Universitat Autònoma de Barcelona

Institut de Biotecnologia i de Biomedicina

Instituto de Biología Molecular de Barcelona - CSIC

Towards the understanding of the Terminal Organelle of Mycoplasma genitalium: structural insights into the wheel complex

TESI DOCTORAL

Luca Martinelli

Luca Martinelli

Directors:

Dr. Ignacio Fita Rodríguez (IBMB-CSIC)

Dr. Bárbara L. Machado Calisto (ESRF)

Barcelona, 2014

Contents

List of Figures

CHAPTER 1

Introduction

Chapter 1 Introduction

1. Mycoplasmas general introduction

Since its first identification in 1889 by a German biologist named Albert Bernhard Frank, who thought was a fungus rather than a bacterium (Frank 1889), Mycoplasma has been an open debate for almost six decades. Following discoveries in the next years only made Mycoplasma classification even harder: because of their mutual relationships with other organisms, they were misleadingly thought to be viruses; and the absence of a cell wall led the scientific community to think Mycoplasma could be related to L-forms bacteria, which resemble mycoplasmas in individual and colonial morphology. It was only with the sequencing of the first Mycoplasma genome that this erroneous relationship was ruled out.

Mycoplasmas, commensals or parasites of vertebrates, are Eubacteria of the class Mollicutes, a group of organisms phylogenetically related to Gram-positive bacteria with strict host range and high host specificity. Their more characteristic features reside in the small size of their genomes, the low guanine (G) plus cytosine (C) content of their genomic DNA and the lack of a cell wall (Garnier, Foissac et al. 2001).

Due its small size genomes, mycoplasmas have limited biosynthetic capabilities. During evolution of the species, several pathways have been lost, such as the capability of synthetize amino acids, co-factors and nucleic acids. Mycoplasmas thus need to obtain most of the nutrient from the host they colonized. To overcome the deficiencies of the mycoplasmas, complex media are used for their cultivation. The media are usually based on beef heart infusion, peptone, yeast extract, and serum with various supplements (Razin, Yogev et al. 1998). Another characteristic that makes Mycoplasmas unique among bacteria is the requirements of sterols to regulate their membrane fluidity, which is done by incorporating large quantities of external cholesterol, either by depleting the cells host or by using the supplemented one in laboratory media used for their growth. Additionally, Mycoplasmas use an alternate genetic code, in which the codon UGA codes for tryptophan

rather than being used for a Stop codon. This makes even more tedious working with Mycoplasma proteins if they have to be expressed as recombinant in E.coli cells, since the UGA tryptophan triplet has to be mutated.

Even considering all the above limitations, Mycoplasma reduced genome size have attracted the attention of the scientific community, struggling to identify what is known as the minimal genome. The concept of minimal genome is based on the idea that a genome can be reduced to a minimal set of genes that should be enough to sustain all necessary biological pathways inside a viable synthetic cell, removing all the non-essential genes while keeping those for replication and metabolism.

2. Mycoplasma genitalium

In 1995, Fraser and co-workers completely unravelled the Mycoplasma genitalium genome sequence (Fraser, Gocayne et al. 1995), which ideally represents a perfect model for a first definition of minimal life, with 58070 bp and an overall G + C content of 32%, unevenly distributed across the genome, with 470 predicted coding regions, of which 374 were putatively identified and 96 had no matches from any other organism.

A genome comparison between Mycoplasma genitalium and Gram-negative Haemophilus influenza bacterium, sequenced the same year, suggested that 256 genes could be a realistic approximation of what is needed to sustain life in a cell (Mushegian and Koonin 1996), which resulted to be a good estimation, since few years later it was demonstrated by global transposon mutagenesis that 265 to 360 of its coding regions are essential under laboratory growth conditions (Hutchison, Peterson et al. 1999). However, proof of gene dispensability need to be proven by isolation of pure clonal populations, which was not done. Further and more detailed analysis pointed out that 387 of 482 Mycoplasma genitalium protein-coding genes are essential (Glass, Assad-Garcia et al. 2006), which really exceeds the theoretical projection presented ten years before by Mushegian and Koonin. These findings permit to define what the minimal gene set could be to sustain a simple synthetic cell, with 387 protein-coding genes and 43 structural RNA genes.

2.1. Mycoplasma genitalium pathogenicity

Since its first isolation in the early 80's from the urethral discharge of two homosexual men with non-gonococcal urethritis (Tully, Taylor-Robinson et al. 1981), Mycoplasma genitalium has been directly associated with sexually transmitted diseases (STD).

In men it has been associated also with balanoposthitis, chronic prostatitis, and acute epididymitis; while in women, even if the data is still limited, its infection has been related with urethritis, cervicitis, bacterial vaginosis, vaginitis, pelvic inflammatory disease, and possibly female infertility. The urogenital tract appears to be the primary tissue infected by M. genitalium, but the adherence does not appear to be restricted to uroepithelial cells in

vitro: it has been already showed that M.genitalium cells can attach, among others, erythrocytes and human fallopian tubes (Jensen 2006).

Since their first isolation, Mycoplasmas species were considered vertebrate commensals or parasites with strict host and tissue specificities. However, considering how mycoplasma species seem to be ubiquitous and how difficult and tedious is their identification (mycoplasma infections can be asymptomatic), more recently the idea that Mycoplasmas pathogens possess this high specificity has changed. In fact several Mycoplasmas were found to cross the expected natural host and thus isolated from vertebrates that are not the usual hosts (Pitcher and Nicholas 2005). For example, M. bovis and M. agalactiae, amongst others, can pass between goats and cattle; M. gatae is found in dogs and cats, and that M. orale and M. salivarum are found both in humans and other primates.

2.2. Terminal organelle

Despite its apparent simplicity and being the organism with the smallest self-replicating genome, Mycoplasma genitalium (like the other members of the pneumoniae cluster) is characterized by a cytoskeleton-like complex ultrastructure.
This polar protrusion at one end of the cell, normally referred as terminal organelle (TO) or attachment organelle (Figure 1-1), has been proven to be implicated in different indispensable functions such as cell division, host recognition, adhesion, structural support for the overall shape and thought to confer the propelling force to move the cell.

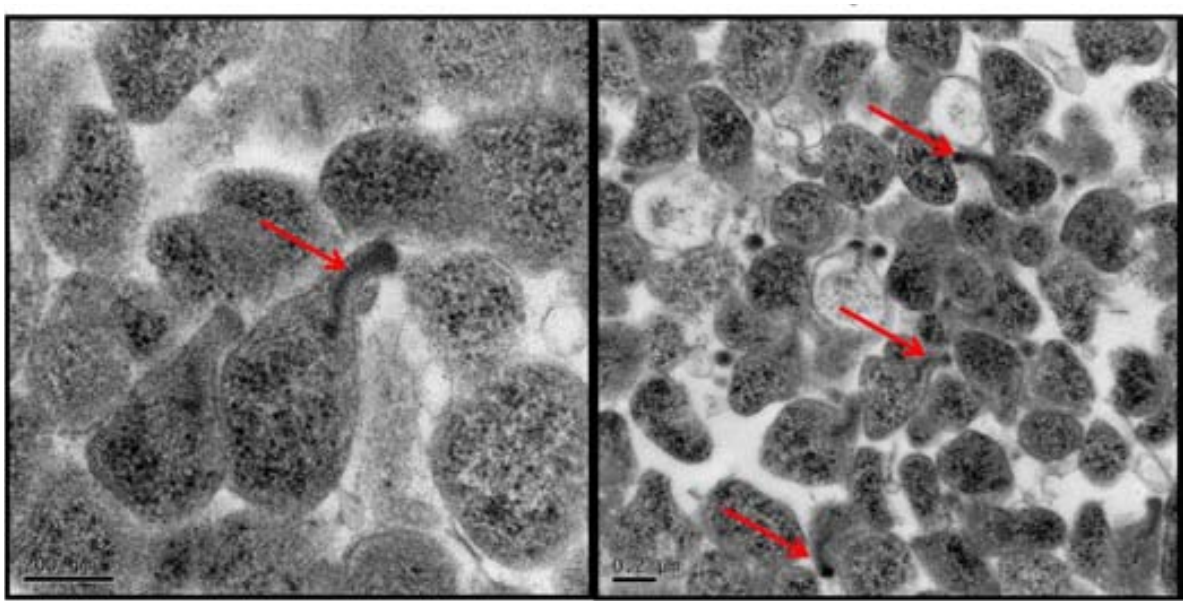

Figure 1-1. Transmission electron microscopy (TEM) micrographs of M. genitalium G37 cells.
Red arrows indicate the unique ultrastructure called terminal organelle (TO). TEM micrographs obtained by Luis Garcia Gonzales at Institut de Biotecnologia i de Biomedicina, UAB, Barcelona.

2.2.1. Cell division

In several Mycoplasma species (including pneumoniae, genitalium and gallisepticum) just before cell division, cells with two or more copies of a TO have been largely observed, supporting the idea that TO duplication event precedes celle division (Seto, Layh-Schmitt et al. 2001). Then the nascent TO, assembled from de novo protein components, migrate at the other pole of the cell and is followed by cytokinesis. These data could indicate a possible role in cell division for the TO in Mycopalsma, perhaps being the responsible for the segregation of daughter cells from the mother. A recent study demonstrated that ftsZ, a gene considerate essential for cell division by binary fission, is non-essential for M. genitalium cell growth and reveal that, in the absence of FtsZ protein, M. genitalium can manage feasible cell divisions and cytokinesis using the force generated by its motile machinery. This is an alternative mechanism, completely independent of the FtsZ protein, to perform cell division by binary fission in a microorganism (Lluch-Senar, Querol et al. 2010).

This is in accordance with the idea that in Mycoplasma, DNA duplication must take place before cell division and almost at the same time with duplication of other biomolecules. However, chromosome segregation in mycoplasma is still not completely understood, making even more complicated and difficult speculating about TO real function in cell division.

2.2.2. Adhesion and overall cell shape

In most bacterial systems, cell shape is dictated by the interaction between the cytoskeleton and the cell wall. On the other hand, lacking a cell wall, mycoplasmas managed to maintain their particular shapes thanks to the support given by those distinctive proteins forming the TO.

Over 100 Mycoplasma different species have been identified in the last century, 57 of which have been completely or partially sequenced to date. In all cases, adhesion of mycoplasmas to host cells is a prerequisite for colonization: the loss of adhesion capacity by mutation results in loss of infectivity, and reversion to the original phenotype is accompanied by recovering adhesion and therefore infectivity (Krause, Leith et al. 1983).

The idea that Mycoplasmas adhere to their host via specific proteins that work as mediator of the attachment was confirmed in the late 70's in Baseman's lab (Hu, Collier et al. 1977): after a mild treatment of M.pneumoniae cells with trypsin, the cells were not able to adhere to tracheal epithelium and erythrocyte cells, while still preserving their availability. Comparison of electrophoretic Mycoplasma pneumonia whole extract cells profiles before and after protease treatment revealed in the latter the absence of two major protein bands, named P1 and P2. Few years later another major adhesin of Mycoplasma pneumonia, P30, was identified and further analysis drawn the attention on the similarities that shares with the other adhesin P1. Both proteins are highly immunogenic, present a proline-rich region at their C-terminus and localize as clusters at the polar end of many members of the M.pneumoniae cluster. This polar end is defined and stabilized by a

peculiar cytoskeleton-like element, firstly identified after blend Triton treatment of M.pneumonia cells (Meng and Pfister 1980) and described as a network of thin filaments present throughout the entire cell body that is interconnected to another Triton X-100 (TX)-insoluble structure characteristic of pneumoniae cells, the electron-dense core (Biberfeld and Biberfeld 1970). These elements, in addition to other components lately identified (Henderson and Jensen 2006), can somehow interact with each other and form the TO ultrastrucure.

2.2.3. Gliding motility

In addition for being responsible for host recognition, adhesion and overall cell shape, the TO has been indicated as the possible propeller for the unknown type of motility that characterize the motile mycoplasmas. Since the first identification of motile Mycoplasma pulmonis in 1946, gliding motility for mycoplasma cells have been reported for several different species, including M.pneumoniae and genitalium.

Interestingly, Mycoplasmas have no surface flagella or pili, and their genomes contain no genes related to known bacterial motility and no homologs of conventional motor proteins that are common in eukaryotic motility have been found, suggesting that the mechanisms by which they move, collectively referred as gliding motility, are a novel type of motility.

Molecular and genomic studies suggested that because of different machineries responsible for gliding motility, it was plausible to divide mycoplasma in two main groups: the fast gliders, represented by M. mobile (a member of the hominis phylogenetic subgroup), and the slow gliders, for example M. pneumoniae and genitalium (two closely related members of the pneumonia subgroup). Sequence comparison of the proteins involved in gliding of slow and fast gliders showed no similarity, confirming the two groups are phylogenetically distant (Miyata 2010).

2.2.3.1 Fast gliders

Fresh water fish pathogen M.mobile, with speed of 2 to 4.5 µm s^{-1} (that means up to seven times the length of the cell per second) is the fastest mycoplasma identified to date. In laboratory, the cells of M.mobile glide without resting, resulting in a perfect model to study this motility mechanism. However, M. mobile is not yet amenable to genetic manipulation, making more complicated the assignment of a role to each protein.

First analysis of intact M. mobile cells by EM revealed a classical cone shape, with three distinguishable parts: the cell body, the neck and the head. Sequent Triton X-100 treatments aimed to better characterized the subcellular structure supporting movement and cell shape revealed a unique structure, named jellyfish structure (Figure 1-2, inset B) (Nakane and Miyata 2007), mainly composed by a solid oval bell, located in the cell head and filled with a hexagonal lattice of 12 nm periodicity, with tens of spikes (or tentacles) connected to it, riddled with 20 nm diameter particles and going throughout the cell body(Figure 1-2, inset A).

More recently, several experimental evidences allowed Miyata and co-workers to identify the protein responsible for gliding in M.mobile, by analysing non-binding and non-gliding mutants. Four proteins were identified, which are coded in tandem in M mobile genome and demonstrated not to be implicated in the jellyfish structure. Three of these four proteins (Gli123, Gli349 and Gli521), localize around the neck in an equimolar ratio, and are thought to form a trimer complex. The fourth, named P42 and which localization is still unknown, is a nucleoside triphosphate that when expressed as a recombinant protein in Escherichia coli showed ATPase activity, for which it is thought to be the motor of M. mobile (Miyata 2010).Structural information available on Gli349, named the leg and responsible for the binding of M. mobile on glass surface, and Gli521, the gear, allowed to elaborate a plausible assembly of these 4 proteins, as shown in Figure 1-2, inset C.

A working mechanism for M.mobile motility has been proposed, in which the gliding occurs through subsequent binding, pull and release of sialylgalactose molecules which are used as scaffold. The energy required for the repeated steps is given by ATP hydrolysis

performed by P42. This model, which is usually referred as centipede or power stroke model, is summarized in Figure 1-2, inset D.

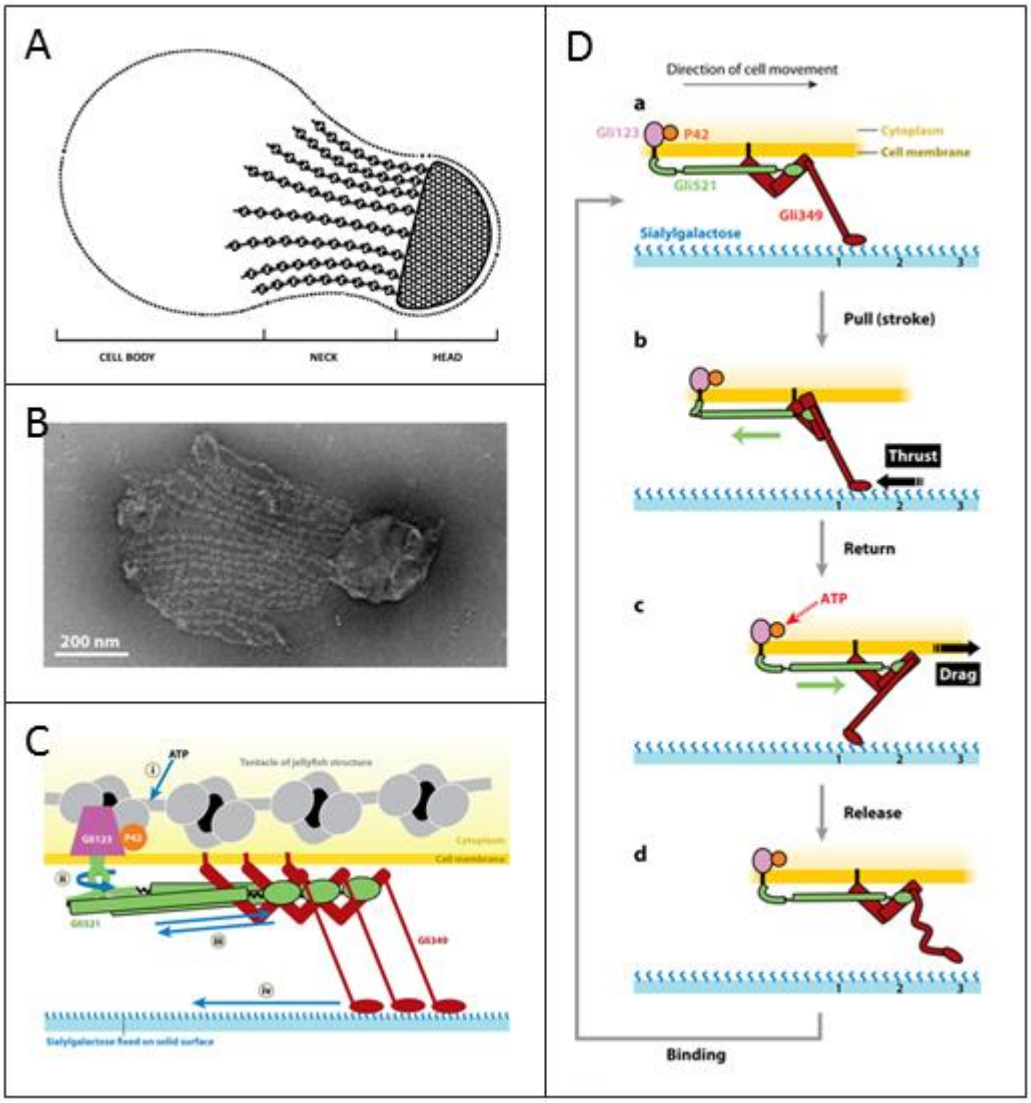

Figure 1-2. Mycoplasma mobile cell and centipede model gliding.
A) Schematic representation of the whole cell body of M. mobil., Dozens of tentacles are connecting the cell body to the head, crossing entirely the neck . B) M. mobile single cell electron microscopy (EM) image after the addition of 0.1% w/v of Triton X-100. C) Centipede model for M.mobile gliding. Images A) and B) have been adapted from (Nakane and Miyata 2007), while images C) and D) have been adapted from (Miyata 2008).

2.2.3.2 Slow gliders

Slow gliders category includes M.pneumoniae and other members of its cluser, such as M.genitalium , M.gallisepticum, M.imitans, M.testudinis, M.amphoriforme, and M.pirum. These species are all capable of gliding motility, with gliding speed ranging from 28 nm s^{-1} for M.pirum to 2970 nm s^{-1} for M.testudinis (Hatchel and Balish 2008) and are thought to share common adhesion and motility mechanisms.

Different independent experiments pointed out that, even if slow and fast gliders seem to use proteins to glide and adhere that do not share any sequence similarities, several analogies between mobile and pneumonaie motility and binding mechanisms are evident.

First evidence that the TO in M. pneumoniae was the cellular molecular motor came from an experiment in which global transposon mutagenesis was used to disrupt genes implicated in cytoadherence or motility (Hasselbring and Krause 2007). In particular, with transposon mapped to MPN311 gene, coding for a cytoskeleton protein, the resulting mutants showed the de-attachment of the TO from the cell body, left motionless, indicating that the TO appears to have the complete set of proteins necessary for gliding. The movements of the sole TO were detectable up to 30 minutes after the de-attachment, demonstrating that the principal components of this molecular motor localize within the TO. Mycoplasma pneumoniae TO's surface is covered by nap structures, reminiscent of Gli349 mobile protein, probably responsible for binding and gliding (Hu, Cole et al. 1982). The terminal organelle is supported by a cytoskeleton-like structure visible when cytosol and membranes are removed by detergent, similar to the jellyfish structure found in M. mobile. The direct binding target in M. pneumoniae (modified oligosaccharide) is closely related to sialylgalactose, binding target of M. mobile (Roberts, Olson et al. 1989).

The terminal button is bound to the cell membrane or any structures embedded in the membrane, and the rods forming the elecontron-dense core have many flexible parts and a characteristic 150° bend. Thus, an inchworm model has been suggested (Henderson and Jensen 2006, Seybert, Herrmann et al. 2006), in which the electron-dense core may repeatedly contract by bending, minimizing the distances between different components

of the terminal organelle and then spring back to a more extended conformation. When the electron-dense core is completely contracted, the wheel complex may provide the resistance necessary to release the generated force in direction of the terminal organelle's tip, forcing the cell to move forward.

2.3. Terminal organelle ultrastructure and components of slow glider member M. genitalium

M. pnumoniae and M. genitalium (two human pathogens preferentially found, respectively, in respiratory and urogenital tract) are, to date, two of the most studied members of slow gliders family, and share morphological characteristics (Figure 1-3).

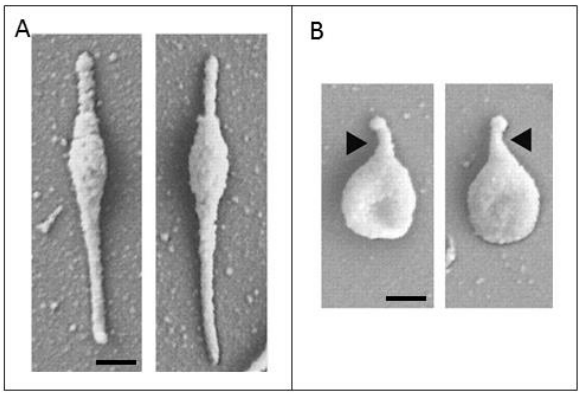

Figure 1-3. Scanning electron microscopy (SEM) of mycoplasma cells grown on glass coverslips.
Two representative cells are shown for M.pneumoanie (A) and M.genitalium (B). Black arrows indicate the curvature of M.geniralium terminal organelle (TO) which is lacking in M.pneumoniae cells. Scale bar, 250 nm. Image adapted from (Hatchel and Balish 2008).

Due to the poor gliding activity of M. pneumoniae and M. genitalium (if compared to M. mobile), studies aimed to unravel the mechanism they use to glide are time consuming and difficult, leaving the gliding mechanism for the slow gliders still unclear. Instead, adherence and binding events have been analysed in details, since they correlate with infection and virulence and can be essential to survival in vivo.

In the past years, electron microscopy (EM) was the most (and probably the only) technique that permitted to show structural features of cells, like organelles. In particular, Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM) are extremely useful tools for the ultrastructural examination of prokaryotic cells as well as for the study of the interaction between bacterial pathogens and host cells (Stadtländer

2007).However, Mollicutes are particularly difficult to handle for SEM and TEM investigation because their cell body can be easily deformed and their cytoplasmic membranes can disintegrate when subjected to the numerous specimen preparation steps required for SEM or TEM investigation. For these reason cryo-electron tomography (cryo-ET), a technique in which tomography is used to obtain 3D reconstructions from 2D tilted images, is nowadays the only technique which leaves the cells in a native state, respecting the diverse and complex cellular environment necessary to stabilize the target macromolecular structures. Recently (Henderson and Jensen 2006, Seybert, Herrmann et al. 2006), Cryo-ET permitted to analyse in details the substructure of M. pneumoniae terminal organelle, revealing that at least 3 different compartments are distinguishable: i) a wheel complex (also called bowl-complex); ii) an electron-dense core mainly composed by two segmented parallel rods of different length and thickness; iii) a terminal button, located between the most distal part of the electron-dense core and the cell membrane. Additionally, exposed proteins were found to form tightly packed 5 nm thick rows on the extracellular surface of the TO, forming the so called nap structure and being probably responsible for the adhesion of the cells. These rows localized above the terminal button and extended through the membrane, forming an additional layer of tightly packed rows which could interact directly with the terminal button itself.

First analysis aimed to characterize the components of M. pneumoniae insoluble fraction, obtained after treatment with non-ionic detergent Triton X-100, revealed that more than 50 proteins were present, including proteins belonging to different categories, such as energy metabolism, translation and heat-shock response (Regula, Boguth et al. 2001). The list of components forming the Triton shell includes, for example, heat shock protein DnaK or the elongation factor Tu, two of the most quantitatively proteins present.
Within these proteins present in the insoluble fraction, 11 ORFs are thought to be implicated in pneumoniae motility or adherence, based on subcellular localization or analysis of non-adherent and non-motile mutants (Table I.1).

Table I.1 Terminal organelle proteins implicated in cytoadherence and motility

M. genitalium		M. pneumonaie		Sequence	Subcellular
Gene name[1]	Protein name	Gene name[1]	Protein name	identity (%) [2]	localization [3]
MG191 (MgpA)	P140	MPN141	P1	44,9	S
MG192 (Mgp3)	P110	MPN142	P40 / P90	48,9	S
MG200	DNAJ-like protein	MPN119	TopJ	34,8	W
MG217	P65 homolog	MPN309 (p65)	P65	40,2	B
MG218 (hmw2)	HMW2	MPN310 (hmw2)	HMW2	57,0	R
MG491 (MG218.1°)	MG491	MPN311	P41	53,2	W
MG219	MG219	MPN312	P24	17,0	W
MG312 (hmw1)	HMW1	MPN447 (hmw1)	HMW1	32,5	R
MG317 (hmw3)	HMW3	MPN452 (hmw3)	HMW3	33,4	B/R
MG318	P32	MPN453 (p30)	P30	42,8	S/B
MG386	P200	MPN567	P200	29,2	W

[1] In parenthesis, alternative gene name.

[2] Using BLOSUM50 scoring matrix.

[3] B: terminal button; R: rod; S: TO's surface; W: wheel complex.

° old gene name.

Mycoplasma pneumoniae P1, a 170-kDa protein found in mycoplasmas, binds to solid surfaces, such as host cells or glass (Hu, Cole et al. 1982), and it is encoded in the same operon with two other open reading frames (ORFs), MPN140 and MPN142. While MPN140 encodes for a putative phosphodiesterase whose role in cytadherence is not known, the translational product of MPN142 gene, predicted to be a 130 KDa protein, undergoes a processing event and results in two proteins, of respectively 40 and 90 KDa, named P40 and P90. Both proteins have been shown to interact with P1 adhesin (Layh-Schmitt, Podtelejnikov et al. 2000, Nakane, Adan-Kubo et al. 2011), and they are thought to form the nap structure. This complex is located in the region of the membrane that forms the TO, and it has been demonstrated necessary for adherence (Seto, Kenri et al. 2005). These results are in agreement with the results obtained in M. genitalium, where P110 and P140 are reciprocally stabilized and required for terminal organelle formation or cell adhesion (Burgos, Pich et al. 2006).

P65, HMW3 (high molecular weight protein 3) and P32 are thought to be the components of the terminal button, a knob-formation located at the distal end of the electron-dense core. Immunogold labelling of MG217 gene product showed that P65 orthologue localized predominantly under the cell surface, in very close proximity to the terminal button (Burgos, Pich et al. 2008), while previous studies performed in M. pneumoniae showed that its MG217 orthologue is, at least, partly surface-exposed (Proft, Hilbert et al. 1995). P65

has been implicated in determining the curvature of the terminal organelle in genitalium, and its rupture by transposon insertion in MPN309 gene resulted in lower gliding velocity, reduced cyto-adherence, and decreased steady-state levels of several terminal organelle proteins, including P30 (Hasselbring, Sheppard et al. 2012). P30 is an integral membrane protein located above the terminal button and is implicated in both motility and adherence, with mutations impacting gliding, virulence and cell morphology. Monoclonal antibody reactive with MPN453 gene product inhibited the attachment of M. pnumoniae cells to chicken erythrocytes, confirming its role in cyto-adherence (Morrison-Plummer, Leith et al. 1986). Like P30, P32 localized to the attachment organelle, and orthologous gene replacement experiments showed that P32 functions in M. pneumoniae about as well as P30, suggesting that the sequence differences between P30 and P32 are not responsible for the obvious morphological differences seen in the attachment organelles of M. pneumoniae and M. genitalium (Relich and Balish 2011). HMW3 is a member of a family of mycoplasma proteins that include P65, P200 and HMW1 (Proft, Hilbert et al. 1995), which share in common an internal domain rich in Pro and acidic residues in repeated motifs. The Pro-rich domains probably impart an extended conformation on the protein backbone, while the hydrophobic surface provided by the Pro residues may contribute to interactions with other mycoplasma proteins (Krause 1996). HMW3 transposon mutant in M. genitalium exhibit a reduced gliding motility, although its ability to attach to solid surfaces was not completely abolished (Pich, Burgos et al. 2008). Analysis of the mg318 null mutant also revealed that HMW3 is involved in the formation of the terminal button and contributes to anchoring the electron-dense core to the cell membrane.

While HMW1 and HMW3 are part of a cluster of genes which constitutes the hmw operon, the other high molecular weight protein, HMW2, is encoded in the P65 operon (crl), together with other cyto-adherence accessory proteins such as P41, P24 and P65. HMW1 and HMW2, the main components of the rods visible after treating mycoplasma cells with detergent, are closely dependent to each other for localization and stability. In the absence of HMW1, HMW2 resulted less stable, with accelerated turnover and failing to properly localized to the attachment organelle, while loss of HMW2 resulted in failure to localize P1 and P30 adhesins normally to the TO, reduced cyto-adherence, altered cell morphology and decreased steady-state levels of HMW1, HMW3, P65, and P30 (Willby, Balish et al.

2004). In Mycoplasma genitalium, MG312 deletion mutant obtained by homologous recombination has revealed that the MG312 protein is required for the assembly of the M. genitalium terminal organelle. Reintroduction of several MG312 deletion derivatives into the MG312 null mutant allowed us to identify two separate functional domains: an N-terminal domain implicated in gliding motility and a C-terminal domain involved in cytadherence and terminal organelle assembly functions (Burgos, Pich et al. 2007).

The proximal end of the electron-dense core is in contact with a wheel-like complex that is connected to the cell periphery by fibrils and it is thought to be the connection between the TO and the cell body. The use of fluorescent protein fusions permitted to localize MPN311 and MPN312 gene products at the wheel complex (Kenri, Seto et al. 2004) and to assess the roles of P41 and P24 individually in terminal organelle development and gliding function (Hasselbring and Krause 2007). P41, essential for P24 function, is necessary for proper positioning of nascent terminal organelle and for gliding velocity, while P24 is required for wild type frequency of new terminal organelles formation. In addition P24 is the only terminal organelle protein which resulted affected when TopJ is deleted. The J-domain co-chaperone TopJ null mutant is non-adherent and non-motile, despite the fact of having a detectable electron-dense core and wild type levels of terminal organelle proteins, except for P24 (Cloward and Krause 2009).M. pneumoniae P200 null mutant exhibited a slower velocity and a higher percentage of time resting than the wild type, but not a deficiency in gliding, which was as high as the wild type (Jordan, Chang et al. 2007). Interestingly, P200 has several features in common with the well characterized cytadherence-associated proteins, HMW1 and HMW3 (Proft, Hilbert et al. 1996). These features consist of abnormal migration in SDS-PAGE, a central acidic domain with a high Pro content and a well-conserved Enriched in Aromatic and Glycine Residues motif, the EAGR box, found so far in only three proteins (and orthologous in pneumoniae and gallisepticum) of the TO, for instance MG386 (which contains five EAGR boxes), MG200 (one EAGR box) and MG312 (one EAGR box) (Calisto, Broto et al. 2012). This motif, uniquely found in members of pneumoniae cluster, has a specific contribution to cell motion, as demonstrated by Burgos and co-workers (Burgos, Pich et al. 2007). In M. genitalium, MG386 and MG200 proteins have been implicated in gliding motility, since

both mg386 and mg200 mutants exhibited a slower velocity and a higher percentage of the cells being non-motile (Pich, Burgos et al. 2006).

A schematic representation of the terminal organelle substructures organization and localization of proteins listed in Table I.1 is given in F.

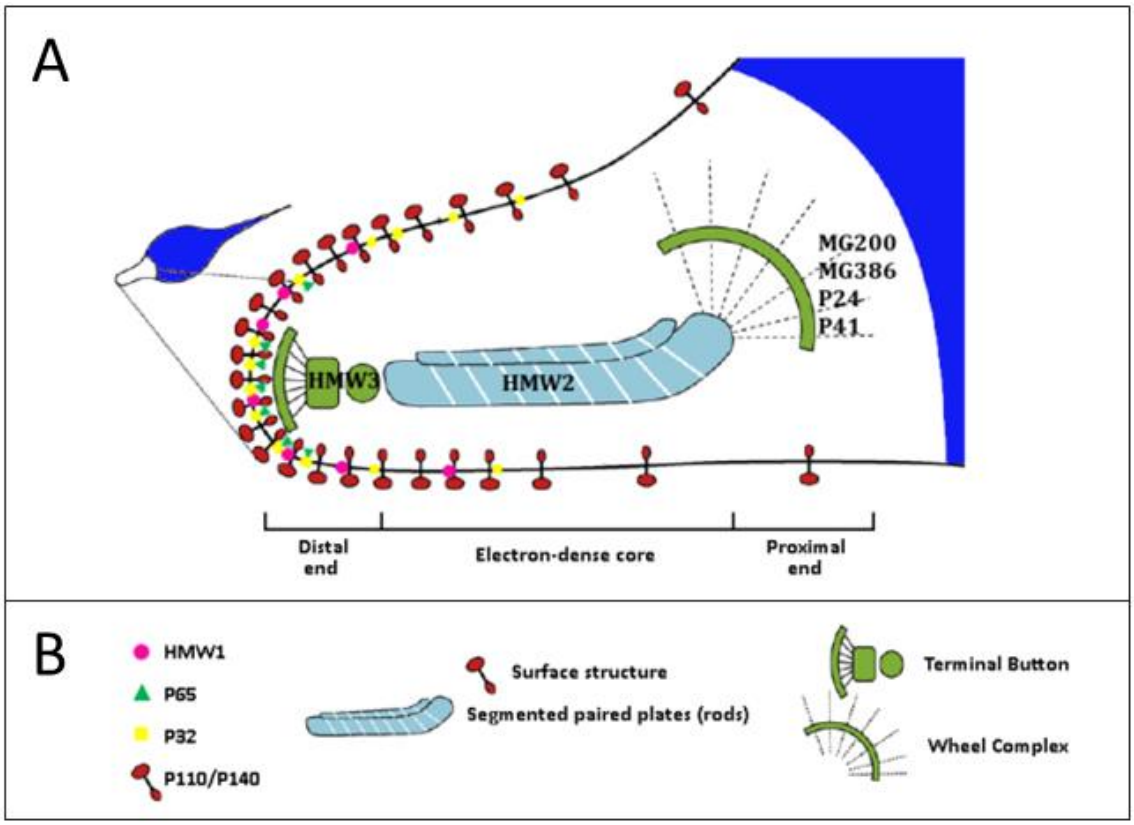

Figure 1-4. Architecture of M.genitalium terminal organelle.
A) Schematic representation of the terminal organelle ultrastructure. The cytoplasm is coloured in blue. The cytoplasmatic part of the TO can be divided into three main regions, which are the distal end, the electron-dense core and the proximal end. Proteins that are known to localize at the terminal organelle are indicated: at the tip HMW1, P32 and the major adhesins P110/P140; in the terminal button P65 and HMW3, HMW2 in the electron-dense core and MG200, MG386, MG491 and MG219 at the wheel complex. B) Representation of the structures and proteins forming the TO. Image has been adapted from (Miyata 2008)

Objectives

Objectives

Mycoplasma genitalium, a human pathogen which complete genome was successfully sequenced in 1995, is a cell wall less bacteria characterized by a reduced genome size and a novel motility system, known as gliding motility, involving proteins that do not share any sequence similarities with other known prokaryotic or eukaryotic motility systems.

Biochemical and structural studies aimed to better characterize the terminal organelle composition permitted to identify at least eleven proteins which are implicated, directly or indirectly, in adhesion and gliding motility.

Since many pathogenic bacteria are able to move and then to colonize their hosts, it is clear that virulence and motility are closely interlinked. The direct or indirect role the proteins identified at the terminal organelle have in motility and adhesion made them an attractive series of target for the development of new therapeutic agents.

Since very little is known on Mycoplasma unique gliding motility system, this thesis focuses on the structural characterization of Mycoplasma genitalium MG491 protein and on the biochemical and structural characterization of the several interactions that have been identified between MG491, MG219, MG200 and MG386 proteins, which are known to co-localize at the wheel complex.

CHAPTER 2

Materials and Methods

Chapter 2 Materials and Methods

1. List of material and equipment:

Cartesian Honeybee dispenser (Hamilton, Reno, NV, USA)

Columns and slurries for protein chromatography were from GE Healthcare Life Sciences (Uppsala, Sweden)

Crystal Phoenix pipetting robot (ARInstruments, Sunnyvale, CA, USA)

Enzymes for DNA restriction and for standard DNA procedures were from Fermentas (Vilnius, Lithuania)

Eppendorf Mastercycler® PCR machine (Eppendorf, Hauppauge, NY, USA)

Gene Genius Bio Imaging System (Syngene, Cambridge, UK)

GFXTM PCR DNA and gel band Purification Kit (GE Healthcare,Buckinghamshire, UK)

iCycler5 real-time PCR instrument (BioRad, Hercules, CA, USA)

Immobilon-PVDF Membrane (Millipore, Billerica, MA, USA)

Nano-drop® spectrophotometer (Thermo Scientific, Wilmington, DE, USA)

NMR spectrometers 600 and 800 MHz Bruker Digital Avance (San Francisco,CA, USA)

Oligonucleotide primers were synthesized by Roche Diagnostics (Mannheim, Germany)

Perfect DNA™ 1 kbp Ladder (Novagen, Madison, WI, USA)

QIAprep® Spin Miniprep Kit (QIAGEN, Hilden, Germany)

SeeBlue® pre-stained molecular weight standards (Invitrogen, Carlsbad, CA,USA)

SYBR® Safe (Invitrogen, Carlsbad, CA, USA)

SYPRO® Orange (Invitrogen, Carlsbad, CA, USA)

Zetasizer Nano DLS instrument (Malvern, Worcestershire, UK)

Vibra-Cell VCX 750W ultrasonic processor (Sonics, Newtown, USA)

2. Bacterial strains and vectors

2.1. Bacterial strains used for plasmidic DNA amplification and cloning

E. coli DH5α. Genotype: F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), λ–

E. coli XL1 Blue. Genotype: endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 $proAB^+$ $lacI^q$ Δ(lacZ)M15] hsdR17(r_K^- m_K^+)

2.2. Bacterial strains used for recombinant protein expression.

E. coli B834(DE3). B834 is the parental strain for BL21. These hosts are methionine auxotrophs and allow high specific-activity labeling of target proteins with selenomethionine for crystallography. Genotype: F^-, ompT, $hsdS_B(r_B^-$ $m_B^-)$, gal, dcm, met(DE3). Genotype: F^- ompT gal dcm lon $hsdS_B(r_B^-$ $m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

E. coli BL21(DE3). Genotype: F^-, ompT, $hsdS_B(r_B^-$ $m_B^-)$, gal, dcm, λ(DE3).

2.3. Vectors for protein expression.

pET21a and pET21d, commercial ampicillin resistant vectors (Novagen, Madison, WI, USA) used for recombinant protein expression in E. coli that add a His6-tag to the C-terminal of the interest protein.

pOPINE, expression vectors ampicillin resistant based on pTriEx2 vector that encodes for a KHHHHHH tag at the C terminus of the cloned protein.

pOPINS, expression vector ampicillin resistant based on pTriEx2 vector that encodes for a MGSSHHHHHH-SUMO N terminus tag, which should help to improve target's solubility, being a potentially solubilising fusion partner. Histagged SUMO could be easily removed by SUMO Protease (Invitrogen, Carlsbad, CA, USA) cleavage followed by a HisTrap™ HP column (GE Healthcare Life Sciences, Uppsala, Sweden).

pOPIN expression vector is based on the In-Fusion™ cloning enzyme (Clontech Laboratories, Mountain View, CA, USA) and is a ligation-indipendent cloning system which is insert sequence independent, applicable to expression in multiple hosts and capable of cloning large PCR fragments (Berrow, Alderton et al. 2007).

3. Molecular biology procedures

3.1. Microbiologic methods

3.1.1. Bacterial culture media composition

Luria-Bertani (LB) broth/agar (1L):
10 g tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar. Autoclave, let the medium cool down to about 60 °C, add the appropriate supplements and antibiotics then distribute 30-35 mL in Petri dishes.

LB broth (1 L):
10 g tryptone, 5 g yeast extract and 10 g NaCl (autoclave).

3.1.2. Antibiotics and supplements

1000× ampicillin stock solution - 100 mg/mL ampicillin was prepared in water and filtered through a 0.22 µm filter.
1000× carbenicillin stock solution - 34 mg/mL carbenicillin was prepared in 96 % (v/v) ethanol.
Isopropyl-β-D-thiogalactoside (IPTG) stock solution - 1M IPTG was prepared inwater and filtered through a 0.22 µm filter.

3.1.3. E. coli culture conditions

E. coli cells were cultivated in LB/agar plates containing appropriate antibiotic and, after growing at 37 °C overnight, a single colony was inoculated into 5 mL LB media and allowed to grow overnight with shaking until a supersaturated culture was obtained. After, this pre-culture was inoculated into fresh LB media (1/200 dilution factor) and grown in 37 °C shaking incubator until reach an OD600nm of 0.6-0.8. Cells were then equilibrated at the final expression temperature, induced with a final concentration of IPTG equal to 1mM and lastly incubated at the final expression temperature from 3 to 16 hours while shaking at 250 rpm. All media contained the appropriate antibiotic depending on the bacterial strain and the expression vector to be used. Protein over-expression was undertaken in cell cultures of 0.5 to 2 L.

3.1.4. Heat-shock transformation of bacterial cells

Bacterial competent cells are thawed on ice. For each sample to be transformed 50 µL of cells are used in a sterilized and pre-chilled eppendorf tube. Add 1 µL of the sample (except when otherwise indicated), mix gently and incubate on ice for 20 min. Heat-pulse the transformation reactions for 45 sec at 42 °C, place on ice for 2 min, add 0.5 mL LB media pre-heated at 37 °C and incubate the reactions for 1 h at 37 °C. Transformation reaction

products are spread on LB/agar plates containing the appropriate antibiotics and colonies allowed to grow overnight at 37 °C.

3.1.5. Site-directed mutagenesis

Mutagenic oligonucleotide primers were designed to introduce single amino acid changes in the native sequence using the method of Kunkel (Kunkel, Roberts et al. 1987), as previously described (Hashimoto, Yamada et al. 1996). Mutagenic PCR reaction mixture was carried out in a final volume of 50 µL containing 1× reaction buffer (0.02 M Tris pH 8.8, 0.01 M KCl, 0.01 M (NH4)2SO4, 2 mM MgSO4 and 0.1 % (v/v) Triton X-100), 10 ng dsDNA template, 0.1 µM of each forward and reverse mutagenic olignucleotide primers, 40 µM dNTPs mix and 0.05 U/µL Pfu DNA polymerase (Fermentas, Vilnius, Lithuania). After an initial denaturation step at 95 °C for 30 sec, the samples were subjected to 16 cycles of 95 °C for 30 sec, 55 °C for 1 min and 68 °C for 12 min (last step duration can change according to the size of the amplification fragment).
The PCR product was placed on ice to cool the reaction to ~37 °C. DpnI (Fermentas, Vilnius, Lithuania) was added to the amplification reaction and incubated for 1 h at 37°C. After digestion, 10 µL of the DpnI-treated DNA was transformed in E. coli DH5 competent cells and lately sequenced to check for positive mutant colonies.

4. Protein production and general analysis methods

4.1. Preparation of total protein extracts

Cells were harvested by centrifugation at 4500 g for 30 minutes at 4 °C, resuspended in lysis buffer (50 mL per 1 L of culture) of 0.02 M Tris-HCl pH 8, 0.5 M NaCl, 0.02 M imidazole and 0.3 mg/mL lysozyme supplemented with a cocktail of protease inhibitors from Roche

Diagnostics (Mannheim, Germany). The extract was incubated for 10 min at room temperature and cells were mechanically disrupted by sonication, for 10 cycles at 250 W (20 seconds on, 59 seconds off), while kept in an ice bath for the entire sonication. The lysate was then centrifuged at 45000 g for 30 minutes (at 4 °C) twice and then filtered through a 0.22 µm filter.

4.2. Protein electrophoresis in denaturing conditions (SDS-PAGE)

Polyacrylamide gel electrophoresis (Shapiro, Vinuela et al. 1967, Weber and Osborn 1969) was preformed in a slab gel following standard procedures. The samples were denatured before loading into the gel by heating at 80 °C for 5 min in the presence of 5 % (w/v) SDS and 0.025 % (w/v) bromophenol blue. Molecular weight SeeBlue® pre-stained protein standards (Invitrogen, Carlsbad, CA,USA) were myosin (250 kDa), BSA (98 kDa), glutamic dehydrogenase (64 kDa), alcohol dehydrogenase (50 kDa), carbonic anhydrase (36 kDa), myoglobin (30 kDa), lysozyme (16 kDa), aprotinin (6 kDa) and insulin B chain (4 kDa).

4.3. Polyacrylamide protein gels staining in CBB

Staining of polyacrylamide protein gels with CBB is a quantitative and protein specific method that can detect 0.3-1.0 µg of protein per band. A completely blue gel is obtained after submerge the gel for at least 15 min in a 0.1 % (w/v) CBB R-250 solution prepared in 7 % (v/v) aceti acid and 15 % (v/v) ethanol. The gel is then distained in 7 % (v/v) acetic acid and 15 % (v/v) ethanol in order to observe the protein bands.

4.4. Electroblot onto PVDF membranes

Samples are separated by SDS-PAGE. Meanwhile, the PVDF membrane is rinsed for a few seconds in 100 % (v/v) methanol for activation, washed in MiliQ water and stored in transfer buffer (0.025 M Tris pH 8.3, 0.192 M glycine, 20 % (v/v) methanol and 0.05 % (w/v) SDS). The gel is sandwiched between the activated PVDF membrane sheet and several sheets of 3MM paper (Millipore), assembled in a wet blotting apparatus and transferred at 4 °C for 2 h at 80 V. The membrane is then stained for 1 min in 0.1 % (w/v) CBB R-250 prepared in 50 % (v/v) methanol, destained for 10 min (50 % (v/v) methanol, 10 % (v/v) acetic acid) and washed with water. The protein bands electrobloted onto the PVDF membrane were used for N-terminal EDMAN sequencing.

4.5. Edman sequencing

Edman sequencing of protein bands were done in the Proteomics platform of the SCT-UB.

4.6. Protein purification chromatographic techniques

In order to study a protein structure the protein must be isolated and purified to homogeneity. The high purity level is generally achieved after several chromatographic purification steps. As each step usually results in some degree of product loss, a purification strategy was defined to reach the highest purification level in the fewest steps. These were selected depending on the target protein size, charge and solubility. Chromatographic methods were applied using automated FPLC equipment.

4.6.1. Immobilized metal ion affinity chromatography

The first purification step involved Ni2+-based immobilized metal ion affinity chromatography (IMAC) of the supernatant solution obtained after cell disruption by sonication. The target protein, which contained a histidine tag, bound specifically to the column beads. Affinity columns coupled with Ni2+ ions (HisTrap™ HP; GE Healthcare Life Sciences, Uppsala, Sweden) were equilibrated with 0.02 M Tris-HCl pH 8, 0.500 M NaCl and 0.02 M imidazole (buffer A) to avoid unspecific binding and improve sample purity. The target protein was then eluted with a linear gradient of buffer B (0.02 M Tris-HCl pH 8, 0.500 M NaCl and 0.500 M imidazole).

4.6.2. Gel filtration chromatography (size-exclusion)

Gel filtration chromatography (GFC) was used as the second and usually last purification step to further purify each protein and evaluate homogeneity and oligomerization state. Prepacked Superdex gel filtration columns (GE Healthcare Life Sciences, Uppsala, Sweden) of appropriate separation range were used to further purify the target protein and separate it from contaminants and eventually present degradation products.

4.7. Protein cross-linking with gluteraldehyde

In order to characterize the possible oligomeric forms of a purified protein in solution 10 μL drops of 0.5-1 mg/mL protein solution were equilibrated from 5 min to 2 h against 40 μL drops of acidified 25 % (v/v) glutaraldehyde (Fadouloglou, Kokkinidis et al. 2008). The protein drops were then analyzed by SDS-PAGE.

4.8. Limited proteolysis

In order to identify protein fragments that can fold properly and then behave as protein domains, limited proteolysis was used to check whether smaller variants of the target proteins were less recalcitrant to crystallize and more stable. The rational of the technique is that a folded protein is rather resistant to proteolysis, allowing the identification of more compact fragments. 100 µl of 1mg/ml protein solution was treated with diluted solutions of trypsin and papain proteases (each one dissolved in its reaction buffer) and incubated at 20 °C. A 15 µL aliquots were taken from the reaction at certain intervals of time, the reaction blocked with SDS loading buffer and then loaded into a SDS gel to check the eventual proteolytic cleavage.
Reaction buffer for trypsin: 0.450 M Tris-HCl pH 8. Reaction buffer for papain: 1 M K_2HPO_4 pH 7.5, 0.025 M EDTA, 0.030 M 2-mercaptoethanol (2-ME or β-ME).

5. Biophysical methods for protein-protein interaction characterisation

Many of the functions exhibited by a protein in the cell require a set of specific protein-protein interactions (PPI). Thus, protein function must be analysed in the larger context of binding events that each protein may form with interacting partners, in a specific localization inside the cell, at a precise time of the cell life. Consequently, it is normal to imagine that identifying possible PPI would make easier to infer a possible function for the interacting partners.

PPI are often divided into two general categories, based on the magnitude of the interaction strength. For this, PPI are often referred as strong interaction (molar dissociation constants K_d down to $10^{-9} \div 10^{-12}$ M) and transient (or weak) interaction ($K_d \geq 10^{-4}$ M).

Over many decades, several biophysical methods have been developed to identify in vitro proteins that interact, and each one of these approaches present limitations (due to the environmental setup or intrinsically associated to the technique) and advantages if compared to the others (Phizicky and Fields 1995, Berggard, Linse et al. 2007).
To better evaluate all the possible interactions between the investigated proteins, three different approaches were used, such as nuclear magnetic resonance (NMR) spectroscopy, surface plasmon resonance (SPR) and size exclusion gel chromatography.

5.1. Nuclear magnetic resonance spectroscopy

NMR spectroscopy is highly suited to investigate molecular interactions at a close physiological condition and is particularly suited for the study of low-affinity, transient complexes. It can provide information on protein surface interaction and their dynamic properties during protein recognition (O'Connell, Gamsjaeger et al. 2009).

The interactions investigated by NMR spectroscopy here presented are the results of collaboration with Daniela Lalli, at CERM (Magnetic Resonance Center, Sesto Fiorentino, Florence, Italy). All NMR experiments were carried out at 298K on Bruker Avance spectrometer operating at proton frequencies of 700 MHz, equipped with cryoprobe. All NMR spectra were processed with Topspin version and analysed with CARA (Keller and Wuthrich 2004).

NMR experiments for resonance assignment were performed on 1 mM ^{13}C, ^{15}N-labelled MG200 EAGR box in 0.02 M Phosphate buffer (pH 6.5) containing 0.1 M Sodium chloride and 10% (v/v) D2O. Backbone resonance assignments of MG200 EAGR box were carried out through conventional multidimensional NMR techniques based on triple resonance experiments. Backbone resonance assignments of MG200 EAGR box were transferred from pH 6.5 to pH 7.4 following the chemical shift differences of the amide peaks occurring at different pH, i.e., 6.5, 7.0 and 7.4. The φ and ψ dihedral angle constraints were derived from the chemical shift analysis by using the TALOS+ program (Shen, Delaglio et al. 2009). Titrations of 50 µM ^{15}N-MG200 EAGR box prepared in 0.02 M Phosphate buffer (pH7.4)

containing 0.1 M Sodium chloride with unlabelled MG491, MG491-Nt_1-308, MG491-Ct_peptide1 and MG491-Ct_peptide2 all in the same buffer as the MG200 EAGR box were followed through ^{1}H-^{15}N HSQC. The titration experiments of MG491, MG491-Nt_1-308, MG491-Ct_peptide1 and MG491-Ct_peptide2 over ^{15}N-MG200 EAGR box were performed until ratios of 1:8, 1:4, 1:7 and 1:4, respectively, were reached. The chemical shift change (in ppm) of the individual amide pairs was defined as the weighted average ^{1}H and ^{15}N chemical shift variations according to (Garrett, Seok et al. 1997).

5.2. Surface plasmon resonance

Surface plasmon resonance is an optical technique used to identify and characterize interaction of two different molecules, in which one is fixed on the surface of a sensor chip and the other is mobile. The binding of the mobile molecule, or analyte, to the immobilized molecule, or bait, causes a change in the refractive index on the surface of the sensor chip. This change in SPR angle is reported as resonance units (RU), with 1 RU roughly equivalent to a surface density of a protein of approximately 1 pg/mm^2.Thus plotting the SPR signal over time during the interaction between the two investigated proteins or molecules results in a sensorgram, a visual representation of the interaction over time (Figure 2-1).

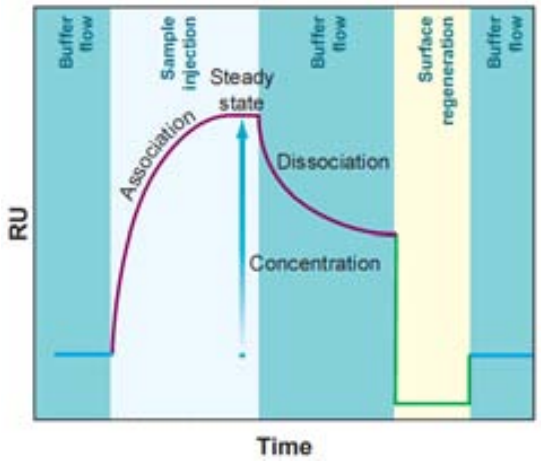

Figure 2-1. Figure 2-2. Schematic representation of a general sensorgram.
The progress of an interaction monitored as a sensorgram is shown. Cyan line at the beginning and at the end represents a continuous flow of buffer. An increase in signal (association) is observed when the analyte binds during the sample injection to the ligand immobilized on the surface of the sensor chip. At the end of the injection the sample is replaced by a continuous flow of the buffer and the decrease in signal (disassociation) reflects the disrupture of the analyte-ligand interaction. Surface regeneration is performed to remove from the chip surface the remaining analyte (regeneration). Image was adapted from (Wilson 2002).

If after the immobilization of the bait on the surface of the sensor chip the free analyte molecule does not interact with it, then the SPR angle will be the same in the reference and sample flow cells, which results in a zero net RU response that indicates no binding occurred between the bait and the analyte (Wilson 2002).

Since the first application of SPR phenomenon to study molecular interactions in real time (Jonsson, Fagerstam et al. 1991), hundreds of publications have reported successful applications for this label-free real-time binding approach. The experimental protocol for examining the binding interaction between two biomolecules by SPR technology can be summarized as a four-step process: (i) immobilize one of the interacting partners (ligand or bait) to a gold-plated chip; (ii) inject the second binding partner (analyte), and record the corresponding sensorgram (a real-time interaction curve); (iii) perform step (ii) using a series of linear concentrations of the analyte; and (iv) choose an appropriate kinetic model that best fits raw data, to extract rate-constant estimates(Szabo, Stolz et al. 1995). The basic technical concept can be summarized as the detection of specific adsorption of an analyte in continuous flow to an immobilized molecule at the same spot (on the chip) and at the same time as it occurs. Analyte-ligand association and dissociation events are monitored by changes in molecular mass accumulation at the solid-liquid interface on the sensor chip surface.

Originally, SPR experiments for the detection of possible PPIs could only be performed by running sequentially the above steps until a full analyte concentration series was measured. However, more recently different and more sophisticated biosensors were introduced into the market to increase the sample throughput of SPR experiments. One of these approaches is the One-shot Kintetics™(Bravman, Bronner et al. 2006), approach fully exploited by the ProteON™ XPR36 (BioRad, Hercules, CA, USA) Protein Interaction Array System. This innovative technique allows the simultaneous monitoring of multiple pairs of interacting partners, including proteins, small molecules and antibodies (Bravman, Bronner et al. 2008).

The ProteON™ XPR36 system made possible the generation of a six per six interaction array for the analysis of up to six ligands with up to six analytes, on the same chip and at the same time. ProteOn HTG sensor chips (BioRad, Hercules, CA, USA) were used to immobilize up to five different proteins or their domains on the chip, according to the manufacturer's instructions. One channel was used to immobilize a histagged recombinant version of human prolactin receptor (PRLR), which represented the negative control to account for possible non-specific interactions of the tested analytes. A general protocol included the activation of the chip using 60uL of 1mM Ni_2SO_4, a subsequent 10 minutes injection of the ligands on the chip at a flow of 25µl/min, a stabilization step in which the chip was washed with running buffer only (20mM potassium phosphate buffer pH7.5, 500mM NaCl, 200µM EDTA and 0.02% Tween 20) to remove any unbound ligand from the its surface. Then analytes injection is performed at 25µl/min for 10 min, to avoid sample consumption. Fine optimizations of a variety of parameters such as association or dissociation time, analyte concentrations and injection flow rates in always needed to obtain high.quality sensorgrams. Lastly, a 10 min injection of a 300mM EDTA solution is used to strip off the ligands and regenerate the chip. These preliminary conditions usually provide a starting point for the next optimization of the experimental conditions, in which the concentrations of the molecules, the flow rate and the contact time are varied systematically, to maximize the complex formation and its detection.

Similarly, positive binary interactions detected during the mid-throughput screening were further investigated using a BIAcore® 3000 (GE Healthcare). In this configuration, the proteins were immobilized as ligands at the carboxymethyldextrane surface of a CM5chip (GE Healthcare) according to the manufacturer's instructions for amine coupling. Protein immobilization to the chip surface was performed at a flow rate of 10µl/min and the same flow rate was applied for sensor regeneration, using a 500mM NaCl solution containing 5µM EDTA. Measurements were performed at 25°C in 20mM potassium phosphate buffer (pH7.4) containing 0.1M Sodium chloride and 0.005% Tween 20.

5.3. Size exclusion gel chromatography

GFC is a method for separating macromolecules based on differences in the hydrodynamic radius of each macromolecules present in solution. The volume at which one protein elutes from the column is directly proportional to the time the macromolecules stays in column, and inversely proportional to their molecular weights (MW). Therefore size exclusion chromatography has been used not only for protein purification purposes but also to estimate protein MW in a coarse but quick and almost inexpensive way. If two (or more) macromolecules inside the column interact, the resulting hydrodynamic radius of the complex will be bigger than the singular ones. This change in the radius will cause a change in the elution volume of the species in solution, giving a rough estimation of the complex MW if a proper calibration curve for the column is available. If compared to the previous two methods used to detect PPI, it's the less sensitive technique, since it can only detect strong interactions, which result in the formation of a stable complex that can be eluted as a single chromatographic peak.

6. Protein crystallography

6.1. Protein crystallization

Protein crystallization is the major bottleneck in X-ray crystallography-based structure determination, which accounts for the majority of solved structures.

Among the different methods employed to obtain crystals, vapour diffusion is the most common. In this method, a droplet of solution containing protein, a precipitating agent and, if necessary to stabilize the protein, salt is equilibrated by vapour diffusion against a larger volume of solution, named reservoir, containing higher concentrations of precipitants and salt. As the droplet and the reservoir equilibrate, water leaves gradually

the drop, while salt and precipitant concentrations slowly increase. During this equilibration time can occur that the protein reaches its saturation point and, thus, comes out of solution as a useless precipitate, or, hopefully, as a well ordered single crystals. Clearly, precipitants and macromolecular concentrations play an important role in this scenario: if the system reaches the saturation level too quickly, a crystalline precipitate or hundreds of tiny useless crystals can appear easily. Contrarily, if the concentrations of protein or precipitants are too low, the supersaturated level necessary for the nucleation and the growth of crystals will never be reached. However, concentrations of precipitants and the target macromolecule are not the only parameters that must be sampled to find the best condition in which a protein can crystalize. Other variables such as the presence or absence of particular ions, changes in pH, temperature at which crystallization experiments take place and the inclusion of additives which should promote the crystallization of biological macromolecule are all parameters that can alter this equilibrium. Two different formats are normally employed with this method, known as sitting drop or hanging drop. Essentially, the only difference is the orientation of the droplet containing the protein relative to the reservoir: in hanging drop setup, the protein is located vertically above the reservoir solution, while in sitting drop setup the protein is positioned a bit higher than the reservoir, on a tiny shelf normally on a side of the microwell.

The initial crystallization screenings were carried out in an automated way using MRC 96-well sitting drop crystallization plates (Molecular Dimensions Ltd., Newmarket, UK) with nanoliter scale drops (range 100-400 µL) dispensed by the crystallization robots Cartesian Honeybee (Hamilton, Reno, NV, U.S.A.) or Crystal Phoenix (ARI,Sunnyvale, CA, USA). To sample the crystallization space of the protein of interest, commercially available spare matrix screens were used. If no promising conditions were found, parameters such as the concentration of the protein stock, the ratio between protein and reservoir in the drop or the temperature were altered one at a time, to avoid screening too many variables at once. If a positive hit was found using these preliminary screens, subsequent optimizations of the condition were carried out in 96-well plates (as the ones used for the initial screenings) and in 24-well plates (VDX™ plate with sealant, Hampton Research, CA, USA), which allow microliter scale drops instead. These optimization steps include increasing or decreasing the initial concentration of the precipitants or salts of the reservoir in small steps, including

in the reservoir particular additives which should promote the crystallization of the target macromolecule, since they can alter protein solubility, or changing the crystallization temperature.

6.1.1. Chemical modification to improve crystallization

It is already well established that chemical modification of certain residues can alter the propensity of a protein to crystalize, reducing its solubility and favouring the formation of intermolecular contacts. Among these modifications, reductive methylation of lysine residues has been proven value for protein crystallization (Walter, Meier et al. 2006), inducing a decrease of the overall solubility, due to the increased hydrophobicity of the modified lysine sidechains. 20µl 1M dimethylamine-borane complex (ABC) freshly prepared and 40µl 1M formaldehyde (from 37% stock) are added per ml of protein solution at 1 mg/ml, and the reaction is mixed and incubated at 4°C for two hours. This treatment is repeated one more time and after the 2 hours incubation a final addition of 10µl of ABC is performed before the final overnight incubation at 4°C. Since no clear precipitation was detected, the final volume was concentrated by centrifugation using Vivaspin ™ centrifugal sample concentrators (GE Healthcare Life Sciences, Uppsala, Sweden) of appropriate volume and molecular weight cutoff and the reaction was then loaded onto a prepacked Superdex gel filtration column (GE Healthcare Life Sciences, Uppsala, Sweden) in presence of 50 mM Tris buffer (to deactivate any unreacted reagents) and crystallization experiments were set up immediately.

6.2. Recovering the lost phases

In addition to the most serious bottleneck in protein crystallography represented by the production of suitable diffracting crystals, another severe limitation is the loss of information, when making a physical measurement, concerning the phase. When collecting X-ray diffraction data from a crystal, we measure the intensities of the diffracted waves scattered from a series of planes that we can imagine slicing through the crystal in all

directions. From these intensities we derive the amplitudes of the scattered waves, but in the experiment we lose the phase information. This is generally known as the 'phase problem' (Taylor 2010)

In X-ray crystallography, there are three different techniques to recover the lost phases and therefore solving a crystal structure: *ab initio* methods, the molecular replacement (MR) method and heavy atom (HA) method.

6.2.1. Ab initio method

The use of *ab initio* methods as a way of solving protein structures is not feasible due to the computational resources needed, although they are routinely employed to solve the structure of small molecules, where the diffraction data usually extend to atomic resolution. This requirement has reduced the use of direct methods in phase determination for proteins, even though has been used successfully to phase small protein, in instance HEW lysozyme, composed of 1001 unique nonhydrogen atoms (Deacon, Weeks et al. 1998).

6.2.2. Molecular replacement method

Another way to obtain a set of starting phases is to use an ensemble of techniques that are nowadays known as molecular replacement (MR), a term lately introduced by Rossmann, the same who first described the use of noncrystallographic symmetry (NCS) as a tool for structure determination (Rossmann and Blow 1962). If coordinates of a homologous protein are known, then it is possible to use them for phase determination. More precisely, the phases from the phasing model can be used directly as initial phases only if the model used has the same orientation and position of the unknown protein in the unit cell. Since this is not the most common case, the searching model has to be moved and rotated within the unit cell to find the best overlap with the target protein. This is achieved with the use of a rotational function, which rotate the search model as a rigid body, to orient the molecule

within the unit cell, and then with a translation function, for the correct positioning of the molecule.Originally, MR techniques were based on the Patterson function, a modified version of the electron density function that can be directly computed from the measured intensities, since in this notation the structure factors do not depend on the phases and the amplitudes, but only on the squared amplitude and then it can be calculated without any prior knowledge of the phases. As more and more structures are available, the chances of finding a homologous structure are becoming higher and MR techniques are now largely used to solve unknown structures.

6.2.3. Anomalous scatters

When a homologous structure is not available, the most used method to solve protein structures in protein crystallography is the HA method, in its various alternatives. In general, the initial phases are derived from the differences in between a native crystal and a crystal which contain heavy atoms, normally referred s anomalous scatters. These anomalous scatters could normally be present in the native protein, without the need of any prior modification of the protein or the crystal. These cases include, for example, a metal cofactor for metallo-proteases or sulphur for proteins in general. However, in most cases, the protein is modified, before or after crystallization, to incorporate such anomalous scatters. Once the anomalous scatters are present in the target protein, the initial phases are derived by finding the positions of HA that have been added to the protein through soaking or are intrinsically present in the unknown structure. Methods such as SAD (Single–wavelength Anomalous Diffraction), MAD (Multi-wavelength Anomalous Diffraction), MIR (Multiple Isomorphous Replacement), SIR (Single Isomorphous Replacement) or a combination of these are all based on the anomalous signal given by some atoms of the investigated protein, and are routinely used as ways of solving novel protein structures.

6.2.3.1. Heavy atom soaking of native crystals

If native crystals are available, one approach is to soak the crystals into a solution containing low concentration of heavy metal salts or coordination compounds for prolonged time. Such technique presents several disadvantages, being highly time consuming and frequently unsuccessful, due to the low binding of the HA and the deterioration of the crystal lattice , which led to crystals with poor diffracting quality after the soaking procedure. Dauter demonstrated that a quick (less than 1 min) soak of protein crystals in a cryo-solution containing bromide or iodide anions leads to incorporation of these anomalous scatters into the ordered solvent region around the protein molecules (Dauter, Dauter et al. 2000). Additionally, Sun and colleagues compared long (12 h) and short (15 min) $HgCl_2$ derivatives of two protein structures and observed that the phasing statistics were comparable, while the shorter soak resulted in a derivative more isomorphous to the native than the longer soak (Sun and Radaev 2002, Sun, Radaev et al. 2002). In addition to the heavy atoms reviewed by Petsko almost twenty years ago (Petsko 1985), lately other heavy atoms or compounds have been employed successfully for phasing novel structure. Among these new compounds it's worth mentioning 5-Amino-2,4,6-triiodoisophthalic acid (I3C)(Sigma), a novel phasing agent in which the three iodine atoms form an equilateral triangle with 6.0 Å side lengths, making it easily to identify in an anomalous electron density map. For heavy-atom derivatization a stock solution with a high concentration of the phasing atom or coordination compound is desirable, especially for soaking experiments. To obtain a concentrated HA stock solution, the compounds were dissolved directly, when possible, in the crystallization condition in which the protein crystalized, for a final concentration of 10 mM. In case of I3C, only lithium hydroxide solution (LiOH) used as base produces an I3C salt with high solubility, thus 1 M I3C solution was prepared as previously described (Beck, Krasauskas et al. 2008), by dissolving 280 mg of I3C salt with 500 µl of 2 M lithium hydroxide solution, used to fully deprotonate the I3C carboxyl groups, and then removing the eventually formed precipitate.

Numerous heavy atoms must typically be screened before one or more is found that binds without damaging the crystal or disrupting the crystal lattice. In addition, carefully optimization of other parameters such as soaking temperature, presence or absence of

light, concentration of the HA used, and length of soaking time is necessary to reduce at minimum levels the unwanted effects of soaking on proteins crystals, possibly resulting in useless poor diffracting crystals.

After the soaking procedure all the crystals were back soaked with the original conditions without HA, to reduce non-specific HA binding and to make easier to determine whether or not the HA has entered the crystal lattice, by a simple fluorescence scan prior to data collection.

6.2.3.2. Production of selenomethionine mutants

Even if a complete derivatization by soaking can be achieved with shorter times, this technique still present some limitations and disadvantages, which make necessary to find an alternative to the soaking approach. Said that, the most general approach to solve the phase problem in protein crystallography is to modify the primary sequence of the investigated protein or, better said, to substitute methionine residues by selenomethionine (SeMet) ones. In view of the fact that selenium is a useful anomalous scatter, SeMet labelled protein production offers a general method for introduction of anomalous scatters into cloned proteins in an easy and robust way. On the contrary of what happen with soaking, where it's impossible to known how many scatters will bind to the investigated structure, when phasing with the anomalous signal provided by selenium atoms it's indispensable to check the number of methionine residues the target protein contains. As a rule of thumb, when dealing with the signal of selenium atoms, for a successful SAD or MAD data experiment, at least one modified methionine per 100 amino acids is required, even if, with favourable conditions, may be possible to succeed with a lower ratio(Hendrickson and Ogata 1997). When no methionine is present in the interest protein sequence, variant proteins were produced by site-directed mutagenesis as explained in 3.1.5. Ohmura and co-workers demonstrated that, in addition to the "safe" methionine for leucine residue substitution by site-directed mutagenesis (Bordo and Argos 1991), also the substitution of methionine for isoleucine in a protein could be performed without any significant change of the overall and local structure (Ohmura, Ueda et al. 2001). Thus, the

DNA oligonucleotides used as mutagenic primers were designed such that the trinucleotide encoding for the chosen isoleucine residues were replaced by methionine residues (changed the target codon from ATC to ATG), using the wild type gene as a template.

The mutants were then cloned into a pOPINE vector and produced as C-terminal histidines tagged proteins in Ecoli B834(DE3) cells, which is auxotrophic met- strain, and BL21(DE3), which is not. In both cases SelenoMet™ base media and nutrients (Molecular Dimensions Ltd., Newmarket, UK) were used to grow the cells and produce the SeMet variants. SelenoMet™ is a synthetic media supplemented with glucose, vitamins and amino acids with the exception of L-methionine, derived from the original M9 minimal media recipe (Ramakrishnan, Finch et al. 1993). When expressing the recombinant protein with BL21(DE3) cells, a 100 mL pre-culture was grown overnight at 37°C in presence of 400 µl of L-methionine at 10mg/ml, 2mL of glucose at 50 % w/v (freshly prepared and filtered on a 0.22µm filter), and appropriate antibiotic. The cells were than pelleted at 4500g for 15 minutes, washed three times with large amount of phosphate buffered saline (PBS) 1x, to remove the L-methionine that has not been incorporated by the cells, and then resuspended with 2mL of PBS 1x, which were used to inoculate 1L of SelenoMet™media, in abundant presence of L-selenomethionine. Instead of using IPTG based cells induction, the medium was supplied with 20 ml per liter of OnEx Solutions 1 (a blend carbon source solution optimized for tightly regulated uninduced growth), 50 ml per liter of OnEx Solutions 2 (a concentrated buffer containing nitrogen, sulphur and phosphorous that mediates metabolic acid production) and 1 ml per liter of OnEx Solutions 3 (to provide magnesium for maximal cell density)(Novagen, Madison, WI, USA). This method of growth and induction relies on medium components that are metabolized differentially to promote high-density growth and automatic induction of target proteins from lac-based promoters (Grabski, Mehler et al. 2003). Cells were grown for 6 hours at 37°C, then the temperature was lowered to 25°C and growth was continued for a further 20 hours before cells harvest. Considering that L-methionine is incorporated preferentially than L-selenomethionine residues , when the recombinant proteins were expressed in BL21(DE3) cells the protocol was modified to block the methionine biosynthesis, by inhibiting aspartokinases in the presence 10 ml per liter of lysine, threonine, phenylalanine (at 10 mg/ml each), and 10 ml per liter of leucine, valine and isoleucine (at 5 mg/ml) amino acids,

all freshly prepared (Doublie 1997), while no extra amino acids were not added when expressing the SeMet variants in B834 (DE3) cells. The final mutants were successfully purified from bacterial cells following the protocol established for the full length protein.

6.3. Single crystal X-ray diffraction

6.3.1. Data collection and data collection strategies

All the datasets were collected at beamline X06SA (PXI) at SLS, at the Paul Scherrer Institut, Villigen, Switzerland, at beamline PROXIMA1 at SOLEIL, Gif-sur-Yvette, France and at beamlines ID23-1 and ID29 at ESRF, Grenoble, France. For the anomalous data (except for I3C), a fluorescence scan was performed every time to locate in an automatic way the absorption edge of the scatter and a peak dataset was collected at the corresponding wavelength. For an optimal measurement of anomalous differences, MiniKappa mounted on beamline ID29 was used to re-orient the investigated crystal before data collection, aligning a crystallographic axis along the rotation axis such that Bijvoet mates (a reflection and the Friedel pair of its symmetry equivalent) can be measured on the same image(Brockhauser, Ravelli et al. 2013).

6.3.2. Data reduction (indexing, integration and scaling)

All the data collected were integrated with XDS (Kabsch 2010, Kabsch 2010). Unless otherwise stated, the output unmerged XDS ASCII file reflection .HKL was then converted to MTZ format by XDSCONV script, the conversion program of the XDS suite. The resulting reflection file was finally scaled with SCALA (Evans 2006, Evans 2011). Datasets with anomalous differences were further processed as follow.

6.3.3. Phasing

The phasing statistics of each dataset containing anomalous differences were assessed with the inspection of alternatively CORRECT.LP logfile (produced by XDS), by SCALA logfile, by Bruker program XPREP (Bruker 1998) or by analysing the statistics reported by SHELXC from SHELX suite (Sheldrick 2008, Sheldrick 2010).The hkl2map GUI interface (Pape and Schneider 2004), which connects SHELXC/D/E programs, was used to generate useful plots and statistics of each run and to edit in a quick way the various input files needed to run the SHELX triad.

After SHELXC was used to prepare the input data with an estimation of the structure factor amplitude for the substructure atoms, these estimated structure factors were used by SHELXD to find the marker atoms coordinates by integrating direct and Patterson methods. Once identified, the marker atom coordinates were refined and the initial set of phases improved by density modification as implemented in SHELXE.

If phase ambiguity still remained, the revised substructures were further refined and eventually completed by log-likelihood gradient maps as implemented in Phaser (McCoy, Grosse-Kunstleve et al. 2007). The resulting phases for both enantiomorphs were then subjected to density modification as implemented in DM (Cowtan and Main 1993). Density modification can significantly improve an electron-density map by incorporating features that are expected to appear in the map, such as flatness or disorder of the solvent region, the similarity of regions related by noncrystallographic symmetry and the similarity of the density-map histogram to histograms of deposited macromolecules(Skubak and Pannu 2011). These real space constraints based on known features of a protein density map are routinely applied to improve the approximate phasing obtained from the diffraction experiment.

6.3.4. Model building, refinement and validation

The initial experimental map was used for initial manual model building with Coot (Emsley, Lohkamp et al. 2010), and the resulting initial model was further improved by alternating cycles of manual model building in Coot associated with refinement in REFMAC (Murshudov, Vagin et al. 1997). The final model was refined by several rounds of restrained refinement in REFMAC with overall and isotropic B-factor for all non-hydrogen atoms. Finally, TLS refinement as implemented in REFMAC was used to account for possible different overall displacements of the different copies of the macromolecule in the asymmetric unit. The quality of the final structure was validated using MolProbity (Chen, Arendall et al. 2010) and PROCHECK (Laskowski, MacArthur et al. 1993)

CHAPTER 3

Results and Discussion

Chapter 3 Results and Discussion

1. MG491

1.1. Cloning and expression of MG491 variants

1.1.1. Cloning and expression of flMG491

After changing the triplet coding for the Trp232 (from tga into tgg), the sequence coding for the mutated version of full length MG491 protein, which consists of 348 amino acids and a computed MW of 39575 Da, was cloned in the laboratory of Dr. Enrique Querol and Dr.Jaume Piñol (Institut de Biotecnologia i de Biomedicina, UAB) by Luis Garcia Morales between NdeI and XhoI restriction sites of a pet21d expression vector (Novagen, Madison, WI, USA), which also codes for a C terminus hexahistidine (His6) tag. The resulting vector was transformed into Escherichia coli BL21(DE3) cells, and the transformant cells were plated on LB agar with ampicillin. Before continuing any further, the correct sequence was checked by DNA sequencing.

The sequenced plasmid was then grown in 1 L of LB medium containing 1 mL of ampicillin the cells induced overnight with 1mM of IPTG at 20 °C. Cells were harvested by centrifugation at 4500 g for 15 minute at 4 °C and the lysed by sonication. The total lysate extract was then centrifuged twice for 20 minutes at 45000 g to remove cells debris and filtered through a 0.22 µm filter.

The histagged MG491 gene product present in the resulting supernatant was firstly purified with a 5 mL HisTrap™ HP column previously equilibrated with a 0.02 M Tris-HCl pH 8, 0.500 M NaCl and 0.02 M Imidazole buffer, concentrated to a suitable volume and then loaded on a Superdex 200 column in a 0.050 M Tris-HCl pH 8, 0.500 M NaCl buffer with a flow rate according to manufacturer's guidelines for the column. The purified protein was collected superdex by SDS page (Figure 3-1).

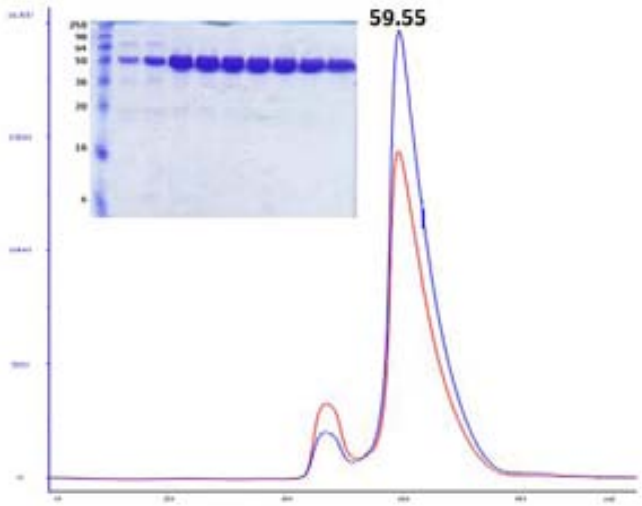

Figure 3-1. Purification of flMG491 by GFC and SDS-PAGE analysis of the eluted peak.
Chromatographic profile of flMG491 protein with a Superdex 200 16/60 GFC column with MG491 elution volume highlighted. 15% (v/v) SDS-PAGE analysis of the fractions corresponding to the GFC peak are also shown. Molecular weights of standard proteins are reported.

According to its elution volume and a previously determined calibration curve, obtained for the same column running standards of known size, MG491 full length elutes as a tetramer in solution. MG491 quaternary structure was also assessed by crosslinking with glutaraldehyde, which confirmed its propensity to form oligomers (Figure 3-2).

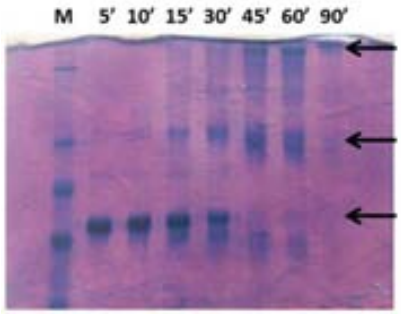

Figure 3-2.Glutharaldehyde chemically-induced crosslinking of flMG491 protein SDS-PAGE analysis.
7.5 % (v/v) SDS-PAGE analysis of flMG491 protein. The sample was equilibrated against a 25%v (v/v) glutaraldehyde solution for the time reported above each lane and then analysed by SDS-PAGE. flMG591 chemically-induced oligomeric states are indicated with black arrows.

The fractions of the peak corresponding to MG491 tetramer were pooled together and concentrated to 10 mg/ml using Amicon Ultra MWCO 30 KDa centrifugal filters (Sigma-Aldrich).

Small crystals appeared in three similar crystallization conditions at 20 °C (out of the hundreds conditions tested) after more than three months from drop depositions. For instance, the conditions were #75 (0.200 M lithium sulphate monohydrate, 25% w/v PEG 3350, 0.100 M BIS-TRIS pH 6.5), #76 (0.200 M lithium sulphate monohydrate, 25% w/v PEG 3350, 0.100 M HEPES pH 7.5) and #77 (0.200 M lithium sulphate monohydrate, 25% w/v PEG 3350, 0.100 M Tris-HCl pH 8.5), all from the Index crystallization screen (Hampton Research) (Figure 3-3). Crystals were diffracted at European Synchrotron Radiation Facility (ESRF) ID23.1 and ID29 beamlines, resulting in protein crystals with poor-quality diffraction.

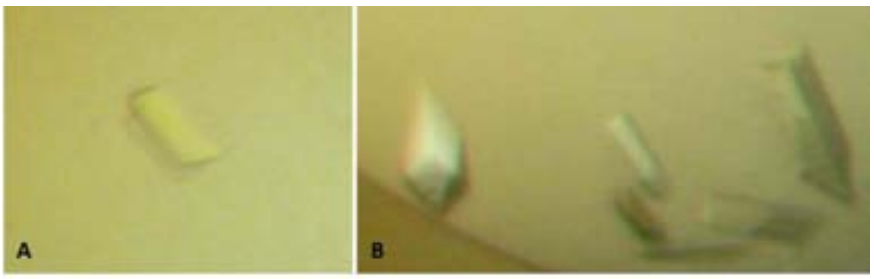

Figure 3-3. flMG491 crystals.
Crystals obtained in nanoliter-scale drops for flMG491 protein at A) 7mg/ml and B) 14mg/ml final concentration. Crystals grown in presence of 0.200 M lithium sulphate monohydrate, 25% w/v PEG 3350, 0.100 M HEPES pH 7.5.

A possible explanation for the long crystallization time needed could be that flMG491 was not completely stable in solution in these conditions, which resulted in its partial degradation to a smaller but more compact protein fragment, which could then be crystallized. To corroborate this hypothesis, analysis of protein crystal content by mass spectrometry (MS) was attempted, but failed due to the high content of PEG in the examined crystals, which resulted in tens of peaks in the mass spectra, making them too crowded to be interpreted.

At this stage, a new approach to structurally characterize MG491 protein was designed, involving the design of different constructs with a better propensity to crystallize than the full length protein.

1.1.2. Cloning of full length MG491 methionine mutant

Since a homologous structure was not available to solve MG491 structure by MR, the production of SeMet labelled protein was considered as an alternative approach to soaking with heavy atom solutions the native crystals eventually obtained. As already stated previously, an optimistic rule of thumb to derive initial phases estimates that at least one methionine residue per 80-100 amino acids is necessary. For that, and considering that wild type MG491 protein contains only three methionine residues (out of 346), including the N-terminal methionine which is usually disordered in the electron density map and not assumed to contribute to the anomalous scattering, four additional methionines were introduced.

Psipred (Jones 1999) v3.3, as implemented in the PSIPRED webserver (http://bioinf.cs.ucl.ac.uk/psipred/)(McGuffin, Bryson et al. 2000), was used to predict the secondary structure of flMG491 (Figure 3-4)

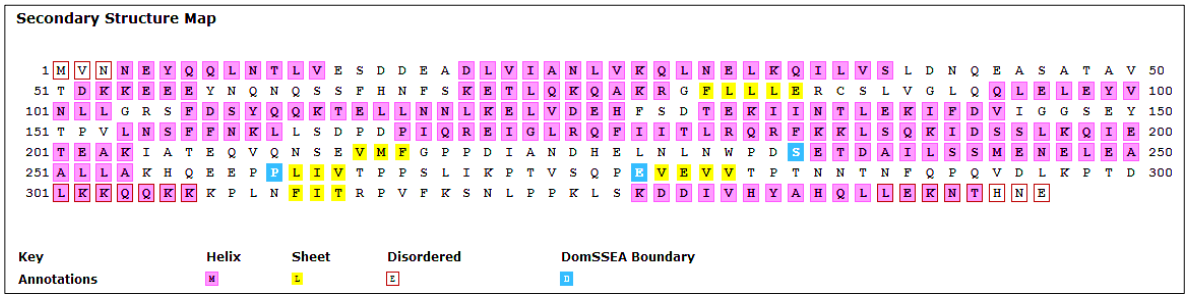

Figure 3-4. Psipred secondary structure prediction of flMG491 protein.

The secondary structure prediction was used to identify 4 isoleucine residues that were likely in well-ordered regions of the protein. Using the vector pet21d-MG491 as template and the oligonucleotide primers listed in Table 1, site-directed mutagenesis were employed to generate a variant with the residues Ile36, Ile168, Ile205 and Ile313 changed for methionines.

Mutation inserted •	Primer name	Primer sequence and direction
I36M	I36M_F ±	5' GAACTAAAGCAAAT***G***CTTGTTTCAC
	I36M_R †	3' CTTGATTTCGTTTA***C***GAACAAAGTG
I168M	I168M_F	5' TGATCCAAT***G***CAACGGG
	I168M_R	3' ACTAGGTTA***C***GTTGCCC
I205M	I205M_F	5' GGCTAAAAT***G***GCCACTG
	I205M_R	3' CCGATTTTA***C***CGGTGAC
I313M	I313M_F	5' CTTAACTTTAT***G***ACCCGTCC
	I313M_R	3' GAATTGAAATA***C***TGGGCAGG
• based on MG491 amino acidic sequence numbering		
± F forward primer	† R reverse primer	

1.1.3. Cloning and expression of MG491-Nt 1-255 and MG491-Nt 1-308 proteins

The secondary structure of flMG491 is predicted to be manly composed of α helices, with a more compact N terminal region and less structured thus more flexible C terminal.
This was in accordance with the results obtained with a limited proteolytic digestion of the full length protein when treated with different proteases. A first attempt was made with trypsin and chymotrypsin proteases, incubating 2µl of MG491 at 10 mg/ml with 2µl of the protease buffer 10x, 2µl of protease at different concentration (10µg/ml and 1 µg/ml) and 14µl of ddH_2O (double distilled water), for a total volume reaction of 20µl and a final MG491 concentration of 1mg/ml. After the addition of the proteases, the reactions were incubated at 20°C and aliquots of each reaction were removed at specific time points, the protease blocked by the addition of SDS loading buffer and the results of the proteolytic

cleavages analysed by SDS-PAGE. Among the proteases tested, the best result was obtained using trypsin. The bands (Figure 3-5) were further analysed by Edman sequencing and resulted to have the same N terminal residues as the untreated MG491 protein, suggesting a C terminal more accessible thus more susceptible to proteolysis.

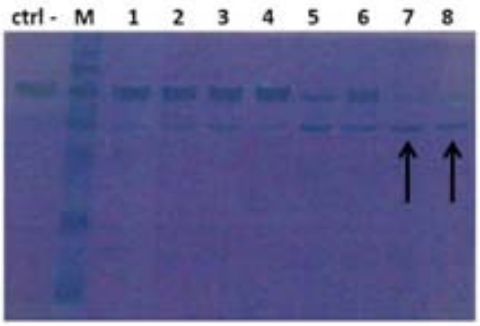

Figure 3-5. 15%(v/v) SDS-PAGE time course analysis flMG491 trypsin-limited proteolysis.
The bands indicated with a black arrow were further analysed by Edman sequencing. Ctrl - , uncleaved flMG491. M, Molecular-weight size markers. Lanes 1 and 2 are replicates of 30 minutes treatment, lanes 3 and 4 of 60 minutes, lanes 5 and 6 of 120 minutes, lanes 7 and 8 of 180 minutes.

Thus, based on the Edman sequencing and the secondary structure prediction computed with Psirepd (figure), two new smaller variants of MG491 were designed, spanning residues 1-255 and 1-308 and named, respectively, MG491-Nt_1-255 and MG491-Nt_1-308. Using as template MG491 gene with the and the primers specified in Table2, both fragments were cloned separately in pOPINE expression vector (Berrow, Alderton et al. 2007), which encodes for an extra lysine and a His6 tag at the C-terminal, which can easily be cleaved with carboxypeptidase A protease for improved crystallizability.

MG491 variant	Primer name	Primer sequence
MG491-Nt_1-255	MG491_F1	5' AGGAGATATACCATGGTTAATAATGAATATCAACAACTAAACACTTTAG
	MG491_R255	3' GTGATGGTGATGTTTTGCTAATAAAGCAGCTTCCAATTCATTTTC
MG491-Nt_1-308	MG491_F1	5' AGGAGATATACCATGGTTAATAATGAATATCAACAACTAAACACTTTAG
	MG491_R308	3' GTGATGGTGATGTTTTTTCTTCTGTTGTTTTTTCAAATCAGTAGG

The sequenced plasmids were then transformed into BL21(DE3) cells and expressed normally as the full length protein. The products of both constructs were purified following the protocol established for flMG491 protein, with comparable yields.

MG491-Nt_1-255 and MG491-Nt_1-308 proteins behaved as flMG491 on gel filtration chromatography (Figure 3-6), eluting as tetramers. The oligomerization state-of the two N-terminal variants were also analysed by nanoelectrospray ionization mass spectrometry (nanoESI-MS), which pointed out that both constructs were predominantly present as monomers, with low percentage of the proteins forming dimers and tetramers (data not shown).

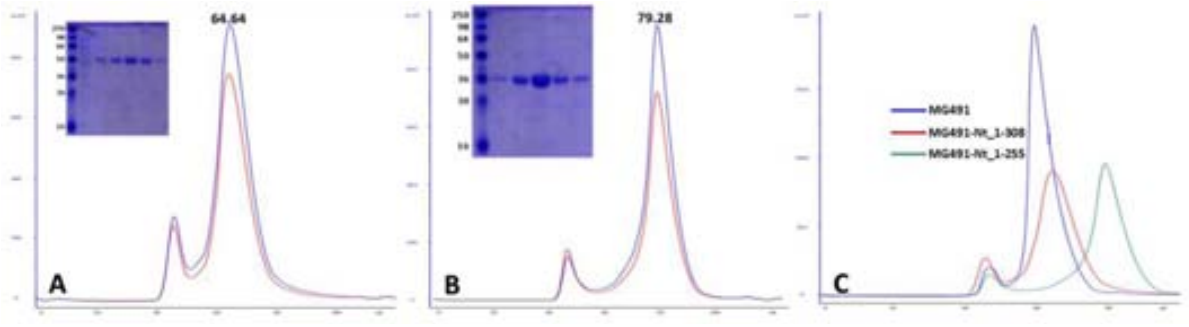

Figure 3-6. Purification of MG491 variants by GFC and SDS-PAGE analysis of the eluted peaks.
Chromatographic profile of A) MG491-Nt_1-308 and B) MG491-Nt_1-255 proteins with a Superdex 200 16/60 GFC column with corresponding elution volumes reported. 15% (v/v) SDS-PAGE analysis of the major GFC peak fractions are shown next to the corresponding chromatographic profile. Molecular weights of standard proteins are reported. C) Superposition of the GFC profiles of flMG491 (blue), MG491-Nt_1-308 (Bravman, Bronner et al.) and MG491-Nt_1-255 (Green, Ingram et al.).

1.1.3.1. Expression and purification of Semet MG491 N-terminal variants

The pOPINE vector containing the constructs MG491-Nt_1-255 and MG491-Nt_1-308 were transformed in BL21(DE3) or B834(DE3) cells and then expressed as explained in 6.2.3.2.

Both SeMet labelled proteins were purified following the same protocol established for flMG491, with comparable yields.

1.2. Crystallization of MG491-Nt_1-255 and MG491-Nt_1-308 variants

MG491-Nt_1-255 and MG491-Nt_1-308 native and SeMet labelled proteins were concentrated in 0.050 M Tris-HCl pH 8 and 0.150 M NaCl at about 8 and 15 mg/ml using Amicon Ultra MWCO 10 KDa centrifugal filters (Sigma-Aldrich). The sitting drop vapour diffusion method was used for preliminary crystallization screenings by mixing 100 µl protein solution with an equal volume of precipitants solution. Likewise MG491 full length protein, tiny crystals appeared at 20 °C for both constructs in conditions #75 (0.200 M lithium sulphate monohydrate, 25% w/v PEG 3350, 0.100 M BIS-TRIS pH 6.5) and #76 (0.200 M lithium sulphate monohydrate, 25% w/v PEG 3350, 0.100 M HEPES pH 7.5) from Index crystallization screen (Hampton Research) within 5-10 days (Figure 3-7).

Further optimization of these crystals in 24-well plates to obtain bigger crystals failed repeatedly, forcing to optimize these crystallization conditions in 96 well plates directly. Several optimization screenings were design to include sensible changes in pH (range screened from 5 to 8.5), percentage of PEG used (ranging from 15 to 32 %), type of PEGs employed (PEG 1000, PEG2000, PEG3350, PEG 4000, PEG8000 and PEG 10000), salt concentration and protein:precipitants drop ratio employed. Even if similar crystals appeared in some of the new screened conditions, no substantial improvements were made in terms of crystal's size or crystallization time. In addition, when setting up replicates of the same crystallization conditions, poor crystal reproducibility and crystal heterogeneity were observed for native and SeMet-labelled MG491-Nt_1-255 and MG491-Nt_1-308 proteins. However, crystals were soaked in a cryoprotectant solution containing 15 % (v/v) glycerol or 15 % (v/v) propylene glycol, and then mounted into a loop of appropriate dimensions straight from the drop , and then flash-frozen in liquid nitrogen (Figure 3-7, inset D).

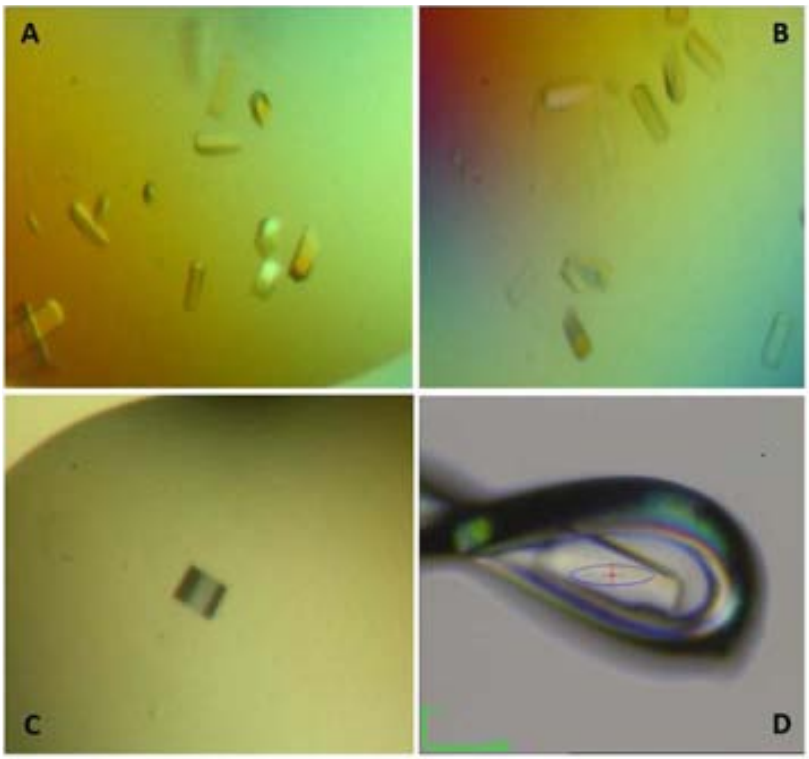

Figure 3-7. MG491-Nt_1-255 and MG491-Nt_1-308 crystals.
Nanoliter-scale drop crystals of A) MG491-Nt_1-308, B) SeMet MG491-Nt_1-308 and C) MG491-Nt_1-255 are shown. Crystals approximate dimension are 30x30x100 µm. D) SeMet MG491-Nt_1-308 crystal mounted into a 0.05 µm loop.

1.2.1. HA derivatization of MG491-Nt_1-255 and MG491-Nt_1-308 crystals

Most of the methods used nowadays to solve an unknown structure when a homologous model is not available involve the use of an anomalous scatters, either present endogenously or introduced in the crystals, by soaking with heavy atom solutions or by incorporating them in the protein sequence, prior to crystallization, using modified amino acids such as selenomethionine. The incorporation of heavy atoms into protein crystals for the purposes of phasing was pioneered in mid-50s, when mercury and silver compounds were used to determine the signs of most of the reflections recorded for a horse haemoglobin crystal (Green, Ingram et al. 1954).

Several heavy atom compounds were selected to derivatize native crystals of MG491-Nt_1-255 and MG491-Nt_1-308 by soaking procedure. Solutions of I3C (1 M), KBr (0.5 M), $HgCl_2$ (0.020 M), $(CH_3CO_2)_3Sm \cdot xH2O$ (0.010 M), K_2PtCl_4 (Winn, Ballard et al.)$_4$ (0.010 M) and K_2PtCl_6 (0.010 M) were used as stock solutions from which all the necessary dilutions were

prepared. To reduce non-specific binding and HAs weakly interacting the soaked crystals were then back soaked with the heavy atom free crystallization condition.

After the back soaking procedure, all the crystals were transferred to a drop of mother liquor containing 15 % (v/v) glycerol or 15 % (v/v) propylene glycol as cryoprotectants and rapidly frozen in liquid nitrogen.

1.3. Data collection

Fluorescence scan at the absorption edge of the investigated anomalous scatter, when reachable, was used to locate the optimal wavelength for a SAD data collection at peak wavelength (Figure 3-8). I3C derivatives, on the other hand, were collected at 1.90745 Å (6500 eV), wavelength at which iodine still exhibits an anomalous scattering effect of almost 10 electron units. By comparing the anomalous signal statistics of I3C soaked crystals with a native reference data set it was possible to identify whether or not iodine was present in the diffracted crystals. However, neither the soaked crystals nor the SeMet crystals were characterized by strong phasing signals, which made necessary to evaluate possible anomalous signal optimizations.

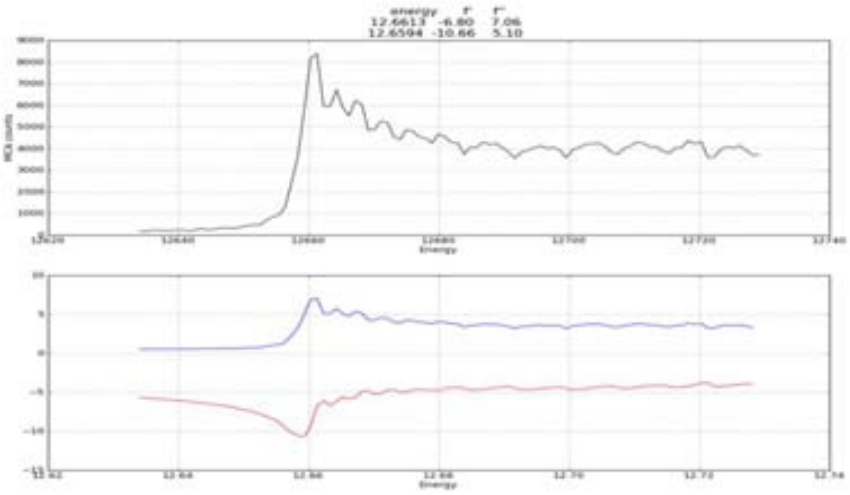

Figure 3-8. SeMet MG491-Nt_1-308 fluorescence scan. Experimentally measured X-ray fluorescence data plot (above) and the outputted f' and f'' spectra computed by CHOOCH (from ccp4 suite).

Firstly “inverse beam” approach was used: after a few degrees of data have been taken, the crystal was rotated 180 degrees and the same amount of data was collected. In this way, the set of images collected after the 180 degrees rotation will contain the anomalous mates of reflections on the first set. Nevertheless, all the data sets collected using the “inverse beam” data collection strategy, except for one I3C derivative which exhibited a slightly stronger anomalous signal, were still characterized by a poor phasing statistics.
Thus, further optimization of the anomalous signal by crystal reorientation was essayed. Mini-kappa goniometer as implemented at the ESRF beamline ID29 was used with an alignment protocol aimed to maximize the anomalous signal by collecting Bijvoet mates simultaneously, on the same image, by aligning the even-fold symmetry axis to the spindle axis (Brockhauser, Ravelli et al. 2013), resulting in an improved anomalous signal for the oriented crystals tested. Nonetheless, all the data sets collected with this additional optimization were still characterized by a quite weak anomalous signal.

A complete native flMG491 dataset was collected using the fully tunable ID29 beamline at ESRF. MG491-Nt_1-255 and MG491-Nt_1-308 native crystals were diffracted at the ESRF beamlines ID23eh1, ID23eh2 and ID29. HA soaked crystals were all diffracted at the ESRF beamlines ID29 and ID23eh1, except for I3C and samarium derivatives, which were diffracted at the beamline PROXIMA1 at SOLEIL.

1.4. Data analysis

Native crystals of flMG491, MG491-Nt_1-255 and MG491-Nt_1-308 were all orthorhombic crystals which could be indexed either as $P2_12_12$ or $P2_12_12_1$. To correctly assign the space group, ID29 mini-kappa goniometer head was used to re-orient the crystal so that one reciprocal cell axis was parallel to the spindle axis, resulting in the other two reciprocal axis moving into the diffraction condition. This alignment allowed to accurately measuring of the systematic absences reflections along each crystallographic axis to unambiguously index the crystals as belonging to the $P2_12_12$ orthorhombic space group. Likewise the native crystals, the use of the mini-kappa goniometer permitted the indexing of the SeMet

derivative crystals as belonging to the $P2_12_12$ orthorhombic space group. The best I3C derivative, which is a 50 mM I3C 60 seconds soak of a MG491-Nt_1-308 crystal, was directly indexed as belonging to the $P2_12_12$ space group. Interestingly, unit cell parameters of the different datasets collected differ "sensibly", being in the range of a=96-98 Å, b=107-112 Å and c=62-70 Å, suggesting an extensive non-isomorphism both between different native crystals and between native and heavy atom soaked crystals.

After a final scaling with SCALA, the data sets were fed into ANISOSCALE server (http://services.mbi.ucla.edu/anisoscale/) (Strong, Sawaya et al. 2006) to diagnose, if any, the degree of anisotropy, and then to correct it by ellipsoidal truncation and anisotropic scaling. Matthews probability analysis (Kantardjieff and Rupp 2003) suggested the presence of two molecules in the asymmetric unit, which gives a solvent content volume in the range of 49 to 60 %. The presence of two molecules in the asymmetric unit raised the possibility of NCS. MOLREP (Vagin and Teplyakov 1999, Vagin and Teplyakov 2000, Vagin and Isupov 2001) was used to calculate Patterson self-rotation function to check for rotational NCS symmetry. No peaks other than the ones arising from crystallographic symmetry were observed for any of the data sets, which could be explained if the NCS axis is parallel or nearly parallel to a crystallographic symmetry axes Then FFT from the CCP4 suite was used to calculate native Patterson map. The peak analysis performed by peakmax revealed the presence of a strong (37%) off-origin peak (fractional coordinates 0.0765, 0.5, 0.5) for the SeMet MG491-Nt_1-308, confirming the presence of tNCS (translational NCS).
The same peak was detected also for the other datasets, in all cases with a relative height lower than MOLREP threshold for NCS detection, resulting in possible undetected off origin peaks.

1.5. Substructure determination

1.5.1. Initial phasing for multi-crystal datasets

Recently Liu and co-workers have demonstrated how, by combining SAD data from multiple crystals with low anomalous signal, it was possible to increase data multiplicity without adding radiation damage, resulting in a merged dataset with better anomalous statistics. This multi crystal averaging approach allowed them to solve a large and poorly diffracting SeMet protein structure (Liu, Zhang et al. 2011) or several native proteins at modest resolution using only intrinsic anomalous scattering sulfur atoms (Liu, Dahmane et al. 2012). Thus, having collected a high number of presumably identical datasets for SeMet MG491-Nt_1-308, phasing with multi-crystal averaging was attempted.

All the six SeMet MG491-Nt_1-308 single crystal data sets were processed individually with XDS. After integration, XDS CORRECT module was used for initial scaling. By POINTLESS (Evans 2006) COPY function the XDS data sets in ASCII format were combined to the mtz reflection file format. The combined data sets were then merged and scaled with SCALA (Evans 2006). Intensities scaled and merged by SCALA were then converted to amplitude by CTRUNCATE (Evans 2011). The resulting mtz files were then used for subsequent analysis and calculations.

Up to six different SeMet MG491-Nt_1-308 complete datasets were merged together, and averaged multi-crystal statistics compared. Even if the anomalous statistics looked better for the averaged multi-crystals than for the single datasets, SHELXD was unable to determine the complete heavy atom substructure. Only few sites with low occupancy were identified, and neither recycling the heavy atoms positions with SHELXE nor using a log-likelihood-gradient (LLG) map for substructure model completion (McCoy and Read 2010) as implemented in Phaser were able to identify any possible additional site. Even if the phasing statistics appeared improved when employing multi crystal averaging approach, the resulting electron density map was of very low quality.

1.5.2. Initial phasing for single-crystal dataset

All anomalous scatters containing single crystal data sets were processed individually by XDS as datasets containing anomalous signal, and then converted to the mtz reflection file format by the XDSCONV conversion program included in the XDS program package (Kabsch 2010). Anomalous difference of structure factor amplitudes and its standard deviation column in the outputted reflection file were obtained choosing CCP4 as format for the output file in the file parameters XDSCONV.INP. The resulting mtz files were then separately merged and scaled with SCALA before converting intensities to amplitudes with CTRUNCATE. SCALA statistics were used to assess the single datasets with stronger anomalous signals and better statistics. Best representative crystals of successful derivatives and their statistics are reported in Table IV.

	Native MG491	**SeMet MG491-Nt_1-255**	**SeMet MG491-Nt_1-308**	**I3C MG491-Nt_1-308**
Data Collection				
Wavelength (Å)	0.979	0.979	0.979	1.907
Resolution (Å)	**35 -3.30 (3.40 - 3.30)**	**35 - 3.10 (3.20 - 3.10)**	**35 - 3.00 (3.10 - 3.00)**	**35 - 3.35 (3.45 - 3.35)**
Unit cell parameters	a = 99.27 Å b = 108.73 Å c = 66.62 Å α = β = γ = 90°	a = 98.86 Å b = 107.99 Å c = 65.63 Å α = β = γ = 90°	a = 98.26 Å b = 108.42 Å c = 62.19 Å α = β = γ = 90°	a = 97.56 Å b = 111.79 Å c = 70.98 Å α = β = γ = 90°
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
I/σ(I)	16.70 (2.63)	20.93 (2.72)	14.21 (2.25)	19.60 (6.20)
Rrim	0.0435 (0.4376)	0.0915 (0.6544)	0.1599 (0.8278)	0.1429 (0.3706)
Rpim	0.0205 (0.1992)	0.0250 (0.147)	0.0365 (0.1835)	0.0285 (0.1004)
Completeness (%)	99.3 (93.0)	99.1 (98.4)	99.4 (95.8)	99.3 (96.4)
Multiplicity	4.2 (4.1)	13.2 (13.4)	19.1 (19.1)	15.1 (7.0)
Phasing Statistics				
d''/σ(d'')		**1.39 (0.80)**	**1.15 (0.83)**	**1.88 (1.46)**
Anomalous completeness		99.3 (98.3)	99.2 (94.7)	99.2 (96.2)
Anomalous multiplicity		7.1 (7.1)	10.2 (10.0)	7.9 (3.7)

Values in parantheses are for the highest resolution shell.

Phasing was attempted for each one of the three best derivative crystals obtained, namely SeMet MG491-Nt_1-255, SeMet MG491-Nt_1-308 and I3C MG491-Nt_1-308.

Substructure determination with SHELX triad was first attempted with I3C soaked MG491-Nt_1-308 crystal, which was the dataset with the strongest anomalous signal among the

three. Additional keyword MAGI in SHELXD was used to assign a higher figure of merit to those solutions which contained heavy atom sites within a certain range, 6 Å for iodine atoms in an I3C molecule. The best solution was further analysed with Professs, from CCP4 suite (Winn, Ballard et al. 2011), to check if the atoms identified were within the desired distance range. The remaining iodine sites were then used to calculate an initial set of phases which were then subjected to density modification as implemented in SHELXE.
Even though after SHELXE density modification cycles the two hands exhibited different contrast, the electron density map obtained using the (supposedly) correct hand was of very poor quality, making impossible any kind of interpretation. Optimization of the most crucial parameters in substructure determination in SHELXD, such as the number of heavy atom sites to be identified and the high resolution cut off to be used did not lead to any sensible improvement regarding the quality of the final electron density map.

In the case of SeMet MG491-Nt_1-308 dataset SHELXD identified only four SeMet sites out of ten expected, with a clear drop in occupancy in between the fourth and fifth site (Figure 3-9). Phaser was then used to iteratively compute SAD LLG maps to search for possible additional sites, without finding any, suggesting that the real substructure could indeed be formed by only four SeMet sites.

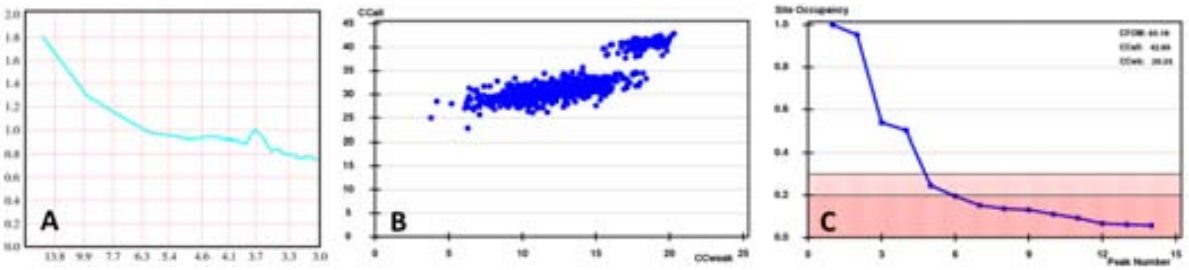

Figure 3-9. SeMet MG491-Nt_1-308 plots.
A) <d''/sig> plot showing anomalous signal profile against resolution. B) CCall/CCweak plot, differentiating wrong (broad cluster) from correct (compact cluster) solutions. C) Site occupancy plot, with a drop in between the 4th and the 5th sites. Xprep was used to compute the plot reported in A), while B) and C) graphics were outputted by SHELXD.

Since either hand of the identified substructure is a possible solution, SHELXE was then used to calculate initial phases for both enantiomorph. Initial electron density maps were subjected to 60 cycles of density modification to break down the substructure enantiomorph ambiguity. However, the electron density maps based on original and inverted sites had similar contrast, which made necessary to visually examine both maps directly with Coot (Emsley, Lohkamp et al. 2010). As clearly visible in Figure 3-9, the map based on the original hand was characterized by a better contrast, higher connectivity and by discernable features such as the presence of several α helices, in agreement with Psipred secondary structure prediction.

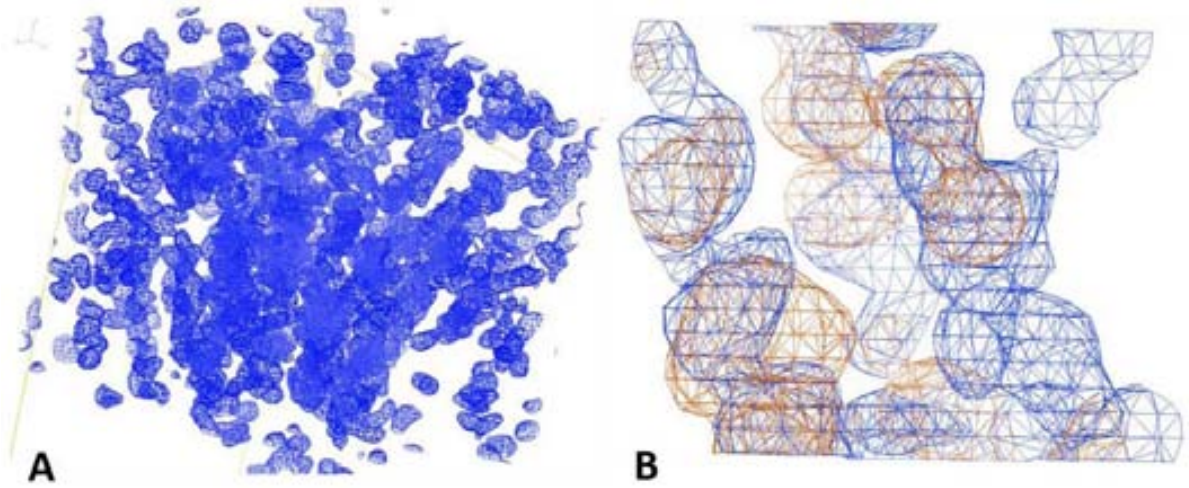

Figure 3-10. SeMet MG491-Nt_1-308 initial electron density.
A) Overall electron density map calculated using the four SeMet sites identified by SHLEXD. B) Comparison between the electron density map obtained using the original hand of the identified substructure, which resulted to be the correct one (in blue) and the map obtained using the inverted substructure (in orange). Electron density maps are countered at 1σ.

1.6. Structure determination of MG491-Nt_1-308 protein

Considering that automated model building procedures using software like Buccaneer (Cowtan 2006, Cowtan 2008) or SHELXE auto tracing algorithm (Sheldrick 2010) resulted in incomplete and very fragmented models, initial model building was done by manually positioning twelve polyAla α helices in Coot. These helices appeared to be organized into a dimer, with the two putative monomers which are related to each other, as well as the four SeMet sites, by a 2-fold NCS, as shown in Figure. Each monomer seemed to be formed by two compact terminal domains, both composed of three helices.

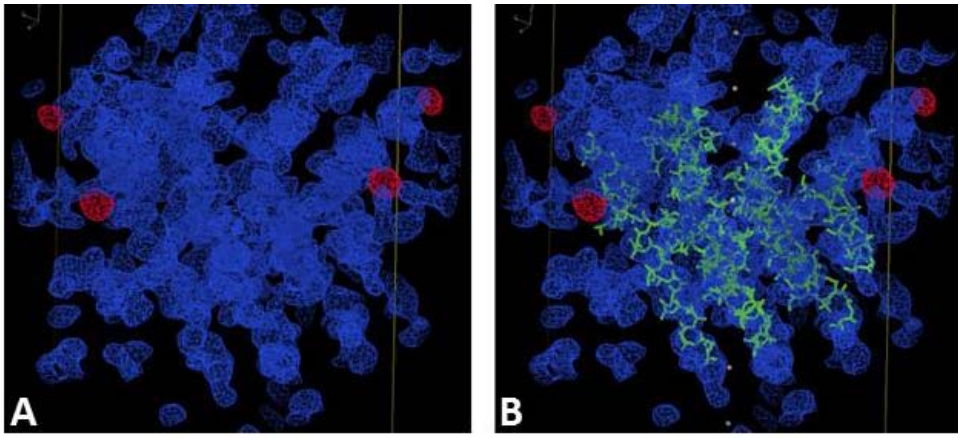

Figure 3-11. SeMet MG491-Nt_1-308 initial manual model building.
A) SeMet MG491-Nt_1-308 electron density map (in blue) countered at 1σ with anomalous map (in red) countered at 3.5 σ. B) Same as A), but with the first helices (in green) manually placed with Coot. Identified 2-fold NCS is shown in grey.

Density modification with solvent flattening, histogram matching and NCS averaging was used to improve the initial electron density map and help in the construction of the linker which should connect the N-terminal domain to the C-terminal. A mask covering one monomer was compute in P1 space group using ncsmask program from CCP4 suite and then used as input mask for NCS averaging in DM. The needed rotation matrix and translation vector (also called NCS operator) relating the two sets of atom were obtained by aligning the Cα atoms of the two molecules using LSQKAB (Kabsch 1976) from the CCP4 suite. These are the operations which map the density covered by the mask onto the other equivalent region, or rather the other monomer (Figure 3-12).

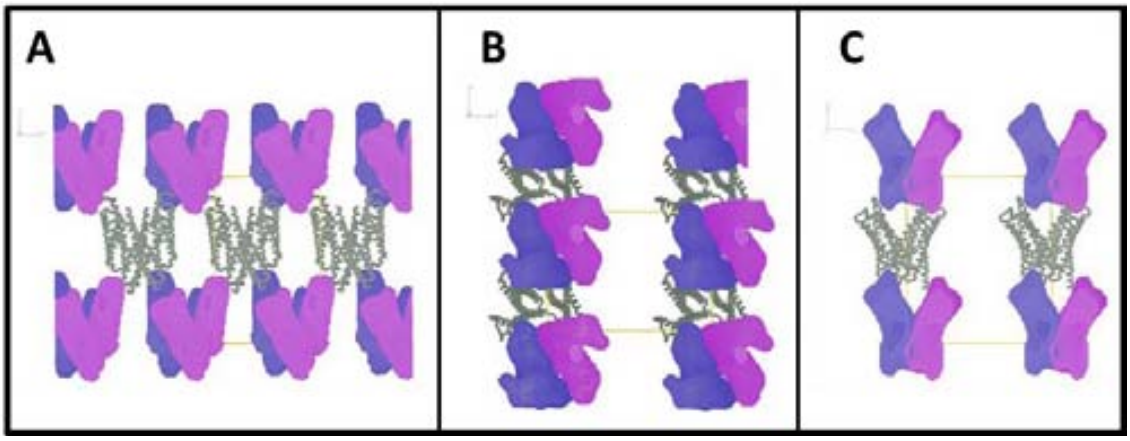

Figure 3-12. SeMet MG491-Nt_1-308 monomer masks used for initial NCS averaging with DM.
The masks covering the two putative SeMet MG491-Nt_1-308 monomers, which are related by a 2-fold NCS, are shown in purple and dark purple and bright pink. The crystal packing is shown as a projection onto the xy plane (A), xz plane (B) and yz plane (C),respectively.

Visual inspection of the outputted electron density map together with the NCS mask used for the averaging were routinely done after each DM run, with manual adjustments of the boundaries of the mask to cover all the density supposedly belonging to the monomer. Even though DM runs converged properly and gave high correlations, the electron density maps obtained after different runs were characterized by a general lack of details, with almost no clear density along the helices. Moreover, density corresponding to the loops connecting the helices within the N-terminal (or C-terminal) were in some cases very confusing, with no evidence indicating the presence of obvious connections between the two three-helix-bundles. For the SeMet MG491-Nt_1-308 dataset, as already stated, Matthews coefficient analysis suggested the presence of two molecules in the asymmetric unit, for a total of 632 residues, with an average solvent content of around 50%. Considering that the twelve helices used as initial model were composed of almost 600 residues in total, the possible loops should have been relatively shorts in length, but the spatial rearrangement of the four three-helix-bundles within the crystal required a minimum length of around 20 Å for a loop to connect two three-helix-bundles.

Since it was not possible to build the loops using SeMet MG491-Nt_1-308 dataset, one monomer of this partial model was used as search model for MR against native, SeMet MG491-Nt_1-255 and I3C MG491-Nt_1-308 datasets using Phaser via the CCP4i GUI, to check for possible differences among the crystals that could help in the building of the

missing loops. In all the cases Phaser was able to find and correctly place the two expected molecules within the asymmetric unit, with Z-scores higher than 8, which imply a definitive solution. However, neither of the resulting electron density maps was of any help for the loops construction. Even though the length of the constructs supposedly crystallized vary sensibly (ranging from 265 residues for the histagged MG491-Nt_1-255 monomer to 354 amino acids for the histagged flMG491 monomer), the unit cell parameters of the different crystals were almost the same, and a visual inspection of their electron density maps after MR revealed they were highly similar, with no remarkable differences (Figure 3-13).

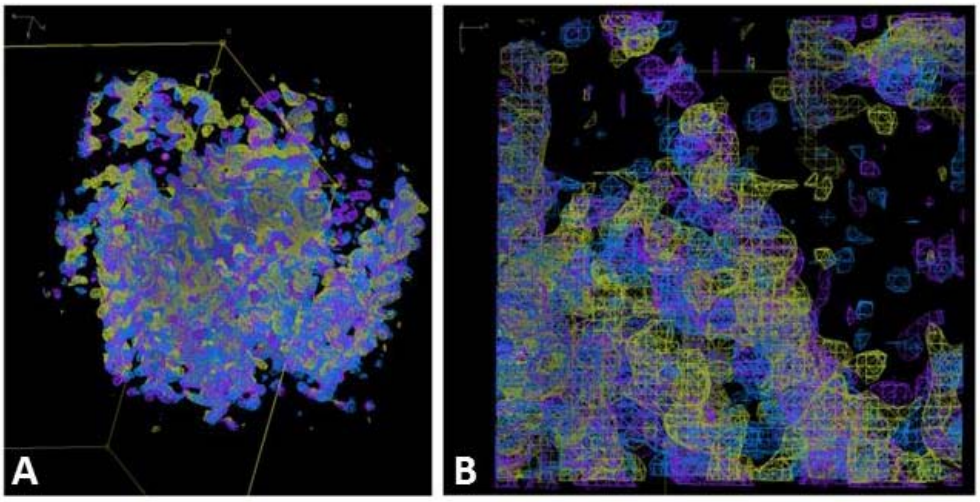

Figure 3-13. Comparison of SeMet MG491-Nt_1-255, SeMet MG491-Nt_1-308 and I3C MG491-Nt_1-308 electron density maps.
A) Maps of SeMet MG491-Nt_1-255 (in cyan), SeMet MG491-Nt_1-308 (in yellow) and I3C MG491-Nt_1-308 (in purple), countered at 1.5σ with common origin imposed. Maps for SeMet MG491-Nt_1-255 and I3C MG491-Nt_1-308 were obtained by MR using as search model the initial 12 helices model manually built for SeMet MG491-Nt_1-308. B) Detailed comparison of the electron density maps of the three datasets. Even though the overall densities were very similar, some regions of the maps didn't match perfectly.

DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/) (Holm and Rosenstrom 2010) was then used to check whether the identified three helices forming the terminal domains resemble any known folding. The server identified the first N-terminal spectrin repeat of human dystrophin as a plausible structural homolog, thus a 3D model based on its crystallographic structure (PDB ID: 3UUN, chain A) was used as template to start building the loops connecting the three α helices clusters.

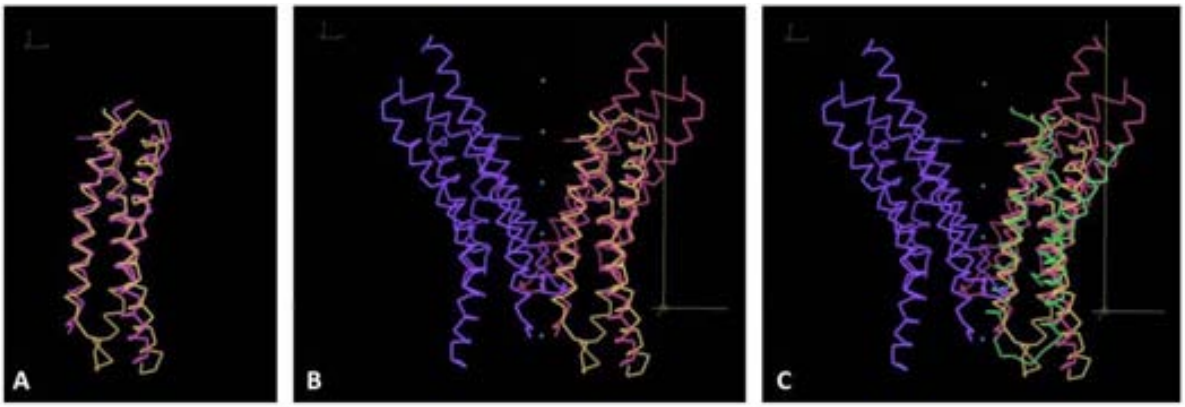

Figure 3-14. Human dystrophin N-terminal spectrin repeat (PDB ID:3UUN, chain A) superposition onto SeMet MG491-Nt-1-308 partial model.
A) Cα representation of 3UUN (in yellow) superposed onto one of the four three-helix-bundles (in pink) manually built of MG491-Nt-1-308. B) As A) but with the 2-fold NCS axys and the remaining three-helix-bundles covering the asymmetric unit shown. C) In green in shown the other three-helix-bundle forming the putative monomer superposed into the structure shown in A)

3UUN model superposed perfectly with one cluster of three helices, as shown in Figure Y, inset A, with a root-mean-square-deviation (RMSD) of 2.39 Å. More surprisingly, an even lower RMSD was obtained superposing 3UUN on the other three helices forming the supposed monomer (2.32 Å). A further analysis of flMG491 sequence didn't reveal any possible genetic duplication that could explain the identical folding for the two putative terminal domains forming what we thought were the monomer. This, in addition to the fact that the different crystallized constructs all exhibited a very comparable electron density and that in neither of the datasets was possible to place any amino acid of the expected loop connecting the two terminal domains, raised the possibility that instead of a dimer formed by two monomers of six helices each, the crystallized constructs were instead organized as a tetramer or as a dimer of dimer. This possibility was also in agreement by MS and GFC results, with which MG491 was detected as a tetramer. The presence of four molecules within the asymmetric unit also implied that this tetramer (or dimer of dimers) could form only if the monomer length did not exceed 210 residues in length, as pointed out by Matthews analysis coefficients (Table).

Monomer length	Nmol/asym	Matthews Coeff	%solvent	P(tot)
150	4	2.45	49.91	0.62
160	4	2.30	46.57	0.70
170	4	2.17	43.23	0.64
180	4	2.04	39.89	0.47
190	4	1.94	36.55	0.25
200	4	1.84	33.21	0.08
210	4	1.75	29.87	0.02

Then DM was used to infer the solvent content that gave the best electron density map after density modification, which resulted to be around 36 %, considering the asymmetric unit composed by four monomers of 190 residues each. With the assumption of having the same crystallized fragment in all the collected datasets, the solvent content for the I3C derivative was a bit higher, around 46%, due to the slightly bigger unit cell of this datasets. Even though all the collected datasets belonged to the same space group, the differences in term of unit cell parameters between SeMet MG491-Nt_1-308 and I3C MG491-Nt_1-308 were exploited for multi-crystal averaging with DM. To facilitate the comparison between the two electron density maps after each run of DM, a common origin was set using SeMet MG491-Nt_1-308 origin as reference. Using phenix.get_cc_mtz_mtz, a command-line utility from Phenix suite (McCoy, Grosse-Kunstleve et al. 2007), which uses RESOLVE (Terwilliger 2000) to find the origin shift compatible with space-group symmetry which has to be applied to the second map to maximize the correlation of the two maps.

Density modification often suffers from an overestimation of the quality of the phases, clearly evincible from very high figures of merit (FOM). Since the FOMs are based on an estimate of the error in the phases, a high FOM implies that the phases are correct and therefore should not be modified, FOMs have been divided by a factor 3 before each run of multi-crystal averaging, using the program SFTOOLS from CCP4 suite on command line.
By combining and alternating manual model rebuilding and refinement in Coot, model refinement in REFMAC including TLS and restrained refinement and the improved phases obtained after each run of DM it was possible to sensibly enhance the quality of SeMet MG491-Nt_1-308 density map, as shown in Figure 3-15, insets A and B.

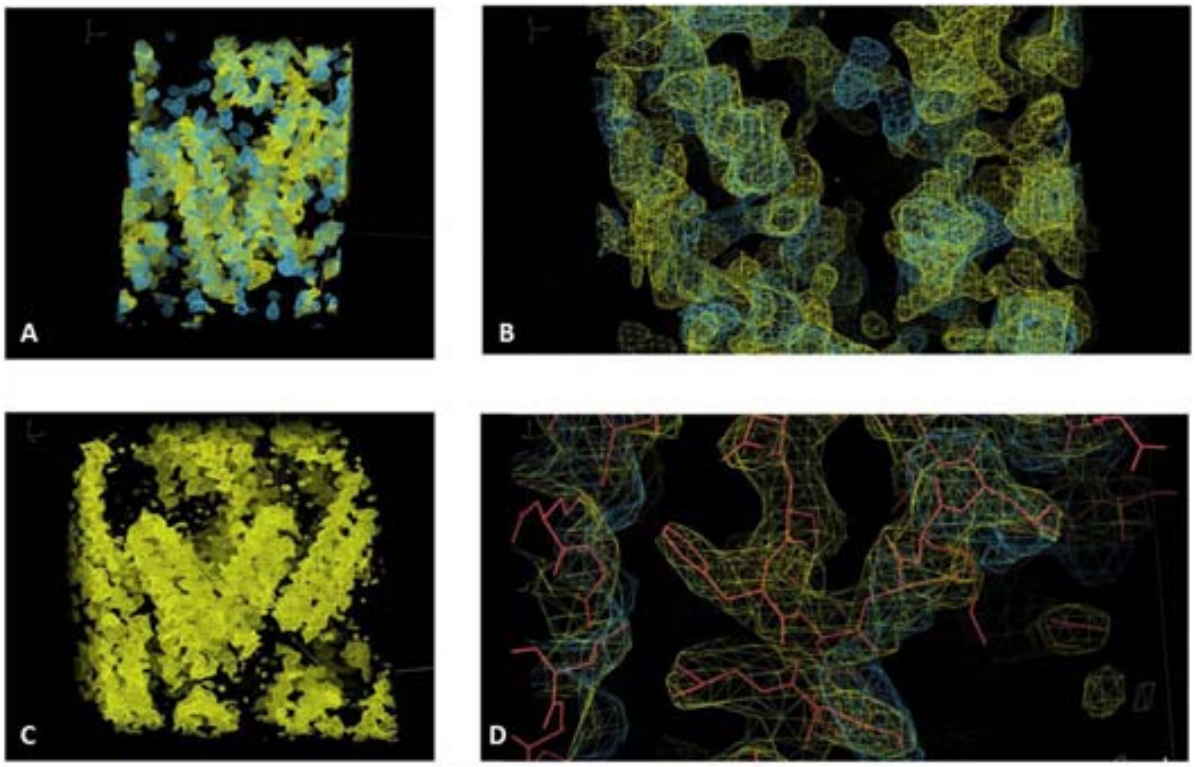

Figure 3-15. SeMet MG491_Nt_1-308 electron density map improvements after DM.
A) Electron density map of SeMet MG491-Nt_1-308 obtained after initial SHELXE density modification (in cyan, countered at 1.5σ) and after several runs of DM (in yellow, countered at 1.5σ). B) Detailed view of the achieved phase improvements after applying DM. C) Final overall electron density map of SeMet MG491-Nt_1-308, countered at 1.0σ. D) Detailed view of the final built model, with the initial (cyan) and final (yellow) electron density maps also represented.

By using MG491 protein sequence between residues Ser63 and Ile205, a model of the molecules that satisfactorily accounts for all the electron density in the map was built. SeMet MG491-Nt_1-308 dataset was used for final refinement at 3 Å in REFMAC with the final model; with the statistics reported in the following Table.

	SeMet MG491-Nt_1-308
Resolution (Å)	50 - 3.00
Unit cell parameters	a = 98.27 Å b = 108.42 Å c = 62.19 Å α = β = γ = 90°
Space group	$P2_12_12$
Reflections (work/test)	13851 / 694
R factor (work/test) (%)	22.58 / 24.91
Number of:	
chains	4
protein atoms	4588
Average B-factor (Å²)	
chain A	41.9
chain B	43
chain C	42.4
chain D	39.9
RMSD bond (Å)	0.0119
RMSD angle (°)	1.4051
Ramachandran plot *	
favoured regions (%)	93.0
allowed regions (%)	98.0
Molprobity score (overall percentile)*	2.42 (96)

* calculated using MOLPROBITY server

1.7. Overall structure

DSSP analysis (Kabsch and Sander 1983, Joosten, te Beek et al. 2011) was used to standardize secondary structure assignment. Each monomer folds into an antiparallel three-helix-bundle, composed by a short N terminal disordered region which comprehends residues 63 to residues 69, followed by the first (α1) of the three helices. The helices α1 and α2 are connected by a short loop of 6 amino acids, L1 (Figure inset B), from Arg105 to Tyr110, while α2 and α3 helices are connected by a longer loop, L2 (Figure inset C), from residues Gly146 to Asp166. L2 loop is characterized by the presence of a 3_{10}-helix turn between Lys160 and Ser163, just before the start of the third helix L3 (fig inset A).

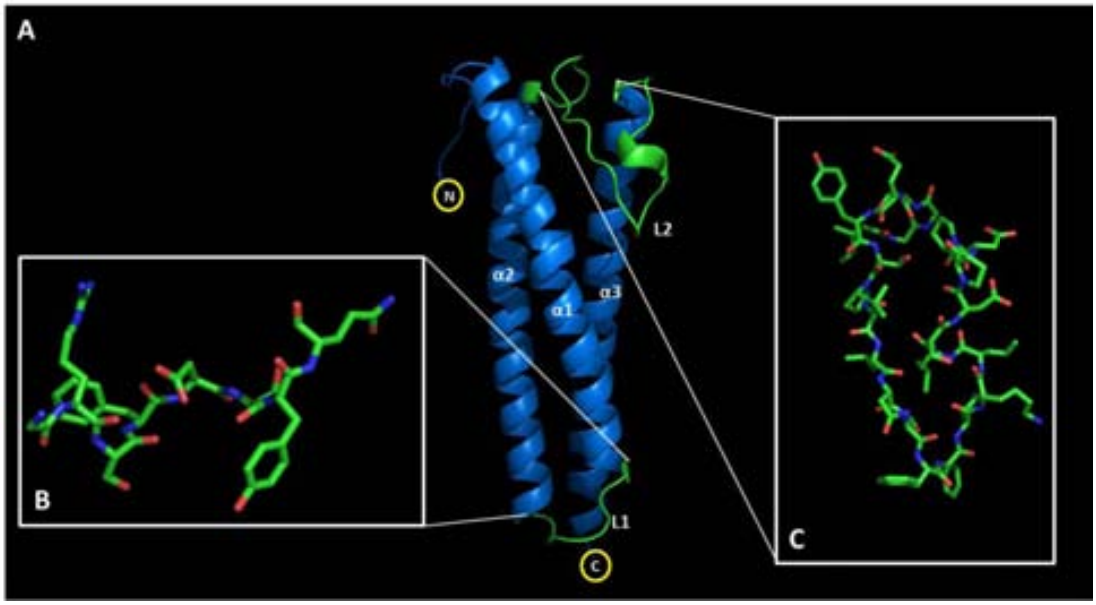

Figure 3-16. MG491-63-205 monomer structure.
A) Ribbon representation of MG491-63-205 monomer, composed by a short and flexible N-terminal, followed by three αhelices, α1, α2 and α3. The short loop L1 (inset B) connects α1 to α2, while the longer loop L2 (inset C) connects α2 to α3.

The final structure includes atomic coordinates for 555 residues in total, from Asn67 to Ala203 for the 4 unique chains present in the asymmetric unit, with additional residues 63-66 and 204-205 only built for D monomer. Monomers A and D, as B and C, form a tight dimer (Fig inset B) in which the two forming subunits are related by a 72° rotation. In turn, these tight dimers are related one to each other by a NCS 2-fold which axis is nearly perfectly parallel to y-axis (fig inset C).

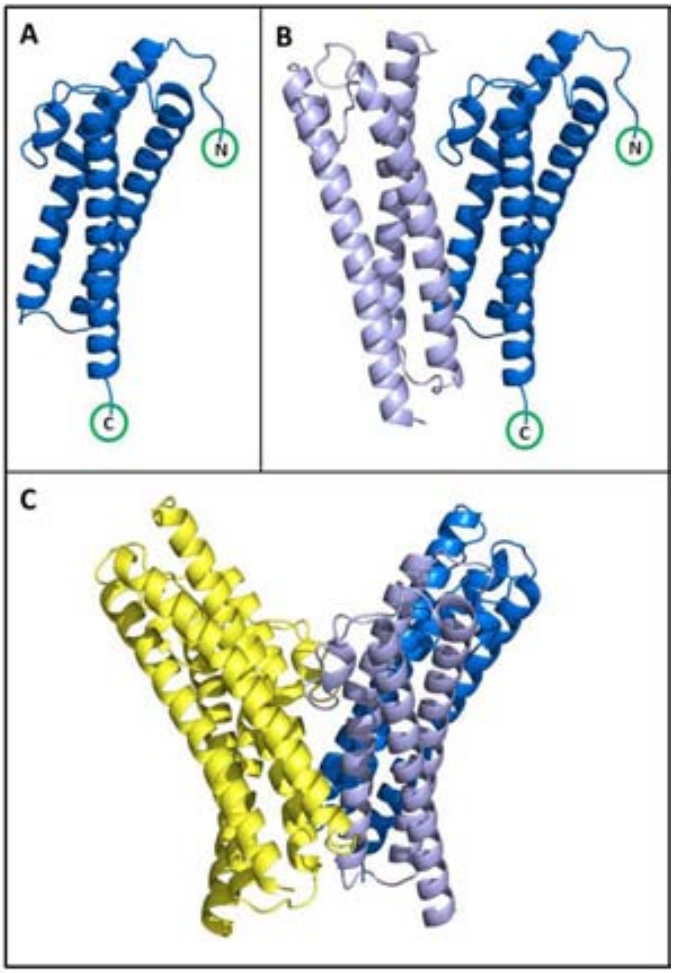

Figure 3-17. MG491-63-205 tetramer organization.
A) Ribbon representation of MG491-63-205 monomer, with the N and C terminal indicated with a green circle. B) Ribbon representation of the tight dimer formed by two MG491-63-205 monomers (in blue and in grey) which are related by a 72° rotation. C) Ribbon representation of MG491-63-205 tetramer. The two tight dimers represented are related by a rotation of 180°, which axis is nearly perfectly parallel to the y-axis. The overall organization of the monomers to form the dimer of dimers places the N- and C-terminals of each subunit at opposite faces of the tetramer.

PISA (Protein Interfaces, Surfaces and Assemblies) (Krissinel and Henrick 2007) web server was used to compute the interface areas between the two subunits forming a tight dimer (A and D or B and C) and between two subunits belonging to a loose dimer (A and B or C and D), with the average interface area between the monomers forming the tight dimers which is approximately the double of the average interface area between the other two monomer pairs (1183 $Å^2$ and 530 $Å^2$, respectively). An obvious consequence of this difference is that each monomer interacts with its two neighbour subunits in a very uncommon asymmetric way, through two different protein-protein interfaces.

The secondary structure element more affected by this asymmetric interaction is the loop L2, in which F157 and F158 residues adopt two different conformations, depending on the monomer to which they are exposed. In the tight dimer formed by the subunits A and D (considering the tight dimer BC would lead to the same observations), F157 of the monomer D is stabilized by an interaction with Gly80 of the subunit A. On the other hand, if considering the interface in between the two subunits forming the loose dimer (for

example the pair AB), F157 of subunit A is interacting with Gly91 of monomer B. Most likely to maintain this stabilizing interaction, loop's backbone between F157 and F158 undergoes a displacement of almost 2Å when this interaction takes place at the loose dimer interface.

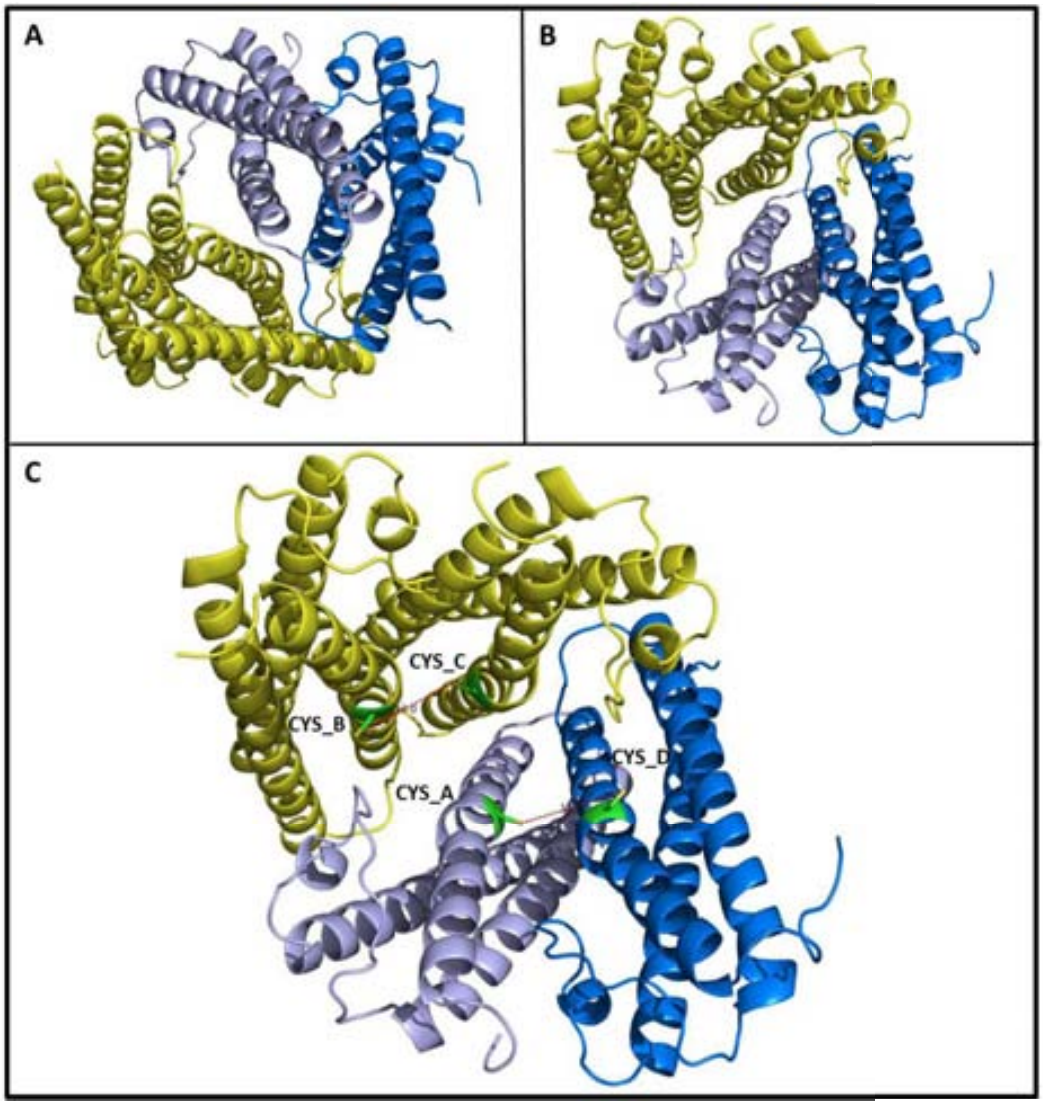

Figure 3-18. MG491-63-205 tetramer views.
Bottom (A) and top (B) view of MG491-63-205 tetramer. In both view, the 2-fold NCS is clearly evincible, with its axis perpendicular to the plane. C) MG491_63-205 tetramer top view shows Cys87 side chains proximity from two neighboring subunits. The very similar distances between the cysteines make difficult to speculate which could be the residues involved in an eventual disulphide bond.

The analysis of the structure also revealed that Cys87 of each monomer (the only cysteine of MG491 protein) is very close to the other cysteines, suggesting that the formation of a disulfide bond would be possible. Even though the solved structure did not possess this SS-bond, the possibility of a SS-bond formation that could somehow stabilize MG491

quaternary structure was assessed, following an established protocol normally used to achieve disulfide bond formation in peptides (Chen, Annis et al. 2001). Briefly, the protein was diluted with 0.01 M PBS pH 7.5 to a final concentration of 1mg/ml and then let reacting (between 2 and 4 hours) with 1% (v/v) DMSO. Then the results of the reaction were analysed by running a SDS-PAGE, which showed a partial conversion of MG491-Nt_1-255 monomer to a dimer (Figure 3-19, similar results for MG491-Nt_1-308 are not shown).

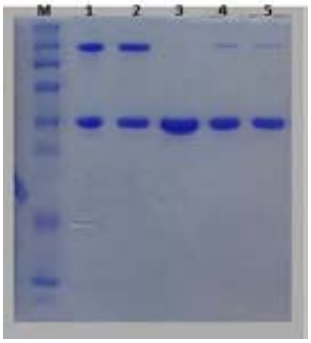

Figure 3-19. 15% (v/v) SDS-PAGE analysis of disulfide bond formation in MG491_Nt-1-255 in presence of DMSO.
Lane 1, sample treated with 1% (v/v) DMSO, without β-mercaptoethanol (2-ME), sample not heated. Lane 2 as lane 1 but heating the sample for 10 minutes at 80 °C before loading the gel. Lane 3, sample treated with 1% (v/v) DMSO and then loaded in presence of 2-ME. Lane 4, sample treated with 1% (v/v) DMSO and then loaded into the gel in presence of DTT. Lane 5 as lane 4, but heating the sample for 10 minutes at 80 °C before loading the gel.

2. MG219

MG219 is a 148 residues protein with a computed MW of 16617 Da characterized by a very high isoelectric point (pI) of 10.14, due to its content of lysine residues (30 out of 148), mainly located at its C terminal region. Psipred secondary structure prediction (Figure) computed an N terminal slightly ordered and a C terminal highly flexible and disordered, mainly because of the several lysine residues, known to be disorder promoting residues.

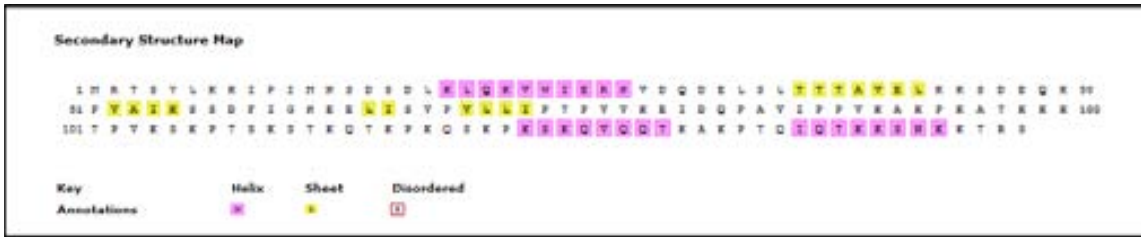

Figure 3-20. Psipred secondary structure prediction of flMG219 protein.

2.1. Cloning and expression of flMG219

MG219 protein was was cloned in the laboratory of Dr. Enrique Querol and Dr.Jaume Piñol (Institut de Biotecnologia i de Biomedicina, UAB) by Luis Garcia Gonzalez between NdeI and XhoI restriction sites of a pet21a expression vector (Novagen, Madison, WI, USA), which also codes for a C terminus hexahistidine (His6) tag. The resulting vector was transformed into Escherichia coli BL21(DE3) cells, and the transformant cells were plated on LB agar with ampicillin. Before continuing any further, the correct sequence was checked by DNA sequencing.

The sequenced plasmid was then grown in 1 L of LB medium containing 1 mL of ampicillin the cells induced for 4 hours with 1mM of IPTG at 37 °C. Cells were harvested by centrifugation at 4500 g for 15 minute at 4 °C and the lysed by sonication. The total lysate extract was then centrifuged twice for 20 minutes at 45000 g to remove cells debris and filtered through a 0.22 µm filter.

The histagged MG219 gene product present in the resulting supernatant was firstly purified with a 5 mL HisTrap™ HP column previously equilibrated with a 0.02 M Tris-HCl pH 8, 0.800 M NaCl and 0.02 M Imidazole buffer, concentrated to a suitable volume and then loaded on a Superdex 75 column in a 0.050 M Tris-HCl pH 8, 0.500 M NaCl buffer with a flow rate according to manufacturer's guidelines for the column. The purified protein was collected in fractions of 0.5 or 1 ml, depending on the bed volume of the column used and analysed by SDS page.

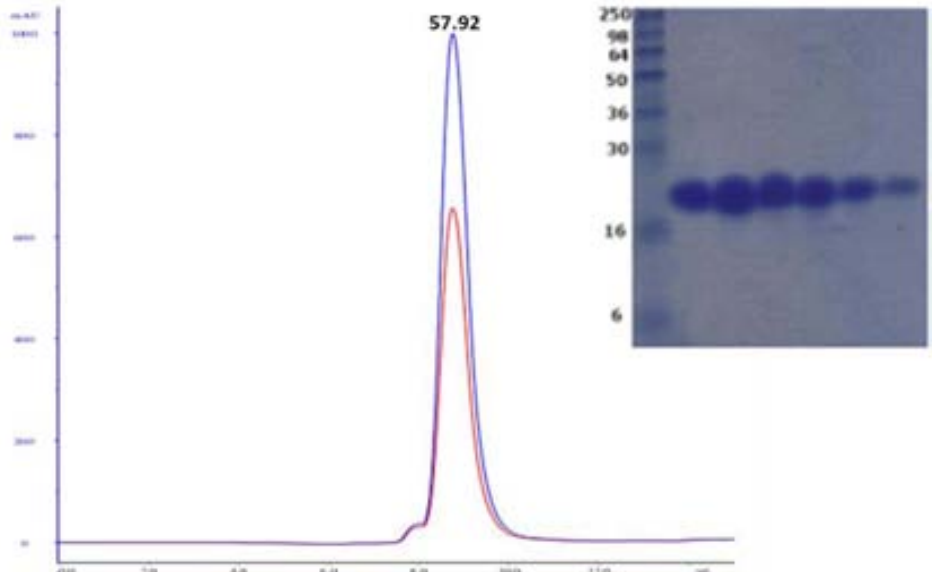

Figure 3-21. Purification of flMG219 by GFC and SDS-PAGE analysis of the eluted peak.
Chromatographic profile of flMG219 protein with a Superdex 75 16/60 GFC column with its elution volume highlighted. 15% (v/v) SDS-PAGE analysis of the fractions corresponding to the GFC peak are also shown. Molecular weights of standard proteins are reported.

According to its elution volume and a previously determined calibration curve, obtained for the same column running standards of known size, MG219 elutes as a tetramer in solution. The oligomerization state-of the expressed protein was also analysed by nanoelectrospray ionization mass spectrometry (nanoESI-MS) and multi angle light scattering (MALS), which instead pointed out the monomeric nature of the protein in solution.

The unusual behaviour of flMG219 in GFC could be explained by the very high content of lysine residues at its C-terminal, most likely causing this region being very flexible and thus not behaving like a hypothetical globular protein, for which the estimation of MW by GFC is still reliable.

The fractions of the peak corresponding were pooled together and concentrated to 20 mg/ml using Amicon Ultra MWCO 3,5 KDa centrifugal filters (Sigma-Aldrich).

2.2. Crystallization of flMG219 protein

Initial crystallization screening at 4 and 20 °C of flMG219 resulted in more than 80% of clear drops, even when screened at relatively high protein concentrations (20 to 50 mg/mL), likely because of the high content of charged residues, which account for the 30% of the protein sequence.

To reduce MG219 solubility, lysine residues were di-methylated following the protocol described in 6.1.1. This should have two additional consequences: i) reduce the pI of the modified protein and ii) due to the hydrophobic nature of the dimethylated lysines, protein-protein interactions should be favoured, leading to possible improvements in term of crystallizability. Due to the high content of lysines, it was possible to check the results directly by running a SDS PAGE, as shown in Figure.

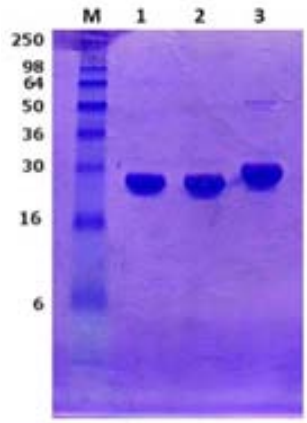

Figure 3-22. 15% (v/v) SDS-PAGE analysis of methylated flMG219.
Lane 1 corresponds to MG219 after the last step of GFC purification. Lane 2 corresponds to the same sample as loaded in lane 1, but after 2 hours dialysis to completely remove any trace of Tris buffer used to purify flMG219, which could interfere with the methylation. Lane 3 represents methylated flMG219, after 3 hours reaction. Molecular weights of standard proteins are reported (lane M).

After the methylation process MG219 was complete, the protein was concentrated to 25 mg/ml using Amicon Ultra MWCO 3,5 KDa centrifugal filters (Sigma-Aldrich), and initial crystallization screenings set up as for the unmethylated. Even with the methylated protein, more than 50% of the crystallization drops were clear, with any sign of nucleation or crystallization after several months.

Limited proteolysis with papain was used to investigate the possible existence of a more compact domain, less recalcitrant to crystallization. The bands obtained after the proteolysis (Figure 3-23) were further analysed by Edman sequencing, which surprisingly revealed that both fragments exhibited the same N terminal residues Ile73, Pro74 and Thr75, corresponding to the start of a C terminal region predicted to be mainly unstructured.

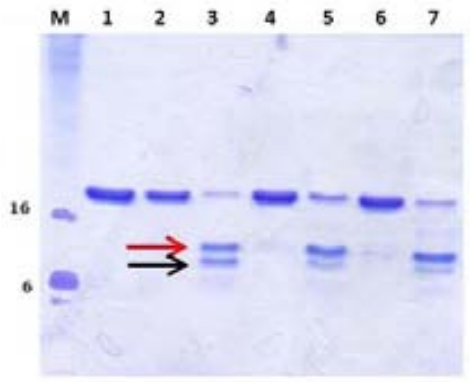

Figure 3-23. 12%(v/v) SDS-PAGE time course analysis flMG219 papain-limited proteolysis.
The bands indicated with a black and red arrow were further analysed by Edman sequencing. Lane 1, uncleaved flMG219.. Lane 2, 60 minutes treatment with papain at 10 µg/ml. Lane 3, 60 minutes treatment with papain at 100 µg/ml. Lane3, 120 minutes treatment with papain at 10 µg/ml. Lane 4, 120 minutes treatment with papain at 100 µg/ml. Lane 5, 180 minutes treatment with papain at 10 µg/ml. Lane 6, 180 minutes treatment with papain at 100 µg/ml. Molecular weights of standard proteins are reported (lane M).

2.3. Cloning and expression of MG219 variants

Thus two new smaller variants of MG219 were designed, spanning residues 1-79 and 75-148 and named, respectively, MG219-Nt_1-75 and MG219-Ct_75-148. Using as template MG219 gene with the and the primers specified in Table, both fragments were cloned

separately in pOPINE expression vector (Berrow, Alderton et al. 2007), which encodes for an extra lysine and a His6 tag at the C-terminal, which can easily be cleaved with carboxypeptidase A protease for improved crystallizability.

MG219 variant	Primer name	Primer sequence
MG219-Nt_1-79	MG219_F1	5' AGGAGATATACCATGCGCACCAGTTACTTGAAAAAAATACC
	MG219_R79	3' GTGATGGTGATGTTTAACAACAGGGGTTGGGATTAGTAAAAC
MG219-Ct_75-148	MG219_F75	5' AGGAGATATACCATGACCCCTGTTGTTAAAGAGATTGATCAAC
	MG219_R148	3' GTGATGGTGATGTTTAGATCTGGTTTTTTTATTGCTTTTTTTTGTTTGAA

The sequenced plasmids were then transformed into BL21(DE3) cells and expressed normally as the full length protein. The products of both constructs were purified following the protocol established for flMG491 protein, with comparable yields.

3. Screening wheel complex protein interactions

The ultrastructure detectable in Mycoplasma genitalium TO, which is thought to be formed by at least 11 proteins, is a multicomponent macromolecular complex constituted by a series of proteins that must interact with each other to permit the normal assembly and functioning of the TO. Thus, possible binary interactions among the proteins which presumably form the M. genitalium TO wheel complex have been analysed.

The complete list of constructs tested at this stage (full length proteins or putative domains) is reported in Table

Construct name
MG219 [1]
MG219-Nt_1-79 [2]
MG491 [1]
MG491-Nt_1-308 [2]
MG491-Nt_1-255 [2]
MG200 [1]
MG200-Ct_304-601 [2]
MG200-DnaJ_1-74 [2]
MG200_DnaJ+EAGR_1-207 [2]
MG200_EAGR_124-207 [2]
MG386_EAGR1-2_148-262 [2]
MG386_EAGR3-4-5_1433-1616 [2]

[1] full length protein, [2] putative domain

SPR multi array system ProteOn™ XPR36 (BioRad) was firstly employed to speed up the initial screening for binary interactions, then the positive hits were further analysed by NMR and GFC, to corroborate the preliminary SPR results. SPR and NMR were also employed to study the kinetic of each one of the identified interaction.

3.1. Screening binary interactions by SPR

The screening of possible binary protein-protein interactions that could be necessary for the proper assembly and functioning of the TO wheel complex permitted the identification of several interactions, which are summarized in Table. This qualitative assay allowed to disclose the interactions that take place between the four proteins forming the wheel complex. MG491 was found to be interacting with MG200 full length protein and, more in particular, with its recently described EAGR domain, which is found exclusively in few proteins forming the TO (Calisto, Broto et al. 2012) in M.genitalium, M.pneumoniae and M.gallisepticum. The construct containing the only MG200 EAGR box (spanning MG200 residues 124-207) was shown to interact with MG491 full length but not with the smaller constructs MG491-Nt_1-308 and MG491-Nt_1-255. The construct expressing the last three consecutive EAGR boxes of MG386 (spanning MG386 residues 1433-1616) was shown to bind to MG219 full length but not to MG219-Nt_1-79, suggesting the measured interaction involved the C terminus of MG219. MG219 was also shown to interact with MG200-Ct_304-601 putative domain through its N terminal domain, evincible from the positive hits that MG200-Ct_304-601 gave with both MG219 and MG219_Nt_1-79 constructs.

		BAIT										
		MG219	MG219 Nt_1-79	MG491	MG491 Nt_1-255	MG491 Nt_1-308	MG200	MG200 Ct_304-601	MG200 DnaJ_1-74	MG200 EAGR_124-207	MG386 EAGR1-2	MG386 EAGR3-4-5
LIGAND	MG219	M	X	X	X	X	✔	✔	X	X	X	✔
	MG219 Nt_1-79		M	X	X	X	✔	✔	X	X	X	X
	MG491			T	X	X	✔	X	X	✔	X	X
	MG491 Nt_1-255				T	X	X	X	X	X	X	X
	MG491 Nt_1-308					T	X	X	X	X	X	X
	MG200						T	X	X	X	X	X
	MG200 Ct_304-601							T	X	X	X	X
	MG200 DnaJ_1-74								M	X	X	X
	MG200 EAGR_124-207									D	X	X
	MG386 EAGR1-2										M	X
	MG386 EAGR3-4-5											M

3.2. MG491/MG200 interaction

No SPR response was detectable when either MG491-Nt_1-308 (Figure 3-24, dashed line) or MG491-Nt_1-255 (data not shown) was used as ligands, with MG200 covalently bound to a CM5 sensor chip as before. On the other hand, a clear positive SPR response was measured when MG491 was injected over MG200, immobilized on a sensor chip by amine coupling (Figure 3-24, solid line).

The preliminary results suggested that MG491 is interacting with MG200 EAGR box using a small domain located at the C terminus of MG491 protein, since two constructs missing the last 40 (MG491-Nt_1-308) or 93 (MG491-Nt_1-255) residues didn't show any binding to the immobilized MG200 EAGR box.

The PPI initially detected between the TO wheel complex proteins MG200 and MG491 was further analysed by SPR, injecting MG491 over MG200, which was covalently bound to a sensor chip by amine coupling to ~6000 RU. SPR response increased in proportion to increasing concentrations of MG491 and data was fit to a 1:1 binding model resulting in an equilibrium dissociation constant (KD) of 78.7 nM, which corresponds to association (ka) and dissociation rates (kd) of 2.48x104 M-1 s-1 and 1.95x10-3 s-1, respectively.

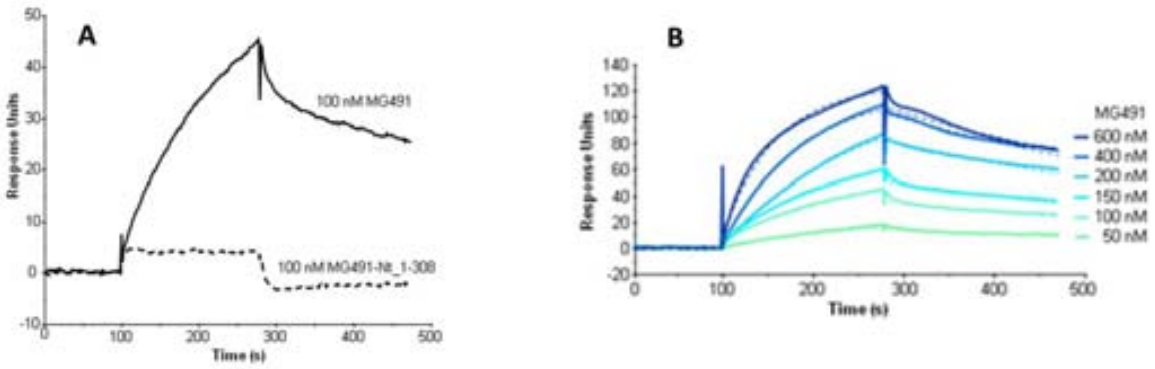

Figure 3-24. Comparative analysis of the interactions between variants of the MG200 and MG491 proteins.
A) SPR sensorgrams of 100 nM MG491 and 100 nM MG491-Nt_1-308 injected over the MG200-immobilized cell and of 40 nM MG200 and 40 nM MG200ΔEAGRbox over the MG491-Ct_peptide1-immobilized cell, respectively. B) Interaction of immobilized MG200 with increasing concentrations of MG491 as measured by SPR. The coloured lines represent the measured response units while the coloured dashed lines represent the calculated fit to a 1:1 Langmuir binding model, producing calculated constants of k_a 2.48x10^4 M^{-1} s^{-1}, k_d 1.95x10^{-3} s^{-1}, K_D 78.7 nM.

To crosscheck these results, the interaction between MG200 EAGR box and MG491 or its variants has also been investigated by GFC and NMR. In an attempt to further distinguish between a stable and a transient complex, gel filtration chromatography was used for the isolation of the detected complex between MG200 full length (or MG200-EAGR_124-207) and MG491 full length. However, nor co-expressing the two proteins neither an overnight incubation at 4 °C allowed the complex formation and isolation, suggesting the interaction between these two proteins could be transient (data not shown). Therefore, in collaboration with Daniela Lalli (Center for Magnetic Resonance CERM, Sesto Fiorentino, Italy) NMR spectroscopy was used to better characterize the interaction between MG200 and MG491 proteins.

First, the same construct used in crystal structure determination of MG200 EAGR box (Calisto, Broto et al. 2012), spanning MG200 residues 124-207 plus a C terminus 6 histidine tag, was $^{13}C/^{15}N$ double-labelled and its NMR spectrum assigned. Combining the information derived from triple resonances experiments (Kay, Ikura et al. 1990, Wuthrich 1990, Wider and Wuthrich 1999) 96% of the Cα, 85% of the CO, 95% of the backbone amides ^{1}H and ^{15}N and 89% of the Cβ were assigned together with 88% of side-chain carbons and hydrogen protons, excluding the 6 histidine tag. Missing resonances correspond to residues Lys124, Gln125, Glu126 and Val136 from the N-terminal region and also to residues from the loop connecting helix α1 with strand β3. Secondary structure elements predicted by TALOS from the backbone torsion angles restraints derived from the chemical shift index analysis match the ones found in the MG200 EAGR box crystal structure (PDB accession code 4DCZ). Therefore, the MG200 EAGR box maintains in solution the same secondary structural organization found in crystals. Briefly, two consecutive β-hairpins composed of β-strands β3-β4 and β5-β6 are preceded by helix α1 and strand β2 which, by flanking the β-hairpin β5-β6, form a hydrophobic core (Figure 3-25).

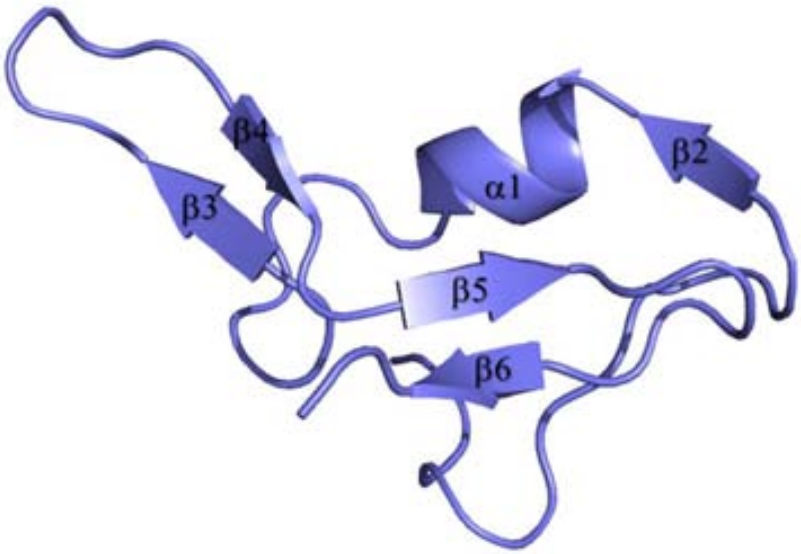

Figure 3-25. Cartoon representation of the crystal structure of the MG200 EAGR box.
The secondary structure elements nomenclature is consistent with that previously reported (Calisto et al., 2012) (PDB accession code 4DCZ). Residues with missing resonances are coloured red.

The interaction between MG200 EAGR box and MG491 has been investigated at the atomic level monitoring the complex formation by ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) NMR spectra. Backbone chemical shift variations of ^{15}N-MG200 EAGR box occurring upon addition of increasing amounts of unlabelled MG491 reveal the formation of a specific albeit transient complex. The interface between proteins was defined following the HSQC spectra resonances of the ^{15}N-MG200 EAGR box during the titration with MG491 (Figure 3-26, inset A). The linearity of the chemical shift changes in the HSQC maps proves that a single binding event is occurring. MG200 EAGR box residues experiencing chemical shift perturbations larger than 0.03 ppm during titration have been mapped on the protein structure (Figure 3-28, inset A). The MG200 EAGR residues that were more affected by the interaction with MG491 upon complex formation cluster in the region belonging to the solvent exposed face of helix α1 (Ala166, Ser167, Glu170 and Ala171), strand β2 (Glu163-Thr165) and to the N-terminal residues Leu156 to Val164. To determine the region of MG491 involved in the interaction with MG200 EAGR box, MG491-Nt_1-308 and MG491-Nt_1-255 constructs were further investigated by NMR. As observed in SPR experiments, no backbone chemical shift variations of ^{15}N-MG200 EAGR box were detected after addition of increasing amounts of either unlabelled MG491-Nt_1-308 and MG491-Nt_1-

255 proteins (Figure 3-26), confirming that a C terminal region of MG491, which is missing in these two constructs, is critical for the interaction with MG200.

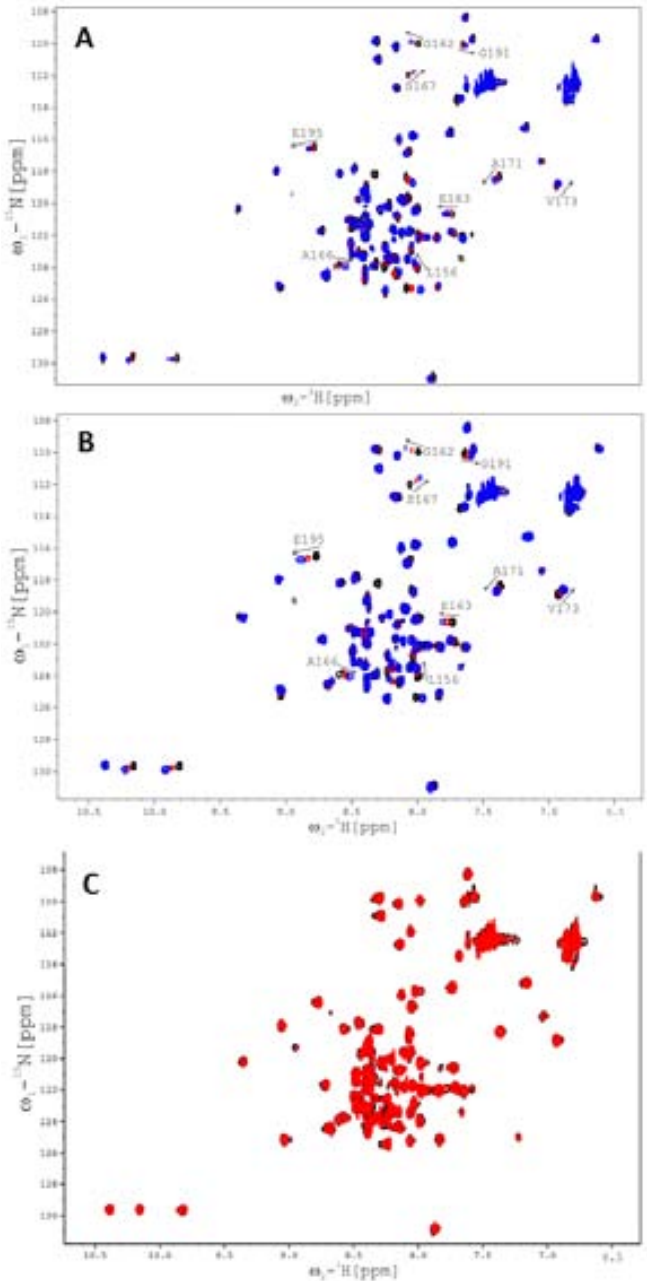

Figure 3-26. Chemical shift perturbation (CSP) mapping of MG200 EAGR box interacting residues.
A) Superimposition of ^{1}H-^{15}N HSQC spectra of MG200 EAGR box in absence and in presence of MG491. MG200 EAGR box:MG491 complex ratios 1:0, 1:4 and 1:8 are represented in black, red and blue traces, respectively. Very low intensity peaks are below the displayed threshold in some of the spectra. B) Superimposition of ^{1}H-^{15}N HSQC spectra of MG200 EAGR box in absence and in presence of MG491-Ct-peptide1. MG200 EAGR box:MG491-Ct_peptide1 complex ratios 1:0, 1:3.5 and 1:7 are represented in black, red and blue traces, respectively. C) Superimposition of ^{1}H–^{15}N HSQC spectra of MG200EAGRbox (black trace) and MG200-EAGRbox:MG491-Nt_1-255 complex at a 1:4 ratio (red trace).

To validate this result, chemical-shift perturbation mapping was used to study the interaction between MG200 EAGR box and the C-terminal end of MG491. The lack of binding of both MG491-Nt_1-308 and MG491-Nt_1-255 proteins to MG200 indicated that the last forty MG491 residues, missing in both constructs, were crucial for the interaction. For that, two peptides from MG491 C terminal region, spanning MG491 residues 299-346 were custom-synthesized by GenScript (Piscataway, NJ, USA). To reduce the peptide syntheses cost and maximize the yields, instead of a 47 residues long peptide, the putative C terminal domain of MG491 was break down into two smaller fragments, which correspond respectively to residues Thr299-Pro323 (MG491-Ct_peptide1) and residues Pro324 to Glu346 (MG491-Ct_peptide2). The break point between Pro323 and Pro324 was chosen since its known Proline acts as a structural disruptor in the middle of regular secondary structure elements such as α helices and β sheets, thus minimizing the risk of altering MG491 secondary structure. The two lyophilized peptides which sequences are shown in Figure 3-27(B) were dissolved in pure ddH_2O to a final concentration of 2 mM. Stability over time for both peptides was assayed by checking their HPLC-PDA profiles at fixed interval of time in ddH_2O (Figure 3-27, insets C and D) and in 20 mM phosphate buffer with 100 mM NaCl (data not shown). HPLC profile of MG491-Ct_peptide1 and MG491-Ct_peptide2 showed a single peak, suggesting the mono-dispersion of both fragments. Finally, MALDI-TOF technique was used to determine MG491-Ct_peptide1 and MG491-Ct_peptide2 oligomerization state: with a detected MW of 2974.91 Da for MG491-Ct_peptide1 (expected MW 2975.57 Da) and 2689.46 Da for MG491-Ct_peptide2 (expected MW 2689.94 Da), the measures were in agreement with a monomeric state for both peptides in solution (Figure 3-27, insets E and F).

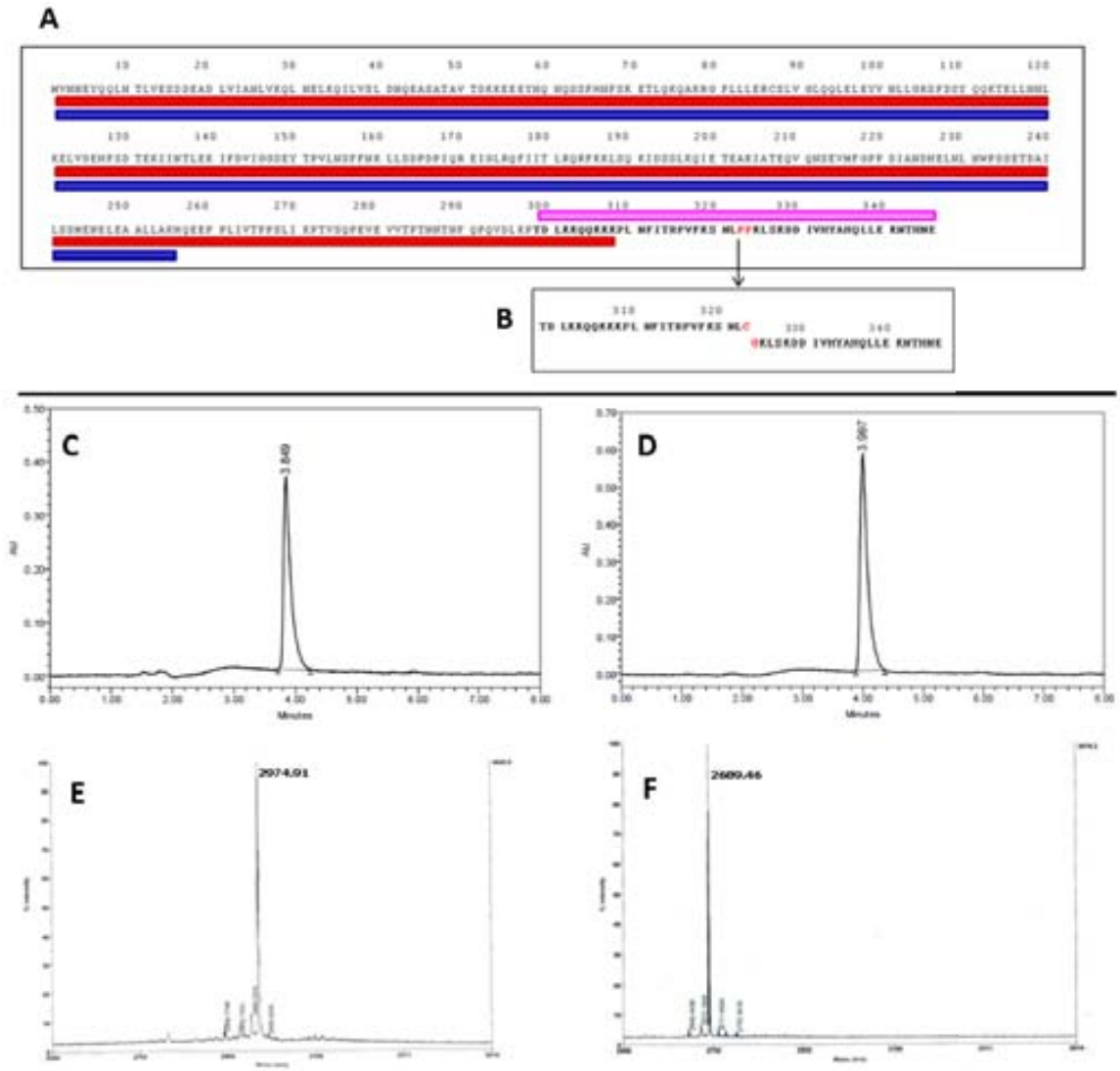

Figure 3-27. Amino acidic sequences of MG491 variants.
A) Amino acidic sequences of flMG491 (residues Met1 to Glu346), MG491-Nt_1-255 (residues Met1 to Lys255, highlighted with a blue box) and MG491-Nt_1-308 (residues Met1 to Lys308, highlighted with a red box). The C-terminal end of MG491 (residues Thr299 to Glu346), which is responsible for the interaction with MG200 EAGR box, is highlighted in magenta. B) Amino acidic sequence of two peptides from the MG491 C-terminal end: MG491-Ct_peptide1 (residues Thr299 to Pro323) and MG491-Ct_peptide2 (residues Pro324 to Glu346). C) HPLC profile of MG491_peptide1 dissolved in ddH_2O. Conditions: SunFireTM C18 analytical column (3.5 µm x 4.6 µm x 100 µm); tR = 3.97 min, from 5% to 100% ACN over 8 min, purity > 97%. Linear gradients of ACN (0.036% TFA) into H2O (0.045% TFA) with a flow rate of 1 mL/min were used. D) HPLC profile of MG491_peptide2 dissolved in ddH_2O. Conditions: SunFireTM C18 analytical column (3.5 µm x 4.6 µm x 100 µm); tR = 3.97 min, from 5% to 100% ACN over 8 min, purity > 98%. Linear gradients of ACN (0.036% TFA) into H2O (0.045% TFA) with a flow rate of 1 mL/min were used. E) and F) MALDI TOF mass spectrum of MG491_peptide1 and MG491_peptide2, respectively., using α-Cyano-4-hydroxycinnamic acid (ACH) as matrix for the analysis.

The titration of ^{15}N-MG200 EAGR box with unlabelled MG491-Ct_peptide2 do not show any significant shift (with the exception of the resonances belonging to the histidine tag) indicating that no interaction is taking place. On the contrary, significant chemical shift variations were observed on the titration with unlabelled MG491-Ct_peptide1 confining

the interaction area of MG491 to a 25-residue-long region on the C-terminal end (Figure 3C). The shifts changes are large and progressive indicative of a weak interaction occurring at a fast rate on the NMR chemical-shift timescale, as confirmed from the calculated KD (550±200 µM). The shift trajectories of the MG200 EAGR box in complex with MG491-Ct_peptide1 are similar to those observed for MG491 (compare 3A and 3C), suggesting that the chemical environment of the protein residues is similar in the two complexes. The weighted average chemical shift perturbations were plotted as function of the MG200 EAGR box backbone resonances and mapped on the protein surface (Figure 3-28). From the comparison of these changes with those observed for MG491 it is apparent that they are significantly similar in magnitude and distribution on the protein surface. Indeed, for MG491-Ct_peptide1 the interacting residues are mainly located in solvent exposed face of helix α1, strand β2 (Glu163 to Thr165) and in the N-terminal loop residues Leu156 to Val164. Some additional shifts were observed for residues Asp194 and Glu195 from the loop connecting the β-hairpin strands β5 and β 6, which can be due to new interactions or to a conformational change in this region. Moreover, there are missing shifts for some residues of the N-terminal end (namely for Val136, Glu137, Asp154 and Glu155) very likely due to the smaller size of MG491-Ct_peptide1 with respect to MG491 and probably responsible for the smaller affinity of MG200 EAGR box-MG491-Ct_peptide1 complex in comparison with the MG200 EAGR box-MG491 complex.

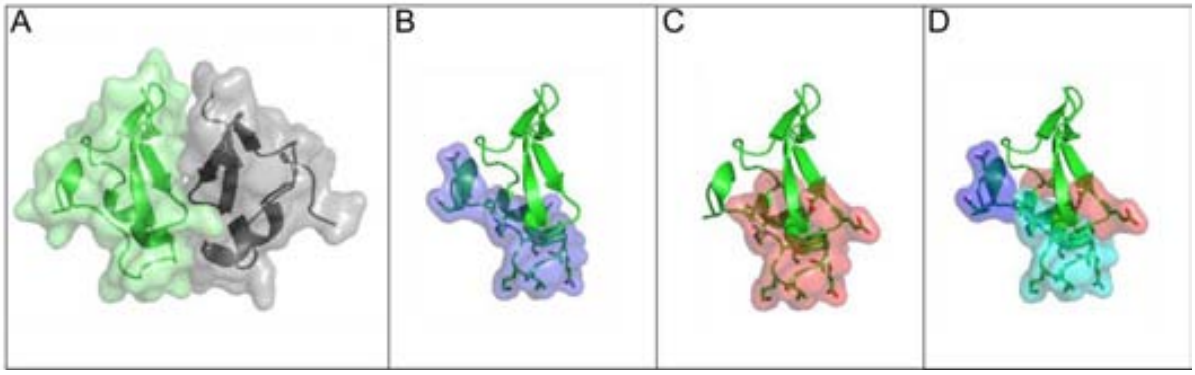

Figure 3-28. Cartoon representation of MG200EAGR tertiary structure and interacting residues.
A) Semi-transparent surface and cartoon representation of EAGRMG200 dimer. B) Map of the interaction surface with fIMG491 on the structure of the MG200 EAGR box monomer. C) Map of the interaction surface with MG491-Ct_peptide1 on the structure of the MG200EAGR box monomer. D) Superimposition of C) and D) interaction surfaces.

To validate the results obtained by NMR the interaction of the MG200 EAGR box with MG491-Ct_peptide1 was investigated by additional SPR measurements. Firstly, MG491-Ct_peptide1 was immobilised on a sensor chip by an amine coupling reaction. After immobilization, MG200 or MG200ΔEAGR box (a construct lacking only MG200 residues 124-207, which correspond to the EAGR box region) were injected over MG491-Ct_peptide1 (Figure 5). Injecting MG200 over MG491-Ct_peptide1 gave an SPR response that was inexistent when MG200ΔEAGR box was injected on the same flow cell. Together, these results clearly demonstrate that MG200 binds MG491 through its EAGR box containing domain and that MG491 binds MG200 through a 25 amino acids long peptide at its C-terminal region. This conclusion completely corroborates the NMR results described above.

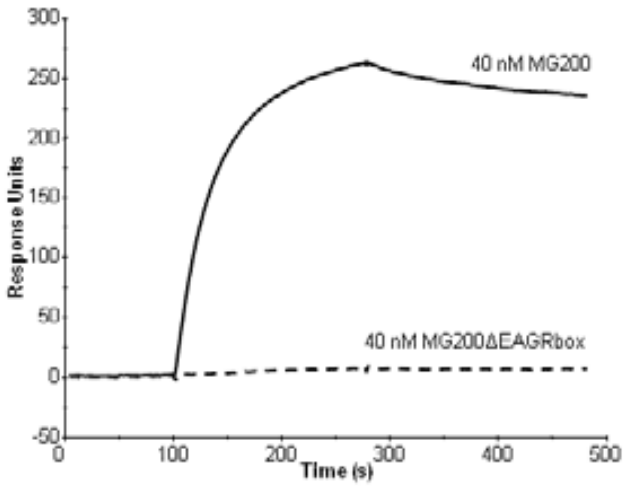

Figure 3-29. SPR sensorgrams of MG200 and MG200ΔEAGRbox over the MG491-Ct_peptide1.
SPR sensorgrams of 40 nM MG200 (solid line) and 40 nM MG200ΔEAGRbox (dashed line) over the MG491-Ct_peptide1, showing the absence of SPR response when injecting MG200ΔEAGRbox, a construct which lacks EAGR box, over immobilized MG491-Ct_peptide1.

To determine the role of this peptide in vivo and to test the relevance of the described interactions a M. genitalium mutant strain lacking the MG491-Ct_peptide1 was engineered by gene replacement by Luis Garcia Morales (Institut de Biotecnologia i de Biomedicina, UAB). Approximately 30% of the recovered clones showed a PCR fragment consistent with the presence of a 75 bp deletion in the 3' region of mg491 (Figure 6A) which was further confirmed by sequencing the PCR products (data not shown) The presence of the intended

deletion was also demonstrated by Western blot on protein profiles from three recombinant clones using a heterologous anti-P41 antiserum (Krause, Proft et al. 1997) (Figure 6B). A shift in the electrophoretic mobility of MG491 protein band consistent with the presence of a 25 amino acids deletion was apparent when comparing the protein profiles of the recombinant clones with the protein profile of the G37 WT strain. Clones 1 and 9, bearing the intended deletion, were also investigated for the existence of polar effects affecting or reducing the expression of the downstream mg219 gene by Western blot. In this blot no differences were detected in the levels of MG219 when comparing the mutant clones with the G37 WT strain (Figure 6C). Furthermore, the protein profile from clone 1 was also compared with the one from G37 WT strain to check for downstream events involving the adhesins and the cytadherence-related proteins, which levels were comparable in both strains, indicating that no downstream events affecting the major TO proteins occur despite the deletion of MG491-Ct_peptide1. Clone 1 was named Δp1c1 and selected for further studies to account for possible gliding motility impairments due to the lack of the C-terminal region of MG491.

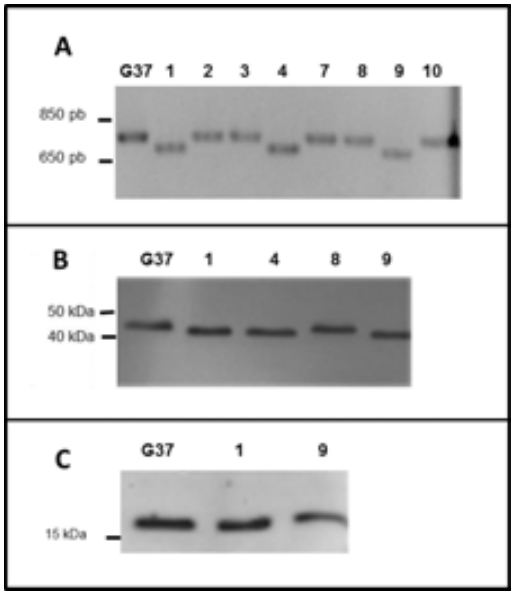

Figure 3-30. Engineering of M. genitalium MG491-Ct_peptide1 deficient mutant (Δp1).
A) PCR screening of mutants/clones using genomic DNA as template. The deletion of the region coding for MG491-Ct_peptide1 was identified by the presence of a PCR band of smaller size (~700 bp; Clones 1, 4 and 9) in comparison with the one obtained for the G37 WT strain. B) Western blot using a polyclonal serum against the P41 protein from M. pneumoniae (homologous to MG491 protein). Clones 1, 4 and 9 show a shift in the MG491 band when compared with the G37 WT strain, consistent with the deletion of the 25 amino acids of the MG491-Ct_peptide1. C) Western blot using protein profiles from G37 WT strain and clones 1 and 9 and a polyclonal antiserum against MG219 protein. All cells show a similar level of the MG219 protein discarding the existence of polar effects derived from the deletion of MG491-Ct_peptide1.

The motile properties of Δp1c1 cells were first investigated by time lapse microcinematography. Several parameters, such as the frequency of motile cells, the mean gliding velocity and the diameter of the circular cell tracks, were quantified and compared to those from WT cells (Figure 7, insets A and B). Even though the ratio of motile cells in the Δp1c1 strain was slightly but significantly lower than the ratio found in G37 WT cells, the mean gliding velocity of the motile cells was very similar in both strains. The most significant differences were found when examining the diameter of the circular tracks, which were dramatically reduced in the Δp1c1 strain. Since most of the M. genitalium cells glide drawing circular tracks, these results suggest that Δp1c1 cells showing very narrow circular tracks could have a lower capacity to disseminate than WT cells, which in majority draw wider tracks.

The presence of MG491-Ct_peptide1 deficient cells showing altered patterns of gliding motility prompted the IBB collaborators to investigate the TO architecture by scanning electron microscopy (SEM). M. genitalium G37 WT cells have curved terminal organelles and their curvature correlates very well with the diameter of the circular tracks(Burgos, Pich et al. 2008). Many cells from the Δp1c1 strain exhibited terminal organelles extremely twisted over the cell body, with a significant reduction (~10°) in the curvature angle when compared to G37 WT cells (Figure 7, insets C-I), providing an explanation to the narrow tracks observed in the microcinematographic studies.

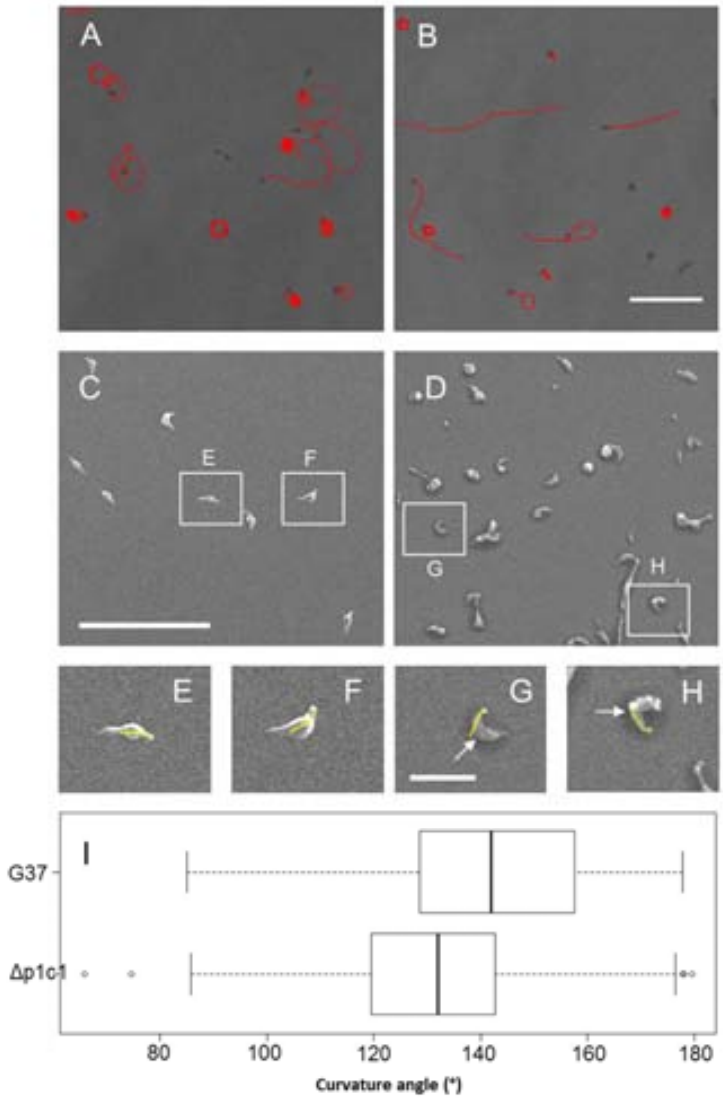

Figure 3-31. Gliding motility and TO morphology of M. genitalium G37 WT and Δp1c1 cells.
A) and B) Microcinematographic studies of G37 WT and Δp1c1 strains, respectively. Bar is 5 µm. C) and D) SEM micrographs of G37 WT and Δp1c1 strains, respectively. Bar is of 5 µm. E) and F) Insets of panel C. G) and H) Insets of panel D. Yellow lines indicate the dihedral angles between the TO and cell body axis. White arrows point to the bipartite bodies found in Δp1c1 cells. Bar is of 1 µm. I) Box and whisker plots of the curvature angles determined for 100 cells of G37 WT and Δp1c1 strains, respectively.

Taken together, these informations indicate that the M. genitalium MG491 protein plays a similar role to that of its M. pneumoniae homologous protein, being the peptide 1 region essential for MG491 function. In consequence, the interactions with EAGR box(es) may be an important factor conferring stability of the mycoplasma cytoskeleton.

3.3. MG219/MG200 interaction

The initial SPR screening aimed to identify possible binary interactions among the four proteins that localize at the wheel complex pointed out that MG200 is not only interacting through its EAGR domain with the C terminal region of MG491, but is also interacting with MG219. In particular, MG200 it seemed to interact through its C-terminal putative domain, named MG200-Ct_304-601, with the N terminal variant of MG219, called MG219-Nt_1-79.
To further investigate and validate these preliminary results, SPR was used to follow the RU signal upon linear addition of MG200-Ct_304-601 on either immobilized MG219 full length (Figure 3-32, inset A) or MG219_Nt_1-79 (inset B). The sensorgrams profiles in both case indicated that the interaction taking place between MG200-Ct_304-601 and MG219 (or MG219-Nt_1-79) is relatively strong, clearly evincible from the almost absent complex dissociation.

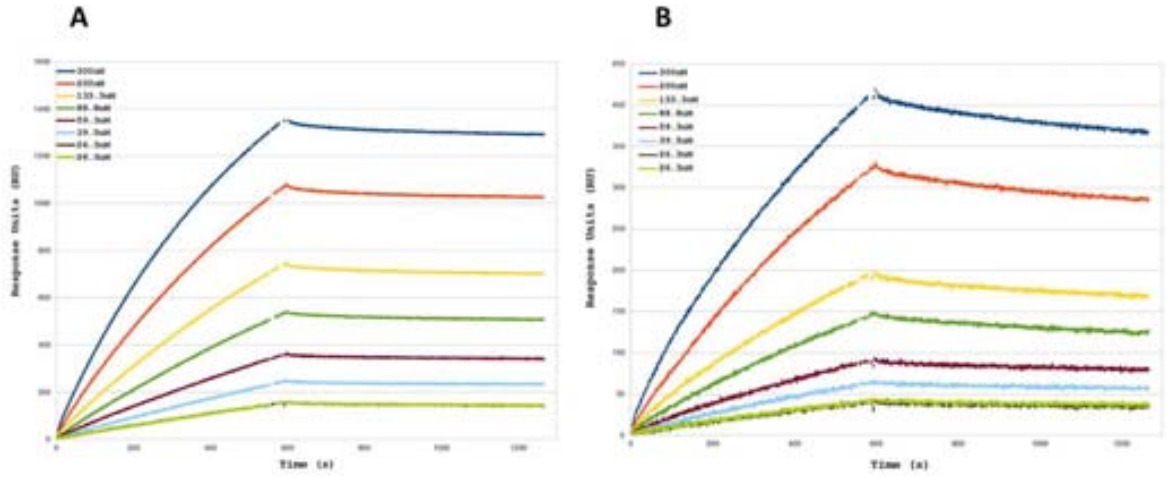

Figure 3-32. SPR sensorgrams for the interaction of MG200-Ct_304-601 with MG219 constructs.
A) Sensorgrams of MG200-Ct_304-601 at concentration from 26.3 to 300 nM over MG219. B) Sensorgrams of MG200_Ct-304-601 at concentration from 26.3 to 300 nM over MG219-Nt_1-79.

GFC profiles of MG200-Ct_304-601 pre-incubated with MG219 (or with MG219-Nt_1-79) corroborate the interaction between MG200 and MG219, confirming the interaction involved MG200 C-terminal residues and a region spanning residues 1-79 of MG219 (Figure 3-33).

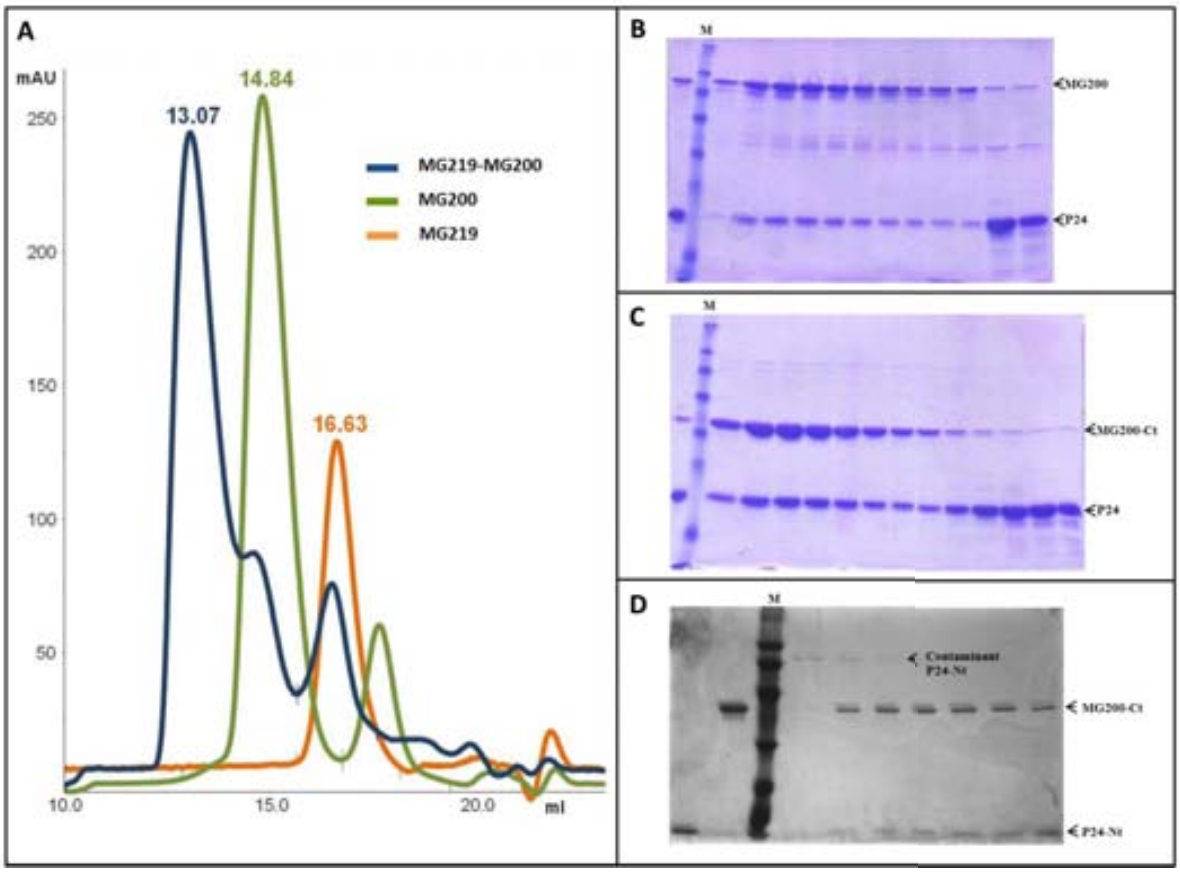

Figure 3-33. GFC and SDS-PAGE analysis for the interaction of MG219 with MG200.
A) Chromatographic profiles with a Superdex 200 10/300 GFC column of flMG219 (orange), flMG200 (Green, Ingram et al.) and MG200:MG219 complex (blue) with the corresponding elution volumes highlighted. 15% (v/v) SDS-PAGE analysis of the GFC peak for B) MG200:MG219 complex, C) MG200-Ct_304-601:MG219 complex and D) MG200-Ct_304-601:Mg219-Nt_1-79 complex are also shown. Molecular weights of standard proteins are reported (Lane M).

3.4. MG219/MG386 interaction

In addition to the interaction with MG200 C terminal domain, SPR mid throughput assay also revealed MG219 is interacting with its C terminal domain, covering residues 75-148, with a C terminal domain of MG386, characterized by the presence of three consecutive EAGR boxes, which share high sequence similarity with the EAGR box of MG200.

In figure 3-34 are shown the GFC profiles of the MG219 and MG386 injected separately (in orange and red, respectively) and then GFC profile of the obtained MG219-MG386 complex (in blue), confirming the initial interaction detected during the mid-throughput SPR assay (Table).

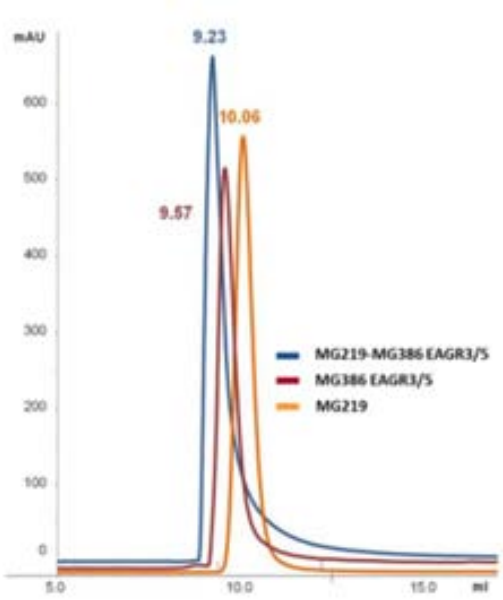

Figure 3-34. GFC analysis for the interaction of MG219 with MG386-EAGR3/5.
Chromatographic profiles with a Superdex 200 10/300 GFC column of flMG219 (orange), MG386-EAGR3/5 (Bravman, Bronner et al.) and MG219:MG386-EAGR3/5 complex (blue) with the corresponding elution volumes highlighted.

In collaboration with Melanie Schwarten and Bernhard Brutscher from the Biomolecular NMR Spectroscopy Group (IBS, Grenoble, France) the interaction between MG219 and MG386 C-terminal was also analysed by NMR taking advantage of the fact that both proteins are shorter than 200 amino acids and so within the ideal size to be analysed by

this method. The proteins were produced in minimal media in presence of 15N-ammonium chloride the interaction was then verified by performing a titration experiment which consisted in acquiring 2D-NMR spectra of the labelled-MG386-Ct protein in absence and presence of increasing concentrations of MG219 (Figure inset A). The peak shifts/peak disappearances appreciated when the different spectra are superposed correspond to perturbations in residues of the MG386-Ct caused by the interaction with MG219. Analogously, the addition of MG386 C-terminal caused the shift or disappearance of several flMG219 peaks. The assignment of MG219 full length NMR spectrum confirmed that the residues affected by the addition of unlabelled MG386 C-terminal are located at the C terminal of MG219, corroborating the initial SPR results.

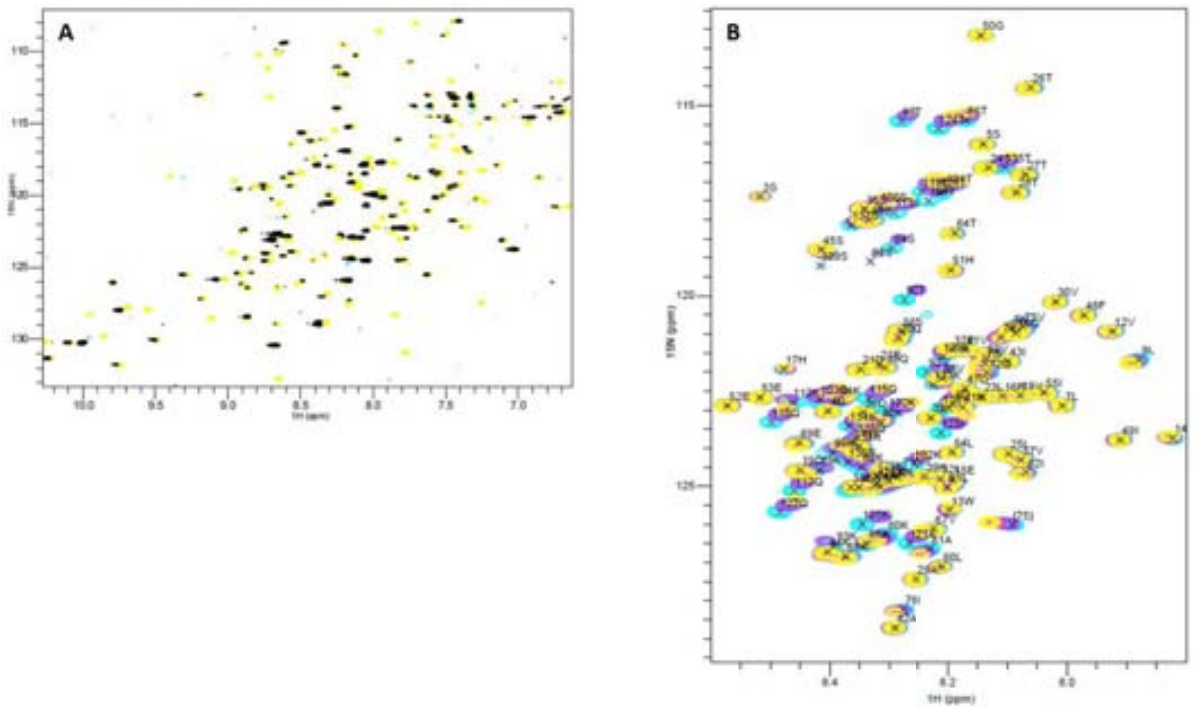

Figure 3-35. Chemical shift perturbation (CSP) analysis of MG219 and MG386-EAGR3/5 interacting residues.
A) MG386 EAGR3/5 $^{1}H,^{15}N$-HSQC spectra in absence (black) and in presence of MG219 (yellow). B) MG219 $^{1}H,^{15}N$-HSQC spectra in absence (yellow) and presence of MG386 EAGR3/5 (violet and light blue);

CONCLUSIONS

Conclusions

Several transient and strong interactions between MG219, MG200, MG386 and MG491 have been identified and characterized by three different biophysical techniques such as surface plasmon resonance (SPR), nuclear magnetic resonance (NMR) and gel filtration chromatography (GFC).

These results, in addition to the structural analysis of the solved MG491-Nt_63-203 structure, contribute in a discrete but accurate way to a better comprehension of the interactions that have to take place at the wheel complex and of possible the role of MG491 in wheel complex ultrastructure, and are now listed in form of conclusions.

1) MG200 EAGR box, a characteristic domain only found in terminal organelle (TO) proteins, interacts in a transient way with a 25 amino acid residues at the C-terminal region of MG491, named MG491-Ct_peptide1.

2) MG200 C-terminal region strongly binds to the N terminal domain of MG219.

3) MG386 interacts through its C-terminal, a region containing three non-identical EAGR boxes, with the C-terminal region of MG219.

4) *In vivo* experiments of MG491-Ct_peptide1 deficient cells showed altered patterns of gliding motility, caused by a 10° decrease of the curvature angle compared to wild type cells.

5) MG491 full length protein behaves in solution as a tetramer, predicted to be composed mainly by α helices, with a more flexible C-terminal region.

6) Trypsin limited digestion helped in the identification of two MG491-Nt constructs, spanning residues 1 to 255 (named MG491-Nt_1-255) or residues 1 to 308 (named MG491-Nt_1-308) for which high quality-diffracting crystals were obtained.

7) A SeMet MG491-Nt_1-308 dataset was successfully used to solve at 3Å resolution, by single-wavelength anomalous dispersion (Glass, Assad-Garcia et al.), the structure of MG491-Nt-63-205.

8) MG491-Nt_63-205 structure forms a dimer of structural hetero-dimers, related by a 2-fold symmetry, in which monomers interact with its two neighbour subunits through two different interfaces. Subunits from the hetero-dimer presenting the largest surface contact are related by a transformation of ~72° rotation and a small screw translation.

9) Two alternative conformations of a 19 residues loop, which locates at the interface between monomers, appear involved in the stability of the heterodimers.

BIBLIOGRAPHY

Bibliography

Beck, T., A. Krasauskas, T. Gruene and G. M. Sheldrick (2008). "A magic triangle for experimental phasing of macromolecules." Acta Crystallogr D Biol Crystallogr **64**(Pt 11): 1179-1182.

Berggard, T., S. Linse and P. James (2007). "Methods for the detection and analysis of protein-protein interactions." Proteomics **7**(16): 2833-2842.

Berrow, N. S., D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart and R. J. Owens (2007). "A versatile ligation-independent cloning method suitable for high-throughput expression screening applications." Nucleic Acids Res **35**(6): e45.

Biberfeld, G. and P. Biberfeld (1970). "Ultrastructural features of Mycoplasma pneumoniae." J Bacteriol **102**(3): 855-861.

Bordo, D. and P. Argos (1991). "Suggestions for "safe" residue substitutions in site-directed mutagenesis." J Mol Biol **217**(4): 721-729.

Bravman, T., V. Bronner, K. Lavie, A. Notcovich, G. A. Papalia and D. G. Myszka (2006). "Exploring "one-shot" kinetics and small molecule analysis using the ProteOn XPR36 array biosensor." Anal Biochem **358**(2): 281-288.

Bravman, T., V. Bronner, O. Nahshol and G. Schreiber (2008). "The ProteOn XPR36™ Array System—High Throughput Kinetic Binding Analysis of Biomolecular Interactions." Cellular and Molecular Bioengineering **1**(4): 216-228.

Brockhauser, S., R. B. Ravelli and A. A. McCarthy (2013). "The use of a mini-kappa goniometer head in macromolecular crystallography diffraction experiments." Acta Crystallogr D Biol Crystallogr **69**(Pt 7): 1241-1251.

Bruker (1998). Bruker AXS Inc., Madison, Wisconsin, USA.

Burgos, R., O. Q. Pich, M. Ferrer-Navarro, J. B. Baseman, E. Querol and J. Pinol (2006). "Mycoplasma genitalium P140 and P110 cytadhesins are reciprocally stabilized and required for cell adhesion and terminal-organelle development." J Bacteriol **188**(24): 8627-8637.

Burgos, R., O. Q. Pich, E. Querol and J. Pinol (2007). "Functional analysis of the Mycoplasma genitalium MG312 protein reveals a specific requirement of the MG312 N-terminal domain for gliding motility." J Bacteriol **189**(19): 7014-7023.
Burgos, R., O. Q. Pich, E. Querol and J. Pinol (2008). "Deletion of the Mycoplasma genitalium MG_217 gene modifies cell gliding behaviour by altering terminal organelle curvature." Mol Microbiol **69**(4): 1029-1040.

Calisto, B. M., A. Broto, L. Martinelli, E. Querol, J. Pinol and I. Fita (2012). "The EAGR box structure: a motif involved in mycoplasma motility." Mol Microbiol **86**(2): 382-393.

Chen, L., I. Annis and G. Barany (2001). Disulfide Bond Formation in Peptides. Current Protocols in Protein Science, John Wiley & Sons, Inc.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Cloward, J. M. and D. C. Krause (2009). "Mycoplasma pneumoniae J-domain protein required for terminal organelle function." Mol Microbiol **71**(5): 1296-1307.

Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." Acta Crystallogr D Biol Crystallogr **62**(Pt 9): 1002-1011.

Cowtan, K. (2008). "Fitting molecular fragments into electron density." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 83-89.

Cowtan, K. and P. Main (1993). "Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints." Acta Crystallogr D Biol Crystallogr **49**(Pt 1): 148-157.

Dauter, Z., M. Dauter and K. R. Rajashankar (2000). "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides." Acta Crystallogr D Biol Crystallogr **56**(Pt 2): 232-237.

Deacon, A. M., C. M. Weeks, R. Miller and S. E. Ealick (1998). "The Shake-and-Bake structure determination of triclinic lysozyme." Proc Natl Acad Sci U S A **95**(16): 9284-9289.

Doublie, S. (1997). "Preparation of selenomethionyl proteins for phase determination." Methods Enzymol **276**: 523-530.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.

Evans, P. (2011). "An introduction to data reduction: space-group determination, scaling and intensity statistics." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 282-292.
Fadouloglou, V. E., M. Kokkinidis and N. M. Glykos (2008). "Determination of protein oligomerization state: two approaches based on glutaraldehyde crosslinking." Anal Biochem **373**(2): 404-406.

Frank, B. (1889). "Uber dir Pilzsymbiose der Leguminosen." Bericht der Deutschen botanischen Gesellschaft **7**: 332-346.

Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd and J. C. Venter (1995). "The minimal gene complement of Mycoplasma genitalium." Science **270**(5235): 397-403.

Garnier, M., X. Foissac, P. Gaurivaud, F. Laigret, J. Renaudin, C. Saillard and J. M. Bove (2001). "Mycoplasmas, plants, insect vectors: a matrimonial triangle." C R Acad Sci III **324**(10): 923-928.

Garrett, D. S., Y. J. Seok, A. Peterkofsky, G. M. Clore and A. M. Gronenborn (1997). "Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system." Biochemistry **36**(15): 4393-4398.

Glass, J. I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, 3rd, H. O. Smith and J. C. Venter (2006). "Essential genes of a minimal bacterium." Proc Natl Acad Sci U S A **103**(2): 425-430.

Grabski, A., M. Mehler and D. Drott (2003). "The Overnight Express Autoinduction System: High-density cell growth and protein expression while you sleep." **2**(3): 233-235.

Green, D. W., V. M. Ingram and M. F. Perutz (1954). "The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method." Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **225**(1162): 287-307.

Hashimoto, Y., K. Yamada, H. Motoshima, T. Omura, H. Yamada, T. Yasukochi, T. Miki, T. Ueda and T. Imoto (1996). "A mutation study of catalytic residue Asp 52 in hen egg lysozyme." J Biochem **119**(1): 145-150.

Hasselbring, B. M. and D. C. Krause (2007). "Cytoskeletal protein P41 is required to anchor the terminal organelle of the wall-less prokaryote Mycoplasma pneumoniae." Mol Microbiol **63**(1): 44-53.

Hasselbring, B. M. and D. C. Krause (2007). "Proteins P24 and P41 function in the regulation of terminal-organelle development and gliding motility in Mycoplasma pneumoniae." J Bacteriol **189**(20): 7442-7449.

Hasselbring, B. M., E. S. Sheppard and D. C. Krause (2012). "P65 truncation impacts P30 dynamics during Mycoplasma pneumoniae gliding." J Bacteriol **194**(11): 3000-3007.

Hatchel, J. M. and M. F. Balish (2008). "Attachment organelle ultrastructure correlates with phylogeny, not gliding motility properties, in Mycoplasma pneumoniae relatives." Microbiology **154**(Pt 1): 286-295.

Henderson, G. P. and G. J. Jensen (2006). "Three-dimensional structure of Mycoplasma pneumoniae's attachment organelle and a model for its role in gliding motility." Mol Microbiol **60**(2): 376-385.

Hendrickson, W. A. and C. M. Ogata (1997). "Phase determination from multiwavelength anomalous diffraction measurements." Methods Enzymol **276**: 494-523.

Holm, L. and P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." Nucleic Acids Res **38**(Web Server issue): W545-549.

Hu, P. C., R. M. Cole, Y. S. Huang, J. A. Graham, D. E. Gardner, A. M. Collier and W. A. Clyde, Jr. (1982). "Mycoplasma pneumoniae infection: role of a surface protein in the attachment organelle." Science **216**(4543): 313-315.

Hu, P. C., A. M. Collier and J. B. Baseman (1977). "Surface parasitism by Mycoplasma pneumoniae of respiratory epithelium." J Exp Med **145**(5): 1328-1343.

Hutchison, C. A., S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith and J. C. Venter (1999). "Global transposon mutagenesis and a minimal Mycoplasma genome." Science **286**(5447): 2165-2169.

Jensen, J. S. (2006). "Mycoplasma genitalium infections. Diagnosis, clinical aspects, and pathogenesis." Dan Med Bull **53**(1): 1-27.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.

Jonsson, U., L. Fagerstam, B. Ivarsson, B. Johnsson, R. Karlsson, K. Lundh, S. Lofas, B. Persson, H. Roos, I. Ronnberg and et al. (1991). "Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology." Biotechniques **11**(5): 620-627.

Joosten, R. P., T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander and G. Vriend (2011). "A series of PDB related databases for everyday needs." Nucleic Acids Res **39**(Database issue): D411-419.

Jordan, J. L., H. Y. Chang, M. F. Balish, L. S. Holt, S. R. Bose, B. M. Hasselbring, R. H. Waldo, 3rd, T. M. Krunkosky and D. C. Krause (2007). "Protein P200 is dispensable for Mycoplasma pneumoniae hemadsorption but not gliding motility or colonization of differentiated bronchial epithelium." Infect Immun **75**(1): 518-522.

Kabsch, W. (1976). "A solution for the best rotation to relate two sets of vectors." Acta Cryst. **32**(5): 922-923.

Kabsch, W. (2010). "Integration, scaling, space-group assignment and post-refinement." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 133-144.

Kabsch, W. (2010). "XDS." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-2637.

Kantardjieff, K. A. and B. Rupp (2003). "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." Protein Sci **12**(9): 1865-1871.

Kay, L. E., M. Ikura, R. Tschudin and A. Bax (1990). "Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins." Journal of Magnetic Resonance **89**(3): 496-514.

Keller, R. and K. Wuthrich (2004). "Computer-aided resonance assignment (CARA)."

Kenri, T., S. Seto, A. Horino, Y. Sasaki, T. Sasaki and M. Miyata (2004). "Use of fluorescent-protein tagging to determine the subcellular localization of mycoplasma pneumoniae proteins encoded by the cytadherence regulatory locus." J Bacteriol **186**(20): 6944-6955.

Krause, D. C. (1996). "Mycoplasma pneumoniae cytadherence: unravelling the tie that binds." Mol Microbiol **20**(2): 247-253.

Krause, D. C., D. K. Leith and J. B. Baseman (1983). "Reacquisition of specific proteins confers virulence in Mycoplasma pneumoniae." Infect Immun **39**(2): 830-836.

Krause, D. C., T. Proft, C. T. Hedreyda, H. Hilbert, H. Plagens and R. Herrmann (1997). "Transposon mutagenesis reinforces the correlation between Mycoplasma pneumoniae cytoskeletal protein HMW2 and cytadherence." J Bacteriol **179**(8): 2668-2677.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-797.

Kunkel, T. A., J. D. Roberts and R. A. Zakour (1987). "Rapid and efficient site-specific mutagenesis without phenotypic selection." Methods Enzymol **154**: 367-382.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Cryst. **26**: 283-291.

Layh-Schmitt, G., A. Podtelejnikov and M. Mann (2000). "Proteins complexed to the P1 adhesin of Mycoplasma pneumoniae." Microbiology **146 (Pt 3)**: 741-747.

Liu, Q., T. Dahmane, Z. Zhang, Z. Assur, J. Brasch, L. Shapiro, F. Mancia and W. A. Hendrickson (2012). "Structures from anomalous diffraction of native biological macromolecules." Science **336**(6084): 1033-1037.

Liu, Q., Z. Zhang and W. A. Hendrickson (2011). "Multi-crystal anomalous diffraction for low-resolution macromolecular phasing." Acta Crystallogr D Biol Crystallogr **67**(Pt 1): 45-59.

Lluch-Senar, M., E. Querol and J. Pinol (2010). "Cell division in a minimal bacterium in the absence of ftsZ." Mol Microbiol **78**(2): 278-289.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McCoy, A. J. and R. J. Read (2010). "Experimental phasing: best practice and pitfalls." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 458-469.

McGuffin, L. J., K. Bryson and D. T. Jones (2000). "The PSIPRED protein structure prediction server." Bioinformatics **16**(4): 404-405.

Meng, K. E. and R. M. Pfister (1980). "Intracellular structures of Mycoplasma pneumoniae revealed after membrane removal." J Bacteriol **144**(1): 390-399.

Miyata, M. (2008). Molecular Mechanism of Mycoplasma Gliding - A Novel Cell Motility System. Cell Motility, Springer New York**:** 137-175.

Miyata, M. (2010). "Unique centipede mechanism of Mycoplasma gliding." Annu Rev Microbiol **64**: 519-537.

Morrison-Plummer, J., D. K. Leith and J. B. Baseman (1986). "Biological effects of anti-lipid and anti-protein monoclonal antibodies on Mycoplasma pneumoniae." Infect Immun **53**(2): 398-403.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.

Mushegian, A. R. and E. V. Koonin (1996). "A minimal gene set for cellular life derived by comparison of complete bacterial genomes." Proc Natl Acad Sci U S A **93**(19): 10268-10273.

Nakane, D., J. Adan-Kubo, T. Kenri and M. Miyata (2011). "Isolation and characterization of P1 adhesin, a leg protein of the gliding bacterium Mycoplasma pneumoniae." J Bacteriol **193**(3): 715-722.

Nakane, D. and M. Miyata (2007). "Cytoskeletal "jellyfish" structure of Mycoplasma mobile." Proc Natl Acad Sci U S A **104**(49): 19518-19523.

O'Connell, M. R., R. Gamsjaeger and J. P. Mackay (2009). "The structural analysis of protein-protein interactions by NMR spectroscopy." Proteomics **9**(23): 5224-5232.

Ohmura, T., T. Ueda, Y. Hashimoto and T. Imoto (2001). "Tolerance of point substitution of methionine for isoleucine in hen egg white lysozyme." Protein Eng **14**(6): 421-425.

Pape, T. and T. R. Schneider (2004). "HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs." Journal of Applied Crystallography **37**(5): 843-844.

Petsko, G. A. (1985). "Preparation of isomorphous heavy-atom derivatives." Methods Enzymol **114**: 147-156.

Phizicky, E. M. and S. Fields (1995). "Protein-protein interactions: methods for detection and analysis." Microbiol Rev **59**(1): 94-123.

Pich, O. Q., R. Burgos, M. Ferrer-Navarro, E. Querol and J. Pinol (2006). "Mycoplasma genitalium mg200 and mg386 genes are involved in gliding motility but not in cytadherence." Mol Microbiol **60**(6): 1509-1519.

Pich, O. Q., R. Burgos, M. Ferrer-Navarro, E. Querol and J. Pinol (2008). "Role of Mycoplasma genitalium MG218 and MG317 cytoskeletal proteins in terminal organelle organization, gliding motility and cytadherence." Microbiology **154**(Pt 10): 3188-3198.

Pitcher, D. G. and R. A. Nicholas (2005). "Mycoplasma host specificity: fact or fiction?" Vet J **170**(3): 300-306.

Proft, T., H. Hilbert, G. Layh-Schmitt and R. Herrmann (1995). "The proline-rich P65 protein of Mycoplasma pneumoniae is a component of the Triton X-100-insoluble fraction and exhibits size polymorphism in the strains M129 and FH." J Bacteriol **177**(12): 3370-3378.

Proft, T., H. Hilbert, H. Plagens and R. Herrmann (1996). "The P200 protein of Mycoplasma pneumoniae shows common features with the cytadherence-associated proteins HMW1 and HMW3." Gene **171**(1): 79-82.

Ramakrishnan, V., J. T. Finch, V. Graziano, P. L. Lee and R. M. Sweet (1993). "Crystal structure of globular domain of histone H5 and its implications for nucleosome binding." Nature **362**(6417): 219-223.

Razin, S., D. Yogev and Y. Naot (1998). "Molecular biology and pathogenicity of mycoplasmas." Microbiol Mol Biol Rev **62**(4): 1094-1156.

Regula, J. T., G. Boguth, A. Gorg, J. Hegermann, F. Mayer, R. Frank and R. Herrmann (2001). "Defining the mycoplasma 'cytoskeleton': the protein composition of the Triton X-100 insoluble fraction of the bacterium Mycoplasma pneumoniae determined by 2-D gel electrophoresis and mass spectrometry." Microbiology **147**(Pt 4): 1045-1057.

Relich, R. F. and M. F. Balish (2011). "Insights into the function of Mycoplasma pneumoniae protein P30 from orthologous gene replacement." Microbiology **157**(Pt 10): 2862-2870.

Roberts, D. D., L. D. Olson, M. F. Barile, V. Ginsburg and H. C. Krivan (1989). "Sialic acid-dependent adhesion of Mycoplasma pneumoniae to purified glycoproteins." J Biol Chem **264**(16): 9289-9293.

Rossmann, M. G. and D. M. Blow (1962). "The detection of sub-units within the crystallographic asymmetric unit." Acta Cryst. **15**: 24-31.

Seto, S., T. Kenri, T. Tomiyama and M. Miyata (2005). "Involvement of P1 adhesin in gliding motility of Mycoplasma pneumoniae as revealed by the inhibitory effects of antibody under optimized gliding conditions." J Bacteriol **187**(5): 1875-1877.

Seto, S., G. Layh-Schmitt, T. Kenri and M. Miyata (2001). "Visualization of the attachment organelle and cytadherence proteins of Mycoplasma pneumoniae by immunofluorescence microscopy." J Bacteriol **183**(5): 1621-1630.

Seybert, A., R. Herrmann and A. S. Frangakis (2006). "Structural analysis of Mycoplasma pneumoniae by cryo-electron tomography." J Struct Biol **156**(2): 342-354.

Shapiro, A. L., E. Vinuela and J. V. Maizel, Jr. (1967). "Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels." Biochem Biophys Res Commun **28**(5): 815-820.

Sheldrick, G. M. (2008). "A short history of SHELX." Acta Crystallogr A **64**(Pt 1): 112-122.

Sheldrick, G. M. (2010). "Experimental phasing with SHELXC/D/E: combining chain tracing with density modification." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 479-485.

Shen, Y., F. Delaglio, G. Cornilescu and A. Bax (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR **44**(4): 213-223.

Skubak, P. and N. S. Pannu (2011). "Reduction of density-modification bias by beta correction." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 345-354.

Stadtländer, C. T. K.-H. (2007). Scanning Electron Microscopy and Transmission Electron Microscopy of Mollicutes: challenges and opportunities. Modern Research and Educational Topics in Microscopy, FORMATEX. **I:** 122-131.

Strong, M., M. R. Sawaya, S. Wang, M. Phillips, D. Cascio and D. Eisenberg (2006). "Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis." Proc Natl Acad Sci U S A **103**(21): 8060-8065.

Sun, P. D. and S. Radaev (2002). "Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures." Acta Crystallogr D Biol Crystallogr **58**(Pt 7): 1099-1103.

Sun, P. D., S. Radaev and M. Kattah (2002). "Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases." Acta Crystallogr D Biol Crystallogr **58**(Pt 7): 1092-1098.

Szabo, A., L. Stolz and R. Granzow (1995). "Surface plasmon resonance and its use in biomolecular interaction analysis (BIA)." Curr Opin Struct Biol **5**(5): 699-705.

Taylor, G. L. (2010). "Introduction to phasing." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 325-338.

Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallogr D Biol Crystallogr **56**(Pt 8): 965-972.

Tully, J. G., D. Taylor-Robinson, R. M. Cole and D. L. Rose (1981). "A newly discovered mycoplasma in the human urogenital tract." Lancet **1**(8233): 1288-1291.

Vagin, A. and A. Teplyakov (1999). "MOLREP: an Automated Program for Molecular Replacement." Journal of Applied Crystallography **30**: 1022-1025.

Vagin, A. and A. Teplyakov (2000). "An approach to multi-copy search in molecular replacement." Acta Crystallogr D Biol Crystallogr **56**(Pt 12): 1622-1624.

Vagin, A. A. and M. N. Isupov (2001). "Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps." Acta Crystallogr D Biol Crystallogr **57**(Pt 10): 1451-1456.

Walter, T. S., C. Meier, R. Assenberg, K. F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart and J. M. Grimes (2006). "Lysine methylation as a routine rescue strategy for protein crystallization." Structure **14**(11): 1617-1622.

Weber, K. and M. Osborn (1969). "The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis." J Biol Chem **244**(16): 4406-4412.

Wider, G. and K. Wuthrich (1999). "NMR spectroscopy of large molecules and multimolecular assemblies in solution." Curr Opin Struct Biol **9**(5): 594-601.

Willby, M. J., M. F. Balish, S. M. Ross, K. K. Lee, J. L. Jordan and D. C. Krause (2004). "HMW1 is required for stability and localization of HMW2 to the attachment organelle of Mycoplasma pneumoniae." J Bacteriol **186**(24): 8221-8228.
Wilson, W. D. (2002). "Tech.Sight. Analyzing biomolecular interactions." Science **295**(5562): 2103-2105.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Wuthrich, K. (1990). "Protein structure determination in solution by NMR spectroscopy." J Biol Chem **265**(36): 22059-22062.