Dottorato in Scienze Biotecnologiche – XIX ciclo
Indirizzo Biotecnologie Molecolari
Università di Napoli Federico II

# “Protein interactions mediated by α-helical domains”

**Dott.ssa Barbara Tizzano**

Dottoranda: Tizzano Barbara

Relatore: Prof. Ettore Benedetti

Coordinatore: Prof. Gennaro Marino

# INDEX

## Riassunto

## “Interazioni proteiche mediate da domini α-elicoidali”

Uno degli aspetti più interessanti ed impegnativi nella moderna ricerca biotecnologica riguarda la comprensione, a livello molecolare delle funzioni cellulari attraverso lo studio delle relazioni struttura-funzione delle proteine. L’importanza nell’ampliare le informazioni sull’attività biologica di singole proteine, risiede nella possibilità di controllarne i meccanismi di azione a livello cellulare e di prevederne gli effetti. Molti processi non sono controllati soltanto dall’abbondanza relativa delle varie proteine, ma anche dalla regolazione transiente della loro attività mediante, ad esempio, modificazioni covalenti reversibili e da interazioni non covalenti tra fattori proteici all’interno della cellula. Gli studi di proteomica funzionale e di biologia strutturale che studiano questi processi, sono dedicati all’approfondimento delle conoscenze dei meccanismi fisiopatologici legati alle interazioni proteiche ed alla progettazione razionale di farmaci da usare per combattere malattie in rapida diffusione. E’ infatti noto che numerose patologie che vanno dalle più comuni disfunzioni metaboliche a malattie quali la tubercolosi e la malaria, alle più svariate forme tumorali, sono strettamente correlate a fattori di natura proteica ed alle loro interazioni con molteplici partner molecolari che propagano i segnali di innesco iniziali (trasduzione del segnale).

Uno dei meccanismi biomolecolari più importanti è proprio la trasduzione di questi segnali, innescati generalmente da specifiche interazioni di proteine extracellulari con recettori transmembrana, propagate all’interno della cellula sia da altre interazioni proteina-proteina sia da più complessi meccanismi enzimatici. Tali interazioni si traducono in eventi molecolari che portano all’attivazione di fattori di trascrizione attraverso complesse cascate amplificative. Alcuni meccanismi regolano l’intero ciclo vitale della cellula e sono la risultante di un complesso “balancing” di opposti segnali di sopravvivenza cellulare (anti-apoptotici) o di morte cellulare programmata (pro-apoptotici) che sono regolati dal riconoscimento tra specifici domini proteici, per questo denominati “Death domains” (DD). La segnalazione dell’apoptosi si avvale per lo più di interazioni omofiliche tra DD di recettori di superficie e proteine adattatrici intracellulari che determinano l’attivazione di proteasi specifiche (caspasi). Tale attivazione determina una cascata di eventi proteolitici e nucleolitici che amplificano il segnale e portano alla morte cellulare. Un’alterazione nei processi mediati da fattori proteici coinvolti nell’apoptosi è responsabile, ad esempio, della formazione di tumori quali il linfoma gastrico di tipo MALT (Mucosa-Associated Lymphoid Tissue). La sovraespressione del fattore Bcl10 nel nucleo e nel citoplasma dei linfociti B porta ad un’alterazione del segnale di attivazione del fattore NF-κB, con conseguente alterazione dei meccanismi di morte cellulare programmata e sviluppo del linfoma. Il meccanismo di azione cellulare di tale fattore è tuttora oggetto di numerosi studi.

Un altro aspetto interessante nello studio dei meccanismi cellulari proteina-mediati risiede nello studio delle interazioni di fattori proteici con la parete cellulare batterica.

E’ noto che la maggior parte di antibiotici e antimicrobici agisce a livello della biosintesi del peptidoglicano, il principale costituente della parete batterica, o modificando la permeabilità di quest’ultima. Recenti studi rivelano che anche le interazioni di alcuni fattori proteici con la parete cellulare rivestono un ruolo cruciale

in processi patologici, quali ad esempio la tubercolosi, una malattia, ancora oggi, di altissimo impatto sociale. Il trattamento della tubercolosi, è seriamente compromesso dai fenomeni di resistenza ai comuni antibiotici sviluppati dal batterio responsabile, *Mycobacterium tuberculosis*. Pertanto, è pressante la richiesta di nuovi farmaci, che consentano di potenziare i trattamenti attualmente in uso.
La capacità di tale batterio di permanere in uno stato di latenza nell'organismo ospite, per poi attivare i propri effetti patologici dopo lungo tempo, ne fa ancora oggi una delle principali cause di morte in molti pesi del mondo.
Studi recenti su *Mycobacterium tuberculosis* hanno identificato fattori proteici in grado di "resuscitare" il batterio dal suo stato di latenza. Tali proteine, mediante un meccanismo di azione simile a quello descritto per il lisozima, idrolizzano parzialmente la parete cellulare determinando la riattivazione delle funzioni patologiche del batterio. Il meccanismo che determina la transizione del batterio dallo stato di latenza a quello patologico non è stato ancora chiarito, e non è noto se la parziale degradazione della parete batterica determini l'attivazione di altri fattori. La possibilità di dare una risposta a questi interrogativi fa delle proteine coinvolte in tali meccanismi un interessante argomento di studio.

Il presente lavoro di tesi si propone lo studio di due tipi di interazioni proteiche coinvolte in meccanismi patologici: le interazioni omotipiche fra domini di proteine coinvolte nella formazione di linfomi e le interazioni di fattori proteici con la parete cellulare del batterio responsabile della tubercolosi, allo scopo di chiarire i meccanismi di riattivazione della patogenesi del batterio nel suo stato di latenza. Tali studi sono volti ad una migliore comprensione dei meccanismi molecolari che regolano i processi patologici citati con lo scopo di individuare nuovi e più specifici "target" di azione farmacologica.

La prima parte del lavoro di tesi è incentrata sulla caratterizzazione strutturale e funzionale del dominio CARD (Caspase-recruitment domain) del fattore proteico Bcl10. Sulla base di recenti dati di letteratura è emerso che la sovraespressione di tale proteina è legata allo sviluppo di linfomi di tipo MALT ed è determinata da una traslocazione genica che pone il gene bcl10 sotto il controllo del gene promotore delle catene pesanti delle immunoglobuline. Il meccanismo attraverso il quale la sovraespressione di Bcl10 conduca all'oncogenesi è attualmente sconosciuto. E' noto, tuttavia, che il gene "wild-type" bcl10 codifica per una proteina contenente un dominio CARD N-terminale, mediante il quale Bcl10 interagisce con domini uguali, presenti in altri fattori proteici, attivando il fattore NF-κB e promuovendo l'apoptosi. Modificazioni morfologiche o mutazioni a livello del dominio CARD si riflettono nella perdita di capacità, da parte della proteina, di indurre apoptosi e, conseguentemente, nello sviluppo del tumore. E' stato inoltre riportato che Bcl10 è in grado di assemblarsi in filamenti all'interno del citoplasma, mediante un processo di omodimerizzazione tra domini CARD di molecole contigue. La formazione di filamenti avviene a seguito della dimerizzazione CARD/CARD e questa organizzazione porterebbe all'attivazione del fattore NF-κB.
Sebbene la struttura del dominio CARD di Bcl10 non sia stata ancora determinata, è nota la struttura cristallografica ed NMR di altri CARD che presentano una discreta identità di sequenza con quello appartenente a Bcl10. Il dominio, presente nella regione N-terminale della proteina, è costituito da sei α-eliche con un arrangiamento tipico a chiave greca ed ha una lunghezza di circa 110 residui amminoacidici. Le interazioni mediate da CARD sono da attribuire ad interazioni di tipo idrofobico tra

eliche, tuttavia, uno studio sulle regioni imputate a tali interazioni non è mai stato riportato in letteratura. Nel presente lavoro di tesi si è studiato un mutante del dominio CARD, ottenuto per via ricombinante, con le uniche due cisteine presenti nella sequenza, sostituite da serine. Tale dominio è stato caratterizzato in soluzione mediante tecniche di dicroismo circolare per verificarne il folding, e, successivamente, è stato condotto uno studio per verificarne la tendenza associativa in forma di omodimero.
Da questi esperimenti preliminari è stato possibile studiare l'equilibrio monomero-dimero in soluzione e derivare il calcolo della costante di dissociazione del dimero. Il valore di tale costante è stato derivato considerando la variazione di un parametro spettroscopico (in questo caso l'ellitticità molare) in dipendenza della concentrazione di CARD. Il valore ottenuto è di 18.6 μM. Si è inoltre osservato che ad un aumento della concentrazione di CARD in soluzione corrisponde uno spostamento dell'equilibrio verso la forma dimerica ma, inaspettatamente, anche una perdita di strutturazione da parte del dominio che passa da un folding prevalentemente di tipo α elicoidale ad una forma di tipo random. Successivamente, l'interazione CARD/CARD è stata verificata con saggi ELISA immobilizzando il dominio su piastra e monitorando l'associazione con lo stesso dominio derivatizzato con la biotina, utilizzando il sistema di rilevazione biotina-streptavidina. Una volta verificata l'associazione, allo scopo di identificare le regioni della sequenza di CARD responsabili di tali interazioni omotipiche, il dominio è stato digerito con tripsina e le frazioni di digerito sono state utilizzate per un saggio di competizione ELISA. I frammenti triptici di CARD, che competono nell'interazione con il dominio intero, sono stati selezionati e caratterizzati mediante LC-MS.
Allo scopo di caratterizzare le regioni di interazione tra domini CARD individuate mediante i saggi sul digerito triptico, tali frammenti sono stati ottenuti per via sintetica con strategie di sintesi peptidica in fase solida. Il frammento CARD [127-134] e due suoi analoghi di sequenza più estesa, CARD [104-134] e CARD [114-134], sono stati sintetizzati e la loro attività come modulatori dell'interazione omotipica è stata verificata con successo mediante un saggio ELISA di competizione analogo a quello utilizzato per testare i frammenti triptici. Successivamente, è stata effettuata la caratterizzazione al dicroismo circolare sui frammenti sintetici per rivelarne le preferenze conformazionali in soluzione. Dagli spettri acquisiti a 20°C e a due differenti valori di pH (4.5 e 7.5), si osserva che il frammento CARD [104-134] adotta in soluzione un folding di tipo α-elicoidale in entrambe le condizioni sperimentate. Si è inoltre osservato che, tale folding è molto stabile, anche a temperature superiori ai 40°C. Un preliminare studio NMR è stato effettuato sul frammento a diverse concentrazioni e diverse temperature ed è stato possibile definire che il peptide rimane stabile e non subisce aggregazione fino a concentrazioni pari a 0.3 mM. Ulteriori esperimenti NMR sono attualmente in corso per la determinazione della struttura di tale frammento il soluzione. Inoltre, esperimenti *in vitro* per testarne la capacità di inibire la formazione dei filamenti citoplasmatici che si formano a seguito di interazioni CARD/CARD, sono stati programmati e forniranno un importante contributo alla comprensione del meccanismo che conduce all'oncogenesi.

La seconda parte del presente lavoro di tesi è dedicata allo studio delle interazioni di proteine con la parete batterica, coinvolte nei meccanismi di resistenza del batterio *M. tuberculosis*. E' noto che tale batterio patogeno è in grado di entrare in uno stato di latenza, in cui non si verifica replicazione e tutte le attività metaboliche sono ridotte. In tale stato il batterio può rimanere inattivo per lunghi periodi, finchè un

evento di “resuscitazione” lo riporta nella sua forma patogenica. I farmaci antibiotici attualmente impiegati per combattere la patologia agiscono solo sul batterio attivo e quindi quando la patologia è già in atto. Una maggiore comprensione dei meccanismi alla base del processo che porta il batterio fuori dallo stato di latenza è in questo caso di fondamentale importanza per poterne contrastare gli effetti. Il primo fattore proteico denominato “Resuscitation-promoting factor” (Rpf) perchè in grado di “resuscitare” colture in stato di latenza è stato individuato in *Micrococcus luteus*. Tuttavia, mentre in *M. luteus* è presente un solo gene codificante per tale fattore, *M. tuberculosis* così come altri organismi appartenenti alla stessa famiglia *M. bovis* e *M. leprae,* possiede cinque geni codificanti per altrettante proteine denominate Rpf A-E. Tali proteine presentano una discreta omologia di sequenza con quella identificata in *M. luteus* ed hanno in comune un dominio C-terminale altamente conservato, di circa 80 residui amminoacidici. La struttura NMR di tale dominio, ottenuta per il fattore Rpf B, rivela la presenza di 4 α-eliche ed un folding analogo a quello del lisozima.
L’attività muralitica sembra essere alla base del meccanismo di “resuscitazione” del batterio. Recenti studi condotti su Rpf di *M. luteus* hanno infatti dimostrato che tale fattore proteico ha attività paragonabile a quella del lisozima e, tale attività, coinvolgerebbe il suo dominio C-terminale mentre, lo stesso dominio, nelle Rpf di *M. tuberculosis* sembrerebbe non possedere attività muralitica.
Nel presente lavoro di tesi si è dimostrato che un’analoga attività è presente anche in Rpf B da *M. tuberculosis*. Allo scopo di approfondire le conoscenze sul meccanismo di azione delle Rpf in *M. tuberculosis*, sono state selezionate due proteine da tale organismo: Rpf B ed Rpf C. La scelta è stata dettata da uno studio preliminare della struttura primaria e su predizioni di struttura secondaria ottenute con metodi bioinformatici. Con l’intento di effettuare una caratterizzazione strutturale, oltre che funzionale, di tali proteine, sono state selezionate le sequenze prive di regioni ricche in proline ed alanine, problematiche per il folding della proteina. Sono state quindi prese in considerazione le proteine Rpf B ed Rpf C che fra tutte le Rpf di *tuberculosis* meglio si prestavano a tali studi. Entrambe le proteine sono state clonate, espresse in *E. coli* e purificate, sia nella forma “full length” che in una forma troncata in cui, un breve segmento N-terminale di 30 residui, identificato come potenziale peptide segnale, è stato rimosso per aumentare le rese di proteina solubile. E’ stato inoltre clonato, espresso e purificato il solo dominio C-terminale della proteina Rpf B, denominato Rpf $B_{cd}$ in quanto costituisce il dominio catalitico della proteina. Una caratterizzazione in soluzione con tecniche di dicroismo circolare è stata effettuata per Rpf B e il suo dominio catalitico e per Rpf C. Gli spettri ottenuti rivelano che Rpf B ed Rpf C hanno entrambe propensione ad adottare un folding di tipo α-elicoidale, sebbene Rpf B sia più strutturata, nelle condizioni di pH e concentrazione sperimentate. Anche il dominio C-terminale Rpf $B_{cd}$ ha un folding di tipo α-elicoidale. Per Rpf B ed Rpf $B_{cd}$ sono stati effettuati esperimenti per verificare la stabilità delle proteine monitorando allo spettropolarimetro la perdita di strutturazione a seguito di un aumento della temperatura da 5°C a 80°C, e la capacità della proteina in esame di riacquistare il folding iniziale in seguito a raffreddamento a 5°C. Da tale studio emerge che, tanto la proteina, quanto il dominio Rpf $B_{cd}$ hanno una buona stabilità in quanto, il carattere α-elicoidale viene perso, a seguito dell’unfolding, in modo quasi del tutto reversibile. E’ possibile infatti paragonare gli spettri di dicroismo circolare delle proteine nello stato iniziale e dopo refolding. Successivamente, allo scopo di verificare l’attività muralitica di Rpf B ed Rpf C, sono stati effettuati esperimenti preliminari di “Dynamic Light Scattering” (DLS) per escludere la presenza di forme aggregate, che avrebbero potuto interferire con l’eventuale catalisi, come riportato in

letteratura per studi analoghi. Tali esperimenti hanno dimostrato che la preparazione di Rpf B non presentava aggregati mentre per Rpf C è stato possibile individuare una minima percentuale di aggregazione. Per questo motivo, i successivi studi di attività, sono stati focalizzati sulla sola Rpf B. L'attività muralitica della proteina, è stata quindi testata, utilizzando il substrato artificiale del lisozima, il 4-metilumbelliferil-β-D-N,N',N"-triacetilchitotrioso (MUF-tri NAG). Tale substrato è costituito da tre unità di N-acetilglucosammina (NAG), uno dei componenti essenziali del peptidoglicano che costituisce la parete cellulare, e dal 4-metilumbelliferone (MUF), un composto fluorogenico la cui presenza può essere rivelata in soluzione monitorando, allo spettrofluorimetro, l'emissione ad una determinata lunghezza d'onda. L'attività catalitica del lisozima è generalmente verificata a seguito della produzione del composto fluorogenico, eccitando la molecola ad una lunghezza d'onda di 360 nm e verificando l'emissione a 455 nm.
Dagli esperimenti di fluorescenza effettuati sulla proteina incubata a 37°C si osserva un incremento dell'intensità di fluorescenza a 455 nm, più evidente nell'arco delle prime tre ore di incubazione. L'incremento dell' intensità di emissione a 455 nm corrisponde ad un progressivo aumento della concentrazione di composto fluorogenico prodotto dall'idrolisi del substrato ad opera della proteina testata.
Ulteriori esperimenti sono in corso per dare una stima quantitativa dell'attività muralitica della proteina Rpf B e del suo dominio Rpf $B_{cd}$, paragonando l'attività osservata a quella del lisozima, sullo stesso substrato.
Gli stessi esperimenti di attività saranno effettuati sulla proteina Rpf C a seguito dell'ottimizzazione delle condizioni di stabilità della proteina in soluzione e, successivamente paragonati con i risultati ottenuti per Rpf B.
Per entrambe le proteine sono attualmente in corso prove di cristallizzazione per determinarne la struttura cristallografica. La determinazione dalla struttura servirà anche ad individuare la regione del dominio α-elicoidale responsabile dell'interazione con il substrato e a fornire ulteriori chiarimenti sul meccanismo di reazione.
Le prospettive future di questo studio sono inoltre rivolte verso la progettazione di test *in vitro* delle proteine Rpf B ed Rpf C su colture in stato di latenza per testarne l'attività promotrice della "resuscitazione" e progettare eventuali test di molecole potenzialmente inibitrici.

## Summary

## “Protein interactions mediated by α-helical domains”

A recent and exciting topic in the biotecnological research is the exploitation of a better knowledge of protein-protein interaction for the design of new pharmaceuticals. An intriguing challenge is to understand how interaction domains cooperate to establish the complex signaling networks, controlling events as fundamental as cell development, homeostasis and immune system function. Furthermore, other interaction types such as protein-membrane and protein-cell wall, are a challenging issue of investigation because of the involvement of these interactions in crucial mechanisms of cell homeostasis and related pathological processes.
In this thesis work, two important issues are developed: the role of protein-protein interactions in dysregulation of the apoptotic signal leading to MALT lymphoma, and the role of protein-cell wall interactions in *Mycobacterium tuberculosis* reactivation from latency.
The first part of this study is dedicated to a complete investigation on the homodimerization of the human protein Bcl10, caused by homophilic interactions between α-helical domains named as “Caspase-Recruitment Domains” (CARD). The structural behaviour of such a domain was studied by focusing the investigation on the domain regions, potentially involved in CARD/CARD self-association. The understanding of the dimerization mechanism, and how it can be modulated, is essential for the developing of tumour suppressor molecules.
The results obtained, provide an investigation on CARD dimerization properties, and the identification of one of the region sequences involved in its self-association, the fragment CARD [78-98]. Preliminary NMR studies were carried out, in order to obtain the solution structure of this fragment.
The second part of this thesis work is focused on the importance of protein-cell wall interactions in the transition of *Mycobacterium tuberculosis* from a latent state to a pathological state. With this aim, the study is devoted to the structural/functional characterisation of proteins from the bacterium *Mycobacterium tuberculosis* which are key players for its pathogenic activity. The main systems under investigation are two proteins (“Resuscitation-Promoting Factors”, named as Rpf). The role of Rpf proteins in the revival of dormant bacteria, still not clear, has been suggested to be the cleavage of oligosaccharides constituting the bacterial cell wall. All the studies reported so far in literature evidenced a muralytic activity, similar to that of the c-type lysozyme.
The proteins Rpf B, Rpf C and the catalytic domain Rpf $B_{cd}$ were cloned, expressed and purified. In view of Rpf B’s structural similarity to lysozyme and in order to verify the muralytic activity of the protein, the synthetic, artificial lysozyme substrate, was used for activity assays. This work has evidenced that Rpf B possesses significant muralytic activity. Further studies will be aimed to design Rpfs’ inhibitors, which may prevent the pathogen to exit its dormant state. *In vitro* assays will be performed by screening a complete library of compounds, in order to identify those able to inhibit the muralytic activity observed so far.

## Abbreviation index

Apaf 1: Apoptotic Protease-Activating Factor 1
BCL10: B-cell lymphoma 10
BSA: Bovine Serum Albumin
CARD: Caspase Recruitment Domain
CC-CARD: coiled-coiled caspase recruitment domain
CD: Circular Dichroism
DLS: Dynamic Light Scattering
CMI: Cell mediated immunity
DD: Death Domain
DED: Death Effector Domain
DMF: *N,N*-dimethylformamide
$D_T$ : Diffusion Coefficient
DTT: 1,4-Dithiothreitol
ELISA: Enzyme-Linked Immunosorbent Assay
FPLC: Fast Protein Liquid Chromatography
HPLC: High Performance Liquid Chromatography
HRP: horseradish peroxidase
IPTG: isopropyl-beta-D-thiogalactopyranoside
LC: long-chain
LC-MS: Liquid Chromatography-Mass Spectrometry
LB: Luria-Bertani
MAGUK: Membrane-Associated Guanylate Kinase
MALT: Mucosa-Associated Lymphoid Tissue
MALT1: Mucosa-Associated Lymphoid Tissue Lymphoma Translocation Protein 1
MTB: *Mycobacterium tuberculosis*
MUF tri-NAG: 4-methylumbelliferyl-β-D-N, N', N"-triacetylchitotriose
NAG: N-acethylglucosamine
NBD: Nucleotide-Binding Domain
NF-kB: Nuclear Factor-kB
NHS: N-hydroxysuccinimide
NMR: Nuclear Magnetic Resonance
NTA: nitrilotriacetic acid
$OD_{600}$: Optical Density at 600 nm
OPD: ortophenyldiamine
PBS: Phosphate Buffered Saline
PCR: Polymerase Chain Reaction
PDB: Protein Data Bank
PMSF: Phenylmethylsulphonylfluoride (Inhibitor of Serine proteases)
$R_H$ : Hydrodynamic Radius
RP-HPLC: Reverse Phase-High Pressure Liquid Chromatography
Rpf: Resuscitation-promoting factor
SD: Shine-Dalgarno
TB: *tuberculosis*
TEV: Tobacco Etch Virus (TEV) NIa protease
TFA: Trifluoroacetic acid
TIS: triisopropylsilane
TNF-α: Tumor Necrosis Factor-α

TPCK: Tosyl Phenylalanyl ChloromethylKetone 1-Chloro-3-tosylamido-4-phenyl-2-butanone (irreversible inhibitor of chimotrypsin)
Tris: Triisopropylsilane
tri-NAG: N, N', N"-triacetylchitotriose

# 1. Introduction

## 1.1 Protein interactions mediated by α-helical domains

The rapidly evolving field of protein science has now come to realize the ubiquity and importance of protein-protein interactions. It had been known for some time that proteins may interact with each other to form functional complexes, but it was thought to be the property of only a handful of key proteins. However, protein-protein interactions at an organism level, are the norm and not the exception. Thus, protein function must be understood in the larger context of the various binding complexes that each protein may form with interacting partners at a given time in the life cycle of a cell. Proteins are now seen as forming sophisticated interaction networks subject to remarkable regulation. The study of these interaction networks and regulatory mechanism, is one of the most challenging fields of proteomics. Characterizing the interactions of proteins in a given cellular proteome, will be the next milestone along the road to understanding the biochemistry of the cell.

The study of proteins impacts directly on human health. Indeed, enzymes, receptors, and key regulator proteins have been targeted for decades for drug discovery. However, a recent and exciting development is the exploitation of a better knowledge of protein-protein interaction for the design of new pharmaceuticals. Furthermore, other interaction types such as protein-membrane and protein-cell wall, are a challenging issue of investigation because of the involvement of these interactions in crucial mechanisms of cell homeostasis and related pathological processes. Most common antibiotics interfere with the bacterial cell wall, consisting of a mesh-type peptidoglycan polymer (murein). Cell wall plays a fundamental role in protecting of the bacteria from outside. In addition, it provides the organism with the necessary rigidity to withstand considerable osmotic pressure. According to the type of bacteria, the cell wall has different architecture and even mediates important mechanisms of resistance. Hence bacterial cell walls are an excellent target for antimicrobial therapy, studies on the interactions occurring with molecules and protein factors are of great interest in molecular and medical biotechnologies.

This thesis work aims to provide insights on two important pathological mechanisms involving protein interactions with either other proteins of with the cell-wall.

The first part of the study is focused on the protein-protein interactions occurring in cellular signal transduction and leading to the activation of specific transcription factors through complex cascades. These pathways result in a fine balancing of opposite signals of cell survival (anti-apoptotic signals) and cell programmed death (pro-apoptotic signals). The signaling key word is in the interactions between proteinaceous factors, mediated by specific α-helical domains, called as “Death Domains”. The interactions occurring between such domains are most likely homophylic and, as the recognition of similar domains as domains oligomerization, allow the activation of transduction signals. Alterations in these pathways leads to alterations in apoptosis signals and, consequently, to oncogenesis, as observed in MALT lymphomas. This part of the thesis work provides a structure-function characterization of a domain of the human protein Bcl10, named as CARD. This domain is involved in dysregulation of the apoptotic signaling and in the development of MALT lymphoma.

Going further in protein interactions investigations, the second part of this thesis work is dedicated to the study of proteins interacting with the bacterial cell wall. The

bacterium under investigation is *Mycobacterium tuberculosis*, the organism responsible for the diffusion of tuberculosis, which is one of the biggest bacterial pathologies in the world. The aim is to clarify how this interaction activates the bacterium from a latent, non replicating state to the pathological state. The understanding of the mechanisms that regulate the route through the latent state is of fundamental importance for the development of new therapies. Indeed, inhibition of the process of exit from latency would keep the bacterium in a non pathologic, “dormant” state.

## 2.I Protein-protein interactions: insights into signal transduction

### 2.I.1 Introduction

Modular protein interaction domains play an important role in signal transduction by mediating the assembly of components into specific signaling complexes. The interchange of protein modules between signaling molecules has allowed nature to evolve new signal transduction pathways responding to specific stress and developmental stimuli.
The cytoplasmic proteins that convey information from cell surface receptors to their intracellular targets are commonly constructed of modular domains, that either have a catalytic function (such as protein or lipid kinase activity), or mediate the interactions of proteins with one another, or with phospholipids, nucleic acids or small molecule second messengers. These latter interaction domains play a critical role in the selective activation of signaling pathways, through their ability to recruit target proteins to activated receptors, and to regulate the subsequent formation of signaling complexes at appropriate subcellular locations. Such interaction domains may control not only the specificity of signal transduction, but also the kinetics with which cells respond to external and intrinsic cues, and can therefore give rise to complex cellular behaviors.
An intriguing challenge is to understand how interaction domains cooperate to establish the complex signaling networks, controlling events as fundamental and as complex as cell development, maintenance of homeostasis and immune system function. The cellular life-cycle is actually governed by of one of the most important regulated processes: the apoptotic cascade pathway [1].
Apoptosis is a conserved, gene-directed mechanism for the elimination of unnecessary or dangerous cells and is essential for maintaining a fine balance an for the well-being of a multicellular organism [2].
Dysregulation in apoptosis mechanisms may contribute to a variety of human diseases such as autoimmune syndromes, lymphomas and solid tumors.
The apoptosis result in a complex cascade of events and signaling pathways involving a large variety of receptors and proteinaceous factors, such as a family of proteins belonging to the so called “Death Domain” superfamily, composed of the “death domain” (DD), the “death effector domain” (DED) and the “caspase recruitment domain” (CARD) proteins. These proteins share a common involvement in signalling events regulating apoptosis by mediating the interactions necessary for transducing a “death signal” in a vital process of normal cell development and maintenance of homeostasis [3].

Homotypic interactions between proteins containing CARDs, DEDs or DDs have been found in divergent signalling pathways resulting in caspase activation and apoptosis. Several CARD-containing caspase, indeed, have been implicated in proteolytic processing of cytokines involved in the inflammatory response. The caspase recruitment domain (CARD) is a protein module that participates in apoptosis signaling through protein-protein interactions and several CARD proteins have been found to assemble into discrete signaling complexes [4].
Proteins containing CARD domains can be divided in subfamilies based on their overall domain structures: NBD-CARDs, coiled-coiled CARDs, bipartite-CARDs and CARD-only proteins.

The NBD-CARDs contain a nucleotide-binding domain in addition to CARD segment that acts as an oligomerisation surface. The coiled-coiled CARD proteins are structurally similar to the NBD-CARDs but the central NBD is replaced by a coiled-coil motif acting as an oligomerisation surface. The bipartite-CARDs contain a CARD domain and one other motif like a kinase domain or a protease domain. The CARD-only proteins may act as positive or negative regulators of the multi-domain CARDs by regulating the recruitment between proteins [5].
All of the proteins belonging to bipartite-CARD subfamily appear to recruit coiled-coiled CARD via CARD-CARD interactions and they often promote the inducible transcription factor NF-kB activation, leading to apoptosis-related defects.

Lymphocyte activation is trigged by a stimulation of antigen receptors on the surface of T or B lymphocytes inducing a series of signal transduction cascades leading to activation of multiple transcription factors, including NF-kB.
NF-kB (nuclear factor-kB) is found in essentially all cell types. It is a dimeric transcription factor structurally and functionally related to the family of DNA-binding proteins that recognize a common sequence motif [6]. The signaling pathway leading to the activation of this transcription factor has been largely investigated. However, how factors involved in the signaling interact with each other is still a debatable issue. Since many nodal signal molecules in this pathway relay more than one of the upstream signals to their downstream targets, it has been speculated that the transmission of signals involves a network, rather than a linear sequence in the activation of NF-kB. In the cytoplasm, NF-kB is normally inhibited by IkB. The activation of NF-kB is dependent on the phosphorylation and subsequent degradation of IkB proteins [7]. IkB degradation occurs after its inducible phosphorylation, which results in ubiquitination and degradation by the proteasome [8]. The free and activated NF-kB can then translocate to the nucleus where it regulates gene transcription.

The inhibition of NF-kB/IkB kinase complex is responsible for the phosphorylation of molecules normally retaining NF-kB in an inactive state in the cytoplasm. In most cells, NF-kB activation protects from apoptosis, through induction of survival genes, although under certain conditions and in certain cell types it may also induce apoptosis [9, 10].
The antigen receptor-mediated signal to activate NF-kB of T and B cells is related to the protein Bcl10. The pathway by which this protein activates NF-kB is still unclear; however, several factors have been shown to participate in this biological process. Recent studies have identified a signalling complex involving Bcl10 and other two proteins: CARMA1, and MALT1 [11]. Such a complex has a key role in antigen-receptor-mediated activation of the NF-kB linking antigen receptors on B and T cells to the IkB kinase complex and is crucial for the activation and proliferation of mature B and T cells. Deletion of any one of these proteins impairs the ability of antigen receptors to activate NF-kB but appears not to affect other pathways [12].

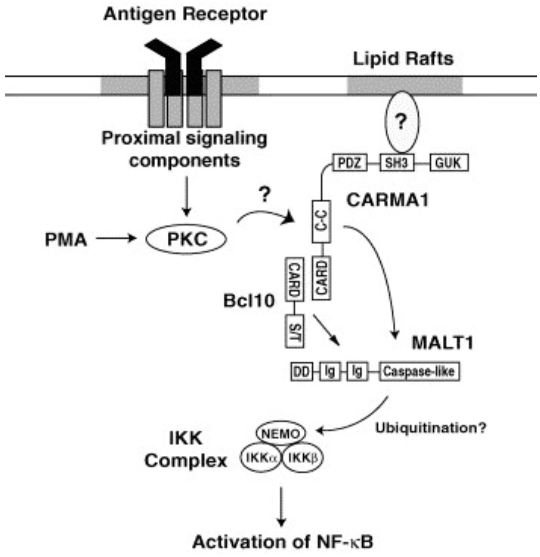



**Figure 2.I.1:** CARMA1, Bcl10 and MALT1 complex scheme

In T lymphocytes, Bcl10 normally resides in the cytoplasm and specifically relays antigen-receptor–mediated signals to activate nuclear factor NF-kB [13].
Bcl10 nuclear expression was found to be more closely associated with the genetic aberration and advanced tumor stages, in gastric MALT lymphoma [14]. Mucosa-associated lymphoid tissue (MALT) lymphomas are the most common type of B-cell human lymphomas virtually affecting every organ in the human body but the majority occur in the stomach.
The pathogenesis of MALT lymphomas is specifically associated with, at least 3 independent chromosomal translocations. Two of them, t(1;14)(q22;q32) and t(14;18)(q32;q21), result in the upregulation of bcl10 gene and its related factor MALT1, leading to a mechanism of overexpression of the gene product and, consequently, to cellular oncogenesis [15].
In particular, molecular cloning of the breakpoint t(1;14)(p22;q32) has identified the translocation of bcl10. This translocation juxtaposes bcl10 gene of chromosome 1 to the immunoglobulin heavy chain gene locus of chromosome 14, resulting in overexpression of Bcl10 protein in the nuclei and cytoplasm, in contrast to the weak cytoplasmic expression of Bcl10 in normal germinal center B.
Oncogene activation is often related to translocations, in which rearrangements often involve the immunoglobulin heavy chain IgH gene. Translocations into the immunoglobulin loci result in dysregulation of the uncoming gene, and may be due to transcriptional factors at these sites or mutation events within these regions [16].
The overexpressed Bcl10 in MALT lymphomas contains a variety of mutations, resulting in protein truncations and plays a critical role in antigen receptor–mediated lymphocyte proliferation and signaling to the nuclear factor NF-κB, controlling the expression of genes critical for cell survival, proliferation, and immune responses [17, 18]. Bcl10 pathway is still to be clarified but several studies have been focused on its interactions with other protein partners mediated by its CARD domain [19].
In normal B-cells, in response to antigen-receptor signalling, Bcl10 oligomerizes, via its N-terminal CARD, and interacts with MALT1. Recent studies have shown that Bcl10 and MALT1 mediate NF-κB activation, by facilitating polyubiquitination of subunit of the IkB kinase complex. MALT1, is a member of the conserved superfamily of caspase-like proteins. Although Bcl10 and MALT1 are two independent targets of

chromosomal translocation in MALT lymphomas, they form a tight complex *in vitro* [20, 21].
CARMA1 binds Bcl10 and induces its phosphorylation and translocation from the cytoplasm into perinuclear structures [22, 23]. Bcl10 phosphorylation is correlated with its ability to activate NF-kB, suggesting that this modification is required for NF-kB activation [4]. CARMA1 is a member of the membrane-associated guanylate kinase (MAGUK) family of scaffold proteins that assemble signal-transduction complexes by binding to both transmembrane and intracellular signalling molecules at sites of cell–cell contact [24].
Bcl10 CARD interacts with both MALT1 caspase-like domain and CARMA1 which is classified as a coiled-coiled CARD protein (figure 2.I.2) [11].

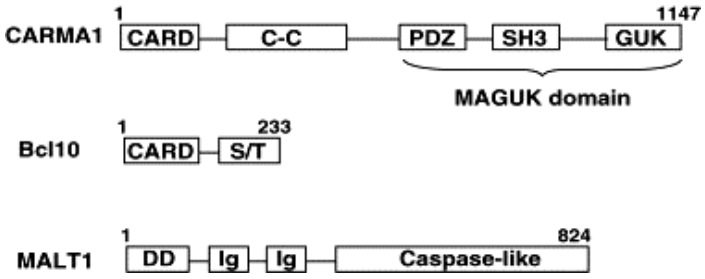
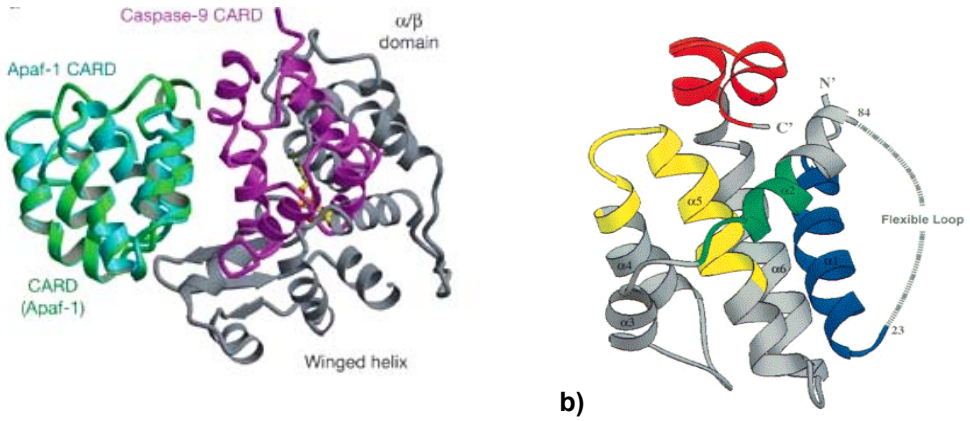



**Figure 2.I.2:** Structure scheme of CARMA1, Bcl10 and MALT1. CARMA1 and Bcl10 possess a CARD domain and their interaction occurs via CARD/CARD. MALT1 possess a "caspase-like" domain which binds the other CARDs

Bcl10 has a bipartite structure consisting of an N-terminal CARD and a C-terminal domain that is rich in serine-threonine residues.
Although there are no structures available for Bcl10 CARD domain, other CARDs have been studied by NMR and X-Ray crystallography [25].
Crystal and solution NMR structure has been reported for other proteins related to apoptosis such as RAIDD and Apaf1, a multi-domain protein with an N-terminal NBD-CARD and C-terminal repeats, responsible for recruitment of caspase9 via CARD-CARD interactions with consequent activation of the apoptosis cascade [26, 27].




**a)** **b)**

**Figure 2.I.3: a)** Crystal structure of the protein Apaf1 CARD domain in complex with caspase 9 CARD domain.The CARD-CARD interaction involves significant charge-charge interactions between the helices 2 and 3 of Apaf1 and the helices 1 and 4 of procaspase9. **b)** CARD domain model for Bcl-$x_L$.

All CARDs possess a remarkably similar secondary structure and often a discrete sequence identity. They are 10-15 kDa domains located at the N-terminus and consisting in antiparallel six helical bundle in the Greek key topology that form highly specific homophilic interactions between signaling partners [28].

A recent study [29] demonstrate that Bcl10 may also undergo homo-dimerization *via* CARD-CARD interactions. This oligomerization event is responsible of Bcl10 assembly in cytoplasmic filaments. The evidences provided in this study show that filaments serve as scaffold for recruitment of NF-κB-activating signal transduction.
Generation of filaments seems to be essential for the activation of a signaling cascade leading to cellular apoptosis and inhibition of filament formation prevents cell death. These observations make the investigation on CARD-CARD self-association an intriguing issue.

In order to provide a better understanding of Bcl10 CARD self-association mechanism, the first part of this thesis work is focused on dimerization studies based on the structural behaviour of the Bcl10 CARD domain in solution.
Further evidences of CARD-CARD association are provided and the specific regions involved in this homophilic interactions have been investigated to clarify how alterations, in these interactions, leads to alterations in apoptosis signals and, consequently, to oncogenesis.

## 2.I.2 Methods

### Protocol for CARD domain cloning, expression and purification

CARD domain (produced by "Dipartimento di Biologia Molecolare, II Policlinico, Napoli) was obtained as mutant of the wild-type human Bcl10 CARD.
The wild-type sequence is:

MEPTAPSLTEEDLTEVKKDALENLRVYL**C**EKIIAERHFDHLRAKKILSREDTEEIS**C**RT
SSRKRAGKLLDYLQENPKGLDTLVESIRREKTQNFLIQKITDEVLKLR

Two cysteine residues (underlined in bold) of this sequence were mutated in serine by site-directed mutagenesis of the wild-type gene (Quik Change Site-Directed Mutagenesis Kit commercially available from Stratagene), and, the mutated sequence, was cloned in pET 28 vector to obtain a fusion-$(His)_6$ protein with a Δ-link of 36 residues (poly-histidine tag included). The expression was carried out in BL21 (DE3) *E. coli* strain by growing the cultures at 37°C and inducing with 1 mM IPTG after 3 hours. The cultures were grown overnight at 37°C and the pellets were lysed in PBS buffer 8 M urea. The purification was performed by affinity chromatography by adding the lysate to a $Ni^{2+}$-chelated resin column, equilibrated in 100 mM phosphate , 10 mM TRIS-HCl buffer pH= 8.0 and 8 M urea. The protein was eluted with a gradient of pH from pH=8.0 to pH= 4.5. The final step of purification was performed by gel filtration on a Sephadex G150 column in buffer 50 mM TRIS, pH=4.0, dialyzed against buffer sodium acetate 100 mM pH= 4.0, lyophilized and stored at 4°C.
The aminoacidic sequence expressed is the following:

**MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEF**MEPTAPSLTEEDLTEVK
KDALENLRVYL**S**EKIIAERHFDHLRAKKILSREDTEEIS**S**RTSSRKRAGKLLDYLQENP
KGLDTLVESIRREKTQNFLIQKITDEVLKLR

where the two mutations are underlined in bold and the Δ-link with the His-tag is in bold characters. The following studies, reported in this thesis, are referred to this Bcl10 CARD mutant sequence even though, the domain cited is simply named as CARD.

### Protocol for biotinilation:

A protein is typically labelled with the water-soluble vitamin biotin to tag it for later recovery or for detection with biotin-binding proteins, such as avidin or streptavidin. The reagent used, NHS-LC-biotin has an N-hydroxysuccinimide (NHS) ester group to increase reactivity with primary amines and long-chain (LC) spacer arm to extend the biotin from the protein surface, improving binding of subsequent reagents.

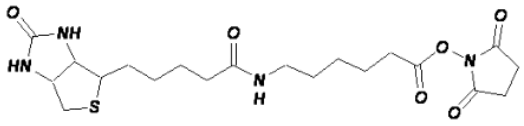



**Figure 2.I.4:** NHS-LC-Biotin

1 mg/mL purified CARD was dissolved in the coupling buffer (Sigma). NHS-LC-biotin solution 2 mg/mL (freshly prepared by dissolving biotin in EtOH to 2 mg/mL) was added to the solution. The mixture was incubated for 1 hour on ice, occasionally mixing. After incubation, 1 volume of 50 mM glycine solution (in coupling buffer) was added to bind unreacted biotin. Then glycine was removed to free biotin by dialysing for 48 hours against coupling buffer. The biotinilated CARD solution was sterilized by passing the solution through a 0.22-µm filter and stored at 4°C.
The yield of CARD biotinilated product is not the same for each preparation. For this reason the OD was measured before use for each new preparation of labeled CARD.

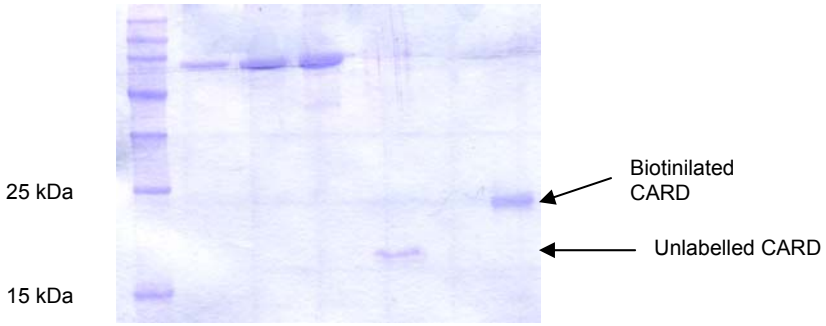



**Figure 2.I.5:** Acrylamide SDS-PAGE gel (15%) in reducing conditions. The gel shows the CARD sample before and after biotinilation.

### *Enzyme-Linked Immunosorbent Assay* (ELISA)

Biotinilated CARD was used as biotin-labelled protein in ELISA assays to test the self-association properties of this domain by using the biotin-streptavidin detection system.

### Protocol

The ELISA system uses polystyrene plates (Falcon) for protein binding. The plates are treated with UV to obtain sterile conditions and to enhance binding properties. The binding occurs between the protein sample and the polystyrene surface.

Lyophilized unlabelled CARD was dissolved in buffer sodium acetate 50 mM pH= 4.0 to obtain three different concentrations: 1.0 µg/mL, 0.50 µg/mL, 0.10 µg/mL.
This sample is insoluble at pH values between 4 and 7.
Each solution was spinned at 12 kRpms for 20 minutes, to avoid any precipitate.

**Coating**: binding of the unlabelled protein to the bottom of each well of the plate.
100 µL unlabelled CARD solution was placed in each well. The plate was incubated at 37°C for 3 hours to allow complete binding.
**Washing**: the wells were washed 3-5 times with PBS buffer 0.04% Tween-20 to remove all the molecules which have not been specifically bound to the plate.

**Blocking**: the remaining sites for protein binding on the plate must be saturated by incubating with blocking PBS buffer containing *Bovine Serum Albumin* (BSA) to avoid non-specific bindings. The wells were filled to the top with 250 µL BSA 1% (Sigma) in PBS buffer 0.04% Tween-20. Then, the plate was incubated for 2 hours at 37°C.
**Washing**: wells were washed twice with PBS buffer 0.04% Tween-20.
**Second coating**: application of the labelled (biotinilated) CARD to the wells (100 µL). Increasing concentrations biotinilated CARD (6.25 µg/mL, 12.5 µg/mL, 25.0 µg/mL, and 50.0 µg/mL) in buffer sodium acetate 50 mM pH= 4.0) were applied to the wells containing unlabelled CARD. The plate was incubated for 1 hour at 37° C.
**Washing**: the plate was washed four times with PBS buffer 0.04% Tween-20.
**Signal amplification with biotin-streptavidine method**: the wells were filled up with 100 µL streptavidine conjugated with horseradish peroxidase (HRP) enzyme (solution 1mg/mL in buffer PBS-sodium phosphate 1X, 137 mM NaCl, KCl 27 mM $Na_2HPO_4$,10 mM $KH_2PO_4$, 2 mM, pH=7.4 provided by Sigma) and incubated for 1 hour at 37°C and then washed 3 times with PBS washing buffer.
**Detection**: the wells were filled up with 100 µL detection buffer (citrate-phosphate 50 mM pH=5.0; $H_2O_2$ [0,4mg/ml]; ortophenyldiamine (OPD) [0,4mg/ml] (prepared by using the Sigma FAST OPD tablet sets KIT). This buffer should be freshly prepared before use. The plate was incubated at 37°C for 1 hour to reveal the binding of biotinilated CARD to the unlabelled CARD. The substrate is sensible to the light, so the plate was covered with aluminium paper to avoid any alteration in the response. OPD reaction produces colorimetric effect: the reaction solution changes from colorless to yellow and this changing can be quantitatively detected by measuring the intensity of the absorbance at 490 nm. The observed intensity is directly related to the amount of protein bound. This reaction went on till sulphuric acid 2.5 M was added (50 µL in each well).
The CARD/CARD binding properties were evidenced in this preliminary assay. The results are reported in the section 2.I.3.

**Proteolytic Digestion**

Proteins are specifically cleaved into small fragments to facilitate their identification. Specific cleavage can be achieved by trypsin which specifically hydrolyzes peptide bonds at the carboxyl side of lysine (Lys, K) and arginine (Arg, R) residues. Tryptic digestion is part of the standard protein analysis by mass spectrometry.

Trypsin is a pancreatic serine protease (Mr ~24 kDa). The aspartate residue in the binding site enables trypsin to form a salt bridge with positively charged residues (Lys and Arg side-chains) of the substrate, thus enabling the cleavage reaction to occur. Trypsin can undergo autolysis and lose activity. A chemically modified trypsin which does not undergo autolysis and retains high enzymatic activity is commercially available and has been used for this studies' purposes.

**Protocol for tryptic digestion**

CARD was dissolved in a solution buffer TRIS 50 mM, pH=7.5, $CaCl_2$ 20 mM and 10% acetonitrile was added to enhance the solubility. 0.1 µg/µL solution of trypsin (TPCK-treated trypsin from Sigma-Aldrich, Milan, Italy) was prepared using the same

buffer. CARD and trypsin solutions were mixed together in 100:1 (w:w) ratio. The reaction mix was incubated overnight at 37°C.

**Selection of tryptic CARD fragments**

In order to identify the fragments produced from the hydrolysis with trypsin, and to test them in a competitive ELISA assay, the digested mix was passed through a preparative RP-HPLC (Shimadzu HPLC equipped with a C18 Jupiter column 250x22mm ID column (15 µm, 300 Å) applying a linear gradient of acetonitrile in 0.1% TFA from 5% to 70% in 50 min.

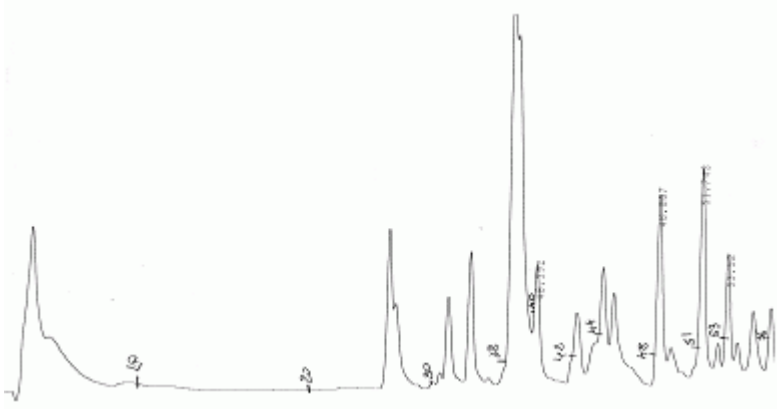


**Figure 2.I.6:** Preparative RP-HPLC chromatogram of digested CARD. The fraction were collected each 10 min. and named as 1F to 11 F.

Peaks were collected each 10 minutes in Falcon tubes. The fractions, collected from the tryptic digestion and named from 1F to 11F, were tested by competitive ELISA assay against the whole CARD.

**Protocol for competitive ELISA assay CARD/CARD digested fractions:**

**Coating**: binding of the unlabelled CARD 12.5 µg/mL (this is the unlabelled CARD concentration previously described for pre-saturation conditions). The plate was incubated at 37°C for 3 hours to allow complete binding.
**Washing**: the wells were washed 3-5 times with PBS buffer 0.04% Tween-20 to remove all the molecules which have not been specifically bound to the plate.
**Blocking**: The wells were filled to the top with 250 µL in BSA 1% (Sigma) in PBS buffer 0.04% Tween-20. Then, the plate has been incubated for 2 hours at 37°C.
**Washing**: wells were washed twice with PBS buffer 0.04% Tween-20.
**Competition:** the competition mix is composed of tryptic fragment and biotinilated CARD in two different ratio: mix 1:1 (w/w) and 10:1 (w/w) in sodium acetate 50 mM pH=4.0. The labelled CARD concentration was chosen, from the previous ELISA test, before the saturation point value and, with this aim, a concentration of 0.5 µg/mL of biotinilated CARD was used. The two reaction mixes were pre-incubated at 37°C for 45 min. To allow the fragments to better interact with labelled CARD.

**Second coating**: application of the biotinilated CARD/tryptic fragments mix to the wells (100 µL). The plate was incubated for 1 hour at 37° C.
**Washing**: the plate was washed four times with PBS buffer 0.04% Tween-20.
**Signal amplification with biotin-streptavidine method**:
The wells were filled up with 100 µL streptavidine conjugated with horseradish peroxidase (HRP) enzyme (solution 1mg/mL in buffer PBS-sodium phosphate 1X, 137 mM NaCl, KCl 27 mM $Na_2HPO_4$,10 mM $KH_2PO_4$, 2 mM, pH=7.4, provided by Sigma) and incubated for 1 hour at 37°C and then washed 3 times with PBS washing buffer.
**Detection**: the wells were filled up with 100 µL detection buffer (citrate-phosphate 50 mM pH=5.0; $H_2O_2$ [0,4mg/ml]; ortophenyldiamine (OPD) [0,4mg/ml] prepared by using the Sigma FAST OPD tablet sets KIT). The plate was incubated at 37°C for 1 hour. The reaction was detected by measuring the absorbance at 490 nm.
From this preliminary assay 3 fractions were selected: 3F, 8F and 9F (ELISA assay results are shown and discussed in section 2.I.3.2. The fraction 11F shows very low binding properties.

**LC-MS characterization of digested CARD:**

The CARD sequence:

**MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEF**MEPTAPSLTEEDLTEVK
KDALENLRVYL**S**EKIIAERHFDHLRAKKILSREDTEEIS**S**RTSSRKRAGKLLDYLQENP
KGLDTLVESIRREKTQNFLIQKITDEVLKLR

**Number of amino acids:** 143, **molecular weight:** 16328.4

with the two cysteines, mutated in serine residues and the Δ-link, including the His-tag was employed for all the experiments but the numbering reported for the identified sequences is referred to the "wild-type", without the Δ-linker. For the identification of the tryptic fragments the program MS-Digest for trypsin digestion cleavage was used on the web page http://www.expasy.org/tools. The following CARD fragments were identified by LC-MS. Near the fragments sequence the corresponding fraction number of the digestion mix is reported.

CARD [32-36]: IIAER (3F)
CARD [50-62]: EDTEEISSRTSSR (3F)
CARD [26-49]: VYLSEKIIAERHFDHLRAKKILSR (3F)
CARD [26-31]: VYLSEK (4F)
CARD [78-87]: GLDTLVESIR (5F)
CARD [78-88]: GLDTLVESIRR (5F)
CARD [37-42]: HFDHLR (6F)
CARD [99-105]: ITDEVLK (6F)
CARD [19-25]: DALENLR (7F)
CARD [91-98]: TQNFLIQK (8F)
CARD [89-98]: EKTQNFLIQK (8F)
CARD [68-77]: LLDYLQENPK (9F)
CARD [46-58]: ILSREDTEEISSR (11F)

LC-MS analyses were performed by using a LCQ DCA XP Ion Trap mass spectrometer, equipped with a narrow bore, $C_{18}$ BioBasic 50x2 mm ID column (Thermo Electron, Milan, Italy) and applying a gradient of solvent B (0.05% TFA in $CH_3CN$) on solvent A (0.08% TFA in $H_2O$) from 30 % to 70 % (v/v). 5 µL of each sample (0.1 mg/mL) were injected for this analysis. The global LC-MS profile and the corresponding identified fragments and their discussion are reported in figure 2.I.9 in section 2.I.3.2.

**CARD peptides synthesis**

The following fragments, previously selected from ELISA assay as CARD binding regions, were obtained by Solid-phase peptide synthesis using Fmoc synthesis strategy. The first fragment, together with its related longer sequence peptides: CARD [78-98] and CARD [68-98], are belonging to the fraction 8F, collected from the tryptic digestion. The fragment CARD [68-77] is belonging to the collected fraction 9F and CARD [26-49] is belonging to the fraction 3F.

Sequences:

CARD [91-98]: TQNFLIQK
CARD [78-98]: GLDTLVESIRREKTQNFLIQK
CARD [68-98]: LLDYLQENPKGLDTLVESIRREKTQNFLIQK

CARD [68-77]: LLDYLQENPK
CARD [26-49]: VYLSEKIIAERHFDHLRAKKILSR

The first three sequences were tested by ELISA competitive assay to determine their behaviour in competing with the whole CARD domain to bind the CARD itself (section 2.I.3.2). The other sequences were synthetized but so far never tested. Further experiments are in progress.

**Protocol**

Peptides N-terminal acetilated and C-terminal amidated, were synthesized in batch by standard Fmoc chemistry protocol on Rink-amide MBHA resins. Acylations were carried out in 50% DCM/DMF for 15 min using PyBOP/DIEA (1:2) as activating agents (without preactivation) with a four-fold amino acid excess. Fmoc removal was achieved by 30% piperidine/DMF treatment for 10 min. After peptide assembling, acetylation was carried out by treatment with a 1 M solution of acetic anhydride in DMF containing 5% DIEA. Cleavage from the resin was achieved by treatment with a trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/water (90:5.0:5.0 v/v/v) mixture for 90 min at room temperature. Then, peptides were precipitated in ether, dissolved in water/acetonitrile (1:1 v/v) mixture, lyophilized, and purified. Chemical and organic reagents were from Sigma-Aldrich (Milan, Italy). N-α-Fmoc protected amino acids and the activating agents were purchased from Inbios (Pozzuoli, Italy). Resins for peptide synthesis were from Novabiochem (Laüfelfingen, CH).

## Protocol for CARD peptides purification

Peptides purification was carried out by RP-HPLC using a Shimadzu purifier equipped with a C18 Jupiter 250x22mm ID column (15 µm, 300 Å) applying a linear gradient of acetonitrile in 0.1% TFA from 10% to 60% in 50 min. Peptide purity and integrity were confirmed by RP-HPLC analysis using analytical columns (Jupiter C18, 250 x 4.6 mm ID,15 µm, 300 Å) and LC-MS on a LCQ DCA XP Ion Trap mass spectrometer (Thermo Electron, Milan, Italy). All solvents were reagent grade. HPLC chemicals were purchased from Lab-Scan (Dublin, Ireland), while other Columns ($C_{18}$) for peptides purification and characterization were from Phenomenex (Torrance, CA).

## Circular Dichroism Spectroscopy of Proteins

Different types of protein secondary structures (helices, sheets, turns and coils) give rise to different CD spectra. The wavelengths of light that are most useful for examining the structures of proteins and peptides are in the ultraviolet (UV) ranges (from 160 to 300 nm) because these are the regions of the electronic transitions of the peptide backbone and side chains in proteins.

**Table 2.I.1:** Far UV-CD transitions for soluble proteins

| | | | | |
|---|---|---|---|---|
| α-helix | positive | π -> π * | 190-195 nm | 60,000 to 80,000 deg $cm^2$ $dmol^{-1}$ |
| | negative | π -> π * | 208 | -36,000 ± 3,000 |
| | negative | n-> π * | 222 | -36,000 ± 3,000 |
| β-sheet | positive | π -> π * | 195 - 200 | 30,000 to 50,000 |
| | negative | n-> π * | 215 - 220 | -10,000 to -20,000 |
| random | negative | π -> π * | ca. 200 | -20,000 |
| | positive | n-> π * | 220 | small |

CD spectroscopy has been used to monitor: 1) secondary structure, 2) conformational changes, 3) environmental effects, 4) protein folding and denaturation, and 5) dynamics and can be used to analyze a number of proteins from which interesting candidates can be selected for more detailed structural analysis like NMR or X-ray crystallography.Monitoring θ222nm of a protein as a function of temperature or chemical denaturant yields information about protein stability. The absorption should be less than 1.0 (usually < 0.3) for cell pathlengths of 0.05 to 1.00 cm in order to maintain reasonable signal-to-noise ratios and accurate CD measurements. Protein concentration is typically 1 mg/mL. Buffers are typically 10 mM phosphate with low salt if any.
Thermal stability is assessed using CD by following changes in the spectrum with increasing temperature. In some cases the entire spectrum in the far-UV CD region can be followed at a number of temperatures. Alternatively, a single wavelength can be chosen which monitors some specific feature of the protein structure, and the signal at that wavelength is then recorded continuously as the temperature is raised. CD is often used to assess the degree to which solution pH, buffers, and additives such as sugars, amino acids or salts alter the thermal stability. Many proteins

aggregate or precipitate quickly after they are unfolded ("melted"), making unfolding irreversible. The reversibility of the unfolding reaction can be assessed by cooling the sample and then heating again to see if the unfolding reaction is duplicated. Only if the melting is fully reversible, the melting temperature is directly related to conformational stability, and the thermodynamics of protein folding can be extracted from the data. The fact that thermal unfolding can generally be measured by CD at much lower concentrations than by DSC increases the probability of reversible reactions and of thermodynamically interpretable data. Fitting the data also allows a more reproducible measurement of the onset of unfolding, a temperature which is often more relevant for formulation and shelf-life considerations than the midpoint. If the protein precipitates or aggregates as it is unfolded, the melting reaction will be irreversible, and the melting temperature will reflect the kinetics of aggregation and the solubility of the unfolded form of the molecule as well as the intrinsic conformational stability. The cooperativity of the unfolding reaction is measured qualitatively by the width and shape of the unfolding transition. A highly cooperative unfolding reaction indicates that the protein existed initially as a compact, well-folded structure, while a very gradual, non-cooperative melting reaction indicates that the protein existed initially as a very flexible, partially unfolded protein or as a heterogeneous population of folded structures.

**Protocol for CARD peptides characterization by Circular Dichroism**

The following synthetic peptides were characterized by far-UV circular dichroism:

CARD [68-98]: LLDYLQENPKGLDTLVESIRREKTQNFLIQK
CARD [78-98]: GLDTLVESIRREKTQNFLIQK

Each peptide was characterized in buffer phosphate 10 mM at two different pH values: pH=7.5 and pH=4.5 by using a Jasco J-810 spectropolarimeter in 1.0 cm quarz cuvettes. CD spectra are reported in molar ellipticity [θ] in section 2.I.3.2.

Parameters:

Peptide concentration: 10 μM, wavelength range: 190 nm - 260 nm, scan Speed: 20 nm/min, band widht: 2.0 nm, resolution: 1.0 nm, sensitivity: 50 mdeg, response: 16 s, accumulation: 3

**Protocol for Circular Dichroism characterization of CARD domain**

Bcl10 CARD far-UV CD spectra for dimerization study were performed in sodium acetate 5 mM buffer solution at pH=4.5 and at room temperature on a Jasco J-810 spectropolarimeter. A set of 7 different pathlenght cuvettes was used (from 0.01 cm to 5 cm). CD spectra , in molar ellipticity [θ], and the used CARD concentrations, are reported in section 2.I.3.1.

Parameters:

Wavelength range: 195 nm - 260 nm, scan Speed: 20 nm/min, band widht: 2.0 nm, resolution: 1.0 nm, sensitivity: 50 mdeg, response: 16 s, accumulation: 3

### *Nuclear Magnetic Resonance* (NMR)

NMR methods are used to determine the primary, secondary and tertiary structure of peptides and proteins in solution. One and two-dimensional (1D and 2D) NMR experiments are used to assign the chemical shifts of all peptide protons. The set of distance and angular constraints are input into molecular modeling programs to calculate a tertiary structure if a unique fold exists. Typical sample concentrations for structure determination are 1 to 2 mM in a 750 microliter 5 mm sample tube for the 600 MHz instrument.
The H-α chemical shift of amino acids in a polypeptide or protein is a widely used indicator of secondary structure. The ”secondary structure shift" is defined as the difference between the observed chemical shift and the appropriate random coil value. There is a relationship between shift and secondary structure context, demonstrating downfield chemical shifts in beta-sheets and upfield shifts in alpha-helices and turns. The origin of these observed chemical shift changes remains a matter of some debate. A large number of studies [30, 31], show that H-α chemical shift is largely determined by neighbouring residue effects and local backbone restriction. Nuclear Magnetic Resonance provides a powerful method for determination of peptides and proteins structures. It is the only technique that can provide detailed information on the exact three-dimensional structure of biological molecules in solution. NMR measurement, generate information about spatial proximity of protons in a molecule. These spatial information can be used to generate models of peptides or polypeptides chains, giving rise to structures comparable with defined 3-Dimensional structures in solution. It is generally observed that such models are very close to the structure determined by X-Ray crystallography in the crystalline state.

### Protocol for NMR experiments

In order to provide the optimal experimental conditions (peptide concentration, pH and temperature), several preliminary experiments were performed on the synthetic fragment CARD[78-98]. NMR experiments were recorded on a Varian Inova 600 spectrometer at temperature of 25°C, located in “Istituto di Biostrutture e Bioimmgini” of CNR, Naples. All the spectra were processed and analyzed with the software MestreC (Mestrelab Research Santiago de Compostela a Coruna, Spain).
CARD[78-98] assayed concentrations in NMR preliminary experiments: 1.0 mM peptide neat $H_2O$ (10 % $D_2O$; 600 µL) at 25°C pH=3.5, 0.60 mM peptide neat $H_2O$ (10 % $D_2O$; 600 µL) at 298 K pH=3.5, 0.30 mM peptide neat $H_2O$ (10 % $D_2O$; 600 µL) at 25°C, pH=3.5, 0.10 mM peptide neat $H_2O$ (10 % $D_2O$; 600 µL) at 25°C pH=3.5.
In these experimental conditions 1D spectra showed changing of intensity and multiplicity of signals. In the second set of experiments, the following conditions were used: 0.1 mM concentration peptide $H_2O$ (10% $D_2O$; 600 µL) 5.0 mM Na acetate buffer pH=5.5. These conditions were chosen as the best for spectra acquirements. Spectra and discussion in section 2.I.3.2.

## 2.I.3 Results and discussion

### 2.I.3.1 Structural and dimerization properties of Bcl 10 CARD

The “Caspase-recruitment domain” (CARD) is a protein-binding module that mediates the assembly of CARD-containing proteins into apoptosis cascade signaling. The inhibition of these interactions between the involved proteins, such as Bcl10, causes pathogenesis. CARD-mediated homophilic interactions between Bcl10 CARDs cause Bcl10 multimerization and filaments formation. This organization is essential for the NF-κB activation and apoptosis signaling activation.
These studies aim to investigate into CARD domains self-association and to identify the specific regions involved. A physical-chemical approach has been used to describe CARD monomer-dimer equilibrium and to calculate the dissociation constant for CARD dimer formation.

Binding interactions between proteins are non-covalent, so they are weak and usually reversible and binding is stabilized by the same nonbonded interactions that stabilize folded proteins, such as the hydrophobic effect, van der Waal’s contacts, hydrogen bonds, and salt bridges. Binding reactions can be described by simple chemical equilibria. For the association of a molecule A with a molecule M, the equilibrium expression is given by the equation:

$M + A \rightleftharpoons MA$

$K_d = [M][A]/ [MA]$

Where $K_d$ is the dissociation constant which is the reciprocal of the corresponding association constant $K_a$. $K_d$ is usually expressed in units of concentration (M), assuming that the activities of the equilibrium components are corresponding to their concentrations.
If a protein, with a specific domain, binds to another protein or self-associate to form a dimer or an oligomer, the interactions usually involve modification of secondary structure elements in the protein folding. In this study the homodimerization, occurring between two identical CARD domains in solution, has been considered and the dissociation constant for its dimer has been calculated. Furthermore it has been calculated the percentage of CARD in monomeric form in dependence of the concentration.
This study starts from the consideration that the CARD domain can be present in solution in two forms: monomer (M) and dimer (D) and between the two forms the following equilibrium is considered:

$2M \rightleftharpoons D$ (1)

The dissociation can be described with a constant $K_d$.

$K_d = [M] \times [M]/ [D]$ (2)

The total protein concentration can be identified as x because is the only variable parameter considered in the experiment. It can be obtained by adding monomer and dimer concentration and by expressing the dimer concentration as the monomer multiplied for two:

$$x = [M] + 2 \bullet [D] \quad (3)$$

If α is the protein fraction present in monomeric form on the total concentration of the CARD domain: $\alpha = \frac{[M]}{x}$

$$1 - \alpha = 2\frac{[D]}{x} \quad (4)$$

transferring in the first equation D and M it becomes:

$$K_d = 2 \bullet \frac{(\alpha \bullet x) \bullet (\alpha \bullet x)}{x \bullet (1-\alpha)} = 2 \bullet x \bullet \frac{\alpha^2}{(1-\alpha)} \quad (5)$$

Solving the equation it is possible to obtain the expression:

$$\alpha = \frac{1}{x}\left(\sqrt{\frac{K_d^2}{16} + \frac{K_d \bullet x}{2}} - \frac{K_d}{4}\right) \quad (6)$$

The experiments here considered are expressed in molar residue ellipticity, [θ].
This parameter can be written as y and is corresponding to the hypothetical value observed when the monomeric form is the only one present in solution (a), multiplied for the monomeric fraction (α) and adding the value that would be observed if only dimer is present (b) multiplied for the dimer fraction (1-α).
The resulting equation can be written as:

$$y = a \bullet \alpha + b \bullet (1-\alpha) \quad (7)$$

Writing α value obtained combining the equation (4) with the equation (5) the global expression can be written as:

$$y = \frac{a}{x} \bullet \left(\sqrt{\frac{K_d^2}{16} + \frac{K_d \bullet x}{2}} - \frac{K_d}{4}\right) + \frac{b}{x} \bullet \left(x - \left(\sqrt{\frac{K_d{}^2}{16} + \frac{K_d \bullet x}{2}} - \frac{K_d}{4}\right)\right) \quad (8)$$

## Circular Dichroism Experiments

CARD domain belonging to the human protein Bcl10 with 2 cysteines mutated in serines was characterized in solution by far-UV circular dichroism to investigate the conformational behaviour of the domain. The behaviour of a protein in specific conditions of pH and temperature, together with the dependence of the spectroscopic parameters from the concentration can be used to study the tendency of the protein to adopt monomeric or oligomeric forms.
Far UV-CD spectra were acquired for Bcl10 CARD in sodium acetate 5 mM buffer solution at pH= 4.5 and at room temperature. In these conditions the domain shows a typical α-helical folding. The spectrum reported in the figure below shows one negative band centered at 208 nm with a shoulder at 222 nm and a positive band at 196 nm wavelength. These values are typical for α-helical structures.

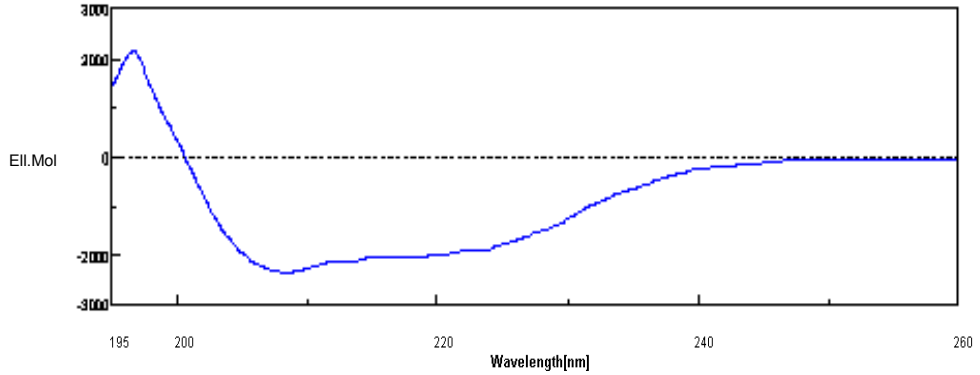



**Figure 2.I.7:** Bcl10 CARD far-UV CD spectrum in buffer sodium acetate 5 mM pH=4.5. The spectrum is a typical α-helical spectrum and shows a positive band at 196 nm wavelength and two negative bands at 208 nm and 222 nm.

In order to study and to verify the self-association between CARD domains it is possible to monitor a change in the spectroscopic properties of the protein. If there is a linear dependence between the chosen spectroscopic parameter and the protein concentration, it might be that there no self-association.
For this study it has been chosen the molecular ellipticity measured by far-UV circular dichroism, but any other spectroscopic parameter dependent on the sample concentration could be used. It has been also calculated the dissociation constant of the dimer $K_d$.
Bcl10 CARD has been studied in sodium acetate buffer 5 mM at pH= 4.5 by using a set of seven cells for UV experiments with different path lengths (0.01 -5.0 cm) and, thus, different volumes of sample solution. For each cuvette, sample concentrations were chosen to have a constant number of molecules crossed by the beam and circular dichroism experiments were carried out at room temperature starting from a stock concentrated solution. CARD concentrations and cell path lengths are reported in table 2.I.2.

**Table 2.I.2:** Cell path lengths and relative sample concentrations used for CD experiments

| Cell path (cm) | 0.01 | 0.05 | 0.10 | 0.20 | 0.50 | 1.00 | 5.00 |
|---|---|---|---|---|---|---|---|
| [CARD] (M) | $2.4 \cdot 10^{-4}$ | $4.8 \cdot 10^{-5}$ | $2.4 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $4.8 \cdot 10^{-6}$ | $2.4 \cdot 10^{-6}$ | $4.8 \cdot 10^{-7}$ |

The spectrum was acquired for each solution and ellipticity was measured after each dilution. The results are expressed as mean residue ellipticity, [θ].
These experiments show a variation in molar ellipticity dependent from CARD concentration and provide the evidences to deduce that a self-association is occurring between CARD molecules in solution.

The figure 2.I.8 shows the curve obtained by fitting seven points. Each point is the molar ellipticity value observed on CD spectrum for each CARD concentration measured in the appropriate path length cell. Molar ellipticity values obtained at 196 nm are here reported. This wavelength has been chosen because the corresponding values obtained give a better evidence of the changes in the spectroscopic parameter in dependence of the CARD concentration, but any other wavelength, different from isodichroic point, gives a similar value for $K_d$. Theoretically, the $K_d$ values calculated for each wavelength should be approximately the same but experimentally, as far from the isodichroic point is the value as higher is the confidence.

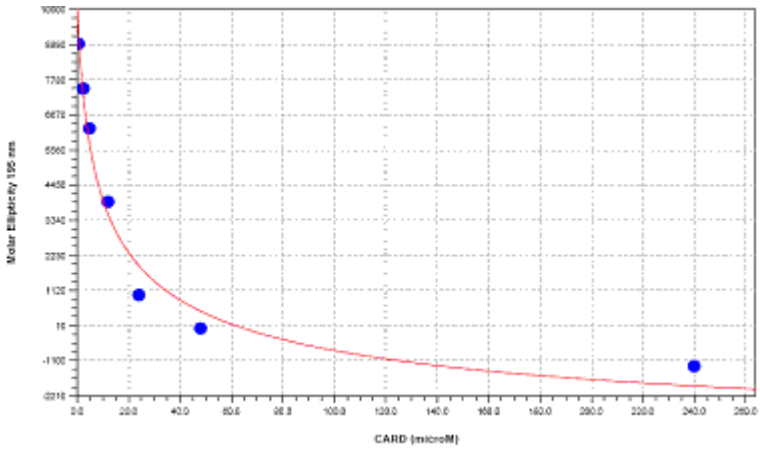



**Figure 2.I.8:** Bcl10 CARD. Molar ellipticity measured at 196 nm for different concentrations in different path length cells. The increasing concentration corresponds to a decreasing signal.

Starting from literature data [32], reporting CARD-like domains in dimeric form, it has been possible to characterize the intrinsic CARD behaviour to dimerize and to extrapolate the dissociation constant by using the equation (8).
By solving such equation, for different values of wavelength from 190 nm to 229 nm, it has been possible to obtain the fitting for a dimeric model. The calculated average value for dimerization constant $K_d$ is 18.6 μM.

The parametric values a and b were considered for each wavelength to extrapolate the dichroic signal, respectively to 0 and to ∞ concentration. In this way it has been possible to obtain the theoretical CD spectra for the CARD domain in the theoretical monomeric and dimeric form.
The curve built at 0 concentration shows a typical α-helical behaviour while, for infinite concentration a typical random spectrum is observed (figure 2.I.9).

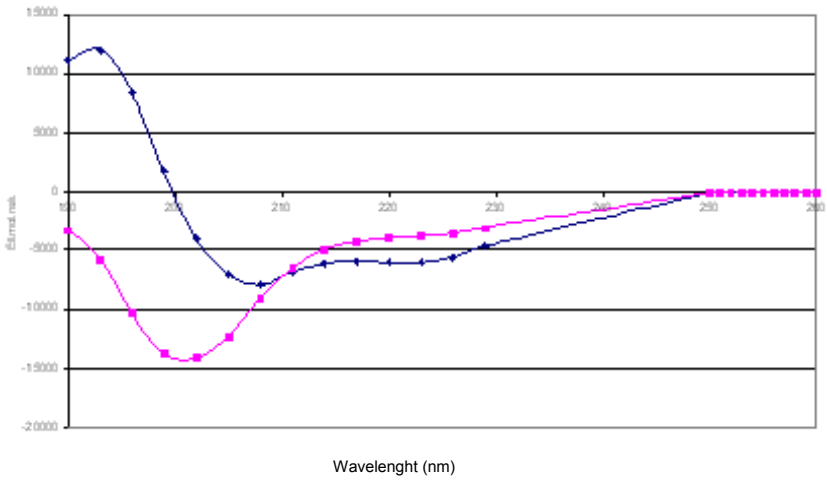



**Figure 2.I.9:** Bcl10 CARD theoretical dichroic behaviour. The curve here reported in blue represents the theoretical curve obtained by extrapolating the dichroic values at 0 concentration of CARD. In such hypothetical condition CARD could be considered as in monomeric form. The curve in magenta represents the extrapolation at infinite CARD concentration and is corresponding to a theoretical dimeric form.

The data extrapolated from these calculations show that, in response to an increasing concentration, a loss of folding is observed and, consequently, a dimeric CARD/CARD association should cause unfolding of the domain. This unusual behaviour has been also described for other homo-dimerizing proteins in literature.

## Dimer formation quantification

In order to calculate the percentage of monomeric and dimeric form, the difference between the theoretical value corresponding to the 100% momomeric form and the theoretical value, corresponding to the 100 % dimeric form was considered.
Each point of the theoretical α-helical curve reported in figure 2.I.13, represents the a value considered for a specific wavelength and the corresponding points of the theoretical random curve represents the value b for the same wavelength considered.
It is possible to calculate the percentage of monomeric form for a given wavelength and for each concentration considered for the experiments so far described.
The equation for 196 nm wavelength is:

$$([\theta]_{196nm} - b) \bullet 100 / (a-b) \qquad (9)$$

The calculated percentages are reported in table 2.I.3.

**Table 2.I.3:** Calculated values of monomeric form percentage for each CARD concentration. Low concentrations correspond to a higher monomeric content and higher concentrations to a lower monomeric and, consequently, higher dimeric form.

| **Cell path (cm)** | 0.01 | 0.05 | 0.10 | 0.20 | 0.50 | 1.00 | 5.00 |
|---|---|---|---|---|---|---|---|
| **[CARD] (M)** | $2.4 \cdot 10^{-4}$ | $4.8 \cdot 10^{-5}$ | $2.4 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $4.8 \cdot 10^{-6}$ | $2.4 \cdot 10^{-6}$ | $4.8 \cdot 10^{-7}$ |
| **%Monomer** | 43% | 48% | 58% | 69% | 81% | 87% | 95% |

From these calculations, it is possible to deduce that CARD domain is present in solution in a monomer-dimer equilibrium which is strongly dependent from the concentration. For higher concentrations the dimer population increases with a concomitant increase in unfolded conformer state. These results are supported by the evidence that CARD domain increasing concentrations also changes the chromatographic behaviour, during CARD purification. Four different gel filtration profiles are reported for CARD applied on the G150 chromatographic column at different concentrations. Each point of the curve is obtained by reporting the absorbance at 280 nm against the elution volume.

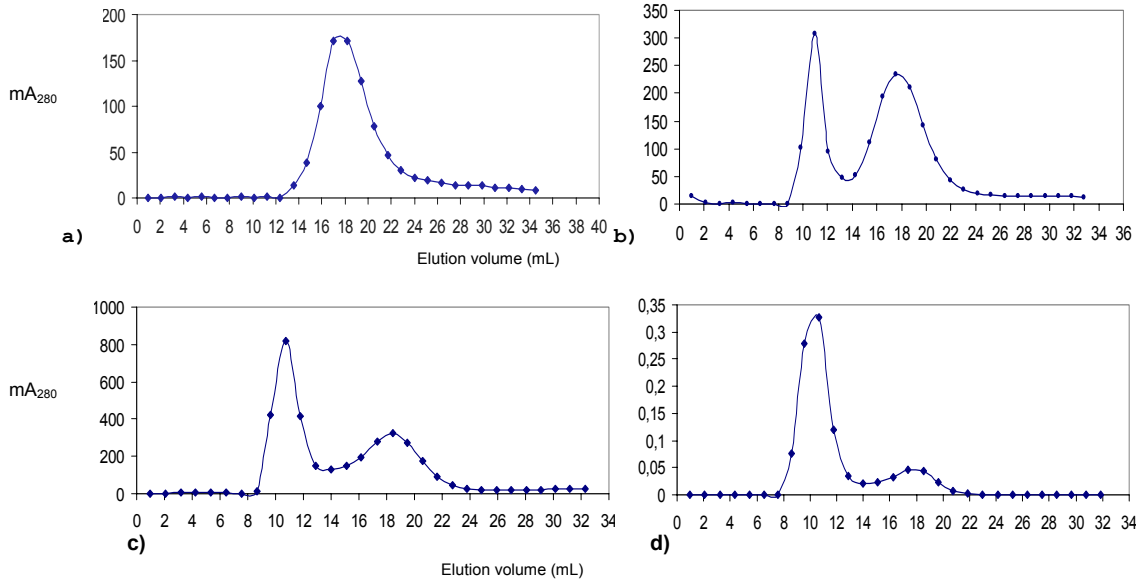



**Figure 2.I.10: a)** Bcl10 CARD 0.20 mg/mL. **b)** Bcl10 CARD 0.66 mg/mL. **c)** Bcl10 CARD 1.78 mg/mL. **d)** Bcl10 CARD 3.50 mg/mL.

As shown in figure 2.I.10, the CARD domain comes out from the gel filtration column as a single species, with an elution approximately corresponding to the molecular weight of the domain in monomeric form. By increasing the concentration from 0.2 mg/mL ($1.2 \cdot 10^{-5}$ M) to 3.50 mg/mL ($2.1 \cdot 10^{-4}$ M), the increasing of a second peak, at higher molecular weight, is observed. The second peak is approximately corresponding to the dimer. According to the dimerization study previously reported, about 70% monomeric form is expected for a concentration corresponding to the first chromatogram (a) and a percentage of monomer of 40% in the last chromatogram

considerated (d), which means to have about 60% dimeric form. The integration of the areas of the peaks obtained from CARD gel filtration, at different concentration, gives approximately the same percentages of monomeric and dimeric species calculated with the equation (9).

### 2.I.3.2 Characterization of protein binding sites

In order to produce more evidences of the self-association occurring between CARD molecules, preliminary ELISA assay experiments were performed by immobilizing the CARD domain on the ELISA plate and verifying its binding to the biotin-labelled CARD, using the biotin-streptavidine system to detect the association. Increasing concentrations biotinilated CARD (up to 50.0 µg/mL, in buffer sodium acetate 50 mM pH=4.0), were applied to the unlabelled CARD in concentration 1.0 µg/mL, 0.5 µg/mL and 0.1 µg/mL, respectively and the absorbance at 490 nm was measured at the UV spectrophotometer. The values obtained are reported in figure 2.I.11.

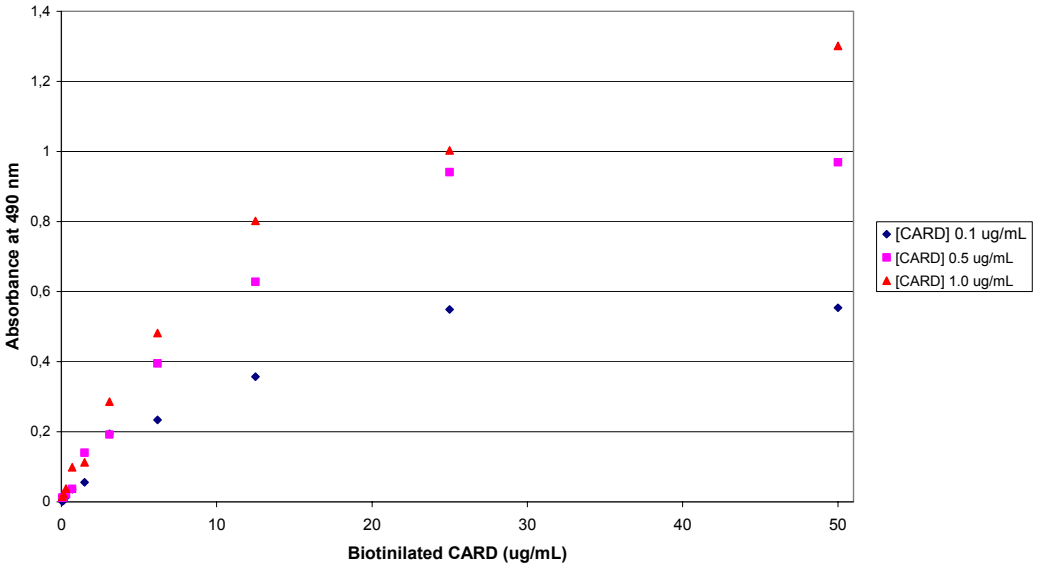



**Figure 2.I.11.** ELISA preliminary assay for CARD/CARD binding. The absorbance intensity measured at 490 nm increases for increasing labelled CARD concentrations. The saturation occurs over a 0.5 µg/mL CARD and 12.5 µg/mL biotinilated CARD (0.63 $Abs_{490}$).

The absorbance, measured at 490 nm, increases by adding increasing concentrations of labelled CARD to the unlabelled domain. This preliminary experiment was due to test the better experimental conditions in which CARD is able to self-associate. Over 12.5 µg/mL labelled CARD added, the curve represents the saturation condition. The condition 0.5 µg/mL CARD has been chosen from the later experiments.

Furthermore, in order to identify the specific regions involved in CARD self-association, a proteolytic digestion with trypsin was carried out for the CARD domain to obtain fragments to conveniently employ in a competition assay against the whole CARD. The tryptic digestion mix was applied on a RP-HPLC column (data reported in

section 2.I.2). The fractions, collected from the tryptic digestion and named from 1F to 11F, were tested by competitive ELISA assay to investigate which one contains fragments binding to the CARD domain. Unlabelled CARD was immobilized on the ELISA plate and the competition mix was applied to the CARD.
The competition mix is composed of each one of the tryptic fragments and biotinilated CARD in two different ratio: 1:1 (w/w) and 10:1 (w/w) in sodium acetate 50 mM pH=4.0. The labelled CARD concentration used in this assay, was chosen, before the saturation point value (the pre-saturation conditions, determined in the preliminary ELISA test, are corresponding to biotinilated CARD concentration of 0.5 µg/mL). In figure 2.I.12 the absorbance at 490 nm is reported for each one of the 11 tryptic fractions assayed.

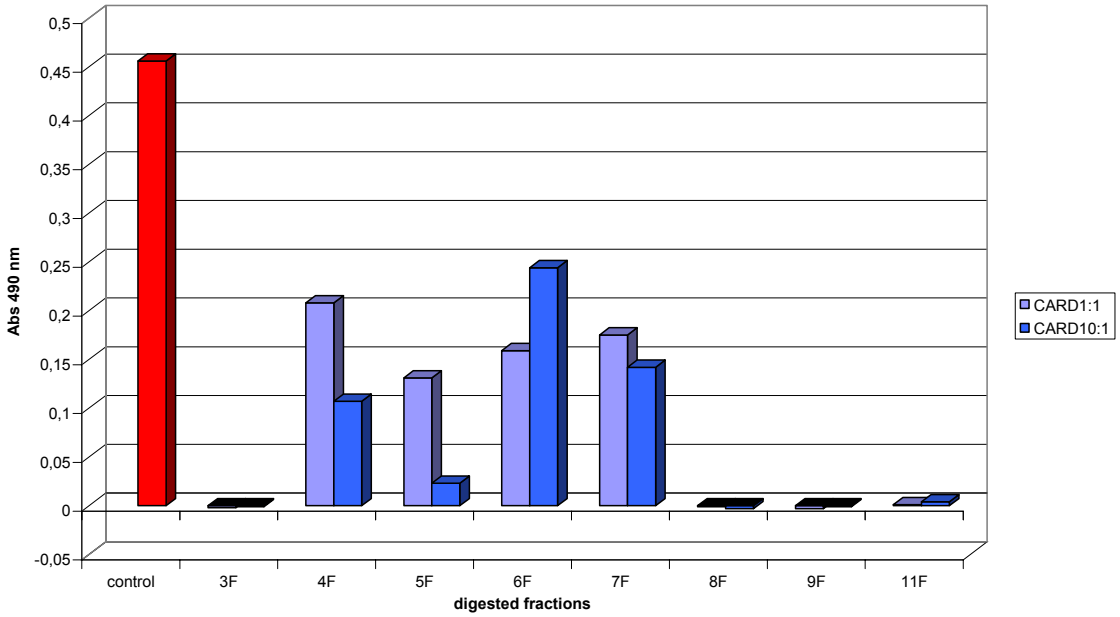



**Figure 2.I.12.** Competitive ELISA assay. The histogram shows the competition assay of eleven CARD fractions obtained from tryptic digestion with CARD domain. For the fractions showing lower absorbance the strongest competition is observed. Peaks in light violet correspond to 1:1 (w/w) CARD/fragment ratio and in blue is shown to 10:1 (w/w) CARD/fragment ratio.

The first peak of the histogram is corresponding to the only labelled CARD applied to the immobilized CARD. It represents the positive “control” for the CARD/CARD binding. The assayed competition mix showing a higher absorbance at 490 nm are corresponding to a positive binding between CARD and its biotin-labelled homologous, detected through the system biotin-streptavidin. The fractions competing in this binding show the lowest absorbance at 490 nm because the biotin-label is not detected. It means that biotinilated CARD, in these competition mix, has been brought away from the washing buffer because it is not bound to the CARD-containing well surface. From this assay the fractions 1F, 3F, 8F, 9F and 11F were selected. The fraction shows lower binding properties. In summary, this competition assay shows the attitude of the CARD fractions, obtained from CARD tryptic digestion, to bind the whole CARD domain. Where a fragment binds the unlabelled CARD, immobilized on the ELISA well surface, the biotinilated CARD cannot bind. This results in a decreasing signal of absorbance measured at 490 nm.

Furthermore, all the digested fractions were characterized by LC-MS to identify the CARD region sequences responsible for this competitive binding. The LC-MS profile and the corresponding peaks identified by using the program MS-Digest for trypsin digestion cleavage (section 2.I.2 table 2.I.1), are reported in the following figure.

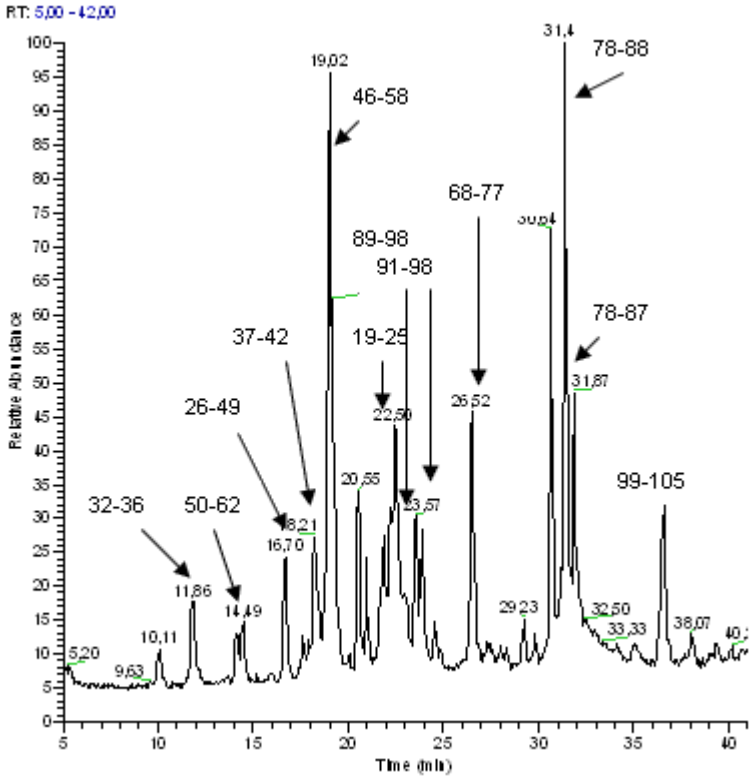



**Figure 2.I.13:** LC-MS chromatogram of digested CARD domain with the identified sequences.

For the samples collected as fractions 1F, 2F and 10F it was not possible to assign any sequence within the "full length" sequence considered. They are most likely impurities. The most interesting samples in this study are the fractions 3F, 8F, 9F and 11F collected from the digestion mix, because they show a competitive binding in competitive ELISA assay. With the aim to identify the sequences of the tryptic fragments contained in these fractions, and to further investigate their binding attitudes, they were characterized by LC-MS.
As reported in the chromatograms in figure 2.I.13, the digestion fractions named as 3F contains three identified sequence fragments: CARD [26-49], CARD [32-36] and CARD [50-62]. The fraction, named as 8F shows two main peaks, identified as CARD [89-98] and CARD [91-98]. For the fraction named as 9F, one sequence was identified: CARD [68-77] and for the fraction 11F the sequence CARD [46-58] was identified.

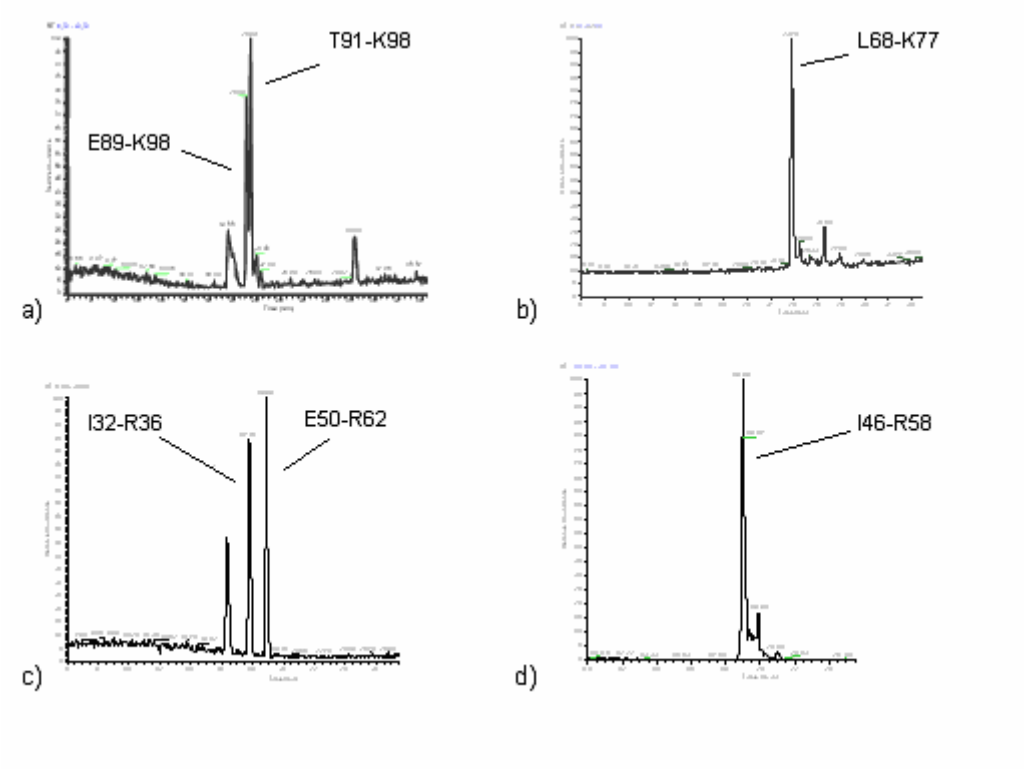



**Figure 2.I.14:** LC-MS chromatograms belonging to the fractions 8F(a), 9F(b), 3F(c) and 11F(d). The fragments contained in these fractions were identified as CARD [89-98], CARD [91-98], CARD [68-77], CARD [26-49], CARD [32-36], CARD [50-62] and CARD [46-58].

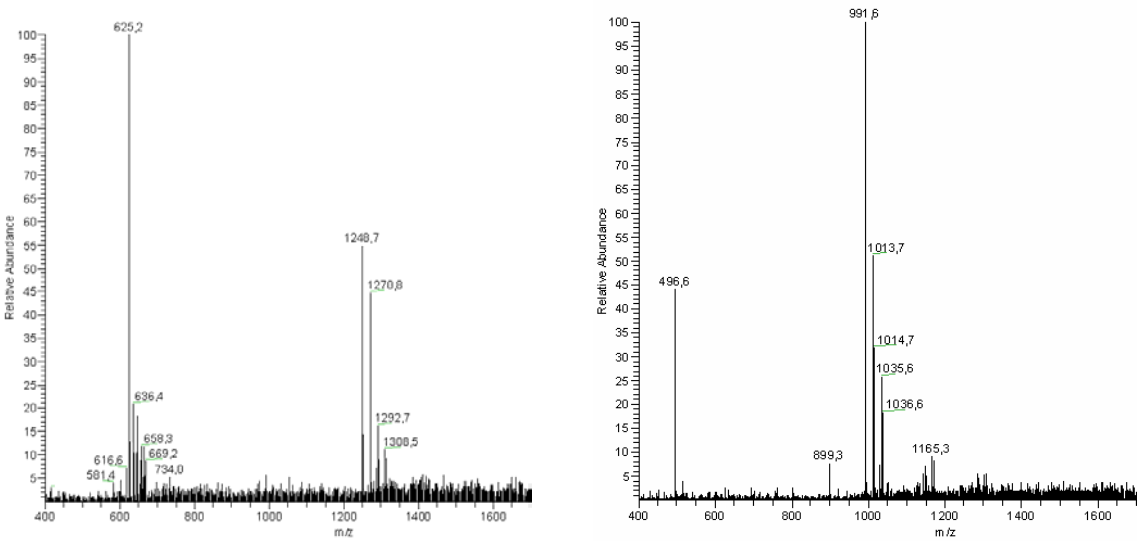



**Figure 2.I.15:** MS spectra of the fraction 8F. a) Fragment E89-K98. $[M+2H]^{2+}$ = 625.2 m/z, $[MH]^{+}$ = 1248 m/z . b) Fragment T91-K98. $[M+2H]^{2+}$ = 496.7 m/z, $[MH]^{+}$ = 991.6 m/z

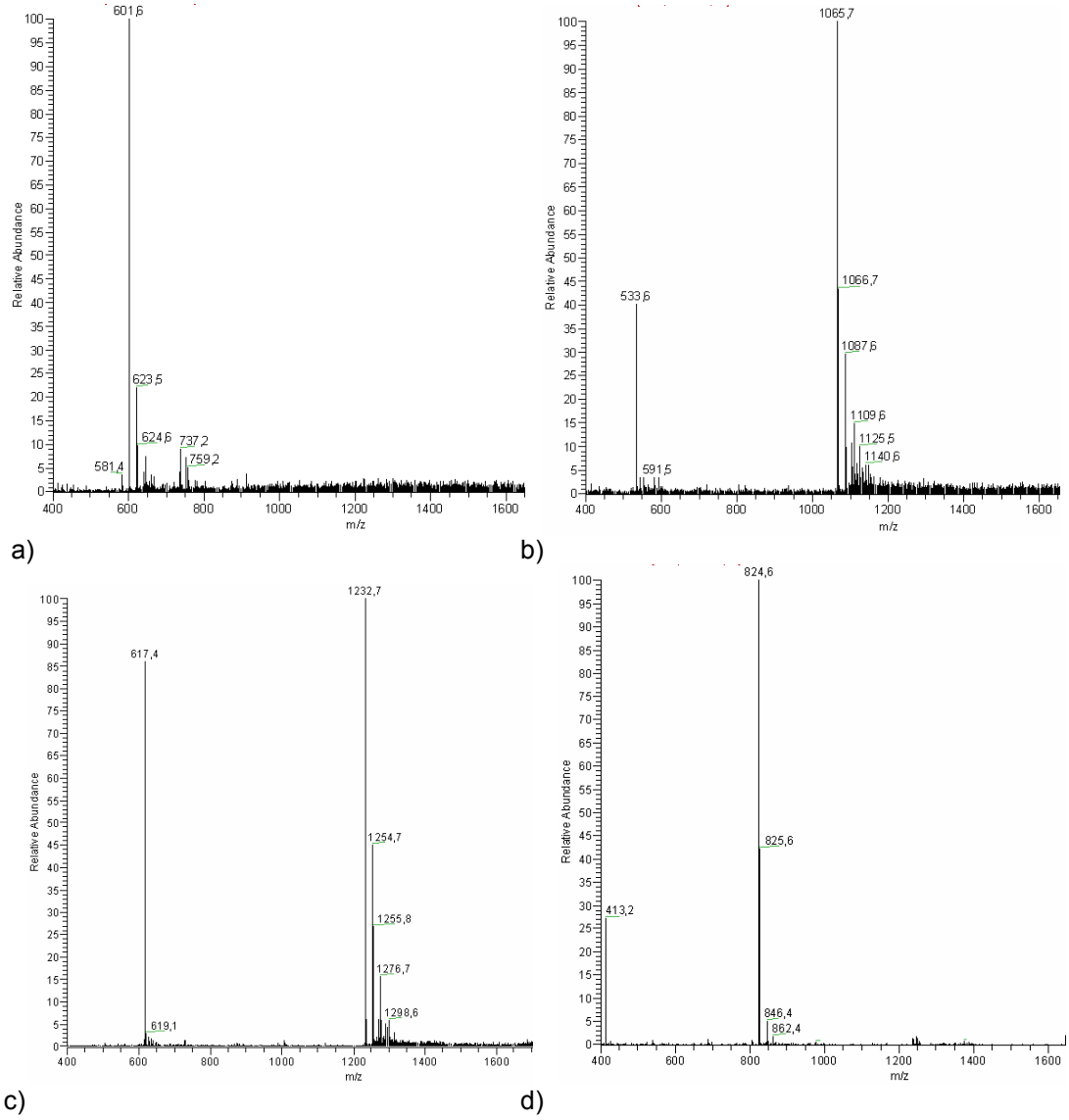



**Figure 2.I.16:** MS spectra of the fractions 3F, 9Fand 11F. a) Fraction 3F, fragment I32-R36, $[MH]^+$ = 601.6 m/z and b) fragment E50-R62, $[M+2H]^{2+}$ = 533.6 m/z, $[MH]^+$ = 1065.7 m/z. c) Fraction 9F, fragment L68-K77, $[M+2H]^{2+}$ = 617.4 m/z, $[MH]^+$ = 1232.7 m/z. d) Fraction 11F, fragment I46-R58. $[M+2H]^{2+}$ = 413.2 m/z, $[MH]^+$ = 824.6 m/z.

Once the “CARD binding fragments” were identified, their sequences were obtained as synthetic peptides in order to test the identified regions against the whole CARD, to verify their binding properties. For the first fragment, belonging to the fraction 8F, its related longer sequence peptides, CARD [78-98] and CARD [68-98] were also synthetized. The fragment CARD [68-77] is belonging to the collected fraction 9F and the fragment CARD [26-49] is belonging to the fraction 3F.

**Synthetized sequences**:

CARD [91-98]: TQNFLIQK
CARD [78-98]: GLDTLVESIRREKTQNFLIQK
CARD [68-98]: LLDYLQENPKGLDTLVESIRREKTQNFLIQK
CARD [68-77]: LLDYLQENPK
CARD [26-49]: VYLSEKIIAERHFDHLRAKKILSR

## Competitive ELISA assay with CARD synthetic fragments

Synthetic fragments CARD [91-98], CARD [78-98], and CARD [68-98], were assayed in a competitive ELISA to test their competing properties in CARD/CARD association. The assay was carried out by following the same procedures reported for the competitive assay of CARD tryptic fractions. The results are shown in figure 2.I.17.

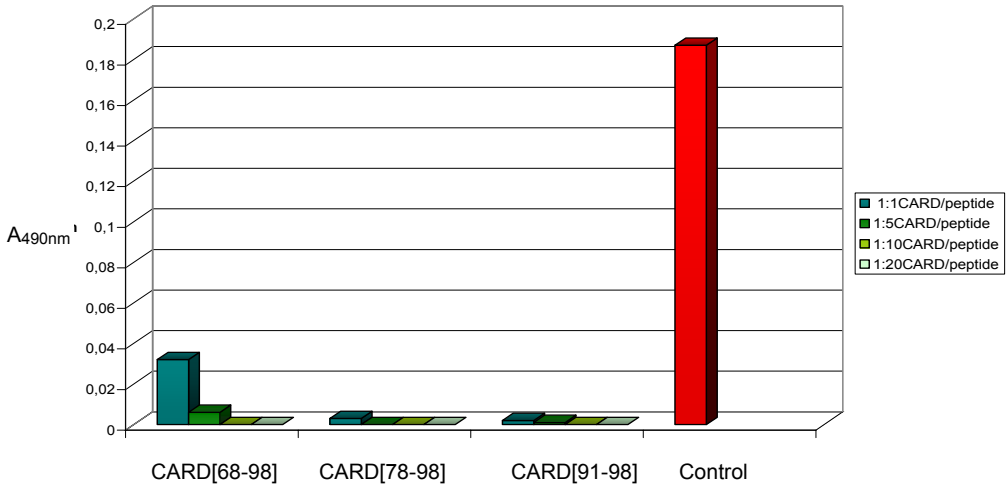



**Figure 2.I.17:** Competitive ELISA assay with synthetic fragments. The fragments CARD[78-98], and CARD[91-98], show stronger competition properties in the whole CARD domain binding.

For this assay, four different concentrations of synthetic fragments were used in the competition mix: 1:1, 1:5, 1:10 and 1:20 ratio CARD-peptide. The results shown in figure 2.I.17 evidenced that the peptide competition is relatively dependent to the concentration used. Higher concentrations peptide in the competition mix result in a higher competitive effect in CARD/CARD association. The longer sequence region is less competitive in the assayed binding, while, the two shorter peptides have stronger competition properties. In order to relate the peptides binding properties to their structural behaviour, a circular dichroism characterization was performed.

## CD characterization of CARD fragments

Once the sequences involved in CARD domain self-association were identified, synthetically produced, and tested again to provide the evidence of their attitude to compete in CARD-CARD interactions, circular dichroism experiments were carried out to investigate the structural behaviour of these fragments. CD experiments were performed in buffer solution at pH=4.5, which is the pH value used for all the experiments with CARD domain, and pH=7.5, the physiological condition. For both the peptides CARD[78-98] and CARD[68-98], the structural behaviour does not

change in dependence of pH. For the peptide CARD [78-98] a typical α-helical folding, with two negative bands centered at 208 nm and 222 nm is observed.

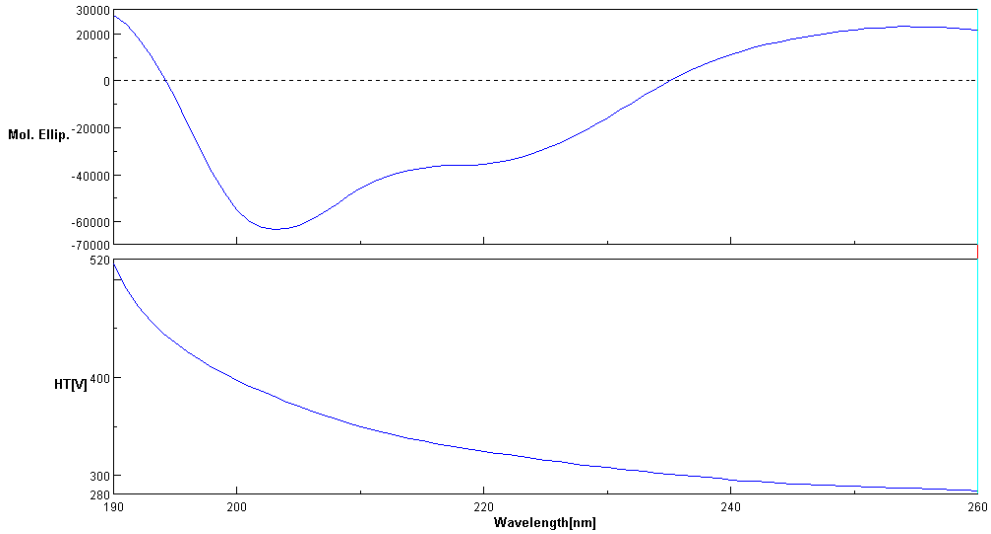



**Figure 2.I.18:** Far-UV CD spectrum of the synthetic fragment CARD [78-98] acquired at room temperature and pH=4.5.

This folding is quite stable. An increase of the temperature up to 40°C does not significantly modify the peptide structural behaviour.
The synthetic fragment CARD [68-98], in the same CD conditions tested for its longer analogue, shows a “random” behaviour, with negative band at 200 nm and no positive bands in the wavelength region between 190 nm and 200 nm.

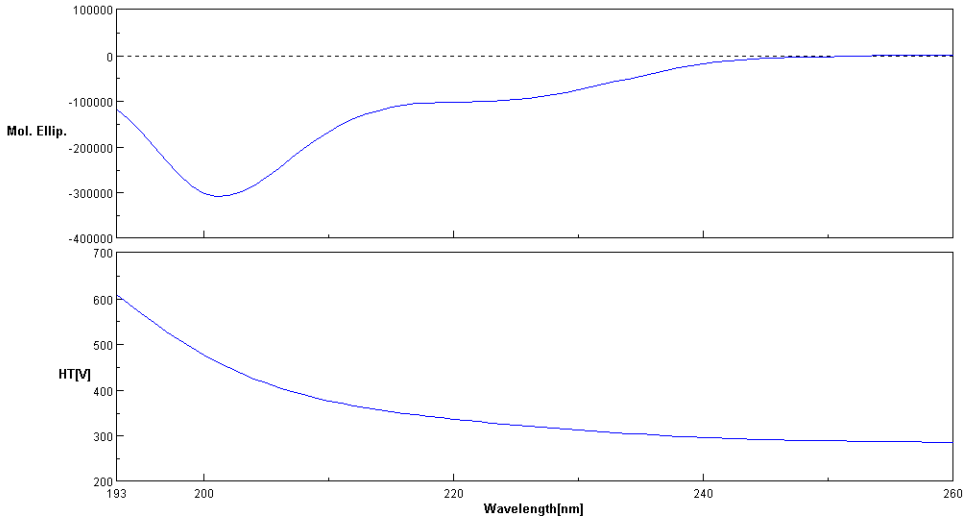



**Figure 2.I.19:** Far-UV CD spectrum of the synthetic fragment CARD [68-98] acquired at room temperature and pH=4.5.

These CD experiments suggest that the longer synthetic CARD fragment, is rather unfolded in solution. It is also the weakest competing in ELISA competition assay of CARD-CARD association. The shorter sequence, CARD [78-98], shows a prevalently α-helical folding and also shows strong competition properties in CARD-CARD self-association assay. CD experiments on the stronger binding peptide in ELISA competition assay, the CARD[91-98] are still in progress.

In order to verify the reliability of the CD experimental results, a sequence aligment was performed between Bcl10 CARD sequence and two similar CARDs (from Apaf1 [25] and RAIDD [33]), whose NMR structure are availeble in Protein Data Bank.
RAIDD CARD domain, shows a higher omology (~26%) with the CARD studied in this thesis work. From this model structure, it was possible to predict the synthetic peptide structural behaviour in the whole CARD sequence and, thus, compare it to the structural preferences evidenced in the CD experiments here reported.

**Figure 2.I.20:** The CARD[78-98] peptide (in blue) location in the CARD domain. The image is obtained by using weblab wiewer Lite program, considering one of the 15 RAIDD CARD domain NMR structures (PDB entry: 3CRD ) as model.

The synthetic peptide CARD [78-98] is predicted to have a prevalently α-helical folding, and should form an helix-loop-helix within the CARD sequence as shown in figure 2.I.20. The results obtained from the circular dichroism characterization of this peptide are in agreement with the expected structural behaviour reported in the model.

**Preliminary NMR studies**

Small peptide systems provide useful models for the elucidation of fundamental principles of protein folding and stability, and the conformational analysis of a peptide can provide insight into relationship between the three dimensional structure of a specific region in solution and its biological activity.
Preliminary experiments were carried out to look for the best conditions for the NMR conformational characterization of CARD[78-98].
Different concentrations of the peptide (1.0 mM, 0.3 mM, 0.1 mM) were previously assayed in neat $H_2O$ (10 % $D_2O$), at pH=3.5 and 25°C. In these experimental conditions, 1D $^1H$ spectra showed changing of intensity and multiplicity of signals in dependance of time. Indeed, spectra acquired for these preparations of the peptide after one, two and three days, showed different signals. This behaviour evidenced that the peptide CARD[78-98], may exist as a complex mixture of different oligomeric states.
In the second set of experiments, the peptide was studied in the lowest concentration possible for NMR experiments at 25°C, in order to avoid any aggregatation. Hence, the synthetic fragment was dissolved in buffer sodium acetate 5.0 mM pH=5.5 to obtain a concentration of 0.1 mM. In this condition 1D spectrum showed well resolved resonances and no aggregation behavior (figure 2.I.21).

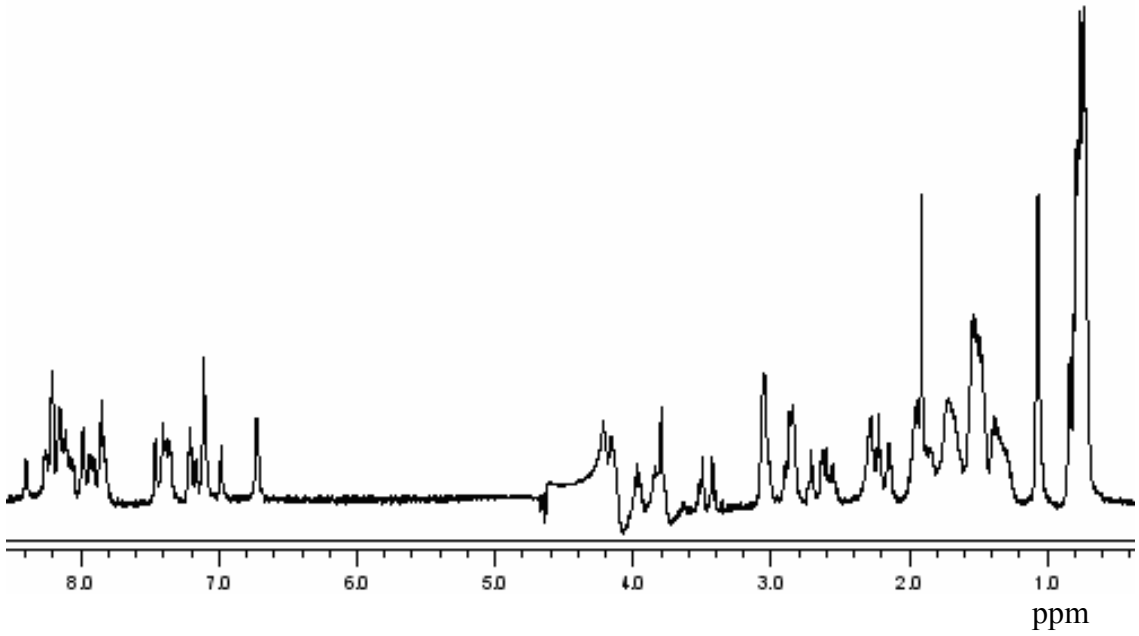



**Figure 2.I.21**: 1D $^1H$ NMR spectrum of CARD[78-98] peptide 0.1 mM in buffer 5.0 mM Na acetate pH=5.5, 10% $D_2O$.

Furthermore, the same preparation was used to acquire spectra after one, two and three days storage at 4°C and no changes were observed in the signals. These conditions were chosen to set up further two dimensional NMR experiments in order to determine peptide structure.

## 2.II. Protein-cell wall interactions: dormancy in *Mycobacterium tuberculosis*

### 2.II.1 Introduction

An interesting challenge is the study of the cellular mechanisms mediated by proteinaceous factors and their interactions with the cell wall. Recent studies, devoted to the pursuit of chemicals that might give rise to the development of new drugs, are focused on these interactions. Many pathologies, such as those caused by *Streptococcus pneumoniae* [34] and *Mycobacterium tuberculosis* are related to interactions between proteins and bacterial cell wall because of their involvement in antibiotical resistance. Because of drug resistance, these pathologies are now more difficult to treat than they were just a few decades ago. The antibiotics may work by interfering in many metabolic activities of the bacterial cell such as cell wall synthesis, protein or nucleic acid synthesis, or alter cell wall permeability. Since bacteria have a cell wall made up of repeating units of peptidoglycan and human cells lack this feature, it would seem that the bacterial cell wall presents an ideal target for chemotherapy.

Among infectious bacterial pathogens, *Mycobacterium tuberculosis* (MTB), is one of the biggest killers, with 2-3 millions deaths per year. One third of the world's population is infected with *tuberculosis*. This is the leading infectious disease cause of death and represents more than a quarter of the world's preventable deaths. In the lung, the organism is taken up by alveolar macrophages and carried to lymph nodes, from where it may spread to multiple organs. Two to eight weeks after infection, cell mediated immunity (CMI) and hypersensitivity develop leading to the characteristic reaction to the tuberculin test and, in immunocompetent individuals, containment of infection. Uncontrolled *M. tuberculosis* growth in human host is associated with extensive lung damage that ultimately causes death by suffocation due to insufficient oxygen [35].

*Mycobacteria* are usually classified as Gram-positive (no outer cell membrane), non-motile, pleomorphic rods, related to the *Actinomyces*. Most *Mycobacteria* are found in habitats such as water or soil. However, a few are intracellular pathogens of animals and humans. *Mycobacterium tuberculosis*, along with *M. bovis*, *M. africanum*, and *M. microti* all cause the disease known as *tuberculosis* (TB) and are members of the *tuberculosis* species complex. Each member of the TB complex is pathogenic, but *M. tuberculosis* is pathogenic for humans while *M. bovis* is usually pathogenic for animals.

Tuberculosis is currently treated by combining 3-4 drugs with different properties: isoniazid, rifampin, streptomycin, as antibacterial activity and a combination of isoniazid, rifampin, ethambutol to inhibit the development of resistance.

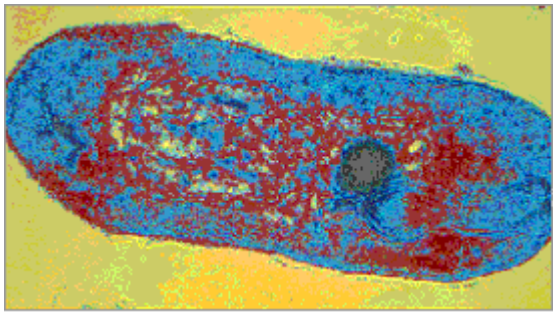


**Figure 2.II.1:** Electron micrograph of *M. tuberculosis*

To rationally develop new antitubercular agents, it is important to understand the MTB-host interaction to learn how these bacteria circumvent host defences and cause disease.
One of the key events of MTB's pathogenity is its ability to persist in its host organism in a latent state for years after the first phase of infection [35].
Actively dividing MTB infection can be cleared by patient's immune response, but once the dormant state has been entered the bacteria are no longer cleared. What are the bacterial determinants necessary for persistent infection and what is the physiological state of tubercle bacilli during latent infection remain unclear. This clinical latent state of infection is usually called "dormancy". Understanding and controlling the entry and exit from dormancy is important in the development of new anti-tubercular therapies. Dormancy has been described as a state of low metabolic activity in which cells are alive but fail to grow under standard culture conditions.
There is ample circumstantial evidence from observation of *tuberculosis* in humans and animals that it is capable of adapting to prolonged periods of dormancy in tissues, becoming responsible for latency of the disease itself. Furthermore, the dormant bacilli are resistant to antimycobacterial agents. Although *tuberculosis* persistence and reactivation play significant roles in the pathogenesis of the disease, relatively little is known regarding the molecular mechanism which regulate mycobacterial physiology in these disease states of persistence [36-38].
It has been suggested that the ability of MTB to shift down into nonreplicating stages is responsible for the ability of tubercle bacilli to lie dormant in the host for long periods of time, with the capacity to revive and activate disease at a later time, when conditions are more favourable [39]. The latency of the bacterium has been extensively studied but reactivation from this state of dormancy is still an important field of investigation. A probable mechanism of exit from the latent state involves, for MTB, an attack of the bacterial cell wall by a protein component, denoted as Resuscitation Promoting Factor (Rpf). Indeed, it has been demonstrated, both for MTB and for the bacterium *Micrococcus luteus* that these Rpfs are responsible of increasing the number of colony-forming units when added to dormant cultures and is active at picomolar concentrations [40-43]. MTB encodes a family of five proteins (Rpf A-E) which share homology with the Rpf found in *M. luteus*. *In vitro* studies of *M. luteus* reactivation show that MTB's Rpfs are potential regulators of persistence and reactivation [40, 44]. All the Rpfs share a common core domain located in the C-terminal region of the proteins.

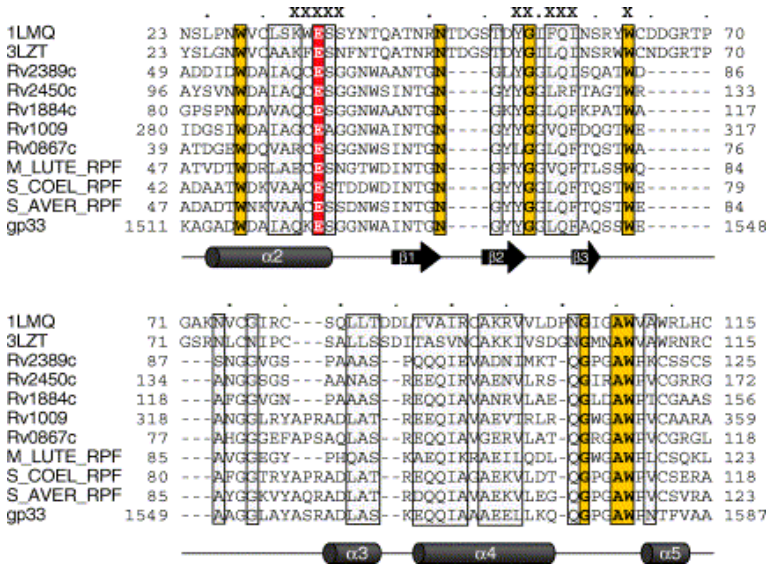


**Figure 2.II.2**. Multiple sequence alignment of Rpf domain region and c-type lysozymes. Structure codes are from the Protein Data Bank (http://www.rcsb.org/pdb/) and sequence codes are from the *Mycobacterum tuberculosis* genome for Rv2389c, Rv2450c, Rv1884c, Rv1009 and Rv0867c.

The solution NMR structure for the 108-residues domain of RpfB from *M. tuberculosis* has been solved by heteronuclear multidimensional NMR. Such a structure has been compared to that of the c-type lysozyme.

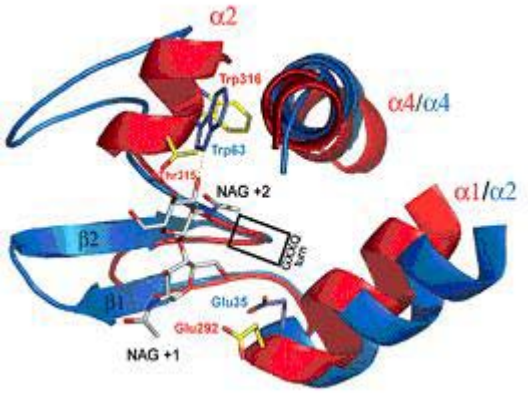



**Figure 2.II.3 .** Comparison of RpfB catalytic C-terminal domain and lysozyme catalytic site .(**a**) Structure superimposition in the cleft area of the RpfBc (red) and a c-type lysozyme (PDB entry 1LMQ; blue). The catalytic glutamate at the C-terminal end of the first helix is at the same position in RpfBc and lysozyme (Glu292 and Glu35, respectively).

The similarity of this region domain to lysozyme suggests that RpfB has a catalytic mechanism common to lysozyme and, thus, a muralytic activity.
Biochemical studies indicate that the conserved active site glutamate is important for the resuscitation of dormant cultures of *M. luteus*.

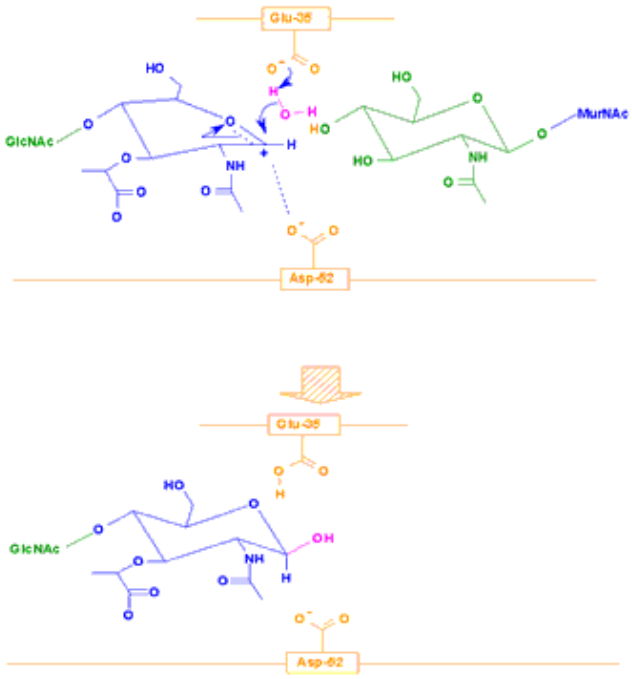



**Figure 2.II.4:** Catalytic mechanism of lysozyme.

It has been also reported that the catalytic site of Rpf B can bind the oligosaccharide N, N', N"-triacetylchitotriose (tri-NAG), the trimer of N-acethylglucosamine (NAG), one of the two sugars that make up the polysaccaride portion of peptidoglycan [43].

This evidence suggests that the oligosaccharide cleavage is probably the signal for revival from dormancy and the glutamate is essential for *in vivo* activity.
One hypothesis about the correlation between muralytic activity and resuscitation, is that the products of such hydrolysis incite cells to divide but what is the nature of the signal that those proteins generates and how is the signal relayed to or detected from the cells is still to be clarified. It has been also proposed that the substrate of Rpfs catalytic domain may be a modified form of peptidoglycan found only in dormant bacteria. Any modification of the peptidoglycan that confers resistance to antibacterial enzymes as lysozymes and amidases could have a role in pathogenesis and such a modification could play a similar role in MTB virulence. So far no crystal structures have been reported in literature for any of the Rpfs. In addition, none of the Rpf forms from *M. tuberculosis* have shown to display muralytic activity.

Several Rpf forms from *M. tuberculosis* have been studied in this thesis work. These studies are aimed at a deeper characterization of both structural and functional properties of these proteins. Understanding the properties of these proteins in relation to their interactions with *M. tuberculosis* cell wall is of fundamental importance for the study of reactivation of the bacterium itself. In turn, this step provides the bases for the identification of molecules which are able to inactivate Rpf and, therefore, to keep the disease in a dormant state.

## 2.II.2 Methods

### Choice of host for protein amplification

The use of recombinant proteins is generally preferred to facilitate purification and detection of the recombinant proteins. The more that is known about the characteristics of a protein, the more easily it can be isolated and purified. Consequently, fusion proteins are simple and convenient to work with and, for many applications, a single purification step, using a commercially available affinity chromatography column, is sufficient. The choice of host depends upon the specific requirements and applications for the recombinant protein. Bacteria like *Escherichia coli* are commonly employed because many references and much experience is available on them, a wide choice of cloning vectors is possible and they are organisms easy to grow with high yields. Therefore, all studies here reported on Rpf proteins have been carried out using *Escherichia coli* as a host.

### PCR protocol used for Rpfs cloning

**Choice of the primers**: PCR primers are usually 15-30 nucleotides in length. Longer primers provide higher specificity. The GC content should be 40-60%. More than three G or C nucleotides at the 3'-end of the primer should be avoided, as nonspecific priming may occur. The primer should not be self-complementary or complementary to any other primer in the reaction mixture, in order to avoid primer-dimer and hairpin formation. The melting temperature of flanking primers should not differ by more than 5°C, so the GC content and length must be chosen accordingly.
The rpf B gene was amplified from MTB genomic DNA using 5 pmol of each primer described in table 2.II.1.
**Estimation of the melting and annealing temperatures of primer**: if the primer is shorter than 30 nucleotides, the approx. melting temperature (Tm) is calculated using the following formula: Tm= 4 (G + C) + 2 (A + T). Annealing temperature should be approx. 5°C lower than the melting temperature.
**$MgCl_2$ concentration**: since $Mg^{2+}$ ions form complexes with dNTPs, primers and DNA templates, the optimal concentration of $MgCl_2$ has to be selected for each experiment. Too few $Mg^{2+}$ ions result in a low yield of PCR product, and too many increase the yield of non-specific products The recommended range of MgCl2 concentration is 1-4mM, under the standard reaction conditions specified.
**dNTPs**: the concentration of each dNTP in the reaction mixture is usually 200 µM. It is very important to have equal concentrations of each dNTP (dATP, dCTP, dGTP, dTTP), as inaccuracy in the concentration of even a single dNTP dramatically increases the misincorporation level.
**Taq DNA Polymerase:** usually 1-1.5u of Taq DNA Polymerase are used in 50µl of reaction mix. Higher Taq DNA Polymerase concentrations may cause synthesis of nonspecific products.
**Cycling Conditions:** the complete denaturation of the DNA template at the start of the PCR reaction is of key importance. The initial denaturation should be performed over an interval of 1-3min at 95°C if the GC content is 50% or less. This interval

should be extended up to 10 min for GC-rich templates. Usually denaturation for 0.5-2min at 94-95°C is sufficient. If the amplified DNA has a very high GC content, denaturation time may be increased up to 3-4min. The primer annealing step is carried out at temperatures 5°C lower than the melting temperature of primer-template DNA duplex. Incubation for 0.5-2min is usually sufficient. The extending step is performed at 70-75°C. The rate of DNA synthesis by Taq DNA Polymerase is highest at this temperature. Recommended extending time is 1min for the synthesis of PCR fragments up to 2 kb. When larger DNA fragments are amplified, the extending time is usually increased by 1min for each 1000 bp. The number of PCR cycles depends on the amount of template DNA in the reaction mix. For less than 10 copies of template DNA, 40 cycles should be performed. If the initial quantity of template DNA is higher, 25-35 cycles are usually sufficient. After the last cycle, the samples are usually incubated at 72°C for 5-15min to fill-in the protruding ends of newly synthesized PCR products. For Rpfs cloning, the following PCR cycle program was used: 95°C for 2 min, then 30 cycles of 95°C for 30 s, 56°C for 30 s and 72°C for 1 min and then 72°C for 10 min. using 200 µM dNTPs mix, 2.5 U standard Taq polymerase, 200 ng of template, 4 µM $MgCl_2$ for a total reaction mixture of 50 µL. The amplified PCR samples were purified using preparative agarose electrophoresis (1% agarose gel).

**Table 2.II.1:** Rpfs used primers and protein information. *Rpf $B_{cd}$ refers to the catalytic domain of RpfB.

| Protein | Gene name and length | Protein length (MW) | P.I. | Primers |
|---|---|---|---|---|
| **Rpf B** | Rv1009 (999bp) | 333 a.a.(35162 Da) | 5.14 | 5'-CAT GCC ATG GGA GTC GAC GGA ACC GCG ATG-3'<br>5'-CCC AAG CTT ATC AGC GCG CAC CCG CTC-3' |
| ***Rpf $B_{cd}$** | Rv 1009 (255 bp) | 85 a.a. (9001 Da) | 4.85 | 5'-CAT GCC ATG GGA ATC GAC GGA AGC ATC TGG GA-3' |
| **Rpf C** | Rv 1884c (441bp) | 146 a.a.(14989 Da) | 9.14 | 5'-GAT GGG TCT CAC ATG ACT TCC GGC GAT ATG TCG AG-3'<br>5'-CCC AAG CTT ATC AGC GCG GAA TAC TTG CC-3' |

## Choice of vectors

In order to clone the gene of interest all engineered vectors have a selection of unique restriction sites downstream of a transcription promoter sequence. The choice of vector family is governed by the host. Once the host has been selected, many different vectors are available for consideration, from simple expression vectors to those that secrete fusion proteins. Fusion His-tag protein production method is the most employed because targeting information can be incorporated into a tag which provide a marker for expression and purification. Purification can be easily obtained by using affinity chromatography under denaturing or non-denaturing conditions. Commonly used vectors for His-tagged fusion proteins are pET vectors containing 6 histidine. In pET vectors, target genes are cloned under control of strong bacteriophage T7 transcription and translation signals, and expression is induced by providing a source of T7 RNA polymerase in the host cell.

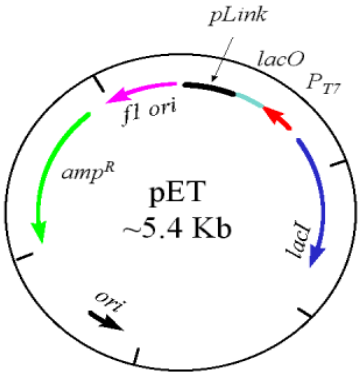



**Figure 2.II.5.** The basic architecture of an *E. coli* expression vector

The plasmid should have an antibiotic-resistance marker and, by supplementing the medium with the appropriate antibiotic, it is possible to kill plasmid-free cells.
The vector should contain a regulatory gene (repressor). The lac-derived promoters are frequently used because they can be controlled by the insertion of a lac-operator sequence downstream the promoter. The origin of replication controls the plasmid copy number. The promotor initiates transcription and is positioned 10-100 nucleotides upstream of the ribosome binding site. The ideal promoter is strong enough to allow product accumulation up to 50% of the total cellular protein, it has a low basal expression level, it is easy to induce. Then the vector contains a transcriptional terminator (that reduces unwanted transcription) and a Shine-Dalgarno sequence which is required for translation initiation and is complementary to the 3'-end of the 16S ribosomal RNA. The consensus sequence is: 5'-TAAGGAGG-3'. It is positioned 4-14 nucleotides upstream the start codon with the optimal spacing being 8 nucleotides. Initiation point of translation is allowed by a start codon (usually ATG or GTG). For the termination of translation there are 3 possible stop codons but TAA is preferred because it is less prone to read-through than TAG and TGA. His-tag fusion proteins vectors have a specific site of seven residues sequence (Glu-X-X-Tyr-X-Gln/Ser) recognised by TEV protease. This cleavage allows tag removal after purification. The cleavage site should be surface exposed. The vector should also contain the restriction enzymes cleavage sites to allow insertion of the gene, which must be cleaved with the same enzymes. Restriction sites are specific target sequences which are palindromic (both strands have the same nucleotide sequence but in antiparallel directions) and this is one of many features that makes them suitable for DNA manipulation. Any DNA molecule will contain the restriction enzyme target just by chance and therefore may be cut into defined fragments of size suitable for cloning.

PCR Rpf B and Rpf $B_{cd}$ products were digested with 1 U Nco I and 1 U Hind III with the appropriate buffer (according to the manufacturer protocols).
Nco I cleaves into the sequence C^CATGG and Hind III cleaves into the sequence A^HGCTT. The digestion mix was incubated for 1 hour at 37°C and purified (Qiagen's QiaQuick purification kit). The gene was cloned into the corresponding sites of pETM-11 vector (previously digested with the same enzymes in order to

obtain the complementary restriction sites) to produce N-terminally His-tagged Rpf protein. Similar conditions were used for Rpf C cloning. However additional cloning using the pET-20 vector was also performed for Rpf C.
pETM-11 is used to obtain $(His)_6$-tag fusion proteins while pETM-20 produces $(His)_6$-tag and thioredoxin fusion proteins. Expressing a protein with a fused thioredoxin is preferable to increase protein solubility. As the yield of protein Rpf C obtained by cloning rpf C gene in pETM-11 was not high because of solubility problems, the same gene was also cloned in pETM-20. The complete map of both the vectors is reported in figure 2.II.6.

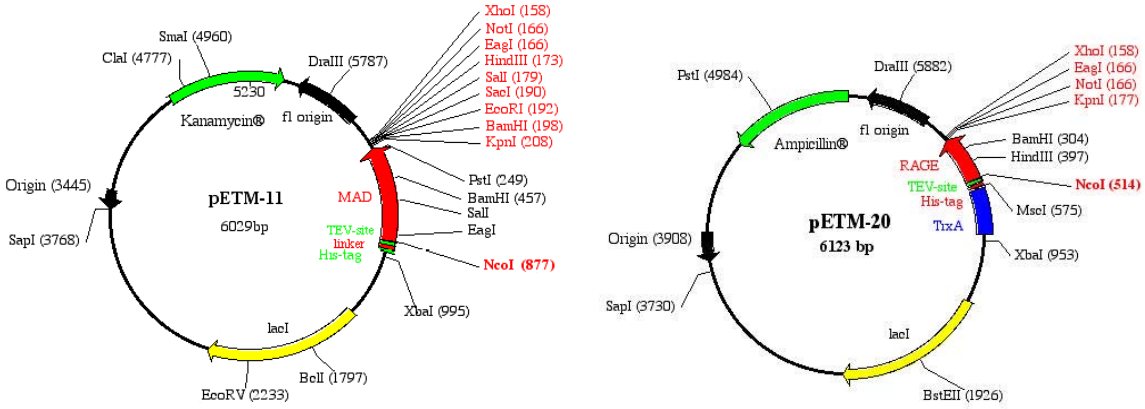



**Figure 2.II.6:** pETM-11 and pETM-20 vectors maps available from the website (http://www.embl hamburg.de/~geerlof/webPP/vectordb/bact_vectors/maps_seqs_mcs/pETM.pdf).

## Expression

A complete expression screening was carried out by transforming rpf B and rpf $B_{cd}$ in pETM-11 in different *E. coli* strains: BL21 (DE3), BL21 RP, BL21 RIL, BL21 pLysS, BL21 pRARE2, Rosetta (DE3) pLysS, BL21 cc1, cc2, cc3, cc4, cc5.
From this screening was selected BL21 (DE3) strain as the best strain to obtain a good expression and higher solubility of the protein. Cultures of Rpf B in BL21 (DE3) were grown in Luria Bertani (LB) medium containing kanamycin. The expression was induced by adding 1 mM isopropyl thio-β-D-galactoside, IPTG (Sigma) after 3 hours incubation at 37°C and at 0.8-1.0 $OD_{600}$. The cultures were grown in a shaking incubator overnight at 25°C up to 6.5-7 $OD_{600}$. Cell suspensions were centrifuged (in a Beckman centrifuge) at 4°C and at 6000 Rpms for 60 min.
The same expression screening was carried out for rpf C cloned in both the plasmids pETM-11 and pETM-20. From this screening was selected BL21 (DE3) strain as the best strain to obtain a good expression and higher solubility of the protein Rpf C in both vectors. Cultures of Rpf C in BL21 (DE3) were grown in LB medium and induced with IPTG after 3-4 hours incubation at 37°C. After induction, the cultures were grown overnight at 25°C up to 5.5-6.0 $OD_{600}$.

## Purification strategy

Protein purification is carried out to isolate the protein of interest from a complex mixture contained in the microbial culture. The protein has to be brought into solution by breaking the tissue or cells containing it (cell lisys). After this extraction process soluble proteins will be in the solvent, and can be separated from cell membranes, DNA etc. by centrifugation. The extraction process also extracts proteases, which will start digesting the proteins in the solution. If the protein is sensitive to proteolysis, it is usually desirable to keep the extract cooled, to slow down proteolysis and use a protease inhibitor. The purification process frees the protein from non-protein parts of the mixture, and finally separate the desired protein from all other proteins. Separation steps exploit differences in protein size, physico-chemical property and binding affinity. Usually a protein purification protocol contains one or more chromatographic steps. Different proteins interact differently with the column material, and can thus be separated by the time required to pass the column and they are detected as they are coming of the column by their absorbance at 280 nm.

## Purification protocol for His-tagged proteins

If the protein of interest has been cloned and expressed as His-tag fusion protein it can be easily selected among the others by using affinity chromatography purification strategy. It consist in chelating the His-tag upon an activated chelating resin by incubating the lisate with affinity media such as NTA-agarose. Chelating Sepharose, when charged with $Ni^{2+}$ ions, selectively retains proteins if histidines are exposed on the protein surface. The polyhistidine-tag binds with micromolar affinity the activated medium. This procedure allows discriminating the selected His-tagged protein while the other proteins that do not specifically interact with the nickel resin are easily eluted with phosphate or TRIS buffer. The washing efficiency can be improved by the addition of 10-20 mM imidazole. His-tag fusion proteins can be desorbed with buffers containing 150-300 mM imidazole. Imidazole competes with his-tags in binding $Ni^{2+}$ and allows protein elution.

## Purification protocol for Rpfs

The pellets were resuspended in buffer 50 mM tris pH= 7.8, 300 mM NaCl, 3 mM imidazol, 10 µg/mL DNase and protease inhibitor (1 mM PMSF protease inhibitor mix to the lysis solution), and lysed by sonication for 5 min. The lysate was centrifuged at 18 000 Rpms for 90 min at 4°C (in a Sorvall centrifuge) to separate the insoluble part (pellet) from the supernatant containing the soluble protein fraction. The supernatant was applied to a $Ni^{2+}$-chelate affinity chromatography column (Hi-trap 5 mL), equilibrated with buffer 50 mM tris pH= 7.8, 300 mM NaCl, 3 mM imidazol, 5% glycerol and washed with 20 column volumes of the same buffer.
To prevent non-specific binding of host cell proteins, 3 mM imidazole was included in the extraction buffer. This procedure is useful when working with $(His)_6$ fusion proteins. The protein was eluted with buffer 50 mM tris pH=7.8, 300 mM NaCl, 300 mM imidazol, 5% glycerol. The eluted fraction were dialyzed against 2 L low salt dialysis buffer 20 mM tris pH= 7.8, 150 mM NaCl, 5% glycerol and treated with TEV protease 1:50 (molar ratio) at 4°C overnight in presence of 1 mM DTT.

TEV protease and His-tag was removed by applying again the protein solution to a $Ni^{2+}$ column and the flow through was collected, concentrated to 5 mL and injected in FPLC system (ÄKTA Explorer) at 4°C on a gel filtration Superdex 200 column (Amersham Pharmacia) equilibrated in buffer 50 mM tris pH= 7.8, 300 mM NaCl, 5% glycerol. The pure fraction was collected and concentrated to 3 mg/mL.
The protein amount was determined spectrophotometrically by measuring the absorbance at 280 nm. The yield of Rpf B protein is about 2 mg for each liter culture.
The purification protocol used for Rpf C is the same used for Rpf B and Rpf $B_{cd}$ but the former was purified on FPLC by using a POROS MC 20µm PEEK 4.6 mm x 100 mm Ni-chelating column and a Superdex 75 column for gel filtration on ÄKTA purifier at 4°C. The yield of Rpf C in pETM-11 is quite poor while cloning the protein in pETM-20 makes it much more soluble. The chromatogram and the acrylamide gel reported in section 2.II.3.1 are referred to Rpf C in pETM-20.

**Circular Dichroism experiments protocols**

CD experiments for Rpf B, its catalytic domain Rpf $B_{cd}$ and Rpf C were carried out in buffer phosphate 10 mM pH= 7.5, 150 mM NaCl by using a protein concentration of 4.0 µM, on a Jasco J-810 spectropolarimeter equipped with Peltier system for temperature control. All the buffers and solutions were previously filtered with a 0.45 µM filter to avoid any dust and placed in 0.1 cm cell.

Parameters:

Wavelength range: 260 nm-195 nm, data pitch: 0.1, acquisition: continuous, scan speed: 100 nm/min, response: 4 s, band width: 1 nm, accumulation: 3

**Activity assay protocols**

The lysozyme-kike activity was tested for the truncated Rpf B in buffer 50 mM sodium citrate, 5 mM $MgSO_4$ at pH= 7.8 by incubating the protein at 37 °C with the substrate 4-methylumbelliferyl-β-D-N, N', N"-triacetylchitotriose (MUF tri-NAG) for 3 hours. The fluorescence emission of the catalytic reaction producing 4-methylumbelliferone, was measured at 455 nm. This wavelength is the maximum emission wavelength for the fluorescent product of this reaction, according to manufacturer information sheet (Sigma-Aldrich, Germany). The experiments were carried out in Labsystems 384 black cliniplates, in a Tecan (Durhan, NC) ULTRA Evolution Multi-Detection Microplate Reader instrument for fluorescence intensity (UV and VIS range) top & bottom detection.

Parameters:

Excitation: 360 nm
Emission: between 400 nm and 500nm
Temperature control: 37°C

Protein concentration: 10 µM in 50 µL total volume
Substrate concentration: from 0.5 µM to 10 µM

The second set of experiments were performed on a Varian Cary Eclipse spectropolarimeter. Spectra were acquired in 10 mm (500 μL) fluorescence cuvettes covered with caps. The protein Rpf B (2 μM) in buffer50 mM sodium citrate, 5 mM $MgSO_4$ at pH= 7.8, was incubated overnight at 37° C with 20 μM MUF tri-NAG.

**Dynamic light scattering experiments**

Hydrodynamic radius ($R_H$) measurements were made with a DynaPro99 MSX instrument (Wyatt Technology) equipped with a gallium aluminium arsenide laser and with Peltier temperature control for accelerated stability studies. Samples (60 μL) were filtered with a 0.22 μm filter, because dust particles may result in large fluctuations in the instrument readings, and placed directly into a black quartz cuvette, waiting for 15 mins for sample stabilization and approximately 30 measurements were collected at 4°C. Histogram of percentage intensity (% I) *versus* $R_H$ were calculated using Dynamics data analysis software (Protein Solutions), and intensity-weighted mean $R_H$ values were obtained for each subpeak.

Conditions:
DLS measurements were performed for both Rpf B and Rpf C in buffer 50 mM TRIS pH= 7.8, 300 mM NaCl, 5% glycerol.

The histograms obtained from these experiments are reported in section 2.II.3.2.

## 2.II.3 Results and discussion

### 2.II.3.1 Cloning, expression and purification of Resuscitation-Promoting Factors

Rpf-like proteinaceous factors are presents in several G+C Gram-positive bacteria such as the non-sporulating *Mycococcus luteus* and in *Mycobacteria* as *M. Smegmantis*, *M. tuberculosis* and its related *M. bovis*. While *M. luteus* possesses one single gene encoding for one Rpf protein, *M. tuberculosis* contains five rpf-like genes, designated as rpfA-E (http://genolist.pasteur .fr/Tuberculist) which are expressed by actively growing cells. The predicted rpf-like gene products of *M. tuberculosis* have very close characteristics and properties to those of Rpf from *M. luteus* and they are also able to resuscitate dormant cultures at picomolar concentrations. It has been proposed [45] that these five Rpf-like proteins from *M. tuberculosis* may be partially redundant and that, none of the five rpf-like genes is essential in MTB, in contrast to the single *M. luteus* rpf.

A secondary structure prediction investigation of all five Rpf forms has been carried out as a preliminary study of feasibility. As a result, we evidenced that both Rpf A and RpfE contain series of Ala-Pro-rich sequence repeat, which are likely unstructured. For this reason, these two proteins were considered as a difficult protein for structural characterization purposes. Similarly, RpfD is predicted to contain a large non-structured region. In contrast, Rpf B and Rpf C are predicted to exhibit a rather defined fold. Furthermore, all the Rpfs contain an N-terminal signal P sequence and/or a membrane anchor. A signal peptide, also called “signal P”, is a short (3-60 amino acids long) peptide chain that directs the post-translational transport of a protein to other locations in the cell. Protein sequences containing signal peptides or membrane anchors are often difficult to express and purify as soluble proteins.
At first, full length forms of Rpf B and Rpf C were cloned and a complete expression screening, in all the available *E. coli* strains, was performed. However, the expression of soluble protein has been too poor to scale up. Therefore, truncated constructs have been designed for Rpf B and Rpf C, avoiding signal P fragments predicted in the N-terminal region between residues 1 and 30.
Truncated versions of these proteins, lacking their N-terminal 30 residues, were cloned.
In order to clone the genes of interest, the host and the vector were previously selected among those available to obtain fusion proteins. *E. coli* was chosen as host because is fast and easy to grow, it is compatible with a wide choice of cloning vectors and gives high yields. The use of pET vectors allowed obtaining the proteins of interest as fusion proteins, containing a tag of known size and biological function. Fusion His-tag protein production method is the most employed because targeting information can be incorporated into the tag which provide a marker for expression and purification. Purification can be then obtained by using affinity chromatography. All rpf genes have been amplified from MTB genomic DNA using the primers and the PCR cycle program reported in table 2.II.1 in section 2.II.2.

Furthermore, a shorter construct was designed, cloned, expressed and purified for the shorter C-terminal domain of Rpf B, the fragment Rpf B [280-362], named as Rpf $B_{cd}$. This fragment corresponds to the cathalytic domain of the protein.
The amplified PCR samples were digested with Nco I and Hind III as the primers were designed with these specific restriction sites and cloned into the corresponding sites of pETM-11 vector to produce N-terminally His-tagged Rpf protein. Additionally, the rpf C gene was cloned in pETM-20, a vector containing thioredoxin as fusion protein and a fusion His-tag in order to improve protein solubility. The first cloning of Rpf C in pETM-11 produced low soluble protein yields. For all proteins, sequences were checked using the sequencing facilities available at EMBL Heidelberg and the plasmid with the gene has been transformed into *E. coli* DH5α.
A complete expression screening was carried out for all the targets in the available *E. coli* strains cited in Methods section. From these screenings the strain BL21 (DE3) was selected among the others because it gave the best yield of soluble protein and the cultures were grown in Luria Bertani medium and induced with IPTG. As proteins were cloned and expressed as His-tag fusion proteins, they were purified by affinity chromatography purification strategy, chelating the His-tag upon an activated chelating resin, charged with $Ni^{2+}$, and desorbing the His-tagged fusion proteins with buffers containing 150-300 mM imidazole. After the first step of purification the His-tag was removed by cleaving it away with TEV protease in a lower salt buffer. Furthermore, an additional step of purification, on a gel filtration column was carried out to check the aggregation of the proteins. The first purifications performed by gel filtration for both Rpf B and Rpf C revealed a significant amount of aggregate in the protein preparation. The aggregate part was eluted out in the void volume of the column, at retention times corresponding from the calibration of the column, to high molecular weights. The purification condition were optimized by adding 5% glycerol to all the buffers used, from the lysis to the final gel filtration and by adding 3% imidazol to the lysis buffer to prevent non-specific binding to the Ni activated resin during the first step of purification. Furthermore, all the purification steps were performed at low temperature (4°C) and in one day to avoid any degradation of the protein occurring after a long time storage. The optimisation, to obtain soluble and much more stable protein, is confirmed by gel filtration ad DLS measurements here reported.
The pure fractions of Rpf B, Rpf $B_{cd}$ and Rpf C, were collected and concentrated to 3 mg/mL. The protein amount was measured spectrophotometrically by measuring the absorbance at 280 nm and the proteins were electrophoretically identified on a acrylamide gradient gel (8-20% acrylamide) and stained with Coomassie. The yield of Rpf B and Rpf $B_{cd}$ is about 2 mg for each liter culture and for Rpf C, 1.5 mg protein was obtained for 1 liter culture. Data here reported for Rpf C are relative to the gene cloned and expressed in pETM-20. The thioredoxin fusion protein increased protein solubility and was removed by the TEV protease, together with the His-tag, before the last step of purification.
As gel filtration is the preferable technique to separate molecules for their size, it was used as the last step of purification to avoid any possible aggregate from the protein preparation.

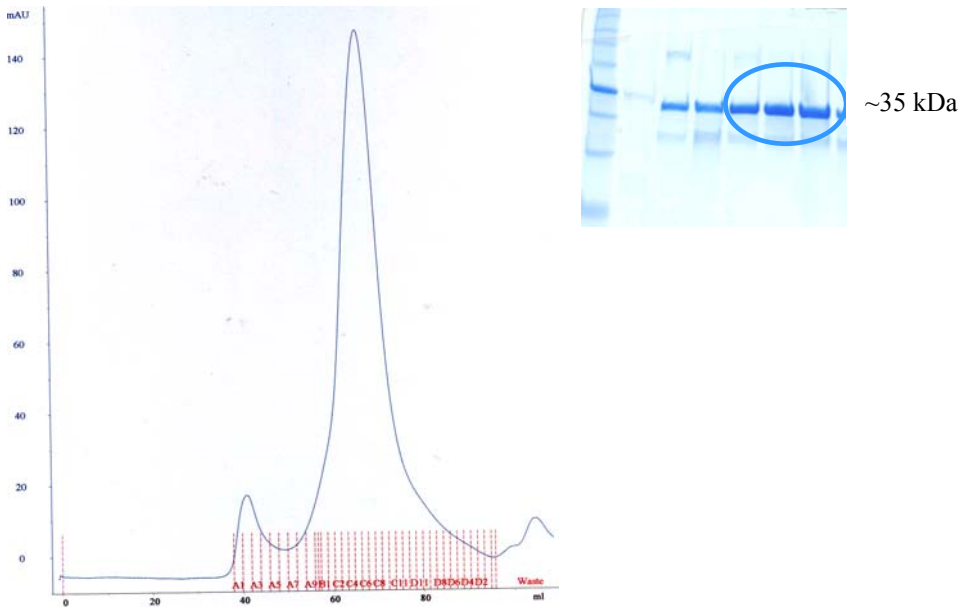



**Figure 2.II.7:** FPLC chromatogram obtained from Rpf B gel filtration after the optimization of the purification and relative acrylamide gel. The main peak corresponds to the molecular weight of the protein and the lower peak is belonging to a small amount of aggregate or high molecular weight impurities.

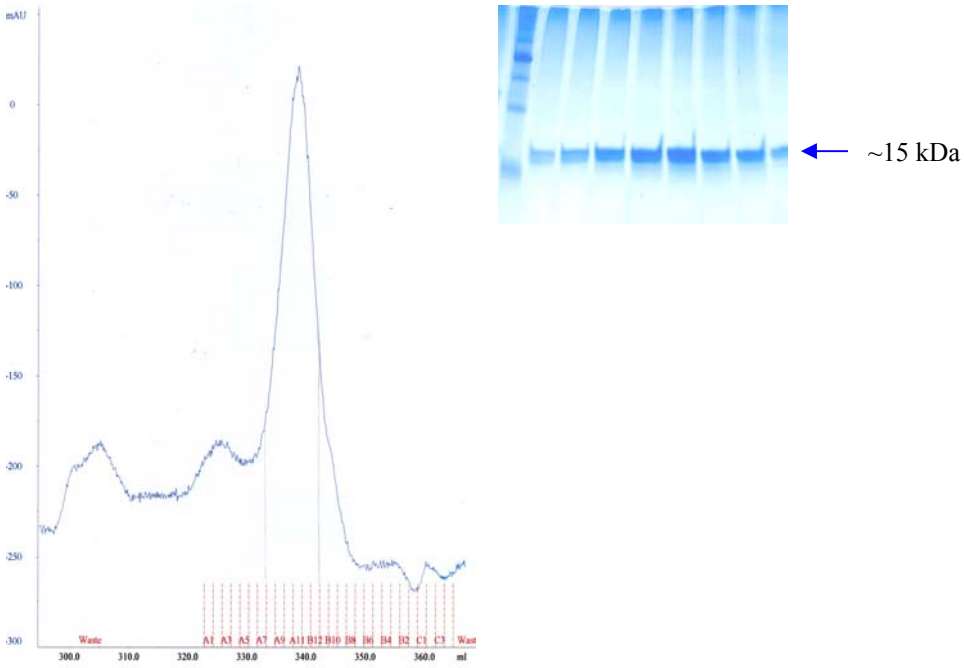



**Figure 2.II.8**: FPLC chromatogram obtained from RpfC gel filtration after the optimization of the purification and relative acrylamide gel. The main peak corresponds to the molecular weight of the protein.

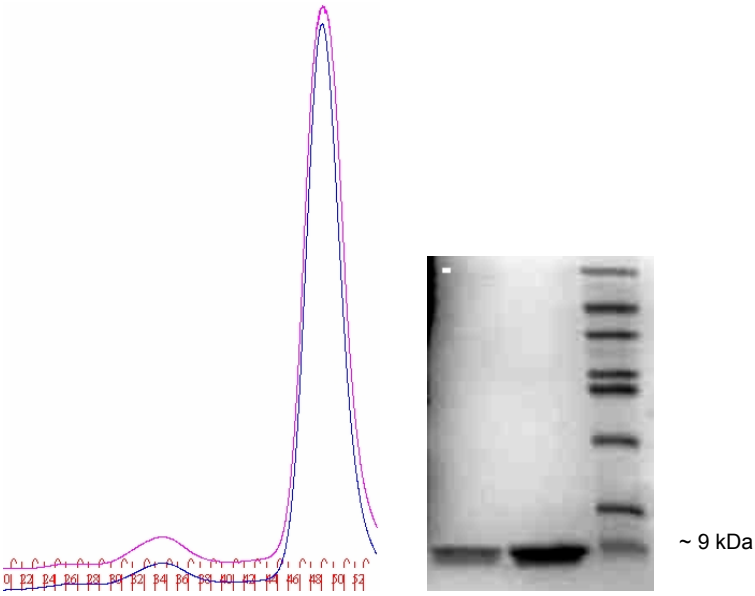



**Figure 2.II.9:** FPLC chromatogram of Rpf $B_{cd}$ gel filtration and its acrylamide gel.

As shown in chromatograms here reported, after the optimization of proteins purification conditions, the amount of aggregate, remains quite low.
The proteins were usually used for further experiments immediately after their purification, to avoid any degradation and/or aggregation. When it has been not possible to employ fresh preparation, the proteins were frozen in 5-10% glycerol and stored at -80°C.

## 2.II.3.2. Dynamic light scattering and circular dichroism studies

### Dynamic Light Scattering (DLS)

In order to determine the aggregation eventually occurring in protein preparation, several experiments were performed by dynamic light scattering. This technique is, in principle, capable of distinguishing whether a protein is a monomer or dimer. It allows calculating the absolute molecular weight for determining the oligomeric state of a protein for a particular solution condition or preparation. Usually, for proteins, absolute molecular weight is measured by chromatography where the elution time of the peak is compared to the elution of a standard in the same column but this technique may require significant amount of sample and several hours to make measurements. With light scattering there is no reference needed. If the concentration and the refractive index increment of the sample are known, then from the absolute amount of light scattered (the intensity) is possible to determine the absolute molecular weight, or molar mass. It is called absolute, because no further assumption has to be made about the sample or its behavior with respect to another measurement technique (as would be the case for a native gel or HPLC run).
Oligomer formation, aggregation or conformation of the protein molecules as well as the amount of impurities in the solution are critical parameters in successful crystallization. Thus, DLS is widely used prior to protein crystallography [46, 47] to assess whether protein samples are able to form crystals. It is also a practical method to investigate conformational changes, folding kinetics, [48] oligomerization, thermostability, [49, 50] and the liquid-phase complex formation [51] of proteins. Together with MS, it can be a useful method to confirm the observed characteristics of the protein molecules. DLS is a quick, non-destructive and sensitive method requiring relatively small amounts of purified material, typically less than 1mg and it is almost all automatized. It also gives the possibility to analyze samples containing broad distributions of species of widely differing molecular masses (a native protein and various sizes of aggregates), and to detect very small amounts of the higher mass species (<0.01% in many cases). However, the analysis might be difficult for non-rigid macromolecules. Commercial "particle sizing" systems mostly operate at only one angle (90°) and use red light (675 nm). Usually in these systems the dependence on concentration is neglected.

The percentage intensity *versus* $R_H$ was calculated using Dynamics data analysis software (Protein Solutions), and intensity-weighted mean $R_H$ values were obtained for each subpeak. These calculations were made by the instrument, which generates the corresponding histogram of the percentage intensity *versus* the hydrodynamic radius. The polydispersion index is a strong indicator of monodispersity of a solution. It can be calculated as the ratio between the dispersity of the peak and the protein hydrodynamic radius. If this index is lower than 0.15, the solution is considered as monodisperse. Above 30-35% it is a polydisperse solution. The peak always come to a higher MW than the true MW, unless the protein is a perfect sphere, due to unaccounted for additional rotation friction. A protein does have higher chances to

crystallize if it is monodisperse. It has been reported about crystallized solutions with a polydispersion index of 20-25%.

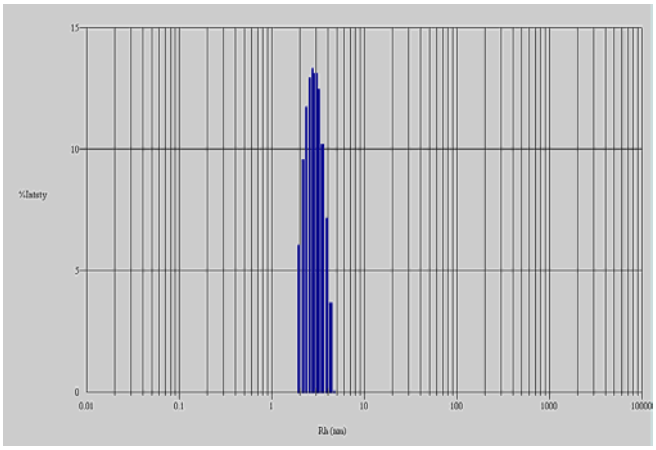


**Figure 2.II.10**: Histogram obtained for Rpf B measurement. The percentage of intensity is reported as a function of the hydrodynamic radius.

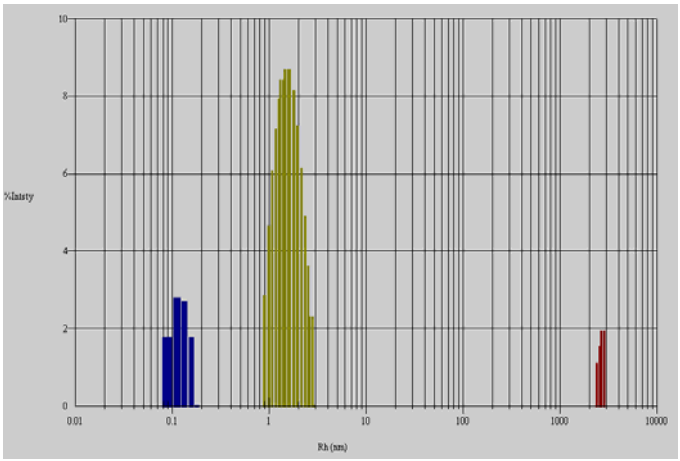


**Figure 2.II.11:** Histogram obtained for Rpf C measurement. The percentage of intensity is reported as a function of the hydrodynamic radius.

Polydispersion indexes were calculated both for Rpf B and Rpf C. As shown in figure 2.II.10, purification of Rpf B resulted in a significantly pure protein preparation, and no aggregation is registered, with a polydispersion index of 20.7 %. A slightly higher polydispersion, close to 29%, was observed for RpfC (figure 2.II.10). Furthermore, beside the existence of a peak in correspondence to low $R_H$ (figure 2.II.11), which is commonly observed in protein preparations, it is also evidenced the presence of a small amount of impurities at high Rh values (close to 2601 nm), which is likely constituted by protein aggregates. These experiments were repeated for each fresh preparation of the protein to check if aggregation problems were occurring, before using the prep for activity assays. An even small amount of aggregate could affect fluorescence activity assays, as reported also in literature, [42] for similar assays.

## Circular Dichroism (CD) spectroscopy experiments

In order to investigate Rpfs' structural behaviour in solution, several CD experiments were carried out for these proteins in 50 mM pH=7.5 phosphate buffer, 150 mM NaCl. Measurements were carried out on a Jasco J-810 spectropolarimeter equipped with Peltier system for temperature control. The first set of experiments was due to previously verify whether Rpf B and Rpf C are folded in buffer solution and in the condition cited. With this aim spectra were acquired for fresh protein preparations at room temperature (20°C).

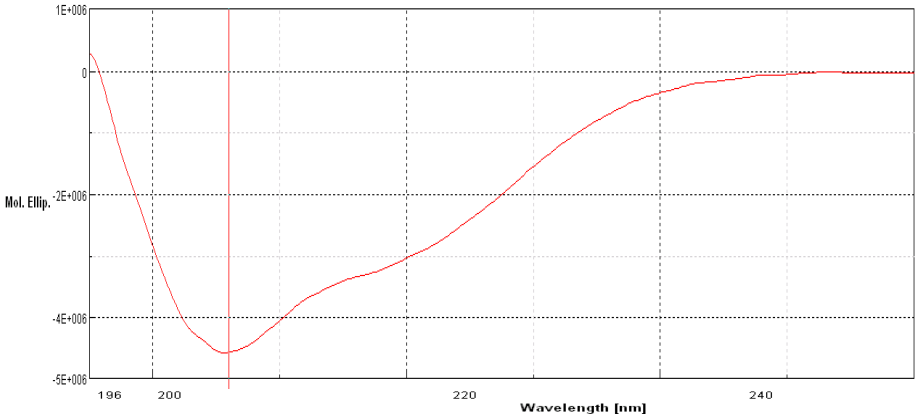



**Figure 2.II.12:** Rpf B 4 µM far UV-CD spectrum at room temperature (20°C)

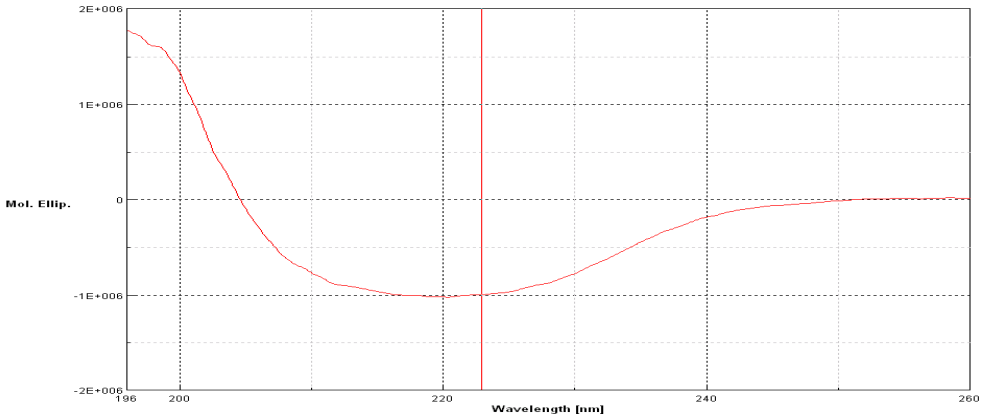



**Figure 2.II.13:** Rpf C 4 µM Far UV-CD spectrum at room temperature.

As shown in figure 2.II.12 and 2.II.13, both the proteins are folded in solution at pH= 7.5 and show prevalently α-helical behaviour. Higher α-helical content in the condition experimented, is observed for Rpf B. Its spectrum shows a strong negative band at 208 nm, a visible shoulder at 222 nm and a positive band at 195 nm. These values are typical for α-helical dichroic behaviours. Rpf C seems to have lower α-helical content (the gap between the minimum at 208 nm and the shoulder at 222 nm

is not so evident), even if it is most likely folded in the experiment conditions. After this previous characterization, further experiments were performed on the protein Rpf B, in order to investigate the stability of this protein. A thermic denaturation of the protein was carried out by following the molar ellipticity at 222 nm (the wavelength that typically represents an α-helical folding), in a range of temperatures between 4°C and 80°C by increasing the temperature of 1°C each minute.

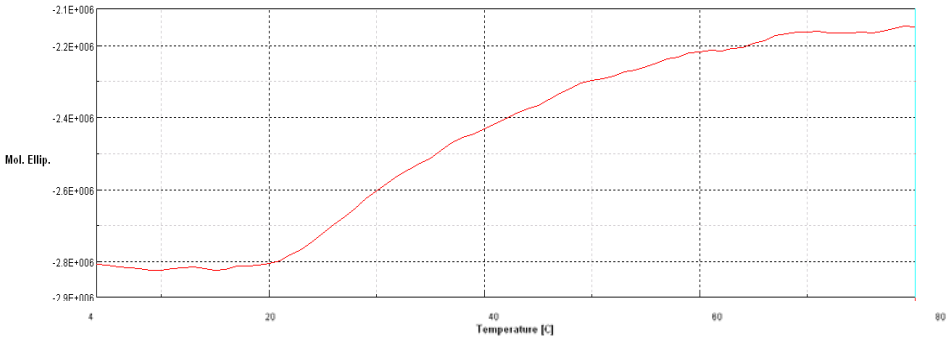



**Figure 2.II.14:** Rpf B 4µM temperature scan measurements from 4°C to 80°C at 222nm

The temperature scan reveals a loss of folding for the protein Rpf B from 23°C to 60°C. A new spectrum was acquired for the protein after the scan measurements at 80°C, where a complete unfolding of the protein was expected. The figure 2.II.14 shows the overlay of the spectra acquired before and after temperature unfolding. The loss of folding is evidenced from the main negative band shift from 208 nm to 202 nm and the gap between the negative bands becomes less relevant.
Furthermore, the positive band around 195 nm became negative.

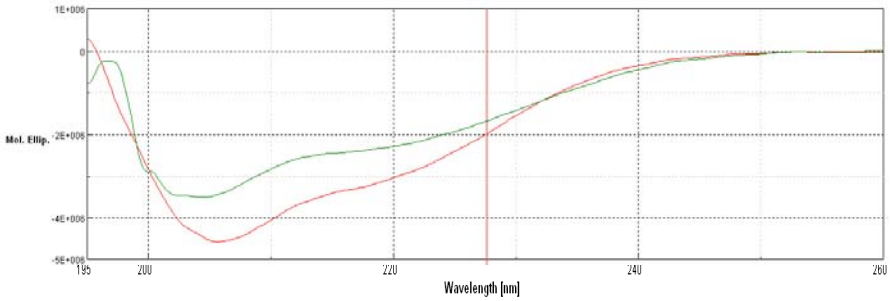



**Figure 2.II.15:** Overlay of Rpf B spectra at 4°C and at 80°C Red: Rpf B spectrum acquired at 4°C Green: Rpf B spectrum acquired at 80°C.

After unfolding the protein was brought back at 4°C, and a new set of temperature scan measurements was acquired to compare the propensity of Rpf B to refold after thermic unfolding. After the temperature scan from 80°C to 4°C a new spectrum was acquired for the protein after refolding. The figure 2.II.15 shows the overlay of the spectra before unfolding and after the protein was brought back in the initial temperature conditions.

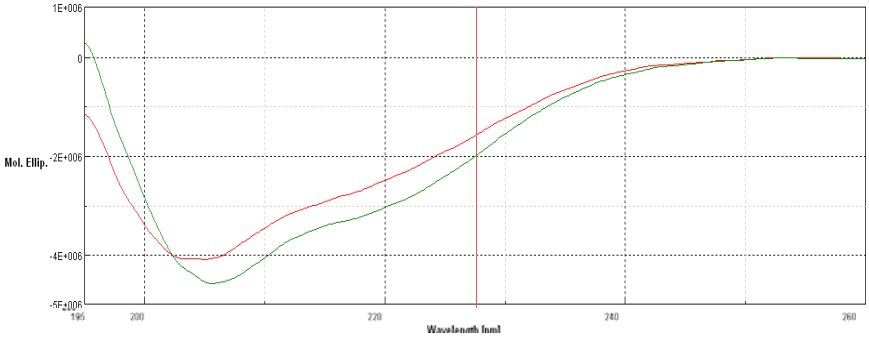



**Figure 2.II.16:** Overlay of Rpf B's spectra at 4°C before and after unfolding by temperature increase to 80°C and refolding by decreasing the temperature to 4°C. In green is reported the spectrum acquired before unfolding and the spectrum acquired after refolding is reported in red.

The spectrum of RpfB after refolding resembles that one measured before unfolding. This indicates that Rpf unfolding is rather reversible and the protein is quite stable.

The same characterization was performed for Rpf $B_{cd}$, the shorter catalytic C-terminal domain of the protein Rpf B. Far-UV CD spectra were acquired in the same conditions used for the previous experiments. The spectrum reported below is for the domain at room temperature.

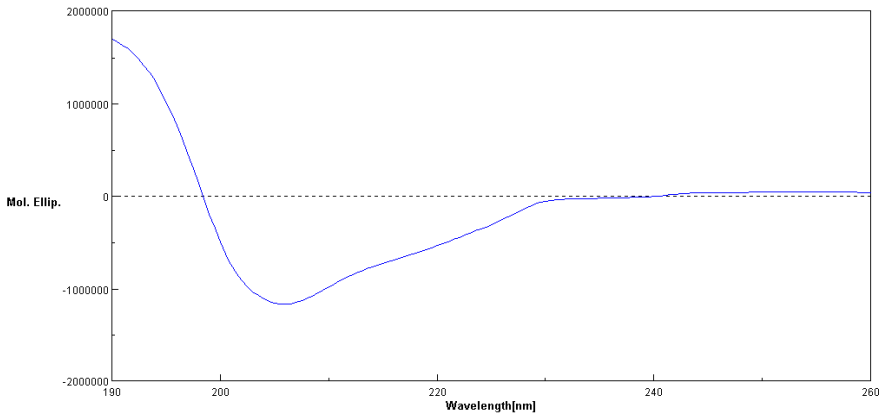



**Figure 2.II.17:** Far UV-CD spectrum of the domain Rpf $B_{cd}$ at 20°C.

Once the folding of the domain has been verified, a temperature scan was performed for this sample to investigate its stability. The wavelength was fixed at 208 nm, a typical α-helical folding value. The measurements were started at 5°C and the temperature increased to 80°C in 1 hour.

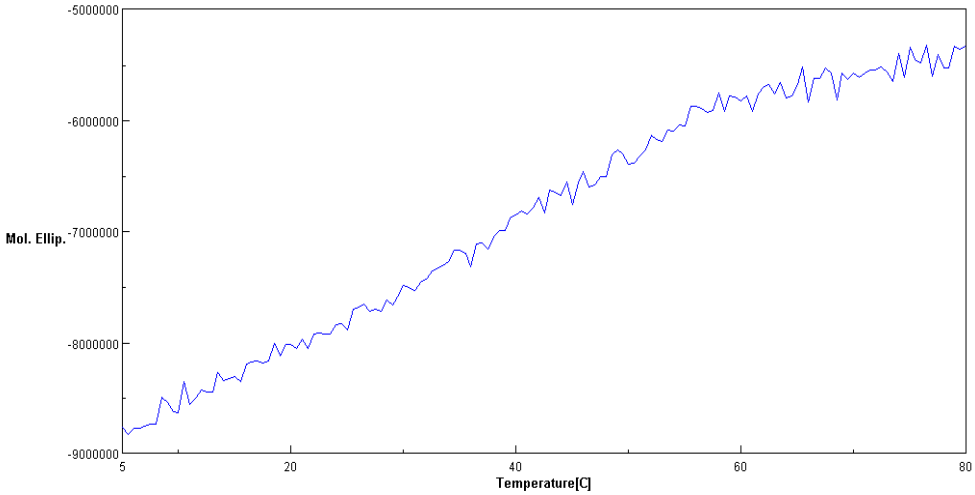



**Figure 2.II.18:** Rpf $B_{cd}$ temperature scan measurements at 208 nm, between 5°C and 80°C. The protein domain is unfolded up to 60°C.

A comparison between the spectrum acquired at 5°C and the spectrum acquired after unfolding at 80°C is reported in figure 2.II.18.

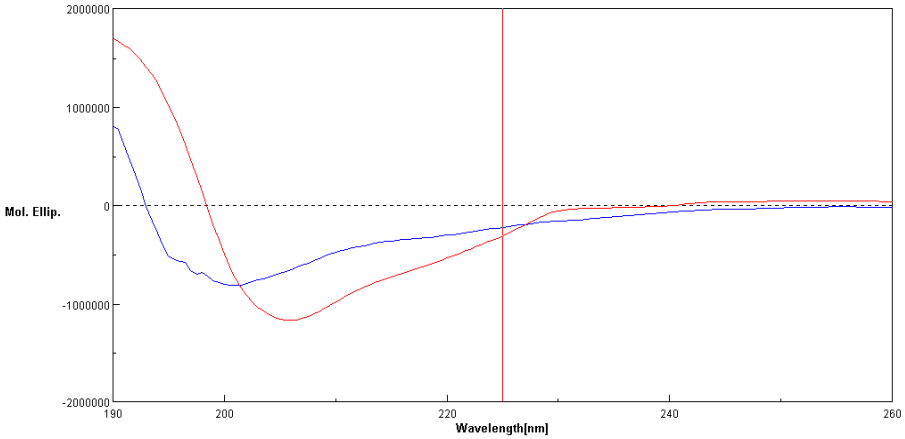



**Figure 2.II.19.** Far UV-CD spectra overlay for the protein domain Rpf $B_{cd}$ before (red) and after (blue) unfolding by increasing the temperature from 5°C to 80°C.

After the unfolding the protein domain was brought back to 5°C and a new spectrum was acquired at 5°C to check protein folding.

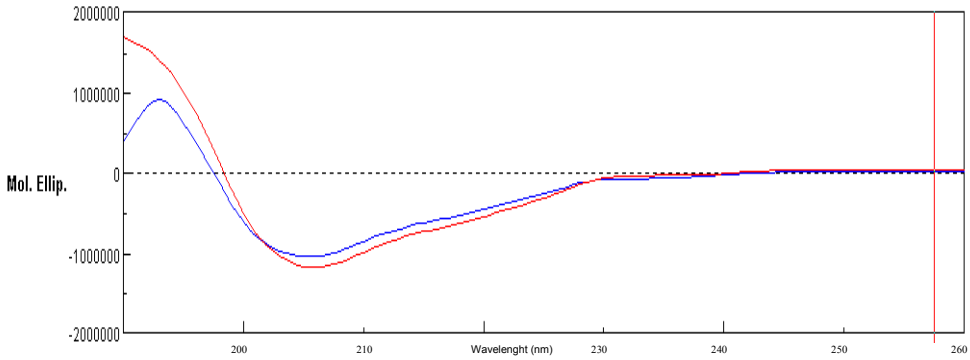



**Figure 2.II.20:** Overlay spectra of Rpf $B_{cd}$ before (red) and after (blue) unfolding (5°C-80°C) and refolding (80°C-5°C).The spectra are rather similar.

From the circular dichroism studies so far reported it is possible to summarize that, as expected, all the Rpfs proteins considered in this study, possess an α-helical folding. As Rpf C preparation, here considered, is the less folded and shows a small amount aggregate, further studies are focused on the Rpf B and its catalytic domain. Furthermore Rpf B and Rpf $B_{cd}$ are rather stable in buffer solution in the experimented conditions. This result is evidenced for these proteins from the unfolding/refolding study performed by CD. After unfolding, induced by temperature increase, the proteins both save the tendency to refold in almost the same α-helical folding, as the overlay of spectra acquired before and after the experiment shows. The folding of the protein, and its stability, is important both for further structural studies and for activity assays.

## 2.II.3.3. Rpf activity

The mechanism of action proposed for lysozyme has been deduced from crystallographic and solution NMR studies by using the polymeric NAG substrate. It has been demonstrated that the active site of lysozyme can accommodate 6 units of NAG, the β(1-4) linked hexasaccharide of N-acetylglucosamine [52].
Interestingly, it has been reported that the catalytic domain of Rpf B has no considerable muralytic activity whereas a significant muralytic activity is shown by an homologous Rpf protein from the bacterium *Micococcus luteus* [43].

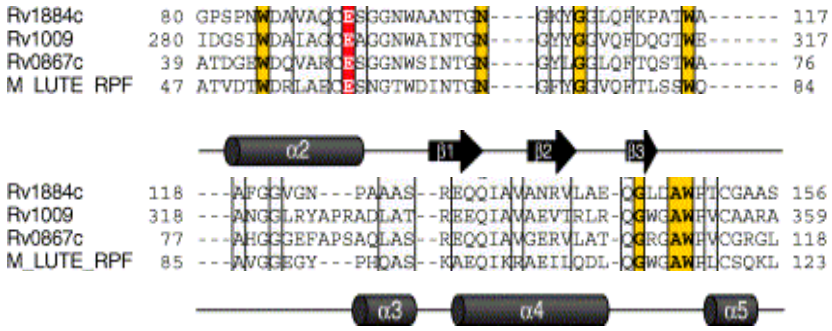
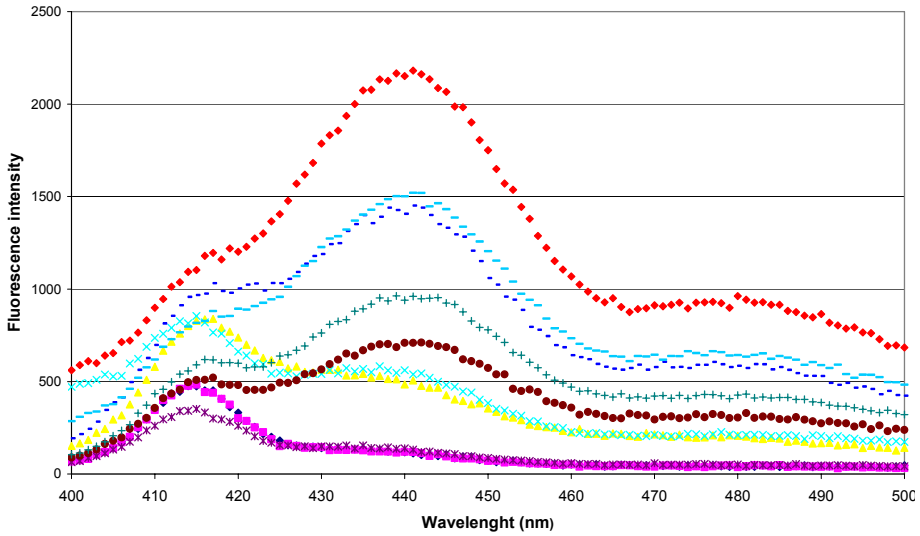
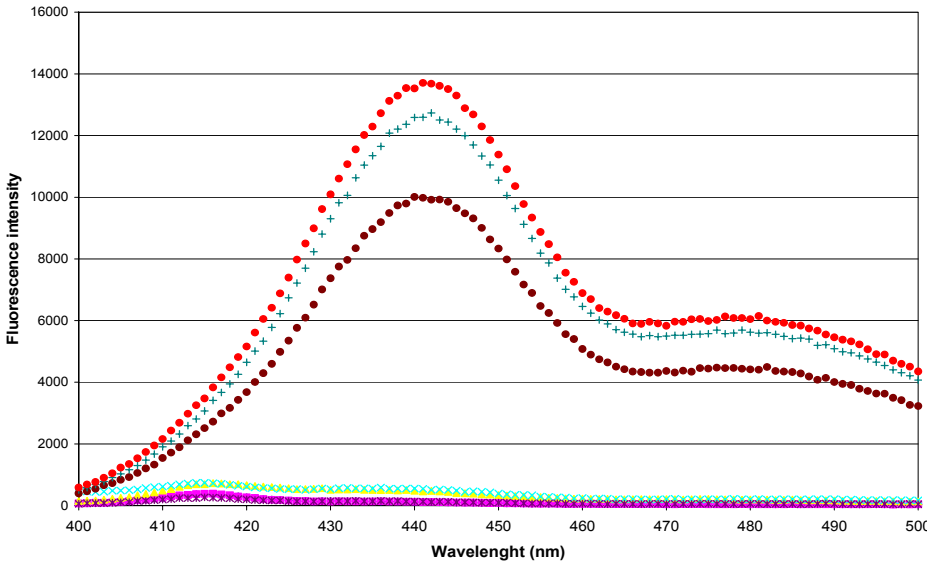



**Figure 2.II.21:** Multiple sequence alignment of Rpfs domain region from *M. Tuberculosis* and from *M. luteus*. Structure codes are from the Protein Data Bank (http://www.rcsb.org/pdb/) and sequence codes are from the *M. tuberculosis* genome for Rpf C (Rv1884c)and Rpf B (Rv1009).

In view of the structural similarity of Rpf B to lysozyme and to Rpf from *M. luteus* the synthetic, artificial lysozyme substrate 4-methylumbelliferyl-β-D-N,N',N"-triacetylchitotriose (MUF tri-NAG) was used in this thesis study as a substrate for activity assays [53]. The choice of this substrate becomes from its fluorogenicity. Its cleavage produces a fluorescent compound and the assay may be performed by following the fluorescence produced by the reaction. The product of the catalytic reaction, 4-methylumbelliferone, has the maximum of fluorescence emission at 455 nm. Therefore, the intensity of fluorescence has been monitored for 3-4 hours at 37°C at 455 nm using an excitation wavelength of 360 nm.




a)




b)

**Figure 2.II.22:** Rpf B activity assay by fluorescence spectroscopy. **a)** The figure shows the spectrum acquired after 7 minutes incubation of the protein with the substrate MUF-tri NAG at 37°C. The higher curve (in red) is referred to the highest concentration of substrate (10 µM). **b**) The picture b shows the spectrum acquired after 3 hours incubation. The fluorescence intensity is reported in arbitrary units (u.a.) and increases up to 13900 u.a.

Each eleven minutes acquisition, the fluorescence intensity measured at 455 nm was reported as function of the time to perform a kinetic of Rpf B activity assay. The points obtained are reported in the graphic in figure 2.II.22.

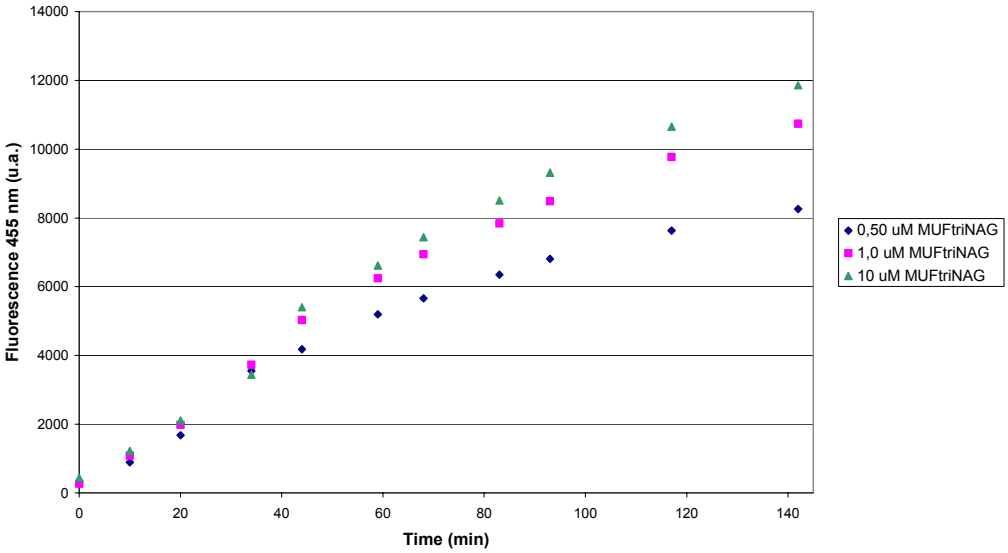



**Figure 2.II.23:** Rpf B muralytic activity measured by following the increasing fluorescence intensity at 455 nm in 3 hours incubation of the protein with substrate at 37 °C.

As shown in figure 2.II.22, fluorescence intensity increased over a three hours incubation at 37°C. This is evidence that 4-methylumbelliferone is released from the reaction and, consequently, that the protein tested is active. Different concentrations protein were also tested and the protein seems to work from 1:1 to 1:3 protein/substrate molar ratio.
During these experiments, for incubation times over 3 hours a decrease in fluorescence intensity was observed, and the samples were affected to evaporation from the plate wells. The plate used in this spectropolarimeter could not be covered and, for long incubation periods, the amount of reaction solution decreased to few microliters. Because of the impossibility to avoid these sample problems, the same experiment was performed in fluorescence cuvettes, with caps, on a Varian Cary Eclipse spectropolarimeter. Spectra acquired in 10 mm path length (500 µL volume) fluorescence cuvettes are reported for the protein Rpf B (2 µM) in the same buffer used for the previous experiment. In this new set of experiments the protein was incubated overnight at 37°C.

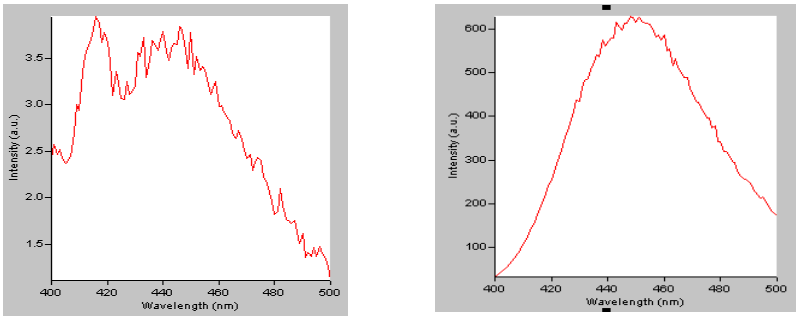



**Figure 2.II.24:** Rpf B 2 µM fluorescence spectrum after and before overnight incubation at 37°C. These spectra were acquired in closed cuvettes in the same conditions used for the previous set of experiments. At the starting time the fluorescence intensity is about 3.7 a.u. After overnight incubation of the protein with the substrate the intensity of fluorescence is over 600 a.u.

The use of capped cuvettes allowed to avoid evaporation problems and to further verify Rpf B activity. In this experiments the fluorescence intensity reaches over 600 u.a. for a protein concentration of 2.0 μM reacting with 20 μM MUF-tri NAG overnight. Further experiments are going to be performed to study Rpf B activity by fluorescence in order to define its reaction kinetic in cuvette and for longer time.

## 3. Conclusions and perspectives

The interactions between biological macromolecules play a key role in almost all the biological mechanisms and pathways. Protein interactions are very specific and represent potential target points of pharmaceutical intervention and for devising new biotechnologically useful molecules to control pathological mechanisms. For this purposes, information are required about protein structural behaviour and their binding properties.

In this thesis work, two important issues are developed: the role of protein-protein interactions in dysregulation of the apoptotic signal leading to MALT lymphoma, and the role of protein-cell wall interactions in *Mycobacterium tuberculosis* reactivation from latency.

The first part of this study is dedicated to a complete investigation on the homodimerization of the human protein Bcl10, caused by homophilic interactions between “Caspase-Recruitment Domains” (CARD). This N-terminal domain is typically found in proteinaceous factors involved in cell signalling and, in particular, in proteins involved in the apoptotic cascade, such as Bcl10. The understanding of the dimerization mechanism, and how it can be modulated, is essential for the developing of tumour suppressor molecules. With this aim, the structural behaviour of the CARD domain was studied by focusing the investigation on the domain regions, potentially involved in CARD/CARD self-association. A preliminary study was carried out by circular dichroism techniques on the whole domain, to establish its global attitude to form homodimers and the influence of protein concentration in dimer formation.

CARD domain exists in solution in two conformational states including a predominantly α-helical monomer, that is in equilibrium with a partially unfolded, dimeric form. This equilibrium is significantly shifted towards the monomeric form at micromolar concentrations. Surprisingly, at higher concentrations, the domain dimerizes and undergoes self-association with a concomitant increasing population of a partially unfolded form. The CARD/CARD interaction in this self-association is quite strong with a calculated $K_d$ value of 18.6 μM. The dimeric form becomes significantly populated at concentrations higher than the $K_d$. CARD self-association, evidenced in this CD study, was in agreement with the gel filtration results, showing a concentration-dependent appearance of dimeric form.

Furthermore, CARD self-association was successfully confirmed by ELISA assays by immobilizing the domain on the ELISA plate and verifying its binding to a biotin-labelled CARD, using the biotin-streptavidin detection system. Then, in order to go deeper into CARD self-association and to determine which regions of the domain were potentially involved in the binding, CARD was hydrolyzed with trypsin and the digestion fractions, obtained by RP-HPLC, were employed in a competition ELISA assay. The competition properties of these fractions were verified in such a test by

comparing the binding attitude of the biotin-labelled CARD to the unlabelled CARD, with and without each digestion fraction. Four of these fractions were selected as positive-competing fractions in CARD-CARD binding and they were identified by LC-MS to determine the regions involved in such interactions. The identified sequences were obtained as synthetic fragments and, so far, three peptide sequences, belonging to a specific domain region, were tested against the whole CARD in a new competition ELISA. The results show that the synthetic fragment, named as CARD [78-98] and its shorter analogue, CARD [91-98], are able to compete in CARD-CARD binding, confirming again their attitude in modulating CARD self-association. Interestingly, the longer sequence CARD [68-98], belonging to the same binding region, shows the lowest competition in CARD binding. This evidence relates the binding property to a very specific region within the domain sequence. In order to provide a structure-function characterization of the identified region, the synthetic fragments CARD [78-98] and CARD [91-98], were characterized by circular dichroism. The longer peptide, which has reduced competition properties, shows no structural preferences and appears as unfolded while, the shorter analogue, CARD [78-98], which is able to compete in CARD binding, is rather folded in solution and shows a stable, α-helical folding.
Even though Bcl10 CARD domain structure is not available, two similar CARDs, belonging to the proteins Apaf1 and RAIDD, were compared with Bcl10 CARD by using a sequence alignment program. The alignment of the fragment CARD [78-98] with RAIDD CARD shows that this fragment lies in a prevalently α-helical region, comprising the helix 5, the helix 6 and a short loop in between, confirming the structural behaviour determined in CD experiments.
The results obtained in this thesis work provide a characterization and identification of one of the region sequences involved in Bcl10 CARD self-association, the fragment CARD [78-98].
Preliminary NMR studies were carried out on this peptide, in order to determine the best conditions for the NMR conformational characterization. In a concentration range between 1.0 mM and 0.1 mM (in $H_2O$) at 25°C, a changing in the intensity and multiplicity of signals was observed, evidencing that, in this conditions, the peptide may exist as a complex mixture of different oligomeric states. $^1$D $^1$H NMR spectra, acquired for the peptide in 5mM Na acetate buffer pH=5.5, showed interestingly well resolved resonances and no aggregation behaviour at 0.1 mM concentration. These conditions were chosen to set up further two dimensional NMR experiments in order to determine peptide structure, and a complete characterization of the smaller fragment CARD [91-98] will be performed to verify the specificity of the sequence in the homophilic interaction. Then, with the aim to provide evidences towards the ability of this specific region to modulate Bcl10 filaments formation, *in vitro* experiments will be performed. Hence it has been demonstrated that filaments formation is due to CARD-CARD association between Bcl10 proteins, the employment of a peptide, specifically modulating such interaction, could provide the possibility to modulate the apoptotic signal in living cells.

The second part of this thesis work is focused on the importance of protein-cell wall interactions in the transition of *Mycobacterium tuberculosis* from a latent state to a pathological state. There is no possibility to clear out the infection when the bacterium is in its dormant state. Therefore, any investigation on the causes of bacterial reactivation is important for the development of new therapies. Five

proteinaceous factors, called as “Resuscitation-Promoting Factors” (Rpf), are involved in the bacterial reactivation from “dormancy”. The role of Rpf proteins in the revival of dormant bacteria, still not clear, has been suggested to be the cleavage of oligosaccharides constituting the bacterial cell wall. All the studies reported so far in literature evidenced a muralytic activity, similar to that of the c-type lysozyme, for the Rpf from *M. Luteus*, but no evidences are reported for the *M. tuberculosis* Rpfs muralytic activity.

In this thesis work, two of these proteins, the Rpf B and Rpf C, were selected. The proteins Rpf B, Rpf C and the catalytic domain Rpf $B_{cd}$ were cloned, expressed and purified. Purification conditions were optimized to increase the solubility of the proteins and to avoid aggregates formation. The optimization is confirmed by gel filtration chromatograms and DLS measurements.

In view of Rpf B’s structural similarity to lysozyme and in order to verify the muralytic activity of the protein, the synthetic, artificial lysozyme substrate 4-methylumbelliferyl-β-D-N,N’,N”-triacetylchitotriose (MUF tri-NAG), was used as substrate for activity assays. This fluorogenic substrate produces a fluorescent compound (4-methylumbelliferone), whose maximum fluorescence emission is observed at 455 nm. The assay was performed by following for 4-5 hours at 37°C the increase of fluorescence intensity at 455 nm using an excitation wavelength of 360 nm. Fluorescence intensity increased over a three hours incubation at 37°C. This is the evidence that 4-methylumbelliferone is released from the reaction and, consequently, that the protein tested possesses muralytic activity. Similar activity assays are in progress for the smaller domain Rpf $B_{cd}$ . Interestingly, despite what is reported in the literature, this work has evidenced that RpfB possesses significant muralytic activity. Further studies will be aimed to design Rpfs’ inhibitors, which may prevent the pathogen to exit its dormant state. *In vitro* assays will be performed by screening a complete library of compounds, in order to identify those able to inhibit the muralytic activity observed so far on the synthetic, fluorogenic substrate of lysozyme, MUF tri-NAG. In addition to that experiments, *in vivo* assays could be made to test the ability of the protein to reactivate dormant cultures of *Mycobacterium tuberculosis.*

A preliminary structural characterization of these proteins by circular dichroism revealed, for all of them, a prevalently α-helical folding in solution. Their fold is rather stable and their unfolding process is rather reversible. For a more detailed structural study, crystallization experiments of all Rpf here produced are in progress. The crystal structure of these proteins and of their complexes with substrate analogs and inhibitors will provide valuable information on the mechanism of action used by Rpfs to resuscitate dormant bacteria.

# 4. Bibliography


1. Steller, H., *Mechanisms and genes of cellular suicide.* Science, 1995. **267**(5203): p. 1445-9.

2. Thompson, C.B., *Apoptosis in the pathogenesis and treatment of disease.* Science, 1995. **267**(5203): p. 1456-62.

3. Weber, C.H. and C. Vincenz, *The death domain superfamily: a tale of two interfaces?* Trends Biochem Sci, 2001. **26**(8): p. 475-81.

4. Bertin, J., et al., *CARD11 and CARD14 are novel caspase recruitment domain (CARD)/membrane-associated guanylate kinase (MAGUK) family members that interact with BCL10 and activate NF-kappa B.* J Biol Chem, 2001. **276**(15): p. 11877-82.

5. Bouchier-Hayes, L. and S.J. Martin, *CARD games in apoptosis and immunity.* EMBO Rep, 2002. **3**(7): p. 616-21.

6. Karin, M. and Y. Ben-Neriah, *Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity.* Annu Rev Immunol, 2000. **18**: p. 621-63.

7. Chen, F., L.M. Demers, and X. Shi, *Upstream signal transduction of NF-kappaB activation.* Curr Drug Targets Inflamm Allergy, 2002. **1**(2): p. 137-49.

8. Zhou, H., et al., *Bcl10 activates the NF-kappaB pathway through ubiquitination of NEMO.* Nature, 2004. **427**(6970): p. 167-71.

9. Baichwal, V.R. and P.A. Baeuerle, *Activate NF-kappa B or die?* Curr Biol, 1997. **7**(2): p. R94-6.

10. Sonenshein, G.E., *Rel/NF-kappa B transcription factors and the control of apoptosis.* Semin Cancer Biol, 1997. **8**(2): p. 113-9.

11. Lin, X. and D. Wang, *The roles of CARMA1, Bcl10, and MALT1 in antigen receptor signaling.* Semin Immunol, 2004. **16**(6): p. 429-35.

12. Narayan, P., et al., *CARMA1 is required for Akt-mediated NF-kappaB activation in T cells.* Mol Cell Biol, 2006. **26**(6): p. 2327-36.

13. Kuo, S.H., et al., *Nuclear expression of BCL10 or nuclear factor kappa B predicts Helicobacter pylori-independent status of early-stage, high-grade gastric mucosa-associated lymphoid tissue lymphomas.* J Clin Oncol, 2004. **22**(17): p. 3491-7.

14. Ye, H., et al., *MALT lymphoma with t(14;18)(q32;q21)/IGH-MALT1 is characterized by strong cytoplasmic MALT1 and BCL10 expression.* J Pathol, 2005. **205**(3): p. 293-301.
15. Willis, T.G., et al., *Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types.* Cell, 1999. **96**(1): p. 35-45.

16. Achuthan, R., et al., *Novel translocation of the BCL10 gene in a case of mucosa associated lymphoid tissue lymphoma.* Genes Chromosomes Cancer, 2000. **29**(4): p. 347-9.

17. Achuthan, R., et al., *BCL10 in malignant lymphomas--an evaluation using fluorescence in situ hybridization.* J Pathol, 2002. **196**(1): p. 59-66.

18. Ruland, J., et al., *Bcl10 is a positive regulator of antigen receptor-induced activation of NF-kappaB and neural tube closure.* Cell, 2001. **104**(1): p. 33-42.

19. Ruefli-Brasse, A.A., et al., *Rip2 participates in Bcl10 signaling and T-cell receptor-mediated NF-kappaB activation.* J Biol Chem, 2004. **279**(2): p. 1570-4.
20. Lucas, P.C., et al., *Bcl10 and MALT1, independent targets of chromosomal translocation in malt lymphoma, cooperate in a novel NF-kappa B signaling pathway.* J Biol Chem, 2001. **276**(22): p. 19012-9.

21. Ruefli-Brasse, A.A., D.M. French, and V.M. Dixit, *Regulation of NF-kappaB-dependent lymphocyte activation and development by paracaspase.* Science, 2003. **302**(5650): p. 1581-4.

22. Gaide, O., et al., *Carma1, a CARD-containing binding partner of Bcl10, induces Bcl10 phosphorylation and NF-kappaB activation.* FEBS Lett, 2001. **496**(2-3): p. 121-7.

23. Gaide, O., et al., *CARMA1 is a critical lipid raft-associated regulator of TCR-induced NF-kappa B activation.* Nat Immunol, 2002. **3**(9): p. 836-43.
24. Thome, M., *CARMA1, BCL-10 and MALT1 in lymphocyte development and activation.* Nat Rev Immunol, 2004. **4**(5): p. 348-59.

25. Zhou, P., et al., *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, 1999. **96**(20): p. 11265-70.

26. Vaughn, D.E., et al., *Crystal structure of Apaf-1 caspase recruitment domain: an alpha-helical Greek key fold for apoptotic signaling.* J Mol Biol, 1999. **293**(3): p. 439-47.

27. Qin, H., et al., *Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1.* Nature, 1999. **399**(6736): p. 549-57.

28. Liang, H. and S.W. Fesik, *Three-dimensional structures of proteins involved in programmed cell death.* J Mol Biol, 1997. **274**(3): p. 291-302.

29. Guiet, C. and P. Vito, *Caspase recruitment domain (CARD)-dependent cytoplasmic filaments mediate bcl10-induced NF-kappaB activation.* J Cell Biol, 2000. **148**(6): p. 1131-40.
30. Dalgarno, D.C., B.A. Levine, and R.J. Williams, *Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins.* Biosci Rep, 1983. **3**(5): p. 443-52.

31. Jimenez, M.A., et al., *1H NMR and CD evidence of the folding of the isolated ribonuclease 50-61 fragment.* FEBS Lett, 1987. **221**(2): p. 320-4.

32. Boatright, K.M. and G.S. Salvesen, *Mechanisms of caspase activation.* Curr Opin Cell Biol, 2003. **15**(6): p. 725-31.

33. Chou, J.J., et al., *Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment.* Cell, 1998. **94**(2): p. 171-80.

34. Fernandez-Tornero, C., et al., *Ofloxacin-like antibiotics inhibit pneumococcal cell wall-degrading virulence factors.* J Biol Chem, 2005. **280**(20): p. 19948-57.

35. Tufariello, J.M., J. Chan, and J.L. Flynn, *Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection.* Lancet Infect Dis, 2003. **3**(9): p. 578-90.

36. Wayne, L.G., *Dormancy of Mycobacterium tuberculosis and latency of disease.* Eur J Clin Microbiol Infect Dis, 1994. **13**(11): p. 908-14.

37. Boon, C. and T. Dick, *Mycobacterium bovis BCG response regulator essential for hypoxic dormancy.* J Bacteriol, 2002. **184**(24): p. 6760-7.

38. Mayuri, et al., *Molecular analysis of the dormancy response in Mycobacterium smegmatis: expression analysis of genes encoding the DevR-DevS two-component system, Rv3134c and chaperone alpha-crystallin homologues.* FEMS Microbiol Lett, 2002. **211**(2): p. 231-7.

39. Wayne, L.G. and L.G. Hayes, *An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence.* Infect Immun, 1996. **64**(6): p. 2062-9.

40. Mukamolova, G.V., et al., *A bacterial cytokine.* Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8916-21.

41. Mukamolova, G.V., et al., *The rpf gene of Micrococcus luteus encodes an essential secreted growth factor.* Mol Microbiol, 2002. **46**(3): p. 611-21.

42. Mukamolova, G.V., et al., *Muralytic activity of Micrococcus luteus Rpf and its relationship to physiological activity in promoting bacterial growth and resuscitation.* Mol Microbiol, 2006. **59**(1): p. 84-98.

43. Cohen-Gonsaud, M., et al., *The structure of a resuscitation-promoting factor domain from Mycobacterium tuberculosis shows homology to lysozymes.* Nat Struct Mol Biol, 2005. **12**(3): p. 270-3.
44. Telkov, M.V., et al., *Proteins of the Rpf (resuscitation promoting factor) family are peptidoglycan hydrolases.* Biochemistry (Mosc), 2006. **71**(4): p. 414-22.

45. Downing, K.J., et al., *Global expression profiling of strains harbouring null mutations reveals that the five rpf-like genes of Mycobacterium tuberculosis show functional redundancy.* Tuberculosis (Edinb), 2004. **84**(3-4): p. 167-79.

46. Chayen, N., et al., *Size and shape determination of proteins in solution by a noninvasive depolarized dynamic light scattering instrument.* Ann N Y Acad Sci, 2004. **1027**: p. 20-7.

47. Papish, A.L., L.W. Tari, and H.J. Vogel, *Dynamic light scattering study of calmodulin-target peptide complexes.* Biophys J, 2002. **83**(3): p. 1455-64.

48. Gast, K., et al., *Application of dynamic light scattering to studies of protein folding kinetics.* Eur Biophys J, 1992. **21**(5): p. 357-62.

49. Nicoli, D.F. and G.B. Benedek, *Study of thermal denaturation of lysozyme and other glubular proteins by light-scattering spectroscopy.* Biopolymers, 1976. **15**(12NA-NA-770103-770104): p. 2421-37.

50. Bohidar, H.B., *Light scattering and viscosity study of heat aggregation of insulin.* Biopolymers, 1998. **45**(1): p. 1-8.

51. Gutierrez, M.M., et al., *Studying low-density lipoprotein-monoclonal antibody complexes using dynamic laser light scattering and analytical ultracentrifugation.* Biochemistry, 1999. **38**(4): p. 1284-92.

52. Banerjee, S.K., et al., *Reaction of N-acetylglucosamine oligosaccharides with lysozyme. Temperature, pH, and solvent deuterium isotope effects; equilbrium, steady state, and pre-steady state measurements*.* J Biol Chem, 1975. **250**(11): p. 4355-67.

53. Cohen-Gonsaud, M., et al., *Resuscitation-promoting factors possess a lysozyme-like domain.* Trends Biochem Sci, 2004. **29**(1): p. 7-10.

## Publication list


1. **Tizzano B**., Palladino P., De Capua A., Marasco D., Rossi F., Benedetti E., Pedone C. Ragone R. and Ruvo M. (2005) "The human prion protein $\alpha$2 helix: a thermodynamic study of its conformational preferences" *Proteins: Structure, Function, and Bioinformatics,* **59**(1):72-9.

2. Palladino P., **Tizzano B**., Pedone C., Ragone R., Rossi F., Saviano G., Benedetti E. and Tancredi T. "Structural determinants of unexpected agonist activity in a retro-peptide analogue of the SDF-1alpha N-terminus" (2005) *FEBS Lett.* **579** (24):5293-8**.**

3. Fiory F., Alberobello A. T., Miele C., Oriente F., Esposito I., Corbo V., Ruvo M., **Tizzano B.**, Rasmussen T.E., Gammeltaft S., Formisano P. and Beguinot F. "Tyrosine phosphorylation of Phosphoinositide-dependent kinase 1 by the insulin receptor is necessary for insulin metabolic signalling" (2005) *Mol. Cell. Biol.* **25** (24):10803-14.

4. Granata V., Palladino P., **Tizzano B**., Negro A., Berisio R. and Zagari A.,"The effect of the osmolyte trimethylamine N-oxide on the stability of the prion protein at low pH" (2006) *Biopolymers* Feb 17 p. 234-240.

5. Ronga L., **Tizzano B**, Palladino P., Ragone R. Urso E., Maffia M., Ruvo M., Benedetti E. and Rossi F. "The Prion Protein: Structural Features and Related Toxic Peptides" (2006), *Chemical Biology and Drug Design,* **68**: 139-147*.*

6. Ronga L., Langella E., Palladino P., Marasco D., **Tizzano B**., Saviano M., Pedone C., Improta R. and Ruvo M. " Does tetracicline bind helix 2 of Prion? An Integrated spectroscopical and Computational Study of the Interaction Between the Antibiotic and a Helix 2 Human Prion Protein Fragments" (2006), *Proteins: Structure, Function and Bioinformatics*, in press*.*

7. Ronga L., Palladino P., **Tizzano B**., Marasco D., Benedetti E., Ragone R., and Rossi F., "Effect of Salts on the Structural Behaviour of hPrP α2-Helix-Derived Analogues: the Counterion Perspective" (2006), *Journal Peptide Science,* in press*.*