Structural and functional studies of NOD1

Florence Manon

UNIVERSITE JOSEPH FOURIER – GRENOBLE I

THESE

Pour obtenir le grade de
Docteur de l'Université Joseph Fourier et du Laboratoire Européen de Biologie Moléculaire

Discipline : Biologie Structurale et Nanobiologie

Présentée et soutenue publiquement le 21 septembre 2006 par

Florence FILIPPETTO-MANON

Etudes structurales et fonctionnelles de NOD1
-
Structural and functional studies of NOD1

Composition du Jury

Présidente	Pr Eva PEBAY-PEYROULA
Rapporteur	Dr Muriel DELEPIERRE
Rapporteur	Dr Michel VERON
Examinateur	Dr Jean-Pierre SIMORRE
Examinateur	Dr Winfried WEISSENHORM
Directeur de thèse	Dr Stephen CUSACK

Thèse préparée au sein du laboratoire européen de biologie moléculaire (EMBL)

A mon mari Sylvain,

à mes parents,

et à tous ceux que je porte dans mon coeur

Acknowledgments

This thesis is by far the most significant scientific accomplishment in my life. It is the result of four years of hard work, during which I have been accompanied and supported by many people. This thesis would never have been accomplished without them and I want to express my sincere gratitude to all of them.

First of all, I am deeply indebted to my supervisor, ***Dr Stephen Cusack****, for allowing me to conduct this thesis, for his stimulating suggestions, his valuable discussion, his encouragements, and also his belief in me sometimes more than myself.*

I also would like to thank the Région Rhône-Alpes and the EMBL PhD program for funding.

My warm thanks are due to ***Dr Jean-Pierre Simorre****, for welcoming me for more than one year in his laboratory, for giving me the possibility to learn this very interesting and almost unknown technique (nuclear magnetic resonance), for his constant help and friendship, and for the great sportive moments we share in Vercors.*

I would like to acknowledge ***M.D. Gabriel Núñez****, for the opportunity he gave me to spend five months in his laboratory in Ann Arbor, Michigan. I have learned a lot during this period, in cellular biology methods but also in "american lab" and in "american life".*

My sincere thanks are due to my two referees, ***Dr Muriel Delepierre*** *and* ***Dr Michel Véron****, for their detailed review of this thesis.*

I also want to express my gratitude to ***Pr Eva Pebay-Peyroula****, for presiding over my thesis jury and to my Joseph Fourier University thesis committee members,* ***Dr Winfried Weissenhorm*** *and* ***Dr Jean-Pierre Simorre****, for their advice and help.*

I don't want to forget ***Dr Luis Serrano****, who together with* ***Pr Eva Pebay-Peyroula****,* ***Dr Winfried Weissenhorm*** *and* ***Dr Stephen Cusack****, is an advisor of my EMBL thesis committee and gave me some very valuable hints during these four years.*

I would like to acknowledge ***Dr Adrien Favier****, for having taught me all the things I know in nuclear magnetic resonance, for his help during and after our collaboration, and his unlimited patience. And also for the great mountain-biking, race biking and ski-mountaineering trips.*

Especially, I am very obliged to my friend and former colleague, ***Dr Chloë Zubieta****, and also to* ***Dr Carlo Petosa*** *and to* ***Dr Andrew Mc Carthy*** *for english style and grammar corrections for a draft version of my introduction, materials and methods and results, respectively. Thank you for your time.*

I would like to acknowledge all my colleagues and friends I have acquired through the years, for the great moments I share with them, for their useful advice, and for their support. Among

the "EMBL people", I want particularly to thank (again) ***Dr Chloë Zubieta*** *for her advice, her scientific help even after she left EMBL, her friendship, the outdoor WEs and the short but great vacations spent together, her husband* ***Dr Max Nanao****, for almost the same reasons,* ***Dr Thibaut Crépin*** *for his moral support the last months,* ***Dr Luc Bousset****, who helped me with GRASP and other software,* ***Delphine Guilligay*** *for her eternal good mood,* ***Dr Cécile Morlot****, to have taken the risk to work on the same difficult project as myself and helped me to feel less isolated,* ***Dr Amy Wernimont*** *and* ***Dr Dave Schibli*** *for their friendship and the unfortunately too short moments spent together during my stay in Michigan, our secretary* ***Françoise Tronel*** *for her great help for administrative papers, and all others members of my group and of EMBL in general.*

I sincerely thank also "LRMN people": in addition to ***Dr Jean-Pierre Simorre and Dr Adrien Favier*** *mentioned above, I want particularly to list* ***Dr Pierre Gans****,* ***Dr Hélène Van Melckebeke****, and* ***Pierre Schanda*** *for the mountain biking, race biking and via-ferrata moments,* ***Dr Laurence Blanchard*** *and* ***Dr Beate Bersch*** *for useful biochemical discussions, all the predocs and post-docs for our great "soirées thésards/post-docs" that I already miss.*

I wish to express my warm and sincere thanks to ***Dr Amal Amer****,* ***Dr Mathilde Body-Malapel*** *and* ***Dr Thirumala Devi Kanneganti*** *from Núñez's lab for their friendship and support during these months of very intensive work spent in Michigan.*

I am also particularly grateful to my friends, particularly ***Estelle****,* ***Ian****,* ***Prisca****,* ***Caroline****,* ***JB****,* ***Lionel****,* ***Natalia****,* ***Brice****,* ***Laurent****,* ***Greg*** *to list just some of them. Thanks a lot for your support and comprehension concerning my lack of news and time during this last year of PhD. I make the promise to be again available for you now!*

My very special gratitude is due to my parents ***Patrick and Annie*** *for their constant loving support and help in the concretisation of my professional and personal projects. Without you, I would never have accomplished all this difficult journey by myself.*

And last but not least, I am deeply grateful to my husband, ***Sylvain****, for his love and understanding of my sometimes very elevated stress level. Particularly I owe him thanks for understanding my research abroad in Michigan, for his patience tested notably by these very long five months of complete separation, and for his help in every day life.*

Table of contents

List of figures

List of tables

Abbreviations

Å	angström
A1	anti-apoptotic factor 1
ADP	adenosine diphosphatase
ADR	ambiguous distance restraints
AIM	absent in melanoma
AP-1	activator protein-1
Apaf-1	apoptotic protease-activating factor 1
AP-ATPase	apoptotic adenosine triphosphatase
ARIA	ambiguous restraints for iterative assignment
ASC	apoptosis-associated speck-like protein containing a CARD
ATP	adenosine triphosphate
BIR	baculovirus inhibitor of apoptosis repeat
β-ME	beta-mercaptoethanol
BMRB	BioMagResBank (http://www.bmrb.wisc.edu/)
CARD	caspases-activating and recruitment domain
CARDIAK	CARD-containing ICE associated kinase
CATERPILLER	CARD, transcription enhancer, R-(purine-) binding, pyrin, lots of LRRs
CC	coiled-coil
CED	cell death defective
c-FLIP	cellular FADD-like interleukin-1-beta-converting enzyme inhibitory protein
c-IAPs	cellular inhibitor of apoptosis protein
CIITA	class II transactivator
CLAN	CARD, LRR and NACHT
CLARP	caspase-like apoptosis-regulatory protein
CLR	CATERPILLER
CNS	crystallography and NMR system
CREB	cAMP-response element binding protein
CSI	chemical shift index
Da	Dalton
dADP	deoxyadenosine diphosphate
DAP	diaminopimelate
DAPIN	pyrin/AIM/ASC/DD like domain
DD	death-domain
DED	death-effector domain
DMEM	Dulbecco's Modified Eagle Medium
DNA	desoxyribonucleic acid
DTT	dithiotreitol
EBD	effector binding domain
EDTA	ethylene diamine tetraacetic acid
EMBL	european molecular biology laboratory
ERK	extracellular signal regulated kinase
FADD	Fas-associated death domain protein
FBS	foetal bovin serum
FID	free induction decay
FPLC	fast protein liquid chromatography
Glu	glucose
GM-triDAP	DAP-containing GlcNAc-tripeptide muropeptide
GTP	guanosine triphosphate
HBS	HEPES-buffered saline
HEAT	huntingtin, elongation factor 3, protein phosphatase 2A, TOR1
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazin-propansulfonic acid
HET-E	plant *het* gene
H^N	amide proton

HSQC	heteronuclear single quantum correlation
IBS	institut de biologie structural
ICE	interleukin-1β-converting enzyme
iE-DAP	DAP-containing GlcNAc-tripeptide muropeptide
IFN	interferon
Ig	immunoglobulin
IKK	inhibitor of IkappaB kinase
IL	interleukin
IPAF	ICE protease activating factor
IPTG	isopropyl-1-beta-D-thio-galactopyranoside
IκB	inhibitor of nuclear factor kappa-B
JNK	JUN amino-terminal kinase
kDA	kilo Dalton
L	liter
LB	luria bertani
LIC	ligation-independent cloning
LPS	lipopolysaccharide
LRD	ligand-recognition domain
LRR	leucin-rich repeat
M	molar
MALDI-TOF	matrix assisted laser desorption ionization – Time of flight
MAPK	mitogen-activated protein kinase
mDAP	meso DAP
MDP	muramyl dipaptide
mg	milligram
MHC	major histocompatibility complex
MHz	mega Hertz
min	minute
mM	millimolar
mm	millimeter
NACHT	domain found in NAIP, CIITA, HET-E, TPA
NAD	NACHT-associated domain
NAG	N-acetyl glucosamine
NAIP	neuronal apoptosis inhibitor protein
NALP	NACHT, LRR and PYD
NAM	N-acetyl muramic acid
NB-ARC	nucleotide-binding adaptor conserved in Apaf-1, apoptosis protease activating factor-1, R gene products, and CED-4
NBD	nucleotide-binding domain
NF-kB	nuclear factor kappa B
ng	nanogram
Ni-NTA	nickel-nitrilotriacetic acid
NK	natural killer
NLR	NACHT-LRRs protein
NMR	nuclear magnetic resonance
NOD	nucleotide-oligomerization domain
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NTP-ase	nucleoside triphosphatase
OD_{280}	optic density measured at 280 nm
ONPG	*o*-nitrophenyl-β-D galactopyranoside
PAAD	pyrin/AIM/ASC/DD like domain
PAMP	pathogen-associated molecular pattern
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
pI	isoelectric point

PRR	host pattern recognition receptor
PVDF	polyvinylidene fluoride
PYD	pyrin domain
R^1	longitudinal relaxation
R^2	transversal relaxation
RAIDD	RIP-associated ICH-1 homologous protein with a death domain
RDC	residual dipolar coupling
R_{ex}	exchange parameter
RICK	RIP-like interacting CLARP kinase
RIL	A*rgU*, I*leY*, and L*eu*W tRNA genes
RIP2	receptor interacting protein 2
r.m.s.d.	root mean square deviation
RNA	ribonucleic acid
rpm	rotation per minute
R-protein	pathogen-resistance protein
S_2	order parameter
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamid gel electrophoresis
SRF	serum response factor
TALOS	torsion angle likelihood obtained from shift and sequence similarity
TBS	tris buffered saline
TEV	tobacco etch virus
TIR	Toll/IL-1 R motif
TLR	Toll-like receptor
TNF	tumor necrosis factor
TOCSY	total correlation spectroscopy
TRADD	TNFRSF1A-associated via death domain
TRAF	TNF receptor-associated factor
TRIP	TRAF interacting protein
TROSY	transverse relaxation optimized spectroscopy
TP1	transporter peptide I
U	unity
µm	micrometer
w	weight

Foreword

This memory is written in English language, according to the agreements established between the EMBL and the University Joseph Fourier. However also appear in it in French language a substantial introduction and conclusion, as well as a short summary of each chapter.

Avant-propos

Ce mémoire est rédigé en langue anglaise, suivant les accords établis entre l'EMBL et l'Université Joseph Fourier. Cependant y figurent également en langue française une introduction et une conclusion substantielles, ainsi qu'un court résumé de chaque chapitre.

Part I

Introduction

Introduction (français)

Les maladies infectieuses et parasitaires causent plus du tiers des décès humains mondiaux. Les agents responsables de ces maladies sont nommés pathogènes. Ils présentent une grande diversité phylogénétique et incluent notamment les bactéries, les protozoaires, les champignons, les vers et les virus.

- Le système immunitaire

La survie des organismes multicellulaires aux attaques des agents pathogènes est assurée par le système immunitaire. Celui-ci est composé d'un ensemble d'organes (ex : moelle osseuse) (cf. Figure 1), de cellules (ex : lymphocytes) (cf. Figure 2) et de molécules (ex : anticorps) (cf. Figure 3 et Table 1) qui interagissent étroitement pour produire une réponse immunitaire complexe. Cette dernière consiste en deux réponses distinctes mais complémentaires : (i) la réponse immunitaire innée, qui est une propriété partagée par tous les métazoaires, et (ii) la réponse immunitaire adaptative, qui est une caractéristique des vertébrés.

La réponse immunitaire innée est capable de détecter et de détruire la plupart des agents infectieux avant leur prolifération. Elle est immédiate car ne reposant pas sur l'expansion clonale de lymphocytes spécifiques d'un antigène. Chez les vertébrés, les mécanismes de défense innée assurent deux rôles fondamentaux : (i) ils constituent la première ligne de défense active de l'organisme contre les infections et (ii) ils activent la réponse immunitaire adaptative.

L'immunité adaptative a la capacité de reconnaître différents antigènes et également de retenir une mémoire immunologique d'eux, de manière à assurer une réponse immunitaire rapide dans le cas d'une nouvelle invasion.

- Les récepteurs du système immunitaire inné

Dans les cellules animales et végétales, il existe des récepteurs spécialisés dans la reconnaissance de motifs conservés présents uniquement chez les pathogènes appelés PAMPs pour "pathogen associated molecular patterns". Ces PAMPs consistent en des structures relativement invariantes présentes par exemple sur l'enveloppe des procaryotes ou autres pathogènes, et absentes des cellules de l'hôte. Le système immunitaire inné est capable de reconnaître plus de 1000 PAMPs différents parmi lesquels figurent notamment les lipopolysaccharides, composants de l'enveloppe des bactéries à Gram-négatif, le peptidoglycane, trouvé abondamment dans l'enveloppe des bactéries à Gram-positif et à

moindre degré dans celle des bactéries à Gram-négatif, l'acide lipoteïchoique, l'ARN double brin présent uniquement chez les virus, etc (cf. Table 2).

Ces récepteurs capables de reconnaître spécifiquement les PAMPs sont appelés PRRs pour "host pattern recognition receptors". Ils diffèrent de ceux de l'immunité adaptative en cela qu'ils permettent la discrimination d'une classe de pathogènes (par exemple les bactéries à Gram-négatif par détection des LPS) plutôt que celle d'un pathogène donné. De plus ils donnent lieu à des réponses rapides puisque n'impliquant pas de délais imposés par une expansion clonale de lymphocytes comme lors des réponses adaptatives.

Ces PRRs se divisent en deux classes fonctionnelles différentes à la surface cellulaire : les récepteurs d'endocytose et ceux de signalisation.

Les récepteurs d'endocytose sont localisés à la surface des phagocytes (ex : macrophages). Ils reconnaissent les pathogènes et permettent leur attachement aux phagocytes, ce qui initie la phagocytose. Cette dernière est un processus actif au cours duquel le pathogène est entouré par de longs prolongements cytoplasmiques. Il se retrouve ainsi internalisé dans une vésicule de phagocytose dans le cytoplasme même du phagocyte. Cette vésicule fusionne avec des lysosomes (granules cytoplasmiques contenant notamment des enzymes digestives) et le pathogène est alors détruit.

Parmi les PRRs d'endocytose, nous pouvons notamment citer les récepteurs au mannose (Stahl and Ezekowitz, 1998), qui reconnaissent les groupes mannose et fucose terminaux des glycoprotéines microbiennes et des glycolipides, et les "scavenger receptors" (récepteurs "éboueurs") (Linehan et al., 2000) qui reconnaissent les polymères anioniques ou les lipoprotéines acétylées de faible densité (lipopolysaccharides ou peptidoglycane par exemple).

La reconnaissance des pathogènes par les macrophages peut également donner lieu à des réponses immunitaires innées autres que la phagocytose et stimuler l'induction de réponses immunitaires adaptées. Les PRRs impliqués dans ces phénomènes sont les PRRs de signalisation.

La voie de signalisation la plus connue est celle induite par des récepteurs possédant des répétitions riches en leucine (LRRs), les Toll-like receptors (TLRs) (Medzhitov et al., 1997). Ces TLRs constituent une famille de récepteurs transmembranaires conservés au cours de l'évolution et semblant fonctionner exclusivement comme des récepteurs de signalisation. Ils jouent un rôle crucial dans les réponses inflammatoires en activant des facteurs de transcription tel notamment le facteur NF-κB, ce qui induit la sécrétion de cytokines (molécules régulatrices intracellulaires) et chemokines (cytokines chemotactiques) et initie la réaction immunitaire. Ces cytokines déclenchent les défenses immunitaires innées telles que

la fièvre, la phagocytose et l'inflammation et permettent ainsi une réponse immédiate. L'inflammation est une réaction de défense immunitaire stéréotypée du corps à une agression. Il est souvent fait référence aux noms latins suivants pour décrire ses manifestations : *rubor* (rougeur), *calor* (chaleur), *tumor* (gonflement), *dolor* (douleur), *functio laesa* (impotence fonctionnelle). Les TLRs participent également à la réponse immunitaire adaptative en transmettant des signaux secondaires variés nécessaires à l'immunité spécifique à médiation humorale (production d'anticorps) et à l'immunité spécifique à médiation cellulaire (production de lymphocytes T cytotoxiques et d'autres cytokines).

- Les protéines NLRs

Des études récentes ont montré l'existence d'une famille de PRRs de signalisation cytosolique jouant un rôle crucial dans la réponse immunitaire innée (cf. Table 4). Cette famille inclut un nombre grandissant de protéines intracellulaires possédant un domaine d'oligomérisation appelé NBD ou NOD (nucleotide-binding oligomerization domain) telles les protéines de résistance des plantes (protéines R), le facteur régulateur de l'apoptose Apaf-1 et les protéines de mammifères NOD (protéines possédant des répétitions riches en leucine ou LRRs). Ces protéines sont dénommées NLRs (NACHT-LRRs), NOD-LRRs ou également CATERPILLER.

Les protéines NLRs assurent une surveillance intracellulaire en détectant des PAMPs provenant de composants de la paroi bactérienne. Elles sont caractérisées par trois domaines fonctionnels différents : un domaine de reconnaissance du ligand carboxy-terminal contenant généralement des séries de LRRs capables de reconnaître les PAMPs provenant des composants de la paroi bactérienne (Martinon and Tschopp, 2005), un domaine central liant les nucléotides (NBD) nécessaire à la dimérisation de la protéine qui est un prérequis pour la transduction du signal par un domaine amino-terminal effecteur (cf. Figure 9). Le NBD des protéines NLRs des mammifères est constitué de domaines NACHT (nommé d'après les initiales des premières protéines identifiées le possédant : NAIP, CIIA, HET-E et TP1) et NAD (NACHT associated domain). Le domaine NACHT contient sept motifs distincts incluant à son extrémité amino-terminale une boucle P spécifique des ATP/GTPases ainsi qu'un site de fixation de l'ion magnésium (Koonin and Aravind, 2000).

Le domaine effecteur des protéines NLRs est composé de domaines hautement diversifiés impliqués dans des interactions protéine-protéine. Cependant la plupart des protéines NLRs humaines contiennent un domaine pyrine ou un domaine CARD (caspase-activating and recruitment domain) impliqué dans des interactions homotypiques avec des protéines

effectrices possédant le même domaine (Girardin et al., 2002; Inohara and Nunez, 2003; Inohara et al., 2002).

- NOD1 et NOD2

NOD1 (ou CARD4) et NOD2 (ou CARD15) sont deux membres représentatifs de la sous-famille NLR-NOD jouant une rôle important dans les phénomènes d'inflammation et d'apotose. L'existence de NOD1 a été rapportée en 1999 par deux groupes (Bertin et al., 1999; Inohara et al., 1999) sur la base d'une approche génomique pour trouver des homologues d'Apaf-1 et de son homologue chez le nématode CED-4. La découverte de NOD2 a suivi de peu (Ogura et al., 2001b).

Ces deux protéines sont principalement exprimées dans le cytoplasme des cellules épithéliales (Inohara and Nunez, 2003) et des cellules de présentation des antigènes telles les macrophages et les cellules dentritiques (Gutierrez et al., 2002; Inohara and Nunez, 2003; Ogura et al., 2001b; Rosenstiel et al., 2003). Cependant leurs profils d'expression diffèrent. NOD1 est exprimée abondamment dans les cellules primaires épithéliales alors que l'expression de NOD2 dans ces cellules est extrêmement basse voire même indétectable (Hisamatsu et al., 2003a; Hisamatsu et al., 2003b; Kim et al., 2004; Lala et al., 2003). En effet, au niveau de ces cellules, l'expression de NOD2 est réduite uniquement aux cellules de Paneth (Ogura et al., 2003), qui sont des cellules spécialisées localisées à la base des cryptes de l'intestin grêle.

NOD1 et NOD2 possèdent un domaine effecteur composé respectivement d'un ou deux CARD, un NBD comprenant un domaine NACHT et trois NADs, et un domaine de reconnaissance du ligand constitué de dix répétitions riches en leucines.

NOD1 et NOD2 reconnaissent des petits motifs appartenant aux peptidoglycanes bactériens, respectivement le GM-Tri_{DAP} (GlcNAc-MurNAc-L-Ala-g-D-Glu-*meso*-DAP) trouvé principalement dans les peptidoglycanes des bactéries à Gram-négatif (Chamaillard et al., 2003b; Girardin et al., 2003a), et le muramyl dipeptide, qui est le motif minimal, bioactif du peptidoglycane, commun à toutes les bactéries (Girardin et al., 2003b; Girardin et al., 2003c; Inohara et al., 2003) (cf. Figure 11). NOD1 agit donc uniquement comme un détecteur des bactéries à Gram-négatif alors que NOD2 agit comme un détecteur général de toutes les bactéries.

La reconnaissance spécifique du ligand bactérien par les répétitions riches en leucine de NOD1 et de NOD2 (Martinon and Tschopp, 2005) induirait la dimérisation de la protéine NOD considérée (Inohara et al., 2000; Tanabe et al., 2004). A l'heure actuelle, la nature de cette reconnaissance est inconnue ; elle peut être directe ou bien nécessiter la présence de

cofacteurs ou d'autres molécules. La dimérisation de la protéine NOD est une étape indispensable pour le recrutement de la sérine-thréonine kinase RICK au moyen d'interactions homotypiques de type CARD-CARD (Inohara et al., 2000; Ogura et al., 2001b). Les deux CARDs de NOD2 sont nécessaires pour cette interaction avec la kinase RICK (Ogura et al., 2001b). RICK médie ensuite l'activation du facteur de transcription NF-κB en interagissant avec la sous-unité régulatrice du complexe inhibiteur des kinases NFκB, la kinase IKKγ. NF-kB initie la transcription de gènes pro-inflammatoires (cf. Table 3), et notamment de la prointerleukine-1β.

NOD1 et NOD2 ont également été reportées pour activer également les kinases MAP, activation résultant en l'expression de gènes proinflammatoires (Schorey and Cooper, 2003).

- Implication de NOD1 et NOD2 dans les maladies inflammatoires et immunodéficientes

Des variations génétiques de plusieurs protéines NLRs ont été associées à des maladies inflammatoires et/ou immunodéficientes (Hugot et al., 2001; Inohara and Nunez, 2003; McGovern et al., 2005; McGovern et al., 2001; Miceli-Richard et al., 2001; Ogura et al., 2001a). Ceci souligne l'importance de ces protéines comme sentinelles intracellulaires de la défense immunitaire (Opitz et al., 2005; Viala et al., 2004a).

Notamment, des études récentes ont corrélé l'existence de polymorphismes d'insertion-délétion localisés dans un des introns de *NOD1* avec la prédisposition aux maladies suivantes : eczéma (Weidinger et al., 2005), asthme (Hysi et al., 2005), et les maladies inflammatoires de l'intestin (McGovern et al., 2005).

Des polymorphismes dans les régions codantes de *NOD2* sont associées à la maladie de Crohn (Hugot et al., 2001; Ogura et al., 2001b), au syndrome de Blau (Miceli-Richard et al., 2001) et la sarcoïdose juvénile (ou maladie de Besnier-Boeck-Schaumann) (Kanazawa et al., 2005).

- Importance du domaine CARD

NOD1 et NOD2 possèdent des domaines CARDs comme domaines effecteurs. Le domaine CARD appartient à une sous-famille des domaines de mort, au même titre que le domaine de mort, le domaine effecteur de mort et le domaine pyrine. Ces domaines sont caractérisés par un repliement en six hélices α antiparallèles, et dérivent probablement du même ancêtre commun. Chacun de ces domaines interagit avec des membres de la même sous-famille à travers des interactions homotypiques (ex : CARD-CARD ou pyrine-pyrine), formant des dimères ou des trimères. Cependant quelques interactions hétérophyliques entre des membres

de sous-familles différentes ont été reportées, notamment l'interaction entre le domaine CARD de RICK et le domaine de mort de p75 (Khursigara et al., 2001).

Le domaine CARD est trouvé au niveau des prodomaines de certaines caspases (famille de protéases proinflammatoires impliquées dans la mort cellulaire) mais également dans une grande variété de protéines, soit isolé, soit associé à d'autres domaines (Bouchier-Hayes and Martin, 2002) (cf. Figure 16). Il se compose d'approximativement 90 résidus. Plusieurs structures ont été résolues par résonance magnétique nucléaire (RMN) ou cristallographie aux rayons X : RAIDD (Chou et al., 1998) (cf. Figure 17), Apaf-1 (Day et al., 1999; Vaughn et al., 1999; Zou et al., 1997), etc. Une seule structure de complexe CARD-CARD est connue à ce jour, celle d'Apaf-1 avec le prodomaine de caspase-9 (Qin et al., 1999). Cette association repose principalement sur des interactions électrostatiques entre un patch acide de Apaf-1 et un patch basique de procaspase-9.

Les protéines possédant un domaine CARD sont impliquées dans un grand nombre de processus intracellulaires cruciaux (cf. Figure 18) tels la croissance cellulaire, la mort cellulaire ou la synthèse de cytokines. En effet, elle jouent un rôle dans (1) la régulation de l'apoptose (ex : Apaf1/procaspase-9), (2) l'activation du facteur de transcription NF-κB (ex : RICK, Bcl10, CARD5, CARD10, CARD11, CARD14, NOD1, NOD2), et (3) l'activation de la procaspase-1 et secrétion de l'interleukine 1-β (ex : NOD1, NOD2, ASC, Apaf-1). De plus, l'expression et la structure d'un grand nombre de protéines possédant un domaine CARD sont altérées lors de tumeurs et de lymphomes. Le ciblage spécifique des protéines possédant un domaine CARD (donc notamment NOD1 et NOD2) peut donc constituer la base pharmacologique pour le traitement à la fois de cancers et de maladies inflammatoires (cf. Figure 18) (Damiano and Reed, 2004).

- Objectifs de thèse

Le mécanisme précis d'activation de NOD1 et de NOD2 (reconnaissance du ligand par les LRRs, oligomérisation *via* le NBD ainsi que le mécanisme de recrutement de la protéine effectrice RICK) restent très peu connus. Etant donné l'importance de ces protéines dans les maladies inflammatoires, les maladies immunodéficientes et les cancers, il est d'un intérêt général de déterminer la base structurale de ce processus.

Le but de cette thèse était de déterminer la structure tridimensionnelle du domaine CARD de NOD1 et de caractériser son interaction avec le domaine CARD de RICK au niveau moléculaire.

- Organisation du manuscript

La première partie (I) est consacrée à l'introduction, la seconde (II) aux matériels et méthodes et la troisième (III) aux résultats obtenus.

Le chapitre III.1 décrit l'expression et la purification du domaine CARD de NOD1, ainsi que les études préliminaires de RMN. La détermination de la structure en solution ainsi que des études de dynamiques constituent le chapitre III.2. L'expression et la purification du domaine CARD de RICK ainsi que la caractérisation de l'interaction CARD-CARD entre NOD2 et RICK sont présentées chapitre III.3. Une comparaison des interactions CARD-CARD entre NOD1 et RICK et entre Apaf-1 et procaspase-9 est discutée chapitre III.4.

1 Overview of the immune system

1.1 Role of the immune system

More than one third of all human deaths in the world are caused by infectious and parasitic diseases. The agents that cause infectious diseases are phylogenetically diverse and include bacteria, protozoa, fungi, worms and viruses. These agents are commonly referred to as pathogens. It is important to note that most bacteria are not pathogenic, only those that contain specific virulence genes.

Viruses and some other pathogens are able to modify or take control of the host's DNA, threatening the integrity of the genetic information contained in the host genome. Moreover, eukaryotic organisms form a source of high energy molecules and nutrients for these faster propagating microorganisms. The pathogens behave like invaders. In a first step, they find a nutritional niche to colonize. They are able to make themselves at home and to take over the host metabolism for their own replication. Finally, they exit the infected host to spread to an uninfected one, causing irreparable damage.

All multicellular organisms need to defend themselves against attacks by these potentially dangerous invaders. Therefore, they have evolved a powerful and sensitive detection system for this purpose: the immune system. The immune system allows fighting off tumor growth.

1.2 Characteristics

The immune system has diverse roles. The first one is to create a barrier that prevents pathogens from entering the body. Secondly, when an infectious agent succeeds in entering the body, the immune system has to try to detect and eliminate it before it begins to propagate. The immune system is also in charge of eliminating the intruder later in the infection, when it is able to reproduce itself and to cause harm.

Due to the considerable biodiversity of microorganisms, the host immune system has to be both very powerful and specific in terms of detection and elimination of pathogens. It must be specific enough to distinguish the self from the non-self, and precise enough to kill the invaders without harming its own tissues.

1.3 Components

1.3.1 Organs

The immune system is not comprised of a distinct set of organs. Instead it is composed of a complex of organs, cells and molecules, which interact to produce a complete immune response. The organs that compose the immune system can be divided into two categories: the primary lymphoid organs, where the cells originate and many of them also mature, and the secondary lymphoid organs, where most of the immune responses occur (cf. Figure 1).

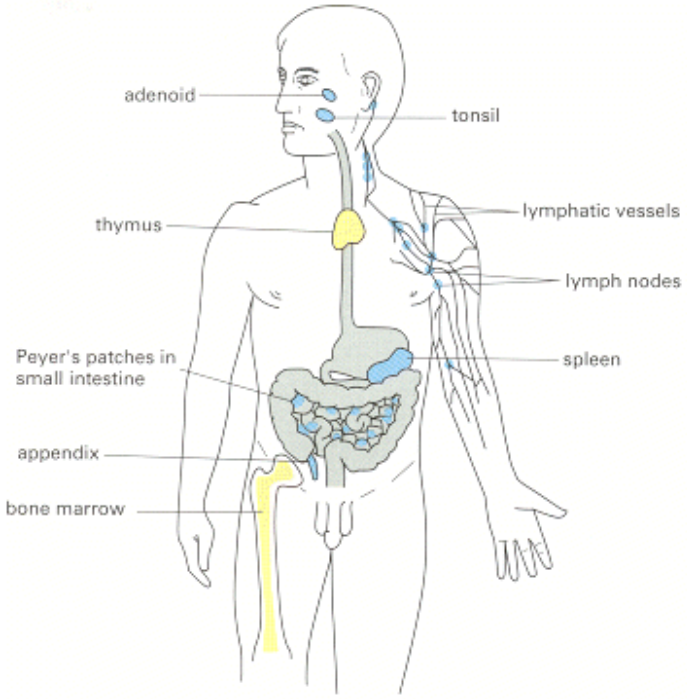

Figure 1. Human lymphoid organs, from Alberts et al., 2002. The bone marrow and thymus, that constitute the primary lymphoid organs, are colored in yellow. Only some secondary lymphoid organs, such for example the spleen, the lymph nodes or the Peyer's patches are shown (blue).

The bone marrow and the thymus, a lobular organ located in the upper chest, form the primary lymphoid organs. The secondary lymphoid organs are composed by the spleen, the lymph nodes and the lymphoid tissues associated with mucosa such for example the Peyer's patches in the small intestine.

1.3.2 Cells

The pluripotent hematopoietic stem cells of the bone marrow give rise to all the cellular elements of the blood (cf. Figure 2). These stem cells differentiate into two categories of stem cells with a more restricted potential: the myeloid stem cells and the lymphoid stem cells.
The myeloid cells differentiate into four types of immune cells: macrophages, dentritic cells, mast cells, and granulocytes. Macrophages form the first type of phagocytes. They derived from monocytes and they are broadly dispersed in the tissues. Dentritic cells act as antigen presenting cells. Mast cells are cells of connective tissues playing roles in allergic reaction, anaphylaxis, wound healing and innate immunity (cf. 1.4.1). Granulocytes (also known as polynuclear leukocytes) are so called because of the multiple staining granules present in their cytoplasm. Granulocytes are subdivided into three types: neutrophils that constitute the second type of phagocytes, eosinophils that help fight infection by parasites, and basophils that may have a function complementary to that of both mast cells and eosinophils.
The lymphoid stem cells give rise to three lineages: the lymphocytes and the natural killer cells. Some lymphocytes differentiate and mature in the bone marrow and are called B cells, whereas others migrate to the thymus where they differentiate into T cells. B and T cells migrate to the secondary lymphoid organs where they can be activated by foreign antigens to become effector B cells and effector T cells. Effector B cells are plasma antibody-secreting cells. Some of them differentiate to memory B cells. Memory B cells present the characteristic to be specific to the antigen that first stimulated their production. These cells are long lived and can respond very quickly upon second exposure to their specific antigen. Effector T cells can be divided into three main classes (depending on the category of receptor they express): the cytotoxic T cells and the suppressor T cells (which both express CD8), and the helper T cells (which express CD4). Cytotoxic T cells function as a killer of cells infected with viruses whereas suppressor T cells control the immune response and maintain the immune system homeostasis. Helper T cells are able to regulate effector lymphocyte function. Natural killer cells (NK cells) are involved in killing virally-infected and malignant cells. They are characterized by the lack of antigen-specific receptors.

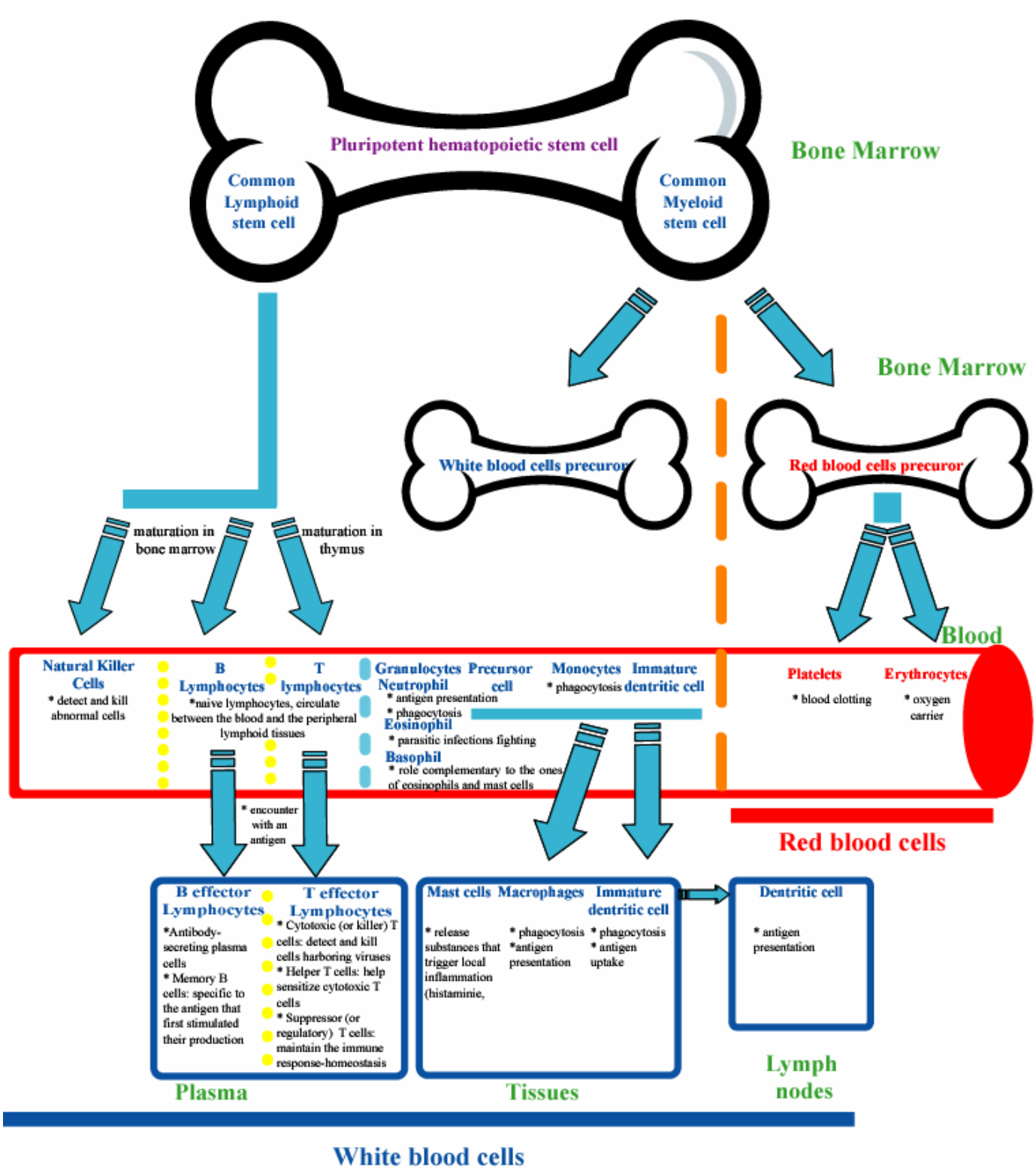

Figure 2. Components of the immune system. All the components of the immune system arise from hematopoietic stem cells in the bone marrow. The pluripotent hematopoietic stem cells of the bone marrow differentiate into two more specificic stem cells: the common lymphoid stem cells that give rise to the different types of lymphocytes, and the common myeloid stem cells that give rise to both the different classes of leucocytes (white blood cells) and also red blood cells.

1.3.3 Molecules

Many molecules playing a role in the immune response are secreted by white blood cells. Among the best known are the antibodies. Antibodies are proteins used by the immune system to recognize and bind an antigen specific to a certain invader and thereby neutralize the invader. They are also commonly referred to as immunoglobulins, even though the terms antibodies and immunoglobulins do not denote exactly the same thing. Immunoglobulins are tissue fluids and blood glycoproteins secreted by B plasma cells that function like antibodies.
The complement system is another important tool of the immune system. It is composed of a series of blood proteins that are originally made in the liver. These proteins are activated by the antibodies and complement their action, hence the name. They are responsible for cell lysis and they also signal to phagocytes when a cell has to be removed.
Several hormones are secreted by the components of the immune system including tymosin, which is able to promote lymphocyte production, and the steroids and corticosteroids, that can suppress the immune system.
In response to an antigen, almost all the cells involved in the immune system produce and secrete low molecular weight (10-30 kDa) soluble proteins critical to the proper functioning of the immune system. These proteins are called cytokines. They are usually secreted but they may be expressed on the cell surface or transiently expressed on the membrane surface and then cleaved. They are often glycosylated to protect them from the harsh environment of the serum. Cytokines are able to bind to cytokine-specific receptors located on other cells and thus to modify the behaviour of these cells. These receptors define the biologic responses of cytokines. Cytokines act as intercellular chemical messenger in the regulation of the immune responses and in the stimulation of hematopoiesis. Their functions also include chemotaxis, immunostimulation, suppression and angiogenesis. They may act in an autocrine manner, in a paracrine manner, and also in an endocrine manner.
Cytokines have diverse characteristics. Firstly, they are functionally redundant (many of them share identical or similar functions). Secondly, they are pleiotropic: they act on numerous cell types and activate different types of responses, eg. differentiation, growth, activation, etc. Furthermore, they can act either in a synergistic or in an antagonistic way.
Cytokines can also be referred to as interleukins (IL) or lymphokines. Due to the diverse and overlapping roles of the cytokines and historical nomenclature schemes, the classification of these proteins is complex (cf. Figure 3).

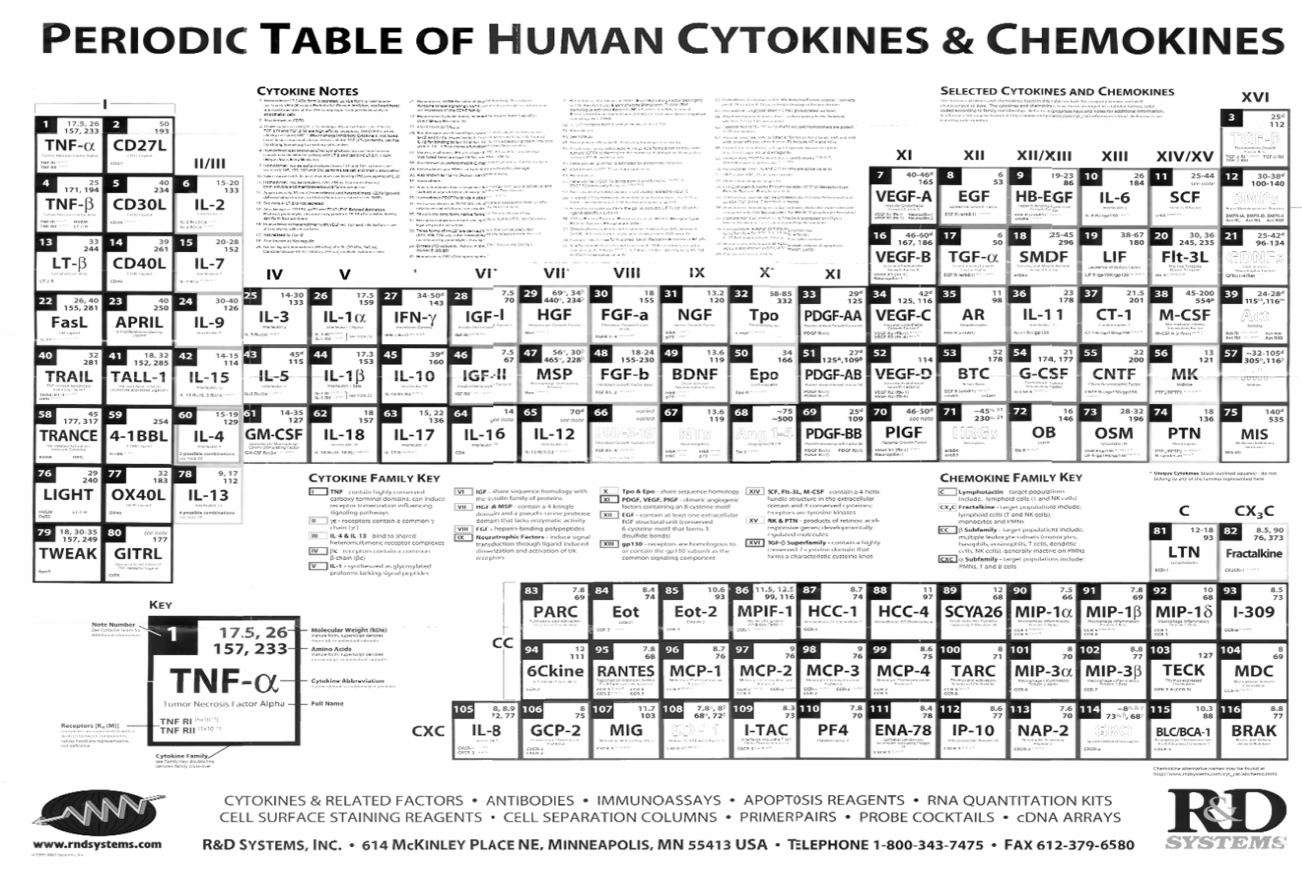

Figure 3. Periodic table of human cytokines and chemokines, from www.rndsystems.com.

Cytokines can be classified into families based on their structure homology: the hematopoietins or four α-helix bundle family (which is composed of three subfamilies known as the interferon subfamily or IFN, the IL-2 subfamily or the IL-10 subfamily); the IL-1 family (in which belong notably IL-1 and IL-18); the IL-17 family; the tumor necrosis factor family (TNF) and the chemokines subfamily. Chemokines are chemotactic cytokines (e.g. the IL-8).

Examples of well characterized cytokines belonging to each of these subfamilies are shown Table 1.

Family	Cytokine	Producer cells	Main actions
Hematopoietins or four α-helix bundle family			
	IFN α	leucocytes	antiviral
	IFN β	fibroblasts	antiviral
	IFN γ	T cells NK cells	inflammation Macrophages activation
	IL-2 (T-cell growth factor)	T cells	T cells proliferation
	IL-10	T and B cells macrophages	suppress macrophages functions
IL-1 family			
	IL-1 α	macrophages and epithelial cells	T cells activation fever inflammation
	IL-1 β	macrophages and epithelial cells	T cells activation fever inflammation
	IL-18	macrophages and Kuppfer cells	T and B cells IFN γ production induction
IL-17 family			
	IL-17	memory B cells	cytokines production promotion
Tumor necrosis factor family			
	TNF-α	macrophages NK cells T cells	inflammation endothelial activation
	TNF-β	T and B cells	tumor killing endothelial activation
Chemokines			
	IL-8	macrophages	neutrophils, basophils and T cells attraction

Table 1. Examples of members of the different cytokines families, their producer cells and actions.

1.4 Immune system responses

The immune system consists of two different but related responses: innate immunity, a property of all metazoans, and adaptive immunity**,** a distinctive quality of vertebrates. Innate immunity is inborn and provides an "all purpose defense" against invaders whereas adaptive immunity has the ability to recognize different antigens and to retain an immunologic memory of them.

1.4.1 Innate immunity

Innate immunity preceded the development of the adaptive immune system in the evolution of vertebrates (Janeway, 1992).

Innate immunity constitutes the first line of host defense after infection. It relies on the ability of the body to recognize the self from the non-self but it is characterized by the lack of specificity to a particular pathogen.

Innate immunity relies on pre-existing barriers or defense mechanisms that a host uses immediately or within the first hours or days following the exposure to almost any new antigen. Innate immunity includes anatomical barriers (e.g. the skin) and mechanical removal (e.g. flushing mechanisms such couching) (cf. 2.2), chemical responses (plasma proteins, e.g. the complement proteins and the cytokines) and cellular responses (e.g. phagocytic cells such neutrophils or monocytes/macrophages, cytotoxic cells such NK cells or cells that release inflammatory mediators such the mast cells, basophils or eosinophils).

Infectious agents that succeed in invading the epithelial surfaces are recognized by cells and molecules of the innate immune system. These pathogens bind to receptors (cf. 2.3.2) located on phagocytic cells, that trigger their subsequent engulfment and destruction by these cells *via* a process called phagocytosis. Furthermore, activated macrophages are also able to secrete cytokines and chemokines. Cytokines and chemokines are inducers of a key process in innate immunity called inflammation.

These two innate defense mechanisms, phagocytosis and inflammation, can also be triggered by the activation of the complement system.

So what exactly is inflammation? This is a means used by the body to restore and maintain homeostasis. We have seen previously that many host defense actors are located in the blood. Inflammation is a localized protective reaction of tissues to irritation, injury or infection which allows the defense cells and molecules to leave the blood and enter the tissues. Inflammation is characterized by heat, pain, stiffness, swelling and sometimes loss of

function. It can be divided in three main steps: (1) local blood flow increases, (2) blood vessel permeability increases and (3) migration of the immune cells and molecules from the blood vessels to the tissues.

Another main role of innate immunity is to activate adaptive immunity using phagocytic cells such macrophages as antigen-presenting cells. The most potent antigen-presenting cells are the immature dentritic cells that localize in many tissues. During an infection, the immature dentritic cells of the infected tissue ingest the invading pathogens or their products by macropinocytosis. They mature and carry the antigens to the secondary lymphoids organs. Then, they present the antigen in association with the major histocompatibility complex (MHC) to the T cells. T cells are effectors of adaptive immunity.

1.4.2 Adaptive immunity

The innate immune system is not always able by itself to fight off infectious agents. In such instances, it calls into play adaptive (or acquired) immunity. The adaptive immune system arose in evolution less than 500 millions years ago. Adaptive immunity consists of responses that are directed toward a specific pathogen. They are more powerful and operate later in the infection. Furthermore, adaptive immunity is also able to retain an immunologic memory of specific pathogens and hence to provide long-lasting protection. This immunity develops throughout life.

Adaptive immunity relies on lymphocytes. It takes several days to become protective because it relies on clonal expansion (so innate responses are important because they act during this time). Clonal expansion refers to the proliferation of B and T cells activated by a specific antigen selection in order to produce a clone of identical cells. This enables the body to have a sufficient number of antigen-specific lymphocytes to mount an effective immune response.

There are two main branches of the adaptive immunity: humoral immunity, which is mediated by B cells, and cell-mediated immunity, which is mediated by T cells. Humoral immunity involves the production of antibodies in response to an antigen and relies upon body fluids, especially the lymph and the blood plasma, to spread the antibodies around the body. These antibodies inactivate the pathogen by blocking its ability to bind receptors on host cells and also mark it for destruction. Cell-mediated immunity involves the production of cytotoxic T cells and cytokines that activate phagocytosis by macrophages in response to an antigen.

Antigen-specific lymphocytes of the adaptive immune response may be activated both by the cytokines released by the cells of the innate immune system, or *via* antigen-presenting cells. When T cells are activated, some of them (cytotoxic or CD8) migrate to the site of infection

where they may kill the infected cells or help macrophages, whereas some others (helper or CD4) remain in the peripheral lymphoid organs where they help B cells.

2 The innate immune system

2.1 Conservation during evolution

Many factors of the innate immune system are very well conserved during evolution. Indeed, there exist a striking conservation of genes encoding recognition molecules, effector molecules as well as signaling pathways involved in innate host defense in organisms as diverse as human, fruit fly (*Drosophila Melanogaster*) and plants (Medzhitov and Janeway, 1998; Medzhitov et al., 1997). This conservation includes the Toll pathway of NF-κB activation of gene function (cf. 2.3.2 and 2.3.3) that is found in vertebrates, invertebrates and plants. This mechanism probably existed in the common ancestor of these three classes of organisms, some six-hundred millions years ago. This conservation underlines the great importance of the innate responses in the defense against microbial pathogens. Furthermore, these ancient components of innate immunity seem to have a crucial role in the development, activation and regulation of the adaptive immune system in vertebrates (Medzhitov et al., 1997).

2.2 Barrier epithelia

Epithelial surfaces include the skin, the respiratory tract, the urinogenital tract and the gastrointestinal tract. They function as a physical shield and prevent microbes from entering tissues. Notably, the tight junctions that hold neighboring cells together constitute a seal against the outside world.

The interior of the epithelial surfaces is composed of a mucus layer. The mucus is a viscous fluid forming a physical barrier that prevents infectious agents from invading the tissues. Additionally, it contains substances that may destroy the pathogens such for example the lyzozyme or antimicrobial peptides named defensins.

Sometimes, some pathogens may be able to cross these epithelial barriers and begin their replication. The cells of the innate immune system, such macrophages, have the capacity to detect them. But how does this recognition work?

2.3 Recognition of the invaders

A big challenge to the innate immune system is the identification, using a small number of receptors (present notably on phagocytic cells), of numerous pathogens possessing a propensity to mutate from self.

In 1989, Charles Janeway Junior (Janeway, 1989) proposed the concept of pattern recognition. He predicted that our immune system would possess pattern recognition receptors (PRR) able to detect conserved motifs on pathogens that are absent in the host. These specific pathogen components are termed pathogen-associated molecular patterns (PAMPs).

2.3.1 Pathogen-associated molecular patterns

The PAMPs are polysaccharides, polynucleotides, polypeptides or complex lipids that are highly conserved and shared by many pathogens but not produced by the host (Schleifer and Kandler, 1972). They may be located internally or externally (Gordon, 2002). They possess crucial roles in the biology of the invading agents, and they are therefore not subject to high mutation rates.

The immune system can recognize more than one thousand different PAMPs. Many of them are products of the bacterial cell wall (cf. Figure 4). The bacterial lipopolysaccharide (LPS) is considered to be the protypical PAMP.

Some of the most relevant PAMPs and the category of pathogens from which they are issued are listed Table 2.

PAMP	Found in	Pathogen
Lipopolysaccharides	Gram-negative cell wall	Gram-negative bacteria
Peptidoclycan	Gram-negative cell wall and Gram-positive cell wall	Gram-negative and Gram-positive bacteria
Lipoteichoic acids	Gram-positive cell wall Fungi cell wall	Gram-positive bacteria Fungi
Pilis	Bacterial pili	Bacteria
Flagellin	Bacterial flagella	Bacteria
N-formylmethionine	Bacterial protein (amino-terminal residue replacing aucaryotic regular protein)	Bacteria
Unmethylated cytosine-guanine dinucleotide	Bacterial and viral nucleic acid	Bacteria Viruses
Double stranded RNA	Viral RNA	Viruses
Glycolipids	Fungi cell wall	Fungi
Zymosan	Fungi cell wall	Fungi
Chitin	Fungi cell wall	Fungi

Table 2. Examples of the most current pathogen-associated molecular patterns from bacteria, fungi and viruses.

Many PAMPs are products of degradation of the peptidoglycan (also known as murein), which is an essential component of the bacterial cell wall. It provides rigidity and structure to the bacterial cell wall (Takeda and Akira, 2001). This bacterial cell wall forms a semi-rigid structure surrounding the cytoplasmic membrane and protects the bacteria from osmotic lysis. Gram-positive bacteria possess many layers of peptidoglycan (20 to 80 nm) (cf. Figure 4(A)), whereas Gram-negative bacteria possess between one and three layers (7 to 8 nm) (cf. Figure 4(B)).

Peptidoglycan is a vast polymer mesh. It presents a crystal lattice structure formed from linear chains of two alternating amino sugars, the N-acetyl glucosamine (NAG) and the N-acetyl muramic acid (NAM), cross-linked by short peptide chains coming off of the NAM (cf. Figure 4(C)) (Van Heijenoort, 2001). These cross-links make the wall strong.

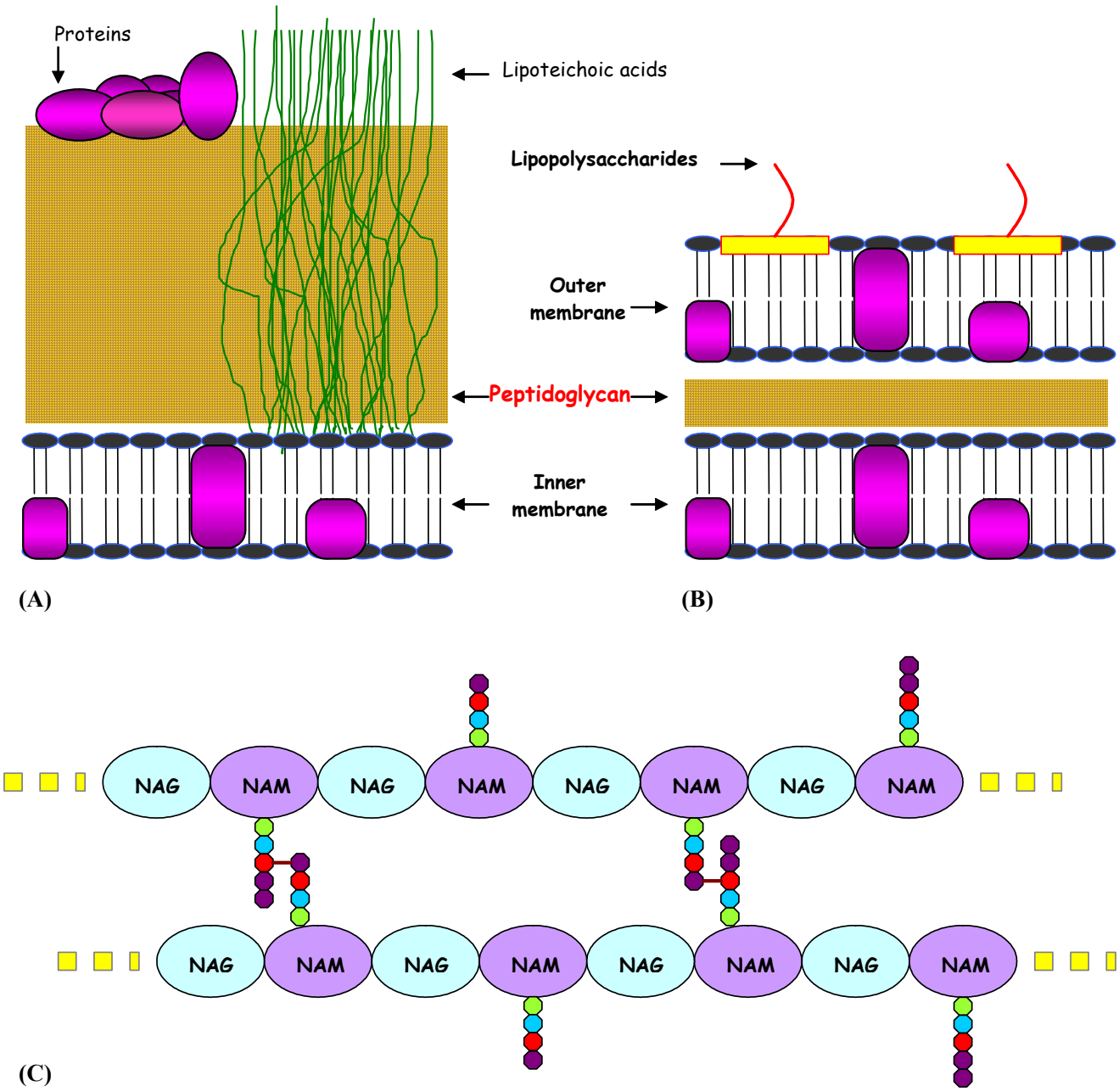

Figure 4. The bacterial cell wall and its components. (A) Cell wall of gram-positive bacteria. (B) Cell-wall of gram-negative bacteria. (C) Peptidoglycan organization.

The nature of the third residue of the stem peptide defines the two main classes of peptidoglycan. Peptidoglycans of Gram-negative bacteria possess a mDAP at this position whereas peptidoglycans of Gram-positive bacteria present a variable residue, which is commonly a lysine (Schleifer and Kandler, 1972).

2.3.2 Pattern-recognition receptors

Animals and plants possess homolog receptors able to sense these PAMPs (Girardin et al., 2002; Inohara and Nunez, 2003; Staskawicz et al., 2001; Takeda and Akira, 2001)}. These receptors can be found in intracellular compartiments, on the cell surface, or secreted into the bloodstream and tissues fluids (Medzhitov and Janeway, 1997). According to functionality, there exist two different classes of PRRs in innate immunity: the endocytic PRRs and the signaling PRRs.
Upon recognition of their PAMPs, these PRRs may initiate opsonization, activation of complement and coagulation cascades, activation of proinflammatory signaling and apoptosis (Barton and Medzhitov, 2002; Janeway and Medzhitov, 2002; Medzhitov, 2001).

2.3.2.1 Endocytic pattern-recognition receptors

The endocytic pattern recognition receptors localize on the surface of phagocytes. They mediate the binding of microorganisms to phagocytes and their subsequent engulfment and destruction. These phagocytic PRRs notably include the mannose receptors (Stahl and Ezekowitz, 1998) and the scavenger receptors (Linehan et al., 2000). Mannose receptors bind to terminal mannose and fucose groups on microbial glycoproteins and glycolipids, whereas scavenger receptors regognize particular anionic polymers and acetylated low-density lipoproteins.

2.3.2.2 Signaling pattern-recognition receptors

Signaling PRRs can identify and bind various PAMPs. This recognition mediates immediate response against the invading pathogen by promoting the synthesis and secretion of intracellular regulatory molecules such the cytokines.
The first signaling PRR was identified in mammals by Charles Janeway Junior, together with Ruslan Medzhitov (Medzhitov et al., 1997), eight years after he firstly published his theory. This receptor was referred as Toll-like receptor 4 (TLR-4), because of its homology to a receptor first discovered and named in fruit fly *Drosophila megalonaster*, the receptor Toll. This founding member of the Toll/Toll-like receptor family was initially described in *Drosophila megalonaster* as a developmental protein essential for embryonic axis establishment (Hashimoto et al., 1988). Its involvement in the fly's immunity to fungal infections and Gram-positive bacterial infection was established in 1996 (Lemaitre et al.,

1996) and in 1997 (Lemaitre et al., 1997). TLR4 was then shown to be implicated in the recognition of LPS and in the consecutive signaling (Poltorak et al., 1998).

Most mammalian species possess between ten and fifteen types of Toll-like receptors (TLRs) (cf. Figure 5). The TLRs are type I transmembrane proteins found in macrophages, dendritic cells, and epithelial cells. They possess a variable number of extracellular amino-terminal leucine-rich repeat (LRR) motifs, followed by a cysteine-rich region, a transmembrane domain, and finally an intracellular Toll/IL-1 R (TIR) motif (Chaudhary et al., 1998; Hashimoto et al., 1988; Medzhitov et al., 1997; Rock et al., 1998).

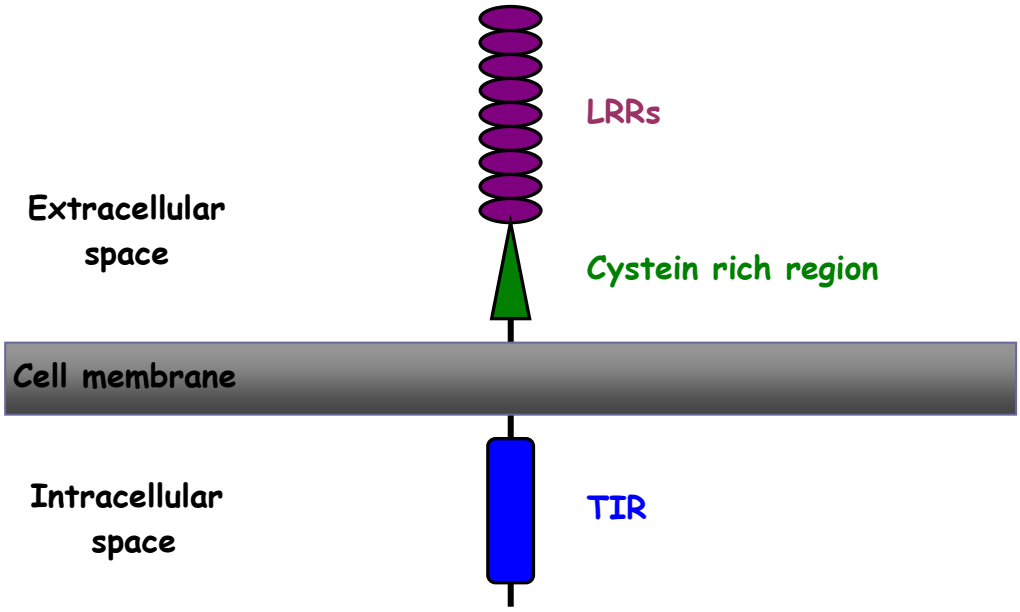

Figure 5. TLR protein structure. TLRs are type I transmembrane proteins possessing a variable number of amino-terminal LRRs followed by a cysteine rich domain, a transmembrane domain, and an intracellular TIR domain

LRRs are short motifs of consensus LxxLxLxxN/CxL. They are found in a diverse range of proteins and are involved in protein-protein or protein-ligand interactions (Kobe and Deisenhofer, 1995). In TLRs, they are thought to be necessary for PAMP binding and the consecutive signaling (Bell et al., 2003; Modlin, 2001). To date, only the X-ray structure of the TLR-3 ectodomain is known (Bell et al., 2005; Choe et al., 2005) (cf. Figure 6). The overall horseshoe-shaped structure of the TLR-3 ectodomain is formed by twenty-three LRRs that are capped at each end by specialized non-LRR domains (called LRRNTT and LRRCT). The extensive β-sheet on the molecule's concave surface forms a platform for several modifications, including insertion in the LRRs and eleven N-linked glycans. This TLR-3 ectodomain structure provides the first indication of how LRR loops can establish distinct pathogen recognition receptors.

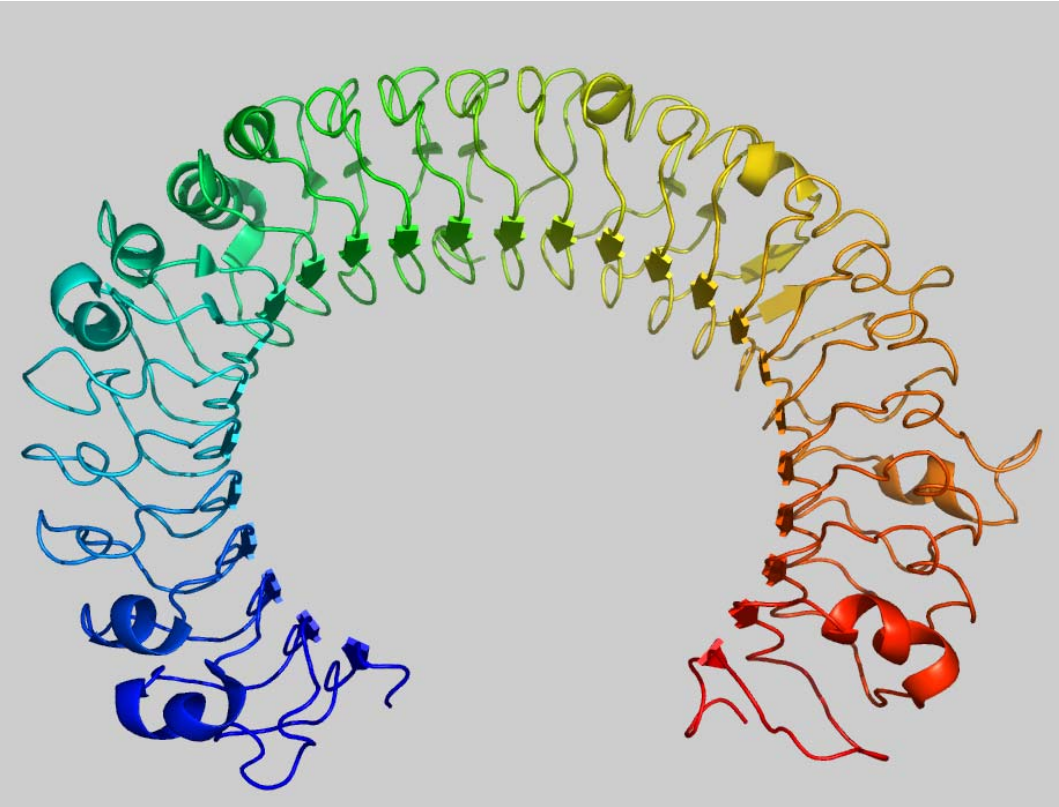

Figure 6. Crystal structure of human TLR-3 ectodomain, from Choe et al., 2005. The human TLR3 ectodomain structure at 2.1 Å forms a large horseshoe-shaped solenoid. This solenoid is stabilized by extensive hydrogen-bonding networks formed by asparagines conserved in the LRRs. TLR-3 is largely masked by carbohydrate, but one face is glycosylation-free, which suggests its potential role in ligand binding and oligomerization. Highly conserved surface residues and a TLR-3 specific LRR insertion form a homodimer interface in the crystal, whereas two patches of positively charged residues and a second insertion would provide an appropriate binding site for double-stranded RNA.

The TIR domain is another important protein-protein interaction domain common in innate immunity and apoptosis signaling (Aravind et al., 2001; Dunne et al., 2003; Dunne and O'Neill, 2003).

To date, ten different TLRs have been identified in humans. Each of them is specialized, often with the help of accessory molecules, in the recognition of a subset of PAMPs (Barton and Medzhitov, 2002; Barton and Medzhitov, 2003; Takeda et al., 2003). For example TLR-2 binds lipoteichoic acid from Gram-positive bacteria, TLR-4 recognizes double-stranded RNA from viruses and LPS from Gram-negative bacteria, TLR-5 recognizes bacterial flagellin, etc. TLRs seem to associate in pairs (e.g. the TLR-1/TLR-2 or the TLR-2/TLR-6 pair) and possess different expression patterns. TLRs play a key role in innate immunity. The recognition of Gram-positive bacteria by TLR-2 and Gram-negative bacteria by TLR-4 enhances phagogytosis and the fusion of the phagosomes with the lysosomes. Furthermore, in all cases, the recognition of the PAMPs by the TLRs initiates a signaling pathway leading to the activation of the transcription factor NF-κB. NF-κB drives the expression of many cytokines

leading, for example, to the synthesis and secretion of some chemokines, TNF-α or IL-1. These effector molecules, as noted earlier, are involved in the pro-inflammatory responses. Notably, IL-1β, which is secreted by macrophages, acts as an "alarm cytokine" and initiates inflammation.

Furthermore, TLRs also initiate adaptive immune responses by triggering diverse secondary signals necessary for humoral immunity leading to antibody production, and for cell mediating immunity such as cytotoxic T-lymphocytes production.

2.3.3 NF-κB signaling pathway

NF-κB is an evolutionarily conserved transcription factor playing a crucial role in many biological processes, notably in the regulation of the immune responses in response to injury, infection or stress (Ghosh et al., 1998). Five NF-κB proteins have been identified in humans: RelA (p65), RelB, c-REL, NF-κB1 (p50/p105) and NF-κB2 (p52/p100) (Chen et al., 1999).

In unstimulated cells, they are found in the cytoplasm as homodimers or heterodimers, each of them able to activate specific genes (Bonizzi and Karin, 2004; Ghosh et al., 1998; Li and Verma, 2002). NF-κB proteins are bound to proteins of the IκB family, which maintain them in an inactive state and prevent their translocation to the nucleus.

Different signaling pathways may induce the activation of NF-κB (cf. Figure 7).

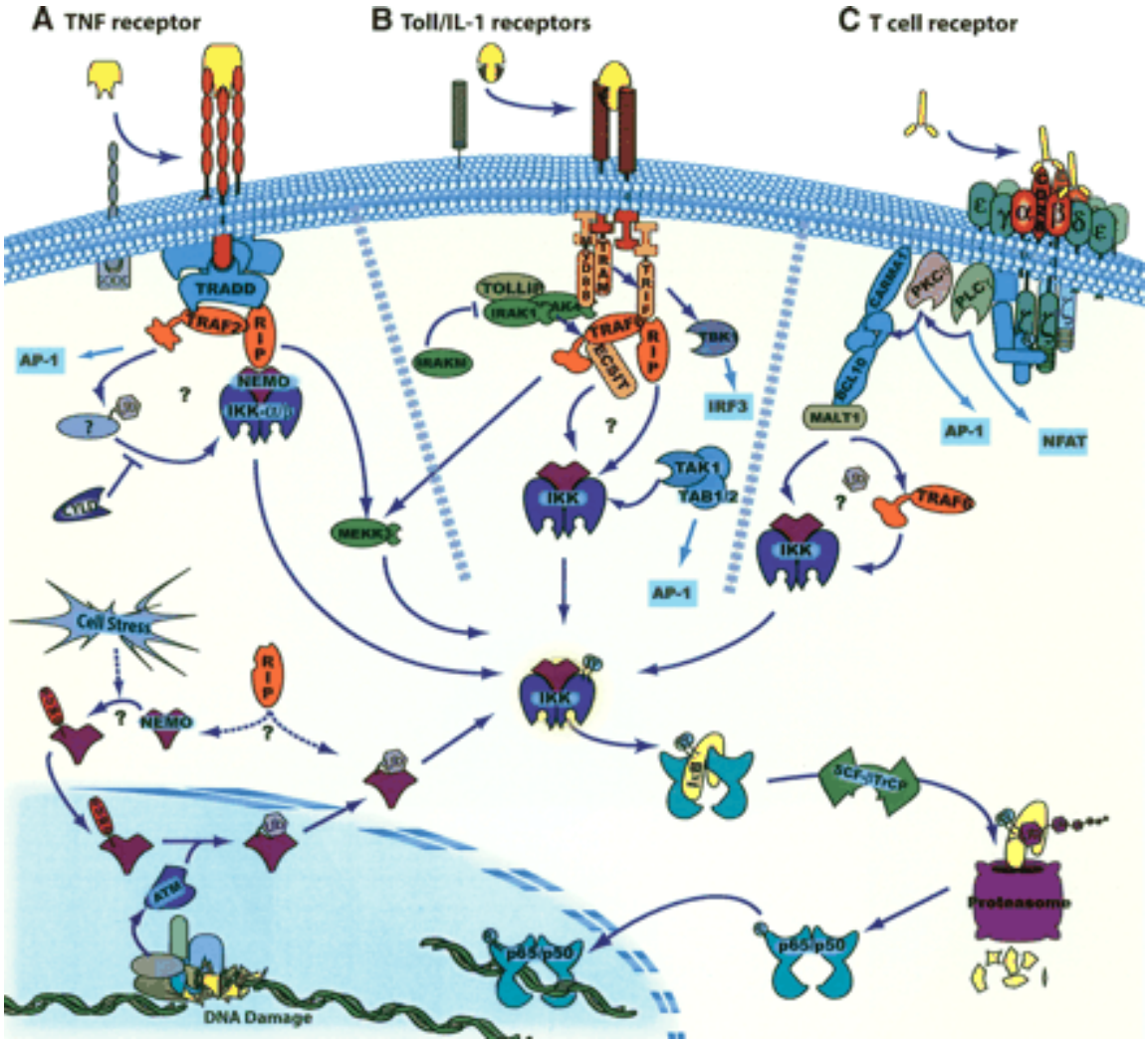

Figure 7. Major signaling pathways leading to the activation of NF-κB, from Hayden and Ghosh, 2004. Signaling transduction pathways emanating from TNF receptor (A), Toll/IL-1 receptor (B), the T-cell receptor (C), intermediary proteins, cell stress or DNA damage may cause phosphorylation of IκB by IKK. This triggers the degradation of IκB through the ubiquitin system, where the target molecule is masked by a chain of ubiquitins for degradation by the proteasome. The freed NF-κB dimer can then translocate to the nucleus where it activates genes transcription.

For example, upon stimuli by the pro-inflammatory cytokines IL-1 or TNF, IκB proteins are phosphorylated on two amino-terminal serine residues by the serine-threonine kinase IκB kinase (IKK), which is composed of two kinase subunits IKKα and IKKβ. IκB phosphorylation signals the ubiquitination and degradation of IκB proteins. This degradation results in the exposure of the nuclear localization signal on the NF-κB dimers and the subsequent translocation of the molecule to the nucleus. Once in the nucleus, NF-κB proteins enhance the transcription of certain genes, such as the pro-inflammatory genes (Baldwin, 1996).

These genes may encode for proteins having key functions in immune responses. Examples are shown Table 3.

Family	Examples of proteins	Function
Rel, IκB proteins	p105, c-Rel, IκBα	Regulation of NF-κB responses
Acute phase response proteins	angiotensinogen, C3 complement, complement factor B	Inflammation
Adhesion molecules	VECAM-1, ELAM-1, ICAM-1	Cell migration and repair
Cytokines	IL-1, IL-2, IL-2R, IL-6, IL-8, TNF-α, TNF-β, IFN-β	Immune cells activation

Table 3. Examples of proteins encoded by genes activable by NF-kB.

3 The NLR proteins

3.1 Similarity with the plant pathogen-recognition receptors

Plants possess both intracellular and extracellular pattern-recognition receptors called pathogen-resistance proteins or R-proteins (Dangl and Jones, 2001; Gomez-Gomez, 2004; Holt et al., 2003; Jones and Takemoto, 2004). These R-proteins are highly polymorphic. However, they are structurally conserved: they possess an amino-terminal effector domain composed of coiled-coil (CC) structures or TLR-IL-1 domains (TIR) (Hulbert et al., 2001; Nimchuk et al., 2003), a central nucleotide binding domain homologous to the NACHT domain, and carboxy-terminal LRRs (Dangl and Jones, 2001). These LRRs are involved in ligand sensing (Staskawicz et al., 2001).

We have previously seen that the body has developed a very elaborate extracellular system to detect the pathogens – the TLRs system, which also uses LRRs for ligand sensing. However, a question has remained unanswered: does our immune system possess intracellular PRRs?

Such PRRs may be very useful for different reasons: (1) intracellular pathogens exist that reside and proliferate inside the cells and the body needs protection against them and (2) some cells, for example those of the intestinal epithelium, lack the expression of some TLRs. This lack of TLR sensing gives them a certain degree of "blindness" to the PAMPs, in order to avoid a dramatic inflammation (Sansonetti, 2004). However, it has been shown that these cells are able to detect *Shigella* and mount an immune response (Philpott et al., 2000), which suggests the existence of an intracellular recognition system.

The NACHT-LRRs (NLRs) proteins have recently been identified as the intracellular pathogen sensing molecules (Carneiro et al., 2004; Philpott and Girardin, 2004; Viala et al., 2004b). The NACHT domain is present in neuronal apoptosis inhibitory protein (NAIP), class II transactivator (CIITA), plant het gene product involved in vegetative incompatibility (HET-E) and human peptide transporter 1 (TP1), hence the name.

NLRs possess numerous alternative names: NOD-LRR, NBS-LRRs, NOD and CATERPILLER (Harton et al., 2002; Inohara et al., 2005; Ting and Davis, 2005). CATERPILLER is an acronym for CARD, transcription enhancer, R-(purine-) binding, pyrin, lots of LRRs (Harton et al., 2002).

NLRs are cytosolic proteins structurally related to the most abundant class of plant R-proteins (Inohara et al., 2005; Ting and Davis, 2005), and also to the apoptotic protease-activating factor-1 (Apaf-1) (cf. Figure 8).

They present a similar tripartite architecture composed of an amino-terminal effector domain (EBD) involved in signaling, a centrally located regulatory nucleotide-binding domain (NBD), and a carboxy-terminal ligand-recognition domain (LRD).

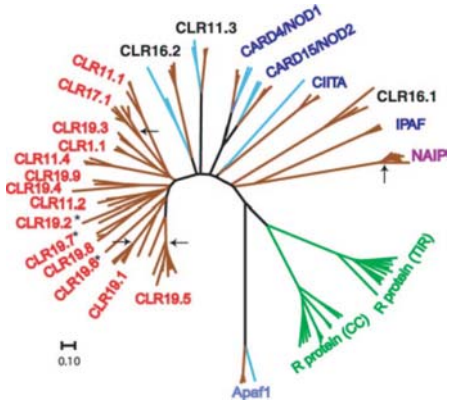

Figure 8. NLR proteins are related to the plant R proteins, from Ting and Davis, 2005. NLR proteins are denoted as CLR (caterpillar). Full length proteins were aligned using clustal X, and used to construct dendrograms. Apaf-1 is shown as a related sequence. Known proteins are putative orthologs from human, mouse, rat and teleost fish are shown. Each branch represents a sequence. Text color code: *green*, plant R proteins; *purple*, NAIPs; *blue*, CARD-containing NLRs; *black*, unrelated N-termini NLRs; *red*, Pyrin-containig NLRs. Line color code: *green*, plant; *brown*, mammalian; *light blue*, teleost fish lineages. Arrows represent murine-specific duplications; asterisks represent human-specific NLRs).

3.2 Tripartite architecture

Domain structures of relevant members of the NLR family of proteins are presented Figure 9.

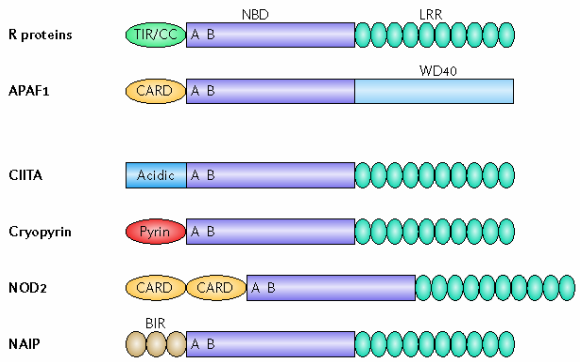

Figure 9. Domain structure of the NLR family and related proteins, from Ting et al., 2006. All NLRs display a modular organization in three functional domains: an amino-terminal effector domain, a centrally located nucleotide binding domain, with A and B depicting respectively the Walker A and Walker B motifs (cf. 3.2.2), and a carboxy-terminal domain composed of LRRs. CIITA, Cryopyrin, NOD2 and NAIP are representative members of the four main subfamilies of NLRs. These subfamilies are characterized by the four possible EBD that are the acidic, the pyrin, the caspase-activating and recruitment domain (CARD) and the baculovirus inhibitor of apoptosis repeat (BIR) domains. The structural organisation of the related proteins R proteins and Apaf-1 are shown for comparison. As R proteins, NLRs possess the NBD-LRRs configuration. However, they display more variable EBD. They also differ from Apaf-1 concerning their carboxy-terminal protein-protein interaction domain, which is composed of LRRs instead of WD40 repeats.

3.2.1 Effector-binding domain

The EBD nature is diverse, in contrast to the highly conserved TIR domains of plants. Mammalian EBDs are mainly composed by domains belonging to the death domain-fold superfamily, the pyrin domain (PYD) or the caspase-activating and recruitment domain (CARD) (Girardin et al., 2002; Inohara and Nunez, 2003; Inohara et al., 2002). Characteristics of the CARD domain are presented in 4. However, other domains can be found, such the baculovirus inhibitor of apoptosis protein repeat (BIR) (Tschopp et al., 2003) or the acidic domain. The combination of different effector domains or truncated activation domains may also be found. A minority of proteins contain an undefined amino-terminal domain unrelated to any known protein, which is predicted to present a fold similar to the one of the death domain superfamily (Liepinsh et al., 2003).

The nature of the EBD determines both the signaling pathway generated and biological functions. Some EBDs may be involved in homotypic interactions with downstream effector molecules carrying the same domain, such as both the PYD and the CARD (Girardin et al., 2002; Inohara and Nunez, 2003; Inohara et al., 2002). However, heterotypic protein-protein interactions may also exist. The BIR is involved in such interactions. The NLRs can be divided in subfamilies depending on their EBD (cf. Table 4). For example, NALPs possess PYD, NODs display CARD and NAIP contain BIR domains.

NLRs	Name	Other names	Domain organization
NALP	NALP1	DEFCAP; NAC; CARD7; CLR17.1	PYD-NACHT-NAD-LRR-FIIND-CARD
	NALP2	PYPAF2; NBS1; PAN1; CLR19.9	PYD-NACHT-NAD-LRR
	NALP3	PYPAF1; CIAS1; Cryopyrin; CLR1.1	PYD-NACHT-NAD-LRR
	NALP4	PYPAF4; PAN2; RNH2; CLR19.5	PYD-NACHT-NAD-LRR
	NALP5	PYPAF8; MATER; PAN11;CLR19.8	PYD-NACHT-NAD-LRR
	NALP6	PYPAF5; PAN3; CLR11.4	PYD-NACHT-NAD-LRR
	NALP7	PYPAF3; NOD12; PAN7; CLR19.4	PYD-NACHT-NAD-LRR
	NALP8	PAN4; NOD16; CLR19.2	PYD-NACHT-NAD-LRR
	NALP9	NOD6; PAN12; CLR19.1	PYD-NACHT-NAD-LRR
	NALP10	PAN5; NOD8; Pynod; CLR11.1	PYD-NACHT-NAD-LRR
	NALP11	PYPAF6; NOD17; PAN10; CLR19.6	PYD-NACHT-NAD-LRR
	NALP12	PYPAF7; Monarch1; RNO2; PAN6; CLR19.3	PYD-NACHT-NAD-LRR
	NALP13	NOD14; PAN13; CLR19.7	PYD-NACHT-NAD-LRR
	NALP14	NOD5; PAN8; CLR11.2	PYD-NACHT-NAD-LRR
NOD	NOD1	CARD4; CLR7.1	CARD-NACHT-NAD-LRR
	NOD2	CARD15; IBD1; PSORAS1; CLR16.3	CARD-CARD-NACHT-NAD-LRR
	NOD3	CLR16.2	(CARD)-NACHT-NAD-LRR*
	NOD4	NOD27; CLR16.1	(CARD)-NACHT-NAD-LRR*
	NOD5	NOD9; CLR11.3	X-NACHT-NAD-LRR
CIITA	CIITA	NHC2TA; C2TA	(CARD)-NACHT-NAD-LRR
IPAF	IPAF	CARD12; CLAN; CLR2.1	CARD-NACHT-LRR
NAIP	NAIP	BIRC1; CLR5.1	BIR-BIR-BIR-NACHT-LRR

Table 4. The human NLR family, from Martinon and Tschopp, 2005. * NOD3 and NOD4 contain an atypical CARD domain.

3.2.2 Nucleotide-binding and oligomerization domain

The NBD of the NLR proteins consists of a NACHT domain, followed in most cases by uncharacterized carboxy-terminal extension termed NAD for NACHT-associated domain. The NACHT domain is a predicted nucleoside triphosphatase (NTP-ase) domain composed of 300 to 400 residues (Koonin and Aravind, 2000) (cf. Figure 10).

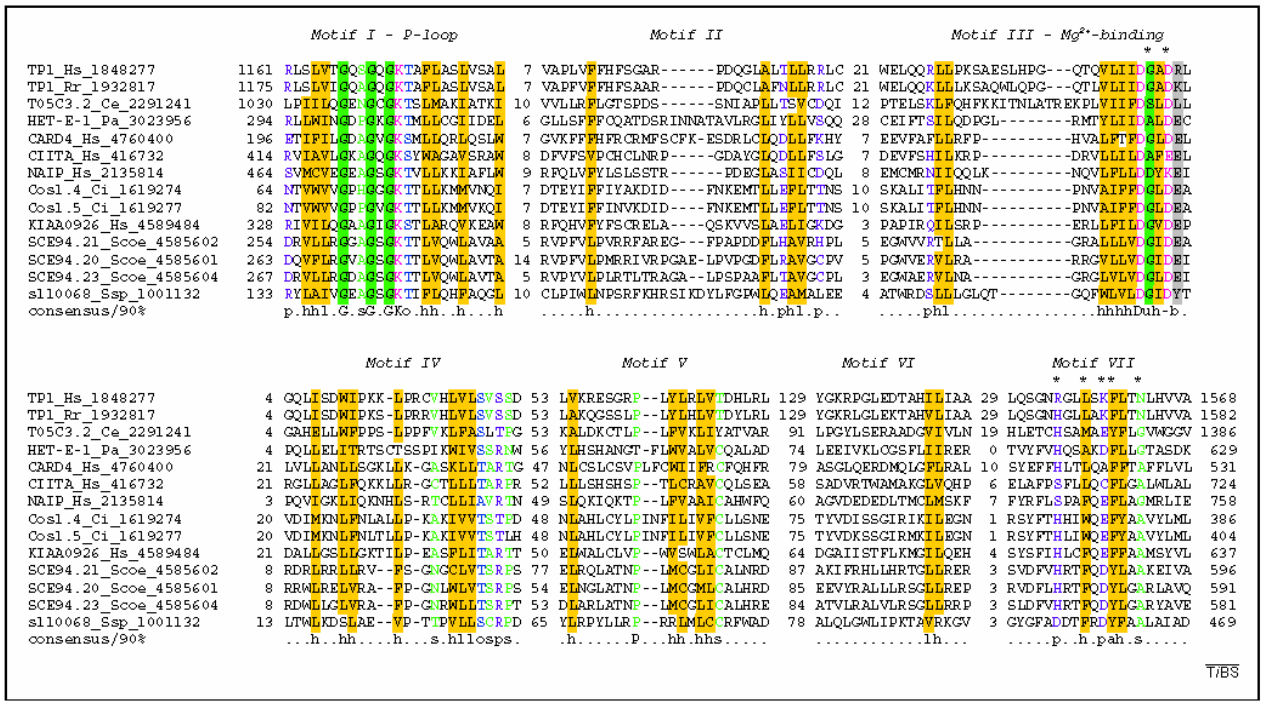

Figure 10. Alignment of NACHT domains, from Koonin and Aravind, 2000. The numbers indicate the distances from the protein termini to the proximal and distal aligned blocks, and the distances between the blocks. The conserved motifs are designated as in the AP-ATPases. The shading and coloring of conserved residues is according to the consensus that is shown below the alignment and includes residues conserved in at least 90 % of the aligned sequences; h indicates hydrophobic residues (A,C,F,I,L,M,V,W,Y; yellow background), s indicates small residues (A,C,S,T,D,N,V,G,P; green), o indicates hydroxy residues (S,T; blue); u indicates 'tiny' residues (G,A,S; green background), b indicates big residues (F,I,L,M,V,W,Y,K,R,E,Q; gray background), p indicates polar residues (D,E,H,K,N,Q,R,S,T; purple), and '-' indicates negatively charged residues (D,E; magenta). The designation of each protein includes the gene name (for uncharacterized proteins, the name in the NR database is indicated) followed by the species abbreviation and the Gene Identification number. Species abbreviations: Ce, Caenorhabditis elegans, Ci, Ciona intestinalis (a urochordate), Hs, Homo sapiens, Pa, Podospora anserina, Rr, Rattus rattus, Scoe, Streptomyces coelicolor, Ssp, Synechocystis sp.

This domain is closely related to the NBD domain found in the human apoptotic protease-activating factor APAF-1 and its nematode homolog CED-4, which is denoted as NB-ARC or AP-ATPase domain (Aravind et al., 1999; Van der Biezen and Jones, 1998).

NB-ARC and NACHT domains are known to mediate self-oligomerization and most of them possess ATP-ase activity (Harton et al., 2002) and belong to the AAA^+ family. However, some may show a GTP-ase activity (Espagne et al., 1997; Harton et al., 1999, Linhoff et al., 2001).

The NACHT domain is found associated with other domains such the CARD domain, WD repeats LRR, BIR, DAPIN or HEAT domains (Koonin and Aravind, 2000). It is composed of seven distinct conserved motifs: an amino-terminal ATP/GTPase specific P-loop (also known as Walker A motif), a Mg^{2+} binding site (Walker B motif), and five more specific motifs.
The NACHT differs from other NTP-ases by unique features including the presence of small residues (e.g. glycine, alanine or serine) at the C-terminus of the Mg^{2}-coordinating aspartate of the Walker B motif instead of a second acidic residue. A second acidic residue is found just two positions downstream. The last motif (motif VII) possesses a unique conserved pattern of polar, aromatic and hydrophobic residues.

3.2.3 Ligand-recognition domain

Similarly to the Toll like receptors and the plant R-proteins, the LRD domain of NLR proteins is composed of a variable number of LRRs thought to sense conserved microbial patterns or other ligands (Martinon and Tschopp, 2005). However the two capping domains LRRNT and LRRCT found in the TLRs (cf. 2.3.2.2) are absent.
LRR recognition is known to rely either on a direct or on an indirect interaction (Akira and Takeda, 2004). It is actually unknown if the LRRs of NLRs bind directly to their ligands, or if they need additional molecules or co-receptors.

3.3 NOD1 and NOD2

NOD1 and NOD2 are two representative members of the NLR-NOD subfamily playing an important role in inflammation and apoptosis. NOD1 (also called CARD4) was firstly reported by two groups in 1999 on the basis of a genomic database mining approach for searching homologs of Apaf-1 and its nematode homolog CED-4 (Bertin et al., 1999; Inohara et al., 1999). The discovery of NOD2 followed (Ogura et al., 2001b).
Apaf-1 is a key component of the apoptosis pathway and is composed of an amino-terminal CARD domain, followed by a NB-ARC motif and carboxy-terminal WD repeats. NOD1 and NOD2 present a similar structural organization (cf. Figure 9), but instead of having WD repeats, they possess carboxy-terminal LRRs. Their NACHT domain is followed by three NADs. At its amino-terminus, NOD1 presents one amino-terminal CARD domain, while NOD2 has two of these domains. Human NOD1 and NOD2 proteins are composed respectively of 953 and 1044 residues.

NOD1 and NOD2 were both shown to respond to PAMPs presented to the cytosol (Philpott and Girardin, 2004). They sense small motifs found in bacterial peptidoglycans (Chamaillard et al., 2003b; Girardin et al., 2003b; Girardin et al., 2003c; Inohara et al., 2003) and activate pro-inflammatory and apoptotic pathways (cf. 3.3.3).

3.3.1 Expression pattern

NOD1 and NOD2 are primarily expressed in the cytoplasmic compartment of both epithelial cells (Inohara and Nunez, 2003) and antigen-presenting cells such macrophages and dentritic cells (Gutierrez et al., 2002; Inohara and Nunez, 2003; Ogura et al., 2001b; Rosenstiel et al., 2003). No expression of these proteins is detected in other hematopoietic cells.
NOD1 and NOD2 expression patterns differ substantially. NOD1 is broadly expressed in the primary epithelial cells and in the intestinal epithelial cell lines, whereas NOD2 expression in these cells is generally very low or undetectable (Hisamatsu et al., 2003a; Hisamatsu et al., 2003b; Kim et al., 2004; Lala et al., 2003). Among the primary epithelial cells, NOD2 expression is restricted to specialised cells located at the base of the crypts of Liberkühn in the small intestine referred as Paneth cells (Ogura et al., 2003). Futhermore, in epithelial cells, NOD2 contains a molecular sequence containing two leucine residues and a tryptophan at its carboxy-terminus that allow its association with the plasma membrane (Barnich et al., 2005).

3.3.2 Bacterial elicitor

Both NOD1 and NOD2 sense products of degradation of the peptidoglycan. The motifs recognized are presented Figure 11.
NOD1 responds to a unique diaminopimelate (DAP)-containing GlcNAc-tripeptide muropeptide (GM-triDAP or iE-DAP) (Chamaillard et al., 2003b; Girardin et al., 2003a). The minimal motif detected is a dipeptide D-Glu-*meso*-DAP containing an exposed *meso*-DAP residue in the terminal position (Chamaillard et al., 2003b; Girardin et al., 2003c). NOD1 is able to recognize these motifs only when the *meso*-DAP is exposed (Girardin et al., 2003a). The *meso*-DAP is absent from Eukaryotes. It is characteristic of Gram-negative proteoglycans as well as few Gram-positive bacteria. Indeed, most Gram-positive bacteria possess a lysine residue at this position. As NOD1 is broadly expressed in tissues, it is thought to act as a general sensor for all Gram-negative bacteria (plus some Gram-positive bacteria) in multiple tissues.

NOD2 senses the muramyl dipeptide (MDP) (Girardin et al., 2003b; Girardin et al., 2003c; Inohara et al., 2003). MDP is the common moiety of all peptidoglycans from both Gram-negative and Gram-positive bacteria. We can thus hypothesize that the role of NOD2 is to sense broadly the two types of bacteria in restricted cell lines such the Paneth cells. NOD2 has also been shown to detect M-triLYS (Girardin et al., 2003b), a motif found solely in Gram-positive bacteria.

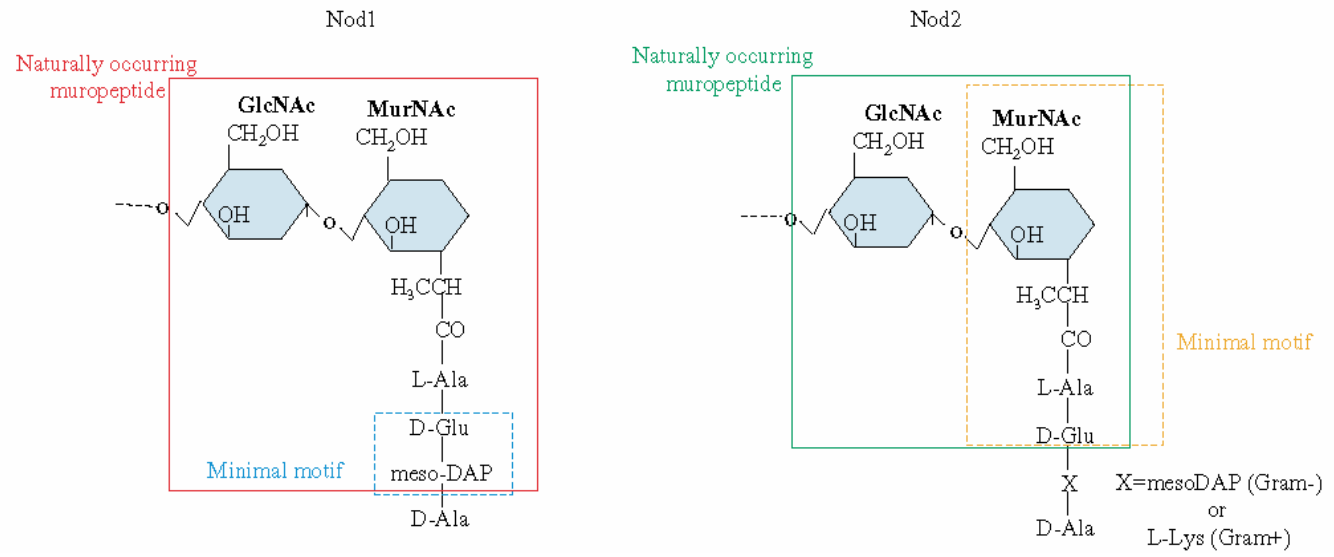

Figure 11. Naturally occurring and minimal peptidoglycan motifs sensed by NOD1 and NOD2, from Philpott and Girardin, 2004. The minimal naturally occurring peptidoglycan motifs detected by NOD1 and NOD2 are outlined respectively in red and green squares. The minimal peptidoglycan structure recognized by NOD1 is the dipeptide D-meso-DAP with the meso-DAP exposed (outlined in blue). The minimal peptidoglycan structure sensed by NOD2 is the MDP (outlined in orange).

3.3.3 Activation

Activation of NOD proteins occurs in three main steps, each of them dependant of a particular domain: (1) ligand sensing through the LRRs, (2) homooligomerization mediated by the NACHT domain (Inohara et al., 2000) and finally (3) recruiting of effector proteins and activation of diverse signaling pathways *via* the CARD(s).

3.3.3.1 Ligand sensing

The molecular mechanisms of the sensing of their respective bacterial ligand by NOD1 and NOD2 are still poorly understood. Specific residues within the LRRs of NOD1 and NOD2 are required for ligand sensing (Tanabe et al., 2004). However, it is still unknown if the GM-triDAP and MDP bind directly or require cofactors for binding and activation.

The LRRs of NOD1 and NOD2 are thought to play a negative role in the regulation of the NOD proteins: indeed, their deletion leads to proteins that are more active in initiating pro-inflammatory pathways (Inohara et al., 1999; Ogura et al., 2001b). In parallel with what is known on CIITA and on the plant R-protein Rx (Linhoff et al., 2001; Moffett et al., 2002; Sisk et al., 2001), intramolecular interactions may play a key role in this negative regulation. The backfolding of the LRRs to the NACHT domain of the NOD proteins is one possibility.
Another interesting question is how exactly the PAMPs reach the LRRs? It is probable that phagocytic cells release the bacterial ligands when they digest the invading bacteria. In epithelial cells, peptide transporters (such PEPT1) may be used (Vavricka et al., 2004). Type III (Cornelis and Van Gijsegem, 2000) and type IV secretion systems (Viala et al., 2004a) may also be involved in the detection of extracellular bacteria. Indeed, NOD1 has recently been shown to detect and respond to the extracellular Gram-negative bacteria *Helicobacter pylori*, through the injection of peptidoglycan through a type IV secretion system.

3.3.3.2 Oligomerization and signal transduction

Due to the structural similarities between Apaf-1 and the NLR-NOD proteins, it is believed that activation of NLR-NOD proteins may rely on similar mechanisms.

- Apaf-1 activation

In the cytosol, inactive Apaf-1 exists as a monomer and cannot interact with its downstream effector procaspase-9 (Li et al., 1997; Zhou et al., 1999). When proapoptotic signaling occurs, cytochrome *c* is released in the cytosol by the mitochondria. Recognition of cytochrome *c* by the WD40 repeats and binding of dADP or ATP to the NB-ARC domain induce the self-oligomerization of Apaf-1. This allows the binding of Apaf-1 to procaspase-9 through homotypic CARD-CARD interactions (Li et al., 1997; Zou et al., 1997) and formation of the apoptosome. The apoptosome is a large protein complex containing heptameric Apaf-1 associated with procaspase-9 (Benedict et al., 2000; Hu et al., 1998; Saleh et al., 1999; Srinivasula et al., 1998). To date, three electron cryo-microscopy structures of the apoptosome have been determined (Acehan et al., 2002; Yu et al., 2006) (cf. Figure 12).

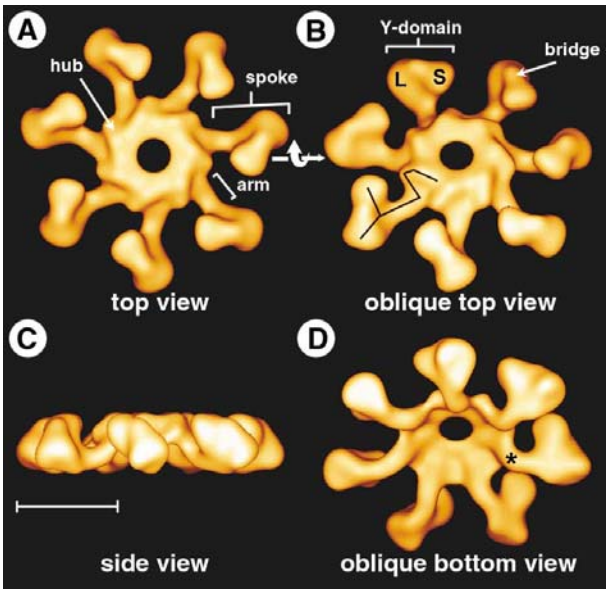

Figure 12. Cryo-electron microscopy three dimensional structure of the apoptosome, from Abbott et al., 2004. (A) Top view. Seven Apaf-1 molecules form the apoptosome, which is a wheel-like particules with seven spokes and a central hub connected by an arm. (B) Oblique top view. The WD40 repeats form the distal Y-shaped region of the spokes, with two lobes that are marked large (L) and small (S) separated by a bridge. The NOD-like region corresponds to most of the hub and the arms. (C) Side view. The scale bar is 100 Å. (D) Oblique bottom view. The arms are bent at an elbow (*) close to the hub.

Formation of the apoptosome causes the proximity and dimerization of procaspase-9, promoting its cross activation in its active form caspase-9 (Li et al., 1997; Zou et al., 1997), leading to apoptosis in the cell.

The crystal structure at 2.2 Å of the ADP-bound, WD40-deleted Apaf-1 was recently solved (Riedl et al., 2005). It gives insight on the molecular mechanism by which Apaf-1 exists in its inactive state. The amino-terminal CARD packs again the NBD domain, resulting in the burial of the procaspase-9 binding surface. This inactive conformation is locked by a buried ADP molecule. Binding and hydrolysis of dATP or ATP allows conformational changes leading to the formation of the apoptosome.

- NOD1 and NOD2 activation model

A model for NOD1 and NOD2 activation was proposed on the basis of the Apaf-1 activation. NOD proteins exist in the cytosol in an inactive autorepressed form, most likely with the LRRs backfolded onto the NACHT domain. Indeed, complementation analysis of NOD2 mutants has shown the existence of functional interactions between residues belonging to the proximal LRRs and residues of the carboxy-terminal part of the NACHT domain (Tanabe et al., 2004). These last residues were proposed to be important in turning an inactive NOD protein to its active state (Tanabe et al., 2004). Ligand recognition is thought to initiate a conformational change, exposing the NACHT domain and thereby allowing the self-

oligomerization of the NOD protein (Inohara et al., 2000; Tanabe et al., 2004). NOD proteins are thought to form hexamers or heptamers, as seen in the AAA^+ family. This oligomerization step is a prerequisite for the signal transduction and probably allows the exposure of the CARD domain(s), that can then interact through homophilic interaction with the CARD domain of the serine-threonine kinase RICK (alternatively named CARDIAK or RIP2) (Inohara et al., 1999; Ogura et al., 2001b) by a mechanism referred as "induced proximity" (cf. Figure 13) (Inohara et al., 2000). Mutational analyses demonstrated that the CARDs of NOD1 and NOD2 are necessary and sufficient for NF-kB activation (Inohara et al., 1999; Ogura et al., 2001b). The two CARDs of NOD2 are necessary for the binding to RICK CARD (Ogura et al., 2001b).

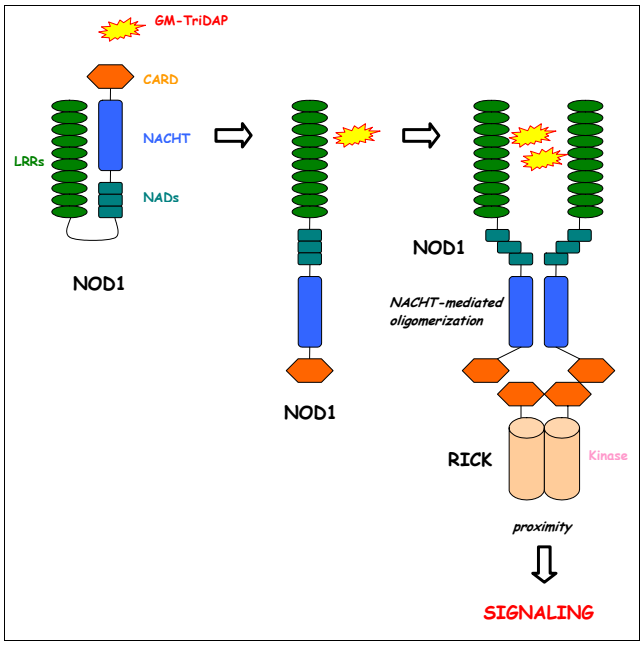

Figure 13. Proximity model for NOD1 activation. Ligand sensing induces self-oligomerization and recruitment of the downstream effector molecule RICK *via* homotypic CARD-CARD interaction. Oligomerization is mediated by the NACHT domain and induces the proximity and activation of RICK, resulting in initiation of signaling.

Furthermore, it is important to notice that NACHT domains may also mediate hetero-oligomerization with other NLR proteins (Damiano et al., 2004b). Details will be given in 3.3.5.

3.3.4 Signaling by NOD1 and NOD2

The main outcomes of NOD1 and NOD2 activation are the activation of the transcription factor NF-κB, the mitogen-activated protein kinases (MAPKs) (cf. Figure 14), and caspases.

- NF-κB activation

NOD1 and NOD2 play a crucial role in activating NF-κB in epithelial cells infected with gram-negative or gram-negative and gram-positive bacteria (Philpott and Girardin, 2004).

This activation of the NF-κB pathway is exclusively mediated through the downstream effector molecule RICK. Indeed, either NOD1 or NOD2 overexpression lead to the activation of NF-κB in wild type embryonic fibroblasts, but not in RICK-deficient embryonic fibroblasts (Kobayashi et al., 2002).

NOD1 has been shown to activate NF-κB and the subsequent pro-inflammatory responses in epithelial cells infected with the gram-negative invasive bacteria *S. flexneri* (Girardin et al., 2001). Girardin and colleagues demonstrated that *S. flexneri* infection enhances NOD1 oligomerization and thus induces NF-kB activation and signaling. Additionally, they showed that dominant-negative NOD1 intestinal epithelial cells are unable to activate NF-κB. Interestingly, they found that NOD1 and RICK form a transient complex with IKK (Girardin et al., 2001).

RICK mediates NF-κB activation by interacting with the inhibitor of the IKK complex, the kinase IKKγ (also referred as NEMO), with the intermediate region located between its CARD and its kinase domain (Inohara et al., 2000). Abbot and colleagues showed that, following its recruitment by NOD2, RICK mediates IKKγ polyubiquitylation to a novel unique ubiquitinylation site at lysine 285 (Abbott et al., 2004). This ubiquitinylation is known to be associated with activation of NF-κB (Zhou et al., 2004). It is followed by the phosphorylation of IKKγ and IKKβ, which signals the phosphorylation and degradation of IkB proteins by the proteasome. NF-κB is translocated into the nucleus (as seen previously in 2.3.3) and initiates the transcription of pro-inflammatory genes (cf. Table 3), and particularly the synthesis of prointerleukin-1 β (proIL-1 β).

- MAPKs activation

MAPKs are a highly conserved family of serine-threonine kinases. They are activated through a Thr-Xxx-Tyr phosphorylation motif by upstream kinases (Martin-Blanco, 2000). They play a role in a broad variety of cells functions such cytoskeletal reorganization, (Guay et al., 1997; Landry and Huot, 1995) proliferation (Mansour et al., 1994; Pages et al., 1993), apoptosis (Cross et al., 2000; Kong et al., 2000; Wang et al., 2000) and also cytokines secretion (Baldassare et al., 1999; Kovalovsky et al., 2000).

NOD1 and NOD2 have also been shown to activate MAPK-stimulated pathways JUN N-terminal kinase (JNK) (Girardin et al., 2001), extracellular-signal-regulated kinase (ERK) and

p38MAPK (Kobayashi et al., 2005; Pauleau and Murray, 2003) through a still unknown mechanism. This results in the activation of specific transcription factors such SRF, CREB, and AP-1 leading also to the expression of pro-inflammatory genes (Schorey and Cooper, 2003).

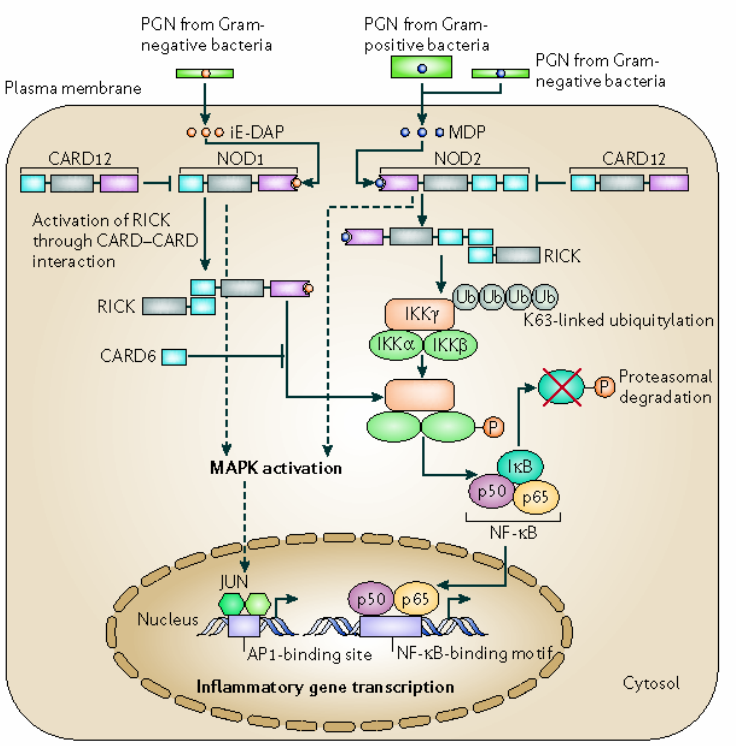

Figure 14. NOD1 and NOD2 signaling pathways, from Strober et al., 2006. Recognition of GM-TriDAP and MDP by the LRRs of respectively NOD1 and NOD2 leads to the activation of these proteins. They recruit the serine-threonine kinase RICK through CARD-CARD interactions, which ultimately activates the transcription factor NF-κB. In the case of NOD1, the mechanism of NF-κBactivation is still unclear. In the case of NOD2, RICK activation leads to the polyubiquitylation of IKKγ, the regulatory subunit of the IKK complex, at Lysine 63. This is followed by the phosphorylation of both IKKγ and IKKβ, signaling the phosphorylation and degradation of IκB the proteasome, and the subsequent translocation of NF-κB into the nucleus. RICK-mediated activation of NF-κB by both NOD1 and NOD2 is negatively regulated by CARD12. CARD6 negatively regulates RICK-mediated activation by only NOD1. NOD1 and NOD2 also activate the MAPKs by unknown mechanisms (dashed arrows). AP1, activator protein 1 (a transcriptional factor); PGN, peptidoglycan.

- Caspases activation

Caspases are a family of cysteine proteases involved in inflammation and apoptosis. They are found in the cell as inactive precursor and are activated by proteolytic processing. Some caspases possess long prodomains containing for example CARD domains or death-effector domains (DEDs) (cf. 4.1). They can thus interact with regulatory CARD-containing proteins, through homophylic interactions.

Caspases are highly specific to certain protein substrates. For example, caspase-1 processes proIL-1 β into active IL-1 β.

Co-immunoprecipitation assays demonstrated that NOD1 can physically interact with several procaspases containing long prodomains such procaspase-1, procaspase-2, procaspase-4, procaspase-8 and procaspase-9 (Inohara et al., 1999; Yoo et al., 2002). Recently, NOD1 was discovered to activate caspase-1 (Yoo et al., 2002). One of the known molecular mechanisms for procaspase-1 proteolytic processing involves the homotypic interaction between its CARD domain and the CARD domain of RICK. Yoo and colleagues demonstrated that NOD1 enhances RICK-mediated procaspase-1 processing into its mature form capase-1, and the subsequent IL-1 β secretion.

NOD2 has also been shown to bind to procaspase-1 using co-immunoprecipitation experiments and to stimulate IL-1 β secretion in cells co-transfected with plasmids encoding NOD2 and procaspase-1 (Liu et al., 2000). NOD2 may bind procaspase-1 through CARD-CARD interactions and initiate proIL-1 β processing and activation. Futhermore, when overexpressed, NOD1 interacts with the CARD domain of procaspase-9, leading to its processing toward its mature form and apoptosis (Inohara et al., 1999). However, it is important to keep in mind that these putative roles of NOD1 and NOD2 in caspases-mediated apoptosis have been shown under overexpression and remain to be verified under physiological conditions. Indeed, caspase-mediated apoptosis induced by NOD1 and NOD2 might be inhibited by the expression of anti-apoptotic factors such A1, c-IAPs and c-FLIP (Inohara et al., 2003; Micheau et al., 2001; Wang et al., 1998) resulting of the simultaneous activation of NF-kB.

3.3.5 NOD1 and NOD2 regulation

To date, two regulators of NOD2 and/or NOD1 have been discovered: CARD12 and CARD6 (also referred as CLAN, IPAF or CLR2.1). These regulators interact with the considered NOD proteins through heterologous CARD-CARD or NACHT-NACHT interactions and thus regulate their function (cf. Figure 14).

When overexpressed in cells, CARD12 binds both NOD1 and NOD2 through NACHT-NACHT interactions (Damiano et al., 2001). This association inhibits NOD1 and NOD2-mediated NF-κB activation (Damiano et al., 2004b). This regulatory effect of CARD12 was confirmed under physiological conditions. Indeed, CARD12 has been demonstrated to associate with ligand-activated NOD1 and NOD2 (Damiano et al., 2004a; Damiano et al., 2004b), negatively regulating their signaling.

CARD6 has been shown to bind both NOD1 and RICK but not NOD2 *via* CARD-CARD interactions. Overexpression of CARD6 has been linked to the inhibition of NOD1 and RICK-mediated NF-κB signaling (Stehlik et al., 2003).

In addition, NOD1 and NOD2 expression is regulated by pro-inflammatory cytokines. NOD1 expression in epithelial cells is upregulated by IFNγ but not by TNF (Hisamatsu et al., 2003a). However, TNF-α upregulates the expression of NOD2 expression synergically with IFNγ (Rosenstiel et al., 2003). In conditions of infection and activation of pro-inflammatory pathways, these cytokines upregulate NOD1 and NOD2 expression in order to amplify the immune response (Chamaillard et al., 2003a).

3.3.6 Cross-regulation of NODs and TLRs

NODs and TLRs are both crucial actors in immunity. An interrelationship between these two families is likely to exist. Numerous elements make this relationship highly possible such as the fact that some NODs and TLRs may recognize the same bacterial elicitors, activate the same signaling pathways, or even use the same molecule RICK as effector (Kobayashi et al., 2002). Furthermore, NODs and TLRs may enhance synergistically the production of cytokines in response to their activation by their respective PAMPs (Inohara et al., 1999; Kobayashi et al., 2002; Lala et al., 2003; Watanabe et al., 2004).

In antigen-presenting cells, the peptidoglycan-mediated activation of TLR2 seems to be negatively regulated by MDP-mediated activation of NOD2 signaling (Watanabe et al., 2004). However two other studies conclude with opposite results (Lala et al., 2003; Zhou et al., 2004).

Furthermore, upon activation by its ligand, NOD1 has been shown to enhance the TLR-induced production of diverse cytokines in human antigen-presenting cells (Chamaillard et al., 2003b; Fritz et al., 2005; Uehara et al., 2005; Van Heel et al., 2005).

3.3.7 Diseases

Genetic variations of several NLRs included NOD1, NOD2, NALP3, CIITA and NAIP have been associated with inflammatory diseases and/or immunodeficiency (Hugot et al., 2001; Inohara and Nunez, 2003; McGovern et al., 2005; McGovern et al., 2001; Miceli-Richard et al., 2001; Ogura et al., 2001a). This fact demonstrates the crucial role of these proteins as host-defense components (Opitz et al., 2005; Viala et al., 2004a). Substitutions or frameshift mutations in NOD1 and NOD2 are correlated with the occurrence of inflammatory and genetic disorders. They provoke a dysregulation of the induction and/or the regulation of the inflammation.

Autoimmune and infectious diseases associated with NOD1 and NOD2 are presented Table 5.

Disease	Mutations*	Comments
NOD1		
Helicobacter pylori infection	No mutation	Delivery of PGN to epithelial cells through type IV secretion system
Chlamydophila pneumoniae infection	No mutation	Activation of NF-κB in endothelial cells
Inflammatory bowel disease	Deletion polymorphism in LRR domain	Risk factor for inflammatory bowel disease
Asthma and high IgE levels	Insertion polymorphism in LRR domain	Risk factor for asthma
NOD2		
Crohn's disease	Arg702Trp, Gly908Arg, Leu1007fsinsCys	Defective NF-κB activation in response to MDP
Blau syndrome	Arg334Trp, Arg334Gln, Leu469Phe	Constitutive NF-κB activation
Early-onset sarcoidosis	Arg334Trp, His496Leu, Thr605Pro	Constitutive NF-κB activation
Graft-versus-host disease	Arg702Trp, Gly908Arg, Leu1007fsinsCys	Risk factor for graft-versus-host disease

*Mutations indicated are substitutions at the indicated amino-acid residue or are frameshifts as a result of insertion of a cysteine residue (fsinsCys).

Table 5. Autoimmune and infectious diseases associated with NOD1 and NOD2, from Strober et al., 2006.

- NOD1

NOD1 is located on chromosome 7p14-p15 on regions that have been genetically linked to atopy and asthma. Recent studies showed that two alleles with insertion-deletion polymorphisms located in an intron of *NOD1* are associated with eczema (Weidinger et al., 2005), atopic asthma (Hysi et al., 2005) and inflammatory bowel diseases (McGovern et al., 2005). Furthermore, splicing variants in the LRRs were also reported in healthy people (Hysi et al., 2005).

Asthma is a familial inflammatory disease of the human respiratory system. The airways of the lung narrow develop an increased responsiveness to various stimuli and the resulting symptoms are coughing, wheezing, increased mucus production, inflammation etc. The allele linked to asthma is correlated with high levels of IgE, even in the absence of allergen-specific IgE (Hysi et al., 2005).

Atopic eczema is a general systemic allergic reaction, which is common in people with related allergic conditions such asthma. Eczema is a form of dermatitism. Dermatitism is a skin irritation characterized by red, flaky skin, sometimes with cracks or tiny blisters (cf. Figure 15).

Figure 15. Atopic dermatitis, from http://www.dermnet.com.

Inflammatory bowel diseases are complex genetic disorders resulting from both genetic and environmental risk factors.

Girardin and colleagues studied the ability of these naturally occurring spliced variants to sense GM-triDAP (Girardin et al., 2005). They demonstrate that the NOD1 spliced variants correlated with susceptibility to the disorders listed above fail to detect the GM-triDAP and thus are unable to activate NF-κB.

Furthermore NOD1 may also be implicated in parasitic traits. For example, infection with *H. pylori* leads to the activation of NOD1 in epithelial cells (Viala et al., 2004a). Abnormalities in NOD1 function during infection with *H. pylori* may lead to chronic infection and development of gastric carcinoma (Uemura et al., 2001).

- NOD2

NOD2 is also associated with several pathologies but, contrary to NOD1, the mutations are located within the coding regions of *NOD2*. NOD2 polymorphism is correlated with Crohn's disease (Hugot et al., 2001; Ogura et al., 2001a), Blau syndrome (Miceli-Richard et al., 2001) and early-onset sarcoidosis (Kanazawa et al., 2005).

Crohn's disease is an autosomal recessive disease characterized by a chronic relapsing auto-inflammation of the digestive tract, primarily localized to the terminal ileum. Multiple rare variants in *NOD2* have been associated with susceptibility to Crohn's disease. However, we can distinguish three major mutations: R702W, G908R and L1007finsC. L1007finsC

describes a frameshift mutation resulting with the insertion of a cysteine residue. These mutations are all localized in the LRRs. Similarly to the ones described in *NOD1*, they are associated with loss of function. NOD2 mutants R702W, G908R and L1007finsC present a reduced ability to sense MDP and hence to activate NF-κB signaling. These three mutations have also be found to be correlated with immune diseases such, for example allergic disease (Kabesch et al., 2003), sepsis in infants (Ahrens et al., 2004), graft-versus host disease (Holler et al., 2004), psoriatic arthritis (Rahman et al., 2003) or radiological sacroilitis (Peeters et al., 2004).

Blau syndrome and early-onset sarcoidosis are both autosomal-dominant diseases, characterized by juvenile-onset systemic granulomatosis syndrome affecting mainly skin, joints and eyes. The main mutations correlated with the susceptibility to these disorders are R334Q, R334W and L469F (Miceli-Richard et al., 2001). They are located in the NACHT domain and cause a gain of function due to a constitutive NF-κB activation (Chamaillard et al., 2003c). Interestingly, NOD2 mutation R334Q corresponds to the mutation R260W in NALP3 which is also associated with two inflammatory diseases (familial cold autoinflammatory syndrome and Muckle Well Syndrome), suggesting a similar molecular mechanism for the development of these inflammatory disorders (Albrecht et al., 2003).

We have just seen the importance of NOD1 and NOD2 in the susceptibility to several disorders and the importance of mutations affecting their LRRs and NACHT domains. However, the CARD domain is also very important and may be correlated with cancers, apoptosis etc (cf. 4).

4 The Caspase-activating and recruitment domain

4.1 The CARD is a member of the Death-domain superfamily

The death-domain superfamily is composed of four families: CARD, death effector domain (DED), death domain (DD), and pyrin/AIM/ASC/DD-like domain (PAAD). Each of these domains possesses a characteristic six antiparallel α-helical bundle fold referred as the "death fold" and probably derives from the same common ancestor. Each module is known for interacting with members of the same subfamily class through homotypic interaction (e.g. CARD-CARD or DD-DD) forming either dimmers or trimers. However, few heterophilic interactions have been reported, such the interaction between the CARD domain of RICK and the DD of the p75 nerve growth factor-receptor (Khursigara et al., 2001).
Members of this superfamily are characterized by a high interfamily structural similarity despite a low sequence identity even within the same family. Interactions between death-domain superfamily members are involved in many processes such as signal transduction in apoptosis, the maintenance of homeostasis, the immune system, cancers etc.

4.2 Implications of CARD domains in crucial intracellular processes

CARD-containing proteins have been reported to be implicated in the following pathways: regulation of caspases both in apoptosis and in inflammation and regulation of NF-κB activation in the context of immune responses.
CARD domains are found in the amino-terminal domains of many caspases, programmed apoptosis and pro-inflammatory proteases. CARD domains may be found in isolation or combinated with other domains (Bouchier-Hayes and Martin, 2002).
CARD domains can be divided into four subfamilies upon their associated domain(s) and likely function (cf. Figure 16): the NBD-CARDs which correspond to our NLRs, the coild-coil CARDs, the bipartite CARDs and the CARD-only proteins.
NBD-CARDs and coiled-coil CARDs possess a similar structural organization. Both the central NBD and corresponding coiled-coil motif mediate oligomerization. Proteins belonging to these two subfamilies are thought to act as a molecular scaffold and initiate the assembly of protein activation complexes. NBD-CARDs have been reported to be involved both in apoptosis (e.g. Apaf-1) and in NF-κB activation. The coiled-coil CARDs appear to play only a role in NF-κB activation through their interaction with the bipartite CARD protein Bcl10 (Willis et al., 1999).

The bipartite-CARD proteins contain a CARD domain associated with an additional domain that can be a kinase domain, a protease domain or another domain belonging to the death-domain superfamily. The bipartite-CARD proteins are generally recruited by NBD-CARDs or coiled-coil CARDs and either become activated (e.g. caspases) or recruit effector molecules (e.g. IKKγ for RICK).

Members of the last subfamily are composed of a single CARD domain. They are thought to regulate the function of NBD-CARDs and coiled-coil CARDs by interacting with bipartite-CARDs. In this connection, Pseudo-Ice has been reported to interact with both the prodomains of caspase-1 and RICK (Druilhe et al., 2001; Lee et al., 2001). It has been shown to inhibit RICK-induced capase-1 oligomerization and processing (Druilhe et al., 2001; Lee et al., 2001).

We can conclude that CARD-containing proteins activate caspases and NF-κB by two main mechanisms: (1) the assembly of multi-protein complexes, which can facilitate dimerisation and (2) as scaffolds on which proteases and kinases are assembled and activated.

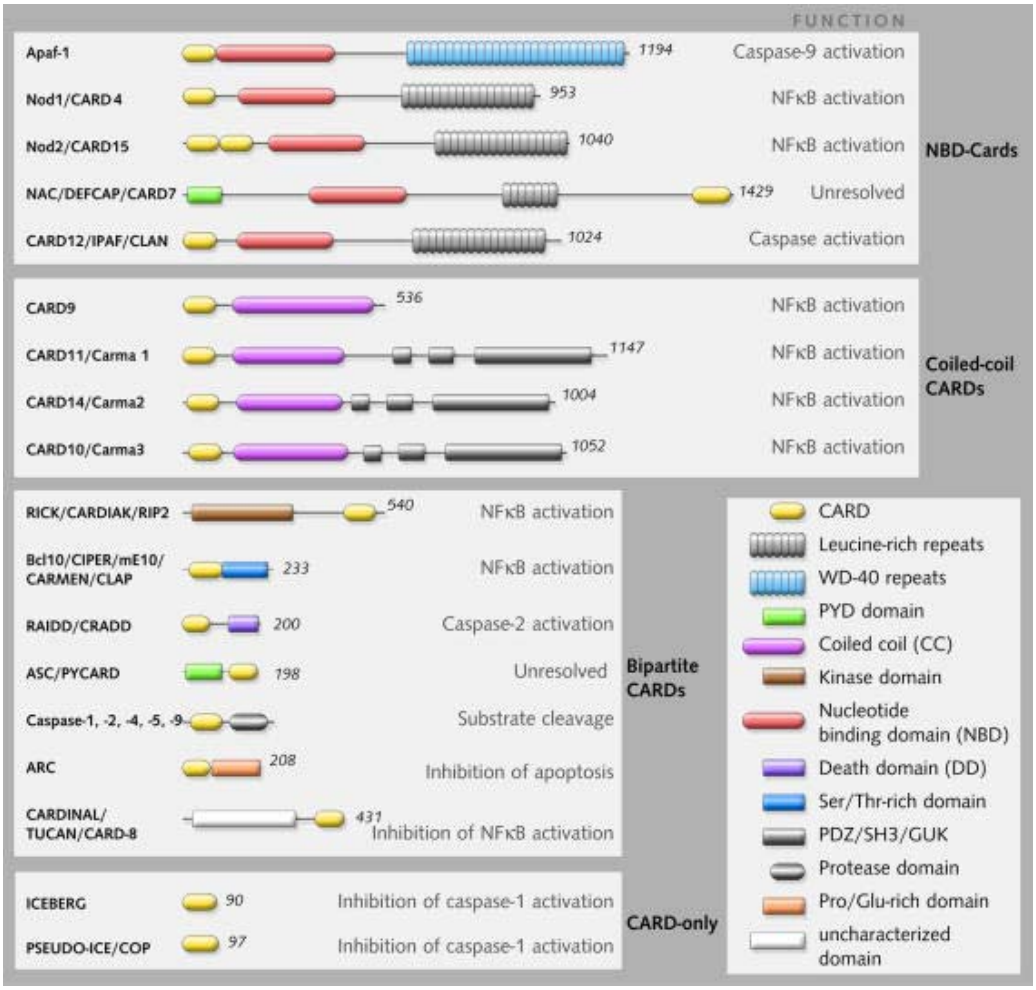

Figure 16. Domain structures of human CARD-containing proteins and associated functions, from Bouchier-Hayes and Martin, 2002.

4.3 Structural and biochemical characteristics of CARD domains

CARD domains comprise approximatively ninety residues. Several CARD structures have been solved. The first structure determined was that of the adaptor protein RAIDD (cf. Figure 16) (Chou et al., 1998), a protein known to interact with caspase-2. It consists of six antiparallel amphipathic helices (cf. Figure 17), closely packed around a hydrophobic core.

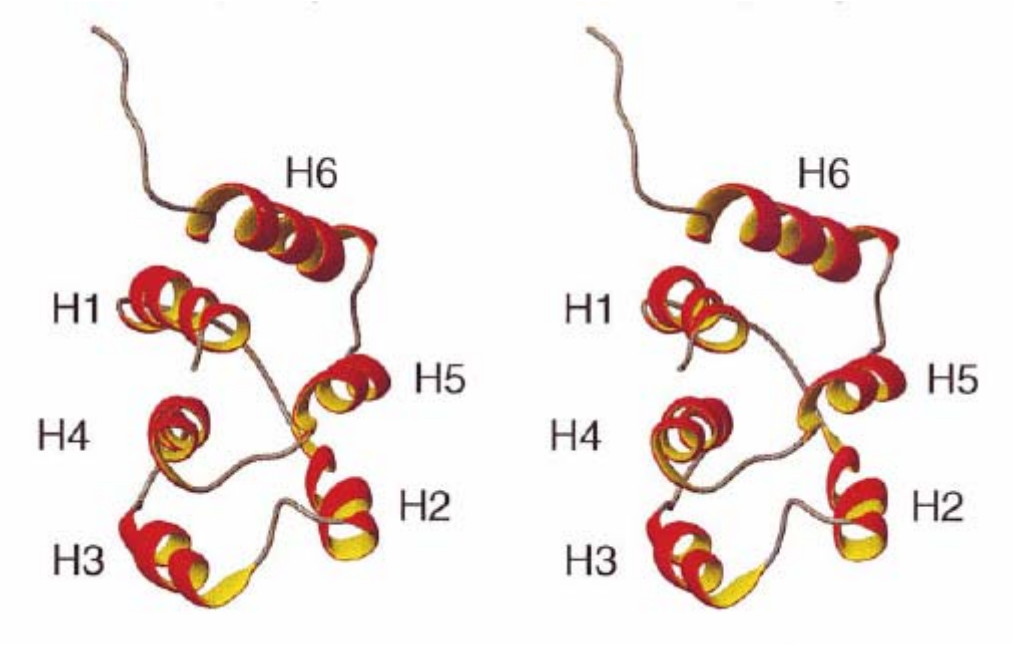

Figure 17. Stereodrawing of the solution structure of RAIDD CARD, from Chou et al., 1998. Helices 1 to 6 are labeled H1-H6.

Three of these helices (H1, H3 and H4) form an acidic patch on the surface of the domain, whereas the three others (H2, H5 and H6) form a basic patch. Modeling of CARD domain of caspases-2 shows a similar distribution of charges. Mutational data (Duan and Dixit, 1997) imply that the binding is mediated by electrostatic interaction between the basic patch of RAIDD and the acidic patch of caspases-2.

Apaf-1 CARD structure has been solved both by NMR and X-ray crystallography (Day et al., 1999; Vaughn et al., 1999; Zhou et al., 1999). It also possesses two charged surfaces: one acidic patch formed by helices 2 and 3 and a basic patch involving residues of helices 1 and 4. Mutagenesis experiments and analysis of the crystal structure of the complex with procaspase-9 (Qin et al., 1999) showed that the acidic patch of Apaf-1 interacts with the basic patch of procaspase-9. This confirms the importance of electrostatic interactions in CARD-CARD associations.

4.4 Targeting CARD-containing molecules

As we have seen earlier, CARD-containing proteins are key regulators in both apoptosis and inflammation. Notably some of them such Iceberg or Pseudo-Ice (cf. Figure 16) regulate caspase-1 activity and thus IL-1β levels. This cytokine plays a crucial role in inflammation, but also contributes to neoplastic progression by promoting angiogenesis, proliferation and the metastasis of tumors (Damiano and Reed, 2004). All these pathological characteristics are common to many cancers. The fact that CARD-containing proteins have the ability to control them provides a new basis for the development of new pharmacological strategies in the treatment of cancers and inflammatory diseases (cf. Figure 18).

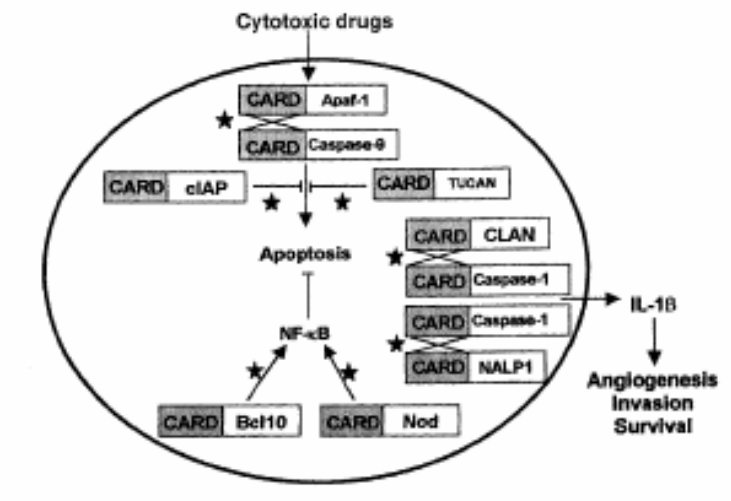

Figure 18. Functions of CARD-containing proteins in cancer-related processes and potential points of drugs intervention (*), from Damiano and Reed, 2004.

5 Perspectives

NOD proteins are crucial components of our innate immune system. Genetic variations of NOD proteins have been correlated with inflammatory and genetic diseases. In spite of the active study of the physiological role of the NOD proteins and of their signaling pathways, many questions remain unanswered concerning their activation mechanism, such (1) how are the bacterial ligands recognized by the LRRs?; (2) how does this binding induce a conformational change allowing the oligomerization through the NBD?; (3) what are the characteristics of the CARD-CARD interactions?; and (4) why are the two CARDs of NOD2 necessary for the activation of the target serine-threonine kinase RICK?

The determination of the 3D structure of different NOD proteins and/or some of their domains alone or in complex with the downstream target partners would help providing a structural basis for a better understanding of these mechanisms on a molecular level.

6 Scope of this thesis

The detailed atomic structure of the CARD domain of NOD1 would help us to better understand its mechanism in the recruitment of the effector molecule RICK. Even if some CARD structures have already been reported, very little is known concerning the basis of their homotypic interaction.

The material and methods used are described in the second part (II) of this thesis and the results are presented in the third part (III).

Chapter III.1 describes the expression, the purification and the preliminarily NMR studies of NOD1 CARD domain. The structure determination of this domain in solution together with dynamics studies and the structure itself are presented chapter III.2. Expression, purification of RICK CARD and characterization of the NOD1/RICK CARD-CARD interaction are described in chapter III.3. Dynamic studies of this domain are presented chapter III.3. A comparison of the NOD1/RICK CARD-CARD interaction with the Apaf-1/procaspase-9 CARD-CARD interaction constitutes chapter III.4.

Thesis deroulement and contributions

This thesis work was carried out thanks to an extensive collaboration between three different laboratories which were involved progressively in this project depending on the results obtained.

I initiated this project of structural studies of NOD proteins in my group at EMBL-Grenoble, under the supervision of Dr Stephen Cusack. I began working on this subject one year after the beginning of my thesis, after having being scooped on my first project on the Nogo receptor structure.

We obtained the cDNA of full length NOD1, NOD2 and RICK proteins from Dr Gabriel Núñez (Department of Pathology and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA).

I made many constructs of the different domains of NOD1, NOD2 and RICK for overexpression either in insect cells or bacteria. In each case, different expression conditions (strains, temperature, coexpression etc.) were tested. Unfortunately, there were problems of expression and/or aggregation. It was therefore extremely difficult to obtain enough soluble materiel for structural studies. A succinct table summarizing the diverse clones realized and problems encountered is presented Appendix 1.

Finally, I obtained some success with a construct of the CARD domain of NOD1 (residues 15-138). We have thus decided to focus our energy on the study of this domain by X-ray crystallography. I tested many cystallogenesis conditions but no crystals nor anything promising were obtained.

At this step, we established a fruitful collaboration with Dr Jean-Pierre Simorre (Laboratoire de Résonance Magnétique et Nucléaire, Institut de Biologie Structural Jean-Pierre Ebel, Grenoble, France). I spent fourteen months in his laboratory during which I learned how to solve a structure by NMR, thanks to the supervision of Dr Adrien Favier. Dr Adrien Favier also helped with the data acquisition, and performed the dynamic studies on NOD1 CARD. We also tried studying the interaction between the CARD domain of NOD1 and the CARD domain of RICK by NMR mapping, but without any success due to problems obtaining soluble RICK CARD in suitable quantities.

After having solved the structure of NOD1 CARD, I went back to my laboratory and began the redaction of an article. I tried to find out how could explore the interaction between NOD1-CARD and RICK-CARD. I hypothesized that the acidic patch on NOD1 CARD may interact through electrostatic interaction with the basic patch of RICK CARD. As it was impossible to study this interaction *in vitro*, we decided to study it in the mammalian cells

system using mutagenesis, immunoprecipitation and immunoblotting. I spent five months in Ann Arbor, MI, in the laboratory of Dr Gabriel Núñez to perform these final experiments.

Part II

Materials and Methods

Résumé (français)

- Chapitre II.1 : biologie moléculaire

Le système d'expression bactérien a été choisi pour la surexpression du domaine CARD de NOD1 et du domaine CARD de RICK. Le clonage des résidus 15-138 du gène *NOD1* humain a été réalisé dans le vecteur pET-M11 (EMBL). Ce vecteur d'expression confère une étiquette poly-histidine clivable (présence du site de reconnaissance de la protéase TEV) en amino-terminal de la protéine d'intérêt. Les résidus 432-540 du gène *RICK* humain ont été clonés dans le vecteur pLX02 (Protein'eXpert). Ce vecteur permet l'expression du domaine CARD de RICK fusionné en son extrémité amino-terminale à une extension poly-histidine non clivable.

Les vecteurs d'expression en cellules humaines pcDNA3-NOD1-HA et pcDNA3-RICK-FLAG, codant respectivement pour les protéines NOD1 et RICK humaines sauvages, ont servi de matrices lors d'expériences de mutagenèse. Dix-sept mutations ont été réalisées dans le domaine CARD de NOD1 et huit dans celui de RICK au moyen du kit de mutagenèse dirigée QuickChange XL (Stratagene).

- Chapitre II.2 : biochimie

Le domaine CARD de NOD1 a été surexprimé dans la souche bactérienne *Escherichia coli* BL21 Star (DE3) (Invitrogen). Les conditions d'expression de la protéine non marquée et de la protéine marquée ^{15}N ou $^{15}N/^{13}C$ sont identiques ; cinq heures à 30°C suite à une induction au moyen d'1 mM d'IPTG. Cependant les milieux de culture utilisés diffèrent. Le milieu LB est utilisé pour le domaine CARD de NOD1 non marqué et le milieu minimum M9 enrichi en isotope ^{15}N et/ou ^{13}C pour la protéine marquée ^{15}N ou $^{15}N/^{13}C$. Le protocole de purification du domaine CARD de NOD1 se décompose en quatre grandes étapes : (1) première colonne de nickel, (2) clivage de l'étiquette poly-histidine par la protéase TEV, (3) deuxième colonne de nickel, (4) concentration et filtration sur gel.

Le domaine CARD de RICK a été exprimé dans la souche bactérienne *Escherichia coli* BL21 (DE3) Codon plus RIL (Stratagene). L'expression a été réalisée de manière similaire à celle du domaine CARD de NOD1 non marqué (milieu LB, induction avec 1 mM d'IPTG) mais durant 20 heures à 15°C. La protéine se retrouvant principalement sous forme d'agrégats solubles, un protocole de purification particulier a été mis au point. Il s'appuie sur une purification d'affinité sur colonne de nickel combinant la technique d'élution au moyen d'imidazole avec celle recourant aux sauts de pH, suivie d'une filtration sur gel.

- Chapitre II.3 : biologie cellulaire

Les cellules humaines HEK-293T ont été utilisées pour toutes les expériences de biologie cellulaire, à savoir co-immunoprécipitations et essais d'activation de NF-κB.

Pour les expériences de co-immunoprécipitations, les cellules HEK-293T ont été co-transfectées par la méthode au phosphate de calcium avec les vecteurs pcDNA3-NOD1-HA et pcDNA3-RICK-FLAG sauvages et/ou mutants (cf. 2.1). Les cellules ont été lysées vingt heures après transfection. Les lysats cellulaires obtenus ont ensuite été incubés avec un anticorps monoclonal anti-FLAG de souris (Sigma). Les immunoprécipitations ont alors été réalisées en utilisant la protéine G-sepharose (Zymed). Les lysats totaux et les produits immunoprécipités ont tous été analysés sur gel d'électrophorèse 10 % par western-blot. Les protéines NOD1 et RICK ont été détectées respectivement au moyen d'un anticorps de lapin anti-HA (Santa Cruz Biotechnology) et d'un anticorps de souris anti-FLAG (Sigma). Des contrôles ont été réalisés avec des lysats de cellules transfectées uniquement avec le vecteur pcDNA3-NOD1-HA sauvage ou le vecteur pcDNA3-RICK-FLAG sauvage (plus pcDNA3 vide pour maintenir constante la quantité d'ADN utilisée), de manière à comparer l'expression des protéines NOD1 et RICK mutantes par rapport aux protéines natives.

Pour les expérience d'activation du facteur de transcription NF-κB, $4x10^4$ cellules HEK-293T ont été co-transfectées avec la lipofectamine (Invitrogen) en utilisant quatre plasmides différents : 5 ng d'un vecteur rapporteur luciférase (pBVIx-luc (Inohara et al., 2000)), 10 ng d'un vecteur rapporteur β–galactosidase (pCMV-βgal, (Inohara et al., 1999)), 5 ng de pcDNA3-NOD1-HA sauvage ou mutant et 395 ng de pcDNA3 vide. Un contrôle négatif a également été réalisé en co-transfectant avec 5 ng de pBVIx-luc, 10 ng de pCMV-βgal et 400 ng de pcDNA3 vide. Les cellules ont été lysées 20 heures après transfection. Les activités β-galactosidase et luciférase ont été déterminées comme décrit précédemment (Inohara et al., 1999). Les résultats ont été normalisés en fonction de l'efficacité de transcription grâce aux valeurs obtenues avec le rapporteur β–galactosidase. Les essais d'activation de NF-κB ont été répétés deux fois en triplicata.

- Chapitre II.4 : résonance magnétique nucléaire

La résonance magnétique nucléaire (RMN) a été la méthode utilisée pour résoudre la structure du domaine CARD de NOD1. Ce chapitre introduit brièvement les concepts de bases de cette technique et explique la méthodologie suivie pour déterminer une structure tridimensionnelle en utilisant la RMN biologique. Il s'adresse plus particulièrement à un public de biologistes ou biochimistes, non expert de cette méthode.

L'étude structurale des macromolécules biologiques par RMN en solution suit un protocole qui peut être décomposé en plusieurs étapes : (1) préparation de l'échantillon, (2) acquisition des données, (3) attribution des fréquences de résonance, (4) extraction des informations structurales, (5) calcul de structure et (6) validation. Les programmes CNS et ARIA ont été utilisés pour le calcul de structures. L'analyse de l'ensemble de structures a été réalisée au moyen des programmes AQUA et PROCHECK-NMR.

La résonance magnétique nucléaire permet également d'étudier de manière approfondie, d'une part, la dynamique interne des protéines par relaxation nucléaire, et d'autre part, les interactions entre molécules.

- Chapitre II.5 : modélisation

Le logiciel de modélisation comparative Modeller8v1 a été utilisé pour la modélisation de la structure du domaine CARD de RICK. La structure tridimensionnelle du domaine CARD de procaspase-9 a servi de matrice.

Le modèle d'interaction entre les domaines CARDs de NOD1 et de RICK a été réalisé sur la base de la structure cristallographique du complexe formé par les domaines CARD d'Apaf-1 et de procaspase-9 (Qin et al., 1999). La structure du domaine CARD de NOD1 et celle modélisée du domaine CARD de RICK ont été superposées sur les structures d'Apaf-1 et de procaspase-9 respectivement.

1 Molecular biology

1.1 Cloning

- NOD1 CARD

In this part, I will present the materials and methods involving only the soluble construct of NOD1 CARD used for structure determination.

Human full length *NOD1* cDNA was used as a template to amplify the region encoding residues 15-138 of the NOD1 protein. This region encompasses the CARD. The two following oligonucleotides were used for the amplification by PCR.

NOD1_15for: 5'-AGTAGGACCCATGGAGTCTCACCCCCACATTCAATTACT-3'

NOD1_138rev: 5'- AGTAGGAGGTACCTCAGCCCAGATGGTGTCGCAGCT-3'

They contain respectively the *Nco* I and *Kpn* I restriction sites (underlined). The PCR product was then cloned into the pETM-11 vector (EMBL) after digestion by the restriction enzymes *Nco* I and *Kpn* I (NEB). This vector results in an amino-terminal six residue histidine-tag with a TEV protease cleavage site fused to the expressed protein. The obtained plasmid was sequenced (MWG). After removal of the His-tag using the TEV protease, the resulting protein fragment possesses three additional amino-terminal residues not present in the wild-type sequence (Gly, Ala and Met).

- RICK CARD

Human full length *RICK* cDNA was used as template for amplifying the region encoding residues 432-540 of the RICK protein. The cloning was carried out at the IBS using the RoBioMol system. Briefly, PCR amplification was achieved with Phusion enzyme (Finnzymes) using the following specific primers:

RICK-432for: 5'-CATCACCATCAATTGCAGCCTGGTATAGCCCAGCAGTGG-3'

RICK-540rev: 5'-TCACCATCCAATTGATTACATGCTTTTATTTTGAAGTAAATTTAAAG-3'

These oligonucleotides contain the sequences required for ligation independent cloning (LIC). The PCR products were then inserted into the vector pLX02 (Protein' eXpert), by annealing. This vector results in the expression of the protein of interest fused to an amino-terminal Met-Ala-$(His)_6$-Gly-$(His)_3$Gln-Leu tag.

The ligation products were transformed into *E. coli*, and the presence of the gene of interest in the vector was verified by *Nco* I restriction analysis and sequencing (Genome Express).

1.2 DNA mutagenesis

Expression plasmid pcDNA3-amino-terminal HA-tagged NOD1 (pcDNA3-NOD1-HA) (Inohara et al., 1999) and pcDNA3-amino-terminal FLAG-tagged RICK (pcDNA3-RICK-FLAG) (Inohara et al., 1998) were used as templates for the mutagenesis experiments. They carry the gene encoding for the full length NOD1 protein and RICK proteins, respectively.

The QuickChange XL site-directed mutagenesis kit (Stratagene) was used to generate single, double and triple amino acid mutants in the CARD domains of NOD1 and RICK, using the protocol given with the kit.

Oligonucleotides used for the mutagenesis PCRs are presented in Appendix 2.

Seventeen mutants were realized in NOD1 CARD: L40A, V41A, V41Q, D42K, L44A, D48K, A52S, E53K, D54K, E56K, I57A, K67A, R69E, L40A-D42K, D42K-L44A, L40A-V41A-D42K, D54A-E56K-I57A ; and eight in RICK CARD: R444E, V448A, I469A, K480E, P481A, R483E, R488E, E500A.

The entire coding sequences of all NOD1 and RICK clones were verified by DNA sequencing (DNA sequencing core, University of Michigan).

2 Biochemistry

2.1 Protein expression in bacteria

2.1.1 NOD1 CARD expression

NOD1 CARD protein was expressed using bacterial strain *Escherichia coli* BL21 Star (DE3) (Invitrogen). Bacteria were transformed with the pETM-11-NOD1-CARD vector and expression was carried out in Erlenmeyer flasks using 1 L of LB medium per flask complemented with 100 µg/ml of kanamycin. Each 1L culture was inoculated using 15 ml of overnight preculture. The cells were grown at 37°C with agitation (180-200 rpm) until an OD_{600} of O.8-1. At this point, protein expression was induced with 1 mM IPTG and the cells were incubated for an additional five hours at 30°C.
Bacteria were then harvested by centrifugation at 7000g for fifteen minutes. Cell pellets were frozen using liquid nitrogen and stored at -80°C until purification.

2.1.2 RICK CARD expression

Escherichia coli BL21 (DE3) Codon plus RIL (Stratagene) bacteria were transformed using the pLX02-RICK-CARD vector. Expression was carried out essentially as for NOD1 CARD. However, there were differences concerning the antibiotics and expression conditions used. Ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml) were used instead of kanamycin, and expression was carried out during twenty hours at 15°C.

2.2 Production of labeled protein for NMR study

Production of ^{15}N-labeled NOD1 CARD and ^{15}N-^{13}C-labeled NOD1 CARD was necessary for the NMR experiments. The three-days procedure described below is for a 1 L culture. In the case of NOD1 CARD, total volume cultures of 5-6 L were necessary to obtain a sufficient quantity of protein.
On day one, *Escherichia coli* BL21 Star (DE3) were transformed with the pETM11-NOD1-CARD vector and plated out on LB-kanamycin-agar plate. On day two, a single colony picked from the fresh plate was used to inoculate 5 ml LB. After a day-long incubation at 37°C with a 200 rpm agitation, 400 µl of this preculture served to inoculate 40 ml of the

adequate M9 minimal medium. The culture was then continued overnight in the same conditions. ^{15}N-M9 and ^{15}N-13C-M9 (cf. Table 6) media were used for the production of ^{15}N-labeled NOD1 CARD and ^{15}N-^{13}C-labeled NOD1 CARD, respectively.

^{15}N-M9 contains 1g/L of $^{15}NH_4Cl$, which constitutes the only source of nitrogen. ^{15}N-^{13}C-M9 also possesses, in addition to $^{15}NH_4Cl$, 2g/L of {U-^{13}C}-glucose. This initial preculture in M9 is grown in order to acclimate the bacteria to this poor medium. It also allows avoiding the phenomena of isotopic dilution.

On day three, the 40 ml of ^{15}N-M9 or ^{15}N-^{13}C-M9 precultures were used to inoculate 1 L of ^{15}N-M9 or 1 L of ^{15}N-^{13}C-M9, respectively. The bacteria were grown for several hours at 37°C with a 200 rpm agitation. When the OD_{600} reached O.8-1, expression was induced with 1 mM IPTG and the cells were incubated for an additional five hours at 30°C with a 200 rpm agitation. Cells were then harvested as described in 2.1.1.

Components	**Amount (1L)**
$Na_2HPO_4.7H_2O$	10g
KH_2PO4	3g
NaCl	0.5g
$^{15}NH_4Cl$	**1g**
pH adjusted to 7.2 with NaOH before autoclaving	
Components (filter-sterilized)	**Amount (1L)**
Glucose (for ^{15}N-M9) or ^{13}C-glucose (for ^{15}N-^{13}C-M9)	**2g** (10ml solution 20 %)
MgSO4	1ml (solution 1M)
$CaCl_2$	0.1ml (solution 1M)
Thiamine	2ml (solution 1 mg/ml)
Kanamycin	1ml (solution 10 mg/ml)

Table 6. ^{15}N-M9 and ^{15}N-^{13}C-M9 compostion. $^{15}NH_4Cl$ constitutes the only nitrogen source available for both ^{15}N-M9 and ^{15}N-^{13}C-M9. Normal glucose is used for ^{15}N-M9, whereas ^{13}C-glucose is used for ^{15}N-^{13}C-M9 as only carbon source.

2.3 Protein purification

2.3.1 NOD1 CARD

The compositions of buffers used for cell lysis and for NOD1 CARD purification are presented in Table 7.
The purification was controlled at each step with 15 % SDS-PAGE gels.

	NOD1 lysis/binding buffer	**NOD1 wash buffer**	**NOD1 elution buffer**	**TEV protease buffer**	**NOD1 size-exclusion buffer**
Tris-Cl pH 7.5 (mM)	25	25	25	50	0
Sodium-phosphate pH 7 (mM)	0	0	0	0	25
NaCl (mM)	500	500	500	500	100
Glycerol (%)	10	10	10	10	0
Imidazole pH 7.5 (mM)	0	10	500	0	0
EDTA (mM)	0	0	0	0.5	0
β-ME (mM)	5	5	5	0	0
DTT	0	0	0	1	5
Roche Complete EDTA-free	1 tablet	1 tablet	0	0	0

Table 7. NOD1 CARD lysis and purification buffers. Composition of buffers used for cell lysis, Ni-NTA column purification, TEV protease cleavage and size exclusion are shown.

Bacterial pellet corresponding to a 5-6 L culture was resuspended in 150-180 ml of NOD1 lysis/binding buffer (cf. Table 7). Cell lysis was achieved by sonication on ice at 4°C (8 pulses of 30 s, separated by a 1 minute pause). The resulting lysate was centrifuged for 45 minutes at 40 000 g to pellet the cellular debris. The supernatant obtained was then clarified through a sterile 0.45 µm filter unit (Nalgene).
NOD1 CARD was first purified by a gravity-flow Ni-NTA (Qiagen) affinity column. The resin is provided charged with Ni^{2+}, as a 50 % suspension in 30 % ethanol. 2 ml of 50 % Ni-NTA slurry resin (1 ml resin total) was manually packed into an empty Econo-Pac column (Biorad). The resin was firstly washed with 10 ml of distilled water to remove any trace of ethanol, and then equilibrated with 10 ml of NOD1 lysis/binding buffer. The clarified protein lysate was then loaded onto the column. The recombinant protein possesses the ability to bind the resin due to the affinity of its His-tag for the nickel. The column was then extensively washed with 10 ml of NOD1 wash buffer (cf. Table 7). This buffer contains 10 mM of imidazole and allows the elimination of most of contaminants binding the column non-

specifically. NOD1 CARD was then eluted using 20 ml of NOD1 elution buffer (cf. Table 7) containing 500 mM of imidazole.

Fractions containing NOD1 CARD were pooled and dialysed overnight at 4°C in a dialysis cassette with a molecular weight cut-off of 7000 Da (Pierce) in the presence of His-tagged TEV protease (1:100 w/w), against 1 L of TEV protease buffer (cf. Table 7). This step allows the cleavage of the His-tag by the TEV protease. The protein was then dialyzed a second time for four hours at 4°C against 1 L of NOD1 lysis/binding buffer to remove EDTA and DTT.

The protein was then applied onto a fresh Ni-NTA column and was eluted using lysis/binding buffer. This step allows the removal of both uncleaved NOD1 CARD and His-tagged TEV protease.

After concentration (Amicon Ultra, Millipore), NOD1 CARD was further purified to homogeneity using size-exclusion FPLC (Superdex 75; Amersham Biosciences) in NOD1 size-exclusion buffer (cf. Table 7). This buffer is suitable for the NMR study.

The protein was finally concentrated (Amicon Ultra, Millipore) to 0.5 or 0.8 mM for the unlabeled NOD1 CARD and 1.4 mM for the ^{15}N and ^{15}N-^{13}C-labeled NOD1 CARD used for NMR experiments. The degree of isotope labeling of the protein after purification was verified to be higher than 98 % using MALDI-TOF mass spectrometry.

2.3.2 RICK CARD

As mentioned previously, RICK CARD possesses a non-cleavable amino-terminal His-tag. Purification on Ni-NTA column as for NOD1 (Qiagen) followed by a size-exclusion column was first attempted. Unfortunately, most of the protein purified after the Ni-NTA column formed soluble aggregates. The final yield obtained after size-exclusion was too low for structural studies.

I have thus adapted this protocol by exploiting a protocol which is sometimes used for the purification of His-tag proteins, involving the use of decreasing pH (Sambrook and Russell, 2001).

Compositions of buffers used for cell lysis and RICK CARD purification are presented Table 8.

The purification was controlled at each step with 15 % SDS-PAGE gels.

	RICK lysis/binding buffer	RICK wash buffer 1	RICK wash buffer 2	RICK low pH elution buffer	RICK size-exclusion buffer
Sodium-phosphate pH 8.5 (mM)	20	20	0	0	0
Sodium-phosphate pH 5.5 (mM)	0	0	20	0	0
Sodium-phosphate pH 4 * (mM)	0	0	0	20	0
Citrate (mM)	0	0	0	0	25
NaCl (mM)	500	500	500	500	200
Glycerol (%)	10	10	10	10	0
Imidazole pH 7.5 (mM)	0	50	0	0	0
β–ME (mM)	5	5	5	5	0
DTT	0	0	0	0	5
Roche Complete EDTA-free	1 tablet	0	0	0	0

Table 8. RICK CARD lysis and purification buffers. Composition of buffers used for cell lysis, Ni-NTA column purification and size exclusion are described. * indicates that for the concerned buffer, the pH was lowered using HCl.

Cells lysis was done similarly as for NOD1 CARD, using RICK lysis/binding buffer (cf. Table 8). RICK CARD was first purified by a gravity-flow Ni-NTA (Qiagen) affinity column. The column was prepared as seen in 2.3.1, but using 4 ml of Ni-NTA slurry resin (i.e. 2 ml resin total), and equilibrated with 20 ml of RICK lysis/binding buffer. After loading of the protein sample, the column was extensively washed with:

- 6 column volumes (12 ml) of RICK lysis/binding buffer
- 6 column volumes (12 ml) of RICK Wash buffer 1: this buffer contains 50 mM of imidazole and allows eliminating most of the contaminants binding the column non-specifically.
- 3 column volumes (6 ml) of RICK lysis/binding buffer, to eliminate the imidazole
- 6 column volumes (12 ml) of RICK wash/buffer 2: the change of pH from 8.5 to 5.5 allows the elution of contaminants still binding the resin

RICK CARD was then eluted using 6-10 column volumes of RICK low pH elution buffer. The very acidic pH used allows one to elute soluble non-aggregated RICK CARD, and to "break up" the soluble aggregates. Fractions of approximatively 1 ml were manually collected in 1.5 ml eppendorf tubes containing 150 µl of RICK wash buffer 2.

After concentration (Amicon Ultra, Millipore), RICK CARD was further purified to homogeneity using size-exclusion FPLC (Superdex 75; Amersham Biosciences) in RICK size-exclusion buffer (cf. Table 8). The final yield is 600 µg for a 6 L culture.

3 Cellular biology

Cell culture components are from Gibco (Invitrogen), except where specified otherwise. Experiments were performed using sterile technique and a hood. Care has been taken to ensure that all solutions were sterilized by autoclaving and/or filtration through a 0.22 μm filter unit (Nalgene), and that the DNAs used for tranfection were not contaminated.

3.1 HEK-293T cells

HEK-293T cells (Human embryonic kidney) were used for all *in vivo* experiments. They were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10 % foetal bovine serum (FBS), 2 mmol/L of L-glutamine, 100 U/mL of penicillin and 100 ug/mL streptomycin. Cells were grown in an incubator at 37°C in presence of 5 % of CO_2.

3.2 Co-immunoprecipitation experiments

3.2.1 Calcium-phosphate transfection of HEK-293T cells

Transfection is the introduction of nucleic acids into eukaryotic cells. The calcium-phosphate method is one of the earliest techniques described (Graham and van der Eb, 1973). DNA is first mixed with calcium chloride. A buffered saline-phosphate solution is then added in a controlled manner and induces the formation of a complex between DNA and calcium-phosphate. The resulting precipitate is then dispersed dropwise onto the culture cells. It is then taken up by the cells *via* endocytosis or phagocytosis. This method has the advantage of being easy to use and inexpensive. However, the quantity of DNA needed is very large.
HEK-293T cells were co-transfected with either pcDNA3-NOD1-HA mutants and wild-type pcDNA3-RICK-FLAG, or with wild-type pcDNA3-NOD1-HA and pcDNA3-RICK-FLAG. Three controls were included: (1) co-transfection with wild type pcDNA3-NOD1-HA and wild-type pcDNA3-RICK-FLAG, (2) co-transfection with wild type pcDNA3-NOD1-HA and empty pcDNA3 and (3) co-transfection with wild-type pcDNA3-RICK-FLAG and empty pcDNA3. In cases (2) and (3), empty pcDNA3 was used as a neutral carrier to keep the quantity of DNA used constant. One plate of cells was set up for each immunoprecipitation experiment.

- Protocol

The protocol presented below is for one co-transfection (one 150 mm cell culture dish).

One day prior to the transfection, HEK-293T cells were plated at 40 % of confluence into a 150 mm cell culture dish (Corning). This dilution allows the cells to be in logarithmically growth on the day of transfection.

The day of transfection, a mix containing 1317 µl of sterile distilled water, 20 µg of the first expression plasmid (e.g. pcDNA3-NOD1-NHA), 20 µg of the second expression plasmid (e.g. pcDNA3-RICK-FLAG) and 183 µl of 2 M $CaCl_2$ was prepared. 1500 µl of HEPES-buffered saline (HBS) was then added dropwise to the tube while mixing. HBS is composed of 50 mM HEPES pH 7.15 and 1.5 mM $NaPO_4$. Great care was taken for the preparation of this buffer because its pH is critical for transfection efficiency. After incubation of the mixture for 10-15 minutes at room temperature, the DNA $CaPO_4$ precipitate obtained was dispersed dropwise onto the cells (1 drop per second). At this point, HEK-293T cells were incubated at 37°C in presence of 5 % CO_2 for 6 hours. The media was then removed. The cells were washed twice with 5 ml DMEM and 25 ml of fresh complete DMEM media was added. The cells were incubated overnight at 37°C.

3.2.2 Co-immunoprecipitation

Co-immunoprecipitation is a method used to test whether two proteins interact *in vivo*. It relies on the principle of immunoprecipitation. Immunoprecipitation is a rapid and simple means commonly used to separate a specific protein from total cell lysates. It is based on the use of an antibody directed against the specific protein. This antibody generates an immune complex with its target protein in the total cell lysate. This complex is then captured on a solid support such as sepharose beads to which either Protein A or Protein G is covalently coupled. Protein A and Protein G have the ability to bind specifically to constant regions (Fc) of many mammalian immunoglobulins. The antibody-antigen complex is then washed to remove contaminants that do not precipitate with the immobilized Protein A or G support (non-immune complex sample components). Finally, the immune complex is eluted from the support and further analysed by gel electrophoresis and immunoblotting.

Co-immunoprecipitation is carried out similarly. However, in this case the target protein recognized by the antibody co-precipitates with a binding partner, also present in the total cell lysate.

All the co-immunoprecipitations performed were repeated three times.

- Protocol

The protocol presented below is for one co-immunoprecipitation experiment (one 150 mm cell culture dish). It can be divided into six main steps:

1. sample preparation and cell lysis: HEK-293T cells were harvested 15-20 hours after transfection and washed twice with 5 ml of PBS. Cells were collected in 10 ml of PBS and pelleted 5 minutes at 1200 rpm at 4°C. The pellet was resuspended in 1 ml of cold 0.2 % Nonided P-40 lysis buffer (10 mM Hepes pH 7.4, 142 mM KCl, 5 mM $MgCl_2$, 1 mM EGTA, 0.2 % NP40, Roche complete EDTA-free) in a prechilled 1.5 ml Eppendorf tube. Nonided P-40 is a non-ionic detergent with a low denaturing level commonly used for immunoprecipitation lysis buffers. Lysis was achieved by a 30 minute incubation on ice. The cell lysate was then centrifuged for 30 minutes at 15000 g at 4°C. 40 µl of the supernatant was saved for the gel-electrophoresis analysis and mixed with 80 µl of 3x SDS-loading buffer (total lysate sample).
2. preclearing: the supernatant was transferred in a new prechilled 1.5 ml Eppendorf tube containing 20 µl of Rec-protein G-Sepharose 4B conjugate (Zymed). The tube was rocked for 30 minutes at 4°C. The sample was then centrifuged for 1 minute at 15000 g and at 4°C. The supernatant was transferred to a new prechilled 1.5 ml Eppendorf tube. This preclear step lowers the amount of non-specific proteins in the cell lysate and also removes proteins that may have possible non-specific interactions and a high affinity for Protein G.
3. Incubation with mouse anti-FLAG antibody (Sigma) and formation of antibody-antigen complexes: 2 µg of mouse anti-FLAG antibody (Sigma) was added to the tube. The antigen recognized by the mouse anti-FLAG antibody is the RICK-FLAG. The tube was rocked at 4°C for two hours.
4. precipitation of the immune complex: 20 µl of Rec-protein G-Sepharose 4B conjugate (Zymed) were added to the tube. The sample was rocked for 1 hour at 4°C to precipitate the immune complex formed by the mouse anti-FLAG antibody and the RICK-FLAG. If NOD1-NHA interacted with RICK-FLAG, it was co-precipitated with it.
5. washing of the antibody-antigen complexes to remove non-specific proteins and contaminants: the beads were recovered by a 1 minute centrifugation step at 1700 g and at 4°C. They were then washed with 1 ml of cold 0.2 % Nonided P-40 lysis buffer. This step of centrifugation/washing was repeated another 7 times. The beads were finally resuspended in 20 µl of 3x SDS-loading buffer (immuno-precipitated proteins sample).

6. SDS-PAGE analysis: total lysate and immuno-precipitated proteins samples were boiled at 100°C for 5 minutes. Electrophoresis on four 10 % SDS-PAGE gels was performed to separate denatured proteins by mass for each co-immunoprecipitation realized. 20 µl of the total lysate sample were run out on 2 gels, whereas 10 µl of the co-immunoprecipitated proteins sample were run out on 2 other gels. Immunoblotting was then performed (cf. 3.2.3)

3.2.3 Immunoblotting

Western-blotting is a method for the detection of a specific protein in a sample using an antibody directed against this particular protein. In our case, we want to detect the presence of NOD1-NHA and RICK-FLAG wild type and/or mutant proteins in both the total lysate sample and in the co-immunoprecipitated proteins sample. In this aim, one gel run with the total lysate samples and one gel run with the co-immunprecipitated samples were immunoblotted with rabbit anti-HA polyclonal antibody (Santa Cruz Biotechnology) to detect NOD1-NHA proteins. RICK-FLAG proteins were similarly detected using mouse anti-FLAG monoclonal antibody (Sigma).

- Protocol for NOD1-NHA proteins detection

This protocol is composed of 4 main steps:

1. Transfer: proteins were first moved from within the gel onto a PVDF membrane (Millipore Immobilon) by semi-dry electrophoresis for 1h45 at 20V. This transfer makes the proteins accessible for antibody detection by exposing then on a thin surface layer.
2. Blocking: the PVDF membrane was shaken for 1 hour at room temperature on a rocker in blocking buffer (5 % non-fat dry milk, TBS, 0.05 % Tween-20). This step prevents non-specific protein interactions between the membrane and the antibody used by blocking non-specific binding sites.
3. Primary antibody detection: the PVDF membrane is then incubated overnight at 4°C on a rocker in blocking buffer containing rabbit anti-HA polyclonal antibody (1/1000 dilution).
4. Washing: the PVDF membrane is then rinsed three times at room temperature in TBS containing 5 % of Tween-20 or Nonided P-40. This step allows the removal of any unbound primary antibody.
5. Secondary antibody detection: the PVDF membrane is then shaken 1 hour at room

temperature in blocking buffer containing a 1/10000 dilution of anti-rabbit antibody (Jackson immunoresearch). This antibody specifically reacts with the heavy chains on rabbit IgG. By interacting with the primary antibody, it allows to enhance the signal.

6. Washing: the PVDF membrane is then rinsed three times at room temperature in TBS containing 5 % of Tween-20 or Nonided P-40 to remove any unbound secondary antibody.
7. Analysis: the secondary antibody used was linked to the horseradish peroxidase. The PVDF membrane is incubated 1 minute at room temperature with the reagents of the ECL western blotting system (Amersham Biosciences). The ECL western blotting system is an extremely sensitive light emitting non-radioactive method for the detection of immobilized specific antigen, directly or indirectly, with horseradish peroxidase. The fluorescence produced is proportional to the amount of protein and detected by a sensitive sheet of photographic film (Kodak BioMax). The film was initially exposed for 1 minute and developed. The exposure time exposure was then adjusted.

- Protocol for NOD1-NHA proteins detection

This protocol is practically identical to the previous one. It only differs concerning the antibodies used. The primary antibody was the mouse anti-FLAG monoclonal antibody mentioned previously. The secondary antibody was an anti-mouse antibody conjugated with the horseradish peroxidase (Jackson immunoresearch).

3.3 NF-κB activation experiments

The ability of NOD1 mutants to activate the transcription factor NF-κB was studied using a luciferase reporter NF-κB assay. The reporter plasmid vector pBVIx-luc (Inohara et al., 2000) was used. It contains six NF-κB recognition sites within the promoter sequence linked to the luciferase reporter gene. The firefly luciferase reporter system is a very sensitive system broadly used in mammalian cells. The luciferase catalyses a bioluminescent reaction in the presence of luciferin and ATP. Lysates of cells transfected with a luciferase reporter will produce a flash of light in the presence of luciferin. The intensity of the emitted light can be detected using a luminometer. This intensity is proportional to the luciferase concentration, and consequently to the NF-κB activity.

A β-galactosidase reporter colorimetric assay was performed as an internal reference to normalize the transfection efficiency variability between the different samples. The pCMV-

βgal vector (Inohara et al., 1999) was used. The β-galactosidase catalyses the hydrolysis of several β-galactosides. We used the *o*-nitrophenyl-β-D-galactopyranoside (ONPG) as substrate. ONPG is colorless. However, its hydrolysis by β-galactosidase results in o-nitrophenol which is yellow in alkaline solution (λ_{max} 420 nm at pH 10.2). In the presence of an excess of ONPG, the OD_{420} of the assayed lysate increases linearly with the concentration of β-galactosidase present.

- Protocol for NF-κB activation assays

One day prior to transfection, $4x10^4$ HEK-293T cells per well were plated into 24 well cell culture microplates (Corning). Transfections were performed using the LipofectAMINE Plus transfection reagent (Invitrogen) according to the protocol given by the supplier. The cells were co-transfected using 5 ng pBVIx-luc, 10 ng pCMV-βgal, 5 ng of wild-type or mutant NOD1 plasmid and 395 ng pcDNA3 per well. A control was also included without any NOD1 plasmid. In this case, 400 ng of pcDNA3 plasmid was used instead of 395 ng to keep the total amount of transfected DNA constant. Cell lysis was performed twenty hours post-transfection using 1x reporter lysis buffer (Promega). This buffer has the advantage to allowing luciferase and β-galactosidase assays to be performed from the same cell extract. Luciferase and β-galactosidase activity were determined as described (Inohara et al., 1999). Luciferase values were normalized for transfection efficiency by dividing by β–galactosidases values. NK-κB assays were performed twice in triplicate.

4 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) was the method used to solve the structure of NOD1 CARD. The main aim of this chapter is not to explain in detail the physical principles behind this technique, but rather to introduce briefly the basic concepts and to explain the methodology followed to determine this three-dimensional structure using biological NMR. The methods concerning the study of protein dynamics by nuclear relaxation and potential intermolecular interactions with RICK CARD are also described.

4.1 Introduction

Two methods can be used to determine the three-dimensional structure of biological macromolecules (e.g. proteins, nucleic acids, sugars, macromolecular complexes) at atomic resolution: X-ray crystallography and NMR.

NMR is a powerful biophysical technique and is the only one which can be applied for the study of the structure of macromolecules in the solution phase. It is also widely used to investigate dynamic processes (Kay, 1998) and to study intermolecular interactions (Zuiderweg, 2002).

The basic phenomenon of NMR was discovered in 1945 by Felix Bloch and Edward Purcell. In 1952, both of them received the Nobel Prize in Physics. NMR spectroscopy relies on an intrinsic magnetic property of atom nuclei (e.g. those of hydrogen) that is called spin. Nuclear properties of atomic nuclei with a ½ spin commonly studied in NMR are presented in table 9

Isotope (I=½)	Frequency at 14T (MHz)	Natural abundance (%)	Relative sensitivity at constant field for equal number of nuclei
^{1}H	600	99.98	1
^{13}C	151	1.11	$1.59\ 10^{-2}$
^{15}N	61	0.37	$1.04\ 10^{-3}$
^{31}P	243	100	$6.63\ 10^{-2}$

Table 9. Nuclear properties of selected isotopes with a ½ spin.

When a sample is placed in a strong magnetic field, the spins align along the magnetic field. This is the magnetization of the sample. It is then possible to transfer energy into the spin system from the ground energy level to a higher energy level by applying electromagnetic pulses in the radiofrequency range (excitation). This process is called resonance absorption. Following excitation, the spins reemit the absorbed energy upon returning to their original

states through a phenomenon named relaxation. By doing this, they send a time-function radiofrequency radiation named free induction decay (FID). FID can be recorded, converted to a function of frequency by Fourier transformation, and then displayed as a spectrum.

The nature of the emitted radiation depends on the chemical environment of each nucleus. Furthermore, every nuclear spin in a molecule also senses the magnetic field of its closest neighbors. It is thus possible to differentiate the signals coming from individual nuclei and to assign them. The study of the resonance frequencies and of the existing correlations between the different spins of the molecule gives insights into the structure and dynamics of the molecule considered.

NMR is a young method compared to X-ray crystallography. The first NMR structure (proteinase inhibitor IIA) was solved only in 1985 by Wüthrich and colleagues (Williamson et al., 1985), whereas the first X-ray structure (myoglobin) was determined in 1958 (Kendrew et al., 1958).

Since 1985, many technological and methodological developments have improved the quality of the structural information available by NMR: the use of high field spectrometers, cryogenic probes, isotopic labeling, new spectroscopic techniques, etc. These advances have improved both the sensitivity (signal-to-noise ratio) and the spectral resolution (resonance linewidth) of the technique, and hence have led to an increase in the size of molecules suitable for NMR solution determination. Size still remains one of the limiting factors. Indeed, NMR spectroscopy can be applied to structure determination by routine NMR techniques only for proteins of size up to 40 kDa and nucleic acids up to 10 kDa. However some solution structures of bigger macromolecules have been successfully solved, such as that of the malate synthase G (723 residues, 82 kDa) (Tugarinov et al., 2005).

4.2 Strategy used to solve a protein structure by NMR

A structural study of a protein in solution involves different experimental and computational phases, namely, sample preparation, data acquisition, resonance assignment, derivation of structural restraints, calculation of a structural ensemble and validation of the structural model (cf. Figure 19). These different steps are detailed in paragraphs 4.3 to 4.7. Dynamics and interaction studies may be performed once the resonance assignment is completed (cf. 4.8 and 4.9).

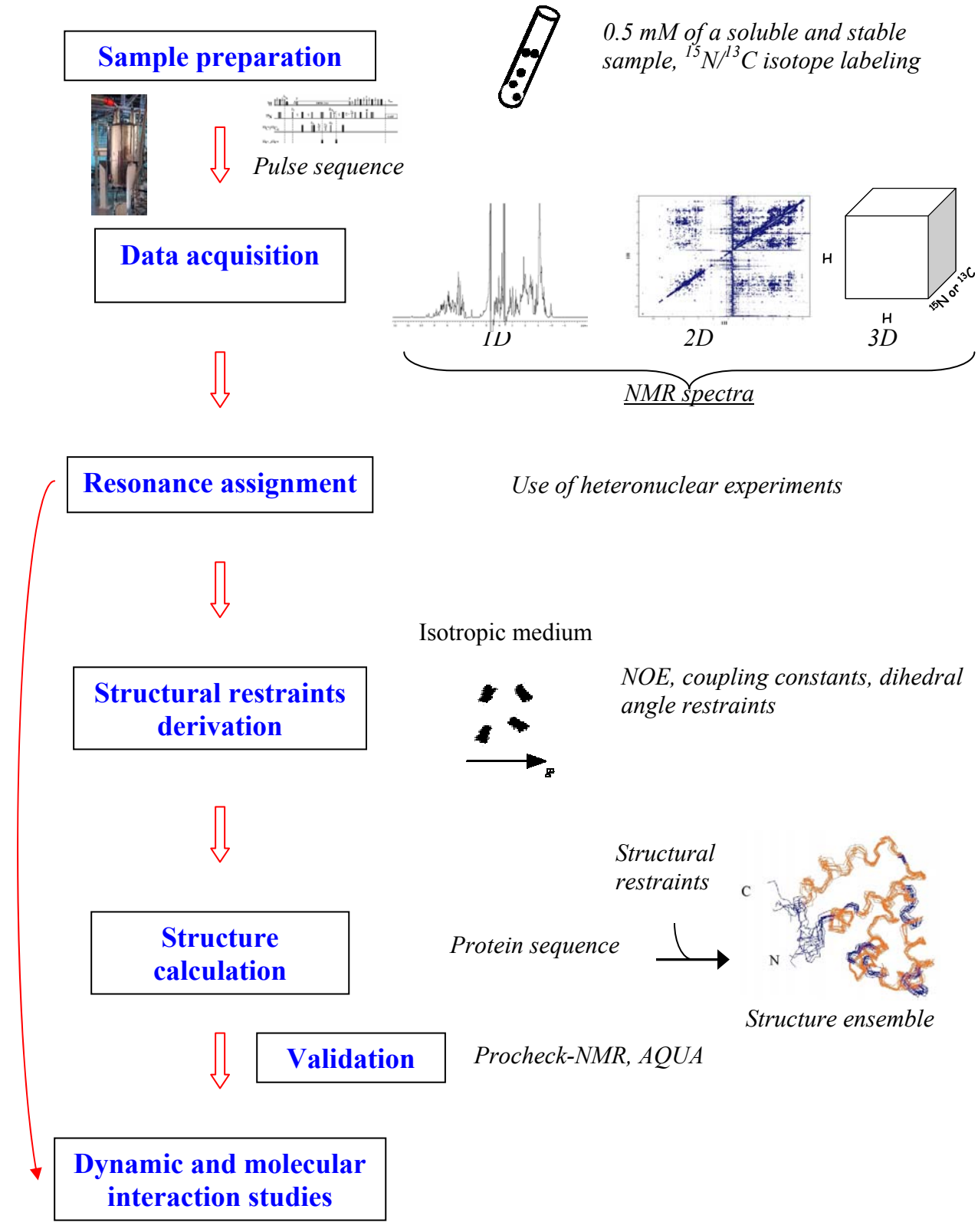

Figure 19. Protein study by NMR.

4.3 Sample preparation

This first step consists of expressing and purifying the protein sample. Because of the low natural abundance of ^{13}C and ^{15}N nuclei (cf. Table 9), protein samples need to be enriched for these isotopes. This isotope labeling is a function of their size. Indeed, for proteins of size below 10 kDA, it is not required and the protein may be studied using only proton NMR. However, as the size increases, labeling with ^{15}N (for proteins between 10 to 14 kDa), with both ^{15}N and ^{13}C (between 14 and 20 kDA) or even with ^{15}N, ^{13}C and ^{2}H is required. This is achieved by recombinant expression of the studied protein in bacteria grown in isotope enriched media, which results in the uniform incorporation of desired isotope(s). The use of isotopes allows one to increase the resolution of NMR spectra and facilitates the chemical identification and assignment of nuclei.

Sample preparation is a crucial step because the quality of the resulting NMR data depends on it. Great care must be taken particularly concerning the quantity of protein used and its stability. Indeed, NMR being a poorly sensitive method, a high quantity of protein is needed: the concentration of the sample should be of at least 0.5 mM in 450 µl (or in 300 µl when Shigami tubes are used). Furthermore, the acquisition time of the triple resonance experiments is several days long. The sample should thus be very stable, if possible for several weeks at room temperature. It can be prepared under inert atmosphere (Ar, N_2) to avoid oxidation of methionines and cysteines.

The expression and purification of the NOD1 CARD samples is described in 2.3.1. To summarize, six samples were prepared for the NMR experiments: 500 µl at 0.5 mM and 500 µl at 0.8 mM for unlabeled NOD1 CARD, 500 µl at 1 mM and 500 µl at 1.4 mM for the ^{15}N-labeled NOD1 CARD, and finally 500 µl at 1 mM and 350 µl at 1.4 mM for the ^{15}N-^{13}C-labeled NOD1 CARD. These samples were prepared in the presence of 5 % D_2O under inert atmosphere.

4.4 NMR spectroscopy, spectra processing and analysis

NMR spectra recorded for the structure determination were acquired at 30°C on Varian INOVA 800 or Varian INOVA 600 spectrometers. They were processed with NMRPipe (Delaglio et al., 1995) and analyzed using NMRView (Bruce A. Johnson, 1994; Johnson, 2004). Smartnotebook (Slupsky et al., 2003) was used for semiautomatic sequential NMR resonance assignment.

4.5 Resonance assignment

4.5.1 Sequential assignment principle

Resonance assignment constitutes the real beginning of the NMR study. Firstly, a 2D NMR experiment named heteronuclear single quantum correlation (HSQC) is performed. It correlates the nitrogen atom of an amide groupe (N) with the directly attached proton ($^{1}H^{N}$). Each signal in a HSQC spectrum represents the ^{15}N-$^{1}H^{N}$ pair of a particular amino acid.
In the case of a $^{15}N/^{13}C$ labeled protein, resonance assignment relies on the use of triple resonance experiments (cf. 4.5.2.). These experiments allow connecting the backbone amide proton (H^{N}) and the backbone amide nitrogen (N) of each residue to the Cα and Cβ of both the preceding residue and its own. This correlation is based on the use of scalar coupling constants existing through one (^{1}J) or two (^{2}J) chemical bonds (cf. Figure 20).

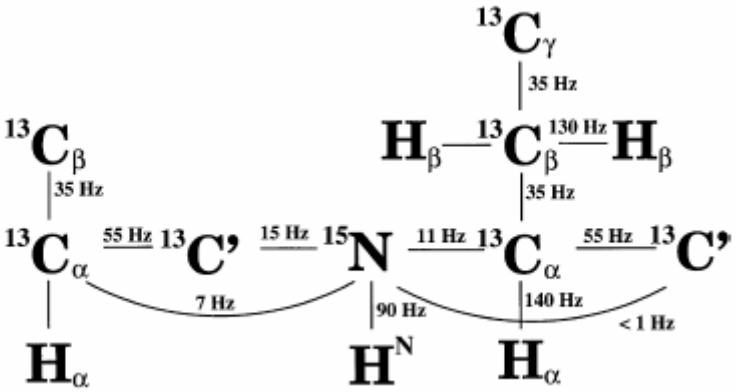

Figure 20. Spin system of the peptide backbone and values in hertz of the ^{1}J and ^{2}J coupling constants used for the magnetization transfer in $^{15}N/^{13}C$-labeled protein, from Sattler et al., 1999.

Neighboring residues are identified by comparison of their Cα and Cβ chemical shift. The next step is to determine the position of the identified residues in the primary sequence of the protein. By adding successively assigned stretches, the sequence of nuclei along the polypeptide backbone can be completely rebuilt. Nuclear Overhauser effect spectroscopy (NOESY) experiments may be employed in a complementary manner to facilitate or confirm this sequential assignment. NOESY experiments allow one to obtain short interproton distances (< 5 Å).
The side-chain resonances can then be assigned (cf. 4.5.3) using both total correlation spectroscopy (TOCSY) under homonuclear coupling ^{1}H-^{1}H (^{3}J) or ^{13}C-^{13}C (^{1}J) and NOESY methods.

4.5.2 Backbone assignment

The nomenclature for triple resonance experiments is characteristic of the magnetization transfer pathway. The experiment is named according to the nuclei involved in this magnetization transfer. However, spin systems involved in this process that only act as relay spins without chemical shift evolution are mentioned in parentheses.

For $^{15}N/^{13}C$-labeled NOD1 CARD, the five standard triple resonance experiments listed below were performed on the 1.4 mM sample in the presence of 5 % D_2O: CBCA(CO)NH, CBCANH, HNCA, HN(CA)CO and HNCO (Sattler et al., 1999) (cf. Figure 21). These experiments allow the complete assignment of the backbone nuclei ^{15}N, $^{1}H^{N}$, ^{13}CO, $^{13}C\alpha$ and $^{1}H\alpha$, and also the assignment of the side-chain nuclei Cβ and $^{1}H\beta$. They are used in pairs (e.g. HN(CO)CA with HNCA) (cf. Figure 21). HN-HN contacts were determined on the 1.4 mM ^{15}N-labeled sample by means of a 3 dimensional ^{1}H-^{15}N NOESY-HSQC spectrum.

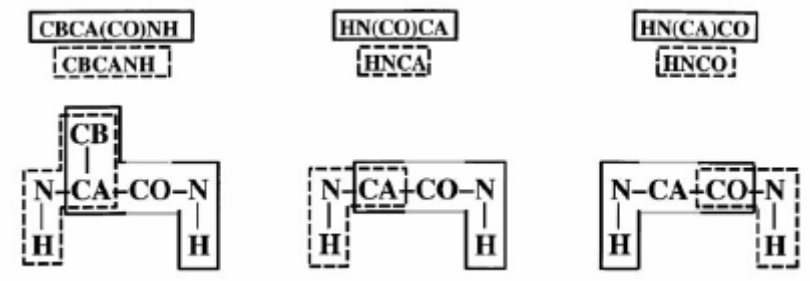

Figure 21. NMR triple resonance experiments for protein backbone resonances assignment, from Sattler et al., 1999. Boxes indicate the correlated spin systems established in each triple resonance experiment. Spin systems used as relay and not edited are depicted in gray.

The combination of two pairs of triple resonance experiments (four experiments in total) is generally sufficient to assign all backbone resonances and to solve any ambiguities. Triple resonance experiments involving a magnetization transfer over the backbone carbonyl carbon (denoted CO) allow connecting a ^{15}N-$^{1}H^{N}$ pair with the backbone carbon atoms of the preceding amino acid. In contrast, experiments without this relay correlate the ^{15}N-$^{1}H^{N}$ both with the carbon atoms of the preceding residues and with those of the same residue.

For example, in the CBCA(CO)NH experiment, the magnetization is transferred from the $^{1}H^{N}$ proton of a residue via its ^{15}N atom and its carbonyl carbon (CO) to the Cα and Cβ nuclei of the previous atom and back the same way. The frequency of the carbonyl carbon is not edited because it acts as a relay nucleus. Frequencies detected are those of the ^{15}N and $^{1}H^{N}$ of the amino acid considered, and those of the Cα and Cβ of the preceding residue. The complementary experiment is named CBCANH. In this case, there is no transfer through the

carbonyl atom. Correlation of the ^{15}N-$^{1}H^{N}$ pair of a given residue with both its own Cα and Cβ (11 Hz coupling constant), and the Cα and Cβ of the preceding residue (7 Hz coupling constant) can be established. Frequencies detected are those of the ^{15}N, $^{1}H^{N}$, Cα and Cβ of the amino acid considered, and those of the Cα and Cβ of the preceding residue. An example of sequential assignment using the pair CBCA(CO)NH/CBCANH is given below (cf. Figure 22).

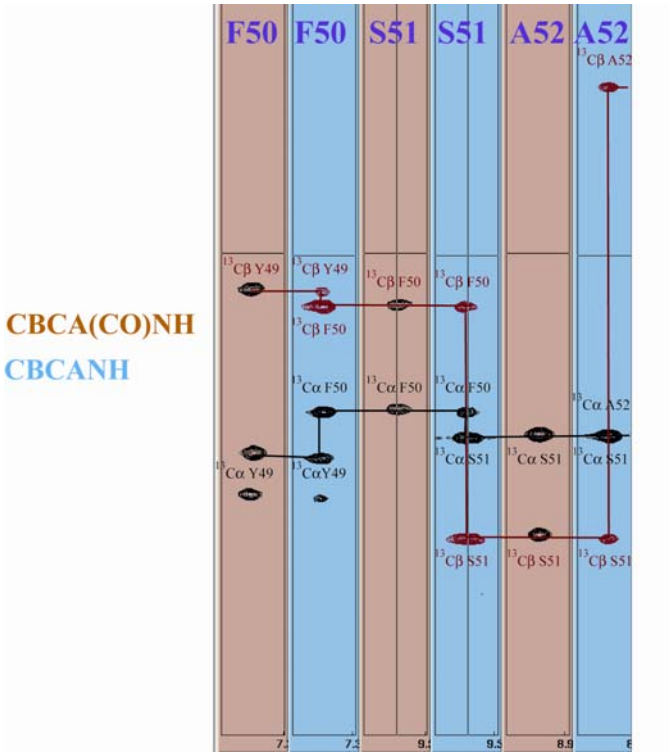

Figure 22. Assignment strategy for 13C/15N-labeled NOD1 CARD using CBCA(CO)NH and CBCANH experiments. CBCA(CO)NH and CBCANH spectra extracted from the ^{15}N dimension for $^{1}H^{N}$ frequencies of F50, S51 and A52. CBCA(CO)NH and CBCA(NH) are colored pink and cyan, respectively. The assignment pathway is indicated with red lines for the $^{13}C\beta$, and with black lines for the $^{13}C\alpha$.

In our NOD1 CARD NMR study, the Smartnotebook program (Slupsky et al., 2003) was used for semi-automated assignment. The software finds and displays the potential connections between residues. They are then manually checked, and either validated or rejected. The program also compares the experimental chemical shifts obtained with those derived from a database. It thus allows fitting the identified polypeptidique chains into the primary sequence of the protein.

It is interesting to point out that knowledge of the backbone resonances gives insights into the structure of the protein assignment using the so-called chemical-shift index (CSI) (Wishart and Sykes, 1994a; Wishart and Sykes, 1994b). Indeed, the chemical shifts of the Hα, Cα, Cβ

and CO are dependent on the local structure. The CSI allows the identification and localization of secondary structures in the protein.

4.5.3 Side-chains assignment

- Aliphatic side-chain assignment

The 1.4 mM NOD1 CARD sample was used for the ^{1}H and ^{13}C aliphatic side-chain assignment. HBHA(CO)NH, (H)CC(CO)NH, H(CCCO)NH, H(C)CH-TOCSY experiments (cf. Figure 23) were performed in the presence of 5 % D_2O. Another H(C)CH-TOCSY was also performed in 100 % D_2O.

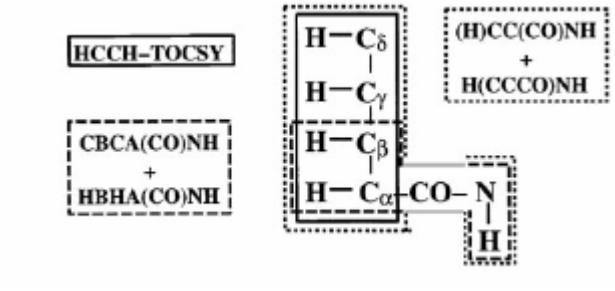

Figure 23. NMR triple resonance experiments for protein side-chain resonance assignment, from Sattler et al., 1999. Boxes indicate the correlated spin systems established in each triple resonance experiment. Spin systems used as relay and not edited are depicted in gray.

An H(C)CH-TOCSY experiment allows the identification of all the ^{1}H and ^{13}C spin systems of the aliphatic side-chains. Correlation with the backbone resonances H^N and N can then be established using CBCA(CO)NH and HBHA(CO)NH experiments. However, the analysis may be problematic. Indeed, the H(C)CH-TOCSY is not directly related to a classical HSQC spectrum. Furthermore, the HBHA(CO)NH does not allow one to assign all the side-chain and ambiguities may remain. This is the reason that another strategy based on (H)CC(CO)NH, H(CCCO)NH was also used. (H)CC(CO)NH and H(CCCO)NH experiments correlate the resonances of the aliphatic carbons or the side-chain protons, respectively, of the preceding residue directly to the backbone amide of a given amino acid.

- Aromatic side-chain assignment

The assignment of aromatic side-chain protons was performed using 1H homonuclear 2D NOESY acquired with 0.8 mM unlabeled protein in the presence of either 5 % D_2O or 100 % D_2O. A 3D ^{13}C-edited NOESY-HSQC with the ^{13}C carrier frequency in the aromatic region was performed similarly on the 1.4 mM $^{15}N/^{13}C$-labeled protein.

4.6 Structural restraints collection

Many different structural restraints may be extracted from NMR experiments, such as: (1) interatomic distances < 5 Å from NOESY experiments; (2) dihedral angle restraints (backbone angle ϕ and ω) from scalar J-coupling ($^3J_{H^N,H\alpha}$) (Karplus and Grant, 1959) or from the CSI (Cornilescu et al., 1999); (3) hydrogen bond restraints, deduced from H-D exchange (Huyghues-Despointes et al., 2001); (4) long-range geometrical information by bond projection angles from residual dipolar couplings (RDC) (Tjandra and Bax, 1997); and (5) long-range distances (10 to 35 Å) or paramagnetic resonance enhancement (PRE) restraints from resonance bleaching experiments induced by spin labels (Battiste and Wagner, 2000; Boisbouvier et al., 1999; Kosen, 1989).
In the context of the NOD1 CARD NMR structural study, NOE, dihedral angles and hydrogen bond restraints were used.

4.6.1 NOEs

NOE restraints constitute the main source of geometrical information used for the NMR structure calculation. The nuclear Overhauser effect is the dipolar interaction of spins for the correlation of protons. NOE restraints are derived from NOESY experiments. The intensity (or volume) of a measured NOE is proportional to $1/r^6$, where r represents the interproton distance.
Characteristic ^1H-^1H distances are observed in regular secondary structure elements (cf. Table 10).

Distance	α-helix	3_{10} helix	β	βp	turn Ia	turn IIa
$d_{\alpha N}$	3.5	3.4	2.2	2.2	3.4	2.2
$d_{\alpha N}$ (i, i+2)	4.4	3.8			3.2	3.2
$d_{\alpha N}$ (i, i+3)	3.4	3.3			3.6	3.3
$d_{\alpha N}$ (i, i+4)	4.2				3.1-4.2	3.8-4.7
d_{NN}	2.8	2.6	4.3	4.2	2.6	4.5
d_{NN} (i, i+2)	4.2	4.1			2.4	2.4

Table 10. Short ^{1}H-^{1}H distances in polypeptide secondary structures, from Wüthrich, 1986.

These characteristic NOEs concern residues that are close in the protein sequence. However, obtaining NOEs between residues that are distant in the protein primary sequence are of particular interest. Indeed, long-range structural information can accelerate and facilitate folding the protein during the structure calculation.

In the NOD1 CARD study, classical NOE restraints were derived from classical 3D ^{15}N-edited NOESY-HSQC performed on the 1.4 mM $^{15}N/^{13}C$-labeled NOD1 CARD. A 3D methyl selected NOESY-HSQC experiment (Van Melckebeke et al., 2004) was also performed on this sample to obtain methyl-methyl distances. Methyl-methyl measurements are a good source of long-range information. The principle is that, even if methyl groups are generally well dispersed in the primary sequence of proteins, they are often found in the protein hydrophobic core. So methyl groups that are distant in the protein sequence may be close in the 3D structure. The knowledge of "long-range" methyl-methyl distances is necessary for the folding of the protein during the structure calculation.

4.6.2 Dihedral angle restraints

Three dihedral angles define the protein backbone: ω, ϕ and ψ. The angle ω corresponds to the peptide bond (usually planar), whereas ϕ and ψ describe the torsional angles about the N and Cα atoms or the Cα and the carbonyl carbon, respectively, of a particular residue (cf. Figure 24).

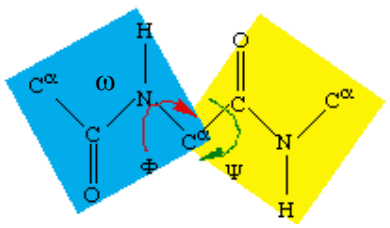

Figure 24. The protein backbone. It can be described with three dihedral angles that are ω, ϕ and ψ.

The dihedral angle restraints can be obtained either by $^{3}J_{HN,H\alpha}$ measurements, or generated based on the ^{1}H and ^{13}C secondary chemical shifts using TALOS (torsion angle likelihood obtained from shift and sequence similarity) (Cornilescu et al., 1999), as was done for NOD1 CARD.

- TALOS

TALOS is database system for the empirical prediction of ϕ and ψ backbone torsion angles based on the fact that secondary chemical shifts are highly correlated with the protein secondary structure (cf. Figure 25).

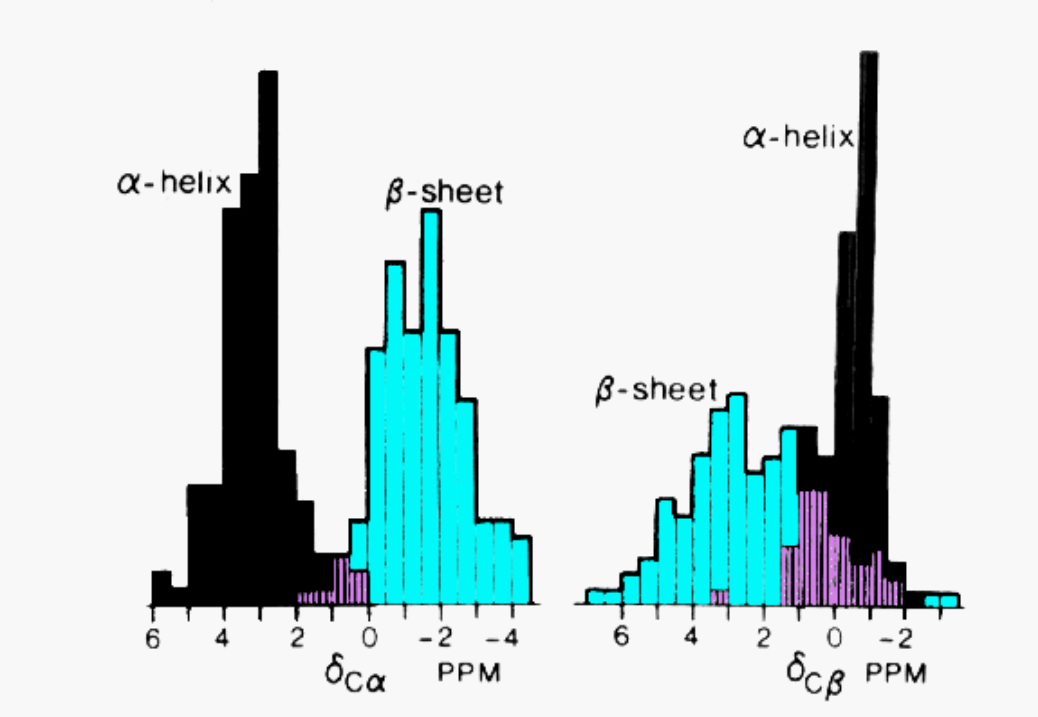

Figure 25. Correlation between the Cα and Cβ chemical shifts and the protein secondary structure, from http://spin.niddk.nih.gov/NMRPipe/talos.

TALOS uses the secondary chemical shifts of three consecutive residues simultaneously to make predictions for the central residue of the triplet. The TALOS database includes data from twenty proteins of known high resolution X-ray crystal structures (2.2 Å or better), representing around three thousand triplets. In practice, TALOS searches in its database the ten best matches to a given triplet in the target protein. If these ten matches have consistent values for ϕ and ψ, then their averages and standard deviations are used as a prediction. In contrast, when the ϕ and ψ values are inconsistent, no prediction is made.

4.6.3 Hydrogen bond restraints

Hydrogen bonding restraints were deduced from backbone chemical shifts and from a H-D exchange experiment analysis. An H-D exchange experiment reveals slowly exchanging amide protons which might form hydrogen bonds (Huyghues-Despointes et al., 2001).
The 1 mM ^{15}N-labeled NOD1 CARD sample was lyophilized and redissolved in D_20. The H-D exchange experiment was then performed by recording a series of ^{15}N-edited HSQC spectra during nineteen hours. Amide protons for which a signal remained visible three hours after the exchange were considered to be involved in a hydrogen bond.

4.7 Structure calculation

4.7.1 CNS and ARIA

To solve a solution structure, all the structural restraints are applied in a simulated annealing dynamics calculation based on a randomized or extended starting structure(s). The result is an ensemble of structures (or structure family) that is consistent with the experimentally derived structural restraints. CNS (crystallographic and NMR system) (Brunger et al., 1998) and ARIA2.0 (ambiguous restraints for iterative assignment) (Habeck et al., 2004) were used for the NOD1 CARD ensemble structure calculation.

- CNS

The simulated annealing strategy of CNS was used. The underlying idea is to use molecular dynamics to heat the starting structure up to a high temperature (1000-2000 K) in a simulation system (i.e. the atoms of the starting structure get a high thermal mobility). The heat causes the system to evolve and wander randomly in conformational space through states of higher energy. During many discrete steps, the system is then slowly cooled down to a minimum for stabilization. This gives the initial structure the possibility of evolving towards the energetically most favourable final state under the influence of a force field derived from the various experimental and geometric constraints. The energy of a molecule can be considered as the sum of three terms: non-bonded interactions, covalent energy, and experimental restraints.

- ARIA

ARIA can be used in combination with CNS for automated NOE assignment and structure calculation. New concepts such the notion of ambiguous distance restraints (ADR) (Nilges, 1995) are exploited. ARIA manages automatically the ambiguous NOEs restraints, contrary to the standard structure calculations which require a long and fastidious manual analysis. Ambiguous NOEs result from the sum of individual contributions (signal overlap). An ADR combines alternative assignment possibilities into one restraint. ARIA allows molecular conformations to be calculated from a list of ADRs. They can thus be used to filter the assignment possibilities. ARIA makes eight iterative calculations with an additional step of refinement in an explicit solvent (water) for the last one (Linge et al., 2003).

At each iteration, ARIA calibrates and assigns the NOESY spectra, merges the restraint lists from different spectra and also calculates a 3D structure ensemble using CNS. 3D structures of lowest energy of iteration i are used by the program for the assignment and calibration of NOEs in the next iteration i+1. It is the role of the spectroscopist to manually verify the NOE assignment generated by ARIA and to decide whether or not they should be introduced in the next iteration. The resulting iterative protocol converges to a structure ensemble and NOE assignments that are consistent.

- NOD1 CARD structure calculation

An initial calculation of the structure of NOD1 CARD was performed with the program CNS using an extended conformation of the protein and the experimental structural restraints described in 4.6. One thousand structures were calculated and the best in terms of energy was used as a starting point for automated NOE assignment of the remaining peaks using the program aria2.0 (Habeck et al., 2004). The program MOLMOL (Koradi et al., 1996) was used to check and display violations of NMR restraints.

The final structure calculation was performed following simulated annealing protocols (Nilges et al., 1988) with CNS and using an extended conformation of NOD1 CARD. NOE intensities were classified as strong (< 2.8 Å), medium (< 3.4 Å) and weak (< 5 Å). One thousand structures were calculated and the forty lowest energy structures were refined in explicit water using CNS.

4.7.2 Structure analysis and quality assignment

The structure family obtained spans a narrow conformational space. The accuracy of an NMR structure can be expressed by the root mean square of deviation (r.m.s.d) of the position of selected atoms in the individual structures from a mean structure (which is calculated previously).

The quality of the structures can be assessed using the AQUA and PROCHECK-NMR programs (Laskowski et al., 1996). They are a means of validating the geometry (e.g. Ramachandran plot) and restraint violations of an ensemble of protein structures solved by solution NMR.

- NOD1 CARD structure analysis

The ten lowest energy structures obtained after explicit refinement were analyzed and validated using the AQUA and PROCHECK-NMR programs.

4.8 ^{15}N relaxation experiments

Protein structures are not rigid. Bonds and even entire segments may bend, move or rotate in a wide variety of amplitudes and time scales, from femtoseconds to minutes (cf. Figure 26).

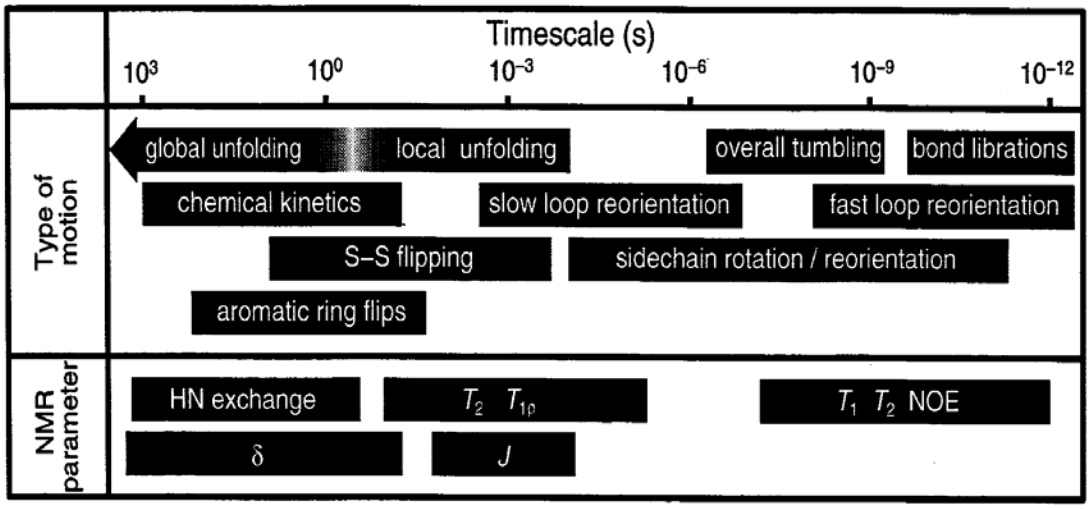

Figure 26. Typical time regions for molecular motions, from http://www.bionmr-1.unl.edu/991/Lectures/NMR-Protein-dynamics.ppt.

The dynamics of protein motion can be investigating by NMR relaxation experiments (Kay, 1998) using isotope-labeled protein. Dynamics can be classified into fast (ns-ps time scale), and slow (µs-ms time scale), depending on the rotational diffusion correlation time. This can

give useful site specific information and functional insights. Indeed, slow time scale motions may be related to biological function.

Relaxation is the post-excitation phenomenon by which spins tend to return to their thermal equilibrium distribution (cf. 4.1). ^{15}N relaxation measurements are an attractive means of studying the main chain dynamics of proteins in solution. ^{15}N are excellent backbone dynamic probes because they are homogeneously dispersed along the polypeptidic chain, and they are linked to one proton.

Dynamics information can classically be obtained by the measure of 3 parameters: R_1, R_2, and heteronuclear ^{15}N $\{^1H\}$-NOE. R_1 and R_2 are respectively the longitudinal relaxation (along the z axis) and the transversal relaxation (along the x and y axes). They are functions of the correlation time τ_c and thus of the size of the protein. R_1 and R_2 depend on the motion and interactions between the spins and the surroundings. ^{15}N $\{^1H\}$-NOEs are measured by saturating the proton signal and observing changes in the ^{15}N signal compared to the unsaturated situation.

Important dynamic parameters can be extracted from R_1, R_2 and heteronuclear heteronuclear ^{15}N $\{^1H\}$-NOEs measurements. The Lipari-Szabo model-free approach (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b) can be used to analyze the measured relaxation and to extract important parameters. This formalism makes a distinction between the global rotational diffusion of the molecule and the time-dependent fluctuations of individual backbone N-H bond vectors. These 2 phenomena are assumed to be completely independent, since they occur on very different time scales (ns vs. ps). The protein is considered as a rigid sphere characterized by a global correlation time τ_c, whereas the N-H vector internal motion is defined both by a local correlation time τ_e and an order parameters S^2 (cf. Figure 27). S^2 parameters describe one amino acid's amplitude of motion. High values indicate rigidity, whereas small values indicate flexibility.

The exchange parameter R_{ex} can be obtained from R_2 measurements. A given residue is in chemical exchange when its measured R_2 value is considerably bigger than the average R_2 value measured.

The methodology used for the NOD1 CARD dynamic studies is presented in Appendix 3.

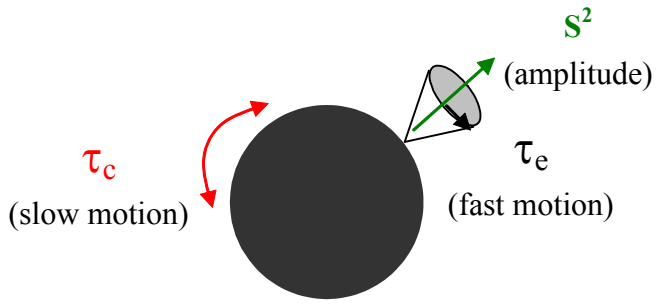

Figure 27. Model-free approach of Lipari-Szabo. The dark grey circle represents the molecule and the green arrow the N-H vector.

4.9 NMR chemical shift mapping

NMR chemical shift mapping experiments may allow one to characterize protein-protein interactions (Clarkson and Campbell, 2003). They rely on the use of HSQC experiments. As seen in 4.5, all amino acids (except proline) of a ^{15}N-labeled protein are represented by single peaks in a ^{1}H-^{15}N-HSQC experiment. When a non-labeled potential binding partner is added, some resonances may change on the assigned spectrum. These perturbations can be correlated with the binding site location if no gross conformational changes occur.

4.10 NOD1 CARD NMR chemical shift mapping

Three NMR chemical shift mapping experiments were performed at 30°C: (1) 0.15 mM of ^{15}N/^{13}C-labeled NOD1 CARD and 0.15 mM of unlabeled RICK CARD in a buffer composed of 25 mM citrate pH 5.6 and 400 mM NaCl; (2) 0.15 mM of ^{15}N-labeled NOD1 CARD and 0.3 mM of unlabeled RICK CARD in a buffer composed of 25 mM citrate pH 5.6 and 60 mM NaCl; 0.15 mM of ^{15}N-labeled NOD1 CARD and 0.3 mM of unlabeled RICK CARD in 25 mM citrate buffer pH 5.6 and 16 mM NaCl. In each case a ^{1}H-^{15}N-HSQC experiment reference was performed using only ^{15}N/^{13}C-labeled NOD1 CARD in the appropriate buffer.

5 Modeling

5.1 RICK CARD modeling

The RICK CARD sequence was aligned with that of procaspase-9 using ClustalW (Thompson et al., 1994). Modeller8v1 (Sali and Blundell, 1993) was then used to model the CARD domain of RICK using both the above sequence alignment and the atomic coordinates of procaspase-9 CARD (obtained from the RCSB protein data bank). Modeller is a useful tool used for homology modeling of protein 3D structures (Marti-Renom et al., 2000). Based on both the sequence alignment given by the user and the known related structure(s), Modeller automatically calculates a model containing all non-hydrogen atoms satisfying the spatial restraints (Sali and Blundell, 1993).

5.2 Modeling of the NOD1/RICK CARD-CARD interaction

The NOD1 CARD structure and RICK CARD model were superimposed on the structure of Apaf-1 and procaspase-9, respectively (Qin et al., 1999). New pdb files were generated using the CCP4 package (Collaborative Computational Project, 1994).

Part III

Results and discussion

Résumé (français)

- Chapitre III.1 : expression, purification et étude préliminaire de résonance magnétique nucléaire du domaine CARD de NOD1

Des alignements de séquences des domaines CARDs de RAIDD, NOD1, NOD2 et RICK ont été réalisés. Ils ont permis de déterminer la construction du domaine CARD de NOD1, utilisée en biologie structurale.

L'expression et la purification de cette protéine sont ensuite illustrées au moyen de gels SDS-PAGE et d'un profil d'élution de filtration sur gel. Le rendement final en protéine pure est approximativement de 3,5 ml / L de culture pour la protéine non marquée et de 2,5 mg / L de culture pour la protéine marquée ^{15}N ou $^{15}N/^{13}C$. Des tests préliminaires de résonance magnétique nucléaire du proton ont été réalisés sur la protéine non marquée : RMN 1D du proton et expérience NOESY. Les spectres obtenus ont permis principalement (1) de vérifier le correct repliement de la protéine et (2) la possibilité d'utiliser la RMN comme méthode structurale pour la détermination de la structure tridimensionnelle.

- Chapitre III.2 : études de résonance magnétique nucléaire du domaine CARD de NOD1

Au début de ce chapitre est résumé, au moyen d'un organigramme, l'ensemble des expériences de RMN effectuées pour déterminer la structure à haute résolution du domaine CARD de NOD1 en solution.

Le spectre de corrélation $^{1}H-^{15}N$ HSQC du domaine CARD de NOD1 est caractéristique d'une protéine bien structurée et montre l'existence de plusieurs conformères. Ce dernier point est particulièrement visible lors de l'étude du spectre de corrélation $^{1}H-^{13}C$ HSQC réalisé dans la région des aromatiques.

L'analyse des déplacements chimiques des atomes du squelette polypeptidique du domaine CARD de NOD1 a permis l'identification des structures secondaires ; la protéine est constituée de six hélices α.

La structure du domaine CARD de NOD1 a été calculée en utilisant le programme CNS en combinaison avec le programme ARIA. Seuls les résidus 15-110 ont été introduits lors du calcul de structure, la partie carboxy-terminale (résidus 111-138) étant extrêmement flexible et manquant de corrélations de distance NOE. Le r.m.s.d moyen calculé sur les structures secondaires est de 0,43 ± 0,07 Å pour les atomes lourds du squelette. Les programmes PROCHECK-NMR et AQUA ont permis de vérifier la bonne qualité de l'ensemble de structures obtenu.

Le domaine CARD de NOD1 se compose d'un paquet compact de six hélices α repliées autour d'un cœur hydrophobe. Ce dernier est composé principalement de résidus hydrophobes hautement conservés dans les domaines CARD des protéines interagissant avec RICK. Bien que possédant une topologie similaire à celle des CARDs de structure connue, le domaine CARD de NOD1 se différencie notamment par la conformation de quelques boucles inter-hélices et par la subséquente orientation adoptée par certaines de ses hélices. En particulier, la sixième hélice présente une orientation inhabituelle, son extrémité carboxy-terminale pointant dans la direction opposée de celle habituellement observée dans les structures connues. De plus cette hélice commence significativement plus tôt dans la séquence et est plus longue.
En dernier lieu, l'étude de dynamique moléculaire réalisée souligne le fort comportement d'échange chimique ou conformationnel de cette sixième hélice.

- Chapitre III.3 : caractérisation de l'interaction CARD-CARD entre NOD1 et RICK

La protéine RICK humaine est constituée de 540 acides aminés. Elle se compose d'un domaine amino-terminal kinase et d'un domaine carboxy-terminal CARD. Le choix de la construction réalisée, utilisée lors des tentatives de caractérisation *in vitro* du complexe NOD1/RICK CARD-CARD, s'est basé sur l'alignement de séquences des domaines CARDs de RAIDD, NOD1, NOD2 et RICK mentionné précédemment (résumé Chapitre III.1). L'expression et la purification du domaine CARD de RICK sont ensuite illustrées au moyen de gels SDS-PAGE et d'un profil d'élution de filtration sur gel. Le rendement final en protéine pure en fin de purification est extrêmement faible, de l'ordre de 100 µg / L de culture. Il s'est avéré impossible de concentrer la protéine au-delà de 0,2-0,3 mM suite à des problèmes d'agrégation. Un spectre de RMN 1D du proton a été réalisé. Il semble vérifier un bon repliement de la protéine mais le signal est extrêmement faible. Il n'a pas été possible de réaliser dans des conditions adéquates des expériences RMN classiques de titration afin de caractériser le site de fixation du CARD de RICK sur le CARD de NOD1. Le domaine CARD de RICK a été modélisé en utilisant la structure cristallographique de procaspase-9. L'interaction CARD-CARD entre NOD1 et RICK a ensuite été également modélisée, de manière très basique, en utilisant la structure cristallographique connue du complexe CARD-CARD entre Apaf-1 et procaspase-9. Ce modèle brut avait simplement pour but de permettre l'indentification de résidus potentiellement importants pour l'interaction. Ces résidus ont alors été mutés dans les protéines entières pour surexpression en cellules humaines HEK-293T. L'effet des diverses mutations a été analysé au moyen d'expériences de co-immunoprécipitation et d'activation du facteur de transcription NF-κB. Les résultats obtenus montrent la nature électrostatique de l'interaction CARD-CARD entre NOD1 et RICK. En

effet, trois résidus du patch acide de NOD1 (E53, D54 et E56) ainsi que trois résidus du patch basique de RICK (R444, R483 et R488) s'avèrent indispensables à l'interaction. De plus, quatre résidus acides (D42) ou hydrophobes du domaine CARD de NOD1 (L40, V41et L44), localisés à l'intérieur ou à proximité immédiate du patch acide mentionné précédemment, empêchent la co-immunoprécipitation avec RICK lorsqu'ils sont mutés en combinaison. Ces acides-aminés forment très probablement une surface d'interaction périphérique à la surface principale formée par E53, D54 et E56. De plus, les expériences d'activation du facteur de transcription NF-κB réalisées ont montré que la mutation ponctuelle d'un de ces résidus, L44, ainsi que celle d'un autre résidu hydrophobe du patch acide de NOD1, I57, affectent l'activation de la voie de signalisation. Ces résidus contribuent probablement à la spécificité de l'interaction CARD-CARD et/ou interagissent probablement avec d'autres régions de RICK.

- Chapitre III.4 : comparaison des interactions CARD-CARD entre NOD1 et RICK et entre Apaf-1 et procaspase-9

Ce chapitre présente une comparaison des interactions CARD-CARD entre NOD1 et RICK d'une part, et entre Apaf-1 et procaspase-9 d'autre part. Cette analyse est basée sur les résultats présentés chapitre III.3, sur le rôle de mutations connues réalisées sur Apaf-1 et procaspase-9, et sur un alignement des séquences de ces différents domaines CARD. Cet alignement a permis la mise en évidence de la conservation, au sein de ces quatre membres de la famille CARD, des résidus nécessaires à l'une ou l'autre de ces interactions. Cette comparaison souligne le fait que tous les résidus nécessaires aux différentes interactions ne correspondent pas forcément.

Note to the readers

Most of the results shown in the coming pages (mainly paragraphs 2, 3 and 4) are part of an article titled "Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK" that was submitted to the "Journal of Molecular Biology" on the 4th July 2006 (cf. Appendix 4). However, additional results, figures and explanations were added to allow a better and more complete understanding.

1 Expression, purification and preliminary NMR study of NOD1 CARD

1.1 Introduction to NOD1 CARD

Human NOD1 is a 953 amino acids protein having a predicted molecular mass of 107.7 kDA and a pI of 6.71 (Expasy ProtParam) (cf. Figure 28). At the beginning of my thesis, there was no reported in the literature of its over expression in bacteria or insect cells. This has prevented any detailed *in vitro* biochemical to be performed. The exact boundaries between the various domains, particularly between the NACHT, NADs and LRD domains, are difficult to predict and are still largely undefined.

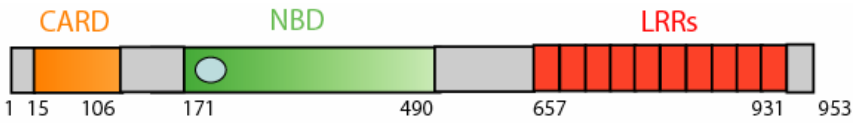

Figure 28. Domain structure of human NOD1, as known from the beginning of this study.

Many constructs encoding for the NOD1 CARD alone or with other domain(s) of the protein were unsuccessfully tested for soluble expression in bacteria and insect cells (cf. Appendix 1). The first construct for bacterial expression encoding the NOD1 CARD alone (residues 15 to 106) was designed based on the below sequence alignment in (cf. Figure 29).

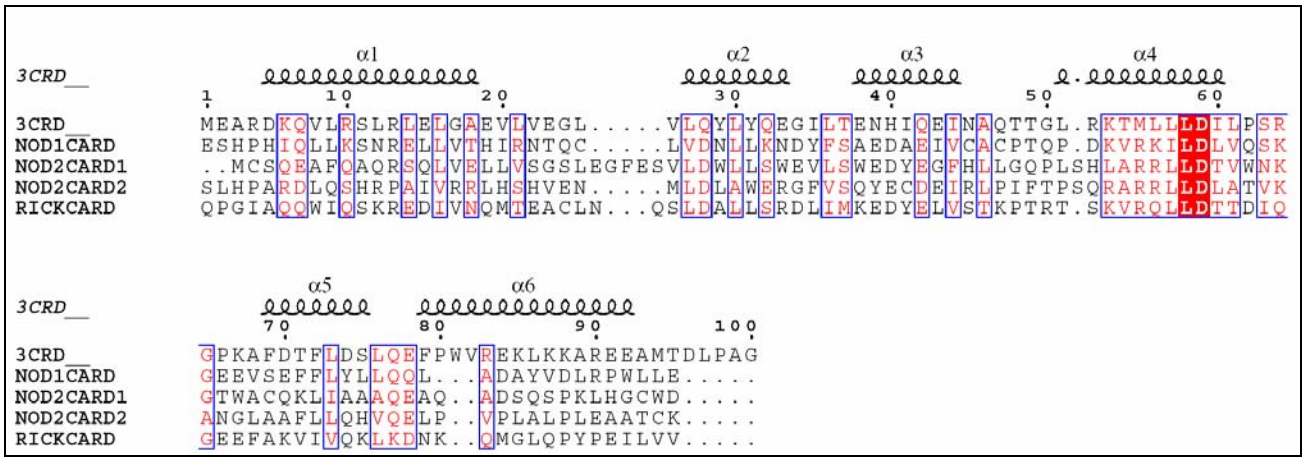

Figure 29. Sequence alignment of NOD1 CARD with other related CARDs. NOD1 CARD is aligned with respectively the first (NOD2CARD1) and the second (NOD2CARD2) CARD of NOD2, the CARD of RICK, and the CARD of RAIDD (referred as 3CRD using its PDB entry). The secondary structure of RAIDD CARD is indicated above the alignement. This alignment was generated using ClustalW and ESPript

This first construct was designed to encompass the characteristic six death-domain α-helical fold. Unfortunately, the resulting protein could only be expressed at levels insufficient for

structural studies. A new construct was therefore designed to try avoiding ending the protein in the middle of a secondary structure. This was based on both a new manual sequence alignment of the CARDs of only NOD1, NOD2 and RICK (cf. Figure 30) and on their secondary structure predictions (not shown).

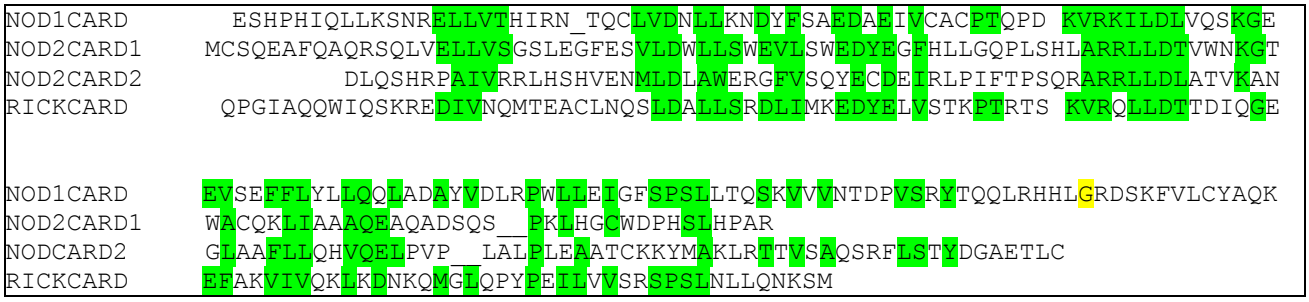

Figure 30. Manual alignment of the CARDs of NOD1, NOD2 and RICK. NOD1 CARD is aligned with respectively the first (NOD2CARD1) and the second (NOD2CARD2) CARD of NOD2 and the CARD of RICK. Similar residues are boxed in green. The NOD1 CARD glycine boxed in yellow symbolises the end of the second NOD1 CARD construct.

This new construct, hereafter named NOD1 CARD, contains 124 residues (15 to 138). It has a predicted molecular weight of 14.6 kDa (Expasy ProtParam) and a pI of 5.7. Its expression level and solubility make it suitable for structural studies.

1.2 NOD1 CARD expression and purification

The protocol used for the expression and purification of NOD1 CARD is explained in details in 2.1, 2.2 and 2.3 of the materiel and method part. The purification is just illustrated here, notably with the corresponding SDS-PAGE 15 % gels.

1.2.1 First Ni-NTA affinity column

The first nickel column allowed purifying NOD1 CARD thanks to the affinity of the His-tag for the nickel. After washing, most of the contaminants binding the resin non-specifically were eliminated. The protein was then eluted already almost pure (cf. Figure 31).

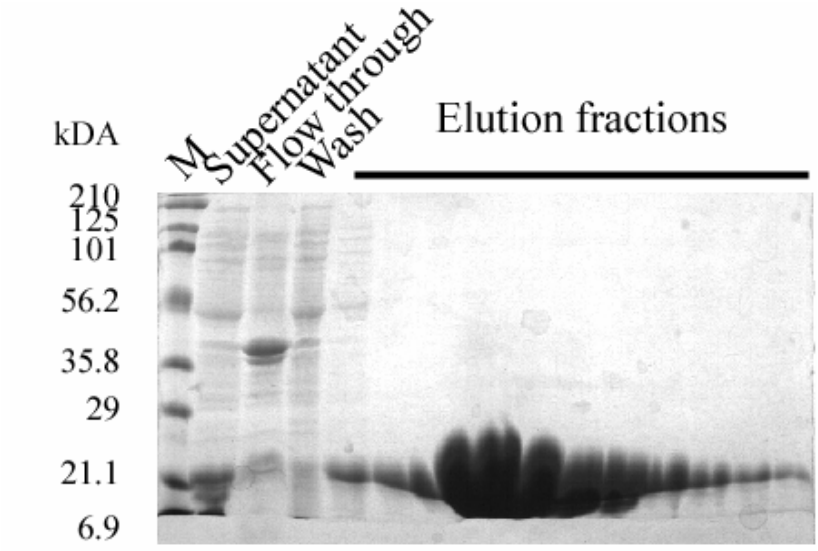

Figure 31. Purification of NOD1 CARD, first nickel column. 15 % SDS-PAGE gel; M =molecular weight markers.

1.2.2 His-tag removal and second Ni-NTA affinity column

NOD1 CARD His-tag was then removed using the TEV protease. This cleavage was not 100 % efficient. It was thus necessary to perform a second affinity nickel column (cf. Figure 32) both to separate the protein with the His-tag from the protein without the His-tag, and also to get rid of the TEV protease which was His-tagged too.

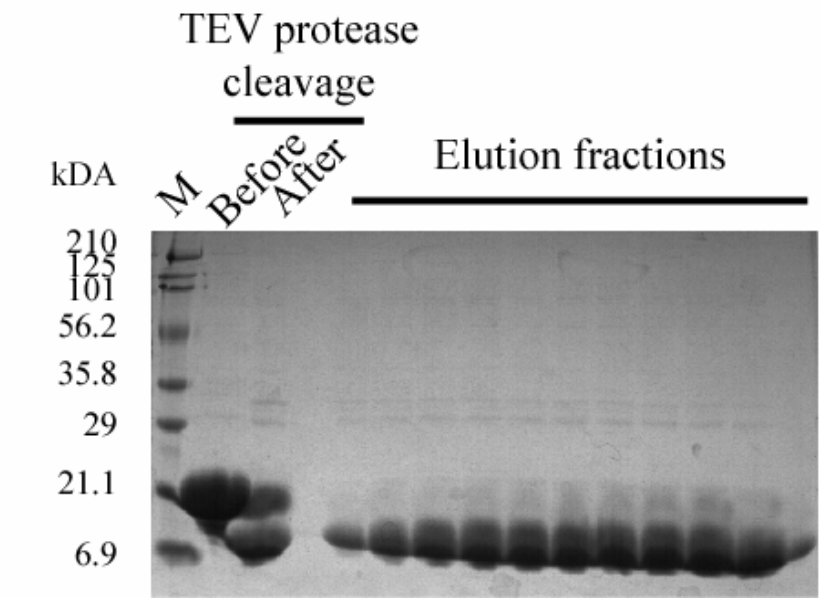

Figure 32. Purification of NOD1 CARD, second nickel column. 15 % SDS-PAGE gel; M =molecular weight markers.

1.2.3 Size-exclusion

After concentration, NOD1 CARD was finally purified for homogeneity through a Superdex 75 column (cf. Figure 33).

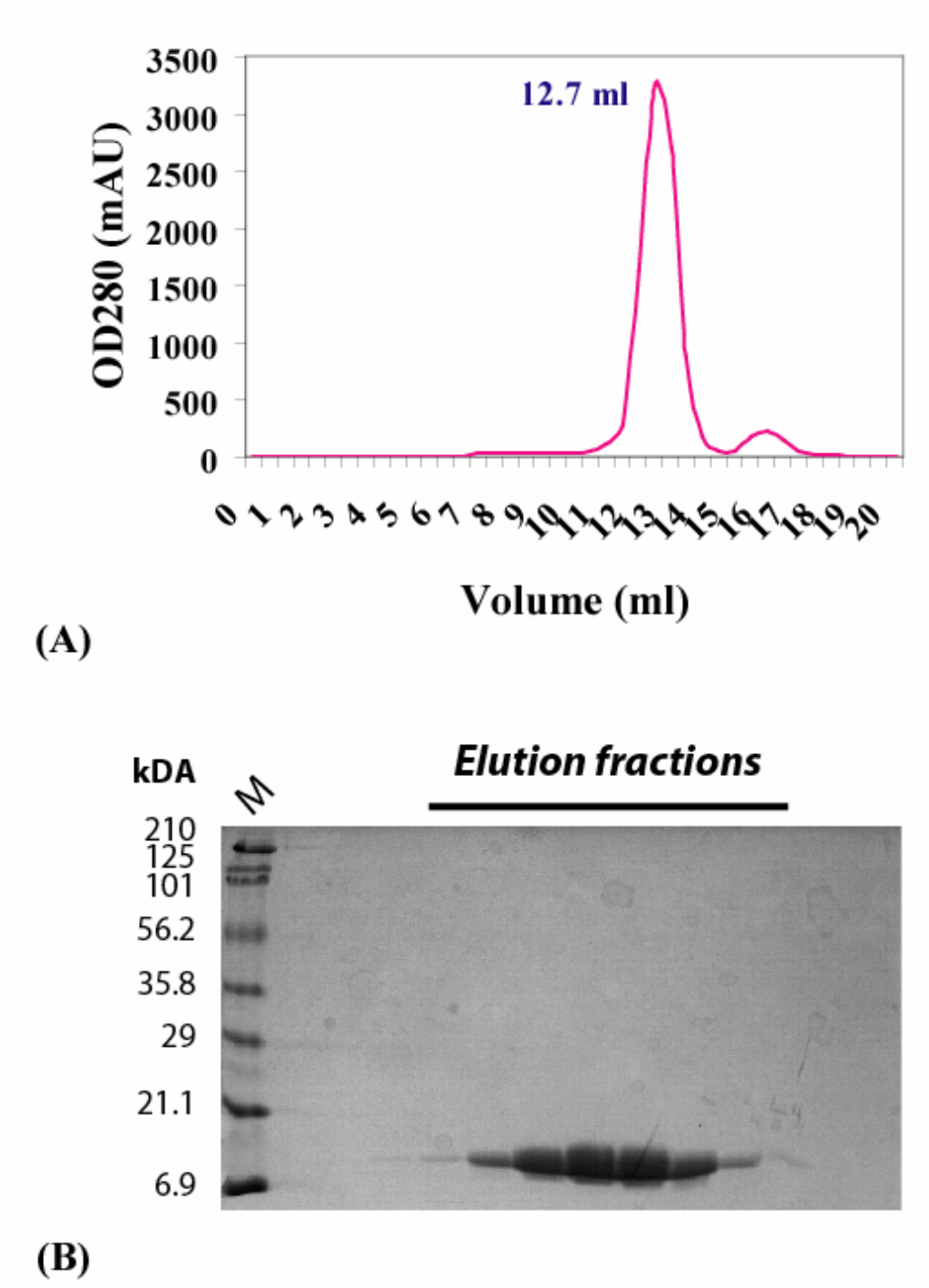

Figure 33. Purification of NOD1 CARD, size-exclusion column. (A) Superdex 75 gel-filtration profile of NOD1 CARD with elution volume of 12.7 ml. (B) 15 % SDS-PAGE gel; M =molecular weight markers.

A single peak was obtained for an elution volume of 12.7 ml. The final yield of pure NOD1 CARD obtained was approximatively 3.5 mg / L for the unlabeled protein expressed in LB,

and 2.5 mg / L of the ^{15}N or $^{15}N/^{13}C$-labeled protein expressed in minimal media M9 enriched with the appropriate isotope(s).

1.3 NMR preliminarily study

Nano-volume crystallisation experiments were attempted using the EMBL HT-Cartesian robo (https://htxlab.embl-grenoble.fr/). This robot allows the screening of 576 different crystallogenesis conditions from classical screens (Hampton etc.) in 90 nl drops. Two NOD1 CARD concentrations (10 and 15 mg / ml) and two temperatures (4°C and 20°C) were tested. Unfortunately, no crystal or suitable crystallisation leads were obtained with any of the 2304 conditions tested.

The NMR method was therefore considered for the 3D NOD1 CARD structure determination, as it is a small and soluble protein. Two samples of unlabeled protein (0.5 mM and 0.8 mM) were tested in 1D NMR and 2D NMR (proton NMR) in the aim to verify that the protein possessed the required qualities for NMR studies

1.3.1 1D NMR

Firstly, two 1D spectra were collected using the two samples at 25 °C (cf. Figure 34). Two different concentrations of NOD1 CARD were measured in order to test if protein aggregation occurs. The resonances cover a spectral window of approximatively 12 ppm with a distribution characteristic of a structured protein; the various types of ^{1}H resonances (e.g. aliphatic protons, amide protons, etc.) are observed. Furthermore, the signal increases accordingly with the protein concentration, meaning that no aggregation of NOD1 CARD was observed while increasing its concentration. This is important because a higher protein concentration allows an increased signal/noise ratio for better sensitivity.

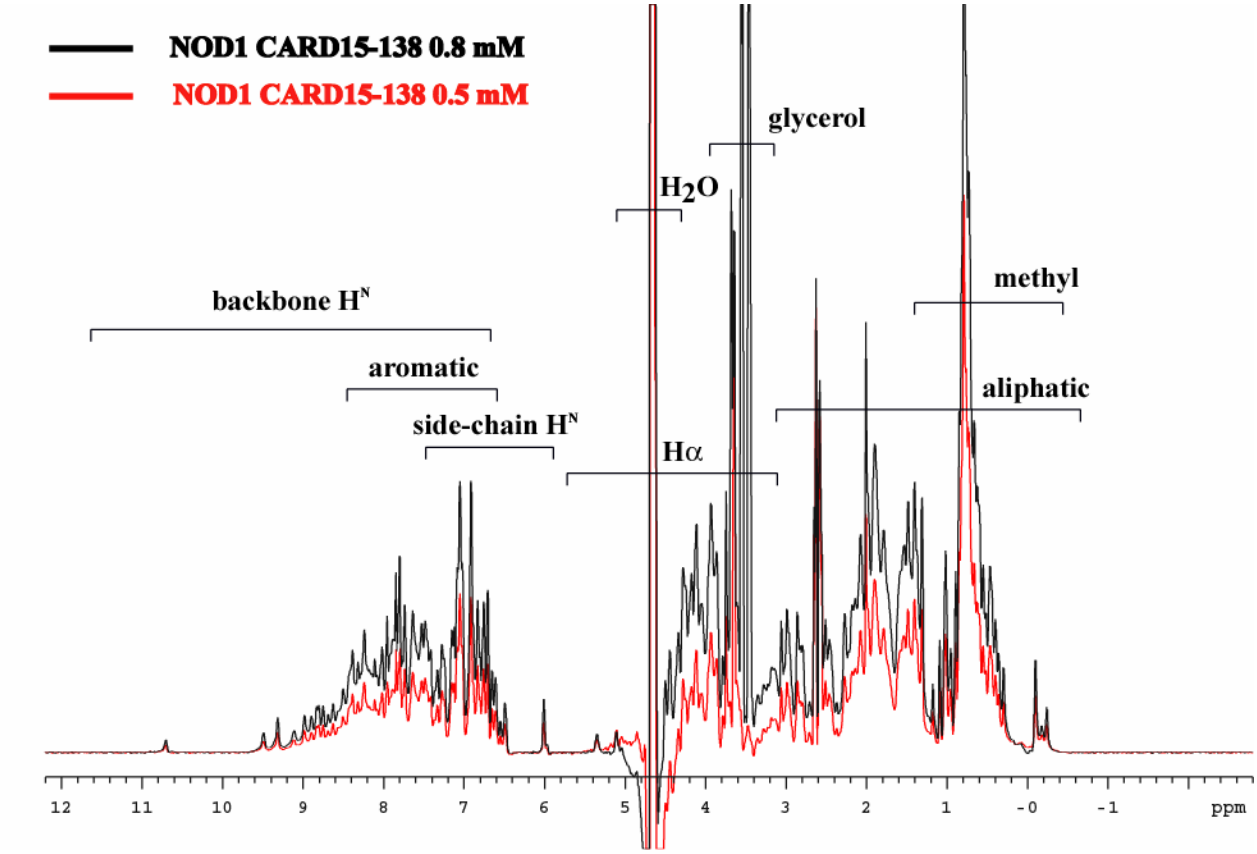

Figure 34. 1D NMR spectra of unlabeled NOD1 CARD at 0.5 mM and 0.8 mM. Chemical shift ranges observed for the various types of protons are indicated on the spectra.

Another important parameter for NMR studies – the protein stability – was also verified by placing the 0.8 mM sample in an oven at 35°C for 10 days. A new 1D NMR experiment was then performed and the spectrum obtained was identical to that presented above, which shows a very good protein stability.

1.3.2 2D NMR

A NOESY spectrum was also performed (cf. Figure 35). This spectrum contains information on the inter-proton distances. A large spectral window is represented and a number of peaks are present, again confirming that the protein is correctly folded.

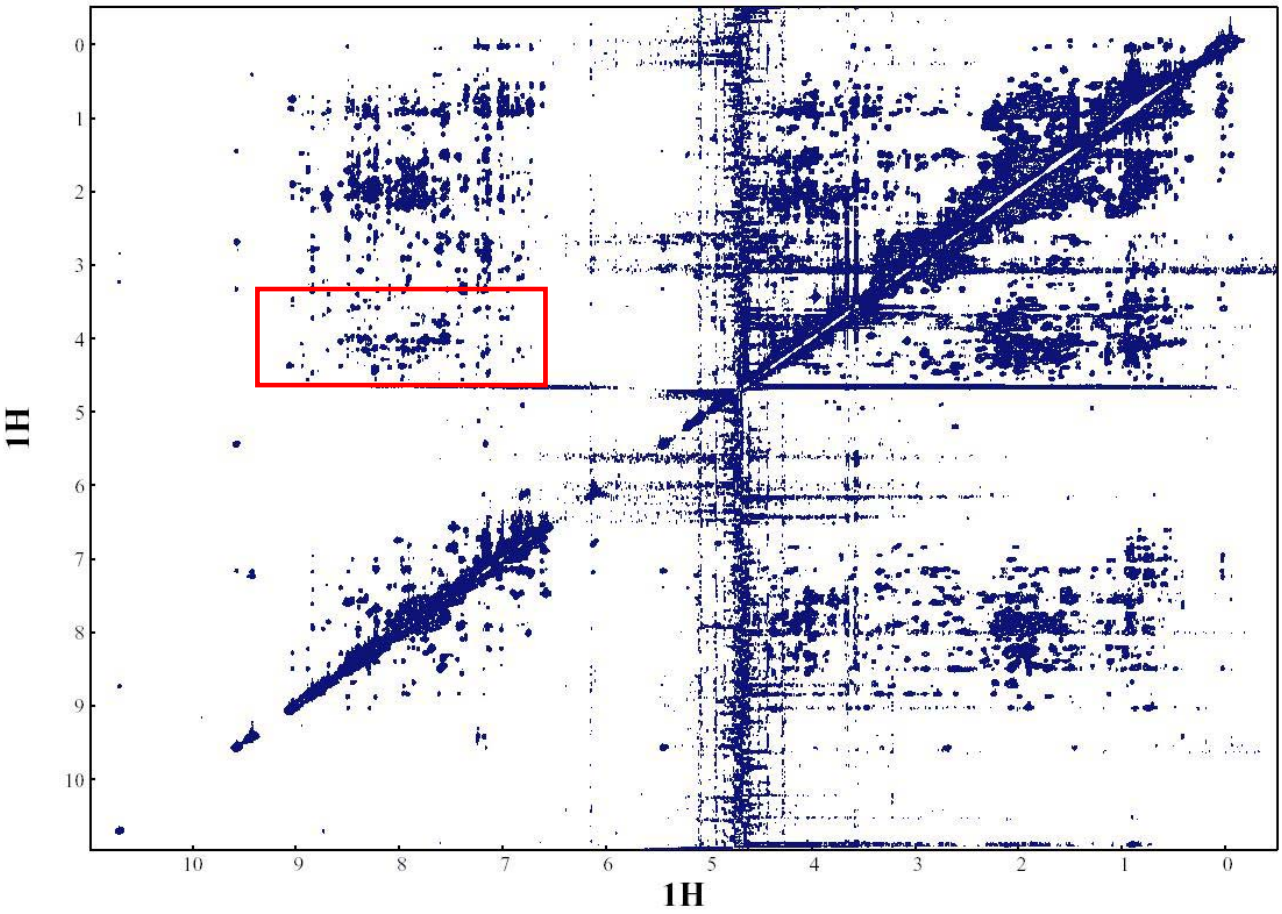

Figure 35. 2D ^{1}H-^{1}H NOESY spectrum acquired with the unlabled NOD1 CARD at 0.8 mM. The red square surrounds the H^{N}-Hα contacts.

Many intense peaks are observed in the red square present on the top left quarter of the spectra. They correspond to the H^{N}-Hα contacts. This shows that the protein is mainly structured in α-helices, as known for CARD domains.

2 NOD1 CARD NMR studies

2.1 Flowchart

All the experimental steps necessary for the resonance assignment and the protein structural restraints are summarized in Figure 36.

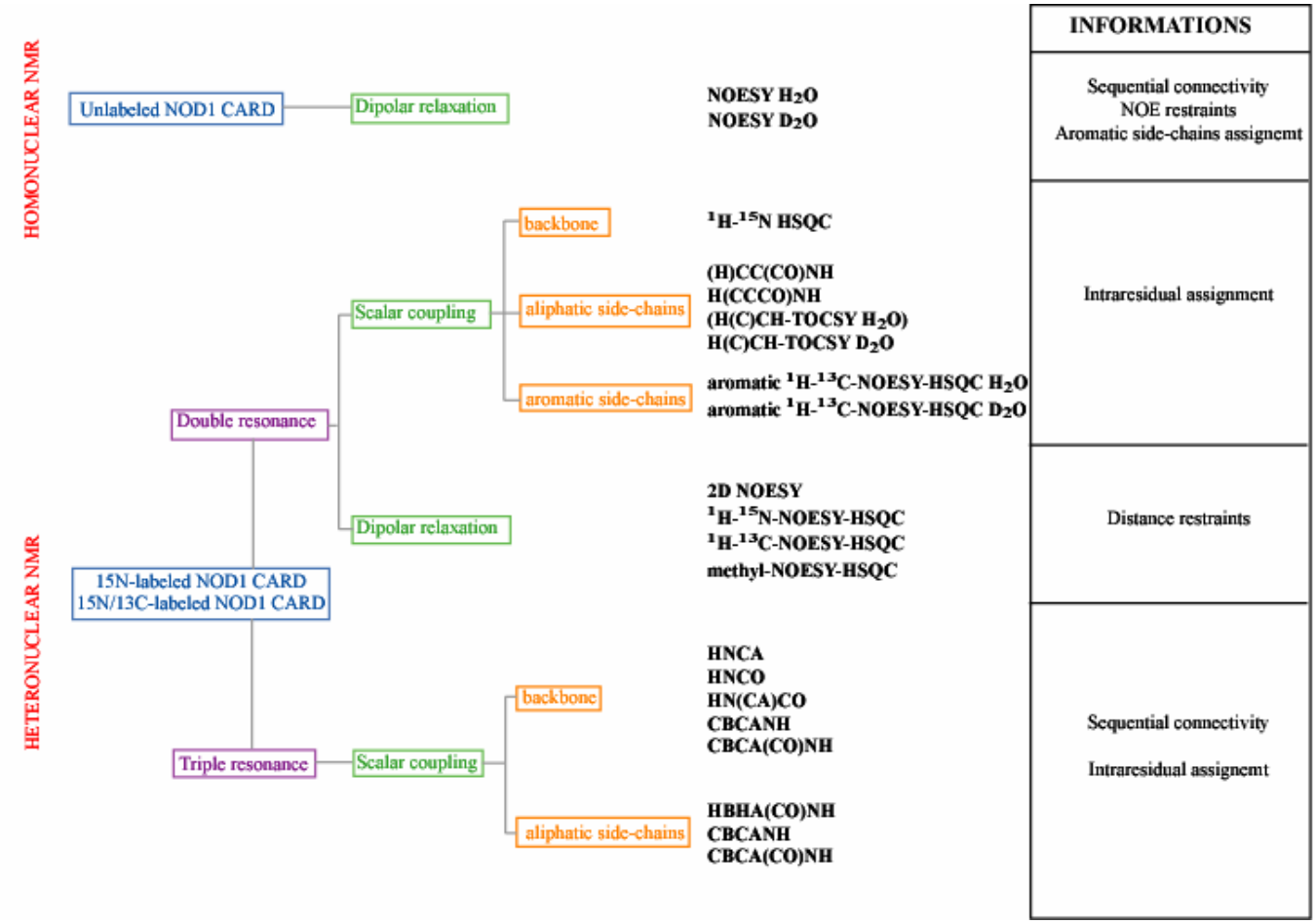

Figure 36. NMR experiments realized to determine the NOD1 CARD structure. All experiments were performed at 30°C on Varian INOVA 600 and Varian INOVA 800 spectrometers. The buffer used for NMR is composed of 25 mM sodium-phosphate pH 7, 100 mM NaCl and 5 mM DTT.

2.2 NOD1 CARD resonances assignment

96.10 % of the backbone resonances were assigned and the completeness of the proton assignment was 87 %. The residues not assigned are mainly located in the carboxy-terminal region (residues 111-138). This region is highly flexible in the picosecond to nanosecond time scale, as shown by low $^{15}N\{^{1}H\}$-NOE values.

Furthermore, the spin systems of L100 and R101 preceding P102 were not observed in the triple resonance spectra used for backbone assignment. The relaxation study detailed in 2.6 shows that these residues are located in the region [Q91-G108] which undergoes conformational or chemical exchange (cf. Figure 48), suggesting that the L100 and R101 NMR lines are broadened by intermediate exchange.

- ^{1}H-^{15}N HSQC spectrum

A ^{1}H-^{15}N-HSQC spectrum was recorded using 1.4 mM ^{15}N-labeled NOD1 CARD (cf. Figure 37). This spectrum correlates all the H^N with their attached N in the protein backbone. This means that a single resonance signal is obtained for each proton-nitrogen pair in the protein backbone. As there is only one H^N per amino acid (proline excluded), each HSQC signal represents a particular residue. However, additional signals corresponding to either the amide groups of the side chains of Asn and Gln, or the aromatic H^N of Trp and His are also present.

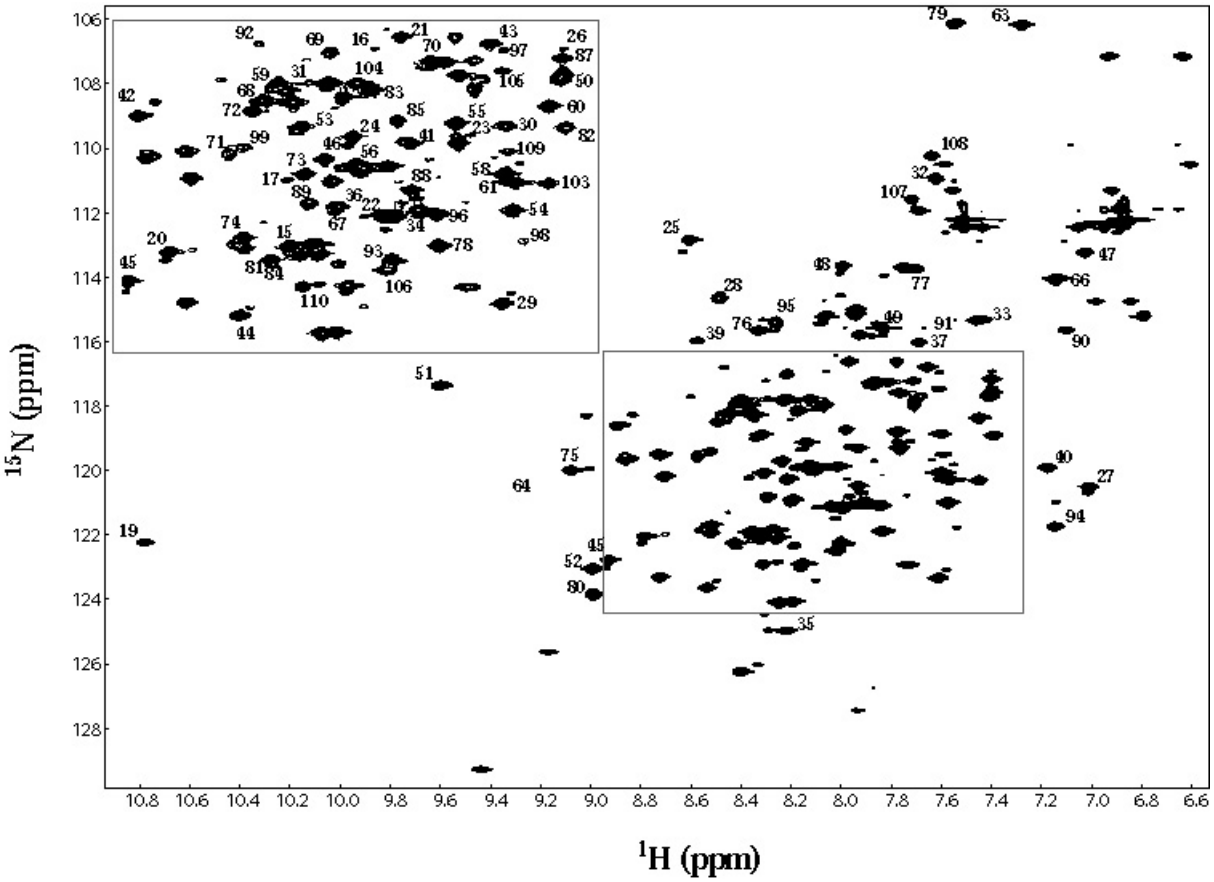

Figure 37. ^{1}H-^{15}N HSQC correlated spectrum of NOD1 CARD recorded at 800 MHz with assignments.

The HSQC spectrum obtained is representative of a correctly folded protein: the peaks are well dispersed, and most of them can be individually resolved. The number of peaks in the spectrum should equal the amount of residues in the protein (plus peaks corresponding to the side-chain amide protons). However, in the case of NOD1 CARD, more peaks were measured than should have been. Furthermore, different peak populations with varying intensities may be observed. This indicates the existence of different conformers, which cannot be explained by cis-trans isomerization of prolines. Indeed, the signals of the different prolines were carefully analysed and no isomerization was observed.

In the region [D95-S110], three sets of resonance lines are present in the NMR spectra for each residue. This pattern is particularly visible in the ^{1}H-^{13}C-NOESY-HSQC correlated spectrum performed with the ^{13}C carrier frequency in the aromatic region (cf. Figure 38). The spectral ranges of the ^{13}C-^{1}H bonds are characteristic of aromatic residues.

- ^{1}H-^{13}C HSQC spectrum

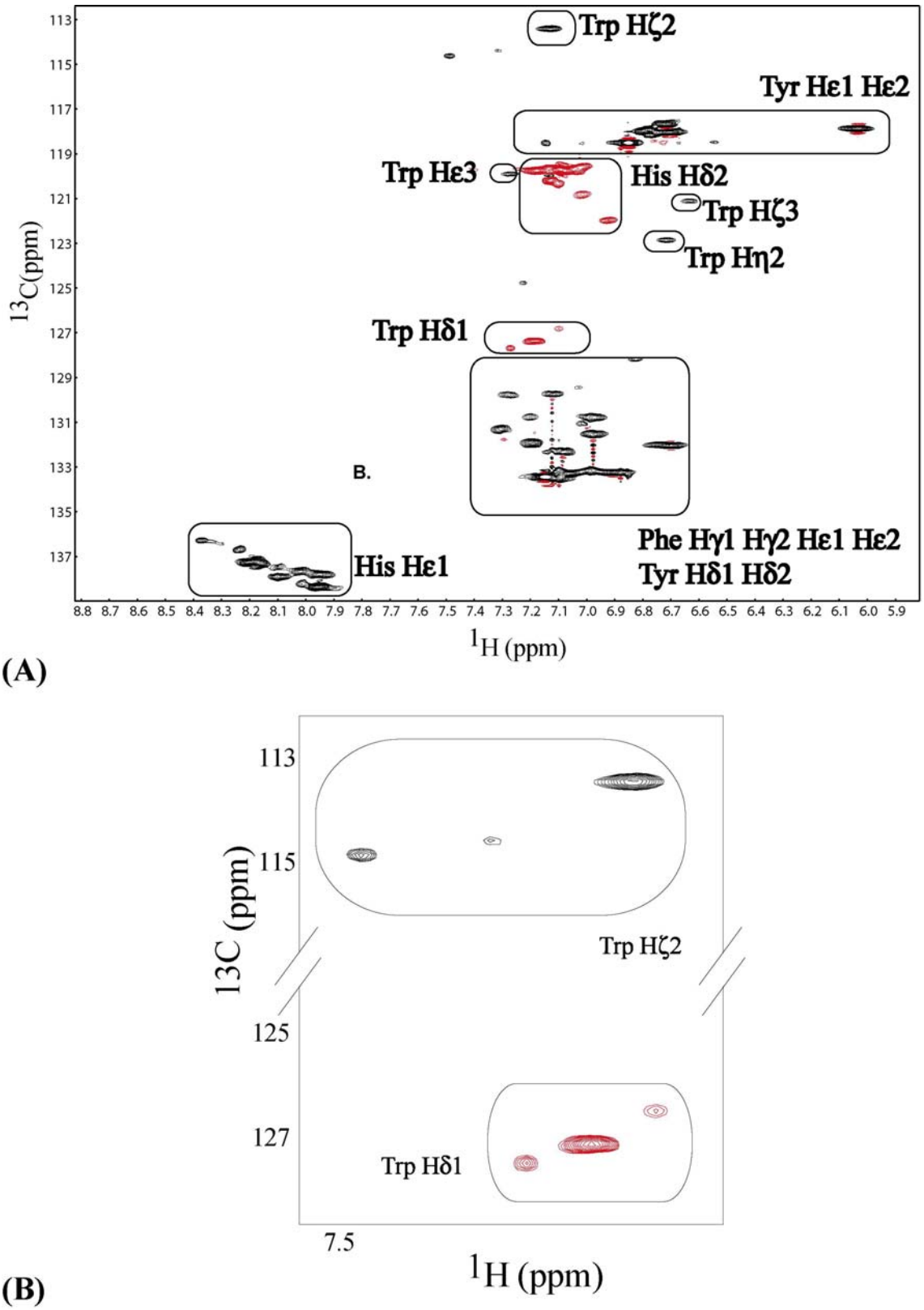

Figure 38. ^{1}H-^{13}C HSQC correlated spectrum in the aromatic region recorded at 600 MHz. (A) whole aromatic region. (B) region of W103.

This experiment is similar to the previous one. It allows linking protons to be coupled with their attached carbons. NOD1 CARD possesses one Trp, four Tyr, four Phe and five His. As previously, there were more experimental peaks than expected. This is particularly obvious in the region of the unique Trp (W103) of the protein. W103 exhibits one major conformation and two less populated conformations with an abundance of approximately 75, 15 and 10 % (see resonances obtained for its Hδ1 or Hζ2, Figure 38(B)).

2.3 Secondary structure

The CSI is a simple technique for identifying protein secondary structures through the analysis of backbone ^{13}C chemical shifts (Wishart and Sykes, 1994a). It is based on the use of four independent chemical-shift measurements, (Hα, Cα, Cβ and C), for the identification and location of secondary structures. This method has been shown to achieve a predictive accuracy of more than 92 % (Wishart and Sykes, 1994a).

The secondary structures of NOD1 CARD identified using the CSI are the following:

- residues 18-25: α-helix
- residues 27-32: α-helix
- residues 38-45: α-helix
- residues 51-59: α-helix
- residues 63-77: α-helix
- residues 80-89: α-helix
- residues 93-106 : α-helix

We can therefore distinguish six α–helices, with the first one (residues 18-32) containing a kink at residue 26, from a comparison with the secondary structures of the known Apaf-1 (Qin et al., 1999; Vaughn et al., 1999; Zhou et al., 1999), procaspase-9 (Qin et al., 1999) and Iceberg (Humke et al., 2000). Furthermore, the last helix (residues 93-106) begins earlier and is longer (fourteen residues) than the helices 6 of Apaf-1 (ten residues), procaspase-9 (ten residues) and Iceberg (five residues)

2.4 3D structure calculation and statistics

The tertiary structure of the CARD domain of NOD1 was refined against a total of 1236 NMR-derived distance constraints (AQUA analysis, (Laskowski et al., 1996)) using CNS in combination with ARIA2.0 (cf. 4.7.1). Only residues 15 to 110 were included in the structure calculations. The remaining parts of the protein were flexible, as shown by ^{15}N relaxation analysis and lack of NOE correlations.

The ten structures of lowest energy obtained after water refinement were analyzed with PROCHECK-NMR to assess their stereochemical quality. Their average r.m.s.d, calculated on the secondary structures, is 0.43 ± 0.07 Å for the backbone heavy atoms and 0.85 ± 0.09 Å for all heavy atoms. The r.m.s.d. obtained from each residue are displayed Figure 39.

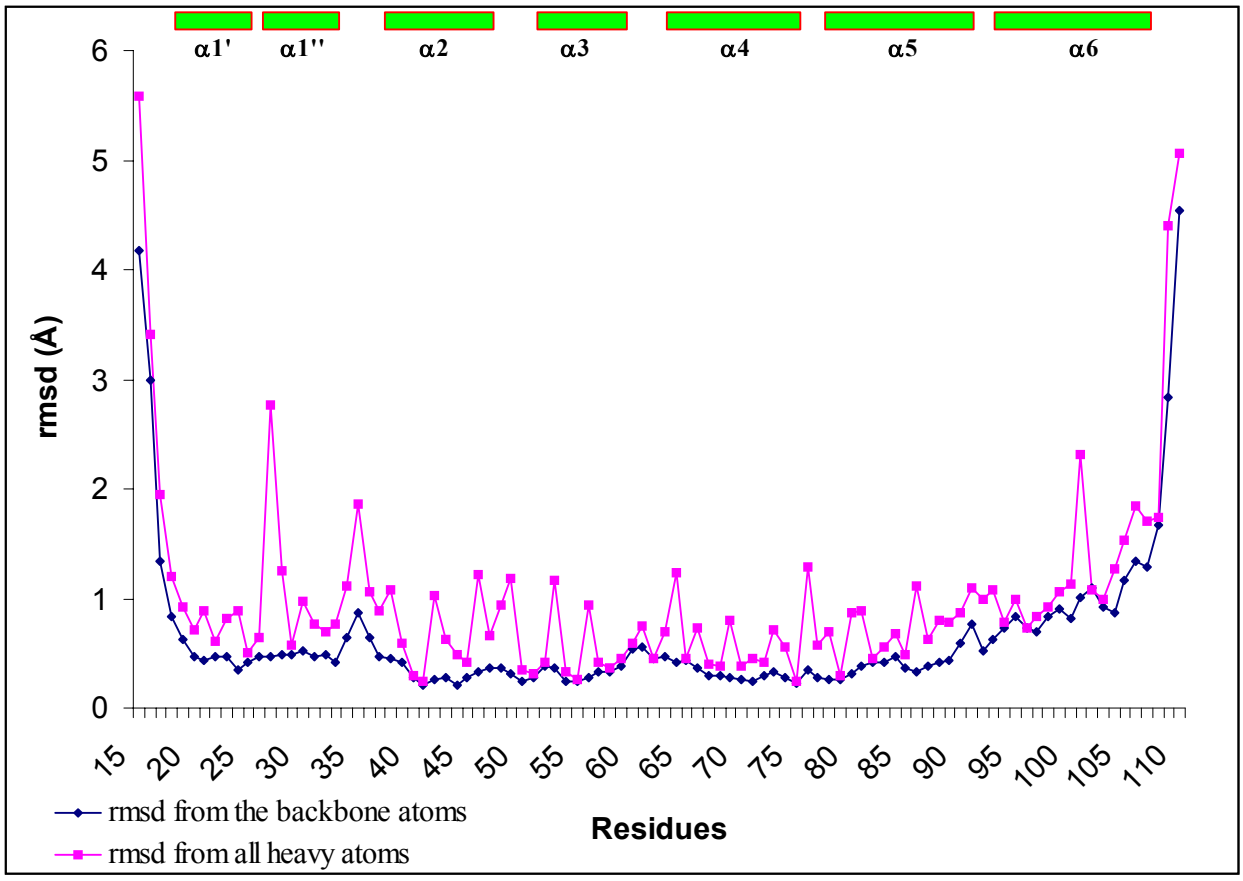

Figure 39. R.m.s.d. calculated for each residue from the ten structures with lowest energy obtained after water refinement.

Residues from the carboxy-terminal helix show a significantly higher r.m.s.d. than residues from others parts of the protein.

The resulting PROCHECK-NMR Ramachandran plot is presented Figure 40. It shows the ϕ-ψ torsion angles for all residues in the ensemble (except those at the chain termini).

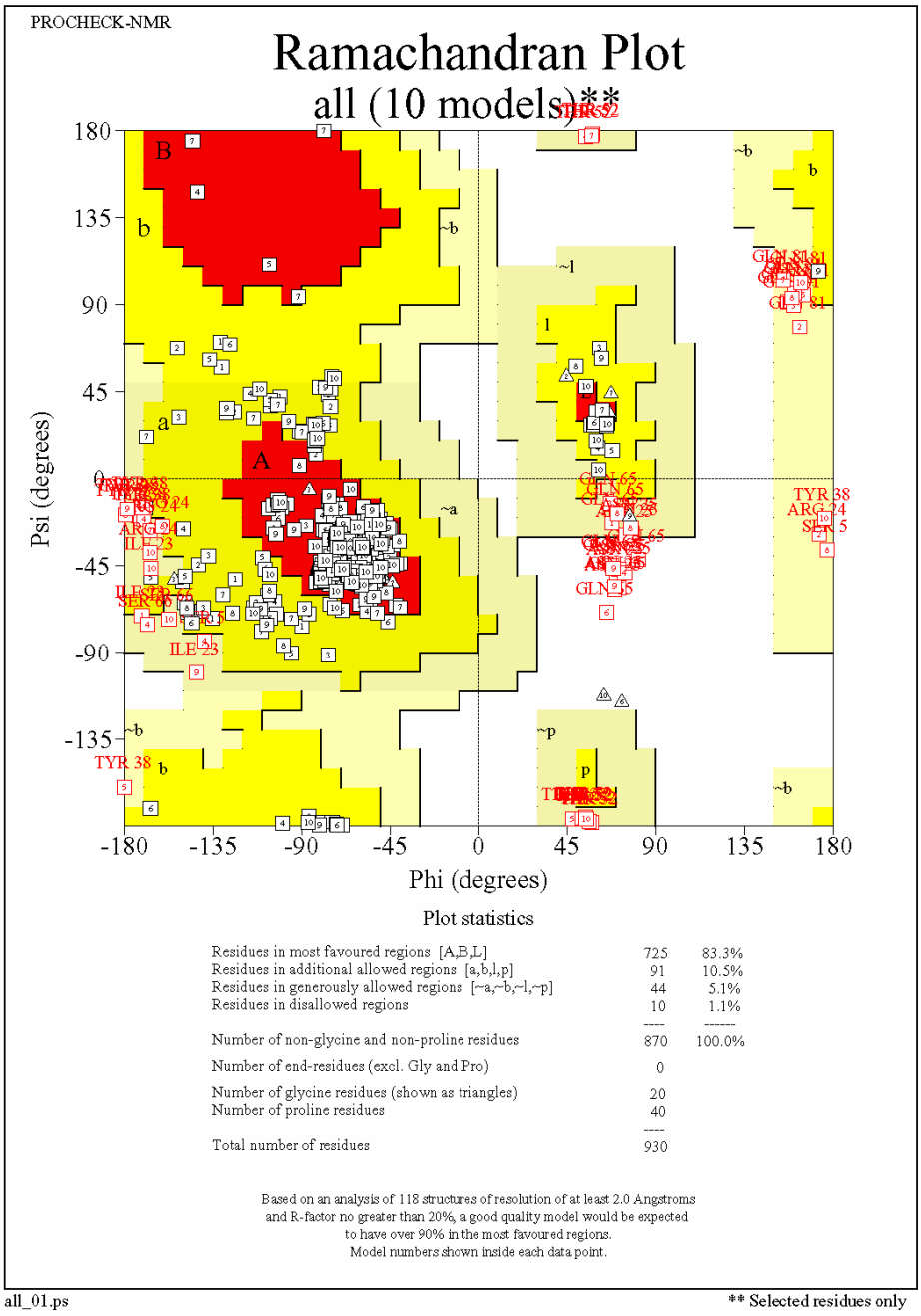

Figure 40. Ramachandran plot of the 10 structures of lowest energy obtained after water refinement. The numbers within each data point correspond to the model number from which that point comes. Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other sidechain types. Red areas represent the most favourable combinations of ϕ-ψ values, yellow areas correspond to other allowed areas, areas colored in light gray describle less favourable regions (generous) regions.

The different regions represented on this plot were described by Morris and colleagues (Morris et al., 1992): A (Core alpha), a (allowed alpha), ~a (generous alpha), B (core beta), b (allowed beta), ~b (generous beta), L (core left-handed alpha), l (allowed left-handed alpha) and ~l (generous left-handed alpha), p (allowed epsilon) and ~p (generous epsilon).

94 % of the residues are found in the two most favoured regions (core and allowed). Furthermore, an AQUA analysis was also performed to validate the quality of the structures obtained. AQUA allows the calculation of the level of completeness of the experimental set of NOEs on the basis of the 3D structures obtained. No NOE restraint was violated more than 0.5 Å (cf. Table 11). The structural validation methods above demonstrate the good quality of the NOD1 CARD structural ensemble.

The complete structural statistics and root-mean-square deviation values are presented in Table 11.

NOE distance restraints	
All	1236
Intraresidue	457
Interresidue	779
Sequential ($\|i - j\| = 1$)	309
Medium range ($2 \leq \|i - j\| \leq 4$)	313
Long range (($\|i - j\| > 4$)	157
Hydrogen bonds (deduced from H-D exchange)	82
Dihedral angle restraints (derived from TALOS)	
ϕ	69
ψ	69
Energies after water refinement	
van der Waals, kcal.mol-1	-895 ± 31
Electrostatic, kcal.mol-1	-4022 ± 47
R.m.s.d from idealized covalent geometry	
Bond lengths, (Å)	0.0045 ± 0.00015
Angles, (°)	0.70 ± 0.02
Improper, (°)	1.67 ± 0.15
R.m.s.d from the mean coordinates (secondary structures)	
Backbone heavy atoms, (Å)	0.43 ± 0.07
All heavy atoms, (Å)	0.85 ± 0.09
Experimental statistics	
Number of distance violations > 0.5 Å	0.8 (± 0.9)
Number of distance violations > 0.3 Å	3.6 (± 1.1)
Number of distance violations > 0.1 Å	47.2 (± 3.0)
Number of hydrogen bond violations > 0.1 Å	3.3 (± 1.2)
Ramachandran plot statistics (PROCHECK)	
Residues in most favorable regions, %	83.3
Residues in additional allowed regions, %	10.5
Residues in generously allowed regions, %	5.1
Residues in disallowed regions, %	1.1

Table 11. Structural statistics for the ten structures of lowest energy.

2.5 Structure

2.5.1 Overview

The NOD1 CARD structure ensemble, composed of the ten lowest energy structures obtained after water refinement, is presented Figure 41.

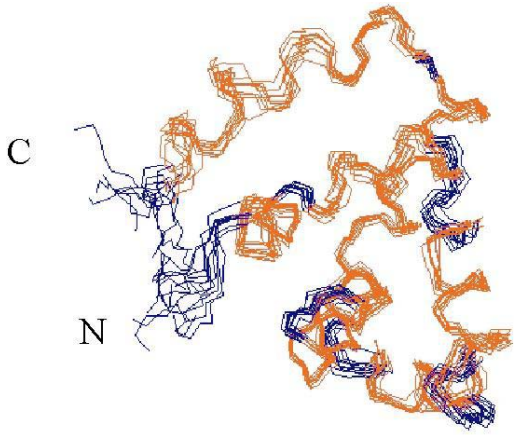

Figure 41. NOD1 CARD structure ensemble. Helices are colored in red. The termini are denoted by the letters N and C.

The NOD1 CARD solution structure consists of six closely packed α-helices (residues 19-33; 38-47; 52-59; 63-75; 78-91; 93-107) (cf. Figures 41 and 42). The boundaries of these secondary structures are very close to the predictions obtained with the CSI. Not surprisingly, the carboxy-terminal helix (helix 6) is less defined than the five others; as seen previously in 2.4, the helix 6 residues show higher r.m.s.d.

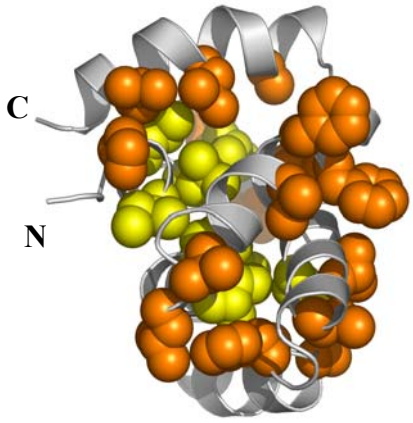

Figure 42. Hydrophobic core of NOD1 CARD. The polypeptide chain of the mean structure is show as a gray ribbon and the hydrophobic residues that make up the core of the protein are shown as spheres. Carbon atoms of the central core, including conserved hydrophobic residues, are colored in yellow. Carbon atoms of hydrophobic residues that form an extensive hydrophobic layer are colored in orange. The termini are denoted by the letters N and C.

The hydrophobic core consists of residues that are highly conserved among members of the CARD family known for interacting with RICK and designed with a star (*) in the alignment below, Figure 43.

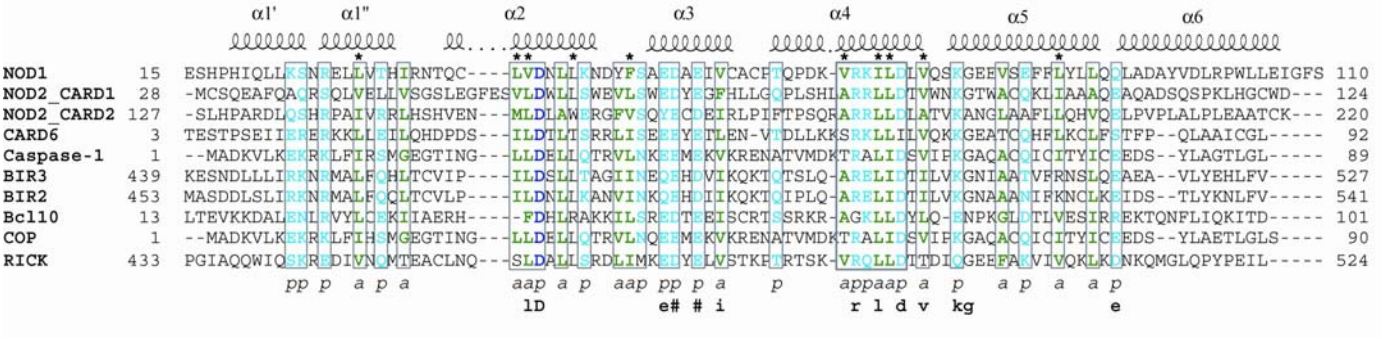

Figure 43. Sequence alignment of RICK CARD with NOD1 CARD and 8 other CARDs shown to interact with it. Common features are identified by color and labeling below the sequences. Polar residues are in cyan and labeled "p", hydrophobic residues are in green and labeled "a". The bottom line indicates the consensus sequence (one-letter code of the amino acid, bold): uppercase is identity, lowercase is consensus level > 0.5, # refers to acidic residues Glu or Asp. Conserved hydrophobic residues that make up the hydrophobic core of this RICK CARD interaction family . The location of the α-helices in the structure of NOD1 CARD is shown above the alignment.

Helices 2-5 form an antiparallel four-helix bundle with helix 6 and helix 1 positioned on the top of helix 5 and helix 4 respectively and crossing each other. Helix 1 is interrupted by a severe kink of 91° between residue Q26 and the conserved R27. It is consequently divided into two smaller helices, helix 1' (N-terminal to the kink) and helix 1" (C-terminal to the kink) with helix 1' parallel to helix 4 and helix 1" parallel to the N-terminal part of helix 6. Helix 6 is also slightly curved at position 100. This is most likely due to the fact that residues 100 and 101 are invisible in the NMR spectra, that residue 102 is a proline, and because only a few long distance NOE restraints define the orientation of its C-terminal part. Helices 3, 5 and 6 cap the exposed hydrophobic core formed by helices 1, 2 and 4.

2.5.2 Comparison with other CARDs

The most similar known structures to NOD1 CARD found using the DALI server (Holm and Sander, 1993) are those of Iceberg (Humke et al., 2000) (Z-score = 6), procaspase-9 (Qin et al., 1999) (Z-score = 5.2) and Apaf-1 (Vaughn et al., 1999) (Z-score = 4.6) with the best-fit superpositions of Cα atoms giving R.M.S.Ds of 3 Å for 81 aligned Cα, 3.5 Å for 83 aligned Cα, and 3.6 Å for 81 aligned Cα respectively. These structural similarities are however lower than that observed between CARD domains generally (e.g. procaspase-9/Apaf-1 Z-score = 14.9, procaspase-9/iceberg Z-score = 10.9, Apaf-1/iceberg Z-score = 11.1). Thus

while the overall topology is similar, NOD1 CARD differs notably in the conformation of some of the inter-helical loops and consequently the length and orientation of some of the helices (cf. Figure 44).

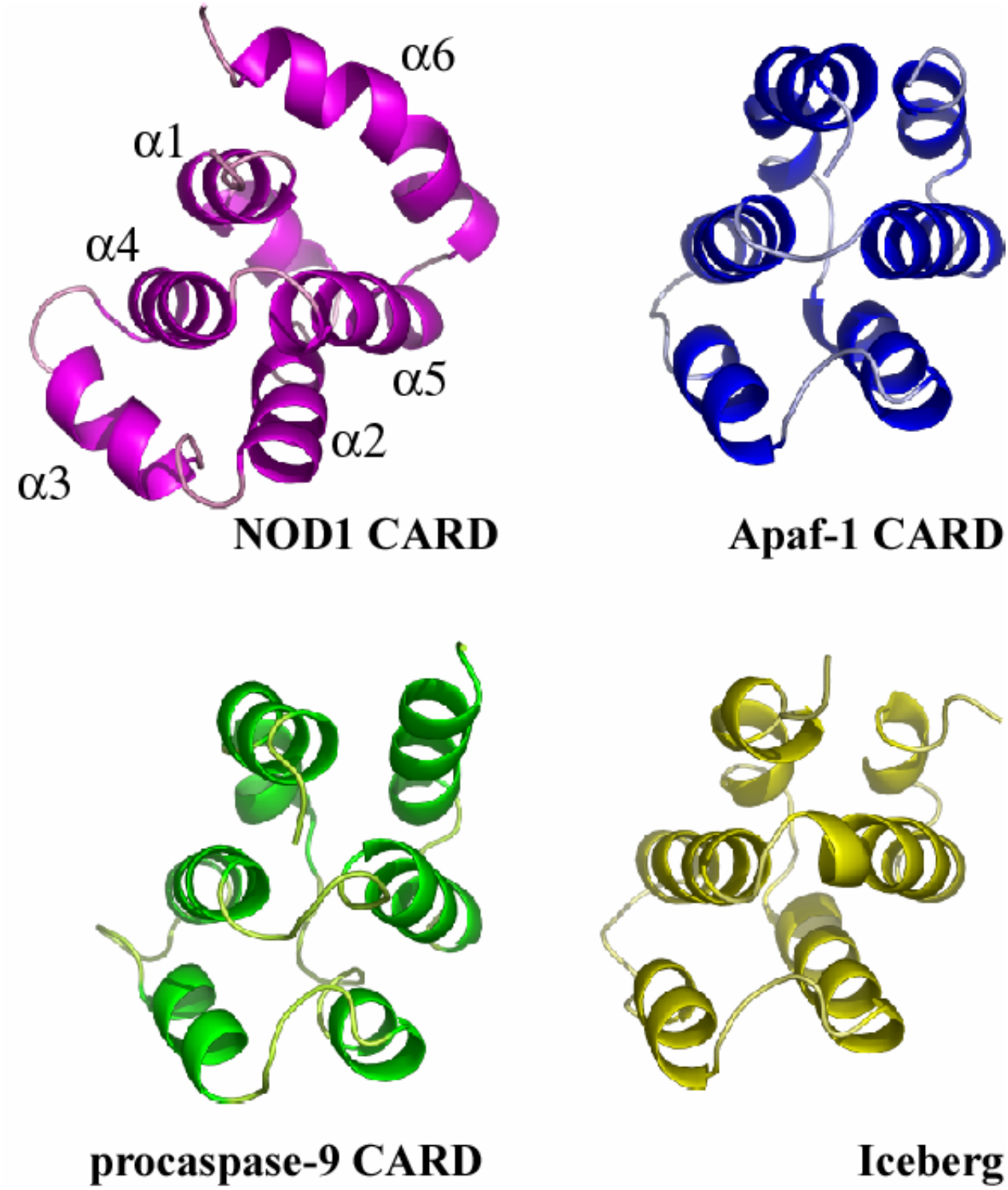

Figure 44. Structure of NOD1 CARD and other homologous CARD structures. The four structures have the same orientation. Helices 1 to 6 are numbered α1-α6 on NOD1 CARD structure.

The main difference concerns helix 6 which begins earlier and is longer in NOD1 CARD than in the other CARDs considered. This difference could be observed already from the backbone assignment step analysing the chemical shift index of the Cα, Cβ and CO resonances (cf. 2.3). Most significantly the final direction of the extended helix 6 differs considerably, as shown in Figure 45.

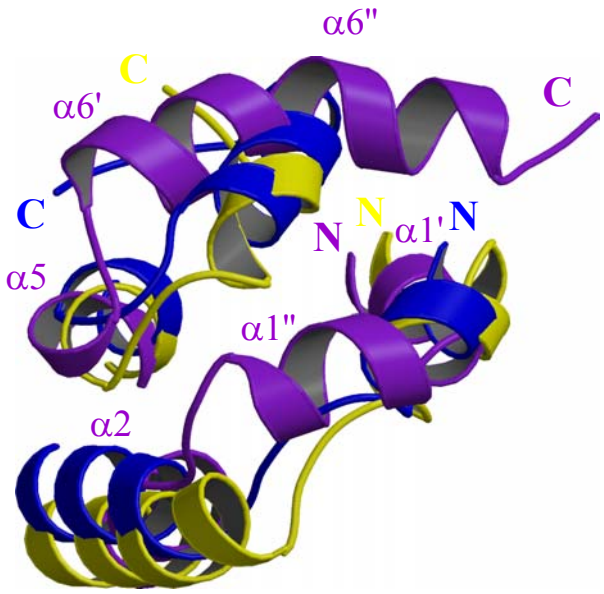

Figure 45. Detail of the structural alignment of the N- and C-terminal regions of NOD1, Iceberg and Apaf-1 CARDs. NOD1 is colored in magenta, Iceberg in gold and Apaf-1 in blue. This structural alignment highlights the major deviations of NOD1 CARD from more prototypical CARDs. Note the additional helix α1" which is antiparallel to the first part of helix 6 (α6') and the extension of the helix 6 (α6") taking the C-terminus of the domain in the opposite direction.

Compared to other CARD domains, the connections between helices 1 and 2 is altered (due to the presence of helix 1") and also that between helices 5 and 6. The C-terminal extension of helix 6 after the kink around P102 leads it to cross over helix 1' and point in the opposite direction to the C-terminus of other CARD domains. However it should be borne in mind that helix 6 is more flexible than the five others as described below by ^{15}N relaxation analysis.

2.5.3 Surfaces

The NOD1 CARD surface is composed of opposing charged and hydrophobic faces (cf. Figure 46).

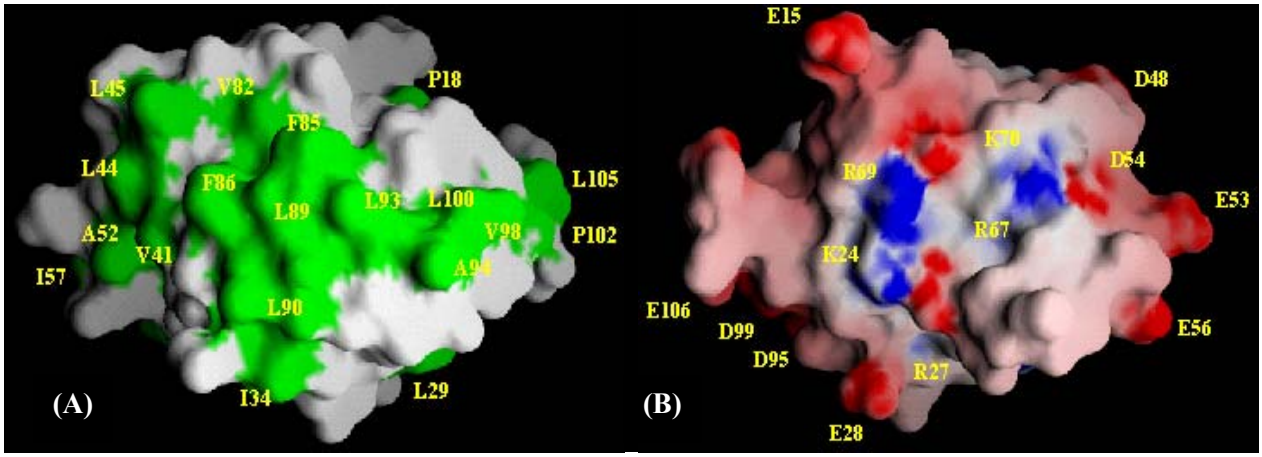

Figure 46. Surface representations of NOD1 CARD. (A) Hydrophobic surface of NOD1 CARD; surface exposed hydrophobic residues including Leu,Val, Ile, Trp and Phe are colored in green, (B) Charged surface of NOD1 CARD; the surfaces are color coded according to electrostatic surface potential: red, -10 kT; white, 0kT; and blue, +10 kT.

The hydrophobic face is mainly composed of hydrophobic residues from helices 5 and 6. The charged face comprises three distinct patches, one basic and two acidic (cf. Figure 46(B)) and one basic.

The basic one is composed of K24, K67, R69 and K70 belonging to helices 1 and 4 (cf. Figure 46(B)). Two acidic patches, on its left and on its right, surround it. The left patch (cf. Figure 47(A)) consists mainly of acidic residues from helices 5 and 6, in particular Glu80, Glu84, Asp95, Asp99 and Glu106. The right and more prominent one (cf. Figure 47(B)) is composed of acidic residues from helices 2 and 3: D42, D48, E53, D54 and E56. These surface characteristics are predicted to be important for the interactions between the NOD1 and RICK CARDs.

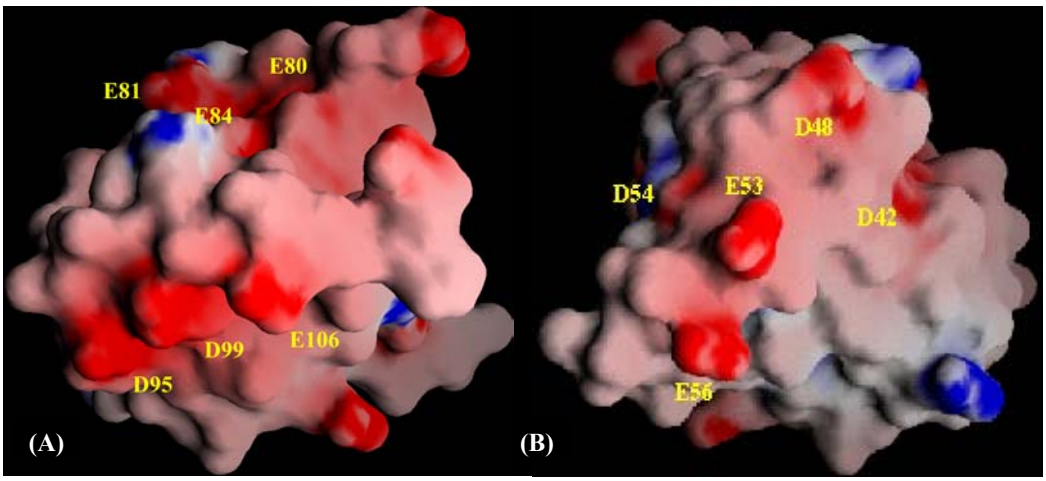

Figure 47. Acidic surfaces of NOD1 CARD. (A) "Left" acidic patch; (B) "right" acidic patch. The surfaces are color coded according to electrostatic surface potential: red, -10 kT; white, 0kT; and blue, +10 kT.

2.6 Dynamic studies

^{15}N spin relaxation data contains information on the time-dependent fluctuations of individual backbone N-H bond vectors. We used the Lipari-Szabo model-free approach (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b) in combination with the description of rotational diffusion anisotropy formulated by Woessner (Woessner, 1962) to analyze the measured relaxation data of the CARD domain of NOD1. Residues with overlapping ^{1}H-^{15}N correlation peaks or weak cross peak intensities due to chemical exchange were removed from the analysis.

The R_2/R_1 value analysis leads to a fully anisotropic diffusion tensor with an overall correlation time for the molecule (τc) of 7.6 ± 0.1 ns, an anisotropy of 1.24 ± 0.03 and a rhombicity of 1.4 ± 0.25. This anisotropy is rather large for the relatively small size of the protein, but this diffusion property is likely caused by the unstructured C-terminal tail of the protein.

The generalized order parameters, S^2, obtained for the 61 remaining residues are displayed in Figure 48. The colour scheme reflects increasing order parameters in the picosecond to nanosecond time scale. Residues located in the helices exhibit mostly small amplitude motions except for L44, V82 and W103. The rest of the more dynamic NH vectors are located in loops or at extremity of helices.

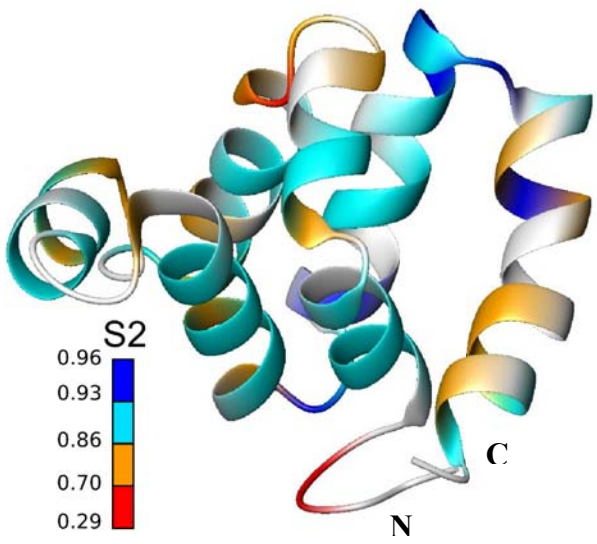

Figure 48. Ribbon representation of the backbone of NOD1 CARD color-coded according to the order parameter values (S^2) from red (more flexible) to blue (more rigid). The termini are denoted by the letters N and C and colored

Residues exhibiting chemical or conformational exchange in the microsecond to millisecond time scale range are colour coded in red Figure 49. R27 and H33 located at the extremities of helix 1', V75 located at the C-terminal of helix 4, and the majority of the residues of the last

helix are involved in such slow events. However, the exchange behaviour of helix 6 may be due to flexible C-terminal tail and suggests that residues L100 and R101 could probably not be assigned because of extensive line broadening in this region.

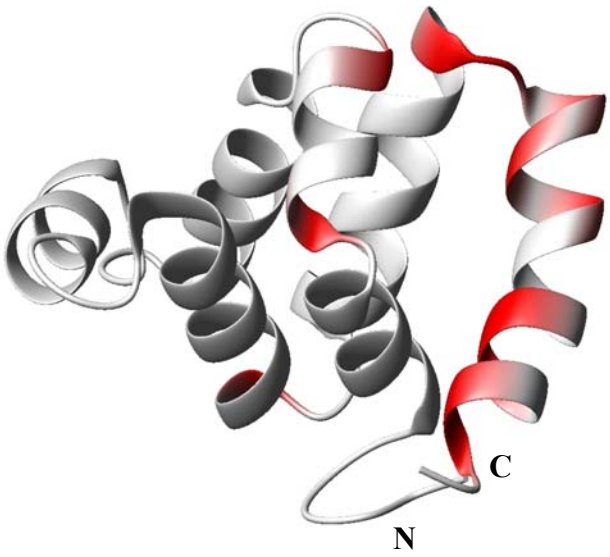

Figure 49. Ribbon representation of the backbone of the NOD1 CARD color coded in red when the exchange rate parameter (Kex in s^{-1}) is present. The termini are denoted by the letters N and C.

3 NOD1/RICK CARD-CARD interaction characterization

3.1 Introduction to RICK CARD

Human RICK is a 540 amino acids protein having a predicted molecular mass of 61 kDA and a pI of 6.63 (Expasy ProtParam). RICK is composed of an amino-terminal kinase domain (residues 18-294) and a carboxy-terminal CARD domain (residues 432-524) (cf. Figure 50).

Figure 50. Domain structure of human RICK.

The first RICK CARD construct was designed using the sequence alignment presented in Figure 29, and encompassing residues 432-524. The resulting protein demonstrated a high level of expression in bacteria but was difficult to work with due to its poor solubity. This problem could not be resolved by co-expression with NOD1 CARD.
Two longer constructs comprising residues 432-533 and 432-540 were designed based on the NOD1 CARD secondary structure. Both have a lower expression level in bacteria and are mainly found as soluble aggregates. The longest construct (432-540) was used for NMR interaction studies with NOD1 CARD. Its predicted molecular weight was 14.2 kDa and its theoretical pI 8.15. This construct is hereafter referred as RICK CARD.

3.2 Expression and purification of RICK CARD

The protocol used for the expression and purification of RICK CARD is explained in details in 2.1, 2.3 of the materiel and method part. The purification is just illustrated here, notably with the corresponding SDS-PAGE 15 % gels.

3.2.1 Ni-NTA affinity column

The nickel column allowed purifying RICK CARD thanks to the affinity of the His-tag for the nickel (cf. Figure 51).

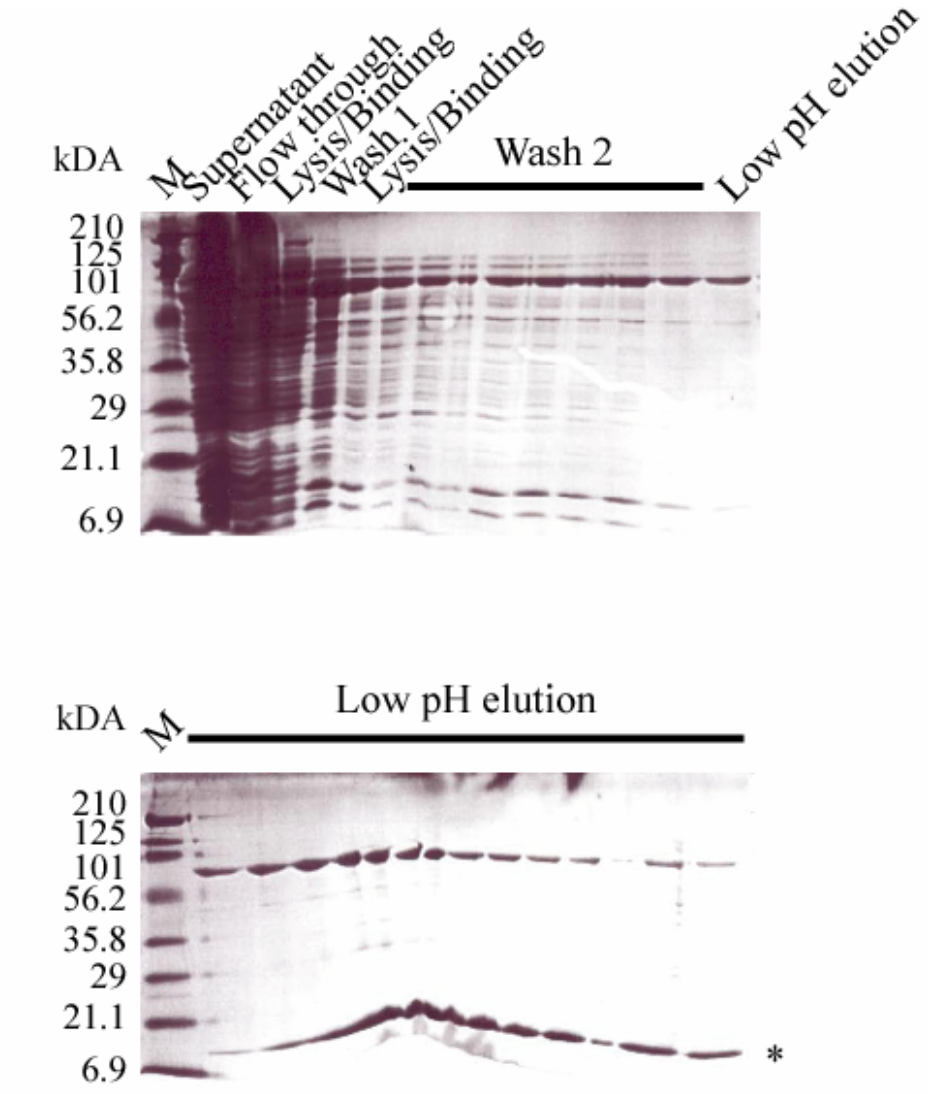

Figure 51. Purification of RICK CARD, nickel column. 15 % SDS-PAGE gels; M =molecular weight markers.

The first washing step with RICK wash buffer 1, which contains 50 mM of imidazole, allowed eliminating many of the contaminants attached non-specifically on the resin. However, a contaminant of approximatively 60 kDa remained. This contaminant may be the GroEL chaperone (molecular weight of 57 kDa). The second washing step with RICK wash buffer 2 (pH 5.5) partially eliminated this contaminant. However, a few quantity of RICK was also lost. The final elution with the low pH buffer (pH 4) allowed obtaining soluble non-aggregated RICK CARD, but still with its contaminant. Size-exclusion column constituted the

next step of purification, since the sizes of RICK CARD and of its contaminant are very different.

3.2.2 Size-exclusion

After concentration, RICK CARD was run on a superdex 75 column (cf. Figure 52).

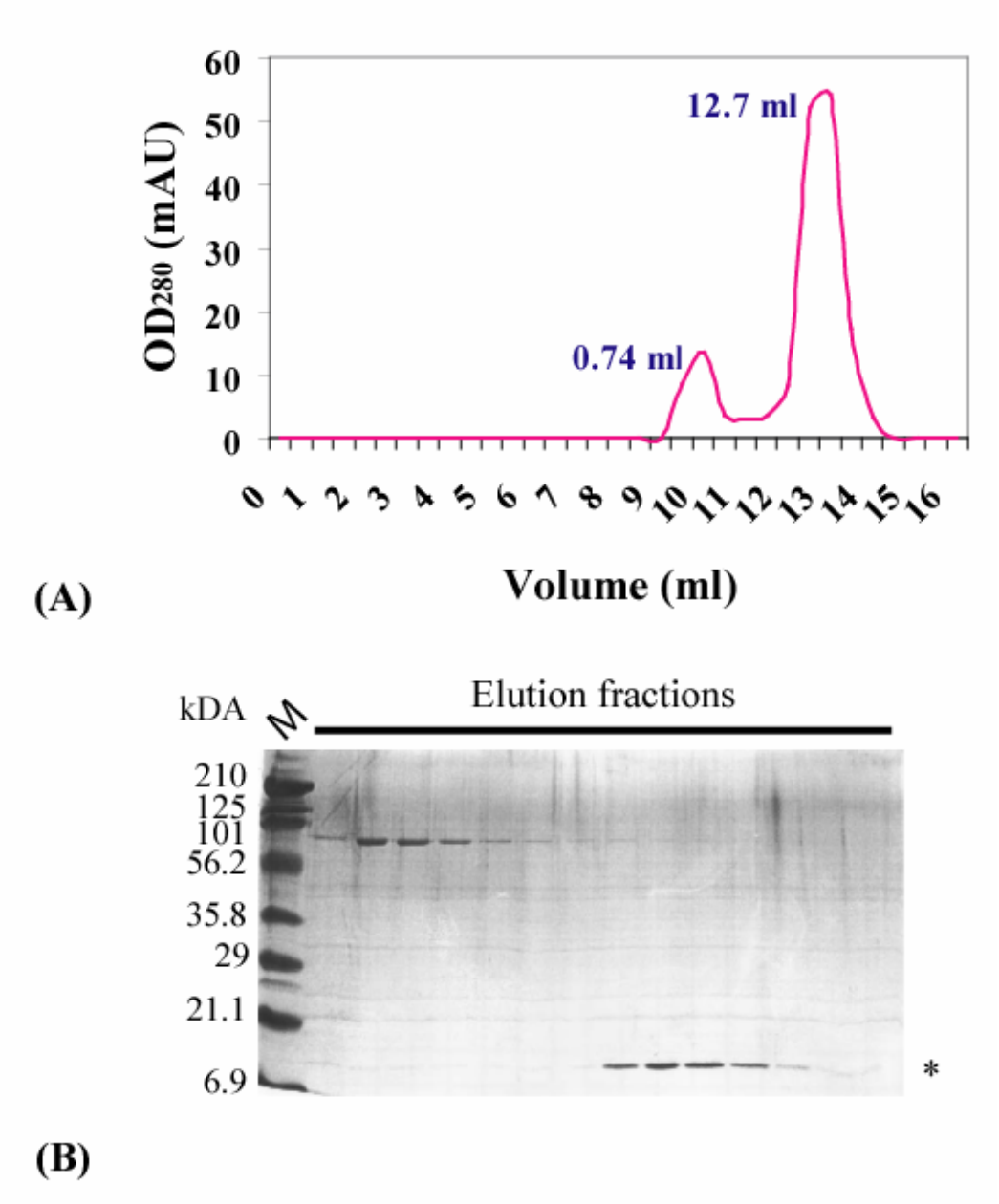

Figure 52. Purification of RICK CARD, size-exclusion column. (A) Superdex 75 gel-filtration profile of RICK CARD with elution volume of 12.7 ml, (B) 15 % SDS-PAGE gel; M =molecular weight markers.

A single peak was obtained for an elution volume of 12.7 ml. The fractions containing RICK CARD were then pooled and concentrated. The concentration was stopped when RICK

CARD began to precipitate. The protein obtained at this step is very pure (cf. Figure 53). The final yield of pure and soluble RICK CARD obtained was very low: 100 µg / L of LB.

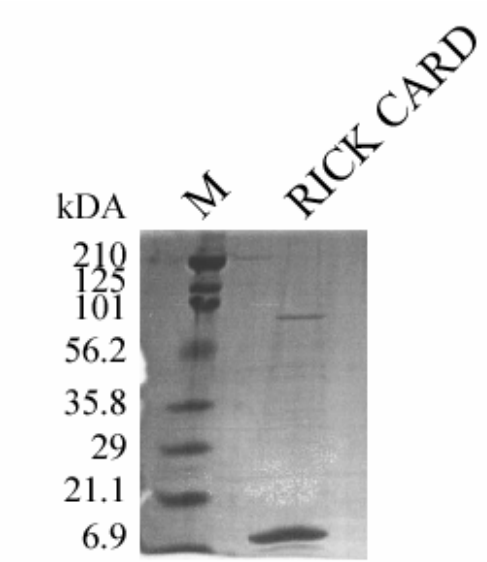

Figure 53. RICK CARD concentration.

The protein aggregating during concentration, it was impossible to obtain a higher concentration than 0.2-0.3 mM.

3.3 Evidence for NOD1/RICK CARD-CARD interaction

Two independent groups demonstrated that NOD1 activates NF-κB *via* its downstream effector, the serine-threonine kinase RICK (Bertin et al., 1999; Inohara et al., 1999). Inohara and colleagues performed co-immunoprecipitation experiments of NOD1 with multiple regulators of apotosis and NF-κB activation. They showed that full length NOD1 was able to interact with several caspases containing long prodomains (e.g. caspases 9) and with the full length RICK protein. To determine which parts of NOD1 and RICK were involved in the binding, Inohara and colleagues tested various construct lengths of NOD1 and RICK. They were able to show that the full length NOD1 protein could bind RICK (374-540), contains the CARD domain, but not RICK (1-374). Therefore they concluded that the CARD domain of RICK is necessary for the interaction and independent of its kinase activity. They also performed co-immunoprecipitation experiments with full length RICK and various NOD1 constructs. A NOD1 construct lacking the LRRs (1-648) was able to bind full length RICK and constructs containing the NBD plus LRRs (126-953) or only the LRRs (649-953) alone did not interact with full length RICK. These results indicate that the NOD1/RICK interaction

is not mediated by the LRRs. They concluded that the NOD1/RICK interaction is mediated through their CARDs because NOD1 (1-648) contains the CARD, which are well known to interact through homotypic CARD-CARD interactions,

Bertin and colleagues used a different method based on the genomic database mining approach for searching homologs of Apaf-1 and its nematode homolog CED-4. They used an amino-terminal construct of NOD1 (1-145) encompassing the CARD domain as bait in a yeast two-hybrid screen of a human breast cDNA library to identify CARD-containing interactors. These studies identified RICK as one of the interactors. They also concluded that the CARD of NOD1 interacts with the CARD of RICK based on two experiments: (1) they showed that the CARD domain of RICK alone (435-540) is sufficient to bind to the bait; (2) they showed by co-immunoprecipitation experiments in mammalian cells that full length NOD1 is unable to bind a construct of RICK lacking this CARD (1-435).

Thus I decided to try and determine the structural basis of this interaction using NMR chemical shift mapping.

3.4 1D NMR study

In order to verify the correct folding of RICK CARD, a 1D NMR spectrum was recorded (cf. Figure 54). The experiment was performed at 25°C with a 0.12 mM sample of unlabeled RICK CARD.

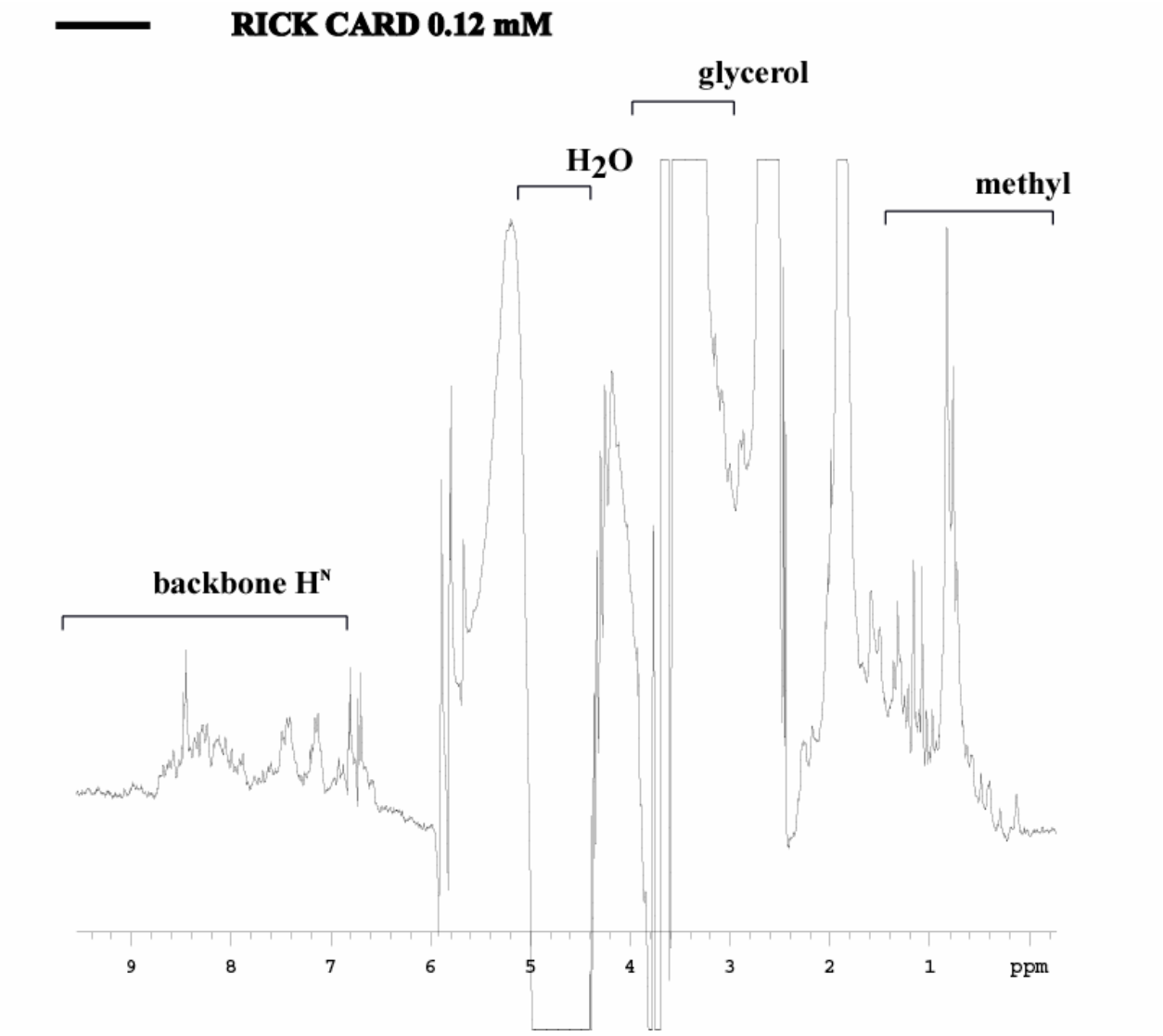

Figure 54. 1D NMR spectra of unlabeled RICK CARD at 0.12 mM. Chemical shift ranges observed for the several types of protons are indicated on the spectra.

The RICK CARD signal appears weak and contains peaks corresponding to elements in the buffer or even to the trace amounts of glycerol coming from the Amicon Ultra centrifugal filter device (Millipore).

However, the protein is structured, supported by the good resolution of the backbone amide protons and the presence of five peaks corresponding to methyl protons that are “outside” the main signal.

3.5 NMR chemical-shift mapping experiments

NMR chemical-shift mapping experiments were attempted (cf. Materials & Methods 4.10), despite the fact that RICK CARD is not ideally suitable (mainly because it was impossible to concentrate the protein higher than 0.3 mM).

RICK CARD and NOD1 CARD were mixed together in the desired ratios. The buffer was then changed/desalted to the appropriate concentration by the method of diafiltration.

The spectrum was recorded at 30°C for 0.15 mM of ^{15}N-labeled NOD1 CARD and 0.3 mM of unlabeled RICK CARD in 25 mM citrate, pH=5.6 and 16 mM NaCl, (cf. Figure 55).

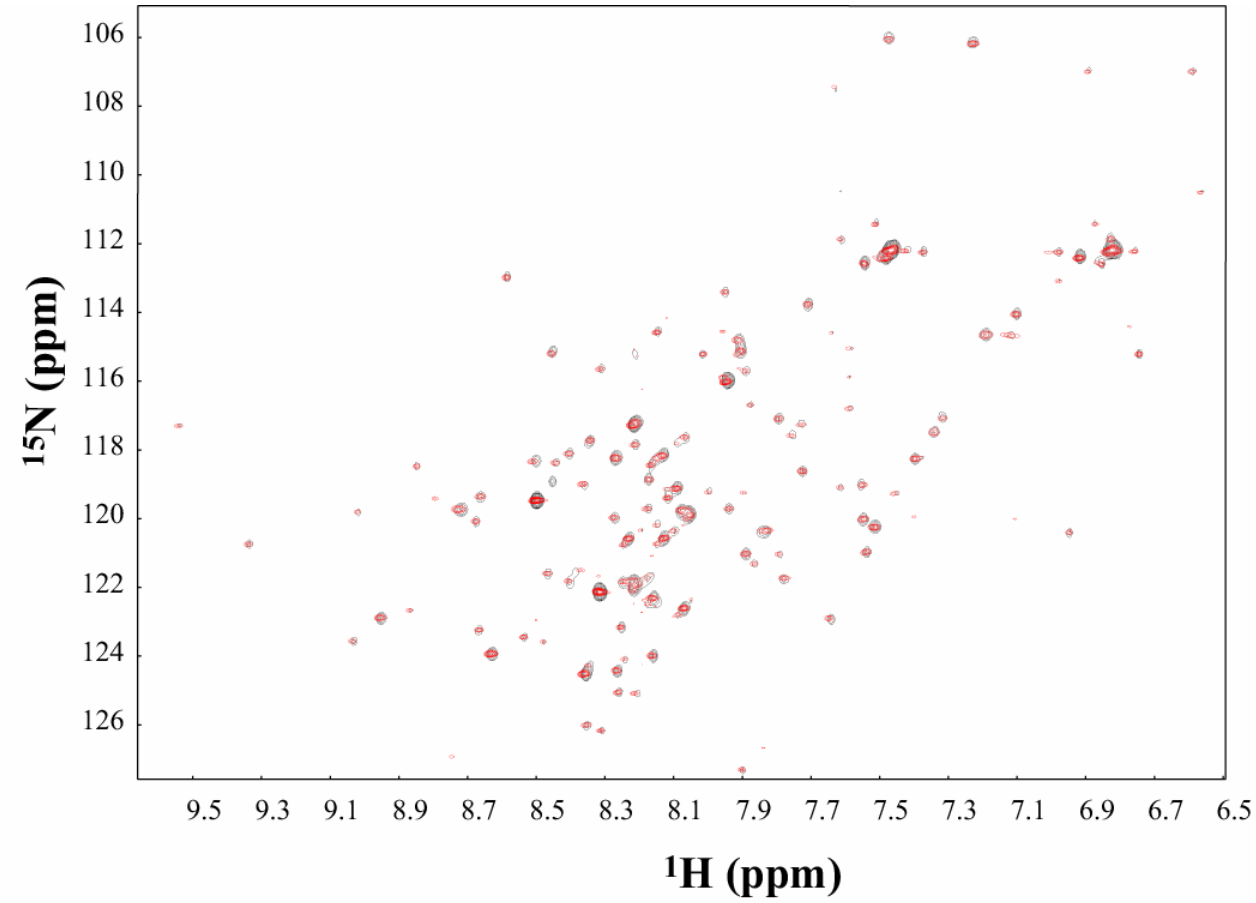

Figure 55. NMR chemical-shift mapping experiments. In black: reference spectrum with ^{15}N-labeled NOD1 CARD alone. In red, spectrum obtained with 0.15 mM of NOD1 CARD and 0.3 mM of unlabeled RICK CARD.

No resonances of NOD1 CARD changed when RICK CARD was added. This means that the electronic environment of the observed nuclei did not vary. However, I certainly lost most of the RICK CARD during the concentration step necessary for mixing the proteins and changing the buffer. Native gels were also attempted but the concentration of RICK CARD was so low that the protein was not even visible. Since it was impossible to characterize the NOD1/RICK CARD-CARD interaction *in vitro* due to the difficulties above, I decided to

construct a crude docking model in order to have a better idea about how this interaction could function.

3.6 RICK CARD model

The amino acid sequence conservation in the CARD family is generally low, although the known structures present a high degree of structural homology. The procaspase-9 CARD was used as a template for RICK CARD homology modeling because it has the highest sequence identity with RICK CARD (about 20 %) among the CARD domains of known structure. Moreover, this CARD domain is known to interact with NOD1 CARD (Inohara et al., 1999), suggesting that the procaspase-9 CARD and RICK CARD interactions with NOD1 might share some common features.

The modeled RICK CARD is composed of the six tightly packed anti-parallel α-helices (cf. Figure 56), similarly to the CARD of procaspase-9.

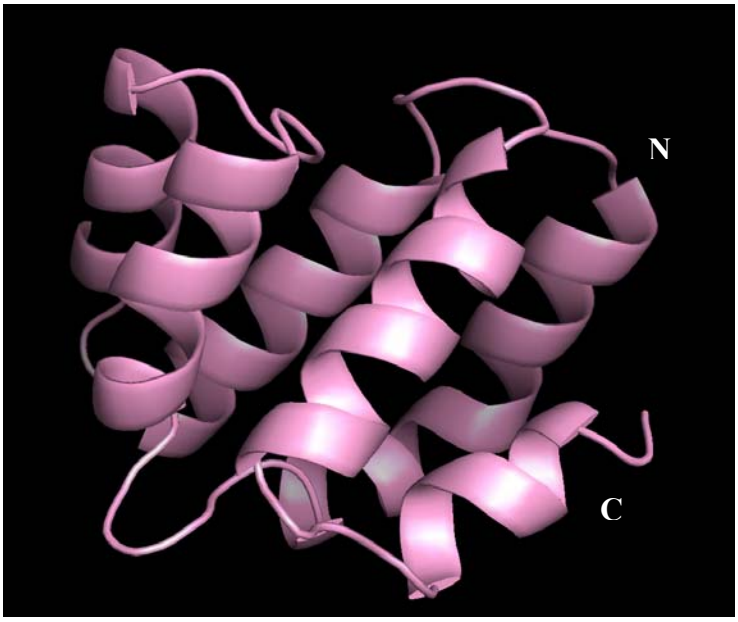

Figure 56. 3D model of RICK CARD. The termini are denoted by the letters N and C.

The modeled RICK CARD has distinct basic and acidic charged surface patches. The acidic patch of the modeled RICK CARD (cf. Figure 57(A)) primarily contains residues from helices 3 and 5. The most extensive basic surface of RICK (cf. Figure 57(B)) comprises K443 and R444 from the turn between helices 1 and 1', K480 from the turn between helices 3 and 4, R483 and R488 from helix 4. It is opposite a rather large hydrophobic patch composed mainly of residues from helices 5 and 6 (cf. Figure 57(C)).

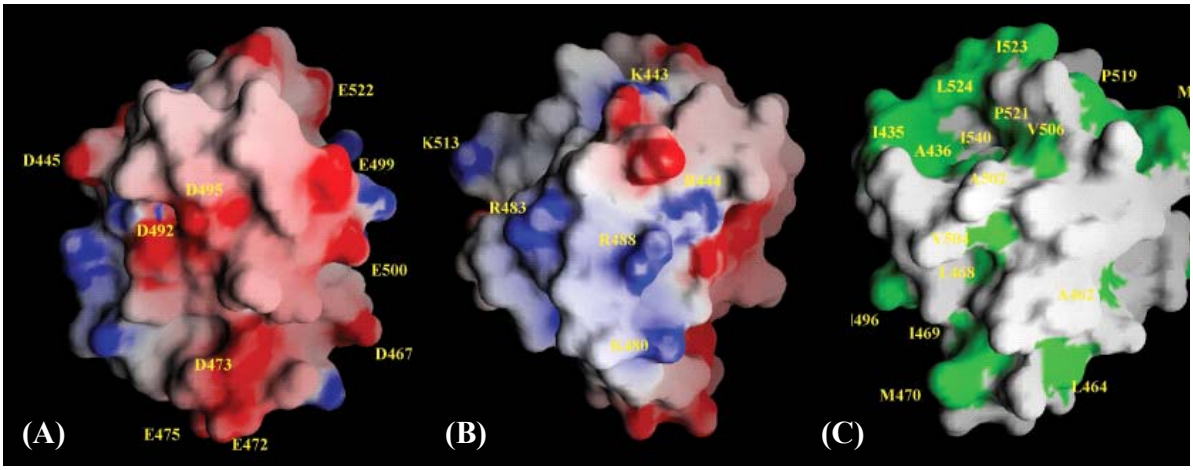

Figure 57. Surface representations of RICK CARD model. In (A) and (B), the surfaces are color coded according to electrostatic surface potential: red, -10 kT; white, 0kT; and blue, +10 kT. (A) Acidic surface of RICK CARD formed primarily by the exposed residues of helices 3 and 5. (B) Basic surface of RICK CARD formed by the exposed residues of helices 1 and 4. (C) Hydrophobic surface of RICK CARD formed primarily by residues of helices 5 and 6; surface exposed hydrophobic residues including Leu,Val, Ile, Trp and Phe are colored in green.

3.7 NOD1 CARD/RICK CARD crude interaction model

Studies of CARD-CARD binding in the Apaf1/caspase-9 (Qin et al., 1999) and RAIDD/caspase-2 (Chou et al., 1998) systems led to the conclusion that the interactions are primarily electrostatic, favouring the initial docking of the two proteins. In the Apaf-1/procaspase-9 CARD-CARD crystal structure, the acidic patch of Apaf-1 interacts with the basic patch of procaspase-9 (Qin et al., 1999; Zhou et al., 1999). Like Apaf-1, NOD1 is able to bind procaspase-9. When over-expressed, NOD1 promotes caspase-9 induced apoptosis through CARD-CARD interactions (Inohara et al., 1999). Both the CARD and the NBD are essential for NOD1 to activate procaspase-9 (Inohara et al., 1999), similarly to Apaf-1 (Hu et al., 1998; Srinivasula et al., 1998). This suggests that NOD1 may interact with procaspase-9 by a mechanism similar to Apaf-1. Thus I hypothesised that in the NOD1/procaspase-9 and in the NOD1/RICK interactions, it is the basic patch of procaspase-9 or RICK that interacts with the acidic patch of NOD1. Moreover, the NOD1 CARD acidic patch contains a number of hydrophobic residues that are potentially able to interact with those in the basic patch of RICK (cf. Figure 58(A) and (B)), as observed in the Apaf-1/procaspase-9 CARD-CARD complex (Qin et al., 1999).

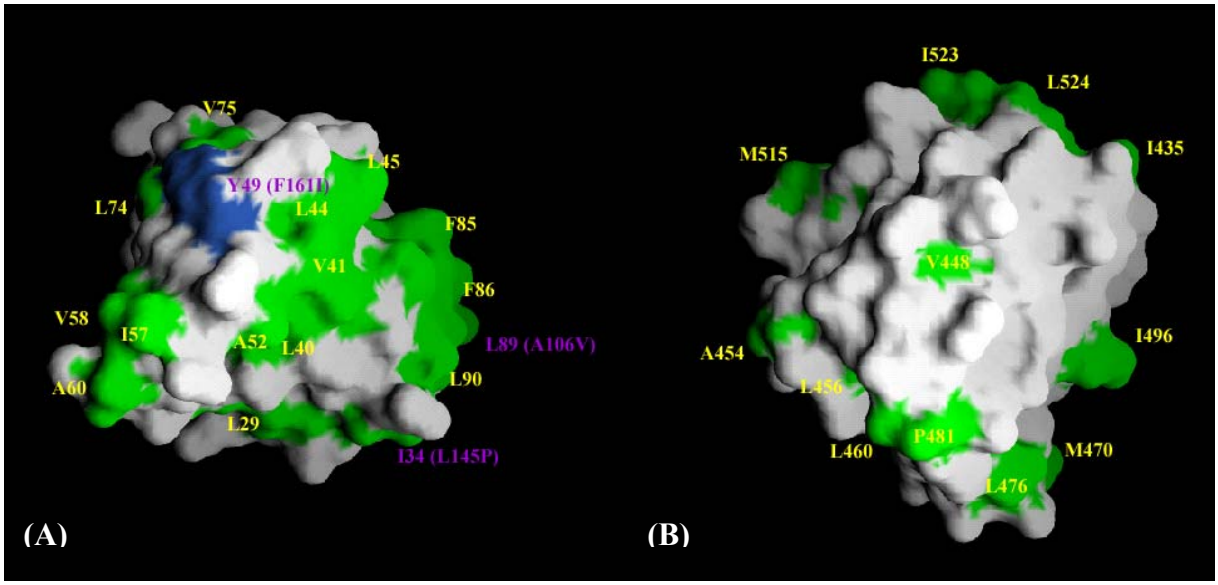

Figure 58. Surface representations of NOD1 CARD and RICK CARD model. In (A) and (B), surface exposed hydrophobic residues including Leu,Val, Ile, Trp and Phe are colored in green. (A) Acidic surface of NOD1 CARD; residues identified in magenta correspond to mutations of hydrophobic residues in NOD2 CARDs that reduce or abolish the interaction with RICK CARD (Tanabe et al., 2004). Residues are colored in green when they correspond to hydrophobic side chains and cyan when they correspond to non-hydrophobic amino acids in NOD1 CARD structure. (B) Basic surface or RICK CARD.

Published mutational analyses on the CARDs of NOD2, a NOD-1 related protein that interacts with RICK and implicated in the same signaling pathway, reinforce this hypothesis (Tanabe et al., 2004). The NOD2 loss of function mutant E69K corresponds to residue E53 in NOD1 CARD and the A106V, L145P and F161I mutations in the NOD2 CARDs reduce or even abolish the interaction with RICK (Tanabe et al., 2004). The corresponding residues in the NOD1 CARD structure are all located in the acidic patch and are either conserved or juxtaposed to conserved hydrophobic residues (cf. Figure 58(A)).

I constructed a crude docking model for the NOD1 and RICK CARDs interaction based on the known structure of the Apaf-1/procaspase-9 CARD-CARD complex (Qin et al., 1999) in order to identify residues potentially involved in the interaction. I superimposed NOD1 CARD on Apaf-1 CARD and my RICK CARD model on procaspase-9 CARD structures respectively (cf. Figure 59).

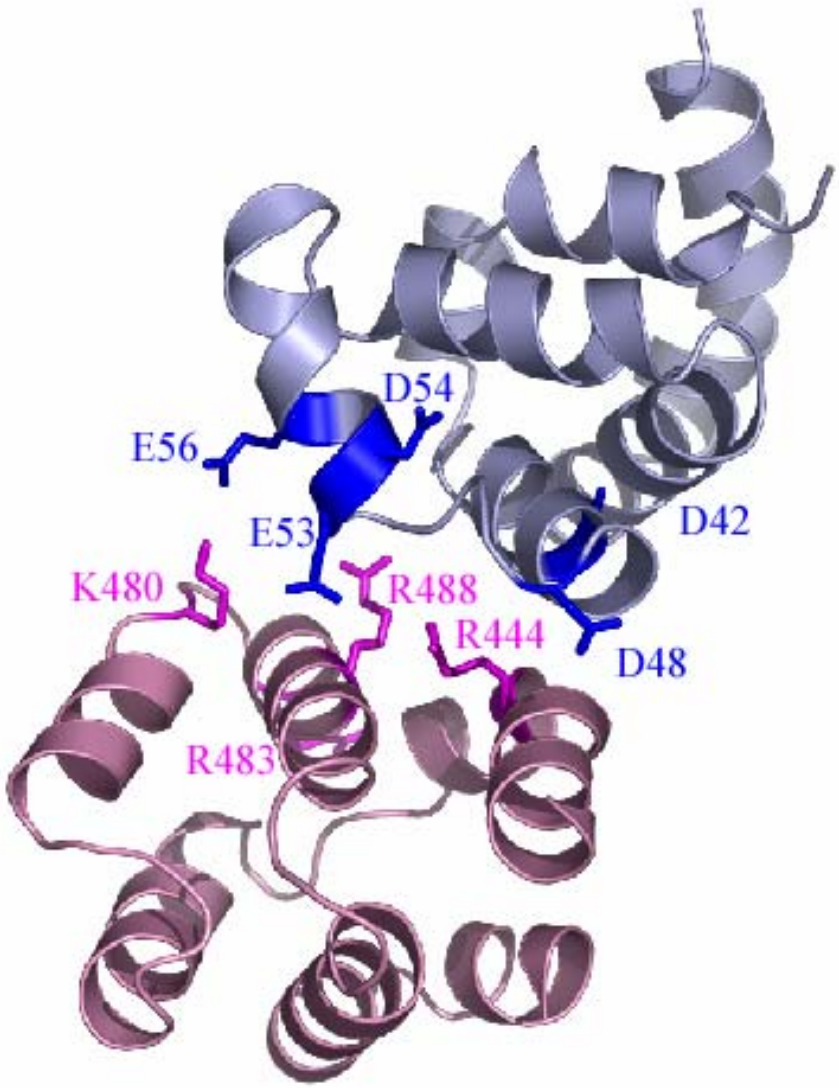

Figure 59. NOD1 CARD/RICK CARD interaction model. NOD1 CARD and RICK CARD are colored lightblue and lightpink, respectively. Side chains of the important interface charged residues are highlighted in blue and magenta, for NOD1 CARD and RICK CARD respectively.

In this model, the acidic α-helices 2 and 3 of NOD1 CARD form an antiparallel four-helix bundle with the basic α-helices 1 and 4 of RICK CARD.

Five acidic residues of NOD1 CARD (D42, D48, E53, D54 and E56) are found at the putative interaction surface. Of these, E53, D54 and E56 are surface exposed. D42 and D54 make intra-molecular interactions and are not in an optimal configuration for interacting with RICK CARD. However in our structure of NOD1 CARD, the side-chains of D42, E53 and E56 exhibit a significantly higher r.m.s.d. than all other side-chains of residues belonging to helices 2 and 3, suggesting that they could flexibly adapt during complex formation. D42 and E56 correspond to D27 and E41 from the acidic patch of Apaf-1. Furthermore, D27 and the neighboring E40 are necessary for Apaf-1 binding to procaspase-9 (Qin et al., 1999). Four basic residues of RICK CARD (R444, K480, R483 and R488) are found at the putative

interaction surface. Of these, R444 and R488 are conserved basic residues in the CARD family corresponding to R13 and R56 in procaspase-9 CARD, two residues known to be indispensable for Apaf-1 binding (Qin et al., 1999). This model allows us to identify important residues to mutate for further experiments.

3.8 In vivo mutagenesis studies

To test this hypothesis, *in vivo* mutational analysis were used to determine which NOD1 and RICK CARD residues are functionally important for binding and signaling. In total, seventeen mutants in full length NOD1 protein and eight mutants in full length RICK protein were constructed and analyzed. All mutations were in the CARD domains with several of them concerning the exposed and/or conserved acidic residues for NOD1 or basic residues for RICK mentioned in 3.6. NOD1 CARD was amino-terminal HA-tagged and RICK CARD was amino-terminal FLAG-tagged.

3.8.1 Co-expression and co-immunoprecipitations experiments

- NOD1 CARD mutations

The seventeen mutations in NOD1 CARD and their effect on RICK CARD binding is schematized Figure 60.

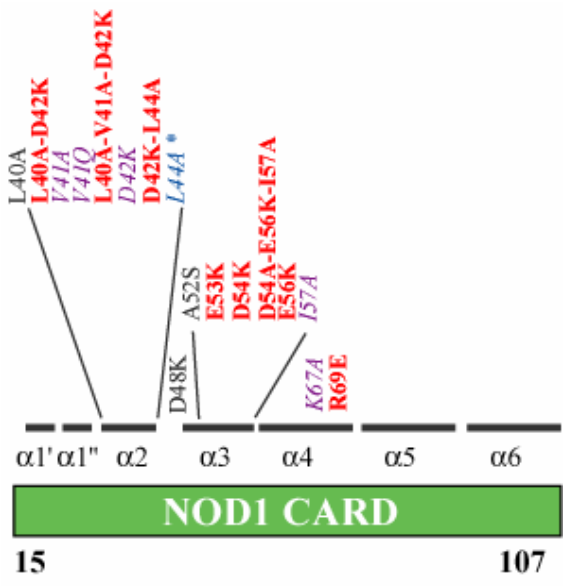

Figure 60. Schematic representation of NOD1 CARD mutants. Numbers indicate the position of amino acid residues. Mutants with wild-type activity are depicted in black. Mutations that abolish NOD1 and RICK interaction are shown in red bold. Mutants with a reduced ability to activate NF-κB are identified in purple italic. The complete loss of function mutant L44A is colored in blue and labeled with the * symbol.

To determine which NOD1 mutants still interact with RICK, I carried out immuno-precipitation experiments in which NOD1 and RICK proteins were transiently co-expressed in HEK-293T cells (cf. Figure 61).

Immuno-blotting experiments with anti-HA antibodies revealed that all NOD1 mutants were expressed at levels similar to those of wild-type NOD1. Immuno-precipitation analysis showed that RICK interacts with NOD1 point mutants D48K, I57A and K67A NOD1 as efficiently as the wild-type, illustrating that the CARD is robust to certain mutations that are presumably not involved in interactions. Three NOD1 point mutants, E53K, D54K and E56K, all exposed residues in the major acidic patch of the CARD domain, did not co-immuno-precipitate with RICK (cf. Figure 61). These residues thus play a crucial role in mediating the interaction between NOD1 and RICK, probably by promoting the CARD-CARD binding. Consistent with this, the NOD1 D54A-E56K-I57A triple mutant does not bind wild-type RICK protein.

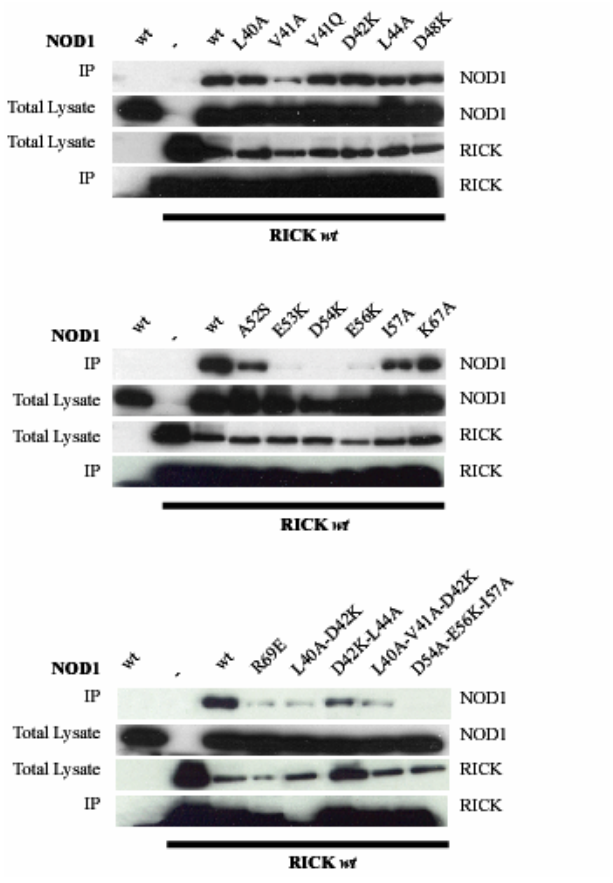

Figure 61. Interactions between mutant NOD1 proteins and wild-type RICK protein. Extracts from HEK-293T cells expressing indicated proteins were immunoprecipitated with anti-FLAG antibody and immunoblotted with both anti-HA antibody to detect NOD1 proteins and anti-FLAG antibody to detect RICK proteins. IP, immunoprecipitation; Lysate, immunoblotting of total lysates. Mutant NOD1 proteins were expressed at levels similar to those of the wild-type protein. NOD1 mutations E53K, D54K, E56K totally abolish the NOD1-RICK interaction. Representative results of mutants with wild-type activity (D48K, I57A and K67A) are shown.

NOD1 R69E mutant, in the basic patch of NOD1, shows a greatly reduced ability to interact with RICK (cf. Figure 61). In our structure, R69 forms a salt-bridge with D73, this likely being a critical stabilising feature of the CARD fold as both residues are highly conserved amongst members of the CARD family (cf. Figure 43). Thus it is possible that the mutation R69E does not affect directly the interaction with RICK, but rather destabilises the overall protein fold. That the basic patch of NOD1 CARD is unlikely to be involved in the RICK interaction is supported by the finding that the K67A mutation does not affect RICK binding.
Our results show that L40A, D42K and L44A single point mutants are able to bind RICK like wild-type and V41A with reduced affinity. However double mutants L40A-D42K and D42K-L44A and the triple mutant L40A-V41A-D42K exhibit an impaired ability to interact with RICK. These four residues form a cluster adjacent to the acidic residues E53K, D54K and E56K (cf. Figure 67) which, as the mutational effects are only exerted in combination, likely forms a peripheral part of the interaction surface. We note that the NOD1 mutant V41Q shows a good ability to bind RICK whereas it has been previously reported to abolish the interaction (Inohara et al., 1999). This difference may be explained by the fact that in this study conditions were used that give more sensitivity for detection of slightly impaired interactions. In the previous study, the binding of the wild-type NOD1 protein appeared itself very weakly and the binding of V41Q may not have been detectable in the conditions used.

- RICK CARD mutations

The eight mutations realized in RICK CARD and their effect on the NOD1/RICK CARD-CARD interaction are schematized Figure 62.

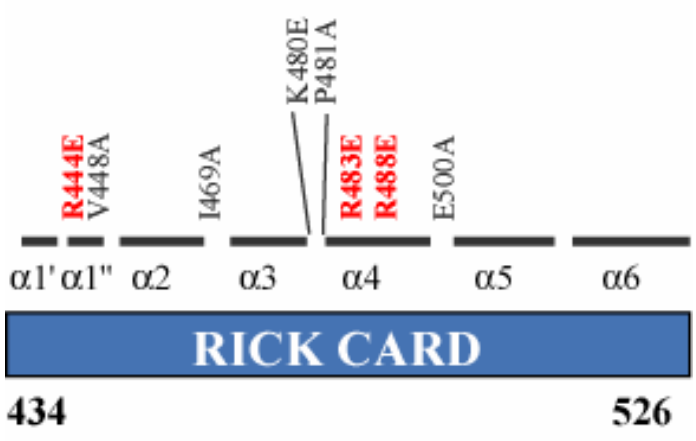

Figure 62. Schematic representation of RICK CARD mutants. Numbers indicate the position of amino acid residues. Mutants with wild-type activity are depicted in black. Mutations that abolish NOD1 and RICK interaction are shown in bold red.

I investigated which residues of RICK CARD are important for binding to NOD1 by co-transfecting HEK-293T cells with wild-type NOD1 and mutant RICK proteins and performing co-immunoprecipitation experiments (cf. Figure 63).

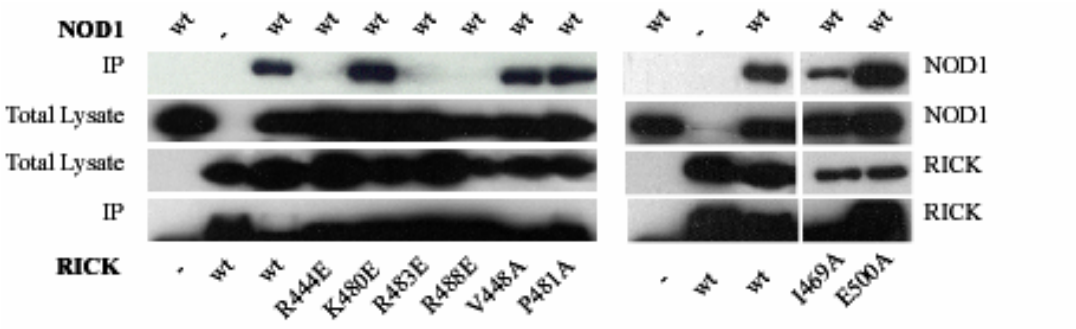

Figure 63. Interactions between wild-type NOD1 protein and mutant RICK proteins. Extracts from HEK-293T cells expressing indicated proteins were immunoprecipitated with anti-FLAG antibody and immunoblotted with both anti-HA antibody to detect NOD1 proteins and anti-FLAG antibody to detect RICK proteins. IP, immunoprecipitation; Lysate, immunoblotting of total lysates. Mutant RICK proteins express at levels similar or slightly lower to those of the wild-type protein. RICK mutations R44E, R483E and R488E totally abolish the NOD1-RICK interaction. Representative results of control mutants (I469A, E500A) with wild-type activity are shown.

Mutant proteins were expressed at levels similar to or slightly lower than those of wild-type RICK. The I469A and E500A RICK mutants were used as positive controls for the immuno-precipitation experiments. These exhibit a similar binding ability to that of the wild-type RICK. These observations indicate that the R444E, K483E and R488E RICK mutants are unable to bind NOD1.

- Conclusion

From these studies I conclude that the basic residues, R444, K483 and R488 from RICK CARD and acidic residues, E53, D54 and E56 from NOD1 CARD are key residues required for the NOD1/RICK CARD-CARD interaction. This suggests that they could mediate an electrostatic interaction between the two CARDs. The NOD1/RICK CARD-CARD complex model helped guide the functional mutational studies to identify residues may have a key role in binding.

3.8.2 NF-κB luciferase assay

Functionally important residues from NOD1 CARD that are involved inNOD1 signaling were also identified by the ability of the NOD1 mutants to activate the transcription factor NF-κB using a luciferase reporter NF-κB assay (cf. Figure 64).

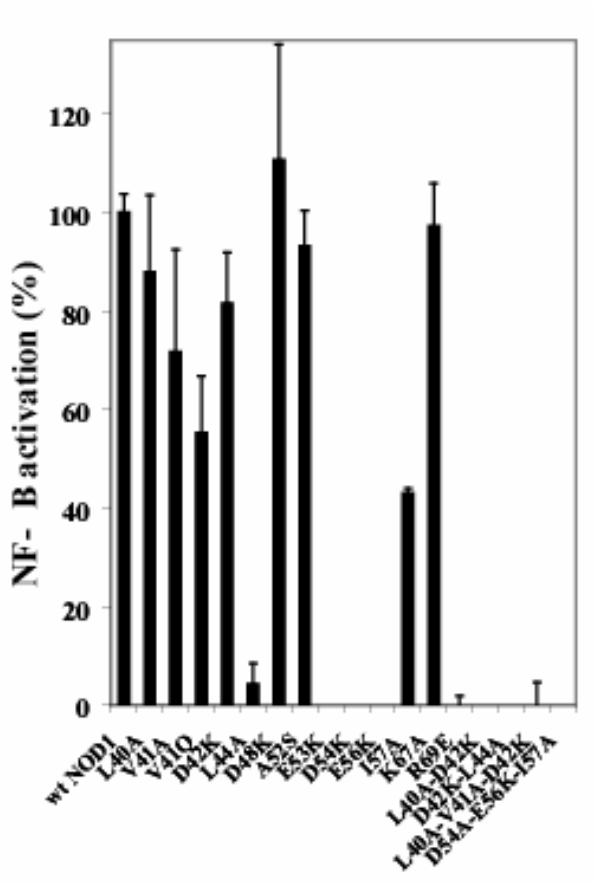

Figure 64. NF-κB activity of mutant NOD1 proteins and wild-type NOD1 protein. NOD1 mutants E53K, D54K, E56K, R69E, L40A-D42K, D42K-L44A, L40A-V41A-D42K, D54K-E56K-I57A inhibite NF-κB activation. Values represent the mean of normalized data ± s.d. of triplicate cultures.

The control NOD1 mutants, D48K, A52S and K67A, present an activation of the NF-κB transcription factor comparable to the wild-type NOD1 protein. As expected, mutants displaying an impaired ability to bind RICK, such E53K, D54K and E56K, fail to activate NF-κB. Interestingly, L44A, I57A and V41Q mutant NOD1 proteins, which have been shown to bind RICK, exhibit a greatly reduced activation of the signaling. L44 and I57 are solvent-accessible hydrophobic residues that may mediate hydrophobic interactions with other residues in RICK CARD. I thus hypothesize that the acidic and basic residues identified in NOD1 and RICK CARD are essential for the binding, while these hydrophobic residues are required for signaling. These may ensure the correct relative alignment of the domains or possibly bind to other regions of RICK.

4 Comparison of NOD1/RICK and Apaf-1/procaspase-9 CARD-CARD interactions

Electrostatic interactions, more than hydrophobic interactions, appear to be crucial for the NOD1/RICK CARD-CARD interaction. Two basic residues, R444 and R488, of RICK CARD are necessary for the interaction with NOD1 CARD. These correspond to two basic residues, R13 and R56, in procaspase-9 that were shown to be indispensable for interaction with Apaf-1 CARD (Qin et al., 1999) (cf. Figure 65). This suggests that these two proteins probably interact with their respective partners in a similar manner, through electrostatic interactions mediated by basic residues from helices 1 and 4 (cf. Figure 66)

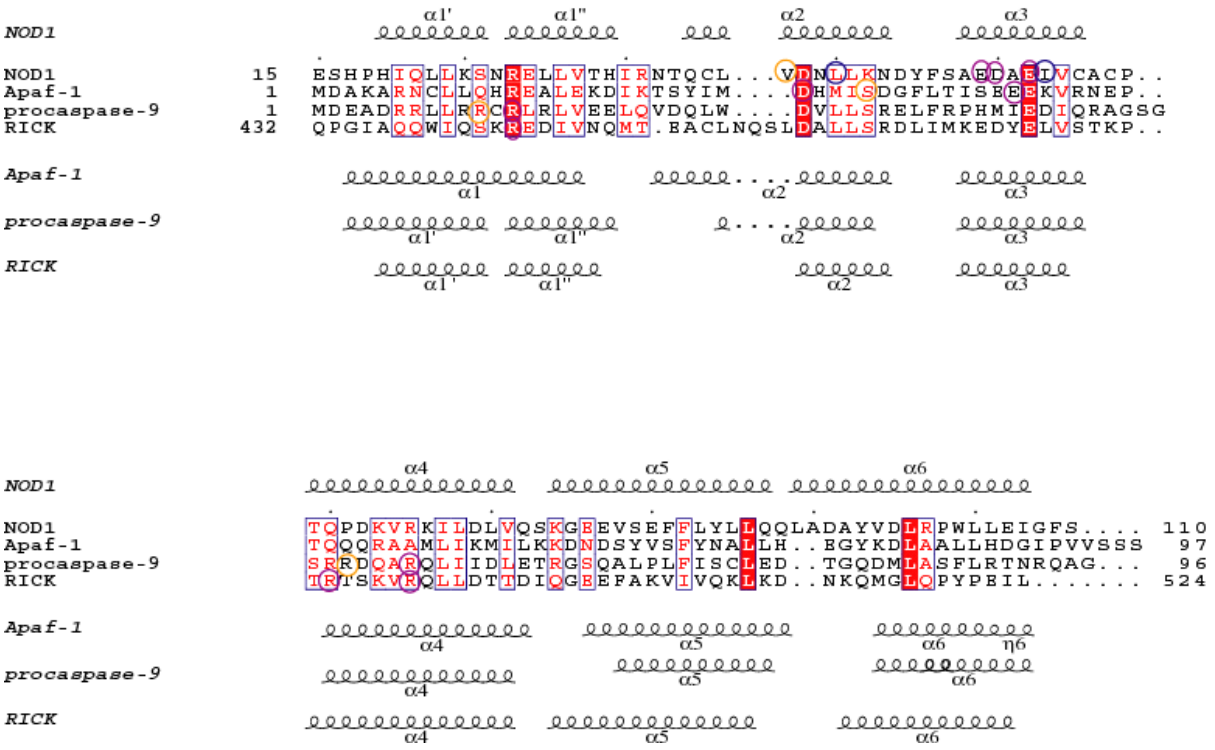

Figure 65. Sequence alignment of the CARD domains of NOD1, Apaf-1, procaspase-9 and RICK model. The secondary structure of NOD1 CARD is pictured above the alignment. Red box and white characters designate identical residues. Red characters are used for similar residues. Important residues abolishing NOD1/RICK or Apaf-1/procaspase-9 CARD-CARD interaction are circled in purple and the ones weakening the interaction in yellow. Residues whose mutation is correlated with a loss of NF-κB activation ability are circled in blue. The secondary structures of Apaf-1, procaspase-9 and the predicted secondary structure of RICK are shown below the alignment.

However, a significant difference concerns R483, the third RICK basic residue necessary for the interaction with NOD1 CARD. This residue corresponds to R51 in procaspase-9, which is not involved in the interaction between Apaf-1 and procaspase-9.

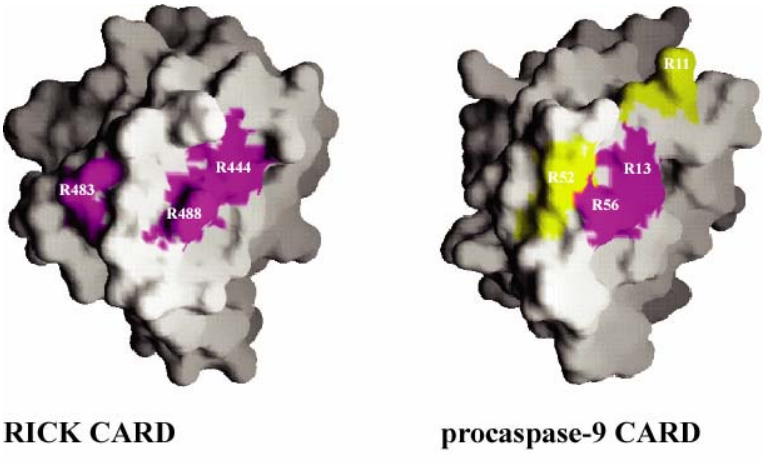

Figure 66. Surface representations of the basic patch of RICK CARD model and procaspase-9. Residues shown to be indispensable for the NOD1/RICK or Apaf-1/procaspase-9 CARD-CARD interaction (Qin et al., 1999) are identified in magenta. Procaspase-9 amino acids whose mutations weaken the Apaf-1/procaspase-9 CARD-CARD interaction (Qin et al., 1999) are colored in yellow.

Although NOD1 CARD and Apaf-1 CARD both interact with their corresponding partners through electrostatic interactions mediated by acidic residues from helices 2 and 3, the exact location of the residues involved differs (cf. Figure 66). For example, Apaf-1 D27 which is necessary for the interaction with procaspase-9 (Qin et al., 1999), corresponds to NOD1 D42 (cf. Figure 65) whose mutation does not affect RICK binding nor NF-κB activity, as shown by co-immunoprecipitation and activity experiments. The D54 and E56 NOD1 residues, which are necessary for the interaction with RICK, correspond to E39 and E41 in Apaf-1 CARD (cf. Figure 65). E39 and E41 are not involved in the interaction with procaspase-9 CARD.

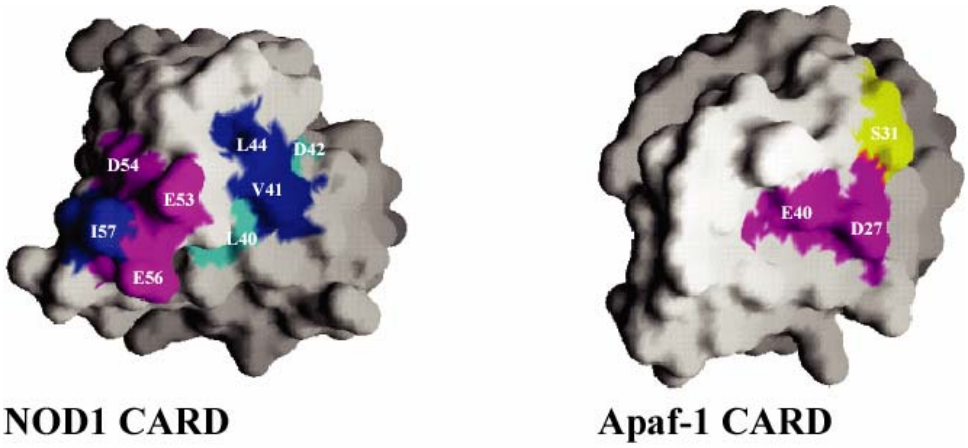

Figure 67. Surface representations of the acidic patch of NOD1 CARD and Apaf-1 CARD. Residues shown to be indispensable for the NOD1/RICK or Apaf-1/procaspase-9 CARD-CARD interaction (Qin et al., 1999) are identified in magenta. Hydrophobic NOD1 residues whose single point alterations affect NF-κB signaling are shown in blue. NOD1 residues whose associated single point mutations affect NF-κB signalling are shown in cyan. Apaf-1 amino acids whose mutations weaken the Apaf-1/procaspase-9 CARD-CARD interaction (Qin et al., 1999) are colored in yellow.

Conclusions

Conclusions and perspectives

I have determined the three-dimensional structure of a construct of NOD1 CARD comprising residues 15-138 using high resolution NMR spectroscopy. I find that residues 15-110 form a well-defined structure (although with some heterogeneity towards the carboxy-terminal end of this region) with a CARD-like fold whilst residues 111-138 are unstructured in solution. Although the structure is homologous to other known CARD structures with its Greek key six-helix bundle fold, it presents some unique features, notably the length and orientation of helix 6. Furthermore, dynamic studies showed a large exchange behaviour of this helix.

Electrostatic interactions were shown to be important for binding in the CARD-CARD interaction between Apaf-1 and procaspase-9 (Qin et al., 1999). The NMR structure shows that the surface of NOD1 exhibits distinct acidic and basic patches and this is also probably true for the RICK CARD, based on my homology model. I have therefore performed mutations in the CARD domains of NOD1 and RICK in order to test the hypothesis that residues in these patches are important for the CARD-CARD interaction and signaling.

My mutagenesis and co-immunoprecipitation experiments on full length proteins clearly show that complementary sets of charged residues, acidic residues from CARD helices 2 and 3 for NOD1 and basic residues helices 1 and 4 from RICK, are crucial for the NOD1/RICK CARD-CARD interaction. Certain hydrophobic residues play a secondary, but still important role, particularly for efficient signaling. This suggests that the NOD1-RICK CARD-CARD interaction is similar in nature to that between the CARDs of procaspase-9 and Apaf-1, in having an important electrostatic component. The E53, D54 and E56 from NOD1 CARD and R444, 483 and R488 from RICK CARD are necessary for this interaction. These residues are highly conserved in many CARD domains, such as Apaf-1 and procaspase-9 and are likely to be generally involved in CARD-CARD interactions. Furthermore, I have also identified several NOD1 residues important for activation of the NF-κB signaling pathway. These include not only the three acidic residues important for binding, but also two hydrophobic residues, L44 and I57, located in or close to the acidic patch of NOD1, not identified by co-immunoprecipation experiments with RICK. These could be involved more in the precise formation of a functional CARD-CARD interaction, although contributing less to the binding. They may also interact with other regions of RICK than the CARD domain, or even other components of the signaling pathway. In this connection, the recent crystal structure of the CARD and NOD domains of Apaf-1 in its inactive, ADP bound form, gives the first structural insight into the activation mechanism of NOD proteins (Riedl et al., 2005). The Apaf-1 CARD domain is observed to pack against the NOD domain using the same interaction

surface as previously shown to interact with the CARD domain of procaspase-9, the downstream effector. Thus, by analogy, I believe that it is probable that correct functioning of NOD1 requires that its CARD domain interacts with other parts of NOD1 in the inactive form and with RICK in the ligand activated state. This highlights the fact that the exact role of RICK in the subsequent steps of NF-κB activation is not entirely elucidated, nor have all the components of the NOD1 signaling complex been identified (Strober et al., 2006). In a recent study on the related NOD2 system, it has been proposed that it is not the serine-threonine kinase activity of RICK that is important in NOD2 signaling but rather the RICK-dependent ubiquitinylation of NEMO, a regulator of IκB kinase (IKK) (Abbott et al., 2004). In a more distantly related pathway, activation of IKK by TNFα requires site-specific ubiquitination of RIP1 (a kinase related to RICK) (Meylan and Tschopp, 2005) and polyubiquitin binding by NEMO (Ea et al., 2006). It has also been shown that TRIP6 is an important interacting partner in RICK associated signaling, including that of NOD1 (Li et al., 2005). It remains to be seen whether ubiquitination of RICK and subsequent recruitment of ubiquitin ligases (e.g. TRAF2, which binds to TRIP6) are involved in NOD1 signaling. Another interesting observation to be resolved is why the two CARDs of NOD2 are necessary for RICK binding.

Much work remains clearly to be done in characterizing the activation mechanism of NOD1 or NOD proteins in general. Structural information of other NOD1 domains would help achieving this goal. Furthermore, additional studies are required to determine all the components from the NOD1 signaling pathway, and more particularly to identify the potential cofactors involved in the ligand recognition by the LRRs.

NOD1 induces the activation of both NF-κB and caspase-1 *via* its interaction with RICK. This leads to inflammation and the processing of pro-inflammatory cytokines such interleukin-1β. This cytokine promotes angiogenesis, proliferation, and the metastasis of tumors. The structure of NOD1 CARD may be useful in developing strategies for the activation or inhibition of RICK be used in the treatment of inflammatory diseases or cancer (Damiano and Reed, 2004).

On a personal point of view, I have learned a lot during these last years. I now have competence in a broad area, from molecular biology, to cellular biology, biochemistry and structural biology. I worked on NOD1 CARD from the initial cloning to the structure determination and the *in vivo* experiments. Most importantly, I have also learnt how to manage a project by myself. NOD1 CARD was a really difficult subject and I have adapted my work, strategies and methods to overcome problems and obtain results or further

experiments to carry out. I have also learnt the importance of communication with other people in science, to discuss, get ideas, advice, and then progress. I have also been initiated in a less wholesome part of scientific research, the competition, being scooped one year after the beginning of my PhD work on my first thesis subject. Actually, I have also learnt that another group has recently solved the same structure of NOD1 CARD.

And last but not least, this thesis allowed me to improve my english by working in an international laboratory and by spending some months in Michigan. I am now fluent in English and this will probably be very useful in the future.

Conclusions et perspectives (français)

Les protéines NODs entières et leurs différents domaines sont extrêmement difficiles à étudier *in vitro* pour une finalité de biologie structurale. En effet, ces protéines possèdent non seulement généralement un faible niveau d'expression dans les systèmes classiques de surexpression (bactéries ou cellules d'insectes), mais présentent également des problèmes de solubilité et de stabilité.

De multiples constructions des différents domaines de NOD1 et de NOD2 ont été réalisées. Seule une construction du domaine CARD de NOD1 (résidus 15-138) s'est avérée posséder les qualités requises pour une étude structurale. Suite à de nombreuses tentatives de cristallogénèse qui se sont avérées infructueuses, la structure tridimensionnelle du domaine CARD de NOD1 (résidus 15-110) a été résolue à haute résolution par résonance magnétique nucléaire. J'ai effectué cette partie de mon travail de thèse au laboratoire de RMN de l'Institut de Biologie Structurale de Grenoble, laboratoire dirigé par le Dr Jean-Pierre Simorre. Les résidus 111-138 du CARD de NOD1 n'ont pas été inclus dans le calcul de structure pour plusieurs raisons. Tout d'abord, les résonances du squelette polypeptidique de cette région n'ont été que partiellement attribuées. De plus, cette partie carboxy-terminale n'est pas structurée (absence d'information de distances (NOEs)), et est hautement flexible sur une échelle de temps allant de la picoseconde à la nanoseconde.

Bien que cette structure soit homologue aux structures connues d'autre domaines CARD avec ses six hélices α repliées en clé grecque, elle présente des caractéristiques particulières inhabituelles concernant notamment l'orientation et la longueur de l'hélice carboxy-terminale.

L'ensemble de structures obtenu présente un r.m.s.d moyen, calculé sur les structures secondaires, de 0.43 ± 0.07 Å pour les atomes lourds de la chaîne principale. La qualité des structures (respect de la géométrie et analyse des violations de distance notamment.) a été validée au moyen des logiciels PROCHECK-NMR et AQUA et s'avère tout à fait satisfaisante.

Des études de dynamiques ont montré un fort comportement d'échange chimique ou conformationnel de la sixième hélice, comportement probablement dû à la présence de l'extension carboxy-terminale non structurée (résidus 111-138).

La surface du domaine CARD de NOD1 possède des patchs basiques et acides distincts. Ceux-ci sont probablement importants pour son interaction avec le domaine CARD de RICK et la signalisation en découlant.

Des tentatives d'expression et de purification du domaine CARD de RICK n'ont pas permis l'obtention de cette protéine sous forme soluble en quantité suffisante pour la caractérisation

du complexe *in vitro*, en particulier au moyen d'expérience RMN de recouvrement des fréquences de résonance.

J'ai alors modélisé grossièrement un modèle d'interaction entre les domaines CARDs de RICK et de NOD1, en me basant sur la structure connue du complexe formé par les domaines CARDs d'Apaf-1 et de procaspase-9. Ce modèle très basique a permis l'identification de résidus potentiellement importants pour cette interaction. Ces résidus ont alors été mutés dans les protéines entières pour surexpression en cellules humaines HEK-293T. L'effet des diverses mutations a été analysé au moyen d'expériences d'une part de co-expression et co-immunoprécipitation et, d'autre part, d'activation du facteur de transcription NF-κB en utilisant un rapporteur luciférase. J'ai réalisé ces expériences à Ann Arbor (Michigan, USA) dans le laboratoire du Dr Gabriel Núñez, au département de pathologie du cancer de l'Université de médecine du Michigan. Ces expériences ont montré clairement qu'un patch acide du domaine CARD de NOD1, formé de résidus des hélices 2 et 3, et un patch basique du domaine CARD de RICK, composé de résidus des hélices 1 et 4 sont cruciaux pour l'interaction NOD1/RICK CARD-CARD. Les résidus acides E53, D54 et E56 du domaine CARD de NOD1 ainsi que les résidus basiques R444, K483 et R488 sont nécessaires à cette interaction. Ces acides aminés sont généralement conservés dans les domaines CARDs, comme notamment dans ceux d'Apaf-1 et de procaspase-9 et sont probablement impliqués de manière générale dans les interactions homotypiques entre les domaines CARDs. De plus, les expériences d'activation de NF-κB ont permis l'identification de plusieurs acides aminés nécessaires pour la signalisation. Ces résidus incluent bien évidemment ceux indispensables à la formation du complexe, présentés ci-dessus, mais également deux acides aminés hydrophobes L44 et I57. Ces résidus sont localisés à proximité ou dans le patch acide du domaine CARD de NOD1 et n'affectent en rien la co-immunoprécipitation de NOD1 avec RICK. Cela signifie qu'ils peuvent être impliqués dans la formation précise d'une interaction CARD-CARD fonctionnelle, bien que n'étant pas indispensable à l'attache. Ces résidus peuvent également interagir avec des régions de RICK autres que le domaine CARD, voire même avec d'autres composants de la voie de signalisation. A ce propos, la structure cristallographique récente d'une construction d'Apaf-1, constituée du domaine CARD et du domaine d'oligomérisation, sous sa forme inactive associée à l'ADP a donné les premiers renseignements structuraux concernant l'activation des protéines NLRs (Riedl et al., 2005). Dans cette structure, le domaine CARD d'Apaf-1 est replié sur le domaine d'oligomérisation en utilisant la même surface d'interaction montrée précédemment pour interagir avec le domaine CARD de son effecteur procaspase-9. Par analogie, je pense qu'il est fortement

envisageable que le domaine CARD de NOD1 interagisse avec d'autres modules de NOD1 sous sa forme inactive, et avec RICK sous sa forme active.

Ceci souligne le fait que, d'une part, le rôle exact de RICK dans les étapes ultérieures d'activation du facteur de transcription NF-κB ne soit pas encore entièrement élucidé et que, d'autre part, tous les composants du complexe de signalisation de NOD1 ne soient pas encore identifiés (Strober et al., 2006). Dans une étude récente du système apparenté NOD2, il a été proposé que l'activité sérine-thréonine kinase de RICK ne serait pas importante pour la signalisation, mais plutôt l'ubiquitinylation médiée par RICK de la protéine IKKγ (NEMO), le régulateur du complexe IKK(Abbott et al., 2004).

Dans une voie de signalisation plus distante, l'activation de IKK par TNFα requiert une ubiquitination site-spécifique de RIP1 (une kinase apparentée à RICK) (Meylan and Tschopp, 2005). Cette ubiquitination permet l'accrochage de RIP1 sur IKKγ {Ea, 2006 #1050}.

Il a également été montré que TRIP6 est un important partenaire d'interaction dans les voies de signalisation médiées par RICK, incluant notamment celle de NOD1 (Li et al., 2005). Il reste donc à vérifier si l'ubiquitination de RICK et le recrutement de ligases à ubiquitine (comme par exemple TRAF2, qui se lie à TRIP6) sont impliqués dans la voie de signalisation de NOD1. De plus, une question importante reste à l'heure actuelle sans réponse : en effet, pour quelle raison les deux domaines CARD de NOD2 sont-ils nécessaires pour l'interaction avec RICK ? La base structurale de cette interaction est à déterminer. Beaucoup de travail reste donc encore à accomplir pour caractériser le mécanisme d'activation de NOD1 et des autres protéines NODs. Résoudre les structures tridimensionnelles des autres domaines de NOD1, l'identification de tous les composants de la voie de signalisation de NOD1 et de NOD2 et en particulier du/des potentiel(s) cofacteur(s) médiant la reconnaissance des ligands bactériens par les répétitions riches en leucine seront d'une grande aide pour atteindre ce but.

NOD1 induit, à travers son interaction avec RICK, l'activation à la fois du facteur de transcription NF-κB et de la caspase-1, conduisant à l'inflammation et à l'activation des cytokines proinflammatoires telles l'interleukine-1β. Cette cytokine promeut les phénomènes d'angiogenèse, de prolifération et de métastase des tumeurs. La structure du domaine CARD de NOD1 pourra éventuellement peut-être permettre ultérieurement l'élaboration de stratégies permettant l'activation ou l'inhibition de RICK, stratégies pouvant s'avérer très utiles dans le traitement des maladies inflammatoires et des cancers (Damiano and Reed, 2004).

- Bilan personnel

D'un point de vue personnel, ces quatre années de thèse ont été très formatrices. J'ai acquis des compétences dans un très large domaine scientifique couvrant les techniques de biologie moléculaire, de biologie cellulaire, de biochimie et également de biologie structurale et biophysique. J'ai eu la chance de pouvoir initier ce projet dans mon laboratoire, de le gérer de manière autonome, de pouvoir décider des stratégies à suivre, et de réaliser moi-même la totalité du travail présenté dans ce mémoire, du clonage initial du domaine CARD de NOD1, à la structure et aux expériences de co-immunoprécipitation et d'activation du facteur de transcription NF-κB. Les nombreuses difficultés rencontrées m'ont appris à savoir m'adapter et réagir efficacement et rapidement en reconsidérant et changeant les stratégies et méthodes envisagées de manière à mener mon projet à bien. Ayant débuté initialement seule cette nouvelle thématique dans mon laboratoire, j'ai également appris qu'il est extrêmement difficile de travailler de manière isolée. Il est primordial en sciences de savoir communiquer, échanger des idées, recevoir conseils et critiques, et collaborer.

Cette thèse m'a également initié à un autre aspect du travail de chercheur, fort peu agréable mais cependant omniprésent : la compétition. En effet, j'ai été « scoopée » en fin de première année de thèse sur mon premier sujet de thèse, portant sur la structure cristallographique du récepteur de Nogo. J'ai donc dû changer de sujet pour travailler sur les protéines NODs. Et là en fin de thèse, j'ai appris que ceci pouvait de nouveau m'arriver ; en effet, un groupe concurrent américain a également résolu la structure tridimensionnelle du domaine CARD de NOD1 et a soumis un article.

Finalement, cette thèse et ces années passées dans cet environnement international qu'est l'EMBL m'ont permis de faire d'immenses progrès en anglais, tant parlé qu'écrit. Cette langue qui est la langue officielle de travail dans mon laboratoire m'est devenue complètement naturelle. Les quelques mois passés au Michigan m'ont permis de vérifier mon aptitude à vivre sans problème au quotidien dans un environnement professionnel et personnel uniquement anglophone. Je pense que cette maîtrise de la langue anglaise me sera certainement fortement utile dans l'avenir.

Bibliography

Abbott, D. W., Wilkins, A., Asara, J. M., and Cantley, L. C. (2004). The Crohn's disease protein, NOD2, requires RIP2 in order to induce ubiquitinylation of a novel site on NEMO. Curr Biol *14*, 2217-2227.

Acehan, D., Jiang, X., Morgan, D. G., Heuser, J. E., Wang, X., and Akey, C. W. (2002). Three-dimensional structure of the apoptosome: implications for assembly, procaspase-9 binding, and activation. Mol Cell *9*, 423-432.

Ahrens, P., Kattner, E., Kohler, B., Hartel, C., Seidenberg, J., Segerer, H., Moller, J., and Gopel, W. (2004). Mutations of genes involved in the innate immune system as predictors of sepsis in very low birth weight infants. Pediatr Res *55*, 652-656.

Akira, S., and Takeda, K. (2004). Toll-like receptor signalling. Nat Rev Immunol *4*, 499-511.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell, Fourth Edition edn: Garland Science).

Albrecht, M., Domingues, F. S., Schreiber, S., and Lengauer, T. (2003). Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2. FEBS Lett *554*, 520-528.

Aravind, L., Dixit, V. M., and Koonin, E. V. (1999). The domains of death: evolution of the apoptosis machinery. Trends Biochem Sci *24*, 47-53.

Aravind, L., Dixit, V. M., and Koonin, E. V. (2001). Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. Science *291*, 1279-1284.

Baldassare, J. J., Bi, Y., and Bellone, C. J. (1999). The role of p38 mitogen-activated protein kinase in IL-1 beta transcription. J Immunol *162*, 5367-5373.

Baldwin, A. S., Jr. (1996). The NF-kappa B and I kappa B proteins: new discoveries and insights. Annu Rev Immunol *14*, 649-683.

Barnich, N., Aguirre, J. E., Reinecker, H. C., Xavier, R., and Podolsky, D. K. (2005). Membrane recruitment of NOD2 in intestinal epithelial cells is essential for nuclear factor-{kappa}B activation in muramyl dipeptide recognition. J Cell Biol *170*, 21-26.

Barton, G. M., and Medzhitov, R. (2002). Toll-like receptors and their ligands. Curr Top Microbiol Immunol *270*, 81-92.

Barton, G. M., and Medzhitov, R. (2003). Toll-like receptor signaling pathways. Science *300*, 1524-1525.

Battiste, J. L., and Wagner, G. (2000). Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry *39*, 5355-5365.

Bell, J. K., Botos, I., Hall, P. R., Askins, J., Shiloach, J., Segal, D. M., and Davies, D. R. (2005). The molecular structure of the Toll-like receptor 3 ligand-binding domain. Proc Natl Acad Sci U S A *102*, 10976-10980.

Bell, J. K., Mullen, G. E., Leifer, C. A., Mazzoni, A., Davies, D. R., and Segal, D. M. (2003). Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol *24*, 528-533.

Benedict, M. A., Hu, Y., Inohara, N., and Nunez, G. (2000). Expression and functional analysis of Apaf-1 isoforms. Extra Wd-40 repeat is required for cytochrome c binding and regulated activation of procaspase-9. J Biol Chem *275*, 8461-8468.

Bertin, J., Nir, W. J., Fischer, C. M., Tayber, O. V., Errada, P. R., Grant, J. R., Keilty, J. J., Gosselin, M. L., Robison, K. E., Wong, G. H.*, et al.* (1999). Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-kappaB. J Biol Chem *274*, 12955-12958.

Boisbouvier, J., Gans, P., Blackledge, M., Brutscher, B., and Marion, D. (1999). Long-Range Structural Information in NMR Studies of Paramagnetic Molecules from Electron Spin-Nuclear Spin Cross-Correlated Relaxation. J Am Chem Soc *121*, 7700-7701.

Bonizzi, G., and Karin, M. (2004). The two NF-kappaB activation pathways and their role in innate and adaptive immunity. Trends Immunol *25*, 280-288.

Bouchier-Hayes, L., and Martin, S. J. (2002). CARD games in apoptosis and immunity. EMBO Rep *3*, 616-621.

Bruce A. Johnson, R. A. B., (1994). NMR View: A computer program for the visualization and analysis of NMR data. Journal of Biomolecular NMR (Historical Archive) *4*, 603-614.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.

Carneiro, L. A., Travassos, L. H., and Philpott, D. J. (2004). Innate immune recognition of microbes through Nod1 and Nod2: implications for disease. Microbes Infect *6*, 609-616.

Chamaillard, M., Girardin, S. E., Viala, J., and Philpott, D. J. (2003a). Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. Cell Microbiol *5*, 581-592.

Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S.*, et al.* (2003b). An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. Nat Immunol *4*, 702-707.

Chamaillard, M., Philpott, D., Girardin, S. E., Zouali, H., Lesage, S., Chareyre, F., Bui, T. H., Giovannini, M., Zaehringer, U., Penard-Lacronique, V.*, et al.* (2003c). Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases. Proc Natl Acad Sci U S A *100*, 3455-3460.

Chaudhary, P. M., Ferguson, C., Nguyen, V., Nguyen, O., Massa, H. F., Eby, M., Jasmin, A., Trask, B. J., Hood, L., and Nelson, P. S. (1998). Cloning and characterization of two Toll/Interleukin-1 receptor-like genes TIL3 and TIL4: evidence for a multi-gene receptor family in humans. Blood *91*, 4020-4027.

Chen, F., Castranova, V., Shi, X., and Demers, L. M. (1999). New insights into the role of nuclear factor-kappaB, a ubiquitous transcription factor in the initiation of diseases. Clin Chem *45*, 7-17.

Choe, J., Kelker, M. S., and Wilson, I. A. (2005). Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science *309*, 581-585.

Chou, J. J., Matsuo, H., Duan, H., and Wagner, G. (1998). Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. Cell *94*, 171-180.

Clarkson, J., and Campbell, I. D. (2003). Studies of protein-ligand interactions by NMR. Biochem Soc Trans *31*, 1006-1009.

Collaborative Computational Project. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Cornelis, G. R., and Van Gijsegem, F. (2000). Assembly and function of type III secretory systems. Annu Rev Microbiol *54*, 735-774.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR *13*, 289-302.

Cross, T. G., Scheel-Toellner, D., Henriquez, N. V., Deacon, E., Salmon, M., and Lord, J. M. (2000). Serine/threonine protein kinases and apoptosis. Exp Cell Res *256*, 34-41.

Damiano, J. S., Newman, R. M., and Reed, J. C. (2004a). Multiple roles of CLAN (caspase-associated recruitment domain, leucine-rich repeat, and NAIP CIIA HET-E, and TP1-containing protein) in the mammalian innate immune response. J Immunol *173*, 6338-6345.

Damiano, J. S., Oliveira, V., Welsh, K., and Reed, J. C. (2004b). Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses. Biochem J *381*, 213-219.

Damiano, J. S., and Reed, J. C. (2004). CARD proteins as therapeutic targets in cancer. Curr Drug Targets *5*, 367-374.

Damiano, J. S., Stehlik, C., Pio, F., Godzik, A., and Reed, J. C. (2001). CLAN, a novel human CED-4-like gene. Genomics *75*, 77-83.

Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. Nature *411*, 826-833.

Day, C. L., Dupont, C., Lackmann, M., Vaux, D. L., and Hinds, M. G. (1999). Solution structure and mutagenesis of the caspase recruitment domain (CARD) from Apaf-1. Cell Death Differ *6*, 1125-1132.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Druilhe, A., Srinivasula, S. M., Razmara, M., Ahmad, M., and Alnemri, E. S. (2001). Regulation of IL-1beta generation by Pseudo-ICE and ICEBERG, two dominant negative caspase recruitment domain proteins. Cell Death Differ *8*, 649-657.

Duan, H., and Dixit, V. M. (1997). RAIDD is a new 'death' adaptor molecule. Nature *385*, 86-89.

Dunne, A., Ejdeback, M., Ludidi, P. L., O'Neill, L. A., and Gay, N. J. (2003). Structural complementarity of Toll/interleukin-1 receptor domains in Toll-like receptors and the adaptors Mal and MyD88. J Biol Chem *278*, 41443-41451.

Dunne, A., and O'Neill, L. A. (2003). The interleukin-1 receptor/Toll-like receptor superfamily: signal transduction during inflammation and host defense. Sci STKE *2003*, re3.

Ea, C. K., Deng, L., Xia, Z. P., Pineda, G., and Chen, Z. J. (2006). Activation of IKK by TNFalpha requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. Mol Cell *22*, 245-257.

Espagne, E., Balhadere, P., Begueret, J., and Turcq, B. (1997). Reactivity in vegetative incompatibility of the HET-E protein of the fungus Podospora anserina is dependent on GTP-binding activity and a WD40 repeated domain. Mol Gen Genet *256*, 620-627.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry *33*, 5984-6003.

Fritz, J. H., Girardin, S. E., Fitting, C., Werts, C., Mengin-Lecreulx, D., Caroff, M., Cavaillon, J. M., Philpott, D. J., and Adib-Conquy, M. (2005). Synergistic stimulation of human monocytes and dendritic cells by Toll-like receptor 4 and NOD1- and NOD2-activating agonists. Eur J Immunol.

Ghosh, S., May, M. J., and Kopp, E. B. (1998). NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. Annu Rev Immunol *16*, 225-260.

Girardin, S. E., Boneca, I. G., Carneiro, L. A., Antignac, A., Jehanno, M., Viala, J., Tedin, K., Taha, M. K., Labigne, A., Zahringer, U*., et al.* (2003a). Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. Science *300*, 1584-1587.

Girardin, S. E., Boneca, I. G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D. J., and Sansonetti, P. J. (2003b). Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. J Biol Chem *278*, 8869-8872.

Girardin, S. E., Jehanno, M., Mengin-Lecreulx, D., Sansonetti, P. J., Alzari, P. M., and Philpott, D. J. (2005). Identification of the critical residues involved in peptidoglycan detection by Nod1. J Biol Chem *280*, 38648-38656.

Girardin, S. E., Sansonetti, P. J., and Philpott, D. J. (2002). Intracellular vs extracellular recognition of pathogens--common concepts in mammals and flies. Trends Microbiol *10*, 193-199.

Girardin, S. E., Tournebize, R., Mavris, M., Page, A. L., Li, X., Stark, G. R., Bertin, J., DiStefano, P. S., Yaniv, M., Sansonetti, P. J., and Philpott, D. J. (2001). CARD4/Nod1 mediates NF-kappaB and JNK activation by invasive Shigella flexneri. EMBO Rep *2*, 736-742.

Girardin, S. E., Travassos, L. H., Herve, M., Blanot, D., Boneca, I. G., Philpott, D. J., Sansonetti, P. J., and Mengin-Lecreulx, D. (2003c). Peptidoglycan molecular requirements allowing detection by Nod1 and Nod2. J Biol Chem *278*, 41702-41708.

Gomez-Gomez, L. (2004). Plant perception systems for pathogen recognition and defence. Mol Immunol *41*, 1055-1062.

Gordon, S. (2002). Pattern recognition receptors: doubling up for the innate immune response. Cell *111*, 927-930.

Graham, F. L., and van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology *52*, 456-467.

Guay, J., Lambert, H., Gingras-Breton, G., Lavoie, J. N., Huot, J., and Landry, J. (1997). Regulation of actin filament dynamics by p38 map kinase-mediated phosphorylation of heat shock protein 27. J Cell Sci *110 (Pt 3)*, 357-368.

Gutierrez, O., Pipaon, C., Inohara, N., Fontalba, A., Ogura, Y., Prosper, F., Nunez, G., and Fernandez-Luna, J. L. (2002). Induction of Nod2 in myelomonocytic and intestinal epithelial cells via nuclear factor-kappa B activation. J Biol Chem *277*, 41701-41705.

Habeck, M., Rieping, W., Linge, J. P., and Nilges, M. (2004). NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol Biol *278*, 379-402.

Harton, J. A., Cressman, D. E., Chin, K. C., Der, C. J., and Ting, J. P. (1999). GTP binding by class II transactivator: role in nuclear import. Science *285*, 1402-1405.

Harton, J. A., Linhoff, M. W., Zhang, J., and Ting, J. P. (2002). Cutting edge: CATERPILLER: a large family of mammalian genes containing CARD, pyrin, nucleotide-binding, and leucine-rich repeat domains. J Immunol *169*, 4088-4093.

Hashimoto, C., Hudson, K. L., and Anderson, K. V. (1988). The Toll gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell *52*, 269-279.

Hayden, M. S., and Ghosh, S. (2004). Signaling to NF-kappaB. Genes Dev *18*, 2195-2224.

Hisamatsu, T., Suzuki, M., and Podolsky, D. K. (2003a). Interferon-gamma augments CARD4/NOD1 gene and protein expression through interferon regulatory factor-1 in intestinal epithelial cells. J Biol Chem *278*, 32962-32968.

Hisamatsu, T., Suzuki, M., Reinecker, H. C., Nadeau, W. J., McCormick, B. A., and Podolsky, D. K. (2003b). CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells. Gastroenterology *124*, 993-1000.

Holler, E., Rogler, G., Herfarth, H., Brenmoehl, J., Wild, P. J., Hahn, J., Eissner, G., Scholmerich, J., and Andreesen, R. (2004). Both donor and recipient NOD2/CARD15 mutations associate with transplant-related mortality and GvHD following allogeneic stem cell transplantation. Blood *104*, 889-894.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol *233*, 123-138.

Holt, B. F., 3rd, Hubert, D. A., and Dangl, J. L. (2003). Resistance gene signaling in plants--complex similarities to animal innate immunity. Curr Opin Immunol *15*, 20-25.

Houben, K., Dominguez, C., van Schaik, F. M., Timmers, H. T., Bonvin, A. M., and Boelens, R. (2004). Solution structure of the ubiquitin-conjugating enzyme UbcH5B. J Mol Biol *344*, 513-526.

Hu, Y., Ding, L., Spencer, D. M., and Nunez, G. (1998). WD-40 repeat region regulates Apaf-1 self-association and procaspase-9 activation. J Biol Chem *273*, 33489-33494.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M.*, et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature *411*, 599-603.

Hulbert, S. H., Webb, C. A., Smith, S. M., and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol *39*, 285-312.

Humke, E. W., Shriver, S. K., Starovasnik, M. A., Fairbrother, W. J., and Dixit, V. M. (2000). ICEBERG: a novel inhibitor of interleukin-1beta generation. Cell *103*, 99-111.

Huyghues-Despointes, B. M., Pace, C. N., Englander, S. W., and Scholtz, J. M. (2001). Measuring the conformational stability of a protein by hydrogen exchange. Methods Mol Biol *168*, 69-92.

Hysi, P., Kabesch, M., Moffatt, M. F., Schedel, M., Carr, D., Zhang, Y., Boardman, B., von Mutius, E., Weiland, S. K., Leupold, W.*, et al.* (2005). NOD1 variation, immunoglobulin E and asthma. Hum Mol Genet *14*, 935-941.

Inohara, Chamaillard, McDonald, C., and Nunez, G. (2005). NOD-LRR proteins: role in host-microbial interactions and inflammatory disease. Annu Rev Biochem *74*, 355-383.

Inohara, N., del Peso, L., Koseki, T., Chen, S., and Nunez, G. (1998). RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. J Biol Chem *273*, 12296-12300.

Inohara, N., Koseki, T., del Peso, L., Hu, Y., Yee, C., Chen, S., Carrio, R., Merino, J., Liu, D., Ni, J., and Nunez, G. (1999). Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB. J Biol Chem *274*, 14560-14567.

Inohara, N., Koseki, T., Lin, J., del Peso, L., Lucas, P. C., Chen, F. F., Ogura, Y., and Nunez, G. (2000). An induced proximity model for NF-kappa B activation in the Nod1/RICK and RIP signaling pathways. J Biol Chem *275*, 27823-27831.

Inohara, N., and Nunez, G. (2003). NODs: intracellular proteins involved in inflammation and apoptosis. Nat Rev Immunol *3*, 371-382.

Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., Fukase, K., Inamura, S., Kusumoto, S., Hashimoto, M.*, et al.* (2003). Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. J Biol Chem *278*, 5509-5512.

Inohara, N., Ogura, Y., and Nunez, G. (2002). Nods: a family of cytosolic proteins that regulate the host response to pathogens. Curr Opin Microbiol *5*, 76-80.

Janeway, C. A., Jr. (1989). Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol *54 Pt 1*, 1-13.

Janeway, C. A., Jr. (1992). The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol Today *13*, 11-16.

Janeway, C. A., Jr., and Medzhitov, R. (2002). Innate immune recognition. Annu Rev Immunol *20*, 197-216.
Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol *278*, 313-352.

Jones, D. A., and Takemoto, D. (2004). Plant innate immunity - direct and indirect recognition of general and specific pathogen-associated molecules. Curr Opin Immunol *16*, 48-62.

Kabesch, M., Peters, W., Carr, D., Leupold, W., Weiland, S. K., and von Mutius, E. (2003). Association between polymorphisms in caspase recruitment domain containing protein 15 and allergy in two German populations. J Allergy Clin Immunol *111*, 813-817.

Kanazawa, N., Okafuji, I., Kambe, N., Nishikomori, R., Nakata-Hizume, M., Nagai, S., Fuji, A., Yuasa, T., Manki, A., Sakurai, Y.*, et al.* (2005). Early-onset sarcoidosis and CARD15 mutations with constitutive nuclear factor-kappaB activation: common genetic etiology with Blau syndrome. Blood *105*, 1195-1197.

Karplus, M., and Grant, D. M. (1959). A Criterion For Orbital Hybridization And Charge Distribution In Chemical Bonds. Proc Natl Acad Sci U S A *45*, 1269-1273.
Kay, L. E. (1998). Protein dynamics from NMR. Biochem Cell Biol *76*, 145-152.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature *181*, 662-666.

Khursigara, G., Bertin, J., Yano, H., Moffett, H., DiStefano, P. S., and Chao, M. V. (2001). A prosurvival function for the p75 receptor death domain mediated via the caspase recruitment domain receptor-interacting protein 2. J Neurosci *21*, 5854-5863.

Kim, J. G., Lee, S. J., and Kagnoff, M. F. (2004). Nod1 is an essential signal transducer in intestinal epithelial cells infected with bacteria that avoid recognition by toll-like receptors. Infect Immun *72*, 1487-1495.

Kobayashi, K., Inohara, N., Hernandez, L. D., Galan, J. E., Nunez, G., Janeway, C. A., Medzhitov, R., and Flavell, R. A. (2002). RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. Nature *416*, 194-199.

Kobayashi, K. S., Chamaillard, M., Ogura, Y., Henegariu, O., Inohara, N., Nunez, G., and Flavell, R. A. (2005). Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science *307*, 731-734.

Kobe, B., and Deisenhofer, J. (1995). Proteins with leucine-rich repeats. Curr Opin Struct Biol *5*, 409-416.

Kong, A. N., Yu, R., Chen, C., Mandlekar, S., and Primiano, T. (2000). Signal transduction events elicited by natural products: role of MAPK and caspase pathways in homeostatic response and induction of apoptosis. Arch Pharm Res *23*, 1-16.

Koonin, E. V., and Aravind, L. (2000). The NACHT family - a new group of predicted NTPases implicated in apoptosis and MHC transcription activation. Trends Biochem Sci *25*, 223-224.

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph *14*, 51-55, 29-32.

Kosen, P. A. (1989). Spin labeling of proteins. Methods Enzymol *177*, 86-121.

Kovalovsky, D., Refojo, D., Holsboer, F., and Arzt, E. (2000). Molecular mechanisms and Th1/Th2 pathways in corticosteroid regulation of cytokine production. J Neuroimmunol *109*, 23-29.

Lala, S., Ogura, Y., Osborne, C., Hor, S. Y., Bromfield, A., Davies, S., Ogunbiyi, O., Nunez, G., and Keshav, S. (2003). Crohn's disease and the NOD2 gene: a role for paneth cells. Gastroenterology *125*, 47-57.

Landry, J., and Huot, J. (1995). Modulation of actin dynamics during stress and physiological stimulation by a signaling pathway involving p38 MAP kinase and heat-shock protein 27. Biochem Cell Biol *73*, 703-707.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and 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.

Lee, S. H., Stehlik, C., and Reed, J. C. (2001). Cop, a caspase recruitment domain-containing protein and inhibitor of caspase-1 activation processing. J Biol Chem *276*, 34495-34500.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., and Hoffmann, J. A. (1996). The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell *86*, 973-983.

Lemaitre, B., Reichhart, J. M., and Hoffmann, J. A. (1997). Drosophila host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. Proc Natl Acad Sci U S A *94*, 14614-14619.

Li, L., Bin, L. H., Li, F., Liu, Y., Chen, D., Zhai, Z., and Shu, H. B. (2005). TRIP6 is a RIP2-associated common signaling component of multiple NF-kappaB activation pathways. J Cell Sci *118*, 555-563.

Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S. M., Ahmad, M., Alnemri, E. S., and Wang, X. (1997). Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. Cell *91*, 479-489.

Li, Q., and Verma, I. M. (2002). NF-kappaB regulation in the immune system. Nat Rev Immunol *2*, 725-734.

Liepinsh, E., Barbals, R., Dahl, E., Sharipo, A., Staub, E., and Otting, G. (2003). The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition. J Mol Biol *332*, 1155-1163.

Linehan, S. A., Martinez-Pomares, L., and Gordon, S. (2000). Mannose receptor and scavenger receptor: two macrophage pattern recognition receptors with diverse functions in tissue homeostasis and host defense. Adv Exp Med Biol *479*, 1-14.

Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics *19*, 315-316.

Linhoff, M. W., Harton, J. A., Cressman, D. E., Martin, B. K., and Ting, J. P. (2001). Two distinct domains within CIITA mediate self-association: involvement of the GTP-binding and leucine-rich repeat domains. Mol Cell Biol *21*, 3001-3011.

Lipari, G., and Szabo, A. (1982a). Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 1. Theory and range of validity. J Am Chem Soc *104*, 4546-4559.

Lipari, G., and Szabo, A. (1982b). Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 2. Analysis of experimental results. 104 *J. Am. Chem. Soc.*, 4459-4570.

Liu, Z., Sun, C., Olejniczak, E. T., Meadows, R. P., Betz, S. F., Oost, T., Herrmann, J., Wu, J. C., and Fesik, S. W. (2000). Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. Nature *408*, 1004-1008.

Mansour, S. J., Matten, W. T., Hermann, A. S., Candia, J. M., Rong, S., Fukasawa, K., Vande Woude, G. F., and Ahn, N. G. (1994). Transformation of mammalian cells by constitutively active MAP kinase kinase. Science *265*, 966-970.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct *29*, 291-325.

Martin-Blanco, E. (2000). p38 MAPK signalling cascades: ancient roles and new functions. Bioessays *22*, 637-645.

Martinon, F., and Tschopp, J. (2005). NLRs join TLRs as innate sensors of pathogens. Trends Immunol *26*, 447-454.

McGovern, D. P., Hysi, P., Ahmad, T., van Heel, D. A., Moffatt, M. F., Carey, A., Cookson, W. O., and Jewell, D. P. (2005). Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. Hum Mol Genet *14*, 1245-1250.

McGovern, D. P., van Heel, D. A., Ahmad, T., and Jewell, D. P. (2001). NOD2 (CARD15), the first susceptibility gene for Crohn's disease. Gut *49*, 752-754.

Medzhitov, R. (2001). Toll-like receptors and innate immunity. Nat Rev Immunol *1*, 135-145.

Medzhitov, R., and Janeway, C. A., Jr. (1997). Innate immunity: impact on the adaptive immune response. Curr Opin Immunol *9*, 4-9.

Medzhitov, R., and Janeway, C. A., Jr. (1998). An ancient system of host defense. Curr Opin Immunol *10*, 12-15.

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A., Jr. (1997). A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature *388*, 394-397.

Meylan, E., and Tschopp, J. (2005). The RIP kinases: crucial integrators of cellular stress. Trends Biochem Sci *30*, 151-159.

Miceli-Richard, C., Lesage, S., Rybojad, M., Prieur, A. M., Manouvrier-Hanu, S., Hafner, R., Chamaillard, M., Zouali, H., Thomas, G., and Hugot, J. P. (2001). CARD15 mutations in Blau syndrome. Nat Genet *29*, 19-20.

Micheau, O., Lens, S., Gaide, O., Alevizopoulos, K., and Tschopp, J. (2001). NF-kappaB signals induce the expression of c-FLIP. Mol Cell Biol *21*, 5299-5305.

Modlin, R. L. (2001). Activation of toll-like receptors by microbial lipoproteins: role in host defense. J Allergy Clin Immunol *108*, S104-106.

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D. C. (2002). Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. Embo J *21*, 4511-4519.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. Proteins *12*, 345-364.

Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. J Mol Biol *245*, 645-660.

Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. FEBS Lett *229*, 317-324.

Nimchuk, Z., Eulgem, T., Holt, B. F., 3rd, and Dangl, J. L. (2003). Recognition and response in the plant immune system. Annu Rev Genet *37*, 579-609.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H.*, et al.* (2001a). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature *411*, 603-606.

Ogura, Y., Inohara, N., Benito, A., Chen, F. F., Yamaoka, S., and Nunez, G. (2001b). Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. J Biol Chem *276*, 4812-4818.

Ogura, Y., Lala, S., Xin, W., Smith, E., Dowds, T. A., Chen, F. F., Zimmermann, E., Tretiakova, M., Cho, J. H., Hart, J.*, et al.* (2003). Expression of NOD2 in Paneth cells: a possible link to Crohn's ileitis. Gut *52*, 1591-1597.

Opitz, B., Forster, S., Hocke, A. C., Maass, M., Schmeck, B., Hippenstiel, S., Suttorp, N., and Krull, M. (2005). Nod1-mediated endothelial cell activation by Chlamydophila pneumoniae. Circ Res *96*, 319-326.

Pages, G., Lenormand, P., L'Allemain, G., Chambard, J. C., Meloche, S., and Pouyssegur, J. (1993). Mitogen-activated protein kinases p42mapk and p44mapk are required for fibroblast proliferation. Proc Natl Acad Sci U S A *90*, 8319-8323.

Pauleau, A. L., and Murray, P. J. (2003). Role of nod2 in the response of macrophages to toll-like receptor agonists. Mol Cell Biol *23*, 7531-7539.

Peeters, H., Vander Cruyssen, B., Laukens, D., Coucke, P., Marichal, D., Van Den Berghe, M., Cuvelier, C., Remaut, E., Mielants, H., De Keyser, F., and Vos, M. D. (2004). Radiological sacroiliitis, a hallmark of spondylitis, is linked with CARD15 gene polymorphisms in patients with Crohn's disease. Ann Rheum Dis *63*, 1131-1134.

Philpott, D. J., and Girardin, S. E. (2004). The role of Toll-like receptors and Nod proteins in bacterial infection. Mol Immunol *41*, 1099-1108.

Philpott, D. J., Yamaoka, S., Israel, A., and Sansonetti, P. J. (2000). Invasive Shigella flexneri activates NF-kappa B through a lipopolysaccharide-dependent innate intracellular response and leads to IL-8 expression in epithelial cells. J Immunol *165*, 903-914.

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C.*, et al.* (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science *282*, 2085-2088.

Qin, H., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S., and Shi, Y. (1999). Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. Nature *399*, 549-557.

Rahman, P., Bartlett, S., Siannis, F., Pellett, F. J., Farewell, V. T., Peddle, L., Schentag, C. T., Alderdice, C. A., Hamilton, S., Khraishi, M.*, et al.* (2003). CARD15: a pleiotropic autoimmune gene that confers susceptibility to psoriatic arthritis. Am J Hum Genet *73*, 677-681.

Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R., and Shi, Y. (2005). Structure of the apoptotic protease-activating factor 1 bound to ADP. Nature *434*, 926-933.

Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A., and Bazan, J. F. (1998). A family of human receptors structurally related to Drosophila Toll. Proc Natl Acad Sci U S A *95*, 588-593.

Rosenstiel, P., Fantini, M., Brautigam, K., Kuhbacher, T., Waetzig, G. H., Seegert, D., and Schreiber, S. (2003). TNF-alpha and IFN-gamma regulate the expression of the NOD2 (CARD15) gene in human intestinal epithelial cells. Gastroenterology *124*, 1001-1009.

Saleh, A., Srinivasula, S. M., Acharya, S., Fishel, R., and Alnemri, E. S. (1999). Cytochrome c and dATP-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. J Biol Chem *274*, 17941-17945.

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol *234*, 779-815.

Sambrook, J., and Russell, D. W. (2001). Molecular Cloning: A Laboratory Manual, Third Edition edn (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sansonetti, P. J. (2004). War and peace at mucosal surfaces. Nat Rev Immunol *4*, 953-964.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog NMR Spectrosc *34*, 93-158.

Schleifer, K. H., and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev *36*, 407-477.

Schorey, J. S., and Cooper, A. M. (2003). Macrophage signalling upon mycobacterial infection: the MAP kinases lead the way. Cell Microbiol *5*, 133-142.

Sisk, T. J., Roys, S., and Chang, C. H. (2001). Self-association of CIITA and its transactivation potential. Mol Cell Biol *21*, 4919-4928.

Slupsky, C. M., Boyko, R. F., Booth, V. K., and Sykes, B. D. (2003). Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments. J Biomol NMR *27*, 313-321.

Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T., and Alnemri, E. S. (1998). Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. Mol Cell *1*, 949-957.

Stahl, P. D., and Ezekowitz, R. A. (1998). The mannose receptor is a pattern recognition receptor involved in host defense. Curr Opin Immunol *10*, 50-55.

Staskawicz, B. J., Mudgett, M. B., Dangl, J. L., and Galan, J. E. (2001). Common and contrasting themes of plant and animal diseases. Science *292*, 2285-2289.

Stehlik, C., Hayashi, H., Pio, F., Godzik, A., and Reed, J. C. (2003). CARD6 is a modulator of NF-kappa B activation by Nod1- and Cardiak-mediated pathways. J Biol Chem *278*, 31941-31949.

Strober, W., Murray, P. J., Kitani, A., and Watanabe, T. (2006). Signalling pathways and molecular interactions of NOD1 and NOD2. Nat Rev Immunol *6*, 9-20.

Takeda, K., and Akira, S. (2001). Roles of Toll-like receptors in innate immune responses. Genes Cells *6*, 733-742.

Takeda, K., Kaisho, T., and Akira, S. (2003). Toll-like receptors. Annu Rev Immunol *21*, 335-376.

Tanabe, T., Chamaillard, M., Ogura, Y., Zhu, L., Qiu, S., Masumoto, J., Ghosh, P., Moran, A., Predergast, M. M., Tromp, G.*, et al.* (2004). Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. Embo J *23*, 1587-1597.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Ting, J. P., and Davis, B. K. (2005). CATERPILLER: a novel gene family important in immunity, cell death, and diseases. Annu Rev Immunol *23*, 387-414.

Ting, J. P., Kastner, D. L., and Hoffman, H. M. (2006). CATERPILLERs, pyrin and hereditary immunological disorders. Nat Rev Immunol *6*, 183-195.

Tjandra, N., and Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science *278*, 1111-1114.

Tschopp, J., Martinon, F., and Burns, K. (2003). NALPs: a novel protein family involved in inflammation. Nat Rev Mol Cell Biol *4*, 95-104.

Tugarinov, V., Choy, W. Y., Orekhov, V. Y., and Kay, L. E. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci U S A *102*, 622-627.

Uehara, A., Yang, S., Fujimoto, Y., Fukase, K., Kusumoto, S., Shibata, K., Sugawara, S., and Takada, H. (2005). Muramyldipeptide and diaminopimelic acid-containing desmuramylpeptides in combination with chemically synthesized Toll-like receptor agonists synergistically induced production of interleukin-8 in a NOD2- and NOD1-dependent manner, respectively, in human monocytic cells in culture. Cell Microbiol *7*, 53-61.

Uemura, N., Okamoto, S., Yamamoto, S., Matsumura, N., Yamaguchi, S., Yamakido, M., Taniyama, K., Sasaki, N., and Schlemper, R. J. (2001). Helicobacter pylori infection and the development of gastric cancer. N Engl J Med *345*, 784-789.

Van der Biezen, E. A., and Jones, J. D. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol *8*, R226-227.

Van Heel, D. A., Ghosh, S., Butler, M., Hunt, K., Foxwell, B. M., Mengin-Lecreulx, D., and Playford, R. J. (2005). Synergistic enhancement of Toll-like receptor responses by NOD1 activation. Eur J Immunol.

Van Heijenoort, J. (2001). Formation of the glycan chains in the synthesis of bacterial peptidoglycan. Glycobiology *11*, 25R-36R.

Van Melckebeke, H., Simorre, J. P., and Brutscher, B. (2004). Amino acid-type edited NMR experiments for methyl-methyl distance measurement in 13C-labeled proteins. J Am Chem Soc *126*, 9584-9591.

Vaughn, D. E., Rodriguez, J., Lazebnik, Y., and Joshua-Tor, L. (1999). Crystal structure of Apaf-1 caspase recruitment domain: an alpha-helical Greek key fold for apoptotic signaling. J Mol Biol *293*, 439-447.

Vavricka, S. R., Musch, M. W., Chang, J. E., Nakagawa, Y., Phanvijhitsiri, K., Waypa, T. S., Merlin, D., Schneewind, O., and Chang, E. B. (2004). hPepT1 transports muramyl dipeptide, activating NF-kappaB and stimulating IL-8 secretion in human colonic Caco2/bbe cells. Gastroenterology *127*, 1401-1409.

Viala, J., Chaput, C., Boneca, I. G., Cardona, A., Girardin, S. E., Moran, A. P., Athman, R., Memet, S., Huerre, M. R., Coyle, A. J.*, et al.* (2004a). Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. Nat Immunol *5*, 1166-1174.

Viala, J., Sansonetti, P., and Philpott, D. J. (2004b). Nods and 'intracellular' innate immunity. C R Biol *327*, 551-555.

Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V., and Baldwin, A. S., Jr. (1998). NF-kappaB antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. Science *281*, 1680-1683.

Wang, T. H., Wang, H. S., and Soong, Y. K. (2000). Regulation and functions of c-Jun N-terminal kinase/stress-activated protein kinase. Chang Gung Med J *23*, 57-72.

Watanabe, T., Kitani, A., Murray, P. J., and Strober, W. (2004). NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. Nat Immunol *5*, 800-808.

Weidinger, S., Klopp, N., Rummler, L., Wagenpfeil, S., Novak, N., Baurecht, H. J., Groer, W., Darsow, U., Heinrich, J., Gauger, A.*, et al.* (2005). Association of NOD1 polymorphisms with atopic eczema and related phenotypes. J Allergy Clin Immunol *116*, 177-184.

Williamson, M. P., Havel, T. F., and Wuthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. J Mol Biol *182*, 295-315.

Willis, T. G., Jadayel, D. M., Du, M. Q., Peng, H., Perry, A. R., Abdul-Rauf, M., Price, H., Karran, L., Majekodunmi, O., Wlodarska, I.*, et al.* (1999). Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. Cell *96*, 35-45.

Wishart, D. S., and Sykes, B. D. (1994a). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR *4*, 171-180.

Wishart, D. S., and Sykes, B. D. (1994b). Chemical shifts as a tool for structure determination. Methods Enzymol *239*, 363-392.

Woessner, D. E. (1962). Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. J Chem Phys *37*, 647-654.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids).

Yoo, N. J., Park, W. S., Kim, S. Y., Reed, J. C., Son, S. G., Lee, J. Y., and Lee, S. H. (2002). Nod1, a CARD protein, enhances pro-interleukin-1beta processing through the interaction with pro-caspase-1. Biochem Biophys Res Commun *299*, 652-658.

Yu, X., Wang, L., Acehan, D., Wang, X., and Akey, C. W. (2006). Three-dimensional structure of a double apoptosome formed by the Drosophila Apaf-1 related killer. J Mol Biol *355*, 577-589.

Zhou, H., Wertz, I., O'Rourke, K., Ultsch, M., Seshagiri, S., Eby, M., Xiao, W., and Dixit, V. M. (2004). Bcl10 activates the NF-kappaB pathway through ubiquitination of NEMO. Nature *427*, 167-171.

Zhou, P., Chou, J., Olea, R. S., Yuan, J., and Wagner, G. (1999). Solution structure of Apaf-1 CARD and its interaction with caspase-9 CARD: a structural basis for specific adaptor/caspase interaction. Proc Natl Acad Sci U S A *96*, 11265-11270.

Zou, H., Henzel, W. J., Liu, X., Lutschg, A., and Wang, X. (1997). Apaf-1, a human protein homologous to C. elegans CED-4, participates in cytochrome c-dependent activation of caspase-3. Cell *90*, 405-413.

Zuiderweg, E. R. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry *41*, 1-7.

Appendices

Appendix 1. Expression and solubility of NOD1, NOD2 and RICK constructs.

Protein	Residues	Vector	Expression	Solubility	Expression system
NOD1	15 to 106	pETM-11	+	-	E. coli
NOD1	15 to 138	pETM-11	+++	++	E. coli
NOD1	15 to 138	pETM-30	+++	++	E. coli
NOD1	15 to 138	pETM-40	+++	++	E. coli
NOD1	15 to 367	pETM-11	++	-	E. coli
NOD1	15 to 367	pETM-30	++	-	E. coli
NOD1	15 to 536	pETM-11	+	-	E. coli
NOD1	15 to 536	pETM-30	++	-	E. coli
NOD1	15 to 634	pETM-30	++	-	E. coli
NOD1	15 to 674	pETM-11	+	-	E. coli
NOD1	15 to 674	pETM-30	+	-	E. coli
NOD1	168 to 534	pETM-11	-	-	E. coli
NOD1	168 to 534	pETM-30	+	-	E. coli
NOD1	168 to 534	pETM-40	+	-	E. coli
NOD1	168 to 674	pETM-11	-	-	E. coli
NOD1	168 to 674	pETM-30	+	-	E. coli
NOD1	168 to 674	pETM-40	+	-	E. coli
NOD1	171 to 490	pETM-11	+	-	E. coli
NOD1	1 to 953	pFastBac™HT B (transfer vector)	-		insect cells (High Five)
NOD1	168 to 953	pETM-11	+	-	E. coli
NOD1	168 to 953	pETM-30	+	-	E. coli
NOD1	168 to 953	pETM-40	+	-	E. coli
NOD1	657 to 931	pFastBac™HT B (transfer vector)	+++++	-	insect cells (High Five)
NOD2	28 to 224*	pLX02 (Protein' eXpert)	+	-	E. coli
NOD2	28 to 242*	pLX02 (Protein' eXpert)	+	-	E. coli
NOD2	28 to 220	pETM-11	+	-	E. coli
NOD2	20 to 245	pETM-11	++	-	E. coli
NOD2	20 to 245	pETM-30	+	-	E. coli
NOD2	20 to 245	pETM-40	+	-	E. coli
NOD2	28 to 264	pETM-11	+	-	E. coli
NOD2	28 to 264	pETM-30	+	-	E. coli
NOD2	28 to 264	pETM-40	+	-	E. coli
NOD2	273 to 577	pETM-11	+	-	E. coli
NOD2	270 to 618	pETM-11	-		E. coli
NOD2	270 to 618	pETM-30	++	-	E. coli
NOD2	270 to 618	pETM-40	+	-	E. coli
NOD2	270 to 761	pETM-11	++	-	E. coli
NOD2	270 to 761	pETM-30	++	-	E. coli
NOD2	270 to 761	pETM-40	++	-	E. coli
NOD2	270 to 1044	pETM-11	-	-	E. coli
NOD2	270 to 1044	pETM-30	+	-	E. coli
NOD2	270 to 1044	pETM-40	+	-	E. coli
NOD2	744 to 1020	pFastBac™HT B (transfer vector)	+++++	-	insect cells (High Five)
RICK	1 to 540	pETM-11	-		E. coli
RICK	1 to 540	pETM-30	+	-	E. coli
RICK	15 to 540	pETM-11	++	-	E. coli
RICK	15 to 540	PETM-30	+++	-	E. coli
RICK	432 to 526	PETM-11	+++	-	E. coli
RICK	432 to 526	pCDFDuet™ (coexpression)	+++	-	E. coli
RICK	432 to 533*	pLX02 (Protein' eXpert)	++	+	E. coli
RICK	432 to 540*	pLX02 (Protein' eXpert)	++	+	E. coli

In the case of bacterial expression, a minimum of two strains and three expression temperatures were tested for each construct. For expression in insect cells, different multiplicity of infection and expression times were tested. In each case, only the best result in term of expression and solubility is shown. Constructs made with the RoBioMol facility at the IBS are designated with an asterisk *****.

Appendix 2. Oligonucleotides used for mutagenesis.

Oligonucleotides for mutagenesis of NOD1

Leu40Ala_for: 5′-CCGCAATACTCAGTGT***GC***GGTGGACAACTTGCTG-3′
Leu40Ala_rev: 5′-CAGCAAGTTGTCCACC***GC***ACACTGAGTATTGCGG-3′
Val41Ala_for: 5′-CAATACTCAGTGT***C***TGGCGGACAACTTGCTGAAG-3′
Val41Ala_rev: 5′-CTTCAGCAAGTTGTCC***G***CCAGACACTGAGTATTG-3′
Val41Gln_for: 5′-GCAATACTCAGTGTCTG***CA***GGACAACTTGCTGAAG-3′
Val41Gln_rev: 5′-CTTCAGCAAGTTGTCC***TG***CAGACACTGAGTATTGC-3′
Asp42Lys_for: 5′-CGCAATACTCAGTGTCTGGTG**AAA**AACTTGCTGAAGAATGACTAC-3′
Asp42Lys_rev: 5′-GTAGTCATTCTTCAGCAAGTT**TTT**CACCAGACACTGAGTATTGCG-3′
Leu44Ala_for: 5′-CAGTGTCTGGTGGACAA**CG**CGCTGAAGAATGACTACTTC-3′
Leu44Ala_rev: 5′-GAAGTAGTCATTCTTCAGC**GC**GTTGTCCACCAGACACTG-3′
Asp48Lys_for: 5′-GTGGACAACTTGCTGAAGAAT**AAA**TACTTCTCGGCCGAAGATGCG-3′
Asp48Lys-rev: 5′-CGCATCTTCGGCCGAGAAGTA**TTT**ATTCTTCAGCAAGTTGTCCAC-3′
Ala52Ser_for: 5′-CTGAAGAATGACTACTTCTCG**AG**CGAAGATGCGGAGATTGTGTG-3′
Ala52Ser_rev: 5′-CACACAATCTCCGCATCTTCG**CT**CGAGAAGTAGTCATTCTTCAG-3′
Glu53Lys_for: 5′-GACTACTTCTCGGCC**A**AAGATGCGGAGATTG-3′
Glu53Lys_rev: 5′-CAATCTCCGCATCTT**T**GGCCGAGAAGTAGTC-3′
Glu54Lys_for: 5′-GACTACTTCTCGGCCGAA**AAA**GCGGAGATTGTGTGTGCC-3′
Glu54Lys_rev: 5′-GGCACACACAATCTCCGC**TTT**TTCGGCCGAGAAGTAGTC-3′
Glu56Lys_for: 5′-CTCGGCCGAAGATGCG**AAA**ATTGTGTGTGCCTGCCC-3′
Glu56Lys_rev: 5′-GGGCAGGCACACACAAT**TTT**CGCATCTTCGGCCGAG-3′
Ile57Ala_for: 5′-CCGAAGATGCGGAG**GCG**GTGTGTGCCTGCCC-3′
Ile57Ala_rev: 5′-GGGCAGGCACACAC**CGC**CTCCGCATCTTCGG-3′
Lys67Ala_for: 5′-CCACCCAGCC**TGA**CGCCGTCCGCAAAATTCTG-3′
Lys67Ala_rev: 5′-CAGAATTTTGCGGAC**GGC**GTCAGGCTGGGTGG-3′
Arg69Glu_for: 5′-CACCCAGCCTGACAAGGTC**GAA**AAAATTCTGGACCTGGTAC-3′
Arg69Glu_rev: 5′-GTACCAGGTCCAGAATTTT**TTC**GACCTTGTCAGGCTGGGTG-3′
Leu40Ala_Asp42Lys_for:
5′- CACATCCGCAATACTCAGTGT**GCGGTGAAA**AACTTGCTGAAGAATGACTAC-3′
Leu40Ala_Asp42Lys_rev:
5′- GTAGTCATTCTTCAGCAAGTT**TTTCACCGC**ACACTGAGTATTGCGGATGTG-3′
Leu40Ala_Val41Ala_Asp42Lys_for:
5′-CATCCGCAATACTCAGTGT**GCGGCGAAA**AACTTGCTGAAGAATGAC-3′
Leu40Ala_Val41Ala_Asp42Lys_rev:
5′-GTCATTCTTCAGCAAGTT**TTTCGCCGC**ACACTGAGTATTGCGGATG-3′
Asp54Ala_Glu56Lys_Ile57Ala_for:
5′-CTACTTCTCGGCCGAA**GCGGCGAAAGC**GGTGTGTGCCTGCCCCAC-3′
Asp54Ala_Glu56Lys_Ile57Ala_rev:
5′-GTGGGGCAGGCACACA**CCGCTTTCGCC**GCTTCGGCCGAGAAGTAG-3′

Oligonucleotides for mutagenesis of RICK

Arg444Glu_for: 5′-CAGCAGTGGATCCAGAGCAAA**GAA**GAAGACATTGTGAACCAAATG-3′
Arg444Glu_rev: 5′-CATTTGGTTCACAATGTCTTC**TTC**TTTGCTCTGGATCCACTGCTG-3′
Val448Ala_for: 5′-CAAAAGGGAAGACATTG**C**GAACCAAATGACAGAAG-3′
Val448Ala_rev: 5′-CTTCTGTCATTTGGTTC**G**CAATGTCTTCCCTTTTG-3′
Ile469Ala_for: 5′-CCCTTCTGTC**CA**GGGACTTGGCCATGAAAGAGGACTATGAAC-3′
Ile469Ala_rev: 5′-GTTCATAGTCCTCTTTCATG**GC**CAAGTCCCTGGACAGAAGGG-3′
Lys480Glu_for: 5′-GACTATGAACTTGTTAGTACC**GAA**CCTACAAGGACCTCAAAAGTC-3′
Lys480Glu_rev: 5′-GACTTTTGAGGTCCTTGTAGG**TTC**GGTACTAACAAGTTCATAGTC-3′
Pro481Ala_for: 5′-GAACTTGTTAGTACCAAG**GCG**ACAAGGACCTCAAAAGTC-3′
Pro481Ala_rev: 5′-GACTTTTGAGGTCCTTGT**CGC**CTTGGTACTAACAAGTTC-3′
Arg483Glu_for: 5′-GTACCAAGCCTACA**GAA**ACCTCAAAAGTCAG-3′
Arg483Glu_rev: 5′-CTGACTTTTGAGGT**TTC**TGTAGGCTTGGTAC-3′
Arg488Glu_for: 5′-CTACAAGGACCTCAAAAGTC**GA**ACAATTACTAGACACTACTG-3′
Arg488Glu_rev: 5′-CAGTAGTGTCTAGTAATTGT**TC**GACTTTTGAGGTCCTTGTAG-3′
Glu500Ala_for: 5′-CTGACATCCAAGGAG**AA**GCCTTTGCCAAAGTTATAGTAC-3′
Glu500Ala_rev: 5′-GTACTATAACTTTGGCAAA**GG**CTTCTCCTTGGATGTCAG-3′

Mismatched bases reslting in point mutations are indicated in bold.

Appendix 3. ^{15}N relaxation experiments

All relaxation experiments were performed at 303 K on a Varian INOVA spectrometer (^{1}H frequency of 600 MHz) using ^{15}N-labeled samples with a concentration of approximately 0.7 mM. ^{15}N R_1 and heteronuclear ^{15}N $\{^{1}H\}$-NOE values were determined using the experiments described by Farrow et al. (Farrow et al., 1994). R_1 times were extracted from 14 spectra with different values for the relaxation delay: 0 (2x), 55, 111, 166, 222, 333, 444, 666, 888, 1110 (2x), 1554 and 1998 ms, applying 180° pulses on protons every 5 ms to suppress CSA(^{15}N)/DD(HN) cross-correlated relaxations. ^{15}N R_2 relaxation times were extracted from $R_{1\rho}$ experiments. The $R_{1\rho}$ experiments were recorded with spin-lock pulses of varying lengths 2, 6 (2x), 10, 20, 30, 40, 60, 80 (2x), 100, 120, 140 and 160 ms (Houben et al., 2004). Relaxation parameters were extracted with the program Curvefit, using a two-parameter fitting and a Monte Carlo simulation to estimate the errors. In the $R_{1\rho}$ experiments, a spin lock field of $|\gamma_N B1|/2\pi$ = 1.6 kHz was applied on the ^{15}N spins to suppress chemical shift and scalar J_{NH} coupling evolution. Transverse relaxation rate constants R_2 were then calculated from the measured $R_{1\rho}$ rate constants, taking into account the frequency offset $\Delta\nu$, using the equation:

$R_{1\rho} = \cos^2(\theta)R_2+\sin^2(\theta)R_1$

with $\theta = \tan^{-1}(2\pi\Delta\nu/\gamma_N B1)$

The ^{15}N heteronuclear relaxation rates were interpreted using the program TENSOR2 which uses the description of the molecular diffusion derived by Woessner, in combination with the Lipari-Szabo model-free analysis of local flexibility (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b). In the model-free approach, internal mobility is characterized using an order parameter S^2, which may be interpreted as the amplitude of the motion and a correlation time τi the characteristic time constant of this motion. The program TENSOR2 has been described elsewhere (Dosset, 2000; Tsan, 2000) and all data analysis was performed as described in these references. Briefly, the analysis uses the following approach: rotational diffusion of the molecule is characterized using the ratio R_2/R_1 measured in regions of the molecule where contributions to this ratio from internal motion are expected to be negligible. Once the diffusion tensor has been characterised, the contribution to R_1, R_2 and the heteronuclear NOE from the global motion is derived from this analysis and the remaining contribution is fitted to the local motional parameters (S^2 and τi). In the case of insufficient reproduction of the experimental data using these parameters, an additional chemical shift exchange contribution to R_2 (Rex), or an extended motion comprising two independent internal motions may be evoked. R_1, R_2 and the heteronuclear NOE are expressed in terms of the angular spectral density function of the inter-nuclear vector at different transition frequencies as described above. The inter-nuclear distance rNH was assumed to average to 1.01, and the chemical shift anisotropy of the ^{15}N nucleus (s// - s^) was approximated to -170 ppm.

Appendix 4. Article submitted to Journal of Molecular Biology

Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK

Florence Manon[1], Adrien Favier[2], Gabriel Núñez [3], Jean-Pierre Simorre[2], and Stephen Cusack[1,*]

[1]Grenoble Outstation, European Molecular Biology Laboratory, 6 rue Jules Horowitz, BP 181, F-38042 Grenoble Cedex 9, France
[2]Institut de Biologie Structurale Jean-Pierre Ebel CNRS-CEA-UJF, Laboratoire de Résonance Magnétique Nucléaire, 41 rue Jules Horowitz, F-38027 Grenoble Cedex 1, France
[3]Department of Pathology and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

* Corresponding author
Telephone (33) 476207238
Telefax (33) 476207199
Email: cusack@embl-grenoble.fr

Running title: NOD1 CARD solution structure

Submitted to Journal of Molecular Biology, July 4th 2006.

Abstract

NOD1 is a cytosolic signaling host pattern-recognition receptor composed of a caspase-activating and recruitment domain (CARD), a nucleotide-binding and oligomerization domain (NOD) and leucine-rich repeats. It plays a crucial role in innate immunity by activating the NF-κB pathway *via* its downstream effector the kinase RICK (RIP2) following the recognition of a specific bacterial ligand. RICK is recruited by NOD1 through interaction of their respective CARDs. This paper presents the high resolution NMR structure of the NOD1 CARD. It is generally similar to other CARDs of known structure, consisting of six tightly packed helices, although the length and orientation of the last helix is unusual. Mutations in both the NOD1 and RICK CARD domains were assayed by immuno-precipitation of cell lysates and *in vivo* NF-κB activation in order to define residues important for CARD-CARD interaction and downstream signalling. The results show that the interaction is critically dependent on three acidic residues on NOD1 CARD and three basic residues on RICK CARD and thus is likely to have a strong electrostatic component, similar to other characterised CARD-CARD interactions.

Keywords: NOD1/CARD4; RICK/RIP2/CARDIAK; NF-κB; inflammation; apoptosis; innate immunity; NMR structure

Introduction

The survival of a multicellular organism to pathogenic attack depends on the sensing of microbial-derived products and the subsequent initiation of immune system responses. Innate immunity relies on the specific recognition of pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs). The NOD proteins are a family of intracellular PRRs implicated in the recognition of PAMPs, which results in activation of pro-inflammatory pathways [1; 2; 3]. Most NOD family members show a modular tripartite domain organization comprising a carboxy-terminal ligand-recognition domain containing leucine-rich repeats and likely involved in the recognition of PAMPs, a centrally located nucleotide-binding oligomerization domain (NBD or NOD) typically consisting of a NACHT domain (domain present in NAIP, CIITA, HET-E, TP-1) and several NACHT-associated domains (NAD) and an amino-terminal effector-binding domain (EBD) that can be a caspase-activating recruitment domain (CARD), a baculovirus inhibitor of apoptosis protein repeat (BIR) or a pyrin domain. The NBD mediates the self-oligomerization of the protein that is a prerequisite for inducing signal transduction *via* the EBD [4].

The NOD family member, NOD1 (or CARD4) acts as a cytosolic receptor for the diaminopimelate-containing GlcNAc-tripeptide muropeptide found mostly in gram-negative bacterial peptidoglycans [5; 6]. It has been shown to activate and regulate pro-inflammatory pathways by mediating the activation of the nuclear factor NF-κB [7; 8] and also triggering interleukin-1β secretion by procaspase-1 activation [9]. NOD1 has been specifically implicated in the response to *Helicobacter pylori* [10]. Recently NOD1 polymorphisms have been associated with eczema, atopic asthma and inflammatory bowel disease [11; 12]. NOD1 activates NF-κB *via* its downstream effector, a serine-threonine kinase named RICK also known as RIP2 or CARDIAK, through homotypic CARD-CARD interactions [7; 8]. The CARD [13] is a protein-protein interaction domain belonging to the death domain superfamily, which also includes the death domain, death effector domain and pyrin domain [14; 15]. It allows the assembly of multi-protein complexes and plays a key role in different cellular processes such inflammation, apoptosis, cancer or auto-inflammatory diseases by activating notably caspase-1, caspase-8 or NF-κB [16]. The CARD structure comprises a bundle of six antiparallel α-helices, similar to other death domain superfamily members, and the atomic structures of several CARD domains have been determined by X-ray crystallography or NMR: Apaf-1 CARD [17; 18; 19], procaspase-9 CARD [19], RAIDD CARD [20], Iceberg [21].

RICK is known to interact through homotypic CARD-CARD interactions with seven other CARD-containing proteins; NOD2 (CARD15) [22], CARD6 [23], caspase-1 [24], BIR2/cIAP1 [25], BIR3/cIAP2 [25], Bcl10 [26] and COP [27; 28]. NOD2 is another NOD family member implicated in the same signaling pathway as NOD1. Its EBD comprises two CARD domains, both of them necessary to recruit RICK [22]. Point mutations or truncations in NOD2 can give susceptibility to certain inflammatory diseases such as Crohn's disease [29], a chronic inflammatory disease of the gastrointestinal tract or Blau syndrome [30], another granulomatous disorder.

In this paper we report the expression and purification of a soluble form of the CARD domain of NOD1 and determination of its solution structure by high resolution NMR. Furthermore, by mutagenesis studies using co-immuno-precipitation we have identified residues in NOD1 and RICK CARDs critical for the interaction between NOD1 and RICK, and by NF-κB activation assays in mammalian cells, residues in the NOD1 CARD that are important for its signalling function. These results support the notion that the interaction between NOD1 and RICK CARDs is primarily electrostatic, similar to that found in other CARD-CARD interactions such as between the Apaf-1 and procaspase-9 CARDS. Understanding the structural basis of the CARD-CARD interaction between NOD1, NOD2 and RICK is of considerable importance for a better knowledge of the activation mechanism of NOD proteins and for potential exploitation of the CARD-CARD interface as a target for therapeutic intervention and treatment of inflammatory diseases [31].

Results

Structure Determination

A number of constructs encompassing the human NOD1 CARD domain were tested for soluble and high levels of expression. Notably, a construct containing residues 15 to 106 showed a very low expression level and solubility, whereas one containing residues 15 to 138 gave high yields of very soluble protein. This suggests that some residues in the region 107-138 are important for protein stability. As extensive crystallisation trials proved negative, the 15-138 construct was used for structure determination by NMR spectroscopy.

More than 96 % of the backbone chemical shifts could be assigned. The residues not assigned are mainly located in the C-terminal extension (residues 111-138) which is highly flexible in the picosecond to the nanosecond time scale, as shown by low $^{15}N\{^{1}H\}$-NOE values. Moreover the lack of intra- and inter- ^{1}H-^{1}H NOESY cross peaks suggests that this region experiences conformational exchange. Therefore, only residues 15-110 were included in the structure calculation.

The assigned ^{1}H-^{15}N HSQC spectrum of NOD1 CARD shows a good dispersion of the ^{1}H-^{15}N cross-peaks (Figure 1a). However, the spin systems of L100 and R101 preceding P102 were not observed in the triple resonance spectra used for backbone assignment. The relaxation study detailed below shows that these residues are located in the region [N91- G108] which undergoes conformational or chemical exchange (Figure 3b) suggesting that L100 and R101 NMR lines are broadened by intermediate exchange. Moreover, in about the same region [D95-S110], three sets of resonance lines are present in the NMR spectra for each residue, particularly visible in the ^{1}H-^{13}C HSQC spectrum of the aromatic residues (Figure 1b). For example, W103, exhibits one major conformation and two less populated conformations with an abundance of approximately 75, 15 and 10 % (see resonances obtained for its Hδ1 or Hζ2, Figure 1b). By careful analysis of the signals from P102, it was found that cis-trans isomerization of this residue is not the cause of the conformational heterogeneity of the C-terminal region.

The structure of the main conformation of the CARD domain of NOD1 was defined by a total of 1235 NMR-derived distance constraints (AQUA analysis, [32]). The complete structural statistics and root-mean-square deviation values are presented in Table 1.

Structure Overview

The solution structure of NOD1 CARD (Figure 2a and 2b) consists of an arrangement of six α-helices closely packed around a hydrophobic core including in particular hydrophobic residues highly conserved among members of the CARD family known for interacting with RICK (Figure 2c). Helices 2-5 form an antiparallel four-helix bundle with helix 6 and helix 1 positioned on the top of helix 5 and helix 4 respectively and crossing each other. Helix 1 is interrupted by a severe kink at residue Q26 neighbouring the conserved R27 and consequently is divided into two smaller helices, helix 1' (N-terminal to the kink) and helix 1" (C-terminal to the kink) with helix 1' parallel to helix 4 and helix 1" parallel to the N-terminal part of helix 6. Helix 6 is slightly curved at position 100. This is most likely due to the fact that residues 100 and 101 are invisible in the NMR spectra, that residue 102 is a proline and because only a few long distance NOE restraints define the orientation of its C-terminal part. Helices 3, 5 and 6 cap the exposed hydrophobic core formed by helices 1, 2 and 4.

The known structures most similar to NOD1 CARD found using the DALI server [33] are those of Iceberg [21] (Z-score = 6, RMSD of 3 Å for 81 aligned Cα), RAIDD [20] (Z-score = 5.3, RMSD 3.8 Å for 85 aligned), procaspase-9 [19] (Z-score = 5.0, RMSD 3.5 Å for 82 aligned) and Apaf-1 [17] (Z-score = 5.0, RMSD 3.5 Å for 82 aligned). These structural similarities are however lower than that observed between CARD domains generally (e.g. procaspase-9/Apaf-1 Z-score = 14.9, procaspase-9/iceberg Z-score = 10.9, Apaf-1/iceberg Z-score = 11.1). Thus while the overall topology is similar, NOD1 CARD differs notably in the conformation of some of the inter-helical loops and consequently the length and orientation of some of the helices (Figure 2d). The main difference concerns helix 6 which begins earlier and is longer in NOD1 CARD than in the other CARDs considered. This difference could be observed already from the backbone assignment step analysing the chemical shift index of the Cα, Cβ and CO resonances. Most significantly the final direction of the extended helix 6 differs considerably, as shown in Figure 2e. Compared to other CARD domains, the connections between helices 1 and 2 is altered (due to the presence of helix 1") and also that between helices 5 and 6. The C-terminal extension of helix 6 after the kink around P102 leads it to cross over helix 1' and point in the opposite direction to the C-terminus of other CARD domains. However it should be borne in mind that helix 6 is more flexible than the five others as described below by ^{15}N relaxation analysis.

^{15}N Relaxation and Backbone Dynamics

^{15}N spin relaxation data contain information on the time-dependent fluctuations of individual backbone N-H bond vectors. To analyze the measured relaxation data of the CARD domain of NOD1, we used the Lipari-Szabo model-free approach [34; 35] in combination with the description of rotational diffusion anisotropy formulated by Woessner [36]. Residues with overlapping ^{1}H-^{15}N correlation peaks or weak cross peak intensities due to chemical exchange were removed from the analysis.

The R_2/R_1 value analysis leads to a fully anisotropic diffusion tensor with an overall correlation time of the molecule (τc) of 7.6 ± 0.1 ns, an anisotropy of 1.24 ± 0.03 and a rhombicity of 1.4 ± 0.25. This anisotropy is rather large for the relatively small size of the protein, but this diffusion property is likely caused by the unstructured C-terminal tail of the protein. The generalized order parameters S^2 obtained for the 61 remaining residues are displayed in Figure 3a). The colour scheme reflects increasing order parameters, i.e. degree of local flexibility along the protein backbone, from red (more flexible) to yellow, light blue and blue (more rigid) in the picosecond to nanosecond time scale. Residues located in the helices exhibit mostly small amplitude motions except for L44, V82 and W103. The rest of the more dynamic NH vectors are located in loops or at extremity of helices. Residues exhibiting chemical or conformational exchange in the microsecond to millisecond time scale range are colour coded in red Figure 3b). R27 and H33 located at the extremities of helix 1' as well as V75 located at the C-terminal of helix 4 are involved in such slow events as well as the majority of the residues of the last helix. The large exchange behaviour of helix 6 may be caused by the connected and flexible C-terminal tail and suggests that residues L100 and R101 could probably not be assigned due to extensive line broadening in this region

Surfaces of NOD1 and RICK CARDs

Studies of CARD-CARD binding in the Apaf1/caspase-9 [19] and RAIDD/caspase-2 [20] systems lead to the conclusion that tight binding is achieved by electrostatic interaction of oppositely charged surfaces on each partner, which would guide the initial docking, reinforced by hydrogen bonds and hydrophobic interactions. For example, the Apaf-1/procaspase-9 CARD-CARD crystal structure, shows that an acidic patch of Apaf-1 interacts with a basic patch of procaspase-9 [18; 19]. We therefore examined whether there were similar patches of charged residues on the NOD1 CARD surface. As shown in Figure 4, the NOD1 surface includes a basic patch composed of K24, K67, R69 and K70 belonging to helices 1 and 4, which is surrounded by two acidic patches on its left and on its right (Figure 4a). The left patch consists mainly of acidic residues of helices 5 and 6 in particular E80, E84, D95, D99 and E106 whereas the right one (Figure 4b), which is the most prominent, is composed of acidic residues from helices 2 and 3: D42, D48, E53, D54 and E56. As shown in Figure 4c there are also a number of hydrophobic residues, mainly from helices 5 and 6, associated with the prominent acidic patch.

We also constructed a homology model of the RICK CARD to identify putative surface features. The amino acid sequence conservation in the CARD family is generally low although the known structures present a high degree of structural homology. Procaspase-9 CARD was used as a template for RICK CARD homology modelling because amongst the CARD domains of known structure, it has the highest sequence identity with RICK CARD (about 20 %). Moreover, it is the only other CARD domain known so far that interacts with NOD1 CARD [8], suggesting that the procaspase-9 CARD and RICK CARD interactions with NOD1 might share common features. The modelled RICK CARD also has distinct positively and negatively charged surfaces. The most extensive basic surface of RICK (Figure 4e) comprises K443 and R444 from the turn between helices 1 and 1', Lys480 from the turn between helices 3 and 4, R483 and R488 from helix 4. The acidic patch of the modelled RICK CARD contains primarily residues from helices 3 and 5 (Figure 4d).

NOD1/RICK CARD-CARD interactions

Convincing evidence for the direct interaction between NOD1 and RICK CARDs has been provided by two groups using different approaches. Inohara and colleagues showed by co-immuno-precipitation experiments that residues 374-540 of RICK, which contains its C-terminal CARD domain, was sufficient to bind full length NOD1 [8]. They also showed that residues 1-648 of NOD1 (i.e. lacking the LRRs but containing the N-terminal CARD) were able to bind full length RICK protein, whereas constructs containing only the LRRs or both the LRRs and NBD (residues 126-953) were not. Bertin and colleagues used an amino terminal fragment of NOD1 (residues 1-145), encompassing the CARD, as bait in a yeast two-hybrid screen of a human breast cDNA library to find interacting proteins [7]. They thus identified RICK and showed that its CARD alone (residue 435-540) was sufficient for binding the bait. They showed by co-immuno-precipitation experiments in mammalian cells that full length NOD1 is unable to bind a construct of RICK l (residues 1-435) lacking the CARD. These experiments show that the CARD domains of NOD1 and RICK are both indispensable for interaction of NOD1 with RICK, independently of the kinase domain of RICK and of the NBD and LRR domains of NOD1.

We therefore asked the question as to whether there is an acidic patch of NOD1 CARD which interacts with a basic patch of RICK CARD or *vice versa*. Like Apaf-1, NOD1 is able to bind procaspase-9. When over-expressed, NOD1 promotes caspase-9-induced apoptosis through CARD-CARD interactions [8]. Both the CARD and the NBD are essential for NOD1 to activate procaspase-9 [8], similarly to Apaf-1 [37; 38]. This suggests that NOD1 may interact with procaspase-9 by a mechanism similar to Apaf-1. Thus we hypothesised that in the NOD1/procaspase-9 and in the NOD1/RICK interactions, it is the basic patch of procaspase-9 or RICK that interacts with the acidic patch of NOD1. Moreover, the NOD1 CARD acidic patch contains a number of hydrophobic residues (Figure 4c) potentially able to interact with those in the basic patch of RICK (Figure 4f), as observed in the Apaf-1/procaspase-9 CARD-CARD complex [19]. Published mutational analyses of the CARDs of NOD2, another NOD-1 related protein interacting with RICK and implicated in the same signalling pathway, reinforce this hypothesis. Firstly the NOD2 loss of function mutant E69K corresponds to residue E53 in NOD1 CARD [39]. Mutations A106V, L145P and F161I in NOD2 CARDs reduce or even abolish the interaction with RICK [39]. The corresponding residues in NOD1 CARD structure are either conserved hydrophobic residues or residues juxtaposed to hydrophobic conserved residues and they are all located in the acidic patch (Figure 4c).

In order to identify residues potentially involved in the interaction, we constructed a crude docking model for the NOD1 and RICK CARDs based on the known structure of the Apaf-1/procaspase-9 CARD-CARD complex [19]. We superimposed NOD1 CARD on Apaf-1 CARD and RICK CARD on procaspase-9 CARD structures respectively. In this model, the acidic α-helices 2 and 3 of NOD1 CARD form an antiparallel four-helix bundle with the basic α-helices 1 and 4 of RICK CARD. Five acidic residues of NOD1 CARD (D42, D48, E53, D54 and E56) are found at the putative interaction surface. Of these E53 and E56 are surface exposed, whereas D42 and D54 make intra-molecular interactions and are not in an optimal configuration for interacting with RICK CARD. However in our structure of NOD1 CARD, the side-chains of D42, E53 and E56 show significantly higher RMSDs than all other side-chains of residues belonging to helices 2 and 3 suggesting that they could flexibly adapt during complex formation. D42 and E56 correspond to D27 and E41 from the acidic patch of Apaf-1 and D27 and neighbouring E40 are necessary for Apaf-1 binding to procaspase-9 [19]. Four basic residues of RICK CARD (R444, K480, R483 and R488) are found at the putative interaction surface. Of these, R444 and R488 are conserved basic residues in the CARD family corresponding to R13 and R56 in procaspase-9 CARD, two residues known to be indispensable for Apaf-1 binding [19].

To test our hypothesis, we first tried to express and purify RICK CARD with the aim of characterising the NOD1-RICK CARD-CARD interaction *in vitro*. The least unsatisfactory construct was of residues 432 to 540 but the expression level was low and the resulting protein was extremely poorly soluble, tending to aggregate. Thus it was not possible to obtain enough soluble RICK CARD to try characterising the CARD-CARD by biochemical and biophysical experiments e.g. by NMR chemical shift mapping experiments.

Identification of critical residues in the NOD1 and RICK CARDs involved in NOD1/RICK interaction and in NOD1 signalling

We therefore turned to mutational analysis using full-length proteins to define NOD1 and RICK CARD residues functionally important for binding and signalling. We constructed and analysed 17 mutants in full-length NOD1 protein and 8 mutants in full-length RICK protein (Figure 5a). All mutations were in the CARD domains with several of them concerning the exposed and/or conserved acidic residues for NOD1 or basic residues for RICK mentioned above.

Immuno-precipitation experiments

To determine which NOD1 mutants still interact with RICK, we carried out immuno-precipitation experiments in which NOD1 and RICK proteins were transiently co-expressed in HEK293T cells (Figure 5b). Immuno-blotting experiments revealed that all NOD1 mutants were expressed as levels similar to those of wild-type NOD1. Immuno-precipitation analysis showed that RICK protein interacts with NOD1 point mutants D48K, I57A and K67A as efficiently for wild-type showing that the CARD domain is robust to certain mutations, presumably not involve in the interface. Three NOD1 point mutants E53K, D54K and E56K, all exposed residues in the major acidic patch of the CARD domain, did not co-immuno-precipitate with RICK (Figure 5b). These residues thus play a crucial role in mediating the interaction between NOD1 and RICK, probably by promoting the CARD-CARD binding. Consistent with this, the NOD1 D54A-E56K-I57A triple mutant does not bind wild-type RICK protein. NOD1 R69E mutant, in the basic patch of NOD1, shows a greatly reduced ability to interact with RICK (Figure 5b). In our structure, R69 forms a salt-bridge with D73, this likely being a critical stabilising feature of the CARD fold as both residues are highly conserved amongst members of the CARD family (Figure 2c). Thus it is possible that the mutation R69E does not affect directly the interaction with RICK, but rather destabilises the overall protein fold. That the basic patch of NOD1 CARD is unlikely to be involved in the RICK interaction is supported by the finding that the K67A mutation does not affect RICK binding. Our results show that L40A, D42K and L44A single point mutants are able to bind RICK like wild-type and V41A with reduced affinity. However double mutants L40A-D42K and D42K-L44A and the triple mutant L40A-V41A-D42K exhibit an impaired ability to interact with RICK. These four residues form a cluster adjacent to the acidic residues E53K, D54K and E56K (Figure 7a) which, as the mutational effects are only exerted in combination, likely forms a peripheral part of the interaction surface. We note that the NOD1 mutant V41Q shows a good ability to bind RICK whereas it has been previously reported to abolish the interaction [8]. This difference may be explained by the fact that in this study conditions were used that give more sensitivity for detection of slightly impaired interactions. In the previous study, the binding of the wild-type NOD1 protein appeared itself very weakly and the binding of V41Q may not have been detectable in the conditions used.

We also investigated which residues of the CARD of RICK are important for binding to NOD1, by co-transfecting HEK293T cells with wild-type NOD1 and mutant RICK proteins. Mutant proteins were expressed at levels similar to or slightly lower than those of wild-type RICK. RICK mutants I469A and E500A were used as positive controls for the immuno-precipitation experiments as they show a binding ability similar to that of the wild-type RICK protein. We found that RICK mutants R444E, K483E and R488E are unable to bind NOD1.

We conclude from these studies that the basic residues R444, K483 and R488 of RICK CARD and acidic residues E53, D54 and E56 from NOD1 CARD are key residues required for the NOD1-RICK interaction, suggesting that they could mediate an electrostatic interaction between the two CARDs.

NF-κB activation Assays

We also identified amino acid residues in NOD1 CARD that are functionally important for NOD1 signalling by determining the ability of certain NOD1 mutants to activate the transcription factor NF-κB using a luciferase reporter NF-κB assay (Figure 5d). Control NOD1 mutants, D48K, A52S and K67A, unimpaired in their interaction with RICK, activate NF-κB as for wild-type NOD1 protein. On the other hand, mutants displaying an impaired ability to bind RICK, such as E53K, D54K and E56K, fail to activate NF-κB. Interestingly, L44A, I57A and V41Q mutant NOD1 proteins, which have been shown to bind RICK, exhibit a greatly reduced activation of the signalling (Figure 5d). L44 and I57 are two solvent-accessible hydrophobic residues that may mediate hydrophobic interactions with other residues in RICK CARD. We hypothesize that the identified acidic and basic residues in NOD1 and RICK CARD are essential for the binding but that these hydrophobic residues, required for signalling, may ensure the correct relative alignment of the domains or possibly bind to other regions of RICK.

Discussion

Previous yeast two-hybrid and co-immunoprecipitation experiments have shown that the interaction between NOD1 and RICK is dependent on their CARD domains [7; 8]. The smallest regions shown to interact by these experiments were 1-145 of NOD1 with 435-540 of RICK. We have determined by high-resolution NMR spectroscopy the three-dimensional solution structure of a construct comprising residues 15-138 of NOD1. We find that residues 15-110 form a well-defined structure (although with some heterogeneity towards the C-terminal end of this

region) with a CARD-like fold whilst residues 111-138 are unstructured in solution. Although the NOD1 CARD is structurally homologous to other known CARD structures, with its Greek key six-helix bundle fold, it presents some unique features, notably the length and orientation of helix-6. In the case of the well-characterised CARD-CARD interaction between Apaf-1 and procaspase-9, interactions between oppositely charged surfaces are important for binding [19]. The NMR structure shows that the surface of NOD1 exhibits distinct acidic and basic patches and this is also probably true for the RICK CARD, based on a homology model. We have therefore performed mutations in the CARD domains of NOD1 and RICK in order to test the hypothesis that residues in these patches are important for the CARD-CARD interaction and for signalling.

Our mutagenesis and co-immuno-precipitation experiments on full-length proteins show that complementary sets of charged residues, acidic residues from CARD helices 2 and 3 for NOD1 and basic residues helices 1 and 4 from RICK, are crucial for the NOD1/RICK CARD-CARD interaction, with certain hydrophobic residues playing a secondary, but still important role, particularly for efficient signalling. This suggests that the NOD1-RICK CARD-CARD interaction is similar in nature to that between the CARDs of procaspase-9 and Apaf-1, in having an important electrostatic component. However, the particular residues employed in each case do not necessarily correspond (Figure 7). R444 and R488, two basic residues of RICK CARD necessary for the interaction with NOD1 CARD, correspond to R13 and R56, in procaspase-9 that were shown to be indispensable for the interaction with Apaf-1 CARD [19](Figure 6; Figure 7b and 7d). On the other hand, R483, the third RICK basic residue necessary for the interaction with NOD1 CARD corresponds to R51 in procaspase-9 (Figure 6), which however is not involved in the interaction between Apaf-1 and procaspase-9. Similarly, the important acidic residues we have identified in the NOD1 CARD differ from those in Apaf-1 CARD. For example, Apaf-1 D27 which is necessary for the interaction with procaspase-9 [19], corresponds to NOD1 D42 whose mutation does not affect RICK binding nor NF-κB activity as shown by our co-immuno-precipitation and activity experiments. NOD1 D54 and E56 residues, which are necessary for the interaction with RICK, correspond to E39 and E41 residues in Apaf-1 CARD. E39 and E41 are not involved in the interaction with procaspase-9 CARD. In addition, we have also identified several NOD1 CARD residues important for activation of the NF-κB signalling pathway. These include not only the three acidic residues important for binding to RICK, but also two hydrophobic residues, L44 and I57, located in or close to the acidic patch of NOD1, but which did not affect co-immunoprecipitation with RICK. These could be required for the precise formation of a functional CARD-CARD interaction, although contributing less to the binding, or they could interact with other regions of RICK than the CARD domain or even other components of the signalling pathway, not all of whose components have been identified [40]. In this connection, we note that the recent crystal structure of the CARD and NOD domains of Apaf-1 in its inactive, ADP bound form has given the first structural insight into the activation mechanism of NOD proteins [41]. The Apaf-1 CARD domain is observed to pack against the NOD domain using the same interaction surface as previously shown to interact with the CARD domain of procaspase-9, the downstream effector. Thus, by analogy, correct functioning of NOD1 probably requires that its CARD domain interacts with other parts of NOD1 in the inactive form and with RICK in the ligand activated state. Clearly more structural studies are required on intact NOD1 and in complex with RICK to more precisely define the conformations and interactions involved.

Experimental procedures

Cloning and Protein Expression of NOD1 CARD
The DNA coding for residues 15 to 138 of the NOD1 protein was cloned in the pETM-11 vector (EMBL) between the *Nco* I and *Kpn* I restriction sites. This vector fuses a cleavable amino terminal six residue histidine tag to the expressed protein. NOD1 CARD was expressed in *Escherichia coli* BL21 Star (DE3) (Invitrogen). The transformed cells were grown either in LB medium for unlabeled NOD1 CARD or M9 minimal medium supplemented with $^{15}NH_4Cl$ (1 g l^{-1}) for ^{15}N-labeled NOD1 CARD and $^{15}NH_4Cl$ (1 g l^{-1}) and [$^{13}C_6$]-glucose (2 g l^{-1}) for ^{15}N-^{13}C-labeled NOD1 CARD. The transformed cells were grown at 37°C and protein expression was induced with 1 mM isopropyl-D-thiogalactoside during 5 hours at 30°C. The degree of isotope labelling of the protein after purification was verified to be higher than 98 % using MALDI-TOF mass spectrometry.

Sample Preparation
After cell lysis the 6 his-tag recombinant protein was purified using a Ni-NTA affinity column (Qiagen) at pH 7.5. The his-tag was removed by overnight TEV protease cleavage at 4°C. The protein was then run over a fresh Ni-NTA column to remove any uncleaved protein aswell as the his-tagged TEV protease. The eluted NOD1 CARD was further purified to homogeneity using size-exclusion FPLC (Superdex®75; Amersham Biosciences) in 25 mM sodium-phosphate pH 7, 100 mM NaCl and 5 mM DTT, a buffer suitable for NMR. The protein was concentrated to approximately 0.5 or 0.8 mM for the unlabeled NOD1 CARD and 1.4 mM for the ^{15}N and ^{15}N-^{13}C-labeled NOD1 CARD used for NMR experiments.

NMR Spectroscopy

All NMR spectra recorded for the structure determination were acquired at 30°C on Varian INOVA 800 or Varian INOVA 600 spectrometers. For sequential backbone assignment, six standard triple-resonance experiments were performed on the ^{15}N-^{13}C-labeled protein in 95 % H_2O/5 % D_2O: HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH and HNCA. An additional 3D ^{1}H-^{15}N-edited NOESY-HSQC spectrum was performed on a ^{15}N-labeled sample of NOD1 CARD in order to identify HN-HN contacts. ^{1}H and ^{13}C aliphatic side-chain assignments were accomplished using a set of five 3D triple-resonance experiments: HBHA(CO)NH, CC(CO)NH, HC(CO)NH and H(C)CH-TOCSY in 95 % H_2O/5 % D_2O plus H(C)CH-TOCSY in 100 % D_2O. The assignment of aromatic side-chain protons was performed using 1H homonuclear 2D NOESY acquired with the unlabeled protein in 95 % H_2O/5 % D_2O and in 100 % D_2O, and with 3D ^{13}C-edited NOESY-HSQC with the ^{13}C carrier frequency in the aromatic region acquired using the ^{15}N-^{13}C-labeled protein in 95 % H_2O/5 % D_2O and in 100 % D_2O.

Distance restraints were derived from 3D ^{15}N-edited NOESY-HSQC, 3D ^{13}C-edited NOESY-HSQC and 3D methyl selected NOESY-HSQC experiments [42]. A mixing-time of 100 ms was used for all NOESY experiments. An H-D exchange experiment was performed at 800 MHz by re-dissolving a lyophilized ^{15}N-labeled sample of NOD1 CARD in D_2O and following the loss of intensity of labile protons by recording a series of ^{15}N-edited HSQC spectra over 19 hours. Amide protons were considered to be involved in a hydrogen bond if signal remained visible three hours after the exchange.

All spectra were processed with NMRPipe [43] and analyzed using NMRView [44; 45]. Smartnotebook [46] was used for sequential NMR resonance assignment.

Structure Calculation

Only residues 15 to 110 were included in the structure calculations, the remaining part of the protein being more flexible as shown by ^{15}N relaxation analysis and lacking any NOE correlations. Assignments of unambiguous NOEs of the ^{15}N-edited NOESY-HSQC and the ^{13}C-edited NOESY-HSQC experiments were done manually. Dihedral angle restraints were generated using TALOS [47] based on secondary chemical shifts. Hydrogen bonding restraints were deduced from backbone chemical shifts and from the H-D exchange experiment analysis. An initial calculation of the structure of NOD1 CARD was performed with the program CNS [48] using an extended conformation of the protein and the aforementioned structural restraints. NOE intensities were classified as strong (< 2.8 Å), medium (< 3.4 Å) and weak (< 5 Å). 1000 structures were calculated and the best in term of energy was used as a starting point for automated NOE assignment of the remaining peaks using the program aria2.0 [49]. The program MOLMOL [50] was used for checking and displaying violations of NMR restraints. The final structure calculation was performed following simulated annealing protocols with CNS [51] and using an extended conformation of NOD1 CARD. 1000 structures were calculated and the 40 lowest energy structures were refined in explicit water using CNS. The 10 lowest energy structures obtained were analyzed and validated in terms of geometry and restraints violations with the AQUA and PROCHECK-NMR programs [32]. Structure representations were made with the program GRASP [52], PYMOL (http://www.pymol.org/) and MOLMOL [50].

^{15}N Relaxation Experiments

All relaxation experiments were performed at 30° on a Varian INOVA 600 spectrometer using ^{15}N-labeled samples with a concentration of approximately 0.7 mM. ^{15}N R_1 and heteronuclear ^{15}N{^{1}H}-NOE values were determined as previously described [53]. R_1 times were extracted from 14 spectra with different values for the relaxation delay: 0 (2x), 55, 111, 166, 222, 333, 444, 666, 888, 1110 (2x), 1554 and 1998 ms, applying 180° pulses on protons every 5 ms to suppress CSA(^{15}N)/DD(HN) cross-correlated relaxations. ^{15}N R_2 relaxation times were extracted from $R_{1\rho}$ experiments. The $R_{1\rho}$ experiments were recorded with spin-lock pulses of varying lengths 2, 6 (2x), 10, 20, 30, 40, 60, 80 (2x), 100, 120, 140 and 160. Relaxation parameters were extracted with the program Curvefit, using a two-parameter fitting and a Monte Carlo simulation to estimate the errors. In the $R_{1\rho}$ experiments, a spin lock field of $|\gamma_N B_1|/2\pi$ = 1.6 kHz was applied to the ^{15}N spins to suppress chemical shift and scalar J_{NH} coupling evolution. Transverse relaxation rate constants R_2 were then calculated from the measured $R_{1\rho}$ rate constants, taking into account the frequency offset $\Delta\nu$, using the equation: $R_{1\rho} = \cos^2(\theta)R_2+\sin^2(\theta)R_1$ where $\theta = \tan^{-1}(2\pi\Delta\nu/\gamma_N B_1)$.

The ^{15}N heteronuclear relaxation rates were interpreted using the program TENSOR2 [54; 55] which uses the description of molecular diffusion derived by Woessner [36], in combination with the Lipari-Szabo model-free analysis of local flexibility [34]. In the model-free approach, internal mobility is characterized using an order parameter S^2, which may be interpreted as the amplitude of the motion and a correlation time, τi, the characteristic time constant of this motion. The inter-nuclear distance rNH was assumed to average to 1.01, and the chemical shift anisotropy of the ^{15}N nucleus ($\sigma_{//}$ - $\sigma_{\perp}$) was approximated to -170 ppm.

Sequence Alignments and Homology Modelling

The amino acid sequences of all proteins were obtained from the National Center for Biotechnology Information database. The sequence alignment in Figure 2c was done using ClustalW [56]. Coordinates from the CARDs of Apaf-1,

procaspase-9, and Iceberg were obtained from the RCSB Protein Data Bank. Modelling of the structure of the RICK CARD was performed with the program Modeller8v1 [57] using the atomic coordinates of procaspase-9 CARD.

Mutagenesis of NOD1 and RICK

Expression plasmid pcDNA3-N-terminal HA-tagged NOD1 (pcDNA3-NOD1-HA) [8] and pcDNA3-N-terminal FLAG-tagged RICK (pcDNA3-RICK-FLAG) [58] were mutagenised using the QuickChange XL site-directed mutagenesis kit (Stratagene, La Jolla, CA) to generate single, double and triple amino acid mutants in the CARD domains of NOD1 or RICK. NOD1 mutants included L40A, V41A, V41Q, D42K, L44A, D48K, A52S, E53K, D54K, E56K, I57A, K67A, R69E, L40A-D42K, D42K-L44A, L40A-V41A-D42K, D54A-E56K-I57A and RICK mutants were R444E, V448A, I469A, K480E, P481A, R483E, R488E, and E500A. The entire coding sequences of all NOD1 and RICK clones were verified by DNA sequencing.

Immuno-precipitation and immuno-blotting

HEK293T cells were co-transfected using the calcium-phosphate method with either pcDNA3-NOD1-HA mutants and wild-type pcDNA3-RICK-FLAG or with wild-type pcDNA3-NOD1-HA and pcDNA3-RICK-FLAG mutants as described [8]. In each case, three controls were performed: co-transfection with wild type pcDNA3-NOD1-HA and wild-type pcDNA3-RICK-FLAG, or co-transfection with wild type pcDNA3-NOD1-HA and empty pcDNA3 or wild-type pcDNA3-RICK-FLAG and empty pcDNA3. Empty pcDNA3 was used to keep constant the amount of DNA used. HEK293T cells were harvested 20 hours after transfection and lysed in 0.2 % Nonided P-40 lysis buffer [59]. Lysates were incubated with mouse monoclonal anti-FLAG antibody (Sigma) and then immuno-precipitations were carried out using Rec-protein G-Sepharose®4B conjugate (Zymed). Total lysates and immuno-precipitated products were resolved by 10 % SDS-PAGE gels, and immunoblotted with rabbit anti-HA antibody (Santa Cruz Biotechnology) to detect NOD1 proteins or mouse anti-FLAG antibody (Sigma) to detect RICK proteins. A reference extract from HEK293T cells transfected with the wild-type NOD1 or RICK cDNA were included in all immuno-blotting experiments to compare the expression of mutant and wild-type NOD1 and RICK proteins.

NF-κB Activation Assays

HEK293T cells ($4x10^4$) were co-transfected with LipofectAMINE Plus transfection reagent (Invitrogen) using 5 ng pBVIx-luc [60], 10 ng pCMV-βgal [8], 400 ng of pcDNA3 plasmid for the control or 5 ng of wild-type or mutant NOD1 plasmid plus 395 ng pcDNA3 to keep constant the total amount of transfected DNA. 20 hours post-transfection, cell extract was prepared and β-galactosidase and luciferase activity (Promega) were determined as described [8]. Results were normalized for transfection efficiency with values obtained with pCMV-βgal. NK-κB assays were performed in triplicate.

RCSB Protein Data Bank Accession Number

The NOD1 CARD coordinates and NMR data have been deposited with the PDB accession number **2B1W** and the BMRB access number **6843**.

Supplementary Data

^{15}N relaxation data and dynamic parameters of NOD1 CARD are available online at http://www.sciencedirect.com.

Acknowledgments

F.M. received a grant Emergence 2002 from the Région Rhône-Alpes and was also supported by the European Molecular Biology Laboratory International PhD programme.

References

1. Inohara, N. & Nunez, G. (2003). NODs: intracellular proteins involved in inflammation and apoptosis. *Nat Rev Immunol* **3**, 371-82.
2. Chamaillard, M., Girardin, S. E., Viala, J. & Philpott, D. J. (2003). Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation. *Cell Microbiol* **5**, 581-92.
3. Kufer, T. A., Fritz, J. H. & Philpott, D. J. (2005). NACHT-LRR proteins (NLRs) in bacterial infection and immunity. *Trends Microbiol*.
4. Inohara, N. & Nunez, G. (2001). The NOD: a signaling module that regulates apoptosis and host defense against pathogens. *Oncogene* **20**, 6473-81.
5. Girardin, S. E., Boneca, I. G., Carneiro, L. A., Antignac, A., Jehanno, M., Viala, J., Tedin, K., Taha, M. K., Labigne, A., Zahringer, U., Coyle, A. J., DiStefano, P. S., Bertin, J., Sansonetti, P. J. & Philpott, D. J. (2003). Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* **300**, 1584-7.
6. Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S., Valvano, M. A., Foster, S. J., Mak, T. W., Nunez, G. & Inohara, N. (2003). An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nat Immunol* **4**, 702-7.
7. Bertin, J., Nir, W. J., Fischer, C. M., Tayber, O. V., Errada, P. R., Grant, J. R., Keilty, J. J., Gosselin, M. L., Robison, K. E., Wong, G. H., Glucksmann, M. A. & DiStefano, P. S. (1999). Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-kappaB. *J Biol Chem* **274**, 12955-8.

8. Inohara, N., Koseki, T., del Peso, L., Hu, Y., Yee, C., Chen, S., Carrio, R., Merino, J., Liu, D., Ni, J. & Nunez, G. (1999). Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB. *J Biol Chem* **274**, 14560-7.
9. Yoo, N. J., Park, W. S., Kim, S. Y., Reed, J. C., Son, S. G., Lee, J. Y. & Lee, S. H. (2002). Nod1, a CARD protein, enhances pro-interleukin-1beta processing through the interaction with pro-caspase-1. *Biochem Biophys Res Commun* **299**, 652-8.
10. Viala, J., Chaput, C., Boneca, I. G., Cardona, A., Girardin, S. E., Moran, A. P., Athman, R., Memet, S., Huerre, M. R., Coyle, A. J., DiStefano, P. S., Sansonetti, P. J., Labigne, A., Bertin, J., Philpott, D. J. & Ferrero, R. L. (2004). Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nat Immunol* **5**, 1166-74.
11. Weidinger, S., Klopp, N., Rummler, L., Wagenpfeil, S., Novak, N., Baurecht, H. J., Groer, W., Darsow, U., Heinrich, J., Gauger, A., Schafer, T., Jakob, T., Behrendt, H., Wichmann, H. E., Ring, J. & Illig, T. (2005). Association of NOD1 polymorphisms with atopic eczema and related phenotypes. *J Allergy Clin Immunol* **116**, 177-84.
12. Hysi, P., Kabesch, M., Moffatt, M. F., Schedel, M., Carr, D., Zhang, Y., Boardman, B., von Mutius, E., Weiland, S. K., Leupold, W., Fritzsch, C., Klopp, N., Musk, A. W., James, A., Nunez, G., Inohara, N. & Cookson, W. O. (2005). NOD1 variation, immunoglobulin E and asthma. *Hum Mol Genet* **14**, 935-41.
13. Hofmann, K., Bucher, P. & Tschopp, J. (1997). The CARD domain: a new apoptotic signalling motif. *Trends Biochem Sci* **22**, 155-6.
14. Weber, C. H. & Vincenz, C. (2001). The death domain superfamily: a tale of two interfaces? *Trends Biochem Sci* **26**, 475-81.
15. Fairbrother, W. J., Gordon, N. C., Humke, E. W., O'Rourke, K. M., Starovasnik, M. A., Yin, J. P. & Dixit, V. M. (2001). The PYRIN domain: a member of the death domain-fold superfamily. *Protein Sci* **10**, 1911-8.
16. Bouchier-Hayes, L. & Martin, S. J. (2002). CARD games in apoptosis and immunity. *EMBO Rep* **3**, 616-21.
17. Vaughn, D. E., Rodriguez, J., Lazebnik, Y. & Joshua-Tor, L. (1999). Crystal structure of Apaf-1 caspase recruitment domain: an alpha-helical Greek key fold for apoptotic signaling. *J Mol Biol* **293**, 439-47.
18. Zhou, P., Chou, J., Olea, R. S., Yuan, J. & Wagner, G. (1999). Solution structure of Apaf-1 CARD and its interaction with caspase-9 CARD: a structural basis for specific adaptor/caspase interaction. *Proc Natl Acad Sci U S A* **96**, 11265-70.
19. Qin, H., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S. & Shi, Y. (1999). Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature* **399**, 549-57.
20. Chou, J. J., Matsuo, H., Duan, H. & Wagner, G. (1998). Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. *Cell* **94**, 171-80.
21. Humke, E. W., Shriver, S. K., Starovasnik, M. A., Fairbrother, W. J. & Dixit, V. M. (2000). ICEBERG: a novel inhibitor of interleukin-1beta generation. *Cell* **103**, 99-111.
22. Ogura, Y., Inohara, N., Benito, A., Chen, F. F., Yamaoka, S. & Nunez, G. (2001). Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. *J Biol Chem* **276**, 4812-8.
23. Stehlik, C., Hayashi, H., Pio, F., Godzik, A. & Reed, J. C. (2003). CARD6 is a modulator of NF-kappa B activation by Nod1- and Cardiak-mediated pathways. *J Biol Chem* **278**, 31941-9.
24. Thome, M., Hofmann, K., Burns, K., Martinon, F., Bodmer, J. L., Mattmann, C. & Tschopp, J. (1998). Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. *Curr Biol* **8**, 885-8.
25. McCarthy, J. V., Ni, J. & Dixit, V. M. (1998). RIP2 is a novel NF-kappaB-activating and cell death-inducing kinase. *J Biol Chem* **273**, 16968-75.
26. Ruefli-Brasse, A. A., Lee, W. P., Hurst, S. & Dixit, V. M. (2004). Rip2 participates in Bcl10 signaling and T-cell receptor-mediated NF-kappaB activation. *J Biol Chem* **279**, 1570-4.
27. Druilhe, A., Srinivasula, S. M., Razmara, M., Ahmad, M. & Alnemri, E. S. (2001). Regulation of IL-1beta generation by Pseudo-ICE and ICEBERG, two dominant negative caspase recruitment domain proteins. *Cell Death Differ* **8**, 649-57.
28. Lee, S. H., Stehlik, C. & Reed, J. C. (2001). Cop, a caspase recruitment domain-containing protein and inhibitor of caspase-1 activation processing. *J Biol Chem* **276**, 34495-500.
29. Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., Achkar, J. P., Brant, S. R., Bayless, T. M., Kirschner, B. S., Hanauer, S. B., Nunez, G. & Cho, J. H. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603-6.
30. Miceli-Richard, C., Lesage, S., Rybojad, M., Prieur, A. M., Manouvrier-Hanu, S., Hafner, R., Chamaillard, M., Zouali, H., Thomas, G. & Hugot, J. P. (2001). CARD15 mutations in Blau syndrome. *Nat Genet* **29**, 19-20.
31. Damiano, J. S. & Reed, J. C. (2004). CARD proteins as therapeutic targets in cancer. *Curr Drug Targets* **5**, 367-74.
32. 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**, 477-86.
33. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-38.
34. Lipari, G. & Szabo, A. (1982). Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4459-4570.
35. Lipari, G. & Szabo, A. (1982). Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559.
36. Woessner, D. E. (1962). Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys* **37**, 647-654.
37. Hu, Y., Ding, L., Spencer, D. M. & Nunez, G. (1998). WD-40 repeat region regulates Apaf-1 self-association and procaspase-9 activation. *J Biol Chem* **273**, 33489-94.
38. Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. & Alnemri, E. S. (1998). Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol Cell* **1**, 949-57.
39. Tanabe, T., Chamaillard, M., Ogura, Y., Zhu, L., Qiu, S., Masumoto, J., Ghosh, P., Moran, A., Predergast, M. M., Tromp, G., Williams, C. J., Inohara, N. & Nunez, G. (2004). Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. *Embo J* **23**, 1587-97.
40. Strober, W., Murray, P. J., Kitani, A. & Watanabe, T. (2006). Signalling pathways and molecular interactions of NOD1 and NOD2. *Nat Rev Immunol* **6**, 9-20.
41. Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R. & Shi, Y. (2005). Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature* **434**, 926-33.
42. Van Melckebeke, H., Simorre, J. P. & Brutscher, B. (2004). Amino acid-type edited NMR experiments for methyl-methyl distance measurement in 13C-labeled proteins. *J Am Chem Soc* **126**, 9584-91.
43. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
44. Bruce A. Johnson, R. A. B., Volume 4, Issue 5, Sep 1994, Pages 603 - 614. (1994). NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR (Historical Archive)* **4**.
45. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
46. Slupsky, C. M., Boyko, R. F., Booth, V. K. & Sykes, B. D. (2003). Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments. *J Biomol NMR* **27**, 313-21.
47. 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.
48. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-21.
49. Habeck, M., Rieping, W., Linge, J. P. & Nilges, M. (2004). NOE assignment with ARIA 2.0: the nuts and bolts. *Methods Mol Biol* **278**, 379-402.
50. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-5, 29-32.
51. Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* **229**, 317-24.
52. Doreleijers, J. F., Rullmann, J. A. & Kaptein, R. (1998). Quality assessment of NMR structures: a statistical survey. *J Mol Biol* **281**, 149-64.
53. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984-6003.
54. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* **16**, 23-8.
55. Tsan, P., Caffrey, M., Daku, M., Cusanovich, M., Marion, D. & Gans, P. (2001). Magnetic susceptibility tensor and heme contact shifts determinations in the Rhodobacter capsulatus ferricytochrome c':NMR and magnetic susceptibility studies. *J Am Chem Soc* **123**, 2231-42.
56. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.
57. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815.

58. Inohara, N., del Peso, L., Koseki, T., Chen, S. & Nunez, G. (1998). RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. *J Biol Chem* **273**, 12296-300.
59. Inohara, N., Ogura, Y., Chen, F. F., Muto, A. & Nunez, G. (2001). Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *J Biol Chem* **276**, 2551-4.
60. Inohara, N., Koseki, T., Lin, J., del Peso, L., Lucas, P. C., Chen, F. F., Ogura, Y. & Nunez, G. (2000). An induced proximity model for NF-kappa B activation in the Nod1/RICK and RIP signaling pathways. *J Biol Chem* **275**, 27823-31.

Figure Legends

Figure 1. NMR Spectra of NOD1 CARD.
(a) ^{1}H-^{15}N HSQC correlated spectrum recorded at 800 MHz with assignments.
(b) ^{1}H-^{13}C HSQC correlated spectrum in the region of W103 recorded at 600 MHz. Both spectra show the presence of multi-conformers of the protein.

Figure 2. NOD1 CARD structure and comparison with other CARD family members.
(a) 10 superimposed NMR-derived structures of NOD1 CARD. These structures are the 10 lowest energy structures obtained after water refinement. Helices are coloured in red. The termini are denoted by the letters N and C.
(b) Stereo view of the backbone atoms of the structure closest to the mean. The polypeptide backbone is drawn in blue. Every tenth residue is denoted by a red ball at the Cα position and numbered. All other Cα atoms are represented by yellow balls. The orientation is the same as in (a).
(c) Sequence alignment of RICK CARD and nine other CARD domains from *Homo sapiens* proteins shown to interact with RICK CARD. Common features are identified by colour and labelling below the sequences. Polar residues are in cyan and labelled "p", hydrophobic residues are in green and labelled "a". The bottom line indicates the consensus sequence (one-letter code of the amino acid, bold): uppercase is identity, lowercase is consensus level > 0.5, # refers to acidic residues Glu or Asp. The stared residues are conserved hydrophobic residues that make up the hydrophobic core of the CARD family. The location of the α-helices in the structure of NOD1 CARD is shown above the alignment.
(d) Structure of NOD1 CARD and other homologous CARD structures. The four structures have the same orientation. Helices 1 to 6 are numbered α1-α6 on NOD1 CARD structure.
(e) Detail of the structural alignment of the N- and C-terminal regions of NOD1 (magenta), Iceberg (gold) and Apaf-1 (blue) CARDs highlighting the major deviations of NOD1 CARD from more prototypical CARDs. Note the additional helix α1" which is antiparallel to the first part of helix 6 (α6') and the extension of the helix 6 (α6") taking the C-terminus of the domain in the opposite direction.

Figure 3. ^{15}N relaxation and backbone dynamics of NOD1 CARD.
(a) Ribbon representation of the backbone of NOD1 CARD colour-coded according to the order parameter values (S^2) determined by the model free approach analysis of the ^{15}N relaxation parameters (R_1, $R_1\rho$, NOE). High S^2 values reflect high rigidity of the NH backbone bond whereas low values reflect more flexible regions. Residues that could not be interpreted because of low peak intensity or spectral overlap are coloured in white. Helices 1 to 6 are numbered α1-α6.
(b) Ribbon representation of the backbone of the NOD1 CARD colour coded in red when the exchange rate parameter (Kex in s^{-1}) determined by the model free approach analysis of the ^{15}N relaxation parameters is present.

Figure 4. Surface representations of NOD1 CARD and RICK CARD. In (a), (b), (d) and (e), the surfaces are color coded according to electrostatic surface potential: red, -10 kT; white, 0kT; and blue, +10 kT. In (c) and (f), surface exposed hydrophobic residues including Leu, Val, Ile, Trp and Phe are coloured in green. Residues identified in magenta correspond to the equivalent position of mutations of hydrophobic residues in NOD2 CARDs that reduce or abolish the interaction with RICK CARD [39]. Residues are coloured in green when they correspond to hydrophobic side chains and cyan when they correspond to non-hydrophobic amino acids in the NOD1 CARD structure.
(a) Basic surface of NOD1 CARD formed by the exposed residues of helices 1 and 4.
(b) The prominent acidic surface of NOD1 CARD formed by the exposed residues of helices 2 and 3 (orientation is rotated by 90° about the vertical axis relative to 4a).
(c) Hydrophobic residues on the acidic surface of NOD1 CARD (same orientation as in b).
(d) Acidic surface of RICK CARD homology model formed primarily by the exposed residues of helices 3 and 5.
(e) Main basic surface of RICK CARD model formed by the exposed residues of helices 1 and 4.
(f) Hydrophobic residues on the basic surface of RICK CARD (same orientation as in e).

Figure 5. Mutational Analyses of full length NOD1 and RICK.
(a) Schematic representation of the CARD domains of NOD1 and RICK. Numbers indicate the position of amino acid residues. Mutants with wild-type activity are depicted in grey. Mutations that abolish NOD1 and RICK interaction are shown in bold black. Mutants with a reduced ability to activate NF-κB are identified in grey italics. Complete loss

of function mutant L44A is labelled with the symbol *.
(b) Interactions between mutant NOD1 proteins and wild-type RICK protein. Extracts from HEK293T cells expressing indicated proteins were immunoprecipitated with anti-FLAG antibody and immunoblotted with both anti-HA antibody to detect NOD1 proteins and anti-FLAG antibody to detect RICK proteins. IP, immunoprecipitation; Lysate, immunoblotting of total lysates. Mutant NOD1 proteins were expressed at levels similar to those of the wild-type protein. NOD1 mutations E53K, D54K, E56K totally abolish the NOD1-RICK interaction. Representative results of mutants with wild-type activity (D48K, I57A and K67A) are shown. (c) Interactions between wild-type NOD1 protein and mutant RICK proteins. Extracts from HEK293T cells expressing indicated proteins were immunoprecipitated with anti-FLAG antibody and immunoblotted with both anti-HA antibody to detect NOD1 proteins and anti-FLAG antibody to detect RICK proteins. IP, immunoprecipitation; Lysate, immunoblotting of total lysates. Mutant RICK proteins express at levels similar or slightly lower to those of the wild-type protein. RICK mutations R44E, R483E and R488E totally abolish the NOD1-RICK interaction. Representative results of control mutants (I469A, E500A) with wild-type activity are shown.
(d) NF-κB activity of mutant NOD1 proteins and wild-type NOD1 protein. NOD1 mutants E53K, D54K, E56K, R69E, L40A-D42K, D42K-L44A, L40A-V41A-D42K, D54K-E56K-I57A inhibite NF-κB activation. Values represent the mean of normalized data ± s.d. of triplicate cultures.

Figure 6. Sequence alignment of the CARD domains of NOD1, Apaf-1, procaspase-9 and RICK. The secondary structure of NOD1 CARD is pictured above the alignment. Red box and white character design strictly identical residues. Red character is used for similar residues. Important residues abolishing NOD1/RICK or Apaf-1/procaspase-9 CARD-CARD interaction are surrounded in purple and the ones weakening the interaction in yellow. Residues whose mutation is correlated with a loss of NF-κB activation ability are surrounded in blue. The secondary structures of Apaf-1, procaspase-9 and the predicted secondary structure of RICK are shown below the alignment.

Figure 7. Comparison of the residues involved in NOD1/RICK and in Apaf-1/procaspase-9 CARD-CARD interaction. Residues shown to be indispensable for the NOD1/RICK or Apaf-1/procaspase-9 CARD-CARD interaction [19] are identified in magenta. Hydrophobic NOD1 residues whose single point alterations affect NF-κB signalling are shown in blue. NOD1 residues whose associated single point mutations affect NF-κB signalling are shown in cyan. Apaf-1 and procaspase-9 amino acids whose mutations weaken the Apaf-1/procaspase-9 CARD-CARD interaction [19] are coloured in yellow.
(a) Surface representation of the acidic patch of NOD1 CARD (same orientation as in figure 4b).
(b) Surface representation of the basic patch of RICK CARD, (same orientation as in figure 4e).
(c) Surface representation of the acidic patch of Apaf-1 CARD (same orientation as NOD1 CARD in a).
(e) Surface representation of the basic patch of procaspase-9 (same orientation as the modelled RICK CARD in b).

Table 1. Structural Statistics for the 10 Structures of Lowest Energy	
NOE distance restraints	
All	1236
Intraresidue	457
Interresidue	779
Sequential (|i – j| = 1)	309
Medium range (2 ≤ |i – j| ≤ 4)	313
Long range ((|i – j| > 4)	157
Hydrogen bonds (deduced from H-D exchange)	82
Dihedral angle restraints (derived from TALOS)	
ϕ	69
ψ	69
Energies after water refinement	
van der Waals, kcal.mol-1	-895 ± 31
Electrostatic, kcal.mol-1	-4022 ± 47
Rmsd from idealized covalent geometry	
Bond lengths, (Å)	0.0045 ± 0.00015
Angles, (°)	0.70 ± 0.02
Improper, (°)	1.67 ± 0.15
Rmsd from the mean coordinates (secondary structures)	
Backbone heavy atoms, (Å)	0.43 ± 0.07
All heavy atoms, (Å)	0.85 ± 0.09
Experimental statistics	
Number of distance violations > 0.5 Å	0.8 (± 0.9)
Number of distance violations > 0.3 Å	3.6 (± 1.1)
Number of distance violations > 0.1 Å	47.2 (± 3.0)
Number of hydrogen bond violations > 0.1 Å	3.3 (± 1.2)
Ramachandran plot statistics (PROCHECK)	
Residues in most favorable regions, %	83.3
Residues in additional allowed regions, %	10.5
Residues in generously allowed regions, %	5.1
Residues in disallowed regions, %	1.1

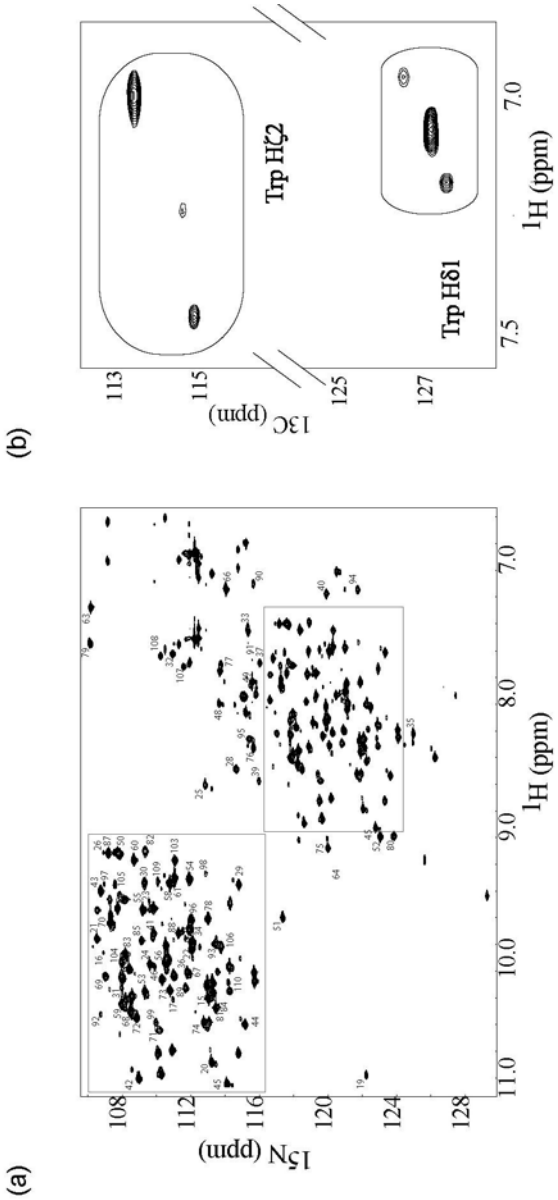

Figure 1

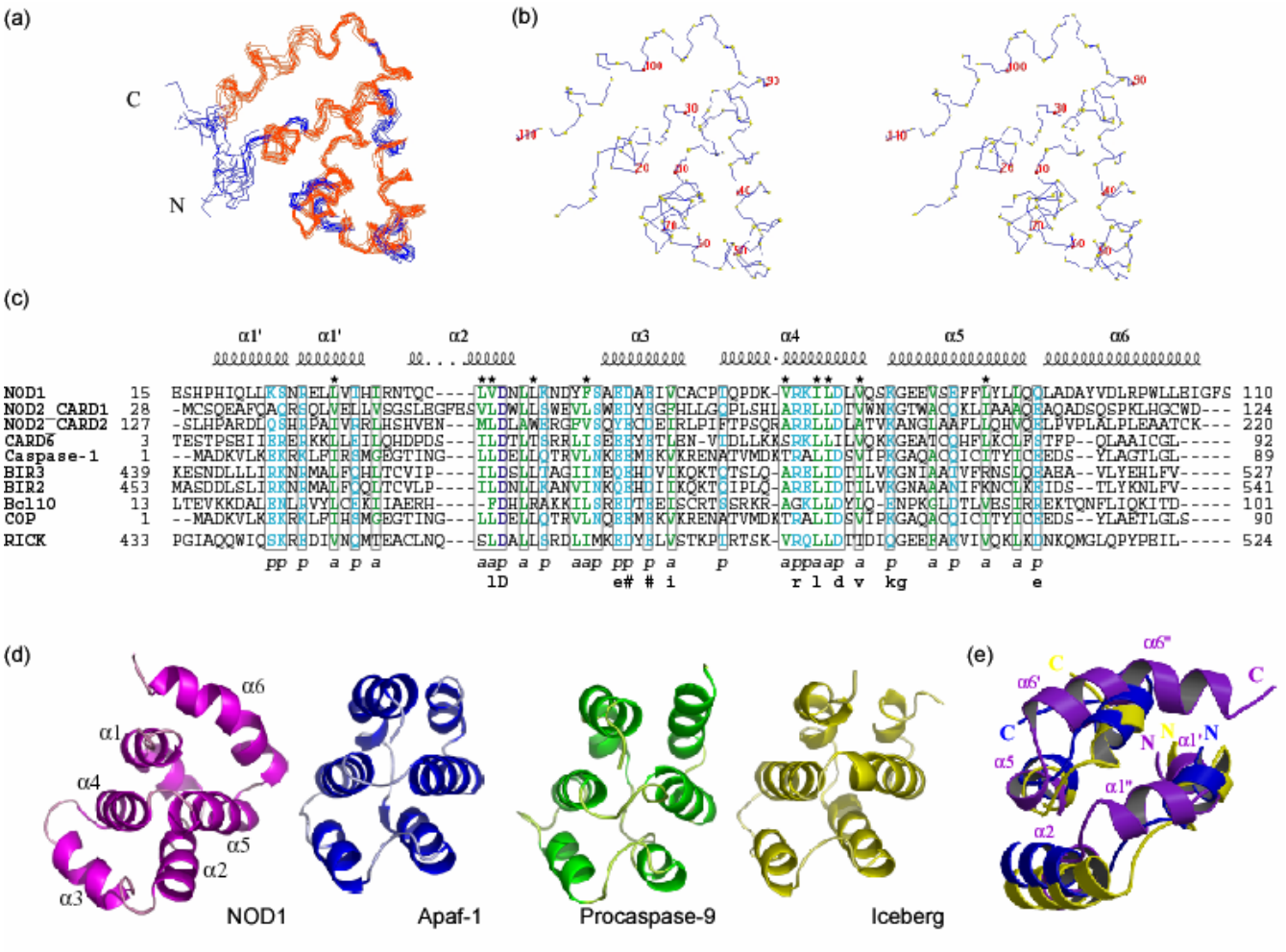

Figure 2

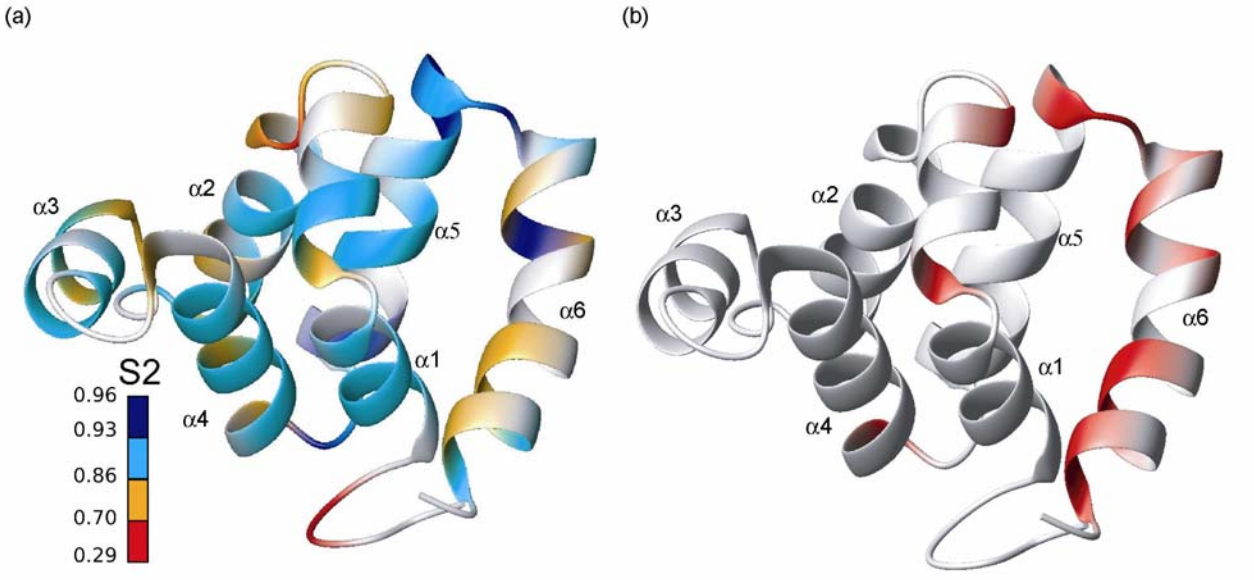

Figure 3

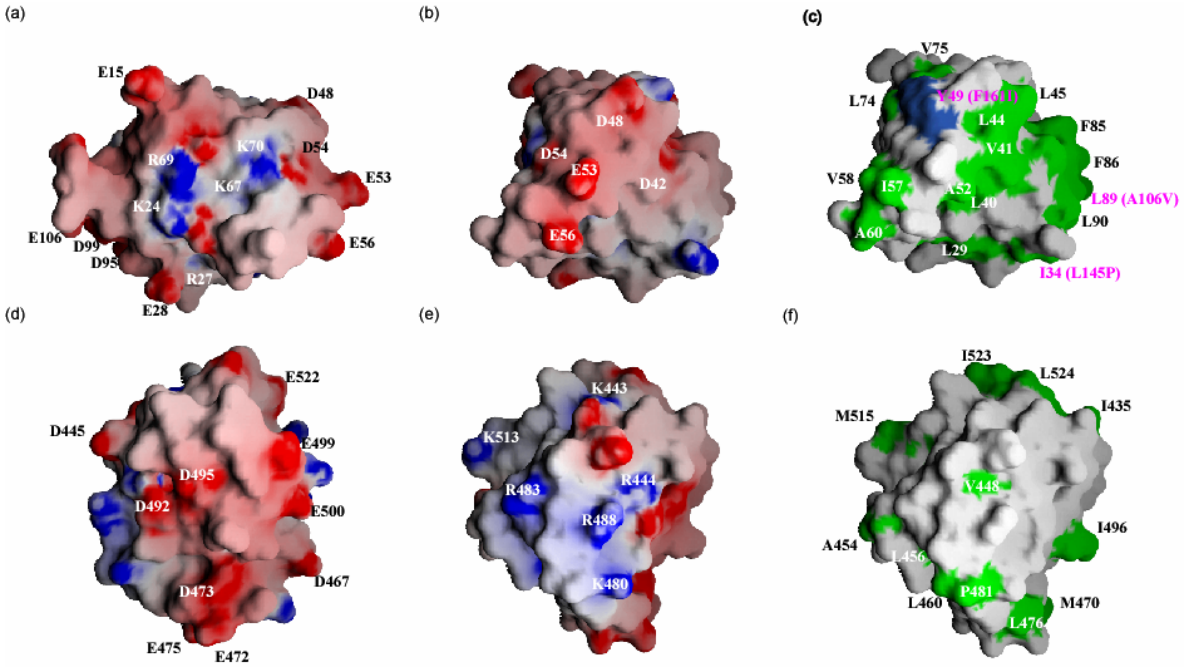

Figure 4

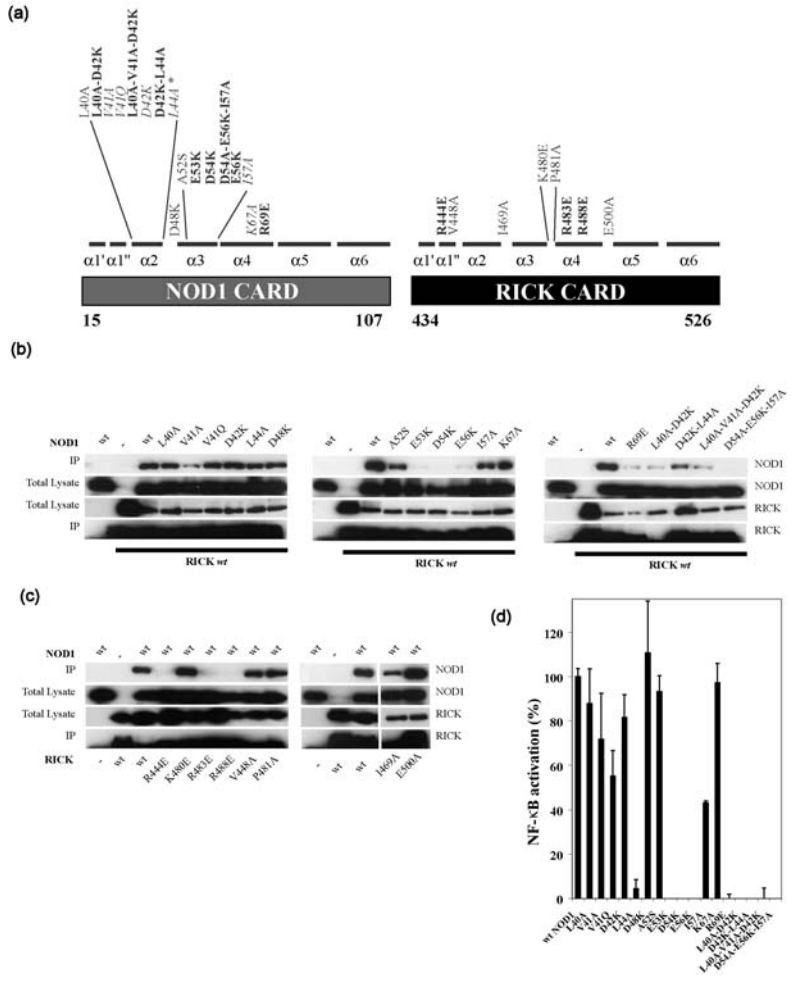

Figure 5

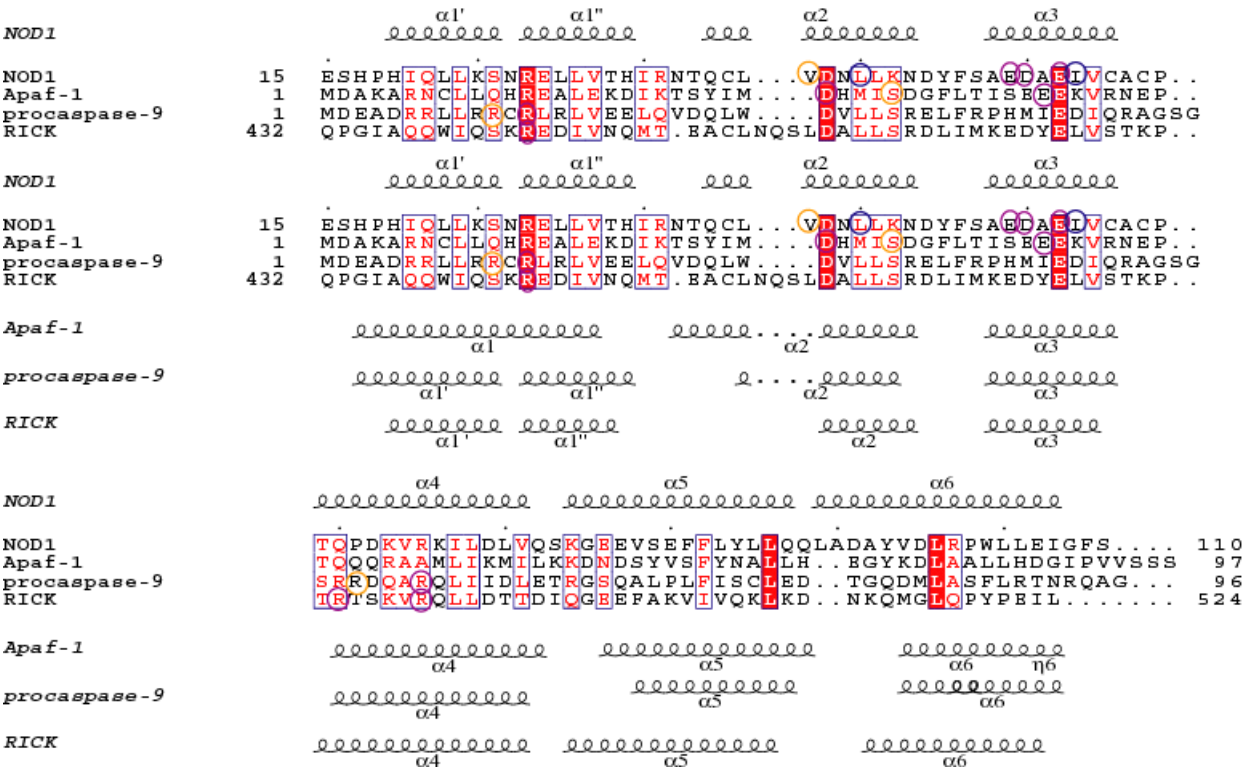

Figure 6

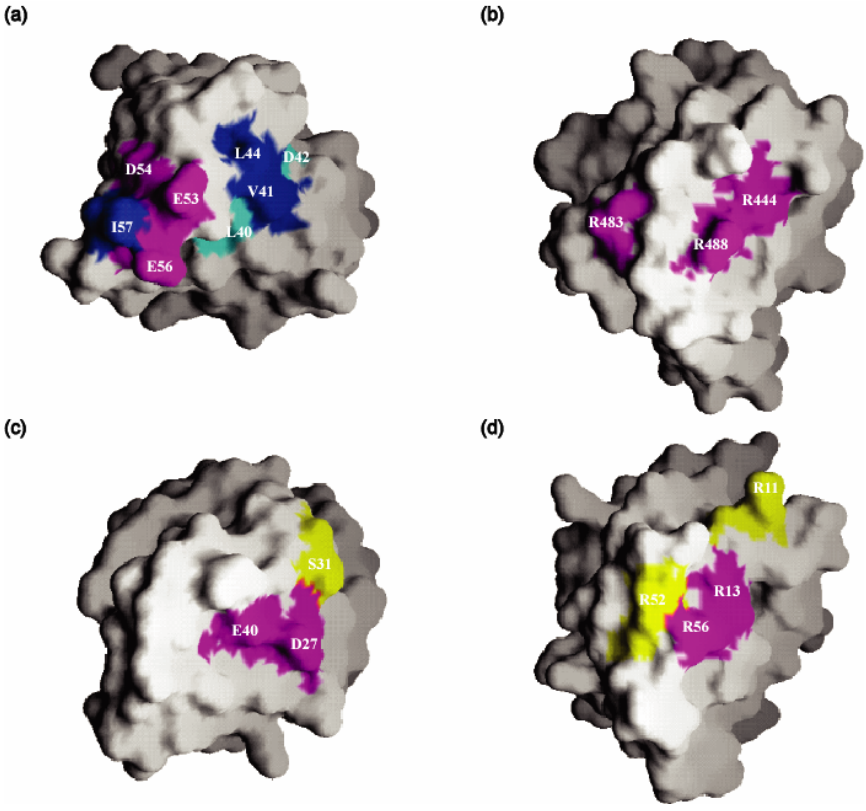

Figure 7

Abstract

NOD1 is a cytosolic signaling host pattern-recognition receptor playing a crucial role in innate immunity. It regulates pro-inflammatory pathways by mediating the activation of the nuclear factor NF-κB *via* its downstream effector RICK following the recognition of a specific bacterial ligand. RICK is recruited by NOD1 through a CARD-CARD interaction. The goal of this thesis was to determine the structural basis of this process. NOD1 CARD was cloned, expressed and purified at the Grenoble EMBL Outstation. Its structure was determined by nuclear magnetic resonance at the IBS in Grenoble. It is generally similar to other CARDs of known structure, consisting of six tightly packed α-helices, although the conformation of some of the inter-helical loops and the orientation and length of the last helix are unusual. Mutations in the CARD domains of both NOD1 and RICK full-length proteins were made in order to define residues important for CARD-CARD interaction. Co-immunoprecipitation of cell lysates and NF-κB activation assays were performed at the University of Michigan Medical School. The results suggest that the interaction is primarily electrostatic and dependent on a patch of three acidic residues on NOD1 CARD (E53, D54, and E56) and three basic residues on RICK CARD (R444, R483 and R488). In addition, two amino acids L44 and I57, located in or close to the acidic patch of NOD1 were found to affect NOD1 signaling. These could be involved in the precise formation of a functional CARD-CARD interaction, or they could interact with other regions of RICK than the CARD domain.

Keywords

NOD1/CARD4, RICK/RIP2/CARDIAK, CARD, innate immunity, NMR, 6-helix bundle, NF-κB, inflammation, apotosis.

Résumé

NOD1 est une protéine de surveillance intracellulaire jouant un rôle crucial dans la défense immunitaire innée. Suite à la reconnaissance d'un ligand bactérien spécifique, NOD1 recrute RICK au moyen d'une interaction CARD-CARD. RICK module ensuite l'activation du facteur de transcription NF-κB et initie une cascade de signalisation pro-inflammatoire. Le but de cette thèse était de déterminer la base structurale de ce processus. Le domaine CARD de NOD1 a été cloné, exprimé et purifié à la station EMBL de Grenoble. La structure de ce domaine a été résolue par résonance magnétique nucléaire à l'IBS de Grenoble. Cette structure est généralement similaire à celle des autres domaines CARD connus, étant constituée d'un paquet compact de 6 hélices α. Cependant elle présente des caractéristiques inhabituelles concernant la conformation de certaines boucles inter-hélices ainsi que la longueur et l'orientation de la sixième hélice. Des mutations dans les domaines CARDs des protéines NOD1 et RICK entières ont été réalisées pour identifier les résidus importants pour l'interaction ou la signalisation. Des expériences de co-immunoprécipitations et d'activation de NF-κB ont été effectuées à l'Université du Michigan. Trois résidus d'un patch acide de NOD1 (E53, D54 et E56) et trois résidus d'un patch basique de RICK (R444, R483 et R488) sont indispensables à l'interaction. Celle-ci serait donc de nature électrostatique. Deux autres résidus du patch acide du CARD de NOD1, L44 et I57, ont été identifiés comme importants pour la signalisation. Ces résidus peuvent potentiellement contribuer à la formation d'une interaction CARD-CARD fonctionnelle, ou bien également interagir avec d'autres régions de RICK.

Mots-clés

NOD1/CARD4, RICK/RIP2/CARDIAK, CARD, immunité innée, RMN, paquet de 6 hélices, NF-κB, inflammation, apotose.