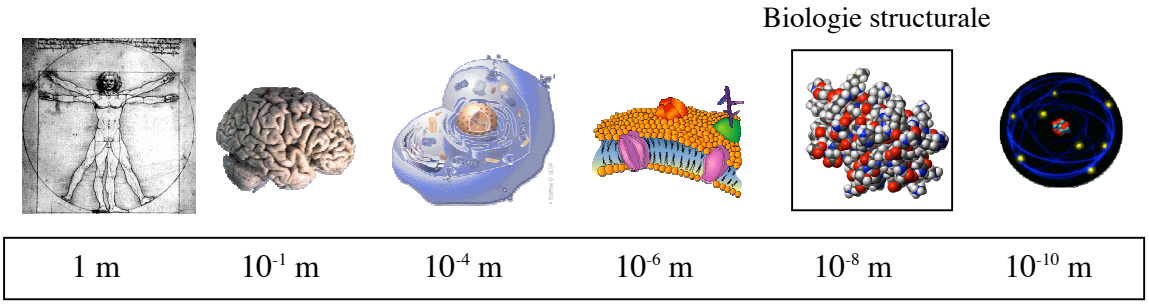


Figure 1: Différentes échelles auxquelles on peut étudier le monde du vivant : les espèces, les organes, les cellules, les organites cellulaires, les molécules, les atomes. La biologie structurale étudie la structure et la dynamique des macromolécules du vivant.

Les propriétés des macromolécules biologiques sont donc étudiées de manière indirecte en analysant la manière dont elles perturbent un rayonnement incident. Ce rayonnement peut être de nature électromagnétique (photon) ou particulaire (électrons, neutrons). Selon la longueur d'onde du rayonnement qu'on utilise, on peut exploiter différentes formes d'interaction matière-rayonnement (Figure 2). Dans les méthodes de diffraction, l'étude de la répartition spatiale de l'intensité diffusée permet de caractériser l'organisation de la matière. Souvent, des cristaux de molécules sont utilisés pour augmenter la sensibilité de la méthode. Dans les méthodes spectrométriques, les atomes absorbent une partie des radiations incidentes et en réémettent d'autres. Les spectres d'absorption et d'émission permettent d'extraire des données sur les énergies de transition de la molécule, qui sont caractéristiques de sa structure et de sa dynamique.

En biologie structurale, de nombreuses méthodes spectroscopiques et de diffraction sont utilisées. Chacune d'elle apporte des informations différentes concernant la structure et la dynamique des molécules, selon la nature du rayonnement utilisé et son énergie. Le choix des méthodes structurales pour l'étude d'une macromolécule biologique dépend en pratique non seulement du type d'informations que l'on veut obtenir, mais également des difficultés pratiques rencontrées. Par exemple, l'obtention d'échantillons cristallins ou sous forme soluble est une tâche souvent difficile, malgré les développements faits dans ce domaine. Les méthodes utilisant la diffraction se heurtent souvent à la fragilité des échantillons biologiques en présence de rayonnements de faible longueur d'onde. Dans les méthodes spectroscopiques, les données obtenues sont généralement trop peu nombreuses pour pouvoir caractériser la molécule dans son ensemble à un niveau atomique. On a alors recours à une modélisation qui utilise des données statistiques concernant la chimie de la molécule. La multiplicité des techniques disponibles actuellement permet donc d'envisager des études structurales sur un plus grand nombre d'échantillons, en apportant un savoir faire et des informations complémentaires.

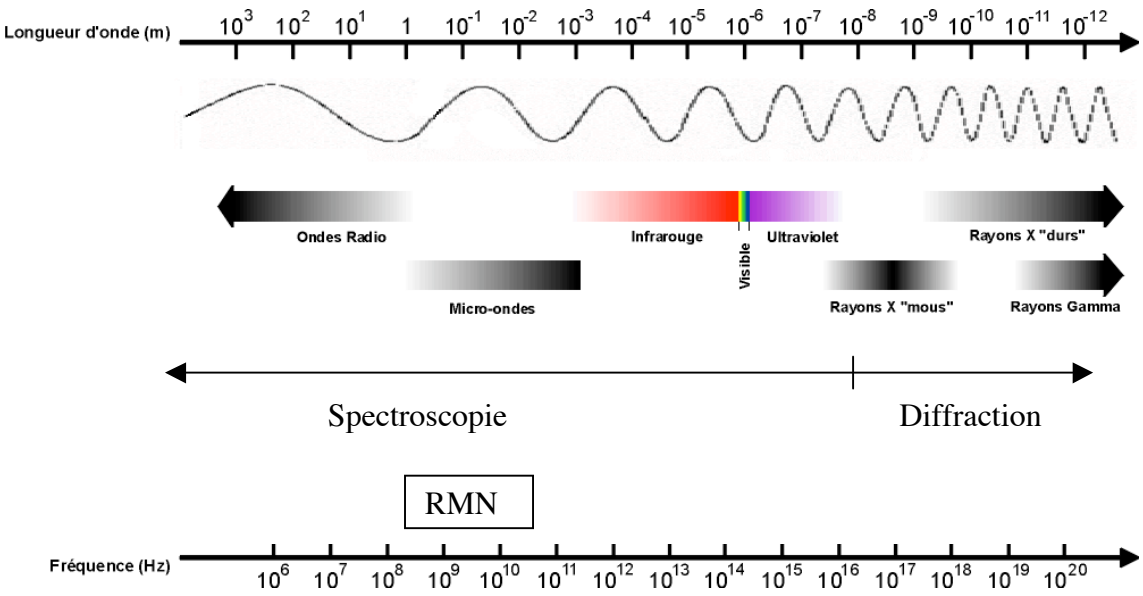


Figure 2: Spectre électomagnétique et caractérisation des rayonnements de différentes longueurs d'onde.

Actuellement, la majorité des structures de macromolécules biologiques à résolution atomique déposées dans la PDB (Protein Data Bank : http://www.rcsb.org/pdb) ont été obtenues par cristallographie par diffraction des rayonx X, et par spectroscopie RMN (80% et 20% pour les protéines, et 60% et 40% pour les acides nucléiques, respectivement). De l'ensemble de ces structures sont nés de nouveaux outils bioinformatiques qui permettent d'obtenir le repliement d'une molécule par homologie avec les molécules dont les structures sont connues. Cette technique, sans remplacer les méthodes expérimentales, permet de guider leur choix et d'accélérer leur étude, que ce soit pour la cristallographie ou pour la RMN.

Le travail de thèse présenté dans ce document se situe dans le contexte de l'utilisation de la Résonance Magnétique Nucléaire (RMN) au service de la biologie structurale. La RMN est une technique spectroscopique qui s'intéresse aux transitions énergétiques des niveaux Zeeman des spins nucléaires placés dans un champ magnétique. Ces transitions se font dans le domaine des ondes radio (Figure 2) et sont caractéristiques de la nature et de l'environnement chimiques du noyau considéré. Les noyaux de spin 1/2 dont les fréquences de résonance sont étudiées en RMN des macromolécules biologiques sont donnés dans le Tableau 1.

Noyau	Sensibilité relative (γ/γ_H)	Fréquence ν_0 dans un champ de 14 T (MHz)	Abondance naturelle (%)
^{1}H	1	600	99.98
^{15}N	$1.04 \quad 10^{-3}$	61	0.37
^{13}C	$1.59 \quad 10^{-2}$	151	1.11
^{31}P	$6.63 \quad 10^{-2}$	243	100

Tableau 1: Propriétés nucléaires des quatre principaux noyaux de spin 1/2 étudiés par RMN dans les macromolécules biologiques.

L'étude des fréquences de résonance et des réseaux de couplages qui existent entre les spins de la molécule permet d'obtenir des informations sur sa structure et sa dynamique. Souvent on a recours à l'enrichissement des macromolécules biologiques avec les isotopes ^{15}N et ^{13}C, ce qui permet d'augmenter le nombre de sondes magnétiques détectables au sein de la molécule.

L'étude structurale des macromolécules biologiques par RMN en solution suit un protocole qui est décrit dans la Figure 3. La première étape consiste à produire et purifier un ou plusieurs échantillons contenant la molécule d'intérêt enrichie avec les isotopes ^{15}N et ^{13}C, voire ^{2}H. Ces échantillons doivent être purs, stables au moins sur plusieurs jours, et concentrés à une concentration de l'ordre de 0.5 mM. Les fréquences de résonance de chacun des atomes de la molécule sont ensuite attribuées. Cette étape est effectuée en enregistrant une série de spectres qui corrèlent les fréquences des différents noyaux. L'étape suivante consiste à extraire des informations sur la distance entre les atomes, sur les différents angles de torsion de la molécule et sur les orientations des vecteurs internucléaires à partir des spectres RMN. Ces informations sont ensuite utilisées sous forme de contraintes dans un calcul de dynamique moléculaire qui reconstruit la structure de la molécule. Ces différentes étapes sont explicitées sur l'exemple de la détermination de structure de la protéine BlaI dans la première partie de ce manuscrit.

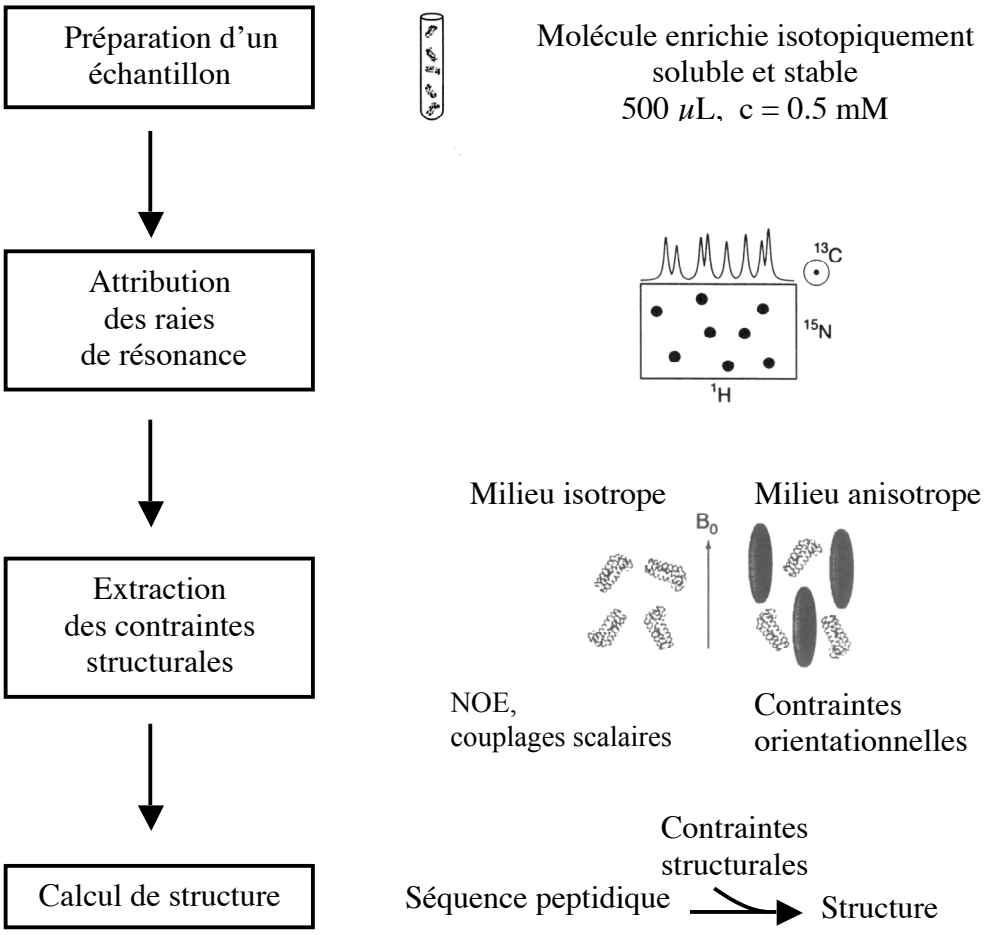


Figure 3: Schématisation des différentes étapes nécessaires à l'étude structurale classique d'une macromolécule biologique par RMN.

La RMN est une technique encore jeune, en particulier dans le domaine de la biologie structurale. Depuis la première résolution de structure d'une petite protéine de 57 résidus en 1985 par Williamson et ses collaborateurs, des progrès faits sur le plan de la préparation des échantillons biomoléculaires, de l'instrumentation, et des techniques spectroscopiques ont amélioré le nombre et la qualité des informations structurales qu'on peut obtenir par RMN. Néanmoins, la taille des molécules que l'on peut étudier par RMN reste limitée. Actuellement, les études effectuées avec des échantillons enrichis isotopiquement de façon uniforme concernent des protéines de moins de 150 résidus (17 kDa) ou des acides nucléiques de moins de 30 nucléotides (10 kDa). La Figure 4 donne quelques exemples des structures de molécule de taille importante qui ont été résolues par RMN en solution récemment.

Pour obtenir de tels résultats, il faut se battre contre les problèmes rencontrés dans l'étude de systèmes plus complexes à différents niveaux. Tout d'abord, il faut réussir à attribuer sans ambiguïté toutes les résonances du système par des techniques spectroscopiques appropriées, ou procéder à un enrichissement sélectif de la molécule avec différents isotopes. La deuxième étape consiste à collecter assez de contraintes structurales pour obtenir la structure à haute résolution de la molécule. En plus des contraintes de distance obtenues de façon classique par effet Overhauser nucléaire (NOE), de nouvelles techniques permettent désormais d'obtenir des contraintes sur l'orientation des vecteurs internucléaires par la mesure de constantes de couplage résiduel dipolaire dans un milieu qui aligne faiblement la molécule. L'utilisation de sondes paramagnétiques au sein de la molécule fournit également une nouvelle source de contraintes structurales. Enfin, le développement de la RMN en milieu solide ouvre de nouveaux horizons vers l'étude de systèmes complexes encore mal caractérisés, tels les systèmes fibrillaires ou les protéines membranaires. Pour toutes ces avancées, l'efficacité des techniques spectroscopiques utilisées est déterminante, et le développement d'outils appropriés est nécessaire.

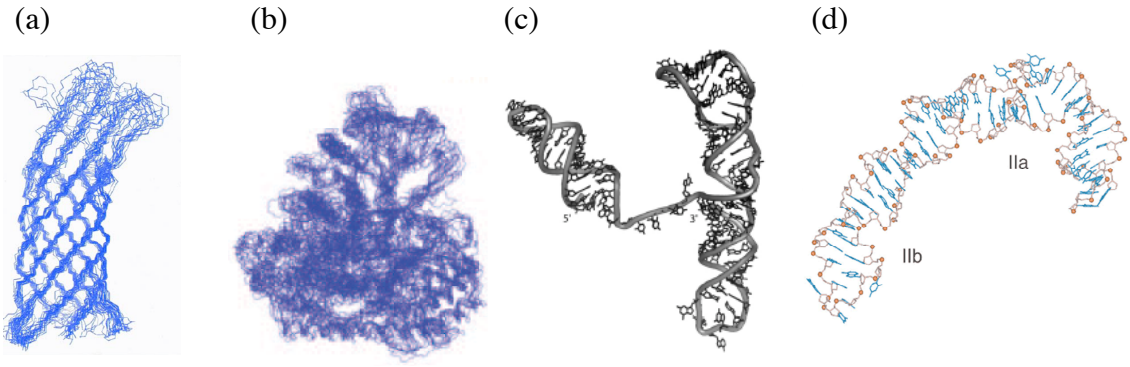


Figure 4 : Exemples de structures résolues par RMN. (a) Protéine membranaire OmpX reconstituée dans des micelles de DHPC : 177 résidus, 60 kDa (tiré de Fernandez *et al*., 2004). (b) Malate synthase G : 723 residues, 82 kDa (extrait de Tugarinov *et al.*, 2005). (c) Séquence d'ARN correspondant au signal d'encapsidation du virus de la leucémie murine monoloney : 101 nucléotides, 33 kDa (tiré de D'Souza *et al.*, 2004). (d) Domaine II de l'IRES de l'hépatite C : 77 nucléotides, 25 kDa (tiré de Lukavsky *et al.*, 2003).

Le travail de thèse présenté dans ce document illustre deux facettes de la RMN biomoléculaire en solution. La première partie décrit l'étude du répresseur du gène de la β-lactamase, seul et en interaction avec la séquence opératrice du gène, par des méthodes de RMN classiques. Cette étude donne un très bon exemple des réponses que peut apporter la RMN à des questions biologiques de très haute importance, ici dans le contexte de la résistance aux antibiotiques des bactéries. La deuxième partie est méthodologique et décrit une nouvelle technique de filtrage Hadamard qui permet d'augmenter la résolution des spectres RMN. Le principe de cette technique sera détaillé, et des applications seront données dans le contexte de l'attribution des résonances et de la recherche de contraintes structurales pour les protéines et les acides nucléiques. Cette technique montre le potentiel qu'offre la RMN pour étudier des systèmes de plus en plus complexes.

Etude structurale de la répression du gène de la β-lactamase chez B. licheniformis 749/I

Introduction: Résistance des bactéries aux antibiotiques de type β-lactame

I. Mode d'action des antibiotiques à noyau β–lactame

I. 1. La chimiothérapie anti-bactérienne

Un antibiotique est une substance produite par des micro-organismes (champignons, levures ou bactéries) ou synthétisée chimiquement, qui est capable d'inhiber la multiplication ou de détruire d'autres bactéries. On parle respectivement d'effet bactériostatique ou bactéricide.

L'utilisation des antibiotiques à des visées thérapeutiques a été initiée au début du vingtième siècle par le médecin allemand P. Ehrlich. Les recherches de Ehrlich aboutiront à un remède à base d'arsénobenzènes contre la syphilis, ce qui lui vaudra le prix Nobel en 1908. L'action bactériostatique de certains micro-organismes avait déjà été observée en 1877 par L. Pasteur et M. Joubert (à propos du bacille charbonneux), mais ce n'est qu'en 1929 que A. Fleming constate que la culture en boîte de Petri d'une souche de *Staphylococcus aureus* est inhibée par une contamination de moisissures du genre *Penicillium*. Le champignon sécrète une substance chimique bactériostatique que Fleming nomme pénicilline. Il prouve par la suite que cette substance n'est pas nocive pour l'Homme et suggère de l'utiliser comme antiseptique. En 1939, Florey et Chain purifient la pénicilline G et, avec Abraham et Heatley, démontrent ses vertus comme médicament interne. Le 12 février 1941, un premier patient atteint d'une septicémie est sauvé par la pénicilline à Oxford. Depuis, l'usage d'antibiotiques variés s'est répandu dans le monde, ce qui a permis de faire d'énormes progrès pour guérir les maladies bactériennes.

Les antibiotiques qui sont utilisés à l'heure actuelle peuvent être classés selon plusieurs critères: l'origine, la nature chimique, le mécanisme d'action et le spectre d'action. Ils sont pour la plupart extraits d'organismes vivant dans le sol. Pour qu'ils soient toxiques pour les bactéries, mais inoffensifs pour l'hôte à qui on administre le traitement, les antibiotiques ont pour cibles principales les constituants moléculaires proprement bactériens. Actuellement, les drogues anti-bactériennes agissent en priorité au niveau de la biosynthèse et du maintien de la paroi cellulaire, des protéines, et de l'ADN bactériens (Walsh, 2000). La biosynthèse de la paroi cellulaire peut être inhibée pour empêcher la croissance des bactéries. Plusieurs classes d'antibiotiques bloquent la synthèse protéique en agissant sur le ribosome bactérien. Les composés qui inhibent la synthèse des acides nucléiques agissent souvent au niveau de l'ARN polymérase ou des topoisomérases bactériennes. Les principales classes d'antibiotiques et leurs cibles sont résumées dans la Figure 5 et le Tableau 2, et une liste plus complète est donnée dans l'Annexe 1.

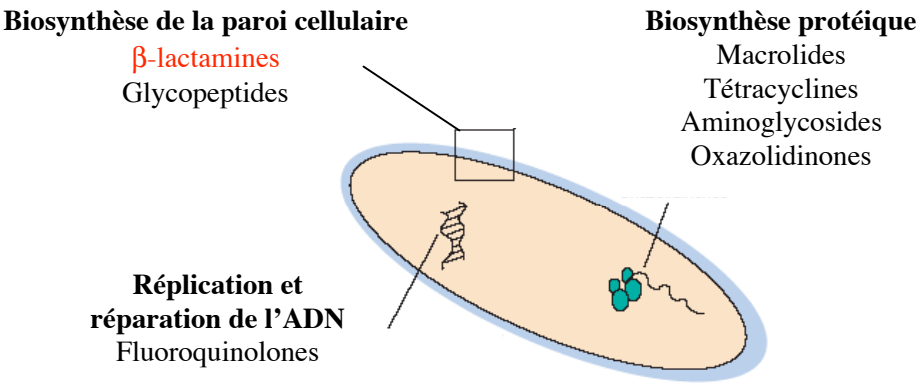


Figure 5: Cibles principales des antibiotiques et exemples de drogues anti-bactériennes (adapté de Walsh, 2000).

Antibiotique	Cible	Mode d'action	Mécanisme résistance
Paroi cellulaire			
β-lactamines	DD-peptidases (PBP)	Inhibition réticulation du peptidoglycane	β-lactamases, mutants PBP
Vancomycine	Peptidoglycane et lipide II (D-Ala-D-Ala)	Inhibition du substrat utilisé pour la réticulation	Production de D-Ala-D-Lac ou D-Ala-D-Ser
Synthèse protéique			
Macrolides (Erythromycine)	Peptidyl transférase Centre ribosome	Inhibition de la synthèse protéique	Méthylation de rARN Systèmes d'efflux
Tétracyclines	Peptidyl transférase	Inhibition de la synthèse protéique	Systèmes d'efflux
Aminoglycosides	Peptidyl transférase	Inhibition de la synthèse protéique	Modification enzymatique
Oxazolidinones	Peptidyl transférase	Inhibition de la synthèse protéique	Non connu
Réplication/maintien ADN			
Fluoroquinolones	ADN gyrase	Inhibition de la réplication de l'ADN	Mutations de la gyrase

Tableau 2 : Cibles et mode d'action de différentes familles d'antibiotiques. Quelques mécanismes de résistance développées par les bactéries contre ces antibiotiques sont explicités. La classe d'antibiotiques qui fait l'objet de ce travail est surlignée en rouge. L'Annexe 2 donne un tableau plus complet des différentes classes d'antibiotiques et de leur composition chimique.

I. 2. Le peptidoglycane

La présence d'une paroi autour de la membrane plasmique est une caractéristique des eubactéries. Cette paroi détermine la morphologie du microorganisme et le protége de sa forte pression osmotique interne (Figure 6). C'est pourquoi l'intégrité de la paroi cellulaire est indispensable à la croissance et à la division cellulaire des bactéries (Green, 2002). La rigidité de cette paroi lui est conférée par le peptidoglycane, qui est un hétéropolymère réticulé. L'épaisseur et la localisation du peptidoglycane ne sont pas constantes pour toutes les bactéries (Figure 6). C'est néanmoins un constituant ubiquitaire de la paroi des bactéries, ce qui fait de lui une cible thérapeutique de choix.

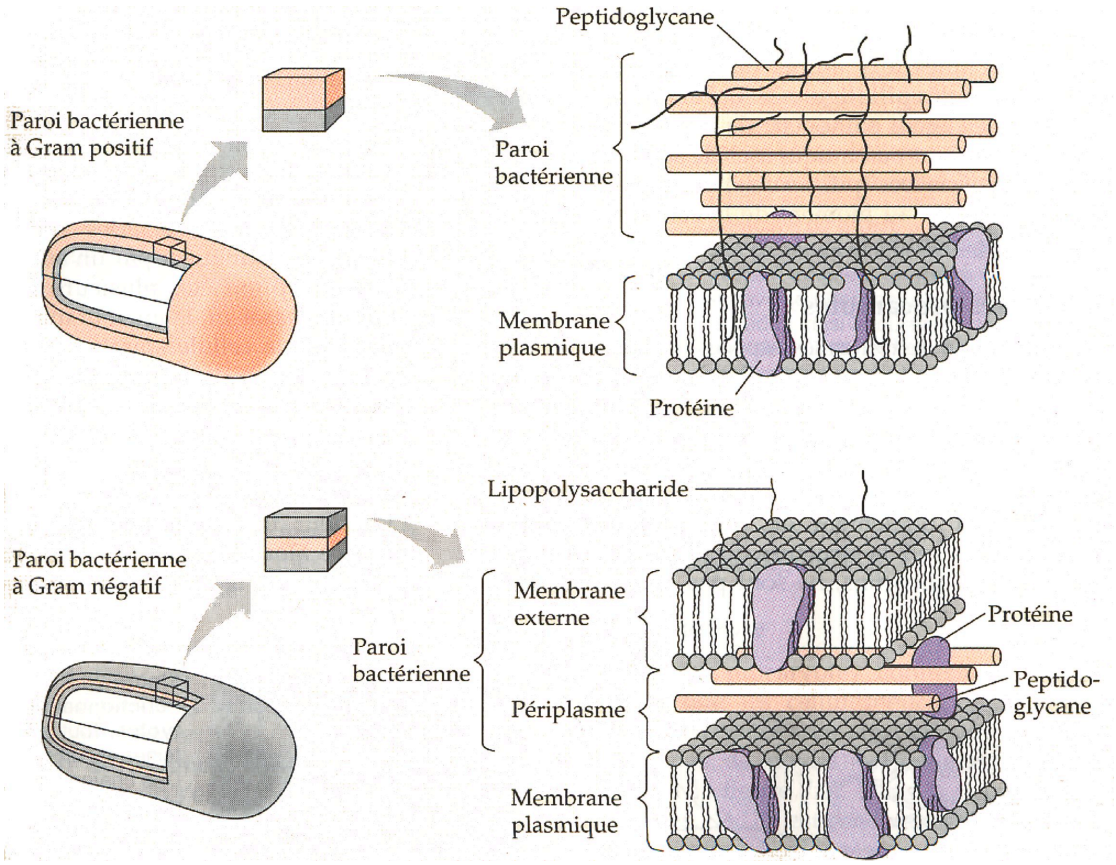


Figure 6: Schéma en coupe de la paroi des eubactéries. (a) Les bactéries dites à Gram-positif possèdent une paroi épaisse (15 à 80 nm) composée à 90% par le peptidoglycane, et à 10% par un polymère anionique appelé acide téichoïque (polymère de glycérol ou de ribitol relié par des groupements phosphates) (Ward et Curtis, 1982). (b) Chez les bactéries à Gram-négatif, la paroi est plus fine (10 à 15 nm) et de constitution plus complexe: une fine couche de peptidoglycane (1 à 3 nm) est surmontée d'une membrane externe, formée par une couche interne de phospholipides et d'une couche externe de lipopolysaccharides. Cette membrane externe constitue une barrière de protection contre les agents toxiques. Des protéines spécialisées appelées porines y sont enchassées et permettent le passage de molécules de masse moléculaire inférieure à 600 Da vers le périplasme (Green, 2002).

Le peptidoglycane est un mucopolysaccharide composé de chaînes linéaires osidiques pontées par de courts peptides (Van Heijenoort, 2001) (Figure 7). Les chaînes osidiques sont formées par une alternance de résidus N-acétylglucosamine (NAG) et d'acide N-acétylmuramique (NAM). Chaque disaccharide porte un tétrapeptide fixé au groupement lactyle de l'unité NAM. La formation de ponts entre les peptides des chaînes glycane voisines crée une véritable réticulation de la paroi (Figure 7). La composition du pont interpeptidique et le degré de pontage varient selon les espèces bactériennes (Van Heijenoort, 2001 ; Rohrer et Berger-Bächi, 2003).

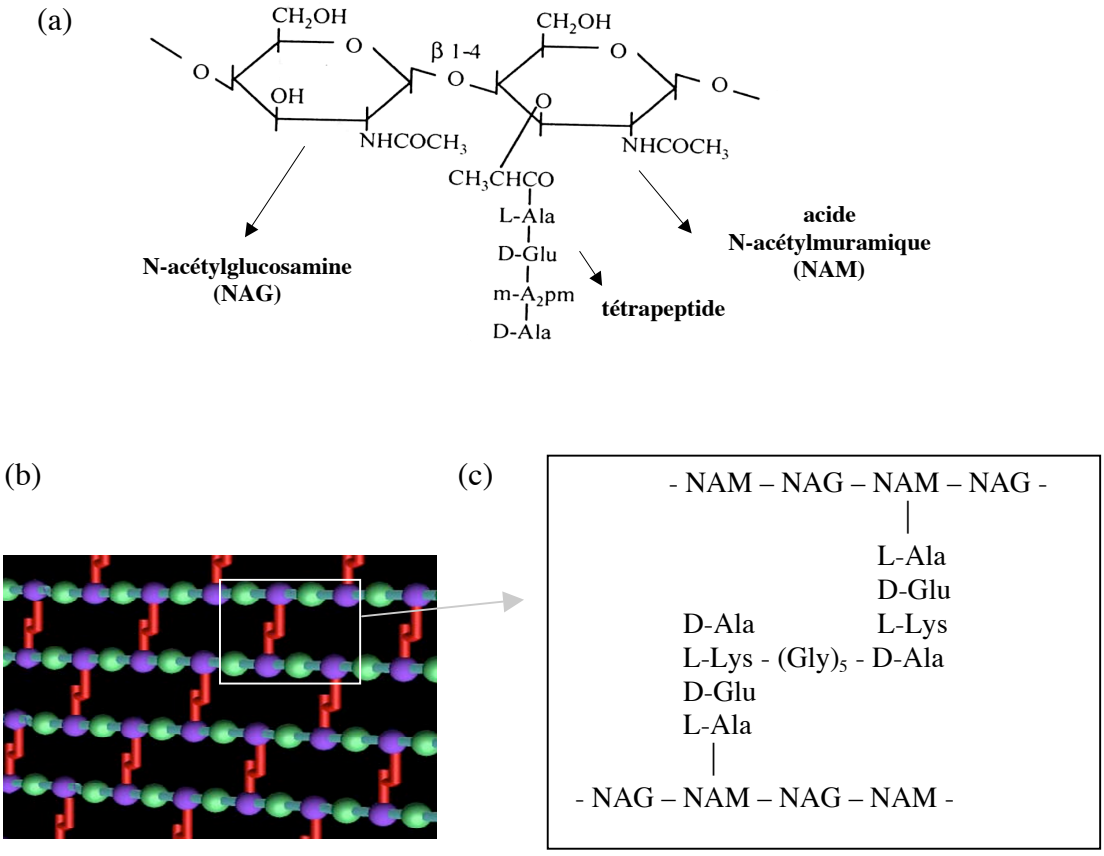


Figure 7 : (a) Composition de la sous-unité de base et (b) structure générale du peptidoglycane (Ghuysen, 1997). Les résidus N-acétylglucosamine (NAG) et acide N-acétylmuramique (NAM) sont liés par une liaison β 1-4. L'acide N-acétylmuramique porte un tétrapeptide composé d'une alternance d'acides aminés de configuration L et D, dont le troisième acide aminé est un résidu diaminé (acide méso-diaminopimélique). Pour la plupart des bactéries, la fonction amine libre de ce résidu est reliée par une liaison amide à l'extrémité carboxyle d'un tétrapeptide d'une chaîne de glycane voisine. Pour d'autres bactéries, un pont interpeptidique permet la liaison entre la fonction amine libre du résidu diaminé et la fonction carboxyle de l'autre peptide. (c) Chez *Staphylococcus aureus*, le troisième acide aminé est une L-Lysine, et un pont pentaglycine permet la liaison entre les deux tétrapeptides (Ghuysen *et al.*, 1984).

La biosynthèse du peptidoglycane se décompose en une série de réactions enzymatiques complexes qui peuvent être regroupées en trois grandes étapes en fonction de leur localisation cellulaire (Ghuysen, 1968). Ces trois étapes correspondent respectivement à la biosynthèse du précurseur soluble du peptidoglycane dans le cytoplasme, au transport de celui-ci à l'extérieur de la cellule via un transporteur lipidique, et enfin à l'assemblage du peptidoglycane à la surface externe de la membrane cytoplasmique (van Heijenoort, 2001; Rohrer et Berger-Bächi, 2003; Green, 2002). On peut noter que ces différentes étapes sont conservées chez toutes les bactéries. Durant la dernière étape, la polymérisation des tétrapeptides de chaque unité de base confère à la paroi du peptidoglycane sa rigidité naturelle (Blumberg et Strominger, 1974; Frère et Joris, 1985; Frère, 1988). La réaction de transpeptidation est catalysée sur la face externe de la membrane cytoplasmique par des DD-transpeptidases (van Heijenoort, 2001). Le degré de pontage du peptidoglycane est contrôlé par les DD-carboxypeptidases (Ghuysen *et al.*, 1996). Les DD-transpeptidases et les DD-carboxypeptidases sont appelées DD-peptidases. Ce sont des enzymes à sérine nucléophile qui peuvent être inhibées par les antibiotiques de type β–lactame.

I. 3. Mode d'action des antibiotiques de type β–lactame

Les antibiotiques à noyau β–lactame (ou β-lactamines) sont caractérisés par un cycle β-lactame à quatre liaisons dont la tension stérique rend la liaison amide plus réactionnelle qu'une liaison amide linéaire. Les β-lactamines sont divisées en quatre grandes classes selon la nature de l'hétérocycle à 5 ou 6 atomes accolé au noyau β–lactame (Figure 8).

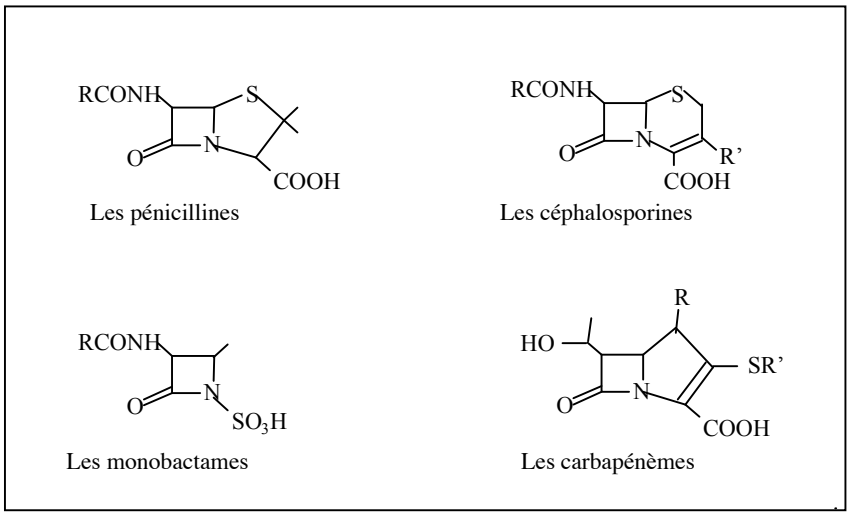
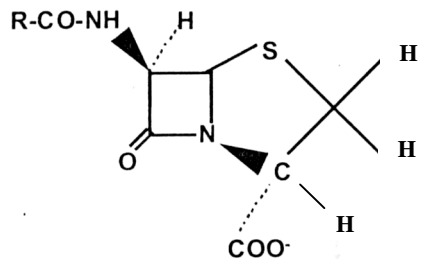
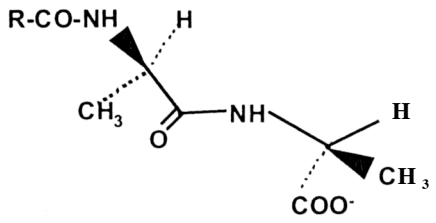


Figure 8: Les différentes classes d'antibiotiques à noyau β-lactame. La variabilité des chaînes latérales R et R' permet d'étendre la spécificité d'action de ces composés et de contrer l'émergence de souches résistantes.

Les β–lactamines agissent sur les bactéries en croissance en inhibant la dernière étape de la synthèse du peptidoglycane, la transpeptidation (Frère *et al.*, 1976). Elles se lient en effet de façon covalente aux DD-peptidases, aussi appelées pour cette raison Penicillin Binding Proteins ou PBP (Blumberg et strominger, 1974; Frère et Joris, 1985; Frère, 1988). PBP est donc un terme générique utilisé pour désigner l'ensemble des protéines possédant un domaine transpeptidase ou carboxypeptidase (Goffin et Ghuysen, 1998). L'inhibition des PBP par les β–lactamines est due à l'analogie biochimique et structurale de ces antibiotiques avec le substrat naturel des DD-peptidases, la D-alanyl-D-alanine (Figure 9) (Tipper et Strominger, 1965; Ghuysen *et al.*, 1984). L'acyl-enzyme stable qui est formé provoque l'inactivation irréversible des PBP (Nicholas et Strominger, 1988; Frère, 1988). Un déséquilibre se crée alors au niveau de la paroi cellulaire car l'activité autolytique naturelle du peptidoglycane n'est plus compensée par la synthèse de nouveaux polymères (Calamita et Doyle, 2002). La dégradation progressive du peptidoglycane qui en découle empêche la bactérie de se diviser et la rend beaucoup plus sensible à la lyse osmotique.



La pénicilline



La D-alanyl-D-alanine

Figure 9: Similarité de structure entre les antibiotiques à noyau β–lactames (ici la pénicilline) et le substrat naturel des DD-peptidases (D-alanyl-D-alanine).

Le succès des β-lactamines en thérapie humaine réside dans le fait qu'elles exercent leur effet antibiotique sur les germes contenant une paroi riche en peptidoglycane, et n'ont en général aucun effet sur les organites dépourvus de peptidoglycane, comme les cellules animales. Les effets secondaires lors de la prise de β-lactamines sont donc très réduits, malgré une bonne efficacité bactéricide. Malheureusement, l'efficacité des β-lactamines a été largement diminuée par l'apparition de souches bactériennes résistantes à ces antibiotiques.

II. Résistance des bactéries aux antibiotiques à noyau β–lactame

II. 1. Un phénomène inéluctable et de grande ampleur

La découverte des agents thérapeutiques antimicrobiens est peut-être l'une des plus grandes avancées médicales du XXème siècle. Les causes de décès dues aux maladies infectieuses bactériennes ont été en effet divisées par 20 entre 1900 et 1980 (Fisher *et al.*, 2005). Cependant, l'émergence de souches résistantes a été constatée en moyenne quelques années après l'utilisation d'une classe d'antibiotiques (Lipsitch, 2001). La Figure 10 donne une illustration de l'ampleur et de la rapidité d'adaptation des bactéries pour contrer la chimiothérapie antibactérienne.

La rapide adaptabilité des bactéries face aux antibiotiques peut s'expliquer tout d'abord par le fait que certaines bactéries possèdent naturellement tout un arsenal de gènes de résistance. Ces gènes de résistance chromosomiques ou plasmidiens ont été acquis par des bactéries qui produisent elles-mêmes des antibiotiques, ou qui vivent de façon ancestrale dans la niche écologique de microorganismes qui en produisent. Par ailleurs, les bactéries Gram-négatives possèdent une membrane externe qui empêche la diffusion des antibiotiques vers leurs cibles. Enfin, certaines espèces bactériennes ne possèdent pas les cibles sur lesquelles la drogue agit. Ces différents types de résistance intrinsèque limitent donc le spectre d'action d'un antibiotique donné.

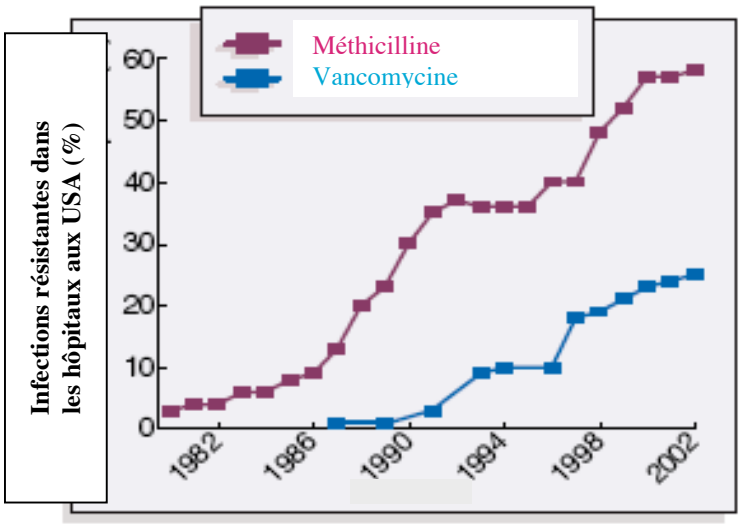


Figure 10: Evolution des résistances aux antibiotiques de type méthicilline et vancomycine au fil des années. Les souches qui ont été analysées proviennent de patients infectés dans les hôpitaux américains (tiré de Nathan, 2004).

De plus, les bactéries sont capables de s'adapter très vite aux pressions de sélection extérieures et d'acquérir de nouveaux gènes (Figure 11). Cette résistance acquise n'apparaît que chez certaines populations d'une même espèce bactérienne. Elle résulte de modifications génétiques dans le génome bactérien.

Ces modifications peuvent avoir pour origine des mutations spontanées conservées par la sélection naturelle et transmises aux générations futures (on parle d'évolution verticale). Dans un environnement où est présent l'antibiotique, les populations sauvages sont détruites, alors que le mutant résistant peut se multiplier dans la niche laissée vacante. Ce genre d'évènement est loin d'être anodin, puisque le taux de mutation pour la plupart des gènes bactériens est de l'ordre de 10^{-8}, voire plus élevé dans certains cas (LeClerc *et al.*, 1996; Schaaff *et al.*, 2002).

L'autre mécanisme permettant l'acquisition de gènes de résistance est l'échange de matériel génétique entre souches ou espèces (on parle d'évolution horizontale). Ce transfert peut se faire par un processus de conjugaison, de transduction, ou de transformation (Figure 11). A l'heure actuelle, trois classes d'intégrons de résistance et plus de 70 cassettes de gènes de résistance aux antibiotiques ont été caractérisées (Mazel and Davies, 1999).

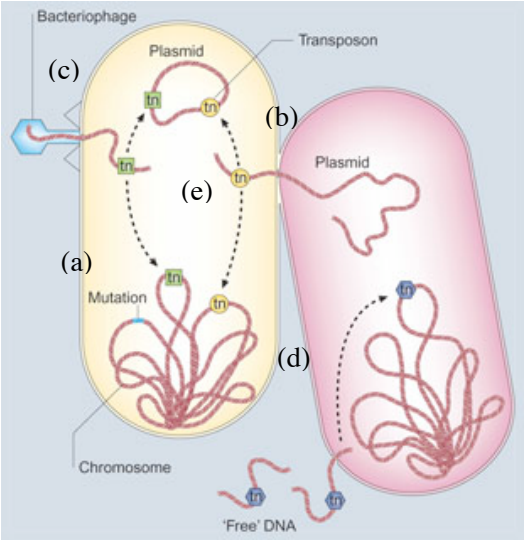


Figure 11: Mécanismes d'acquisition de gènes de résistance par les bactéries (adapté de Levy et Marshall, 2004). (a) Mutation ponctuelle. (b) La conjugaison permet l'échange de plasmides entre les bactéries par l'intermédiaire des pili. (c) Mécanisme de transduction. L'ADN de la bactérie contenant les gènes de résistance est transféré d'une bactérie à une autre par l'intermédiaire d'un phage lors de l'infection. (d) Mécanisme de transformation. L'ADN du milieu environnant, provenant par exemple d'une bactérie lysée, peut pénétrer à l'intérieur d'une cellule vivante. (e) Les plasmides qui sont échangés entre les bactéries peuvent contenir des transposons et des intégrons. Les transposons sont des fragments d'ADN qui contiennent des gènes codant pour des enzymes permettant leur transfert d'un ADN à un autre. Ils peuvent véhiculer des gènes de résistance. Les intégrons sont des éléments d'ADN constitués de deux segments conservés situés de part et d'autre d'une région centrale dans laquelle sont insérées des cassettes de gène de résistance.

Enfin, l'utilisation massive des antibiotiques non seulement en médecine humaine et vétérinaire, mais également dans l'industrie agroalimentaire à une large échelle, accélère l'émergence de mécanismes de résistance. La transmission des bactéries pathogènes est également facilitée par la proximité des personnes infectées avec des personnes affaiblies dans les hôpitaux.

Si on considère le nombre d'atouts que possèdent les bactéries pour s'adapter à leur environnement, on se rend compte que le phénomène de résistance à un nouvel antibiotique est inévitable à plus ou moins long terme. Nos seuls moyens d'action consistent à (Levy et Marshall, 2004):

- retarder l'apparition de nouvelles résistances, grâce à une utilisation modérée et ciblée des antibiotiques.

- éviter la propagation des souches résistantes, par l'isolement des sujets malades.

- développer de façon ininterrompue de nouveaux antibiotiques ou des drogues restaurant l'activité d'un antibiotique. Pour cela, la compréhension des mécanismes moléculaires de résistance est fondamentale.

II. 2. Mécanismes de résistance aux antibiobiotiques à noyau β-lactame

De façon très générale, les mécanismes de résistance des bactéries aux antibiotiques sont nombreux et variés, mais ils reposent sur trois principes de base qui sont schématisés dans la Figure 12: l'acquisition de moyens physico-chimiques pour empêcher l'antibiotique d'atteindre sa cible, la dégradation biochimique de l'antibiotique, et la modification des cibles des antibiotiques de façon à éviter leur inhibition (Walsh, 2000; Figure 5). Les paragraphes suivants détaillent ces trois types de résistance dans le cas des antibiotiques à noyau β–lactame (Frère *et al*., 1991; Fisher *et al*., 2005).

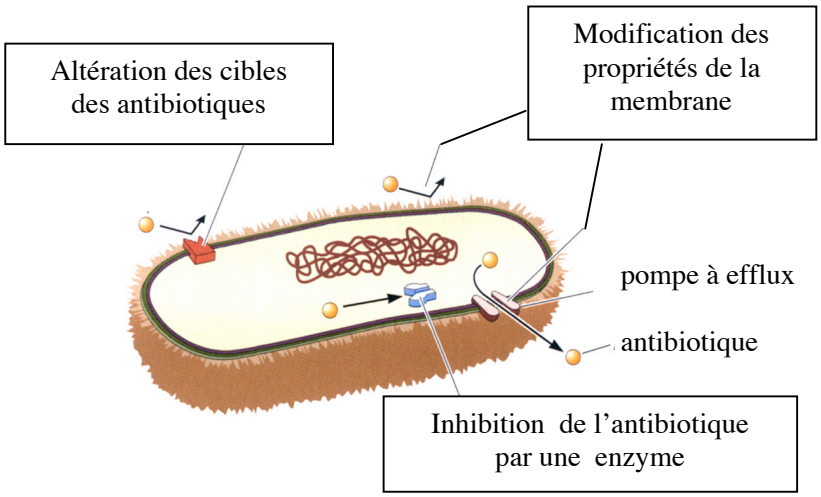


Figure 12 : Représentation schématique des différents mécanismes de résistance des bactéries aux antibiotiques.

- **Perméabilité de la membrane et transport actif**

Le premier mode de résistance aux β−lactamines concerne la modification de la perméabilité de la membrane externe des bactéries à Gram-négatif à la suite d'une mutation ou de la perte des gènes codant pour les porines (Parr *et al.*,1987; Nikaido, 1994; Kar *et al.*, 1997). Les porines sont des protéines transmembranaires qui assurent le passage des antibiotiques dans le périplasme à travers la membrane externe (Figure 6). Pour les bactéries résistantes, la membrane externe devient imperméable aux β−lactamines, ce qui rend inefficace les antibiotiques dont les cibles se trouvent sur la face externe de la membrane cytoplasmique.

Des systèmes d'efflux peuvent également être responsables de la résistance (Levy, 2002). La concentration en β−lactamine dans le périplasme est rendue faible grâce à une excrétion active de l'antibiotique hors de la cellule (Nikaido, 1996).

- **Dégradation de l'antibiotique par les β-lactamases**

Le mécanisme de résistance le plus efficace et le plus répandu dans le monde bactérien est la production de β−lactamase**,** une enzyme capable d'hydrolyser la liaison amide des antibiotiques à noyau β−lactame (Figure 13) (Richmond, 1965 ; Livermore, 1995). Les gènes codant pour les β−lactamases, qu'ils soient chromosomiques ou plasmidiens, ont été mis en évidence chez un grand nombre de bactéries à Gram-positif et négatif (Foster *et al.*, 1983). Chez les bactéries à Gram-positif, ces enzymes sont sécrétées dans le milieu extérieur (Nielsen et Lampen, 1982). Chez les bactéries à Gram-négatif, les β−lactamases sont maintenues et agissent au sein du périplasme, ce qui rend leur présence très efficace.

Sur la base de leur stucture primaire, on distingue quatre classes de β−lactamases: A, B, C et D (Frère, 1995 ; Philippon, 1998). Les enzymes de classe A, C et D sont des enzymes à sérine active, celles de classe B sont des métallo-enzymes qui ont un ou deux atomes de zinc dans leur site actif (Waley, 1992).

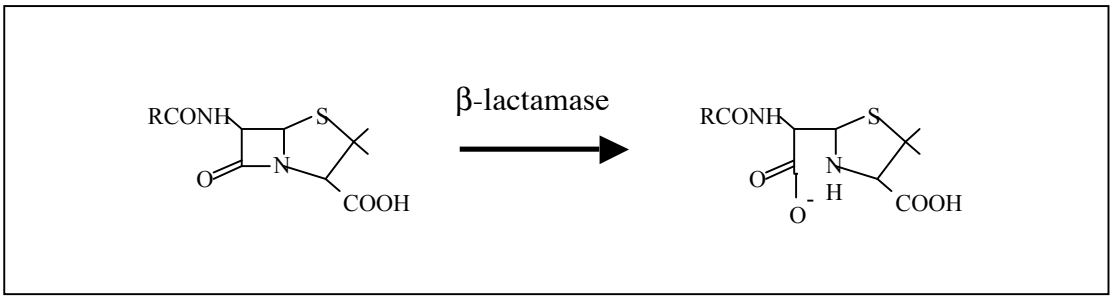


Figure 13 : La β-lactamase est excrétée dans le périplasme, où elle hydrolyse le cycle β-lactame des antibiotiques.

Le mécanisme catalytique des β–lactamases à sérine en interaction avec les β–lactamines est semblable à celui des PBP (ou DD-peptidases). Il semblerait d'ailleurs que ces deux types de protéines soient liées au sens évolutif (Kotra et Mobashery, 1998 ; Joris *et al*., 1985). La protéine forme un acyl-enzyme avec la pénicilline, dans lequel la sérine est acylée par l'antibiotique. Cependant, contrairement au cas des DD-peptidases, l'acyl-enzyme est instable chez les β–lactamases, ce qui se caractérise par une constante de déacylation élevée. La rupture du complexe acyl-enzyme aboutit alors à la libération de l'antibiotique inactivé, et à la régénération de l'enzyme. Ce processus est très efficace (une enzyme peut inactiver jusqu'à 1000 β–lactamines par seconde) et permet d'éviter l'inhibition des PBP par l'antibiotique (Frère, 1995).

- **Modification des PBP**

Un autre mécanisme de résistance est basé sur la production de PBP alternatives qui, contrairement aux PBP sauvages, n'interagissent pas avec l'antibiotique. Elles permettent donc la synthèse du peptidoglycane même en présence d'antibiotique, ce qui assure la survie de la bactérie (Hartman, 1984). Cet effet est observé chez un nombre croissant d'espèces bactériennes et la diversité de ce phénomène est importante. Pour certaines souches, les mutations apparues dans le gène codant pour la DD-transpeptidase affectent uniquement quelques résidus (Malouin et Bryan, 1986; Spratt, 1994). Par exemple, l'insertion de quatre acides aminés dans la PBP3 de *Escherichia coli* ou d'un seul résidu dans la PBP2 de *Neisseria gonorrhoea* réduit fortement l'affinité des PBP pour les β–lactamines (Hedge et Spratt, 1985 ; Brannigan *et al*., 1990). Par ailleurs, des fragments importants de gènes sont substitués par des fragments qui proviennent d'autres bactéries, formant des gènes en mosaïque. C'est le cas, par exemple, chez *Streptococcus pneumoniae* (Dowson *et al*., 1989 ; Dessen *et al*., 2001).

Chez les souches *Staphylococcus aureus* et *Staphylococcus epidermidis* résistantes à la méthicilline (antibiotique de la famille des β–lactamines), les PBP sensibles aux antibiotiques ont été conservées. En présence de méthicilline, leurs fonctions enzymatiques sont cependant assurées grâce à la production d'une PBP supplémentaire de faible affinité pour l'antibiotique, appelée PBP2' ou PBP2a (Ubukata *et al.* 1985; Joris *et al.*, 1994).

III. Régulation des gènes de résistance chez *B. licheniformis* et *S. aureus*

III. 1. Les systèmes de régulation

- **Régulation des gènes de résistance**

Pour assurer sa survie en présence d'antibiotique, la bactérie peut selon les souches produire les protéines nécessaires à sa défense de façon constitutive, ou posséder des systèmes de régulation qui initient la synthèse protéique uniquement en présence d'agents anti-bactériens. Cette dernière alternative a l'avantage d'optimiser les dépenses énergétiques de la bactérie en absence d'antibiotique.

Ce phénomène est par exemple observé pour la production inductible de certaines β-lactamases (Bennet et Chopra, 1993). Actuellement, quatre mécanismes de régulation distincts ont été reportés dans la littérature pour la production des β-lactamases (deux modèles pour les bactéries à Gram-positif, et deux pour les bactéries à Gram-négatif). Un des modèles est commun aux souches *Staphylococcus aureus* et *Bacillus licheniformis*, qui produisent des β-lactamases de classe A, BlaP et BlaZ respectivement (Sherrat et Collins, 1973). Ces β-lactamases présentent une grande similarité de structures primaire et tertiaire (Knox et Moews, 1991 ; Herzberg, 1991). L'expression de la β-lactamase de chacune de ces souches est induite par la présence d'une β-lactamine dans le milieu extra-cellulaire (Salerno et Lampen, 1986 ; McKinney, 2001). De plus, les gènes impliqués dans le mécanisme de régulation sont homologues pour les deux souches (Zhang et al ., 2001).

Un phénomène de régulation similaire à celui des β-lactamases de *S. aureus* et *B. licheniformis* a été observé chez certaines souches de *S. aureus* présentant un mode de résistance supplémentaire à la méthicilline (Chambers, 1997). Le gène chromosomique *mecA* responsable de cette résistance code une PBP alternative (PBP2' ou PBP2a) de faible affinité pour les β-lactamines (Brown et Reynolds, 1980; Ubukuta *et al*., 1985; Matsushashi *et al*., 1986; Lim et Strynadka, 2002). Cette PBP assure la fonction de transpeptidation dans la biosynthèse du peptidoglycane quand les autres PBP sont inactivées par l'antibiotique. Les souches possédant ce type de résistance (MRSA) posent de graves problèmes de santé publique (Levy et Marshall, 2004 ; Figure 10). La synthèse des PBP2' est régulée par des éléments génétiques analogues à ceux intervenant dans la régulation des β-lactamases BlaP et BlaZ (Himeno *et al*., 1986; Kobayashi *et al*., 1987; Rowland et Dyke, 1990).

Dans les paragraphes qui suivent, nous allons décrire en détail le mode de régulation de la β-lactamase BlaP de *B. licheniformis*, qui fait l'objet de ce travail de thèse. Ce mode de régulation sera comparé aux systèmes de régulation de la β-lactamase et des PBP2' de la souche pathogène *S. aureus* en soulignant les nombreuses similarités et les quelques différences.

- **Eléments génétiques**

La Figure 14 donne une représentation schématique de l'organisation des gènes impliqués dans la régulation des gènes de résistance aux antibiotiques chez *B. licheniformis* et *S. aureus.* Les β−lactamases BlaP et BlaZ de *B. licheniformis* et *S. aureus* sont codées par les gènes *blaP* et *blaZ,* respectivement. La régulation des gènes est très similaire pour les deux souches, et fait intervenir au moins trois gènes: le gène *blaR1* qui code pour un récepteur membranaire BlaR1, le gène *blaI* qui code pour le répresseur BlaI, et le gène *blaR2* (Gregory *et al.*, 1997; Philippon *et al.*, 1998; Sherratt et Collins, 1973). Les gènes *blaI* et *blaR1* sont organisés en un opéron, et l'opéron *blaI-blaR1* forme un divergeon avec le gène *blaP* ou *blaZ*. Le gène *blaR2* n'est pas lié au divergeon, et son rôle n'a pas encore été identifié (Filée *et al.*, 2002). On peut noter que, dans la littérature, le système *bla* est parfois également appelé *pen* chez *B. licheniformis* (Wittman *et al.*, 1993).

De façon similaire, les souches *S. aureus* présentant une résistance à la méthicilline possèdent un divergeon *mecA*/*mecI-mecR1* au sein de leur chromosome. Le gène *mecA* code pour la PBP2' de faible affinité pour les antibiotiques. Il forme un divergeon avec les deux autres gènes régulateurs *mecI* et *mecR1*. Les protéines codées par ces gènes sont similaires à celles des gènes *blaI* et *blaR1*. Le gène *mecR2* semble également être nécessaire à l'induction du gène *mecA* (Archer et Bosevilac, 2001).

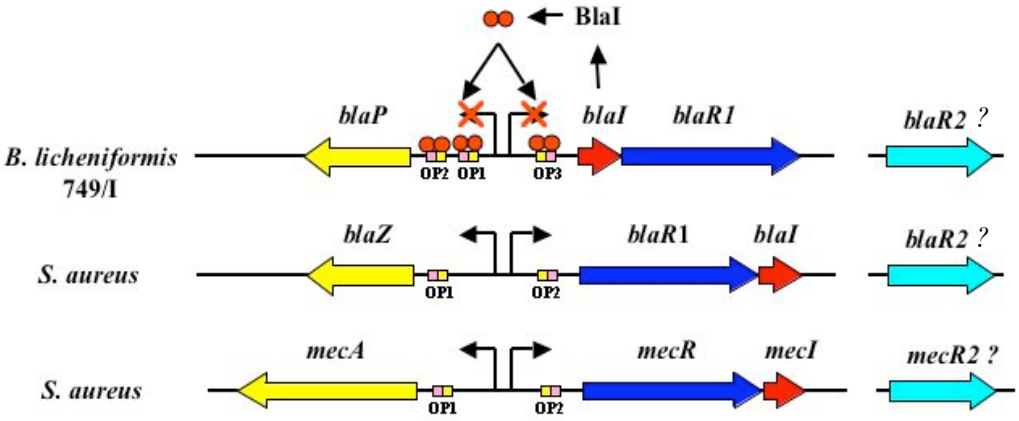


Figure 14 : Représentation schématique de l'organisation des divergeons impliqués dans la synthèse de la β-lactamase BlaP chez *B. licheniformis* et BlaZ chez *S. aureus*, et dans la synthèse de PBP2' chez *S. aureus*. En absence d'antibiotique, la transcription est inhibée par la fixation des répresseurs BlaI et MecI sur les régions opératrices (Op).

Chez *B. licheniformis*, les gènes *blaP/blaI/blaR1* sont chromosomiques. Les gènes *blaR1* et *blaI* sont organisés en un opéron dans lequel *blaI* est situé en amont de *blaR1*. Dans la région intergénique entre *blaP* et l'opéron *blaI-blaR1,* il existe trois régions opératrices (Op_1, Op_2, et Op_3) où la protéine BlaI peut venir se fixer sous forme de dimère. Les opérateurs Op_1 et Op_2 contiennent les promoteurs du gène *blaP*, Op_3 celui de l'opéron *blaI/blaR1*. La fixation de BlaI sur ces régions se fait de façon indépendante, non coopérative, et n'implique pas de courbure prononcée de l'ADN (Clarke *et al.*, 2001; Gregory *et al.*, 1997; Vreuls, 2004).

Chez *S. aureus*, l'organisation des gènes *blaR1-mecR1* et *blaI-mecI* dans l'opéron est inversée. Les gènes *mecA/mecI/mecR1* sont chromosomiques, alors que les gènes *blaZ/blaI/blaR1* sont soit chromosomiques, soit plasmidiens. La région intergénique contient uniquement deux opérateurs Op_1 et Op_2 qui chevauchent les promoteurs du divergeon. La fixation de MecI sur les différents promoteurs se fait de façon coopérative et induit une courbure de l'ADN (Sharma *et al.*, 1998).

La Figure 15 donne l'alignement de séquence des opérateurs pour les trois systèmes. On peut y remarquer l'existence de motifs palindromiques conservés.

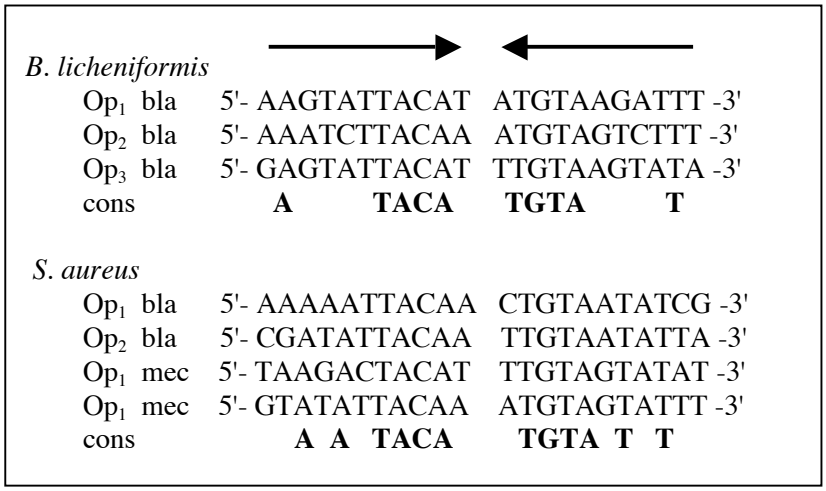


Figure 15 : Alignement de séquences des opérateurs Op_1, Op_2 et Op_3 de *B. licheniformis* et des opérateurs *bla* et *mec* de *S. aureus*. Les doubles flèches →← représentent les séquences opératrices palindromiques répétées inverses. La séquence consensus (cons) des opérateurs de *B. licheniformis* est similaire à celle rencontrée chez *S. aureus*. La répression du gène *blaP* chez *B. licheniformis* est plus stricte que celle des gènes *blaZ/mecA* chez *S. aureus* car elle s'effectue par l'intermédiaire de trois opérateurs.

III. 2. Acteurs protéiques

Cette partie décrit les acteurs protéiques impliqués dans la régulation de la β-lactamase BlaP de *B. licheniformis*. Les homologies avec les protéines impliquées dans la régulation de la β-lactamase BlaZ et des PBP2' de *S. aureus* seront soulignées.

- **Les récepteurs membranaires BlaR1 et MecR1**

Le gène *blaR1* code pour un récepteur transmembranaire BlaR1 (Kobayashi *et al.*, 1987; Joris *et al.*, 1990) qui a pour fonction de détecter la présence de pénicilline dans le milieu extracellulaire et de transmettre un signal à l'intérieur de la cellule. Le récepteur BlaR1 de *B. licheniformis* est une protéine de 601 acides aminés (Imanaka *et al.* 1987; Hardt *et al.*, 1997) composée de deux domaines fonctionnels (Figure 16).

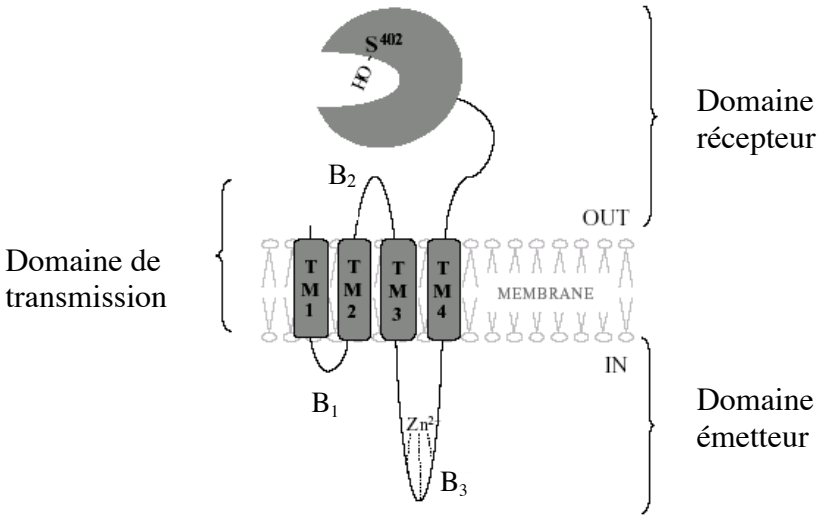


Figure 16: Topologie du récepteur BlaR1 de *B. licheniformis* (Hanique *et al.*, 2004).

Le domaine C-terminal (BlaR-CTD : résidus 346-601) possède les motifs protéiques SXXK, Y/SGN, et KTG, qui sont caractéristiques des protéines liant la pénicilline (Joris *et al.*, 1990 ; Joris *et al.*, 1994). Il interagit de manière covalente avec les β-lactamines du milieu extérieur (Duval *et al.*, 2003; Joris *et al.*, 1990). La structure de BlaR-CTD de *B. licheniformis* a été résolue par cristallographie et est similaire à celle d'une β-lactamase de classe D (Kerff *et al.*, 2003 ; Wilke *et al.*, 2004).

Le domaine N-terminal (BlaR-NTD) comprend quatre segments hélicoïdaux TM_1-TM_4 formant une superstructure à quatre hélices (Figure 16). Ces segments transmembranaires délimitent trois boucles B_1-B_3. B_1 et B_3 sont exposées dans le cytoplasme alors que B_2 pointe à l'extérieur de la cellule. La boucle cytoplasmique B_3 contient les motifs HEXXH et EXXXD caractéristiques des métalloprotéases à zinc (Hardt *et al.*, 1997). Ce domaine est responsable de la transmission d'un signal intracellulaire lorsque l'antibiotique a été détecté par la partie C-terminale.

L'alignement de la structure primaire de BlaR1 de *B. licheniformis* avec ses deux homologues BlaR1 et MecR1 de *S. aureus* montre que le domaine BlaR-CTD est bien conservé chez les trois récepteurs membranaires. Le pourcentage d'acides aminés conservés dans le couple BlaR1 *S. aureus*/MecR1 *S. aureus* est de 30 %. Il est de 21% pour le couple BlaR1 *B. licheniformis*/BlaR1 *S. aureus* et de 23 % pour le couple BlaR1 *B. licheniformis*/MecR1 *S. aureus*. De plus, les profils d'hydrophobicité de ces trois récepteurs sont superposables ce qui suggère que leur topologie est identique. Au niveau de la boucle B_3, on constate que le motif métalloprotéase à zinc HEXXH est strictement conservé parmi les trois récepteurs (Joris *et al.*, 1994).

- **Les répresseurs BlaI et MecI**

La protéine BlaI de *B. licheniformis* peut venir se fixer spécifiquement sous forme de dimère sur les régions opératrices Op_1, Op_2 et Op_3 qui portent les promoteurs des gènes *blaZ*, *blaI* et *blaR1* (Figure 14; Filée *et al.*, 2001). BlaI est donc un répresseur qui régule la synthèse de la β-lactamase BlaP, ainsi que sa propre synthèse et celle de BlaR1 (Salerno et Lampen, 1986 ; Grossman *et al.*, 1987 ; Wittman et Wong, 1988 ; Gregory *et al.*, 1997). Chez *S. aureus*, les protéines BlaI et MecI jouent également le rôle de répresseur en venant se fixer sur les régions opératrices Op_1 et Op_2 (Figure 14).

Le répresseur BlaI de *B. licheniformis* possède 128 acides aminés et une masse moléculaire de 15 034 Daltons (Himeno *et al.*, 1986). Des expériences de délétion ont montré que BlaI est composé de deux domaines fonctionnels distincts qui peuvent être séparés par digestion à la papaïne (Wittman *et al.*, 1993). Le domaine N-terminal (BlaI-NTD : résidus 1-82) est responsable de la liaison à l'ADN. Il reconnaît ses régions opératrices avec une affinité moindre (500 à 1000 fois) que la protéine entière et n'est plus capable de réprimer la synthèse de la β-lactamase (Wittman *et al.*, 1993). Au vu de la fonction et de la séquence de la protéine, il a été proposé que ce domaine contienne un motif hélice-coude-hélice (Nicholls et Lampen, 1987 ; Wittman *et al.*, 1993). Le domaine C-terminal (BlaI-CTD: résidus 83-128) est impliqué dans l'homodimérisation du répresseur (Wittman *et al.*, 1993 ; Filée *et al.*, 2003). Le répresseur BlaI de *B. licheniformis* est capable de former un homodimère mais également des formes multimériques d'ordre plus élevé. Il a été montré que la dimérisation est essentielle pour conférer à BlaI son activité de répresseur (Wittman *et al.*, 1993 ; Filée *et al.*, 2003).

La concentration cytoplasmique de BlaI *in vivo* a été estimée dans les souches *B. licheniformis* 749/I et *B. subtilis* 168, cette dernière ayant été transformée par le plasmide pMK4RIP porteur du locus β-lactamase *blaP-blaI-blaR1* à haut nombre de copies. Elle est égale à 1.9 μM chez *B. licheniformis* 749/I et à 75 μM chez *B. subtilis* 168 (Filée *et al.*, 2003). En comparant la concentration cytoplasmique en BlaI à la valeur de la constante de dissociation du dimère estimée par tamis moléculaire (25 μM), on constate que BlaI est principalement sous forme de monomère dans le cytoplasme de *B. licheniformis* (21 % dimère), et essentiellement sous forme de dimère dans le cytoplasme de *B. subtilis* (75 % dimère).

BlaI fait partie des répresseurs de la famille LacI (Vartak *et al.*, 1991). Les protéines homologues à BlaI de *B. licheniformis* sont données dans la Figure 17.

Souches	Séquence 1 10 20 30 40 50 60
BlaI (*B. licheniformis*)	MKKIPQISDAELEVMKVIWKHSSINTNEVIKELSKTSTWSPKTIQTMLLRLIKKGALNHH
BlaI (*B. cereus*)	--QLPKISEAELEIMKVLWSSSPQTANEIIEELEDAMDWKPKTIRTLINRLVQKEAVSYH
BlaI (*B. anthracis*)	--QLPKISEAELEIMKVLWLKSPQTANEIIEELEDPMDWKPkTIRTLINRLVQKEAVSYH
BlaI (*S.aureus*)	-----EISMAEWDVMNIIWDKKSVSANEIVVEIQKYKEVSDKTIRTLITRLYKKEITKRY
BlaI (*E. faecalis*)	- KLATITDAEWEVMRVIWTQGKTISREVHTLLNEKKEWKSTTVKTLLSRLTDKGLVGTE
MecI (*S. sciuri*)	--KIPQISDAELEVMKVIWKHSSINTNEVIKELSKTSTWSPKTIQTMLLRLIKKGALNHH
MecI (*S. aureus*)	--KTYEISSAEWEVMNIIWMKKYASANNIIEEIQMQKDWSPKTIRTLITRLYKKGFIDRK
MecI (*S. epidermidis*)	--KTYEISSAEWEVMNIIWMKKYASANNIIEEIQMQKDWSPKTIRTLITRLYKKGFIDRK
CopY (*Strept*)	--QMTSISNAEWEVMRVIWVAKQMTSSSKIIAILSRTYCWSATIKTLITRLSEKGYLTSQ
CopY (*E. hirae*)	-----KISDSEWEVMRVIWTLGQANAQQITQILADSMDWKVATVKTLIGRLVKKEAIWTE

Souches	Séquence 70 80 90 100 110 120
BlaI (*B. licheniformis*)	KEGRVFVYTPNIDESDYIEVKSHSFLNRFYNGTLNSMVLNFLENDQLSGEEINELYQILEEHKNRKKE
BlaI (*B. cereus*)	DKGRMYAYYPLVSQHNYLQVETKSLLKRFCGAAFKPLLVNFLKEEKLSSEDINELKLRLDEKTEENKRK
BlaI (*B. anthracis*)	QDKGRMYYYPLVSQDNYLQVETKSLLKRFCGAAFKPLLVNFLKEEKLSSEDINELKRLLDEKTEENKRK
BlaI (*S.aureus*)	KSENIYFYSSNIKEDDIKMKTAKTFLNKLYGGDMKSLVLNFAKNEELNNKEIEELRDILND-------
BlaI (*E. faecalis*)	KAGNKYIYYPLVEERQSVETVADELLAKVCSRKVGSVVENFLTESQLSYDDIDRLEELLKEKRKTAVE
MecI (*S. sciuri*)	KEGRVFVYTPNIDESDYIEVKSHSFLNRFYNGTLNSMVLNFLENDQLSGEEINELYQILEEH
MecI (*S. aureus*)	KDNKIFQYYSLVEESDIKYKTSKNFINKVYKGGFNSLVLNFVEKEDLSQDEIEELRNILNKK*N----
MecI (*S. epidermidis*)	KDNKIFQYYSLVEESDIKYKTSKNFINKVYKGGFNSLVLNFVEKEDLSQDEIEELRNILNKK*N----
CopY (*Strept*)	RQGRKYIYSSLISEEEALEQQVSEVFSRICVTKHQALIRRLVEETPMTLSDIEKLEALLLSKK-----
CopY (*E. hirae*)	QEGKKFIYHPAVSEMENVRSATENLFSHICAKRVGATIADLVEEATLTQEDIQQLMKQLNK-------

Figure 17: Alignement de structure primaire des protéines homologues à BlaI de *B. licheniformis*. Les résidus strictement conservés sont indiqués en rouge. Les résidus fortement et faiblement similaires sont surlignés en vert et bleu, respectivement. Les répresseurs BlaI et MecI de *S. aureus* présentent une identité de 60 % tandis qu'ils présentent une identité respective de 31 et 41 % avec le répresseur BlaI de *B. licheniformis* (Lewis et Dyke, 2000).

L'analyse de l'alignement de séquence de la Figure 17 nous montre que les répresseurs BlaI et MecI identifiés dans divers génomes bactériens sont très homologues au répresseur BlaI de *B. licheniformis*. On peut noter également que les protéines CopY présentent un fort taux d'homologie avec les répresseurs BlaI et MecI. CopY est un répresseur de l'opéron *cop* inductible au cuivre (Strausak et Solioz, 1997). Jusqu'en 2003, aucune structure n'avait été résolue pour l'ensemble de ces protéines.

- **Les autres acteurs protéiques**

Parmi les acteurs protéiques essentiels pour la répression et l'induction des gènes codant pour la β–lactamase et les PBP2', les protéines BlaI / MecI et BlaR1 / MecR1 sont de mieux en mieux caractérisées. L'implication d'autres protéines non encore identifiées est fortement probable, en particulier en ce qui concerne la levée de la répression des gènes (Filée et al, 2002). La transformation de la souche *Bacillus subtilis* BD224 par un plasmide de *B. licheniformis* contenant le divergeon *bla* conduit à la production régulée de la β-lactamase, indiquant que tous les gènes nécessaires à l'induction sont présents dans cette souche recombinante. En revanche, une souche recombinante de *E. coli* transformée par le même plasmide conduit à une production de β-lactamase non inductible (Kobayashi *et al.*, 1987).

Par ailleurs, le rôle des gènes *blaR2* / *mecR2* n'a pas encore été identifié. Des expériences de mutagenèse ont cependant montré que *blaR2* est indispensable à l'induction de la synthèse de la β–lactamase (Dubnau et Pollock, 1965; Sherratt et Collins, 1973; Filée *et al.*, 2002). Chez *B. licheniformis*, le mutant blaR2$^-$ se caractérise par une β–lactamase non-inductible. Chez *S. aureus*, le gène *mecR2* semble également être nécessaire à l'induction du gène *mecA* (Archer et Bosevilac, 2001). On peut donc soupçonner que des acteurs protéiques pas encore connus (BlaR2, MecR2, et/ou d'autres) jouent un rôle dans la régulation des systèmes *bla* et *mec*.

III. 3. Mécanismes de répression et d'induction

- **Description générale**

La Figure 18 donne une vision schématique du mécanisme de répression et d'induction du gène *blaP* chez *B. licheniformis*. En l'absence d'antibiotique à noyau β–lactame, le répresseur BlaI est lié sous forme de dimère à ses trois régions opératrices localisées dans la région intergénique entre *blaP* et l'opéron *blaI-blaR1,* et réduit notablement la transcription des gènes *blaI, blaR1* et *blaP* (Himeno *et al.*, 1986; Salerno et Lampen, 1986; Grossman *et al.*, 1987).

Lorsque les β–lactamines arrivent en contact avec la bactérie, elles se lient au domaine C-terminal extracellulaire de BlaR1. Cette interaction provoque la transmission d'un signal transmembranaire qui a pour conséquence le clivage du motif métalloprotéase de la boucle B_3 entre les résidus R293-R294 (Zhang *et al.*, 2001). La métalloprotéase ainsi activée conduit, par un mécanisme qui n'est pas encore bien décrit, au détachement du répresseur BlaI des trois régions opératrices du divergeon. Le produit du gène *blaR2* est peut-être impliqué dans la transmission de ce message intracellulaire. Dès que les régions promotrices des gènes sont libérées du répresseur, la transcription des gènes est activée, et un haut niveau d'expression de la β–lactamase est observé (Filée *et al.*, 2003).

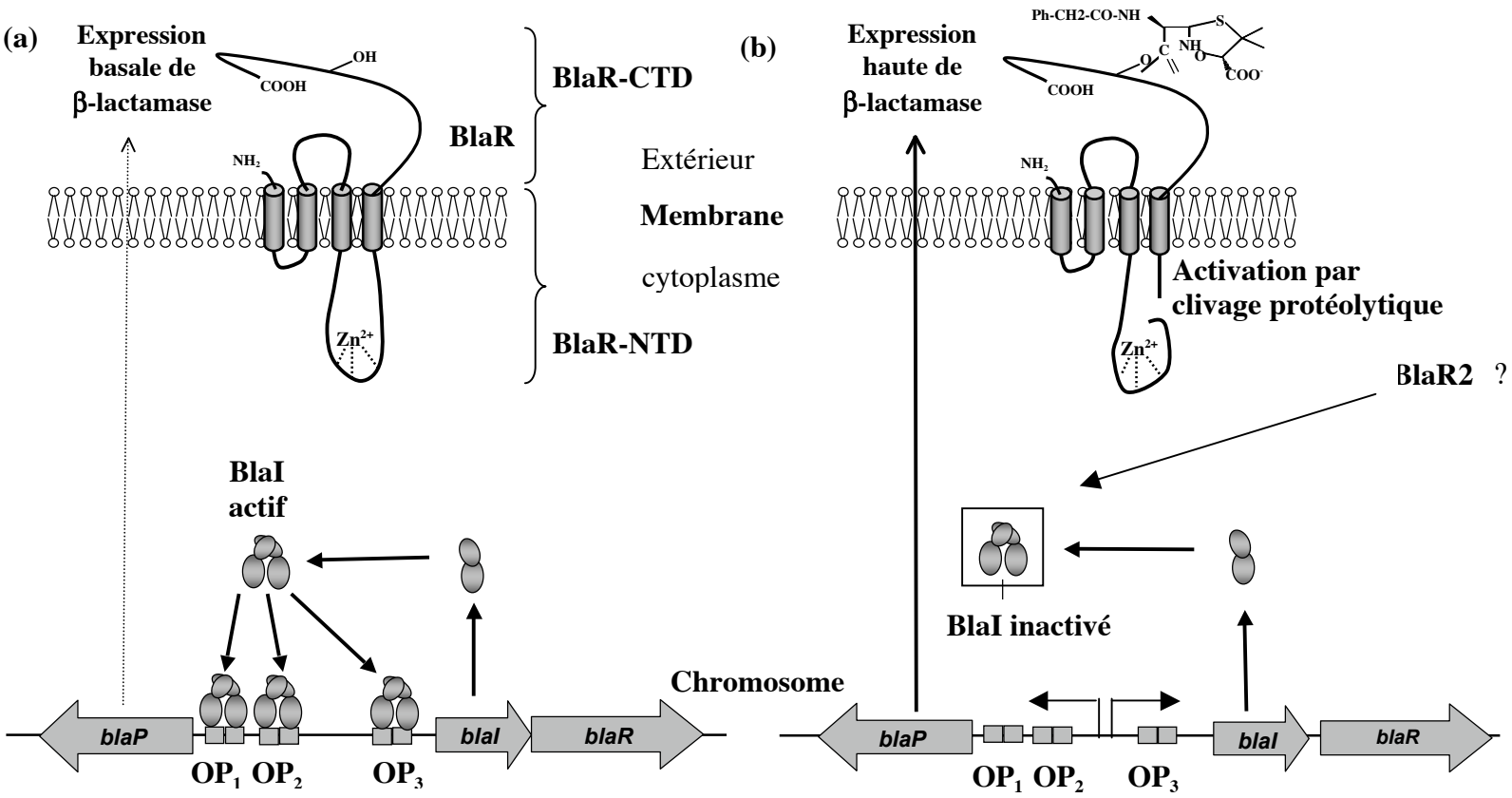


Figure 18: Mécanisme de régulation de la β–lactamase BlaP chez *B. licheniformis* : (a) Répression du gène *blaP* en l'absence d'antibiotique dans le milieu, et (b) Induction du gène en présence de β–lactamine.

- **Une grande similarité de régulation …**

Chez *S. aureus*, le mécanisme d'induction de la synthèse de la β-lactamase BlaZ est similaire au mécanisme d'induction de la β–lactamase BlaP rencontré chez *B. licheniformis.* Pour la production des PBP2', la transcription du gène *mecA* est sous le contrôle du répresseur MecI et du récepteur transmembranaire MecR1 (Song *et al.*, 1987; Hiramatsu, 1992). La protéine MecR2 semble également être nécessaire à l'induction du gène *mecA* (Archer et Bosevilac, 2001).

Le mécanisme de transduction du signal par les récepteurs membranaires est encore mal connu, mais il semble être similaire pour les β-lactamases de *B. licheniformis* et *S. aureus*. Pour les deux souches, une interaction entre le domaine C-terminal de BlaR1 et sa boucle B_2 a été mise en évidence en l'absence d'antibiotique (Hanique *et al.*, 2004 ; Wilke *et al.*, 2004). Cette interaction maintient les quatre segments transmembranaires dans une conformation telle que la boucle B_3 est inactive. En présence d'antibiotique, BlaR-CTD est acylé, mais ne change pas de conformation (Hanique *et al.*, 2004 ; Wilke *et al.*, 2004). L'interaction entre BlaR-CTD et la boucle B_2 est relâchée, ce qui entraîne un mouvement des hélices transmembranaires et la transmission de l'information dans le cytoplasme par l'intermédiaire de l'activation de la boucle métalloprotéase B_3 (Hanique *et al.*, 2004). Chez *S. aureus,* les résultats de Wilke sont néanmoins différents de ceux observés par Golemi-Kotra, pour qui un changement significatif de conformation du domaine BlaR-CTD est observé lors de son acylation par l'antibiotique (Golemi-Kotra *et al.*, 2003).

L'homologie entre les systèmes de répression des gènes *bla* et *mec* de *S. aureus* est telle que les répresseurs MecI et BlaI sont capables de coréprimer la production de PBP2' et de β–lactamase (Lewis *et al.*, 2000; Clarke *et al.*, 2001). De plus, il est possible de former des hétérodimères BlaI-MecI *in vitro* (Mc Kinney *et al.*, 2001 ; Sharma *et al.*, 1998). On peut cependant noter que chez *S. aureus*, il n'existe que deux sites opérateurs, contre trois pour *B. licheniformis.* Le site supplémentaire répresse la transcription du gène *blaP*. La fuite de production de β–lactamase est donc contrôlée plus sévèrement chez *B. licheniformis* que chez *S. aureus* en l'absence d'antibiotique.

- **… Mais deux mécanismes d'inactivation du répresseur différents ?**

Le mécanisme de levée de la répression des gènes *blaP*, *blaZ* et *mecA* de *B. licheniformis* et *S. aureus* est encore peu caractérisé. Récemment, quelques éléments de réponse ont été apportés, qui semblent montrer que les gènes de résistance des deux souches ont des mécanismes d'induction différents, malgré la grande homologie de leur régulation. La différence entre les deux modèles qui ont été proposés pour *S. aureus* et *B. licheniformis* se situe au niveau du mécanisme d'inactivation du répresseur.

Chez *S. aureus*, Zhang *et al.* ont proposé une inhibition du répresseur BlaI par protéolyse de celui-ci au cours de l'induction. Le clivage de BlaI a été localisé entre les résidus N_{101} et F_{102} (Lewis *et al.*, 1999; Zhang *et al.*, 2001). Il est à noter que ce site de clivage est conservé pour les protéines BlaI et MecI de toutes les souches connues, mais qu'il est absent chez CopY (Figure 17). Selon Zhang, l'acylation de BlaR-CTD par une β-lactamine entraîne le clivage de la boucle B_3 de BlaR-NTD entre les résidus R293 et R294. La métalloprotéase de la boucle B_3 ainsi activée entraine l'inactivation de BlaI par protéolyse ce qui permet de libérer la transcription du gène *blaZ* codant pour la β-lactamase. La transduction du signal chez *S. aureus* implique donc la protéolyse de BlaR1, puis celle de BlaI. Ce mode d'induction a déjà été rapporté, par exemple pour le répresseur λcI (Johnson *et al.*, 1980) ou la protéine σ^E (Hofmeister *et al.*, 1995).

Chez *B. licheniformis*, l'induction de la β–lactamase ne résulte pas d'une inactivation du répresseur BlaI par un clivage protéolytique. En effet, aucune dégradation de la protéine BlaI n'est observée lors de l'induction de la β–lactamase chez *B. licheniformis* (Wittman, *et al.*, 1993), ni dans la souche recombinante *B. subtilis* contenant le divergeon *blaP/blaI-blaR1* (Filée *et al.*, 2002). Ces résultats suggèrent la présence d'un ligand comme coactivateur (Filée *et al.*, 2002; Filée *et al.*, 2003; Duval *et al.*, 2003). Ce coactivateur, généré suite à l'acylation de BlaR1 et au stress de la pénicilline, induit un changement de conformation du répresseur BlaI dimère, diminuant ainsi l'affinité du dimère de BlaI pour son ADN cible. Ce mécanisme a déjà été décrit dans de nombreux mécanismes de régulation, par exemple pour la régulation de la tétracycline chez les bactéries à Gram négatif (Orth *et al.*, 2000).

IV. Objectifs

L'objectif de ce travail est l'étude structurale du répresseur BlaI de *B. licheniformis* 749/I, seul et en interaction avec la séquence opératrice du gène de la β-lactamase BlaP, en vue de mieux comprendre le système de répression et d'induction de la production de la β-lactamase chez *B. licheniformis*. L'étude est également motivée par la similitude du mode de régulation de la β-lactamase de *B. licheniformis* avec ceux de la β-lactamase et des PBP2' de la souche hautement pathogène *S. aureus.*

Ce travail a été fait en collaboration très étroite et interactive entre le laboratoire de RMN de l'IBS de Grenoble et le groupe de Bernard Joris du Centre d'Ingénierie des Protéines de l'université de Liège (CIP). La partie biochimique de ce projet ainsi que les expériences de spectroscopie autres que celles de Résonance Magnétique Nucléaire ont été faites exclusivement à Liège, en particulier par Christelle Vreuls.

Mon implication dans les résultats que je vais présenter concerne toutes les parties faisant intervenir la RMN, ma participation à la production et à la purification d'un échantillon de protéine à Liège, et aussi mon implication constante dans la réflexion commune sur les expériences à mener. Certaines expériences réalisées par nos collaborateurs, et dont le résultat a eu un impact sur l'étude effectuée à Grenoble, seront données et analysées brièvement dans la partie 'Résultats' en mentionnant le nom de leurs auteurs.

N. B. Les parties 'Introduction' et 'Résultats' de ce chapitre ne font pas mention des structures cristallographiques obtenues en 2003 et 2004 pour MecI et BlaI de *S. aureus* par les groupes de X. Gomis-Rüth et de G. Archer, respectivement. Ces résultats, qui ont été publiés conjointement ou après la publication de nos propres résultats pour BlaI de *B. licheniformis*, apportent un éclairage complémentaire à notre travail qui sera commenté dans la partie 'Discussion' de ce chapitre.

Résultats : Etude structurale du répresseur BlaI du gène de la β-lactamase chez *B. licheniformis* 749/I

I. Production et caractérisation des échantillons protéiques

Cette partie décrit comment les différents échantillons utilisés dans ce travail ont été produits et caractérisés de façon préliminaire pour orienter l'étude structurale.

I. 1. La protéine BlaI sauvage

La protéine BlaI de *B. licheniformis* est produite sous forme soluble dans le cytoplasme de la souche *E . coli* B121(DE3) ou B121(DE3)plysS transformée par le plasmide pET-22b(+) contenant le gène *blaI*. Pour la production des protéines enrichies en ^{15}N , ^{13}C et ^{2}D, des milieux M9 ou OD2-CN de Silantes ont été utilisés (Vreuls, 2004). Après la lyse des bactéries, l'échantillon protéique est purifié en deux étapes, par chromatographie sur colonne échangeuse de cations et par chromatographie d'affinité, selon un protocole décrit précédemment (Filée *et al*., 2002 ; Vreuls, 2004). Les échantillons sont conditionnés dans un tampon phosphate pH 7.6 : phosphate de sodium 50 mM – KCl 200 mM – azoture de sodium 1 mM – péfabloc 0.1 mM.

Une analyse par électrophorèse sur gel de polyacrylamide (SDS-PAGE) montre que la protéine obtenue est pure (Figure 19a). Aux concentrations en protéine nécessaires pour effectuer une étude RMN, BlaI présente néanmoins des problèmes de solubilité dans le tampon utilisé pour la purification. L'homogénéité de la solution de protéine a été analysée par diffusion lumineuse dynamique et ultracentrifugation analytique (Vreuls, 2004). Les résultats montrent que la solution est hétérogène, et que des formes monomériques, dimériques et tétramériques coexistent en solution. Une étude en microgouttes a été réalisée (Lepre et Moore, 1998), dont les résultats ont montré que l'addition au tampon de 300 mM de glycine et de 240 mM de NaCl permet d'augmenter la solubilité de BlaI à des concentrations de l'ordre de 1 mM.

Une première caractérisation structurale a été effectuée par RMN en enregistrant le spectre ^{1}H-^{15}N HSQC de la protéine BlaI enrichie en ^{15}N et placée dans les conditions de tampon optimisées (Figure 20). Les spectres ^{1}H-^{15}N HSQC représentent en effet la signature propre d'une protéine, comportant pour chaque acide aminé de la séquence un pic de corrélation aux fréquences proton et azote des atomes ^{1}H-^{15}N de la liaison peptidique. Le nombre de pics de corrélation ^{1}H-^{15}N attendus est normalement égal au nombre de résidus non prolines dans la séquence. Leur éclatement, leur intensité et leur largeur de raie permettent d'avoir des informations sur l'état de structuration et d'oligomérisation de la protéine en solution. Dans le spectre ^{1}H-^{15}N HSQC de BlaI, on peut constater le manque de sensibilité du spectre et la largeur des raies de résonance. De plus, il n'est pas possible de dénombrer le nombre de pics attendus (125), alors que le centre du spectre comporte une superposition de pics non résolus dont les fréquences de résonance sont caractéristiques d'un fragment de protéine dénaturé ou présentant une dynamique défavorable à une étude RMN.

Des essais de changement de température (de 5 à 35°C) et de pH (pH 5.5 à 7.5) ont été effectués, de manière à modifier l'état d'oligomérisation de la protéine et la structuration ou la gamme de dynamique des résidus dont les pics sont non résolus. Toutes les conditions testées ont mené à des spectres similaires à celui de la Figure 20a. La qualité des spectres RMN de la protéine sauvage BlaI n'a pas été jugée suffisante pour envisager son étude structurale par RMN à l'état natif.

I. 2. Les deux domaines fonctionnels de BlaI

Le répresseur BlaI de *B. licheniformis* est composé de deux domaines fonctionnels: un domaine N-terminal (BlaI-NTD) impliqué dans l'interaction avec l'ADN et un domaine C-terminal (BlaI-CTD) responsable de la dimérisation (Wittman *et al.*, 1993). Pour mieux comprendre l'allure du spectre de la protéine sauvage, nous avons produit et caractérisé les deux domaines fonctionnels de BlaI.

La protéine sauvage purifiée est digérée à la papaïne. Après une nuit, le clivage est quasi total et deux fragments de 10 et 5 kDa sont générés (Figure 19a; Vreuls, 2004). Le fragment N-terminal (BlaI-NTD) dont le point isoélectrique est égal à 10 est purifié par chromatographie sur une colonne échangeuse de cations SP Sepharose Fast Flow (Figure 19b). Le rendement en BlaI-NTD purifié est de 8 mg/L de culture. Le fragment BlaI-CTD n'est pas élué, en dépit de son point isoélectrique largement inférieur au pH du tampon (pI = 5.1). Diverses méthodes ont été testées pour produire et purifier BlaI-CTD (Vreuls, 2004). La solution qui a été retenue est de fusionner un hexapeptide d'histidine à l'extrémité C-terminale de la protéine sauvage pour pouvoir la purifier. La production de la protéine dans *E. coli*, sa digestion par la papaïne et la purification de la partie C-terminale ont conduit à des rendements très faibles. Un tampon contenant un mélange imidazole – CHAPS 13 mM a été choisi pour éviter l'agrégation de la protéine. L'étude du domaine C-terminal est décrite dans la partie II. 2 de ce chapitre.

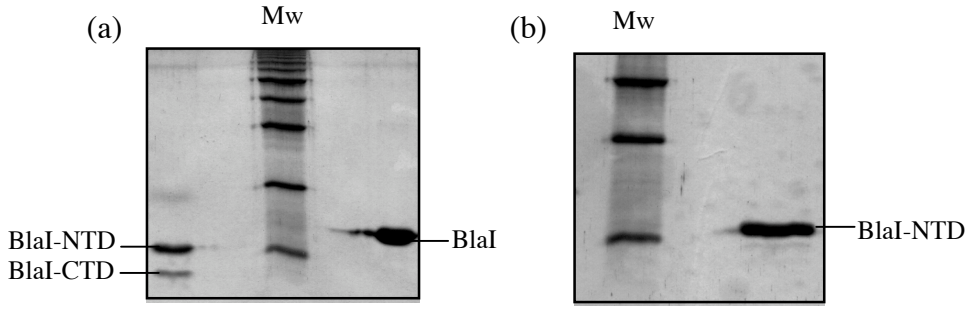


Figure 19 : Gel SDS-PAGE de la protéine BlaI digérée à la papaïne et du fragment BlaI-NTD purifié (Vreuls, 2004). (a) BlaI (5μg) est digéré durant une nuit à 28 °C par 50 ng de papaïne dans un tampon Hepes 50 mM - NaCl 100 mM - EDTA 1 mM - pH 7.6. Les fragments BlaI-NTD (10 kDa) et BlaI-CTD (5 kDa) sont générés. (b) BlaI-NTD concentré après purification sur la colonne SP Sepharose Fast Flow. Mw = marqueur de poids moléculaire.

Le séquençage NH_2-terminal de BlaI-NTD et l'analyse par spectrométrie de masse MALDI-TOF révèlent que le site de clivage à la papaïne de la protéine BlaI se situe entre les résidus Ser82 et His83 et non entre Leu86-Asp87, comme le prédisait Wittman (Wittman *et al.*, 1993).

Une première caractérisation structurale de BlaI-NTD a été faite par RMN en comparant le spectre ^{1}H-^{15}N HSQC de la protéine entière à celui de BlaI-NTD (Figure 20). La sensibilité du spectre BlaI-NTD est nettement supérieure à celle de la protéine entière. Les expériences de pontage intermoléculaire au DSP (ou dithiobis(succinimidylpropionate)) et d'ultracentrifugation analytique (Vreuls, 2004) confirment que le domaine N-terminal du répresseur BlaI a perdu sa capacité à dimériser (Wittman *et al.*, 1993).

Une autre caractéristique doit être soulignée dans la comparaison entre les deux spectres : la majorité des pics de corrélation présents dans le spectre de BlaI-NTD se retrouvent à la même position dans le spectre de la protéine entière. Seuls les quelques résidus proches du site de clivage à la papaïne diffèrent entre les deux spectres. Cette similarité démontre que le domaine N-terminal de BlaI est structuralement indépendant du domaine C-terminal. Ce module fonctionnel de liaison à l'ADN (Wittman *et al.*, 1993) est donc également un module structural à part entière. Ce résultat a été confirmé par l'étude de dénaturation faite dans la partie II. 2.

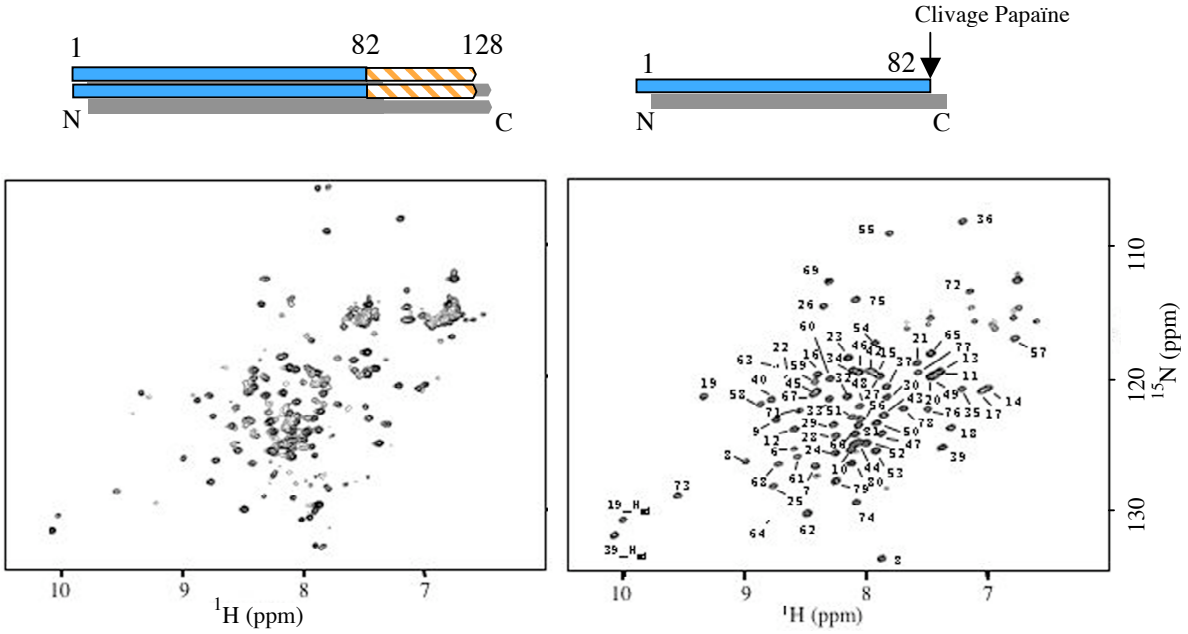


Figure 20: Comparaison des spectres ^{1}H-^{15}N HSQC (a) de la protéine BlaI entière et (b) du domaine N-terminal de BlaI. L'attribution des résidus 6 à 81 est annotée sur le spectre de BlaI-NTD. Les spectres ont été enregistrés à 25°C sur un spectromètre Varian INOVA 800. Une représentation schématique de la longueur des molécules étudiées est donnée au-dessus des spectres.

I. 3. Mutagenèse dirigée

Pour modifier l'état d'oligomérisation et de dynamique de l'échantillon, et améliorer les spectres de l'extrémité C-terminale de BlaI, des études de mutagenèse dirigée ont été entreprises. La position des mutations a été guidée par les prédictions de structures secondaires effectuées pour BlaI-CTD (voir II. 2).

Dans un premier temps, nous avons voulu caractériser la protéine sous forme monomérique. Pour cela, un mutant 112IA/115LA/119LA de BlaI a été produit, où les acides aminés I112, L115 et L119 ont été mutés en alanine de manière à déstabiliser l'interface de dimérisation. Les expériences d'ultracentrifugation analytique et de pontage au DSP réalisées sur ce mutant montrent qu'il a effectivement perdu sa capacité à dimériser (Vreuls, 2004). Cependant, le spectre de dichroïsme circulaire dans l'UV lointain montre que le mutant est partiellement dénaturé, et son étude a été abandonnée.

Dans un deuxième temps, nous avons testé si la mauvaise qualité des spectres RMN de BlaI-CTD provenait de l'existence d'un échange conformationnel au niveau du domaine de dimérisation. Pour figer BlaI-CTD dans une unique conformation, un mutant BlaI115LC dont l'acide aminé L115 a été substitué en cystéine a été produit. Un pont disulfure intermoléculaire entre les résidus cystéines 115 des deux monomères a été formé en oxydant les cystéines par dialyse de l'échantillon pendant 24 h en présence d'oxygène contre le tampon RMN additionné d'iode à 0.5 mM. Le dosage des SH libres au DTNB révèle qu'environ 80 % des SH sont pontés. Les monomères résiduels sont éliminés par passage sur un tamis moléculaire Superdex 75 conditionné en iode 0.5 mM (Vreuls, 2004).

Le spectre ^{1}H-^{15}N HSQC d'un échantillon du mutant BlaI115LC enrichi en ^{15}N a été enregistré et analysé. Les pics de corrélation de la partie N-terminale sont superposables à ceux du spectre de la protéine sauvage ou de BlaI-NTD (Figure 20). Concernant la partie C-terminale, aucun pic nouveau n'a été détecté par rapport au spectre du sauvage, et la majorité des corrélations a été observée au centre du spectre. La production de ce mutant n'a donc pas permis d'envisager l'étude structurale par RMN de la forme longue de la protéine. En revanche, des conclusions structurales ont été tirées de son étude, et seront discutées dans le paragraphe II. 2.

Au vu de la qualité des spectres de la protéine BlaI sous sa forme sauvage et des résultats de l'étude de mutagenèse dirigée, nous avons choisi d'étudier les deux domaines fonctionnels de BlaI séparément. Cette étude est justifiée par l'indépendance structurale des deux domaines. La suite de ce manuscrit décrit comment la structure de la partie N-terminale de BlaI a été résolue par RMN. La caractérisation de la partie C-terminale a ensuite été effectuée par un ensemble de techniques complémentaires. La dernière partie de ce chapitre décrit les résultats structuraux et thermodynamiques concernant le mode d'interaction de BlaI avec l'ADN.

II. Etude structurale du répresseur BlaI

L'étude structurale du répresseur BlaI a été faite en deux étapes. La structure du domaine d'interaction à l'ADN a tout d'abord été résolue par RMN. Le domaine de dimérisation a été ensuite caractérisé par différentes techniques.

II. 1. Structure de BlaI-NTD

- **Détermination de la structure de BlaI-NTD par RMN**

L'attribution des résonances et la détermination de la structure de BlaI-NTD ont été effectuées de façon classique par RMN sur des échantillons marqués isotopiquement en ^{15}N, $^{15}N/^{13}C$, et $^{15}N/^{13}C/^{2}H$. La liste des expériences RMN réalisées pour l'ensemble de l'étude est donnée Figure 21. L'originalité de l'étude concerne l'utilisation d'une expérience 4D ^{1}H-^{13}C HSQC-NOESY-HSQC (Kay *et al.*, 1990), qui a permis de résoudre l'ambiguïté de certains pics de corrélation des spectres NOE. Les détails expérimentaux de l'ensemble de cette étude peuvent être consultés dans l'article donné en Annexe 3 (Van Melckebeke *et al.*, 2003).

L'attribution des résidus de la chaîne principale de BlaI-NTD a été réalisée grâce aux expériences triple résonance HNCA, HNCO, HN(CA)CO, CBCANH, CB(CO)NH enregistrées sur des échantillons enrichis en ^{15}N et en $^{15}N/^{13}C$. Une expérience HNCA a également été enregistrée sur un échantillon marqué $^{15}N/^{13}C/^{2}H$ en vue de confirmer l'attribution du domaine N-terminal de la protéine entière. Les fréquences de résonance $^{1}H/^{13}C$ des chaînes latérales de BlaI-NTD ont été attribuées grâce à l'analyse simultanée des spectres H(CCO)NH-TOCSY, HC(CO)NH-TOCSY, CBCANH, CBCA(CO)NH enregistrés sur la protéine enrichie en $^{15}N/^{13}C$. Pour attribuer les fréquences de résonance des résidus aromatiques, un spectre 2D-NOESY dans D_2O et un spectre H(C)CH-TOCSY de la protéine enrichie en $^{15}N/^{13}C$ ont été utilisés.

Ces expériences ont permis d'attribuer la quasi-totalité des fréquences de résonance $^{15}N/^{13}C/^{1}H$ de la chaîne principale et des chaînes latérales de BlaI-NTD (Gans *et al.*). Seuls quelques résidus correspondant aux extrémités N et C-terminales de BlaI-NTD (résidus 1-4, 82) n'ont pas été identifiés. L'attribution des pics de corrélation du spectre ^{1}H-^{15}N HSQC de BlaI-NTD est indiquée dans la Figure 20, et celle du spectre ^{1}H-^{13}C HSQC des méthyles dans la Figure 34. L'ensemble des déplacements chimiques de BlaI-NTD a été déposé dans la banque de donnée BioMagResBank (http://www.bmrb.wisc.edu/) sous le code **5873**.

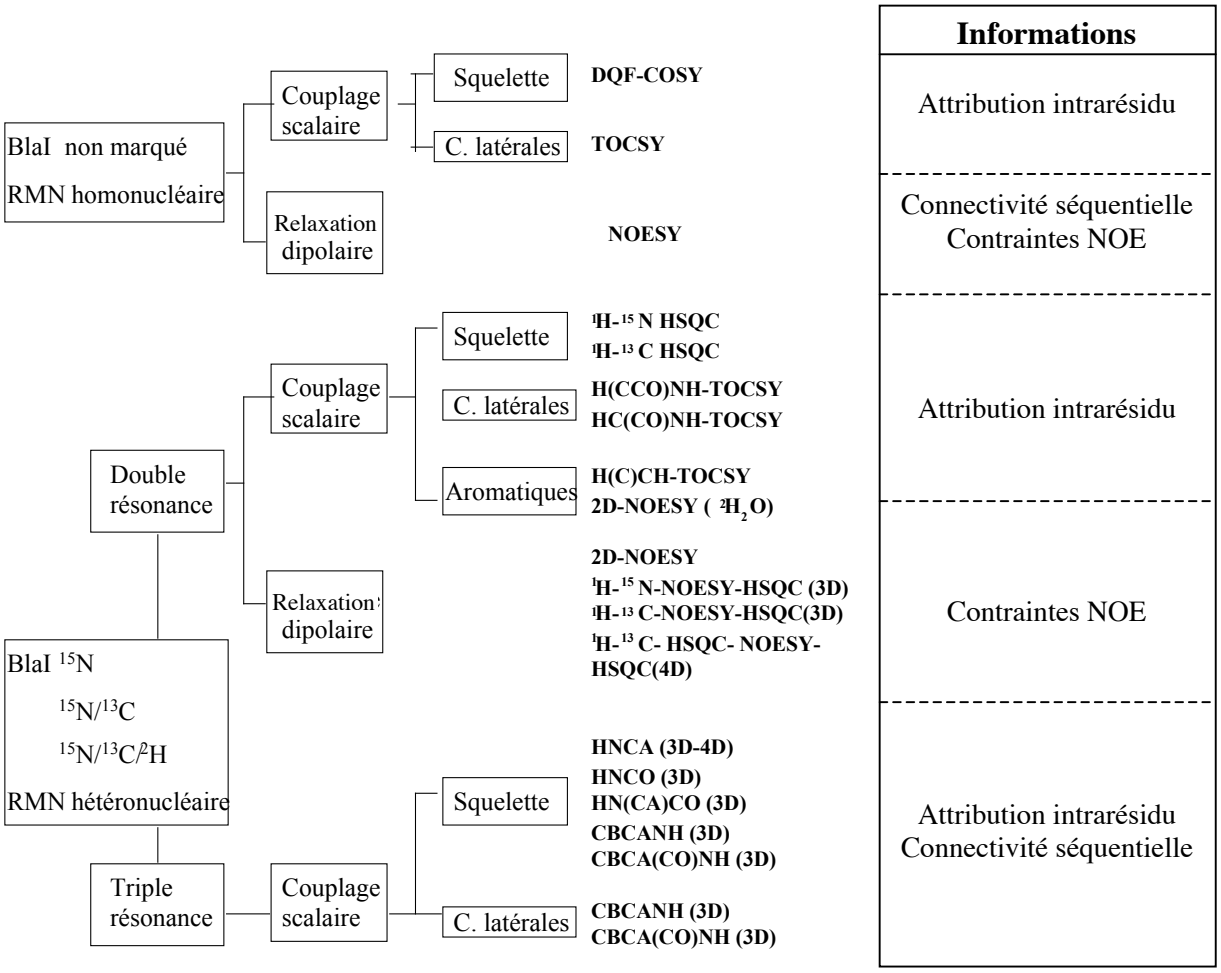


Figure 21 : Liste des expériences RMN réalisées pour déterminer la structure tridimensionnelle de BlaI-NTD. L'ensemble des expériences a été effectué à 25°C, sur des spectromètres Varian INOVA 600 et INOVA 800. Le tampon de conditionnement des échantillons est un tampon phosphate pH 7.6 : phosphate de sodium 50 mM – KCl 200 mM – azoture de sodium 1 mM – péfabloc 0.1 mM. Pour les échantillons contenant la forme longue de la protéine, 300 mM de glycine et 240 mM de NaCl ont été ajoutés.

Les éléments de structures secondaires de la protéine ont été identifiés par calcul d'index de déplacement chimique (Wishart et Sykes, 1994). L'analyse statistique des déplacements chimiques permet en effet de détecter les motifs de structure secondaire d'une protéine par comparaison du déplacement chimique de chaque noyau ^{13}C, ^{15}N, et $^{1}H_{\alpha}$ de la protéine à un déplacement chimique de référence correspondant au même noyau placé dans une conformation aléatoire. La différence observée est spécifique de la famille de structure secondaire, hélicoïdale ou en brin, à laquelle appartient le résidu où se trouve ce noyau. Les calculs d'index de déplacement chimique (CSI) obtenus pour BlaI-NTD sont illustrés à la Figure 23 pour les noyaux $^{13}C_{\alpha}$, ^{13}CO, et $^{13}C_{\beta}$. Une valeur de CSI consensus a été calculée à partir des valeurs de CSI de ces différents atomes. La structuration secondaire de BlaI-NTD déduite de celle valeur consensus comprend trois hélices α (9-20, 26- 36, 41-53) ainsi que deux brins β (57-62, 65-70).

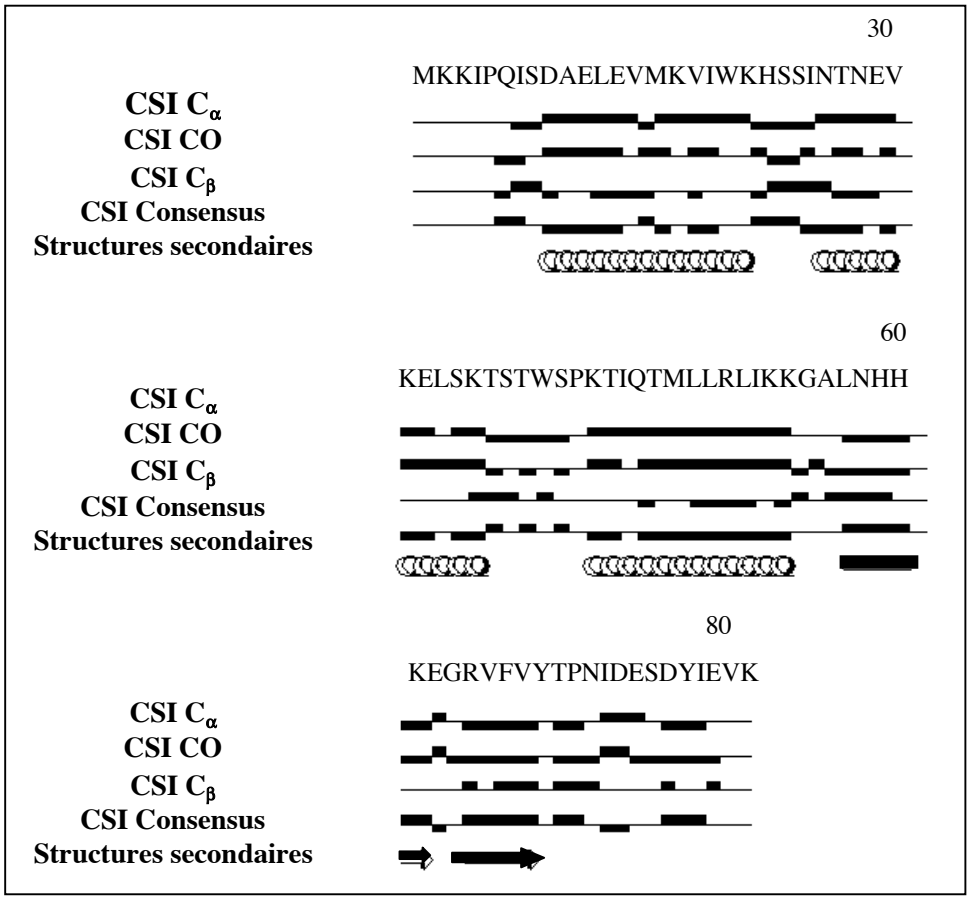


Figure 22 : Variation des fréquences de résonance des noyaux $^{13}C_{\alpha}$, ^{13}CO, $^{13}C_{\beta}$ de BlaI-NTD par rapport aux valeurs moyennes de leur déplacement chimique dans des conformations aléatoires. Par convention, une valeur CSI (Chemical Shift Index) égale à (1) ou (–1) est attribuée à un atome si sa valeur de déplacement chimique est significativement supérieure ou inférieure à son déplacement chimique de référence (Wishart et Sykes, 1994). Pour les noyaux $^{13}C_{\alpha}$ et ^{13}CO, un groupe possédant au minimum quatre valeurs consécutives de (1) est caractéristique d'une hélice α et un groupe d'au moins trois valeurs consécutives de (-1) est caractéristique d'un brin β. Pour les noyaux $^{13}C_{\alpha}$, un groupe possédant au minimum quatre valeurs consécutives de (-1) est caractéristique d'un brin β. Les petits rectangles au-dessus et au-dessous des lignes représentent les valeurs (1) et (-1) de l'index de déplacement chimique. La structure secondaire de BlaI-NTD déduite du consensus CSI est représentée le long de la séquence en acides aminés.

La détermination de la structure de BlaI-NTD a été effectuée en utilisant un procédé de dynamique moléculaire classique en RMN des biomolécules (Discover, Accelrys). Cette technique repose sur la minimisation d'une fonction d'énergie cible composée d'un terme d'énergie théorique associé à un champ de force (Amber) et d'un terme d'énergie expérimentale dictée par les données RMN. Le champ de force contraint la géométrie de la protéine en tenant compte des données chimiques connues pour un polymère peptidique (séquence en acides aminés, longueur des liaisons, forces de Van der Waals et de Coulomb,…). Les données RMN sont traduites en termes de contraintes structurales d'angles ou de distances et permettent de déduire le repliement global de la protéine. La structure de la protéine à haute résolution est obtenue en combinant les contraintes issues du champ de force et les contraintes expérimentales.

L'ensemble des contraintes expérimentales utilisées pour le calcul de structure de BlaI-NTD est donné dans le Tableau 3. Elles sont de trois types :

- Nous avons fixé la chiralité des 97 centres chiraux des 82 acides aminés de BlaI-NTD à celle qui est toujours observée dans la nature.

- Des contraintes de distance ont été déduites de la quantification des pics de corrélation proton-proton dans des expériences NOESY. L'effet NOE permet en effet de transférer l'aimantation entre deux protons situés à moins de 5 **Å** l'un de l'autre par interaction dipolaire, et cet effet peut être quantifié. Une expérience homonucléaire 2D-NOESY, des expériences 3D NOESY-HSQC éditées en ^{15}N ou en ^{13}C, et une expérience 4D HSQC-NOESY-HSQC éditée en ^{13}C ont été enregistrées avec un temps de mélange de 120 ms. Cette dernière a permis de résoudre des ambiguïtés qui se sont révélées cruciales pour obtenir le premier repliement de la protéine. À partir de ce premier repliement, l'attribution des NOE ambigus a été facilitée. Au final, ces expériences ont permis de déterminer un total de 1252 distances interprotons, dont 596 intrarésidus, 205 séquentielles, 87 de distance moyenne, 145 à longue portée, et 219 ambiguës.

- Les valeurs des angles dièdres ϕ et ψ définissant l'orientation des plans peptidiques entre eux ont été déduites des valeurs des déplacements chimiques grâce au programme TALOS (Cornilescu *et al.*, 1999). Ce programme s'appuie sur une étude statistique de la corrélation entre les valeurs des déplacements chimiques et les valeurs des angles ϕ et ψ, sur l'ensemble des structures à haute résolution de la PDB. 110 Angles ont ainsi été définis de façon non ambiguë. Ces contraintes locales sont importantes dans l'étape d'affinement de la structure.

Ces contraintes ont été introduites dans le calcul de structure de BlaI-NTD. La première étape du calcul de dynamique moléculaire a été répétée 80 fois. Les 19 structures de plus basse énergie ont été affinées par dynamique moléculaire et recuit simulé. Elles forment l'ensemble RMN final de BlaI-NTD, visualisé Figure 23. Les coordonnées des 19 conformères RMN de plus basse énergie de BlaI-NTD ont été déposées dans la banque de donnée PDB sous le code **1P6R**.

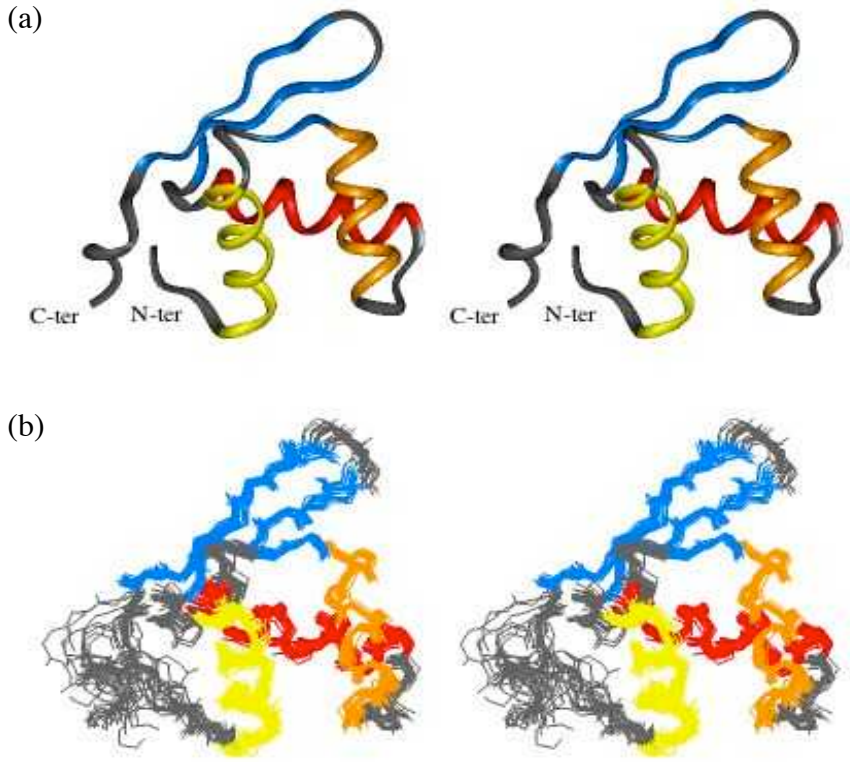


Figure 23: Structures RMN de BlaI-NTD. Les structures secondaires H_1, H_2 et H_3 sont représentées respectivement en jaune, orange et rouge tandis que les feuillets β sont représentés en bleu. (a) Vues en stéréo de la structure de plus faible énergie. (b) Vues en stéréo de la superposition des 19 structures de plus faible énergie. Seuls les atomes lourds de la chaîne principale sont représentés.

Quand on analyse les structures de l'ensemble RMN, on s'aperçoit que la conformation des résidus 1 à 4 et 79 à 82 est mal résolue. Ceci est dû au manque d'attribution des fréquences de résonance des résidus 1-4 et 82, et au faible nombre de contraintes NOE identifiées pour les résidus 79-81. Les valeurs des NOE héteronucléaires ^{1}H-^{15}N pour les résidus 79-81 laissent supposer que la partie C-terminale (79-82) est déstructurée (Figure 24).

La qualité de l'ensemble final de la structure RMN de BlaI-NTD a été analysée à l'aide du programme PROCHECK-NMR (Laskowski *et al.*, 1996). Plus de 99 % des angles dièdres ϕ et ψ de BlaI-NTD sont situés dans les régions énergétiquement favorables du diagramme de Ramachandran. L'énergie expérimentale est faible ce qui suggère une bonne coïncidence avec les contraintes expérimentales. Aucune violation de distance supérieure à 0.1 Å n'est observée. Pour les résidus 5 à 78 qui correspondent au corps rigide de la protéine, 16.4 contraintes de distance ont été utilisées en moyenne par résidu. Pour ces résidus, le r.m.s.d (root mean square deviation) des atomes du squelette peptidique et de tous les atomes lourds par rapport à la structure moyenne de l'ensemble final RMN est respectivement de 0.63 ± 0.17 Å et 1.07 ± 0.18 Å. Le résumé des statistiques structurales est donné dans le Tableau 3.

Contraintes RMN	1362
Interprotons	1252
Intrarésidu	596
Interrésidu	437
séquentiel	205
distance moyenne	87
longue distance	145
liaison hydrogène	0
ambigu	219
Angle dièdre	110
Chiralité	97
r.m.s.d (déviation par rapport aux coordonnées moyennes (Å)[a,b])	
Atomes lourds du squelette peptidique	0.63 ± 0.17
Tous les atomes lourds	1.07 ± 0.18
Diagramme de Ramachandran (%)[b,c]	
Résidus dans régions favorables	85.6
Résidus dans régions autorisées	13.7
Résidus dans régions supplémentaires autorisées	0.2
Résidus dans régions interdites	0.5

[a] Statistique sur les résidus 5-78
[b] Moyenne des 19 structures de plus faible énergie
[c] Analyse avec le programme PROCHECK-NMR (les glycines et les prolines ne sont pas prises en considération)

Tableau 3: Contraintes et statistiques structurales pour le calcul de structure de BlaI-NTD de *B. licheniformis*

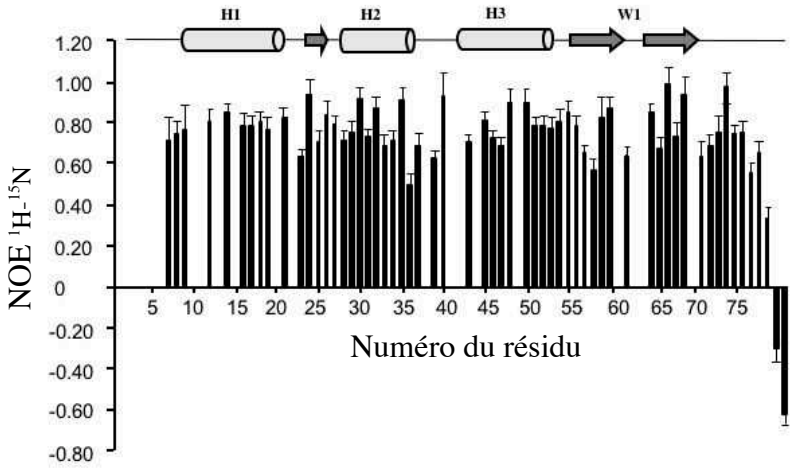


Figure 24: Valeurs des NOE hétéronucléaires ^{1}H -^{15}N du fragment BlaI-NTD à 25°C. Les expériences sont enregistrées sur un spectromètre Varian INOVA 600 avec saturation des protons pendant une période de 3 s selon la séquence WALTZ-16. Le temps de récupération est de 7 s et la durée totale de l'expérience de 42 h pour un échantillon de concentration 0.5 mM.

- **BlaI, une protéine à motif 'winged helix'**

La structure de BlaI-NTD (Figure 25) est composée de trois feuillets β (S_1:Ser23-Asn25, S_2: Leu57-Glu62, S_3: Val65-Pro70), de trois hélices α (H_1: Asp9-Lys20, H_2: Thr26-Thr36, H_3: Pro41-Lys53) et d'une boucle W_1 (Gly63-Arg64) organisés de la manière suivante: H_1 - S_1 - H_2 - T_1 - H_3 - S_2 - W_1 - S_3, où le coude T_1 relie H_2 à H_3 (Van Melckebeke *et al.*, 2003 : Annexe 3). L'extrémité N-terminale de BlaI-NTD est largement hélicoïdale alors que l'extrémité C-terminale est composée de deux brins β antiparallèles (S_2 et S_3) reliés par une boucle W_1 à la manière d'une aile de papillon. On peut noter que le premier brin S_1 du feuillet est court, antiparallèle au brin S_3, et n'avait pas été prédit sur la base des déplacements chimiques (voir plus haut).

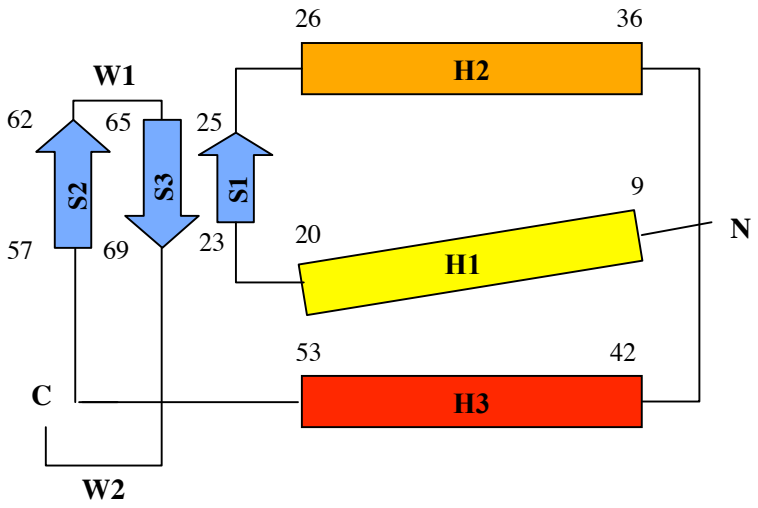


Figure 25: Organisation des structures secondaires du motif 'winged-helix' du domaine N-terminal de la protéine BlaI. Les hélices α sont représentées en jaune orange et rouge, et les feuillets β en bleu.

Cette structure montre que le répresseur BlaI de *B. licheniformis* appartient à la famille des protéines à motif 'winged-helix' (Pour revue : Gajiwala et Burley, 2000). Ce motif a été identifié pour la première fois chez la protéine eucaryote HNF/3 (Lai *et al.*, 1993). Le motif 'winged helix' contient un motif central hélice-coude-hélice, identifié ici à H_2-T_1-H_3. Les protéines à motif 'winged helix' constituent donc une sous-famille de la famille hélice-coude-hélice (Brennan et Matthews, 1989). La structuration de BlaI en motif hélice-coude-hélice proposée par Wittman sur la base de la séquence de la protéine est donc bien confirmée. Cependant, la longueur du coude T_1 reliant H_2 et H_3 différencie les protéines à motif 'winged helix' des protéines à motif hélice-coude-hélice classiques.

La structure 3D de BlaI-NTD a été comparée avec celle d'autres protéines à motif 'winged-helix' grâce à l'algorithme DALI (Holm et Sander, 1993). Malgré l'absence totale d'homologie de séquence, BlaI-NTD présente une homologie de structure importante avec les protéines E2F4 (Zheng *et al.*, 1999), ADAR1 (Schade *et al.*, 1999), LexA (Knegtel *et al.*, 1995), MarR (Alekshun *et al.*, 2001), FokI (Wah *et al.*, 1997), DTXR (Chen *et al.*, 2000), SmtB (Cook *et al.*, 1998) et Genesis (Jin *et al.*, 1998). Pour toutes ces protéines, le domaine contenant le motif 'winged-helix' est un domaine de liaison à l'ADN. Certaines protéines à motif 'winged-helix' ont été identifiées dans les bactéries (DtXr, MarR, FokI, SmtB, LexA), et d'autres dans les organismes eucaryotes (E2F4, Genesis). Ce sont des régulateurs de la transcription des gènes, impliqués dans des mécanismes variés comme par exemple la résistance aux antibiotiques (MarR) ou aux métaux lourds toxiques (SmtB), la production de toxines diphtériques (DtXr), la restriction de l'ADN (FokI), ou la régulation des gènes du cycle cellulaire (E2F4). L'homologie de structure importante entre BlaI-NTD et la protéine E2F4 est illustrée Figure 26. D'une manière générale, lorsque l'on compare les structures des protéines à motif 'winged-helix', on constate que l'orientation relative des structures secondaires est variable. Ceci est dû à la différence de longueur du coude T_1 (reliant H_2 et H_3) et/ou de la boucle W_1 (Gajiwala et Burley, 2000). Citons par exemple, la protéine Genesis qui a une boucle W_1 longue et flexible tandis que la boucle W_1 de BlaI-NTD (Gly63 et Arg64) ne présente pas une dynamique particulière (voir NOE hétéronucléaires, Figure 24).

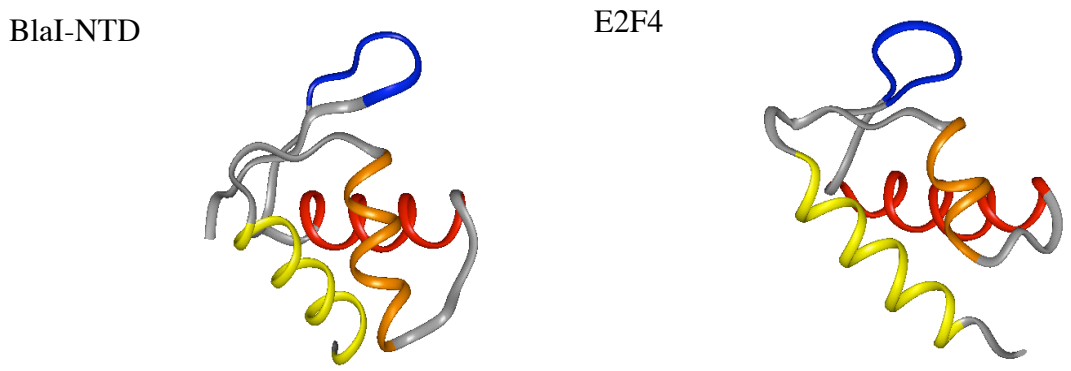


Figure 26: Comparaison des domaines à motif 'winged-helix' de BlaI (PDB 1P6R) et de la protéine E2F4 (PDB 1CF7).

Au vu de l'homologie de séquence de BlaI-NTD de *B. licheniformis* avec les domaines N-terminaux des protéines BlaI, MecI et CopY d'autres organismes (Figure 17), nous proposons que ces protéines possèdent elles aussi un repliement à motif 'winged-helix'. Les acides aminés impliqués dans le cœur hydrophobe sont en effet très bien conservés dans l'ensemble de ces protéines. L'étude structurale de BlaI de *B. licheniformis* permet donc la caractérisation structurale de toute une famille de répresseurs. Une comparaison des résultats de ce travail avec les résultats obtenus pour BlaI et MecI de *S. aureus* sera faite en détail dans la partie 'Discussion' de ce chapitre.

II. 2. Les mystères du domaine de dimérisation C-terminal de BlaI

Plusieurs options ont été envisagées pour l'étude du domaine de dimérisation de BlaI. La résolution des spectres RMN de BlaI sauvage n'étant pas suffisante pour envisager la détermination de la structure de l'ensemble de la protéine (voir paragraphe I), nous avons tout d'abord caractérisé la structuration du domaine de dimérisation isolé. Une étude de dénaturation a ensuite permis d'en étudier la stabilité et les propriétés hydrophobes. Enfin, l'étude du montant BlaI115LC a permis d'obtenir des informations supplémentaires sur la structure et la dynamique de BlaI-CTD.

• Structuration de BlaI-CTD

L'alignement de séquence des homologues de BlaI révèle la présence de motifs répétés de sept acides aminés au niveau des domaines C-terminaux (voir Introduction, Figure X). Ces motifs sont caractérisés par trois cycles de huit acides aminés dont le premier est une leucine (L100-L107-L115). Des prédictions de structures ont été faites pour BlaI-CTD à l'aide du programme GORIV (Garnier *et al*., 1996). Les résultats prédisent la présence d'hélices amphiphiles pour les segments Ser96-Glu103 et Gln106-Lys123 de BlaI (Vreuls, 2004). Cette hypothèse est en accord avec les résultats biochimiques concernant la production et la purification de BlaI-CTD$(His)_6$ décrites dans la partie I. Dans le tampon de l'étude, on suppose que BlaI-CTD multimérise et devient insoluble en l'absence de la partie N-terminale de la protéine. Le détergent (CHAPS) qui est ajouté pour solubiliser l'échantillon protéique permet de masquer les interactions hydrophobes présentes dans BlaI-CTD de façon à éviter la multimérisation.

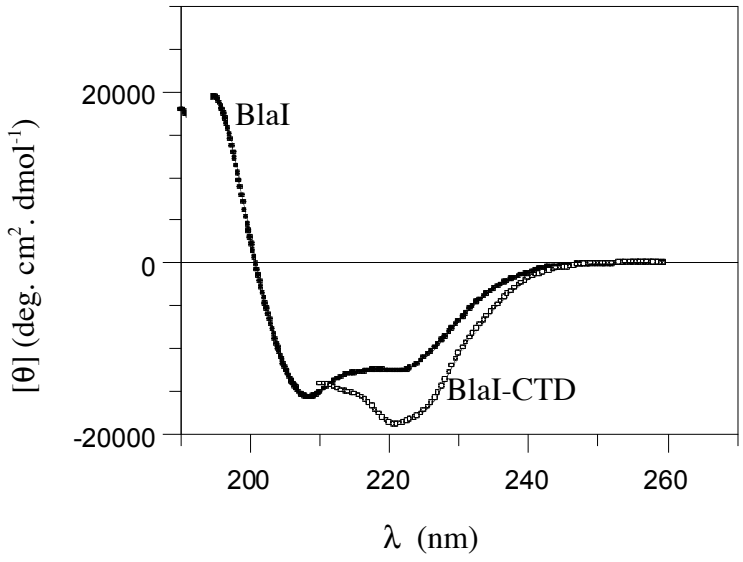


Figure 27: Spectres de dichroïsme circulaire de BlaI et BlaI-CTD. Les spectres ont été enregistrés sur un spectromètre Jobin-Yvon CD6 à 20 °C à concentration en protéine de 0.5 mg/mL.

Le spectre de dichroïsme circulaire dans l'UV lointain montre que BlaI-CTD(His)$_6$ est structuré (Figure 27). Le contenu en hélices α calculé sur la base de l'ellipticité molaire à 222 nm (Chang *et al.*, 1978) est de 53% pour BlaI-CTD(His)$_6$, contre 33% pour BlaI (Vreuls, 2004). De plus, les spectres de dichroïsme circulaire de BlaI et de BlaI-NTD sont similaires, et caractérisés par un maximum à 195 nm et deux minima à 208 et 222 nm caractéristiques des protéines ayant une proportion élevée de résidus en hélices α (Woody, 1985). Les minima sont néanmoins beaucoup plus marqués pour la protéine sauvage. On peut donc conclure que BlaI-CTD est structuré en grande partie sous la forme d'hélices α.

Une étude RMN du fragment BlaI-CTD(His)$_6$ a été envisagée. Malheureusement, des expériences de spectrométrie de masse MALDI-TOF et de pontage au DSP mettent en évidence la présence de formes multimériques de BlaI-CTD(His)$_6$ (Vreuls, 2004). Devant la difficulté de produire BlaI-CTD(His)$_6$, et l'hétérogénéité des échantillons même en présence de détergent, nous avons abandonné l'étude de la partie C-terminale de BlaI par RMN.

- **Etude de dénaturation au GdnHCl**

Pour obtenir des informations sur la stabilité des différents domaines structuraux de BlaI, une étude de dénaturation par le chlorure de guanidinium a été effectuée pour la protéine BlaI native et les deux domaines fonctionnels séparés. La dénaturation des protéines est provoquée par l'augmentation progressive de la concentration en dénaturant, et elle a été suivie par différentes techniques spectroscopiques. Cette étude a été publiée (Vreuls *et al.*, 2004), et l'article correspondant est donné dans l'Annexe 4. Nous allons rappeler ici certains résultats qui ont été obtenus à Grenoble par RMN, donner les conclusions principales de l'étude, et discuter ce qu'elles apportent pour une meilleure caractérisation de BlaI-CTD.

La dénaturation du domaine N-terminal de BlaI a été suivie à l'échelle de chaque résidu en comparant les spectres ^{1}H-^{15}N HSQC de BlaI et BlaI-NTD à diverses concentrations en GdnHCl (Figure 29). La comparaison des spectres montre que le domaine N-terminal de BlaI a le même comportement vis-à-vis de la dénaturation dans la protéine entière et dans le fragment BlaI-NTD. Ce résultat confirme que BlaI-NTD et BlaI-CTD sont structuralement indépendants (voir partie I).

L'ajout d'une concentration de 1 M en GdnHCl induit des variations de déplacement chimiques pour le proton de l'ordre de 0.8 ppm pour les extrémités N et C-terminales de BlaI-NTD (6-9, et 73-80), la boucle W1 (61-63), et le feuillet β3 (68-69). Cette modification est due à la variation d'environnement électrostatique des résidus exposés au solvant en présence de concentrations croissantes en sels. Le déplacement chimique des résidus du cœur de la protéine n'est pas modifié jusqu'à une concentration en GdnHCl de 4 M. À 5 M, tous les pics de corrélation de BlaI-NTD disparaissent pour laisser place à un spectre caractéristique de l'état dénaturé. Ces données montrent une dénaturation totale et coopérative du domaine de liaison à l'ADN de BlaI entre 4 et 5 M de GdnHCl. On peut donc en conclure que BlaI-NTD est un module structural stable et globulaire.

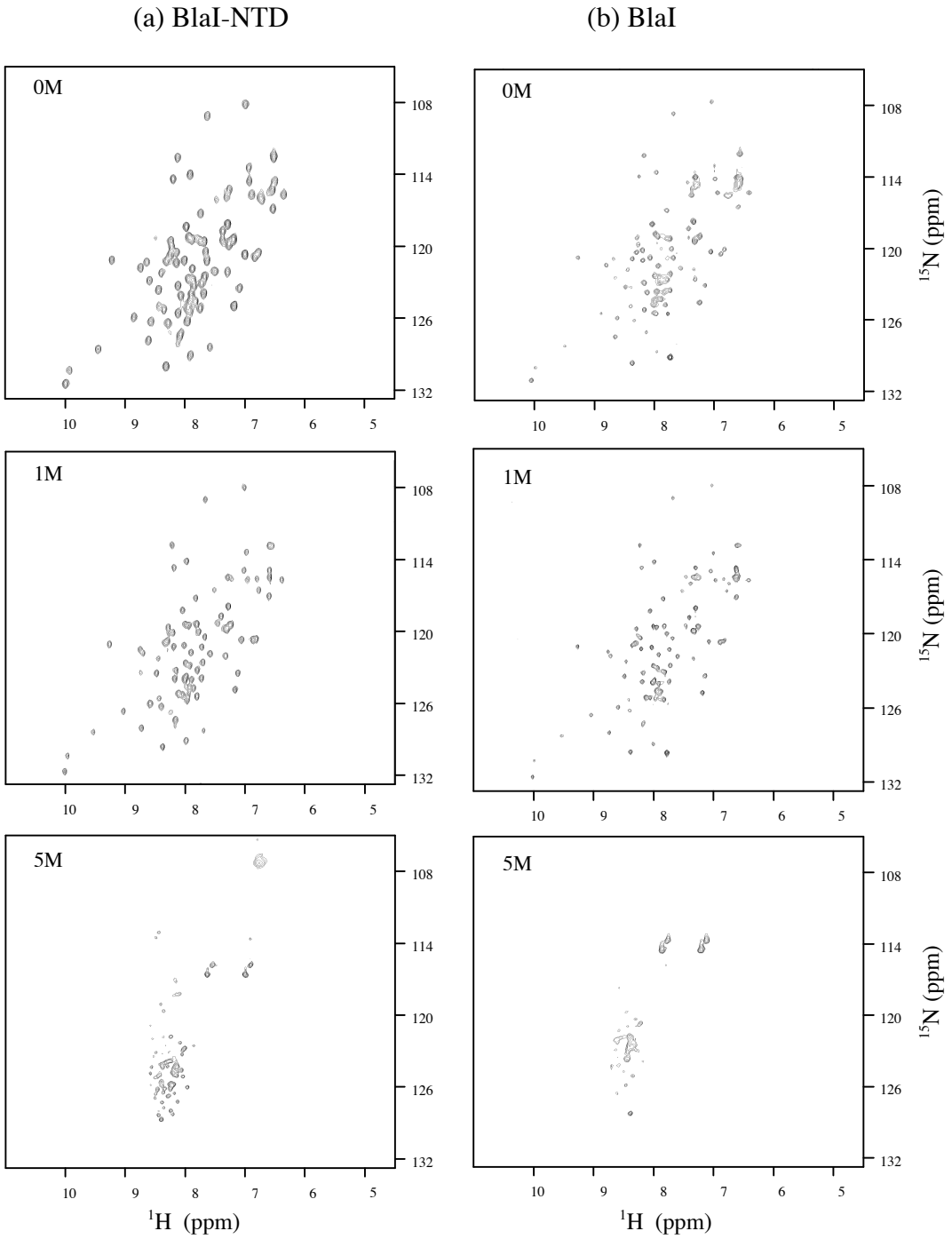


Figure 28 : Spectres ^{1}H-^{15}N HSQC de BlaI et BlaI-NTD. (a) BlaI (0.4 mM) et (b) BlaI-NTD (0.3 mM) sont dans un tampon 50 mM phosphate - 200 mM KCl – pH 7.6 additionné de la concentration adéquate en GdnHCl : 0 M GdnHCl (spectres en haut à gauche), 1 M GdnHCl (spectres en haut à droite) et 5 M GdnHCl (spectres en bas à droite). Les derniers spectres (en bas à gauche) montrent que la renaturation des protéines après élimination du GdnHCl est complète. L'intensité des pics de corrélation diminue pour des concentrations en agent dénaturant plus élevées, suite à la variation du facteur de qualité Q en présence de grandes quantités de sels (Gadian, 1984).

La dénaturation a également été suivie par fluorescence intrinsèque et fluorescence de l'ANS (8-anilino-1-naphtalène-sulfonique), et par dichroïsme circulaire. L'état d'oligomérisation lors de la dénaturation a été suivi par pontage au DSP et analyse par gel SDS-PAGE, tamis moléculaire, et ultracentrifugation analytique (Vreuls *et al.*, 2004). L'étude détaillée peut être consultée dans l'Annexe 4. La Figure 29 résume les différents états de transition observés pour la protéine BlaI au cours de sa dénaturation par les différentes techniques.

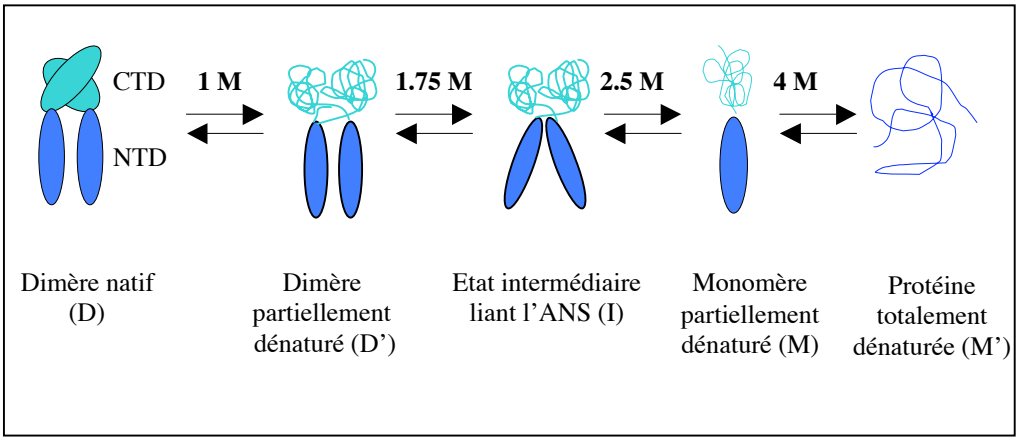


Figure 29: Schématisation du mécanisme de dénaturation du répresseur BlaI par le chlorure de guanidinium. Les domaines fonctionnels de BlaI se dénaturent de manière séquentielle et indépendante. Trois états intermédiaires sont formés : (i) un état où l'extrémité C-terminale est dénaturée mais encore capable d'assurer la dimérisation, et où la partie N-terminale est native (D'), (ii) un état intermédiaire dimérique semblable au précédent, où la partie C-terminale continue d'évoluer vers un état liant l'ANS (I), (iii) un état monomérique où la partie C-terminale est totalement dénaturée, mais où la partie N-terminale est encore intacte (M).

Les résultats de cette étude de dénaturation montrent que le domaine de dimérisation C-terminal de BlaI impliqué dans sa dimérisation est structuré à l'état natif, et qu'il se dénature à une concentration d'environ 1 M en GdnHCl. Il est donc thermodynamiquement moins stable que BlaI-NTD. Les expériences de pontage au DSP, de sédimentation de vitesse et de tamis moléculaire montrent que la transition dimère-monomère s'effectue autour de 2.5 M en agent dénaturant.

Enfin, l'étude de la fluorescence de l'ANS suggère que BlaI-CTD possède des cavités hydrophobes accessibles au solvant. Ces propriétés sont des témoins de la capacité de BlaI-CTD à pouvoir réagir facilement aux changements d'environnement, ce qui est sans doute mis en œuvre lors de l'induction.

- **Pontage covalent des domaines de dimérisation**

Malgré la structuration en hélices α de BlaI-CTD, les pics de corrélation des résidus de la partie C-terminale sont très peu résolus dans les spectres RMN de BlaI (Figure 20). Ce domaine présente donc une gamme de dynamique ou d'échange conformationnel qui peut jouer un rôle important lors de l'interaction du répresseur avec l'ADN, ou avec un potentiel inducteur pour la levée de la répression. Pour caractériser ce processus dynamique, nous avons essayé de le figer en pontant de façon covalente deux acides aminés appartenant à la surface de dimérisation. Le mutant BlaI115LC dont l'acide aminé leucine 115 a été substitué en cystéine a été produit dans cet objectif. La formation d'un pont disulfure entre les résidus cystéines 115 de chaque monomère a été obtenue par réduction des cystéines (voir I. 3).

Ce dimère covalent est capable de lier l'ADN avec la même affinité que la protéine sauvage, et le spectre de dichroïsme circulaire du mutant est superposable à celui de BlaI (Figure 27; Vreuls, 2004). Les structures secondaires n'ont donc pas été affectées par la mutation et la présence du pont disulfure, ce qui confirme que l'acide aminé 115 fait partie de la surface d'interaction du dimère.

Le spectre ^{1}H-^{15}N HSQC de cet échantillon a été enregistré et analysé. Les pics de corrélation de la partie N-terminale sont superposables à ceux du spectre de la protéine sauvage ou de BlaI-NTD (Figure 20). Concernant la partie C-terminale, aucun pic nouveau n'est détecté par rapport au spectre du sauvage, et la majorité des corrélations est observée au centre du spectre. On peut donc conclure que l'échange conformationnel ou la dynamique responsable du manque de résolution des spectres de BlaI-CTD n'est pas influencée par la liaison covalente entre les deux domaines de dimérisation.

L'étude de ce mutant a donc tout d'abord permis de montrer que la leucine 115 est impliquée dans la surface de dimérisation du répresseur BlaI. De plus, nous avons prouvé que la gamme de dynamique responsable du manque de résolution des spectres de BlaI-CTD n'implique pas des mouvements de grande amplitude autour de la position 115 de la séquence peptidique.

III. Caractérisation par RMN de l'interaction de BlaI avec l'ADN

Ce paragraphe décrit comment l'interaction du répresseur BlaI avec ses séquences opératrices a été caractérisée par RMN. Les résultats concernant le mode d'interaction de BlaI avec l'ADN et l'affinité de l'interaction monomère / demi-opérateur seront discutés.

III. 1. Matériel et méthode

- **Cartographie des variations de déplacement chimique**

Par RMN, on peut caractériser l'interaction entre deux partenaires très simplement si on connaît l'attribution des fréquences de résonance d'au moins un des partenaires. Pour cela, il suffit de faire une titration de ce partenaire enrichi avec les isotopes ^{15}N, ou ^{15}N-^{13}C par l'autre partenaire non marqué. Le suivi de l'évolution des pics de corrélation du partenaire marqué donnera une liste de résidus dont les atomes ont changé d'environnement chimique lors de l'interaction. Cette méthode, appelée cartographie des variations de déplacement chimique, a été utilisée pour caractériser l'interaction de BlaI et BlaI-NTD avec différents fragments d'ADN en titrant la protéine marquée par l'ADN non marqué.

Les séquences d'ADN qui ont été choisies pour l'étude sont données dans la Figure 30.

Op1 *B. lich* (30 pb)	5' – G AAA GTA T**TA CA**T A**TG TA**A GAT TTA AAT GC - 3'
Palindrome parfait (24 pb)	5' – AAA GTA T**TA CA**T A**TG TA**A TAC TTT - 3'
Demi-palindrome (12 pb)	5' – A**TG TA**A TAC TTT - 3'

Figure 30 : Séquences d'ADN utilisées pour l'étude de l'interaction avec BlaI. La première séquence est l'opérateur Op1 de *B. licheniformis,* pour lequel BlaI a la plus haute affinité (Filée *et al.*, 2003). La deuxième représente la séquence consensus symétrisée des opérateurs *bla* et *mec* de *B. licheniformis* et de *S. aureus* (Figure 15). Cette séquence est composée de deux sites palindromiques parfaits. Les séquences surlignées en gras (le motif TACA et son palindrome TGTA) sont strictement conservées entre tous les opérateurs chez *B. licheniformis* et *S. aureus*. La troisième séquence représente le demi-palindrome de la séquence consensus.

- **Délimitation des domaines d'interaction**

Différentes interactions ont été étudiées par cartographie des variations de déplacement chimique pour délimiter les séquences d'ADN et protéiques les plus courtes permettant de caractériser totalement l'interaction.

Interaction de BlaI avec le 24 mer et le 30 mer

La protéine sauvage BlaI a été titrée par des concentrations croissantes d'ADN (24 mer et 30 mer) pour obtenir un rapport final (BlaI dimère / ADN double brin) égal à 1. Les résultats sont similaires pour les deux ADN utilisés, et les spectres ^{1}H-^{15}N HSQC obtenus pour le complexe avec le 24 mer sont donnés Figure 31. L'existence et la spécificité de l'interaction ont été également vérifiées par gel retard (Vreuls, 2004). Les résultats montrent que BlaI interagit de la même façon avec la séquence consensus symétrisée et l'opérateur de plus haute affinité Op1. Cette séquence d'ADN plus courte a été retenue pour les études postérieures.

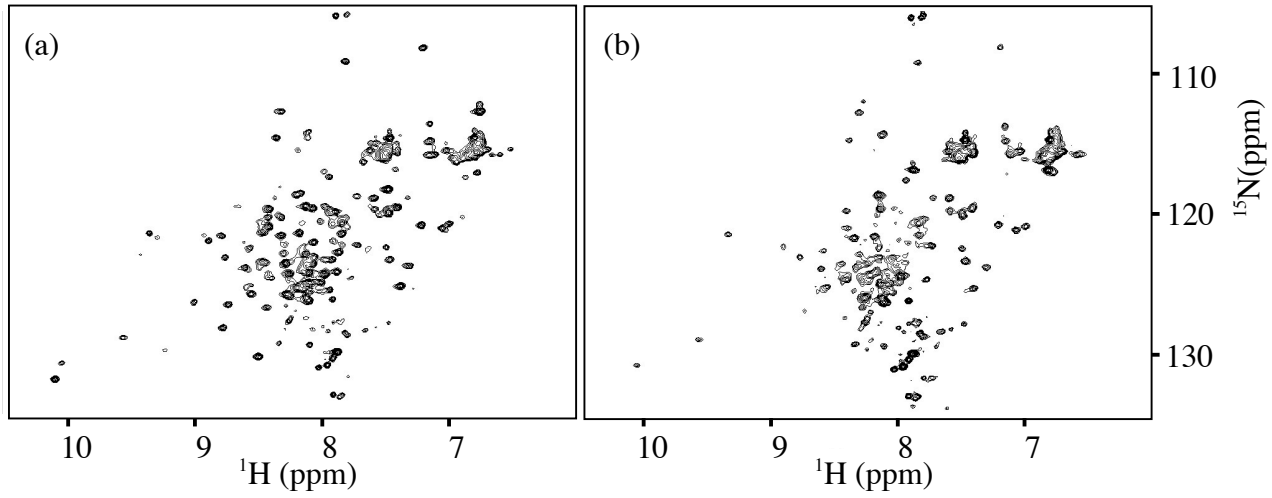


Figure 31: Spectres ^{1}H-^{15}N HSQC de BlaI, 1mM, tampon pH 7.6, 25° C. (a) En absence d'ADN (b) L'ADN palindromique de 24 pb a été ajouté jusqu'à atteindre un rapport [$BlaI_{dimère}$]/[$ADN_{double\ brin}$] de 1/1. L'ajout d'ADN supplémentaire ne modifie plus le spectre de BlaI.

La comparaison des spectres ^{1}H-^{15}N HSQC de BlaI en absence et en présence d'ADN permet de tirer plusieurs conclusions. Tout d'abord, l'allure des spectres de la protéine n'est pas dramatiquement modifiée en présence d'ADN. On peut donc en conclure que la structure de BlaI est peu modifiée par son interaction avec l'ADN. En particulier, nous n'observons pas une restructuration du domaine N-terminal, ni un changement au niveau du comportement dynamique du domaine de dimérisation lors de la fixation de l'ADN. En étudiant le devenir de chaque pic de corrélation du spectre, on s'aperçoit qu'une partie des pics reste inchangée, que d'autres subissent des variations de déplacement chimiques, et que les derniers disparaissent. On constate également que tous les pics qui subissent des changements apparents appartiennent au domaine N-terminal. Cette observation est conforme aux résultats désignant BlaI-NTD comme le domaine de liaison à l'ADN de BlaI (Wittman *et al.*, 1993).

Interaction de BlaI-NTD avec le 24 mer et le 12 mer

Le domaine BlaI-NTD a été titré par des concentrations croissantes en ADN pour obtenir un rapport final (dimère / ADN double brin) égal à 10. L'affinité du monomère pour l'ADN est en effet moins grande que celle du dimère (Wittman *et al.*, 1993), et il faut se mettre en excès d'ADN pour obtenir un équilibre où le complexe est majoritaire. La spécificité de l'interaction est néanmoins conservée, ce qui a été vérifié par gel retard (Vreuls, 2004). La titration de BlaI-NTD avec la séquence palindromique (24 mer) et le demi-palindrome (12 mer) donnent des résultats similaires à ceux observés pour le domaine N-terminal de la protéine entière. Cette observation prouve que BlaI-NTD interagit avec le demi-palindrome. Cette interaction a été confirmée par des expériences de spectrométie de masse nanospray (Vreuls *et al.*, 2004). Les spectres obtenus pour le complexe BlaI-NTD / 12 mer sont donnés Figure 32.

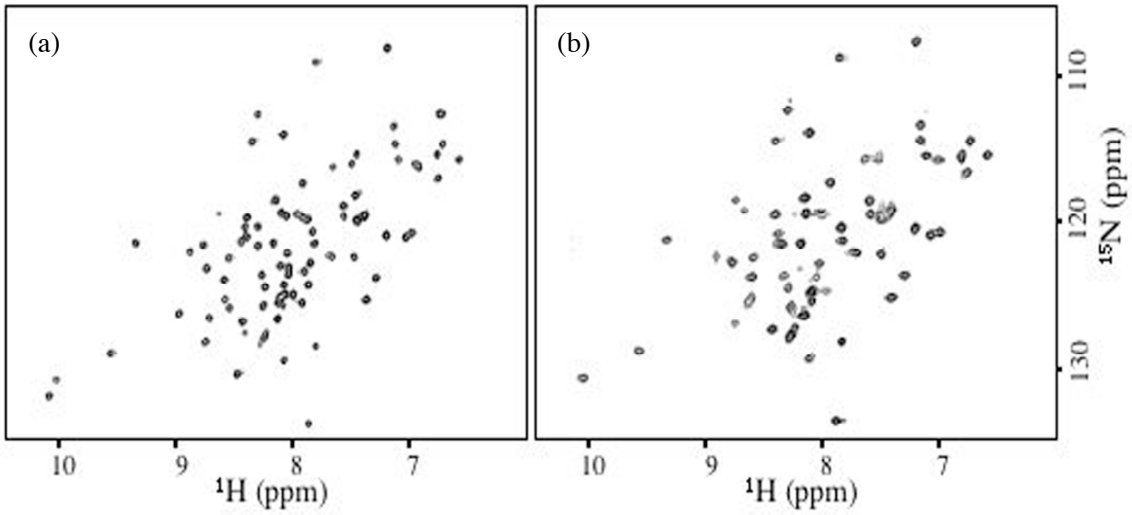


Figure 32: Spectres ^{1}H-^{15}N HSQC de BlaI-NTD, 0.1 mM, tampon pH 7.6, 25° C. (a) En absence d'ADN. (b) L'ADN palindromique de 12 pb a été ajouté jusqu'à atteindre un rapport [BlaI-NTD]/[ADN$_{\text{double brin}}$] de 1/10.

Au vu des résultats, nous avons montré que l'étude du système réduit BlaI-NTD / demi-palindrome (12 pb) est un bon modèle pour décrire l'interaction de BlaI avec son opérateur (Figure 33). C'est ce système réduit qui sera donc caractérisé par la suite.

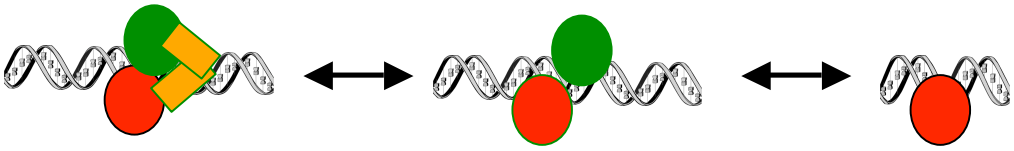


Figure 33: L'étude de l'interaction de BlaI avec ses séquences opératrices peut être réduite à celle de BlaI-NTD avec le demi-palindrome de 12 paires de base. Les domaines N-terminaux de BlaI sont schématisés par des boules, et les domaines C-terminaux par des rectangles.

III. 2. Caractérisation structurale de l'interaction

- **Délimitation de la surface d'interaction**

Dans le but de déterminer le mode de reconnaissance à l'ADN de la protéine BlaI de *B. licheniformis*, nous avons étudié la surface d'interaction entre la protéine et l'ADN par cartographie des déplacements chimiques des spectres HSQC de BlaI-NTD en absence et en présence de l'ADN palindromique de 12 pb (Figure 32 et Figure 34). Le suivi des pics de corrélation du spectre ^{1}H-^{15}N HSQC permet d'avoir des informations sur la chaîne principale de la protéine, alors que l'étude des spectres ^{1}H-^{13}C HSQC des méthyles permet de localiser les interactions des chaînes latérales avec l'ADN. Ces deux spectres sont bien résolus et apportent des informations complémentaires.

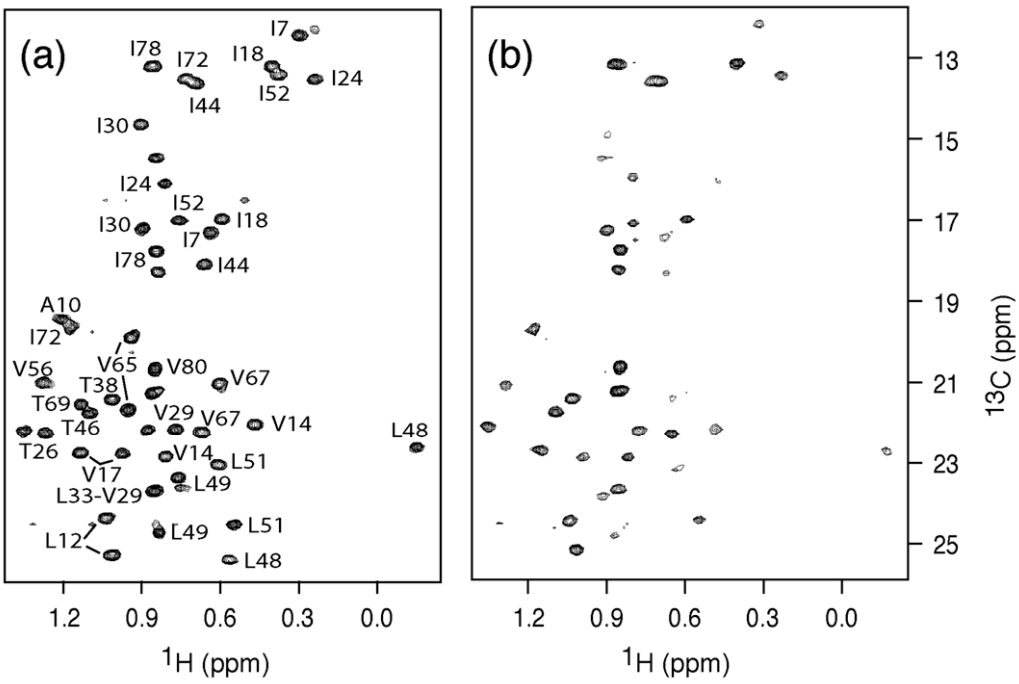


Figure 34: Spectres ^{1}H-^{13}C CT-HSQC de la région des méthyles de BlaI-NTD, 0.1 mM, tampon pH 7.6, 25°, enregistrés à 600 MHz. (a) En absence d'ADN. (b) L'ADN palindromique de 12 pb a été ajouté jusqu'à atteindre un rapport [BlaI-NTD]/[$ADN_{double\ brin}$] de 1/10.

Les spectres de BlaI-NTD, seul et en complexe avec l'ADN, présentent de nombreuses similitudes. Les déplacements chimiques des ^{1}H, ^{15}N et ^{13}C de BlaI-NTD dans le complexe protéine-ADN ont été attribués par référence aux déplacements chimiques de BlaI-NTD. Les variations de déplacement chimique observées dans les spectres ^{1}H-^{15}N HSQC et ^{1}H-^{13}C HSQC des méthyles sont données Figure 35.

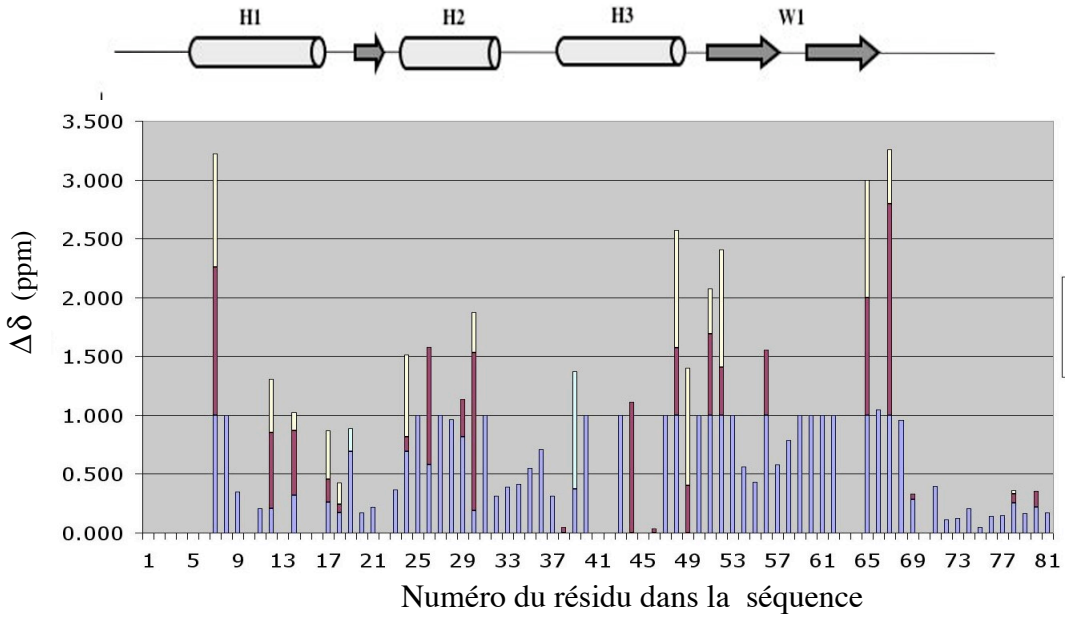


Figure 35: Graphique des variations de déplacement chimique de BlaI-NTD en interaction avec le demi-palindrome. Les barres colorées en bleu représentent les variations de déplacement chimique des pics de corrélation ^{1}H-^{15}N, les barres rouges et jaunes ceux des groupements méthyles, et les barres en turquoise ceux des groupements amides aromatiques des tryptophanes. L'axe vertical représente $|\delta^{1}H| \times |\gamma_H/\gamma_N| + |\delta\ ^{15}N|$ ou $|\delta^{1}H| \times |\gamma_H/\gamma_C| + |\delta\ ^{13}C|$. Pour la visualisation, une valeur de 1 a été attribuée aux résidus dont les pics disparaissent. L'absence de barres indique la présence d'un résidu proline ou d'un déplacement chimique non mesurable à cause des recouvrements. Les structures secondaires de BlaI-NTD sont rappelées au-dessus du graphique.

Des modifications significatives de déplacement chimique et/ou d'intensité sont observées pour les résidus de l'extrémité N-terminale de la protéine (Ile7, Ser8), de l'hélice H_3 (Ser40, Thr43, Met47, Leu48, Arg50, Leu51, Ile52) et des feuillets β proches de la boucle W_1 (Asn25, His59, His60, Lys61, Glu62, Val65, Val67). On peut donc conclure que la protéine BlaI de *B. licheniformis* interagit avec ses régions opératrices par l'intermédiaire de son hélice de reconnaissance H_3, de sa boucle W_1 et de son extrémité N-terminale.

Les résidus de BlaI-NTD affectés par la liaison à l'ADN forment une surface d'interaction qui a été colorée sur la Figure 36. Cette surface est localisée sur une des faces de la protéine et se superpose à la surface électrostatique positive de la protéine. Ces résultats démontrent l'importance de l'extrémité N-terminale de BlaI dans le processus d'interaction avec l'ADN, et sont en parfait accord avec les expériences de mutagenèse dirigée montrant que les mutants de BlaI dont les acides aminés Lys3 à Ser8 ont été supprimés (Wittman *et al.*, 1993), ou dont l'acide aminé Lys4 a été muté en Ala (Gregory *et al.*, 1997), ont perdu leur capacité à lier l'ADN.

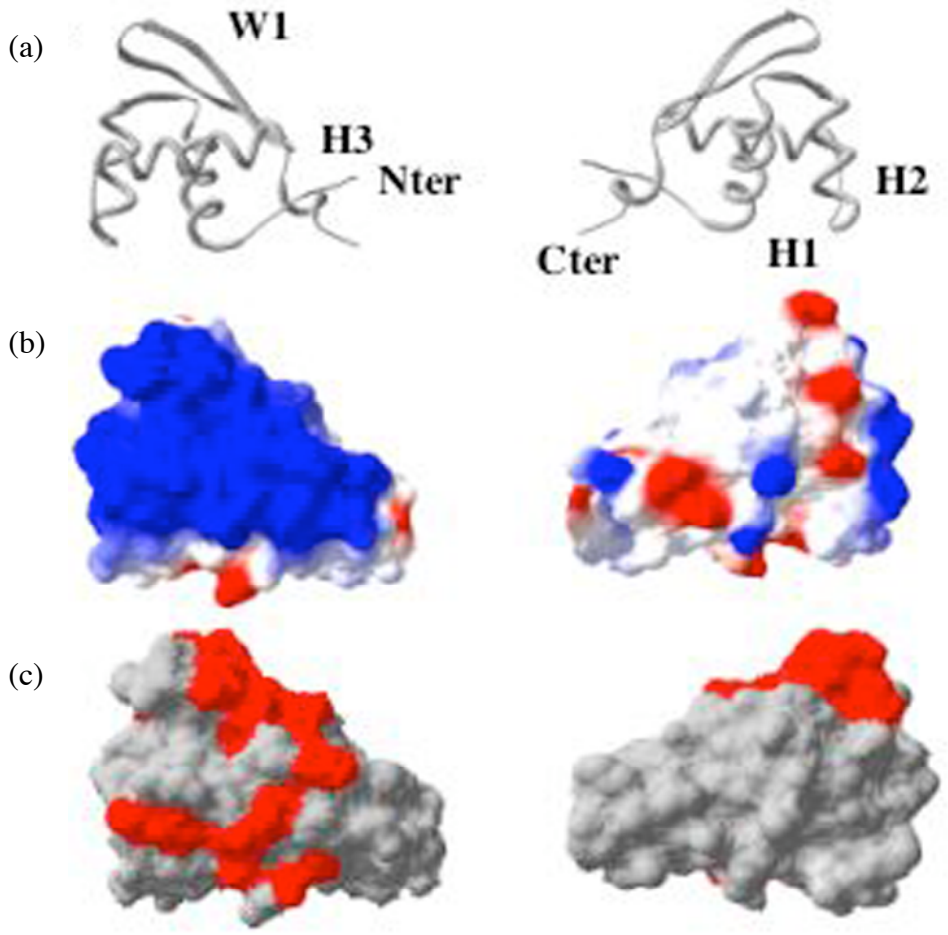


Figure 36 : Structure du fragment BlaI-NTD. (a) Représentation en ruban de la face avant et arrière de BlaI-NTD. (b) Représentation de la surface électrostatique de BlaI-NTD chargée positivement (en rouge) et négativement (en bleu). (c) Les variations de déplacement chimique $|\delta^1H| \times |\gamma_H/\gamma_N| + |\delta\ ^{15}N|$ supérieures à 0.7 ppm pour les résidus de BlaI-NTD en interaction avec l'ADN sont indiquées en rouge sur la surface de BlaI-NTD.

- **Reconnaissance classique des protéines à motif 'winged helix'**

La structure du complexe protéine-ADN de certaines protéines à motif 'winged helix' a été résolue par cristallographie des rayons X ou par résonance magnétique nucléaire (Gajiwala et Burley, 2000). Les deux modes de reconnaissance à l'ADN rencontrés chez les protéines à motif 'winged-helix' sont décrits dans la Figure 37.

Le premier a été découvert pour la première fois pour le complexe HNF/3 avec l'ADN (Lai *et al*., 1993), mais il est rencontré chez la majorité des protéines à motif 'winged-helix', comme par exemple chez les régulateurs transcriptionnels LexA et MarR de *E. coli* (Fogh, et al, 1994 ; Alekshun *et al*., 2001).

Le deuxième est complètement différent, et il n'a été identifié jusqu'à présent que pour la protéine RFX1 (Iwana *et al*., 1999 ; Cornille *et al*., 1998).

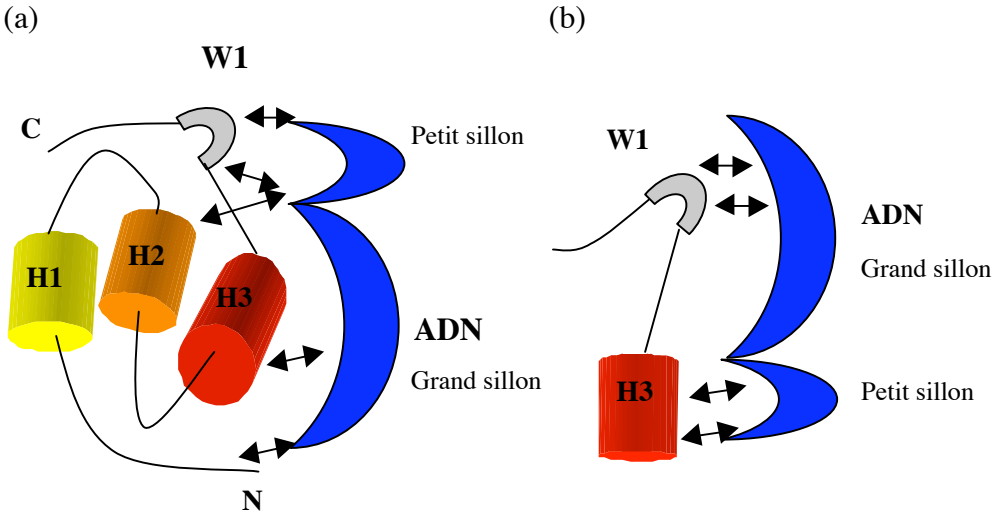


Figure 37: Schéma illustrant le mode de reconnaissance à l'ADN des protéines à motif 'winged-helix'. (a) Mode de reconnaissance classique observé chez la majorité des protéines à motif 'winged-helix'. L'hélice de reconnaissance H_3 interagit de manière spécifique avec les bases nucléotidiques du grand sillon de l'ADN. La boucle W_1 et l'extrémité N-terminale établissent des contacts avec le petit sillon de l'ADN. (b) mode de reconnaissance à l'ADN atypique rencontré chez la protéine RFX1. La boucle W_1 interagit avec le grand sillon de l'ADN, tandis que l'hélice H_3 interagit avec le petit sillon. Il n'y a aucun contact entre l'extrémité N-terminale de la protéine et l'ADN.

Le rôle de l'extrémité N-terminale de la protéine dans la reconnaissance de l'ADN et la topologie de la surface d'interaction protéine/ADN montrent que BlaI présente toutes les caractéristiques du mode de reconnaissance à l'ADN classique des protéines à motif 'winged helix'.

Ce mode de reconnaissance a été décrit en détail lors de la détermination de la structure à haute résolution du complexe E2F4-ADN (Zheng *et al.*, 1999). De plus, les distances séparant l'hélice H_3, la boucle W_1 et l'extrémité N-terminale de BlaI sont semblables à celles de E2F4. Etant données les nombreuses similitudes entre la protéine BlaI et la protéine E2F4, nous proposons dans la Figure 38 un modèle pour la structure du complexe BlaI-ADN par comparaison avec la structure du complexe E2F4-ADN. Dans ce modèle, l'hélice H_3 interagit avec le motif de reconnaissance conservé TACA de l'opérateur et les autres contacts protéine-ADN ont lieu par l'intermédiaire de l'extrémité N-terminale et de la boucle W_1. Les résidus de l'hélice H_3 sont chargés positivement et sont capables de former des liaisons hydrogènes avec les bases nucléotidiques. Cependant, la position précise des contacts entre les acides aminés de BlaI et les bases nucléotidiques de l'ADN nécessite la détermination de la structure 3D à haute résolution du complexe BlaI-ADN.

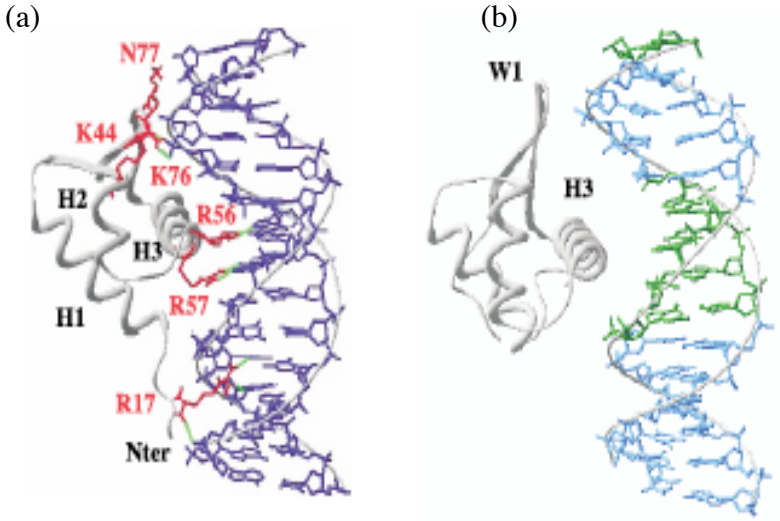


Figure 38: Comparaison des régions impliquées dans l'interaction protéine-ADN pour les protéines E2F4 et BlaI-NTD (a) Structure du complexe E2F4-ADN (Zheng *et al.*, 1999). Les chaînes latérales impliquées dans l'interaction avec l'ADN sont indiquées en rouge, et les liaisons hydrogène sont indiquées en vert. L'hélice H_3 (Arg56, Arg57) et l'extrémité N-terminale (Arg17) forment directement des liaisons hydrogène avec les bases de l'ADN. La boucle W_1 ainsi que le coude T_1 reliant H_2 et H_3 (Lys44, Asn77) forment des liaisons hydrogène avec les groupements phosphates et les sucres de l'ADN. On peut noter que la numérotation des résidus de E2F4 est différente de celle de BlaI. (b) Modèle pour la structure du complexe BlaI-NTD/ADN dérivé de la structure du complexe E2F4/ADN. La structure de BlaI-NTD a été superposée à la celle de E2F4. Les bases de l'ADN qui sont conservées sont représentées en vert. Dans ce modèle, le motif TACA interagit avec l'hélice H_3.

III. 3. Détermination de la constante de dissociation BlaI-NTD / demi-opérateur

La titration de BlaI-NTD par le demi-palindrome a été suivie en repérant la position des pics des spectres ^{1}H-^{15}N et ^{1}H-^{13}C HSQC de la protéine (Figure 32 et Figure 34) à chaque ajout d'ADN. Ceci est possible car l'équilibre du complexe est rapide pour la majorité des pics dans les conditions de l'étude. Une moyenne a été faite sur l'ensemble des variations de déplacement chimique des pics bien résolus de ces spectres. L'évolution de cette variation en fonction de la concentration d'ADN est donnée Figure 39.

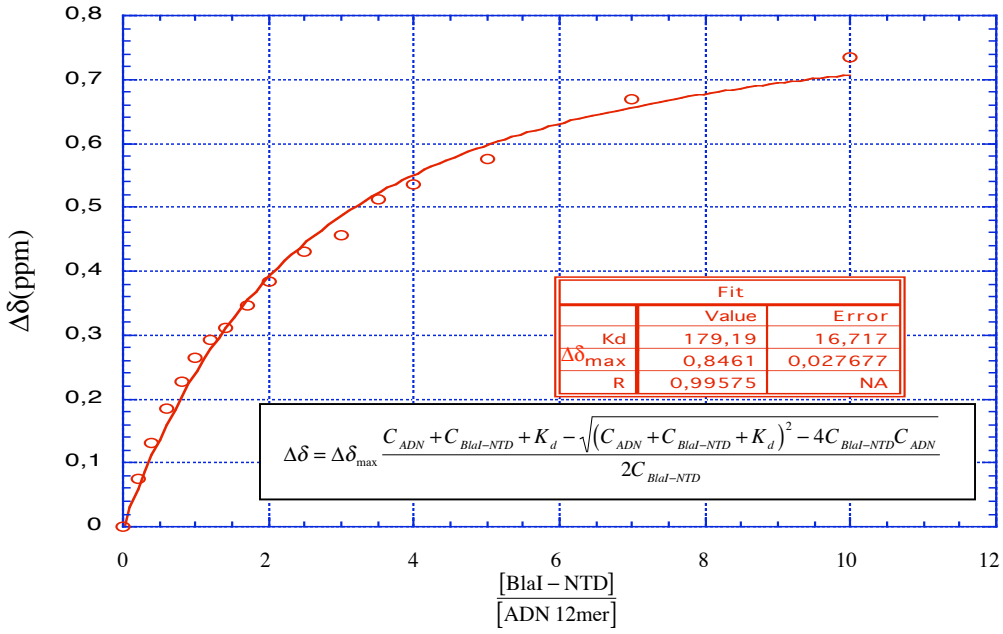


Figure 39 : Titration de BlaI-NTD (0.1 mM) par des concentrations croissantes d'ADN 12mer jusqu'à un rapport des concentrations en ADN et en protéine égal à 10. L'ordonnée est une moyenne des variations de déplacements chimiques protons, azotes et carbones des pics bien résolus des spectres ^{1}H-^{15}N HSQC et ^{1}H-^{13}C HSQC des méthyles, pondérés par leurs rapports gyromagnétiques. L'encadré donne l'équation et les paramètres obtenus pour la modélisation de la courbe (dessinée en rouge) par le logiciel Kaleidagraph (Synergy software).

La courbe de la Figure 40 est caractéristique d'un équilibre chimique de la forme :

BlaI-NTD + ADN <-> BlaI-NTD/ADN

$$K_d = \frac{[BlaI - NTD][ADN]}{[BlaI - NTD/ADN]}$$

La constante de dissociation Kd décrivant cet équilibre peut être obtenue en effectuant une optimisation des paramètres de la formule décrivant l'équilibre chimique, donnée dans la Figure 40. La modélisation obtenue pour cette courbe par le logiciel Kaleidagraph donne le résultat suivant:

Kd = 180 +/- 80 µM

L'erreur a été estimée en prenant en compte l'erreur statistique donné par la modélisation, et également les erreurs d'estimation des concentrations en protéine et en ADN. On peut également noter qu'une modélisation faite séparément pour les variations de déplacement chimique NH et CH_3 donne des résultats équivalents.

Cette étude montre que l'affinité du domaine d'interaction isolé et monomérique est faible par rapport à celle du dimère. En effet, la constante de dissociation entre le dimère et l'ADN a été estimée par gel retard de 4.10^{-11} à 3.10^{-10} M selon les opérateurs (Filée *et al.*, 2003). Cette étude confirme donc que la dimérisation est essentielle pour conférer à la protéine BlaI une affinité élevée pour l'ADN en plaçant les deux domaines de liaison à l'ADN dans une conformation favorable.

Discussion et perspectives: Répression et induction des gènes de résistance chez *B. licheniformis* 749/I et *S. aureus*

Cette partie fait le bilan de nos résultats concernant le répresseur BlaI de *B. licheniformis,* et les compare aux résultats récemment publiés pour les répresseurs BlaI et MecI de *S. aureus*. Tous les résultats concernant MecI de *S. aureus* sont tirés des articles du groupe de X. Gomis-Rüth (Garcia-Castellanos *et al.*, 2003 ; Garcia-Castellanos *et al.*, 2004 ; Mallorqui–Fernandez *et al.*, 2004). Les données qui concernent BlaI de *S. aureus* sont extraites de l'article du groupe de G. Archer (Safo *et al.*, 2004). Cette comparaison permettra d'apporter des conclusions structurales sur les modes de répression et d'induction des β-lactamases et des PBP2', et de proposer des perspectives à ce travail.

I. Structure des répresseurs BlaI et MecI

I. 1. Domaine de liaison à l'ADN

Notre étude a montré que la protéine BlaI de *B. licheniformis* est un répresseur dimérique dont les deux domaines structuraux sont indépendants. La structure du domaine N-terminal de liaison à l'ADN a été résolue par RMN, montrant que BlaI fait partie des protéines à motif 'winged-helix'.

Le domaine de liaison à l'ADN du répresseur BlaI de *B. licheniformis* présente une haute homologie de séquence avec celui des protéines BlaI et MecI de *S. aureus*. De nombreux résidus sont conservés, et plus particulièrement les résidus situés au niveau du coeur hydrophobe de la protéine: (H_1) Ala10, Val14, Met15, Ile18, Trp19 ; (H_2) Val/Ile29, Ile/Val30, Glu32, Leu/Ileu33 ; (H_3) Ile44, Leu/Ile48, Leu51 ; (feuillet β) Leu/Ileu57, Tyr68, Ile72. Nous avons proposé que les répresseurs de *S. aureus* présentent la même structure en 'winged-helix' que BlaI de *B. licheniformis* (Van Melckebeke *et al.*, 2003).

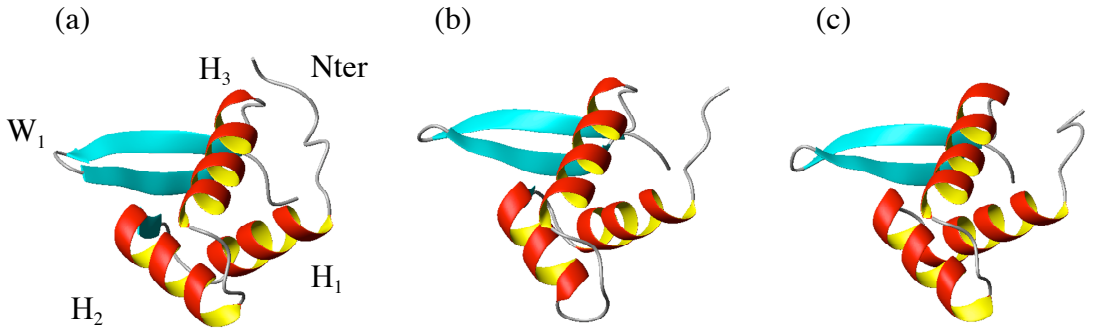


Figure 40: Comparaison des domaines de liaison à l'ADN de (a) BlaI de *B. licheniformis* (PDB 1P6R), (b) MecI de *S. aureus* (PDB 1OKR), (c) BlaI de *S. aureus* (PDB 1SD7). Le rmsd calculé sur la chaîne principale pour les résidues 1 à 78 est 0.52 Å entre 1OKR et 1SD7, 1.22 Å entre 1OKR et 1P6R, et 1.27 Å entre 1SD7 et 1P6R.

Cette proposition a été confirmée par les structures cristallographiques de BlaI et MecI de *S. aureus* parues récemment, qui montrent que les répresseurs appartiennent effectivement à la famille des protéines à motif 'winged helix' (Garcia-Castellanos *et al.*, 2003 ; Safo *et al.*, 2004). La structure 3D de leurs domaines de liaison à l'ADN est en effet très proche de celui de BlaI de *B. licheniformis* (Figure 40). La seule différence est la longueur du brin β1, qui est légèrement plus long pour BlaI de *B. licheniformis* (3 résidus) que pour les répresseurs de *S. aureus* (1 résidu).

I. 2. Domaine de dimérisation

Nous avons également montré que le domaine C-terminal de BlaI de *B. licheniformis* impliqué dans la dimérisation est structuré en hélices α, et qu'il présente une grande flexibilité et des cavités hydrophobes au niveau de sa surface de dimérisation.

Les structures cristallographiques de BlaI et MecI de *S. aureus* montrent que l'architecture du domaine C-terminal responsable de la dimérisation est composée de 3 hélices α consécutives H_4-H_5-H_6 maintenues par un cœur hydrophobe s'étendant de H_4 à H_6 et comprenant deux cavités hydrophobes de 22 et 117 $Å^3$ (Figure 41). Les données cristallographiques montrent également que ce domaine est assez flexible (Garcia-Castellanos *et al.*, 2003 ; Safo *et al.*, 2004). On peut noter que la structure tertiaire de ce domaine en 'spiral staircase' n'avait été observée pour aucune autre protéine auparavant.

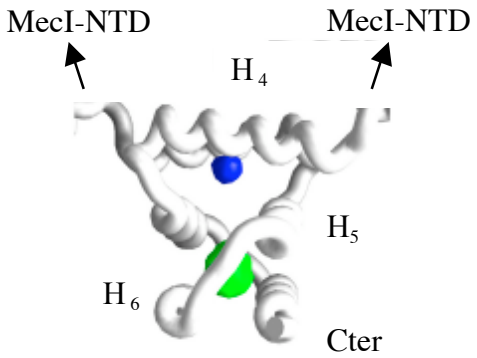


Figure 41: Zoom sur les domaines de dimérisation de l'homodimère de MecI de *S. aureus* (extrait de Garcia-Castellanos *et al.*, 2003). Les deux cavités hydrophobes sont données en bleu et en vert. Les deux domaines de dimérisation en interaction ont des repliement similaires, mais ne sont pas tout à fait superposables.

Ces structures sont en accord avec nos résultats. Ceci n'est pas étonnant car les trois domaines de dimérisation sont très homologues. Les leucines participant à la surface d'interaction sont en particulier totalement conservées chez *B. licheniformis* et *S aureus* (Figure 17). On peut aussi noter que le site de clivage (Asn101 et Phe102) proposé par Zhang (Zhang *et al.*, 2001) est conservé pour les trois répresseurs, et qu'il appartient à l'hélice H_5. La position de ce site dans la structure du répresseur dimérique est telle qu'il n'est pas très accessible au solvant. Il est néanmoins situé dans une région présentant une dynamique particulière (Mallorqui–Fernandez *et al.*, 2004 ; Safo *et al.*, 2004).

I. 3. Structure quaternaire : la pince à ADN

Il a été montré que l'interaction de BlaI avec l'ADN se produit lorsque le répresseur est sous forme dimérique, et que la formation du dimère est antérieure à la liaison avec l'ADN (Filée *et al.*, 2003). Les résultats fournis par les études cristallographiques donnent la topologie des dimères de BlaI et MecI de *S. aureus*. On peut noter que le dimère est structuré en forme de pince même en absence d'ADN, comme le montre la Figure 42.

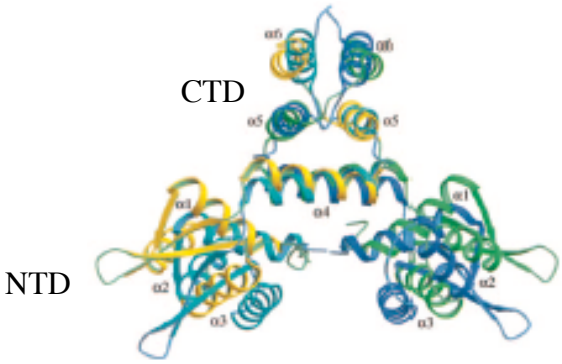


Figure 42 : Superposition de la structure quaternaire des dimères de BlaI et MecI de *S. aureus* en absence d'ADN (extrait de Safo *et al.*, 2004). MecI est représenté en bleu et turquoise, et BlaI en jaune et vert.

La comparaison de la structure quaternaire des dimères de BlaI et MecI de *S. aureus* en absence d'ADN montre que la pince qui est formée peut avoir une ouverture différente. Pour BlaI, le triangle est en effet plus ouvert que pour MecI (Figure 42). De plus, des contacts entre les domaines N- et C-terminaux de MecI ont été observés. Ce résultat est différent de celui que nous avons eu pour BlaI de *B. licheniformis*, où les deux domaines fonctionnels sont structuralement indépendants. Ces deux résultats sont à analyser avec prudence, car ils peuvent être dus au fait que le cristal impose des contraintes supplémentaires.

On peut noter cependant une certaine propension du domaine C-terminal à adopter des conformations multiples, ce qui pourrait expliquer l'allure du spectre RMN de BlaI de *B. licheniformis*. Cette flexibilité conformationnelle a déjà été observée pour d'autres régulateurs transcriptionnels bactériens comme MexR (Lim *et al.*, 2002) et MtaN (Godsey *et al.*, 2001), qui ont des similitudes structurales avec BlaI et MecI. Cette propriété a vraisemblablement un rôle biologique concernant la levée de la répression des gènes de résistance.

II. Interaction des répresseurs avec l'ADN

II. 1. Interaction du domaine de liaison avec l'ADN

Les expériences de cartographie des variations de déplacement chimique effectuées dans ce travail mettent en évidence que BlaI interagit avec l'ADN selon le mode de reconnaissance classique rencontré chez la majorité des protéines à motif 'winged helix'. Le domaine N-terminal interagit avec l'ADN via son hélice de reconnaissance H_3, sa boucle W_1 et son extrémité N-terminale. Le domaine C-terminal de la protéine n'est pas impliqué dans l'interaction avec l'ADN. La mesure de la constante d'affinité du monomère a néanmoins montré que la dimérisation est essentielle pour que BlaI assure sa fonction de répresseur avec une affinité suffisante.

L'alignement de séquence des répresseurs BlaI et MecI de *B. licheniformis* et de *S. aureus* montre que les résidus de l'hélice H_3 (Lys42, Thr46, Arg50, Lys53) et les résidus proches de la boucle W_1 (His59, His60, Lys61) impliqués dans l'interaction avec l'ADN sont totalement conservés. L'alignement de séquence entre les opérateurs *bla* et *mec* de *B. licheniformis* et de *S. aureus* a mis en évidence un motif palindromique TACA également totalement conservé (Figure 15). Nous avons donc proposé un mécanisme de répression dans lequel l'hélice H_3 du domaine N-terminal de chaque monomère établit des contacts spécifiques avec le motif TACA de l'opérateur, ce qui place les deux domaines C-terminaux dans une conformation favorable pour la dimérisation.

Cette hypothèse a été confirmée lors de la détermination de la structure 3D des répresseurs MecI et BlaI de *S. aureus* en complexe avec l'ADN par cristallographie (Garcia-Castellanos *et al.*, 2004 ; Safo *et al.*, 2004). La Figure 43 donne par exemple le mode d'interaction de MecI avec l'ADN. L'hélice H_3 de MecI fait des interactions spécifiques avec le motif conservé TACA. La boucle W_1 et la partie N-terminale ont également des contacts avec l'ADN.

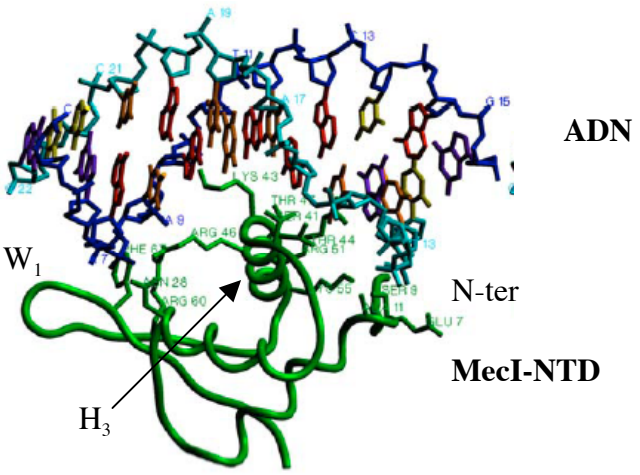


Figure 43: Structure de MecI en complexe avec son opérateur (extrait de Garcia-Castellanos *et al.*, 2004). Un agrandissement a été fait sur le domaine de liaison à l'ADN, et le domaine de dimérisation n'est pas représenté.

II. 2. Modification de la pince

Deux informations différentes ont été apportées par les structures cristallographiques des dimères MecI et BlaI de *S. aureus* en présence d'ADN. Les structures du dimère de MecI en absence et en présence d'ADN sont superposables, alors qu'on observe un changement de l'ouverture de la pince pour BlaI (Figure 44). Par RMN, nous n'avons pas observé de modifications des spectres de BlaI-CTD de *B. licheniformis* lors de la fixation du dimère sur l'ADN. On peut néanmoins noter que, lorsqu'ils sont liés à un même ADN, les structures des répresseurs BlaI et MecI sont superposables (Figure 44).

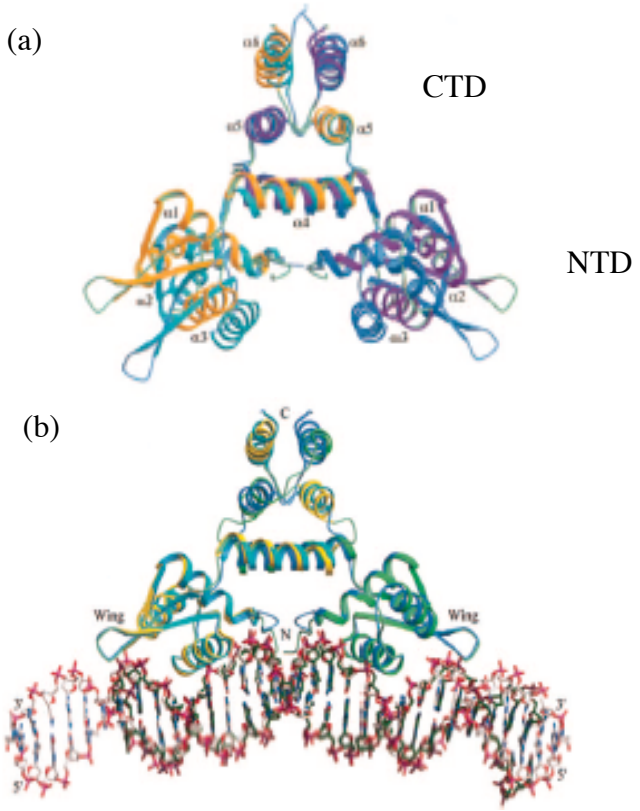


Figure 44: (a) Superposition des formes libres (orange et rose) et liées (bleu et turquoise) du dimère de BlaI de *S. aureus* (extrait de Safo *et al.*, 2004). (b) Superposition des dimères de BlaI et MecI de *S. aureus* en interaction avec l'ADN (extrait de Safo *et al.*, 2004).

II. 3. Organisation des régions intergéniques

Au vu des structures cristallographiques et des données de protection aux DNases (Clarke *et al.*, 2001; Sharma *et al.*, 1998), il a été proposé un modèle d'organisation du positionnement des répresseurs sur les différents opérateurs de la région intergénique des systèmes *bla* et *mec* de *S. aureus* (Figure 46).

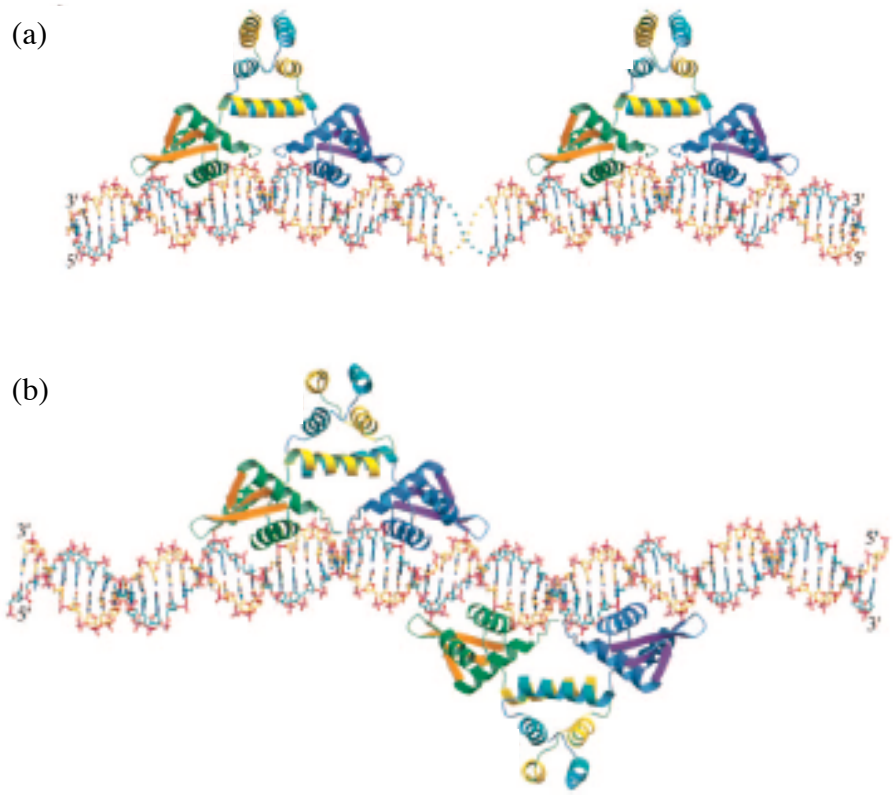


Figure 45 : Organisation des régions intergéniques des divergeons *mec* et *bla* chez *S. aureus* lorsque les répresseurs y sont fixés (extrait de Safo *et al.*, 2004). (a) Modèle où deux répresseurs dimériques sont fixés sur les deux opérateurs de *bla*. (b) Modèle où deux répresseurs dimériques sont fixés sur les deux opérateurs de *mec*.

Cette organisation pourrait expliquer pourquoi la fixation des répresseurs est coopérative dans le cas de l'opérateur *mec* (Sharma *et al.*, 1998), où les deux dimères sont proches et de part et d'autre de l'ADN, alors qu'elle se fait de façon indépendante pour *bla*, où les régions de fixation des deux dimères sont distantes de 13 paires de bases (Gregory *et al.*, 1997). On peut noter également qu'une certaine courbure de l'ADN est nécessaire pour la fixation des dimères dans le système *mec*, ce qui est observé dans les structures cristallographiques (Garcia-Castellanos *et al.*, 2004).

III. Bilan et conclusion sur l'induction

III. 1. Bilan structural

La comparaison des trois répresseurs montre qu'ils sont très fortement homologues au point de vue structural. En particulier, ils possèdent tous les trois un domaine de liaison à l'ADN à motif 'winged-helix', et interagissent avec le motif TACA des régions opératrices selon le mode classique, où l'hélice H_3 fait des contacts spécifiques avec le grand sillon de l'ADN, et où des contacts supplémentaires s'établissent entre la boucle W_1, la partie N-terminale, et l'ADN. La partie C-terminale est structurée en trois hélices α, et elle assure la dimérisation des répresseurs par des interactions de type essentiellement hydrophobe entre les hélices. Ce domaine est flexible, capable d'avoir plusieurs conformations, et il présente des cavités hydrophobes. Le site de clivage entre les résidus N_{101} et F_{102} (Lewis *et al.*, 1999) est peu accessible au solvant, mais il se trouve dans une région où une grande dynamique est observée.

Cette homologie de structure importante entre les répresseurs BlaI de *B. licheniformis* et MecI de *S. aureus* et les similarités dans leur mode de reconnaissance à l'ADN permettent d'expliquer la corépression des opérateurs *bla* et *mec* observée pour BlaI et MecI chez *S. aureus* (Mc Kinney *et al.*, 2001). Elle explique en particulier pourquoi il est possible de former des hétérodimères BlaI-MecI, et comment BlaI et MecI peuvent se fixer aux mêmes séquences d'ADN (Sharma *et al.*, 1998; Safo *et al.*, 2004). Ce résultat est également observé pour BlaI de *B. licheniformis,* qui peut interagir avec les opérateurs *bla* et *mec* de *S. aureus* (Vreuls, résultats de gels retards non présentés).

Quelques différences ressortent également de cette étude entre les différents systèmes. Tout d'abord, il existe des contacts entre les domaines C- et N-terminaux de MecI, qui n'existent pas chez BlaI de *B. licheniformis.* Ensuite, il peut exister des différences dans l'ouverture de la pince du dimère en l'absence d'ADN. Enfin, le nombre et l'organisation quaternaire des répresseurs sur les régions opératrices intergéniques varient d'un système à l'autre.

III. 2. Conclusion et perspectives

L'objectif de la détermination de la structure 3D du répresseur BlaI de *B. licheniformis* et de son mode de reconnaissance à l'ADN était de mieux comprendre le mécanisme de régulation des gènes *blaP/blaZ/mecA* de *B. licheniformis* et de *S. aureus*.

Les résultats obtenus pour BlaI de *B. licheniformis* et leur comparaison aux données obtenues pour *S. aureus* ont permis une description très précise du mode de répression des gènes codant pour la β-lactamase et les PBP2'. En particulier, la structure et l'interaction du domaine N-terminal avec l'ADN ont été caractérisées en détail. Le rôle du domaine de dimérisation en ce qui concerne l'affinité de l'interaction a également été mis en évidence. Cette corrélation étroite entre la dimérisation, la fixation à l'ADN, et l'activité, est présente chez de nombreuses protéines régulatrices (Johnson *et al.*, 1980).

La baisse d'affinité menant à la levée de la répression peut avoir deux causes. Le clivage du répresseur au niveau du site conservé de la partie C-terminale pourrait empêcher la dimérisation (Zhang *et al.*, 2001). Un autre mécanisme ferait intervenir un inducteur qui vient se fixer sur la partie C-terminale, et la bloque dans une conformation où le terme entropique n'est plus favorable pour une liaison de forte affinité avec l'ADN (Filée *et al.*, 2002). Pour l'instant, nos données structurales ne permettent pas de trancher entre ces deux explications. Le site de clivage, bien que très conservé entre les souches et situé dans une région présentant une certaine flexibilité, est peu accessible au solvant. Dans l'autre schéma, aucun inducteur n'a encore été identifié. On ne peut également pas écarter l'hypothèse que l'induction se fasse différemment selon le type de souche bactérienne, ou selon le divergeon.

La répression des gènes *bla* et *mec* a été caractérisée dans ce travail et ceux de Garcia-Castellanos et Safo. Les questions qui restent en suspens concernent la transmission du signal intracellulaire par BlaR1 et la levée de la répression des gènes. Plusieurs perspectives sont proposées pour répondre à ces questions en s'aidant de la Résonance Magnétique Nucléaire.

Tout d'abord, la détermination de la structure 3D à haute résolution du complexe BlaI-NTD avec l'ADN est en cours de réalisation au laboratoire (Boudet *et al.*, en préparation). Pour cela, les informations tirées des variations des déplacements chimiques observés sur l'ADN et sur la protéine lors de l'interaction sont utilisées dans des méthodes d'arrimage (docking). Cette étude permettra de déterminer la position précise des contacts entre les acides aminés de BlaI de *B. licheniformis* et les bases nucléotidiques de ses régions opératrices, ce qui permettra de finaliser la comparaison entre les différents répresseurs à un niveau atomique.

Concernant la partie C-terminale des répresseurs, nous voulons faire une comparaison du comportement des répresseurs BlaI et MecI par RMN. Pour cela, des échantillons de MecI et MecI-NTD enrichis en ^{15}N seront produits et caractérisés. La comparaison des spectres RMN permettra de voir si les domaines N- et C-terminaux de MecI comportent effectivement des contacts structuraux en solution, comme c'est le cas dans les cristaux obtenus pour l'étude cristallographique (Garcia-Castellanos *et al.*, 2003). De plus, une étude des pics de corrélation des résidus du domaine de dimérisation permettra de conclure si le domaine de dimérisation de MecI présente la même dynamique ou le même échange conformationnel que celui de BlaI de *B. licheniformis*.

A l'avenir, il sera intéressant d'identifier la nature d'un éventuel inducteur de BlaI de *B. licheniformis,* et de déterminer la structure 3D en solution de la protéine en interaction avec cet inducteur. La nature de l'inducteur de certaines protéines à motif 'winged helix' a été identifiée. C'est le cas, par exemple, du répresseur SmtB Zn (II) de *Synechoccus*. L'affinité de SmtB pour l'ADN est du même ordre de grandeur que celle de BlaI de *B. licheniformis*. La liaison d'un ion Zn^{2+} à l'extrémité C-terminale de SmtB entraîne une diminution de l'affinité pour l'ADN d'un facteur 1000, ce qui induit la levée de la répression (Cook *et al.*, 1998 ; Vanzile *et al.*, 2002). Des inducteurs potentiels pour BlaI ont déjà été testés par RMN, sans résultats pour l'instant. Cette recherche va se poursuivre aux laboratoires de Liège et de Grenoble.

Augmentation de la résolution des spectres RMN multidimensionnels des protéines et des acides nucléiques par filtrage Hadamard

La spectroscopie Hadamard dans le contexte de l'acquisition rapide des spectres multidimensionnels en RMN

La Résonance Magnétique Nucléaire est une technique spectroscopique dont la réalisation expérimentale a été mise en évidence en 1946 (Purcell *et al.* , 1946 ; Bloch *et al* ., 1946). Depuis, c'est devenu une technique très répandue pour analyser la matière. La qualité des informations que l'on peut obtenir sur la structure et la dynamique des molécules étudiées dépend en grande partie de la qualité des spectres que l'on enregistre. Ce chapitre donne une comparaison des différentes techniques d'acquisition des données RMN. En particulier, la spectroscopie Hadamard sera présentée comme une technique alternative ou complémentaire à la spectroscopie par transformée de Fourier pour accélérer l'acquisition des spectres multidimensionnels. Les applications seront discutées dans le contexte de l'étude des macromolécules biologiques.

I. Spectroscopie RMN des macromolécules biologiques

I. 1. La RMN: une technique spectroscopique

Un spectre RMN représente l'intensité du rayonnement électromagnétique réémis après excitation d'un échantillon donné en fonction de la fréquence de ce rayonnement (Figure 46). Ce spectre présente des raies aux fréquences de résonance des spins excités.

Le nombre, la position, l'intensité et la largeur des raies sont les paramètres qui sont extraits des spectres pour obtenir des informations sur la structure et la dynamique des molécules présentes dans l'échantillon. La qualité de ces informations dépend de plusieurs facteurs. Tout d'abord, la manière d'exciter les spins de l'échantillon va conditionner la forme du signal obtenu. Depuis les débuts de la RMN, et jusqu'à aujourd'hui, les spectroscopistes proposent toujours de nouvelles séquences d'impulsions de manière à obtenir des informations supplémentaires sur le système de spins. Pour une technique de manipulation des spins donnée, la qualité d'un spectre peut alors être caractérisée par sa *sensibilité*, sa *résolution*, et la présence d'éventuels artéfacts. L'existence d'artéfacts dépend de la qualité de l'échantillon et du spectromètre, et de la séquence d'impulsions utilisée. Les concepts de sensibilité et de résolution sont cruciaux en RMN des macromolécules, et vont être discutés plus en détail.

I. 2. Sensibilité des spectres

La *sensibilité* d'une expérience RMN désigne le rapport signal sur bruit du spectre par unité de temps expérimental. Le signal correspond à l'intensité électrique provenant de la bobine de détection du spectromètre, et le bruit de ce signal est défini par son écart-type. La RMN est une technique intrinsèquement peu sensible, car elle dépend de la faible différence entre les populations des niveaux d'énergie Zeeman selon une distribution de Boltzman. Par exemple, pour les protons à la température ambiante et à un champ de 10 T, la différence entre les populations est uniquement de 1 spin sur 10^4, ce qui signifie que les 9 999 autres ne contribuent pas au signal RMN. La sensibilité a donc toujours été une préoccupation en spectroscopie RMN. Dans de nombreuses expériences, le rapport signal sur bruit est jugé suffisant quand il est possible de distinguer une raie de résonance du bruit dans un spectre (Figure 46). Pour d'autres expériences quantitatives, on exige que le rapport signal sur bruit atteigne une valeur seuil dépendant de la précision requise.

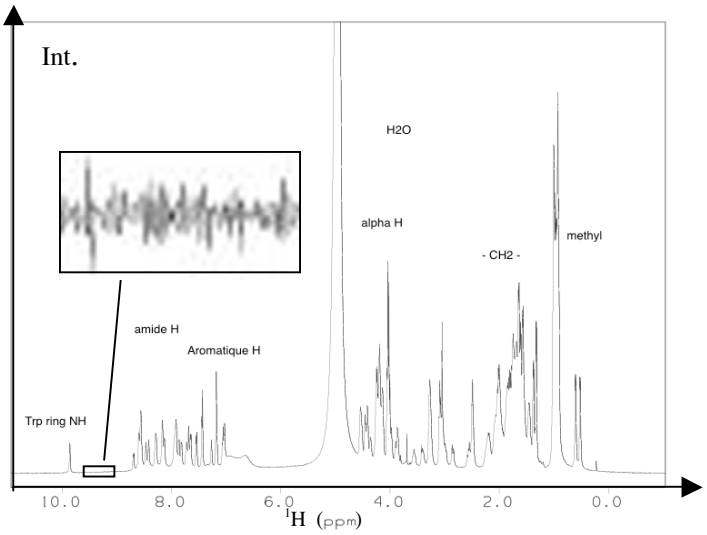


Figure 46: Exemple d'un spectre RMN ^{1}H d'une petite protéine (30 acides aminés). L'intensité du rayonnement réémis après excitation du système de spins est représenté en fonction de la fréquence ν de ce rayonnement. L'unité de cette fréquence peut être exprimée en Hertz, mais souvent on préfère la donner sous forme de déplacement chimique en ppm (partie par million), défini par : $\delta_{ppm} = 10^6 \frac{\nu - \nu_{ref}}{\nu_{ref}}$, où ν_{ref} est une fréquence de référence, et ν la fréquence de résonance du spin considéré. Cette unité a l'avantage d'être indépendante du champ magnétique auquel a été enregistré le spectre. La gamme de déplacements chimiques des différents groupements chimiques est indiquée au-dessus du spectre. L'encadré donne un agrandissement de la ligne de base du spectre, où on visualise le bruit thermique.

La sensibilité d'un spectre peut être optimisée en jouant sur différents paramètres. L'équation 1 explicite la dépendance du rapport signal sur bruit d'une expérience RMN unidimensionnelle avec le nombre de molécules situées dans le volume actif de l'échantillon (N_m), les rapports gyromagnétiques des spins utilisés pour l'excitation et la détection (respectivement γ_{exc} et γ_{det}), le champ magnétique (B_0), le nombre de répétitions de l'expérience (N_S), la largeur de raie de chacune des raies de résonance ($1/T_2$), et la température (T).

(Equation 1) $$\frac{Signal}{Bruit} \propto N_m \gamma_{exc} \gamma_{\det}^{3/2} B_0^{5/2} N_S^{1/2} \frac{T_2}{T}$$

Au cours des années, les spectroscopistes se sont battus pour optimiser chacun de ces paramètres. L'augmentation de la concentration des molécules étudiées est toujours favorable, à condition que la molécule reste dans un état natif. En pratique, on est souvent limité par l'agrégation qui se produit à des concentrations supérieures à quelques millimoles par litre pour les macromolécules biologiques. Le proton est souvent choisi comme noyau d'excitation et de détection à cause de son rapport gyromagnétique élevé. La force du champ magnétique statique a été augmentée tout au long des années dernières (Figure 47). L'augmentation de la sensibilité peut néanmoins être réduite par des propriétés de relaxation moins favorables des spins à champ élevé (T_2 moins élevé).

Par ailleurs, la sensibilité intrinsèque d'une séquence d'impulsions dépend du nombre et de l'efficacité des transferts d'aimantation et des propriétés de relaxation des spins considérés. Récemment, l'utilisation de l'effet TROSY (Pervushin *et al.* , 1997) a permis d'optimiser les propriétés de relaxation transversales des spins détectés. Les propriétés de relaxation des spins dépendent également de la température, de la taille et de la dynamique de la molécule considérée.

Ensuite, le niveau du bruit doit être minimisé (Figure 46). Le bruit thermique provenant des composants électroniques, qui est la cause majeure du bruit spectral, a récemment été abaissé grâce au développement de sondes cryogéniques (Figure 47). D'autres voies pour améliorer la sensibilité sont en cours d'exploration. Une technique très prometteuse consiste à augmenter la différence de populations des niveaux d'énergie Zeeman en transférant aux spins électroniques et nucléaires le moment angulaire de photons laser par pompage optique. Les méthodes d'hyperpolarisation des spins nucléaires (DNP) peuvent ainsi augmenter la sensibilité intrinsèque d'une expérience RMN de plusieurs ordres de grandeur.

Le dernier facteur déterminant le rapport signal sur bruit d'un spectre est le temps expérimental pendant lequel il a été enregistré. En effet, si on additionne N_S expériences de même nature, le signal RMN est multiplié par N_S. Le bruit de fond, quant à lui, est multiplié par $\sqrt{N_S}$, car c'est une variable aléatoire. Le rapport signal sur bruit d'un spectre résultant de la superposition de N_S signaux équivalents est donc augmentée par un facteur $\sqrt{N_S}$ par rapport à un spectre enregistré en un scan. L'augmentation du temps expérimental par un facteur N_S permet donc une augmentation du rapport signal sur bruit d'un facteur $\sqrt{N_S}$.

En cumulant l'ensemble des avancées technologiques et méthodologiques décrites plus haut, la sensibilité d'un grand nombre d'expériences RMN a été augmentée d'un facteur 3 à 10 ces dernières années, ce qui équivaut à une réduction du temps expérimental nécessaire pour obtenir le même rapport signal sur bruit d'un facteur 10 à 100 (Figure 47). Actuellement, pour un grand nombre d'expériences RMN et pour un échantillon concentré de macromolécules biologiques, un temps d'acquisition de quelques minutes à quelques heures est souvent suffisant pour obtenir un rapport signal sur bruit adéquat.

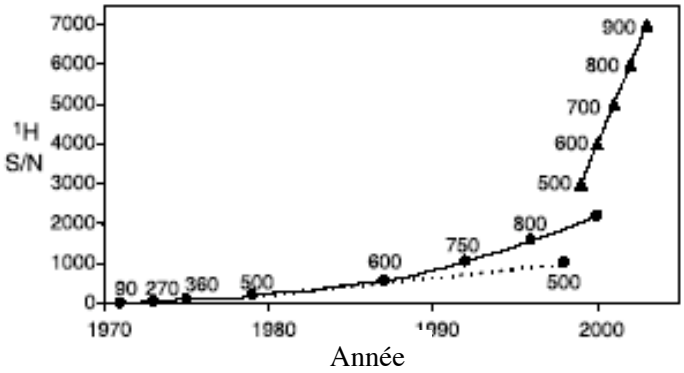


Figure 47: Rapport signal sur bruit (S/N) des résonances du proton de 0.1% d'ethylbenzène dans $CDCL_3$ en fonction de l'année de fabrication du spectromètre et du champ magnétique. Le champ magnétique est indiqué au-dessus de chaque point. Les ronds et les triangles donnent les valeurs obtenues avec une sonde normale ou une cryosonde, respectivement (adapté de Kovacs *et al.*, 2005). Les lignes pointillées donnent l'augmentation en sensibilité obtenue ces dernières années pour une sonde conventionnelle à une fréquence proton de 500 MHz.

I. 3. Résolution spectrale

La *résolution* d'un spectre est définie par l'écart en fréquence minimal $\Delta\nu$ nécessaire pour séparer deux raies de résonance (Figure 48). La résolution d'un spectre est conditionnée par la position relative des raies de résonance, et par leur largeur. La largeur de raie peut être dictée par les propriétés de relaxation de la transition observée dans les conditions de l'étude, ou élargie par les procédés d'acquisition du spectre (on parle de résolution digitale).

En RMN des macromolécules biologiques, la résolution des spectres est cruciale pour l'attribution non ambiguë des résonances et l'extraction de contraintes expérimentales. La séparation des pics est un problème d'autant plus présent que la taille du système étudié est importante. En effet, le nombre de raies de résonance augmente avec la taille du système. De plus, les propriétés de relaxation sont moins favorables pour les systèmes de grosse taille, ce qui élargit les pics de résonance. Enfin, comme nous l'avons vu dans l'introduction de ce document, les macromolécules biologiques sont constituées par la répétition d'un petit nombre d'entités chimiques. L'environnement chimique des différents noyaux est donc très semblable, ce qui conduit à une superposition des différents signaux dans des gammes spectrales réduites. Cette constatation est encore plus flagrante pour les acides nucléiques que pour les protéines, pour deux raisons. Tout d'abord, il n'existe que quatre

nucléotides de base, contre vingt types d'acides aminés pour les protéines. De plus, les acides nucléiques adoptent très souvent des structures en hélice qui placent les atomes des différents nucléotides dans des environnements chimiques similaires. Les protéines, quant à elles, présentent couramment une structuration en hélice α ou en feuillet β. La liste des plages de déplacements chimiques ^{1}H, ^{13}C, et ^{15}N usuels trouvés dans les protéines et les acides nucléiques est donnée dans l'Annexe 1.

Pour augmenter la résolution des spectres RMN, il existe plusieurs stratégies qui peuvent être combinées. La première repose sur l'affinement des raies de résonance (Figure 48).

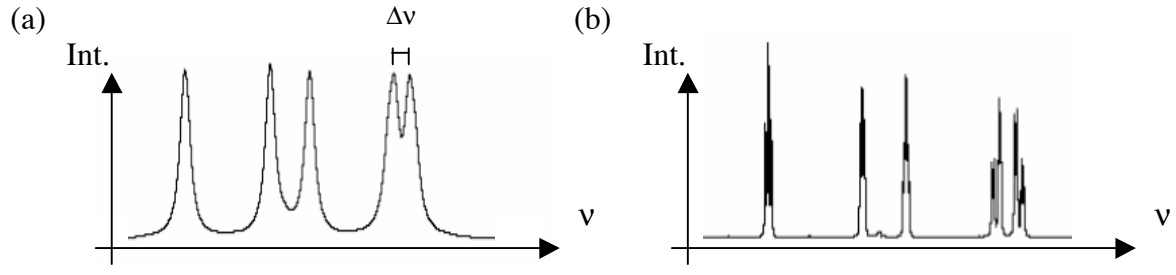


Figure 48: Spectres unidimensionnels enregistrés avec différentes largeurs de raie. (a) La résolution est tout juste suffisante pour séparer les deux pics de la droite du spectre. (b) Une résolution meilleure permet de séparer une structure fine de ces pics.

La largeur minimale d'un pic est déterminée par les propriétés de relaxation de l'état de spin qui est édité. Pour rendre plus favorables les propriétés de relaxation d'un système de spins, on peut agir à plusieurs niveaux. Tout d'abord, on peut choisir la composante qui relaxe le moins vite, par exemple en utilisant des expériences de type TROSY. Des solutions biochimiques sont également utilisées, où un enrichissement sélectif en isotopes de spin 1/2 est pratiqué, de manière à supprimer certains chemins de relaxation (Lian et Middleton, 2001). Les conditions expérimentales (température, viscosité) peuvent également être optimisées pour réduire le temps de corrélation de la molécule.

A cette largeur de raie naturelle peuvent s'ajouter des élargissements supplémentaires dus au mode d'acquisition, et aux inhomogénéités du champ magnétiques. Ces causes doivent être minimisées. Des techniques de traitement du signal peuvent également affiner les raies (prédiction linéaire, ou non linéaire). Cependant, malgré ces techniques d'amélioration de la largeur de raie des résonances, un grand nombre de pics restent superposés dans les spectres des macromolécules biologiques (Figure 47 et Figure 49).

La deuxième stratégie pour résoudre les problèmes de résolution est apparue en 1971 sous le nom de *spectroscopie multidimensionnelle* (Jeener, 1971; Aue *et al.* , 1976). Cette technique consiste à séparer les pics selon une dimension supplémentaire. Le principe est simple : on transfère l'aimantation du spin dont on veut éditer la fréquence sur un autre spin, par couplage scalaire ou dipolaire par exemple. Ensuite, on édite la fréquence de résonance de ce spin supplémentaire. On obtient dans ce cas un spectre bidimensionnel, comportant des pics de corrélation aux fréquences de résonance de chaque paire de spins couplés (Figure 49).

Cette technique peut être étendue, et on obtient ainsi des spectres 3D, 4D, nD corrélant les fréquences de résonance de tout un réseau de spins couplés. Cette approche profite largement de l'enrichissement des macromolécules biologiques en isotopes de spin 1/2, en particulier en ^{15}N et ^{13}C.

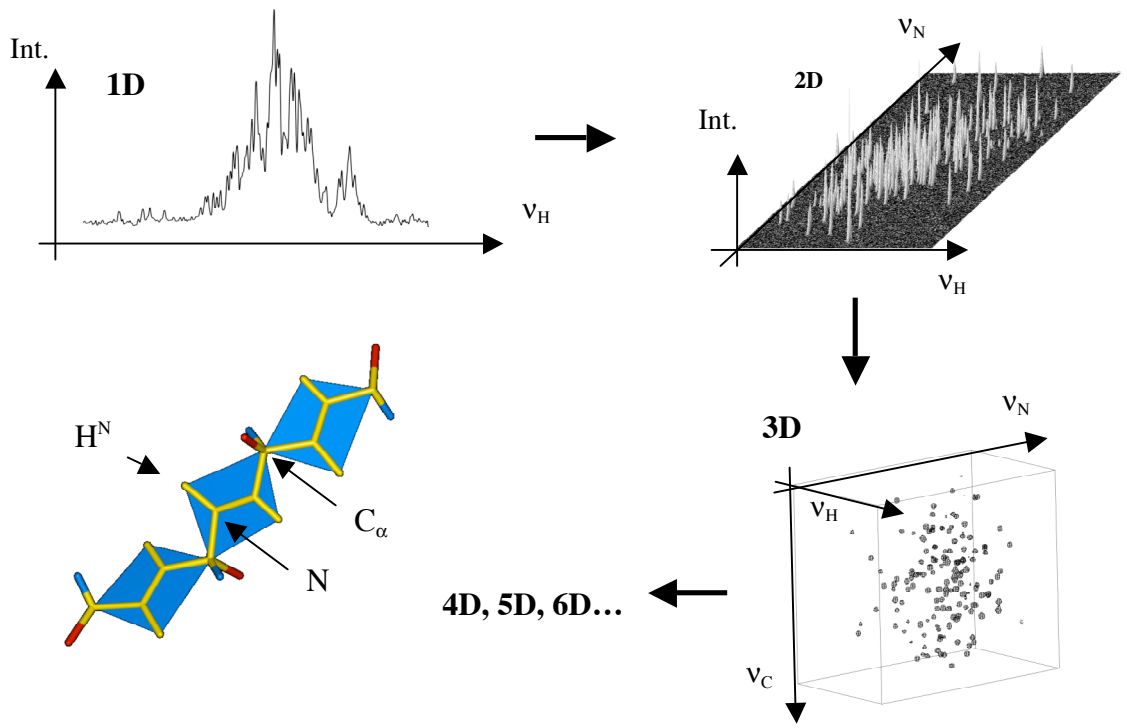


Figure 49: Figure illustrant les principes de la RMN multidimensionnelle sur l'exemple des spectres de corrélation HNCO utilisés pour l'attribution des résonances de la chaîne principale des protéines (ici pour une protéine de 167 résidus). L'aimantation des spins à éditer (ici le proton amide de la chaîne principale des protéines) est transférée sur un autre spin (ici l'azote auquel il est lié covalemment, par couplage scalaire). La fréquence de ce spin est éditée le long d'une dimension supplémentaire. Ce procédé peut être répété en transférant l'aimantation à un troisième spin : une troisième dimension fréquentielle est alors créée (dans notre exemple, l'aimantation est transférée au carbone alpha de la chaîne peptidique). Des expériences de dimensionalité plus élevée (4D, 5D…) peuvent être obtenues en éditant les fréquences d'autres spins couplés.

La séparation des pics superposés peut également se faire grâce à d'autres propriétés que la fréquence de résonance d'un spin couplé à celui qu'on observe. Par exemple, dans le cas d'un mélange de molécules, on peut séparer les pics des différents spins en fonction de la taille de la molécule où ils se trouvent en tirant profit des différentes propriétés de relaxation ou de diffusion. Dans de nombreux cas, la dimension supplémentaire qui est créée est discrète, et on parle de *filtrage* (Figure 50). Le marquage sélectif d'uniquement certains sites avec des isotopes de spin 1/2 constitue également un moyen de réduire le nombre de signaux dans un spectre. De plus, des filtres spectroscopiques ont été développés pour séparer les signaux des protons des molécules enrichies en ^{15}N et ^{13}C de ceux qui proviennent de molécules non marquées. Enfin, la valeur de couplages scalaires et la multiplicité d'un atome peuvent être utilisées pour séparer certaines raies de résonance.

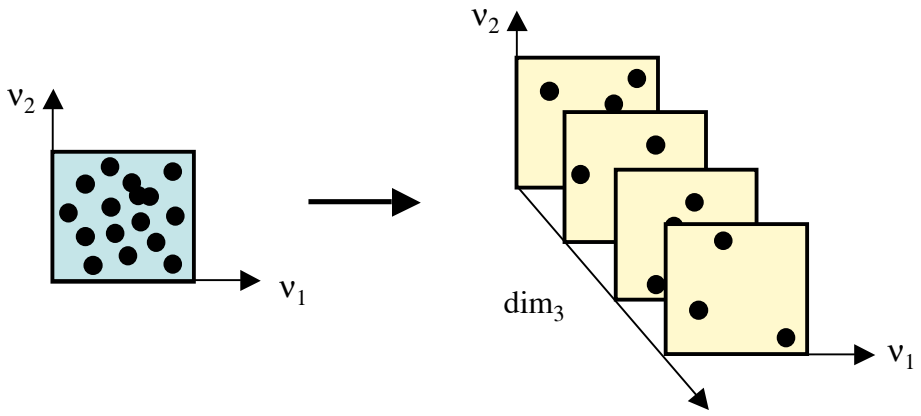


Figure 50: Principe du filtrage spectral sur l'exemple d'un spectre bidimensionnel. Les pics de corrélation sont dispersés dans plusieurs sous-spectres. La dimension supplémentaire qui est créée est discrète et comporte un nombre fini de points.

La taille des protéines et des acides nucléiques et leurs propriétés biochimiques conduisent à des spectres pour lesquels l'optimisation de la sensibilité et de la résolution est cruciale (Figure 46). Le but de ce chapitre est de montrer comment le choix de la technique d'acquisition des spectres RMN peut influencer la qualité du spectre que l'on obtient. En particulier, la spectroscopie par TF et la spectroscopie Hadamard seront comparées dans le contexte de l'acquisition de spectres multidimensionnels pour l'étude des macromolécules biologiques.

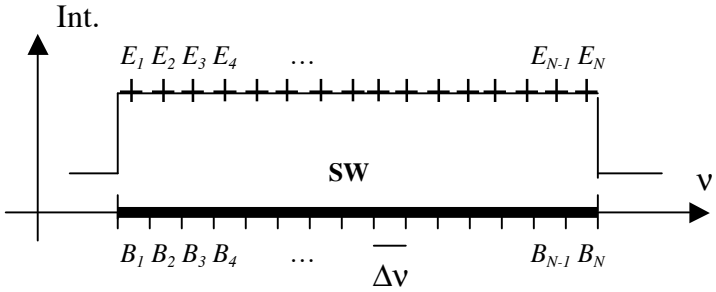
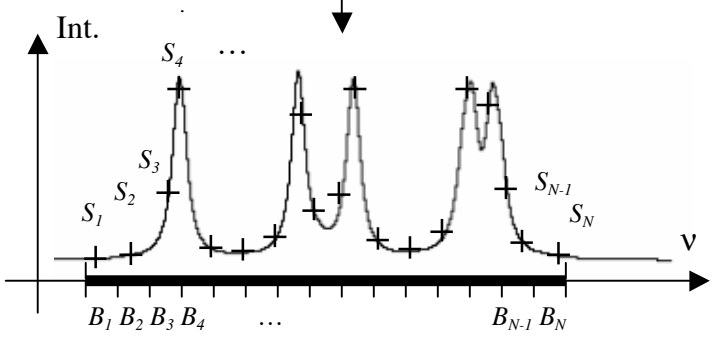
II. Principe des techniques d'acquisition des spectres RMN

L'obtention d'un spectre RMN passe par une phase expérimentale où les différents spins sont excités par l'intermédiaire de champs radiofréquences et de délais, et où leur réponse est enregistrée sous forme d'un signal électrique. Ce chapitre décrit les trois techniques majeures qui ont été proposées jusqu'à présent pour enregistrer un spectre RMN.

II. 1. Principe général

Toutes les méthodes qui ont été proposées pour faire l'acquisition des spectres RMN reposent sur un principe commun qui consiste à digitaliser la plage spectrale à éditer en *N* bandes de fréquence (Figure 51). Le nombre *N* est imposé par la largeur spectrale *SW* et la résolution Δν souhaitées par la relation : $N = \frac{SW}{\Delta\nu}$. Obtenir le spectre, c'est trouver l'intensité du signal RMN pour chacune des *N* bandes de fréquence. De façon pratique, les spins de l'échantillon sont excités par des champs radiofréquences, et leur réponse est enregistrée par le spectromètre sous forme d'une intensité électrique qui est la superposition des intensités de l'ensemble des bandes de fréquence excitées. *N* données expérimentales sont donc nécessaires pour réussir à séparer les signaux des différentes bandes. Le procédé de codage et de décodage utilisé pour obtenir l'intensité de chaque bande dépend du type de spectroscopie qu'on utilise, et conditionne la manière dont l'acquisition et la transformation des données vont être effectuées. La Figure 51 explicite le schéma expérimental général d'obtention d'un spectre RMN, ainsi qu'une description matricielle des différentes étapes. Les paragraphes qui suivent donneront des illustrations concrètes de cette modélisation générale.

Figure 51 (page suivante): Schéma de principe décrivant l'édition d'une dimension spectrale. La plage spectrale à éditer est divisée en *N* bandes de fréquence $B_1, B_2, ..., B_N$, où *N* est lié à la largeur spectrale *SW* et à la résolution Δν souhaitées par $N = \frac{SW}{\Delta\nu}$. Le vecteur *E* décrit l'intensité $E_1, E_2, ..., E_N$ de l'excitation imposée à chacune des bandes $B_1, B_2, ..., B_N$ pour obtenir le spectre final. Le spectre final est décrit par le vecteur *S* contenant l'ensemble des intensités $S_1, S_2, ..., S_N$ de chacune des bandes $B_1, B_2, ..., B_N$ après une excitation homogène sur la plage de fréquences *SW* que l'on veut observer. Pour obtenir ce spectre en pratique, on effectue *N* expériences, où un certain nombre de bandes sont excitées avec un codage particulier. Le vecteur *Exp* décrit quelles bandes de fréquence doivent être excitées dans chacune des *N* expériences à réaliser. Ce codage est effectué en appliquant la matrice de codage *H* au vecteur *E*. Les *N* expériences sont appliquées successivement sur le système de spin, ce qui fournit un jeu de *N* données expérimentales, qui sont rangées dans le vecteur *Data*. Ces données sont ensuite transformées au moyen de la matrice de décodage H^{-1} pour construire le spectre final *S*. Les différentes matrices sont explicitées dans les paragraphes suivants dans le cas des spectroscopies par onde continue, par TF, et Hadamard.

Etapes	Schéma expérimental	Représentation matricielle
Choix du spectre d'excitation		$E = \begin{bmatrix} E_1 \\ E_2 \\ \dots \\ E_N \end{bmatrix}$
Codage	Choix des séquences d'impulsions	Matrice de codage H
Excitation du système de spins	Réalisation de N expériences successives (voir Figure 54)	$Exp = \begin{bmatrix} Exp_1 \\ Exp_2 \\ \dots \\ Exp_N \end{bmatrix} = H \times E$
Réponse du système de spins Acquisition	Enregistrement de N données expérimentales	$Data = \begin{bmatrix} Data_1 \\ Data_2 \\ \dots \\ Data_N \end{bmatrix}$
Décodage	Traitement des données	Matrice de décodage H^{-1}
Spectre RMN final		$S = \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} = H^{-1} \times Data$

Pour l'implémentation pratique de ce principe, trois techniques différentes ont été proposées en RMN: la 'Spectroscopie par onde continue', la 'Spectroscopie par TF', qu'on pourrait désigner de façon plus complète par 'Spectroscopie RMN dans le domaine des temps utilisant la Transformée de Fourier', et la 'Spectroscopie Hadamard', qui est une abréviation utilisée pour désigner le concept de 'Spectroscopie RMN dans le domaine des fréquences avec codage Hadamard'. Cette partie détaille l'historique et les principes de ces trois techniques spectroscopiques utilisées en RMN.

II. 2. RMN par onde continue

Pour connaître la réponse d'un système de spins à une excitation sur N bandes de fréquence $B_1, B_2, \ldots, B_N$, la technique la plus simple est d'exciter les bandes une à une, d'enregistrer la réponse de chaque bande, et de reconstruire le spectre par superposition des réponses de toutes les bandes. N expériences doivent donc être réalisées pour obtenir un spectre de résolution $\Delta\nu = \frac{SW}{N}$. Cette technique est appelée spectroscopie par onde continue dans le domaine des fréquences, car chaque fréquence est excitée individuellement et successivement. Elle est utilisée dans de nombreux types de spectroscopies, par exemple en spectroscopie UV, IR, ou RPE, et a été utilisée en RMN à ses débuts (dans les années 1960).

Malheureusement, la sensibilité intrinsèque de cette méthode est pauvre, puisque seul le signal d'une unique bande de fréquence est enregistré à chaque passage. La sensibilité est d'ailleurs d'autant plus mauvaise que la résolution du spectre (et donc le nombre N de bandes du spectre) est grande, ce qui est un compromis toujours très cornélien.

Cette technique spectroscopique peut être décrite de façon matricielle. En spectroscopie par onde continue, chaque bande de fréquence n'est excitée que dans une unique expérience. Le vecteur Exp_{OC} décrivant quelles bandes de fréquence doivent être excitées dans les N expériences à réaliser est donc liée au vecteur E décrivant les intensités d'excitation de chaque bande $B_1, B_2, \ldots, B_N$ par la relation :

(Equation 2)
$$\begin{bmatrix} Exp_1 \\ Exp_2 \\ \ldots \\ Exp_N \end{bmatrix}_{OC} = \begin{bmatrix} E_1 \\ E_2 \\ \ldots \\ E_N \end{bmatrix}$$

Cette équation peut encore s'écrire :

(Equation 2') $$\begin{bmatrix} Exp_1 \\ Exp_2 \\ \dots \\ Exp_N \end{bmatrix}_{OC} = H_{OC} \times \begin{bmatrix} E_1 \\ E_2 \\ \dots \\ E_N \end{bmatrix}, \text{ avec } H_{OC} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

La deuxième écriture fait intervenir la matrice de codage H_{OC} qui permet de décrire la façon dont on va exciter chacune des bandes dans chacune des N expériences. Dans le cas de la spectroscopie par onde continue, la matrice H_{OC} est simplement égale à la matrice unité.

Cette écriture matricielle permet également de décrire le décodage des données expérimentales $Data_{OC}$ grâce à la matrice de décodage H_{OC}^{-1}, pour obtenir le vecteur spectre final S :

(Equation 3) $$\begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} = H_{OC}^{-1} \times \begin{bmatrix} Data_1 \\ Data_2 \\ \dots \\ Data_N \end{bmatrix}_{OC}, \text{ } i.e. \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} = \begin{bmatrix} Data_1 \\ Data_2 \\ \dots \\ Data_N \end{bmatrix}_{OC}$$

Dans notre cas, H_{OC}^{-1}est aussi égale à la matrice unité. Le signal enregistré dans la première expérience est en effet directement égal à celui de la première bande de fréquence.

N. B. Le formalisme matriciel semble un peu artificiel sur cet exemple simple. Il va néanmoins permettre de décrire les autres techniques spectroscopiques grâce à un formalisme très compact.

II. 3. Spectroscopie Hadamard dans le domaine des fréquences

Pour résoudre la perte de sensibilité inhérente à l'excitation successive des bandes de fréquence, il a été proposé dans les années 1960 d'effectuer une combinaison intelligente des bandes à exciter (Golay, 1949 ; Anderson 1966).

- **Quelques exemples simples**

Imaginons le cas simple de deux bandes de fréquence B_1 et B_2. Plutôt que d'enregistrer deux expériences contenant chacune uniquement le signal de B_1 et celui de B_2 (E_1 et E_1, respectivement), on peut enregistrer une première expérience où les signaux des deux bandes sont de même signe, par exemple ($E_1 + E_2$), et une autre où les signaux sont inversés, par exemple (E_1-E_2). En additionnant et en soustrayant les deux expériences, on obtient les signaux correspondant respectivement aux bandes B_1 et B_2 dans deux spectres différents. L'avantage de cette méthode est que le signal correspondant à chaque bande est contenu dans chacune des expériences, ce qui fait que dans les spectres finaux les intensités s'ajoutent et qu'il n'y a pas de perte en sensibilité.

Dans notre cas simple, nous pouvons écrire le procédé de codage des expériences à enregistrer sous la forme matricielle suivante:

(Equation 4) $\begin{bmatrix} Exp_1 \\ Exp_2 \end{bmatrix}_{HAD} = H_2 \times \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}$, où $H_2 = \begin{bmatrix} + & + \\ + & - \end{bmatrix}$ est appelée matrice Hadamard d'ordre 2.

Dans cette écriture, le signe '+ ' équivaut à +1 et le signe '–' à –1. Cette convention d'écriture sera conservée tout au long du texte.

Pour obtenir les spectres correspondant à B_1 et B_2, il suffit alors d'appliquer le procédé de décodage des signaux suivant:

(Equation 5) $\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = H_2^{-1} \times \begin{bmatrix} Data_1 \\ Data_2 \end{bmatrix}_{HAD}$ où $H_2^{-1} = \begin{bmatrix} + & + \\ + & - \end{bmatrix}$ est l'inverse de la matrice H_2.

Si on veut effectuer le même procédé avec quatre bandes B_1, B_2, B_3, et B_4, il suffit d'enregistrer quatre expériences où le signe des quatre bandes varie selon la matrice H_4, appelée matrice Hadamard d'ordre quatre :

(Equation 6) $$\begin{bmatrix} Exp_1 \\ Exp_2 \\ Exp_3 \\ Exp_4 \end{bmatrix}_{HAD} = H_4 \times \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix}, \quad \text{où} \quad H_4 = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix}$$

Le codage se fait de nouveau en lisant les lignes de la matrice. Dans la première expérience, les signaux de toutes les bandes seront positifs. Dans la deuxième expérience, les signaux des bandes B_2 et B_4 seront inversés en signe par rapport aux deux autres. Dans la troisième, les bandes B_3 et B_4 seront inversées, et dans la quatrième, les bandes B_2 et B_3.

L'obtention des signaux propres à chaque bande de fréquence se fait par un procédé de décodage qui consiste à faire la somme ou la différence de chacune des expériences enregistrées. Cela se traduit sous forme matricielle par :

(Equation 7) $$\begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{bmatrix} = H_4^{-1} \times \begin{bmatrix} Data_1 \\ Data_2 \\ Data_3 \\ Data_4 \end{bmatrix}_{HAD}$$

De façon pratique, on fait une sommation de toutes les expériences pondérées par les signes des lignes de la matrice Hadamard H_4^{-1}. Les matrices Hadamard ont la propriété que leur matrice inverse est égale à un facteur près à leur transposée. Pour le décodage des bandes, il faut donc s'intéresser aux lignes de la matrice transposée de H_4, ce qui revient à lire les colonnes de la matrice H_4. Par exemple, pour obtenir le signal propre à la bande B_1, il faut faire la somme des quatre expériences (colonne 1 de la matrice H_4). Pour la bande B_2, les signaux de l'expérience 1 et 3 sont ajoutés, et ceux des expériences 2 et 4 soustraits (colonne 2). De même, les signaux correspondant aux bandes B_3 et B_4 seront obtenus en lisant les colonnes 3 et 4 de la matrice H_4.

Par ce procédé, la sensibilité correspondant à la bande sélectionnée est quatre fois plus grande que celle d'une unique expérience, et les signaux des trois autres bandes s'annulent. On voit sur cet exemple que le problème de perte en sensibilité observée pour la spectroscopie par onde continue est totalement résolu par ce codage ingénieux.

- **Généralisation : les matrices Hadamard**

Cette stratégie peut être généralisée pour un nombre de bandes N plus élevé (Figure 52). Il suffit pour cela d'enregistrer N expériences où les signes des signaux des différentes bandes B_1, B_2, ..., B_N varient selon une matrice H_N appelée matrice Hadamard. Le décodage du signal de chacune des bandes s'effectue par addition ou soustraction des différentes expériences selon la matrice inverse H_N^{-1}.

(Equation 8) $$\begin{bmatrix} Exp_1 \\ Exp_2 \\ \dots \\ Exp_N \end{bmatrix}_{HAD} = H_N \times \begin{bmatrix} E_1 \\ E_2 \\ \dots \\ E_N \end{bmatrix}$$ (Equation 9) $$\begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} = H_N^{-1} \times \begin{bmatrix} Data_1 \\ Data_2 \\ \dots \\ Data_N \end{bmatrix}_{HAD}$$

Figure 52: Représentation matricielle de la technique de codage Hadamard pour N bandes spectrales B_1, B_2, ..., B_N. (Equation 8) Codage des signes des signaux des différentes bandes à exciter pour les N expériences à enregistrer Exp_1, Exp_2, …, Exp_N. (Equation 9) Technique de décodage des signaux des différentes bandes à partir des N expériences enregistrées.

Certaines des matrices Hadamard H_N ont été découvertes par James Joseph Sylvester en 1867, et d'autres par Jacques Hadamard en 1893. Ces matrices n'existent que pour des ordres $N = 2^j$ ou $N = 4j$, où j est un nombre entier. La sous-famille des matrices Hadamard contenant les matrices d'ordre $N = 2^j$ peut être créée de manière itérative à partir de la matrice H_2 donnée plus haut en appliquant le développement suivant (Hadamard, 1893):

(Equation 10) $$H_{2^{j+1}} = \begin{bmatrix} H_{2^j} & H_{2^j} \\ H_{2^j} & -H_{2^j} \end{bmatrix}$$

où le signe '-' devant une matrice signifie que le signe de tous les éléments de matrice est inversé. Les propriétés des matrices Hadamard sont telles que la matrice inverse d'une matrice Hadamard est égale à sa transposée, à un facteur $1/N$ près : $H_N^{-1} = \frac{1}{N} H_N^T$.

L'utilisation de ce codage impose de travailler avec un nombre d'expériences N multiple de quatre. De plus, les N expériences doivent être toutes enregistrées pour rendre possible le décodage des signaux des différentes bandes. L'intérêt de cette méthodologie est un gain en sensibilité de $\sqrt{N}$ par rapport au mode par onde continue. En effet, chacune des N expériences enregistrées contribue au signal final de toutes les bandes de fréquence. On peut noter que cette propriété est due au fait que tous les termes des matrices Hadamard sont de module égal à un.

II. 4. La RMN impulsionnelle par Transformée de Fourier

Dans les années 1970, la spectroscopie impulsionnelle a révolutionné l'utilisation pratique de la RMN (Ernst et Anderson, 1966). Le grand progrès a été d'utiliser une impulsion polychromatique large bande pour exciter l'ensemble des spins dans chaque expérience (RMN impulsionnelle). En plus du gain en sensibilité qu'elle offre par rapport à la RMN par onde continue, la spectroscopie impulsionnelle a permis le développement de la spectroscopie multimensionnelle (Jeener, 1971; Aue *et al.* , 1976), ce qui a fait d'elle la méthode reine de la RMN moderne.

Pour mettre en pratique la spectroscopie impulsionnelle, il a fallu adapter le mode d'excitation et de traitement des expériences RMN, et la transformée de Fourier a été choisie comme moyen d'effectuer le codage et le décodage décrits dans la Figure 51. La spectroscopie par TF est utilisée actuellement dans toutes les expériences RMN classiques. Elle comporte beaucoup d'avantages pratiques qui seront détaillés plus loin. Dans la littérature, on a souvent tendance à opposer la 'spectroscopie dans le domaine des temps' utilisant un codage par TF, et la 'spectroscopie dans le domaine des fréquences'. La spectroscopie par TF peut néanmoins être décrite par le même formalisme que la spectroscopie Hadamard. Cette description matricielle n'est pas celle qui est utilisée habituellement pour expliquer la spectroscopie par TF. Elle permet cependant d'unifier les concepts et de pouvoir mieux comparer les différents types de spectroscopies.

Dans le formalisme de la Figure 51, la matrice de codage *H* et le vecteur *Exp* utilisés en RMN par TF pour un codage à *N* bandes de fréquence peuvent être décrits par les équations suivantes:

(Equation 11)
$$H_{TF} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{2i\pi\frac{1}{N}} & e^{2i\pi\frac{2}{N}} & \dots & e^{2i\pi\frac{N-1}{N}} \\ 1 & e^{2i\pi\frac{2}{N}} & e^{2i\pi\frac{4}{N}} & \dots & e^{2i\pi\frac{2(N-1)}{N}} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & e^{2i\pi\frac{N-1}{N}} & e^{2i\pi\frac{2(N-1)}{N}} & \dots & e^{2i\pi\frac{(N-1)^2}{N}} \end{bmatrix}$$

(Equation 12)
$$Exp_{TF} = H_{TF} \times E = \begin{bmatrix} E_1 + E_2 + E_3 + \dots + E_N \\ E_1 + e^{2i\pi\frac{1}{N}} E_2 + e^{2i\pi\frac{2}{N}} E_3 + \dots + e^{2i\pi\frac{N-1}{N}} E_N \\ E_1 + e^{2i\pi\frac{2}{N}} E_2 + e^{2i\pi\frac{4}{N}} E_3 + \dots + e^{2i\pi\frac{2(N-1)}{N}} E_N \\ \dots \\ E_1 + e^{2i\pi\frac{N-1}{N}} E_2 + e^{2i\pi\frac{2(N-1)}{N}} E_3 + \dots + e^{2i\pi\frac{(N-1)^2}{N}} E_N \end{bmatrix}$$

A partir de ces matrices, nous pouvons retrouver les propriétés connues de la spectroscopie par TF :

La matrice Exp_{TF} donnée dans l'équation 12 explicite les différentes expériences à effectuer pour obtenir un spectre de résolution $\Delta\nu = \frac{SW}{N}$. Chaque ligne décrit en effet comment doivent être excitées les différentes bandes de fréquence dans chaque expérience. Par exemple, toutes les bandes de fréquence sont excitées de la même façon dans la première expérience, décrite par la première ligne. Pour les autres expériences, on voit que le codage des différentes bandes s'effectue en changeant la *phase* de chaque bande. Ce codage en phase permet de laisser l'amplitude de l'excitation de chaque bande égale à 1 dans chaque expérience. Le signal de chacune des bandes est donc accumulé avec une amplitude maximale dans chaque expérience, ce qui engendre un gain en sensibilité de $\sqrt{N}$ par rapport à la spectroscopie par onde continue. Ce gain en sensibilité peut vite devenir important quand N augmente. Tout comme pour la spectroscopie Hadamard, la résolution d'un spectre obtenu par la technique de spectroscopie impulsionnelle par TF ne se fait plus au prix de la sensibilité.

D'un point de vue pratique, ce codage en phase est obtenu en laissant évoluer l'hamiltonien pour les spins de chacune des bandes pendant un incrément de temps donné. Par exemple, le premier point expérimental est enregistré à un temps d'édition nul, de façon à ce que toutes les bandes soient en phase (1e ligne du vecteur Exp_{TF}). Les autres points expérimentaux (obtenus après excitation selon les différentes lignes du vecteur Exp_{TF}) sont enregistrés après des délais croissants qui font évoluer l'état de spin des différentes bandes avec une phase qui dépend de leur fréquence de résonance. C'est pour cette raison que la spectroscopie par TF est également appelée *spectroscopie dans le domaine des temps*. Les points expérimentaux, dont l'intensité est la superposition des signaux des différentes bandes avec différentes phases, sont enregistrés et rangés dans le vecteur $Data_{TF}$. Pour obtenir les signaux de chacune des bandes (rangés dans S_{TF}), on effectue un décodage du vecteur $Data_{TF}$ en lui appliquant une transformée de Fourier inverse grâce à la matrice H_{TF}^{-1}.

Pour conclure, nous avons vu que la spectroscopie par TF peut se décrire de façon matricielle, tout comme la spectroscopie Hadamard. Les matrices de codage et de décodage sont cependant différentes pour les deux types de spectroscopie. Cette description va être utile pour effectuer une comparaison théorique et pratique de ces deux options pour l'acquisition des spectres RMN.

N.B. La transformée de Fourier n'est donc pas uniquement un moyen de *transformer* les données RMN. Le codage de l'édition par l'utilisation de délais incrémentables est en effet également nécessaire et propre à la spectroscopie par TF.

.

III. Spectroscopie par TF et spectroscopie Hadamard : concurrentes ou complémentaires?

La spectroscopie par TF et la spectroscopie Hadamard offrent toutes deux un gain en sensibilité de $\sqrt{N}$ par rapport à la spectroscopie par onde continue. Ce chapitre donne quelques autres éléments de comparaison théoriques et pratiques sur l'utilisation de la spectroscopie par TF et la spectroscopie Hadamard. Ce bilan pourra être utile pour guider un utilisateur dans le choix du mode d'acquisition de chacune des dimensions du spectre RMN qu'il veut enregistrer.

III. 1. Fondements théoriques

Pour expliquer pourquoi on peut utiliser indifféremment la spectroscopie par onde continue, la spectroscopie Hadamard, ou la spectroscopie par TF, pour obtenir un spectre RMN, il faut se référer au théorème d'algèbre linéaire suivant :

'Toute fonction appartenant à un ensemble de fonctions peut être décomposée en une série de fonctions de base de cet ensemble'.

Soir *s* la fonction qui associe à chaque fréquence son intensité dans le spectre. Elle peut être décomposée sur une base de fonctions *f* sous forme intégrale (équation 13) ou discrète (équation 14).

(Equation 13) $$s(\nu) = \int k(\nu,t) f(t) dt$$, pour toute fréquence ν.

(Equation 14) $$s_m = \frac{1}{N}\sum_{n=1}^{N} k(m,n) f_n$$

pour tout *m* compris entre 1 et *N*, où s_m représente l'intensité du spectre pour la bande de fréquence B_m.

La transformation inverse, de la nouvelle base vers l'ancienne, s'écrit :

(Equation 15) $$f_n = \frac{1}{N}\sum_{m=1}^{N} h(n,m) g_m$$ pour tout *n* compris entre 1 et *N*.

Les coefficients *h(m,n)* et *k(m,n)* définissent des matrices de changement de base, inverses l'une de l'autre. Le procédé de codage et de décodage décrits dans la Figure 51 est donc équivalent à un changement de base pour décomposer la fonction spectrale *s*. On passe de la base naturelle pour enregistrer le spectre (celle de la spectroscopie par onde continue : $B_1, B_2, ..., B_N$) à une base nouvelle par les matrices de changement de base *H*. L'excitation des spins *Exp* et l'enregistrement des données expérimentales *Data* sont effectués dans cette nouvelle base. Pour revenir dans la base d'origine ($B_1, B_2, ..., B_N$) de façon à visualiser le spectre de façon conventionnelle, on effectue un changement de base inverse (Figure 51).

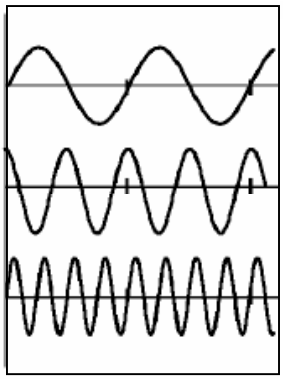
La fonction s et le vecteur S, qui représentent le spectre de l'échantillon qu'on cherche à obtenir, sont des quantités physiques indépendantes de la base choisie pour les décomposer. En revanche, la matrice de changement de base H, le vecteur d'excitation Exp et le vecteur contenant les résultats de la détection $Data$ dépendent du choix de la base expérimentale. D'un point de vue purement théorique, le choix de la base utilisée pour décomposer et mesurer le vecteur spectre S est complètement indifférent. En pratique, le choix du spectroscopiste va se faire en considérant les avantages pragmatiques de chacune des bases. Les points importants dans cette démarche sont la recherche de sensibilité, l'absence d'artéfacts lors de l'implémentation pratique, et l'obtention des spectre dans un temps expérimental le plus court possible.

Concernant la sensibilité, on peut généraliser ce qui a été décrit pour la spectroscopie Hadamard et par TF. Pour que chaque expérience contienne le signal de chacune des bandes, il faut faire un codage en phase, et non en amplitude, de façon à ce que tous les termes soient codés avec un module égale à un. La matrice de changement de base peut alors s'écrire :

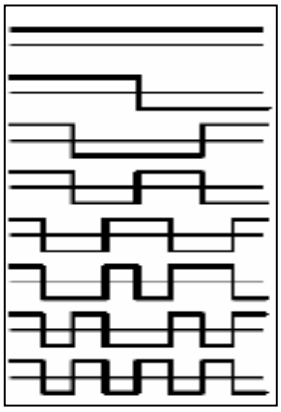
(Equation 16) $H = \left[e^{i\Delta\phi_{mn}} \right]$, où m et n indiquent les indices des lignes et des colonnes de la matrice.

Pour la spectroscopie par TF: $\Delta\phi = 2\pi \dfrac{(m-1)(n-1)}{N}$, et pour la spectroscopie Hadamard : $\Delta\phi = 0 \; ou \; \pi$. La Figure 53 donne quelques exemples graphiques des fonctions de base utilisées dans la décomposition en série de Fourier ou en série Hadamard, ainsi que leur représentation matricielle. D'autres codage en phase pourraient néanmoins être utilisées (Kupce et Freeman, 2003a).

(a)



(b)



$$H_{TF} = \left[h(m,n) \right] = \left[e^{2i\pi \frac{(m-1)(n-1)}{N}} \right]$$

$$H_{Had} = \left[h(m,n) \right] \text{ avec } h(m,n) = +/-1.$$

Figure 53: Géométrie de quelques fonctions de base, et matrice de codage (de changement de base) utilisées (a) en décomposition de Fourier, (b) en décomposition Hadamard.

Contrairement à d'autres codages en phase, les fonctions de Fourier et les fonctions Hadamard ont l'avantage d'avoir un lien direct avec la géométrie des fonctions spectrales *s* que l'on étudie. En effet, après leur excitation, les spins émettent un signal oscillant à leur fréquence de résonance lors de leur retour à l'équilibre. Les spectres RMN sont par définition constitués par la superposition de signaux monochromatiques provenant des différentes transitions énergétiques des spins de la molécule. La décomposition selon une série de Fourier, composée de fonctions sinus et cosinus, semble donc être très naturelle (Figure 53). Une autre idée peut naître si on considère qu'un spectre RMN est un objet contenant, certes, quelques endroits où il existe du signal, mais souvent entourés par de larges régions vides. L'information qui est recherchée dans les spectres est très souvent de type binaire: le pic existe (1), ou pas (0). Dans cette optique, une base de décomposition adaptée est celle des fonctions Hadamard. Ces fonctions constituent une sous-classe des fonctions de Walsh (Walsh, 1923), et ne peuvent prendre que les valeurs 1 ou –1. Ces sont des fonctions qui sont souvent utilisées en logique binaire.

Le deuxième critère pour le choix de la base de décomposition concerne l'implémentation pratique de chaque type de spectroscopie, qui conditionne le temps expérimental nécessaire pour obtenir un spectre RMN et la présence éventuelle d'artéfacts. Les paragraphes suivants détaillent dans cette perspective les avantages pratiques de la spectroscopie par TF et de la spectroscopie Hadamard.

III. 2. Spectroscopie par TF et problème du temps expérimental

• Forces et faiblesses de la spectroscopie par TF

L'ingéniosité de la spectroscopie par TF réside dans le fait qu'on peut obtenir les *N* points expérimentaux nécessaires à l'enregistrement d'un spectre (Figure 51) en un *unique* scan dans la dimension de la détection (Figure 54). Cela est rendu possible car le codage en phase des différentes bandes de fréquence selon la matrice de Fourier est effectué en laissant agir l'hamiltonien sur le système de spins pendant un délai déterminé, ce délai étant incrémenté. Les *N* points expérimentaux sont donc enregistrés successivement, incrément de temps après incrément de temps, pendant la période d'acquisition.

Pour la dimension directe, la TF est donc la plus efficace des méthodes spectroscopiques au niveau de la durée minimale de l'expérience. En effet, pour toutes les autres méthodes existantes, l'acquisition des *N* points expérimentaux nécessite *a priori* la répétition de *N* scans, ce qui est *N* fois plus long (par exemple pour la spectroscopie Hadamard : Figure 54).

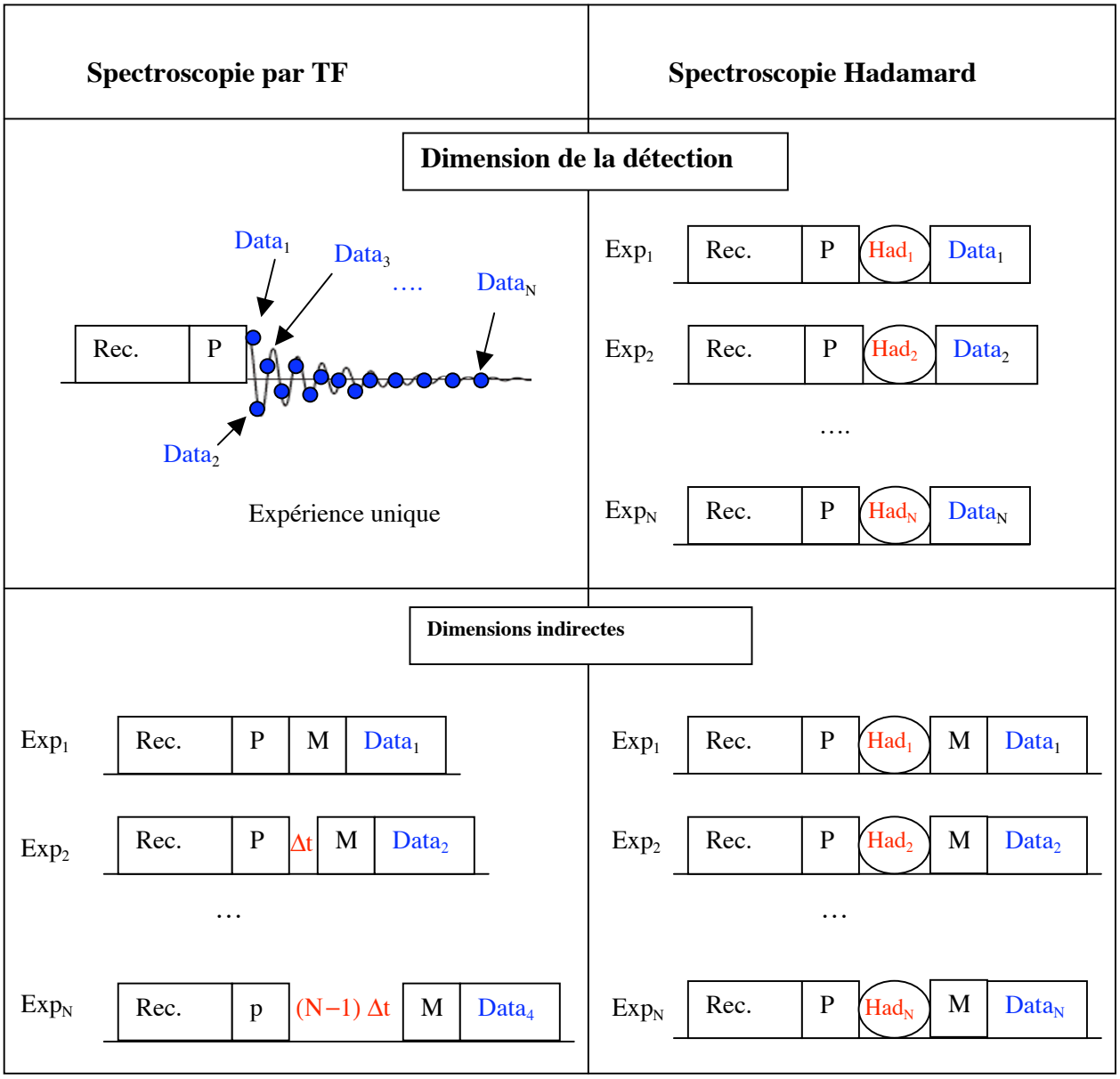


Figure 54: Description schématique des différentes expériences à réaliser pour obtenir les *N* points expérimentaux nécessaires à une édition spectrale (Figure 51). Ce schéma est présenté pour la spectroscopie par TF et la spectroscopie Hadamard dans les dimensions directe et indirectes. A condition d'utiliser le traitement approprié, i.e. la TF et le décodage Hadamard, respectivement, ces deux éléments de séquences sont équivalents en termes de résolution. Les blocs P et M sont identiques dans les deux types de spectroscopie, et représentent une succession de délais et d'impulsions radiofréquences nécessaires pour la création d'états de spins voulus (P : préparation, M : mélange). Le signal enregistré en spectroscopie Hadamard est un signal dont seule l'intensité est importante. Le spectroscopiste habitué à la spectroscopie par TF sera donc peut-être surpris de l'absence de délais temporels incrémentables en spectroscopie Hadamard, même dans la dimension de la détection. Le temps de recyclage (Rec.) est en revanche similaire pour les deux options, pour permettre à l'aimantation initiale d'être restaurée. La durée d'un scan est donc équivalent dans les deux spectroscopies.

Ce type de codage dans le domaine des temps a un prix : on n'a pas de prise directe dans le domaine des fréquences. En particulier, le codage par le temps définit une infinité de bandes de fréquence, puisque chaque spin va évoluer avec sa propre fréquence de résonance. Si le nombre de points expérimentaux enregistrés N n'est pas assez grand, l'échantillonnage n'est pas correct, et le décodage produit des artéfacts appelés 'artéfacts de troncature'. Des traitements post-acquisition permettent d'adoucir ces artéfacts, au prix d'une perte en sensibilité et en résolution.

De plus, l'excitation très large bande appliquée sur les spins génère éventuellement l'existence d'un pic intense du solvant et le repliement des résonances situées en dehors de la largeur spectrale sur le spectre d'intérêt. Des techniques spectroscopiques permettant de filtrer le signal du solvant, et le choix de la largeur spectrale permettent néanmoins de contourner ces inconvénients.

Actuellement, le problème majeur de l'utilisation de la spectroscopie par TF en RMN des macromolécules concerne le temps expérimental nécessaire pour effectuer des expériences multidimensionnelles. En effet, quand plusieurs dimensions sont éditées par TF dans une même expérience, on ne peut malheureusement utiliser le temps d'acquisition comme moyen d'éditer les N points expérimentaux que pour une seule des dimensions (le temps est unidimensionnel). Les points expérimentaux des dimensions supplémentaires sont alors enregistrés en répétant la séquence : excitation-acquisition-recyclage dans N expériences, en incrémentant le délai d'édition à chaque passage (Figure 55).

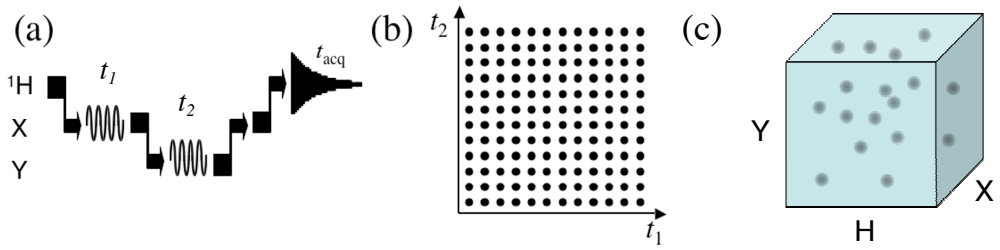


Figure 55: (a) Schéma expérimental d'une expérience de corrélation 3D H-X-Y. Les spins ^{1}H sont souvent excités au début et détectés à la fin d'une expérience pour des raisons de sensibilité. (b) Les deux variables de temps t_1 et t_2 des deux dimensions indirectes sont incrémentées indépendamment. Chaque point représente l'acquisition d'une expérience avec des incrément t_1 et t_2 particuliers. (c) Le spectre $S(\omega_1,\omega_1,\omega_{acq})=S(\omega_X,\omega_Y,\omega_H)$ est ensuite obtenu par transformation de Fourier successives le long des trois axes t_1, t_1 et t_{acq}. La RMN multidimensionnelle par TF constitue la base de la majorité des applications RMN actuelles dans le domaine de l'étude des biomolécules.

Le temps expérimental minimal T_{nD} nécessaire pour l'enregistrement d'un spectre de dimension n en utilisant la spectroscopie par TF est donc déterminé par le nombre d'incréments N_i dans les différentes dimensions indirectes, le nombre de répétitions N_S choisi pour moyenner le signal dans la dimension de la détection, et le délai de répétition τ_{rep} (équation 17). τ_{rep} est largement conditionné par le temps de recyclage nécessaire entre deux acquisitions pour permettre le retour à l'équilibre thermodynamique des spins.

(Equation 17) $$T_{nD} = 2^{n-1} \tau_{rep} N_S \prod_{i=1}^{n-1} N_i$$

Le facteur 2 de cette équation provient du fait que deux jeux de données sont nécessaires pour obtenir le signe des fréquences dans chacune des dimensions indirectes. La nécessité d'effectuer une détection en quadrature est l'inconvénient d'avoir enregistré l'expérience à l'aide d'un codage en phase, où les composantes réelles et imaginaires doivent être détectées.

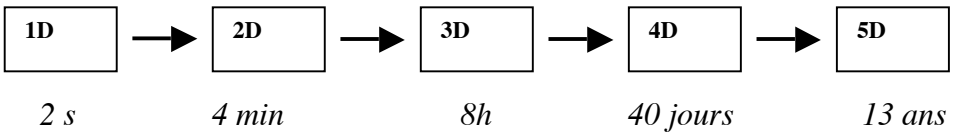


Figure 56: Temps minimal requis pour une expérience nD avec τ_{rep}=1s, N_i=60 et N_S=2 pour chaque dimension indirecte.

La Figure 56 donne un ordre de grandeur des temps d'acquisition pour les spectres RMN multidimensionnels des macromolécules biologiques. On s'aperçoit que la gamme de temps des expériences 2D ou 3D est compatible avec celle qui est imposée par la sensibilité des spectres des macromolécules biologiques, mais que l'acquisition de spectres de dimensionalité supérieure est irréaliste. De plus, pour 60 incréments de temps par dimension, la résolution du spectre enregistré est en général très inférieure à la résolution spectrale maximale imposée par la largeur de raie des résonances. La réalisation d'expériences 4D ou 5D par la méthode n'est donc possible qu'en compromettant très fortement la résolution dans les différentes dimensions. Pour résoudre ce problème, des techniques alternatives sont développées actuellement pour accélérer l'acquisition des spectres multidimensionnels.

- **Techniques d'acquisition rapide des spectres multidimensionnels**

Les techniques d'acquisition rapide ont pour but d'obtenir des spectres multidimensionnels en utilisant des moyens plus performants en termes de temps expérimental que ceux qui sont proposés par les protocoles implantés actuellement. De nombreuses techniques sont en cours de développement, dont aucune n'a encore imposé sa domination (Pour revue : Kupce et Freeman, 2003d ; Malmodin et Billeter, 2005). Ce paragraphe n'a pas l'ambition d'être exhaustif sur le sujet, mais uniquement d'illustrer la multiplicité des principes qui peuvent être utilisés, voire combinés, pour réduire le temps expérimental d'une expérience multidimensionnelle. Toutes ces techniques d'acquisition rapide s'intéressent à diminuer un ou plusieurs des facteurs qui influent sur le temps expérimental d'une expérience multidimensionnelle, donnés dans l'équation 15. On peut remarquer que toutes ont conservé le codage par TF qui est très efficace dans la dimension de la détection.

Tout d'abord, le nombre de scans *Ns* doit être minimisé. Il est souvent lié au cyclage de phase que l'on doit effectuer pour supprimer les artéfacts des spectres, et peut être réduit en utilisant par exemple une suppression des voies de cohérence non désirées par des techniques de gradients.

Ensuite, le délai de recyclage des expériences est dicté par la constante de retour à l'équilibre de l'aimantation longitudinale des spins de l'échantillon. Des méthodes rapides développées récemment ont permis de réduire ce temps de plusieurs ordres de grandeur en utilisant une excitation sélective des spins selon un angle optimal, qui favorise le retour à l'équilibre de l'aimantation longitudinale (Pervushin *et al.* , 2002 ; Schanda et Brutscher, 2005). Cette technique ne peut malheureusement pas être implémentée dans toutes les séquences d'impulsion.

Pour diminuer le nombre d'expériences à effectuer pour échantillonner les différents points des dimensions indirectes, plusieurs stratégies ont été proposées. La première consiste à faire un échantillonnage non uniforme des données dans le domaine des temps (Figure 57). Pour reprendre la description générale de la Figure 51, ces méthodes consistent à utiliser un codage en phase similaire à celui de la transformée de Fourier, mais où les intervalles de temps échantillonnés ne sont plus égaux (Marion, 2005). Le procédé de décodage utilisé dans ce cas est souvent non linéaire. Parmi les procédés de transformation des données, on peut citer par exemple la prédiction linéaire (Gesmar *et al.*, 1990), la méthode d'entropie maximale (Hoch et Stern, 2001), ou la méthode de filtre de diagonalisation (FDM) (Hu *et al.* , 2000). Ces techniques sont basées sur une optimisation itérative de reconstruction des spectres en relation avec les données expérimentales. Elles permettent d'accélérer l'acquisition d'un spectre de résolution donnée d'un ordre de grandeur quand la sensibilité des spectres le permet. La non-linéarité des procédés de transformation génère cependant des artéfacts qui dépendent de l'ajustement d'un certain nombre de paramètres.

Une des techniques d'échantillonnage non uniforme utilise un codage linéaire qui consiste à enregistrer les données de façon radiale dans l'espace des temps (Figure 57b). Elle est inspirée de la spectroscopie 'accordéon' (Bodenhausen et Ernst, 1982), et est désignée sous les noms de 'RMN à dimensionalité réduite' (Szyperski *et al.* , 1993 ; Simorre *et al.* , 1994), ou 'GFT-RMN' (Kim et Szyperski, 2003). Le décodage s'effectue par la combinaison de la transformée de Fourier et d'une matrice appelée matrice G. La visualisation des spectres projetés peut être faite directement en fonction de l'angle de projection (Brutscher *et al.* , 1995), mais des méthodes de reconstruction des spectres le long des dimensions fréquentielles ont également été proposées (Kupce et Freeman, 2004).

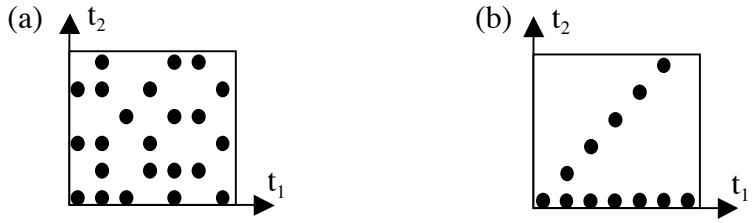


Figure 57: Principes de l'acquisition non uniforme des données dans deux dimensions indirectes t_1 et t_2. (a) Sur la grille de points d'enregistrement des données par TF (donnée Figure 56), seuls une partie d'entre eux ont été conservés. (b) Cas particulier de la RMN projetée : les points qui sont sélectionnés pour l'acquisition sont rangés sur une droite qui fait un angle donné avec les axes.

La méthode la plus révolutionnaire qui a été proposée pour effectuer l'acquisition des spectres multidimensionnels en un seul scan s'inspire des techniques d'imagerie (Frydman *et al.* , 2002). Le codage des dimensions indirectes est effectué de façon classique par TF, mais l'échantillonnage des différents points est effectué selon un axe de l'espace (x, y, z). L'échantillon est divisé en petits volumes qui subissent la même séquence d'impulsion, sauf en ce qui concerne la valeur des délais incrémentables utilisés pour l'édition des différentes dimensions indirectes. Les procédés de codage et décodage spatiaux sont implémentés à l'aide d'impulsions sélectives et de gradients de champ magnétique. Cette technique impose néanmoins une perte de sensibilité qui est proportionnelle au nombre d'incréments de temps dans chacune des dimensions. Une idée similaire de codage spatial a été proposée récemment en utilisant une sonde comportant plusieurs bobines (Wang *et al.* , 2005).

La dernière technique d'acquisition rapide proposée récemment est la spectroscopie Hadamard, qui a été remise au goût du jour pour l'occasion (Pour revue : Kupce *et al.* , 2003). Dans les paragraphes qui suivent, nous allons discuter comment et dans quelles conditions la spectroscopie Hadamard peut remplacer avantageusement la spectroscopie par TF pour accélérer l'acquisition des spectres RMN multidimensionnels.

III. 3. Implémentation pratique de la spectroscopie Hadamard

Depuis l'apparition de la spectroscopie impulsionnelle par TF, d'abord unidimensionelle puis multidimensionnelle, des séquences d'impulsions nombreuses et variées ont été développées, permettant d'attribuer les résonances et d'extraire des contraintes structurales et de dynamique. Pour utiliser la spectroscopie Hadamard plutôt que par TF dans toutes ces séquences, il suffit de remplacer n'importe quelle édition dans le domaine des temps par un codage Hadamard, et de changer conjointement le procédé de transformation du signal obtenu. Ce changement de technique d'acquisition/transformation peut être mis en œuvre aussi bien dans la dimension directe que dans les dimensions indirectes (Figure 54). De plus, le choix de la spectroscopie Hadamard ou par TF peut être fait de manière indépendante pour l'acquisition des différentes dimensions, ce qui laisse une grande liberté d'utilisation. Les paragraphes qui suivent décrivent différentes options pour l'implémentation pratique de la spectroscopie Hadamard.

- **Le codage binaire**

La spectroscopie Hadamard impose la condition pratique de faire un codage binaire +/- efficace des signaux provenant des différentes bandes à encoder. Plusieurs solutions ont été proposées au cours du temps pour mettre en place le codage Hadamard. En effet, l'idée de combiner astucieusement les bandes de fréquence à exciter a été imaginée et utilisée assez tôt en spectroscopie. En spectroscopie Infra-Rouge, Golay fut le premier à effectuer l'excitation de différentes fréquences simultanément (Golay, 1949). Ayant à sa disposition un émetteur et un détecteur large-bande, il mit au point un système de codage et de décodage basé sur l'utilisation de masques permettant de sélectionner certaines bandes de fréquence. En RMN, Anderson inventa un dispositif similaire de codage par multiplexage optique, couplé à l'utilisation de détecteurs monochromatiques (Anderson, 1966). Cependant, la facilité d'utilisation et les perspectives qu'offrait l'arrivée de la RMN impulsionnelle consigna cet équipement technique délicat dans une vitrine du musée du Smithsonian Institute à Washinton, DC. Depuis, la spectroscopie Hadamard a été proposée en imagerie par résonance magnétique (IRM) en utilisant des techniques de gradients (Bolinger et Leigh, 1988 ; Goelman *et al* ., 1990 ; Goelman *et al* ., 1992).

En spectroscopie RMN moderne, la création d'une véritable librairie d'impulsions sélectives a permis de reconsidérer en toute simplicité l'implémentation du codage Hadamard sur des équipements classiques. Pour chaque bande de fréquence, une impulsion monochromatique est générée par une succession de courtes impulsions, dont le nombre, l'amplitude et la phase sont déterminés par le profil d'excitation (Morris et Freeman, 1978). La fréquence effective de l'impulsion sélective est décalée de la porteuse en augmentant ou diminuant linéairement la phase de la radiofréquence au cours du temps (Kupce et Freeman, 1993a ; Kupce et Freeman, 1993b ; Kupce et Freeman, 1994). Une excitation polychromatique peut être obtenue par addition vectorielle de plusieurs séquences d'impulsions monochromatiques avec le même taux de répétition mais différentes rampes de phase (Kupce et Freeman, 1993a ; Kupce et Freeman, 1993b). Ces avancées méthodologiques ont été rendues possibles grâce aux développements technologiques réalisés sur les spectromètres en matière d'électronique et d'informatique (pour revue : Freeman, 1998).

Le principe du codage binaire utilisé actuellement consiste à manipuler les différentes bandes de fréquence à l'aide d'impulsions sélectives de façon à inverser le signe de l'aimantation de certaines bandes. Le chemin de cohérence dans le reste de la séquence sera le même pour l'ensemble des spins, à un signe près. Le signal RMN final sera donc la superposition des signaux positifs ou négatifs des différentes bandes. Le procédé de décodage décrit plus haut permettra alors de séparer les différentes bandes simplement par addition ou soustraction des différents signaux. Concrètement, l'inversion des signes de chaque bande de fréquence peut se faire par différents moyens spectroscopiques qui sont résumés dans le Tableau 3.

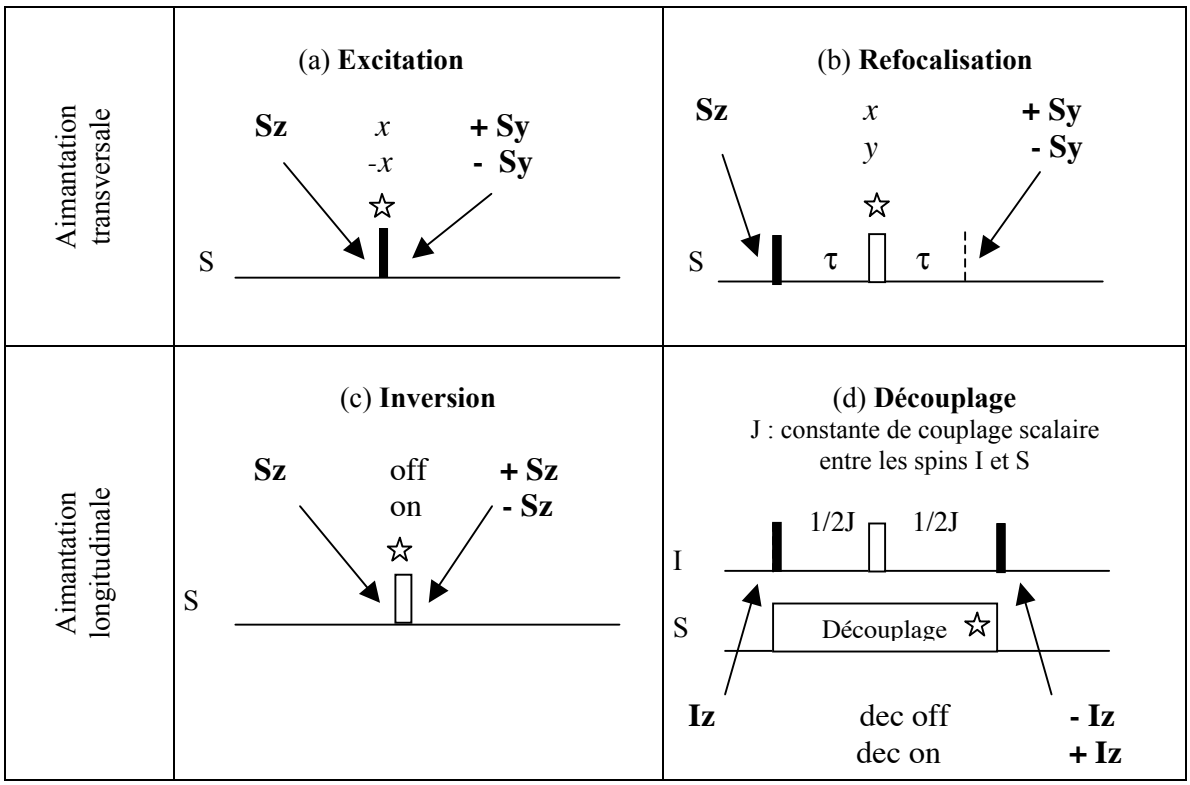


Tableau 3: Représentation des éléments de séquences qui peuvent être appliquées sur le spin S à encoder (et éventuellement sur un spin I couplé à S) pour effectuer le codage binaire +/- . Les rectangles pleins représentent des impulsions à 90°, les rectangles vides des impulsions à 180°. La phase des impulsions est *x* si elle n'est pas spécifiée en italique. Les états de spin sont indiqués en gras. L'étoile indique l'impulsion sélective sur laquelle s'effectue le codage Hadamard. Les deux options permettant d'effectuer le codage binaire +/- sont explicitées. L'appellation '*on*' signifie que l'impulsion est appliquée, '*off*' qu'aucune impulsion n'est appliquée.

Les techniques de codage binaire utilisées en spectroscopie Hadamard peuvent être divisées en deux catégories. La première concerne la manipulation sélective d'une cohérence transversale du spin à encoder. Dans cette catégorie entrent les expériences d'excitation (Tableau 3a) et de refocalisation (Tableau 3b) sélectives d'une cohérence transversale. Dans ces deux techniques, toutes les bandes de fréquence subissent le même type d'impulsion, à une phase près : le changement de phase de l'impulsion sélective inverse le signe de certaines bandes de fréquence par rapport aux autres.

On peut également agir au niveau de l'aimantation longitudinale du spin à encoder. La technique la plus simple est l'inversion sélective de l'état de spin longitudinal (Tableau 3c). La dernière technique de codage utilise l'existence d'un couplage scalaire entre un spin I et le spin S à encoder (Tableau 3d). Une aimantation transversale de I subit ou non une inversion de signe pendant un délai égal à l'inverse de la constante de couplage entre I et S selon qu'on laisse évoluer ou non ce couplage. Ce délai peut être mis à profit pour faire une édition à temps constant du spin I. En pratique, on peut utiliser un découplage sélectif continu, ou une refocalisation sélective du découplage. Ces deux méthodes seront discutées dans le contexte des filtres Hadamard présentés plus loin. Toutes les techniques de codage agissant au niveau de l'aimantation longitudinale s'appuient sur un procédé pratique où certaines des bandes subissent des impulsions, alors que d'autres ne sont pas touchées (procédé on / off).

- **Qualité des spectres**

En pratique, l'option qui est retenue parmi les méthodes proposées dans le Tableau 3 pour éditer la fréquence du spin S est celle qui donne la meilleure qualité de spectre. S'il n'est pas possible de donner un guide précis pour le choix de ces options, on peut néanmoins discuter certains critères qu'il est important de prendre en compte pour optimiser la séquence.

Concernant la sensibilité des spectres, on peut remarquer que, contrairement à la spectroscopie dans le domaine des temps, le signal enregistré ici est un signal dont seule l'intensité est intéressante. En particulier, il n'est d'aucun besoin d'enregistrer deux composantes (réelle et imaginaire) pour obtenir la phase du signal, comme c'est le cas pour la spectroscopie par TF. Dans le cas de l'acquisition d'une dimension indirecte, la spectroscopie Hadamard offre donc un gain théorique en sensibilité de $\sqrt{2}$ par rapport à la spectroscopie par TF lorsque la détection en quadrature est obtenue en effectuant la projection du signal sur deux axes orthogonaux (ce qui est le cas de la majorité des expériences utilisées actuellement, sauf pour quelques expériences qualifiées de « sensitivity enhanced »).

En pratique, la sensibilité des spectres Hadamard peut être affaiblie par les effets de relaxation qui agissent durant les délais qui sont ajoutés pour effectuer le codage Hadamard, en particulier durant les impulsions sélectives. Ces effets sont généralement d'autant plus grands que ces impulsions sont longues, i.e. que les bandes à exciter sont fines ou sélectives. Dans certains cas, des délais existant dans la séquence peuvent être néanmoins utilisés pour effectuer le codage Hadamard, ce qui réduit les pertes en sensibilité.

L'autre point critique concerne la minimisation des artéfacts dans les spectres.

Concernant l'utilisation des impulsions sélectives, la première source d'artéfacts provient de l'évolution des couplages scalaires entre deux spins pendant la durée de l'impulsion, connu sous le nom d'effet TSETSE (Ewing *et al.* , 1990 ; Nuzillard et Freeman, 1995). Cet effet est d'autant plus marqué que les impulsions sont sélectives, donc longues, et que le couplage est fort.

Le signal d'une bande est obtenu en faisant la somme ou la différence des diverses expériences enregistrées. Cela peut être utilisé pour réduire certains artéfacts systématiques provenant du spectromètre, un peu comme un cyclage de phase. Dans cette optique, la colonne où toutes les bandes de fréquence sont ajoutées doit être évitée, car les imperfections dues au spectromètre s'additionnent, alors qu'elles peuvent se compenser quand on fait des soustractions.

Il est également nécessaire, pour avoir un spectre final de bonne qualité pour une bande donnée, que les signaux des autres bandes s'annulent parfaitement. En particulier, les différentes expériences doivent posséder une grande précision quant à la position de chacun des pics, le cas contraire conduisant à l'apparition de pics artéfactuels qui ressemblent à des pics anti-phase. Ceci impose de corriger l'effet Bloch-Siegert qui intervient lors de l'irradiation de bandes proches les unes des autres. En effet, les bandes de fréquence qui sont excitées simultanément sont souvent très proches, ce qui conduit à des effets d'interférence, créant des variations de déplacement chimique et l'apparition de bandes latérales similaires à celles qu'on obtient en découplage homonucléaire (Kupce *et al.* , 1999).

De plus, l'inversion de signe des signaux doit se faire sans perte d'intensité, sinon l'annulation des pics non voulus est partielle, ce qui introduit des artéfacts sous forme de petits pics en phase aux mauvaises fréquences. Pour cela, les techniques utilisant un codage par la phase présentent l'avantage que tous les spins sont dans des états de spin similaires, à une phase près. Au contraire, dans les procédés d'encodage on / off, les propriétés de relaxation des spins dont l'aimantation longitudinale a été inversée sont différentes de celles d'un spin dont l'aimantation longitudinale est restée positive, ce qui peut créer des différences d'intensité. De plus, si l'impulsion n'est pas parfaite, une différence d'intensité va être observée entre les spins touchés par l'impulsion et ceux qui ne subissent rien. Cependant, des systèmes de purification par gradients peuvent être employés dans le cas d'impulsions sélectives selon l'axe z, ce qui n'est pas possible dans le cas des impulsions transversales.

Pour conclure, on voit que la qualité et la longueur des impulsions sélectives utilisées sont deux critères déterminants pour minimiser les problèmes de sensibilité et d'artéfacts. Le choix de l'expérimentateur devra donc se faire en considérant le type de séquence d'impulsions dans laquelle il veut implémenter le codage, la largeur des bandes de fréquence, et les propriétés de relaxation des spins à encoder, en fonction de la longueur et de la qualité des impulsions sélectives pour ces bandes dans les différentes options. Deux exemples de codage seront discutés en détail pour des cas pratiques dans la partie 'Résultats' de ce chapitre.

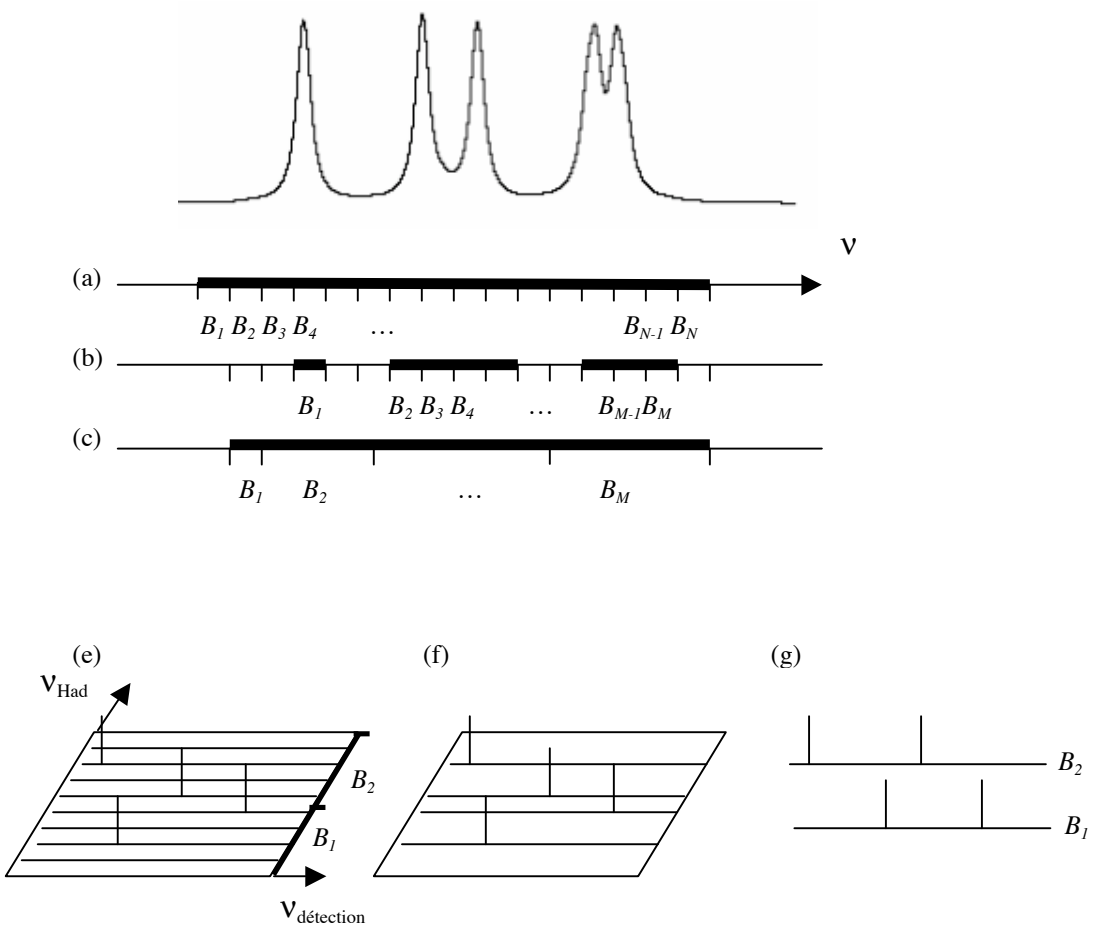


Figure 58: (a), (b) et (c) Exemples de décomposition d'un spectre RMN de largeur spectrale *SW* en bandes de fréquence pour le codage Hadamard. (a) Division de la largeur spectrale *SW* en *N* bandes dont la largeur est égale à la résolution du spectre souhaitée. (b) Uniquement *M* de ces bandes sont sélectionnées pour le codage Hadamard. Les bandes où le spectre ne comporte pas de signal sont éliminées. (c) On choisit un petit nombre *M* de bandes larges qui peuvent contenir plusieurs pics. (e), (f) et (g) : Illustration pratique du choix de bandes (a) (b) et (c), respectivement, pour l'acquisition d'une dimension indirecte ν_{Had} d'un spectre bidimensionnel contenant quatre pics de corrélations. La dimension de la détection a été enregistrée de façon classique par TF. Dans (e), l'ensemble des 8 bandes de fréquence a été échantillonné. Dans (f), uniquement les quatre bandes contenant les signaux ont été enregistrées. Après reconstruction, les spectres (f) et (e) sont identiques, et de même résolution, alors que le temps expérimental a été divisé par 2. Dans (g), un filtrage spectral a été effectué suivant les deux bandes B_1 et B_2 décrites dans le spectre (e). Dans (g), les corrélations de (e) ont donc été réparties en deux sous-spectres.

- **Choix des bandes de fréquence**

L'avantage de la spectroscopie Hadamard est d'avoir un pouvoir d'action direct dans l'espace des fréquences. Le nombre des bandes de fréquence utilisées pour l'encodage, la largeur de chacune d'elles, et leur espacement sont des paramètres qui peuvent être choisis selon l'allure du spectre de la molécule étudiée et l'application visée par l'utilisateur. Il est donc tout à fait concevable d'éviter certaines régions vides du spectre, de se focaliser sur les régions qui ont besoin de résolution, ou d'éviter la plage spectrale du solvant. Cela permet de réduire le nombre de bandes excitées, donc le temps expérimental, tout en conservant la même résolution. La Figure 58 donne quelques exemples de décomposition d'un spectre en plusieurs bandes qui vont être discutés au cas par cas.

Dans la Figure 58a, la largeur spectrale *SW* à irradier est divisée en *N* canaux de même largeur. Le nombre $N = \frac{SW}{\Delta\nu}$ impose la résolution du spectre $\Delta\nu$, et il est égal au nombre de points complexes qu'il faut enregistrer pour obtenir la même résolution avec la spectroscopie par TF (Figure 54). La Figure 58e et la Figure 59 illustrent le principe et la réalisation pratique de ce procédé dans les dimensions indirecte et directe, respectivement.

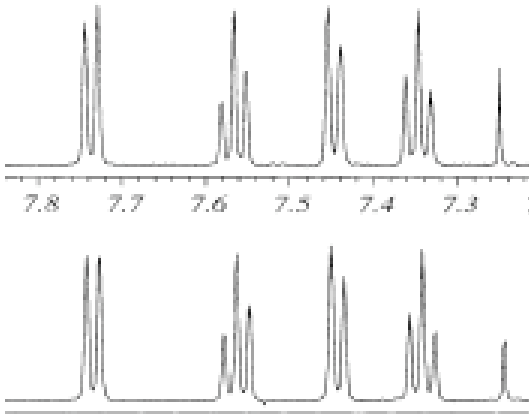


Figure 59: Comparaison des spectres ^{1}H à 500 MHz du 2-ethyl-1-indanone enregistrés avec 512 points expérimentaux (a) par spectroscopie par TF et (b) par spectroscopie Hadamard (extrait de Kupce et Freeman, 2003c). Les fréquences sont données en ppm.

Si on a une connaissance *a priori* du spectre qu'on doit obtenir, on peut s'affranchir d'acquérir les signaux de bandes où le spectre est vide. Les fréquences du solvant peuvent être également évitées. Ceci est schématisé dans la Figure 58b, où *M* bandes sont choisies parmi les *N* bandes initiales.

Le gain de temps expérimental d'un facteur $\frac{M}{N}$ qui en découle est très intéressant si on veut effectuer un spectre multidimensionnel. Après avoir fait au préalable l'acquisition 1D des différentes fréquences à éditer, on effectue un codage Hadamard uniquement aux fréquences intéressantes pour l'acquisition du spectre multidimensionnel. Par ce biais, on peut reconstruire le spectre total de l'échantillon (Figure 58f), ou choisir uniquement quelques résonances d'intérêt. Dans ce cas, la spectroscopie Hadamard est un moyen de filtrer des signaux superposés, sans avoir besoin d'éditer une dimension fréquentielle complète. Le filtrage sera naturellement fait dans le domaine des fréquences (là où la spectroscopie Hadamard peut agir).

La Figure 60 illustre expérimentalement ce procédé de filtrage sur l'exemple du spectre ^{15}N-^{1}H HSQC d'une protéine où quelques sites ont été sélectionnés, ce qui accélère l'acquisition du spectre par un facteur 110 sur cet exemple.

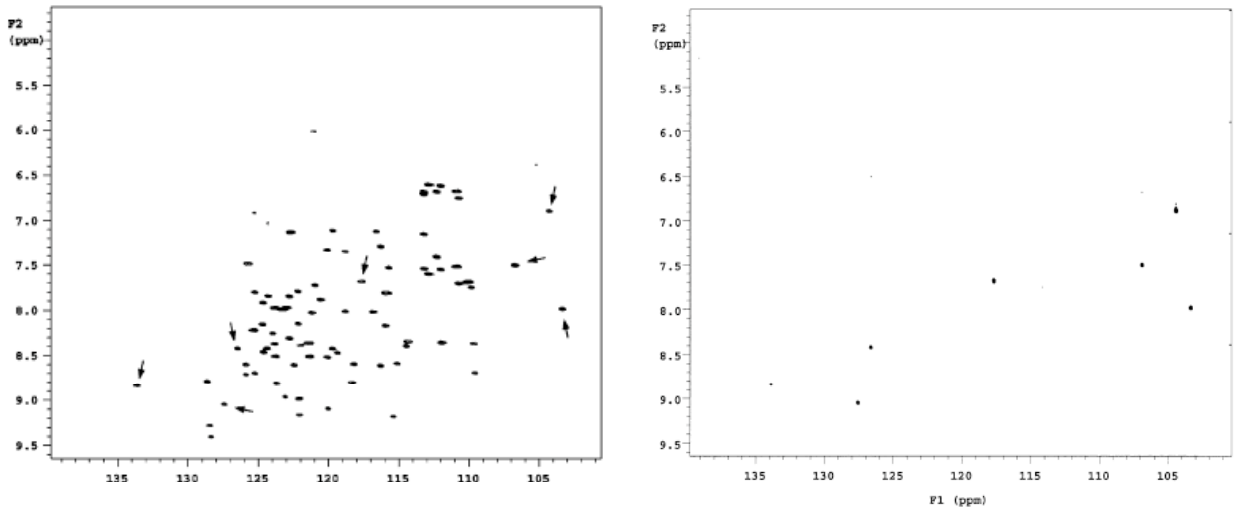


Figure 60: Spectre ^{15}N-^{1}H HSQC de l'ubiquitine humaine (800 MHz). (a) Spectre enregistré de façon conventionnelle en 24 min. avec un cyclage de phase de 4 et une détection en quadrature dans la dimension indirecte. (b) Les 7 résonances qui ont été sélectionnées par des flèches dans le spectre (a) ont été enregistrées par spectroscopie Hadamard en 13 s (ordre 8). Des impulsions gaussiennes de largeur spectrale 20 Hz ont été utilisées pour l'inversion sélective (tiré de Kupce *et al.*, 2003).

Un autre type de filtrage (que nous appellerons 'filtrage Hadamard par bandes' par la suite) peut être fait si on accepte de ne plus séparer toutes les résonances de la dimension à encoder, mais qu'on les regroupe par bandes. Un petit nombre *M* de bandes larges sont choisies dans le spectre comme décrit Figure 58c. *M* sous-spectres seront édités, chacun contenant les signaux appartenant à une des bandes (Figure 58g). La Figure 61 illustre ce principe sur l'exemple de la séparation des corrélations ^{1}H-^{15}N de la chaîne principale des protéines selon quatre bandes de fréquence ^{13}CO.

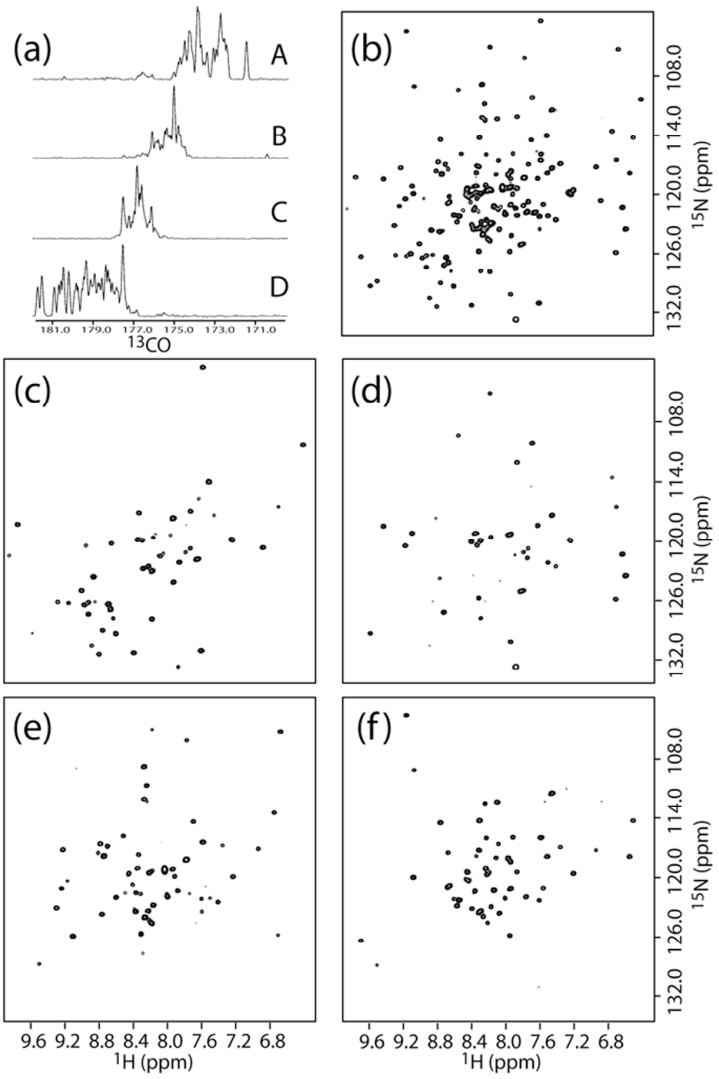


Figure 61: Filtre Hadamard d'ordre 4 agissant par bande sur la fréquence ^{13}CO, appliqué sur un échantillon de Sir-FP18 (167 résidus), 600 MHz, 30°C. (a) Spectres 1D ^{13}CO illustrant la performance du filtre pour séparer les quatre bandes. (b) Spectre de corrélation ^{1}H-^{15}N standard. (c-f) Spectres de corrélation ^{1}H-^{15}N extraits de la dimension Hadamard, correspondant aux bandes A, B, C, et D, respectivement (extrait de Brutscher, 2004).

III. 4. Bilan

Nous avons montré en théorie et en pratique que la spectroscopie par TF et la spectroscopie Hadamard sont deux techniques alternatives pour l'enregistrement des données RMN. L'implémentation pratique des deux méthodes offre des contraintes et des avantages qui peuvent être différents. La spectroscopie par TF est une technique bien implantée historiquement, qui est très efficace dans de nombreux cas. Nous allons discuter certaines situations où il peut être cependant avantageux d'utiliser la spectroscopie Hadamard plutôt que la spectroscopie par TF pour l'acquisition d'une dimension donnée.

Dans la dimension de la détection, la spectroscopie Hadamard a été utilisée en RMN stochastique (Tomlinson et Hill, 1973 ; Kaiser, 1974 ; Blümich et Ziessow, 1982), en RMN solide (Ashida *et al.* , 1990 ; Greferath *et al.* , 1993), et en RMN liquide (Krishnamurthy, 2001 ; pour revue : Kupce *et al.* , 2003). L'intérêt de la spectroscopie Hadamard dans ce contexte est de pouvoir s'affranchir des signaux du solvant et d'envoyer des excitations de faible puissance uniquement sur les régions d'intérêt, sans perte de sensibilité. Dans le contexte de l'accélération des données RMN pour l'acquisition des spectres multidimensionnels, la spectroscopie par TF est néanmoins plus efficace pour l'acquisition de la dimension de la détection, puisqu'un unique scan suffit à enregistrer l'ensemble du spectre.

Dans les dimensions indirectes, la spectroscopie Hadamard peut remplacer avantageusement la spectroscopie par TF dans certains cas où le temps expérimental pose un problème pour l'implémentation de spectres multidimensionnels. En effet, l'intérêt essentiel de la spectroscopie Hadamard par rapport à la spectroscopie par TF est que le temps expérimental nécessaire à l'acquisition d'une dimension spectrale indirecte de résolution donnée peut être diminué par un facteur $\frac{M}{2N}$, où M est le nombre de bandes de fréquence à exciter (Figure 58), et où le facteur 2 correspond à l'absence de détection en quadrature en spectroscopie Hadamard. Ce facteur peut facilement atteindre un ordre de grandeur. Ce gain peut se révéler très utile si on cherche à effectuer très rapidement une expérience 2D ou 3D. Il permet également d'envisager des expériences à dimensionalité plus élevée dans un temps raisonnable, les facteurs gagnés sur chaque dimension indirecte étant multiplicatifs.

Applications du filtrage Hadamard pour les protéines et les acides nucléiques

Cette partie décrit l'implémentation et les applications de deux filtres Hadamard que nous avons développés pour augmenter la résolution des spectres des macromolécules biologiques. Le premier concerne le filtrage des corrélations ^{1}H-^{13}C des méthyles selon le type d'acide aminé dans les protéines, et le deuxième une expérience HCN filtrée selon le type de base pour l'attribution des ARN.

I. Filtrage des corrélations ^{1}H-^{13}C des méthyles selon le type d'acide aminé

Les groupements méthyles d'une protéine sont souvent bien dispersés le long de la chaîne peptidique. Ils sont généralement situés dans le coeur de la protéine, et forment un réseau d'interactions hydrophobes qui assure une partie de la stabilité de la molécule.

De plus, même pour des protéines de taille élevée, les spectres de corrélation ^{1}H-^{13}C des méthyles ont un bon rapport signal sur bruit, ce qui s'explique par des propriétés de relaxation transversale favorables et la présence de trois protons équivalents (Figure 62).

Un avantage pratique supplémentaire est l'existence de procédés de marquage spécifique permettant de deutérer l'ensemble de la protéine sauf les groupements méthyles, ce qui permet de simplifier les spectres ^{1}H-^{13}C et de limiter les pertes de sensibilité par relaxation.

Enfin, parmi les pics de corrélation ^{1}H-^{13}C observés dans une protéine, ceux des méthyles sont relativement bien résolus, et séparés du reste des corrélations (Figure 62). L'ensemble de ces propriétés fait des groupements méthyles des sondes structurales et dynamiques de choix pour l'étude des protéines.

I. 1. Attribution des corrélations ^{1}H-^{13}C des méthyles

Les spectres de corrélation ^{1}H-^{13}C des méthyles sont très importants dans l'étape d'attribution des résonances de la protéine, mais également pour la recherche de contraintes structurales (NOE, constantes de couplage dipolaire résiduel) et l'étude de la dynamique (mesure de la relaxation ^{2}H ou ^{13}C). De plus, une cartographie des variations de déplacement chimique peut être faite sur les spectres de corrélations ^{1}H-^{13}C pour définir les contacts des chaînes latérales d'une protéine avec un partenaire (voir l'étude de BlaI en interaction avec l'ADN, partie III des résultats du premier chapitre).

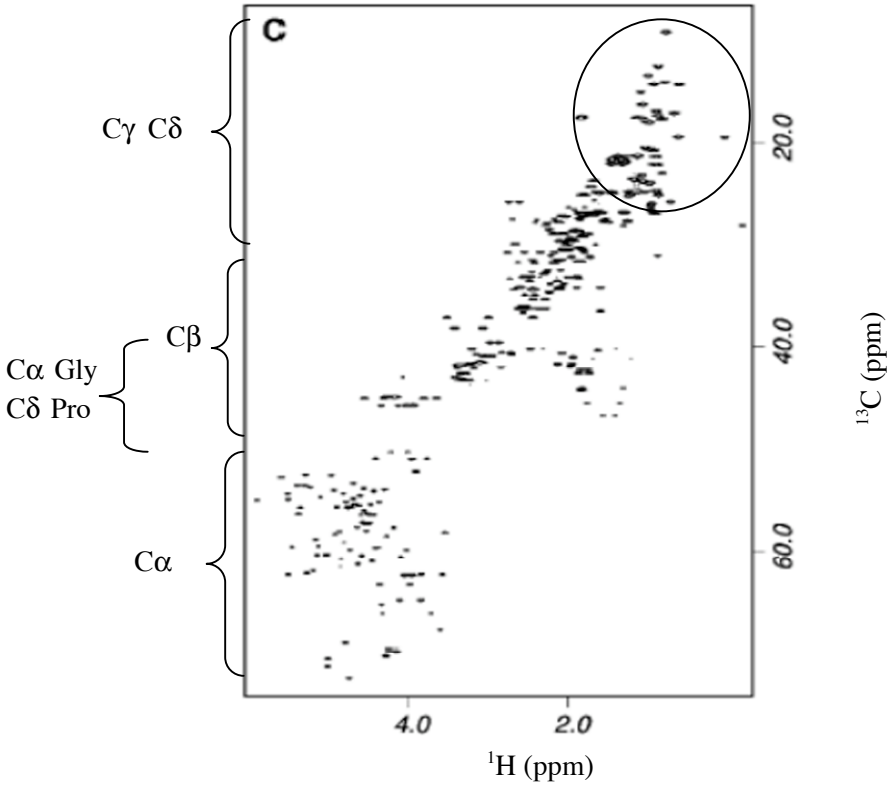


Figure 62: Spectre de corrélation ^{1}H-^{13}C typique d'une protéine. La région du spectre contenant les corrélations des groupements méthyles est entourée. A gauche sont indiquées les plages de fréquences usuelles des carbones à différentes positions le long des chaînes latérales.

La résolution des spectres de corrélation ^{1}H-^{13}C des méthyles est donc cruciale pour l'attribution non ambiguë de l'ensemble des pics de corrélation. Un premier moyen d'améliorer la résolution des spectres est d'utiliser une période d'édition à temps constant pour affiner les raies des spectres dans la dimension carbone (Bax *et al.*, 1979).

Nous proposons ici l'implémentation d'un filtre Hadamard dans la séquence d'impulsion ^{1}H-^{13}C-CT-HSQC qui profite de l'existence d'un unique carbone, que nous appellerons C^{next}, qui est lié de façon covalente au carbone C^{met} du groupement méthyle (Figure 63a). Ce carbone C^{next} est couplé au carbone C^{met} avec une constante de couplage scalaire J_{CC} dont la valeur varie très peu entre les différents acides aminés. Il est donc possible d'effectuer un transfert d'aimantation de C^{met} vers C^{next} de façon homogène pour tous les acides aminés possédant un groupement méthyle. Le principe du filtre repose sur l'existence de quatre plages de fréquences bien séparées pour le carbone C^{next}. Ces plages ont la caractéristique supplémentaire d'être spécifiques du type d'acide aminé (Figure 63b).

L'application d'un filtre Hadamard d'ordre quatre sur les bandes de fréquence du C^{next} permet donc d'effectuer un filtrage des pics de corrélation ^{1}H-^{13}C des méthyles en quatre sous-spectres selon leur type d'acide-aminé (Figure 63c).

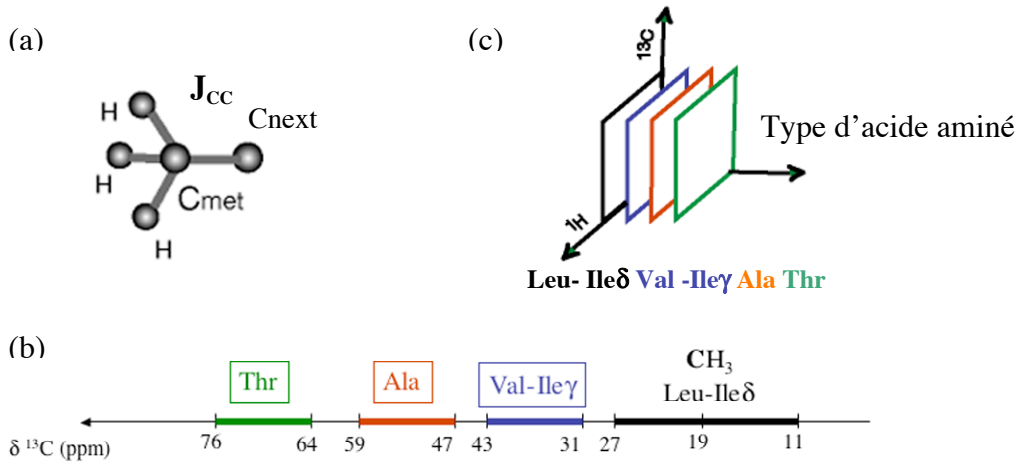


Figure 63: (a) Le carbone du groupement méthyle C^{met} est couplé avec son unique voisin C^{next} par couplage scalaire de valeur J_{CC}. (b) Représentation des quatre plages de déplacement chimique utilisées pour le codage Hadamard sur le carbone C^{next}. On peut noter que la plage contenant les leucines et les isoleucines δ est superposée avec la gamme de déplacement chimique du carbone C^{met}, et qu'elle ne sera jamais découplée. (c) Le résultat du filtre est la séparation des corrélations ^{1}H-^{13}C des méthyles selon les quatre bandes du C^{next} données en (b), ce qui revient à ajouter une dimension spectrale supplémentaire où la coordonnée discrète donne le type d'acide aminé.

L'implémentation pratique et les résultats du filtre Hadamard que nous avons développé pour l'expérience ^{1}H-^{13}C-CT-HSQC des méthyles sont présentés dans l'article qui suit (Van Melckebeke *et al.*, 2004).

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 170 (2004) 199–205
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Communication

Suppression of artifacts induced by homonuclear decoupling in amino-acid-type edited methyl ^{1}H–^{13}C correlation experiments

Hélène Van Melckebeke, Jean-Pierre Simorre, Bernhard Brutscher *

Institut de Biologie Structurale, Jean-Pierre Ebel C.N.R.S.-C.E.A.-UJF 41, rue Jules Horowitz, 38027 Grenoble Cedex, France

Received 26 May 2004; revised 24 June 2004
Available online 24 July 2004

Abstract

A detailed theoretical and experimental analysis of the artifacts induced by homonuclear band-selective decoupling during CT frequency labeling is presented. The effects are discussed in the context of an amino-acid-type editing filter implemented in ^{1}H–^{13}C CT-HSQC experiments of methyl groups in proteins. It is shown that both Bloch–Siegert shifts and modulation sidebands are efficiently suppressed by using additional off-resonance decoupling as proposed by Zhang and Gorenstein [J. Magn. Reson. 132 (1998) 81], and appropriate adjustment of a set of pulse sequence parameters. The theoretical predictions are confirmed by experiments performed on ^{13}C-labeled protein samples, yielding artifact-free amino-acid-type edited methyl spectra.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Amino-acid-type editing; Bloch–Siegert shifts; Hadamard spectroscopy; Homonuclear decoupling; Methyl; NMR; Proteins; Resolution enhancement; Sidebands

1. Introduction

Because of their hydrophobic properties methyl groups are important probes of molecular structure and dynamics inside the core of proteins or protein complexes. The favorable relaxation properties of the ^{1}H and ^{13}C spins in methyl groups, and the availability of specific isotope-labeling strategies at the methyl positions [1,2] allow application of methyl-based NMR methods to high molecular weight systems [3–5]. In the past NMR methods focused on methyl groups have been proposed for the study of molecular structure [6–8], side chain dynamics [9,10], molecular interfaces [11,12], or ligand screening for drug design [13]. Because all these experiments are based on ^{1}H–^{13}C correlation spectra, it is crucial to have methods that provide the required high spectral resolution. We have therefore designed a pulse sequence, shown in Fig. 1A that allows amino-acid-type editing in ^{1}H–^{13}C CT-HSQC experiments of methyl groups. Amino-acid-type and ^{13}C frequency editing are both performed during the constant time (CT) delay T. The additional amino-acid-type editing filter exploits the fact that the carbon directly attached to the methyl group (C^{next}) resonates in an amino-acid-type-dependent frequency range. The C^{next} spectrum can thus be divided into four distinct bands as shown in Fig. 1B. Selective decoupling of different C^{next}-bands allows switching 'on' or 'off' the homonuclear C^{next}–C^{met} scalar coupling (J_{CC}) evolution during the CT delay T. This results in a sign change of the observed NMR signal for the corresponding methyl groups that can be exploited by a binary ('plus' and 'minus') Hadamard-type frequency encoding [14,15], according to the experimental scheme of Fig. 1C, to achieve the amino-acid-type editing.

The basic element of the amino-acid-type editing filter of Fig. 1A consists in a time period $(T + \lambda^{n} t_1)/2$ of C^{met} spin evolution in the presence of C^{next}-band-selective decoupling, followed by a 180° ^{13}C pulse, and a

* Corresponding author. Fax: +33-4-76-88-54-94.
E-mail address: Bernhard.Brutscher@ibs.fr (B. Brutscher).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.06.016

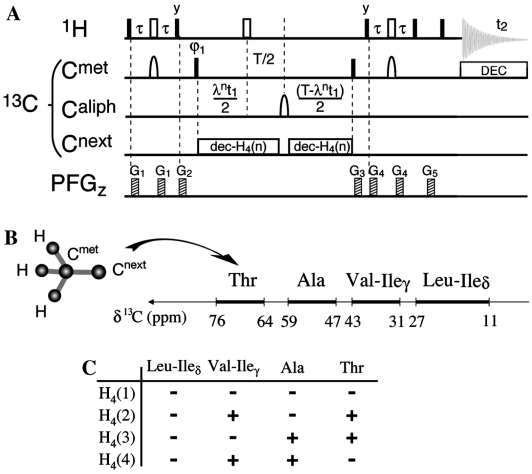


Fig. 1. Pulse sequence of the amino-acid-type edited methyl ^{1}H–^{13}C CT-HSQC. All radiofrequency (rf) pulses are applied along the x-axis unless indicated. 90° and 180° rf pulses are represented by filled and open pulse symbols, respectively. The proton carrier is set to the water resonance (4.7 ppm), and the ^{13}C carrier to the center of the methyl region (19 ppm) throughout the experiment. ^{13}C pulses drawn on the C^{aliph} or C^{next} lines are frequency shifted to 35 ppm (C^{aliph}) and the center of the corresponding C^{next} bands, respectively, by a linear phase ramp. The selective 180° C^{met} pulses are applied with an i-SNOB-5 shape [22] covering a bandwidth of 18 ppm. For application at high magnetic field strengths, an offset-compensating C^{aliph} refocusing pulse $58_x - 140_{-x} - 344_x - 140_{-x} - 58_x$ [23] is used during the CT delays T in order to properly invert all C^{met} and C^{next} carbon spins over a bandwidth of 80 ppm. The CT delay was set to $T = 28$ ms, and the INEPT transfer delays to $\tau = 1.6$ ms. ^{13}C decoupling during acquisition is achieved using a WALTZ-16 sequence [24] at a field strength of $\gamma B_1/2\pi = 3$ kHz. Pulsed field gradients, G_1–G_5 are applied along the z-axis (PFG_z) with a gradient strength of approximately 20 G/cm and lengths ranging from 100 to 500 μs, followed by a recovery delay of 100 μs. The phase cycle employed is $\varphi_1 = x, -x$, and $\varphi_{rec} = x, -x$. Quadrature detection in t_1 is obtained by incrementing the phase φ_1 according to STATES-TPPI. Constant-adiabaticity (CA) WURST-2 pulses are used as the basic elements for the C^{next}-decoupling. The CA-WURST-2 pulses have a length of $\tau_p = 5$ ms and cover a bandwidth of 12 ppm. They are centered at 37 ppm (1 ppm), 53 ppm (−15 ppm), and 70 ppm (−32 ppm) for the Val-Ile-γ, Ala, and Thr C^{next}-bands, respectively (B). The values in parentheses indicate the center of the symmetric off-resonance decoupling applied for each band using the same amplitude modulation but opposite frequency sweep [17]. Multiple-band-selective decoupling is achieved by vector addition of the individual WURST pulses [25]. The final decoupling sequence is phase cycled according to a TPG-5 supercycle [26]. The C^{next}-decoupling is restarted at the beginning of the decoupling waveform after the 180° C^{aliph} refocusing pulse. For the four-step Hadamard amino-acid-type selection filter, four experiments are recorded using C^{next}-decoupling of the bands indicated by a 'plus' in (C). A Hadamard transformation [14,15] is applied along the filter dimensions to disentangle the NMR signals from the individual frequency bands.

second time period $(T - \lambda^n t_1)/2$ of C^{met} spin evolution in the presence of C^{next}-band-selective decoupling. For the following discussion we will ignore the scaling factors ($\lambda^n = 1$). Homonuclear decoupling induces a number of undesired side effects, such a Bloch–Siegert frequency shifts and modulation sidebands [16]. Or, it is crucial for the performance of the present amino-acid-type editing filter, which relies on the addition and subtraction of different sub-spectra, that Bloch–Siegert shifts are absent from the spectrum. Otherwise anti-phase-like residual peaks will be observed due to incomplete peak cancelation. Modulation sidebands give rise to spurious artificial peaks, which may interfere with real cross-peaks, making the analysis of complex spectra difficult. We have thus undertaken a detailed theoretical and experimental analysis of the effects of multiple-band-selective C^{next} decoupling on the C^{met} spectrum, in order to devise appropriate pulse schemes to minimize or compensate these undesired peak shifts and sidebands.

If for each C^{next} band a second decoupling field is applied, symmetrically with respect to the center of the methyl spectrum with the same amplitude modulation $f(t)$ and opposite frequency sweep [17], spin evolution in the rotating frame is governed by a time-dependent first-order average Hamiltonian $H^{(1)}(t)$, as demonstrated by Zhang and Gorenstein [18]:

$$H^{(1)}(t) = \alpha f_{\mathrm{rms}}^2(t) I_z \tag{1}$$

with

$$\alpha = -2\pi \frac{v}{\Delta v_{\mathrm{rf}}^2} \tag{2}$$

and

$$f_{\mathrm{rms}}^2(t) = \frac{1}{t}\int_0^t f^2(\tau)\,\mathrm{d}\tau, \tag{3}$$

where Δv_{rf} is the difference of the average decoupling frequency and the spectral center frequency, and v is the frequency offset of a given spin from the spectral center.

The evolution of the spin density operator $\sigma(t)$ during the CT delay T in the sequence of Fig. 1A is then schematically described by:

$$\sigma(0) \overset{H^{(1)}(t)}{\rightarrow} \sigma((T+t_1)/2) \overset{\pi I_x}{\rightarrow} \overset{H^{(1)}(t)}{\rightarrow} \sigma(T), \tag{4}$$

which can be simplified to

$$\sigma(0) \overset{H^{(1)}(t)}{\rightarrow} \sigma((T+t_1)/2) \overset{-H^{(1)}(t)}{\rightarrow} \sigma(T). \tag{5}$$

We can now introduce a phase modulation function $\theta(t_1, T)$ defined as follows:

$$\sigma(T) = \exp\{-\mathrm{i}\theta(t_1,T)I_z\}\sigma(0)\exp\{+\mathrm{i}\theta(t_1,T)I_z\}. \tag{6}$$

From Eqs. (1), (5), and (6) we obtain the following expression for the phase modulation function $\theta(t_1, T)$:

$$\begin{aligned}\theta(t_1,T) &= \alpha\left[\int_0 \frac{T+t_1}{2} f^2(\tau)\,\mathrm{d}\tau - \int_0^{\frac{T-t_1}{2}} f^2(\tau)\,\mathrm{d}\tau\right] \\ &= \alpha \int_{\frac{T-t_1}{2}}^{\frac{T+t_1}{2}} f^2(\tau)\,\mathrm{d}\tau.\end{aligned} \tag{7}$$

Eq. (7) takes into account that the decoupling waveform is restarted at the beginning of the decoupling waveform after the 180° C^{aliph} refocusing pulse (see Fig. 1A).

For the amino-acid-type editing filter of Fig. 1A, a constant-adiabaticity WURST-2 pulse shape [19] is chosen as the repeated element of the decoupling sequence for its relatively sharp inversion profile and the moderate radiofrequency (rf) power required. For this special case of WURST-2 decoupling the amplitude modulation function is given by

$$f(t) = f_{\max}\sin^2\left(\pi\frac{t}{\tau_p}\right) \tag{8}$$

with τ_p the length of a single WURST pulse and $f_{\max}$ the peak amplitude of the rf field. For WURST-2 decoupling, the integral in Eq. (7) can be evaluated analytically, and we obtain the following phase modulation function

$$\theta(t_1, T) = \omega_{BS}t_1 - a\cos\left(\pi\frac{T}{\tau_p}\right)\sin\left(\pi\frac{t_1}{\tau_p}\right) + \frac{a}{8} \times \cos\left(2\pi\frac{T}{\tau_p}\right)\sin\left(2\pi\frac{t_1}{\tau_p}\right) \tag{9}$$

with the coefficients $\omega_{BS} = -\frac{3\pi}{4}\frac{f_{\max}^2}{\Delta\nu_{rf}^2}\nu$, and $\alpha = -\frac{f_{\max}^2}{\Delta\nu_{rf}^2}\nu\tau_p$.

1.1. Bloch–Siegert shift compensation

The first term in Eq. (9) gives rise to a Bloch–Siegert shift $\Delta\nu_{BS}$ with $\omega_{BS} = 2\pi\Delta\nu_{BS}$. If we introduce the root-mean square rf amplitude f_{rms} for the WURST-2 decoupling sequence defined as

$$f_{rms} = f_{rms}(\tau_p) = \sqrt{\frac{1}{\tau_p}\int_0^{\tau_p} f^2(\tau)\,d\tau} = \sqrt{\frac{3}{8}}f_{\max}, \tag{10}$$

we obtain an expression for the Bloch–Siegert shift which is well known for the case of homonuclear decoupling during real-time frequency labeling [17]

$$\Delta\nu_{BS} = -\left(\frac{f_{rms}}{\Delta\nu_{rf}}\right)^2\nu. \tag{11}$$

The two decoupling fields applied symmetrically with respect to the center of the observed spectrum with the same shape and opposite frequency sweep, induce a Bloch–Siegert shift $\Delta\nu_{BS}$ which is, within a certain frequency range, a linear function of the frequency offset ν. Therefore, Bloch–Siegert shift compensation in this linear regime is achieved by using a dilated evolution time $\lambda^n t_1$ (see Fig. 1A) with the scaling factor λ^n set to

$$\lambda^n = 1 + (f_{rms}/\Delta\nu_{rf})^2. \tag{12}$$

The use of additional symmetric off-resonance decoupling, and appropriate scaling factors λ^n compensates the Bloch–Siegert shifts, and ensures identical peak positions along the ^{13}C frequency dimension.

We have experimentally evaluated the Bloch–Siegert shifts as a function of the chemical shift offset ν and the magnetic field strength B_0. Bloch–Siegert shifts, measured for the two small proteins, *human* ubiquitin (76 residues) and *Ralstonia metallidurans* MerAa (68 residues), are plotted in Fig. 2 as a function of the C^{met} chemical shift. CT-HSQC spectra were recorded at Varian INOVA 800 and INOVA 600 spectrometers using the pulse sequence of Fig. 1A. The Bloch–Siegert shifts were extracted from the relative cross-peak positions along the ^{13}C dimension in the spectra acquired with and without band-selective C^{next}-decoupling, and the scaling factors set to $\lambda^n = 1$ for all experiments. For the spectra recorded with simultaneous Ala and Thr C^{next}-decoupling the observed offset dependence is well described by Eq. (11), and the scaling factors calculated from Eq. (12) thus provide a good estimate of the required correction. For experiments with C^{next}-decoupling of the (Val-Ile$_\gamma$) band, however, a deviation from linearity is observed (Figs. 2A and C). This band is closest to the observed methyl ^{13}C, and thus responsible for most of the spectral perturbation. For optimal filter performance over the methyl ^{13}C bandwidth, it is therefore necessary to experimentally optimize the scaling factors λ^n, taking into account the non-linear behavior of the observed Bloch–Siegert shifts. Because the Bloch–Siegert shifts depend on the B_0-field strength, a different set of scaling factors λ^n has to be used for different magnetic fields. For the application presented here, we have empirically optimized the following scaling factors: $\lambda^2_{600} = 1.0046$, $\lambda^3_{600} = 1.0181$, and $\lambda^4_{600} = 1.0162$ at 600 MHz, and $\lambda^2_{800} = 1.0028$, $\lambda^3_{800} = 1.0102$, and $\lambda^4_{800} = 1.0091$ at 800 MHz, which yield residual Bloch–Siegert shifts of less than 1.0 Hz within a spectral width of about 15 ppm (Figs. 2B and D). Outside this spectral range the frequency shift increases very quickly, and as a consequence the filter performance decreases. The ^{13}C frequency range of 11.5–26.5 ppm, however, covers the large majority of methyl resonances (≈90% of the methyls in the BMRB). We conclude from these results that a similar Bloch–Siegert shift-compensation performance is obtained for experiments conducted at currently available high magnetic field strengths (400–900 MHz ^1H frequency), using the pulse sequence of Fig. 1A and optimized scaling factors λ^n.

1.2. Modulation sideband suppression

The periodic amplitude-modulated decoupling field also induces cyclic irradiation sidebands, commonly referred to as *modulation sidebands* [16]. These sidebands are described by the second and third terms of Eq. (9), which correspond to first- and second-order harmonic phase modulations in the rotating frame. Similar to the Bloch–Siegert shifts, a linear offset dependence is obtained. This leads to sideband suppression for all C^{met}

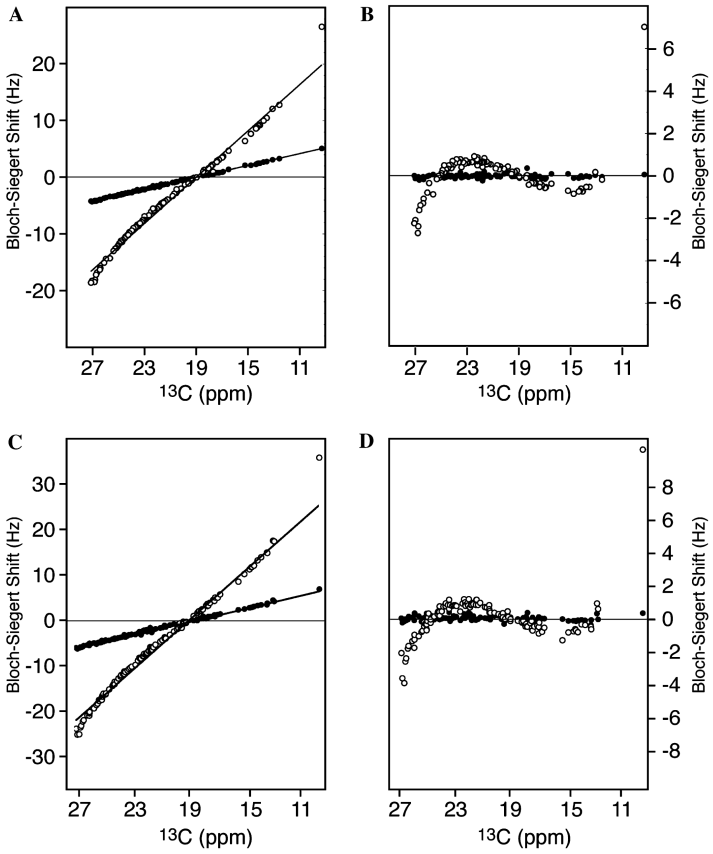


Fig. 2. Bloch–Siegert shifts measured at 800 MHz (A) and at 600 MHz (C) ^{1}H frequency with the pulse sequence of Fig. 1A setting $\lambda^n = 1$. Open circles correspond to measurements with simultaneous C^{next}-decoupling of the Ala and (Val-Ile-γ) bands, and filled circles represent the corresponding results for Ala and Thr C^{next}-decoupling. Residual Bloch–Siegert shifts using optimized scaling factors λ^n are plotted in (B) for 800 MHz ($\lambda^1_{800} = 1.0$, $\lambda^2_{800} = 1.0026$, $\lambda^3_{800} = 1.0102$, $\lambda^4_{800} = 1.0091$), and in (D) for 600 MHz ($\lambda^1_{600} = 1.0$, $\lambda^2_{600} = 1.0046$, $\lambda^3_{600} = 1.0181$, $\lambda^4_{600} = 1.0162$).

resonances close to the spectral center. Far from the center, first- and second-order sidebands are predicted with an eight-times higher intensity for the first-order sidebands. In the following, we will focus on the larger first-order effects, which are observable at frequencies $\nu_{\pm 1} = \nu_0 \pm 1/(2\tau_p)$, with ν_0 the central peak frequency and τ_p the periodicity of the decoupling sequence. Interestingly, a periodic modulation of the sideband amplitudes $A_{\pm 1}$ with the CT delay T is predicted from Eq. (9), $A_{\pm 1} \propto \pm\cos(\pi T/\tau_p)$, yielding sideband suppression by adjusting the constant time delay T and the decoupling pulse length τ_p in such a way that

$$\frac{T}{\tau_p} = \frac{(2m-1)}{2} \tag{13}$$

with m an integer.

To confirm these theoretical results we have recorded a series of methyl CT ^{1}H–^{13}C correlation spectra of *human* ubiquitin at an INOVA 800 spectrometer using the pulse sequence of Fig. 1A with simultaneous C^{next}-decoupling of the Ala and (Val-Ile$_\gamma$) bands, as well as symmetric off-resonance decoupling. Spectra were recorded for two different WURST-2 pulse lengths of $\tau_p = 3$ ms and $\tau_p = 5$ ms, and a series of CT delays T in the range $26\,\text{ms} \leqslant T \leqslant 33\,\text{ms}$. A single cross-peak extracted at a C^{met} frequency of 26.9 ppm is shown in Figs. 3A and B to highlight the position and intensity of the detected modulation sidebands. Only first-order sidebands are observed in the spectra of Fig. 3 located 167 (Fig. 3A) and 100 Hz (Fig. 3B) away from the central peak, which exactly corresponds to the predicted frequencies of $\nu_{\pm 1} = \nu_0 \pm 1/(2\tau_p)$. The observed asymmetry of the sideband intensity is explained by the fact that the harmonic phase modulation in the rotating frame also yields a small symmetric contribution to the sideband intensity, which adds to one sideband and subtracts from the other one [18]. The intensity modulation as a function of the CT delay T is confirmed by the experimental results shown in Fig. 3. As expected from our theoretical treatment, the sideband intensities are modulated with

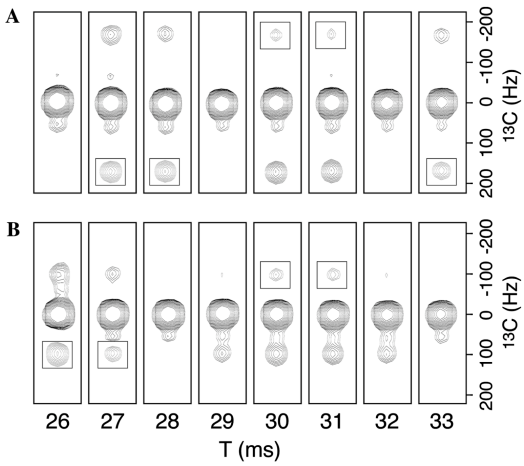


Fig. 3. Small spectral region extracted from a CT-HSQC methyl ^{1}H–^{13}C correlation spectra of ubiquitin recorded at 800 MHz ^{1}H frequency using the pulse sequence of Fig. 1A. C^{next}-decoupling of the Ala and (Val-Ileγ) bands is applied during the CT time delay T, which is varied from 26 to 33 ms. The experiments were repeated for a WURST pulse length of (A) $\tau_p = 3$ ms, and (B) $\tau_p = 5$ ms. A single methyl peak is observed in the plotted spectral region extracted at $\omega_C = 26.9$ ppm. A box indicates a peak with negative intensity.

a periodicity of τ_p. Therefore first-order sideband suppression is achieved over the whole C^{met} spectral range by properly adjusting the constant time delay T and the decoupling pulse length τ_p. If we consider the additional experimental conditions $\tau_p < 1/5J_{CC}$ and $T \approx 1/J_{CC}$ ($J_{CC} \approx 35$ Hz), the optimal parameter combination is found to be $T = 28$ ms and $\tau_p = 5$ ms for the amino-acid-type editing filter of Fig. 1A. Note that the experimentally optimized ratio $T/\tau_p = 5.6$ slightly deviates from the theoretical value $T/\tau_p = 5.5$ predicted by Eq. (13).

Other experimental schemes for the suppression of sidebands induced by homonuclear decoupling have been presented before [18,20]. The common idea of these approaches is to average out the modulation of the effective magnetic field in the rotating frame by shifting the origin of the decoupling waveform for different scans. This contrasts with the method presented here, where sideband suppression is achieved in a 'single scan' by exploiting the inherent properties of homonuclear decoupling during constant time spin evolution. This result is not limited to the particular case of WURST-2-based decoupling sequences. For the general case of an amplitude-modulated waveform, which can be decomposed into a Fourier series, one obtains after integration in Eq. (7) terms of the form

$$k_1^n\left\{k_2^n \sin\left(\frac{n}{2}\pi\frac{T}{\tau_p}\right) + k_3^n \cos\left(\frac{n}{2}\pi\frac{T}{\tau_p}\right)\right\} \sin\left(\frac{n}{2}\pi\frac{t_1}{\tau_p}\right)$$

in the phase modulation function $\theta(t_1, T)$, with n an integer, and k_1^n, k_2^n, and k_3^n constants depending on the chosen waveform. It is thus still possible to suppress the dominant term yielding the highest sideband intensity by adjusting the ratio T/τ_p.

1.3. Experimental application of methyl amino-acid-type editing

For amino-acid-type editing in the pulse sequence of Fig. 1A, four experiments $H_4(1)$, $H_4(2)$, $H_4(3)$, and $H_4(4)$ are performed with simultaneous decoupling of different C^{next} bands (Fig. 1B) indicated by a 'plus' in the scheme of Fig. 1C. In addition, 'off'-resonance decoupling is applied for each C^{next} band symmetrically to the C^{met} spectral center. The use of different B_0-field-dependent scaling factors λ^n for the four experiments ensures identical peak positions along the ^{13}C dimension. Note that the correction factors λ^n are additive for the different C^{next} bands, and that they have been optimized experimentally as explained above. A Hadamard transformation [14], instead of a Fourier transformation, then allows disentangling the four bands along the additional amino-acid-type dimension. This yields a ($2_{cs} + 1_{aa}$)-D methyl correlation spectrum, with two chemical shift and one amino-acid-type dimensions.

An example of an amino-acid-type edited methyl ^{1}H–^{13}C correlation spectrum, recorded using the pulse sequence of Fig. 1A, is shown in Fig. 4. The data were acquired at 800 MHz ^{1}H frequency on a sample of uniformly ^{13}C-, ^{15}N-labeled calmodulin, a 163-residue protein with 90 methyl groups that has been extensively studied by NMR in the past. The four 2D ^{1}H–^{13}C spectra, plotted in Figs. 4A–D, were extracted along the amino-acid-type dimension of the ($2_{cs} + 1_{aa}$)-D methyl correlation spectrum. Overall, the spectra are artifact-free, and the different C^{next} frequency bands are well separated in the four spectra. Some cross-talk is observed between the Leu and Val peaks, for which the actual C^{next} frequency ranges slightly overlap. Therefore some Val peaks (Fig. 4C) are also detected in the spectrum of Fig. 4A, although with a significantly lower intensity. The methyl group at C$^{\varepsilon}$ position of methionine side chains, which has no directly attached carbon, gives rise to a negative cross-peak in the subspectrum shown in Fig. 4A. The amino-acid-type editing filter enhances spectral resolution, and allows additional amino-acid-type identification for the purpose of resonance assignment. Many NMR experiments, based on the spectral resolution of ^{1}H–^{13}C methyl correlation spectra, may be improved by the use of additional amino-acid-type editing. Recently, we have demonstrated that amino-acid-type editing in methyl NOESY experiments helps for unambiguous NOE peak assignment, and allows fast acquisition of methyl NOE data [21]. Other potential applications

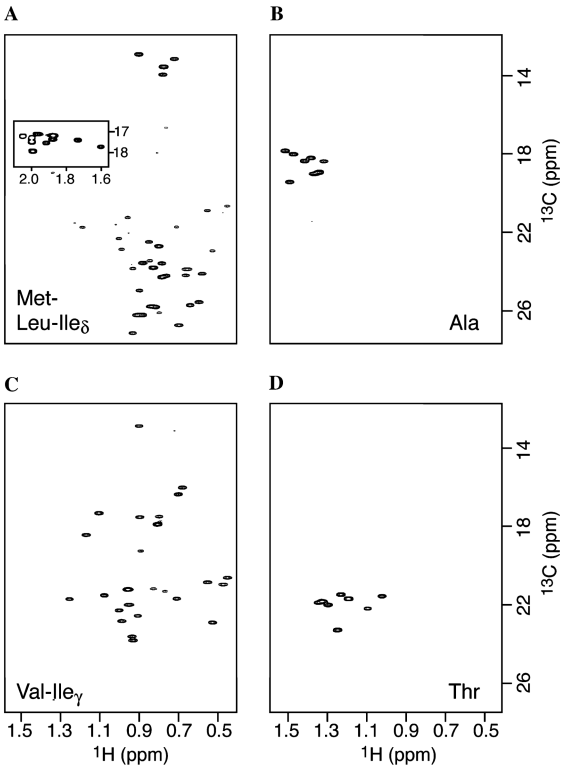


Fig. 4. Amino-acid-type edited methyl CT-HSQC spectrum recorded at 800 MHz ^{1}H frequency using the pulse sequence of Fig. 1A on a sample of ^{13}C-labeled calmodulin (1 mM, 27 °C, pH 6.5). $4(H_4) \times 110(^{13}C) \times 512(^{1}H)$ complex points were recorded for spectral widths of 4000 Hz (^{13}C) and 9000 Hz (^{1}H) in an overall experimental time of 2 h. The four ^{1}H–^{13}C spectra are extracted along the amino-acid-type dimension. They show methyl cross-peaks from (A) Met-Leu-Ile$_{\delta}$, (B) Ala, (C) Val-Ile$_{\gamma}$, and (D) Thr residues. The methionine cross-peaks in the inset of (A) have negative intensity.

of this filter include the study of side chain dynamics by ^{13}C or ^{2}H spin relaxation measurements, the quantification of spin–spin couplings, and the study of molecular interfaces by chemical shift mapping.

In summary, we have shown that both, Bloch–Siegert shifts and cyclic irradiation sidebands induced by symmetric band-selective decoupling during CT frequency labeling can be suppressed by the use of a dilated evolution time λt_1, and proper adjustment of the T/τ_p ratio. This has allowed us to design a filter sequence based on multiple-band-selective homonuclear decoupling, which yields clean amino-acid-type editing in methyl ^{1}H–^{13}C correlation experiments of proteins. Other NMR experiments may benefit as well from artifact-free band-selective homonuclear decoupling during CT frequency labeling. Examples include base-type editing in HCN- or HCC-type experiments of nucleic acids, and C^{β}-decoupling during C^{α} frequency labeling in HNCACO-type experiments of proteins. The implementation of such experimental schemes is currently under investigation in our laboratory.

Acknowledgments

This work was supported by the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique. The authors thank B. Bersch, J. Covès, D. Bouvier, and P. Gans for the preparation of the labeled protein samples. H.V.M. acknowledges the receipt of a fellowship from the C.E.A.

References

[1] K.H. Gardner, L.E. Kay, The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins, Annu. Rev. Biophys. Biomol. Struct. 27 (1998) 357–406.

[2] N.K. Goto, K.H. Gardner, G.A. Mueller, R.C. Willis, L.E. Kay, A robust and cost-effective method for the production of Val, Leu, Ile(δ1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins, J. Biomol. NMR 13 (1999) 369–374.

[3] K.H. Gardner, M.K. Rosen, L.E. Kay, Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR studies, Biochemistry 36 (1997) 1389–1401.

[4] V. Tugarinov, L.E. Kay, Ile, Leu, and Val methyl assignment of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods, J. Am. Chem. Soc. 125 (2003) 13868–13878.

[5] V. Tugarinov, P.M. Hwang, J.E. Ollerenshaw, L.E. Kay, Cross-correlated relaxation enhanced ^{1}H–^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes, J. Am. Chem. Soc. 125 (2003) 10420–10428.

[6] M.K. Rosen, K.H. Gardner, R.C. Willis, W.E. Parris, T. Pawson, L.E. Kay, Selective methyl group protonation of perdeuterated proteins, J. Mol. Biol. 263 (1996) 627–636.

[7] W.J. Metzler, M. Wittekind, V. Goldfarb, L. Mueller, B.T. Farmer, Incorporation of ^{1}H-^{13}C-^{15}N-{Ile, Leu, Val} into a perdeuterated, ^{15}N-labeled protein: potential in structure determination of large proteins by NMR, J. Am. Chem. Soc. 118 (1996) 6800–6901.

[8] G.A. Mueller, W.Y. Choy, D. Yang, J.D. Forman-Kay, R.A. Venters, L.E. Kay, Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein, J. Mol. Biol. 300 (2000) 197–212.

[9] R. Ishima, J.M. Louis, D.A. Torchia, Transverse ^{13}C relaxation of CHD_2 methyl isotopmers to detect slow conformational changes of protein side chains, J. Am. Chem. Soc. 121 (1999) 11589–11590.

[10] O. Millet, D.R. Muhandiram, N.R. Skrynnikov, L.E. Kay, Deuterium spin probes of side-chain dynamics in proteins. 1. measurement of five relaxation rates per deuteron in ^{13}C-labeled and fractionally ^{2}H-enriched proteins in solution, J. Am. Chem. Soc. 124 (2002) 6439–6448.

[11] E.F. DeRose, T. Darden, S. Harvey, S. Gabel, F.W. Perrino, R.M. Schaaper, R.E. London, Elucidation of the epsilon–theta subunit interface of *Escherichia coli* DNA polymerase III by NMR spectroscopy, Biochemistry 42 (2003) 3635–3644.

[12] W.P. Shao, S.C. Im, E.R.P. Zuiderweg, L. Waskell, Mapping the binding interface of the cytochrome *b*(5)–cytochrome *c* complex by nuclear magnetic resonance, Biochemistry 42 (2003) 14774–14784.

[13] P.J. Hajduk, D.J. Augeri, J. Mack, R. Mendoza, J. Yang, S.F. Betz, S.W. Fesik, NMR-based screening of proteins containing ^{13}C-labeled methyl groups, J. Am. Chem. Soc. 122 (2000) 7898–7904.

[14] J. Hadamard, Résolution d'une question relative aux déterminants, Bull. Sci. Math. 17 (1893) 240–248.

[15] E. Kupce, T. Nishida, R. Freeman, Hadamard NMR spectroscopy, Prog. NMR Spectrosc. 42 (2003) 95–122.

[16] E. Kupce, H. Matsuo, G. Wagner, Homonuclear decoupling in proteins, Biol. Magn. Reson. 16 (1999) 149–193.

[17] S. Zhang, D.G. Gorenstein, Bloch-Siegert shift compensated and cyclic irradiation sidebands eliminated, double-adiabatic homonuclear decoupling for ^{13}C- and ^{15}N-double-labeled proteins, J. Magn. Reson. 132 (1998) 81–87.

[18] S. Zhang, D.G. Gorenstein, "BEST" homonuclear adiabatic decoupling for ^{13}C- and ^{15}N-double-labeled proteins, J. Magn. Reson. 138 (1999) 281–287.

[19] E. Kupce, R. Freeman, Optimized adiabatic pulses for wideband spin inversion, J. Magn. Reson. A 118 (1996) 299–303.

[20] J. Weigelt, A. Hammarström, W. Bermel, G. Otting, Removal of zero-quantum coherence in protein spectra using SESAM decoupling and suppression of decoupling sidebands, J. Magn. Reson. B 110 (1996) 219–224.

[21] H. Van Melckebeke, J.-P. Simorre, B. Brutscher, Amino acid-type edited NMR experiments for methyl–methyl distance measurements in ^{13}C-labeled proteins, J. Am. Chem. Soc. (2004) (in press).

[22] E. Kupce, J. Boyd, I.D. Campbell, Short selective pulses for biochemical applications, J. Magn. Reson. B 106 (1995) 300–303.

[23] A.J. Shaka, A.J. Pines, Symmetric phase-alternating composite pulses, Magn. Reson. 71 (1987) 495–503.

[24] A.J. Shaka, J. Keeler, T. Frenkiel, R. Freeman, An improved sequence for broadband decoupling: Waltz-16, J. Magn. Reson. 52 (1983) 335–338.

[25] E. Kupce, R. Freeman, Techniques for multisite excitation, J. Magn. Reson. A 105 (1993) 234–238.

[26] R. Tycko, A. Pines, R. Gluckenheimer, Fixed point theory of iterative excitation schemes in NMR, J. Chem. Phys. 83 (1985) 2775–2802.

I. 2. Application à la recherche de contraintes structurales de type NOE

La mesure des NOE méthyles-méthyles est une source importante d'information structurale qui peut être utilisée avec succès pour obtenir le repliement d'une protéine lors de la détermination de sa structure, ou valider un repliement proposé par homologie (Figure 64).

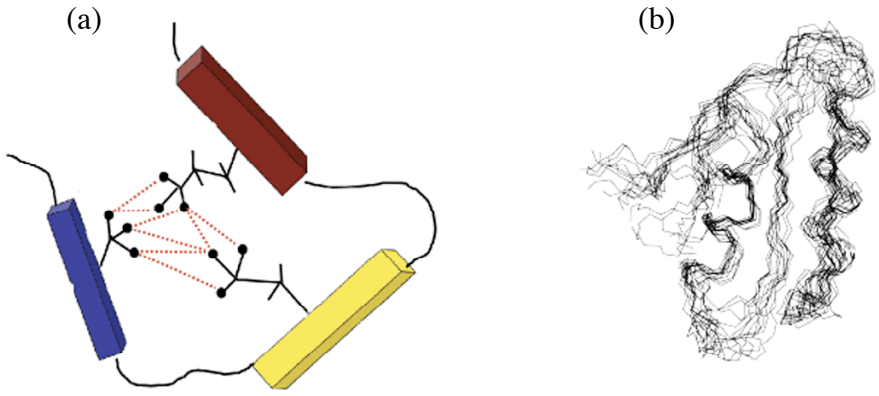


Figure 64: (a) Schéma illustrant comment la donnée d'un jeu restreint de contraintes structurales à longue distance entre groupements méthyles peut permettre de trouver le repliement de la protéine. (b) Ensemble RMN (15 structures) décrivant le repliement de la protéine MerAa, calculé avec les contraintes suivantes : 56 NOE méthyle-méthyle, 62 angles dièdres et 47 liaisons hydrogènes (éléments de structures secondaires et appariement des brins β anti-parallèles) (extrait de Bersch *et al.*, 2003).

Pour obtenir ces contraintes, des expériences de type 3D-NOESY-HSQC ou 3D-HSQC-NOESY-HSQC avec édition des fréquences des carbones en temps réel ou à temps constant ont été proposées (Clore *et al.*, 1991; Zwahlen *et al.*, 1998). Le filtre qui sépare les corrélations ^{1}H-^{13}C des méthyles peut être implémenté avantageusement dans les expériences 3D-HSQC-NOESY-HSQC pour augmenter la résolution des spectres sans augmenter de façon significative le temps de l'expérience. La Figure 65 résume les options spectrales que nous proposons pour caractériser les deux groupements méthyles dont les protons sont en interaction NOE.

La description de l'implémentation du filtre et des différentes options proposées est faite dans l'article donné ci-après (Van Melckebeke *et al*., 2004). L'application des séquences NOESY à des protéines de taille moyenne sera discutée dans le contexte de la détermination rapide du repliement des protéines. Un jeu d'options concernant l'étude de protéines de grosse taille sera illustré dans le contexte de l'augmentation de la résolution des spectres pour l'étude de complexes moléculaires de plus en plus gros par RMN.

Méthyle 1

- Edition du carbone à temps constant
- Filtre Hadamard type acide aminé

Méthyle 2

- Edition du proton (détection)
- Edition du carbone *en temps réel ou à temps constant*
- *Filtre Hadamard type acide aminé*

Figure 65: Liste des dimensions spectrales (continues ou dicrètes) qui peuvent être éditées pour chacun des deux groupements méthyles dont les protons sont en interaction NOE. Les dimensions données en italique sont optionnelles. Le choix des options se fera en fonction du compromis entre sensibilité, résolution, et temps expérimental qui est souhaitable pour l'échantillon observé.

J|A|C|S
ARTICLES
Published on Web 07/17/2004

Amino Acid-Type Edited NMR Experiments for Methyl−Methyl Distance Measurement in ^{13}C-Labeled Proteins

Hélène Van Melckebeke, Jean-Pierre Simorre, and Bernhard Brutscher*

Contribution from the Institut de Biologie Structurale, Jean-Pierre Ebel CNRS-CEA-UJF, 41, rue Jules Horowitz, 38027 Grenoble Cedex, France

Received February 24, 2004; E-mail: Bernhard.Brutscher@ibs.fr

Abstract: New NMR experiments are presented for the measurement of methyl−methyl distances in ^{13}C-labeled proteins from a series of amino acid-type separated 2D or 3D NOESY spectra. Hadamard amino acid-type encoding of the proximal methyl groups provides the high spectral resolution required for unambiguous methyl−methyl NOE assignment, which is particularly important for fast global fold determination of proteins. The experiments can be applied to a wide range of protein systems, as exemplified for two small proteins, ubiquitin and MerAa, and the 30 kDa BRP−Blm complex.

Introduction

Liquid-state NMR spectroscopy has been established as a powerful method for structure determination of proteins. The development of many multidimensional triple-resonance NMR experiments over the past decade, combined with advanced isotope-labeling strategies, has significantly enlarged the scope of NMR problems that can now be studied by NMR. On one hand, TROSY-type experiments[1] have increased the size of molecules amenable to an NMR investigation. Recently, near-complete backbone assignments have been reported for the 723-residue malate synthase G protein from *Escherichia coli*,[2] the 110 kDa homo-octameric protein 7,8-dihydroneolpterin aldolase,[3] and the membrane proteins OmpX and OmpA.[4] On the other hand, the development of new acquisition schemes has sped up NMR data collection.[5] Such techniques are especially promising in the context of fast NMR fold determination of moderately sized proteins (<200 residues). Recently, we have obtained complete backbone and methyl side-chain assignment of two proteins, the 68-residue protein fragment MerAa from *Ralstonia metallidurans* and the 167-residue protein fragment SiR−FP18 from *E. coli*, from a set of two-dimensional NMR spectra acquired in an overall experimental time of less than 48 h.[6,7] With these experimental advances, the most limiting and time-consuming step of the NMR structure determination process remains the measurement and unambiguous assignment of NOE cross-peaks to specific, proximal pairs of protons.

Methyl groups are extremely valuable probes of molecular structure and dynamics. Methyl-containing residues are generally well-dispersed throughout the primary sequence, and the methyl groups are preferably located in the hydrophobic protein core. Therefore, methyl−methyl distances derived from NOE measurements provide a rich source of long-range structural information.[8] For NMR studies, methyls are advantageous because they give rise to reasonably well-resolved ^{1}H−^{13}C correlation spectra with good sensitivity even for larger molecules.[9] The high sensitivity of methyl correlation experiments, as compared to that of CH or CH_2 groups, is explained by favorable ^{1}H and ^{13}C transverse relaxation due to the fast methyl rotation and the presence of three equivalent protons attached to the methyl carbon. In addition, robust labeling strategies developed for obtaining uniformly deuterated proteins with protonated methyl groups[10] allow application of methyl-based NMR methods to high molecular weight systems. Methyl resonance assignments have been obtained for molecular systems with a tumbling correlation time of up to 20 ns[11] using standard ^{13}C-TOCSY-based magnetization transfer experiments. More recently, Kay and co-workers have achieved near-complete methyl assignments for Ile, Val, and Leu side chains of the 723-

(1) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371.
(2) Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 10025−10035.
(3) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *J. Am. Chem. Soc.* **2000**, *122*, 7543−7548.
(4) (a) Fernandez, C.; Hilty, C.; Bonjour, S.; Fernandez, C.; Pervushin, K.; Wüthrich, K. *FEBS Lett.* **2001**, *504*, 173−178. (b) Fernandez, C.; Fernandez, C.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2358−2363. (c) Arora, A.; Abilgaard, F.; Bushweller, J. H.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8*, 334−338.
(5) (a) Freeman, R.; Kupce, E. *J. Biomol. NMR* **2003**, *27*, 101−113. (b) Kupce, E.; Nishida, T.; Freeman, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *42*, 95−122.
(6) (a) Bersch, B.; Rossy, E.; Covès, J.; Brutscher, B. *J. Biomol. NMR* **2003**, *27*, 57−67. (b) Brutscher, B. *J. Biomol. NMR* **2004**, *29*, 57−64.
(7) Brutscher, B. *J. Magn. Reson.* **2004**, *167*, 178−184.
(8) (a) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 627−636. (b) Metzler, W. J.; Wittekind, M.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *J. Am. Chem. Soc.* **1996**, *118*, 6800−6901. (c) Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197−212.
(9) Gardner, K. H.; Rosen, M. K.; Kay, L. E. *Biochemistry* **1997**, *36*, 1389−1401.
(10) (a) Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357−406. (b) Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E. *J. Biomol. NMR* **1999**, *13*, 369−374.
(11) (a) Gardner, K. H.; Zhang, X.; Gehring, K.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 11738−11748. (b) Hilty, C.; Fernandez, C.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2002**, *23*, 289−301.

10.1021/ja0489644 CCC: $27.50 © 2004 American Chemical Society

residue malate synthase G protein using a new isotopic labeling strategy in combination with a set of COSY-type transfer experiments.[12] ^{13}C-edited NOESY experiments using constant time (CT) ^{13}C frequency labeling have been proposed in the past[13] to record high-resolution methyl−methyl correlation spectra. Such experiments have been successfully applied to fully or methyl-only protonated protein samples in the molecular weight range of up to 40 kDa. The main limitations of the experiment of Zwahlen et al.[13] are that it does not allow fast data collection and that the spectral resolution may still not be sufficient for unambiguous NOE assignment in the case of larger molecules. As demonstrated here, these limitations may be overcome by the addition of amino acid-type editing filters to the original sequence, which spreads the methyl−methyl correlation peaks along additional dimensions according to the amino acid type of the methyl-containing residue.

In this article, we propose new ^{13}C-edited HSQC−NOESY−HSQC-type experiments, where one amino acid-type editing filter is added to the standard ^{1}H and ^{13}C frequency dimensions for each of the two methyls involved in the NOE interaction. This yields ($n_{cs} + n_{aa}$)-dimensional NOESY spectra, with n_{cs} and n_{aa} the number of chemical shift and amino acid-type dimensions, respectively. These amino acid-type edited methyl NOESY experiments are equally useful for application to small and large molecular systems. For smaller molecules, ($2_{cs} + 2_{aa}$)-D methyl NOESY spectra are recorded in a short acquisition time, typically a few hours on a modern high field NMR spectrometer. In these spectra, one of the proximal methyl groups is labeled by its ^{13}C frequency and amino acid type, whereas the other one is identified by its ^{1}H frequency and amino acid type. The resulting ($2_{cs} + 2_{aa}$)-D NOESY correlation maps provide a set of mostly unambiguous methyl−methyl distance restraints, which is important for the validation of a structural homology model or for de novo protein fold determination. The amino acid-type selective methyl NOESY experiment is demonstrated on ^{13}C, ^{15}N-labeled samples of two small proteins: (i) MerAa, a protein fragment consisting of the 68 N-terminal residues of the cytosolic mercuric reductase MerA from *R. metallidurans* and (ii) the 76-residue *human* ubiquitin. For the two proteins, the set of intermethyl distance restraints extracted from the ($2_{cs} + 2_{aa}$)-D NOESY spectra, recorded in a short experimental time, proved sufficient for molecular fold determination. For larger molecular systems, acquisition of ($3_{cs} + 1_{aa}$)-D or ($3_{cs} + 2_{aa}$)-D correlation maps provides the necessary spectral resolution for methyl−methyl NOE assignment and an attractive spectroscopic alternative to amino acid-type selective isotope labeling approaches. Such spectra can still be recorded in a reasonable experimental time of a few days, thus presenting an attractive way of collecting long-range distance restraints of larger protein systems with assigned methyl resonances. A ($3_{cs} + 1_{aa}$)-D methyl NOESY spectrum, acquired on a ^{13}C, ^{15}N-labeled sample of the 30 kDa bleomycine-resistance protein−bleomycine (Zn^{2+}) complex (BRP−Blm) from *Streptoalloteichus hindustanus*, demonstrates the performance of the new experiment for application to larger molecular systems.

(12) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 13868−13878.
(13) Zwahlen, C.; Gardner, K. H.; Siddharta, P. S.; Horita, D. A.; Byrd, R. A.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 7617−7625.

Materials and Methods

NMR Samples. Uniformly ^{13}C, ^{15}N-labeled *human* ubiquitin was purchased from VLI (Southeastern, PA). A sample was prepared containing 2 mM ubiquitin in 45 mM sodium acetate buffer at pH 4.7 (90% H_2O, 10% D_2O). ^{13}C, ^{15}N-labeled MerAa from *R. metallidurans* was prepared as described previously.[30] Sample conditions were 1.5 mM protein in 50 mM Tris buffer at pH 7 (90% H_2O, 10% D_2O). A total of 1.2 molar equivalent of $HgCl_2$ was added to form the metal−protein complex. A ^{13}C, ^{15}N-labeled sample of the complex between *S. hindustanus* Bleomycin resistance protein (BRP) and Zn^{2+}-ligated bleomycine (Blm) was prepared as described previously.[14] The final BRP−Blm concentration was 1 mM in 20 mM deuterated MES buffer at pH 6.5 (∼99% D_2O), supplemented with 100 mM NaCl and 0.02% sodium azide.

NMR Data Acquisition and Processing. All NMR experiments were performed on a Varian INOVA 800 spectrometer, equipped with a triple-resonance (^{1}H, ^{15}N, ^{13}C) probe and shielded three-axes gradients. Additional ($2_{cs} + 1_{aa}$)-D CT-HSQC spectra were acquired on a Varian INOVA 600 spectrometer to calibrate the Bloch-Siegert frequency shifts at 600 MHz ^{1}H frequency. The sample temperature was set to 25 °C for MerAa and to 30 °C for ubiquitin and the BRP−Blm complex. Data processing and determination of peak positions were performed using the FELIX program version 2000 (Accelrys, Inc.). Mirror image linear prediction[15] was applied to the ^{13}C CT dimensions. Squared cosine apodization was used in all dimensions prior to zero-filling and Fourier transformation.

For the ($2_{cs} + 1_{aa}$)-D CT-HSQC spectra recorded for ubiquitin and MerAa, the pulse sequence of Figure 1a was used without the NOE mixing and the second ^{1}H−^{13}C HSQC transfer bloc. Data sets of 4 times (AA_i) 110(^{13}C) × 512(^{1}H) complex points were recorded for spectral widths of 4000 Hz (^{13}C) and 9000 Hz (^{1}H) in an overall experimental time of 30 min.

The amino acid-type editing filters were implemented using continuous C^{next} band-selective decoupling. As shown in Figure 1 dilated evolution times, $\lambda^n t_1$ and $\lambda^n t_2$ were used to compensate Bloch-Siegert shifts induced by the band-selective C^{next}-decoupling. The B_0 field-dependent scaling factors λ^n were calibrated experimentally from ($2_{cs} + 1_{aa}$)-D CT-HSQC spectra of ubiquitin and MerAa recorded at two B_0 field strengths of 14.4T and 18.8T, corresponding to ^{1}H frequencies of 600 and 800 MHz. The following optimized scaling factors were used for the four Hadamard experiments $H_4(n)$: $\lambda_{600}^1 = 1.0$, $\lambda_{600}^2 = 1.0046$, $\lambda_{600}^3 = 1.0181$, and $\lambda_{600}^4 = 1.0162$ at 600 MHz, and $\lambda_{800}^1 = 1.0$, $\lambda_{800}^2 = 1.0026$, $\lambda_{800}^3 = 1.0102$, and $\lambda_{800}^4 = 1.0091$ at 800 MHz. The same C^{next}-decoupling patterns and scaling factors were then used

(14) Vanbelle, C.; Brutscher, B.; Blackledge, M.; Muhle-Goll, C.; Rémy, M.-H.; Masson, J.-M.; Marion, D. *Biochemistry* **2003**, *42*, 651−663.
(15) Zhu, G; Bax, A *J. Magn. Reson.* **1990**, *90*, 405−410.
(16) Weigelt, J.; Hammarström, A.; Bermel, W.; Otting, G. *J. Magn. Reson.* **1996**, *110B*, 219−224.
(17) (a) Clore, G. M.; Kay, L. E.; Bax, A.; Gronenborn, A. M. *Biochemistry* **1991**, *130*, 12−18. (b) Zuiderweg, E. R. P.; Petros, A. M.; Fesik, S. W.; Olejniczak, E. T. *J. Am. Chem. Soc.* **1991**, *113*, 370−371.
(18) (a) Schubert, M.; Oschkinat, H.; Schmieder, P. *J. Magn. Reson.* **2001**, *148*, 61−72. (b) Dötsch, V.; Oswald, R. E.; Wagner, G. *J. Magn. Reson.* **1996**, *110B*, 107−111.
(19) Hadamard, J. *Bull. Sci. Math.* **1893**, *17*, 240−248.
(20) Kupce, E.; Freeman, R. *J. Magn. Reson.* **1996**, *118A*, 299−303.
(21) Kupce, E.; Boyd, J.; Campbell, I D. *J. Magn. Reson.* **1995**, *106B*, 300−303.
(22) Kupce, E.; Freeman, R; Wider, G.; Wüthrich, K. *J. Magn. Reson.* **1996**, *122A*, 81−84.
(23) Zhang, S.; Gorenstein, D. G. *J. Magn. Reson.* **1999**, *138*, 281−287.
(24) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439−6448.
(25) Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W. *J. Am. Chem. Soc.* **2000**, *122*, 7898−7904.
(26) Tjandra, N.; Bax, A. *J. Magn. Reson.* **1997**, *124*, 512−515.
(27) Sibille, N.; Bersch, B.; Covès, J.; Blackledge, M.; Brutscher, B. *J. Am. Chem. Soc.* **2002**, *124*, 14616−14625.
(28) Pervushin, K.; Vögeli, B. *J. Am. Chem. Soc.* **2003**, *125*, 9566−9567.
(29) Meissner, A.; Sørensen, O. W. *J. Biomol. NMR* **2001**, *19*, 69−73.
(30) Rossy, E.; Champier, L.; Bersch, B.; Brutscher, B.; Blackledge, M.; Covès, J. *J. Biol. Inorg. Chem.* **2004**, *9*, 49−58.

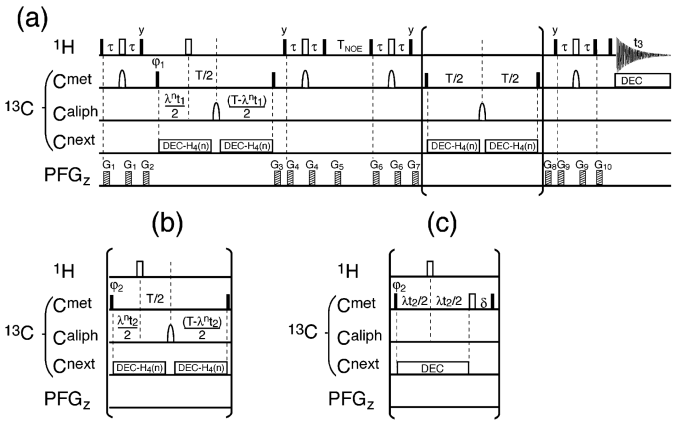


Figure 1. Pulse sequences for amino acid-type edited ($n_{cs} + n_{aa}$)-D methyl NOESY experiments. The pulse sequence in (a) allows recording of ($2_{cs} + 2_{aa}$)-D NOESY spectra. Higher-dimensional experiments, ($3_{cs} + 1_{aa}$)-D or ($3_{cs} + 2_{aa}$)-D NOESY, are generated by inserting one of the pulse sequence blocks labeled (b) and (c) into the brackets of sequence (a). All radio frequency (rf) pulses are applied along the x-axis unless indicated. 90° and 180° rf pulses are represented by filled and open pulse symbols, respectively. The proton carrier is set to the water resonance (4.7 ppm), and the ^{13}C carrier is set to the center of the methyl region (19 ppm) throughout the experiment. ^{13}C pulses drawn on the C^{aliph} or C^{next} lines are frequency-shifted to 35 ppm (C^{aliph}) and the center of the corresponding C^{next} bands, respectively, by a linear phase ramp. The selective 180° C^{met} pulses are applied with an i-SNOB-5 shape[21] covering a bandwidth of 18 ppm. For application at high magnetic field strengths, an offset-compensating C^{aliph} refocusing pulse $58_x{-}140_{-x}{-}344_x{-}140_{-x}{-}58_x$[32] is used during the CT delays T to properly invert all C^{met} and C^{next} carbon spins over a bandwidth of 80 ppm. The CT delay was set to $T = 28$ ms, and the INEPT transfer delays were set to $\tau = 1.6$ ms. The small delay δ in (c) compensates for the length of the 180° ^{1}H pulse. ^{13}C decoupling during acquisition is achieved using a WALTZ-16 sequence[33] at a field strength of $\gamma B_1/2\pi = 3$ kHz. Pulsed field gradients, G_{1-G10}, are applied along the z-axis (PFG_z) with a gradient strength of approximately 20 G/cm and lengths ranging from 100 to 500 μs, followed by a recovery delay of 100 μs. The phase cycle employed is $\varphi_1 = x,-x$, and $\varphi_{rec} = x,-x$. Quadrature detection in t_1 and t_2 is obtained by incrementing the phases φ_1 and φ_2 according to STATES-TPPI. Constant-adiabaticity (CA) WURST-2 pulses are used as the basic elements for the C^{next}-decoupling. The CA-WURST-2 pulses have a length of $\tau_p = 5$ ms and cover a bandwidth of 12 ppm. They are centered at 37(1), 53 (−15), and 70 ppm (−32 ppm) for the Val-Ile$_\gamma$, Ala, and Thr C^{next} bands, respectively. The values in parentheses indicate the center of the symmetric off-resonance decoupling applied for each band using the same amplitude modulation but opposite frequency sweep.[22] Multiple-band-selective decoupling is achieved by vector addition of the individual WURST pulses.[34] The final decoupling sequence is phase-cycled according to a TPG-5 supercycle.[35] The C^{next}-decoupling is restarted at the beginning after the 180° C^{aliph} refocusing pulse. For the four-step Hadamard amino acid-type selection filter, four experiments are recorded using C^{next}-decoupling of different bands: H_4(1) none, H_4(2) Ala and Thr, H_4(3) Ala and (Val-Ile$_\gamma$), H_4(4) Thr and (Val-Ile$_\gamma$). At 800 MHz ^{1}H frequency, the scaling factors $\lambda_{800}{}^n$ were set to $\lambda_{800}{}^1 = 1.0$, $\lambda_{800}{}^2 = 1.0026$, $\lambda_{800}{}^3 = 1.0102$, and $\lambda_{800}{}^4 = 1.0091$. At 600 MHz ^{1}H frequency, the following optimized scaling factors were used: $\lambda_{600}{}^1 = 1.0$, $\lambda_{600}{}^2 = 1.0046$, $\lambda_{600}{}^3 = 1.0181$, and $\lambda_{600}{}^4 = 1.0162$. A Hadamard transformation[5,19] is applied along the filter dimensions to disentangle the NMR signals from the individual frequency bands. C^{next}-decoupling during t_2 in (c) is realized using the same WURST-2 based sequence described above applied to the Ala, Thr, and (Val-Ile$_\gamma$) C^{next}-frequency bands simultaneously. Bloch-Siegert shift compensation is achieved by setting $\lambda_{800} = 1.0111$ (or $\lambda_{600} = 1.0195$). For modulation sideband suppression the decoupling sequence is shifted in time by an amount of $\tau_p/2$ for each second scan.[16] Pulse sequences in Varian pulse sequence language and transformation protocols in Felix macro language can be obtained from the authors upon request.

without any further optimization for the amino acid-type edited NOESY experiments, described below.

($2_{cs} + 2_{aa}$)-D methyl NOESY spectra were recorded using the pulse sequence of Figure 1a. Different NOE mixing times were tested ranging from 75 to 200 ms. The highest NOE cross-peak intensities were observed in the spectra recorded with $T_{NOE} = 200$ ms. Four-step amino acid-type editing was used during the two CT ^{13}C-evolution periods before and after the NOE mixing. Chemical shift evolution was sampled with 110(t_1) and 512(t_3) complex points for spectral widths of 4000 and 9000 Hz, respectively, yielding a final matrix of 220(C_i) × 1024(H_j) × 4(AA_i) × 4(AA_j) data points. The total acquisition time was set to 2 h for ubiquitin and 4 h for MerAa.

A ($3_{cs} + 1_{aa}$)-D methyl NOESY spectrum was acquired for the BRP−Blm complex using the pulse sequence of Figure 1a with the insert of Figure 1c. The NOE mixing time was set to 100 ms. Multiple-band-selective C^{next}-decoupling was applied during t_2 using two different decoupling schemes for alternating scans.[16] Chemical shift evolution was sampled with 90(t_1), 90(t_2), and 512(t_3) complex points for spectral widths of 4000 (C_i), 4000 (C_j), and 9000 Hz (H_j), respectively, yielding a final matrix of 180(C_i) × 180 (C_j) × 1024 (H_j) × 4(AA_i) data points. A simple two-step phase cycle, two scans per FID, and a repetition delay of 0.9 s were used to yield a total acquisition time of 80 h.

Results and Discussion

Figure 1 illustrates the new ^{13}C and amino acid-type edited methyl NOESY experiment developed to record high-resolution NOESY spectra for methyl−methyl distance measurement. The pulse sequence is of the ^{13}C-HSQC-NOESY-^{13}C-HSQC type[13,17] with the flow of magnetization for two proximal methyl groups (i) and (j) given as follows:

$$H_i \xrightarrow{J_{CH}} C_i[t_1, AA_i - \text{filter}] \xrightarrow{J_{CH}} H_i \xrightarrow{NOE} H_j \xrightarrow{J_{CH}} C_j[t_2, AA_j - \text{filter}] \xrightarrow{J_{CH}} H_j[t_3] \quad (1)$$

The active spin−spin couplings are indicated above the arrows. While C_i frequency labeling is performed in a CT manner during t_1, with the constant time delay set to the inverse of the scalar carbon−carbon coupling constant ($T \cong 1/J_{CC}$), for C_j either CT (Figure 1b), real-time (Figure 1c), or no (Figure 1a) frequency labeling is used in t_2. Two AA filters can be applied, one before and one after the NOE mixing, thus allowing additional amino acid-type identification of both interacting methyl groups. The pulse sequence of Figure 1a allows recording of fast ($2_{cs} + 2_{aa}$)-D (C_i, AA_i, H_j, AA_j)-NOESY spectra for application to small and medium-sized proteins under high-sensitivity experimental conditions (concentrated protein sample, high magnetic field strength, cryogenic probe, etc.). For application to larger proteins, both methyl carbons C_i and C_j

are frequency-labeled, using the inserts of Figure 1b or 1c to record ($3_{cs} + 2_{aa}$)-D (C_i, AA_i, C_i, H_j, AA_j)-NOESY or ($3_{cs} + 1_{aa}$)-D (C_i, AA_i, C_i, H_j)-NOESY spectra, respectively. The major modification with respect to pulse sequences proposed previously[13] is the addition of amino acid-type editing filters (AA filters), which will be discussed in detail in the following section.

Amino Acid-Type Editing of Methyl Groups. Amino acid-type selective experiments have been proposed in the past[18] to simplify the $^1H-^{15}N$ correlation spectra of $^{13}C,^{15}N$ labeled proteins. These experiments exploit the different spin-coupling topologies of amino acid side chains. Here, a different concept is used for amino acid-type editing based on frequency-selective Hadamard spectroscopy.[5,19] Methyl groups are located at the side-chain ends, and thus the methyl C^{met} is only bound to one other carbon, which we will refer to as C^{next}. The C^{next} corresponds to carbon sites with distinct chemical shift ranges for different amino acid types. This feature is exploited in the Hadamard amino acid-type selection filters of Figure 1. Four C^{next} frequency bands were defined as follows: AA = 4: 64−76 ppm (Thr), AA = 3: 31−43 ppm (Val-Ile-γ), AA = 2: 47−59 ppm (Ala), and AA = 1: 10−30 ppm (Leu-Ile$_\delta$). The ^{13}C spins resonating within these frequency bands can be manipulated independently by means of band-selective radio frequency (rf) pulses. A binary Hadamard-type C^{next}-frequency encoding scheme, consisting in a series of "plus" and "minus" signs, is used to disperse the methyl correlations along an additional amino acid-type dimension. The sign encoding is realized by the presence ("minus") or absence ("plus") of $C^{met} - C^{next}$ scalar-coupling (J_{CC}) evolution during the CT delay T. In the absence of C^{next}-decoupling, the J_{CC} coupling evolution is active during $T = 1/J_{CC}$, resulting in a sign inversion of the detected NMR signal ("minus"-encoding):

$$C_x^{met} \xrightarrow{J_{CC}} - C_x^{met}$$

Additional band-selective C^{next}-decoupling allows switching off the $C^{met} - C^{next}$ scalar coupling evolution for a given amino acid type ("plus"-encoding):

$$C_x^{met} \xrightarrow{dec} + C_x^{met}$$

Four spectra are recorded, where the sign of the NMR signals originating from methyl groups with C^{next} corresponding to different frequency bands is changed according to a Hadamard matrix of order 4: $H_4(1)$: without C^{next}-decoupling, $H_4(2)$: with Ala-Thr C^{next}-decoupling, $H_4(3)$: with Ala-(Val-Ile$_\gamma$) C^{next}-decoupling, and $H_4(4)$: with Thr-(Val-Ile$_\gamma$) C^{next}-decoupling. The four bands (AA = 1, 2, 3, and 4), corresponding to different amino acid types, are then disentangled by a Hadamard transformation performing linear combinations of the four recorded data sets $H_4(1)$, ..., $H_4(4)$. In this Hadamard filter, the methionine methyl groups, which have no directly attached carbon, are detected in the AA = 1 band together with the methyls from Leu and Ile$_\delta$ residues, but with negative intensity. Hadamard spectroscopy provides a time-saving alternative to C^{next} chemical shift labeling along an additional frequency dimension, because only four points need to be recorded for the Hadamard AA-type dimension. It is also more sensitive because no additional spin evolution period is required, and the full NMR signal is detected (sign-modulated) in each of the four Hadamard experiments. Therefore, the amino acid-type edited $^1H-^{13}C$ correlation experiment retains the full sensitivity of the conventional experiment of the same duration.

C^{next}-decoupling can be achieved in two alternative ways: either by continuous band-selective decoupling or by band-selective spin inversion. The latter implementation has the advantage that no Bloch-Siegert shifts or modulation sidebands are created, but it reduces the available time for C^{met} chemical shift labeling, and thus the spectral resolution. We have tested both approaches experimentally, using a 5 ms constant adiabaticity (CA) WURST-2 pulse[20] as the repeated element of the decoupling sequence and an i-SNOB-5[21] pulse shape for C^{next} spin inversion. We obtained a much better filter performance (proper selection of the individual AA bands) in the case of continuous band-selective decoupling, provided that the Bloch-Siegert shifts and modulation sidebands were correctly suppressed. Therefore, we have chosen the continuous decoupling option for all the experiments presented here. Bloch-Siegert shift compensation is crucial for the amino acid-type editing filter, which relies on the addition and subtraction of different subspectra. Additional off-resonance decoupling is applied for each C^{next} band, symmetrically with respect to the center of the methyl spectrum with the same shape and opposite frequency sweep.[22] This eliminates the Bloch-Siegert shifts for C^{met} resonances close to the spectral center. The remaining shifts, which are a linear function of the C^{met} frequency offset, can be compensated by appropriately adjusting the B_0-field-dependent scaling factors λ^n in the pulse sequence of Figure 1.[23] Experimentally optimized $\lambda_{600}{}^n$ and $\lambda_{800}{}^n$ values are given in the Materials and Methods section and in the caption to Figure 1. Modulation sidebands are suppressed by properly adjusting the WURST-2 pulse length and the constant time delay T. A detailed theoretical and experimental evaluation of these effects on the filter performance is presented elsewhere (Van Melckebeke, H.; Simorre, J.-P.; Brutscher, B. *J. Magn. Reson.* **2004**, in press).

To demonstrate the performance of the new amino acid-type methyl-editing filter, a ($2_{cs} + 1_{aa}$)-D CT-HSQC spectrum was recorded on the MerAa sample. The (C_i, H_i, AA_i) correlation spectrum was acquired using the pulse sequence of Figure 1a without the NOE mixing and the second $^1H-^{13}C$ HSQC transfer block. The four 2D planes extracted along the AA_i dimension are shown in Figure 2a−d. A few small residual peaks (marked by a star) are observed in the spectra of Figure 2a,c with an intensity of less than 5% of the normal cross-peak intensity. These peaks are due to slightly different peak intensities of the particular cross-peak in the four data sets recorded for the filter, because of small variations in the J_{CC} coupling constants and differential relaxation of C_x^{met} and $2C_x^{met}C_z^{next}$ coherence. A clean separation of methyls from Ala and Thr side chains is obtained by the filter sequence. While the alanine (AA = 2) and threonine (AA = 4) bands cover more than 95% of the corresponding C^{next} chemical shifts found for diamagnetic proteins in the BMRB data bank, only about 70% of the valine and 85% of the Ile$_\gamma$ methyls have their corresponding C^{next} frequencies inside the AA = 3 band. This leads to some "cross-talk" between the AA = 1 (Met-Leu-Ile$_\delta$) and AA = 3 (Val-Ile$_\gamma$) bands, as shown in Figure 2a,c for MerAa. Therefore, some ambiguities may remain when using such methyl correlation spectra for amino acid-type identification.

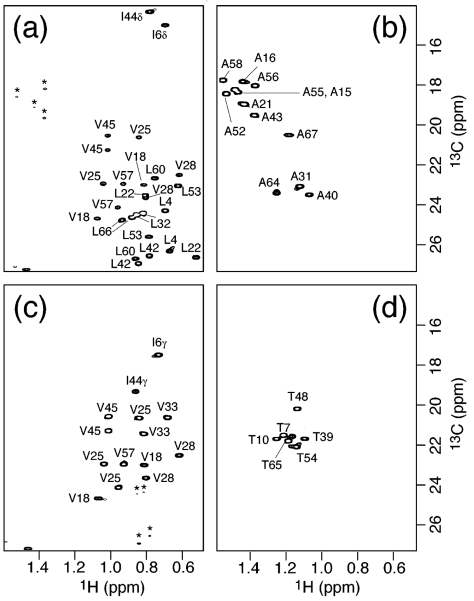


Figure 2. ($2_{cs} + 1_{aa}$)-D methyl CT-HSQC spectrum recorded for MerAa using the pulse sequence of Figure 1a without the NOE mixing and the second HSQC. The four (C_i, H_i) spectra extracted along the Hadamard filter dimension are shown in (a) (Met-Leu-Ile$_\delta$) band, (b) Ala band, (c) (Val-Ile$_\gamma$) band, and (d) Thr band. The corresponding amino acid type and residue number assigns the bona fide cross-peaks, and a star indicates a residual peak due to filter imperfections.

In conclusion, spectral resolution of the methyl ^{1}H−^{13}C correlation spectrum is significantly enhanced by the amino acid-type selection filter, without a significant increase in experimental time. We therefore expect that the amino acid-type editing filter, presented here, will improve many existing NMR experiments based on the resolution of ^{1}H−^{13}C methyl correlation spectra, especially for application to larger molecules. Examples are the study of side-chain dynamics,[24] molecular interfaces by chemical shift mapping,[25] or the measurement of residual dipolar or scalar coupling constants for structure determination.[26−28]

Application to Smaller Molecules: Fast Methyl NOE Measurement. As a first application of interest, highly resolved methyl ($2_{cs} + 2_{aa}$)-D NOESY spectra of moderately sized proteins can be recorded in a short experimental time, using the pulse sequence of Figure 1a. Acquisition times of 2 h for ubiquitin and 4 h for MerAa proved sufficient on a standard 800 MHz spectrometer to detect a large number of methyl−methyl correlation peaks. The experiment is basically a 2D NOESY experiment where one of the proximal methyls (methyl-i) is identified by its ^{13}C, and the other one (methyl-j) is identified by its ^{1}H frequency. CT ^{13}C editing removes line splitting due to scalar J_{CC} couplings yielding narrow lines. In general, such 2D (C_i, H_j)-NOESY spectra suffer from the presence of strong diagonal peaks with intensities 1 order of magnitude higher than observed for the cross-peaks and, in addition, from severe signal overlap among cross-peaks. Both problems were addressed in the present experiment by amino acid-type spectral editing filters, described in the previous section, and applied before and after the NOE mixing. In the

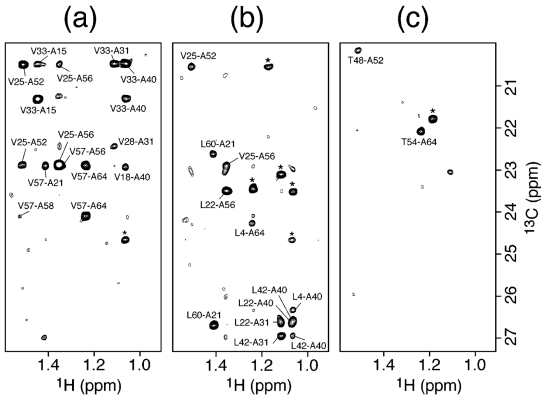


Figure 3. 2D (C_i, H_j) planes extracted from a ($2_{cs} + 2_{aa}$)-D methyl NOESY spectrum recorded for MerAa using the pulse sequence of Figure 1a. An NOE mixing time of 200 ms was used. The different sub spectra were extracted at (a) $AA_i = 3$, $AA_j = 2$, (b) $AA_i = 1$, $AA_j = 2$, and (c) $AA_i = 3$, $AA_j = 2$, with 1, 2, 3, and 4 corresponding to the (Met-Leu-Ile$_\delta$), Ala, (Val-Ile$_\gamma$), and Thr bands, respectively. Peaks labeled by a star are residual diagonal peaks, whereas assigned peaks correspond to methyl−methyl NOEs.

resulting ($2_{cs} + 2_{aa}$)-D spectrum, the NOE diagonal and cross-peaks are spread along two additional dimensions (AA_i, AA_j) according to the amino acid-type of the involved methyl groups (see above). Diagonal peaks are detected in only four out of the 16 (C_i, H_j) planes, while the 12 remaining planes are (mostly) diagonal-free. The concept of double-editing for diagonal suppression has been proposed in the past for CH_n-multiplicity editing in HCCH correlation spectra of proteins.[29] Examples of 2D (C_i, H_j) planes, extracted from the ($2_{cs} + 2_{aa}$)-D NOESY spectrum of MerAa at different (AA_i, AA_j) coordinates, are shown in Figure 3a−c. The observed correlation peaks correspond to NOEs between methyl groups of Ala and either (a) Val-Ile$_\gamma$, (b) Met-Leu-Ile$_\delta$, or (c) Thr residues. Most of the peaks are well-resolved, and the few residual diagonal peaks (indicated by a star in Figure 3) are of the same or lower intensity than the observed cross-peaks. The high spectral resolution allows fast and straightforward NOE peak assignment. The remaining ambiguities are resolved by exploiting the intrinsic symmetry of the methyl NOESY spectrum. The spatial proximity of two methyls gives rise to a first cross-peak at (C_i, H_j, AA_i, AA_j) and a second one at (C_j, H_i, AA_j, AA_i) in the ($2_{cs} + 2_{aa}$)-D NOESY data set. A simple protocol allows extraction of (almost) unambiguous methyl−methyl distance restraints from peak picking lists of the ($2_{cs} + 2_{aa}$)-D NOESY data set and the assigned ($2_{cs} + 1_{aa}$)-D CT-HSQC spectrum (C-program available from the authors upon request). The low time requirement for recording a ($2_{cs} + 2_{aa}$)-D NOESY spectrum also makes it attractive to run a series of experiments varying the NOE mixing time T_{NOE} to detect and quantify a larger number of long-range methyl NOEs.

Figure 4 shows connectivity plots of observed short methyl−methyl distances in the proteins (a) ubiquitin and (b) MerAa. Different symbols represent NOE peaks detected in distinct (C_i, H_j) planes of the ($2_{cs} + 2_{aa}$)-D NOESY spectra. The cross-peaks were assigned without relying on the existing structures of the two molecules. A total of 52 and 58 medium- and long-range contacts ($i \geq j + 2$) were observed for ubiquitin and MerAa, respectively. A large number of the relevant NOE peaks were

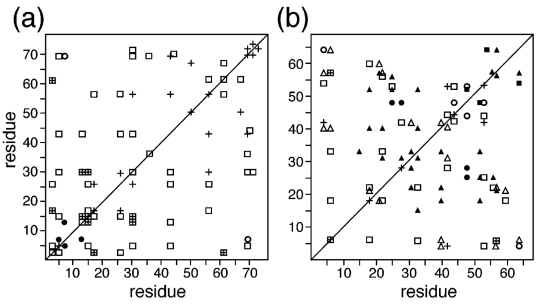


Figure 4. Methyl NOE connectivity plot for (a) ubiquitin and (b) MerAa. All NOEs were extracted and assigned from a single ($2_{cs} + 2_{aa}$)-D methyl NOESY spectrum recorded in 2 h for ubiquitin and 4 h for MerAa. The different symbols indicate in which 2D (C, H) planes the particular NOE has been detected: Open triangles (1, 2), open squares (1, 3), open circles (1, 4), filled triangles (2, 3), filled squares (2, 4), filled circles (3, 4), with (k, l) the coordinates ($AA_i = k$, $AA_j = l$) in the ($2_{cs} + 2_{aa}$)-D methyl NOESY spectrum. The numbers 1, 2, 3, and 4 correspond to the (Met-Leu-Ile$_\delta$), Ala, (Val-Ile$_\gamma$), and Thr bands, respectively. Note that all NOE peaks were also detected in the symmetric (l, k) plane. A cross indicates a diagonal plane (k, k).

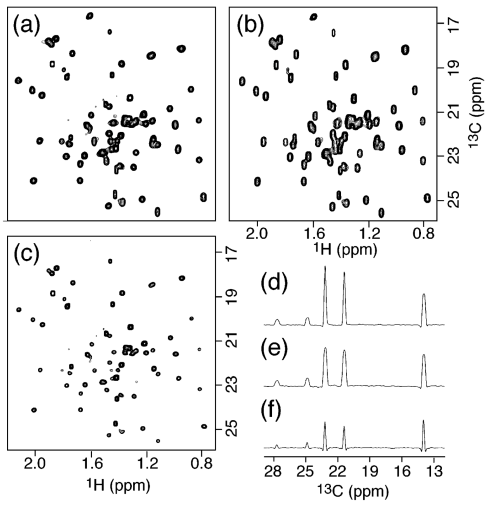


Figure 5. $^{1}H-^{13}C$ methyl correlation spectra of the BRP−Blm complex recorded at 800 MHz ^{1}H frequency using (a) real-time ^{13}C labeling and C^{next}-decoupling of the Ala, (Val-Ile$_\gamma$), and Thr bands, (b) real-time ^{13}C labeling, and (c) constant time ^{13}C labeling. Identical acquisition and processing parameters were used for all three experiments. The t_1 acquisition time was set to $t_1^{max} = 28$ ms, which ensures reasonably well-resolved correlation spectra even for non-CT frequency labeling. The three spectra were plotted using the same contour levels. For a better appreciation of the relative line width and intensity, 1D spectra were extracted along the ^{13}C dimension at 0.8 ppm ^{1}H frequency from (d) C^{next}-decoupled HSQC, (e) HSQC, and (f) CT-HSQC spectra.

extracted from "diagonal-free" NOESY planes. Because of the different amino acid-type composition of the two proteins, the dispersion of the NOE cross-peaks along the additional Hadamard AA dimensions differs for the two proteins. While for MerAa a large number of observed NOEs involve alanine residues, for ubiquitin mostly valine, leucine, and isoleucine residues are found in the hydrophobic protein core. The experiment works well for both proteins, yielding a large number of long-range distance restraints in a short experimental time. The set of proximal methyl−methyl contacts obtained for MerAa (Figure 4b) is very similar to previous results obtained from a standard 3D CT-HSQC-NOESY-CT-HSQC data set acquired in an overall time of 60 h,[30] as compared to 4 h for the new ($2_{cs} + 2_{aa}$)-D NOESY experiment. Combined with additional restraints defining the local structure of the two α-helices and the β-sheet, the methyl−methyl NOEs proved sufficient to define the fold of the protein with an rmsd of ∼1.2 Å calculated over the heavy atoms of the protein backbone.[30] Compared to standard 3D NMR techniques, measurement times are reduced by roughly 1 order of magnitude, as long as sensitivity is not the limiting factor. The new ($2_{cs} + 2_{aa}$)-D methyl NOESY experiment complements the set of NMR experiments proposed for fast backbone and methyl side-chain resonance assignment.[6,7] This makes it possible to record a complete set of NMR data required for resonance assignment and molecular fold determination in typically a few days or less on a modern high field NMR spectrometer. Fast data acquisition will be particularly important for high-throughput molecular fold screening of small proteins and for the study of unstable protein systems by NMR.

Application to Larger Molecules: Optimized Resolution and Sensitivity of Methyl NOE Experiments. For larger proteins, a third chemical shift dimension (C_j) is added to the methyl NOE experiment (inserts of Figure 1b,c) to retain the high spectral resolution required for unambiguous NOE assignment. Because the sensitivity of the pulse sequence of Figure 1a decreases very quickly with increasing molecular weight, both sensitivity and spectral resolution of the experiment need to be optimized to detect and unambiguously assign a sufficient number of methyl NOEs for molecular fold determination.

The relative sensitivity of real-time versus CT techniques depends on the molecular tumbling correlation time τ_c and the total acquisition time t_1^{max}.[13] For small molecules or highly mobile side chains, CT ^{13}C-labeling presents a resolution and sensitivity advantage with respect to real-time frequency labeling because the relaxation loss during the CT delay is compensated by the removal of the line splitting due to the J_{CC} scalar coupling. For larger proteins, CT frequency labeling ensures narrow lines in the ^{13}C dimensions, albeit at the expense of sensitivity. We have experimentally evaluated the relative sensitivity of $^{1}H-^{13}C$ methyl HSQC spectra, using either CT or real-time ^{13}C labeling, for a sample of the Bleomycin resistance protein (BRP, 2 × 14 kDa) from *S. hindustanus* in complex with Zn^{2+}-ligated bleomycine (Blm). This 30 kDa protein complex was used as a model system for a larger protein. Spectra acquired at 30 °C ($\tau_c = 12.7$ ns) yield an average signal gain of a factor of 1.5 for real time ^{13}C labeling (Figure 5a,c). At 10 °C ($\tau_c = 22.4$ ns), an average signal enhancement of a factor of 2.5 is obtained (data not shown).

Although the homonuclear J_{CC} couplings are not resolved in the spectra acquired without CT labeling (Figure 5b), they represent a significant contribution to the ^{13}C line width. Homonuclear decoupling thus offers an attractive way to further increase the resolution and sensitivity of the experiment. C^{next}-decoupling during t_2 in the pulse sequence element of Figure 1c is applied as described above for the amino acid-type selection filter, with simultaneous decoupling of the Ala, Thr, and (Val-Ile$_\gamma$) bands. Symmetrically applied off-resonance decoupling combined with a dilated evolution time λt_2, using

an optimized scaling factor λ, compensates the Bloch-Siegert shifts in the methyl ^{13}C spectrum.[23] To suppress the modulation sidebands, a different decoupling sequence is used for odd- and even-numbered scans. As shown previously,[16] the sign of the sidebands can be inverted by a $\tau_p/2$ time shift of one decoupling sequence with respect to the other, where τ_p is the length of a single inversion pulse as defined above. Addition of the two spectra yields efficient sideband suppression, as illustrated in Figure S1 of the Supporting Information. The C^{next}-decoupled $^1H-^{13}C$ methyl HSQC correlation spectrum, recorded for BRP−Blm, is shown in Figure 5a. Compared to the standard HSQC spectrum shown in Figure 5b, a considerable resolution enhancement is observed in the crowded central spectral region that comprises mostly alanine, threonine, valine, and isoleucine-γ methyl cross-peaks. This resolution enhancement is accompanied by an increase in signal intensity, which is on average 50%. This gain is slightly reduced to 30% for spectra recorded at 10 °C (data not shown). Cross-peaks from methyl groups in methionine, leucine, and the δ-position of isoleucines are not affected by the C^{next}-decoupling. The multiple-band C^{next}-decoupled $^1H-^{13}C$ HSQC sequence (Figure 1c) offers a good compromise between spectral resolution and sensitivity for application to larger proteins, as long as the J_{CC} coupling presents a nonnegligible contribution to the ^{13}C line width.

The structure of BRP solved by X-ray crystallography[31] and confirmed by solution NMR[14] highlights a homodimeric fold stabilized by an exchanged β-strand, binding one bleomycine molecule per monomer. BRP is rich in methyl-containing side chains (50 out of 124) resulting in a total of 77 methyl groups per monomer, which are mainly located in the hydrophobic pockets of the protein. The detection of long-range methyl−methyl contacts is thus essential for structure determination of this 30 kDa protein complex by NMR. To demonstrate the performance of the new methyl NOESY experiment for application to a larger protein system, we have recorded a (3_{cs} + 1_{aa})-D data set of the BRP−Blm complex at 30 °C using the pulse sequence of Figure 1a with the insert of Figure 1c. A basic two-step phase cycle allows recording of spectra with high resolution in the ^{13}C dimensions in a reasonable experimental time of about 3 days. In the (3_{cs} + 1_{aa})-D spectrum, one of the proximal methyls is identified by its ^1H and ^{13}C chemical shift (H_i, C_i), whereas the other one is characterized by its ^{13}C frequency and amino acid type (C_j, AA_j). A selection of (ω_{Ci}, ω_{Hj}) strips extracted form the (3_{cs} + 1_{aa})-D data set is shown in Figure 6a. Each slice is taken at the ω_{Cj} frequency corresponding to the ^{13}C chemical shift of the methyl group indicated at the top. Only AA_i slices corresponding to the (Met-Leu-Ile$_\delta$), Ala, and (Val-Ile$_\gamma$) bands are shown, as no threonine residues are found in this particular hydrophobic patch (Figure 6b). A number of long-range contacts are observed in the spectra of Figure 6a, defining the spatial proximity of the first α-helix (α_1) and part of the β-sheet comprising strands β_3 and β_4, forming one of the structural subunits of the BRP monomer.[31] The additional amino acid-type selection filter clearly provides a gain in spectral resolution by dispersing the NOE peaks along the additional AA_j dimension. This allows, for example, to distinguish the cross-peak between the methyls of Val20$_{\gamma2}$ and Ala39 (highlighted by bold letters in Figure 6a) from the strong diagonal peak, although both methyl ^{13}C have very similar resonance frequencies. For the BRP−Blm complex, the high spectral resolution allowed unambiguous and straightforward assignment of most of the observed methyl NOEs in the (3_{cs} + 1_{aa})-D spectrum.

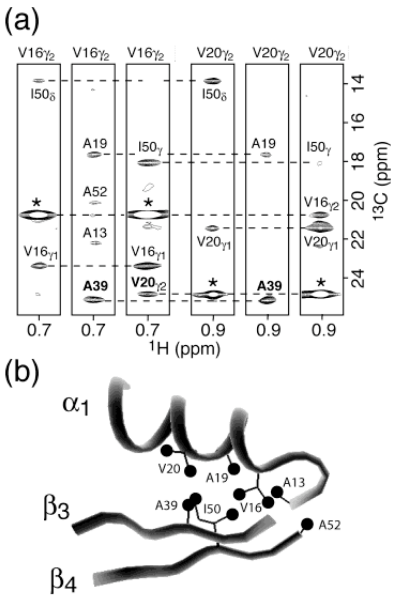


Figure 6. (a) (C_i, H_j) strips extracted from a (3_{cs} + 1_{aa})-D methyl NOESY spectrum recorded for the BRP−Blm complex. The spectrum was recorded with the pulse sequence of Figure 1a and the insert of Figure 1c. The NOE mixing time was set to 100 ms. The strips were taken at the C_j frequencies of valine methyls V16$_{\gamma2}$ and V20$_{\gamma2}$ and from different AA_i planes. Only AA_i planes with detected NOE peaks are shown. Diagonal peaks are marked by a star, and cross-peaks detected at (C_i, AA_i) are assigned to the corresponding methyl group. The methyl NOEs detected in (a) define a hydrophobic patch formed by helix α_1, and the two strands β_3 and β_4 shown in (b). The schematic representation was derived from the X-ray structure of the BRP dimer[31] using the INSIGHT software (Accelrys).

If sensitivity allows for a second CT ^{13}C-evolution period (insert of Figure 1b), amino acid-type selection for both proximal methyls further increases spectral resolution. Because sensitivity may be increased by the use of higher field spectrometers and cryogenic probes, the high spectral resolution provided by the new amino acid-type edited NOESY experiment may become particularly helpful for measuring a set of unambiguous long-range distances for larger molecular systems.

Conclusions

Methyl groups are valuable probes of molecular structure and dynamics. It is therefore important to have NMR tools to record high-resolution spectra of methyl groups. Here, a new NMR filter sequence has been presented that allows amino acid-type editing of methyls in $^1H-^{13}C$ correlation experiments. This filter presents a spectroscopic alternative to expensive and time-consuming amino acid-type-specific isotope labeling of the protein to increase spectral resolution. As a particular important

(31) Dumas, P.; Berdoll, M.; Cagnon, C.; Masson, J.-M. *EMBO J.* **1994**, *13*, 2483−2492.
(32) Shaka, A. J.; Pines, A. *J. Magn. Reson.* **1987**, *71*, 495−503.
(33) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. *J. Magn. Reson.* **1983**, *52*, 335−338.
(34) Kupce, E.; Freeman, R. *J. Magn. Reson.* **1993**, *105A*, 234−238.
(35) Tycko, R.; Pines, A.; Gluckenheimer, R. *J. Chem. Phys.* **1985**, *83*, 2775−2802.

application of this filter, we have introduced new NMR experiments to record ($n_{cs} + n_{aa}$)-dimensional methyl−methyl correlation spectra for NOE measurements. These experiments present interesting new features for application to small and large molecules. We have shown that ($2_{cs} + 2_{aa}$)-D NOESY experiments are most appropriate for fast methyl−methyl NOE measurement in smaller proteins. Fast acquisition schemes are especially important in the context of high-throughput structural genomics or the study of unstable molecular systems by NMR. Higher-dimensional ($3_{cs} + 1_{aa}$)-D or ($3_{cs} + 2_{aa}$)-D NOESY spectra provide the high resolution required for unambiguous NOE assignment of larger proteins, which is a crucial step for protein fold determination by NMR. We therefore expect that the methyl NOESY experiments proposed here will become widespread in liquid-state NMR of proteins.

Acknowledgment. This work was supported by the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique. We thank B. Bersch, E. Rossy, and J. Covès for the preparation of the MerAa sample, and C. Vanbelle, M.-H. Rémy, and J.-M. Masson for the BRP−Blm sample. H.V.M. acknowledges the receipt of a fellowship from the C.E.A.

Supporting Information Available: One figure illustrating sideband suppression for C^{next}-decoupling during real-time ^{13}C labeling using a time-shifted decoupling sequence for alternating scans. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0489644

II. Filtrage des corrélations selon le type de base dans l'expérience HCN

II. 1. Importance de l'expérience HCN pour l'attribution des ARN.

A cause de la faible dispersion des raies de résonance des acides nucléiques, il faut avoir recours au marquage isotopique ^{15}N et ^{13}C pour attribuer un ARN de taille supérieure à 15 nucléotides. La stratégie employée pour l'attribution d'un ARN marqué peut être décomposée en plusieurs étapes (pour revue : Wijmenga *et al.*, 1993 ; Fürtig *et al.*, 2003). La Figure 66 donne une représentation des différents noyaux dont les fréquences sont corrélées à chaque étape.

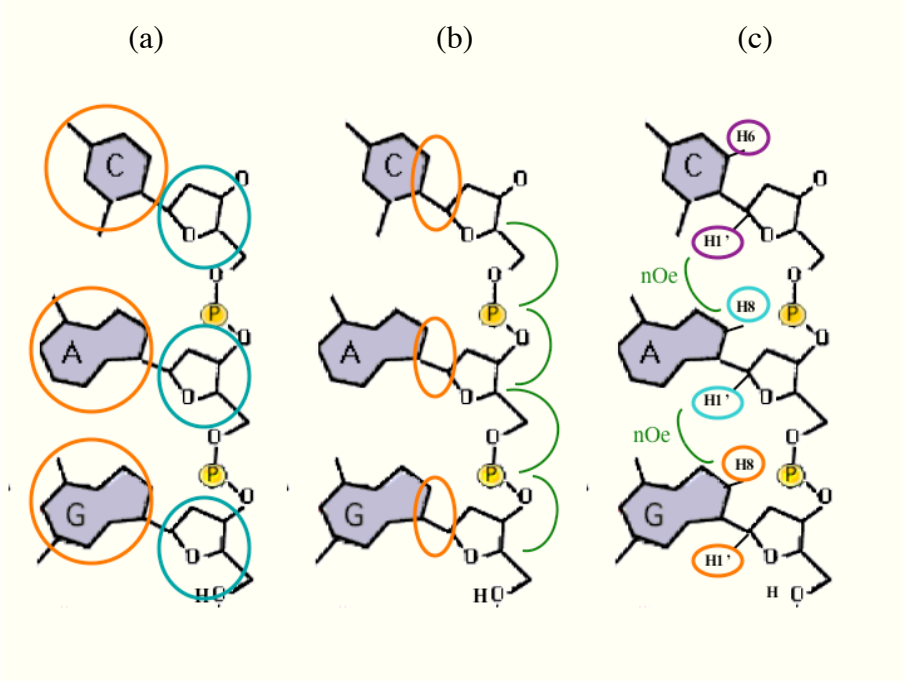


Figure 66: Mise en corrélation des différents noyaux à chaque étape de l'attribution des ARN. Les étapes A, B, et C correspondent à celles qui sont décrites dans le texte. (a) Etape A. (b) Etape B et étape C dans le cas d'une attribution séquentielle passant par la chaîne principale. (c) Etape C en utilisant le transfert d'aimantation séquentiel par effet NOE.

(A) Les résonances des protons du ribose sont tout d'abord corrélées entre elles par des expériences de type HCCH-TOCSY, HCCH-RELAY, ou HCCH-COSY. De même, on corrèle entre elles les fréquences des protons des bases. Les corrélations H_5-H_6 s'obtiennent par exemple par une expérience HCCH-COSY. Le transfert d'aimantation des protons échangeables vers les protons non échangeables s'effectue à travers la base par un réseau de petits couplages.

(B) La deuxième étape consiste à relier le système de spins du sucre au réseau de spins de la base. Pour cela, on peut utiliser des expériences du type HCNCH qui corrèlent le proton $H_{1'}$ du sucre au proton H_6 (pour les pyrimidines) ou H_8 (pour les purines) de la base. Ces expériences souffrent néanmoins d'une faible sensibilité, et sont souvent remplacées par des expériences de type HCN. Dans ces expériences, l'aimantation des protons $H_{1'}$ et H_6/H_8 est transférée simultanément aux carbones qui les portent, puis à l'azote qui assure la liaison glycosidique entre le sucre et la base (N_1 pour les pyrimidines, et N_9 pour les purines) (Figure 67a). Les spectres de corrélation à deux dimensions H(C)N présentent donc pour chaque nucléotide deux pics à la même fréquence azote et aux fréquences proton correspondant à $H_{1'}$ et H_6/H_8 (Figure 67b). Ajouter l'édition de la dimension carbone permet de séparer les signaux qui sont superposés dans la dimension proton. Dans cette expérience, la résolution dans la dimension azote est essentielle pour pouvoir corréler sans ambigüité les pics de corrélation provenant de $H_{1'}$ à ceux de H_6/H_8. On peut cependant déplorer un fort taux de superposition, surtout pour les purines (Figure 67b).

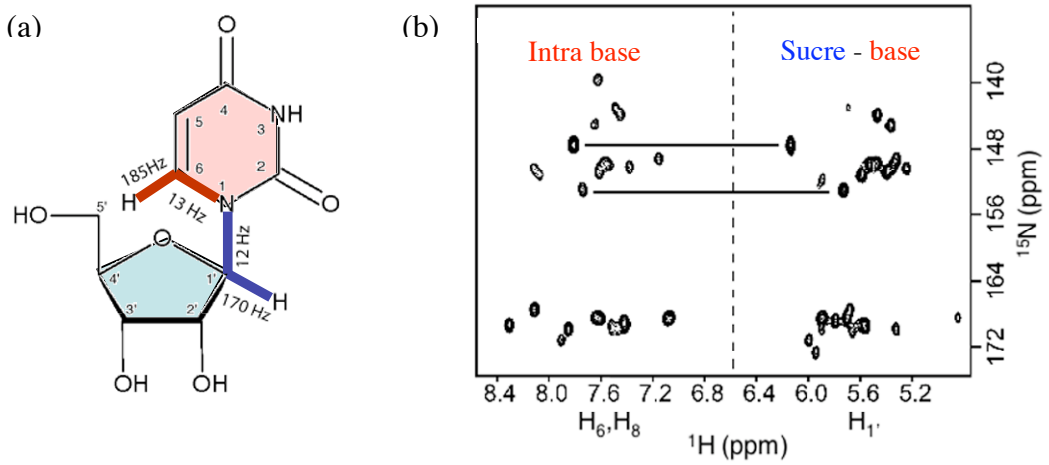


Figure 67: (a) Schéma de principe des transferts d'aimantation simultanés du sucre (bleu) ou de la base (rouge) vers l'azote de la liaison glycosidique. (b) Spectre H(C)N de l'aptamère de 33 nucléotides en complexe avec la théophylline (tiré de Brutscher *et al.*, 2001). Les lignes indiquent comment on peut corréler les fréquences de résonances des bases (partie gauche du spectre) avec celles des sucres (partie droite).

(C) La dernière étape est une étape d'attribution séquentielle. Il s'agit de corréler le système de spins de chaque nucléotide avec celui de ses voisins. Par analogie avec ce qui est fait pour l'attribution des protéines, on peut utiliser un transfert d'aimantation par couplage scalaire à travers la chaîne principale pour corréler les protons de sucres d'un nucléotide avec ceux du nucléotide voisin (Figure 66). Cette approche souffre en pratique de deux inconvénients. Tout d'abord, les couplages ^{1}H-^{31}P ^{13}C-^{31}P sont faibles, de l'ordre de 3 à 10 Hz, ce qui fait que l'aimantation des protons des sucres ne peut pas être transférée directement d'un nucléotide à l'autre. L'attribution se fait alors par la fréquence du ^{31}P qui lie les deux nucléosides, grâce à des expériences de type HCP. Le deuxième inconvénient concerne le manque de résolution des spectres du phosphore, dû non seulement à une faible dispersion des fréquences de résonance du phosphore, mais également à ses propriétés de relaxation non favorables (dues en particulier à un fort CSA). La technique d'attribution qui est utilisée quand les recouvrements sont trop importants utilise un transfert d'aimantation par couplage dipolaire entre les protons $H_{1'}$ et $H_{2'}$ du sucre, et les protons H_6/H_8 de la base suivante (Figure 66c). Des expériences de type ^{1}H-^{13}C-HSQC-NOESY sont utilisées.

II. 2. Résultats.

Les expériences de type HCN sont donc cruciales pour effectuer l'attribution des ARN doublement marqués. Les résultats présentés dans l'article qui suit (Van Melckebeke *et al*., 2005) concernent l'implémentation d'un filtre Hadamard par bandes qui permet de séparer les corrélations des spectres HCN dans une dimension spectrale discrète supplémentaire. Ce filtre permet non seulement d'augmenter la résolution des spectres, mais également de séparer les pics de corrélations selon le type de base. La réalisation du filtre et son application sur plusieurs échantillons d'ARN seront présentées.

Journal of Biomolecular NMR (2005) 32: 263–271
DOI 10.1007/s10858-005-8872-1
© Springer 2005

Article

Resolution-enhanced base-type-edited HCN experiment for RNA

Hélène Van Melckebeke[a], Arthur Pardi[b], Jérôme Boisbouvier[a], Jean-Pierre Simorre[a] & Bernhard Brutscher[a,*]
[a]*Laboratoire de RMN, Institut de Biologie Structurale - Jean-Pierre Ebel, UMR, 5075 CNRS-CEA-UJF, 41, rue Jules Horowitz, 38027, Grenoble Cedex, France;* [b]*Department of Chemistry and Biochemistry, UCB 215, Boulder, CO, 80309-0215, USA*

Received 28 January 2005; Accepted 7 June 2005

Key words: base-type selection, HCN, NMR, resonance assignment, RNA, spectral editing

Abstract

New base-type-edited transverse-relaxation optimized CT-HCN(C) experiments are presented that yield intra-base and sugar-to-base correlations for $^{13}C-^{15}N$ labeled RNA. High spectral resolution in the ^{13}C and ^{15}N dimensions is achieved by constant time (CT) frequency editing. A spectral editing filter applied during the CT ^{15}N labeling period separates the correlation peaks arising from G/U and A/C nucleotide bases. This provides the increased spectral resolution required to unambiguously connect the ^{1}H and ^{13}C resonances in sugar and base moieties of RNA nucleotides. In addition, the experiment allows base type identification for each residue, and therefore presents an attractive spectroscopic alternative to nucleotide-specific isotope labeling. Application to a 33-nucleotide RNA aptamer demonstrates the performance of the new pulse scheme.

Introduction

HCN-type correlation experiments (Sklenar et al., 1993a; Farmer et al., 1994) are important for resonance assignment of nucleic acids (Wijmenga and van Buuren, 1998; Furtig et al., 2003). They provide intranucleotide sugar-base connectivity information by correlating the base C_6, H_6 (pyrimidines) or C_8, H_8 (purines) with the sugar $C_{1'}$, and $H_{1'}$, nuclei. Recently, HCN experiments have also shown to be useful for the measurement of scalar and residual dipolar spin–spin couplings in RNA (Yan et al., 2002; Jaroniec et al., 2005), or of cross-correlated relaxation rates related to the torsion angle χ_1 (Ravindranathan et al., 2003; Duchardt et al., 2004) to obtain complementary structural information to NOE-derived distance restraints. Different implementations of the HCN sugar-to-base correlation experiment have been proposed in the past. In the most common approach, two separate correlation peaks are detected for the sugar and the base moieties. The two peaks corresponding to the same nucleotide are then identified via the shared ^{15}N chemical shift of the nitrogen in the glycosidic bond, N_1 in pyrimidines and N_9 in purines (Figure 2a). To increase the intrinsically low sensitivity of the out-and-back HCN experiment (due to the small $^{1}J_{CN} \approx 12$ Hz scalar coupling constant), transverse-relaxation optimized pulse sequences have been proposed exploiting either the favorable relaxation properties of multiple-quantum coherence (Marino et al., 1997; Fiala et al., 1998), or the interference of the ^{13}C CSA and $^{1}H-^{13}C$ dipolar relaxation mechanisms, also known as the TROSY effect (Fiala et al., 2000; Riek et al., 2001), or a combination of both (Brutscher and Simorre, 2001).

*To whom correspondence should be addressed: E-mail: Bernhard.Brutscher@ibs.fr

For application to larger RNA, however, increased ^{15}N chemical shift overlap makes unambiguous peak assignment difficult. To overcome this problem, unidirectional HCNCH experiments have been proposed (Farmer et al., 1993; Sklenar et al., 1993b; Tate et al., 1994; Hu et al., 2001) which correlate directly the sugar ^{1}H and ^{13}C with those of the base. Unfortunately, these experiments are even less sensitive than the out-and-back HCN experiment due to the additional transfer steps. An HC(N)C-type out-and-back correlation experiment has also been proposed recently (Furtig et al., 2004). In this pulse scheme magnetization is further transferred from $N_{1/9}$ to the quaternary C_2 or C_4 base carbons, which are frequency labeled to provide the required frequency match between sugar and base correlations. Other alternatives to increase spectral resolution include the preparation of different nucleotide-specific isotope-labeled RNA samples. This solution offers the advantage of reduced spectral overlap and provides additional base type information, but the production of four selectively isotope-labeled RNA samples is costly and labor intensive. In addition, NMR data have to be acquired for each of the 4 samples. Here we present new 3D transverse-relaxation optimized CT-HCN(C) experiments for uniformly ^{13}C, ^{15}N labeled RNA. CT frequency editing provides high spectral resolution in both, the ^{13}C and ^{15}N frequency dimensions. A base-type editing filter further increases spectral resolution and facilitates the assignment of sugar and base HCN correlation peak pairs. The experiment also allows base-type identification without the need of selective isotope labeling. The performance of the new pulse sequence is demonstrated for a 33-nucleotide RNA aptamer in complex with theophylline.

Materials and methods

Sample preparation

The uniformly ^{13}C, ^{15}N-labeled 33-nucleotide RNA aptamer – theophylline complex was prepared as described elsewhere (Zimmermann et al., 1997). Two samples were used for this study containing a 1: 1.2 RNA complex with theophylline at concentrations of about 1.0 and 0.7 mM dissolved in a D_2O solution at pH 6.8 containing 20 mM sodium phosphate, 30 mM NaCl, and 2 mM $MgCl_2$.

NMR data acquisition and processing

NMR experiments were performed on a Varian INOVA 800 spectrometer, equipped with a triple-resonance (^{1}H, ^{13}C, ^{15}N) probe and shielded z-axis gradients. The sample temperature was set to 25 °C. For the 2D CT-H(C)N(C) spectrum shown in (Figure 4), the time variable t_2 was set to $t_2 = 0$. A data set of 2-times (base-type editing filter) 180 (^{15}N) × 512 (^{1}H) complex points were recorded in an overall experimental time of 14 h. The spectral widths were set to 3600 Hz (^{15}N), and 10000 Hz (^{1}H). Additional 2D CT-HC(NC) spectra recorded using either the sensitivity enhanced TROSY (Figure 1a) or the MQ-TROSY version (Figure 1b) of the experiment were recorded with 78 (^{13}C) × 512 (^{1}H) complex points for spectral widths of 5000 Hz (^{13}C), and 10000 Hz (^{1}H) in an experimental time of 1.5 h. No base-type selection was performed, as the aim of this experiment was to compare the relative sensitivity of the two implementations of the CT-HCN(C) experiment. Data processing was performed using the FELIX program version 2000 (Accelrys Inc.). Mirror image linear prediction was applied (Zhu and Bax, 1990) to double the acquisition times of the ^{15}N CT dimension prior to squared cosine apodization, zero-filling and Fourier transformation.

Results and discussion

Figure 1 shows two implementations of the new CT-HCN(C) experiment developed to record high-resolution HCN correlation spectra for resonance assignment of RNA. The flow of magnetization is given as follows:

$$\begin{aligned} &H_{6/8/1'} \xrightarrow{J_{CH}} C_{6/8/1'} \xrightarrow{J_{CN}} N_{1/9}[CT - t_1, base - type\ filter] \\ &\xrightarrow{J_{CN}} C_{6/8/1'}[CT - t_2] \xrightarrow{J_{CH}} H_{6/8/1'}[t_3] \end{aligned} \quad (1)$$

with the active spin–spin couplings indicated above the arrows. Both pulse sequences are of the out-and-back type with the standard coherence transfer pathway of HCN experiments. In (a) TROSY-type (Pervushin et al., 1997) ^{13}C evolution is implemented for both the base and sugar moieties during the $^{13}C \rightarrow {}^{15}N$ transfer steps,

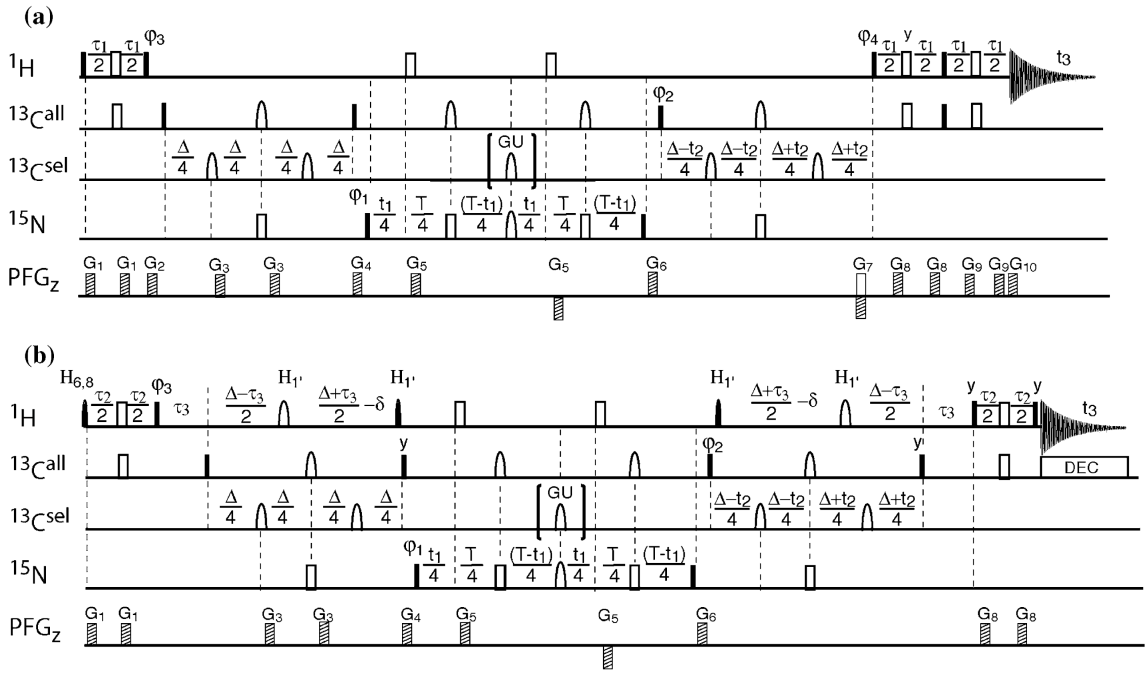


Figure 1. Pulse sequences for base-type-edited (a) TROSY-CT-HCN(C) and (b) MQ-TROSY-CT-HCN(C) experiments of RNA. Two data sets are recorded with and without the GU-labeled pulse in brackets labeled (GU) on the $^{13}C^{sel}$ channel to perform base-type editing. The carrier frequencies are set to 4.7 ppm (^{1}H), 112 ppm ($^{13}C^{all}$), and 153 ppm (^{15}N) throughout the experiment. The pulses applied on the $^{13}C^{sel}$ channel, and the band-selective ^{1}H pulses (denoted $H_{6,8}$ and $H_{1'}$) in (b) are frequency shifted by time-proportional phase incrementation to the center of the desired frequency bands. All radio-frequency (rf) pulses are applied along the x-axis unless indicated. The phases φ_1 and φ_2 are initially set to $\varphi_1 = \varphi_2 = x$. The phase φ_3 is adjusted to y or $-y$ depending on the spectrometer (see text). 90° and 180° rf pulses are represented by filled and open pulse symbols, respectively. Shaped pulse symbols indicate band-selective or composite pulses: the shaped $^{13}C^{all}$ pulses are applied as symmetric composite refocusing pulses of the form $58_x - 140_{-x} - 344_x - 140_{-x} - 58_x$ (Shaka and Pines, 1987) adjusted to cover a band width of 120 ppm. Selective inversion pulses on the $^{13}C^{sel}$ channel applied during the transfer delays δ_2 have an IBURP-2 shape (Geen and Freeman, 1991) and are centered at 72 ppm ($C_{2'}$) and 160 ppm (C^{arom}), both covering a band width of 16 ppm. These two pulse shapes are combined to a single shape by vector addition (Kupce and Freeman, 1993). The $^{13}C^{sel}$ inversion pulse labeled (GU) required for base-type filtering is applied with an i-SNOB-5 shape (Kupce et al., 1995) covering a bandwidth of 5 ppm centered at 153.4 ppm. The selective ^{15}N pulse is applied with an r-SNOB profile (Kupce et al., 1995) centered at 155 ppm and covering a band width of 40 ppm. The $H_{6,8}$ excitation pulse is applied with an e-SNOB profile centered at 7.7 ppm and covering a bandwith of 2.0 ppm. The $H_{1'}$ refocusing and flipback pulses are applied with r-SNOB, and time- and phase-inverted e-SNOB pulse shapes, respectively, centered at 5.6 ppm and covering a bandwith of 1.4 ppm. The transfer delays were set to $\tau_1 = 2.65$ ms, $\tau_2 = 2.5$ ms, $\tau_3 = 2.8$ ms, $\Delta = 30$ ms, and $T = 57$ ms (including pulse lengths of ^{13}C pulses). The delay δ is set to the difference in pulse length between the composite 180° and the 90° $^{13}C^{all}$ pulses, and ensures complete refocusing of $H_{1'}$ transverse coherence. ^{13}C composite decoupling during acquisition is achieved using adiabatic WURST-2 decoupling (Kupce and Freeman, 1996) covering a band width of 120 ppm. Pulsed field gradients G_1–G_{10} are applied along the z-axis (PFG_z) with gradient strengths ranging from 10 to 20G/cm and lengths from 100 to 500 µs, followed by a recovery delay of 100µs. The relative strengths of the gradients G_7 and G_{10} have to be adjusted to the gyromagnetic ratios of ^{1}H and ^{13}C as $G_7/G_{10} = \gamma_H/\gamma_C$. A simple 2-step phase cycle is used with $\varphi_2 = x, -x$, and $\varphi_{rec} = x, -x$. Quadrature detection in t_1 (^{15}N) is obtained by incrementing the phase δ_1 according to STATES-TPPI. In (a) sensitivity-enhanced echo/antiecho quadrature detection in t_2 (^{13}C) is obtained by recording two data sets with the following phase settings: (I) $\varphi_2 = x$, $\varphi_4 = -y$, and (II) $\varphi_2 = y$, $\varphi_4 = y$. In addition the sign of G_7 is changed from minus to plus for the second data set. In (b) quadrature detection in t_2 (^{13}C) is obtained by incrementing the phase φ_2 according to STATES. Pulse sequences in Varian pulse sequence language and transformation protocols in Felix macro language can be obtained from the authors upon request.

whereas the pulse sequence shown in (b) is of the MQ-TROSY type (Brutscher and Simorre, 2001) using $^{1}H-^{13}C$ MQ coherence evolution for the sugar moieties and TROSY-type spin evolution for the base moieties.

In short, the coherence transfer pathway of the TROSY CT-HCN(C) experiment of Figure 1a is given as follows: the excited ^{1}H coherence H_y is transferred to the attached carbon via the $^{1}J_{CH}$ coupling yielding an anti-phase coherence $2H_zC_x$.

The same INEPT sequence also transfers the ^{13}C steady state magnetization C_z into an inphase coherence C_x. Both, inphase and anti-phase coherences are further transferred to the glycosidic nitrogen $N_{1,9}$. The transfer delay δ_2 is set to $2/J_{C6C5} \cong 30$ ms to refocus spin evolution due to the J_{C6C5} scalar coupling in pyrimidine residues. The band-selective inversion pulses applied on the $^{13}C^{sel}$ channel in Figure 1a remove the effects of other carbon–carbon couplings ($J_{C1'C2'}$, $J_{C8C6/4}$, J_{C6C2}). Then, the ^{15}N frequency is edited in a CT manner, with the CT delay T adjusted as explained below to allow the implementation of a base-type filter. ^{1}H and ^{13}C 180° pulses are used during T to refocus J_{NH} and J_{NC} coupling evolution. In addition the small homonuclear J_{N9N3} scalar coupling is refocused by a band-selective $N_{1,9}$ pulse applied at the centre of the CT delay. Finally, in the second half of the pulse sequence magnetization is transferred back to ^{1}H for final detection. Additional CT ^{13}C frequency editing is realized during the $^{15}N \rightarrow {}^{13}C$ back-transfer delay δ_2. No ^{1}H pulses are applied during the $^{13}C \rightarrow {}^{15}N$ transfer steps thus providing TROSY-type signal enhancement for both, the base and the sugar correlations. The double S^3CT transfer element (Pervushin et al., 1997; Sørensen et al., 1997), used for the $^{13}C \rightarrow {}^{1}H$ back-transfer, allows sensitivity-enhanced echo/antiecho quadrature detection in t_2 and selection of the slowly relaxing ^{13}C doublet line. As usual the phase ϕ_3 of the ^{1}H 90° pulse in the initial $^{1}H \rightarrow {}^{13}C$ INEPT transfer is spectrometer dependent and has to be adjusted ($\varphi_3 = y$ or $\varphi_3 = -y$) to add up the signals originating from the ^{1}H and ^{13}C steady-state magnetizations (Brutscher et al., 1998).

The major part of the MQ-TROSY CT-HCN(C) pulse sequence shown in (Figure 1b) is identical to the above-described TROSY version of (Figure 1a). The major modifications concern the $^{1}H \rightarrow {}^{13}C$ out and back-transfer steps. A specially designed pulse sequence element transforms ^{1}H spin polarization into ^{1}H–^{13}C MQ coherence for the sugar moieties, and into a single-transition state for the base moieties as described previously (Brutscher and Simorre, 2001). Additional band-selective $H_{1'}$ pulses are required to refocus ^{1}H chemical shift evolution during the transfer delays τ_3 and Δ, and to create antiphase single-quantum coherence $4N_y\ C_z\ H_z$ for ^{15}N CT frequency editing. A different zero-order phase correction along the ^{1}H dimension is required for the sugar and base moieties to account for the 90° phase shift between the $H_{1'}$ and $H_{6,8}$ coherences (Brutscher and Simorre, 2001) prior to detection. The MQ-TROSY version of the CT-HCN(C) experiment presents the additional advantage that the $^{1}H \rightarrow {}^{13}C$ transfer delays can be adjusted separately for the sugar (τ_3) and base (τ_2) ^{1}H—^{13}C spin pairs. However, it lacks the possibility of sensitivity-enhanced ^{13}C quadrature detection.

An experimental comparison of the two CT-HCN(C) experiments is shown in (Figure 2a–d) showing 2D ^{1}H–^{13}C spectra of the theophylline-apatamer RNA complex recorded at 800 MHz using the pulse sequences of (Figure 1a and 1b). The experiments were performed without base-type filtering, and setting the CT delay to $T = 2$ms, and $t_1 = 0$. The TROSY-only based version yields slightly higher peak intensity for the base moieties, whereas the MQ-TROSY implementation significantly enhances intensities of most cross peaks in the sugar region. However, for some cross peaks with $H_{1'}$ frequencies outside or et the edge of the frequency band covered by the $H_{1'}$ pulses the peak intensity is reduced with respect to the TROSY version, as illustrated for nucleotide U23 (Figures 2b and d).

The principal modifications of the new CT-HCN(C) with respect to existing implementations of the HCN experiment are the addition of a CT ^{15}N editing period and a base-type spectral editing filter. In general, CT frequency editing allows adjusting the spectral resolution $\Delta\ \nu_{res}$ by choosing a constant time delay $T = 1/\Delta\nu_{res}$, albeit at the expense of the overall sensitivity of the experiment. The relative sensitivity of the CT versus non-CT HCN experiment depends on the $N_{1/9}$ transverse relaxation rates which are dominated by the ^{15}N chemical shift anisotropy (CSA) contribution, and the desired spectral resolution in the ^{15}N dimension. Additional base-type spectral editing is realized during the CT delay T. The filter implemented in the pulse sequence of Figure 1 exploits different properties of the involved spin systems: (i) the presence of a scalar coupling between $N_{1,9}$ to a quaternary carbon C_4 (purines) or C_2 (pyrimidines) (see Figure 3a), with a coupling constant of $^{1}J_{N9C4} = {}^{1}J_{N1C2} \approx 20$ Hz for A, G, and U, and $^{1}J_{N1C2} \approx 12$ Hz for C nucleotides (Wijmenga and van Buuren, 1998), (ii) the fact that the quaternary C_4 and C_2 carbons resonate in distinct frequency ranges for the different nucleotides.

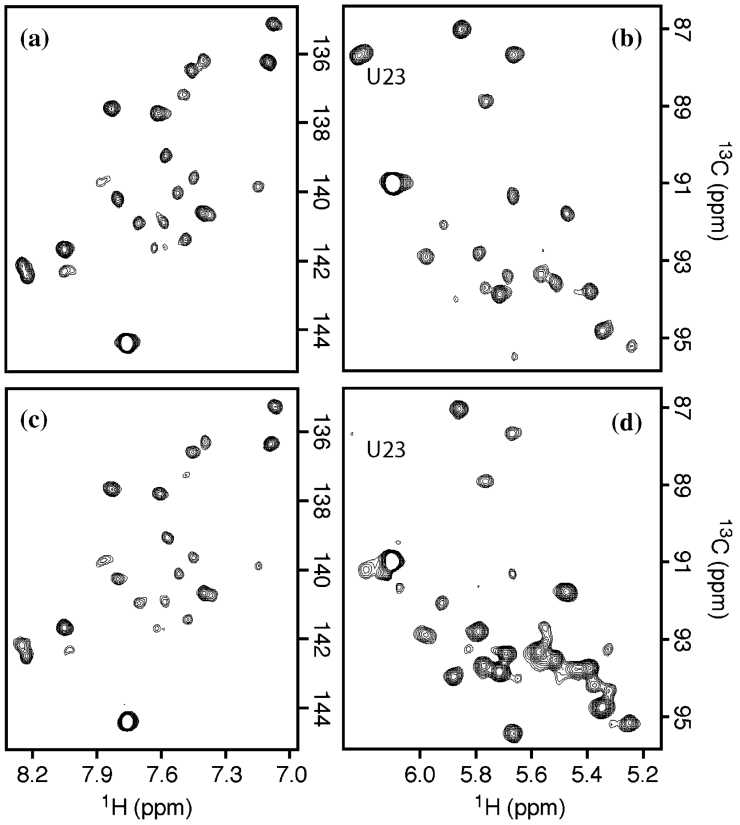


Figure 2. Comparison of 2D $^{1}H-^{13}C$ correlation spectra recorded (a) and (b) using the TROSY CT-HCN(C) sequence of Figure 1a, and (c) and (d) using the MQ-TROSY CT-HCN(C) sequence of Figure 1b without base-type spectral editing and setting $t_1 = 0$. Both spectra were recorded in the same experimental time of 1.5 h. The regions of the spectra corresponding to the H_8-C_8 and H_6-C_6 correlations are shown in (a) and (c), whereas the regions corresponding to the $H_{1'}$-$C_{1'}$ correlations are shown in (b) and (d).

As shown in (Figure 3b) 3 frequency bands can be distinguished, thus allowing for Hadamard-type spectral editing (Hadamard, 1893; Kupce and Freeman, 2003; Van Melckebeke et al., 2004). Because of their overlapping $C_{2,4}$ chemical shift ranges, a frequency-based spectral editing filter does not allow to discriminate between G and U nucleotides. We have designed a filter acting on only two bands corresponding to either G + U or A + C resonances. As will be seen later, this simple 2-step filter is sufficient to discriminate between all four nucleotide types. The filter requires the acquisition of only two data sets. In the first experiment, the pulse denoted GU on the $^{13}C^{sel}$ channel in Figure 1 is omitted, and no scalar $^{1}J_{N9C4}$ or $^{1}J_{N1C2}$ coupling evolution occurs during the CT delay T. In the second experiment, a selective inversion pulse is applied covering the $C_{2,4}$ band of G and U bases (see Figure 3b). This induces an additional $^{1}J_{N9C4}$ and $^{1}J_{N1C2}$ coupling evolution during the delay T for spin systems corresponding to U or G nucleotides, while spin evolution for C and A nucleotides is not changed by this pulse. If the delay T is adjusted to $T = 1/J_{N9C4}(G) = 1/J_{N1C2}(U)$, this results in a sign change of the observed NMR signal for U and G residues. Addition of the two experiments leads to a spectrum containing exclusively HCN correlation peaks from A and C, whereas subtraction provides the corresponding spectrum for G and U nucleotides. Note that the base-type-edited CT-HCN(C) experiment retains the full sensitivity of the CT-HCN experiment of the same duration, because no additional spin evolution period is required, and the full NMR signal is detected (sign-modulated) in

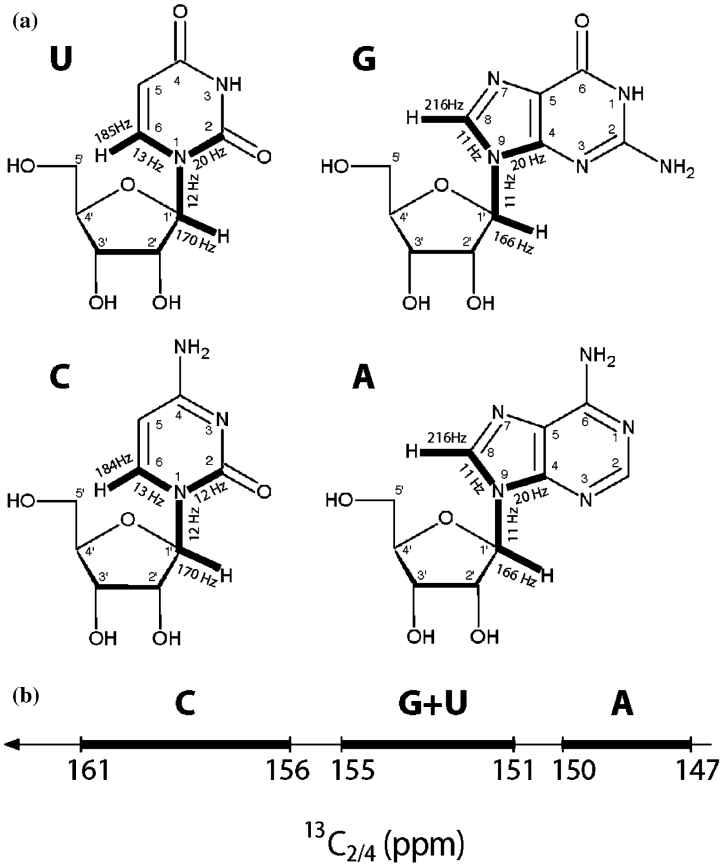


Figure 3. (a) Schematic representation of the 4 RNA nucleosides. Thick lines highlight the nuclei involved in the coherence transfer pathway of the CT-HCN(C) experiment. The coupling constants are given according to Wijmenga and Van Buuren (1998). (b) ^{13}C chemical shift ranges for the quaternary carbons involved in the base-type spectral-editing filter: C_4 for G and A, and C_2 for U and C nucleotides. The frequency bands were defined on the basis of the available chemical shift data for four different RNA, the 33-nucleotide RNA aptamer used in the present study, a 25-nucleotide stem-loop RNA (Fiala et al., 2004), a 14-nucleotide RNA tetraloop (Furtig et al., 2004), and a 16-nucleotide stem loop currently studied in our laboratory.

both experiments, as usual for Hadamard-type spectral editing.

An example of a base-type-edited 2D MQ-TROSY CT-H(C)N(C) correlation spectrum, recorded using the pulse sequence of Figure 1a (with $t_2 = 0$), is shown in Figure 4a and b. The data were acquired at 800 MHz ^1H frequency on a sample of the uniformly ^{13}C, ^{15}N labeled 33-nucleotide RNA aptamer-theophylline complex. The corresponding spectrum recorded with the sequence of Figure 1b is shown in Figure S1 of the Supporting Information. The CT delay has been adjusted experimentally to $T = 57$ ms (including the duration of the $^{13}C^{all}$ and $^{13}C^{sel}$ pulses applied during the CT delay). This corresponds to slightly smaller coupling constants $J_{N9C4}(G) = J_{N1C2}(U) \approx 17.5$ Hz compared to the values (see Figure 3a) given by Wijmenga et al. (1998). This finding is also confirmed by a recent experimental work of Jaroniec et al. (2005). Using this optimized CT delay, the addition and subtraction of the 2 data sets recorded for base-type filtering results in a clean separation of cross peaks from C and A nucleotides (Figure 4a) and from U and G nucleotides (Figure 4b). This is illustrated by 1D traces, extracted at the N_1 frequency of residue C27, shown in the spectra of Figure 4a and b. The high signal to noise ratio in the C/A sub-spectrum, and the absence of any residual peak intensity detected in the U/G sub-spectrum clearly demonstrate the

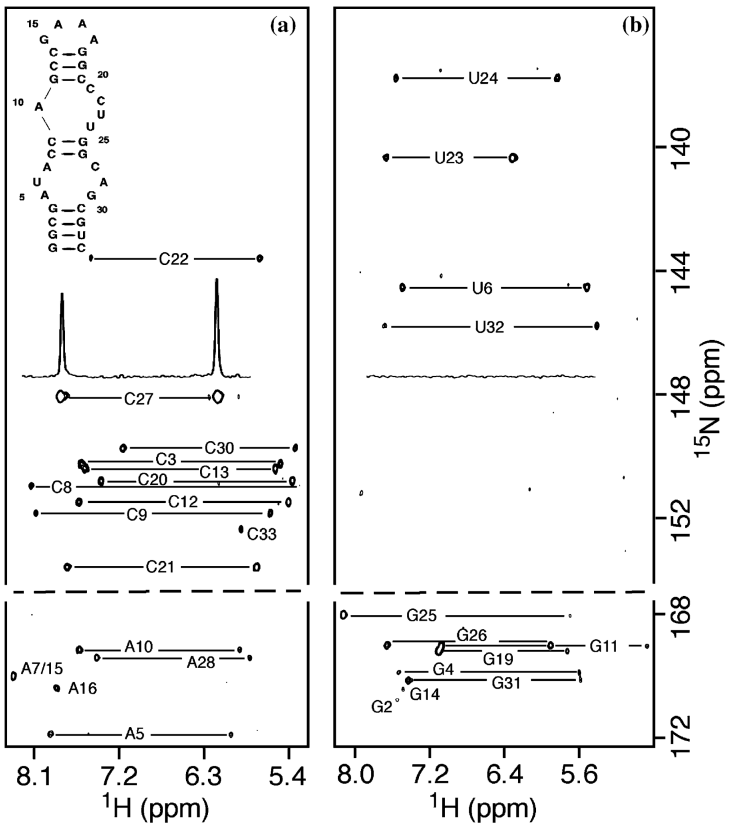


Figure 4. 2D H(C)N correlation spectra recorded using the TROSY CT-HCN(C) pulse sequence of Figure 1a with base-type spectral editing. The data were acquired on a 1.0 mM uniformly $^{13}C,^{15}N$-labeled sample of the 1:1.2 33-nucleotide aptamer-theophylline complex at 800 MHz ^{1}H frequency and 25 °C in an overall experimental time of 14hours. The spectrum shown in (a) contains only cross peaks from C and A nucleotides, whereas the spectrum in (b) shows those from U and G. For simplicity the ^{15}N frequency gap between N_9 (purines) and N_1 (pyrimidines) is not shown. Assigned peak pairs are annotated by their base type and residue number. The 1D ^{1}H traces shown in (a) and (b) were extracted at the ^{15}N frequency of C27 illustrating the good filter performance for this nucleotide. In addition, the secondary structure of the RNA aptamer is drawn in the upper left corner in (a).

good filter performance. This base-type filter is particularly useful to improve spectral resolution in the crowded spectral region corresponding to the overlapping A and G cross peaks, as illustrated for the two nucleotides A10 and G26 resonating at the same $N_{1,9}$ frequency. The long acquisition time in the ^{15}N dimension ($t_1^{max} = 50$ ms) that has been further doubled during processing by mirror image linear prediction (Zhu and Bax, 1990) ensures high spectral resolution ($\Delta\nu_{res} \approx 10$ Hz). Both, the enhanced resolution and the base-type discrimination allows unambiguous assignment of all observed peak pairs for this RNA. Note that the missing cross peaks are also absent (or of very low intensity) from a standard TROSY-HCN spectrum most likely due to conformational exchange line broadening. Additional ^{13}C frequency labeling in the 3D version of the experiment provides the required spectral resolution to discriminate between the sugars and bases of different nucleotides. For the 33-nucleotide RNA aptamer studied here, unambiguous intranucleotide base-sugar connectivity assignment ($H_{1'}$-$C_{1'}$-$H_{6,8}$-$C_{6,8}$) was obtained from a 3D MQ-TROSY CT-HCN(C) data set (data not shown). The sensitivity of TROSY or MQ-TROSY CT-HCN(C) is somewhere in between relaxation-optimized HCN and HC(NC)H experiments. It therefore presents a good compromise between sensitivity and spectral resolution for application to larger RNA molecules at high field NMR spectrometers, especially if a cryogenic probe is available that provides the required higher sensitivity.

The base-type edited CT-TROSY HCN experiment also provides a convenient method for base-type identification, which is crucial for sequential RNA resonance assignment. Pulse sequence elements to select or separate NMR signals originating from different types of nucleotides in RNA have been proposed in the past. For example, the H_6-C_6 correlation peaks of pyrimidines can be separated from H_8-C_8 correlation peaks of purines based on the presence or absence of a scalar $^1J_{CC}$ coupling. The distinct chemical shift ranges of the C_5 carbons in U and C bases can be exploited to further separate the H_6-C_6 correlation peaks (Brutscher et al., 2001). Alternatively, experiments designed to correlate different protons within the same base allow selection of a single base type by exploiting a base-type-specific coherence transfer pathway (Simorre et al., 1995; Simorre et al., 1996a; Simorre et al., 1996b). The spectral sediting filter in the CT-HCN(C) experiments proposed here creates two sub-spectra. Signals originating from C and A nucleotides are detected in one sub-spectrum, while signals arising from U and G nucleotides are observed in the second sub-spectrum (Figure 4). For each spectrum, the distinct ^{15}N chemical shift ranges for N_9 (A, G) and N_1 (C, U) allow to further distinguish between purine and pyrimidine cross peaks. Therefore, as long as the C_2 of pyrimidines and C_4 of purines resonate within the frequency bands (Figure 3b) used for the filtering, the base-type edited CT-HCN(C) experiment allows discrimination between all four nucleotides. The performance of the base-type filter is further demonstrated by application to a second RNA. Figure S2 of the Supporting Information shows 2D ^1H–^{15}N spectra recorded with the pulse sequence of Figure 1b for a 16-nucleotide ^{15}N,^{13}C labeled stem loop RNA in complex with a 16-nucleotide unlabeled stem-loop RNA. Again a clean separation of cross peaks from the four different nucleotide types is obtained. Therefore, in view of the quite different secondary and tertiary structures of the two RNA molecules, one may expect similar results for other RNA samples. The CT-HCN(C) experiment is to the best of our knowledge the first spectroscopic method providing base-type identification of base ($H_{6/8}$, $C_{6/8}$) and sugar ($H_{1'}$, $C_{1'}$) resonances within a single experiment.

In summary, we have shown that highly resolved HCN correlation spectra can be recorded using new transverse relaxation optimized CT-HCN(C) pulse sequences including a base-type spectral editing filter. The base-type filter not only increases spectral resolution, but it also provides a simple way of base-type identification, an information which otherwise is difficult to obtain. Therefore we expect that this experiment will become a widespread tool for NMR studies of larger RNA molecules. This pulse scheme may prove equally useful for application to DNA provided that the constant time delay T and the frequency band of the inverted quaternary ^{13}C are properly adjusted.

Supporting information available

One figure showing 2D ^1H–^{15}N correlation spectra of the 33-nucleotide aptamer-theophylline complex recorded with the MQ version of the CT-HCN(C) pulse sequence of Figure 1b with base-type spectral editing. One figure showing the 2D base-type edited ^1H–^{15}N correlation spectra of a 16 nucleotide ^{13}C,^{15}N-labeled RNA in complex with a 16 nucleotide unlabeled RNA. This material is available in electronic form at http://dx.doi.org/10.1007/s10858-005-8872-1.

Acknowledgements

We thank Ad Bax for the sample of a complex formed by two RNA stem loops, and Marella Canny for sample preparation of the 33-nucleotide aptamer. This work was supported by the Commissariat à l'Energie Atomique, the Centre National de la Recherche Scientifique, and NIH (AI33098). H.V.M. acknowledges the receipt of a fellowship from the C.E.A. J.B. acknowledges support from the HFSP Organization.

References

Brutscher, B., Boisbouvier, J., Kupce, E., Tisne, C., Dardel, F., Marion, D. and Simorre, J.P. (2001) *J. Biomol. NMR*, **19**, 141–151.

Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. and Simorre, J.P. (1998) *J. Am. Chem. Soc.*, **120**, 11845–11851.

Brutscher, B. and Simorre, J.P. (2001) *J. Biomol. NMR*, **21**, 367–372.

Duchardt, E., Richter, C., Ohlenschlager, O., Gorlach, M., Wohnert, J. and Schwalbe, H. (2004) *J. Am. Chem. Soc.*, **26**, 1962–1970.

Farmer, B.T., Muller, L., Nikonowicz, E.P. and Pardi, A. (1994) *J. Biomol. NMR*, **4**, 129–133.
Farmer, B.T., Müller, L., Nikonowitz, E.P. and Pardi, A. (1993) *J. Am. Chem. Soc.*, **115**, 11040–11041.
Fiala, R., Czernek, J. and Sklenar, V. (2000) *J. Biomol. NMR*, **16**, 291–302.
Fiala, R., Jiang, F. and Sklenar, V. (1998) *J. Biomol. NMR*, **12**, 373–383.
Fiala, R., Munzarova, M.L. and Sklenar, V. (2004) *J. Biomol. NMR*, **29**, 477–490.
Furtig, B., Richter, C., Bermel, W. and Schwalbe, H. (2004) *J. Biomol. NMR*, **28**, 69–79.
Furtig, B., Richter, C., Wohnert, J. and Schwalbe, H. (2003) *Chembiochem.*, **4**, 936–962.
Geen, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93–141.
Hadamard, J. (1893) *Bull. Sci. Math.*, **17**, 240–248.
Hu, W.D., Gosser, Y.Q., Xu, W.J. and Patel, D.J. (2001) *J. Biomol. NMR*, **20**, 167–172.
Jaroniec, C.P., Boisbouvier, J., Tworowska, I., Nikonowicz, E.P. and Bax, A. (2005) *J. Biomol. NMR*, **31**, 231–241.
Kupce, E., Boyd, J. and Campbell, I.D. (1995) *J. Magn. Reson.*, **106**, 300–303.
Kupce, E. and Freeman, R. (1993) *J. Magn. Reson.*, **105A**, 234–238.
Kupce, E. and Freeman, R. (1996) *J. Magn. Reson.*, **118A**, 299–303.
Kupce, E. and Freeman, R. (2003) *J. Magn. Reson.*, **162**, 158–165.
Marino, J.P., Diener, J.L., Moore, P.B. and Griesinger, C. (1997) *J. Am. Chem. Soc.*, **119**, 7361–7366.
Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.
Ravindranathan, S., Kim, C.H. and Bodenhausen, G. (2003) *J. Biomol. NMR*, **27**, 365–375.
Riek, R., Pervushin, K., Fernandez, C., Kainosho, M. and Wuthrich, K. (2001) *J. Am. Chem. Soc.*, **123**, 658–664.
Shaka, A.J. and Pines, A. (1987) *J. Magn. Reson.*, **71**, 495–503.
Simorre, J.P., Zimmermann, G.R., Mueller, L. and Pardi, A. (1996a) *J. Biomol. NMR*, **7**, 153–156.
Simorre, J.P., Zimmermann, G.R., Mueller, L. and Pardi, A. (1996b) *J. Am. Chem. Soc.*, **118**, 5316–5317.
Simorre, J.P., Zimmermann, G.R., Pardi, A., Farmer, B.T. and Mueller, L. (1995) *J. Biomol. NMR*, **6**, 427–432.
Sklenar, V., Peterson, R., Rejante, M. and Feigon, J. (1993a) *J. Biomol NMR*, **3**, 721–727.
Sklenar, V., Peterson, R.D., Rejante, M.R., Wang, E. and Feigon, J. (1993b) *J. Am. Chem. Soc.*, **115**, 12181–12182.
Tate, S., Ono, A. and Kainosho, M. (1994) *J. Am. Chem. Soc.*, **116**, 5977–5978.
Van Melckebeke, H., Simorre, J.P. and Brutscher, B. (2004) *J. Am. Chem. Soc.*, **126**, 9584–9591.
Wijmenga, S.S. and Buuren, B.N.M.van (1998) *Prog. in Nucl. Magn. Reson. Spectroc.*, **32**, 287–387.
Yan, J.L., Corpora, T., Pradhan, P. and Bushweller, J.H. (2002) *J. Biomol. NMR*, **22**, 9–20.
Zhu, G. and Bax, A. (1990) *J. Magn. Reson.*, **90**, 405–410.
Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P. and Pardi, A. (1997) *Nat. Struct. Biol.*, **4**, 644–649.

III. Bilan et perspectives

L'implémentation d'un filtre Hadamard par bandes durant une période d'édition à temps constant a été montrée sur deux exemples. L'usage du filtre suppose que les conditions expérimentales offrent la sensibilité nécessaire pour l'utilisation d'une édition à temps constant plutôt qu'une édition en temps réel. Le temps expérimental minimal nécessaire pour réaliser l'expérience augmente par rapport au temps minimal requis pour faire l'expérience à temps constant sans le filtre, d'un facteur égal au gain de résolution obtenu (ici, quatre ou deux). En revanche, l'implémentation du filtre se fait sans perte de sensibilité à temps expérimental donné.

L'intérêt d'implémenter un tel filtre dans une séquence peut se situer à deux niveaux. Tout d'abord, la résolution gagnée grâce à la dimension spectrale offerte par le filtre peut éviter l'enregistrement d'une dimension fréquentielle (complète) supplémentaire. Une expérience 2D filtrée peut par exemple remplacer une expérience 3D, ce qui réduit le temps d'acquisition d'un ordre de grandeur. Ce gain en temps est très intéressant dans le contexte de l'étude RMN de molécules instables, de l'étude de processus dynamiques, ou de recherche d'interactions avec des ligands à haut débit.

Le deuxième intérêt concerne l'étude de molécules de taille importante, pour lesquelles la corrélation de trois noyaux n'est pas suffisante pour résoudre les ambiguïtés d'attribution des pics spectraux. Dans ce cadre, l'implémentation d'un ou deux filtres Hadamard permet d'ajouter une ou deux dimensions spectrales supplémentaires pour séparer les pics de corrélations dans un nombre discret de sous-spectres. L'implémentation d'expériences 4D ou 5D est dans ce cas envisageable dans un temps d'expérience n'excédant pas quelques jours. Le gain de résolution obtenu rend possible l'étude des systèmes de plus en plus complexes.

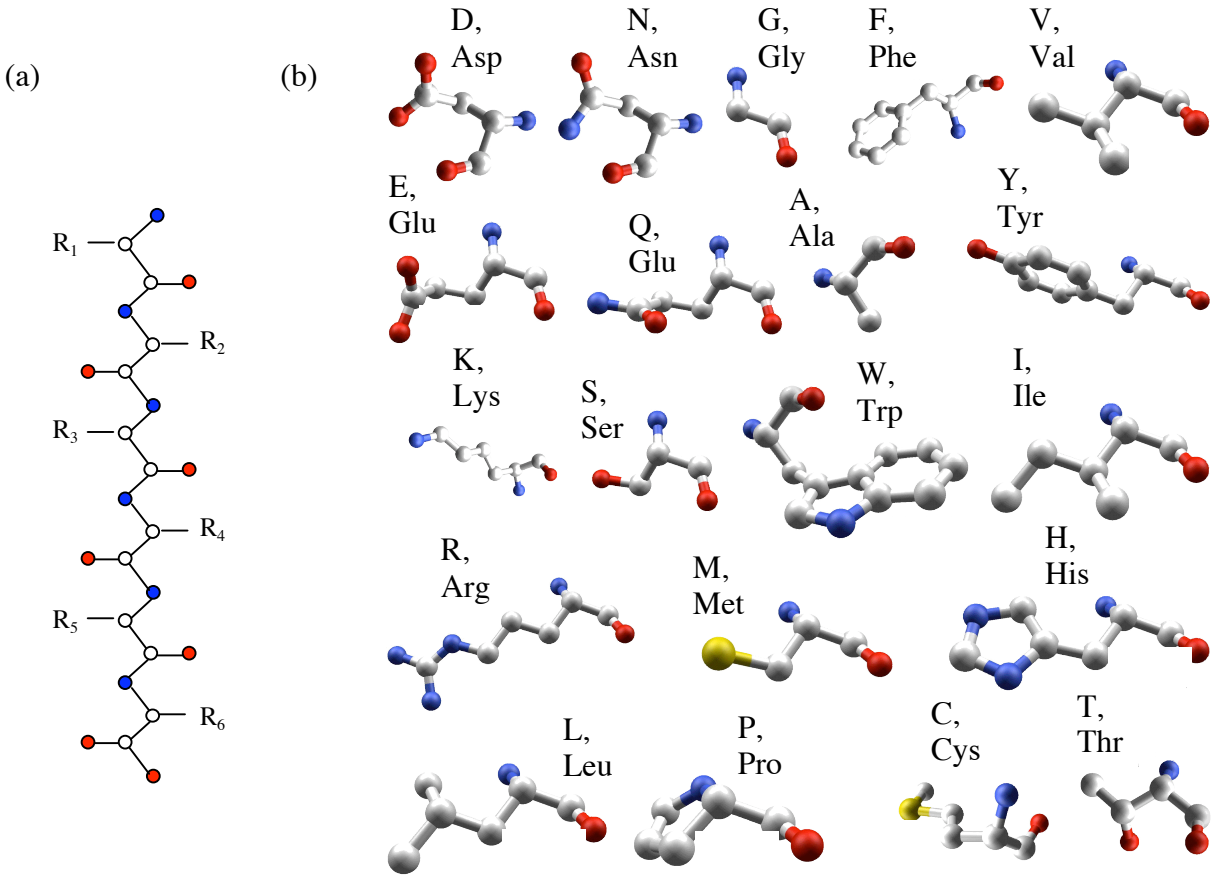
Enfin, l'intérêt des deux filtres que nous avons présentés ne se limite pas à l'augmentation de la résolution des spectres. En effet, l'information concernant le type d'acide aminé ou le type de base constitue une aide précieuse dans les processus d'attribution des résonances. Ces filtres spectroscopiques permettent de s'affranchir d'une étape de marquage isotopique spécifique, qui est coûteuse en termes de préparation biochimique et en temps expérimental. De plus, les informations spectrales sont obtenues sur un unique échantillon.

Les expériences qui ont été présentées sont actuellement utilisées de façon routinière au laboratoire de RMN de l'IBS. L'attribution et le repliement de quatre protéines (un dimère de 2*122 acides aminés, et trois protéines de 83, 100 et 128 acides aminés, respectivement) ont en effet déjà profité de la séparation des spectres des méthyles selon le type d'acide aminé. L'attribution de deux ARN de 16 nucléotides a également été effectuée en bénéficiant des expériences HCN filtrées selon le type de base.

Le principe du filtrage Hadamard par bandes pourra sans doute trouver d'autres applications dans le contexte de l'étude à haute résolution des protéines ou des acides nucléiques. L'implémentation de nouveaux filtres peut être envisagée par exemple dans les expériences de type HCC pour les bases des acides nucléiques, permettant d'augmenter la résolution et d'attribuer les quatre types de bases. Dans le domaine de l'attribution des protéines, il serait également intéressant d'implémenter des expériences de type 4D ou 5D, dont deux des dimensions seraient des filtres Hadamard par bandes bien choisis. Ces idées sont en cours d'expérimentation au laboratoire.

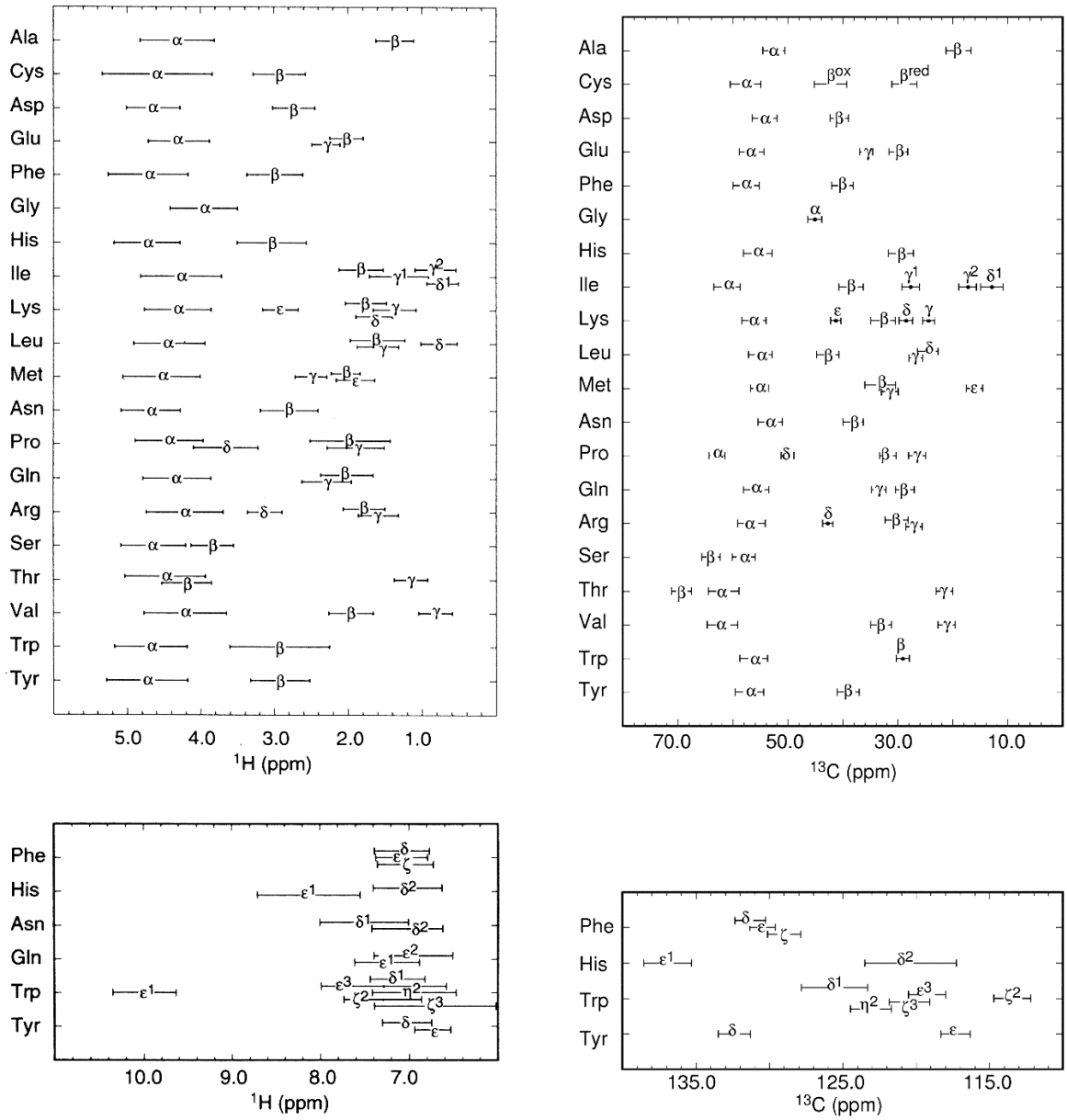
Annexes

Annexe 1
Les protéines



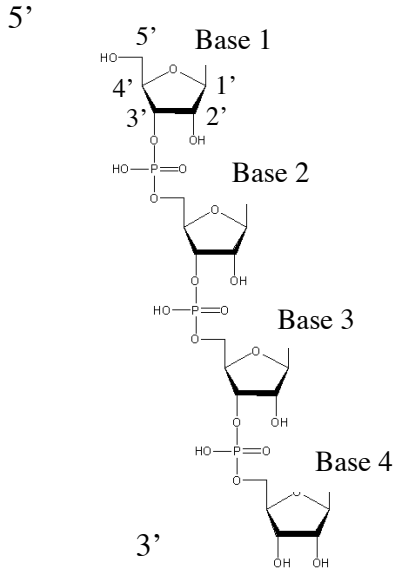
Composition chimique des protéines.

Les atomes de carbone sont représentés en blanc, les atomes d'azote en bleu, les atomes d'oxygène en rouge, et les atomes de soufre en jaune. Les atomes d'hydrogène ne sont pas représentés. (a) Représentation de la chaîne principale d'une protéine de six acides aminés. Les chaînes latérales R sont propres à chaque acide aminé. (b) Les vingt acides aminés naturels. On retrouve dans chaque acide aminé le motif de base de la chaîne peptidique : N-C-CO. On peut noter que les chaînes latérales des différents acides aminés ont des propriétés chimiques différentes en ce qui concerne l'hydrophobicité, l'encombrement stérique, l'acidité, et la polarité.

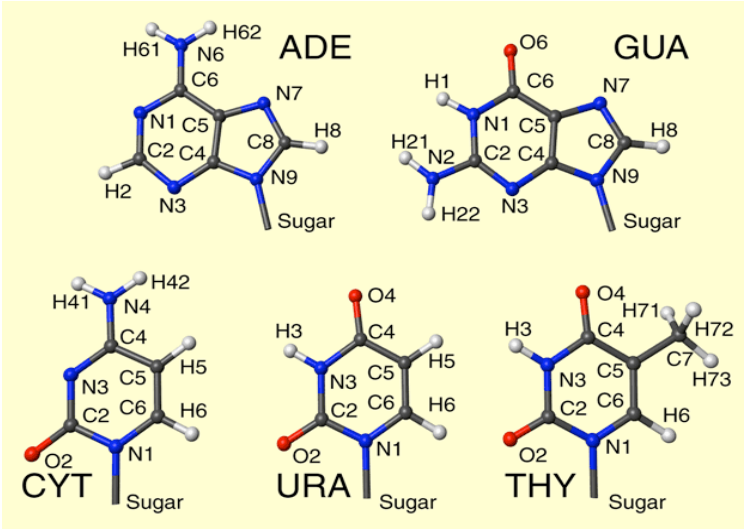


Liste des déplacements chimiques usuels des noyaux ^{1}H, ^{13}C trouvés dans les protéines pour les différents acides aminés.

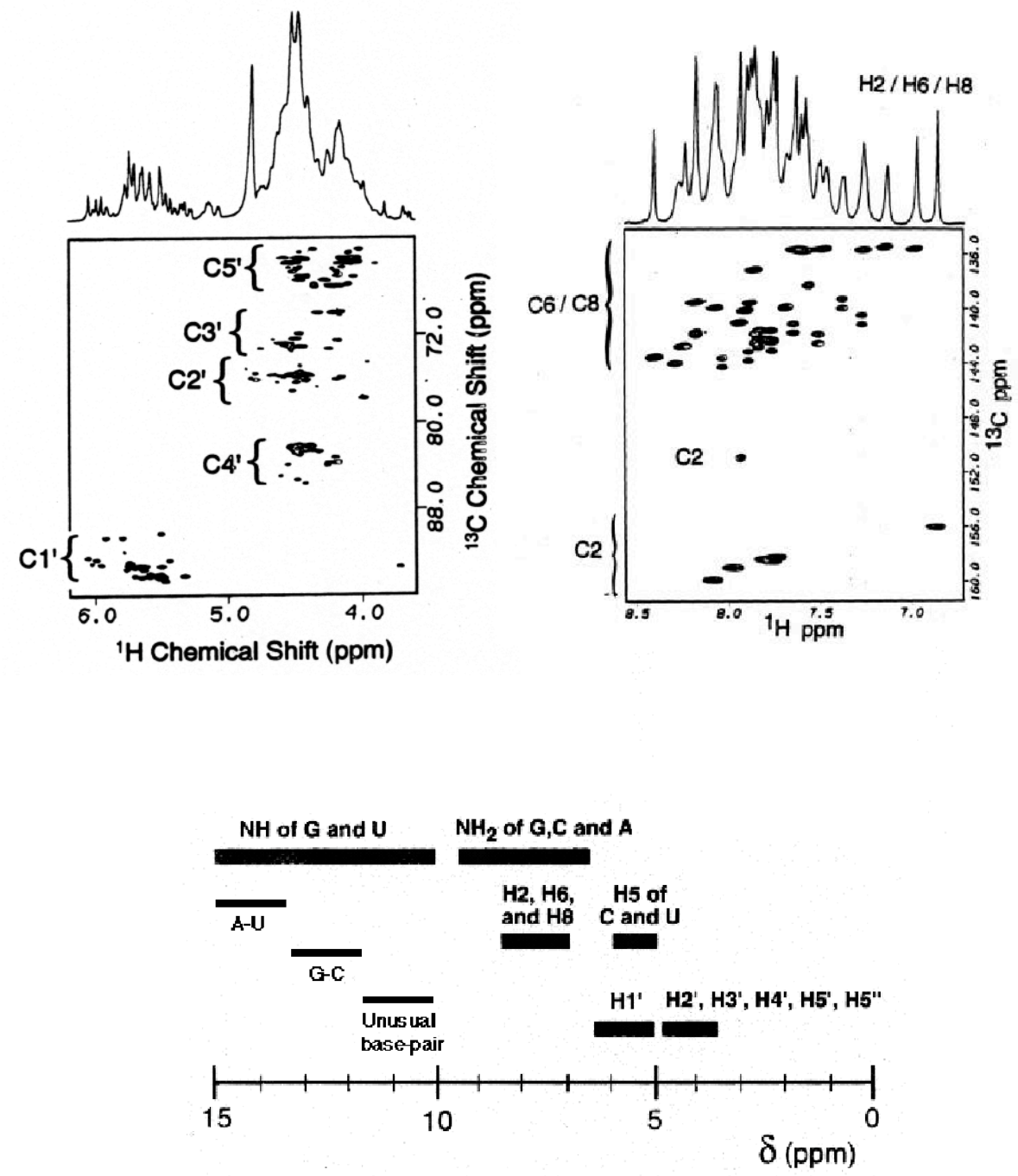
Les acides nucléiques



Chaîne principale d'un acide nucléique (ici un ARN) formé par un enchaînement de quatre nucléotides. Les sucres des différents mucléotides sont reliés par des liaisons phosphodiesters. Dans l'ADN, le groupement OH situé sur le carbone 2' du sucre est remplacé par un proton.

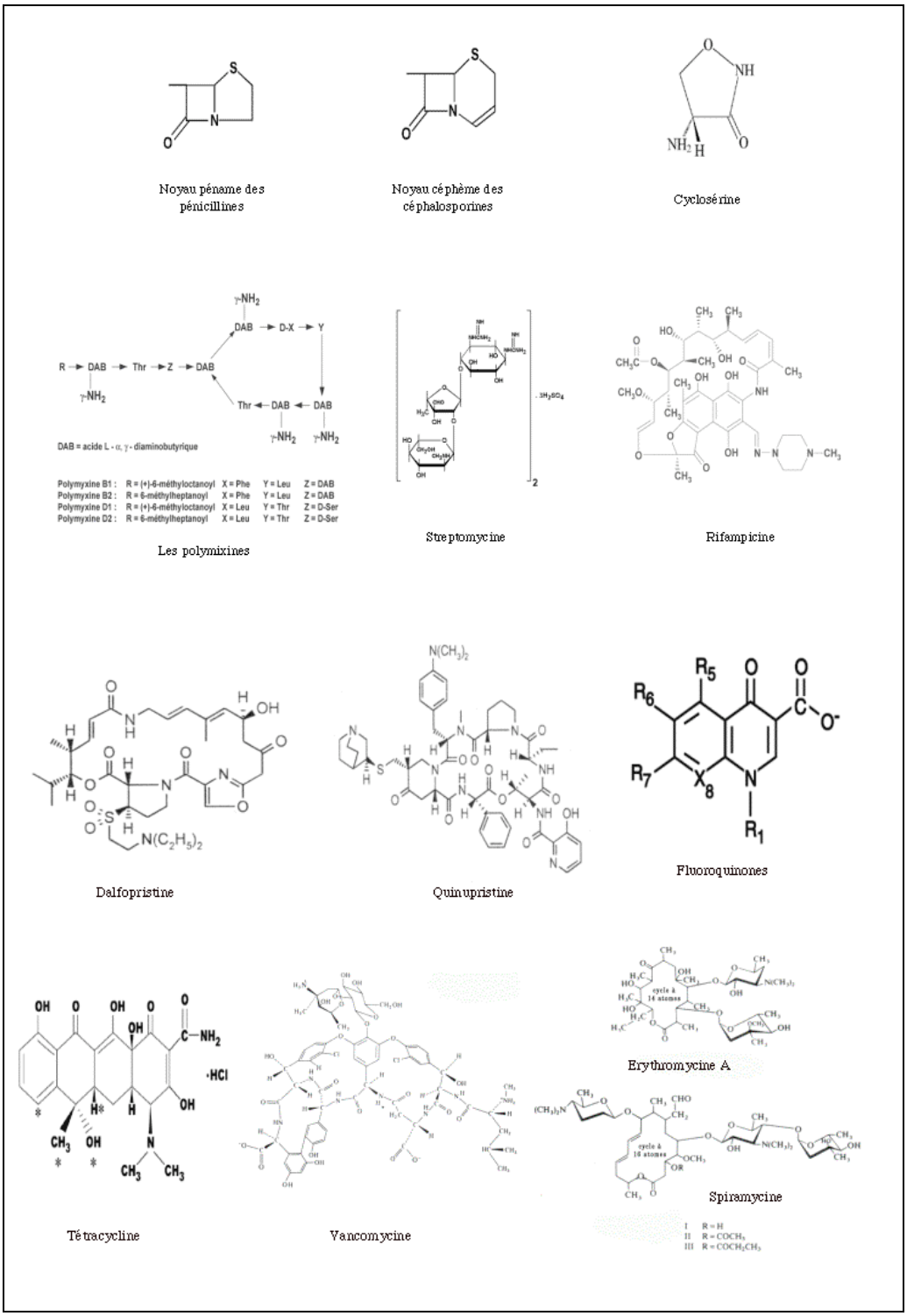


Les cinq bases naturelles: l'Adénosine (Ade ou A), la Guanosine (Gua ou G), la Cytosine (Cyt ou C) et l'Uracyle (Ura ou U). Les bases présentes dans l'ADN sont A, T, C et G, alors que dans l'ARN, on trouve les bases A, U, C et G.



Exemples de déplacements chimiques typiques des noyaux ^{1}H, ^{13}C dans les ARN.

Annexe 2
Les antibiotiques



Structure chimique de divers antibiotiques

Classe chimique	Exemples	Source biologique	Spectre d'action	Mode d'action
Beta-lactame (pénicillines et céphalosporines)	Pénicilline G, Céphalothine	*Penicillium notatum* et *Cephalosporium*	Bactéries Gram-positives	Bloque certaines étapes de la synthèse de la paroi cellulaire (peptidoglycane, muréine)
Pénicilline semi-synthétique	Ampicilline, Amoxycilline		Bactéries Gram-positives et Gram-négatives	Bloque certaines étapes de la synthèse de la paroi cellulaire (peptidoglycane, muréine)
Acide Clavulanique	Clavamox : acide clavulanique + amoxycilline	*Streptomyces clavuligerus*	Bactéries Gram-positives et Gram-négatives	Inhibiteur des beta-lactamases
Monobactame	Aztréoname	*Chromobacter violaceum*	Bactéries Gram-positives et Gram-négatives	Bloque certaines étapes de la synthèse de la paroi cellulaire (peptidoglycane, muréine)
Carboxypénème	Imipénème	*Streptomyces cattleya*	Bactéries Gram-positives et Gram-négatives	Bloque certaines étapes de la synthèse de la paroi cellulaire (peptidoglycane, muréine)
Aminoglycosides	Streptomycine	*Streptomyces griseus*	Bactéries Gram-positives et Gram-négatives	Bloque la traduction (synthèse protéique)
	Gentamicine	*Micromonospora*	Bactéries Gram-positives et Gram-négatives (*Pseudomonas*)	Bloque la traduction (synthèse protéique)
Glycopeptides	Vancomycine	*Streptomyces orientales*	Bactéries Gram-positives (*Staphylococcus aureus*)	Bloque certaines étapes de la synthèse de la paroi cellulaire (peptidoglycane, muréine)
Lincomycines	Clindamycine	*Streptomyces lincolnensis*	Bactéries Gram-positives et Gram-négatives (Bactéroides anaérobies)	Bloque la traduction (synthèse protéique)
Macrolides	Erythromycine	*Streptomyces erythreus*	Bactéries Gram-positives, Gram-négatives (*Neisseria*,*Legionella*, *Mycoplasma*)	Bloque la traduction (synthèse protéique)
Polypeptides	Polymyxine	*Bacillus polymyxa*	Bactéries Gram-négatives	Dégrade la membrane cytoplasmique
	Bacitracine	*Bacillus subtilis*	Bactéries Gram-positives	Bloque certaines étapes de la synthèse de la paroi cellulaire (peptidoglycane, muréine)
Polyènes	Amphotéricine	*Streptomyces nodosus*	Champignons	Inactive les membranes contenant des stérols
	Nystatine	*Streptomyces noursei*	Champignons (*Candida*)	Inactive les membranes contenant des stérols
Rifamycines	Rifampicine	*Streptomyces mediterranei*	Bactéries Gram-positives et Gram-négatives (*Mycobacterium tuberculosis*)	Bloque la transcription (ARN polymérase eubactérienne)
Tétracyclines	Tétracycline	*Streptomyces*	Bactéries Gram-positives et Gram-négatives (*Rickettsias*)	Bloque la traduction (synthèse protéique)
Tétracycline semi-synthétique	Doxycycline		Bactéries Gram-positives et Gram-négatives (*Rickettsias Ehrlichia*, *Borellia*)	Bloque la traduction (synthèse protéique)
Chloramphénicol	Chloramphénicol	*Streptomyces venezuelae*	Bactéries Gram-positives et Gram-négatives	Bloque la traduction (synthèse protéique)

Exemples d'antibiotiques avec leur origine, leurs cibles et modes d'action (informations extraites du site : http://www.bact.wisc.edu/microtextbook/Controlgrowth/antibiotic.html)

doi:10.1016/j.jmb.2003.09.005 **Annexe 3** *J. Mol. Biol.* (2003) **333**, 711–720

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Solution Structural Study of BlaI: Implications for the Repression of Genes Involved in β-Lactam Antibiotic Resistance

Hélène Van Melckebeke[1]†, Christelle Vreuls[2,3]†, Pierre Gans[1] Patrice Filée[2], Gabriel Llabres[3], Bernard Joris[2] and Jean-Pierre Simorre[1]*

[1]*Institut de Biologie Structurale-Jean-Pierre Ebel CEA-CNRS-UJF, 41 Avenue Jules Horowitz, 38027 Grenoble Cedex 1, France*

[2]*Centre d'ingénierie des protéines, Institut de Chimie B6A, Université de Liège Sart-Tilman B4000, Belgium*

[3]*Laboratoire de Physique Expérimentale, Département de Physique, Bat B5, Université de Liège, Sart Tilman B4000 Belgium*

**Corresponding author*

β-Lactamase and penicillin-binding protein PBP2′ mediate staphylococcal resistance to β-lactam antibiotics, which are otherwise highly clinically effective. Two repressors (BlaI and MecI) regulate expression of these inducible proteins. Here, we present the first solution structure of the 82 amino acid residue DNA-binding domain of *Bacillus licheniformis* BlaI which is very similar in primary sequence to the medically significant Staphyloccocal BlaI and MecI proteins. This structure is composed of a compact core of three α-helices and a three-stranded β-sheet typical of the winged helix protein (WHP) family. The protein/DNA complex was studied by NMR chemical shift comparison between the free and complexed forms of BlaI. Residues involved in DNA interaction were identified and a WHP canonical model of interaction with the operators is proposed. In this model, specific contacts occur between the base-pairs of the TACA motif and conserved amino acid residues of the repressor helix H3. These results help toward understanding the repression and induction mechanism of the genes coding for β-lactamase and PBP2′.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: β-lactamase repressor BlaI; NMR spectroscopy; winged helix protein; protein–DNA interaction; antibiotic resistance

Introduction

To date, the nosocomial spread of the bacterial strains resistant to antibiotics has become a major public health concern.[1,2] Many strains have become resistant to β-lactam antibiotics, by either (i) the expression of a specific hydrolase, the β-lactamase, which inactivates β-lactam antibiotics by hydrolyzing their endocyclic amide bond, or (ii) the production of an alternative transpeptidase penicillin-binding protein (PBP2′), insensitive to penicillin inhibition.[3] Although β-lactamase and PBP2′ have different structures and functions, their synthesis is regulated by a similar repression mechanism.

The expression of the β-lactamase, BlaZ in *Staphylococcus aureus* or BlaP in *Bacillus licheniformis* 749/I, is under the control of three genes: *blaI*, *blaR*1 and *blaR*2. The first two genes are clustered in an operon and form a divergon with the *blaZ*/*blaP* gene (Figure 1A). The *blaR*2 gene is not linked to the divergon, and it has not been identified yet.[4] The BlaR1 membrane-bound protein is a penicillin receptor, which detects the presence of the β-lactam antibiotic outside the cell.[5] BlaI is a cytoplasmic protein, which specifically recognizes DNA operator sequences in the intergenic region between the *blaP*/*blaZ* and *blaI-blaR1* genes (Figure 1A). In absence of a β-lactam antibiotic, BlaI is bound to its operator sequence as a dimer and acts as a repressor for the *blaP*/*blaZ* gene expression. The presence of β-lactam antibiotics leads to the acylation of BlaR1[5] and in turn to the transmission of an intracellular signal whose final target is BlaI itself. Consequently, BlaI is released from its operator and this leads to a high level of β-lactamase production. The *S. aureus* and

Supplementary data associated with this article can be found at doi: 10.1016/j.jmb.2003.09.005

† H.V.M. and C.V. contributed equally to this work.

Abbreviations used: WHP, winged helix protein.

E-mail address of the corresponding author: jean-pierre.simorre@ibs.fr

0022-2836/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.

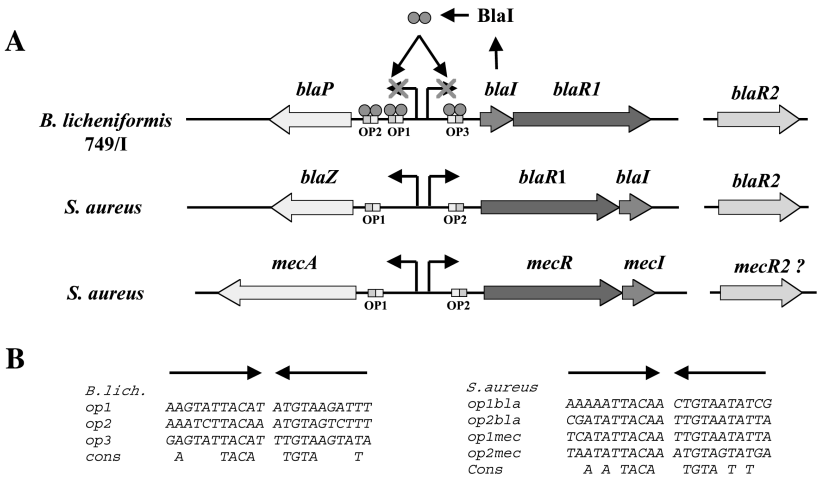


Figure 1. Schematic representation of *bla* and *mec* operon organization in *B. licheniformis* and *S. aureus*. A, The *mec* and *bla* regulators, *mecR1-mecI* and *blaR1-blaI*, respectively, encode inducer-repressor systems with a high level of sequence similarity. The *S. aureus* BlaI and MecI repressors are 60% identical when compared to each other and 31–41% identical when compared to the *B. licheniformis* BlaI repressor. The percentages of highly conserved amino acid residues in the transducers are 30 for the *S. aureus* BlaR–*S. aureus* MecR pair, 21 for the *B. licheniformis* BlaR–*S. aureus* BlaR pair and 23 for the *B. licheniformis* BlaR–*S. aureus* MecR pair.[8] B, Alignment of *B. licheniformis bla*, *S. aureus bla*, and *S. aureus mec* operators.

B. licheniformis BlaI repressors are known to form dimers and multimers in solution, and are composed of two domains.[6] The N-terminal domain (BlaI-NTD) is involved in the DNA recognition and the C-terminal domain (BlaI-CTD) is responsible for BlaI dimerization. The two domains of the *B. licheniformis* BlaI repressor can be separated by papain digestion. BlaI-NTD is still able to bind its operator with reduced affinity, but has completely lost its ability to dimerize.[7]

In the methicillin-resistant *S. aureus* strain, the low-affinity penicillin-binding protein PBP2′ (or PBP2a or MRSA) is encoded by the *mecA* gene. Production of this protein is regulated by similar sensory-transducer and repressor proteins as β-lactamase production. The presence of MecR2 is also postulated in the induction mechanism of *mecA* (Figure 1(A)). The *S. aureus* BlaI and MecI repressors are 60% identical when compared to each other and 31–41% identical when compared to the *B. licheniformis* BlaI repressor, respectively. The percentages of highly conserved amino acid residues in the transducers are 30 for the *S. aureus* BlaR, *S. aureus* MecR pair, 21 for the *B. licheniformis* BlaR, *S. aureus* BlaR pair and 23 for the *B. licheniformis* BlaR, *S. aureus* MecR pair.[8] To date, no structural data concerning BlaI, MecI or homologous proteins have been reported.

DNase footprinting experiments with the BlaI and MecI proteins revealed that in the corresponding strains the different repressors recognize 22 base-pair long symmetry dyads, which exhibit a high degree of similarity (Figure 1B). Furthermore, *S. aureus* BlaI and MecI were shown to corepress PBP2′ and β-lactamase production.[9,10] These studies suggest that BlaI and MecI share the same structural properties concerning DNA interaction. The inactivation mechanism is, however, different for the two repressors, allowing a fine regulation process.[9,11]

To elucidate the BlaI/MecI interaction with its DNA operator sequence, and its mechanism of inactivation during the induction process, the structural study of the *B. licheniformis* BlaI has been undertaken by heteronuclear NMR spectroscopy. Here, we report (i) the high-resolution solution structure of the *B. licheniformis* BlaI-NTD, (ii) the characterization of the BlaI interaction zone with its DNA operator by chemical shift mapping, (iii) a general DNA interaction model describing the interaction of BlaI/MecI with their DNA operators. These results provide important insights about the regulation mechanism of the genes encoding β-lactamase and PBP2′.

Results and Discussion

BlaI repressor belongs to the WHP family

The full-length *B. licheniformis* BlaI protein was cleaved using papain digestion. The position of the unique cleavage site, unambiguously obtained from N-terminal sequencing and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was localized between the residues Ser82 and His83. The full-length BlaI protein as well as the N-terminal domain 1–82 (BlaI-NTD) were studied by heteronuclear NMR. The similarity between the BlaI-NTD and BlaI ^{15}N-heteronuclear single quantum coherence (HSQC) spectra (Figure 2) demonstrates

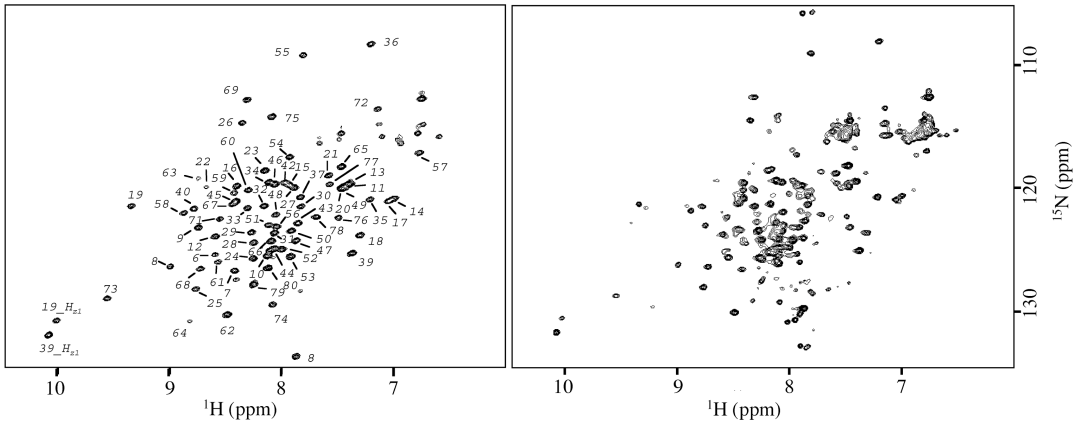


Figure 2. Comparison of the ^{15}N HSQC spectra at 25 °C of BlaI N-terminal domain (BlaI-NTD) and of full-length BlaI. Left: BlaI-NTD spectrum with assignment obtained for residues 6–81. Right: spectrum of the complete protein BlaI. Except for residue 81, all BlaI-NTD resonances are observed in the same positions, indicating that the N-terminal part of the protein forms an independent and structured domain. The resonances corresponding to the C-terminal domain (83–128) are poorly resolved and are mainly located in a region characteristic of extended structure.

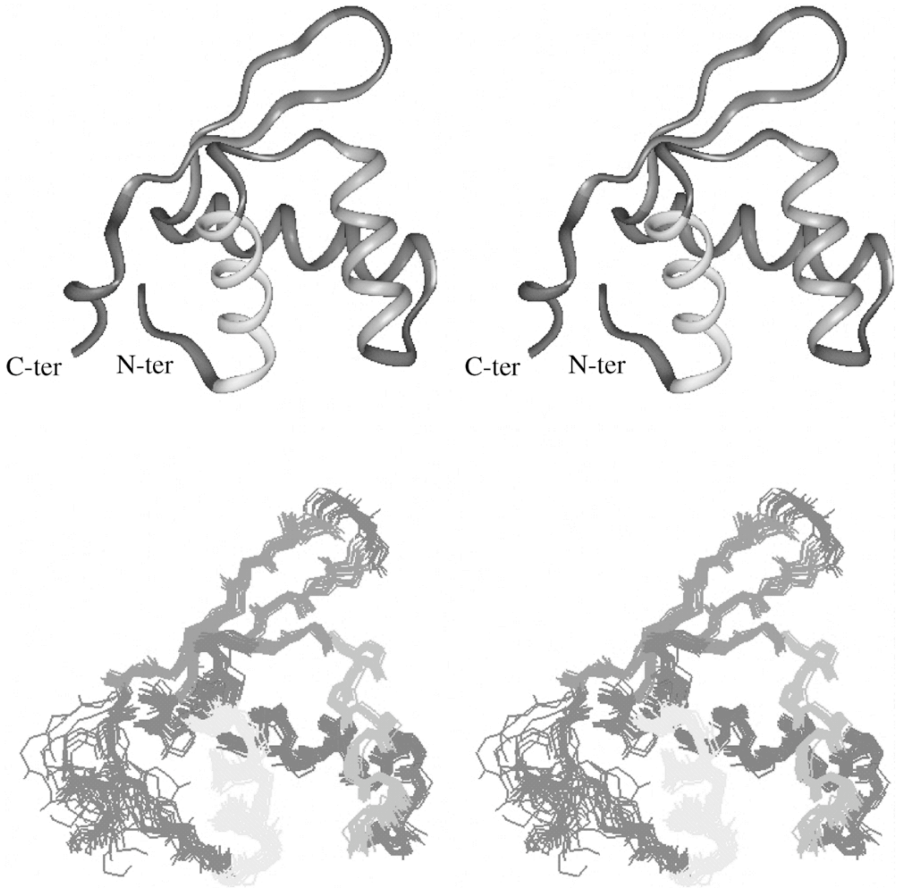


Figure 3. NMR structures of BlaI-NTD (1–82). Secondary structures are colored: H1 in yellow, H2 in orange, H3 in red and the three strands of the β-sheet in blue. Top: stereo view of the ribbon backbone diagram of the lowest-energy structure. Bottom: stereo view of the backbone diagram of the 19 refined low-energy structures. Backbone heavy atoms were used for superimposition.

that the N-terminal part constitutes an independent structural domain. Only a few resonances of BlaI-NTD differ between the two spectra, and those correspond to the residue neighboring the cleavage site (Lys81). The resonances corresponding to the C-terminal domain (82–128) are broad and poorly resolved, suggesting that this domain could be less structured or more mobile than the N-terminal domain.

The BlaI-NTD structure was solved using two- and multidimensional NMR spectroscopy, making use of uniformly ^{15}N, $^{13}C/^{15}N$ and $^{2}H/^{13}C/^{15}N$-labeled proteins. Nearly complete ^{1}H, ^{13}C and ^{15}N assignments were obtained using standard resonance assignment procedures. The few missing assignments correspond to the N and C-terminal ends (residues 1–4, and 82).

For structure calculation, a total of 1252 interproton, 110 dihedral and 97 chiral restraints were used. The distance constraints contain 596 intraresidue, 205 sequential, 87 medium-range, 145 long-range and 219 ambiguous correlations derived from 2D, 3D, and 4D nuclear Overhauser effect spectroscopy (NOESY) spectra. The dihedral ψ and ϕ angle restraints were deduced from TALOS, using backbone chemical shift values. At the final stage, 80 conformers were calculated, and the conformers with the lowest-energy functions were selected to be refined. A ribbon diagram of the lowest-energy structure and a superposition of the final ensemble of 19 simulated annealing structures are shown (Figure 3).

Ensemble analysis of BlaI-NTD backbone stereochemical quality was done with NMR-PROCHECK.[12] Over 99% of the dihedral angles were found in the most favored and additional allowed regions of the Ramachandran plot. Structural conformations of residues 1–4 and 79–82 are poorly defined, due to the lack of assignment for residues 1–4 and 82, or to the small number of NOE constraints for residues 79–81. Heteronuclear NOEs (Supplementary Material) confirm that the N and C-terminal parts of the protein are rather mobile. For statistics, we consider only the rigid core of BlaI-NTD (residues 5–78). For these residues, we obtained an average of 16.4 restraints per residue. A good agreement with the experimental restraints is reflected by the low experimental energy level. There is no distance restraints violation greater than 0.1 Å. The backbone and heavy-atom root-mean-square deviation (rmsd) values calculated for BlaI-NTD 5–78 with respect to the mean coordinates are 0.63(±0.17), and 1.07(±0.18) Å, respectively. A summary of the restraint and structural statistics is presented in Table 1.

The BlaI-NTD 3D structure consists of a three-stranded β-sheet (S1, Ser23-Asn25; S2, Leu57-Glu62; and S3, Val65-Pro70) packed against three α-helices (H1, Asp9-Lys20; H2, Thr26-Thr36; and H3, Pro41-Lys53) arranged in the order H1-S1-H2-T1-H3-S2-W1-S3. S2 and S3 form an antiparallel hairpin (loop called wing W1: Gly63-Arg64) and S1 is connected antiparallel to S3. T1 (Ser37-Ser40) is a type I turn connecting H2 and H3. The α-helices and the β-sheet form a well-defined and compact core. This tertiary arrangement belongs to the winged helix proteins (WHP) family, which is a member of the DNA recognition helix-turn-helix superfamily.[13,14] Among the WHP, the relative orientation of the secondary structures is variable due to the differences in the length of the T1 turn connecting H2 and H3, and in the length of the wing W1. Comparison of the structure of BlaI-NTD with other WHP proteins was made using the DALI search algorithms.[15] The most similar 3D structures are those of E2F4, SMTB, ADAR1, LexA, MarR, FokI, DTXR, and Genesis, all of them being DNA recognition domains. It is interesting to note that the MarR protein is also involved in a mechanism of antibiotic resistance.[16] Although no sequence homology could be detected by primary sequence similarity searches, all these proteins present a central core similar to that of BlaI-NTD. Whereas their helices can be of significantly different lengths, angles between the three α-helices are conserved. The main differences observed for these proteins are localized in the wing region. Some of the WHP proteins like Genesis have a long and flexible W1 wing, as revealed by heteronuclear NOE measurements.[17] In the case of BlaI-NTD, wing W1 is short (residues 63–64). Heteronuclear NOE data (Supplementary Material) show no particular mobility for residues neighboring W1.

A high degree of sequence homology exists between *B. licheniformis* BlaI and *S. aureus* MecI, especially for the non-polar core residues. In particular, many of the buried residues forming the core of the structure are identical or similar (*BlaI/MecI*) between BlaI and MecI: (H1) Ala10, Val14,

Table 1. Structural statistics

NMR-derived restraints	1362
Interproton restraints	1252
Intraresidue restraints	596
Interresidue restraints	437
Sequential restraints	205
Medium-range restraints	87
Long-range restraints	145
Hydrogen bond restraints	0
Ambiguous restraints	219
Dihedral angle restraints	110
Chiral restraints	97
Average rms deviations from the mean coordinates (Å)[a,b]	
Backbone heavy-atoms:	0.63 ± 0.17
All heavy-atoms:	1.07 ± 0.18
Ramachandran analysis (%)[b,c]	
Residues in most favoured regions:	85.6
Residues in additional allowed regions:	13.7
Residues in generously allowed regions:	0.2
Residues in disallowed regions:	0.5

[a] Statistics were made for residues 5–78 (see the text).
[b] Average for the 19 lowest-energy structures.
[c] Only non-Gly and non-Pro residues were assessed with PROCHECK-NMR.

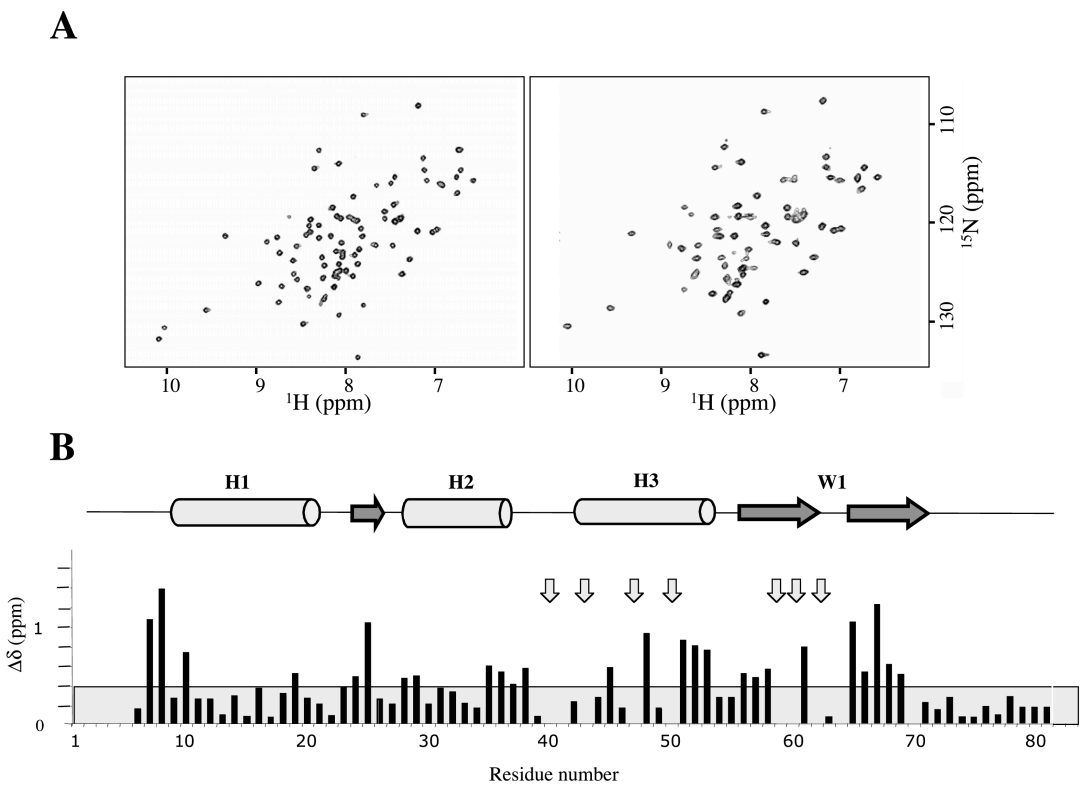


Figure 4. Chemical shift mapping of BlaI-NTD in interaction with DNA. A, ^{15}N HSQC spectra recorded at 25 °C. Left: BlaI-NTD spectrum in absence of DNA. Right: BlaI-NTD/DNA complex spectrum. The 24-mer palindromic DNA was added to the sample to reach a protein/DNA$_{duplex}$ ratio of 1/2. (B) Graph of the chemical shift variations between the DNA free BlaI-NTD and the BlaI-NTD/DNA complex *versus* residue number. The vertical axis corresponds to ($|\delta^1H| \times |\gamma_H/\gamma_N| + |\delta^{15}N|$) measured on the spectra presented in A. Arrows represent amino proton resonances that have been broadened and not assigned in the complexed form. The absence of a bar indicates the presence of a proline residue or an unmeasured shift due to overlap. The limit between weak and strong chemical shift variations has been fixed to 0.4 ppm.

Met15, Ile18, Trp19, (H2) *Val/Ile29*, *Ile/Val30*, Glu32, *Leu/Ile33*, (H3) Ile44, *Leu/Ile48*, Leu51, (β-sheet) *Leu/Ile57*, Tyr68, Ile72. Since there is no deletion or insertion necessary to align BlaI and MecI, the secondary structures predicted for MecI have the same length as for BlaI. All these data support the conclusion that MecI also belongs to the WHP family, and that its fold is very close to the BlaI structure.

DNA recognition by the repressors BlaI and MecI

For some of the WHP proteins with a BlaI-NTD similar fold listed above, the protein–DNA complex has been investigated by X-ray crystallography, NMR or computational tools.[13] For all of them, the same DNA interaction mode has been reported. Helix H3, called the recognition helix, is presented to the major groove of the DNA, and makes specific contacts with the base-pairs. The wings (in particular W1) and the protein N-terminal part also make contacts with the minor groove of the DNA. Unlike this WHP canonical mode of DNA recognition, the WHP RFX1 makes most of the contacts with the DNA major groove *via* wing W1.[18] The so-called recognition helix H3 overlies the minor groove, and there is no contact between the N-terminal part of the protein and the DNA.

To investigate the BlaI DNA recognition mode, the BlaI–DNA interaction surface was determined using the NMR chemical shift perturbation method. The method detects protein residues that are interacting directly, or that undergo conformational changes upon binding. ^{15}N-HSQC spectra of *B. licheniformis* BlaI-NTD were recorded with increasing amounts of DNA for both the 30-mer DNA corresponding to the *B. licheniformis* BlaI operator op1, and a palindromic 24-mer DNA based on the symmetric consensus sequence of the *mec* and *bla* operators. Results obtained with the palindromic 24-mer DNA are shown in Figure 4 and are similar to the results obtained with the 30-mer DNA (data not shown).

Since the ^{15}N-HSQC spectra of DNA-free BlaI-NTD, and of DNA-bound BlaI-NTD share many similarities (Figure 4A), the amide ^{1}H and ^{15}N chemical shifts of the N-terminal domain of DNA-bound

BlaI-NTD were assigned by reference to the free protein assignments. This comparison shows that BlaI-NTD does not undergo major structural rearrangements upon binding to DNA. The ^{15}N, ^{1}H chemical shift and intensity changes for backbone amide groups are given (Figure 4B). Significant chemical shift or intensity changes can be observed in the protein N-terminal part (Ile7 and Ser8), all along helix H3 (Ser40, Thr43, Met47, Leu48, Arg50, Leu51, and Ile52), and in the β-sheet, especially next to wing W1 (Asn25, His59, His60, Lys61, Glu62, Val65, and Val67). All the residues affected upon binding can be highlighted on the BlaI-NTD structure. They define a DNA-binding surface located on a single side of BlaI-NTD. This surface can be superimposed nicely with the positively charged region of the protein electrostatic surface (Figure 5A). The residues of the BlaI-NTD N-terminal part that are affected upon binding demonstrate the functional importance of the repressor N-terminal part in DNA-binding. This result is consistent with former DNA footprinting studies showing that a BlaI repressor mutant carrying a six amino acid deletion (Lys3-Ser8) within its N-terminal region had lost its ability to bind the operator.[7] Mutation of lysine 4 to alanine also severely reduces the BlaI DNA-binding ability.[19] The crucial role of the N-terminal domain and the shape of the electrostatic and interaction surfaces are not consistent with RFX1-DNA recognition mode, where there is no interaction between the N-terminal part of the protein and the DNA. On the contrary, they are in good agreement with a canonical WHP–DNA interaction, as described for E2F4.[20] Furthermore, the distances separating helix H3, wingW1, and the N-terminal ends that interact with DNA are quite equal for BlaI and E2F4. We therefore propose a model for the interaction between the repressors BlaI and MecI and their operators by comparison with the structure of the E2F4–DNA complex (Figure 5B). In this model, helix H3 interacts with the major groove of the operator. Other DNA–protein contacts occur *via* wing W1 and in the N-terminal part. Precise position of the contacts between the amino acid residues of BlaI and MecI and the base-pairs must still be investigated *via* a high-resolution structure of the DNA–repressor complexes.

Biological implication: insights about the BlaI and MecI repression mechanism

It has been reported that BlaI and MecI could corepress β-lactamase and PBP2′ gene expression in *S. aureus*.[9] In the same way, *B. licheniformis* BlaI interacts with *S. aureus mec* operator, as highlighted by gel mobility-shift assay (C. V., unpublished results). To explain this *mec/bla* corepression, the sequence homology for both the repressors and the DNA operator sequences of *S. aureus* and *B. licheniformis* was investigated. The primary structure alignment between *B. licheniformis* BlaI and *S. aureus* MecI and BlaI (Figure 5D) shows that all H3 residues pointing outside the protein, and therefore probably implicated in DNA recognition (Lys42, Thr46, Arg50 and Lys53), are totally conserved among the BlaI and MecI repressors. These residues are either positively charged or able to form specific hydrogen bonds with the base-pairs. The existence of a totally conserved DNA palindromic motif (TACA) in the *mec* and *bla* operators in *S. aureus* and *B. licheniformis* has been highlighted previously (Figure 1).[7] In our interaction model, the conserved residues in BlaI and MecI helix H3 make specific contacts with this TACA motif (Figure 5C). Interestingly, this model locates well-conserved positive residues (His59, His60, and Lys61, next to W1) facing an also conserved AT base-pair. Furthermore, locating two BlaI-NTD monomers on the two symmetric TACA motifs along the operator places the C-terminal parts of the two monomers pointing in a convenient orientation for dimerization contacts. Thus, we propose a repression mechanism, in which the repressor helix H3 makes specific contacts with the TACA motif of the operator sequence. This model is able to explain the corepression of the *mec* and *bla* genes.

Although the *S. aureus* and *B. licheniformis* BlaI share the same DNA-binding features, they exhibit different induction mechanisms. In *S. aureus*, the presence of the β-lactam inducer outside the cell results in the proteolytic cleavage of the repressor between residues N101 and F102,[21] leading to the loss of its ability to dimerize. As pointed out above, the BlaI-NTD 3D-structure is not or slightly modified when compared to that of the wild-type BlaI, whereas the decrease of affinity between BlaI and BlaI-NTD in *B. licheniformis* is about 500–1000 times.[7] The SmtB Zn(II) sensing metallo-regulated repressor is a WHP whose DNA-binding affinity is in the same range as BlaI–DNA affinity. The binding of a zinc ion to its C-terminal domain promotes the disassembly of the SmtB–DNA complex through a loss of affinity of about 1000.[22,23] The binding of a co-activator to the C-terminal domain, as in SmtB, is not the only way to inhibit the repressor activity of a winged-helix transcriptional regulator. For example, in the case of *Escherichia coli* MarR, the binding of two salicylate groups close to the DNA-binding α-helix inhibits the MarR activity.[16] To investigate the role of the C-terminal domain in BlaI–DNA interaction, a chemical shift mapping experiment was carried out for the dimeric full-length BlaI in the same conditions as for the monomeric BlaI-NTD (Supplementary Material). Two interesting observations can be derived from this study. First, a similar chemical shift perturbation pattern is obtained for the residues of the N-terminal domain. Second, no significant modifications of the resonances of the C-terminal domain are observed. This suggests that the interaction mode between BlaI-NTD and the DNA is not influenced by the presence of the C-terminal domain. The BlaI/MecI dimerization C-terminal domain is,

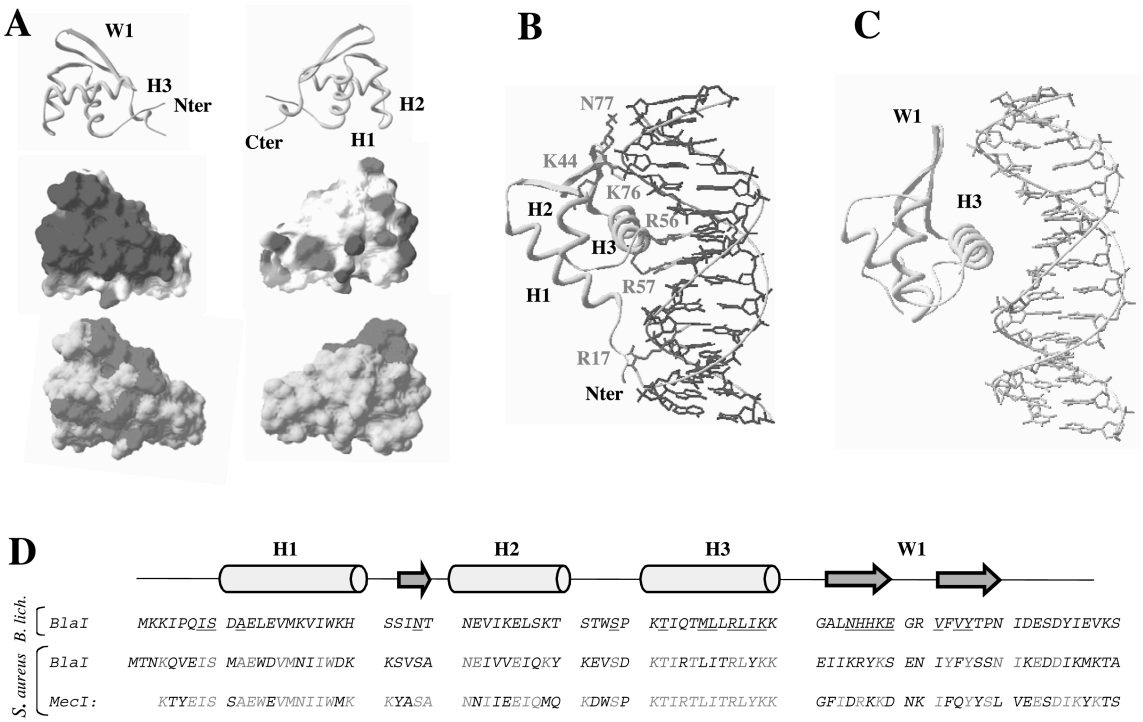


Figure 5. Comparison of the region involved in the protein/DNA complex for E2F4 and BlaI-NTD. A, Electrostatic surface of BlaI and surface mapping of the region affected by the DNA interaction. On top, ribbon representation of BlaI-NTD structure in a front and back orientation. In the middle, surface electrostatic potential is colored in red (positive) and blue (negative). The calculations were performed using the Swiss-PDB viewer (version 3.7). At the bottom, chemical shift variations ($|\delta^1H| \times |\gamma_H/\gamma_N| + |\delta^{15}N|$) superior to 0.4 ppm (see Figure 4B) are indicated in red on the BlaI-NTD surface. B, Structure of E2F4/DNA complex.[20] Side-chains involved in a hydrogen bond with the DNA are indicated in red and the H-bonds in green. Helix H3 forms direct H-bonds with the bases of the DNA (Arg56 and Arg57) as well as the N terminus part (Arg17). The winged helix and the turn between H1 and H2 form H-bonds with the phosphate and the sugar of the DNA (Lys44 and Asn77). Note that the residue numbering of E2F4 is quite different from that of BlaI where the H3 helix extends between the residues 41 and 53. C, Interaction model of the complex BlaI-NTD/DNA derived from the E2F4/DNA complex. BlaI-NTD structure was superimposed with E2F4 structure. Totally conserved nucleotides presented in the consensus sequence of Figure 1 are colored in green D, Sequence alignment of *B. licheniformis* BlaI, and *S. aureus* BlaI and MecI. Residues indicated in blue correspond to amino acid residues conserved between BlaI of *B. licheniformis* and *S. aureus*. Residues indicated in red are conserved between BlaI and MecI of *S. aureus*. The highlighted residues correspond to the residues affected by the DNA interaction (determined by the *B. licheniformis* BlaI chemical shift mapping).

however, essential to obtain a DNA-binding protein of high affinity, probably placing the two DNA-binding domains in a favorable conformation for DNA contacts.

Conclusion

The 3D solution structure of *B. licheniformis* BlaI DNA-binding domain presented in this study reveals that BlaI is a member of the WHP family. *S. aureus* BlaI and MecI share also a similar WHP fold. Chemical shift mapping experiments highlighted that BlaI presents the canonical DNA recognition mode. The C-terminal part of the protein is involved in dimerisation, but not in DNA interaction. The N-terminal part is a winged helix motif, interacting *via* its N-terminal part, helix H3 and wing W1. A model of the BlaI/MecI DNA interaction is proposed. In this model, specific contacts occur between the base-pairs of the TACA motif and conserved amino acid residues of the repressors helix H3. These results help understanding the mechanism by which the BlaI/MecI repressor regulates *blaP/blaZ/mecA* expression.

Materials and Methods

Plasmids and DNA manipulations

PET22b (Novagen) was used as vector for the over-expression of the BlaI and BlaIGM2 products. The construction of plasmids pET22bblaIWT and pET22b-blaIGM2 has been described.[4] The GM2 mutation ($M^{97}V^{98}$/IL) is located in the C-terminal domain of BlaI. For BlaI-NTD production, the GM2 mutant was preferred due to its better over-expression yield.

Protein overexpression and purification

The uniformly ^{15}N-labeled BlaI sample was prepared by growing *E. coli* BL21(DE3) pLys containing the corresponding recombinant pET22b plasmid in an optimal M9 minimal medium without Casamino acids. In this medium, the only source of nitrogen was [^{15}N]ammonium chloride and the culture medium contained 18 mM $^{15}NH_4Cl$, 35 mM unlabeled glucose, 100 μg/ml of ampicillin, 30 μg/ml of chloramphenicol, 5 μM biotin, 7 μM thiamin, 13 μM $FeSO_4$, 0.5 mM $MgSO_4$, 0.01% (w/v) cholin chloride, 0.001% (w/v) folic acid, 0.002% (v/v) pyridoxal, 0.0001% (w/v) riboflavin and 0.01% (w/v) niaccinamid. Cells were grown at 37 °C to an $A_{600\ nm}$ of 0.7 and isopropyl β-D-galactosidase (IPTG) was then added to a final concentration of 1 mM. Overnight-induced cells were then harvested by centrifugation, lysed and purified.[4] The final yield of purified ^{15}N-labeled protein was 6 mg per liter of cell culture and the isotopic labeling was 94% as determined by mass spectrometry.

The uniformly ^{15}N/^{13}C-labeled BlaIGM2 sample was produced according to the same procedure, apart from the fact that M9 minimal medium was replaced by the Silantes OD2-CN medium (Silantes GmbH, Gollierstrasse 70C, D-80339 Munchen). The final yield of purified labeled protein was 20 mg/l of cell culture and the percentage of isotopic labeling was 99%.

Uniformly ^{15}N/^{13}C/^{2}H-labeled BlaIGM2 sample was produced using the optimized M9 medium in which water had been replaced by $^{2}H_2O$ and with $^{15}NH_4Cl$ and ^{13}C glucose concentrations of 18 mM and 10 mM, respectively. The final yield of purified labeled protein was 4 mg per liter of cell culture and the percentage of deuterium labeling was approximately 75%.

Papain digestion and BlaI-NTD purification

Purified BlaI or BlaIGM2 were dialysed against a buffer containing 50 mM Hepes (pH7), 100 mM NaCl, 1 mM EDTA and was digested overnight using 1% mol/mol papain at 28 °C. After this time, the cleavage was complete and the protein digest was applied to an S-Sepharose-Fast-Flow column equilibrated in buffer A (50 mM Hepes (pH7), 100 mM NaCl, 1 mM EDTA, 5% (v/v) glycerol). BlaI-NTD was eluted with 1 M NaCl in the same buffer. The BlaI-NTD containing fractions were pooled, dialyzed exhaustively against 50 mM sodium phosphate (pH7.6), 200 mM KCl, 1 mM NaN_3, 0.1 mM Pefabloc and then concentrated to a final protein concentration of 0.5 mM by ultrafiltration with a Centricon (cut-off 5 kDa, Amicon, Brussels, Belgium).

Amino N-terminal sequencing of BlaI-NTD was carried out using a Procise 492 pulsed liquid phase protein sequencer (Applied Biosystems) with 20–30 pmol of protein.

Band-shift assays

Gel retardation experiments were realized as described.[24]

NMR sample preparation

NMR samples of BlaI-NTD were prepared at a concentration of 0.75 mM in 75 mM sodium phosphate buffer (90% H_2O, 10% $^{2}H_2O$), 300 mM KCl, pH 7.6. For the full-length protein, an extensive microdrop study was carried out to optimize the buffer conditions and prevent aggregation observed at high concentration.[25] Aggregation was avoided by adding 300 mM glycine and 240 mM NaCl to the buffer used for BlaI-NTD. BlaI samples were thus prepared in the same condition at 0.75 mM for ^{15}N-labeled BlaI, 0.5 mM for ^{15}N/^{13}C-labeled BlaI, and 0.5 mM for ^{15}N/^{13}C/^{2}H-BlaI. For protein interaction studies, two DNA sequences were chemically synthesized: a 30-mer corresponding to the *B. licheniformis* BlaI operator op1 5′-GAA AGT ATT ACA TAT GTA AGA TTT AAA TGC-3′, and a consensus palindromic 24-mer DNA: 5′-AAA GTA TTA CAT ATG TAA TAC TTT-3′. DNA was dissolved in the same buffer as the protein. Samples of ^{15}N BlaI and ^{15}N/^{13}C BlaI-NTD were prepared at a concentration of 0.1 mM. DNA solution was progressively added by amounts of 20 μl to reach $BlaI_{dimer}/DNA_{duplex}$ ratios of 0.5, 1.0, 1.5 and 2.0, and a BlaI-NTD/DNA_{duplex} ratios of 0.5, 1.0, 1.5, 2.0, and 2.5.

NMR spectroscopy

For sequential assignment, standard triple resonance experiments were carried out on ^{15}N and ^{15}N/^{13}C-labeled samples of BlaI-NTD: HNCA, HNCO, HN(CA)CO, CBCANH, and CBCA(CO)NH for backbone assignment and (H)C(CO)NH-TOCSY, H(CCO)NH-total correlated spectroscopy (TOCSY) for aliphatic side-chain assignment. For resonance assignment of aromatics, HSQC and H(C)CH-TOCSY experiments with the ^{13}C carrier set to 125 ppm, and 2D-NOESY spectrum performed on a sample dissolved in $^{2}H_2O$ were used. An HNCA experiment was performed on a ^{15}N/^{13}C/^{2}H-labeled sample of BlaIGM2 to confirm the assignment of the N-terminal part in the full-length protein.

Inter-protons restraints were derived from NOESY spectra: 2D-NOESY, 3D ^{15}N-edited NOESY-HSQC, regular 3D ^{13}C-edited NOESY-HSQC, 3D ^{13}C-edited NOESY-HSQC optimized for CH groups and 4D ^{13}C-edited HSQC-NOESY-HSQC.[26] Mixing times were set to 120 ms.

For chemical shift mapping, ^{15}N-HSQC, ^{13}C-HSQC optimized for aromatics and CT-HSQC were performed at different protein/DNA ratios.

All NMR experiments were performed on Varian INOVA 600 and INOVA 800 spectrometers, both equipped with a triple-resonance (^{1}H, ^{15}N, ^{13}C) probe and shielded *z*-gradients. The temperature was set to 25 °C. All triple-resonance experiments used the pulse sequences provided by the Varian Protein Pack.† All data processing, peak picking and peak intensity measurements were performed using the FELIX program version 2000 (Accelrys).

Structure calculation

The NOE distance restraint calibration was performed using cross-relaxation volumes between atoms of known separation in rigid parts of the molecule. A factor of 35% was added to the resulting upper distance, to account for the inherent uncertainty in the distance calculation. Classical pseudo atoms corrections were applied. The lower distance limit was set to the sum of the van der Waals radii of the two protons. Assignment, restraint generation, and structure calculation were performed iteratively. Restraints that were violated, resulting from

† ftp.nmr.varian.com

misassignments or overlapped peaks, were corrected after the first steps of calculation.

The ψ and ϕ angles calculated by TALOS from backbone chemical shift values were used as constraints with lower and upper bounds of ±20°.[27] Structure calculation was carried out using the Discover programs (Accelrys) interfaced to INSIGHTII for visualization and analytical purposes. The force-field used was AMBER.[28] The structure determination protocol used a simulated annealing calculation starting from randomized Cartesian coordinates to explore the conformational space, and a restrained molecular dynamics calculation to refine each structure as described.[29]

Data Bank accession numbers

Chemical shift assignments of BlaI-NTD have been deposited with BioMagResBank (accession number 5873). The coordinates of the 19 BlaI-NTD refined structures have been deposited in the Protein Data Bank under the accession code 1P6R.

Acknowledgements

This work was supported by the Belgian program of Interuniversity Poles of Attraction initiated by the Federal Office for Scientific Technical and Cultural Affaires (PAI no. P5/33), the "Fonds National de la Recherche Scientifique" (FNRS, Crédit aux chercheurs no 1.5201.02 and FRFC no. 2.4530.03) and the "Communauté Française de Belgique" (projet Tournesol 2003, 03/013). We thank Martin Blackledge for the gift of protocols for structure calculation and for many useful discussions. We also thank Laurence Blanchard for helpful discussion and protein expression protocols. We thank Edwin De Pauw for mass spectrometry analysis.

References

1. Charlier, P., Coyette, J., Dehareng, D., Dive, G., Duez, C., Dusart, J. *et al.* (1998). Resistance bactérienne aux β-lactamines. *Medecine/Sciences*, **14**, 544–549.
2. Joris, B., Hardt, K. & Ghuysen, J. M. (1994). Analysis of the *penA* gene of *Pseudomonas cepacia* 249. *New Comput. Biochem.* **27**, 505–510.
3. Lim, D. & Strynadka, C. J. (2002). Structural basis for the β-lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Nature Struct. Biol.* **9**, 870–876.
4. Filée, P., Benlafya, K., Delmarcelle, M., Moutzourelis, G., Frere, J. M., Brans, A. & Joris, B. (2002). The fate of the BlaI repressor during the induction of the *Bacillus licheniformis* BlaP beta-lactamase. *Mol. Microbiol.* **44**, 685–694.
5. Duval, V., Swinnen, M., Lepage, S., Brans, A., Granier, B., Fransen, C. *et al.* (2003). The kinetic properties of the carboxy terminal domain of the *Bacillus licheniformis* 749/I BlaR penicillin-receptor shed a new light on the derepression of β-lactamase synthesis. *Mol. Microbiol.* **48**, 1553–1564.
6. Filée, P., Vreuls, C., Hermann, R., Thamm, I., Frere, J. M. & Joris, B. (2003). Dimerization and DNA-binding properties of the *Bacillus licheniformis* 749/I BlaI repressor. *J. Biol. Chem.* **278**, 16482–16487.
7. Wittman, V., Lin, H. C. & Wong, H. C. (1993). Functional domains of the penicillinase repressor of *Bacillus licheniformis*. *J. Bacteriol.* **175**, 7383–7390.
8. Joris, B., Hardt, K. & Ghuysen, J. M. (1994). Induction of β-lactamase and low affinity penicillin binding protein 2′ synthesis in Gram-positive bacteria. *Bacterial Cell Wall*, **27**, 505–515.
9. McKinney, T. K., Sharma, V. K., Craig, W. A. & Archer, G. L. (2001). Transcription of the gene mediating methicillin resistance in *Staphylococcus aureus* (*mecA*) is corepressed but not coinduced by cognate *mecA* and beta-lactamase regulators. *J. Bacteriol.* **183**, 6862–6868.
10. Lewis, A. & Dyke, G. H. (2000). MecI represses synthesis from β-lactamase operon of *Staphylococcus aureus*. *J. Antimicrob. Chemother.* **45**, 139–144.
11. Salerno, A. J. & Lampen, J. O. (1988). Differential transcription of the *bla* regulatory region during induction of beta-lactamase in *Bacillus licheniformis*. *FEBS Letters*, **227**, 61–65.
12. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
13. Gajiwala, K. S. & Burley, S. K. (2000). Winged helix proteins. *Curr. Opin. Struct. Biol.* **10**, 110–116.
14. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, **364**, 412–420.
15. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
16. Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A. & Head, J. F. (2001). The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nature Struct. Biol.* **8**, 710–714.
17. Jin, C. & Liao, X. (1999). Backbone dynamics of a winged helix protein and its DNA complex at different temperatures: changes of internal motions in genesis upon binding to DNA. *J. Mol. Biol.* **292**, 641–651.
18. Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. & Burley, S. K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature*, **403**, 916–921.
19. Gregory, P. D., Lewis, R. A., Curnock, S. P. & Dyke, K. G. (1997). Studies of the repressor (BlaI) of beta-lactamase synthesis in *Staphylococcus aureus*. *Mol. Microbiol.* **24**, 1025–1037.
20. Zheng, N., Fraenkel, E., Pabo, C. O. & Pavletich, N. P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* **13**, 666–674.
21. Lewis, R. A., Curnock, S. P. & Dyke, K. G. (1999). Proteolytic cleavage of the repressor BlaI of beta-lactamase synthesis in *Staphylococcus aureus*. *FEMS Microbiol. Letters*, **178**, 271–275.
22. Cook, W. J., Kar, S. R., Taylor, K. B. & Hall, L. M. (1998). Crystal structure of the cyanobacterial metallothionein repressor *Smtb*: a model for metalloregulatory proteins. *J. Mol. Biol.* **275**, 337–346.
23. Vanzile, M. L., Chen, X. & Giedroc, D. P. (2002). Structural characterization of distinct alpha3N and

alpha5 metal sites in the cyanobacterial zinc sensor *Smtb*. *Biochemistry*, **41**, 9765–9775.
24. Filée, P., Delmarcelle, M., Thamm, I. & Joris, B. (2001). Use of an ALFexpress DNA sequencer to analyse protein–nucleic acid interactions by band shift assay. *Biotechniques*, **30**, 1044–1051.
25. Lepre, C. A. & Moore, J. M. (1998). Microdrop screening: a rapid method to optimize solvent conditions for NMR spectroscopy of proteins. *J. Biomol. NMR*, **12**, 493–499.
26. Kay, L. E., Clore, G. M., Bax, A. & Gronenborn, A. M. (1990). Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1β in solution. *Science*, **249**, 411–414.
27. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
28. Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P. & Kollman, P. A. (1995). *Amber 4.1*, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA.
29. Cordier, F., Caffrey, M., Brutscher, B., Cusanovich, M. A., Marion, D. & Blackledge, M. (1998). Solution structure, rotational diffusion anisotropy and local backbone dynamics of *Rhodobacter capsulatus* cytochrome c2. *J. Mol. Biol.* **281**, 341–361.

Edited by P. Wright

(Received 21 July 2003; received in revised form 1 September 2003; accepted 5 September 2003)

SCIENCE DIRECT®
www.sciencedirect.com

Supplementary material comprising two figures is available on Science Direct

Biochem. J. (2004) **384**, 179–190 (Printed in Great Britain)

Annexe 4

Guanidinium chloride denaturation of the dimeric *Bacillus licheniformis* BlaI repressor highlights an independent domain unfolding pathway

Christelle VREULS*†, Patrice FILÉE*, Hélène VAN MELCKEBEKE‡, Tony AERTS§, Peter DE DEYN§, Gabriel LLABRÈS†, André MATAGNE*, Jean-Pierre SIMORRE‡, Jean-Marie FRÈRE* and Bernard JORIS*[1]

*Centre d'ingénierie des Protéines, Institut de Chimie B6a, Université de Liège, Sart-Tilman, B4000 Liège, Belgium, †Laboratoire de Physique Biomédicale, Institut de Physique B5, Université de Liège, Sart-Tilman, B4000 Liège, Belgium, ‡Institut de Biologie Structurale Jean-Pierre Ebel, CEA-CNRS-UJF, 38027 Grenoble, France, and §Department of Biomedical Sciences, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium

The *Bacillus licheniformis 749/I* BlaI repressor is a prokaryotic regulator that, in the absence of a β-lactam antibiotic, prevents the transcription of the *blaP* gene, which encodes the BlaP β-lactamase. The BlaI repressor is composed of two structural domains. The 82-residue NTD (N-terminal domain) is a DNA-binding domain, and the CTD (C-terminal domain) containing the next 46 residues is a dimerization domain. Recent studies have shown the existence of the monomeric, dimeric and tetrameric forms of BlaI in solution. In the present study, we analyse the equilibrium unfolding of BlaI in the presence of GdmCl (guanidinium chloride) using different techniques: intrinsic and ANS (8-anilinonaphthalene-l-sulphonic acid) fluorescence, far- and near-UV CD spectroscopy, cross-linking, analytical ultracentrifugation, size exclusion chromatography and NMR spectroscopy. In addition, the intact NTD and CTD were purified after proteolysis of BlaI by papain, and their unfolding by GdmCl was also studied. GdmCl-induced equilibrium unfolding was shown to be fully reversible for BlaI and for the two isolated fragments. The results demonstrate that the NTD and CTD of BlaI fold/unfold independently in a four-step process, with no significant co-operative interactions between them. During the first step, the unfolding of the BlaI CTD occurs, followed in the second step by the formation of an 'ANS-bound' intermediate state. Cross-linking and analytical ultracentrifugation experiments suggest that the dissociation of the dimer into two partially unfolded monomers takes place in the third step. Finally, the unfolding of the BlaI NTD occurs at a GdmCl concentration of approx. 4 M. In summary, it is shown that the BlaI CTD is structured, more flexible and less stable than the NTD upon GdmCl denaturation. These results contribute to the characterization of the BlaI dimerization domain (i.e. CTD) involved in the induction process.

Key words: BlaI *Bacillus licheniformis*, domain structure, folding intermediate, guanidinium chloride, β-lactamase induction, protein folding.

INTRODUCTION

The *Bacillus licheniformis 749/I* BlaI repressor is a prokaryotic regulator that, in the absence of a β-lactam antibiotic, prevents the transcription of the *blaP* gene, which encodes the BlaP β-lactamase. In the presence of an antibiotic, BlaI is inactivated and the BlaP β-lactamase is induced [1,2]. The model which has been proposed to explain the induction of BlaP β-lactamase in *B. licheniformis 749/I* is summarized in Figure 1 [3]. BlaI is a cytoplasmic 128-amino-acid protein composed of two distinct domains: an N-terminal domain (BlaI-NTD: residues 1–82), which contains the DNA-binding motif, and a C-terminal domain (BlaI-CTD: residues 83–128), which is involved in BlaI homodimerization [8]. The dissociation constant (K_d) of the purified BlaI dimer was estimated to be 25 μM by AU (analytical ultracentrifugation) [9]. DNase footprinting experiments have revealed the presence of three regulatory regions that are specifically recognized by BlaI [2]. Two adjacent operators are present in the *blaP* promoter (OP_1 and OP_2), and one in the *blaI–blaR1* promoter (OP_3). These operators possess a 23-bp-long dyad symmetry, and exhibit the consensus sequence 5′-aaAgTaTTACAtaTGTAagNtTt-3′ [3] (where the upper- and lower-case letters show conserved and non-conserved bases respectively). By using fluorescent band-shift assays, we suggested that the BlaI dimer recognizes its operators with a distinct global dissociation constant of 10^{-15}–10^{-14} M^2 [9].

BlaI can be hydrolysed by papain, yielding the NTD and CTD domains. BlaI-NTD has been purified by ion-exchange chromatography [10,11]. During this purification step, BlaI-CTD is lost. It is supposed that, during the separation process, BlaI-CTD multimerizes and becomes insoluble (C. Vreuls, unpublished work).

The full-length BlaI, as well as BlaI-NTD, has been studied by NMR spectroscopy [11]. The similarities between BlaI and BlaI-NTD $^1H/^{15}N$-HSQC (heteronuclear single-quantum coherence spectra) show that the N-terminal part constitutes an independent structural domain. The BlaI-NTD structure was solved by NMR, which revealed that BlaI-NTD belongs to the winged-helix proteins subfamily, which are member of the DNA-recognition helix–turn–helix superfamily [12]. On the other hand, all resonances corresponding to the CTD (amino acids 82–128) were broad, poorly resolved and located in the region of the spectra corresponding to an unfolded polypeptide chain. These results suggest that BlaI-CTD is largely unfolded or highly flexible in comparison with the NTD. Recent X-ray crystallographic data obtained with the *Staphylococcus aureus* MecI repressor, a protein homologous with BlaI, show that MecI forms a dimer composed

Abbreviations used: ANS, 8-anilino-1-naphtalene sulphonic acid; AU, analytical ultracentrifugation; CTD, C-terminal domain; DSP, dithiobis(succinimidyl propionate); GdmCl, guanidinium chloride; HSQC, heteronuclear single-quantum coherence; MALDI–TOF MS, matrix-assisted laser-desorption ionization–time-of-flight MS; Ni-NTA, Ni^{2+}-nitrilotriacetate; NTD, N-terminal domain; Rs, Stokes radius; SEC, size-exclusion chromatography; 3D, three-dimensional; Ve, elution volume.

[1] To whom correspondence should be addressed (email bjoris@ulg.ac.be).

© 2004 Biochemical Society

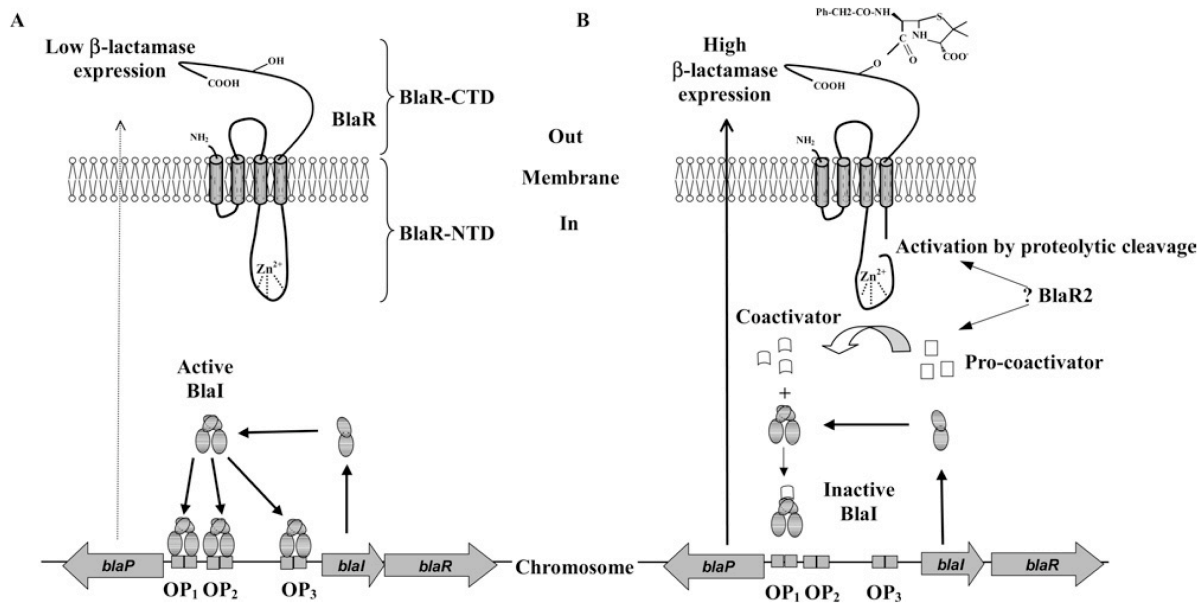


Figure 1 **β-Lactamase induction mechanism in *B. licheniformis* 749/I**

The *B. licheniformis* BlaP β-lactamase is inducible by a β-lactam antibiotic (the inducer) [2]. The regulation of β-lactamase production involves three regulatory genes, *blaI*, *blaR1* and *blaR2*. The first two genes encode a repressor and a penicillin-sensory transducer respectively ([4], but see [4a]), whereas the third gene is not identified yet. (**A**) In the absence of a β-lactam antibiotic, the BlaI repressor is bound as a dimer to three operator sequences (OP_1, OP_2 and OP_3) located in the intergenic region between *blaP* and *blaI-blaR* operon. Bound to DNA, BlaI prevents the transcription of the *blaP*, *blaI* and *blaR1* genes ([4], but see [4a]), [5]. (**B**) When a β-lactam antibiotic is added in the medium, the extracellular penicillin-binding domain of BlaR (BlaR-CTD) is acylated by the antibiotic [6], and a signal is transduced through the transmembrane segment resulting in the activation of the intracellular metalloprotease domain by proteolysis [7]. In *B. licheniformis*, it has been postulated that the activated metalloprotease converts a pro-coactivator into a co-activator, whose final target is BlaI itself. The BlaI dimer complexed with the co-activator is then released from its operator and β-lactamase is produced at high level ([4], but see [4a]). The product of *blaR2* is necessary for the induction process; it is postulated that it can be involved in the activation of the intracellular domain of the BlaR receptor or in the production of the pro-coactivator [3].

of two independent winged-helix domains and two intertwining dimerization domains. Each NTD binds a palindromic DNA operator half-site, and each CTD is composed of three α-helices held together by hydrophobic forces (Figure 2) [13,14].

In order to obtain more structural information about the BlaI CTD, we have investigated the GdmCl (guanidinium chloride)-dependent unfolding pathway of BlaI, and of its separated domains. GdmCl-induced unfolding was monitored by several techniques: intrinsic and ANS (8-anilinonaphthalene-l-sulphonic acid) fluorescence, far- and near-UV CD spectroscopy, cross-linking, AU, SEC (size-exclusion chromatography) and NMR spectroscopy. These experiments have generated complementary data that lead us to propose a multi-step process for BlaI unfolding. All changes were found to be reversible, and the proposed model contains four different steps in which the two BlaI domains sequentially and independently unfold. The implications of these results on the induction process are also discussed.

MATERIALS AND METHODS

Reagents

GdmCl (> 99 %) and ANS were purchased from Aldrich. Urea (> 99 %) was from Merck, and CHAPS was from Boehringer. DSP [dithiobis(succinimidyl propionate)] was from Pierce. Ni-NTA (Ni^{2+}-nitrilotriacetate) agarose was obtained from Affiland. Superdex 75 for gel filtration was purchased from Amersham Biosciences. All solutions were prepared with Milli-Q water and filtered through 0.22 μm pore size membranes (Millipore) before use. All others reagents were of analytical grade.

Expression and purification of BlaI and BlaI(His)$_6$

BlaI and BlaI(His)$_6$, a BlaI recombinant protein with a C-terminal hexahistidine tag, were produced and purified as described previously [3,10]. The uniformly ^{15}N-labelled BlaI sample was prepared as described previously [11]. The protein concentrations were determined using the bicinchoninic acid assay (Pierce), and by measuring the absorbance at 280 nm using a molar absorption coefficient of 16 600 $M^{-1} \cdot cm^{-1}$.

Papain digestion and BlaI-NTD and BlaI-CTD(His)$_6$ purification

Purified BlaI(His)$_6$ in buffer A (50 mM phosphate buffer, pH 8.0, in the presence of 500 mM NaCl) was subjected to papain digestion at 28 °C (1 % mol/mol) to generate BlaI-NTD (residues 1–82) and BlaI-CTD(His)$_6$ (residues 83–134) [11]. After an overnight incubation, cleavage was complete, and the two fragments were applied to a Ni-NTA column pre-equilibrated in buffer A. Under these conditions, BlaI-NTD was eluted in the void volume, whereas BlaI-CTD(His)$_6$ was retained on the column. The elution of BlaI-CTD(His)$_6$ was achieved by using buffer B (400 mM imidazole, pH 8.0, supplemented with 13 mM CHAPS). BlaI-NTD was dialysed against buffer C (50 mM phosphate, pH 7.6, supplemented with 200 mM KCl) and concentrated by ultrafiltration using a 5000 Da cut-off membrane (Millipore). BlaI-CTD(His)$_6$ was concentrated, and buffer B was exchanged against buffer C supplemented with 13 mM CHAPS by ultrafiltration using a 1000 Da cut-off membrane. Both fragments were characterized by MALDI–TOF MS (matrix-assisted laser-desorption ionization–time-of-flight MS), as described

© 2004 Biochemical Society

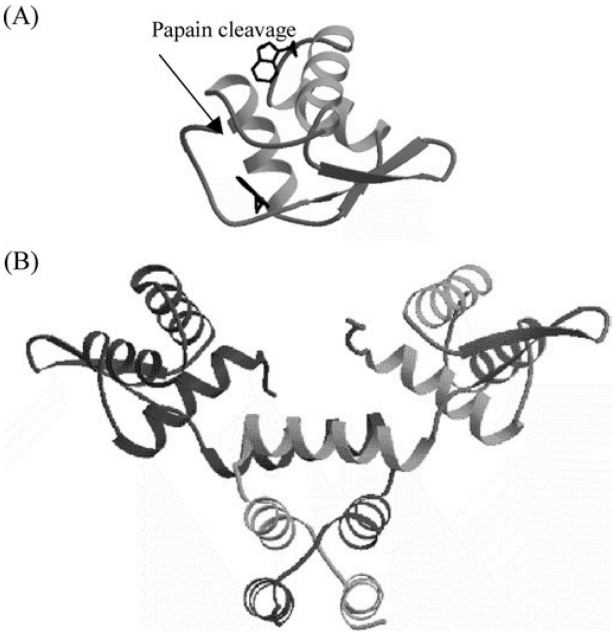


Figure 2 Views of *B. licheniformis* BlaI-NTD and *S. aureus* MecI

(**A**) *B. licheniformis* BlaI-NTD (residues 1–82) belongs to the winged-helix subfamily and consists of a three-stranded β-sheet (S1, Ser^{23}–Asn^{25}; S2, Leu^{57}–Glu^{62}; S3, Val^{65}–Pro^{70}) packed against three α-helices (H1, Asp^{9}–Lys^{20}; H2, Thr^{26}–Thr^{36}; H3, Pro^{41}–Lys^{53}) arranged in the order H1-S1-H2-T1-H3-S2-W1-S3 [11]. S2 and S3 form an antiparallel hairpin (loop called wing W1: Gly^{63}–Arg^{64}) and S1 is connected in parallel with S3. T1 (Ser^{37}–Ser^{40}) is a type 1 turn connecting H2 and H3. The α-helices are in yellow the β-sheet is in red and the tryptophan residues are in black. (**B**) The *S. aureus* MecI dimer is composed of two independent winged-helix domains and of two intertwining dimerization domains. The overall fold topology of the NTD is H1-H2-H3-S1-W1-S2, and the CTD is composed of three consecutive α-helices H4-H5-H6 held together by two hydrophobic cores between H4 and H6 [13]. The first monomer is in blue. The NTD of the second monomer (residues 1–73) is in the same colours as BlaI-NTD, and the CTD (residues 74–123) is in green. The *S. aureus* BlaI and MecI repressors are 60 % identical when compared with each other, and 31–41 % identical when compared with the *B. licheniformis* BlaI repressor [14].

previously [11]. N-terminal sequencing of BlaI-NTD and BlaI-CTD$(His)_6$ was performed using a Procise 492 pulsed liquid-phase protein sequencer (Applied Biosystems) with 20–30 pmol of protein.

Unfolding and refolding studies

Unfolding experiments were performed at room temperature by incubating BlaI, BlaI-NTD and BlaI-CTD$(His)_6$ (at concentrations ranging from 2 μM to 400 μM) in their respective buffers in the presence of various GdmCl concentrations (for further details, see the Figure legends). The GdmCl concentration was determined from refractive index measurements [15] using an Atago R5000 hand refractometer. Refolding studies were initiated by diluting 10 μl of the BlaI, BlaI-NTD or BlaI-CTD$(His)_6$ GdmCl-denatured samples (prepared as above at a 10-fold-higher protein concentration) with 90 μl of appropriate buffer.

Fluorescent band-shift assay

The fluorescent band-shift assays were performed as described previously [16] with the fluorescent, Cy5-labelled double-stranded oligonucleotide OP_1 (5′-Cy5-GCATTTAAATCTTACA-TATGTAATACTTTC-3′).

Fluorescence measurements

Both intrinsic and ANS fluorescence emission spectra were recorded on a PerkinElmer LS50B spectrofluorimeter. Excitation and emission slit widths were 4 and 5 nm respectively, and the scan speed was 350 nm · min^{-1}. Cuvettes with 1 cm path-length were used.

Intrinsic fluorescence measurements were performed using an excitation wavelength of 280 or 295 nm, and the emission spectra were recorded from 300–420 nm. The protein (2 μM) was conditioned in buffer C supplemented with increasing concentrations of GdmCl. The spectra were measured ten times, averaged and corrected by subtraction of the solvent spectrum obtained under identical conditions. Two fluorescence parameters [17] were considered in the present study: the fluorescence intensity at single excitation and emission wavelengths, and the maximum emission wavelength (λ_{max}). The fraction of unfolded protein (F_U) at increasing GdmCl concentrations was calculated using the following equation [15]:

$$F_U = (f_N - f)/(f_N - f_U)$$

where f is the observed fluorescence intensity at a given GdmCl concentration, f_U the fluorescence intensity when the protein is completely unfolded and f_N the fluorescence intensity of the native protein.

ANS fluorescence measurements were performed using a protein concentration of 25 μM with an excitation wavelength of 370 nm. Emission spectra for BlaI-CTD$(His)_6$ were recorded from 420–600 nm in buffer C supplemented with 13 mM CHAPS. The fluorescence spectra were corrected for the background fluorescence of ANS. The ANS concentration (determined from the molar absorption coefficient of 5000 $M^{-1} \cdot cm^{-1}$ at 350 nm) was estimated to be 500 μM [18].

CD measurements

CD measurements were performed using a Jobin-Yvon CD6 spectropolarimeter under constant nitrogen flow. In the far-UV region (195–260 nm), spectra were recorded in a 0.1 cm cell at protein concentrations of 0.5 mg · ml^{-1}, whereas in the near-UV region (250–300 nm), a 1 cm cell was used with a protein concentration of 1 mg · ml^{-1}. For BlaI-CTD$(His)_6$ experiments, Buffer C was used supplemented with 13 mM CHAPS. Spectra were acquired at a scan speed of 20 nm · min^{-1}, with a 2 nm bandwidth and 1 s integration time. The spectra were measured five times, averaged and corrected by subtraction of the solvent spectrum obtained under identical conditions. Results are expressed in terms of molar ellipticity ($[\theta]$). The fraction of unfolded protein F_U at various GdmCl concentrations was calculated using the following equation [15]:

$$F_U = (\theta_N - \theta)/(\theta_N - \theta_U)$$

where θ is the observed molar ellipticity at a given GdmCl concentration, θ_U is the molar ellipticity when the protein is completely unfolded and θ_N is the molar ellipticity of the native protein.

Cross-linking experiments

The stock solution of DSP (2 mM) was prepared in DMSO. Cross-linking experiments were performed in buffer C supplemented with the required GdmCl concentration. Protein (20 μmol) was

© 2004 Biochemical Society

mixed with 50 μmol of DSP in a total volume of 200 μl. The mixture was incubated for 2 h at 25 °C. The excess of free cross-linking reagent was eliminated by adding 10 μl of 1M Tris/HCl, pH 8.0 and incubating for 15 min on ice. GdmCl-denatured and cross-linked samples were then dialysed for 2 h against buffer D [50 mM Hepes, pH 7.6, supplemented with 200 mM KCl, 1 mM EDTA and 5 % (v/v) glycerol] and subjected to SDS/PAGE under non-reducing conditions (i.e. without 2-mercaptoethanol) [19].

AU experiments

Sedimentation velocity experiments were performed with a Beckman Optima XLA instrument at 20 °C using optical absorption detection. BlaI and BlaI-NTD concentrations were 0.9 mg/ml in all experiments. For BlaI-OP_1 experiments, one equivalent of DNA (OP_1) was added for two equivalents of BlaI. Buffer D was supplemented with the appropriate GdmCl concentration. BlaI and BlaI-NTD samples (500 μl) and the buffer against which the protein had been dialysed were placed in cells fitted with double sector centrepieces and quartz windows. Sedimentation was accomplished at a rotor speed of 40 000 rev./min. Cells were scanned at 280 nm (or 292, 295 and 298 nm for the BlaI DNA-binding study); 10–20 scans were collected for each cell. The scans were analysed by the method of Stafford [20]. Apparent sedimentation coefficients S were corrected to $S_{20,w}$ for both viscosity and temperature, as described elsewhere [20]. The densities and relative viscosities of the solutions were determined experimentally, and the partial specific volumes of the proteins were calculated from the amino acid sequences of the proteins.

SEC experiments

GdmCl-induced unfolding of BlaI was analysed by SEC using a Superdex 75 pre-packed FPLC column (10 mm × 300 mm, gel volume 24 ml; Amersham Biosciences) calibrated as described by Uversky [21]. The mobile phase was buffer D added with the required GdmCl concentration. Before injection, samples (50 μM) were filtered through a 0.22 μm filter (Millipore). Samples were eluted at a flow rate of 1 ml/min. All measurements were made at room temperature. The protein elution profile was monitored by recording the absorbance at 280 nm. Stokes radii (R_s) were estimated from the elution volumes (V_e) according to the equation: $1000/V_e = 11.71 + 2.97\,R_s$ [21].

NMR studies

Two-dimensional $^1H/^{15}N$-HSQC spectra were recorded at 25 °C on a Varian INOVA 800 spectrometer equipped with a triple resonance (^1H, ^{15}N and ^{13}C) probe and shielded gradients. Samples contained 0.4 mM 98 % ^{15}N-enriched BlaI or 0.3 mM 98 % ^{15}N-enriched BlaI-NTD in buffer C supplemented with the required GdmCl concentration. The initial sample contained no GdmCl and the $^1H_2O/^2H_2O$ ratio was 9:1. GdmCl salt was added progressively in order to obtain concentrations of 1 M, 1.5 M, 2 M, 3 M, 4 M and 5 M. A $^1H/^{15}N$-HSQC spectrum was recorded at each GdmCl concentration. To verify the reversibility of BlaI and BlaI-NTD unfolding processes, $^1H/^{15}N$-HSQC spectra of BlaI and BlaI-NTD were recorded after complete removal of the GdmCl by dialysing against buffer C. ^1H and ^{15}N chemical shifts are expressed relative to sodium 2.2-dimethyl-2-silapentane-5-sulphonate (DSS) and liquid ammonia respectively. Maximum t_1 and t_2 values were 128 ms and 154 ms, and the spectral widths were set to 2000 Hz and 6666 Hz for ^{15}N and ^1H dimensions respectively. Data processing was performed using the FELIX program version 2000 (Accelrys).

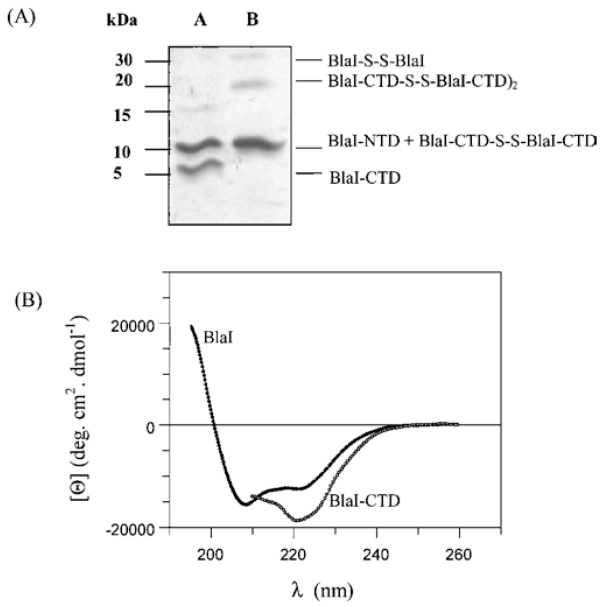


Figure 3 Characterization of the BlaI CTD

(**A**) Study of the oligomerization state of the BlaI CTD by cross-linking experiments. Purified BlaI was subjected to papain digestion. Cross-linking was performed by mixing 20 μmol of the digested protein with 50 μmol of DSP, and the results were analysed by SDS/PAGE (see the Materials and methods section). Lane A corresponds to the digested protein without DSP, and lane B to the digested protein incubated with DSP. Lane B highlights the dimeric (10 kDa) and tetrameric (20 kDa) forms of the BlaI CTD. (**B**) Far-UV CD spectra of BlaI and BlaI-CTD$(His)_6$. Proteins were incubated in 50 mM phosphate buffer, pH 7.6, supplemented with 200 mM KCl and 13 mM CHAPS for BlaI-CTD$(His)_6$. The protein concentrations used were 0.5 mg · ml^{-1} in a 0.1 cm cell.

RESULTS

Purification and characterization of BlaI N- and CTDs

From a papain digest, BlaI-NTD (Met^1–Ser^{82}) could easily be purified on an SP Sepharose Fast Flow column [11], whereas BlaI-CTD (His^{83}–Glu^{128}) could not be eluted. This was unexpected on the basis of the pI value of 5.1 computed according to the BlaI-CTD amino acid sequence. A hexahistidine tag was added to the C-terminal end of BlaI in order to purify the CTD [10]. Fragments generated by papain digestion of the purified BlaI$(His)_6$ were applied on to an Ni-NTA affinity column pre-equilibrated in buffer A. As expected, BlaI-NTD was eluted in the void volume, whereas BlaI-CTD$(His)_6$ was adsorbed and subsequently eluted with buffer B, which contained imidazole and 13 mM CHAPS. Addition of a mild zwitterionic detergent (CHAPS) was necessary to prevent BlaI-CTD$(His)_6$ aggregation both upon elution and during all subsequent experiments. The yield of BlaI-CTD$(His)_6$ purification was very poor (5 %). N-terminal sequencing and MALDI-TOF experiments revealed that BlaI-NTD extends from amino acids 1–82 (molecular mass 9490 Da) and BlaI-CTD$(His)_6$ from residues 92–134 (molecular mass 4220 Da). MALDI-TOF-MS also highlighted the presence of multimeric forms of the BlaI-CTD$(His)_6$ fragment. These results are in good agreement with those obtained by sedimentation velocity and cross-linking experiments. Indeed, in the stoichiometric mixture of BlaI-NTD and BlaI-CTD present in a papain digest, only BlaI-CTD could be cross-linked by DSP (Figure 3A), and during velocity sedimentation experiments BlaI-NTD behaved as a monomeric protein (results not shown). To verify that BlaI-CTD$(His)_6$ was

© 2004 Biochemical Society

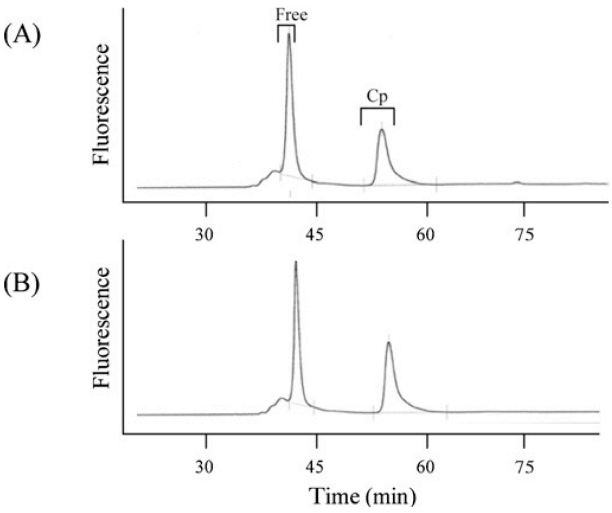


Figure 4 Study of the reversibility of the BlaI repressor GdmCl-induced denaturation by band-shift assays

(**A**) Binding of BlaI to the OP_1 target under native conditions (0 M GdmCl). 'Free' and 'Cp' represent the free DNA target and the DNA–protein complex respectively. The fluorescence is expressed in arbitrary units. DNA probe OP_1 (0.5×10^{-8} M) was incubated with purified BlaI repressor (3.3×10^{-8} M) and treated as described previously [16]. (**B**) Binding of BlaI to the OP_1 target after denaturation–renaturation of the protein. Protein and DNA concentrations are the same as in (**A**).

folded in the presence of CHAPS, far-UV CD spectra were recorded. As shown in Figure 3(B), BlaI-CTD$(His)_6$ exhibits a high percentage of α-helix. The α-helical contents of BlaI, BlaI-NTD and BlaI-CTD$(His)_6$ were calculated as a function of the molar ellipticities at 222 nm (Figure 3B; also see Figures 7A and 7B) [22], and found to be 33 %, 25 % and 53 % respectively. These values are lower than those obtained from the known X-ray structure of the *S. aureus* MecI repressor (64 % for MecI, 53 % for MecI-NTD and 81 % for MecI-CTD) [13] and from the solution structure of the *B. licheniformis* BlaI-NTD (43 %). MS and CD spectroscopy confirm that BlaI-CTD$(His)_6$ is well folded in the presence of CHAPS, and that it retains its ability to form oligomers. On the other hand, the purified NTD binds its DNA operator with a 500–1000-fold reduced affinity [11], despite the fact that it is no longer able to dimerize, as shown by sedimentation velocity experiments (results not shown).

GdmCl-induced reversible denaturation of BlaI and of its separated fragments

The reversibility of the GdmCl-induced denaturation of purified BlaI, BlaI-NTD and BlaI-CTD$(His)_6$ was studied by several analytical methods and by band-shift assay (for details, see the Materials and methods section). It was shown that, after removal of GdmCl by dilution concentration of the denatured protein: (i) 95 % of BlaI and BlaI-NTD intrinsic fluorescence intensity was recovered (results not shown); (ii) BlaI DNA-binding activity was regained, as revealed by band-shift assay experiments (Figure 4); and (iii) the BlaI-CTD$(His)_6$ far-UV CD spectrum (results not shown), as well as the BlaI and BlaI-NTD $^1H/^{15}N$-HSQC NMR spectra, were fully recovered (see the NMR studies section below).

Intrinsic fluorescence of BlaI and BlaI-NTD

The primary structure of BlaI highlights two tryptophan residues at positions 19 and 39, and four tyrosine residues at positions 68,

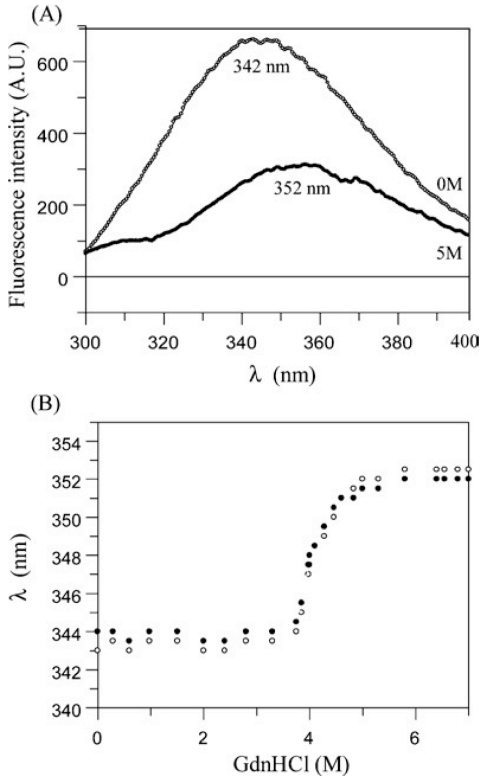


Figure 5 BlaI and BlaI-NTD intrinsic fluorescence emission spectra at various GdmCl concentrations

Spectra were recorded at 25 °C. Proteins (2 μM) were in 50 mM phosphate, pH 7.6 buffer supplemented with 200 mM KCl. The excitation wavelength was 280 nm. (**A**) Fluorescence emission spectra of BlaI under native and unfolded conditions (0 M and 5 M GdmCl ('GdnHCl') respectively). A.U., arbitrary units. (**B**) Variation of the λ_{max} of BlaI (○) and BlaI-NTD (●) as a function of GdmCl concentration.

77, 90 and 116. Only the latter two tyrosine residues are located in BlaI-CTD. Excitation wavelengths of 280 and 295 nm resulted in similar emission spectra, indicating that the tyrosine residues and, hence, the C-terminal part of the protein do not significantly contribute to the fluorescence spectra.

The BlaI unfolding transition was monitored by following the changes of fluorescence emission intensity and of the maximum fluorescence emission wavelength (λ_{max}) in the presence of different GdmCl concentrations. Time-dependent changes in fluorescence emission intensity at increasing GdmCl concentrations showed significant changes within the first 5 min of incubation and no further modification after the next 12 h, suggesting that an incubation time of 2 h was sufficient to achieve equilibrium at any GdmCl concentration (results not shown).

Intrinsic fluorescence emission spectra of native and unfolded proteins are shown in Figure 5(A). The native state is characterized by a λ_{max} at 342 nm. The denaturation process is accompanied by both a decrease in fluorescence intensity and a red-shift of the λ_{max} to 352 nm, which indicates a complete exposure of the tryptophan residues to the aqueous solvent. Figure 5(B) shows the changes in the λ_{max} of BlaI and BlaI-NTD in the presence of increasing GdmCl concentrations, and revealed that BlaI and BlaI-NTD have the same behaviour upon GdmCl unfolding, with their transition midpoints at approx. 4 M GdmCl. Experiments performed in the 1–10 μM protein concentration range gave identical results.

© 2004 Biochemical Society

Urea-induced unfolding of BlaI was also monitored by following the changes in intrinsic tryptophan fluorescence intensity at different denaturant concentrations, and showed that urea-induced unfolding took place at approx. 8 M urea (results not shown).

Unfolding of the C-terminal part of the protein, which cannot be monitored by intrinsic fluorescence, was analysed specifically by far-UV CD and ANS fluorescence.

ANS fluorescence

ANS has been widely used as a sensitive reporter of apolar regions in proteins and as a probe for protein non-native, partially unfolded conformations [23–25]. Such intermediates are characterized by the presence of solvent-exposed hydrophobic clusters. The binding of ANS to apolar regions of proteins results in a significant enhancement of ANS fluorescence intensity and in a pronounced blue-shift of the λ_{max} [23].

The existence of partially unfolded intermediates was investigated by monitoring the binding of ANS to BlaI, BlaI-NTD and BlaI-CTD(His)$_6$ in the presence of increasing GdmCl concentrations (Figures 6A, 6B and 6C respectively). As shown in Figure 6(A), a significant binding of ANS to BlaI occurs at 1.5–1.75 M GdmCl, as revealed by a blue-shift of the λ_{max} from 500 to 445 nm, concomitant with a marked increase in ANS fluorescence intensity (results not shown). A further increase in the GdmCl concentration results in the transfer of the ANS molecules from a hydrophobic to a hydrophilic environment. The binding of ANS to BlaI-NTD (Figure 6B) and BlaI-CTD(His)$_6$ (Figure 6C) was also analysed. BlaI-NTD binds ANS at the same GdmCl concentration as BlaI, but the blue-shift of the λ_{max} value is much smaller. On the other hand, BlaI-CTD(His)$_6$ exhibits quite a different ANS-binding curve measured against GdmCl concentration. The binding of ANS to the native protein fragment suggests the presence of solvent-accessible, non-polar clusters in the C-terminal part of the native BlaI protein, as reported for the homologous *S. aureus* MecI protein [13]. Indeed, the crystallographic 3D (three-dimensional) structure of the dimeric MecI protein reveals the presence of two hydrophobic cavities between the CTDs of each monomer (Figure 2). On the other hand, the variation of ANS binding to BlaI and BlaI-NTD with increasing GdmCl concentrations suggests that BlaI unfolding is accompanied by the accumulation of at least one intermediate with solvent-accessible, non-polar clusters. In order to obtain a further insight into this aspect, the secondary, tertiary and quaternary structures of the protein were studied by CD spectroscopy, gel-filtration chromatography, cross-linking experiments, AU and NMR spectroscopy.

CD studies

Far-UV CD spectroscopy was used to monitor changes in BlaI and BlaI-NTD secondary structures upon GdmCl-induced unfolding. The far-UV CD spectra of BlaI and BlaI-NTD in their native and unfolded states are shown in Figures 7(A) and 7(B) respectively. Both spectra are very similar, and exhibit strong positive maxima at approx. 195 nm and two negative minima at 208 and 222 nm, typical of proteins containing α-helical secondary structures. Incubation of BlaI in GdmCl solutions results both in changes in the shape of the CD spectra and in a GdmCl-concentration-dependent loss in ellipticity. The changes in BlaI ellipticity values at 222 nm with GdmCl concentration highlight two sharp and well-separated transitions (Figure 7A, inset). The first transition occurs at approx. 1 M GdmCl, and is followed by a plateau between 1 and 3.5 M, whereas the second transition occurs

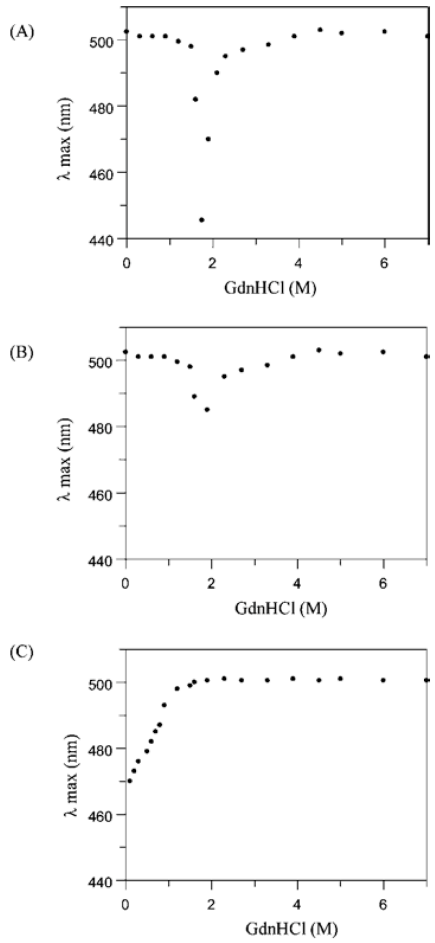


Figure 6 ANS-bound fluorescence emission spectra of BlaI, BlaI-NTD and BlaI-CTD(His)$_6$

Changes of the emission λ_{max} of ANS bound to (**A**) BlaI, (**B**) BlaI-NTD and (**C**) BlaI-CTD(His)$_6$ as a function of GdmCl ('GdnHCl') concentration. Spectra were recorded at 25 °C. Proteins (25 μM) were incubated in 50 mM phosphate buffer, pH 7.6, supplemented with 200 mM KCl and also 13 mM CHAPS for BlaI-CTD(His)$_6$. The ANS concentration was 500 μM. The excitation wavelength was 350 nm.

at approx. 4 M GdmCl. Only approx. 15 % loss of secondary structure is associated with the first transition, indicating only partial unfolding of BlaI molecules under these conditions. The unfolding transition of the isolated BlaI-NTD fragment monitored by CD at 222 nm is monophasic (Figure 7B, inset) with a transition midpoint identical with that of the second transition of the full-length BlaI.

Intrinsic fluorescence data (Figure 5B) suggest that the BlaI-NTD 3D structure is not modified at GdmCl concentrations below 4 M. Thus the transition monitored by far-UV CD at approx. 1 M GdmCl (Figure 7A) is likely to be due to the unfolding of the C-terminal part of the protein. This hypothesis is in agreement with the ANS fluorescence experiments performed with BlaI-CTD(His)$_6$ (Figure 6C), which show that ANS binds to the native CTD and is completely released after the addition of 1 M GdmCl.

As an alternative probe for tertiary structure changes, the BlaI CD spectrum was also measured in the 250–300 nm range. In the absence of denaturant, the spectrum shows a broad negative

© 2004 Biochemical Society

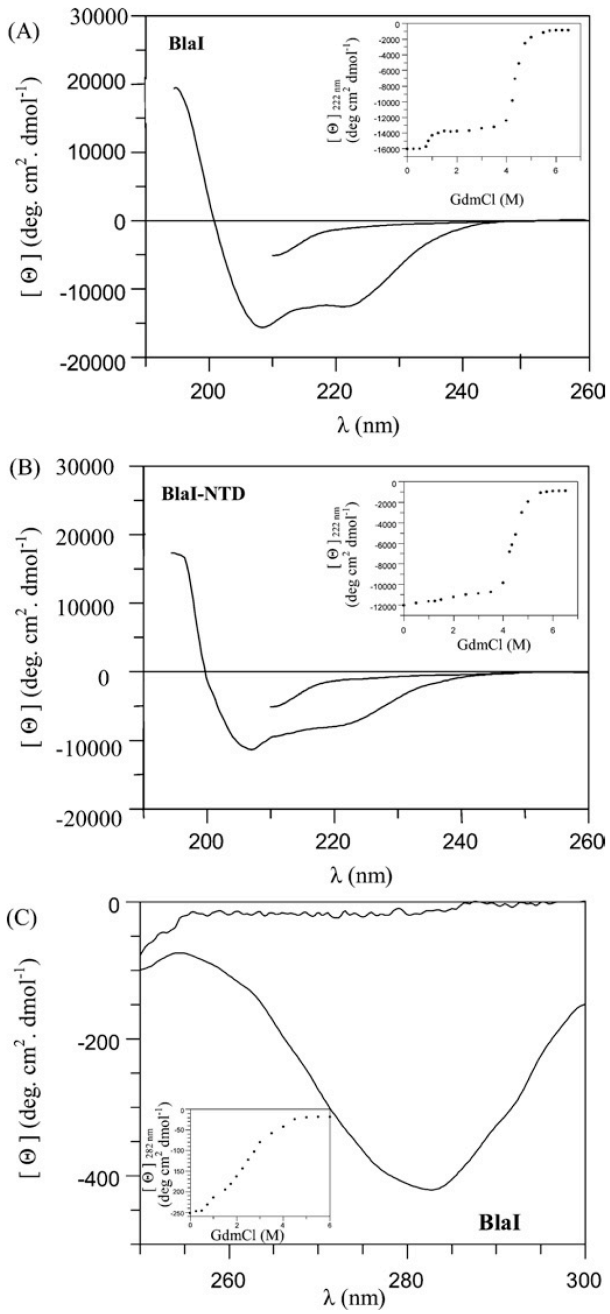


Figure 7 CD studies of BlaI and BlaI-NTD

Far-UV CD spectra of (**A**) BlaI and (**B**) BlaI-NTD under native conditions (50 mM phosphate buffer, pH 7.6, supplemented with 200 mM KCl) and in 5 M GdmCl ('GdnHCl') (same buffer). The protein concentration was 0.5 mg · ml^{-1} in a 0.1 cm cell. The inset shows GdmCl-induced unfolding transition followed at 222 nm. (**C**) Near-UV CD spectra of BlaI under native conditions and in 5 M GdmCl (same buffers). The protein concentration was 1 mg · ml^{-1} in a 1 cm cell. The inset shows GdmCl-induced unfolding transition followed at 282 nm.

band with a minimum at 282 nm (Figure 7C). In the presence of increasing GdmCl concentrations, a progressive decrease in the CD signal is observed between 1–5 M GdmCl (Figure 7C, inset). A possible explanation for this apparent discrepancy between the two optical techniques is that the near-UV CD signal arises from aromatic residues other than those responsible for the intrinsic fluorescence of the protein. Thus the two tyrosine residues (Tyr^{90} and Tyr^{116}) belonging to the CTD might well contribute significantly to the near-UV CD spectrum of the native BlaI protein, and hence explain how the unfolding transitions observed by fluorescence (Figure 5B) and near-UV CD (Figure 7C, inset) measurements are not superimposable. Note that with both techniques, complete unfolding of the protein is achieved in the presence of 5 M GdmCl.

NMR studies

The influence of increasing the GdmCl concentration on the structural changes of BlaI and BlaI-NTD has been monitored by recording ^{1}H/^{15}N-HSQC NMR spectra (Figures 8A and 8B) [26–28]. These NMR spectra highlighted the fact that the BlaI NTD behaves exactly the same upon GdmCl denaturation in the full-length protein or in BlaI-NTD.

In the BlaI native HSQC spectrum (Figure 8A, top left panel), only the cross-peaks corresponding to the 82 N-terminal amino acids can be assigned [11]. CD studies showed that the CTD contains elements of secondary structure, but the dynamic features are such that it cannot be resolved by NMR spectroscopy [11]. Upon increasing the GdmCl concentration (Figures 8A and 8B, top right panels), chemical shift variations of 0.08 p.p.m. (δ_H) maximum are observed in solvent-exposed parts of the protein, i.e. residues 6–9, 73–80 (NTD N- and C-terminal ends respectively), 61–63 (Wing 1) and 68 and 69 (β-strand 3). These shifts are due to the modification of the electrostatic environments of these residues caused by an increasing salt concentration. Chemical shifts of residues belonging to the core of the protein remained unchanged up to 4 M GdmCl (Figures 8A and 8B, lower right panels). At 5 M GdmCl, all signals have been replaced by signals characteristic of the denatured state (Figures 8A and 8B, lower right panels). These data show that a total and abrupt denaturation of the N-terminal part occurs for both BlaI and BlaI-NTD between 4 and 5 M GdmCl.

After removal of GdmCl by extensive dialysis against buffer C, the BlaI and BlaI-NTD native spectra are fully recovered (Figures 8A and 8B, lower left panels). These data, combined with the full recovery of the biological activity after an unfolding/refolding cycle, unambiguously demonstrate that the BlaI and BlaI-NTD unfolding processes are fully reversible.

The NMR results confirm that the NTD of the entire BlaI protein is natively folded at 1 M GdmCl, and hence that the transition monitored by far-UV CD spectroscopy at approx. 1 M GdmCl is most likely to be due to the unfolding of the CTD. Conversely, the transition monitored by far-UV CD, intrinsic fluorescence and NMR at approx. 4 M GdmCl is associated with the unfolding of the NTD. However, these results do not supply any information about the oligomerization state of BlaI. To study this phenomenon, three other experiments were performed: cross-linking experiments with DSP [30,31], sedimentation velocity experiments [32] and SEC [33,34].

Cross-linking with DSP and SDS/PAGE analysis

Figure 9(A) shows the electrophoretic behaviour of the BlaI protein after covalent cross-linking with DSP in the presence of various GdmCl concentrations. In the 0–1 M GdmCl concentration range, the tetrameric, dimeric and monomeric forms coexist. At 2 M GdmCl, both dimeric and monomeric species are present, but the proportion of dimer decreases. At 3 M GdmCl and above, only the monomeric species can be detected.

© 2004 Biochemical Society

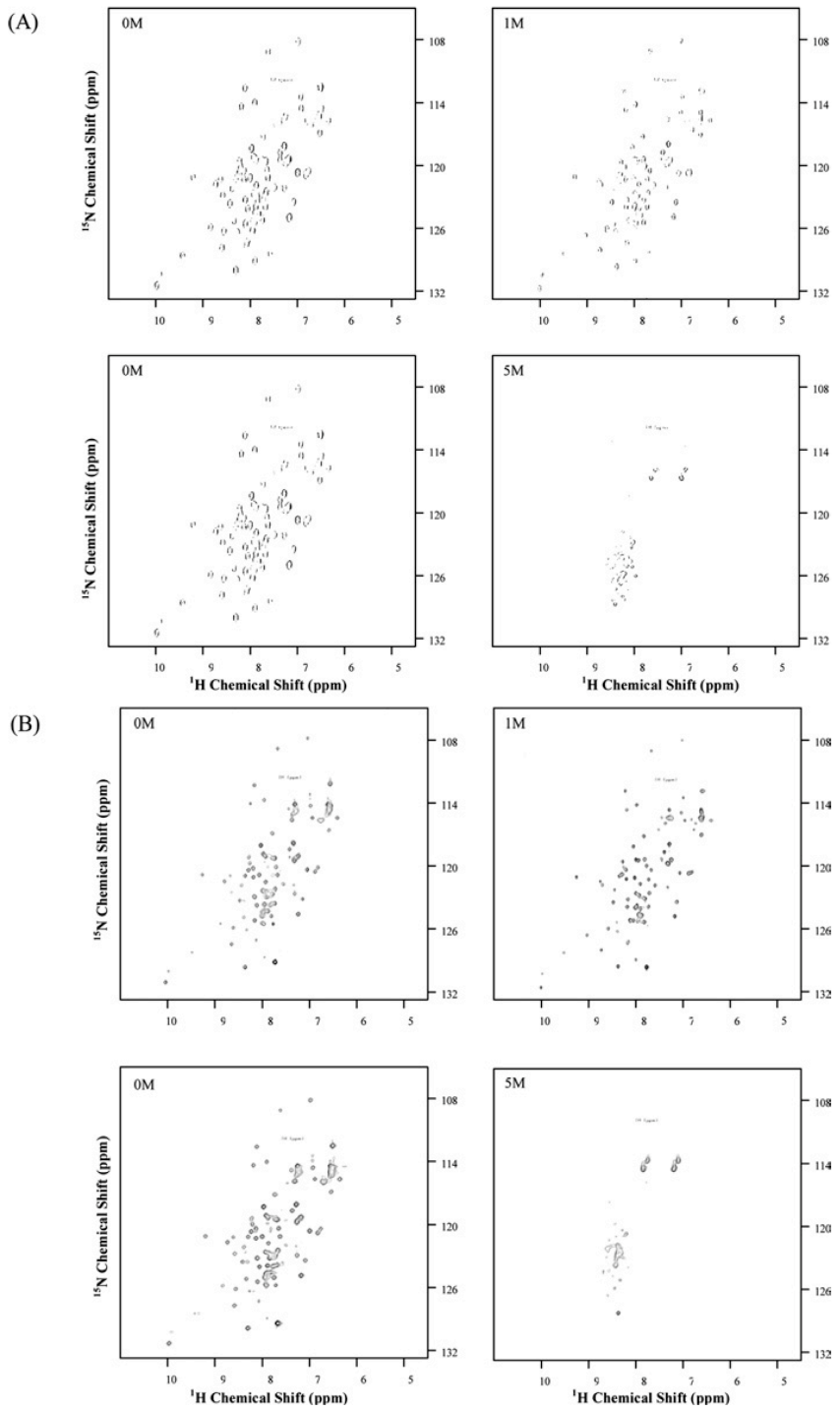


Figure 8 Two-dimensional ^{1}H/^{15}N-HSQC NMR spectra of BlaI and BlaI-NTD

(**A**) BlaI (0.4 mM) and (**B**) BlaI-NTD (0.3 mM) were in 50 mM phosphate buffer, pH 7.6, supplemented with 200 mM KCl and the appropriate GdmCl concentration: 0 M GdmCl (upper left panels), 1 M GdmCl (upper right panels) and 5 M GdmCl (lower right panels). The latter spectra (lower left panels) demonstrate the full renaturation of the unfolded proteins after removal of the denaturant (for further details, see the text and the Materials and methods section). A decrease of peak intensity was observed for all ^{1}H/^{15}N cross-peaks upon increasing the GdmCl concentration. This was due to the change in the quality factor of the probe (Q factor) linked to the presence of high amounts of salt [29].

© 2004 Biochemical Society

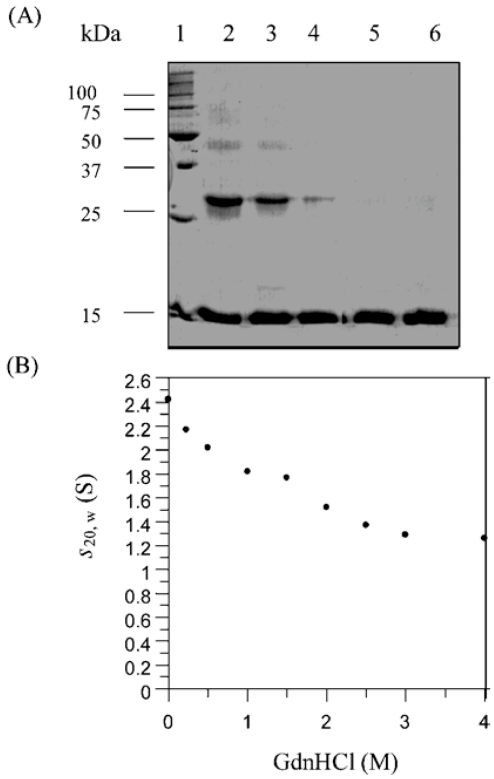


Figure 9 Effect of GdmCl ('GdnHCl') on the oligomerization state of BlaI

(**A**) SDS/PAGE after covalent cross-linking of the BlaI protein at increasing GdmCl concentrations. Protein (20 μM) was incubated for 2 h with DSP and subsequently the reaction was quenched 15 min on ice with 10 μl of 1 M Tris, as described in the Materials and methods section. Samples were electrophoresed on an SDS/4–15 % acrylamide gel. Shown are standard markers (lane 1), BlaI after cross-linking in the absence of GdmCl (lane 2) and protein pre-incubated in a GdmCl concentration of 1, 2, 3 or 4 M before cross-linking (lanes 3–6 respectively). (**B**) Sedimentation velocity experiments: GdmCl concentration-dependence of the sedimentation coefficients $s_{20,w}$ (S). BlaI (60 μM) was incubated in 50 mM Hepes buffer, pH 7.6 containing 200 mM KCl, 1 mM EDTA, 5 % glycerol and supplemented with the required GdmCl concentration. The samples were centrifuged at 40 000 rev./min and scans were collected every 15 min at a wavelength of 280 nm. Sedimentation coefficients were calculated according to the method of Stafford (see [20]).

AU studies

Equilibrium AU experiments were performed in a BlaI concentration range of 10–30 μM, and revealed that monomer, dimer and tetramer co-exist in solution with a [tetramer]/[dimer] ratio equal to 10^{-5}:1 [4]. Sedimentation coefficients ($s_{20,w}$) determined in the 0–5 M GdmCl concentration range are shown in Figure 9(B). Between 0 and 1.5 M GdmCl, a decrease in the $s_{20,w}$ values was observed, probably due to the release of water bound to the protein. Between 1.5 and 2.5 M, a further decrease in $s_{20,w}$ values occurred, probably corresponding to the dissociation of the dimers into monomers. Above 2.5 M GdmCl, only BlaI monomers were observed. In conclusion, cross-linking and sedimentation velocity experiments reveal that the dissociation of the dimeric protein into monomers takes place at approx. 2.5 M GdmCl.

SEC experiments

The behaviour of BlaI upon gel filtration was analysed under both native and denatured conditions [21,33,34]. Figure 10 shows BlaI elution profiles at different denaturant concentrations. Native

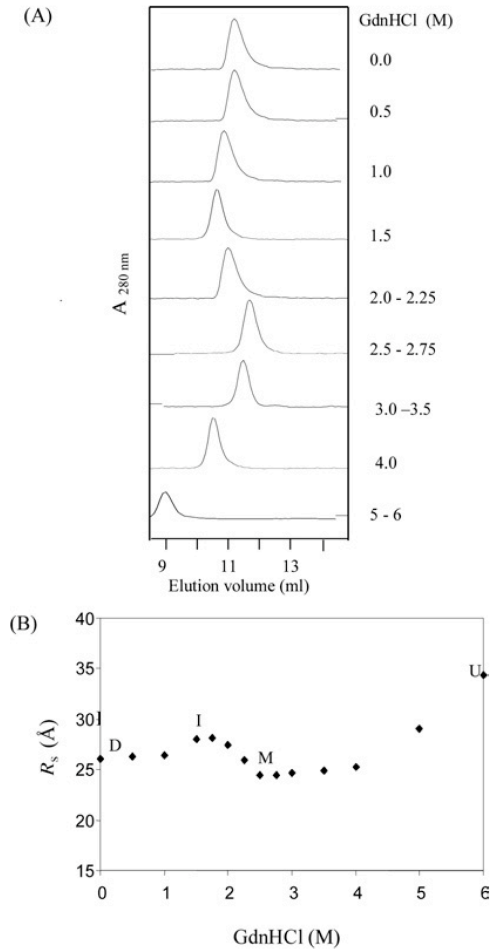


Figure 10 SEC elution profiles of BlaI

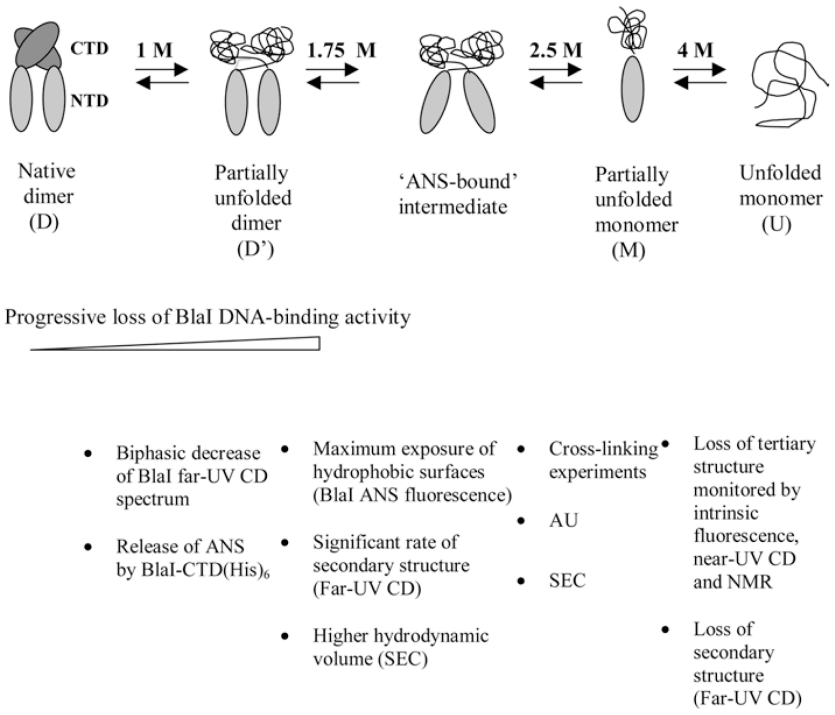
BlaI (50 μM) was incubated for 2 h in 50 mM Hepes buffer, pH 7.6, containing 200 mM KCl, 1 mM EDTA, 5 % glycerol and supplemented with the required GdmCl ('GdnHCl') concentration, and then deposited on to a Superdex 75 column pre-equilibrated with the same GdmCl concentration. Other conditions were as described in the Materials and methods section. (**A**) Elution profiles of BlaI at different GdmCl concentrations. (**B**) R_s of BlaI as a function of GdmCl concentration. D, I, M and U represent respectively the native dimeric protein, the 'ANS-bound' intermediate state, the partially unfolded monomer and the unfolded BlaI protein.

BlaI elutes as a single peak with a V_e of 11.2 ml, corresponding to an R_s of 26.1 Å (1 Å = 0.1 nm). When the GdmCl concentration is increased to 1.5 M, the peak is shifted to a smaller V_e (10.5 ml). This 10 % increase in the R_s value to 28.1 Å suggests a transition to a slightly more expanded compact intermediate state [21]. At approx. 2.5 M GdmCl, the peak is shifted to a higher V_e of 11.9 ml, corresponding to an R_s value of 24.4 Å. This peak should correspond to the monomeric protein, as revealed by cross-linking and sedimentation velocity experiments. The CTD of this monomeric protein should be unfolded, and the N-terminal part 'native-like' (Figures 7A and 7B). At 4 M GdmCl and above, the peak is shifted to a smaller V_e, and at 6 M GdmCl, the V_e corresponds to an R_s value of 34.3 Å (V_e = 8.8 ml).

GdmCl-induced loss of BlaI DNA-binding activity

To highlight the loss of the BlaI DNA-binding property upon GdmCl addition, sedimentation velocity experiments were

© 2004 Biochemical Society



Scheme 1 Representation of the GdmCl-induced unfolding pathway of the BlaI repressor

BlaI denaturation proceeds through the independent sequential unfolding of the two domains and involves three distinct intermediate states: (i) an altered dimer (D') with its CTD in a random state and its NTD in a native conformation; (ii) a partially unfolded 'ANS-bound dimeric intermediate state' with highly exposed hydrophobic surfaces; and (iii) a partially unfolded monomer (M) with an intact NTD. It should be noted that the NTD remains folded in all the intermediate states, showing the high stability of this independent domain.

performed by mixing 60 μM BlaI with 30 μM OP_1 in the appropriate buffer supplemented with increasing GdmCl concentration. Sedimentation coefficients ($s_{20,w}$) corrected for both viscosity and temperature were determined at three different wavelengths (292, 295 and 298 nm), and at 14 different denaturant concentrations. The DNA-binding capacity of BlaI was progressively lost between 0.75 and 1.75 M GdmCl (results not shown).

DISCUSSION

Previous NMR studies suggested that the CTD of the BlaI repressor was not well folded, or that it may exhibit a high degree of flexibility [11]. For better characterizing the BlaI CTD, we purified BlaI-CTD$(His)_6$ in the presence of CHAPS to prevent its aggregation. Far-UV CD studies of BlaI-CTD$(His)_6$ highlighted the presence of α-helices in the purified BlaI-CTD domain (Figure 3). The GdmCl-induced unfolding of BlaI monitored by far-UV CD revealed two transitions at 1 and 4 M GdmCl respectively, whereas the far-UV CD GdmCl-induced unfolding of BlaI-NTD revealed one single transition at 4 M GdmCl. Moreover, ANS binding to BlaI-CTD$(His)_6$ showed that the λ_{max} of emission increased with the GdmCl concentration, and reached a constant maximum value at approx. 1 M GdmCl. At this point, it was assumed that the CTD of BlaI was well folded in the native repressor, and that the unfolding of this domain occurred at 1 M GdmCl. Cross-linking, sedimentation velocity and gel-filtration experiments showed that the dimer–monomer transition occurred at 2.5 M GdmCl. On the basis of all the experiments, the unfolding of BlaI can be described by the multi-step pathway depicted by Scheme 1.

In absence of GdmCl, BlaI is mainly present as a dimer (D state) [9]. After addition of 1 M GdmCl, unfolding of the CTD takes place as shown by comparing the far-UV CD spectra of BlaI and BlaI-NTD (Figures 7A and 7B) and by ANS fluorescence experiments (Figure 6C). The NTD retains its native structure, as shown by NMR experiments (Figure 8B). Cross-linking, sedimentation velocity and gel-filtration experiments show that this first intermediate state (D' state) remains dimeric.

At approx. 1.5–1.75 M GdmCl, the formation of an 'ANS-bound' intermediate state occurs. This intermediate is characterized by the exposure of hydrophobic surfaces as indicated by ANS binding (Figure 6A), an increase in the hydrodynamic volume compared with the folded state (Figure 10) and a high content of native-like secondary structure (Figure 7A). On the basis of these experimental results, one cannot be sure that the first and second steps occur successively and not simultaneously. In fact, the small increase in size of the 'ANS-bound' intermediate state should be concomitant with the unfolding of the CTD. In this case, the first two intermediate states should be replaced by one single intermediate state.

Further increase in the GdmCl concentration (at approx. 2.5 M) results in the dissociation of the dimeric 'ANS-bound' intermediate state into partially unfolded monomers (M state), as shown by cross-linking, sedimentation velocity and gel-filtration experiments. The M state is characterized by a native NTD (see the results for the intrinsic fluorescence and NMR experiments described above) and an unfolded CTD.

At denaturant concentrations above 4 M, intrinsic fluorescence, near- and far-UV CD and NMR spectroscopies showed that BlaI completely loses its secondary and tertiary structures, and exists as

© 2004 Biochemical Society

an expanded monomer (U state). Moreover, $^{1}H/^{15}N$-HSQC NMR spectra and intrinsic fluorescence experiments highlighted the fact that the N-terminal part of BlaI and the isolated BlaI-NTD have exactly the same behaviours upon GdmCl unfolding, i.e. a co-operative and total denaturation of the whole domain. In conclusion, the GdmCl-induced unfolding pathway of the dimeric BlaI repressor occurs in a four-step process. During the first step, the unfolding of the BlaI CTD occurs at approx. 1 M GdmCl, followed in the second step, by the formation of an 'ANS-bound' intermediate state at approx. 1.5–1.75 M GdmCl. The third step at approx. 2.5 M GdmCl involves the dissociation of the dimeric intermediate, and generates two partially unfolded monomers. Finally, the unfolding of the BlaI NTD leads to the complete denaturation of the whole protein at approx. 4 M GdmCl.

These results confirm that the N- and C-terminal BlaI domains are totally independent from each other, and show that they are both well-folded domains [11]. Similar sequential, independent unfolding events of distinct domains have been reported for many proteins, such as SpnHL (*Streptococcus pneumoniae* hyaluronate lyase) [35], papain [36], thermolysin [37], urokinase [38], human interferon-α [39], streptokinase [40], subunit IIABman of the mannose transporter of *Escherichia coli* [41] and aminoacylase [19].

Whereas the NTD is compact and relatively stable upon denaturation, the CTD is more sensitive to the denaturant conditions and exhibits a larger flexibility. On the basis of the sequence similarities between the CTDs of *B. licheniformis* BlaI and *S. aureus* MecI, it can be supposed that BlaI-CTD presents a fold similar to that of MecI-CTD. In particular, most of the hydrophobic residues involved in the C-terminal helix–helix dimeric contacts of the 3D-structure of the MecI repressor are conserved between BlaI and MecI. Garcia-Castellanos et al. [13] reported that two side chains of α-helix 1 (residues Trp^{13} and Asn^{17}) of each NTD contribute to the dimer stabilization by interacting with helices α4 and α5 of the CTD. These interactions do not exist in BlaI, as shown by the NMR spectra, or by comparing the behaviour of the NTD and CTD fragments to that of the whole protein. Moreover, the residues involved in these interactions are not conserved in BlaI (Trp^{13} is a leucine residue in BlaI, and Asn^{17} is a lysine residue). In conclusion, the BlaI CTD must have a fold similar to that of MecI-CTD, but without contacts with the NTD. However, these results do not explain why the (D') intermediate state remains dimeric, although its CTD responsible for dimerization is unfolded. It can be assumed that the dimer is stabilized by hydrophobic interactions between the two unfolded CTDs [42].

As described previously [9], the complex between a BlaI monomer and its DNA target cannot be detected due to the high co-operativity of the binding reaction. By contrast, a BlaI dimer interacts with its DNA target with a high affinity (K_d 4×10^{-11} M). Moreover, Van Melckebeke et al. [11] have shown that BlaI-NTD, which has lost its ability to dimerize, has a DNA-binding affinity 500–1000 times lower than that of the full-length protein. In the present study, we confirm the importance of the CTD of BlaI for its DNA interaction and for its inactivation during the β-lactamase induction. Indeed, our sedimentation velocity experiments showed that BlaI progressively lost its DNA-binding activity at concentrations of between 0.75 and 1.75 M GdmCl. At 1 M GdmCl, the BlaI CTD is unfolded, and approx. 35% of BlaI DNA-binding activity is lost. At 1.75 M GdmCl, BlaI exists in solution as an 'ANS-bound' intermediate state, and has completely lost its high affinity for its DNA target. The two NTDs are probably far from each other, and this prevents the co-operative interactions necessary for the high affinity of BlaI for its DNA target. In conclusion, perturbation of the CTD of BlaI could initiate the disassembly of the BlaI–DNA complex and the loss of the repressor activity. This scenario is in agreement with the hypothesis that a co-activator could destabilize the CTD of the repressor [3]. The high sensitivity to GdmCl and the great flexibility of the CTD reflects thermodynamical and dynamical properties that may be essential for the inactivation of BlaI by the binding of the co-activator during the induction process. In *S. aureus*, the elimination of the BlaI/MecI dimerization domain by a point proteolytic cleavage results in the inactivation of the *S. aureus* repressors during the induction process [43]. Examination of the MecI 3D structure shows that the NTD and CTD interact with each other. On the other hand, in *B. licheniformis* BlaI, NMR experiments highlighted the complete independence of the two domains. Supposing that the contacts between the NTD and CTD of MecI are not crystallographic artefacts, we may surmise that this structural difference between *B. licheniformis* BlaI and *S. aureus* MecI could play a role in the different induction processes of these repressors. Identification of the putative co-activator and analysis of its effect on the quaternary structure of BlaI should completely clarify the induction mechanism.

This work was supported by the Belgian programme of Interuniversity Poles of Attraction initiated by the Federal Office for Scientific Technical and Cultural Affaires (PAI no. P5/33), the 'Fonds National de la Recherche Scientifique' (FNRS, Crédit aux chercheurs no. 1.5202.02 and FRFC no. 2.4530.03 and 2.4545.01) and the 'Communauté Française de Belgique' (projet Tournesol 2003, 03/013). We also thank Eric Sauvage for his assistance in drawing Figure 2, Edwin De Pauw for MS analysis. A. M. and B. J. are Research Associates of the National Fond for Scientific Research (FNRS, Belgium).

REFERENCES

1 Charlier, P., Coyette, J., Dehareng, D., Dive, G., Duez, C., Dusart, J., Fonze, E., Fraipont, C., Frère, J. M., Galleni, M. et al. (1998) Résistance bacterienne aux β-lactamases. Medecine/sciences **14**, 544–549

2 Joris, B., Hardt, K. and Ghuysen, J. M. (1994) Induction of β-lactamase and low-affinity penicillin binding protein 2′ synthesis in Gram-positive bacteria. New Compr. Biochem. **27**, 505–515

3 Filée, P., Benlafaya, K., Delmarcelle, M., Moutzourelis, G., Frère, J. M., Brans, A. and Joris, B. (2002) The fate of the BlaI repressor during the induction of the *Bacillus licheniformis* BlaP β-lactamase. Mol. Microbiol. **44**, 685–694

4 Himeno, T., Imanaka, T. and Aiba, S. (1986) Nucleotide sequence of the penicillinase repressor gene *penI* of *Bacillus licheniformis* and regulation of *penP* and *penI* by the repressor. J. Bacteriol. **168**, 1128–1132

4a Erratum (1987) J. Bacteriol. **169**, 3392

5 Salerno, A. J. and Lampen, J. O. (1986) Transcriptional analysis of β-lactamase regulation in *Bacillus licheniformis*. J. Bacteriol. **166**, 769–778

6 Duval, V., Swinnen, M., Lepage, S., Brans, A., Granier, B., Frère, J. M. and Joris, B. (2003) The kinetic properties of the carboxy terminal domain of the *Bacillus licheniformis* 749/I BlaR penicillin-receptor shed a new light on the derepression of β-lactamase synthesis. Mol. Microbiol. **48**, 1553–1564

7 Zhang, H. Z., Hackbarth, C. J., Chansky, K. M. and Chambers, H. F. (2001) A proteolytic transmembrane signaling pathway and resistance to β-lactams *Staphylococcus aureus*. Science **291**, 1962–1965

8 Wittman, V., Lin, H. C. and Wong, H. C. (1993) Functional domains of the penicillinase repressor of *Bacillus licheniformis*. J. Bacteriol. **175**, 7383–7390

9 Filée, P., Vreuls, C., Hermann, R., Thamm, I., Frère, J. M. and Joris, B. (2003) Dimerization and DNA binding properties of the *Bacillus licheniformis* BlaI repressor. J. Biol. Chem. **278**, 16482–16487

10 Gabelica, V., Vreuls, C., Filée, P., Duval, V., Joris, B. and De Pauw, E. (2002) Advantages and drawbacks of nanospray for studying noncovalent protein-DNA complexes by mass spectrometry. Rapid Commun. Mass Spectrom. **16**, 1723–1728

11 Van Melckebeke, H., Vreuls, C., Gans, P., Filée, P., Llabres, G., Joris, B. and Simorre, J. P. (2003) Solution structural study of BlaI: implications for the repression of genes involved in β-lactam antibiotic resistance. J. Mol. Biol. **333**, 711–720

12 Gajiwala, K. S. and Burley, S. K. (2000) Winged helix proteins. Curr. Opin. Struct. Biol. **10**, 110–116

13 Garcia-Castellanos, R., Marrero, A., Mallorqui-Fernandez, G., Potempa, J., Coll, M. and Gomis-Ruth, F. X. (2003) Three-dimensional structure of MecI. Molecular basis for transcriptional regulation of *staphylococcal* methicillin resistance. J. Biol. Chem. **278**, 39897–39905

© 2004 Biochemical Society

14 Rowland, S. J. and Dyke, K. G. (1990) Tn552, a novel transposable element from *Staphylococcus aureus*. Mol. Microbiol. **4**, 961–975
15 Pace, C. N. (1986) Denaturation and analyses of urea and guanidine hydrochloride curves transition. Methods Enzymol. **131**, 266–280
16 Filée, P., Delmarcelle, M., Thamm, I. and Joris, B. (2001) Use of an Alfexpress DNA sequencer to analyze protein–nucleic acid interactions by band shift assay. BioTechniques **30**, 1044–1051
17 Dumoulin, M., Conrath, K., Van Meirhaeghe, A., Meersman, F., Heremans, K., Frenken, G. L., Muyldermans, S., Wyns, L. and Matagne, A. (2002) Single domain antibody fragments with high conformational stability. Protein Sci. **11**, 500–515
18 Khurana, R. and Udgaonkar, J. B. (1994) Equilibrium unfolding studies of barstar: evidence for an alternative conformation which resembles a molten globule. Biochemistry **33**, 106–115
19 Bai, J. H., Xu, D., Wang, H. R., Zheng, S. Y. and Zhou, H. M. (1999) Evidence for the existence of an unfolding intermediate state for aminoacylase during denaturation in guanidine solutions. Biochim. Biophys. Acta **1430**, 39–45
20 Stafford, W. F. (1992) Methods for obtaining sedimentation coefficient distributions. In Analytical Ultracentrifugation in Biochemistry and Polymer Sciences (Harding, S. E., Rowe, A. J. and Horton, J. C., eds.), pp. 359–393, Royal Society of Chemistry, Cambridge, U.K.
21 Uversky, V. N. (1993) Use of fast protein size exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry **32**, 13288–13298
22 Chang, C. T., Wu, C. S. and Yang, J. T. (1978) Circular dichroic analysis of protein conformation: inclusion of the beta turns. Anal. Biochem **91**, 13–31
23 Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. and Gilmanshin, R. I. (1991) Study of the molten globule intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers **31**, 119–128
24 Stryer, L. (1965) The interaction of naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. J. Mol. Biol. **13**, 482–495
25 Ptitsyn, O. B. (1995) Molten globule and protein folding. Adv. Protein Chem. **47**, 83–229
26 Assfalg, M., Banci, L., Bertinin, I., Turano, P. and Vasos, P. R. (2003) Superoxide dismutase folding/unfolding pathway: role of the metal ions in modulating structural and dynamical features. J. Mol. Biol. **330**, 145–158
27 Van Nuland, N. A., Meijberg, W., Warner, J., Forge, V., Scheek, R. M., Robillard, G. and Dobson, C. M. (1998) Slow cooperative folding of a small globular protein HPr. Biochemistry **37**, 622–637
28 Desmadril, M. and Yon, J. M. (1984) Evidence for intermediates during unfolding and refolding of a two-domain protein, phage T4 lysozyme: equilibrium and kinetic studies. Biochemistry **23**, 11–19

Received 21 April 2004/12 July 2004; accepted 30 July 2004
Published as BJ Immediate Publication 30 July 2004, DOI 10.1042/BJ20040658

29 Gadian, D. G. (ed.) (1984) Nuclear Magnetic Resonance and its Application in Living Systems, Clarendon Press, Oxford
30 Mendoza-Hernandez, G., Minauro, F. and Rendon, J. L. (2000) Aggregation, dissociation and unfolding of glucose dehydrogenase during urea unfolding. Biochim. Biophys. Acta **1478**, 221–231
31 Hermann, R., Jaenicke, R. and Rudolph, R. (1981) Analysis of the reconstitution of oligomeric enzymes by cross-linking with glutaraldehyde: kinetics of reassociation of lactic dehydrogenase. Biochemistry **20**, 5195–5201
32 Behal, R. H., Debuysere, M. S., Demeler, B., Hansen, J. C. and Olson, M. S. (1994) Pyruvate dehydrogenase multienzyme complex. Characterization of assembly intermediates by sedimentation velocity analysis. J. Biol. Chem. **269**, 31372–31377
33 Ackers, G. K. (1970) Analytical gel chromatography of proteins. Adv. Protein Chem. **24**, 343–446
34 Uversky, V. N. and Ptitsyn, O. B. (1994) "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of β-lactamase at low temperature. Biochemistry **33**, 2782–2791
35 Akhtar, M. S. and Bhakuni, V. (2003) *Streptococcus pneumoniae* hyaluronate lyase contains two non-cooperative independent folding/unfolding structural domains: characterization of functional domain and inhibitors of enzyme. J. Biol. Chem. **278**, 25509–25516
36 Sharma, Y. V. and Jagannadham, M. V. (2003) N-terminal domain unfolds first in the sequential unfolding of papain. Protein Pept. Lett. **10**, 83–90
37 Corbett, R. J., Ahmad, F. and Roche, R. S. (1986) Domain unfolding and the stability of thermolysin in guanidine hydrochloride. Biochem. Cell. Biol. **64**, 953–961
38 Bogusky, M. J., Dobson, C. M. and Smith, R. A. (1989) Reversible independent unfolding of the domains of urokinase monitored by 1H NMR. Biochemistry **28**, 6728–6735
39 Skamlova, Z., Kontsekova, E. and Kontsek, P. (1997) Different stabilities of the N- and C-terminal domains of human interferon alpha. Immunol. Invest. **26**, 453–458
40 Conejero-Lara, F., Parrado, J., Azuaga, A. I., Smith, R. A., Ponting, C. P. and Dobson, C. M. (1996) Thermal stability of the three domains of streptokinase studied by circular dichroism and nuclear magnetic resonance. Protein Sci. **5**, 2583–2591
41 Markovic-Housley, Z., Cooper, A., Lustig, A., Flukiger, K., Stolz, B. and Erni, B. (1994) Independent folding of the domains in the hydrophilic subunit IIABman of the mannose transporter of *Escherichiae coli*. Biochemistry **33**, 10977–10984
42 Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. Adv. Protein. Chem. **14**, 1–63
43 Zhang, H. Z., Hackbarth, C. J., Chansky, K. M. and Chambers, H. F. (2001) A proteolytic transmembrane signaling pathway and resistance to β-lactams *Staphylococcus aureus*. Science **291**, 1962–1965

© 2004 Biochemical Society

Bibliographie

Bibliographie BlaI

Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A. & Head, J. F. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 angstrom resolution. *Nat. Struct. Biol.* **8**, 710-714 (2001).

Archer, L. A. & Bosevilac, J. M. Signaling antibiotics resistance in *Staphylococci*. *Sciences* **291**, 915-916 (2001).

Bennett, P. M. & Chopra, I. Molecular basis of beta-lactamase induction in bacteria. *Antimicrob. Agents Chemother*. **37**, 153-158 (1993).

Blumberg, P. M. & Strominger, J. L. Interaction of penicillin with the bacterial cell : penicillin binding proteins and penicillin sensitive enzymes. *Bacteriol. Rev*. **38**, 291-335 (1974).

Brannigan, J. A., Tirodimos, I. A., Zhang, Q. Y., Dowson, C. G. & Spratt, B. G. Insertion of an extra amino acid is the main cause of the low affinity of penicillin-binding protein 2 in penicillin-resistant strains of *Neisseria gonorrhoeae*. *Mol. Microbiol*. **4**, 913-919 (1990).

Brennan, R. G. & Matthews, B. W. The helix-turn-helix DNA binding motif. *J. Biol. Chem*. **264**, 1903-1906 (1989).

Brown, D. F. J. & Reynolds, P. E. Intrinsic resistance to beta-lactam antibiotics in *Staphylococcus aureus*. *FEBS Lett*. **122**, 275-278 (1980).

Calamita, H. G. & Doyle, R. J. Regulation of autolysins in teichuronic acid-containing *Bacillus subtilis* cells. *Mol. Microbiol*. **44**, 601-606 (2002).

Chambers, H. F. Methicillin resistance in staphylococci: Molecular and biochemical basis and clinical implications. *Clin. Microbiol. Rev*. **10**, 781-& (1997).

Chang, C. T., Wu, C. S. & Yang, J. T. Circular dichroic analysis of protein conformation: inclusion of the beta-turns. *Anal. Biochem*. **91**, 13-31 (1978).

Chen, C. S. Y., White, A., Love, J., Murphy, J. R. & Ringe, D. Methyl groups of thymine bases are important for nucleic acid recognition by DtxR. *Biochemistry* **39**, 10397-10407 (2000).

Clarke, S. R. & Dyke, K. G. H. Studies of the operator region of the *Staphylococcus aureus* beta-lactamase operon. *Antimicrob. Agents Chemother*. **47**, 377-389 (2001).

Clarke, S. R. & Dyke, K. G. H. The signal transducer (BlaRI) and the repressor (Blal) of the *Staphylococcus aureus* beta-lactamase operon are inducible. *Microbiology-Sgm* **147**, 803-810 (2001).

Cook, W. J., Kar, S. R., Taylor, K. B. & Hall, L. M. Crystal structure of the cyanobacterial metallothionein repressor SmtB: A model for metalloregulatory proteins. *J. Mol. Biol*. **275**, 337-346 (1998).

Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289-302 (1999).

Cornille, F., Emery, P., Schuler, W., Lenoir, C., Mach, B., Rogues, B. P. & Reith, W. DNA binding properties of a chemically synthesized DNA binding domain of hRFX1. *Nucl. Acids Res*. **26**, 2143-2149 (1998).

D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101-nucleotide core encapsidation signal of the *Moloney murine leukemia* virus. *J. Mol. Biol.***337**, 427-442 (2004).

Dessen, A., Mouz, N., Gordon, E., Hopkins, J. & Dideberg, O. Crystal structure of PBP2x from a highly penicillin-resistant *Streptococcus pneumoniae* clinical isolate - A mosaic framework containing 83 mutations. *J. Biol. Chem.* **276**, 45106-45112 (2001).

Dowson, C. G., Hutchison, A. & Spratt, B. G. Nucleotide sequence of the penicillin-binding protein 2B gene of *Streptococcus pneumoniae* strain R6. *Nucl. Acids Res.* **17**, 7518 (1989).

Dubnau, D. A. & Pollock, M. R. The genetics of *Bacillus licheniformis* penicillinase: a preliminary analysis from studies on mutation and inter-strain and intra-strain transformations. *J. Gen. Microbiol.* **41**, 7-21 (1965).

Duval, V., Swinnen, M., Lepage, S., Brans, A., Granier, B., Franssen, C., Frere, J. M. & Joris, B. The kinetic properties of the carboxy terminal domain of the *Bacillus licheniformis* 749/I BlaR penicillin-receptor shed a new light on the derepression of beta-lactamase synthesis. *Mol. Microbiol.* **48**, 1553-1564 (2003).

Dyke, K. G. H. in *beta-Lactamases* (eds. Hamilton-Miller, J. M. T. & Smith, J. T.) 291-310 (Academic Press, London, 1979).

Fernandez, C., Hilty, C., Wider, G., Guntert, P. & Wuthrich, K. NMR structure of the integral membrane protein OmpX. *J. Mol. Biol.***336**, 1211-1221 (2004).

Filee, P., Delmarcelle, M., Thamm, I. & Joris, B. Use of an ALFexpress (TM) DNA sequencer to analyze protein-nucleic acid interactions by band shift assay. *Biotech.* **30**, 1044 (2001).

Filee, P., Benlafya, K., Delmarcelle, M., Moutzourelis, G., Frere, J. M., Brans, A. & Joris, B. The fate of the Blal repressor during the induction of the *Bacillus licheniformis* BlaP beta-lactamase. *Mol. Microbiol.* **44**, 685-694 (2002).

Filee, P., Vreuls, C., Herman, R., Thamm, I., Aerts, T., De Deyn, P. P., Frere, J. M. & Joris, B. Dimerization and DNA binding properties of the *Bacillus licheniformis* 749/I BlaI repressor. *J. Biol. Chem.* **278**, 16482-16487 (2003).

Fisher, J. F., Meroueh, S. O. & Mobashery, S. Bacterial resistance to beta-lactam antibiotics: Compelling opportunism, compelling opportunity. *Chem. Rev.***105**, 395-424 (2005).

Fogh, R. H., Ottleben, G., Ruterjans, H., Schnarr, M., Boelens, R. & Kaptein, R. Solution Structure of the Lexa Repressor DNA-Binding Domain Determined by H-1-Nmr Spectroscopy. *EMBO J.* **13**, 3936-3944 (1994).

Foster, T. J. Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. *Microbiol. Rev.* **47**, 361-409 (1983).

Frere, J. M. Beta-Lactamases and Bacterial-Resistance to Antibiotics. *Mol. Microbiol.* **16**, 385-395 (1995).

Frère, J. M., Duez, C., Ghuysen, J. M. & Vandekerkhove, J. Occurence of a serine residue in the penicillin-binding site of the exocellular DD-carboxypeptidase-transpeptidase from *Streptomyces R61*. *FEBS Lett.* **70**, 257-260 (1976).

Frère, J. M. & Joris, B. Penicillin sensitive enzymes in peptidoglycan biosynthesis. *CRC Crit. Rev. Microbiol.* **11**, 299-396 (1985).

Frère, J. M. Mécanismes d'action de la pénicilline et des antibiotiques à noyau beta-lactame. *J. Pharm. Belg*. **43**, 107-115 (1988).

Frère, J. M., Joris, B., Granieer, B., Matagne, A., Jacob, F. & Bourguignon-Bellefroid, C. Diversity of the mechanisms of resistance to antibiotics. *Res. Microbiol*. **142**, 705-710 (1991).

Gadian, D. G. *Nuclear magnetic resonance and its application in living systems* (Clearend press, Oxford, 1984).

Gajiwala, K. S. & Burley, S. K. Winged helix proteins. *Curr. Opin. Struct. Biol*. **10**, 110-116 (2000).

Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. & Burley, S. K. Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916-921 (2000).

Garcia-Castellanos, R., Marrero, A., Mallorqui-Fernandez, G., Potempa, J., Coll, M. & Gomis-Ruth, F. X. Three-dimensional structure of MecI - Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J. Biol. Chem*. **278**, 39897-39905 (2003).

Garcia-Castellanos, R., Mallorqui-Fernandez, G., Marrero, A., Potempa, J., Coll, M. & Gomis-Ruth, F. X. On the transcriptional regulation of methicillin resistance - MecI repressor in complex with its operator. *J. Biol. Chem*. **279**, 17888-17896 (2004).

Garnier, J., Gibrat, J. F. & Robson, B. in *Computer Methods for Macromolecular Sequence Analysis* 540-553 (1996).

Ghuysen, J. M. Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. *Bacteriol. Rev*. **32**, 425-464 (1968).

Ghuysen, J. M., Frère, J. M., Leyh-Bouille, M., Nguyen-Distèche, M., Coyette, J., Dusart, J., Joris, B., Duez, C., Dideberg, O., Charlier, P., Dive, G. & Lamotte-Brasseur, J. Bacterial wall peptidoglycan, DD-peptidas abd beta-lactam antibiotics. *Scand. J. Infect. Dis*. **42**, 17-37 (1984).

Ghuysen, J. M., Charlier, P., Coyette, J., Duez, C., Fonze, E., Fraipont, C., Goffin, C., Joris, B. & NguyenDisteche, M. Penicillin and beyond: Evolution, protein fold, multimodular polypeptides, and multiprotein complexes. *Microbial Drug Resist*. **2**, 163-175 (1996).

Ghuysen, J. M. Penicillin-binding proteins. Wall peptidoglycan assembly and resistance to penicillin: Facts, doubts and hopes. *Int. J. Antimicrobial Agents* **8**, 45-60 (1997).

Godsey, M. H., Baranova, N. N., Neyfakh, A. A. & Brennan, R. G. Crystal structure of MtaN, a global multidrug transporter gene activator. *J. Biol. Chem*. **276**, 47178-47184 (2001).

Golemi-Kotra, D., Cha, J. Y., Meroueh, S. O., Vakulenko, S. B. & Mobashery, S. Resistance to beta-lactam antibiotics and its mediation by the sensor domain of the transmembrane BlaR signaling pathway in *Staphylococcus aureus*. *J. Biol. Chem*. **278**, 18419-18425 (2003).

Green, D. W. The bacterial cell wall as a source of antibacterial targets. *Expert Opin. Ther. Targets* **6**, 1-19 (2002).

Gregory, P. D., Lewis, R. A., Curnock, S. P. & Dyke, K. G. H. Studies of the repressor (Blal) of beta-lactamase synthesis in *Staphylococcus aureus*. *Mol. Microbiol*. **24**, 1025-1037 (1997).

Grossman, M. J. & Lampen, J. O. Purification and DNA binding properties of the blaI gene product, repressor for the beta-lactamase gene, blaP, of *Bacillus licheniformis*. *Nucl. Acids Res.* **15**, 6049-6062 (1987).

Hanique, S., Colombo, M. L., Goormaghtigh, E., Soumillion, P., Frere, J. M. & Joris, B. Evidence of an intramolecular interaction between the two domains of the BlaR1 penicillin receptor during the signal transduction. *J. Biol. Chem.* **279**, 14264-14272 (2004).

Hardt, K., Joris, B., Lepage, S., Brasseur, R., Lampen, J. O., Frere, J. M., Fink, A. L. & Ghuysen, J. M. The penicillin sensory transducer, BlaR, involved in the inducibility of beta-lactamase synthesis in *Bacillus licheniformis* is embedded in the plasma membrane via a four-alpha-helix bundle. *Mol. Microbiol.* **23**, 935-944 (1997).

Hartman, B. J. & Tomasz, A. Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. *J. Bacteriol.*. **158**, 513-516 (1984).

Hedge, P. J. & Spratt, B. G. Resistance to beta-lactam antibiotics by re-modelling the active site of an E. coli penicillin-binding protein. *Nature* **318**, 478-480 (1985).

Herzberg, O. Refined crystal structure of beta-lactamase from *Staphylococcus aureus* PC1 at 2.0 A resolution. *J. Mol. Biol.* **217**, 701-719 (1991).

Himeno, T., Imanaka, T. & Aiba, S. Nucleotide sequence of the penicillinase repressor gene penI of *Bacillus licheniformis* and regulation of penP and penI by the repressor. *J. Bacteriol.* **168**, 1128-1132 (1986).

Hiramatsu, K., Asada, K., Suzuki, E., Okonogi, K. & Yokota, T. Molecular cloning and nucleotide sequence determination of the regulator region of mecA gene in methicillin-resistant *Staphylococcus aureus* (MRSA). *FEBS Lett.* **298**, 133-136 (1992).

Hofmeister, A. E. M., Londonovallejo, A., Harry, E., Stragier, P. & Losick, R. Extracellular Signal Protein Triggering the Proteolytic Activation of a Developmental Transcription Factor in *Bacillus-Subtilis*. *Cell* **83**, 219-226 (1995).

Holm, L. & C., S. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138 (1993).

Imanaka, T., Himeno, T. & Aiba, S. Cloning and nucleotide sequence of the penicillinase antirepressor gene penJ of *Bacillus licheniformis*. *J. Bacteriol.*. **169**, 3867-72 (1987).

Iwama, A., Pan, J., Zhang, P., Reith, W., Mach, B., Tenen, D. G. & Sun, Z. J. Dimeric RFX proteins contribute to the activity and lineage specificity of the interleukin-5 receptor alpha promoter through activation and repression domains. *Mol. Cell. Biol.***19**, 3940-3950 (1999).

Jin, C. W., Marsden, I., Chen, X. Q. & Liao, X. B. Sequence specific collective motions in a winged helix DNA binding domain detected by N-15 relaxation NMR. *Biochemistry* **37**, 6179-6187 (1998).

Johnson, A. D., Pabo, C. O. & Sauer, R. T. Bacteriophage lambda repressor and cro protein: interactions with operator DNA. *Methods Enzymol.* **65**, 839-56 (1980).

Joris, B., Ledent, P., Kobayashi, T., Lampen, J. O. & Ghuysen, J. M. Expression in *Escherichia coli* of the carboxy terminal domain of the BLAR sensory-transducer protein of *Bacillus licheniformis* as a water-soluble Mr 26,000 penicillin-binding protein. *FEMS Microbiol. Lett.* **58**, 107-113 (1990).

Joris, B., Galleni, M., Frere, J. M. & Labia, R. Analysis of the Pena Gene of *Pseudomonas-Cepacia-249*. *Antimicrob. Agents Chemother*. **38**, 407-408 (1994).

Joris, B., Hardt, K. & Ghuysen, J. M. Induction of beta-lactamase and low-affinity penicillin binding proteins 2' synthesis in gram-positive bacteria. *New comprehensive biochem*. **27**, 505-515 (1994).

Joris, B., Hardt, K. & Ghuysen, J. M. in *Bacterial cell wall* 505-515 (Elsevier science, New York, N.Y., 1994).

Kar, A. K., Ghosh, A. S., Chauhan, K., Ahamed, J., Basu, J., Chakrabarti, P. & Kundu, M. Involvement of a 43-kilodalton outer membrane protein in beta-lactam resistance of *Shigella dysenteriae*. *Antimicrob. Agents Chemother*. **41**, 2302-2304 (1997).

Kay, L. E., Clore, G. M., Bax, A. & Gronenborn, A. M. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1beta in solution. *Science* **249**, 411-414 (1990).

Kerff, F., Charlier, P., Colombo, M. L., Sauvage, E., Brans, A., Frere, J. M., Joris, B. & Fonze, E. Crystal structure of the sensor domain of the BlaR penicillin receptor from *Bacillus licheniformis*. *Biochemistry* **42**, 12835-12843 (2003).

Knegtel, R. M. A., Fogh, R. H., Ottleben, G., Ruterjans, H., Dumoulin, P., Schnarr, M., Boelens, R. & Kaptein, R. A Model for the Lexa Repressor DNA Complex. *Proteins-Structure Function and Genetics* **21**, 226-236 (1995).

Knox, J. R. & Moews, P. C. Beta-lactamase of *Bacillus licheniformis* 749/C. Refinement at 2 A resolution and analysis of hydration. *J. Mol. Biol*. **220**, 435-455 (1991).

Kobayashi, T., Zhu, Y. F., Nicholls, N. J. & Lampen, J. O. A second regulatory gene, blaR1, encoding a potential penicillin-binding protein required for induction of beta-lactamase in *Bacillus licheniformis*. *J. Bacteriol*.. **169**, 3873-3878 (1987).

Kotra, L. P. & Mobashery, S. beta-Lactam antibiotics, beta-lactamases and bacterial resistance. *Bulletin de l'Institut Pasteur* **96**, 139-150 (1998).

Lai, E., Clark, K. L., Burley, S. K. & Darnell, J. E. Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function. *Proc Natl Acad Sci U S A*. **90**, 10421-10423 (1993).

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486 (1996).

LeClerc, J. E., Li, B. G., Payne, W. L. & Cebula, T. A. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**, 1208-1211 (1996).

Lepre, C. A. & Moore, J. M. Microdrop screening: A rapid method to optimize solvent conditions for NMR spectroscopy of proteins. *J. Biomol. NMR* **12**, 493-499 (1998).

Levy, S. B. Active efflux, a common mechanism for biocide and antibiotic resistance. *Journal of Applied Microbiology* **92**, 65S-71S (2002).

Levy, S. B. & Marshall, B. Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine* **10**, S122-S129 (2004).

Lewis, R. A., Curnock, S. P. & Dyke, K. G. H. Proteolytic cleavage of the repressor (BlaI) of beta-lactamase synthesis in *Staphylococcus aureus*. *FEMS Microbiology Letters* **178**, 271-275 (1999).

Lewis, R. A. & Dyke, K. G. H. MecI represses synthesis from the beta-lactamase operon of *Staphylococcus aureus*. *Antimicrob. Agents Chemother*. **45**, 139-144 (2000).

Lim, D. & Strynadka, N. C. J. Structural basis for the beta-lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Nat. Struct. Biol.* **9**, 870-876 (2002).

Lipsitch, M. The rise and fall of antimicrobial resistance. *Trends in Microbiology* **9**, 438-444 (2001).

Livermore, D. M. Beta-Lactamases in Laboratory and Clinical Resistance. *Clinical Microbiology Reviews* **8**, 557-& (1995).

Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.* **10**, 1033-1038 (2003).

Mallorqui-Fernandez, G., Marrero, A., Garcia-Pique, S., Garcia-Castellanos, R. & Gomis-Ruth, F. X. Staphylococcal methicillin resistance: fine focus on folds and functions. *FEMS Microbiology Letters* **235**, 1-8 (2004).

Malouin, F. & Bryan, L. E. Modification of penicillin-binding proteins as mechanisms of beta-lactam resistance. *Antimicrob. Agents Chemother*. **30**, 1-5 (1986).

Matsuhashi, M., Song, M.D., Ishino, F., Wachi, M., Doi, M., Inoue, M., Ubukata, K., Yamashita, N., Konno, M. Molecular cloning of the gene of a penicillin-binding protein supposed to cause high resistance to beta-lactam antibiotics in *Staphylococcus aureus*. *J. Bacteriol.*. **167**, 975-980 (1986).

Mazel, D. & Davies, J. Antibiotic resistance in microbes. *Cell. Mol. Life Sci*. **56**, 742-754 (1999).

McKinney, T. K., Sharma, V. K., Craig, W. A. & Archer, G. L. Transcription of the gene mediating methicillin resistance in *Staphylococcus aureus* (*mecA*) is corepressed but not coinduced by cognate mecA and beta-lactamase regulators. *J. Bacteriol.***183**, 6862-6868 (2001).

Nathan, C. Antibiotics at the crossroads. *Nature* **431**, 899-902 (2004).

Nicholas, R. A. & Strominger, J. L. Relations between beta-lactamases and penicillin-binding proteins: beta-lactamase activity of penicillin-binding protein 5 from *Escherichia coli* . *Rev. Infect. Dis.* **10**, 733-738 (1988).

Nicholls, N. J. & Lampen, J. O. Repressor gene, blaI, for *Bacillus licheniformis* 749 beta-lactamase. *FEBS Lett*. **221**, 171-183 (1987).

Nielsen, J. B. & Lampen, J. O. Membrane-bound penicillinases in Gram-positive bacteria. *J. Biol. Chem.* **257**, 4490-4495 (1982).

Nikaido, H. Prevention of Drug Access to Bacterial Targets - Permeability Barriers and Active Efflux. *Science* **264**, 382-388 (1994).

Nikaido, H. Multidrug efflux pumps of gram-negative bacteria. *J. Bacteriol.***178**, 5853-5859 (1996).

Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews* **67**, 593 (2003).

Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.* **7**, 215-219 (2000).

Parr, T. R., Moore, R. A., Moore, L. V. & Hancock, R. E. Role of porins in intrinsic antibiotic resistance of *Pseudomonas cepacia*. *Antimicrob. Agents Chemoter*. **31**, 121-123 (1987).

Philippon, A., Dusart, J., Joris, B. & Frere, J. M. The diversity, structure and regulation of beta-lactamases. *Cell. Mol. Life Sci*. **54**, 341-346 (1998).

Richmond, M. Dominance of the inducible state in strains of *Staphylococcus aureus* containing two distinct penicillinase plasmids. *J. Bacteriol*. **90**, 370-374 (1965).

Rohrer, S. & Berger-Bachi, B. FemABX peptidyl transferases: a link between branched-chain cell wall peptide formation and beta-lactam resistance in gram-positive cocci. *Antimicrob. Agents Chemother*. **47**, 837-846 (2003).

Rowland, S. J. & Dyke, K. G. H. Tn552, a novel transposable element from *Staphylococcus aureus*. *Mol Microbiol*. **4**, 961-975 (1990).

Safo, M. K., Zhao, Q. X., Ko, T. P., Musayev, F. N., Robinson, H., Scarsdale, N., Wang, A. H. J. & Archer, G. L. Crystal structures of the BlaI repressor from *Staphylococcus aureus* and its complex with DNA: Insights into transcriptional regulation of the *bla* and *mec* operons. *J. Bacteriol*.**187**, 1833-1844 (2005).

Salerno, A. J. & Lampen, J. O. Transcriptional analysis of beta-lactamase regulation in *Bacillus licheniformis*. *J. Bacteriol*.. **166**, 769-778 (1986).

Salerno, A. J. & Lampen, J. O. Differential transcription of the bla regulatory region during induction of beta-lactamase in *Bacillus licheniformis*. *FEBS Lett*. **227**, 61-65 (1988).

Schaaff, F., Reipert, A. & Bierbam, G. An elevated mutation frequency favors development of vancomycin resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother*. **46**, 3540-3548 (2002).

Schade, M., Turner, C. J., Kuhne, R., Schmieder, P., Lowenhaupt, K., Herbert, A., Rich, A. & Oschkinat, H. The solution structure of the Z alpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc. Natl. Acad. Sci. U.S.A*. **96**, 12465-12470 (1999).

Sharma, Y. V. & Jagannadham, M. V. N-terminal domain unfolds first in the sequential unfolding of papain. *Protein Pept. Lett*. **10**, 83-90 (2003).

Sherratt, D. J. & Collins, J. F. Analysis by transformation of the penicillinase system in *Bacillus licheniformis*. *J Gen Microbiol* **76**, 217-230 (1973).

Spratt, B. G. Resistance to Antibiotics Mediated by Target Alterations. *Science* **264**, 388-393 (1994).

Strausak, D. & Solioz, M. CopY is a copper-inducible repressor of the *Enterococcus hirae* copper ATPases. *J. Biol. Chem*. **272**, 8932-8936 (1997).

Tipper, D. & Strominger, J. L. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc. Natl. Acad. Aci. USA* **54**, 1133-1141 (1965).

Tugarinov, V., Choy, W. Y., Orekhov, V. Y. & Kay, L. E. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. U.S.A*. **102**, 622-627 (2005).

Ubukata, K., Yamashita, N. & Konno, M. Occurrence of a beta-lactam-inducible penicillin-binding protein in methicillin-resistant *staphylococci*. *Antimicrob. Agents Chemother*. **27**, 851-857 (1985).

van Heijenoort, J. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology* **11**, 25R-36R (2001).

Van Melckebeke, H., Vreuls, C., Gans, P., Filee, P., Llabres, G., Joris, B. & Simorre, J. P. Solution structural study of BlaI: Implications for the repression of genes involved in beta-lactam antibiotic resistance. *J. Mol. Biol.***333**, 711-720 (2003).

VanZile, M. L., Chen, X. H. & Giedroc, D. P. Structural characterization of distinct alpha 3N and alpha 5 metal sites in the cyanobacterial zinc sensor SmtB. *Biochemistry* **41**, 9765-9775 (2002).

Vartak, N. B., Reizer, J., Reizer, A., Gripp, J. T., Groisman, E. A., Wu, L. F., Tomich, J. M. & Saier, M. H. Sequence and evolution of the FruR protein of *Salmonella typhimurium*: a pleiotropic transcriptional regulatory protein possessing both activator and repressor functions which is homologous to the periplasmic ribose-binding protein. *Res. Microbiol*. **142**, 951-963 (1991).

Vreuls, C. Thèse de doctorat (Université de Liège, Faculté des sciences, CIP, Belgique, 2004).

Vreuls, C., Filee, P., Van Melckebeke, H., Aerts, T., De Deyn, P., Llabres, G., Matagne, A., Simorre, J. P., Frere, J. M. & Joris, B. Guanidinium chloride denaturation of the dimeric *Bacillus licheniformis* Blal repressor highlights an independent domain unfolding pathway. *Biochem. J*. **384**, 179-190 (2004).

Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97-100 (1997).

Waley, S. G. in *In the chemistry of beta-lactams*. (ed. Hall, C. a.) 198-228 (Glasgow, 1992).

Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775-781 (2000).

Ward, J. & Curtis, C. A. The biosynthesis and linkage of teichuronic acid to peptidoglycan in *Bacillus licheniformis*. *Eur. J. Biochem*. **122**, 125-132 (1982).

Wilke, M. S., Hills, T. L., Zhang, H. Z., Chambers, H. F. & Strynadka, N. C. J. Crystal structures of the apo and penicillin- acylated forms of the BlaR1 beta-lactam sensor of *Staphylococcus aureus*. *J. Biol. Chem*. **279**, 47278-47287 (2004).

Wishart, D. S. & Sykes, B. D. The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *J. Biomol. NMR* **4**, 171-180 (1994).

Wittman, V. & Wong, H. C. Regulation of the penicillinase genes of *Bacillus licheniformis*: interaction of the pen repressor with its operators. *J. Bacteriol.*. **170**, 3206-3212 (1988).

Wittman, V., Lin, H. C. & Wong, H. C. Functional domains of the penicillinase repressor of *Bacillus licheniformis*. *J. Bacteriol.*. **175**, 7383-7390 (1993).

Woody, R. W. *Peptides*. **7**, 15-114 (1985).

Zhang, H. Z., Hackbarth, C. J., Chansky, K. M. & Chambers, H. F. A proteolytic transmembrane signaling pathway and resistance to beta-lactams in *staphylococci*. *Science* **291**, 1915-1916 (2001).

Zheng, N., Fraenkel, E., Pabo, C. O. & Pavletich, N. P. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev*. **13**, 666-674 (1999).

Bibliographie Hadamard

N.B. Les références citées uniquement dans les articles ne sont pas répertoriées dans cette liste.

Anderson, W. A. in *Encyclopedia of Nuclear Magnetic Resonance* (eds. Grant, D. M. & Harris, R. K.) 168-176 (1996).

Ashida, J., Nakai, T. & Terao, T. ID NMR Separation of overlapping powder patterns by selective fr irradiation and switching-angle spinning. *Chem. Phys. Lett.* **168**, 523-528 (1990).

Aue, W. P., Bartholdi, E. & Ernst, R. R. 2-Dimensional Spectroscopy - Application to Nuclear Magnetic-Resonance. *J. Chem. Phys.***64**, 2229-2246 (1976).

Bax, A., Mehlkopf, A. F. & Smidth, J. *J. Magn. Reson.* **35**, 167-169 (1979).

Bersch, B., Rossy, E., Coves, J. & Brutscher, B. Optimized set of two-dimensional experiments for fast sequential assignment, secondary structure determination, and backbone fold validation of C-13/N-15-labelled proteins. *J. Biomol. NMR* **27**, 57-67 (2003).

Bloch, F., Hansen, W. & Packard, M. E. *Phys. Rev.* **69**, 127 (1946).

Blümich, B. & Ziessow, D. Saturation in Hadamard NMR spectroscopy and its description by a correlation expansion. *J. Magn. Reson.* **46**, 385 (1982).

Bodenhausen, G. & Ernst, R. R. Direct Determination of Rate Constants of Slow Dynamic Processes by Two-Dimensional Accordion Spectroscopy in Nuclear Magnetic-Resonance. *J. Am. Chem. Soc.* **104**, 1304-1309 (1982).

Bolinger, L. & Leigh, J. S. Hadamard spectroscopic imaging (HSI) for multivolume localization. *J. Magn. Reson.* **80**, 162-167 (1988).

Bougault, C., Feng, L., Glushka, J., Kupce, E., & Prestegard, J. Quantitation of rapid proton-deuteron amide exchange using Hadamard spectroscopy. *J. Biomol. NMR.* **109**, 385-390 (2004).

Brutscher, B., Morelle, N., Cordier, F. & Marion, D. Determination of an Initial Set of Noe-Derived Distance Constraints for the Structure Determination of N-15/C-13-Labeled Proteins. *J. Magn. Reson. B* **109**, 238-242 (1995).

Brutscher, B., Boisbouvier, J., Kupce, E., Tisne, C., Dardel, F., Marion, D. & Simorre, J. P. Base-type-selective high-resolution C-13 edited NOESY for sequential assignment of large RNAs. *J. Biomol. NMR* **19**, 141-151 (2001).

Brutscher, B. Combined frequency- and time-domain NMR spectroscopy. Application to fast protein resonance assignment. *J. Biomol. NMR* **29**, 57-64 (2004).

Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. M. 4-Dimensional C-13/C-13-Edited Nuclear Overhauser Enhancement Spectroscopy of a Protein in Solution - Application to Interleukin 1-Beta. *Biochemistry* **30**, 12-18 (1991).

Ernst, R. R. & Anderson, W. A. *Rev. Sci. Instrum.* **37**, 93-102 (1966).

Ewing, B. *Chem. Phys.* **147**, 121 (1990).

Freeman, R. Shaped radiofrequency pulses in high resolution NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 59-106 (1998).

Frydman, L., Scherf, T. & Lupulescu, A. The acquisition of multidimensional NMR spectra within a single scan. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15858-15862 (2002).

Furtig, B., Richter, C., Wohnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *Chembiochem* **4**, 936-962 (2003).

Gesmar, H., Led, J. J. & Abildgaard, F. Improved methods for quantitative spectral analysis of NMR data. *Prog. Nucl. Magn. Reson. Spectrosc.* **22**, 255-288 (1990).

Goelman, G., Subramanian, V. H. & Leigh, J. S. Transverse Hadamard Spectroscopic Imaging technique. *J. Magn. Reson.* **89**, 437 (1990).

Goelman, G., Walter, G. & Leigh, J. S. Hadamard spectroscopic imaging technique as applied to the study of human muscles. *Magn. Reson. Med.* **25**, 349-354 (1992).

Golay, M. J. E. Muti-slit spectrometry. *J. Opt. Soc. Am.* **39**, 439-444 (1949).

Greferath, M., Blümich, B., Griffith, W. M. & Hoatson, G. L. Saturation in deuteron Hadamard spectroscopy of solids. *J. Magn. reson. A* **102**, 73-80 (1993).

Hadamard, J. *Bull. Sci. Math.*, 240-248 (1893).

Hoch, J. C. & Stern, A. S. in *Nuclear Magnetic Resonance of Biologica Macromolecules, Pt A* 159-178 (2001).

Hu, H. T., De Angelis, A. A., Mandelshtam, V. A. & Shaka, A. J. The multidimensional filter diagonalization method - II. Application to 2D projections of 2D, 3D, and 4D NMR experiments. *J. Magn. Reson.* **144**, 357-366 (2000).

Jeener, J. *Ampère International Summer School (Basko Polje)* (1971).

Kaiser, R. Application of the Hadamard transform to NMR spectrometry with pseudonoise excitation. *J. Magn. Reson.* **15**, 44-63 (1974).

Kim, S. & Szyperski, T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J. Am. Chem. Soc.* **125**, 1385-1393 (2003).

Kovacs, H., Moskau, D. & Spraul, M. Cryogenically cooled probes - a leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.* **46**, 131-155 (2005).

Krishnamurthy, K. Hadamard excitation sculpting. *J. Magn. Reson.* **153**, 144-150 (2001).

Kupce, E. & Freeman, R. Multisite Correlation Spectroscopy with Soft Pulses - a New Phase-Encoding Scheme. *J. Magn. Reson. Series A* **105**, 310-315 (1993a).

Kupce, E. & Freeman, R. Techniques for Multisite Excitation. *J. Magn. Reson.* **105A**, 234-238 (1993b).

Kupce, E. & Freeman, R. Wide-Band Excitation with Polychromatic Pulses. *J. Magn. Reson. A* **108**, 268-273 (1994).

Kupce, E., Matsuo, H. & Wagner, G. Homonuclear decoupling in proteins. *Biol. Magn. Reson.* **16**, 149-193 (1999).

Kupce, E. & Freeman, R. Fast multi-dimensional Hadamard spectroscopy. *J. Magn. Reson.* **163**, 56-63 (2003a).

Kupce, E. & Freeman, R. Two-dimensional Hadamard spectroscopy. *J. Magn. Reson.* **162**, 300-310 (2003b).

Kupce, E. & Freeman, R. Frequency-domain Hadamard spectroscopy. *J. Magn. Reson.* **162**, 158-165 (2003c).

Kupce, E. & Freeman, R. Fast multi-dimensional NMR of proteins. *J. Biomol. NMR* **25**, 349-354 (2003d).

Kupce, E., Nishida, T. & Freeman, R. Hadamard NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **42**, 95-122 (2003).

Kupce, E. & Freeman, R. Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 6429-6440 (2004).

Lian, L. Y. & Middleton, D. A. Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **39**, 171-190 (2001).

Malmodin, D. & Billeter, M. High-throughput analysis of protein NMR spectra. *Prog. Nucl. Magn. Reson. Spectrosc.* **46**, 109-129 (2005).

Marion, D. Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. *J. Biomol. NMR.* **32**, 141-150 (2005).

Morris, G. A. *J. Magn. Reson.* **29**, 433 (1978).

Nishida, T., Widmalm, G. & Sandor, P. Hadamard Long-Range Proton Carbon Coupling-Constant Measurements with Band-Selective Proton Decoupling. *Magn. Reson. Chem.***33**, 596-599 (1995).

Nishida, T., Widmalm, G. & Sandor, P. Hadamard long-range proton-carbon coupling constant measurements with pulsed field gradients. *Magn. Reson. Chem.***34**, 377-382 (1996).

Nuzillard, J. M. & Freeman, R. Correlation Spectroscopy with 2 Simultaneous Soft Pulses (Tsetse). *J. Magn. Reson. A* **112**, 72-82 (1995).

Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12366-12371 (1997).

Pervushin, K. Impact of Transverse Relaxation Optimized Spectroscopy (TROSY) on NMR as a technique in structural biology. *Quarterly Reviews of Biophysics* **33**, 161-197 (2000).

Pervushin, K., Vogeli, B. & Eletsky, A. Longitudinal H-1 relaxation optimization in TROSY NMR spectroscopy. *J. Am. Chem. Soc.* **124**, 12898-12902 (2002).

Purcell, E. M., Torrey, H. G. & Pound, R. V. *Phys. Rev.* **69**, 37-38 (1946).

Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014-8015 (2005).

Simorre, J. P., Brutscher, B., Caffrey, M. S. & Marion, D. Assignment of Nmr-Spectra of Proteins Using Triple-Resonance 2-Dimensional Experiments. *J. Biomol. NMR* **4**, 325-333 (1994).

Szyperski, T., Wider, G., Bushweller, J. H. & Wuthrich, K. 3d C-13-N-15-Heteronuclear 2-Spin Coherence Spectroscopy for Polypeptide Backbone Assignments in C-13-N-15-Double-Labeled Proteins. *J. Biomol. NMR* **3**, 127-132 (1993).

Tomlinson, B. L. & Hill, H. D. W. Fourier synthesized excitation of nuclear magnetic resonance with application to homonuclear decoupling and solvent line suppression. *J. Chem. Phys.* **59**, 1775–1784 (1973).

Van Melckebeke, H., Simorre, J. P. & Brutscher, B. Amino acid-type edited NMR experiments for methyl-methyl distance measurement in C-13-labeled proteins. *J. Am. Chem. Soc.* **126**, 9584-9591 (2004).

Van Melckebeke, H., Simorre, J. P. & Brutscher, B. Suppression of artifacts induced by homonuclear decoupling in amino-acid-type edited methyl 1H-13C correlation experiments. *J. Magn. Reson.* **170**, 199-205 (2004).

Van Melckebeke, H., Pardi, A., Boisbouvier, J., Simorre, J. P. & Brutscher, B. Resolution-enhanced base-type-edited experiment for RNA. *J. Biomol. NMR*, accepté (2005).

Walsh, J. L. A Closed Set of Orthogonal Functions. *Ann. J. Math.* **55**, 524 (1923).

Wang, H., Ciobanu, L. & Webb, A. Reduced data acquisition time in multi-dimensional NMR spectroscopy using multiple-coil probes. *J. Magn. Reson.* **173**, 134-139 (2005).

Wijmenga, S. S. & van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 287-387 (1998).

Zwahlen, C., Vincent, S. J. F., Gardner, K. H. & Kay, L. E. Significantly improved resolution for NOE correlations from valine and isoleucine (C-gamma 2) methyl groups in N-15,C-13- and N-15,C-13,H-2-labeled proteins. *J. Am. Chem. Soc.* **120**, 4825-4831 (1998).

Résumé : La RMN est une méthode de choix pour la détermination de la structure tridimensionnelle des protéines et des acides nucléiques en solution. Cependant, la taille des systèmes que l'on peut étudier actuellement par RMN est limitée. Dans la première partie de ce travail, la structure du répresseur BlaI de la β-lactamase de *B. licheniformis* 749/I et son interaction avec l'ADN ont été étudiées par RMN avec des méthodes classiques. Ces résultats ont permis de mieux caractériser la répression des gènes de plusieurs mécanismes de résistance aux antibiotiques, incluant la résistance à la méthicilline de la souche pathogène *S. aureus*. Le deuxième volet de ce travail concerne l'implémentation de filtres Hadamard qui augmentent la résolution des spectres dans certaines expériences de RMN multidimensionnelle. Ces filtres permettent de séparer les pics de corrélation des protéines et des acides nucléiques selon le type d'acide aminé et le type de base, respectivement. Ces développements ouvrent de nouveaux horizons vers l'étude de macromolécules biologiques de plus grosse taille par RMN.

Discipline : biologie structurale, spectroscopie RMN, biophysique et biochimie.

Mots clés : spectroscopie RMN, protéine, ARN, attribution des résonances, structure 3D, interactions moléculaires. Antibiotiques à noyau β-lactame, résistance aux antibiotiques, β-lactamase, PBP2', répression des gènes, répresseur BlaI, répresseur MecI, Winged helix protein, interactions protéine/ADN. Spectroscopie Hadamard, RMN multidimensionnelle, filtres de séparation selon le type d'acide aminé et selon le type de base, augmentation de la résolution, méthodes d'acquisition rapides, méthyle, HCN, CT-HSQC, Effet Overhauser Nucléaire.

Abstract : Liquid state NMR is a method of choice for the determination of three-dimensional structure of proteins and nucleic acids. The size of molecules that can be studied nowadays with NMR is however limited. In the first part of this work, the structure of the β-lactamase repressor BlaI of *B. licheniformis* 749/I and its interaction with DNA have been studied with standard NMR methods. These results provide new insights about the mechanisms of genes repression involved in bacterial resistance to antibiotics, including the methicillin resistance observed for the highly pathogen *S. aureus* strain. The second part of this manuscript concerns the implementation of two new Hadamard filters that enhance the resolution of spectra in certain multidimensional NMR experiments. These filters allow to separate the correlation peaks of proteins and nucleic acids according to the amino-acid type and to the base type, respectively. This opens the way for the study of larger biological macromolecules by NMR.

Discipline : structural biology, NMR spectroscopy, biophysics and biochemistry.

Key words : NMR spectroscopy, protein, RNA, resonance assignment , 3D structure, molecular interaction. β-lactam antibiotic resistance, β-lactamase, PBP2', gene repression, repressor BlaI, repressor MecI, Winged helix protein, protein/DNA interaction. Hadamard spectroscopy, multidimensional NMR, amino-acid-type and base-type filters, enhanced resolution, fast acquisition methods, methyl, HCN, CT-HSQC, Nuclear Overhauser Effect.