Ana Catarina da Silva Pereira

BS Biochemistry

Structural investigation of the *Bacillus subtilis* morphogenic factor RodZ

A thesis to obtain a Master degree in
Structural and Functional Biochemistry

Supervisor: Doctor Manolis Matzapetakis
Principal Investigator, Biomolecular NMR Group, ITQB, Oeiras, Portugal

Board Members

Committee Chair: Doctor José Ricardo Ramos Franco Tavares
Assistant Professor, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Examiner: Doctor Jorge da Silva Dias
Assistant Investigator, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Nova University of Lisbon

Lisbon, Portugal

2013

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Ana Catarina da Silva Pereira

BS Biochemistry

Structural investigation of the *Bacillus subtilis* morphogenic factor RodZ

A thesis to obtain a Master degree in
Structural and Functional Biochemistry

Supervisor: Doctor Manolis Matzapetakis
Principal Investigator, Biomolecular NMR Group, ITQB, Oeiras, Portugal

Board Members

Committee Chair: Doctor José Ricardo Ramos Franco Tavares
Assistant Professor, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Examiner: Doctor Jorge da Silva Dias
Assistant Investigator, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Nova University of Lisbon

Lisbon, Portugal

2013

Acknowledgments

I would have never been able to finish my dissertation without the guidance of
my supervisor, help from friends, and support from my family.

After this time I do come to a conclusion that I overcame major doubts about myself and my work. This would have never been possible without the help of my supervisor, Manolis Matzapetakis, who invited me to work again with him knowing beforehand about my limitations but also recognizing and encouraging my major strengths. I would like to express my sincere gratitude for his continuous support, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Also, knowing that it could delay the progress of this project, he met my request to get my hands "dirty" on the wet lab and by that allowing me to gain new molecular biology skills, which I thank for.

I must mention and acknowledge Adriano Henriques and the Microbial Development Laboratory members Teresa Costa and Ana Paiva for providing all the required RodZ samples and detailed information of undergoing developments, without which I couldn't have written this report.

I thank my fellow colleagues from the Biomolecular NMR laboratory: Meire Almeida, Mariana Palma and Ivo Saraiva, for the stimulating discussions during the many group meetings we had for the past year. Meire Almeida was also a key person during my training in the wet lab, being always available to help me solve any question.

My sincere thanks also goes for Claudia Almeida for helping me with equipment training and Isabel Pacheco and all members from NMR and Inorganic Chemistry laboratory for always being available to help me in any way possible.

I would like to thank my family for standing by me through good and bad times: my parents Leonor and Aires Pereira for giving me emotional and financial support, allowing me to complete another cycle of studies; my sister Márcia Pereira whose organizational skills were crucial to help me finish writing this thesis.

I would also like to thank Vicente Canhoto for always being there to cheer me up and for always supporting me and encouraging me with his best wishes.

I couldn't finish without mentioning my friends and college colleagues, which I have to thank for their time and patience.

Abstract

RodZ is a protein widely conserved in bacteria and a core component of the morphogenic apparatus of the cell. It is known to be required for assembly of the bacterial actin homologue, MreB, that controls cell wall synthesis and cell shape. The domain organization of RodZ consists of a well-conserved N-terminal (RodZn) with helix-turn-helix motif (HTH), a conserved transmembrane domain, and a conserved C-terminal domain (RodZc). RodZn, located in the cytoplasm, has been shown to interact with MreB actin-homologue by x-ray studies in *T. maritima*. However, the structure of RodZn from gram-positive *B. subtilis* showed low homology with the published one from gram-negative *T. maritima*. Here we present the solution structure of RodZn from *B. subtilis* determined for the first time, by NMR spectroscopy. Compared to previous structural data obtained from the crystallized RodZn from *T. maritima* and more recently from *S. aureus*, several differences could be observed, namely the length of the alpha-helices and the presence of an extended coil. Interaction studies were preformed between RodZn domain and MreB from which no significant results could be extrapolated. Since HTH motif is frequently associated with DNA interaction, the involvement of RodZn in DNA organization is being investigated. At the same time, RodZc domain, which structure has never been reported, was subject of study. Bioinformatic, biophysical and biochemical methodologies were employed to study this domain. A model based in a pseudo-*ab initio* methodology was built, revealing an Ig-like fold. The Ig superfamily is a large group of cell surface and soluble proteins that are involved in the recognition, binding, or adhesion processes of cells. Therefore, RodZ is thought to be a protein that establishes a link between the inner side of the cell membrane and the outer side, promoting spatiotemporal coordination between peptidoglycan synthesis and cell division.

Key words: *B. subtilis*, Dynamics, Homology Modelling, HTH motif; Ig-like fold, NMR.

Resumo

RodZ é uma proteína amplamente conservada em organismos bacterianos, fazendo parte do complexo aparelho morfogénico celular. Em estudos anteriormente publicados, RodZ foi considerada necessária à organização celular da proteina MreB, um homólogo da actina igualmente conservado em bactérias, responsavél por controlar a síntese da parede celular e a morfologia adoptada pela célula . A proteina RodZ é composta por três domínios distintos: um domínio N-terminal (RodZn) detentor de um motivo estrutural de hélice-volta-hélice (HTH) localizado no citoplasma, um domínio transmembranar (TM), e um domínio C-terminal (RodZc) localizado na região periplasmática da célula. Em *T. maritima* (gram-negativo), a interacção entre o domínio RodZn e a proteina MreB foi demonstrado através de estudos de cristalografia de raios-x. No entanto , a estrutura do domínio RodZn presente em *B. subtilis* (gram-positivo) revelou uma baixa homologia com a estrutura publicada. No presente trabalho, foi determinada pela primeira vez através de Espectroscopia de RMN, a estrutura em solução do domínio RodZn de *B. subtilis*,. Comparando-a com os dados estruturais anterioremente publicados a partir do domínio RodZn presente em *T. maritima* e, mais recentemente, em *S. aureus*, várias diferenças foram observadas , nomeadamente, o comprimento das hélices alfa e a presença de uma zona alongada não estruturada. Estudos de interacção entre o domínio RodZn e a proteina MreB não levaram a resultados significativos. Sendo que o motivo estrutural HTH é frequentemente associado ao estabelecimento de interacções com a molécula de DNA, o envolvimento de RodZn em organização do DNA plasmídico encontra-se a ser investigado. Ao mesmo tempo , o domínio RodZc , cuja estrutura nunca antes fora revelada, foi objecto de estudo. Ferramentas bioinformáticas conjugadas com técnicas biofísicas e bioquímicas foram utilizadas para estudar este domínio. Foi construido um modelo tridimensional para o domínio RodZc. A metodologia usada baseou-se na utilização de ferramentas de modelação com um princípio de pseudo-*ab initio*, revelando um motivo estrutural rico em folhas beta, característico da super-família de imunoglobulinas. Esta super-família é composta por um variado grupo de proteinas que embora possuam um motivo estrutural semelhante, variam bastante em termos de função celular (processos de reconhecimento, ligação e adesão celular, entre outras funções). Sendo assim, uma hipótese foi levantada em que a proteina RodZ possa ser o elo de ligação entre o lado interno e o lado externo da parede da célula promovendo, desta forma, a coordenação espaço-temporal entre a síntese do peptidoglicano e a divisão celular.

Palavras-chave: *B. subtilis*, Dinânimca, Modelação por homologia, motivo HTH, super-família de imunoglobulinas; RMN.

INDEX

LIST OF FIGURES

LIST OF TABLES

ABREVIATIONS

^{1}HA	Also Hα; proton atom attached to CA
^{1}HB	Also Hβ; proton atom attached to CB
^{1}NH	Also NH or HH; proton atom attached to N
^{13}CA	Also CA or Cα; carbon alpha from protein backbone
^{13}CB	Also CB or Cβ; carbon beta from protein side-chain
^{13}CO	Also C'; carbonyl carbon from protein backbone
^{15}N	Also N; nitrogen atom
ASEC	Analytical Size Exclusion Chromatography
BSA	Bovine serum albumin
CD	Circular Dichroism Spectroscopy
D, ASP	Aspartic Acid
Da	Dalton
ddH2O	Bi-distilled Water
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
g	g force
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lisogeni - Broth
M	Molar
mAU	mili units of absorbance
MW	Molecular Weight
MWCO	molecular weight cut off
NMR	Nuclear Magnetic Resonance
O.D.	Optical Density
PMSF	phenylmethylsulfonyl fluoride
ppm	Parts per million
rpm	Rotations per minute
RT	Room Temperature
SDS- PAGE	Polyacrylamide gel electrophoresis in denaturing conditions
Tc	Also τ_c; Rotational correlation time

Chapter 1

Biological significance

Almost all bacteria are surrounded by a giant cell wall. The bacterial cell wall is a complex three-dimensional structure mainly composed by a peptidoglycan layer (PG) that plays a central role in the maintenance of bacterial shape, prevention of cellular stress due to differences in osmotic pressure, and formation of daughter cells during cell division. This biopolymer also known as murein is an important target of several classes of antibiotics. The clinical value of many of these antibiotics diminishes in the face of an increasing prevalence of various resistance mechanisms. As more is learned about it, new roads for research are being opened for novel drug targets and design of antimicrobials. In this light, there has been a huge effort in the science community to understand the highly complex enzymatic machinery that synthesizes the cell wall and how its activity is coordinated with cell growth and division. Spherical-shaped bacteria only synthesize PG at the plane of division (the septum), which bulges out after cell division leaving a round cell. Cells with a more complex shape, such as rod-shaped cells, exhibit an additional growth mode responsible for cell elongation. The rod-shaped bacteria such as *Bacillus subtilis* has been extensively used as a model in cell wall synthesis studies (**reviewed by Scheffers and Pinho, 2005, and Young, 2010**).

Although the existence of cell wall division-specific and elongation-specific factories has not been conclusively shown yet and their precise composition remains unknown this machineries are hypothesized to exist. The latter has been extensively studied and through the years some key proteins have been reported, namely: MreB, MreC, MreD and RodA. More recently, RodZ has been discovered to be a common player in bacterial cell morphogenesis (Gerdes et al., 2009).

In this opening chapter I will attempt to summarize the wide range of information that is now available about the bacterial cytoskeleton, focusing on the Gram positive *B. subtilis* organism whose bacterial cell wall machinery is addressed in this report. Other Gram negative rod-shaped bacteria, such as *E. coli* and *C. crescentus*, will be mentioned for comparison purposes. The structure of the peptidoglycan of *E. coli* and *C. crescentus* and *B. subtilis* is very similar except for a few minor modifications. For instance, the major cell wall constituents, such as teichoic acids, are not attached to the peptidoglycan in *E. coli* and in *C. crescentus*, contrary to *B. subtilis*. The thickness of the cell walls in *B. subtilis* and the presence of an additional cell membrane in *E. coli* and in *C. crescentus* are other important differences in the cell wall (**Figure 1**). *E. coli* and *C. crescentus* possess an inner and outer membrane, with the space between the membranes (called the periplasmic space) containing one to two layers of peptidoglycan. In contrast, *B. subtilis* do not possess an outer membrane, but contains a thick peptidoglycan with 10-20 layers. The structure of the peptidoglycan though is very similar in both *E. coli* and *B. subtilis* (**Foster et al., 2002; Vollmer et al., 2008a; Vollmer et al., 2008b; Archibald et al., 2002**).

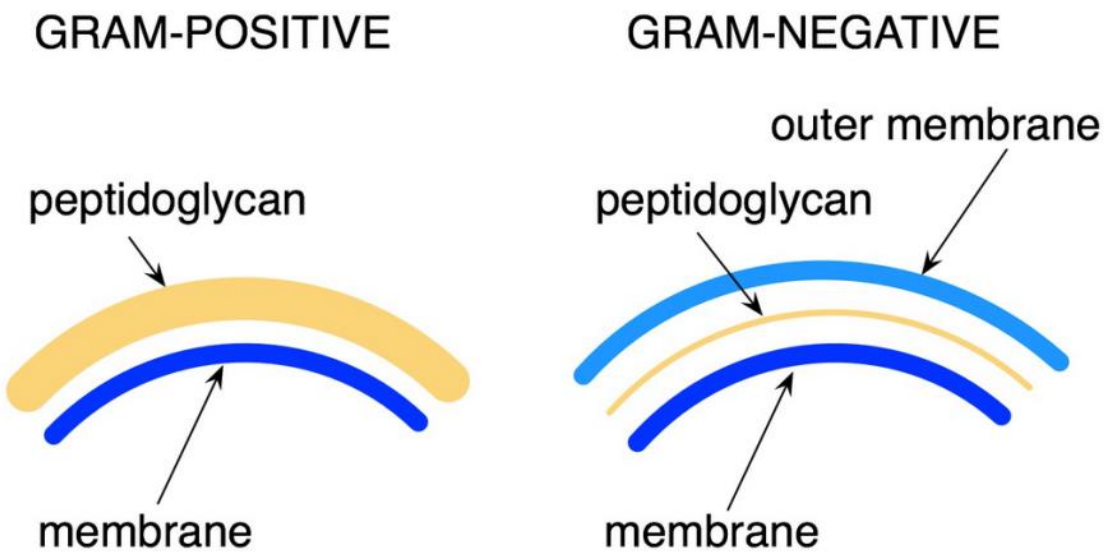

Figure 1 – Gram-positive and Gram-negative bacteria are differentiated by their cell wall structure; Picture comparing Gram-positive and -negative cell wall.

It will be covered some details of the biological role of specific proteins that are related with the regulation of the cell wall synthesis, emphasizing on the relationship between their cellular position/organization with their specific function. When applicable, I will refer recent reviews on these topics.

1.1 Cell Wall

The biosynthesis of the main component of the cell wall, the peptidoglycan, involves the coordination of the activity of proteins present in the cytoplasm, the membrane, and the periplasm. The high complexity of the cell wall elongation and division process has created a significant challenge for the study of the macromolecular interactions that regulate peptidoglycan biosynthesis. The availability of new structural and biochemical data on a number of components of peptidoglycan assembly machineries now provide novel insight into the basis of a complex molecular machinery.

The peptidoglycan layer is formed by glycan strands of repeating disaccharide residues, cross-linked via peptide side chains (**Archibald et al., 1993**). It has a dynamic structure, continuously being synthesized, modified, and hydrolyzed to allow for cell growth and division, among many other roles (**Foster et al., 2002**). In gram-positive bacteria such as *B. subtilis*, PG is presented as a thick layer to which teichoic acids and cell wall specific proteins are covalently bound. Peptidoglycan synthesis in *B. subtilis* is summarized in **Figure 2**. Precursors are synthesized in the cytoplasm, linked to the transport lipid and flipped across the inner membrane followed by attachment of the newly synthesized chain (**Bhavsar et al., 2006**).

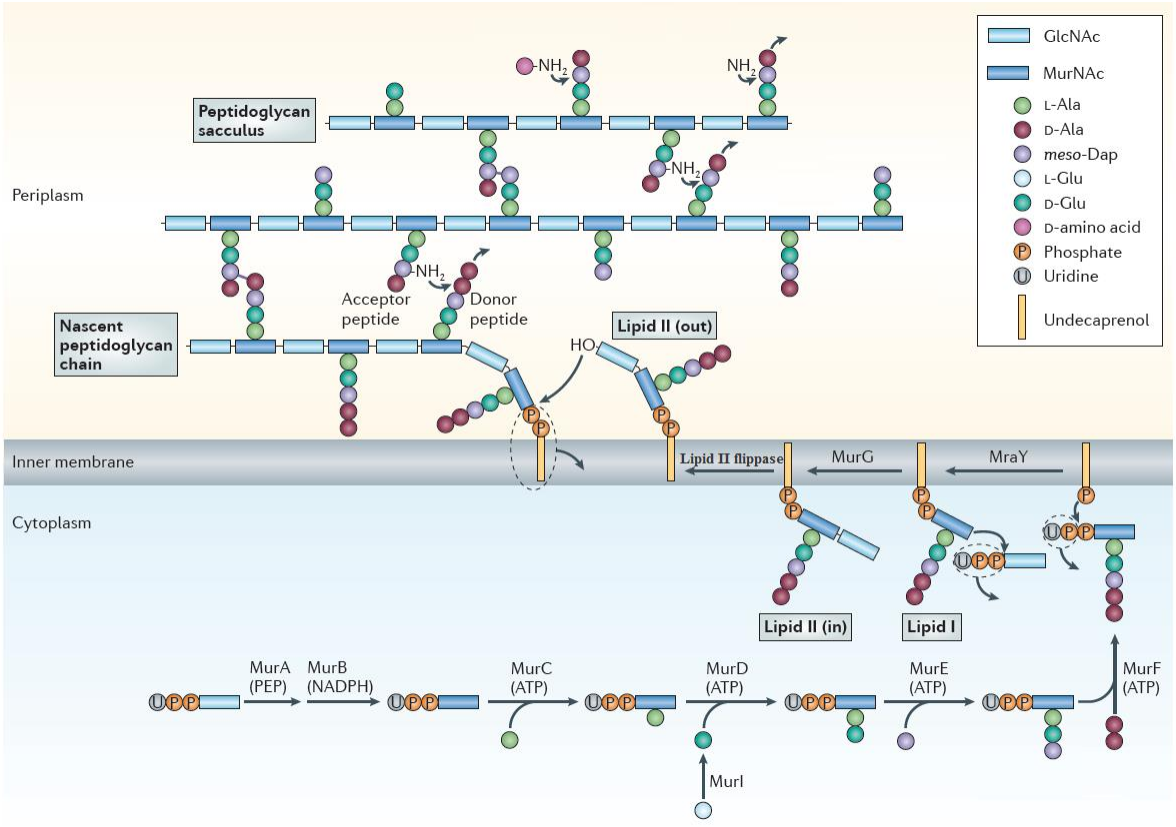

Figure 2 - Peptidoglycan synthesis in *B. subtilis*. The peptidoglycan layer is formed by polymerized chains of repeating disaccharide subunits named GlcNAc (Nacetylglucosamine) and MurNAc (N-acetylmuramic acid) cross-linked by peptides (Holtje et al., 1998; den Blaauwen et al., 2008). Peptidoglycan synthesis occurs

in several stages. Initially, the UDP-MurNAc-pentapeptide precursor is synthesized in six cytoplasmic reactions catalysed by the MurA to MurF synthetases. MraY transferase then catalyses the reaction of this precursor with the membrane acceptor, undecaprenyl phosphate, to yield lipid I. Lipid II, which comprises the complete disaccharide-pentapeptide unit, is formed by the addition of N-acetylglucosamine to lipid I in a reaction catalysed by MurG. Lipid II is then transferred to the outside of the membrane by lipid II flippase. The next stage of peptidoglycan synthesis involves polymerization reactions on the outside surface of the cytoplasmic membrane catalysed by penicillin binding proteins (PBPs) and the incorporation of the newly formed material into the existing peptidoglycan by lytic transglycosylases (MltA). Not all intervenients are characterized in this scheme. Notes: meso-Dap, meso-diaminopimelic acid; MraY, UDP-MurNAc-pentapeptide phosphotransferase; MurA, UDP-GlcNAc enolpyruvyl transferase; MurB, UDP-MurNAc dehydrogenase; MurC, UDP-MurNAc–l-Ala ligase; MurD, UDP-MurNAc-l-Ala–d-Glu ligase; MurE, UDPMurNAc- l-Ala-d-Glu–meso-Dap ligase; MurF, UDP-MurNAc-tripeptide–d-alanyl-d-Ala ligase; MurG, UDP-GlcNAcundecaprenoyl- pyrophosphoryl-MurNAc-pentapeptide transferase; MurI, Glu racemase; PEP, phosphoenolpyruvate. Adapted from Typas et al., 2011.

The rod shape of the *B. subtilis* cell is maintained during its whole life cycle, being clear that all factors that control cell shape must be present in all phases of its growth. The coordinated action of two mechanisms of cell wall synthesis, one specific for cell elongation and the other for cell division is thought to be responsible for maintaining the rod shape. During cell division, the tubulin homologue FtsZ is the main player, whereas elongation is driven by the actin-homologue MreB and its paralogues MreBH and Mbl (**Reviewed by Young, 2010**).

1.2 Cell elongation and division

Cell division in most bacteria is carried out by a contractile protein ring, known as the divisome, which is made up of about a dozen different polypeptides. This sophisticated macromolecular machine, which is centred on FtsZ, is capable of promoting the coordinated invagination of the cell membrane and cell wall to create the division septum (**Adams and Errington, 2009**).

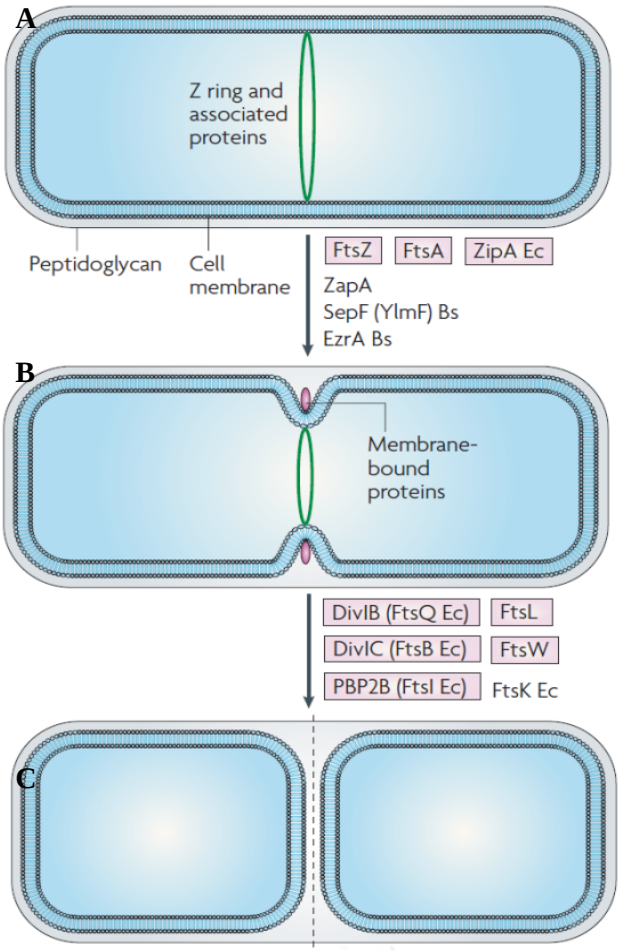

Figure 3 - Cell division in Gram-positive rod-shaped bacteria. A) Z ring formation occurs at midcell, and recruits various FtsZ-binding proteins. B) The membrane-bound cell-division proteins are then recruited, resulting in invagination of the cell wall and membrane to form a division septum. C) Septum formation is complete and peptidoglycan hydrolases hydrolyse the completed cross wall, producing two newborn cells. Proteins listed are the major ones known for *B. subtilis* and/or *E. coli*, and those in purple are potential targets. Adapted from Rowena et al., 2008.

While cocci rely exclusively on the division machinery driven by FtsZ to grow as spheres, with each division producing two new hemispheres, rod-shaped bacteria like *B. subtilis* add an elongation phase before each division. The actin homolog MreB typically plays an essential role in this elongation stage by maintaining a constant cell width (**Jones et al., 2001; Figge et al., 2004; Gitai et al., 2005; Kruse et al., 2003;Cabeen et al., 2011; den Blaauwen., 2008; White et al., 2011**).

During the elongation stage, MreB form helical structures that are thought to guide the insertion of new peptidoglycan (PG) cell wall along the cell circumference. When MreB function is lost, cells become progressively larger as they grow and adopt spheroid morphology over time (**Jones et al., 2001; Kawai et al., 2009**).

How this occurs is not fully understood, but other proteins of the core morphogenic apparatus, such as MreC, MreD, and RodA are likely to be involved. The latter are essential membrane proteins

with both cytoplasmatic and periplasmatic domains thought to be part of this multienzyme complex, mediating peptidoglycan synthesis on the lateral walls of *B. subtilis* cells. (**Henriques et al., 1998; White et al., 2011; Dominguez-Escobar et al., 2011;Garner et al., 2011**). This hypothesis was supported by the observation of MreC forming helical structures that alternate with the MreB helices (**Dye et al., 2005**). And experiments showed that MreC interacts with the penicillin-binding proteins that synthesize the cell wall. These results raised the possibility that the MreB filaments interact with MreCD complexes located in the inner cell membrane and thereby control the activity of the external cell wall-synthesizing protein complexes (**Divakaruni et al., 2007; Van den Ent et al., 2006**). This link between peptidoglycan synthesis and the cytoskeletal system was also confirmed by the observation of interactions between MurG and MreB in *E. coli* (**Mohammadi et al., 2007**). It was also shown that MurG localization is dependent on MreB in *C. crescentus* (**Divakaruni et al., 2007**). The interaction of MraY with MreD and the dependence of its localization on MreB in *C. crescentus* indicate that the morphogenic proteins MreD and MreB play a role in the organization of cell wall synthesis complexes (**White et al., 2010**).

While the exact function of these membrane-bound proteins remains somewhat unclear, evidence suggests that they regulate PG growth by linking MreB to cell wall enzymes or by working in concert with MreB to spatially restrict cell wall activities (**Leaver et al., 2005; Levin et al., 1992; Varley et al., 1992; Wagner et al., 2005; Divakaruni et al., 2007; Kruse et al., 2005; Dye et al., 2005**). The first observations of MreB were performed through fluorescence microscopy, which showed MreB filaments forming bundles moving continuously through growing *B. subtilis* cells (**Defeu et al., 2004**). However, recent reports came out suggesting that all three MreB paralogs would rather form patches moving independently (**Courtney et al., 2012**).

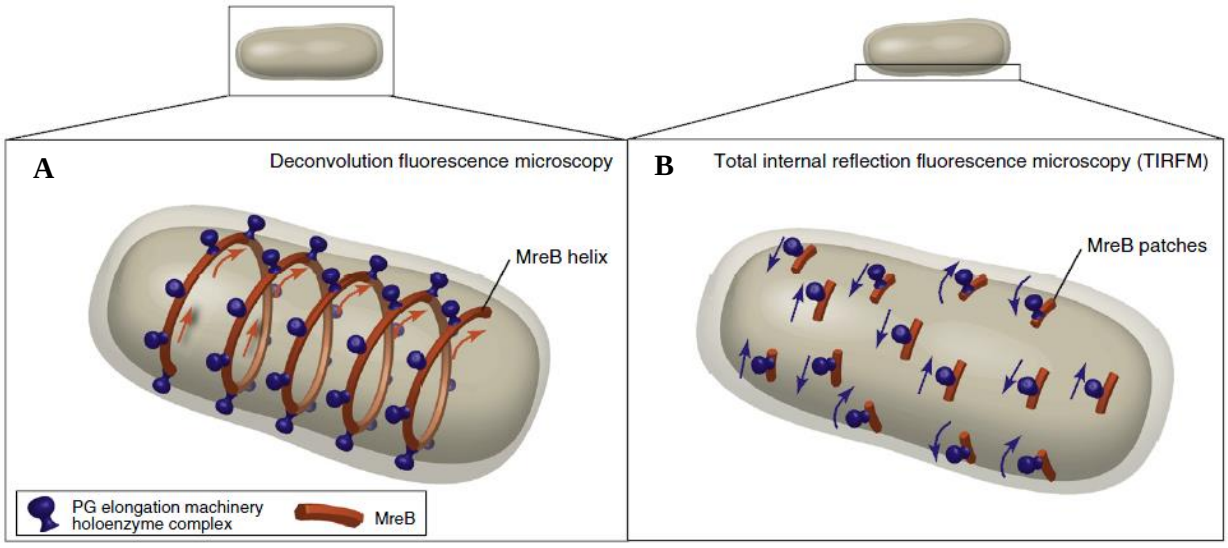

Figure 4 - Cartoon comparing the two microscopic techniques used to visualize MreB structure and movement. MreB is seen in both cases coupled with the peptidoglycan (PG) elongation machinery, which

represents cell wall synthetic enzymes and cell shape determining proteins. In A) is represented what is seen using decovolution fluorescence microscopy, where a stack of images taken through the cell body described MreB to form an helical structure along the bacterial cell wall. In B) through TIRFM microscopy, a high resolution technique capable of capturing the surface of one side of a bacterium found both MreB and a selection of several PG elongation proteins that move in short patches as opposed to long helical filaments with bidirectional motility. Schematic representations are not drawn to scale. Adapted from Courtney et al., 2012.

These observations along with the dynamic directional MreB movement led to a model where MreB serves as a spiral track spanning the cell length, acting as a scaffold to organize cell wall synthesis.

1.3 Protein RodZ

Any missing players of the core morphogenic apparatus would represent a significant limitation to our understanding of cell morphogenesis. However, in recent years a new common player in bacterial cell morphogenesis has been discovered (**Shiomi et al., 2008; Bendezu et al., 2008; Alyahya et al., 2009**).

Named RodZ, this morphogenic factor is widely conserved in bacteria, as shown in **Figure 5.** Its broad conservation across bacterial phyla may indicate an important and ancient function in cell shape determination. Depletion or disruption of RodZ has been shown to result in misshapen cells in rod-shaped bacteria such as Caulobacter crescentus, *E. coli*, Shigella sonnei and *B. subtilis* (**Alyahya et al., 2009; Bendezu et al., 2008; Shiomi et al., 2008; Mitobe et al., 2011**).

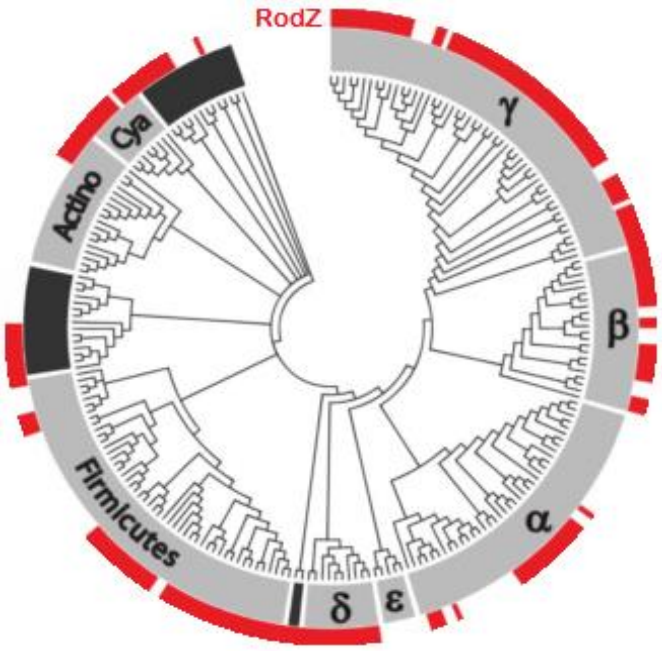

Figure 5 – Phylogenetic tree of bacterial species showing the conserved presence of RodZ (inred). Subtrees indicate phyla, except for the proteobacterial phylum, which is further subdivided by class (Alyahya et al., 2009).

Protein RodZ from *B. subtilis* is encoded in ymfM gene and is composed by 304 residues. The domain organization of RodZ consists of a well-conserved N-terminal helix-turn-helix domain (HTH; residues 19-90), a conserved transmembrane domain (TM; residues 102-132) rich in hydrophobic residues, and a conserved C-terminal domain (residues 204-304) whose structure has not been determined. The conserved domains are capped by N and C terminal extensions and are separated by cytoplasmic and periplasmic linkers of variable length (29 and 117 residues, respectively). The linker separating the conserved C-terminal domain from the TM is enriched in prolines and small residues, such as glycine and alanine (**Gerdes et al., 2009**). The domain organization of RodZ by itself has led to the proposal that this protein could provide a direct link between the cytoplasmatic and periplasmatic peptidoglycan elongation machinery (**Alyahya et al., 2009; Bendezu et al., 2009; Shiomi et al., 2008; Mitobe et al., 2011**).

RodZ exhibits a localization pattern during the cell cycle corresponding to sites of active peptidoglycan synthesis. The temporal transition of RodZ from uniformly distributed patched-pattern to mid-cell localization depends on the actin-like MreB cytoskeleton. More recently the cytoplasmic domain of RodZ has been, in fact, reported to be required for assembly of MreB. And the cytoplasmic domain of RodZ (RodZn) has been shown to interact with MreB by functional and crystallographic studies in Thermotoga maritima as shown in **Figure 6** (**Van den Ent et al., 2010**).

The interaction of the cytoplasmic part of RodZ with MreB enforces the hypothesis that this protein may be an additional transmembrane stabilizing factor of the bacterial cell wall elongation complex.

However, the cytoplasmic domain of RodZ form of Bacillus subtilis has a low homology with the published one from Thermotoga maritima (around 22%) (**Gerdes et al., 2009; Mattei et al., 2010**). And there is no structural evidence of direct interaction between RodZ and MreB in *B. subtilis* (**Alyahya et al., 2009**).

E. coli cells lacking the rodZ gene turned round or misshapen and exhibited a highly reduced growth rate. Cell width is maintained by the MreBCD and PBP2/RodA complexes and the diameters of the majority of the rodZ-null cells were similar to that of the width of wild-type. Therefore it was suggested that RodZ is a primary determinant of cell length. In *Caulobacter crescentus*, RodZ is essential for viability and is involved in all aspects of this organism's complex morphology (**Alyahya et al., 2009**). Overproduction of RodZ resulted in an increased cell length with little or no change of cell width, consistent with previous results. On the other hand, co-overexpression of MreB and RodZ showed

that maintenance of cell shape depended critically on a proper MreB/RodZ ratio. (**Shiomi et al., 2008, Bendezu et al., 2009**).

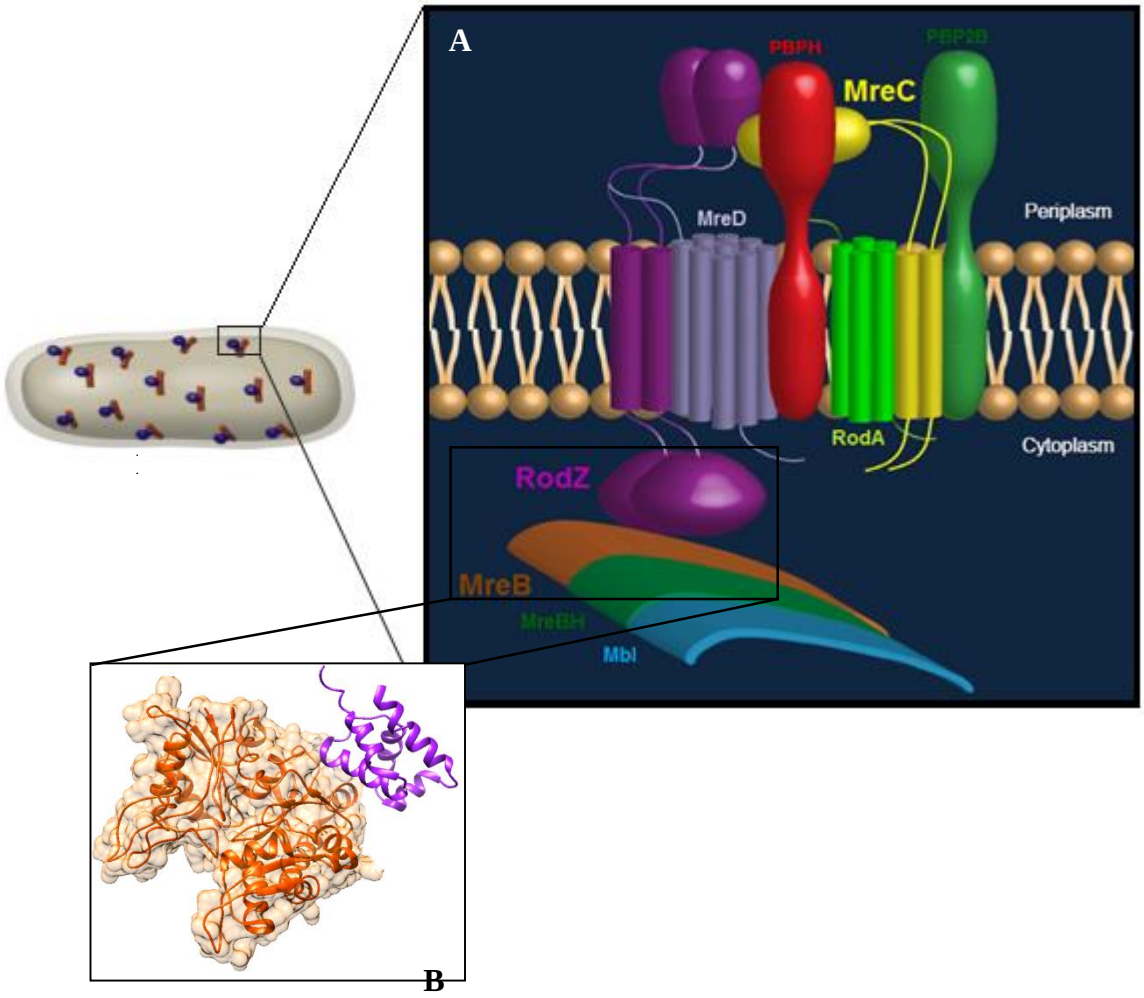

Figure 6 – A) Cartoon of the proposed PG elongation machinery in Gram-positive rod-shaped bacteria like *B. subtilis*. Peptidoglycan layer is not shown. The elongation machinery is represented by proteins shown in various colours. Not all protein from the complex are portrayed. Adapted from Ana Paiva, 7th International Conference on Gram-positive Microorganisms Poster (June, 2013); Scheme credits: Patrícia Amaral); B) Xray structure of MreB complexed with the cytoplasmic domain of RodZ (RodZn) from rod-shaped Gram-negative *T. maritima*; PDB code: 2WUS.

The cytoplasmic domain of RodZ (RodZn) alone is required for proper localization of the protein and mutations that affect the interaction between RodZn and MreB also result in mislocalization of the protein and cell shape defects (**Van den Ent., 2010**). In *B. subtilis* cells expressing RodZ without the cytoplasmic domain, the remain protein lost its characteristic localization pattern (patches), becoming uniformly distributed along the cell membrane, from cell division septa to the poles of the cell, a location

rarely seen for the wild type (**unpublished data from The Microbial development Laboratory from ITQB**). The function of the transmembrane domain of RodZ (TM) has also been investigated. Through mutation of TM domain, this resulted in the production of short and wide rod cells, with asymmetric division. Besides, the mutated RodZ protein was shown to be distributed along the cell membrane, with only a few cells displaying the patch-like pattern. These observations led to the conclusion that the TM domain may be essential for the function and localization of *B. subtilis* RodZ.

As for the periplasmic domain of RodZ (RodZc), this domain is also conserved and is considered to have an important role in cell shape control, possibly through interactions with extracytoplamic components of the cell wall elongation machinery, such as MreC and the PBPs (**Alyahya et al., 2009; Shiomi et al., 2008; Bendezu et al., 2009**).

Unpublished results from the Microbial development laboratory from ITQB also showed that in *B. subtilis*, rodZ mutants lacking the periplasmic domain led to enlarged cells, roundish or with irregular shapes. And cells that still remained rod-shaped were shorter and wider than those of the wild type, similar to those of the rodZ deletion mutants. Proper localization of RodZ requires its cytoplasmic domain via interaction with MreB. But these new studies indicate that the periplasmic domain may also contribute to the sub-cellular localization of RodZ, thus inferring that interactions on either side of the cell membrane are required for the correct sub-cellular positioning of RodZ.

So, both RodZn and TM domains are essential to maintain the normal pattern of midcell division, and the right positioning of the nucleoid. Nevertheless, the overall rod-shape of the cell is maintained in mutants lacking these two domains. In contrast, deletion of RodZc also causes a severe change in the cell shape phenotype. So, the presence of the periplasmic domain in most RodZ orthologues suggests an important function but no meaningful hint about its role has been published so far (**Alyahya et al., 2009**).

Studies in *E. coli* have revealed direct interactions of RodZ with MreC and possibly with MreD and the cell wall elongation PBP2 (**Bendezu et al., 2009**). Therefore, RodZ seems to interact with proteins of the cell elongation machinery on both sides of the cytoplasmic membrane. MreC and PBPs are likely to interact with the periplasmic portion of RodZ (RodZc).

In the end, cell wall synthesis, breakdown and insertion activities seem to be directed by MreC in the periplasm, MreD, RodA and RodZ in the inner membrane, and MreB **Courtney et al., 2012**). The discovery of this highly conserved morphogenic factor opened new possibilities in the difficult but essential analysis of the bacterial cell wall puzzle.

1.4 Bacillus subtilis

B. subtilis has approximately 4,100 genes. Of these, only 192 were shown to be indispensable and 79 were predicted to be essential as well. A vast majority of essential genes were categorized in relatively few domains of cell metabolism, with about half involved in information processing, one-tenth related to cell energetics and one-fifth involved in the synthesis of cell envelope and the determination of cell shape and division (**Kunst et al., 1997; Kobayashi et al., 2003**).

Originally named *Vibrio subtilis* in 1835 by Christian Gottfried Ehrenberg, this organism was renamed by Ferdinand Cohn *Bacillus subtilis* in 1872 (**Ehrenberg, 1835; Cohn, 1872**). Bacillus subtilis is a bacterial microorganism commonly found in the environment, mainly in soil, being categorized as a saprophyte organism (**Brock et al., 2005**). Nevertheless, this microorganism is well known by modern science to be very friendly to the human system, being able to promote dramatic healing benefits, even though it isn't one of the native microbes that normally inhabit the human body (**Hong et al., 2009**). Even though *B. subtilis* has been historically classified as a strictly aerobe microorganism, recent research shows that this species can actually leave under anaerobic conditions (**Nakano et al., 1998**).

Similar to Gram-negative Escherichia *coli* and Caulobacter *crescentus* bacteria, Gram-positive *B. subtilis* has a rod-shape, as shown in **Figure 7** , being 3-5 μm long, of about 1 μm width (**Sargent, 1975**) and with hemispherical cell poles (**Burdett et al., 1978**).

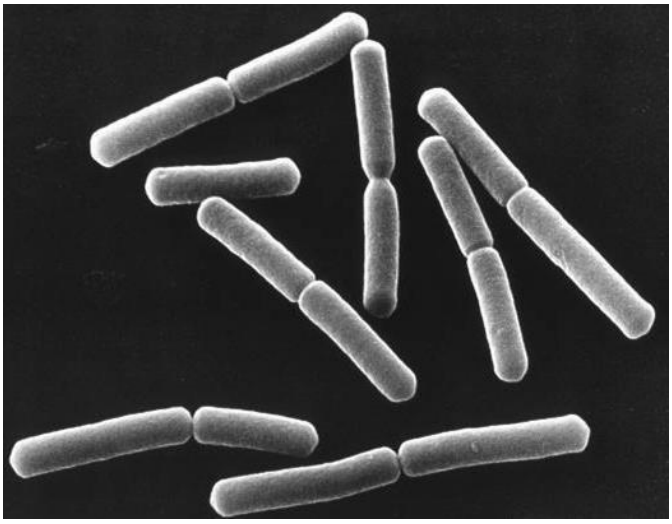

Figure 7 – Electron micrograph picture showing rod-shaped Bacillus subtilis cells. Image credit from NASA.

Growth of rod-shaped bacteria like *B. subtilis* occurs via successive cycles of elongation and division, producing two symmetric daughter cells (binary fission) during which cell shape and its integrity are maintained by the peptidoglycan layer of the cellular envelope (**Nanninga, 1991; Donachie, 1993**).

But *B. subtilis* has also the ability to multiply in an asymmetrical fashion, producing a single endospore that can remain viable for decades, being resistant to unfavourable environmental conditions such as drought, salinity, extreme pH, radiation and solvents (**Errington, 2003**) (**Figure 8**).

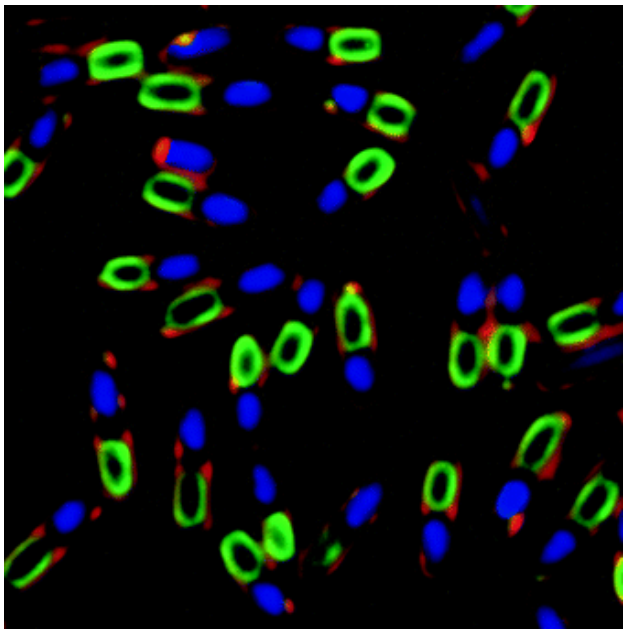

Figure 8 - Fluorescence micrograph of sporulating *Bacillus subtilis* cells, showing nucleoids (blue), membranes (red) and YwcE protein (green; protein required for spore morphogenesis and germination). Cover photograph, American Society for Microbiology. Copyright © 2005. All Rights Reserved.

Prior to the process of sporulation, cells might become motile by producing flagella, take up DNA from the environment, or produce antibiotics. These responses are viewed as attempts to seek out nutrients by searching a more favourable environment, enabling the cell to make use of new beneficial genetic material or simply by killing of competition. (**Bandow et al., 2002**).

B. subtilis turned into an attractive organism to be a model for peptidoglycan shynthesis studies because of its life cycle and how nutrient deprivation leads to a differentiation process resulting in the production of spores. Thus *B. subtilis* is one of the best understood prokaryotes in terms of molecular biology and cell biology. Its superb genetic amenability and relatively large size have provided powerful tools to investigate a bacterium in all possible aspects. Recent improvements in technology have provided novel and amazing insights into the dynamic structure of this single cell organism. The organism is a model for differentiation, gene/protein regulation and cell cycle events in bacteria.

1.5 Aims

As reviewed, recent work suggests that the widely conserved protein RodZ affects the processes of cell division site selection and chromosome positioning, in addition to its central role as a cell shape determinant. RodZ is a multidomain transmembrane protein, responsible for synchronizing inner and outer cell processes of the bacterial cell wall synthesis.

In the present thesis, bioinformatic, biophysical and biochemical tools/techniques are used to study the cytoplasmatic (RodZn) and periplasmatic (RodZc) domains of RodZ individually from *B. subtilis*.

Our first goal is to determine for the first time through NMR spectroscopy the solution structure of RodZn. NMR data acquisition, processing and analysis of RodZn has already been reported in 2011, in my bachelors' graduation thesis. There I also reported the chemical shifts assignment routine. At the end of that work period we were able to complete the assignment of RodZ and from the chemical shifts index (CSI) obtained we were able to determine the secondary structure of $RodZn_{1-98}$ construct. In the current work we concluded the studies of that system with the NMR structure elucidation of that N-terminal domain. From our analysis we concluded that our construct has similarities with previously determined structures but also some meaningful differences (**Pereira, 2011**). The functional role of RodZn *in B. subtilis* hasn't been fully identified yet, but recent data indicate that it may interact with MreB. So, after determining the tertiary structure of RodZn, we intend to use the identified chemical shift resonances of the protein to perform protein-protein interaction studies with MreB.

Next, we'll focus our studies in the uncharacterized periplasmic domain of RodZ (RodZc) from *B. subtilis.* Very little has been published regarding the function/structure of this domain. Since no structure has ever been determined for this domain, our main goal will be to calculate a 3-dimensional structure of RodZc through solution state NMR experiments. We will also study the dynamics of this domain, by collecting ^{15}N relaxation data of the protein backbone and understand if this domain may possibly promote protein-protein interaction with other key players from the peptidoglycan synthesis machinery.

This thesis represents a collaborative work between various research groups from ITQB, in which the ultimate goal is to shed a new light into a better understanding of the biological importance of this multimeric domain.

Levinthal's Paradox

"The length of time in which a protein chain finds its folded state is many orders of magnitude shorter than it would be if it freely searched all possible configurations."

Levinthal, C. 1969. Mossbauer Spectroscopy in Biological Systems. Proceedings of a meeting held at Allerton House. P. Debrunner, J. C. M. Tsibris, and E. Munck, editors. University of Illinois

The Central Dogma

"The three-dimensional structure of a protein is determined by its sequence and its environment without the obligatory role of extrinsic factors".

Anfinsen CB (1973). Principles that govern the folding of protein chains. Science 181 (4096): 223–230.

Chapter 2

Methodology

Proteins are linear chains of amino acids that adopt a three-dimensional structure in their native environment. The biological role of a protein is determined by its function, which is in turn largely determined by its structure (**Dickerson et al., 1969; Petsko, 2000**). Thus there are enormous benefits in knowing the three dimensional structure of all the proteins.

Levinthal's paradox raised the question why and how a sequence of amino acids can fold into its functional native structure given the abundance of geometrically possible structures (**Levinthal, 1969**). The pioneering experiments of Anfinsen shed light on this problem. According to Anfinsen's thermodynamic hypothesis, proteins are not assembled into their native structures by a biological process, but folding is a purely physical process that depends only on the specific amino acid sequence of the protein and the surrounding fluid. Anfinsen's hypothesis implies that a protein structure can be predicted if a model of the free energy is available, and if the global minimum of this function can be identified.

The first three-dimensional protein structure in crystalline state, namely Myoglobin, was experimentally determined in 1959 by John Kendrew using x-ray crystallography (**Kendrew, 1959**).

After that, the structures of many other proteins in crystalline state were also determined. However, in order to determine the structure of proteins in solution state, other methods for structure determination were developed. One of those is NMR, which is based on the use of a strong magnetic field to create the conditions for nuclei to absorb and re-emit electromagnetic radiation. In solution, proteins are expected to have some freedom and flexibility thus, the possibility of determining the structure of proteins in solution brought key advantages to better describe and understand the behaviour

of proteins inside the cells. Therefore, x-ray crystallography and NMR are in many aspects complementary and are still, nowadays, the two major techniques used for protein structure determination at high resolution.

The last RCSB protein data bank (PDB; http://www.rcsb.org) annual report ((http://www.rcsb.org/pdb/statistics/holdings.do accessed in 07/10/2013) listed 94,336 atomic coordinate entries available in the PDB repository, from which 88% were determined by x-ray crystallography and 11% determined by NMR (**Berman et al., 2000**). However, the gap between known protein sequences and structures is increasing rapidly. Statistics released in October 2013 show that UniProtKB/Swiss-Prot contains 540,958 sequence entries (http://web.expasy.org/docs/relnotes/relstat.html). This gives an idea how experimental methods alone will not be able to fill in this gap. Therefore it is necessary to use computational methods to predict protein structures (**Dill et al., 2007**).

Template based homology modelling methods could be used for sequences that have detectable relationship with sequences of experimentally determined protein structures. On the other hand, for predicting the structure of proteins that do not share a detectable sequence relationship with experimental structures, *Ab initio* protein structure prediction techniques must be used. The methods under *Ab initio* protein structure prediction category aim to predict the structure of a protein from the sequence information alone, without any explicit use of previously known structures. These methods use thermodynamic principles and try to identify the native structure of a protein as the global minimum of a potential energy landscape.

Since such methods are computationally complex and are extraordinarily challenging, over the past few years there has been a huge effort in developing *ab inito* protein structure prediction methods. A major milestone in computer-based native structure prediction is the creation of CASP (Critical Assessment of Techniques for Structure Prediction) by John Moult (**Moult, 2006**). In the CASP experiments, research groups apply their prediction methods to amino acid sequences for which the native structure has not been publicly disclosed but has been determined and are to be published soon. These competitions provide a good measure to benchmark methods and progress in the field in an arguably unbiased manner (http//:www.forcasp.org).

These *Ab initio* prediction methods are based on physical and energetic principles that perform the search through the conformational space. Models used are usually simplified and search methods that are often used are Monte Carlo Algorithms.

The *de novo* methods that participate in CASP are not pure *ab initio* methods since they use sequence homology in some way: secondary structure is predicted by using data-base derived potentials, fragments from the existing protein structures, as well as multiple sequence alignment.

In the following chapter will be described the protein structure prediction methods available so far. They will be divided into four categories: **1) Homology Modelling, 2) Pattern-based Homology Modelling, 3) Ab-initio, and 4) Data-driven structure prediction.**

Essentially, the classification reflects the degree to which different methods utilize the information content available from the known structure database and experimental data. I will focus on the specific methods that were used on the present work. I will go over the accuracy, applicability and shortcomings of each kind of tool. Protein structure calculation, refinement and evaluation methods are also discussed. In addition, for simplification purposes, I will be referring to the available not pure-*ab initio* tools (pseudo-*ab initio*) just as *ab initio*.

2.1 Homology Modelling

Homology Modelling relies on the principle that similar sequences exhibit similar three-dimensional structures (**Floudas, 2007**). Strong sequence similarity often indicates strong structure similarity, although the opposite is not necessarily true.

There are usually four steps in homology based protein structure prediction methods: (1) identify one or more suitable structural templates from the known protein structure databases; (2) align the target sequence to the structural template; (3) build the backbone from the alignment, including the loop region and any region that is significantly different from the template; and (4) place the side-chains.

In homology modelling, local sequence comparison methods are usually used since the sequence similarity is most likely over segments of the two sequences. The local sequence comparison can either be pair wise or profile based. Pair wise comparisons, such as the **BLAST** can detect sequence similarities better than 30% (**Altschul et al., 1990**). To increase the chance of detecting weak homologues, **PSI-BLAST** (Position Specific Iterated BLAST) was build to search the database iteratively until no new hits are found (**Altschul et al., 1997**). Methods such as PSI-BLAST encode the information about a whole protein family for the target sequence in a model to increase the chance of detecting remote homologies. To further increase the detection sensitivity, the sequences in the structure database can also be encoded in profiles. This forms the basis of the pattern-based homology modelling methods that will be explained in detail in the next chapter (**Koehl and Levitt, 2002**).

The accuracy of predictions by homology modelling depends on the degree of sequence similarity. If the target and the template sequence have more than 50% of their sequences similar, predictions are of high quality and have been shown to be more accurate than low-resolution X-ray predictions, with the RMSD (root mean square deviation) of the aligned portion between the two structures lower than 1. For 30-50% sequence identity, 80% of the CA-atoms can be expected to be within 3.5 Å of their true positions, while for less than 30% sequence identity, the prediction is likely to contain significant errors (**Kopp and Schwede, 2004**; **Vitkup et al., 2001**).

On the other hand, structurally similar proteins can have low sequence identities (8-10%) and still be identified with sensitive profile-profile based comparison, but the RMSD can be as large as 3-6. The error largely comes from the misalignment from sequence comparison. At such low sequence identity, comparison method that can detect the remote homology as well as align the sequences close to the optimal from structure-structure alignment would be desirable.

- **SWISS-MODEL**

SWISS-MODEL (http://swissmodel.expasy.org) is one of the most widely used web-based servers for automated homology modelling-based structure prediction (**Peitsch et al., 2003**). In addition to a fully automated mode requiring minimum user input, i. e. protein sequence only, SWISS-MODEL offers two more advanced user modes in which users can submit their own multiple sequence alignment or manually adjust the modelling parameters (**Guex and Peitsch, 1997; Schwede et al., 2003**).

To generate a 3D structure from a provided sequence, SWISS-MODEL utilizes rigid body assembly, in which sections from aligned regions of the template are connected together by separately constructed non-conserved regions to form the model backbone. Suitable template structures, those with similar sequences to the query, are first identified by a gapped BLAST search of the SWISS-MODEL template library ExPDB, a subset of PDB. The selected templates are then superimposed using an iterative algorithm, the backbone atom positions averaged, and the query sequence fitted to the template to optimize placement of insertion and deletion regions. Fragments that cannot be modeled by homology to the template are computed based on energy considerations or, if the region cannot be solved, searched against a library of loop structures to find an appropriate match. Finally, side chain conformations and intermolecular interactions are adjusted to minimize conformational energy and correct any irregularities in overall 3D structure that resulted from the assembly process.

Assessments of the SWISS-MODEL prediction algorithm reveal variable accuracy dependent on the degree of query-template sequence similarity. If the predicted protein models show 40% sequence identity between query and template, the RMSD will be less than 3 Å from their experimental structures. However, proteins with lower sequence identities failed to exhibit such modelling accuracy.

Nevertheless, SWISS-MODEL predictions demonstrated the lowest deviations (2 Å CA RMSD) from experimental structures when compared to other servers models. However, this apparently greater accuracy in modelling may be due to the relatively shorter regions that the program modeled in cases of low homology (**Arnold et al., 2006**).

A more recent benchmark study showed that SWISS-MODEL was relatively poor at producing reliable models; SWISS-MODEL was unable to generate predictions for 10% of the provided alignments due to difficulties in loop modelling that crashed the program. SWISS-MODEL also produced more models with poor stereochemistry for difficult query proteins and relatively higher

numbers of 3D models that failed to converge (i.e. >3 Å RMSD) with the backbone structure, reducing the fidelity of the final model to its original template (**Wallner and Elofsson, 2005**).

2.2 Pattern-based Homology Modelling

For evolutionary related proteins, even if the sequence similarity is difficult to detect with sequence comparison methods, there could still be identifiable structural similarity. Structure alignment has been shown to be able to identify homologous protein pairs with sequence similarities less than 10% (**Rost, 1997; Brenner et al., 1998; Gerstein et al., 1998**). Pattern recognition-based methods involve similar steps as in homology modelling. The difference is in the pattern identification step. First of all, a structure library needs to be defined (**Sitbon and Pietrokovski, 2007**). Natural choices for building the library of protein clusters are protein families (**Henikoff et al., 1997**), sequence motifs that separate proteins into either conserved or random regions (**Henikoff et al., 2000**), and the four basic secondary structure elements namely alpha helices, beta strands, turns, and loops (**Henikoff et al., 1995**). Once the library is defined, the target sequence will be fitted to each library entry and an energy function is used to evaluate the fit between the target sequence and the library entries to determine the best possible templates.

Protein sequence comparison has become one of the most powerful tools for characterizing protein sequences because of the enormous amount of information that is preserved throughout the evolutionary process. A general approach for functional characterization of unknown proteins is to infer protein functions based on sequence similarity. One of the approaches is to define signatures of known families of biologically related proteins. These signatures usually identify conserved regions among the family of proteins, revealing the importance for the function of their structural properties. A representative example of this approach is the well-known **PROSITE** database, which gathers protein sequence patterns and profiles for a large number of families (**Hulo et al., 2004**). Another web tool based in pattern homology modelling is PSIPRED.

- **PSIPRED**

Efficient automatic methods for predicting topologies of both globular and membrane bound proteins have been publicly available as individual programs in individual servers. However, in order to make methods more accessible, structure prediction web servers incorporating these programs are becoming more prevalent. The PSIPRED protein structure prediction server incorporates several recently developed methods for predicting proteins structural information with the amino acid sequence as its sole input (**McGuffin et al., 2000**; www.bioinf.cs.ucl.ac.uk).

PSIPRED carries the secondary structure prediction on a protein and gives its name to the prediction server itself. α-helical (H), β-stranded (E) and Coiled-coil (C) protein motifs are predicted based on the analyses of the output obtained from PSI-BLAST combined with the DSSP algorithm (Define Secondary Structure of Proteins). To these predictions are assigned individual confidences (**Altschul et al., 1997**). Using a rigorous cross validation procedure to evaluate performance, PSIPRED has been shown to be capable of achieving an average Q3 score (% of 3-state symbols that are correct) of 81.6% (PSIPRED v3.2). This is one of the highest levels of accuracy published for any method to date (**Jones, 1999**).

PSIPRED server allows users to select a variety of prediction methods they may find useful. Here will be reviewed two recently developed tools: MENSAR-SVM which predicts the topology of a protein and DISOPRED that predicts disordered regions from the amino acid sequence.

Transmembrane protein prediction: Transmembrane proteins (TM) are involved in a wide range of important biological processes such as cell signalling, transport of membrane-impermeable molecules, cell-cell communication, cell recognition and cell adhesion. However, due to experimental difficulties, this class of proteins is severely under-represented in structural databases. In the absence of structural data, sequence-based prediction methods allow TM protein topology to be investigated. PSIPRED server also supports the vector machine-based TM protein topology predictor **MENSAT-SVM**.

This web-tool integrates both signal peptide and re-entrant helix prediction, cross validated on a novel data set of 131 sequences with known crystal structures. The method achieves topology prediction accuracy of 89%, while signal peptides and re-entrant helices are predicted with 93% and 44% accuracy respectively. An additional SVM trained to discriminate between globular and TM proteins detected zero false positives, with a low false negative rate of 0.4%. The high accuracy of TM topology prediction, which is able to discriminate signal peptides and identify the cytosolic and extra-cellular loops, makes this method ideally suited to whole genome annotation of alpha-helical transmembrane proteins (**Nugent and Jones, 2009**).

Most efforts in structural bioinformatics have been directed to the prediction of globular protein structures but there is an increasing interest in disordered regions for studying the function of proteins.

Order/Disorder prediction: Interest in intrinsically disordered proteins (IDPs) grew as a result of the realization that such proteins are unexpectedly common in human and other genomes (**Ward et al., 2004; Oldfield et al., 2005; Tompa et al.,, 2006**).

The structural flexibility of IDPs allows them to achieve functional modes that otherwise would be unfeasible to globular proteins. Moreover, by having this feature of plasticity, IDPs can fold upon binding, forming transient complexes through nonspecific interactions or interactions with rapid dissociation rates, allowing IDPs to remain dynamic (**Wright and Dyson, 2009**). That's one of the

reasons why IDPs are frequently associated with cellular control mechanisms and signalling, and have been identified at the "heart" of protein interaction networks (**Dunker et al., 2005**).

In recent years there has been a growing interest in intrinsically unstructured proteins and their role in biology. However, it's yet poorly understood the relationship between the primary sequence of a protein and its susceptibility to a disordered conformation. A signature of probable intrinsic disorder is the presence of low sequence complexity and amino-acid compositional bias, with a low content of bulky hydrophobic amino acids (Val, Leu, Ile, Met, Phe, Trp and Tyr), and a high proportion of polar and charged amino acids (Gln, Ser, Pro, Glu, Lys, Gly and Ala) (**Romero et al., 2001; Vucetic et al., 2003**).

A number of computer programs are now available for the prediction of unstructured regions from amino acid sequences. Disordered regions are dynamically flexible and are distinct from irregular loop secondary structures, which are static in solution. But many disordered proteins do adopt more highly ordered conformations upon interactions with other cellular components (**Dyson and Wright, 2002; Iakoucheva et al., 2002; Dunker and Obradovic, 2001**).

The **DISOPRED** server present in the PSIPRED platform uses a knowledge-based method to predict dynamically disordered regions from the amino acid sequence. This server allows users to submit a protein sequence, and returns a probability of each residue in the sequence being disordered. The DISOPRED method is developed from the original DISOPRED predictor (**Jones and Ward, 2003**).

2.3 Ab-initio

Only when no suitable structure templates can be found, *Ab initio* methods can be used to predict the protein structure from the sequence information only. In addition, they can also predict protein mechanisms, motions, folding processes, conformational transitions and other situations in which protein behaviour requires more than just knowledge of the static native structure (**Dill et al., 2007**). In recognition of the developments and the importance of this field, the Nobel Prize of Chemistry of 2013 was focused on such approaches.

This class of methods can be applied to any given target sequence using only theoretical-physics restraints. Common to all *Ab initio* methods is first defining a protein representation and corresponding protein conformational space in that representation, second finding a compatible energy function with the protein representation and lastly applying efficient and reliable algorithms to search the conformational space to minimize the energy function. The conformations that minimize the energy function are taken to be the structures that the protein is likely to adopt at native conditions. The folding of the protein sequence is ultimately dictated by the physical forces acting on the atoms of the protein and thus the most accurate way of formulating the protein folding or structure prediction problem is in terms of all-atom model subject to the physical forces. Two main issues to be taken into account for a

successful prediction of a protein structure are generation of a vast number of conformations and accurate scoring functions.

The *Ab initio* methods are rigorous in calculations but are limited by the compute power and time involved which emphasizes the need for faster structure prediction methods. Also, the accuracy of these methods is dependent upon the potential energy functions used during simulations (**Karplus and McCammon, 2002**).

For practical reasons, most *Ab initio* prediction methods use reduced representations of the protein to limit the conformational space to convenient size and use empirical energy functions that capture the most important interactions that drive the folding of the protein sequence toward the native structures.

- **TASSER**

TASSER is a successful free modelling approach that constructs 3D models based on a hybrid between *ab initio* and homology modelling approach (**Zhang and Skolnick, 2004**). The target sequence is first threaded through a set of representative protein structures to search for possible folds. Contiguous fragments (>5 residues) are then excised from the threaded aligned regions and used to reassemble full-length models, while unaligned regions are built by *ab initio* modelling (**Zhang et al., 2003**).

The protein conformation in TASSER is represented by a trace of Cα atoms and side chain centres of mass, and the reassembly process is conducted by Monte Carlo simulations (**Zhang et al., 2002**). The energy terms of TASSER include information about predicted secondary structure propensities, backbone hydrogen bonds, a variety of short- and long-range correlations and hydrophobic energy based on the structural statistics from the PDB library. Weights of knowledge-based energy terms are optimized using a large-scale structure decoy set (computer-generated set of protein structures) which coordinates the correlations between various interaction terms (**Zhang et al., 2003**).

There are several new developments of TASSER. One is **I-TASSER** which refines TASSER cluster centroids (centroid is a simplified model of the polypeptide chain in which mainly CA atoms are considered the center of mass of the side chain) by iterative Monte Carlo simulations (**Wu et al., 2007**).

The spatial restraints are extracted from the first round TASSER models and the template structures searched by TM-align from the PDB library, which are exploited in the second round simulations. TM-align is a highly optimized algorithm for protein structure comparison and alignment. For two protein structures of unknown equivalence, TM-align first generates the residue-to-residue alignment based on structural similarity using dynamic programming iterations. An optimal superposition of the two structures, as well as the TM-score value which scales the structural similarity, will in the end be returned (**Zhang and Skolnick, 2005**). The purpose is to remove the steric clashes from the first round models and refine the topology. Although the procedure uses structural fragments and spatial restraints from threading templates, it often constructs models of correct topology even when topologies of constituting templates are incorrect.

- **ROSETTA**

ROSETTA has been extensively reviewed as being the best *Ab initio* method so far (**Bonneau and Baker, 2001; Bonneau et al., 2001; Simons et al., 2001**).

The ROSETTA method also illustrates many features and techniques that are common to the majority of the *Ab initio* based-methods. The ROSETTA method, like many others, uses a reduced representation of the protein as short segments, since local segments of the protein sequence have statistically important preferences for specific local structures and that the tertiary structure has to be consistent with this preference (**Go, 1983; Simons et al., 1997**).

The energy function is defined as probabilistic model of protein sequence/structure matches in terms of structural segments to formulate secondary structure prediction and this forms the basis of the Monte Carlo sampling of the reduced protein conformational space (**Simons et al., 1999**). The non-local potential, which drives the protein toward compact folded structure, includes terms that favour paired strands and buried hydrophobic residuals. The solvation effect can also be incorporated into the energy function.

A problem intrinsic to the reduced representation of the protein and the simplified empirical potential is that the energy function is not sensitive enough to differentiate the correct native structures from conformations that are structurally close to the native state. The energy landscape calculated from such energy functions will not be properly funnelled but flattened around the native structure. In fact, as the native state is approached, the correlation between the calculated energy and the measure of similarity between predicted and native structures are no longer valid. The usual practice is then to produce a large number of decoy structures and then use various filtering and clustering techniques to pick up the more native like structures. Filters can be used to eliminate structures with poorly formed secondary structures and low contact orders compared with that for sequences with compatible length (**Bonneau et al., 2001**).

2.4 Data-driven structure prediction

Even small amounts of experimental data can dramatically improve the quality and reliability of *ab initio* structure prediction with the application of spatial constraints. For instance, the Rosetta method can produce moderate to high-resolution structures when combined with limited NMR constraints (**Standley et al., 1999; Bowers et al., 2000; Rohl and Baker, 2002**).

- **CS-ROSETTA**

CS-ROSETTA (www.csrosetta.org) is a framework for structure calculation of biological macromolecules on the basis of structural information derived from NMR data, which is build on top of the biomolecular modelling and design software called ROSETTA. The name arise from its origin in combining NMR chemical shift data (CS) with ROSETTA prediction protocols (**Shen et al., 2008**). Further advantages of using chemical shifts are that these are among the most reliable parameters that can be obtained from NMR spectroscopy and that they can potentially be obtained for larger macromolecules for which NOEs become impractical.

Structure determination using CS-ROSETTA requires as only input the amino acid sequence and a list of chemical shifts and a number of parameters to control the process that can be changed from the default values. Backbone chemical shifts for ^{13}CA, ^{13}CB, ^{13}CO, ^{1}HA, ^{1}HN, and ^{15}N that are provided by the user, are validated and stored as the target shifts. These chemical shifts are first used to select a fragment library that contains many possible conformations (fragments) for a given set of degrees of freedom specific for the protein backbone. This selection is made from a structure database, e.g. the RCSB Protein Data Bank (PDB), based on the list of chemical shifts as predicted with **SPARTA**. SPARTA is a database system for empirical prediction of backbone chemical shifts (N, HN, HA, CA, CB, CO) using a combination of backbone phi, psi torsion angles and side chain chi1 angles from a given protein with known PDB coordinates (**Yang and Bax, 2007**).

Then the regular ROSETTA protocol for Monte Carlo assembly and relaxation is used to reassemble the protein from the fragments. For the resulting models the chemical shifts are back predicted using SPARTA and the deviations between the predicted and target values are used as a pseudo-energy term in the scoring of the models, yielding a ranking based on both overall structural quality as well as on the match with the experimental data.

The computationally most expensive step in the process is the construction of a model using Monte Carlo assembly and relaxation. To obtain a reliable prediction, a set of 10,000 to 50,000 models has to be built, each starting from the same fragment library. Using different seeds for generation of random numbers ensures independence of the results from different runs. The computational cost involved in chemical shift based structure determination makes CS-ROSETTA a typical example of a program that is beyond the capacity of most local sites.

More recently, a new protocol was added to CS-ROSETTA. **RASREC** protocol was designed to improve sampling close to the native structure by recombination of frequently occurring structural features such as α-helix or β-strand topologies.

RASREC is an iterative conformational sampling protocol that seeks to pool knowledge gained about the conformational space in previous trajectories to efficiently guide further space exploration. The protocol is inherently parallel and requires inter process communication which is realized with the Message Passing Interface (MPI). (**Andreas et al., 2006**).

The RASREC protocol is characterized by first using chemical shift data to pick and choose a fragment library, as described previously. Then, individual structure calculations employ the **abrelax protocol** that combines *ab initio* fragment assembly in centroid mode with the all-atom refinement using the Rosetta full-atom force field (**relax**), by interleaving packing of side chains with gradient based minimization of torsional degrees of freedom. The advantage of this protocol is that it is relatively time efficient since *Ab initio* folding is faster and the relax step is more time-consuming (**Das et al., 2007**).

Around 200-1000 of independent structure calculations run in parallel and a pool of the best output decoy set is continuously updated from the results of the independent structure calculation. Specific features (like α-helix or β-strand topologies) in the pooled structures are then used to focus further sampling (**Raman et al., 2010; Lange and Baker, 2012**).

The CS-ROSETTA software package has recently been extended to include additional NMR conformational parameters, such as Residual Dipolar Couplings (RDC) and NOE distance restraints (**Raman et al., 2010; Lange et al., 2012**).

2.5 Fundaments of NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a very versatile tool that can detect the frequency at which certain nuclei are resonating under the influence of a magnetic field. Slight variations in this resonant frequency give us detailed information about the molecular structure in which the atom resides. It has a very wide variety of applications and it can also be used to study the structure and properties of biomolecular compounds.

The available nuclei suitable for nuclear magnetic resonance are those with non-zero spin quantum number. For technical reasons, mostly unpaired nuclear spins with I=1/2 are used for NMR spectroscopy. The most widely used nuclei in biomolecular NMR are ^{1}H, ^{15}N and ^{13}C, being the main components of proteins.

Although NMR was discovered in 1946, its application to biological systems only started in the late 1970s. Major breakthroughs happened in this period that revolutionized the field, allowing the use of NMR in determining the three-dimensional structures of biomolecules with low molecular weight (<10kDa) (**Teng, 2005**). More recently, in the early 1990s, the NMR limit of analysis was shifted to an upper limit of 35 kDa due to the development of multidimensional heteronuclear methods as well as advances of molecular biology that allow overexpression of isotope labelled proteins such as $^{15}N/^{13}C$ double labelled proteins. Despite its size limitation for macromolecular structure determination, NMR holds advantages in respect to other protein structure determination techniques, such as X-Ray crystallography. Namely, NMR can provide complementary structural information in a more physiologically-like environment solution, avoiding experimental artefacts such as poor crystal packing seen in some crystal structures. Moreover, it allows structural studies of biomolecules that are difficult

to crystallize such as proteins with flexible domains and weakly bound protein complexes. NMR can also provide information about protein dynamics, flexibility and folding/unfolding transitions.

The protocol for proteins structure determination by NMR spectroscopy usually starts with sample preparation followed by NMR data acquisition and processing. After that, is performed the chemical shifts assignment for each residue of the protein sequence, NOEs assignment and assignments of other conformational restraints such as hydrogen bonds and J-couplings. Lastly, a structure calculation is made using various structure refinement protocols.

The data analysis uses a series of parameters that allow the determination of important values used as pieces in a puzzle to build the tertiary structure of a protein.

The Chemical shifts provide the secondary structural information for proteins, the J couplings constants provide the dihedral angles of the peptide plane which in turn is related to the secondary structure and the nuclear Overhauser effects (NOEs) provide the ^{1}H-^{1}H distance within a 5 Å maximum distance. The NOE data are considered to be the most important because of the long range structural information they contain which leads to tertiary structural information (**Wuthrich, 1986; Malcolm and Levitt, 2008**). More recently, residual dipolar couplings (RCDs) are also used to provide information about the orientation of chemical bonds in respect to each other.

2.5.1 Protein sample preparation for NMR

Recombinant protein expression and purification remains a fundamental issue for biotechnology. Every protein behaves differently so it's fundamental to develop a proper protocol for expression and purification in each case. In addition, depending on the intended application for each sample, additional considerations have to be taken into account. These are the total protein yield, the degree of purity, incorporation of specific isotopes and also the total cost of production.

For NMR studies, proteins need to be produced within high purity percentages (>90%) and stability (>5 days at room temperature). The resulting proteins also need to be isotopically enriched in NMR compatible isotopes such as ^{13}C and ^{15}N.

There are a number of steps that must be followed to properly express and purify a protein. When the appropriate plasmid is available, the first step is the insertion (transformation) of the plasmid containing the gene of interest in the appropriate competent cells and growth of the resulting culture in agar plates supplemented with the appropriate antibiotic for the plasmid in use. Upon selection of a colony, a small scale culture (pre-culture) is made. Growing cells with antibiotic, such as kanamycin, will allow bacteria to grow containing plasmid with antibiotic resistance, making this a mechanism for selection of the plasmid bearing cells. The preculture is used in the next step to inoculate a large scale culture. At this stage the culture growth is monitored and the cells are usually incubated until the culture reaches an optical density between 0.4 and 0.6 that corresponds to the mid-log phase of cells growth, in which

nutrients and vitamins are still available for bacteria to grow. At the optimal O.D., we would have a sufficient number of healthy cells, being the right point to induce the expression of the protein of interest. The induction step is achieved by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to the bacterial culture. For plasmids employing the lac operon, both induction and cells growth must be done in either rich Lysogeny Broth medium (LB) or minimal medium (M9). Both LB and M9 media contain ample amounts of all the essential inorganic nutrients needed for cell growth. Minimal medium is the most commonly used and cheap medium to overexpress isotope labeled proteins for NMR sample preparation. The minimal medium is a mixture of M9 salts, glucose as the carbon source and ammonium chloride or ammonium sulphate as the nitrogen source. Because of the lower nutritional content of Minimal medium, bacterial growth in it is often slow and sometimes irreproducible.. The induction of protein expression is a very important step for the final expression yield so there are many parameters that must be considered, such as the optimal concentration of IPTG as well as the temperature and length of induction.

After the appropriate induction period, begins the process of protein purification. Firstly, bacteria must be separated from its growth medium. Then the cells need to be disrupted, this can be achieved mechanically (e.g. French Press) or chemically (osmotic shock or by detergents). For proteins that are located in the soluble part of the cell extract, the supernatant of a subsequent centrifugation must be kept to continue the purification process. Then, depending on the protein size, isoelectric point (pI), hydrophobicity and other key factors, protein purification methods must be carefully chosen (e.g. the size of the pores of a dialyses membrane; the type of resin of a FPLC column; etc.).

The final protein yield is then determined using either, UV, or chemical indicators (e.g., Bradford). Finally the buffer suitable for either storage or subsequent studies is introduced and the protein is concentrated to the desired degree. So if the purpose of expressing and purifying a protein is to be later studied by NMR, those proteins need to be labeled with different isotopes. These elements can be introduced using a variety of different isotopically labeled carbon and nitrogen sources (supplemented with 15NH4Cl and non-labeled glucose for 15N labeling, and with 15NH4Cl and 13C-glucose for 15N-13C double labeling).

2.5.2 NMR assignment methodology

The first step in any NMR based study of biomolecules is the identification of the nature of the observed signals. This process in which the observed chemical shift resonances are associated with the chemical entities that they represent is called assignment (**Michael et al., 1999**).

A multitude of different methodologies exist for the achievement of this step and the selection of the suitable one depends on a number of factors. First the nature of the sample has to be considered. Small protein or peptides can be assigned without the use of any isotopic labeling, using methods that

rely exclusively on ^{1}H homonuclear data. However, as the biomolecular weight increases, the signals overlapping in ^{1}H 1D spectra becomes too great and the ambiguity of the assignments increases dramatically. The solution to this problem is to simultaneously observe additional nuclei, namely ^{13}C and ^{15}N. The combined interpretation of ^{1}H, ^{13}C and ^{15}N spectral data, allows the assignment of the resonances of proteins up to 25kDa in size.

Using a ^{15}N-labelled protein, it is possible to record the standard ^{1}H-^{15}N-HSQC 2D spectrum. It provides an initial set of HN resonances of backbone assignment that will work as a guide through the 3D spectra. This spectrum shows one peak for each ^{1}H-^{15}N pair connected by a single covalent bond, where the magnetization is transferred from ^{1}H to ^{15}N via J-coupling. The chemical shift is evolved on the nitrogen atom and the magnetization is then transferred back to the hydrogen atom for detection. Since the backbone of a protein only contains one ^{15}N, this experiment can be used to directly identify the HN and N chemical shifts of each residue, except for Prolines. Besides, the backbone amide groups from Asn (NDHD), Gln (NEHE), and Trp (NEHE2) can also be observed. Arg, Lys and His side chains are not usually observed in a regular ^{1}H-^{15}N-HSQC spectrum as their ^{15}N resonances are far away from those of the amide resonances.

With a $^{13}C/^{15}N$ labeled protein, commonly referred to as double labeling, it's possible to record 3D spectra that simultaneously use the information of ^{15}N and ^{13}C nuclei. These are called triple resonance experiments and enable straight forward assignment of both backbone and side-chain. For more details, see the description in Material and Methods present in Chapter 3 and 4.

When backbone sequential assignment is complete, the next step is so assign the remaining side chain atoms. First, to assign the Hα of all residues, a ^{15}N-TOCSY-HSQC 3D experiment must be recorded. To record this experiment, a uniform mixing step transfers magnetization from the amidic to all ^{1}H spins of the same residue. Because the chemical shift is evolved on all three nuclei (^{1}H, ^{15}N and side chain ^{1}H) the result is a 3D spectrum. This spectrum is useful to determine amino acid types. It can also help with the assignment not only of Hα atoms but also the rest of side-chain hydrogen atoms, though normally an HCCH-TOCSY spectrum is necessary for unambiguous side-chain hydrogen assignment. To record an HCCH-TOCSY 3D spectrum, magnetization is transferred from the side-chain hydrogen nuclei to their attached ^{13}C nuclei. This is followed by a uniform ^{13}C mixing and finally transfers back to the side-chain hydrogen atoms for detection. This spectrum is used for side-chain assignment. It yields strips at each carbon frequency in the side chain in which all side-chain hydrogen resonances are visible (**Markley et al., 1998; Higman and Vicky, Protein NMR – A practical guide. http://www.protein-nmr.org.uk/index.html**).

To successfully finish this step, aromatic side chain protons need to be assigned. This task is accomplished with the analysis of the ^{13}C-NOESY and ^{1}H-^{13}C-HSQC. The NOE of each proton in the aromatic region on the ^{13}C-NOESY can be correlated with a specific carbon atom in the ^{1}H-^{13}C-HSQC.

After the assignment of the backbone and side chain atoms, this information is used to identify the NOE peaks in the 2D ^{1}H-^{1}H NOESY, ^{13}C-NOESY and ^{15}N-NOESY spectra. In order to obtain a high resolution structure it is necessary to complete the assignment for a sufficient number of atoms in the sequence.

Finished the assignment of the NOESY spectra, the NOE intensities can be converted into proton-proton distances and used for structure determination. It is also possible to extrapolate information from the chemical shifts to improve the structure determination step. For instance, dihedral angle constraints of the peptide plane can be calculated since they have a strong relationship with the chemical shifts of the HA, CA, CB and CO atoms.

2.5.3 Structure calculation

NMR spectroscopy is an established technique for macromolecular structure determination at atomic resolution. However, the majority of the current structure determination approaches require a large set of experiments and use large amount of data to elucidate the three dimensional protein structures. While current structure determination protocols may perform well in data-rich settings, protein structure determination still remains to be a difficult task in a sparse-data setting, thereby requiring novel algorithms that can compute structures with provable guarantees on solution quality and running time.

UNIO automated protocol ensembles several algorithms and computational tools that enable protein structure determination from sparse NMR data, allowing the minimization of the number of NMR experiments, hence the amount of time and cost to perform them, and still be able to determine protein structures accurately from a limited set of experimental data.

- **UNIO automated protocol**

The UNIO protocol combines UNIO package and **CYANA** algorithm to perform automated NMR data analysis for protein structure determination (**Serrano et al., 2012**).

In the program CYANA, automated NOESY assignment is performed by the CANDID algorithm (**Herrmann et al., 2002**) that proceeds in iterative cycles of ambiguous NOE assignment followed by structure calculation using torsion angle dynamics. This algorithm combines the use of 3D structure-based filters and ambiguous distance constraints, with the new concepts of network anchoring and constraint combination that enable an efficient and reliable search for the correct fold already in the initial cycle of de novo NMR structure determinations (**Mumenthaler et al, 1995; Mumenthaler et al., 1997; Nilges et al., 1997; Nilges and O'Donoghue, 1998; Linge et al., 2001; Linge et al., 2003**). Automated NOESY assignment with CYANA algorithm has no disadvantage compared to the

conventional, interactive approach but is a lot faster and more objective. With CYANA, the evaluation of NOESY spectra is no longer the time-limiting step in protein structure determination by NMR.

The standard UNIO data analysis protocol requires only a minimal set of NMR spectra. But a new version has been released, UNIO10, with new features, namely ATNOS/CANDID has a new Restraint Inspector for NOE upper distance restraint file and supports RDCs data as external restraint files (**Herrmann et al., 2002**).

- **RECOORD**

The quality of NMR structural models is typically assessed by various quality indicators. These indicators can be related to the NMR data themselves, such as restraint violations and NOE completeness (**Doreleijers et al., 1999**), to the precision of the ensembles. Several studies have pointed out that the quality of NMR structures is very heterogeneous (**Doreleijers et al., 1998; Spronk et al., 2002**). Consequently, NMR models are often not the first choice for use in database approaches and drug design (**Laskowski, 2003**). However, the protocols and force fields that have been used for calculating the NMR structures have improved greatly during the last years.

For this purpose, RECOORD protocol was used to recalculate coordinate database of proteins from the PDB. This was built as an attempt to standardize NMR entries with respect to the calculation protocols used. CNS and CYANA algorithms are used to recalculate the structures of proteins providing a unified database of NMR protein structures in which potential user- and software-dependent biases are kept as small as possible.

More specifically, the CNS algorithm was implemented with a version of the ARIA protocols consisting of four MD simulated annealing stages, both in torsion angle and Cartesian space: (1) a high-temperature torsion angle dynamics phase, (2) torsion angle dynamics cooling phase and (3) Cartesian dynamics cooling phase. 200 structures are generated and sorted the structures with respect to the total energy.

The set of 50 lowest-energy structures obtained is water-refined using a scheme similar to that used in ARIA (**Linge et al., 2003**). The explicit solvent refinement consists of the following steps: (1) immersion in a 7.0 Å shell of water molecules and energy minimization, (2) slow heating temperature steps with 200 MD steps per temperature step, (3) refinement step with 2,000 MD steps, (4) slow cooling in temperature step with 200 MD steps per temperature and (5) final energy minimization (200 steps).

2.5.4 Structure refinement

With the improvement of protein structure calculation methods, the protein model refinement problem is becoming increasingly important. A natural choice for a refinement protocol would involve

a detailed atomic model and the use of all-atom physics based potentials. Over the last decade there has been some work in the direction of both structure ranking and refinement using all-atom potentials.

- **Amber**

Amber (Assisted Model Building with Energy Refinement) is the collective name for a suite of programs that allows users to carry out and analyze molecular dynamics simulations, particularly for proteins. None of the individual programs carries this name, but the various parts work reasonably well together, providing a powerful framework for many common calculations (**Bertini et al., 2011**). The term Amber also refers to is a family of force fields for biomolecular dynamics that are implemented in this program.

SANDER (Simulated Annealing with NMR-Derived Energy Restraints) is the central simulation program of Amber suite and provides facilities for energy minimization and molecular dynamics with NMR-derived restraints. Sander incorporates a variety of restraints into an optimization procedure that includes energy minimization and dynamical simulated annealing. The NMR-derived restraints, such as NOE and J-coupling data, are added using aliased syntax which allows definition of pseudo-atoms, connection with peak numbers in spectra, and the use of ambiguous constraints from incompletely-assigned spectra.

More advanced features include the direct refinement against NOESY intensities, use of paramagnetic and diamagnetic chemical shifts, or residual dipolar couplings. In addition, a key strength of the program is its ability to carry out the refinements (usually near the final stages) using an explicit-solvent representation that incorporates force fields and simulation protocols that are known to give pretty accurate results in many cases for unconstrained simulations. This ability should improve predictions in regions of low constraint density and should help reduce the number of places where the force field and the NMR constraints are clashing with one another.

2.5.5 Structure validation

The process of structure validation is used to evaluate the reliability of the 3- dimensional atomic models of biomolecules such as proteins that were built during the structure prediction or structure calculation step. These models provide 3D coordinates for each atom in the molecule and its validation comprises three phases: 1) checks the validity of the thousands to millions of measurements in the experiment, 2) checks how consistent the atomic model is with those experimental data and 3) checks consistency of the model with known physical and chemical properties. Depositors of protein structures in the RCSB Protein Data Bank are also strongly encouraged to use the various validation software's

available to check their NMR experimental data and structure files before uploading them in the database.

- **CING**

CING, which stands for Common Interface for NMR Structure Generation, is a suite of programs, which provides for a residue-based, integrated validation of the structural NMR ensemble in conjunction with the experimental restraints and other input data. **iCing** is a secure web portal (https://nmr.cmbi.ru.nl/icing/) to the CING server that allows users to validate their own data.

External validation programs, like PROCHECK, WHAT IF, Wattos, and new CING internal validation routines compare the NMR-derived models with empirical data, measured chemical shifts, distance- and dihedral restraints. A red/orange/green score is used for residues and restraints to direct the user to those critiques that warrant further investigation. Overall green scores below 20 % accompanied by red scores over 50 % are strongly indicative of poorly modelled structures.

CING performs analysis and evaluation of the experimental data. Various potential assignment issues are evaluated, such as the presence of multiple assignments and/or missing assignments. In addition, the chemical shifts are compared to the BMRB-derived distributions. Peaks typically represent an abstracted stage of the experimental data. What CING does is stores and analyses peak entities for consistency with valid assignments. The distance and dihedral angle restraints are validated to show the RMSD in the models ensemble in which a violation occurs. The distance restraints are then clustered into the following classes: intra-residual, sequential, medium range (between 2 and 4 residues apart), long-range, or ambiguous.

The CING package also implements several tools that evaluate the structural results. CING is able to derive residue-specific statistical preferences for the dihedral angles. Visualization of statistical preferences of dihedral angles through a Ramachandran plot provides information that can aid the assessment of the conformations in the structure ensemble. CING also uses an automated method for outlier data analysis distribution of parameters, such as the dihedral angle distributions of individual residues. Distribution models are colour coded in the corresponding plot.

Chapter 3

High-resolution NMR solution structure of RodZ cytoplasmic domain

3.1 Introduction

To date, only two structures of RodZ have been reported. They correspond to the cytoplasmic domain of RodZ (RodZn) from *Thermotoga maritima and Staphylococcus aureus*. The article where the structure of RodZn from *T. maritima* is described is yet to be published (**Xu et al**). As for the RodZn protein domain from *T. maritima*, its structure was published in 2010, and was shown to interact with the MreB actin-homologue, and it was the first crystal structure of RodZ to be released to the scientific community (**Van den Ent et al., 2010**).

The cytoplasmic domain of RodZ in *T. maritima* consists of 5 α-helices (H1 to H5), with the HTH motif formed by helices H2 and H3, followed by two additional helices (H4 and H5). An N-terminal helix, H1, forms a triangular arrangement with H2 and H3. Residues from this characteristic HTH motif of RodZ were shown to be the ones promoting interaction with MreB, as shown in **Figure 9**. It was reported that a single alanine substitutions of specific residues from the H3 and H4 α-helices (K36 and Y53/Y57 respectively) led to loss of binding between RodZ and MreB. Single alanine substitutions was also performed for the corresponding residues in the *E. coli* RodZn domain, which caused cell shape alterations, and the mislocalization of the protein along the membrane (**Van den Ent et al., 2010**).

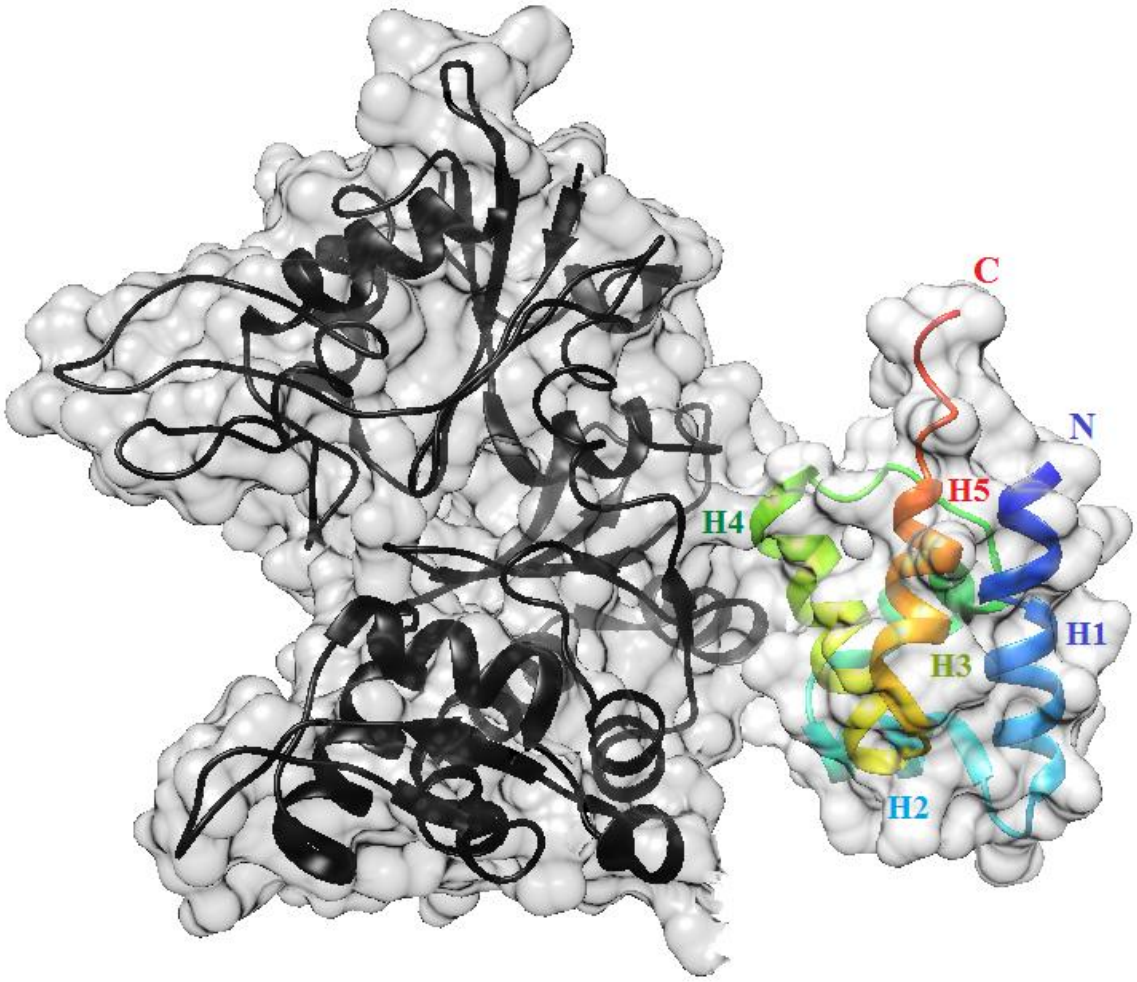

Figure 9 - Ribbon representation of the cocrystal structure of MreB (in black) and RodZ$_{2–88}$ (rainbow coloured) from *T. maritima*. A surface representation of the contacts between the two proteins is shown. Interaction between RodZ and MreB involves residues K36, Y53 and Y57 from H3 and H4 helixes. PDB code: 2WUS (Van den Ent et al., 2010).

However, up until now no structural data has been published for the RodZ from rod-shaped gram-positive bacteria. So, in the present chapter we report the process of structure calculation and refinement of the cytoplasmic domain of RodZn from *B. subtilis*. We extensively analyzed all the NOESY spectra that were collected (2D ^{1}H-^{1}H, and 3D ^{13}C and ^{15}N NOESY spectra) and completed its resonances assignment. The resulting NOEs were used to identify distance restrains in a series of structure minimization steps.

To understand the function of RodZn domain, changes in the NMR spectra were studied upon addition of different ratios of MreB: RodZn to the NMR tube. By acquiring simple 2D ^{1}H-^{15}N-HSQC spectra we could see, by comparison with the same spectrum of pure RodZ, if the protein interacts with MreB. If it does, we would be expecting to see localized changes, i.e., some different chemical shifts for the residues that might be directly interacting with MreB.

3.2 Material and Methods

3.2.1 NMR sample preparation

All the purified protein samples used for NMR were concentrated (**Stoscheck., 1990**) until reaching a final concentration around 1 mM. For reasons of protein stability, the buffer used contained 150 mM Sodium chloride and 50 mM phosphate at pH 7.6 with a final concentration of 10% D2O for signal lock. Because of the high ionic strength from phosphate and sodium chloride, the pulses applied were longer than average, however no noticeable effects on spectrum quality were observed. In addition the solution pH was too high, increasing the solvent exchange effect, leading to the loss of some fast exchanging amide signals. The high buffer pH though, was necessary for reasons of protein stability (**Gasteiger et al., 2005**).

Samples were loaded into a 5 mm NMR tubes (SL-5 from New Era Enterprise). These tubes are made of Type 1, Class B Borosilicate glass specific to work at the highest magnetic fields (from 700 up to 900+ MHz) and ideal for structural biology research (Specifications: 4.960 ± 0.006 mm OD; 0.40 ± 0.001 mm nominal wall; 0.0025 ± 0.002 mm roundness; 178 mm in Length).

3.2.2 NMR data acquisition and structure calculation

NMR experiments were performed on a Bruker AvanceIII spectrometer (Bruker, Rheinstetten, Germany) with 1H operating frequency of 800.33 MHz at the ITQB - António Xavier Magnetic Resonance Center, CERMAX, Nova University of Lisbon, Portugal. The spectrometer was equipped with a triple resonance pulsed field TXI Z axis gradient H C/N/-D (5 mm) probe. All data was aquired at 298.15 K unless otherwise stated, and processed with Bruker's Topspin 2.1 software (Bruker Biospin). Both CARA (v1.8.4.2) and CCPN (v2.2.1 to v2.2.3) software were used to analyze the collected NMR spectra.

Initial investigation of the samples quality was made with the use of 1D ^{1}H and 2D ^{1}H- ^{15}N-HSQC spectra that were collected using the unlabelled and the ^{15}N single labelled samples, respectively. From the latter sample was also recorded a ^{15}N-NOESY (3D ^{15}N -NOESY) and 15N-TOCSY (3D ^{15}N TOCSY) spectra. The ^{15}N/^{13}C double labelled sample allowed us to record HNCO (b_hncogp3d), HN(CA)CO (hncacogp3d), HNCACB (hncacbgp3d), HNcoCACB (hncocacbgp3d), hCCH-TOCSY (hcchdigp3d2) and ^{1}H-^{13}C-HSQC spectra.

RodZn1-101 backbone sequential assignments were obtained from the analysis of ^{1}H-^{15}N HSQC and triple resonance HNCA, HNCO, HN(CO)CA, CBCA(CO)NH, and HNCACB experiments. Side-chains resonances were obtained with the analysis of ^{15}N HSQC- TOCSY, ^{15}N HSQC-NOESY, (H)CCH-TOCSY, and ^{1}H-^{13}C HSQC-NOESY spectra. All the assignment procedure has been described in detail in my final graduation thesis (**Pereira., 2011**).

The protons and carbons chemical shifts obtained from the previous assignment were compared with the values from the Biological Magnetic Resonance Data Bank (BMRB), a repository for data from NMR spectroscopy, and with the values from the Re-referenced Protein Chemical shift Database (RefDB), a data base of carefully corrected chemical shifts derived from the BioMagRes Bank. Secondary structural elements were predicted using the chemical shift index (CSI) for the measured values of CA, CB, CO and proton nuclei chemical shifts for each assigned residue (**Zhang et al., 2011; Ulrich et al., 2008**).

NOE spectra (^{1}H-^{1}H NOESY, 3D ^{1}H-^{15}N HSQC-NOESY and 3D ^{1}H-^{13}C HSQC-NOESY) combined with the previously assigned resonances were used as an input for structure determination. Initial models of the structure were calculated with UNIO 10' automated protocols that implements CYANA 2.1 algorithm (**Hermann et al., 2002; Guntert, et al,. 1997**).

In this procedure, residues ranging from 3-43 and 53-74 were used for the calculation of the RMSD. At the end of seven iterative calculation cycles, bundles of 20 structures were generated, together with upper limit constraint lists. These constraint lists were then used for further calculations of structures in explicit water using Amber suite and CNS-RECOORD protocols (**Nederveen, et al., 2005**). The quality of the calculated structures was evaluated using the iCING suit program (https://nmr.cmbi.ru.nl/icing/iCing.html) (**Jurgen et al., 2012**). The process of structure calculation, refinement and validation is described in more detail in the Methodology chapter (Chapter 2).

3.2.3 NMR titration for protein-protein interaction studies

To perform protein-protein interaction studies, a set of 2D ^{1}H-^{15}N HSQC (hsqcetfpf3gpsi) spectra were collected with different ratios of RodZ:MreB. The first spectrum was collected from a pure 1.33 mM RodZc1-101 solution sample, dissolved in a buffer containing 150 mM Sodium chloride and 50 mM phosphate at pH 7.6 with 5% D2O. Then, a solution of 100 uM MreB buffered with 100 mM Tis-HCl, 150 mM NaCl, 1mM EDTA and 2.5 mM desthibiotin with final pH of 8.0 was added to the RodZn solution with the following RodZ:MreB concentration ratios: 0.5:1; 1:1; 2.5:1; 5:1.

This set of data was collected using 2k (2048) ^{15}N and 128 ^{1}H complex points, and a spectral width of 12820.5 Hz (16 ppm) in the ^{15}N dimension and 24233.2 (30 ppm) in the ^{1}H dimension (**Palmer et al., 1991; Grzesiek et al., 1993; Kay, 1993); Schleucher et al., 1994**).

3.3 Results and Discussion

3.3.1 RodZn Structure analysis

RodZn1-101 construct is composed by the HTH moiety (1-78) plus a linker region and a 6 Histidine-tag (79-101), the latter one added for purification purposes.

Using a combination of the spectra described above (NMR data acquisition and structure calculation section), we were able to identify 76%/60% of the resonances, considering only the HTH domain residues (residues from 1 to 78) or the complete construct, respectively. The assignment summary is shown in **Table 1**.

A large number of residues could not be assigned, in particular 1-2, 47-50, 52, 79-83, 86-88 and 90-101. In the ^{15}N HSQC we could not detect signals that could be attributed for these residues. In the ^{13}C HSQC, a number of unassigned spin systems were identified, and their number and type was consistent with the missing residues. However, they did not excibit any NOE signals in the ^{13}C HSQC-NOESY and therefore they could not be unambiguously assigned to specific residues. Moreover, the absence of NOEs suggests that they are not structured and their presence would not have aided in the structure determination process.

In terms of the assignment of the remaining sidechaisn there is the notable, mention of the phenolic, HH proton of Y79 at 11.066 ppm that could be only be identified from the ^{13}C-NOESY and the 1D ^{1}H spectra. Its identity was confirmed by the absence of the heteronuclear splitting in the 1D-1H spectra of the ^{15}N and the ^{13}C-^{15}N samples implying that it was attached to an oxygen atom.

	RodZn (1 – 78)			RodZn (1 – 101)		
Shifts	**Found**	**Missing**	**Complete (%)**	**Found**	**Missing**	**Complete (%)**
1H	422	103	80	438	239	65
1H aliphatics	405	81	83	421	184	70
1H aromatics	17	22	44	17	55	24
15N (backbone)	64	14	82	67	34	66
15N (side chain)	4	23	15	4	40	40
15N (total)	68	37	65	71	74	70
13C'	67	11	86	87	14	86
13C aliphatics	304	51	86	318	127	72
13C aromatics	10	35	22	10	62	14
13C (Total)	314	86	79	328	189	63
CA	70	8	90	73	28	72
Backbone (HN H N C CA HA HA1 HA2 HA3)	338	55	86	353	156	69
Side chain	466	171	73	484	36	58
Triple shifts (HN H N C CA CB)	333	50	87	348	149	70
Total	804	226	78	837	502	63
Total (without pseudo atoms shifts = 78)	**726**	**226**	**76**	**759**	**502**	**60**

Table 1 - Report of the completeness of the Assignments of RodZn resonances from residue 1 to78, and for the full construct, separately.

Evaluation of the secondary structural elements of RodZn was first performed using only chemical shift information. This is possible because the values of CA, CB, CO and HA chemical shifts are sensitive to different secondary structures. We were able to probe the secondary structure of each aminoacid using various analysis, such as CSI and DANGLE, which are commonly accepted procedure to establish the secondary structure of proteins based on chemical shift differences with respect to some predefined standardized values that corresponds to specific secondary structure elements (**Wishart et al., 1994**). The helical regions of RodZn (in black) and its comparison with that of the homology model (in blue) is shown in **Table 2.**

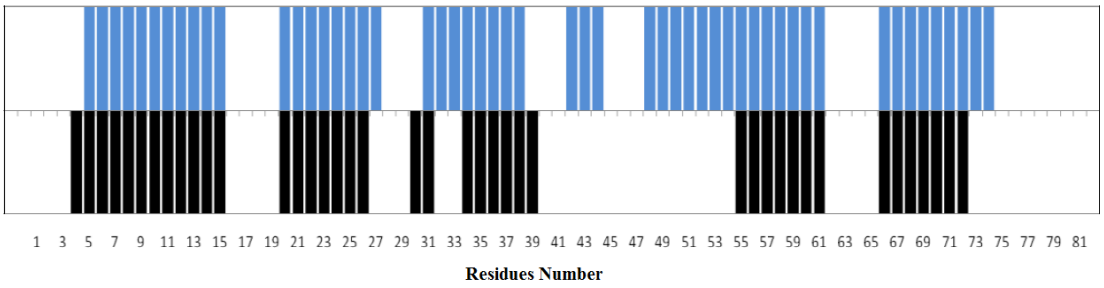

Table 2 - Comparison of alpha helical secondary structure of the Homology model obtained from SWISS MODEL (shown in blue colour) and the NMR based CSI derived of RodZn$_{1-101}$ (shown in black colour). Each bar represents the tendency of each assigned residues for a helical secondary motif. Residues that show no bars, either no assignment was available or CSI predicted it to be in a coiled-coil region of the protein.

As expected, we see that RodZn has no β-sheet secondary elements. Instead, it shows six putative α-helix motifs intercalated with loops regions. The CSI result for RodZn is in agreement with the secondary structure of the predicted model for a putative HTH motif by Swiss Model (shown in **Figure 10**). HTH is a common motif for DNA binding proteins, which gives an indication that this domain may interact with DNA. In my graduation thesis we were able to conclude that the predicted model was actually very close to the RodZn secondary structure determined with CSI.

Since we were unable to find backbone resonances for residues from 45 to 55, CSI didn't give any prediction about the secondary structure in this region. So, contrary to the predicted model from Swiss Model, this region might be indeed unstructured. To make further analysis, we moved on to the determination of the tertiary structure of RodZn.

After the assignment of the backbone and side chain atoms, this information was used to identify the NOE peaks in the 2D ^{1}H-^{1}H NOESY, ^{1}H-^{15}N HSQC-NOESY and 3D ^{1}H-^{13}C HSQC-NOESY spectra. We used all the NOESY spectra mentioned above (in NMR data acquisition and structure calculation

section) and the chemical shifts assignment obtained as input in the automated structure determination protocol of UNIO10 that used CYANA for the structure calculation.

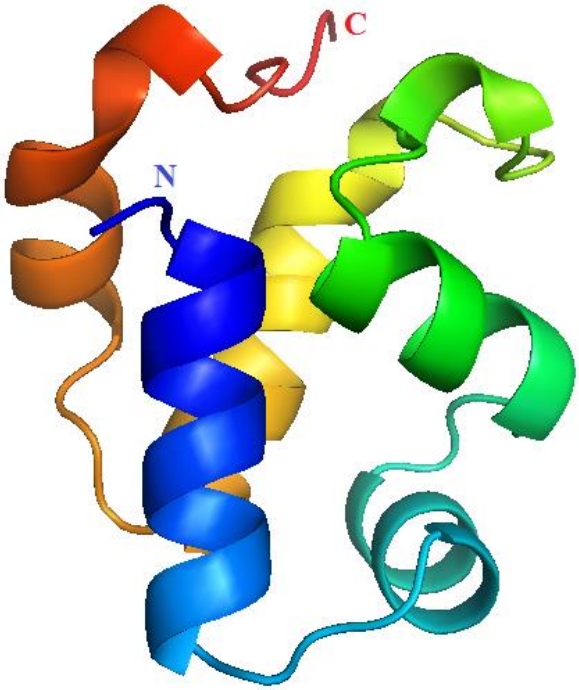

Figure 10 - RodZ schematic structure predicted with SWISS-MODEL (an automated protein structure homology-modeling server; Schwede et al., 2003).

This automated structure determination protocol generated a list of distance restraints which were then used as constraints in a series of 100 simulated annealing calculations. The 20 structures with the least number of NOE violations were kept for evaluation. However, the quality of the structure at this stage was low since the manual assignment contained many incorrect restraints due to the large overlap of ^{1}H chemical shifts. So, the bundle of structures previously generated was used as model for a more correct assignment of the NOEs in a second iteration of the calculation. This in turn led to a more correct assignment of the NOEs and to better quality structures. This process was repeated 7 times before the assembly of the final structure.

Further refinement of the structure was achieved with the introduction of Hydrogen bonds restraints that were added after the previously obtained structure-bundle was used as a guide for identification of h-bond candidates. These distances were picked very conservatively in respect to the obtained structural elements for which the NOE data had driven the models. Together with the dihedral angles calculated by TALOS+, these two new sets of constraints were explicitly included in future UNIO calculations. TALOS+ is an algorithm that predicts protein *phi* and *psi* backbone torsion angles using a combination of chemical shift assignments for a given residue sequence (HN, HA, CA, CB, CO and N chemical shifts).

In summary, 1404 NOEs were used in the calculation, of which 695 were short-range (354 intra-residual and 341 i_{+1}), 342 were medium-range and 367 were long-range. This corresponded to an average

of 16 NOEs per residue for the structured regions. In addition, 22 H-bond and 84 dihedral angle constraints were included in the calculation. Remaining restraints and RMSD values are shown in **Table 3**.

Number of residues	101 (1-101)
Molecular weight	11683.93 Da
Number of models	20
Setup−given RMSD range	3-43, 53-74
• **Backbone RMSD [A]**	0.36 +/− 0.11 (0.24..0.64)
• **Heavy atom RMSD [A]**	0.77 +/− 0.08 (0.66..0.96)
Optimal RMSD range	1−47,51−74
• **Backbone RMSD [A]**	0.67 +/− 0.24 (0.35..1.26)
• **Heavy atom RMSD [A]**	1.02 +/− 0.19 (0.80..1.53)
NOE restraints	1404
• **intraresidual (|i−j|=0)**	354 (25.21%)
• **sequential (|i−j|=1)**	341 (24.29%)
• **medium−range (1<|i−j|<5)**	342 (24.36%)
• **long−range (|i−j| >4)**	367 (26.14%)
NOE restraints per residue	13.90
RMS NOE restraint violation [A]	0.0254
Dihedral restraints	479
RMS dihedral restraint violation [o]	0.5951
Ramachandran statistics	
• **most favoured**	59.51 %
• **additionally allowed**	32.86 %
• **generously allowed**	5.66 %
• **disallowed**	1.98 %

Table 3 - Final statistics of RodZn structure calculation. Output from UNIO10 software (Serrano et al., 2012).

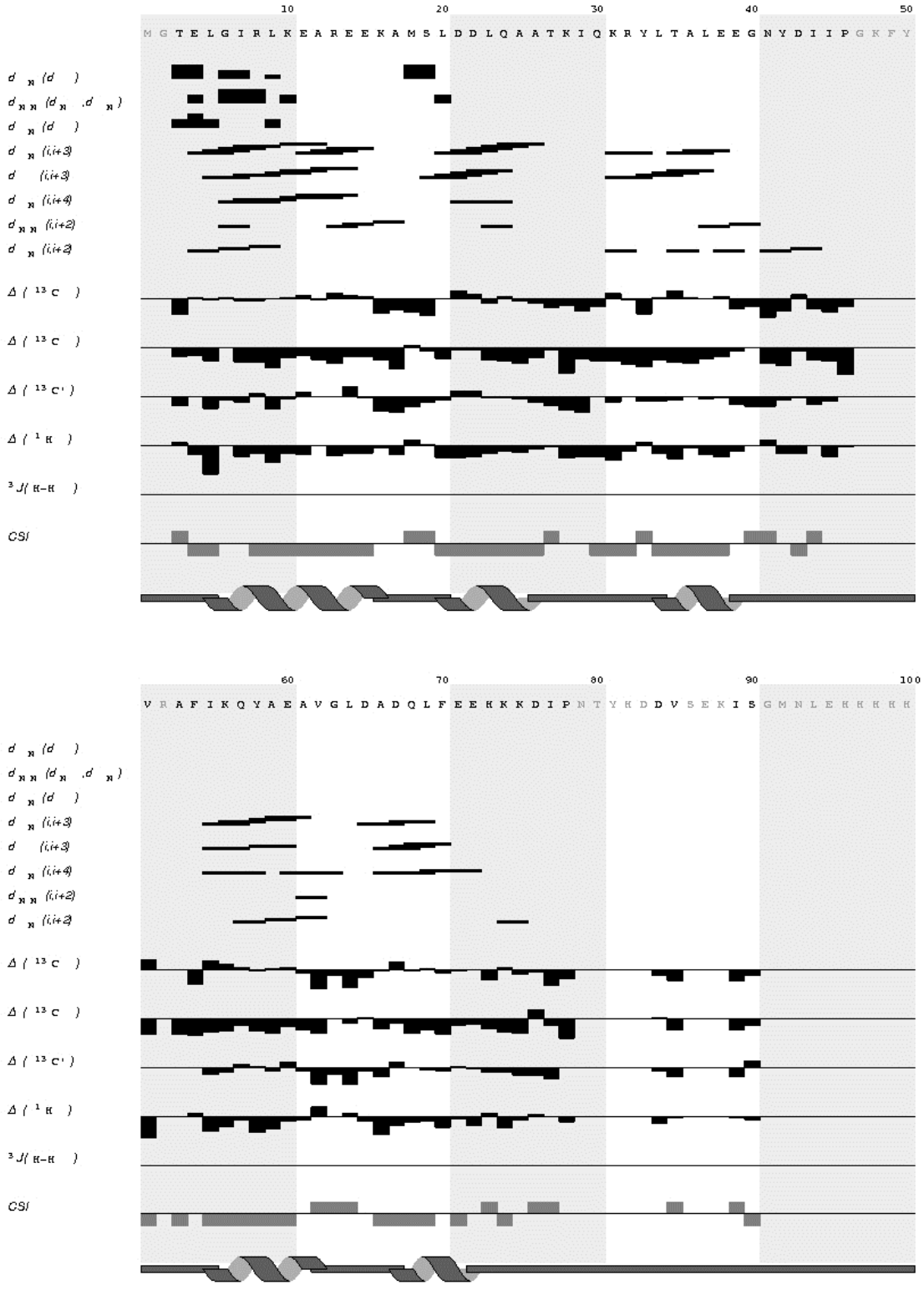

Table 4 - Experimental restraints for RodZn protein, including sequential, short- and medium-range NOEs and HA, CA, CB and CO secondary shifts along with the secondary structure deduced from the data. The amino acid sequence and numbering are shown at the top. Sequential N-N NOEs are indicated

by black bars; the thickness of the bar represents the strength of the observed NOE. The presence of medium-range N-N NOEs is indicated by solid lines. The chemical shift indices calculated from Cα, Cβ, CO and Ha are also shown by black bars at the bottom. The locations of the secondary structure elements identified in the calculated family of structures are shown at the bottom.

At this stage we obtained a bundle of structures with good RMSD. In **Table 4** we can see some of the restraints used for RodZn structure calculation and their correlation with secondary structure

The resulting structures were refined in explicit water calculations using Amber suite and CNS-based RECOORD protocols, with distance and dihedral angle restraints as additional input. The lowest energy structures obtained with RECOORD (bundle of 20 structures) and Amber (one final structure) corresponds to the models with the lowest number of NOE violations, thereby composing the final bundle of NMR structures. These structures were subject of evaluation using CING and are to be de deposited in the PDB database.

In **Table 5** we see a comparison of the quality of the obtained structures between the different refinement methodologies.

Output evaluated	RECOORD	AMBER1	AMBER2
Residues range		**3 - 43 , 53 - 74**	
Number of models	20	1	1
	CING		
RMSD			
Backbone Average	0.49 +/- 0.11	-/-	
Heavy atoms average	0.95 +/- 0.13		
	WHAT IF		
Z-scores			
Ramachandran plot appearance	-2.944 +/- 0.622	-3.346	-3.683
Backbone conformation	- 0.581 +/- 0.608	-0.946	-1.034
chi-1/chi-2 rotamer normality	-5.465 +/- 0.429	-7.942	-7.535
RMS Z-scores			
Bond lengths	1.151 +/- 0.009	1.164	1.162
Bond angles	0.560 +/- 0.017	0.913	0.941
Side chain planarity	0.902 +/- 0.116	1.418	1.200
Improper dihedral distribution	0.947 +/- 0.048	1.410	1.303
	PROCHECK		
Ramachandran statistics			
core	85.9 %	81.7 %	83.3 %
allowed	11.4 %	15.0 %	16.7 %
generous	2.1 %	3.3 %	0.0 %

disallowed	0.6 %	0.0 %	0.0 %

Table 5 – Resume of the evaluation of the obtained structures with RECOORD and AMBER refinement methodologies performed with different sets of restraints. Evaluation output from ICING server. RECOORD and AMBER1 structures were obtained with the full set of restraints (NOE distance restraints, hydrogen bonds restraints and dihedrals restraints in a total of 1777 restraints). AMBER2 structure was obtained with the full set of restraints except the dihedral angles restraints.

The Ramachandran plot obtained from the evaluation with PROCHECK tool shows the phi-psi torsion angles for all residues in the structure (charts not shown), with colour coded rankings where red areas correspond to the "core" regions representing the most favourable combinations of phi-psi values (**Morris et al. (1992)**. Ideally, one would hope to have over 90% of the residues in these "core" regions. The percentage of residues in the "core" regions is one of the best guides to stereochemical quality. According to PROCHECK, RECOORD refinement shows the highest statistics for residues in the "core" region (85.9% of the total residues). However, AMBER2 structure refinement corresponds has all residues in the "core" and "allowed" regions (83.3 % in "core" region and 16.7 % in "allowed" region).

WHAT Ifs output shown in **Table 5** is represent by Z-scores and RMS Z-scores. When the exact values for all bond lengths and bond angles are used as input, there is a set of possible deviations to these values. If assumed that the variation in-between these values is random, then we can determine how many standard deviation each bond length or angle deviates from the ideal value, and report those deviations. The number of standard deviations that any observed value deviates from the ideal average is commonly called the Z-score of that observed value. Another useful characteristic of a normal distribution is that the RMS-Z score is exactly 1.0. So, if the determined RMS-Z score for all bond lengths is higher than 1.0, it means that the restraints on the bond lengths were to weak during refinement. On the other hand, when the RMS-Z score is lower than 1.0, it means that the restraints should have been relaxed. So, a RMS-Z score >1.0 shows an error and a score <1.0 only issue a warning. Z-scores of the structure refined with RECOORD are unquestionably the best, showing lower deviations from the mean in comparison with the structures refined with AMBER. The same happens with RMS Z-scores. Even though bond lengths RMs Z-scores are higher than 1.0 for all the refinements (meaning that there might be errors in the structure), RECOORD refined structure bundle has the best RMS Z-scores for bond angles, side chain planarity and dihedral distribution. So, the bundle of 20 structures refined with RECOORD will be subject of further validation and uploaded into PDB database.

By inspecting the resulting mean-3D model from the bundle of 20-low energy structures refined with RECOORD shown in **Figure 11 A)**, we see that RodZn is composed of 5 α-helices (H1 to H5). From H1 to H3 we can observe an arrangement in a trigonal fashion which makes this domain very stable due to electrostatic interactions between them. H5 forms an antiparallel trigonal motif with H1 and H4. Also, H3 is connected to H4 via an extended unstructured peptide chain. This extended loop

composed by residues ranging from 41 to 54, approximately, corresponds to a region where very few assignments could be made resulting in low confidence in the calculated structure in that region. For the same reasons, the region from 76 to 101 is completely unstructured (the latter residues were removed from the figure for better interpretation of the resulting structure).

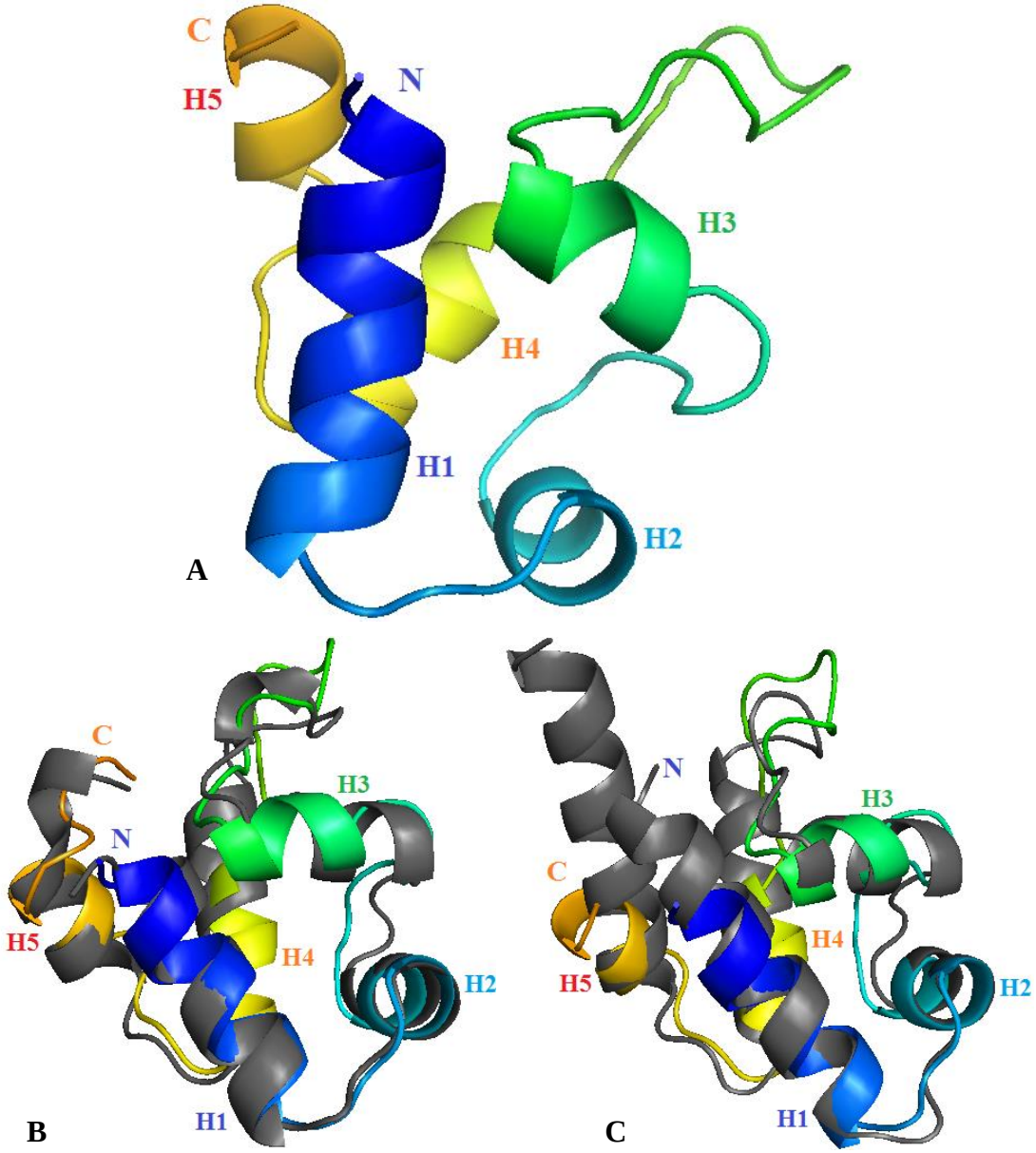

Figure 11 – A) Lowest energy-3D model retrieved from the calculated bundle of 20-low energy structures of RodZn from *B. subtilis* refined with RECOORD. Notations: N corresponds to the N-terminal of RodZn; C corresponds to the C-terminal of RodZn; the helices from the helical moiety are identified (from H1 to H5); this model is shown overlapped with HTH moiety from RodZ of *S. aureus* and in B) with the one from RodZ of *T. maritima* in C). RMSD of the refined model to the RodZ HTH moiety from *S. aureus* and *T. maritima* are 1.351A (for 61 aligned atoms) and 1.497 A (for 56 aligned atoms), respectively.

As previously mentioned, there are two published structures of homologous proteins to RodZ-N originating from the gram-positive *S. aureus* (PDB code: 3FYM) and a complex with MreB from gram-negative *T. maritima* (PDB code: 2WUS).

The structure obtained for RodZn from *B. subtilis* is very similar to both from *S. aureus* and *T. maritima*, with an RMSD of 1.351A (for 61 aligned atoms) and 1.497 A (for 56 aligned atoms), respectively. The RMSD values were determined by CEalign algorithm built into PyMOL molecular graphics software (**Bramucci et al., 2012**). This software performs the optimal superposition of two objects of equal length.

The main difference between RodZn from *B. subtilis* and *S. aureus* is located in the extended loop between H3 and H4, especially in the region comprised between residues 51 to 58 where the RodZn structure from *S. aureus* shows a tendency for a helical secondary structure, while in RodZn from *B. subtilis* appears to be a fully extended coil (**Figure 11 B**). It is possible that this region exists in equilibrium between helical and random-coil in solution resulting in the poorly defined result that we have obtained, while the crystallized structure from *S. aureus* stabilized the helical form. By comparing *B. subtilis* RodZn structure with the one from *T. maritima* complexed with MreB, it revealed slightly more differences. H1, H4 and H5 of *B. subtilis* RodZ are shorter than their structural equivalents in the *T. maritima* protein (**Figure 11 C**). In conclusion, *S. aureus* and *B. subtilis* RodZn are more similar with each other than the homologous structure from *T. maritima*. It is not clear however if these differences are due to the interaction with MreB or if they represent a more widespread difference between the gram positive and the gram negative species.

The calculated RodZn structure shows that residue Y33 is located in the beginning of H3, as for the homologous residue of RodZn from *T. maritima* (K36) (**Van den Ent et al., 2010**). However, while Y53 and Y57 residues in *T. maritima* RodZ are found in H4, the homologous residues Y50 and F54 in *B. subtilis* are located in a flexible loop, connecting H3 and H4 (**Figure 12**). Also, the side chains of Y33, Y50 and F54 in *B. subtilis* RodZn are solvent-exposed (**Figure 12 A**). In the microbial development laboratory from ITQB, each of these residues from RodZn of *B. subtilis* was individually substituted by an alanine and the effect of the various mutations investigated by fluorescence microscopy (data not shown; to be published).

None of the alanine substitutions compromised the accumulation of RodZn in *B. subtilis* cells. It assumed that none of the substitutions grossly affected the overall folding of the protein. However, all of the mutations led to a production of shorter and wider rod cells, compared to the wild type strain phenotype. Moreover, unlike the H3 substitutions (Y33A), the Y50A and F54A substitutions in H4 caused delocalization of RodZ, which was shown to be distributed uniformly along the membrane, including polar regions and division septa of the cell. These results indicate that the Y33, Y50 and F54 residues are functionally important, as their substitution to alanine led to a phenotype similar to the one obtained from deletion of full *rodZ* or upon deletion of its cytoplasmic domain.

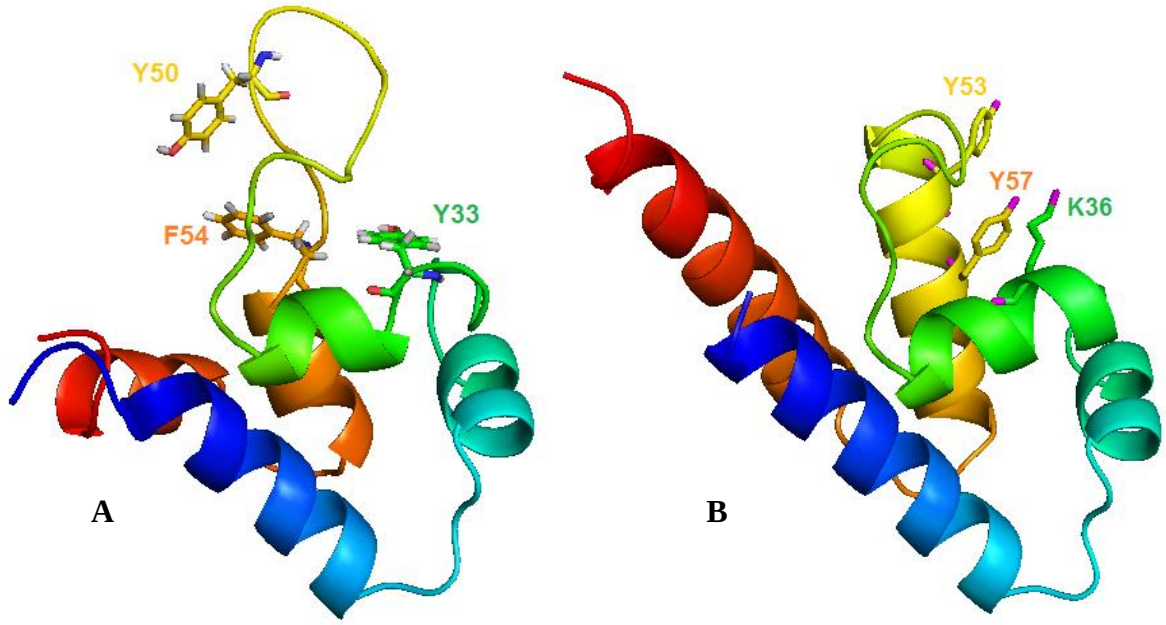

Figure 12 – A) Lowest energy-3D model retrieved from the calculated bundle of 20-low energy structures of RodZn from *B. subtilis* refined with RECOORD. Residues Y33, Y50 and F54 are highlighted from the ribbon cartoon as sticks; B) X-ray structure of RodZ HTH moiety from *T. maritima* (PDB code: 2WUS). Residues K36, Y53 and Y57 are highlighted from the ribbon cartoon as sticks. (Van den Ent et al., 2010).

Residues K36 (in H3), and Y53 and Y57 (in H4) are critical for the interaction of RodZ with MreB in *T. maritima*.(V**an den Ent et al., 2010**). Unpublished results from the Microbial development laboratory show that H4 of *B. subtilis* RodZn also carries essential determinants for the localization of the protein, and it is likely that H4 also mediates an interaction with MreB. Because the mutations in H3 did not affect the ability of RodZ to localize properly, it is possible that these residues make minor contributions to the RodZ-MreB interaction or that in *B. subtilis* different residues are involved.

In order to understand which residues of RodZn from *B. subtilis* may be promoting an interaction with MreB, a study was conducted where changes in the 2D ^{1}H-^{15}N HSQC NMR spectra were analyzed through NMR titration upon addition of different ratios of RodZ:MreB to the NMR tube. Despite the fact that the binding of MreB with RodZ has a low affinity (K_d = 5 – 10 µM), this study was performed with higher concentrations of both proteins. Also, the different buffering pHs allowed a higher stabilization for each protein in solution. Even so, no direct interaction between the two purified proteins was detected, i. e. no localized chemical shifts changes were observed (data not shown). These results ended up being inconclusive.

3.4 Conclusion

It is clear that the interaction of RodZ with MreB in Gram-negative organisms involves the cytoplasmic N-terminal domain of the protein (**Bendezu et al., 2009; Van den Ent et al., 2010**). The structure of the complex between RodZ and MreB from *T. maritima* shows that residues in H3 and H4 of the HTH domain mediate the interaction with MreB, with H4 contributing most of the contacts (**Van den Ent et al., 2010**). In particular, residues Y53 and Y57 in H4 were found critical for the interaction of the *T. maritima* proteins in vitro, and the homologous residues in RodZ from *E. coli* (F60 and Y64) were found essential for cell shape, proper localization of the protein, and direct binding to MreB (**Van den Ent et al., 2010**).

Keeping with a possible interaction of *B. subtilis* RodZ with MreB, deletion of the entire RodZn domain, or single alanine substitutions of the H4 residues Y50 and F54, also caused the evenly localization of the protein along the cell membrane. Even though it seems likely that gram-positive *B. subtilis* RodZ directly interacts with MreB, we were unable to detect a direct interaction between the two purified proteins through NMR titration experiments.

We cannot presently exclude that an interaction between RodZ and MreB only occurs in the context of the membrane. Lending support to this idea, the transmembrane domain (TM) of *B. subtilis* RodZn as well as its counterpart in *E. coli*, is important for protein localization and cell elongation (**Bendezu et al., 2009; Shiomi et al., 2008**). Also, insertion into the membrane may lead to dimerization of RodZ via the TM domain and/or additional parts of the protein, which may allow proper interaction with MreB. On the other hand, the cytoplasmic domain of *T. maritima* RodZ is monomeric, and forms a 1:1 complex with MreB (**Van den Ent et al., 2010**).

Regardless of the oligomeric state of RodZ, it is possible that other unidentified factor may be promoting the interaction of RodZn with MreB in *B. subtilis*. Unpublished results from the Microbial development laboratory in ITQB show that in the absence of RodZ, the nucleoid is expanded. These observations, together with recent findings that in *E. coli* RodZ is required for DNA secretion, cell-to-cell transfer of non-conjugative plasmids, and mRNA processing at the membrane level, suggests that RodZ could indeed promote a link between the chromosome and the cell membrane (**Kurono et al., 2012; Mitobe et al., 2011; Sanchez-Torres et al., 2010**). An obvious limitation to this interpretation is that in the complex between RodZ and MreB from *T. maritima*, the HTH motif (H2 and H3) in RodZ and the adjacent H4 are used to promote interaction with MreB. Consequently, upon binding to MreB, this actin-homologue ends up blocking the HTH motif thereby preventing a putative interaction with DNA (**Van den Ent et al., 2010**).

Even so, previously published data shows that in *E. coli* the linker that connects RodZn HTH moiety to TM domain is sufficient to mediate an interaction with MreB and so, we do not exclude the hypothesis that in vivo, RodZ may interact with both MreB and DNA (**Bendezu et al., 2009**).

As we've seen in the structure obtained of RodZn from *B. subtilis*, residues from 51 to 58 are in an extended fashion, contrary to RodZn structure from *S. aureus*. It is possible that this region exists in a dynamic equilibrium between helical and coil structure upon binding to MreB. Moreover, the interaction between RodZ and MreB may obey to a somewhat different geometry in *B. subtilis.* In *T. maritima*, the K36A substitution severely impaired the interaction in vitro between RodZ and MreB proteins and the corresponding substitution in *E. coli* impaired the interaction with MreB and also affected the localization of RodZ (**Van den Ent et al., 2010**). However, the equivalent substitution in *B. subtilis* (Y33A in H3) did not cause mislocalization.

In contrast, the H4 substitutions Y50 and F54, like their equivalents in *E. coli*, caused the evenly distribution of the protein along the membrane. It is seems plausible that in *B. subtilis*, RodZ forms a triple complex with MreB and DNA in which H3 would be mainly involved in RodZ-DNA contacts with H4 providing critical determinants for the interaction with MreB.

The *Bacillus thuringiensis* FtsZ-like NTPase TubZ and the DNA-binding protein TubR, both involved in plasmid segregation establish a precedent for a triple complex formed by DNA, a DNA-binding, and a cytoskeletal protein. TubR consists of a dimer with an HTH motif. The TubR recognition helices mediate dimerization, making canonical HTH-DNA interactions impossible. Instead, DNA binding involves insertion of the N-terminus of the recognition helices in each of the dimers into a single DNA groove and insertion of the wings into adjacent DNA grooves. TubZ then interacts with TubR-DNA, promoting segregation of the plasmid bound TubR through polymerization of TubZ (**Ni et al, 2010)**.

Since, RodZ appears to be much more abundant in *B. subtilis* than in *E. coli* (**Bendezu et al., 2009; unpublished data from Microbial Development Lab. From ITQB),** another possible view is that some molecules will bind to MreB whereas other bind to DNA. The prospect that RodZ binds directly to both MreB and the DNA is attractive and remains an important research goal.

Chapter 4

RodZ periplasmic domain: Structural and functional insight

4.1 Introduction

As reviewed in the introductory chapter (see Chapter 1), the interaction between MreB and the cytoplasmic domain of RodZ in the inner surface of the cell membrane is required for maintaining cell shape. However, additional interactions of MreB with membrane proteins such as MreC and MreD appear as important factors in its ability to link the internal and external parts of the cell wall elongation machinery. Like MreB, MreC and MreD are encoded in the mreBCD operon and are both required for cell shape maintenance (**White and Gober, 2012**). MreD is an integral membrane protein. As for MreC, it has a transmembrane domain followed by an extracytoplamic moiety, and is thought to form polymeric structures outside the cell (**Van den Ent et al., 2006; Lovering and Strynadka, 2007**).

Interactions of MreC and MreD with components of the elongation system that lie outside the cell membrane have also been documented (**Divakaruni et al., 2005; Ghachi et al., 2011; van den Ent et al., 2006; Kleinschnitz et al., 2011**). MreD has been shown to interact not only with MreB but also with MurG enzyme which is involved in the peptidoglycan synthesis (as described in Chapter 1). And dimeric forms of MreC from *B. subtilis* have been show to directly interact with MreD (**Van den Ent et al., 2006**). Recent studies in *E. coli* and *S. coelicolor* have revealed direct interactions of RodZ with MreC and MreD (**Bendezu et al., 2009;Kleinschnitz et al., 2011**).

Therefore, RodZ seems to engage proteins of the cell elongation machinery on both sides of the membrane, as well as in the plane of the membrane. Interactions within the membrane are likely to be mediated, at least in part, by the TM domain, which is essential for the function of RodZ in *B. subtilis*. On the other hand, interactions with MreC are likely to involve the extracytoplamic region of RodZ protein (RodZc).

To date, the structure of RodZc has never been reported. Therefore the determination of its structure is already a challenge in itself. In this chapter we will present our data driven structural model for the C-terminal moiety of RodZc, predicted with CS-ROSETTA as well as a description of its dynamic nature as studied by relaxation measurements.

4.2 Material and methods

4.2.1 Bioinformatic pre-studies

A profile of RodZc protein was made using bioinformatic tools available on internet. The sequence of RodZc was analyzed using PSIPRED server (www.bioinf.cs.ucl.ac.uk), predicting its secondary structure and disordered regions, as well as recognition of a potential fold (**McGuffin et al., 2000**).

SWISS MODEL (http://swissmodel.expasy.org/) and I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) servers were used as an attempt to predict a 3-dimensional model of RodZc (**Schwede et al., 2003; Roy et al., 2010**).

All bioinformatic tools used in these initial studies are described in detail in the methodology chapter (Chapter 2).

4.2.2 Bacterial Growth and Protein Expression

All cloning and bacterial transformation steps were performed by Ana Paiva from the Microbial and Development Laboratory, ITQB. *B. subtilis* strains and plasmids used in this study are provided in detail in the Appendix chapter. The protocol for expression and purification provided by the aforementioned laboratory was subject of improvement in the present work.

The *Escherichia coli* strain BL21(DE3) was employed for the over-production of all recombinant proteins. A glycerol stock (8% glycerol) of BL21(DE3) *E. coli* cells carrying pTC248 (AH4268 strain) was produced and stored at -80ºC for further expression and purification trials. Plasmid pTC248 codes for a fusion between the His_6 tag and residues 131-304 of RodZ (His_6-RodZc131-304). Expression of the fusion proteins coded in pTC248 is under control of the $T7_{lac}$ promoter (**pET System Manual, Novagen**) therefore it is inducible by IPTG.

The various steps regarding the expression and purification of this protein construct were analysed using Tris – Glycine 12% polyacrylamide gel electrophoresis under denaturing conditions (SDS-PAGE). Protein sample stability was routinely checked by both 12% Tris-Glycine SDS-PAGE electrophoresis and analytical size exclusion chromatography (ASEC).

A pre-culture was prepared by putting 1ml of the glycerol stock in 100 ml of LB medium containing 50µg/ml kanamycin. The culture was grown overnight in a shaker (Gallen Kaup Orbital Incubator) with controlled agitation of 100 rpm at 37ºC. The pre-culture was then used to make a large scale culture in a ratio of 1:100 of pre-culture in a fresh LB medium enriched with 50µg/ml kanamycin. The culture was placed in a shaker (Minitron Infors HT) at 150 rpm at 37ºC. The cells were grown to an OD600 between 0.4 and 0.6, at which time the cultures were induced with 1mM IPTG for 4 hours.

The cells were collected by centrifugation (Avanti J-25I and J-26I centrifuges with JA-10 rotor in Polycarbonate bottles from BECKMAN-355605), at 8000 g at 4ºC and the pellet was stored at -80ºC for future work.

Over expression of isotope-labeled proteins: For overexpression of RodZc construct uniformly labeled with single ^{15}N isotope and double $^{13}C/^{15}N$ isotopes, bacterial cultures from the glycerol stock were grown and expressed in the same condition as previously described but in Minimal medium in which ^{15}N labeled (99%) Ammonium chloride ($^{15}NH4Cl$) and ^{13}C labeled (99%) Glucose was the only nitrogen and carbon sources respectively. Both isotopically labeled reagents were acquired from Cortecnet.

4.2.3 Protein Purification

After defrosting, cells were re-suspended in Lysis Buffer (**Appendix, Table 1**) and then disrupted three times in a Thermo Scientific French Pressure Cell with 1000 psi pressure at room temperature. The obtained lysate was centrifuged for 30 minutes at 50,000 g at 4ºC (Ultracentrifuge Optima LE-80K with rotor 70TI in Polycarbonate bottles from BECKMAN - 355618). The pellet was discarded and the clear supernatant was diluted with IMAC Buffer A until it reached 80 mL of volume.

Immobilized metal affinity chromatography (IMAC) : The target protein was purified using a 5 mL HisTrap FF crude from GE Healthcare ready to use column, pre-packed with a highly cross-linked spherical agarose matrix (Sepharose 6%) with an immobilized chelating group. The medium had already been precharged with Ni^{2+} ions. The column was firstly cleaned with water and subsequently equilibrated with IMAC Buffer A. The supernatant was then loaded into the column and a linear gradient of IMAC Buffer B containing Imidazole was applied to elute the sample components from the column.

A GE Healthcare Akta Purifier 10 FPLC device was used as a controller to optimize this purification step, following the manufactures instructions. The column was operated at a flow rate of 1.5 ml/min with column maximum pressure of 0.3 MPa.

Size Exclusion Chromatography (SEC): A second purification step was preformed. The relevant protein fractions from the IMAC step were pooled together and concentrated using a 3kDa Vivaspin 15R device from Sartorius stedim biotech with a relative centrifugal force of 5000 xg at 4ºC until it reached an appropriate sample loading volume (<5 mL). After that the sample was injected in the HiLoad 16/60 Superdex 75 prep grade size exclusion column from GE Healthcare previously equilibrated with SEC Buffer. This column is prepacked with a dextran matrix covalently bound to highly cross-linked agarose with separation range of globular proteins between 3,000 and 70,000 Da. The same FPLC device previously described was used as a controller of this purification step. The column was operated at a flow rate of 1.0 ml/min with column maximum pressure of 0.3 MPa.

4.2.4 NMR sample preparation

The final sample collected from the size exclusion chromatography was concentrated to produce a suitable NMR sample. The sample was concentrated using a 3kDa MWCO Vivaspin 15R device previously described. And the protein concentration measured spectrophotometrically with Nanodrop ND-2000C equipment (**NanoDrop 2000/2000c Spectrophotometer. V1.0 User Manual. Thermo Scientific**), at a wavelength of 280 nm using an extinction coefficient of 12950 M-1 cm-1 (or ε2780.1%=0.572 mg-1 cm2) determined with ExPASy ProtParam tool (**Wilkins et al., 1999**). The absorbance was corrected for the contributions from the corresponding buffering solution (**Stoscheck, 1990**).

Through Analytical Size Exclusion Chromatography (ASEC), the concentration conditions were also optimized to reduce aggregation and degradation of the sample to achieve the highest possible protein concentration. RodZc131-304 protein in ASEC buffer, was applied in a range of concentrations (from 110 to 982 µM), to a Superdex 75 HR10/30 column (GE Healthcare) using an AKTA Purifier UPC system. The column was operated at a flow rate of 0.7 ml/min with column pressure of 1.8 MPa during operation. (Protein MW markers from GE healthcare instructions manual)

Sample preparation for structure determination and relaxation experiments: The protein samples were set to contain approximately 1mM RodZc133-304 in 90% H2O and 10% D2O needed for signal lock and 1mM of NaN_3 as a bacteriostatic agent (**Lichstein et al., 1943**), with a final pH of 6.8. The buffer used for sample preparation had the same composition as the SEC Buffer.

For Relaxation measurements a 0.5mM sample was used to reduce the possibility of sample aggregation and decrease the solution viscosity

Purified protein samples were loaded into a 5 mm NMR tubes with model SL-5 from New Era Enterprise. These tubes are made of Type 1, Class B Borosilicate glass specific to work at the highest magnetic fields (between 700-900+ MHz) and ideal for structural biology research (Specifications: 4.960 ± 0.006 mm OD; 0.40 ± 0.001 mm nominal wall; 0.0025 ± 0.002 mm roundness; 178 mm of Length).

Sample preparation for RDCs measurements: The protein samples were set to contain approximately 0.5 mM RodZc in 90% H2O and 10% D2O needed for signal lock and 1mM NaN_3, with a final pH of 6.8. The buffer used for sample preparation had the same composition as above.

The produced sample was used to soak a dried piece of polyacrylamide gel. The polyacrylamide dried gel originated from a 4.5% gel that was prepared followed the protocol improved by Martina Fonseca (Biomolecular NMR Laboratory, ITQB) with some modifications. A 4.5% polyacrylamide solution was prepared by diluting a 30% (w/v) stock (Rotiphorese Gel 30-37.5:1- ROTH) with a TRIS buffer solution and then degassed with N2 for 30 minutes. The polymerization reaction was initiated by adding the required volume of a 10% ammonium persulfate (APS) freshly prepared solution and of the catalyst, N,N,N′,N′-Tetramethylethane-1,2-diamine (TEMED). The polymerizing solution was quickly transferred and cast in a gel cylinder device with an internal diameter of 6mm. When polymerization was complete, the gels cylinder were removed and washed repeatedly with water overnight, in order to remove unpolymerized reagents and remove the buffer that might interfere with biological samples. The gel was cut into pieces of approximately 1.0 cm in length and laid in a petri dish to completely dry.

For the stretched sample preparation, the gel was left to swell to its original size in a 2 mL eppendorf tube containing 0.500 ml of the purified protein sample for 48 hours for the protein molecules to diffuse into the polyacrylamide gel. The gel was placed in an open-ended NMR tube by forcing the cylindrical gel to go into the NMR tube with a 4.25mm internal diameter achieving a compression ratio of 4.2: 6.0= 0.7. After the gel has been squeezed into the NMR tube, the pores within the gel on average will be elongated along their long axis parallel to the NMR tube. Thus, when placed in the magnet, the protein molecules in the aqueous phase of the gel will tend to have their long axis parallel to the magnetic field. In **Figure 13** is presented a scheme of the apparatus used in the described process

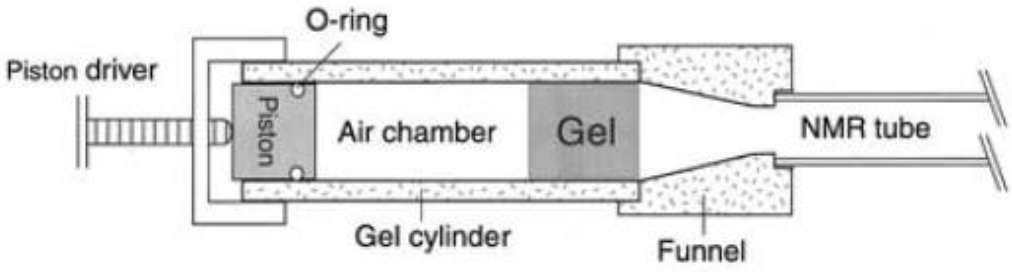

Figure 13 - Schematic drawing of the apparatus for stretching the gel and inserting it in the open-ended NMR tube. The funnel-like device used for radial compression of the gel consists of four pieces: the funnel, the gel cylinder, and the piston, all made of Teflon, and a brass piston driver. Loading apparatus developed by Chou et al., 2001.

4.2.5 NMR data acquisition and 3D-model prediction

NMR experiments were preformed on a Bruker AvanceIII spectrometer (Bruker, Rheinstetten, Germany) with ^{1}H operating frequency of 800.33 MHz at the ITQB - António Xavier Magnetic Resonance Center, CERMAX, Nova University of Lisbon, Portugal. The spectrometer was equipped with a triple resonance pulsed field TXI Z axis gradient H C/N/-D (5 mm) probe.

All data was acquired at 298.15 K unless otherwise stated, and processed with Bruker's Topspin 2.1 software (Bruker Biospin). Both CARA (v1.8.4.2) and CCPN (v2.2.1 to v2.2.3) software were used to analyze the collected NMR spectra.

Data acquisition for structure determination: In order to calculate the NMR solution structure of RodZc131-304 protein construct, several NMR spectra were recorded using unlabeled and both uniformly single 15N labeled and double $^{15}N/^{13}C$ labeled samples.

Initial investigation of the samples quality was made with the use of 1D ^{1}H spectra collected using unlabelled protein samples.

The backbone sequential assignments were obtained from the analysis of a 2D ^{1}H-^{15}N HSQC experiment (hsqcetfpf3gpsi), collected using 2k (2048) ^{15}N and 180 1H complex points, and a spectral width of 12820.5 Hz (16 ppm) in the ^{1}H dimension and 2757.6 (34 ppm) in the ^{15}N dimension (**Palmer et al., 1991; Schleucher et al., 1994; Grzesiek et al., 1993**).

And also through analysis of triple resonance HNCO (b_hncogp3d) (**Schanda et al., 2009; Melckebeke and Brutscher et al., 2006; Lescop et al., 2007; Grzesiek and Bax, 1992**), H(CA)CONH (**Folmer et al., 2000**) , HN(CO)CA (b_hncocagp3d) (**Schleucher et al., 1993**) HNCA (b_hncagp3d) (Schleucher et al., 1993) , and HN(CO)CACB (cbcaconhgp3d) (**Grzesiek et al., 1993; Muhandiram et al., 1994**) experiments.

Side-chains resonances were obtained with the analysis of triple resonance ^{15}N-TOCSY-HSQC spectrum (dipsihsqcf3gpsi3d) (**Davis et al., 1992**) recorded with a spin lock time of 20ms, and using 120 and 1 points in the indirectly detected ^{1}H and ^{15}N dimensions, respectively, and 2k points in the directly detected ^{1}H dimension. The spectral width used were 10004.2 and 2757.6 Hz in the indirectly detected ^{1}H and ^{15}N dimensions and 12820.5 Hz in the ^{1}H direct dimension; ^{15}N NOESY-HSQC spectrum (noesyhsqcetf3gp3d) recorded with a mixing time of 100 ms, and using 192 and 48 points in the indirectly detected ^{1}H and ^{15}N dimensions, respectively, and 2k points in the directly detected ^{1}H dimension. The spectral width used were 10004.2 and 2757.6 Hz in the indirectly detected ^{1}H and ^{15}N dimensions and 12820.5 Hz in the 1H direct dimension. (H)CCH-TOCSY (hcchdigp3d2) and ^{1}H-^{13}C HSQC-NOESY spectra were also recorded.

The protons and carbons chemical shifts were compared with the values from the Biological Magnetic Resonance Data Bank (MRDB), a repository for data from NMR spectroscopy, and with the values from the Re-referenced Protein Chemical shift Database (RefDB), a data base of carefully corrected chemical shifts derived from the BioMagRes Bank (**Zhang et al., 2003; Ulrich et al., 2007**).

Secondary structural elements were predicted using the chemical shift index (CSI) tool from CCPN program suite for the measured values of CA, CA, CO and proton nuclei chemical shifts for each assigned residue. The assigned chemical shifts and the residual dipolar couplings measured were used as an input for structure determination. Models of the structure were calculated with CS-ROSETTA using the RASREC protocol (see Chapter 2)

Data acquisition for RDCs measurements: Dipolar couplings were measured at 298.15 K using a two-dimensional IPAP (In-Phase and Anti-Phase) ^{1}H-^{15}N-HSQC correlation experiment (hsqcf3gpiaphwg) (**Ottiger et al., 1998**), with 2k and 128 complex points and spectral widths of 12820.5 Hz and 2433.2 Hz for the ^{15}N and the ^{1}H dimensions, respectively.

Using the IPAP approach, two separate HSQC experiments were recorded to generate in-phase and anti-phase cross-peaks, measured for the samples in the anisotropic (gel oriented protein sample) and isotropic (protein sample in solution) phase. The sum and difference of the in-phase and anti-phase spectra enabled the separation of the upfield and downfield components of the ^{15}N doublets, in the indirect dimension, into two subspectra. The splittings in the ^{15}N dimension were then directly measured allowing the determination of residual ^{1}H-^{15}N dipolar coupling constants with high accuracy.

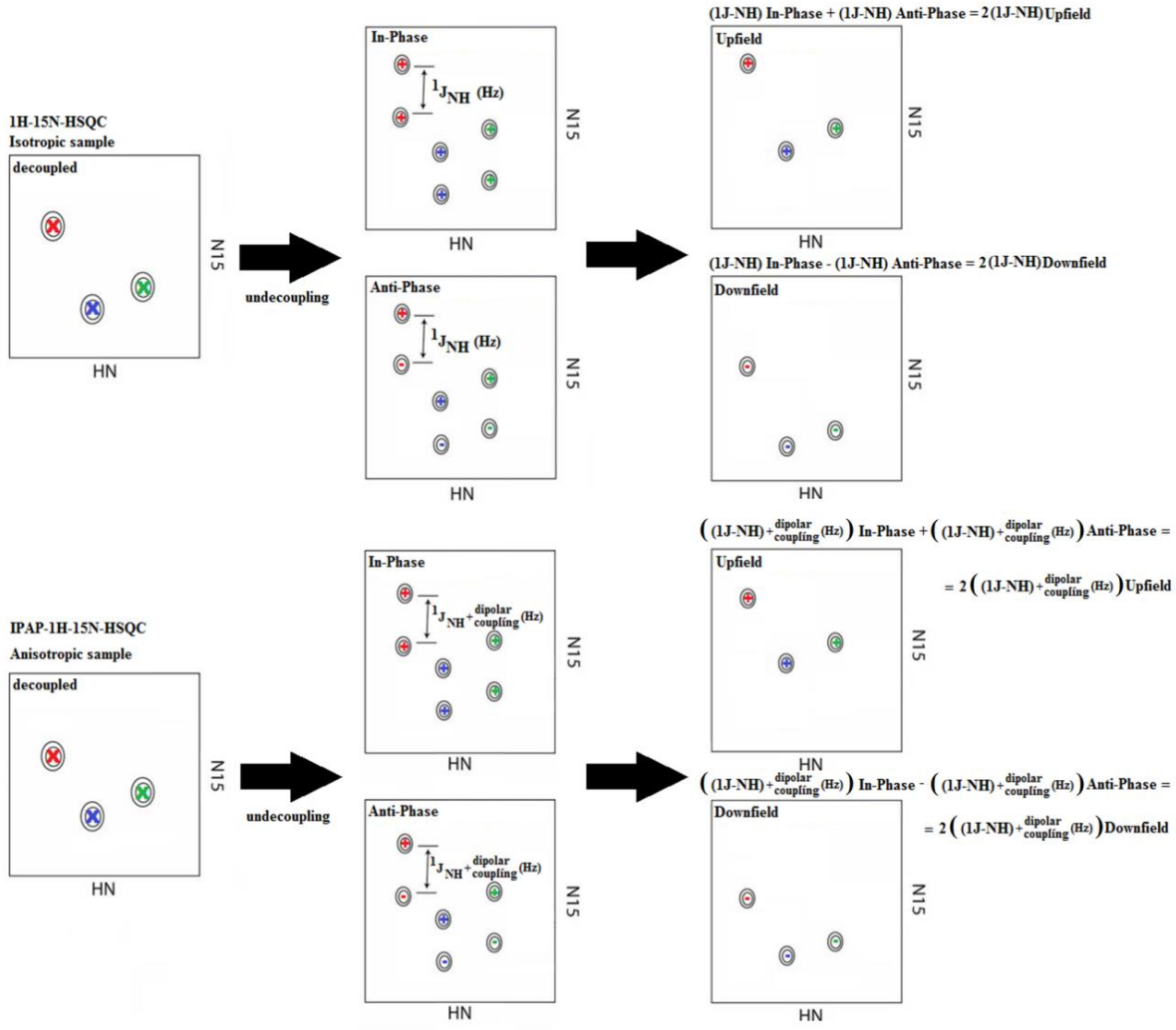

Figure 14 – Scheme of the IPAP approach for determining 15N-1H residual dipolar couplings.

In the isotropic protein solution, the observed splitting correspond the one-bond J-coupling ($^1J_{NH}$), while in the anisotropic protein solution, the observed splitting are the sum of the one-bond J-coupling ($^1J_{NH)}$ and the residual dipolar coupling (D_{NH}). Therefore, ^{1}H-^{15}N residual dipolar couplings (RDCs) can be obtained from the measured difference between the observed splitting, in the ^{15}N dimension, in the anisotropic and isotropic solution (**Figure 14**).

^{1}H-^{15}N RDCs were measured for 98 residues. The RDC histogram and values were plotted against the amino acid sequence (data not shown). Measurements of the RDCs were not possible to carry out for the remaining residues, either as a result of peak overlap or the absence of peaks due to proton exchange with the solvent. RDCs determine the orientation of the N-H bonds vectors in respect to the molecular alignment tensor. The degree of alignment can be estimated from the minimum and maximum values of the RDCs observed, while the tensor can be evaluated from the data and a structure using specialized software such as MODULE (**Dosset et al., 2001**).

Data acquisition for ^{15}N relaxation measurements: ^{15}N NMR relaxation measurements were performed on uniformly ^{15}N-labeled RodZc samples. Longitudinal relaxation time (T1), transverse relaxation time (T2) and the heteronuclear NOEs (^{15}N-^{1}H NOE) were measured using inversion recovery (**Vold et al., 1968**), Carr-Purcell-Meiboom-Gill (CPMG) (**Carr & Purcell, 1954; Meiboom & Gill, 1958**) and steady-state NOE (**Noggle & Shirmer, 1971**) experiments, respectively. The pulse sequences used have been previously described (**Kay et al., 1989; Kay et al., 1992**).

NMR spectra for the determination of T1 and T2 relaxation times were carried out with sweep widths of 2676.5 Hz and 12820.5 Hz for the ^{1}H and ^{15}N dimensions, respectively. A total of 2k complex data points were used in the ^{1}H dimension and 32 increments in the ^{15}N dimension were collected. To obtain the T1 relaxation rates, a list of variable delays (Vd) was employed with eight delays of 10, 50, 100, 300, 600, 1000, 1500 and 2000 ms. 16 scans were acquired per t1 increment. For the measurement of T2, a list of constant delays (Vc) was employed with eight delays of 15.84, 31.68, 63.36, 95.04, 126.72, 158.40, 237.60 and 316.80 ms (Kay et al., 1992). 16 scans were acquired per t2 increment. A recycle delay was used to avoid the saturation effects from not completely relaxed signals.

For ^{15}N-^{1}H NOE measurements, two sets of spectra were acquired, one where the ^{1}H protons were saturated prior to the initiation of the sequence and one without. Sweep widths of 2757.6 Hz and 12820.5 Hz were used in the ^{1}H and ^{15}N dimensions, respectively. 16 scans were acquired. In the saturated experiment NOE, the ^{1}H signals were saturated by applying a train of non-selective 135º pulse.

Cross peaks intensities were evaluated as peak heights. The relaxation T1 and T2 were obtained by non-linear least-squares fitting of single exponential decays to the experimental data. The error propagation in T1 and T2 experiments were calculated using the standard deviation approach.

The NOE effect was calculated as the ratio of the peak intensities measured in the presence and absence of saturation of the proton magnetization, i.e., the peaks heights in the spectra with and without saturation.

4.3 Results and Discussion

4.3.1 Bioinformatic analysis

In **Figure 15** is shown the full RodZ sequence, were the specific periplasmic residues are in blue. RodZc from *Bacillus subtilis* starts at residue His133 and ends at residue Ser304, making a total of 172 residues. It has no cystein residues and has only one tryptophan.

MTELGIRLKEAREEKAMSLDDLQAATKIQKRYLTALEEGNYDIIPGKFYVRAFIKQ

YAEAVGLDADQLFEEHKKDIPNTYHDDVSEKISGMNLQKEMPKPASKALELLPTI

LVILGVIVVIAIVYAIIQFANHKNSDDHNAASEKAITQSESKYEIPKDSTLKENQNNS

SEKETDTKKETKENEDKKKENDSEKLEIKAAGTEGSLTTYEVSGADKIELELKASD

SSWIRVRDENSSSLKEGTLKKDETYKKDITDQKQVDIRTGYAPNLKIKINGKVLSY

ELDPKKVMAQTIKIVNKKEEKSS

Figure 15 – Full aminoacid sequence of RodZ from *B. subtilis*.The periplasmic domain of RodZ (RodZc) is highlighted in blue. The remainin residues (in black) belong to the transmembrane and cytoplasmatic domain.

An initial profile of this protein was made using a number of bioinformatic tools. The sequence of the RodZc domain was first analyzed using the PSIPRED server (www.bioinf.cs.ucl.ac.uk) for secondary structure, disorder prediction and fold recognition. SWISS MODEL and I-TASSER servers were used to predict a model of RodZc.

As shown in **Figure 16,** the PSIPRED server predicted the C-terminal moiety to be fully beta-sheet (approximately nine Beta-sheets) with a coiled coil linker that connects the Beta-sheet rich moiety to the transmembrane domain (TM). A more detailed analysis of different types of amino acids that build this sequence, led us to realize that the predicted Beta-sheet moiety is in fact rich in hydrophobic residues such as Valines, Isoleucines and Leucines, in constast to the exclusively hydrophilic residues like Lysines and Asparagines found in the unstructured coiled-coil linker. The same observation has been reported in Alyshaya et al, 2009, where they suggest an extended conformation for this linker.

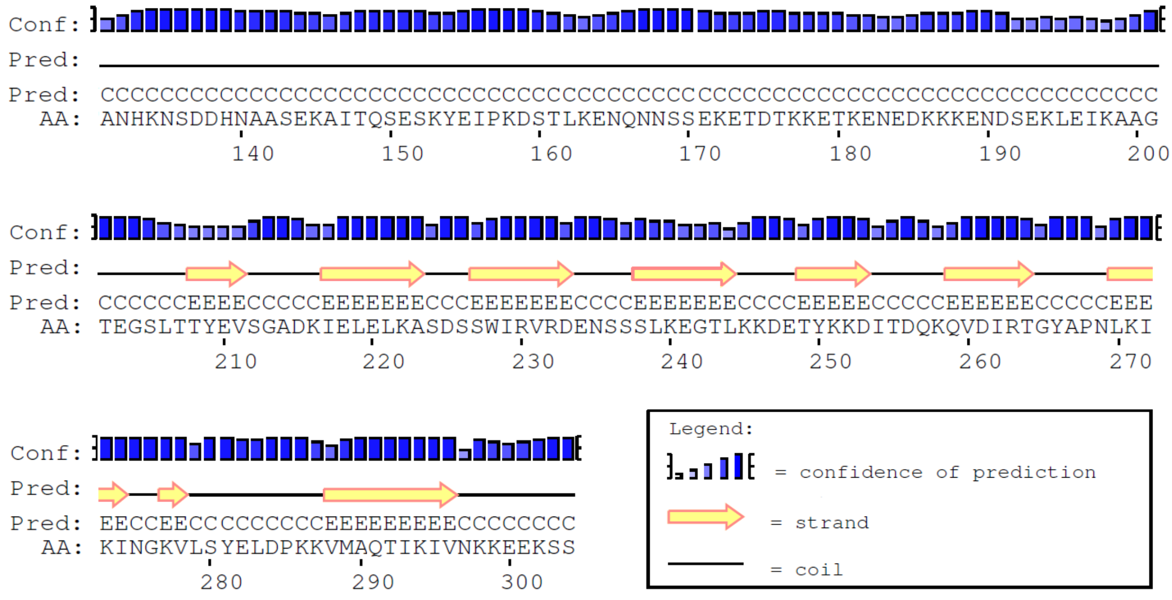

Figure 16 - PSIPRED results for the secundary structure prediction of the full RodZ protein. (www.psipred.org). Predicted β-strands are shown as yellow arrows, and unstructured regions as black lines.

From the PSIPRED Protein Sequence Analysis Workbench, DISOPRED was also used to predict the disorder profile of the RodZc domain. In **Figure 17** is presented the resulting plot for the disorder probability in percentages for each residue of the given sequence. Residues from A200 till the end of the protein have very low disorder probabilities (less than 10%). This information corroborates with the secondary structure prediction, reinforcing the hypothesis of the C-terminal moiety being well packed and rich in Beta-sheet content. In contrast, residues comprised between residue H133 and A200 show high disorder probabilities, with maximum percentages near 100% of disorder. This information raised the possibility of RodZc protein domain being an intrinsically disordered protein (IDP).

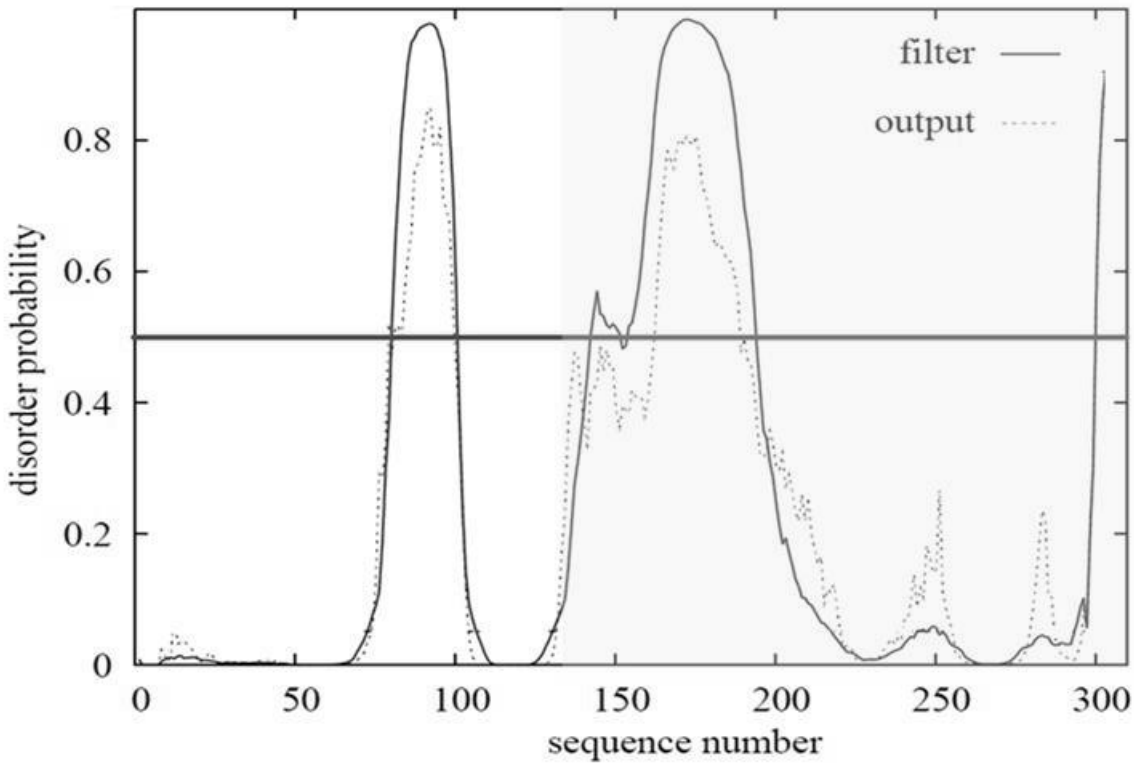

Figure 17 – Order/disorder profile of the full RodZ protein plotted with DISOPRED from the PSIPRED server. The disorder prediction is built against each protein residue. Region squared in grey corresponds to residues from the periplasmic domain of RodZ.

In recent years there has been a growing interest in intrinsically unstructured proteins and their role in biology. However, it's yet poorly understood the relationship between the primary sequence of a

protein and its susceptibility to a disordered conformation. The structural flexibility of IDPs allows them to achieve functional modes that otherwise would be unfeasible to globular proteins. Moreover, by having this feature of plasticity, IDPs can fold upon binding, forming transient complexes through nonspecific interactions or interactions with rapid dissociation rates, allowing IDPs to remain dynamic (**Wright and Dyson, 2009**). That's one of the reasons why IDPs are frequently associated with cellular control mechanisms and signaling, and have been identified at the "heart" of protein interaction networks (**Dunker et al., 2005**).

After this analysis, we wanted to know if structural homologues of RodZc could be identified in current version of protein structure databases. So, a run was made to produce a model of RodZc using Swiss Model Automatic Modelling tool from SWISS MODEL workspace which is a web-based integrated service dedicated to protein structure homology modelling. However, no model was obtained from this run, as no suitable templates were found when BLAST searched for highly similar template structures. So it was not able to detect any similar aminoacid sequence with a tri-dimensional structure available in the databases (**Bordoli et al., 2009**).

Then we analyzed the RodZc sequence using I-TASSER server. As described in the Methodology chapter (see Chapter 2), this bioinformatic tool allows an automatic prediction of the 3-dimensional structure of protein molecules from their amino acid sequences (**Zhang, 2008; Ambrish et al., 2010; Roy et al., 2012**). In **Figure 18** is shown the predicted protein model with the lowest C-score. C-score is a confidence score for estimating the quality of the predicted models, calculated based on the significance of template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range from -5 to 2, making the predicted model with a -3.18 C-score of good confidence.

Two other parameters were retrieved from I-TASSER simulation, namely RMSD and TM-score (or TM-align, described in Methodology, Chapter 2). TM-score is a scale for measuring the structural similarity between two structures, being sensitive to local errors. Because RMSD is an average distance of all residue pairs in two structures, a local error will increase the RMSD value although the global topology is correct. In TM-score, however, the small distance is weighted stronger than the big distance which makes the score insensitive to the local modeling error. A TM-score >0.5 indicates a model of correct topology and a TM-score<0.17 means a random similarity. The Tm-score value retrieved for the best model is 0.36±0.12, which makes this model close to the correct topology of RodZc (**Zhang and Skolnick, 2004**).

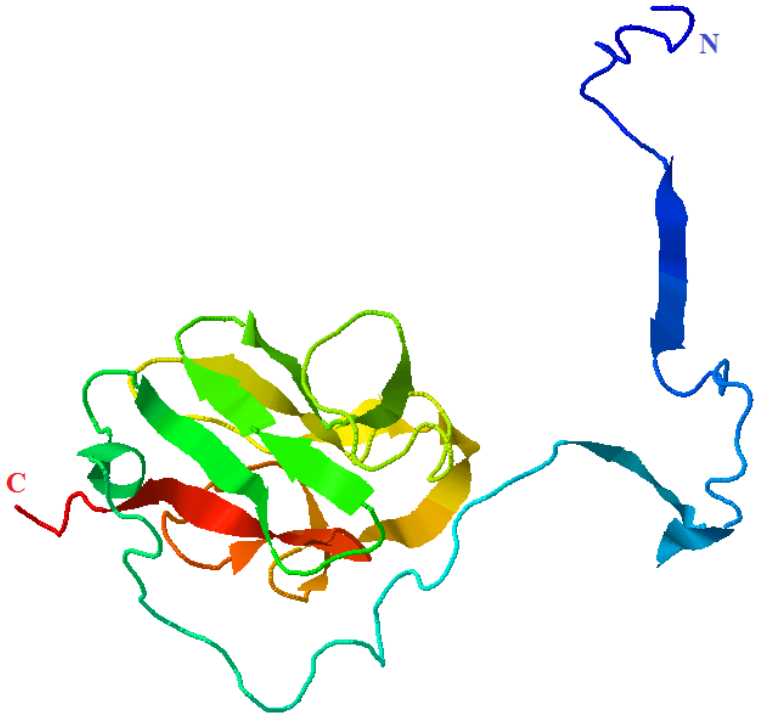

Figure 18 – RodZc model predicted with I-TASSER server. C-score: -3.18; RMSD: 12.4±4.3; TM-Score: 0.36±0.12.

Finally, to analyze the putative fold of RodZc, pDomThreader protein fold recognition tool available in the PSIPRED suite was also run. This tool performs a highly sensitive homologous domain recognition using profile-profile comparison from a domain library (**Lobley et al., 2009**). The NMR solution structure of the immunoglobulin like domain of mouse nuclear lamin (LTD; PDB code: 1ufg) produced the highest confidence fold prediction score (p-value of 1e-4; data not shown). Lamins are the major architectural determinants of the animal cell nucleus. They line the inside of the nuclear membrane, forming a mesh that confers mechanical stability and serves as a platform for functional interactions between the nucleus and the cytoplasm (**Dittmer and Misteli, 2011**). The LTD domain consists of 9 β-strands, forming two β-sheets of 4 and 5 strands, respectively, connected by short loops, and packed into a classical β-sandwich (**Dhe-Paganon et al., 2002; Krimm et al., 2002**).

The LTD domain belongs to the immunoglobulin (Ig) fold family (**Bork et al., 1994; Williams and Barclay, 1988; Potapov et al., 2004**). Proteins belonging to this family share a common structural core of four B-sheets embedded in an antiparallel "curled" B-sheet sandwich with a total of three to five additional strands. The number, position and length of the additional β-strands attached at the edges of the core domain are highly diverse, as is the sequence of the connecting loops. However, these proteins are involved in a variety of functions, including cell-cell recognition, cell-surface receptors, muscle structure and the immune system. They interact with extremely different proteins or ligands, and with different binding modes. Nevertheless they have a common feature of interaction with other Ig-like domains via the B-sheets. In addition, homo- and heterodimers can be formed (**Bork et al., 1994**).

Current classification of classical Ig-like domains is mainly based on the number of strands and sequence similarity. Reflecting their structural and functional diversity, sequence identity of proteins within the Ig domain family is low (10% or less) (**Bork et al., 1994; Halaby et al., 1999; Williams and Barclay, 1988; Potapov et al., 2004**).

Based on the number of strands and their location, however, four distinct types were produced. In **Figure 19**, each topology is described in more detail.

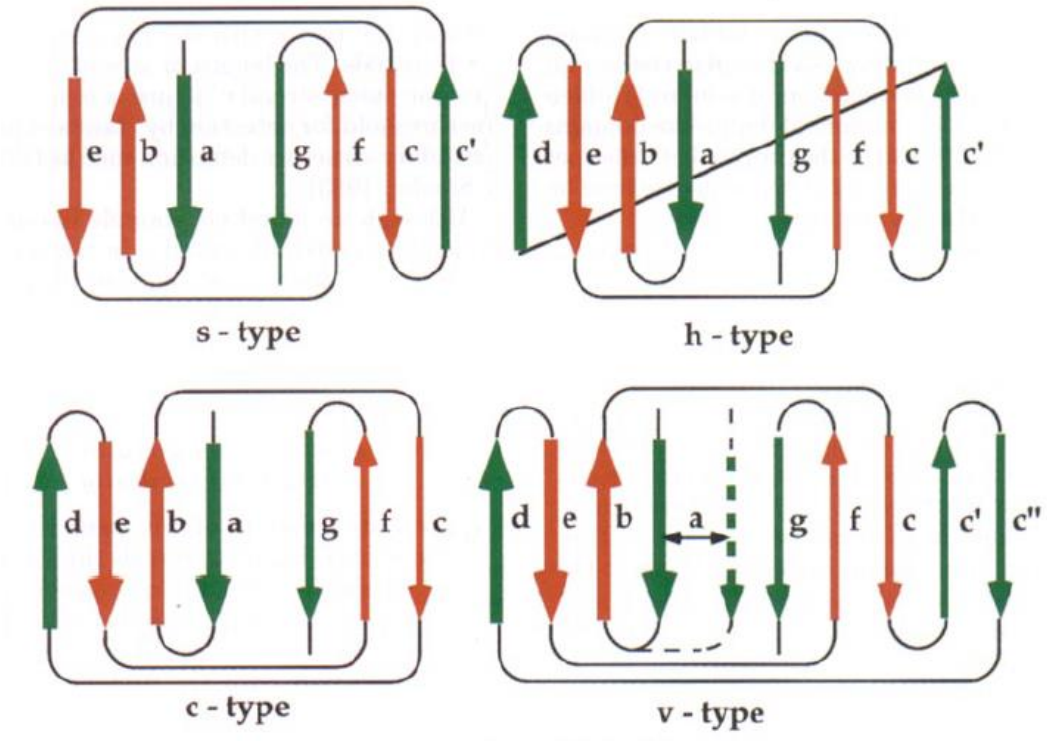

Figure 19 - Topology diagrams of observed hydrogen bonding patterns. The 7—9 strands (a, b, c, c', c'', d, e, f, g) form a sandwich of 2 sheets. The common core is shown in red. Immunoglobulin constant domains have 7 strands in a c-type topology. Immunoglobulin variable domains have an additional hairpin (c'-c'') between strands e and d, with a total of 9 strands (v-type, for variable). Strand a has two alternative locations in v-type domains, being antiparallel with strand b or parallel with strand g. Other Ig-like domains also have 7 strands, but are different from c-type; 4th strand has switched c' and d sheets (s-type).The last type represents an 8-stranded hybrid between c- and s-type that has both c' and d strands (h-type, for hybrid) (Bork et al., 1994).

Although with lower scores, other structures found in our search of the PSIPRED database correspond to all-beta proteins or protein domains (not show). Collectively, these observations suggest that the C domain of RodZ may adopt an Ig-like fold.

Gathering all the relevant information provided by the bioinformatic initial studies, a conclusion was made that would only be relevant to determine the structure of the predicted well folded C-terminal moiety since the linker is pointed to be unstructured. Therefore, our collaborators from the Microbial

and Development laboratory (ITQB) produced the strain AH4334 by transforming BL21(DE3) with plasmid pAP2, created to express His6-RodZc201-304 fusion under control of T7lac promoter. All the procedures are described in **Costa et al, 2012 (to be published)**. Thus, upon addition of IPTG, this strain is prepared to express RodZc residues from G201 till the end of the C-terminus, thereby excluding the first 68 residues of the RodZ periplasmic domain, as shown in **Figure 20**.

MTELGIRLKEAREEKAMSLDDLQAATKIQKRYLTALEEGNYDIIPGKFYVRAFIKQ
YAEAVGLDADQLFEEHKKDIPNTYHDDVSEKISGMNLQKEMPKPASKALELLPTI
LVILGVIVVIAIVYAIIQFANHKNSDDHNAASEKAITQSESKYEIPKDSTLKENQNNS
SEKETDTKKETKENEDKKKENDSEKLEIKAA|GTEGSLTTYEVSGADKIELELKAS
DSSWIRVRDENSSSLKEGTLKKDETYKKDITDQKQVDIRTGYAPNLKIKINGKVLS
YELDPKKVMAQTIKIVNKKEEKSS|

Figure 20 – Full aminoacid sequence of RodZ from *B. subtilis*.The periplasmic domain of RodZ (RodZc) is highlighted in blue. The remaining residues (in black) belong to the transmembrane and cytoplasmatic domain. Residues underlined (dark blue) are present in RodZc201-304 construct.

4.3.2 Loss of structure after elimination of the predicted linker

After expression and purification of His6-RodZc201-304 construct, several studies were made in order to understand its structural properties. Several far-UV CD spectroscopy spectra were collected by P.I. Claudio Gomes from the Protein Biochemistry Folding and Stability Group, ITQB/UNL. RodZc201-304 yielded the characteristics of a poorly structured soluble polypeptide with absorption minima at 208 nm and no more additional features, compatible with a denatured protein CD profile. (**Appendix, Figure IA**). Thermal denaturation was also investigated using CD spectroscopy. This protein showed no transition and was found to be irreversible since the spectrum obtained upon cooling down the sample back to room temperature did not restore its initial spectroscopic signature (**Appendix, Figure IB**). So, a hypothesis was raised in which the linker predicted to be unstructured may be promoting the stability of the predicted folded rich in beta-sheet moiety. However, since CD spectroscopy is not very sensitive to beta-sheet secondary structure motif, NMR was used to gain further insight of the structure of this protein.

1D ^{1H}-NMR was collected (**Figure 22**) and surprisingly the spectrum obtained shows features of an exclusively unfolded protein, with all the amide signals clustered between 8.0 and 8.5 ppm and no methyl resonances observed around 0 ppm. In addition, no signals were observed around 5 ppm, region which is consistent with Hα signals in the B-sheet conformation. Therefore, ^{1}H-NMR data showed to be

in good agreement with CD analysis and that the linker region might be indeed required for the overall folding and/or compactness of RodZc.

Following these observations, we came to a conclusion that this protein is in fact fully unstructured so it became imperative to study the full periplasmic domain of RodZc. Our collaborators then produced a new strain, AH4268 by transforming BL21(DE3) with plasmid pJR2 which codes for a fusion between the 6-Histidine tag and residues 131 to 304 of RodZ. Thus, upon addition of IPTG, this strain is prepared to express RodZc residues from A131 till the end of the C-terminus, including two additional residues that belong to the transmembrane domain (A131 and N132), as shown in **Figure 21.**

MTELGIRLKEAREEKAMSLDDLQAATKIQKRYLTALEEGNYDIIPGKFYVRAFIKQYAE
AVGLDADQLFEEHKKDIPNTYHDDVSEKISGMNLQKEMPKPASKALELLPTILVILGVI
VVIAIVYAIIQF|**AN**HKNSDDHNAASEKAITQSESKYEIPKDSTLKENQNNSSEKETDT
KKETKENEDKKKENDSEKLEIKAAGTEGSLTTYEVSGADKIELELKASDSSWIRVR
DENSSSLKEGTLKKDETYKKDITDQKQVDIRTGYAPNLKIKINGKVLSYELDPKKV
MAQTIKIVNKKEEKSS|

Figure 21 – Full aminoacid sequence of RodZ from *B. subtilis*.The periplasmic domain of RodZ (RodZc) is highlighted in blue. The remaining residues (in black) belong to the transmembrane and cytoplasmatic domain. Residues underlined (both light and dark blue) are present in RodZc131-304 construct. Residues A131 and N132 from the transmembrane domain are also present in this construct.

The same studies described previously were also employed to the purified RodZc131-304 protein. Regarding the circular dichroism experiments, this protein yielded a similar result as the one obtained for RodZc201-304, with a slight difference between 215 and 230 nm. The resulting spectrum suggests it to have a mixture of disordered and beta-sheet structure, being in agreement with the bioinformatic predictions previously described (**Appendix, Figure IA**). However, thermal unfolding of RodZc131-304 showed a reversible transition of small magnitude with a melting temperature (Tm) around 45ºC, indicating that this protein may have secondary structure content (**Appendix, Figure IB).**

A 1D ^{1}H-NMR spectrum was also recorded for RodZc131-304 protein (**Figure 22B**). This simple NMR experiment was enough to distinguish the two constructs. Even though the spectrum of RodZc131-304 shared some features with the one collected from RodZc201-304, namely the amide peaks centred around 8.2 ppm, this spectrum also contains amide resonances from 6 to 10.5 ppm and methyl resonances around 0 ppm, which are consistent with a folded protein. More importantly, the 1H

NMR spectrum of RodZc131-304 shows signals between the water signal (4.6 ppm) and 5.7 ppm, which are consistent with Hα signals in the B-sheet conformation. So this spectrum shows a mixture of folded and unfolded structural features. Therefore, the ^{1}H-NMR data is in good agreement with the CD analysis, in which the linker region might be important for stabilizing the structured rich in β-sheet moiety.

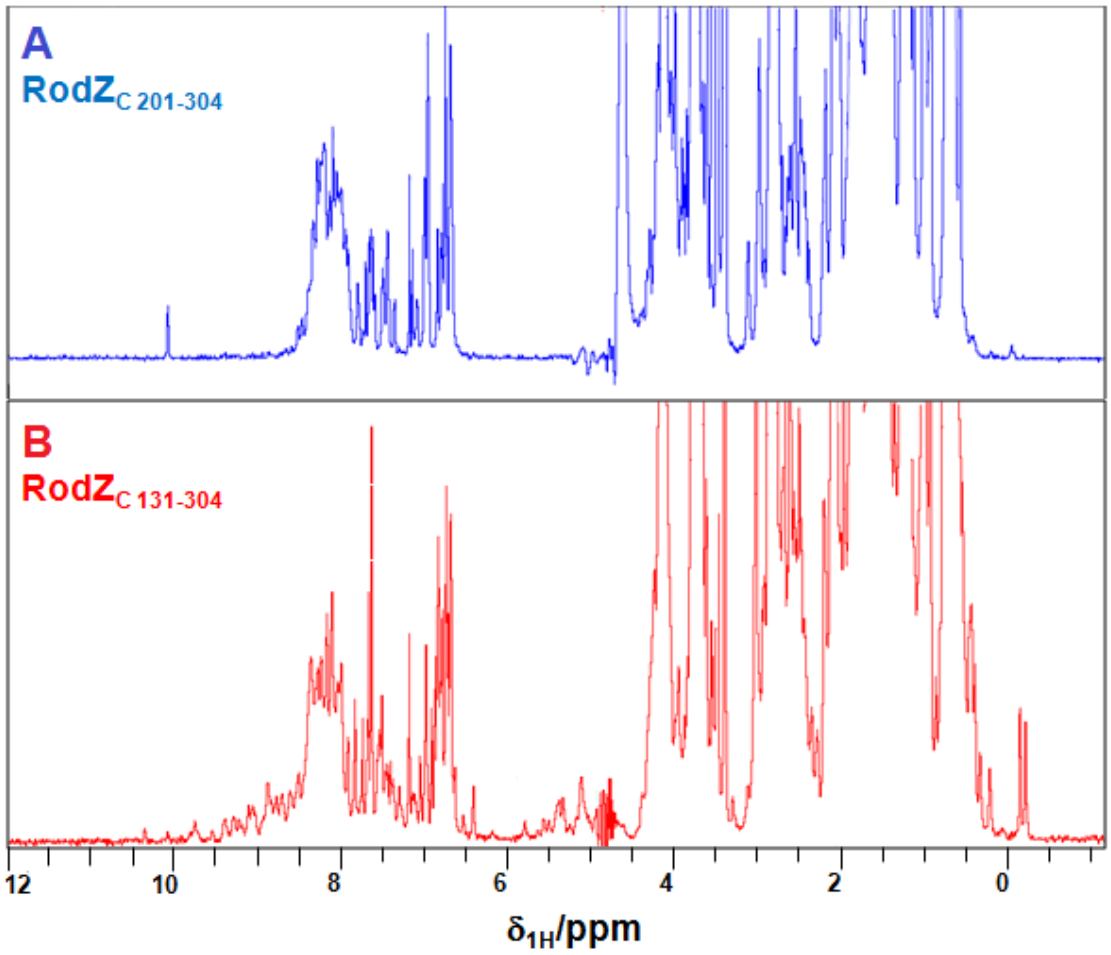

Figure 22 – Proton NMR spectra of RodZn and RodZc. ^{1}H NMR spectra of RodZ$_{C\ 201-304}$ (A, in blue) and RodZ$_{C\ 131-304}$ (B, in red). The resonances from 6 to 10.5 ppm in (B) are consistent with a predominantly folded protein. The resonances around 5 ppm in (B) are consistent with proton signals in the beta sheet conformation; the amide signals are very well resolved spanning more than a 2 ppm window. Also for B, the signals of methyl resonances close or below 0 ppm are indicative of a folded protein. Both groups of resonances are absent from RodZ$_{C\ 201-304}$ (A). The group of peaks around 8.2 ppm in panel B are attributed to an unfolded section in RodZ$_{C\ 131-304}$, presumably corresponding to the linker region.

Further evidence for the presence of two different types of structural elements in RodZc 131-304 came from the differential effects of two water suppression techniques on the spectra. The first one, water pre-saturation, saturates the water signal effectively destroying it for the duration of the experiment. As a result, all the signals that are in fast exchange with water are also suppressed. In the

second, excitation sculpting, the water signal is selectively converted to a non-observable form but is not destroyed and so, signals in fast exchange with water are not suppressed. Using these two approaches we obtained two spectra that are qualitatively similar, yet under pre-saturation conditions, the peak cluster at 8.2ppm is significantly attenuated when compared to the excitation sculpting data (**Appendix, Figure II**). This implies the presence of a number of amidic protons that are in fast exchange with the bulk water, which is consistent with an unfolded structure. In contrast, this phenomenon is not expected in the excitation sculpting case where, in the absence of suppression, the water signal is not selectively removed. Strikingly, the attenuation of the spectrum under pre-saturation conditions, suggests that a significant portion of the protein is in contact with water. This in turn strengthens the view that RodZc131-304 has either a simple fold or an extended structure.

So, from the analysis of the ^{1}H-NMR spectrum of RodZc131-304, we saw that it holds some features of a folded protein. At this stage it is imperative that we determine the structure of the folded region of this construct. However, as we began to analyze the NMR data more carefully, more specifically the 1D ^{1}H and ^{15}N-HSQC spectra, two signals were identified for the only tryptophan residue in the sequence (**Appendix, Figure III**). Since only one tryptophan signal was expected, we started questioning if the sample might not be pure and/or stable. Regarding these questions and to get a bigger insight of what could be happening with these NMR samples, we moved back and tried to optimize the protocol of purification

4.3.3 RodZc protein present in solution in a 3-state

The expression and purification protocol of RodZc$_{131-304}$ (**to be published in Costa et al 2012**) was tested in our laboratory. Since our laboratory is specialized in producing highly pure and stable protein samples specific for NMR essays, we saw that there were many steps in the expression and purification procedure that could be improved. Starting with the expression step, the cell cultures were induced with 1mM IPTG for 4 hours instead of 3 hours allowing the cells to express more of the heterologous protein. After disrupting the cells, the supernatant was diluted in the same phosphate buffer (IMAC Buffer A), but without Imidazole. By doing this, it made the interaction between the different molecules present in the sample and the matrix of the column to be less specific, allowing the column to retain more biological material. Also, IMAC chromatography was performed through linear gradient of increasing IMAC Buffer B concentration containing Imidazole, instead of a stepwise gradient with addition of different elution buffers with different percentages of Imidazole. Using the linear gradient, the eluent composition is changed continuously toward conditions favouring dissociation from the chromatography medium, being more powerful when trying to separate between closely related peaks. Moreover, an Akta Laboratory-scale Chromatography Systems from GE Healthcare was used as a controller to optimize this purification step, unlike the previous protocol. But

most importantly, a second purification step was added to the initial protocol, namely size exclusion chromatography (SEC). This is a rather advantageous last step purification method since it performs a good separation of large molecules from the small molecules with a minimal volume of eluate, while preserving the biological activity of the particles to be separated. There is also no sample loss because solutes do not interact with the stationary phase. In addition, this chromatographic method was also controlled by an Akta Purifier system. Also, dialysis, which is a time consuming procedure, was removed from the main protocol since we used always the same phosphate-based buffer and before every chromatographic step the sample was passed through a filter of 0.22 µm that retained not only small particles like dust that would damage our columns, but also bacteria that could be contaminating and degrading our samples.

From the size exclusion chromatogram shown in **Figure 23**, we can immediately see the presence of two major components, one with a much higher than expected apparent molecular weight and another with two time the expected size for RodZc131-304. SDS PAGE revealed that both species correspond to our construct, while 1D NMR revealed that the higher molecular weight fraction is completely unfolded. It's obvious that with IMAC alone we would never be able to identify the possibility of having more than one species in the sample solution. Isolation of the second peak and reinjection in an analytical SEC column resulted in a spectrum with a single component implying that the higher molecular weight species corresponds to an irreversibly unfolded form of the protein that can be easily removed from solution

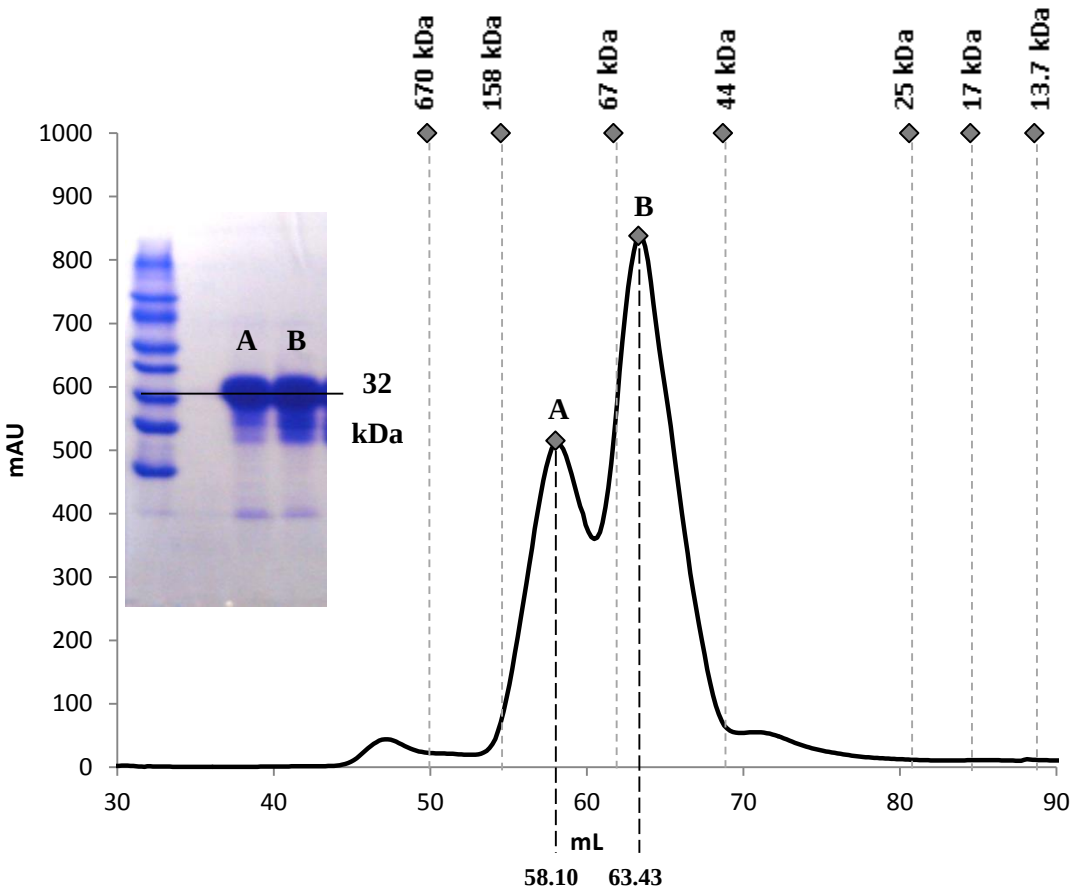

Figure 23 - Size exclusion chromatography chart of protein sample collected from IMAC chromatography, with the corresponding 15%Tris-Gly SDS-PAGE of the two peaks eluted at 58.10 mL and 63.43 mL after being loaded into the separating column. The molecular weight values are standard calibration curved values for Superdex 75HiLoad column.

In order to study if any concentration-dependent aggregation of the construct can take place we then injected the sample in a range of concentrations (from 110 µM to 982 µM; See **Figure 24**), to a Superdex 75 HR 10/30 column (GE Healthcare). The column was operated at a flow rate of 1.5 mL/min, and fractions of 1mL were collected. We observed a single peak in ASEC for all concentrations with an estimated molecular mass (MW) of 44 kDa, i.e., 2x the mass expected for RodZc 131-304 (predicted MW of 22.3 KDa).

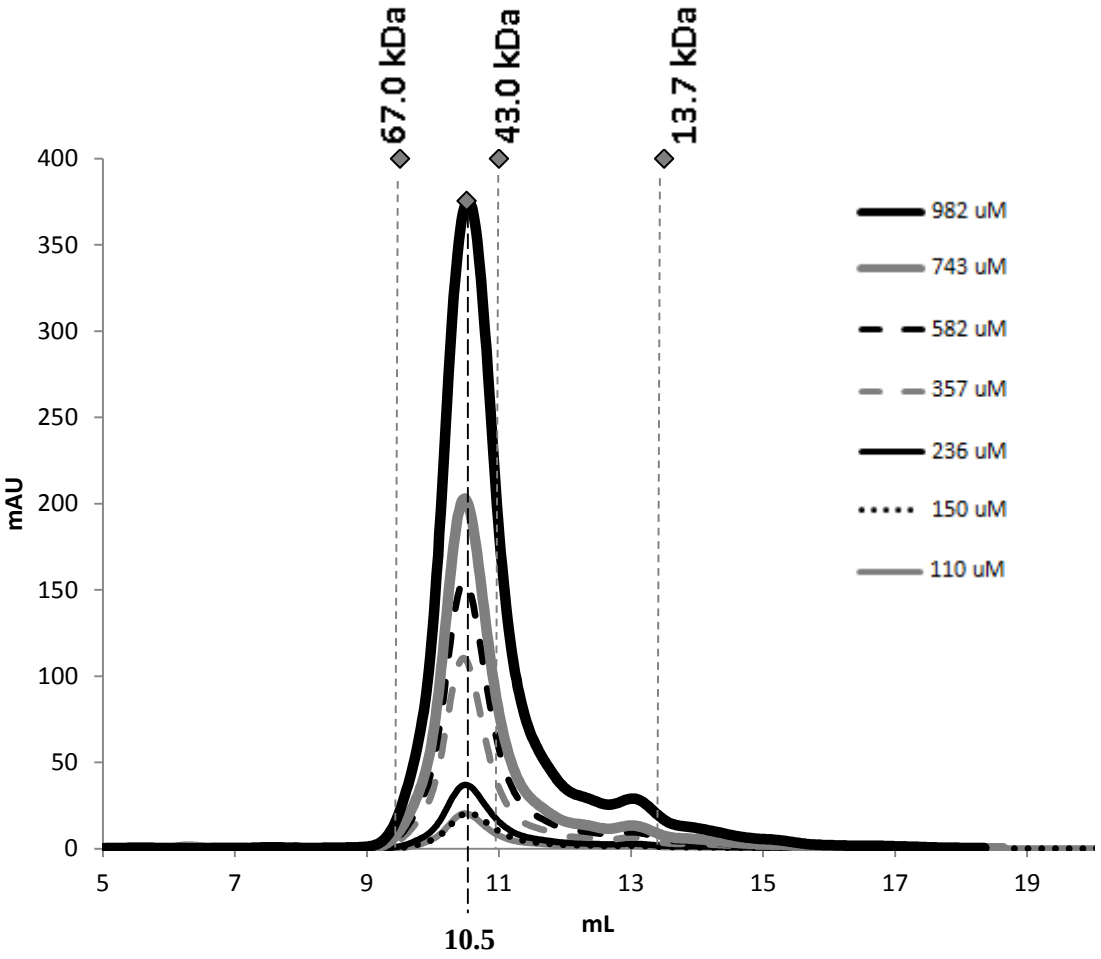

Figure 24 – Analytical size exclusion chromatography chart of protein sample collected from SEC chromatography after being loaded into the separating column. The molecular weight values are standard calibration curved values for Superdex 75 small column.

The higher apparent MW observed for, RodZc is consistent with the idea that it contains large partially disordered stretches of residues. Disordered proteins also show an anomalous behaviour in SDS-PAGE as well as in SEC columns, leading to the overestimation of the MW by a factor of 1.2-1.8 (Tompa, 2002; Weinreb et al., 1996).

Because SEC studies were not conclusive in respect to the MW estimation, NMR DOSY experiments were performed in order to obtain an independent MW estimate. (**Groves et al., 2004; Groves and da Silva, 2010**).

DOSY measurements resulted in diffusion rates that are consistent with MW of approximately 47 kDa for RodZc 131-304, 2x times the expected. (data not shown). In the end, DOSY estimate did not change the view that RodZc is an elongated and/or partially disordered monomer.

MW estimated from both SEC and DOSY are affected by the hydrodynamic radius of the protein that is assumed to be a rigid sphere. In the case of a partially unfolded protein that cannot be

approximated with rigid spheres these estimates are widely overestimating the size of the biopolymer. In contrast, measurements that are sensitive to the tumbling rates of samples can be more sensitive to the actual size of the structured portion of the construct. ^{15}N relaxation studies (described later) offer such an alternative way of determining protein size of a construct by evaluating its rotational correlation time (Tc) that is dependent on its tumbling time. As described later, using this methodology we were able to determine that our construct is indeed monomeric.

However, even though we were able to isolate the folded from of the protein from its completely unfolded state, when we repeated the 1D ^{1}H NMR measurements, the peak corresponding to the side chain of the tryptophan in the unfolded form was still present, albeit at a significantly lower concentration. Given that SEC had shown that the unfolded form was irreversibly removed from the solution, we suspected that a second, reversible unfolding state may exist in solution. To study this we explored the temperature dependence of the 1D ^{1}H spectrum to see if we can modulate the relative concentration of these two species,

1D ^{1}H NMR spectra were collected, with a range of temperatures between 283.15 K and 318.5 K, as shown in **Figure 25**. In these experiments the amidic proton peaks from the side chain of two forms of the tryptophan residue remained in the spectra, however we notice that as we were applying higher temperatures, the ratio of the area of the downfield peak (10.25 ppm) corresponding to the folded form of the protein remained unchanged in respect to the remaining spectral signals, while that of the upfield peak (9.97 ppm) initially decreased up to 303.15 K followed by an increase at higher temperatures. This observation can be explained by the fact that the partially unfolded form is initially destabilized by the increasing temperature. Above 303K the protein starts to globally unfold in a manner consistent with the denaturation temperature determined using CD spectroscopy. And therefore the unfolded signal started increasing again. Since no new TRP peak appears for the partially and the globally unfolded forms, we concluded that the reversible partially unfolded and the completely unfolded TRP side chain forms must have the same resonance leading to the increase of the downfield peak area.

From the observation of the spectra collected, 303.15K was chosen as the optimal temperature to collect all the remaining data necessary to assign RodZc.

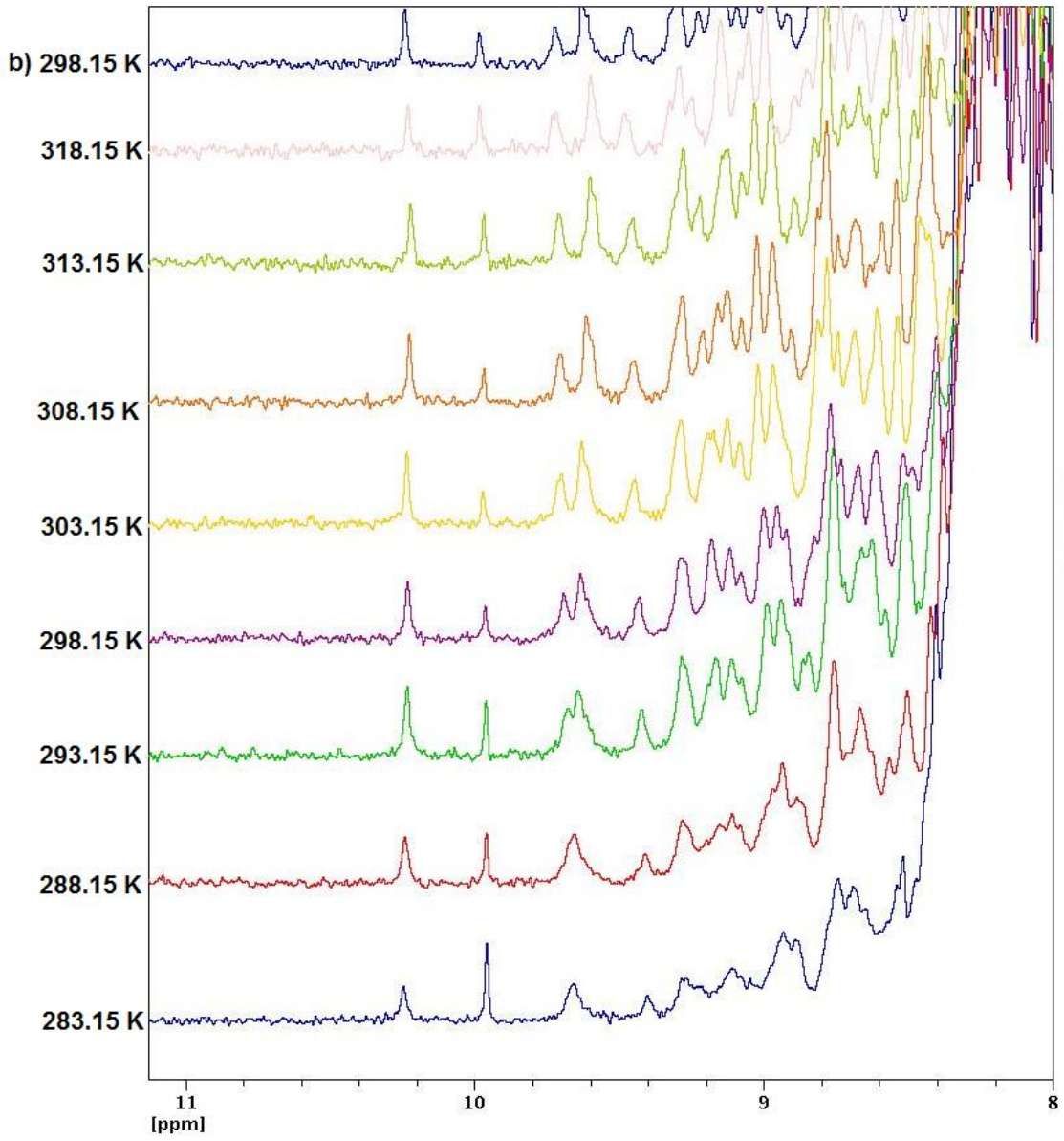

Figure 25 – 1D ^{1}H NMR spectra collected, with a range of temperatures between 283.15 K and 318.5 K, and back to 298.15 K (b). Investigation of the downfield (10.25 ppm) and upfield (9.97 ppm) amidic proton peaks from the side chain of two forms of the only tryptophan present in RodZc domain.

4.3.4 Structure analysis

RodZc131-304 construct is composed by a linker region and a putative rich in b-sheet moiety, and 6 histidine-tag and extra-residues added due to the cloning strategy selected. Using a combination of the spectra described previously (NMR data acquisition and structure calculation Chapter), we were able to identify 86% of ^{15}N, 90% of Carbons and a total of 93% of assigned resonances from the backbone (HN, N, CO, CA and HA chemical shits), considering only the putative-structured region of

RodZc (residues from 190 to 304). The fully assignment table is described in **Table 6.** The assigned ^{15}N-^{1}H-HSQC acquired from $RodZc_{131-304}$ is present in **Figure IV, Appedix chapter.**

Not all signals for all the expected amino acids could be identified however, although they were present in our spectra, more specifically residues comprised between 1-190, which corresponds to the predicted unstructured region. No assignment was possible for the predicted unstructured region since these residues are mostly highly exposed to solvent with high probability for solvent exchange.

Shifts	Found	Missing	Complete (%)
1H	418	71	85.481
1H aromatics	18	10	64.286
1H aliphatics	400	61	86.768
13C	248	27	90.182
13C aromatics	12	13	48.000
13C aliphatics	236	14	94.400
15N	67	31	68.367
Total Backbone	273	21	92.857
Total Side chain	460	108	80.986
Total	733	129	85.035
Total (w/o pseudo atoms shifts)	657	129	83.588

Table 6 - Report of the completeness of the Assignments of RodZc considering only residues from 190-304.

Evaluation of the secondary structural elements of RodZn was initially performed using the available chemical shift data (**Wishart and Sykes, 1994**). The secondary structure of RodZc is shown in **Figure 26**.

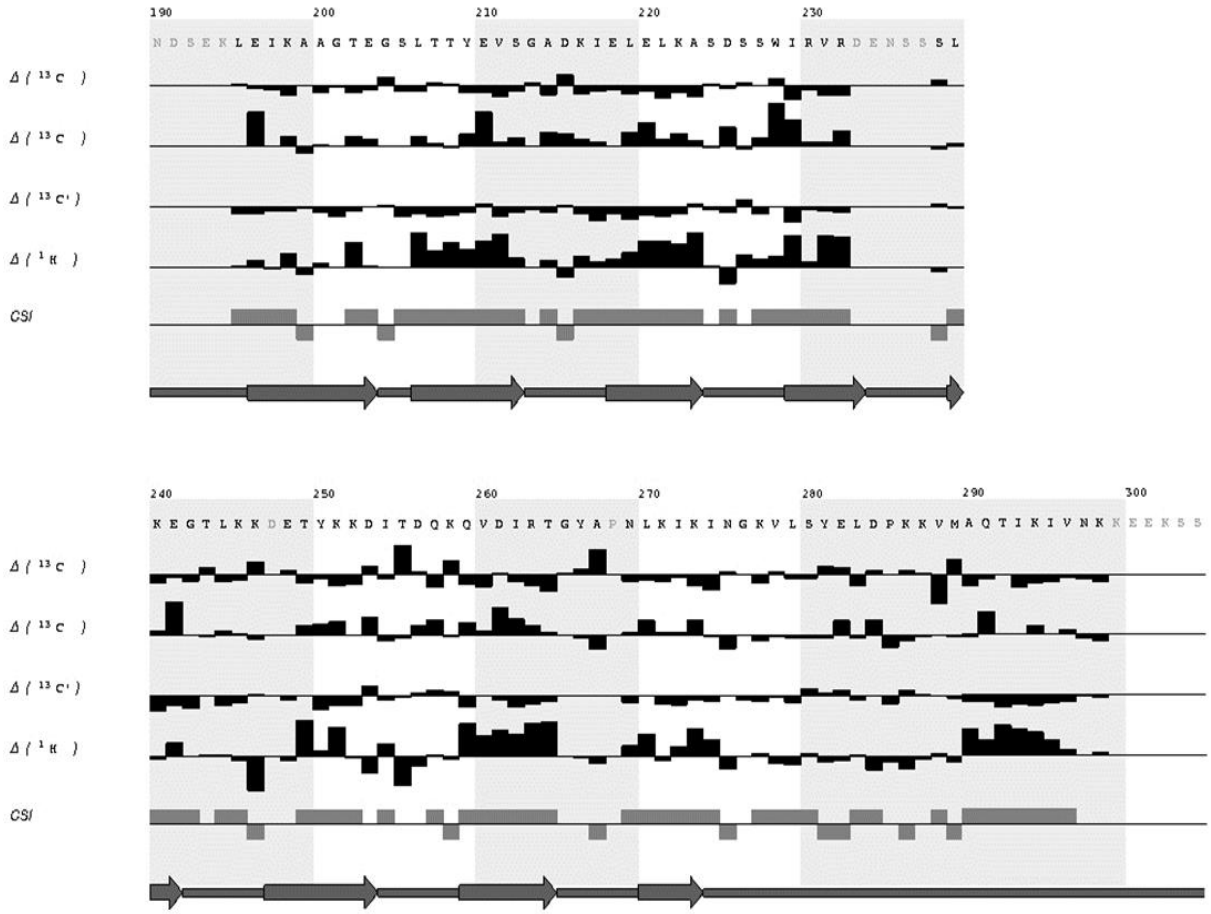

Figure 26 - Experimental restraints for RodZc protein, including HA, CA, CB and CO secondary shits along with the secondary structure deduced from the data. The amino acid sequence and numbering are shown at the top. The chemical shift indices obtained from CA, CB, CO and HA are also shown by black bars at the bottom. The locations of the secondary structure elements identified in the calculated family of structures are shown at the bottom.

Examining the result of CSI, we see that RodZc has only β-sheet as secondary structural motif, being in agreement with the predicted secondary structure with PSIPRED. The predominance of b-sheet structure as expected for a putative-member of the IgI fold is evident in the CSI result that indicates the presence of up to nine b-strands.

If we see more closely, the first beta-sheet starts in residue 190. And taking into account that the first construct was built to begin in residue 201, it means that the first beta-sheet was actually cut in the middle. This may have been the trigger that led to a complete unstructured protein, contrary to the assumption that the unstructured linker region could be stabilizing the folded region. So, from CSI

analysis alone, we were able to understand why the first construct of RodZc comprising residues from 201 to 304 was completely unstructured.

Since we were unable to find backbone resonances for residues from 233 to 237, CSI didn't give any prediction about the secondary structure in this region. So, contrary to the prediction from PSIPRED, this region might be indeed unstructured.

To make further analysis, we moved on to the determination of the tertiary structure of RodZc with CS-ROSETTA *ab-initio* prediction tool, using as input not only the chemical shifts list but also the residual dipolar couplings measured for each backbone N-H bond. The family of ten high-quality structures calculated with CS-ROSETTA is shown in **Figure V, Appendix chapter**. In **Figure 27** is shown the lowest-energy 3D-model retrieved from this prediction. The main body of the protein, namely the b-strands and the hydrophobic core is seen to be well defined, except two parallel aminoacid strands, S9 and S10, which comprise the last from the protein domain residues that don't exactly fit a perfect beta-sheet motif.

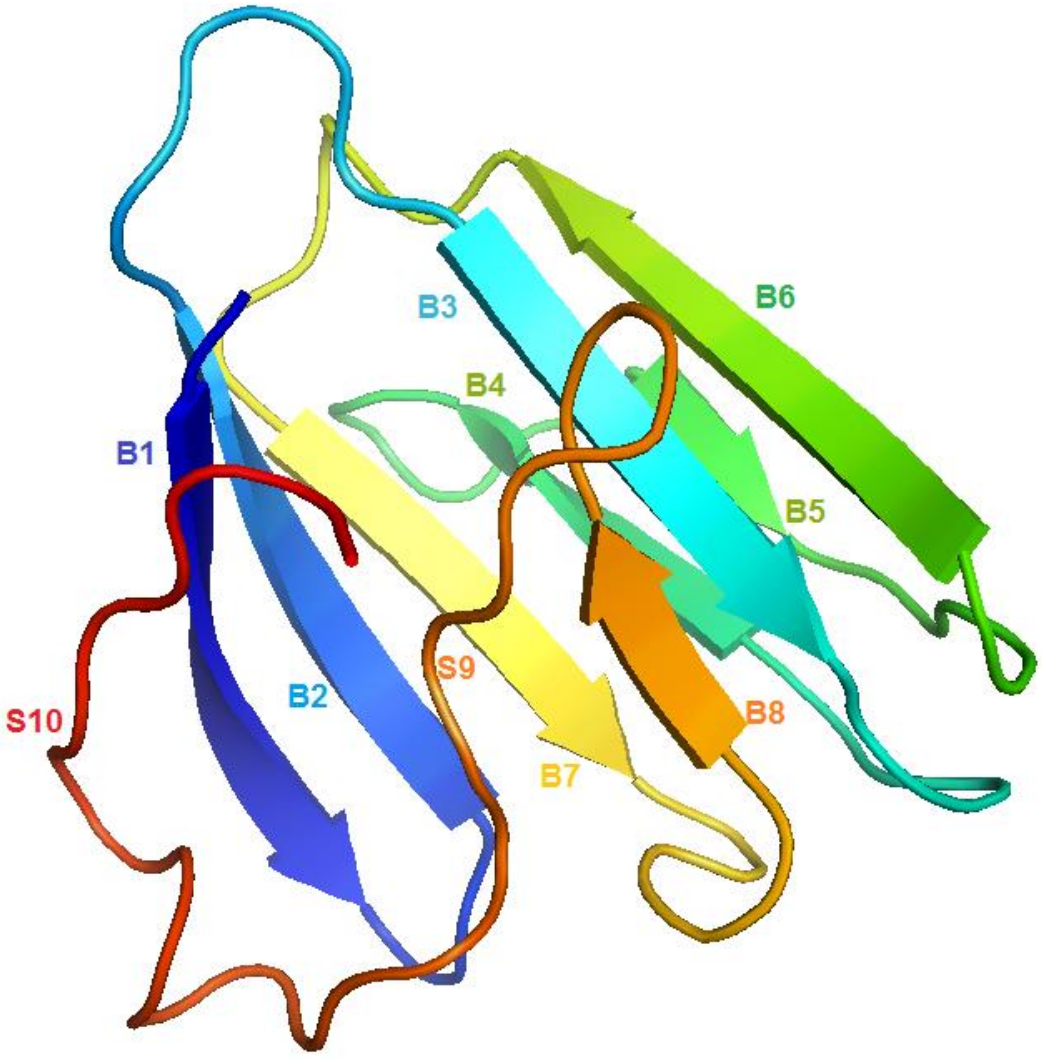

Figure 27 – Lowest-energy 3D-model calculated for RodZc with CS-ROSETTA *ab-initio* based prediction tool. Beta-strands are numbered from B1-B8. Two additional strands that didn't fold in Beta-sheet, are named S9 and S10.

Based in the Ig-like secondary structure classification criteria (Kabsch and Sander 1983), eight beta-strands are identified, namely B1 to B8, plus two unstructured strands (S9 and S10). The pairing pattern involves strands B1, B2, B7, B4, B5 and S10 on one hand, and the remaining strands on the other hand, foming a beta-sandwich motif.

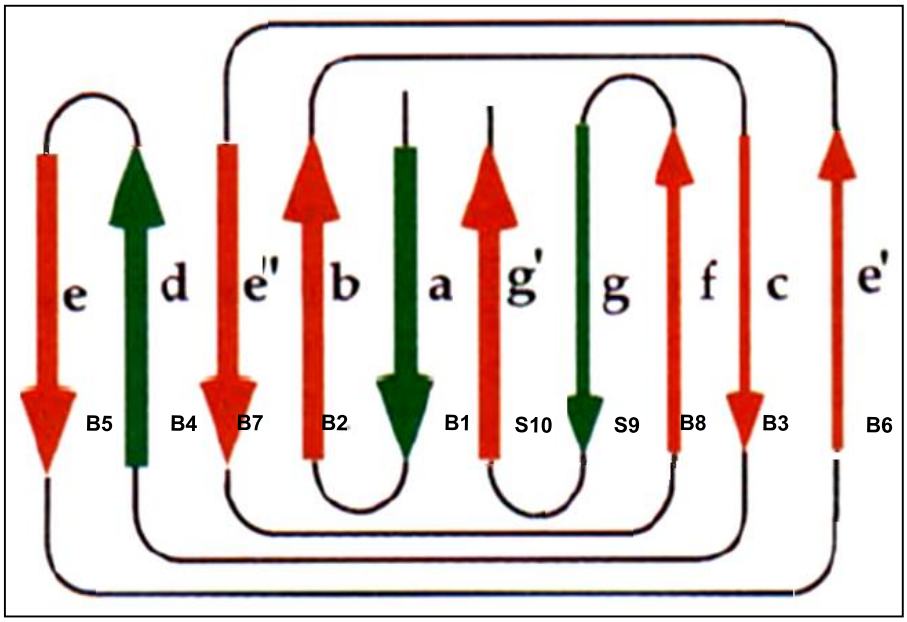

Figure 28 – Ig-like topology of RodZc moiety built based on the analysis of the bundle of structures obtained with CS-ROSETTA prediction. This topology is based in the V- type IG like fold.

It is worth noting, at this point, that the strand numbering described in **Figure 28,** does not exactly coincide with the notation of a v-type Ig-fold (**Figure 19**). Moreover, the two additional parallel unstructured strands that resemble Beta-strands were counted for the Beta-sheet motif. Consistent with the secondary chemical shifts, the protein model shows very distinct features, that deviate considerably from the v-type Ig-fold. So, we may be facing a new variation of the v-type Ig-fold.

4.3.5 Protein dynamics studies

The measurements of the 15N relaxation times T1 and T2 and the ^{15}N-^{1}H NOE for RodZc131-304 provided information about the backbone dynamics of this protein. We were able to quantify the peak intensities for the major assigned residues. The remaining signals for which no assignment was available were also analysed for their T1 and T2 times in order to compared with those of the folded regions. However, these were not evaluated in a quantitative way.

The experimental T1, T2 and ^{1}H-^{15}N Heteronuclear NOE values obtained for RodZc131-304 protein are plotted against aminoacid sequence in **Figure 29**.

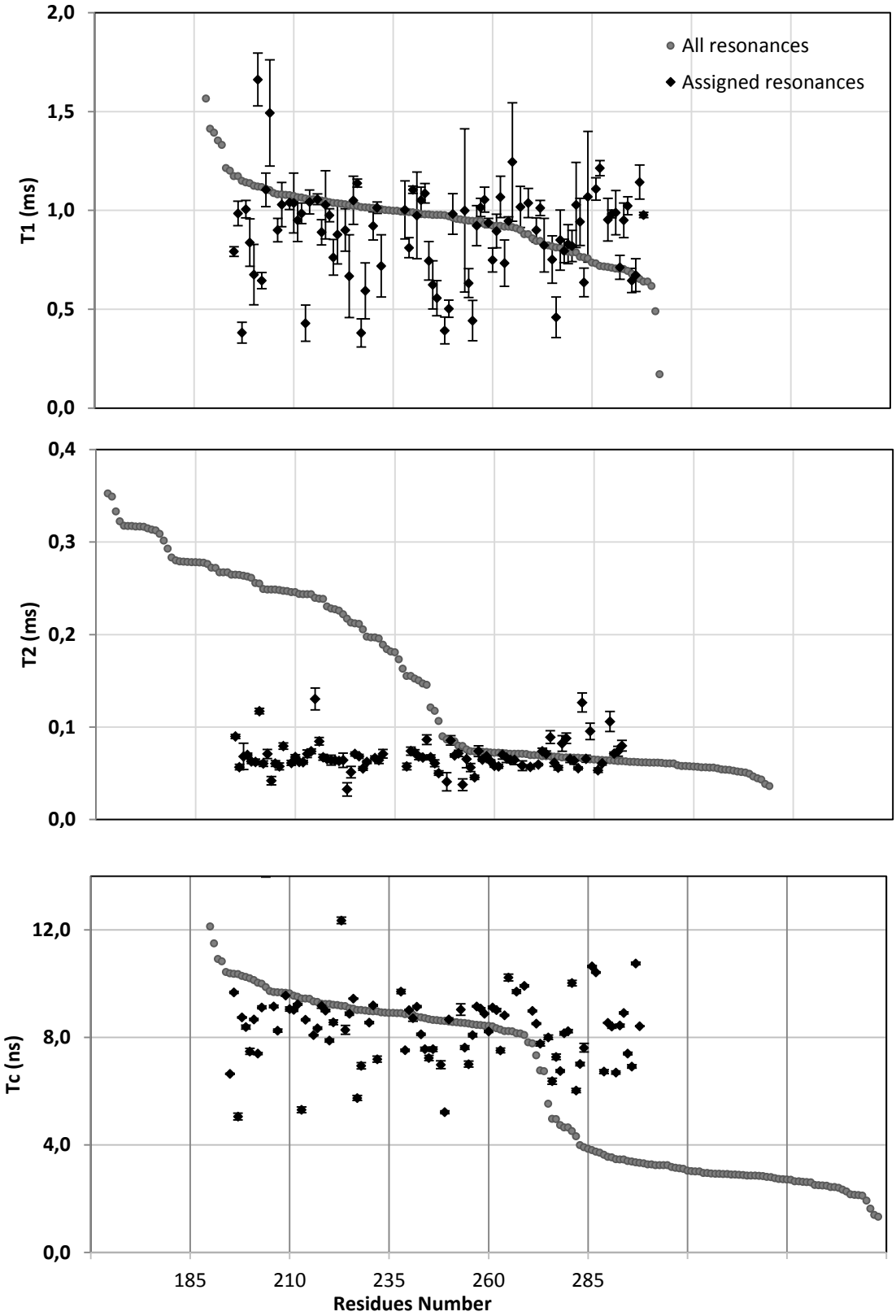

Figure 29 – T1, T2 and Tc values for each residue assigned from the Beta-rich RodZc moiety (residues from 190 to 301: black rhombus dots). The grey dots correspond to the full set of signals present in the corresponding NMR spectrum plotted in descending order.

There are two types of relaxation in NMR: T1, also known as Spin-Lattice, longitudinal relaxation, or relaxation in the z-direction) and T2, also known as Spin-Spin, transverse relaxation, or relaxation in the x-y plane. T1 relaxation corresponds to the process of r re-establishing the normal Gaussian population distribution of α and β spin states in the magnetic field, while T2corresponds to loss of phase coherence among nuclei.

In proteins, T2 relaxation is faster than T1 relaxation, since return of magnetization to the z-direction inherently causes loss of magnetization in the x-y plane. T1 and T2 relaxation are primarly affected by molecular tumbling.

We can define a correlation time (Tc) for a molecule, assuming it behaves more or less spherically. This is the average time it takes the molecule to rotate through one radian. The correlation time for small molecules is of the order of 10-12 s in solution (longer in viscous solvents). On the other hand, larger molecules are usually moving too slowly (Tc is too long), and they have the opposite relationship between molecular motion and T1 (i.e., relaxation is more effective when the molecule moves faster). For globular proteins a spherical approximation can be used and the rotational correlation time is given by Stoke's law:

$$\tau_c = \frac{4\pi\eta r^3}{3kT}$$

where η is the viscosity of the solvent, *r* is the effective hydrodynamic radius of the protein molecule, *k* is the Boltzmann constant and *T* is the temperature.

Since our beta-sheet moiety behaves like a rigid protein, with $\tau_c >> 0.5$ ns making it in the limit of slow molecular motion, a closed-form solution for τ_c as a function of the ratio of the longitudinal (T_1) and transverse (T_2) ^{15}N relaxation times exists:

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2} - 7},$$

where ν_N is the ^{15}N resonance frequency (in Hz). This equation diminishing ^{15}N T2 becomes difficult for larger proteins and cross-correlated relaxation rates are measured instead.

Average ^{15}N T1 and T2 relaxation times for a given protein were measured using pseudo 3D 15N-HSQC edited relaxation experiments as described in the Material and Methods section.

After determining the correlation time using the aforementioned equation (a simplified form of Eq. 8 from **Kay et al., 1989**, neglecting higher frequency terms), Tc values obtained were compared your Tc to a standard plot of Tc (ns) vs. MW (kDa) for known monomers (**Figure 30**). Since temperature is an extremely important variable, all data was acquired at the same temperature as the one used to plot this curve (298K).

The T1/T2 method is suitable for RodZc protein domain since it has a molecular weight lower than 25 kDa (22.3kDa).

These sets of charts were plotted against all resonances present in each edited 2D ^{15}N-^{1}H- HSQC experiments (grey circles). This way we can evaluate all ^{15}N relaxation values of the assigned residues in respect to the complete set of ^{1}H-^{15}N resonances belonging to the full construct. The full set of ^{1}H-^{15}N resonances are plotted in a decreasing order, not respecting any residue-order.

As discussed before, in proteins, T2 is always smaller than T1, and is the limiting factor for most NMR experiments. In our case the range of the values of T1 is small, meaning that T1 relaxation is almost constant for the full construct. In contrast, most of the T2 values present in the second chart (black rhombus) are below 0.1 s. In this chart, values for the full set of T2 measurements, including the indentified flexible linker peaks (grey circles) are somewhat partitioned in three sets of different relaxation regimes: values lower than 0.1s for fast relaxation, values between 0.1-0.2 s for an intermediate relaxation regime and values higher than 0.2 s for slow relaxation. This means that we are in fact in the presence of a protein with two different and almost independent relaxation regimes. So T2 measurements gave the biggest input to the correlation time (Tc) calculations than T1.

As we have explored before in this chapter, τc is a time constant that corresponds to the time it takes for a molecule to make a full rotation and is related to the molecule size. So we are expecting that the rigid beta-sheet moiety of RodZc will be tumbling with the same rate while the unfolded linker will have a tumbling time that is almost independent of the protein size.

Small τc values are usually related to small molecules. Since the average Tc corresponding to the unassigned linker residues is low (lower than 4 ns), this evidence corroborates the hypothesis of the linker region being unstructured and independent of the b-sheet moiety. This is also shown by the low T2 relaxation values for the linker region, thereby behaving like a small protein (lower than 8 kDa, as show in **Figure 30,** grey box).

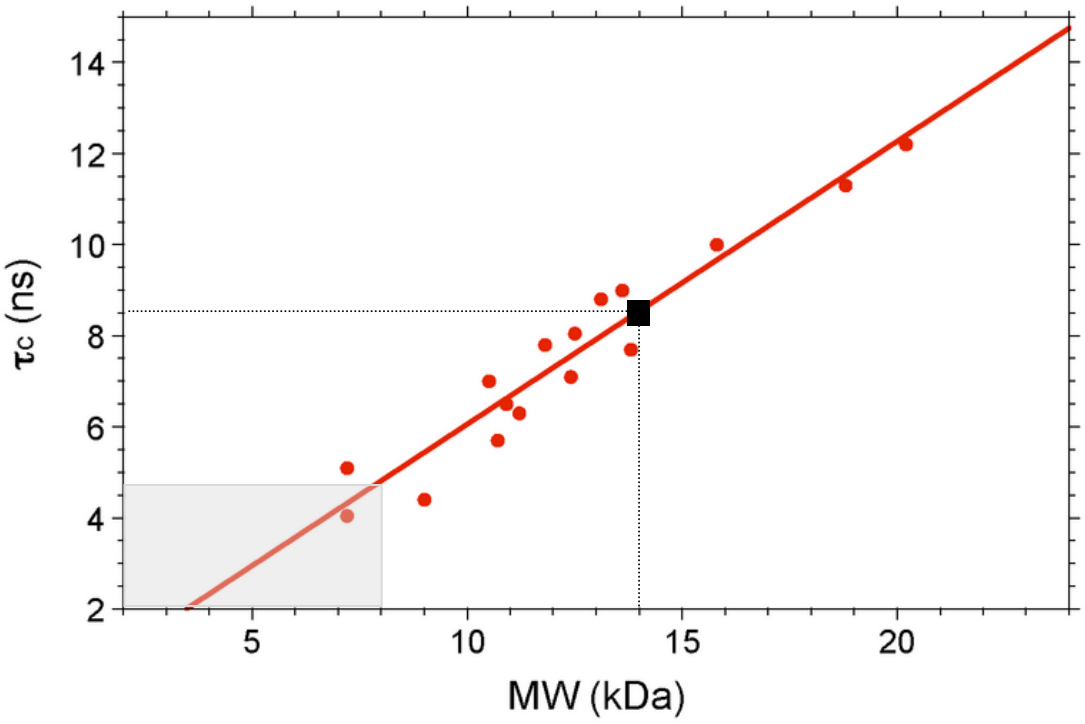

Figure 30 - Standard curve of τ_c vs. protein molecular weight (MW) obtained at the same temperature on a series of known monomeric proteins of varying size. Tc values were compiled for known monomeric NESG targets (Raman and Srivatsan, 2010). All data was recorded on a Bruker 600 NMR instrument at 298 K. The molecular weight for each target takes into account isotopic enrichment and the presence of affinity purification tags (if any).

As for the Tc values of the assigned residues, they can be divided into two sets: Tc values ranging from 4-8 ns and Tc values ranging from 8-12 ns. High Tc values are related to large, slow moving molecules. The average for the Tc values for the 8-12ns range is 8.5 ns (as shown in **Figure 29**), corresponding to a tumbling of a folded 14 kDa molecule. This favours the observation that the beta-sheet rich moiety and the linker tumble separately and don't interact with each other.

By comparing with the secondary structure of the predicted model of RodZc with CS-ROSSETA, we can assign the intermediate Tc values to residues from random-coil regions, in between the beta-sheets.

As for the ^{1}H-^{15}N-NOE determined values, they were also plotted against the amino acid sequence (**Figure 31**). With an average of 0.79 at 800 MHz, they presented a similar profile. Major deviation from the average value can be identified on the unstructured coiled regions, being consistent with the predicted model from CS-ROSETTA.

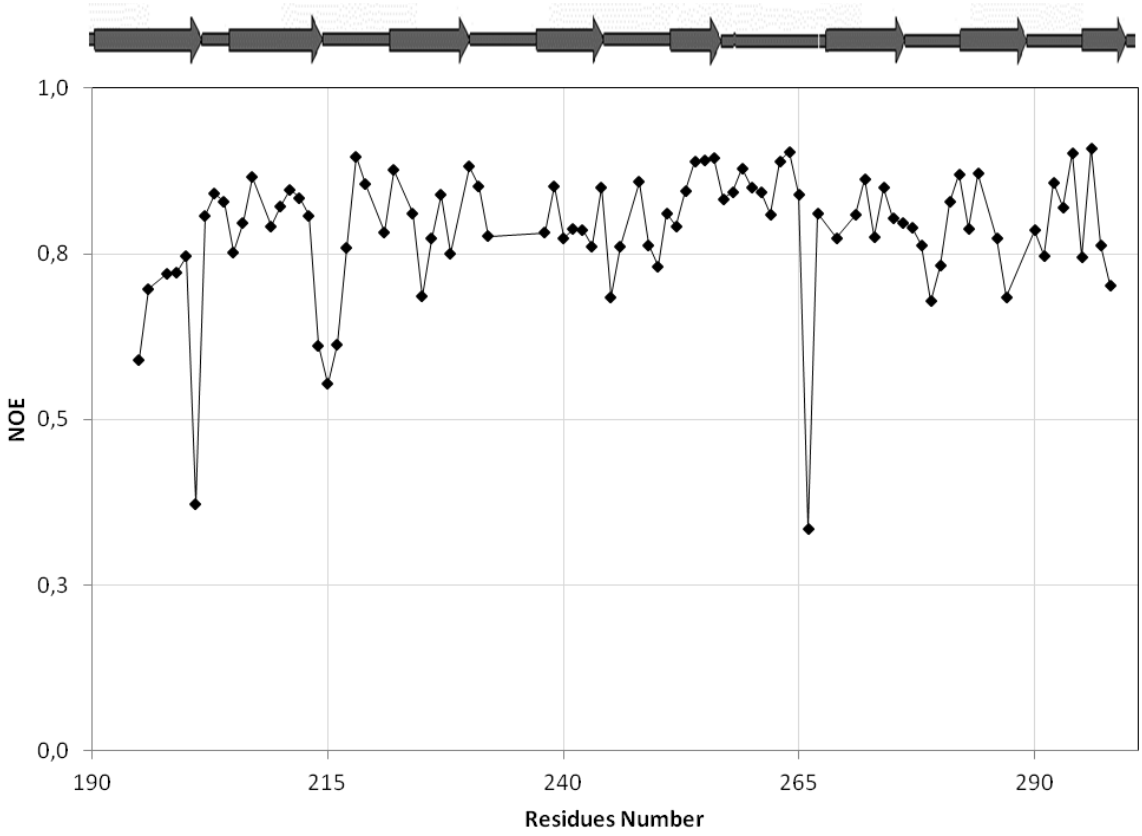

Figure 31 – ^{1}H-^{15}N-NOE values plotted against the amino acid sequence. Comparison of the overall chart with the rich in beta-strands secondary structure moiety of RodZc determined with CS-ROSETTA.

4.4 Conclusions

In this investigation, we combined bioinformatics, biophysical and biochemical methodologies to the analysis of the structure and function of RodZc in *B. subtilis*.

RodZc showed hydrodynamic and diffusion properties of a protein with MW about 1.5-2x larger than predicted. However, we found no evidence for a monomer-dimer equilibrium, and we favour the scenario that RodZc is a monomeric, elongated molecule. We cannot presently exclude though that in vivo RodZ functions as a dimer.

While far-UV CD data indicated that RodZc131-304 is poorly folded, showing a slight tendency for beta-sheet structures, this technique is not sensitive enough to distinguish between beta-sheet and random-coil secondary structure motif. So, all of these assumptions were confirmed by determining a model of RodZc using NMR data as input in the calculations. Also, all the ^{15}N relaxation measurements allowed us to understand the dynamics of RodZc protein and conclude that the linker region doesn't interact with the beta-sheet rich moiety, thereby having two completely different relaxation regimes.

The identification of the c-terminal moiety has having an Ig-like fold has several implications. The Ig fold is normally associated to protein-protein or protein-ligand interactions, in which any surface of the domain may be involved (**Dhe-Paganon et al., 2002; Krimm et al., 2002**). The extent and sequence of the connecting loops are extremely variable; the fringes of the domain are also variable in position, length, and number of beta-strands that are attached to the four-strand common core. These features explain the low sequence identity among members of the family, and the flexibility of the Ig fold as an interaction surface (**Bork et al., 1994**).

Interactions of Ig domains with their ligands can occur through the loop region or by parts of the beta-sheets (**Bork et al., 1994; Halaby et al., 1999**). For instance, Lamins bind to proteins, including heavily glycosylated proteins, through its Ig-like LTD (**Dittmer and Misteli, 2011**).

Evidence suggests that the strands at the fringes of the core Ig domain may be structurally flexible and it seems possible that the poorly structured RodZc only acquires a stable structure in the presence of a binding ligand. We hypothesize that the features of an Ig-fold for the RodZc moiety are directly linked to the ability of RodZ to bind multiple components of the cell wall elongation machinery. These include MreC and MreD and several other proteins of unknown function in *E. coli* and *S. coelicolor* (**Benzedu et al., 2009 (Kleinschnitz et al., 2011**).

A more rigorous analysis of the structure of RodZc using traditional NMR protocols is currently underway and is expected to shed more light into these issues.

Chapter 5

Concluding remarks and future perspectives

In this investigation, we combined bioinformatics, biophysical, biochemical and cell biological methodologies to the analysis of the structure and function of both RodZn and RodZc domains of RodZ protein from *B. subtilis*.

Even though it seems likely that gram-positive *B. subtilis* RodZ directly interacts with MreB, we were unable to detect a direct interaction between the two purified proteins through NMR titration experiments. These experiments however, need to be repeated and studied at higher detail. Furthermore, they should also attempt to investigate the interaction between RodZ and MreB in the context of the membrane, since the transmembrane domain (TM) of *B. subtilis* RodZn as well as its counterpart in *E. coli*, is important for protein localization and cell elongation (**Bendezu et al, 2009; Shiomi et al, 2008**).

Since the HTH motif is a common motif for DNA binding proteins, the question of a potential function of RodZn related with DNA organization should also be examined. For this purpose we intend to conduct new NMR titration experiments with MreB in the presence of bacterial DNA fragments in order to study if RodZ in *B. subtilis* can form complexes with MreB and DNA in which different regions of the full-helical domain would be involved in RodZ-DNA and RodZ-MreB interactions.

Any potential Protein-DNA interaction studies will also be complemented with data driven computational docking tools such as HADDOCK (High Ambiguity Driven protein-protein DOCKing) which is an information-driven flexible docking approach for the modeling of biomolecular complexes (http://haddock.science.uu.nl/).

Regarding the periplasmic domain of Rodz (RodZc) very little information was previously available. Study of its hydrodynamic and diffusion properties were consistent with the behaviour of a

protein with MW about 1.5-2x larger than expected. However, we found no evidence for a monomer-dimer equilibrium. In fact, NMR relaxations measurements conclusively showed that it is a monomer in solution.

Regardinig its secondary structure, far-UV CD data indicated that RodZc131-304 is only partially folded, showing a slight tendency for beta-sheet structures. However, this technique is not sensitive enough to distinguish between beta-sheet and random-coil secondary structure motif. All of these assumptions were confirmed by determining a 3-D model of RodZc using NMR data, which tertiary structure is consistent with an Ig-like fold. Also, ^{15}N relaxation measurements allowed us to understand the dynamics of RodZc protein and conclude that the linker region doesn't interact with the beta-sheet rich moiety, thereby having two completely different relaxation regimes.

The identification of the c-terminal moiety as having an Ig-like fold has several implications. The Ig fold is normally associated to protein-protein or protein-ligand interactions, in which any surface of the domain may be involved (**Dhe-Paganon et al., 2002; Krimm et al., 2002**). In fact, interactions of Ig domains with their ligands can occur through the loop region or by parts of the beta-sheets (**Bork et al., 1994; Halaby and Mornon, 1999**).

Evidence suggests that the strands at the fringes of the core Ig domain may be structurally flexible and it seems possible that the poorly structured RodZc only acquires a stable structure in the presence of a binding ligand. Our conjecture is that the features of an Ig-fold for the RodZc moiety are directly linked to the ability of RodZ to bind multiple components of the cell wall elongation machinery. These include MreC and MreD and several other proteins of unknown function in *E. coli* and *S. coelicolor* (**Benzedú et al., 2009; Kleinschnitz et al., 2011**).

A new construct for RodZc190-304 has already been cloned and transformed and is ready for expression, purification and NMR analysis. By collecting new data sets for this new construct and with the data we have already collected for the full length domain, we aim to produce a high quality NMR structure of this domain. We will also collect new ^{15}N relaxation data at different external magnetic fields and use model free analysis for an accurate analysis of the dynamics of the domain since it has been shown that flexibility often plays key role in the mediation of protein-protein interactions, and in substrate recognition.

Chapter 6

Appendix

Table I – Relevant data regarding solutions composition, buffers composition and bacterial strains used.

Solution	**Composition**
Culture Media (Autoclaved)	
LB medium	10 g/L Tryptone; 5 g/L Yeast extract; 5 g/L NaCl
Minimal medium M9	1X M9 salt mix; 0.002 M MgSO4 Sln; 0.025 uM CaCl2; 0.3X metal mix solution: 50 mg/mL Kanamycine: 0.08% (v/v) Carbon source (Glucose, 13C); 0.005% (v/v) Nitrogen source (NH4Cl, 15N)
10X M9 salt mix	12.8% (w/v) Na2HPO4.7H2O; 3% (w/v) KH2PO4; 0.5% (w/v) NaCl in ddH2O; pH 7.0
10X metal mix solution	0.4M HCl; 8.3% (w/v) FeSO4.7H2O; 0.6% (w/v)H3BO4; 0.02% (w/v) CoCl2.6H2O; 0.004% (w/v) CuCl2.2H2O; 0.34% (w/v) ZnCl2; 0.61% (w/v) Na2MoO4.2H2O; 0.04% (w/v) MnCl2.4H2O

Expression and Purification Protocol (all Filtrated with 0,45 µm cut-off)	
Phosphate Buffer pH 7.4	0.4149% (w/v) Na2HPO4.2H2O; 0.0623% (w/v) NaH2PO4.2H2O
Phosphate Buffer pH 6.5	0.1616% (w/v) Na2HPO4.2H2O; 0.1928% (w/v) NaH2PO4.2H2O
Lysis Buffer	20 mM Phosphate Buffer pH 7.4, 0.5 M NaCl, 10% Glycerol, 1mM NaN3, 1mM PMSF, 1 pill ROCHE protease inhibitor cocktail EDTA free
IMAC Buffer A	20 mM Phosphate Buffer pH 7.4, 0.5 M NaCl, 10% Glycerol, 1mM NaN3
IMAC Buffer B	20 mM Phosphate Buffer pH 7.4, 0.5 M NaCl, 10% Glycerol, 1mM NaN3, 0.5M Imidazole
SEC Buffer	20 mM Phosphate Buffer pH 6.5, 100 mM NaCl, 1mM NaN3
Strains used in this study	
AH4333	BL21 (DE3) pAP1 / Ap^R Km^R (overproduction of $RodZ_{N\ 1\text{-}101}$-His_6)
AH4334	BL21 (DE3) pAP2 / Ap^R Km^R (overproduction of His_6-$RodZ_{C\ 201\text{-}304}$)
AH4268	BL21 (DE3) pTC248 / Ap^R Km^R (overproduction of His_6-$RodZ_{C\ 131\text{-}304}$)

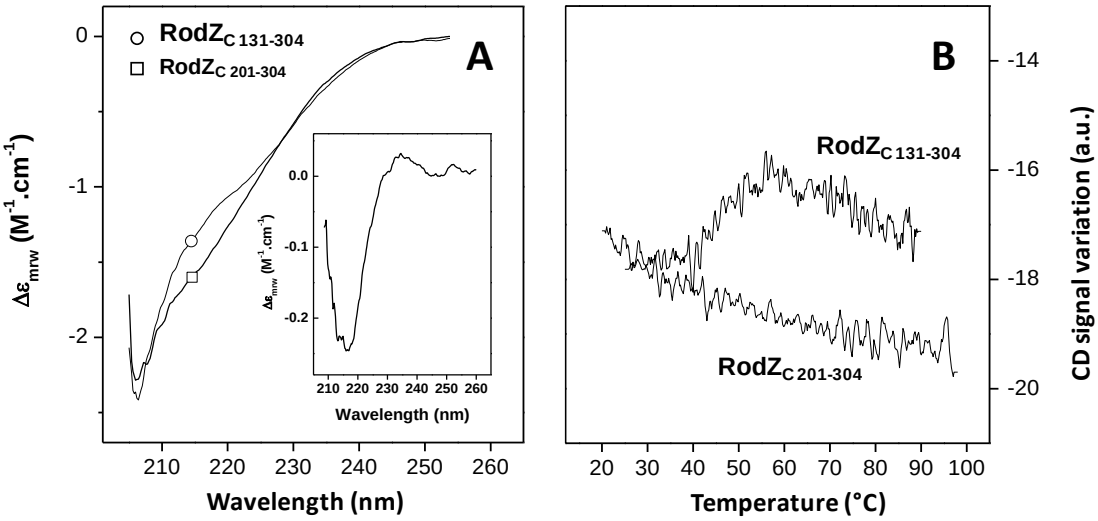

Figure I - Circular dichroism spectroscopy of RodZc. A: Far-UV circular dichroism (CD) spectra of His6-RodZc131-304 and His6-RodZ201-304. The CD spectrum was determined for the two proteins in the range of 190-250 nm and is the average of 3 measurements. The insert shows the computed difference spectrum between those of His6-RodZc131-304 and His6-RodZc201-304. B: thermal unfolding of His6-RodZc131-304 and His6-RodZc201-304 as followed by CD spectroscopy.

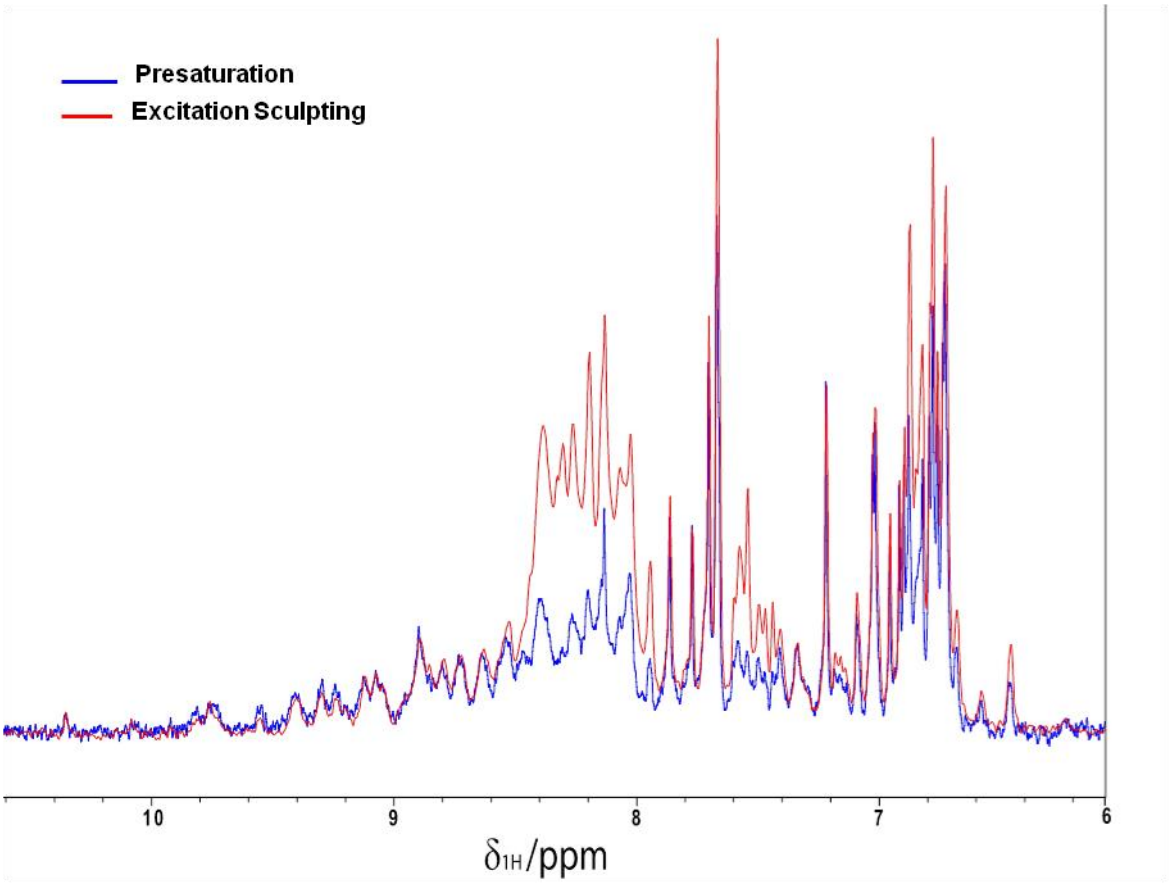

Figure II - 1D ^{1}H NMR spectra collected with different pulse sequences for water suppression, namely presaturation (blue) and excitation sculpting (red).

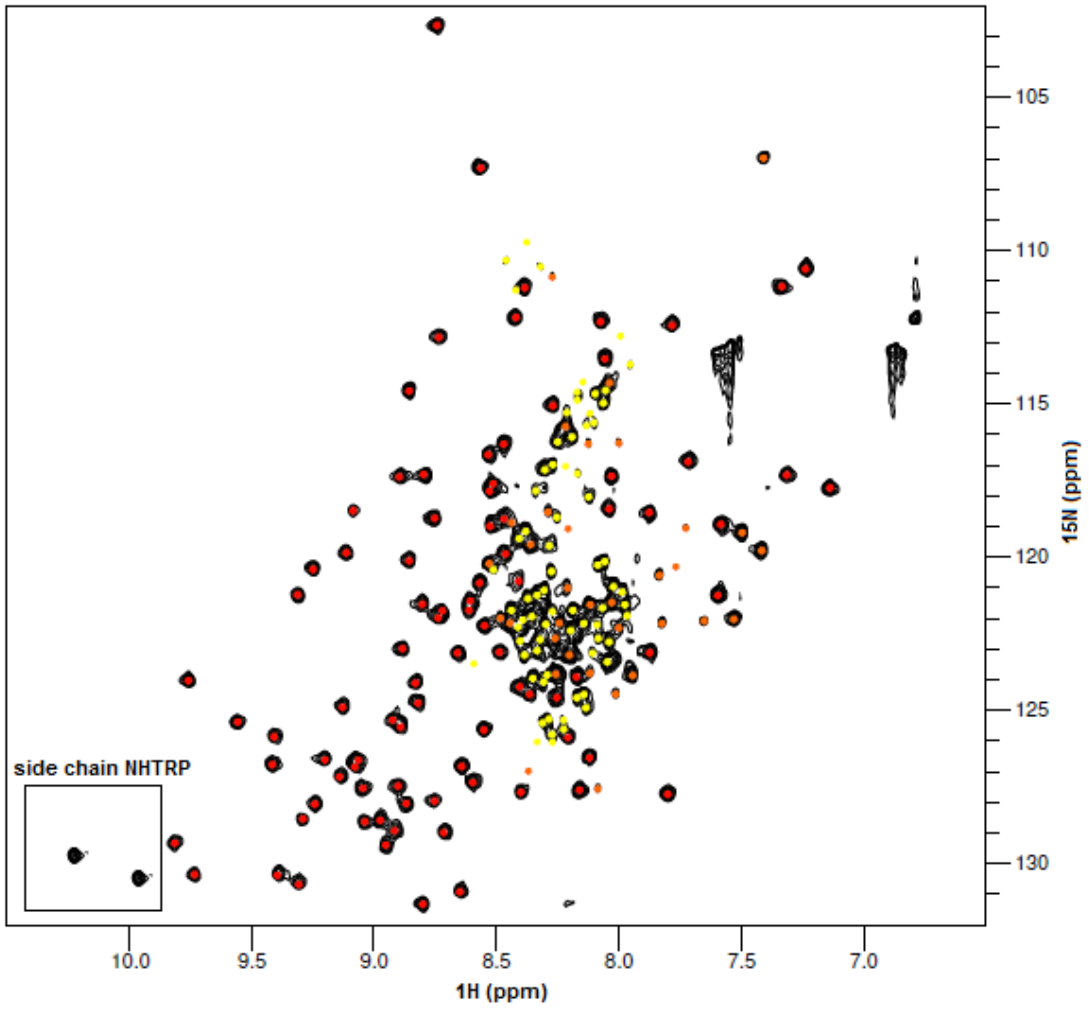

Figure III - 2D ^{1}H-^{15}N HSQC spectrum of RodZc131-304. NH signals are coloured coded in respect to the probability of being in: a folded region (red), intermediate region like random-coils in between folded regions (orange) and unstructured region (yellow). The signals of the NH from the tryptophan side chain are highlighted (black box on the left corner).

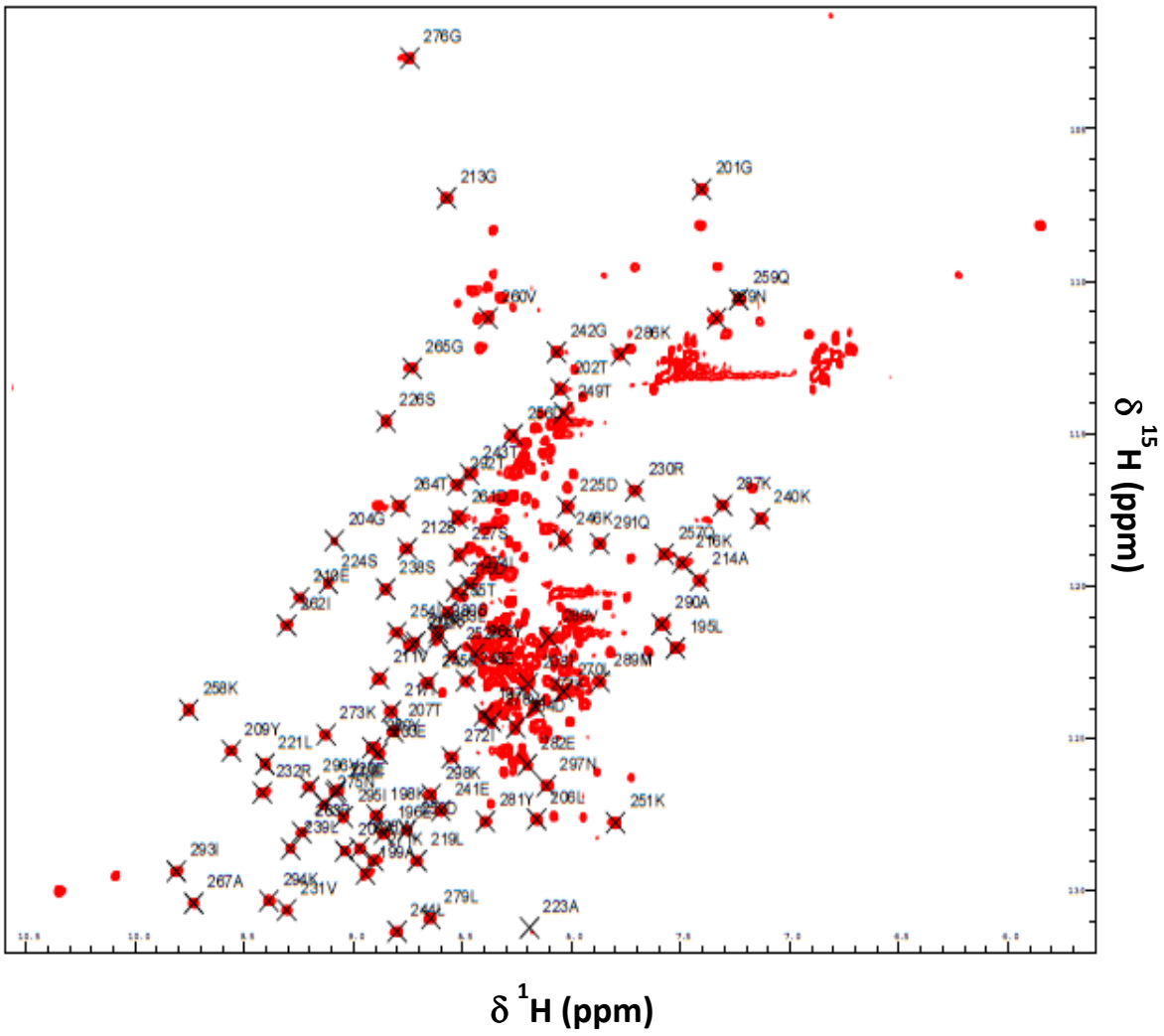

Figure IV - 2D ^{1}H-^{15}N HSQC spectrum of RodZc$_{131-304}$ showing all the assigned residues belonging to the structured region of the periplasmic domain of RodZ.

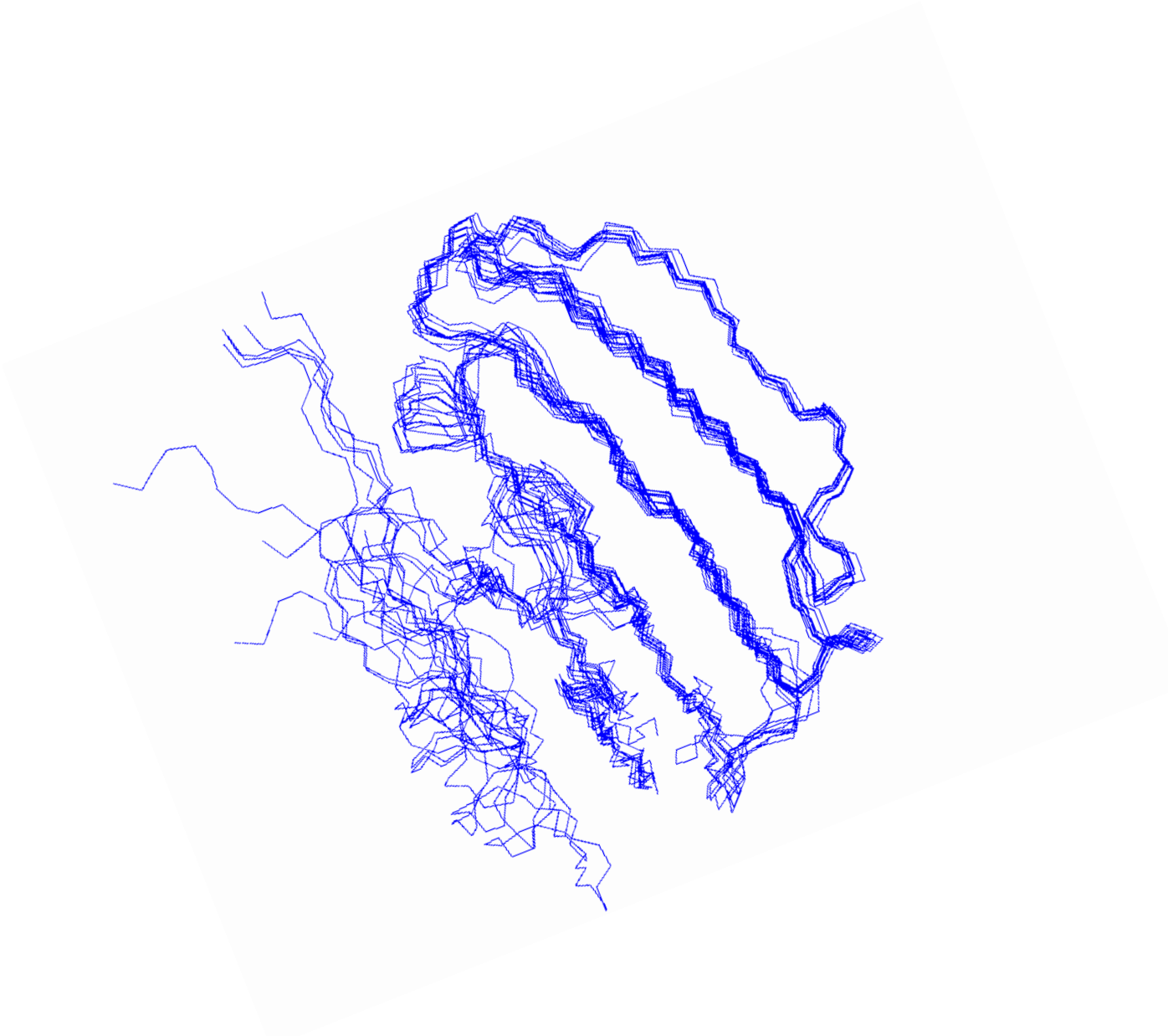

Figure IV – Bundle of ten high-quality 3D-models predicted with CS-ROSETTA *ab-initio* based tool for RodZc$_{190-304}$.

Chapter 7

Bibliography

Adams, David W, e Jeff Errington. «Bacterial cell division: assembly, maintenance and disassembly of the Z ring.» *Nature reviews. Microbiology* 7, n 9 (Sep. 2009): 642–53. doi:10.1038/nrmicro2198.

Altschul, S. «Gapped BLAST and PSI-BLAST: a new generation of protein database search programs». *Nucleic Acids Research* 25, n 17 (Sep. 1997): 3389–3402. doi:10.1093/nar/25.17.3389.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, e David J. Lipman. «Basic local alignment search tool». *Journal of Molecular Biology* 215, n 3 (1990): 403–410.

Alyahya, S Anisah, Roger Alexander, Teresa Costa, Adriano O Henriques, Thierry Emonet, e Christine Jacobs-Wagner. «RodZ, a component of the bacterial core morphogenic apparatus.» *Proceedings of the National Academy of Sciences of the United States of America* 106, n 4 (Jan. 2009): 1239–44. doi:10.1073/pnas.0810794106.

Arnold, Konstantin, Lorenza Bordoli, Jürgen Kopp, e Torsten Schwede. «The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.» *Bioinformatics (Oxford, England)* 22, n 2 (Jan. 2006): 195–201. doi:10.1093/bioinformatics/bti770.

Bandow, J. E., H. Brotz, e M. Hecker. «Bacillus subtilis Tolerance of Moderate Concentrations of Rifampin Involves the B-Dependent General and Multiple Stress Response». *Journal of Bacteriology* 184, n 2 (Jan. 2002): 459–467. doi:10.1128/JB.184.2.459-467.2002.

Barbato, G, M Ikura, L E Kay, R W Pastor, e A Bax. «Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible.» *Biochemistry* 31, n 23 (June 1992): 5269–78.

Bendezú, Felipe O, Cynthia a Hale, Thomas G Bernhardt, e Piet a J Boer. «RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in *E. coli*.» *The EMBO journal* 28, n 3 (Feb. 2009): 193–204. doi:10.1038/emboj.2008.264.

Bendezú, Felipe O, Cynthia A Hale, Thomas G Bernhardt, e Piet A J Boer. «RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in *E. coli*.» *The EMBO journal* 28, n 3 (Feb. 2009): 193–204. doi:10.1038/emboj.2008.264.

Berman, H. M. «The Protein Data Bank». *Nucleic Acids Research* 28, n 1 (Jan. 2000): 235–242. doi:10.1093/nar/28.1.235.

Bertini, Ivano, David A Case, Lucio Ferella, Andrea Giachetti, e Antonio Rosato. «A Grid-enabled web portal for NMR structure refinement with AMBER.» *Bioinformatics (Oxford, England)* 27, n 17 (Sep. 2011): 2384–90. doi:10.1093/bioinformatics/btr415.

Bhavsar, Amit P, e Eric D Brown. «Cell wall assembly in Bacillus subtilis: how spirals and spaces challenge paradigms.» *Molecular microbiology* 60, n 5 (June 2006): 1077–90. doi:10.1111/j.1365-2958.2006.05169.x.

Biegert, Andreas, Christian Mayer, Michael Remmert, Johannes Söding, e Andrei N Lupas. «The MPI Bioinformatics Toolkit for protein sequence analysis.» *Nucleic acids research* 34, n Web Server issue (July 2006): W335–9. doi:10.1093/nar/gkl217.

Bonneau, R, e D Baker. «Ab Initio Protein Structure Prediction: Progress and Prospects.» *Annual Review of Biophysics and Biomolecular Structure* 30 (Jan. 2001): 173–89. doi:10.1146/annurev.biophys.30.1.173.

Bonneau, R, J Tsai, I Ruczinski, D Chivian, C Rohl, C E Strauss, e D Baker. «Rosetta in CASP4: progress in ab initio protein structure prediction.» *Proteins* Suppl 5 (Jan. 2001): 119–26.

Bordoli, Lorenza, Florian Kiefer, Konstantin Arnold, Pascal Benkert, James Battey, e Torsten Schwede. «Protein structure homology modeling using SWISS-MODEL workspace.» *Nature protocols* 4, n 1 (Jan. 2009): 1–13. doi:10.1038/nprot.2008.197.

Bork, P., L. Holm, e C. Sander. «The Immunoglobulin Fold». *Journal of Molecular Biology* 242, n 4 (1994): 309–320.

Bourne, Philip E., e Helge Weissig, eds. *Structural Bioinformatics*. Methods of Biochemical Analysis. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003.

Bowers, Peter M., Charlie E.M. Strauss, e David Baker. «De Novo Protein Structure Determination Using Sparse NMR Data». *Journal of Biomolecular NMR* 18, n 4 (Dec. 2000): 311–318. doi:10.1023/A:1026744431105.

Bramucci, Emanuele, Alessandro Paiardini, Francesco Bossa, e Stefano Pascarella. «PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL.» *BMC bioinformatics* 13 Suppl 4 (Jan. 2012): S2. doi:10.1186/1471-2105-13-S4-S2.

Brenner, S. E., C. Chothia, e T. J. P. Hubbard. «Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships». *Proceedings of the National Academy of Sciences* 95, n 11 (May 1998): 6073–6078. doi:10.1073/pnas.95.11.6073.

Burdett, I. D. J., e M. L. Higgins. «Study of Pole Assembly in Bacillus subtilis by Computer Reconstruction of Septal Growth Zones Seen in Central, Longitudinal Thin Sections of Cells». *J. Bacteriol.* 133, n 2 (Feb. 1978): 959–971.

Cabeen, Matthew T, Harald Herrmann, e Christine Jacobs-Wagner. «The domain organization of the bacterial intermediate filament-like protein crescentin is important for assembly and function.» *Cytoskeleton (Hoboken, N.J.)* 68, n 4 (Ap. 2011): 205–19. doi:10.1002/cm.20505.

Carr, H., e E. Purcell. «Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments». *Physical Review* 94, n 3 (May 1954): 630–638. doi:10.1103/PhysRev.94.630.

Den Blaauwen, Tanneke, Miguel a Pedro, Martine Nguyen-Distèche, e Juan a Ayala. «Morphogenesis of rod-shaped sacculi.» *FEMS microbiology reviews* 32, n 2 (Mar. 2008): 321–44. doi:10.1111/j.1574-6976.2007.00090.x.

Chou, James J., Sander Gaemers, Bernard Howder, John M. Louis, e Ad Bax. «A Simple Apparatus for Generating Stretched Polyacrylamide Gels, Yielding Uniform Alignment of Proteins and Detergent Micelles*». *Journal of Biomolecular NMR* 21, n 4 (Dec. 2001): 377–382. doi:10.1023/A:1013336502594.

Daragan, Vladimir A., e Kevin H. Mayo. «Motional model analyses of protein and peptide dynamics using and NMR relaxation». *Progress in Nuclear Magnetic Resonance Spectroscopy* 31, n 1 (July 1997): 63–105.

Das, Rhiju, Bin Qian, Srivatsan Raman, Robert Vernon, James Thompson, Philip Bradley, Sagar Khare, et al. «Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home.» *Proteins* 69 Suppl 8 (Jan. 2007): 118–28. doi:10.1002/prot.21636.

Davis, Adrian L, James Keeler, Ernest D Laue, e Detlef Moskau. «Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients». *Journal of Magnetic Resonance (1969)* 98, n 1 (June 1992): 207–216. doi:10.1016/0022-2364(92)90126-R.

Defeu Soufo, Hervé Joël, e Peter L Graumann. «Dynamic movement of actin-like proteins within bacterial cells.» *EMBO reports* 5, n 8 (Aug. 2004): 789–94. doi:10.1038/sj.embor.7400209.

Dhe-Paganon, Sirano, Eric D Werner, Young-In Chi, e Steven E Shoelson. «Structure of the globular tail of nuclear lamin.» *The Journal of biological chemistry* 277, n 20 (May 2002): 17381–4. doi:10.1074/jbc.C200038200.

Dickerson, Richard Earl, e Irving Geis. *The structure and action of proteins*, 1969.

Dill, Ken A, S Banu Ozkan, Thomas R Weikl, John D Chodera, e Vincent A Voelz. «The protein folding problem: when will it be solved?» *Current opinion in structural biology* 17, n 3 (June 2007): 342–6. doi:10.1016/j.sbi.2007.06.001.

Dittmer, Travis A, e Tom Misteli. «The lamin protein family.» *Genome biology* 12, n 5 (Jan. 2011): 222. doi:10.1186/gb-2011-12-5-222.

Divakaruni, Arun V, Cyril Baida, Courtney L White, e James W Gober. «The cell shape proteins MreB and MreC control cell morphogenesis by positioning cell wall synthetic complexes.» *Molecular microbiology* 66, n 1 (Oct. 2007): 174–88. doi:10.1111/j.1365-2958.2007.05910.x.

Divakaruni, Arun V, Rachel R Ogorzalek Loo, Yongming Xie, Joseph A Loo, e James W Gober. «The cell-shape protein MreC interacts with extracytoplasmic proteins including cell wall assembly complexes in Caulobacter crescentus.» *Proceedings of the National Academy of Sciences of the United States of America* 102, n 51 (Dec. 2005): 18602–7. doi:10.1073/pnas.0507937102.

Donachie, W D. «The Cell Cycle of Escherichia Coli.» *Annual Review of Microbiology* 47 (Jan. 1993): 199–230. doi:10.1146/annurev.mi.47.100193.001215.

Doreleijers, J F, J A Rullmann, e R Kaptein. «Quality assessment of NMR structures: a statistical survey.» *Journal of molecular biology* 281, n 1 (Aug. 1998): 149–64. doi:10.1006/jmbi.1998.1808.

Doreleijers, Jurgen F, Alan W Sousa da Silva, Elmar Krieger, Sander B Nabuurs, Christian a E M Spronk, Tim J Stevens, Wim F Vranken, Gert Vriend, e Geerten W Vuister. «CING: an integrated residue-based structure validation program suite.» *Journal of biomolecular NMR* 54, n 3 (Nov. 2012): 267–83. doi:10.1007/s10858-012-9669-7.

Doreleijers, Jurgen F., Mia L. Raves, Ton Rullmann, e Robert Kaptein. «Completeness of NOEs in Protein Structures: A Statistical Analysis of NMR Data». *Journal of Biomolecular NMR* 14, n 2 (June 1999): 123–132. doi:10.1023/A:1008335423527.

Doreleijers, Jurgen F., Johan A.C. Rullmann, e Robert Kaptein. «Quality assessment of NMR structures: a statistical survey11Edited by P. E. Wright». *Journal of Molecular Biology* 281, n 1 (1998): 149–164.

Dunker, A K, e Z Obradovic. «The protein trinity–linking function and disorder.» *Nature biotechnology* 19, n 9 (Sep. 2001): 805–6. doi:10.1038/nbt0901-805.

Dunker, A Keith, Marc S Cortese, Pedro Romero, Lilia M Iakoucheva, e Vladimir N Uversky. «Flexible nets. The roles of intrinsic disorder in protein interaction networks.» *The FEBS journal* 272, n 20 (Oct. 2005): 5129–48. doi:10.1111/j.1742-4658.2005.04948.x.

Dunker, A. Keith, Celeste J. Brown, J. David Lawson, Lilia M. Iakoucheva, e Zoran Obradović. «Intrinsic Disorder and Protein Function †». *Biochemistry* 41, n 21 (May 2002): 6573–6582. doi:10.1021/bi012159+.

Dye, Natalie A, Zachary Pincus, Julie A Theriot, Lucy Shapiro, e Zemer Gitai. «Two independent spiral structures control cell shape in Caulobacter.» *Proceedings of the National Academy of Sciences of the United States of America* 102, n 51 (Dec. 2005): 18608–13. doi:10.1073/pnas.0507708102.

Dyson, H.Jane, e Peter E Wright. «Coupling of folding and binding for unstructured proteins». *Current Opinion in Structural Biology* 12, n 1 (2002): 54–60.

Ed., P. Debrunner, J. C. M. Tsibris Ed., e E. Münck Ed. *Mössbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, March 17 and 18, 1969, Monticello, Ill., Organized by the University of Illinois at Urbana-Champaign, Departments of Chemistry and Physics*, 1969.

El Ghachi, Meriem, Pierre-Jean Matteï, Chantal Ecobichon, Alexandre Martins, Sylviane Hoos, Christine Schmitt, Frédéric Colland, et al. «Characterization of the elongasome core PBP2 : MreC complex of Helicobacter pylori.» *Molecular microbiology* 82, n 1 (Oct. 2011): 68–86. doi:10.1111/j.1365-2958.2011.07791.x.

Errington, Jeff. «Regulation of endospore formation in Bacillus subtilis.» *Nature reviews. Microbiology* 1, n 2 (Nov. 2003): 117–26. doi:10.1038/nrmicro750.

Figge, Rainer M, Arun V Divakaruni, e James W Gober. «MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in Caulobacter crescentus.» *Molecular microbiology* 51, n 5 (Mar. 2004): 1321–32. doi:10.1111/j.1365-2958.2003.03936.x.

Floudas, C A. «Computational methods in protein structure prediction.» *Biotechnology and bioengineering* 97, n 2 (June 2007): 207–13. doi:10.1002/bit.21411.

Folmer, R, e G Otting. «Sensitivity enhancement in (HCA)CONH experiments.» *Journal of biomolecular NMR* 16, n 3 (Mar. 2000): 229–33.

García la Torre, J, M.L Huertas, e B Carrasco. *HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations*. Vol 147. 1, 2000.

Gerdes, Kenn. «RodZ, a new player in bacterial cell morphogenesis.» *The EMBO journal* 28, n 3 (Feb. 2009): 171–2. doi:10.1038/emboj.2008.287.

Gerstein, M, e M Levitt. «Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins.» *Protein science : a publication of the Protein Society* 7, n 2 (Feb. 1998): 445–56. doi:10.1002/pro.5560070226.

Gitai, Zemer. «The New Bacterial Cell Biology: Moving Parts and Subcellular Architecture». *Cell* 120, n 5 (2005): 577–586.

Gitai, Zemer, Natalie Anne Dye, Ann Reisenauer, Masaaki Wachi, e Lucy Shapiro. «MreB Actin-Mediated Segregation of a Specific Region of a Bacterial Chromosome». *Cell* 120, n 3 (2005): 329–341.

Go, N. «Theoretical Studies of Protein Folding.» *Annual Review of Biophysics and Bioengineering* 12 (Jan. 1983): 183–210. doi:10.1146/annurev.bb.12.060183.001151.

Groves, Patrick, Martin Ohsten Rasmussen, M Dolores Molero, Eric Samain, F Javier Cañada, Hugues Driguez, e Jesús Jiménez-Barbero. «Diffusion ordered spectroscopy as a complement to size exclusion chromatography in oligosaccharide analysis.» *Glycobiology* 14, n 5 (May 2004): 451–6. doi:10.1093/glycob/cwh037.

Groves, Patrick, e Mateus Webba da Silva. «Rapid stoichiometric analysis of G-quadruplexes in solution.» *Chemistry (Weinheim an der Bergstrasse, Germany)* 16, n 22 (June 2010): 6451–3. doi:10.1002/chem.200901248.

Grzesiek, S, e A Bax. «Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins.» *Journal of biomolecular NMR* 3, n 2 (Mar. 1993): 185–204.

Grzesiek, S, e A Bax. «Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR». *Journal of the American Chemical Society* 114, n 16 (July 1992): 6291–6293. doi:10.1021/ja00042a003.

Guex, N, e M C Peitsch. «SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.» *Electrophoresis* 18, n 15 (Dec. 1997): 2714–23. doi:10.1002/elps.1150181505.

Halaby, D.M., A. Poupon, e J.-P. Mornon. «The immunoglobulin fold family: sequence analysis and 3D structure comparisons». *Protein Engineering Design and Selection* 12, n 7 (July 1999): 563–571. doi:10.1093/protein/12.7.563.

Hartmann, Guido. «The Structure and Action of Proteins. VonR. E. Dickerson undI. Geis. Harper and Row, Publishers, New York-Evanston-London 1969. 1. Aufl., VIII, 120 S., zahlr. Abb., Paperback DM 20.50». *Angewandte Chemie* 82, n 18 (Sep. 1970): 780–780. doi:10.1002/ange.19700821817.

Heijenoort, J. v. «Formation of the glycan chains in the synthesis of bacterial peptidoglycan». *Glycobiology* 11, n 3 (Mar. 2001): 25R–36R. doi:10.1093/glycob/11.3.25R.

Henikoff, J. G. «Increased coverage of protein families with the Blocks Database servers». *Nucleic Acids Research* 28, n 1 (Jan. 2000): 228–230. doi:10.1093/nar/28.1.228.

Herrmann, Torsten, Peter Güntert, e Kurt Wüthrich. «Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA». *Journal of Molecular Biology* 319, n 1 (2002): 209–227.

Holtje, Joachim-Volker. «Growth of the Stress-Bearing and Shape-Maintaining Murein Sacculus of Escherichia coli». *Microbiol. Mol. Biol. Rev.* 62, n 1 (Mar. 1998): 181–203.

Hong, Huynh A., Reena Khaneja, Nguyen M.K. Tam, Alessia Cazzato, Sisareuth Tan, Maria Urdaci, Alain Brisson, Antonio Gasbarrini, Ian Barnes, e Simon M. Cutting. «Bacillus subtilis isolated from the human gastrointestinal tract». *Research in Microbiology* 160, n 2 (2009): 134–143.

Hulo, Nicolas, Christian J A Sigrist, Virginie Le Saux, Petra S Langendijk-Genevaux, Lorenza Bordoli, Alexandre Gattiker, Edouard Castro, Philipp Bucher, e Amos Bairoch. «Recent improvements to the PROSITE database.» *Nucleic acids research* 32, n Database issue (Jan. 2004): D134–7. doi:10.1093/nar/gkh044.

Iakoucheva, Lilia M., Celeste J. Brown, J.David Lawson, Zoran Obradović, e A.Keith Dunker. «Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins». *Journal of Molecular Biology* 323, n 3 (2002): 573–584.

James, Thomas L., Volker Dötsch, Uli Schmitz, J.P. Linge, S.I. O'Donoghue, e Michael Nilges. «[5] Automated assignment of ambiguous nuclear overhauser effects with ARIA». *Methods in Enzymology* 339 (2001): 71–90.

John;Madigan, e Michael John M.;Martinko. *Brocks Biology of Microorganisms*. Lebanon, Indiana, U.S.A.: Pren, 2005.

Jones, David T. «Protein secondary structure prediction based on position-specific scoring matrices1 1Edited by G. Von Heijne». *Journal of Molecular Biology* 292, n 2 (1999): 195–202.

Jones, David T, e Jonathan J Ward. «Prediction of disordered regions in proteins from position specific score matrices.» *Proteins* 53 Suppl 6 (Jan. 2003): 573–8. doi:10.1002/prot.10528.

Jones, Laura J.F., Rut Carballido-López, e Jeffery Errington. «Control of Cell Shape in Bacteria». *Cell* 104, n 6 (2001): 913–922.

Kabsch, W, e C Sander. «Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.» *Biopolymers* 22, n 12 (Dec. 1983): 2577–637. doi:10.1002/bip.360221211.

Karplus, Martin, e J Andrew McCammon. «Molecular dynamics simulations of biomolecules.» *Nature structural biology* 9, n 9 (Sep. 2002): 646–52. doi:10.1038/nsb0902-646.

Kawai, Yoshikazu, Kei Asai, e Jeffery Errington. «Partial functional redundancy of MreB isoforms, MreB, Mbl and MreBH, in cell morphogenesis of Bacillus subtilis.» *Molecular microbiology* 73, n 4 (Aug. 2009): 719–31. doi:10.1111/j.1365-2958.2009.06805.x.

Kay, Lewis E. «Pulsed-field gradient-enhanced three-dimensional NMR experiment for correlating 13C.alpha./.beta., 13C', and 1H.alpha. chemical shifts in uniformly carbon-13-labeled proteins dissolved in water». *Journal of the American Chemical Society* 115, n 5 (Mar. 1993): 2055–2057. doi:10.1021/ja00058a072.

Kay, Lewis E., Dennis A. Torchia, e Ad Bax. «Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease». *Biochemistry* 28, n 23 (Nov. 1989): 8972–8979. doi:10.1021/bi00449a003.

Kay, Lewis, Paul Keifer, e Tim Saarinen. «Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity». *Journal of the American Chemical Society* 114, n 26 (Dec. 1992): 10663–10665. doi:10.1021/ja00052a088.

Kendrew, J C. «Structure and function in myoglobin and other proteins.» *Federation proceedings* 18, n 2, Part 1 (July 1959): 740–51.

Kleinschnitz, Eva-Maria, Andrea Heichlinger, Kathrin Schirner, Juliane Winkler, Annette Latus, Iris Maldener, Wolfgang Wohlleben, e Günther Muth. «Proteins encoded by the mre gene cluster in Streptomyces coelicolor A3(2) cooperate in spore wall synthesis.» *Molecular microbiology* 79, n 5 (Mar. 2011): 1367–79. doi:10.1111/j.1365-2958.2010.07529.x.

Kleinschnitz, Eva-Maria, Annette Latus, Steffen Sigle, Iris Maldener, Wolfgang Wohlleben, e Günther Muth. «Genetic analysis of SCO2997, encoding a TagF homologue, indicates a role for wall teichoic acids in sporulation of Streptomyces coelicolor A3(2).» *Journal of bacteriology* 193, n 21 (Nov. 2011): 6080–5. doi:10.1128/JB.05782-11.

Kobayashi, K, S D Ehrlich, A Albertini, G Amati, K K Andersen, M Arnaud, K Asai, et al. «Essential Bacillus subtilis genes.» *Proceedings of the National Academy of Sciences of the United States of America* 100, n 8 (Ap. 2003): 4678–83. doi:10.1073/pnas.0730515100.

Koehl, Patrice, e Michael Levitt. «Improved recognition of native-like protein structures using a family of designed sequences.» *Proceedings of the National Academy of Sciences of the United States of America* 99, n 2 (Jan. 2002): 691–6. doi:10.1073/pnas.022408799.

Kojima, C, A Ono, M Kainosho, e T L James. «Quantitative measurement of transverse and longitudinal cross-correlation between 13C-1H dipolar interaction and 13C chemical shift anisotropy: application to a 13C-labeled DNA duplex.» *Journal of magnetic resonance (San Diego, Calif. : 1997)* 136, n 2 (Feb. 1999): 169–75. doi:10.1006/jmre.1998.1656.

Kopp, Jurgen, e Torsten Schwede. «Automated protein structure homology modeling: a progress report.» *Pharmacogenomics* 5, n 4 (June 2004): 405–16. doi:10.1517/14622416.5.4.405.

Krimm, Isabelle, Cecilia Östlund, Bernard Gilquin, Joël Couprie, Paul Hossenlopp, Jean-Paul Mornon, Gisèle Bonne, Jean-Claude Courvalin, Howard J Worman, e Sophie Zinn-Justin. «The Ig-like Structure of the C-Terminal Domain of Lamin A/C, Mutated in Muscular Dystrophies, Cardiomyopathy, and Partial Lipodystrophy». *Structure* 10, n 6 (2002): 811–823.

Kroenke, Christopher D., J. Patrick Loria, Larry K. Lee, Mark Rance, e Arthur G. Palmer. «Longitudinal and Transverse 1 H− 15 N Dipolar/ 15 N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules». *Journal of the American Chemical Society* 120, n 31 (Aug. 1998): 7905–7915. doi:10.1021/ja980832l.

Kruse, Thomas, e Kenn Gerdes. «Bacterial DNA segregation by the actin-like MreB protein». *Trends in Cell Biology* 15, n 7 (2005): 343–345.

Kruse, Thomas, Jakob Mø ller-Jensen, Anders Lø bner -Olesen, e Kenn Gerdes. «Dysfunctional MreB inhibits chromosome segregation in Escherichia coli.» *The EMBO journal* 22, n 19 (Oct. 2003): 5283–92. doi:10.1093/emboj/cdg504.

Kunst, F, N Ogasawara, I Moszer, A M Albertini, G Alloni, V Azevedo, M G Bertero, et al. «The complete genome sequence of the gram-positive bacterium Bacillus subtilis.» *Nature* 390, n 6657 (Nov. 1997): 249–56. doi:10.1038/36786.

Kurono, Naomi, Ayako Matsuda, Rika Etchuya, Rina Sobue, Yumi Sasaki, Miki Ito, Tsuyako Ando, e Sumio Maeda. «Genome-wide screening of Escherichia coli genes involved in execution and promotion of cell-to-cell transfer of non-conjugative plasmids: rodZ (yfgA) is essential for plasmid acceptance in recipient cells.» *Biochemical and biophysical research communications* 421, n 1 (Ap. 2012): 119–23. doi:10.1016/j.bbrc.2012.03.127.

Lange, Oliver F, Paolo Rossi, Nikolaos G Sgourakis, Yifan Song, Hsiau-Wei Lee, James M Aramini, Asli Ertekin, et al. «Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples.» *Proceedings of the National Academy of Sciences of the United States of America* 109, n 27 (July 2012): 10873–8. doi:10.1073/pnas.1203013109.

Lange, Oliver F., e David Baker. «Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation». *Proteins: Structure, Function, and Bioinformatics* 80, n 3 (Mar. 2012): 884–895. doi:10.1002/prot.23245.

Laskowski, Roman A. «Structural quality assurance.» *Methods of biochemical analysis* 44 (Jan. 2003): 273–303.

Leaver, Mark, e Jeff Errington. «Roles for MreC and MreD proteins in helical growth of the cylindrical cell wall in Bacillus subtilis.» *Molecular microbiology* 57, n 5 (Sep. 2005): 1196–209. doi:10.1111/j.1365-2958.2005.04736.x.

Lescop, Ewen, Paul Schanda, e Bernhard Brutscher. «A set of BEST triple-resonance experiments for time-optimized protein resonance assignment», 2007.

Levin, P A, P S Margolis, P Setlow, R Losick, e D Sun. «Identification of Bacillus subtilis genes for septum placement and shape determination.» *J. Bacteriol.* 174, n 21 (Nov. 1992): 6717–6728.

Levitt, Malcolm H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2008.

Lichstein, H C. «Studies of the Effect of Sodium Azide on Microbic Growth and Respiration: III. The Effect of Sodium Azide on the Gas Metabolism of *B. subtilis* and P. aeruginosa and the Influence of Pyocyanine on the Gas Exchange of a Pyocyanine-Free Strain of P. aerugino». *Journal of bacteriology* 47, n 3 (Mar. 1944): 239–51.

Linge, J P, S I O'Donoghue, e M Nilges. «Automated assignment of ambiguous nuclear overhauser effects with ARIA.» *Methods in enzymology* 339 (Jan. 2001): 71–90.

Linge, J. P., M. Habeck, W. Rieping, e M. Nilges. «ARIA: automated NOE assignment and NMR structure calculation». *Bioinformatics* 19, n 2 (Jan. 2003): 315–316. doi:10.1093/bioinformatics/19.2.315.

Linge, Jens P, Mark A Williams, Christian A E M Spronk, Alexandre M J J Bonvin, e Michael Nilges. «Refinement of protein structures in explicit solvent.» *Proteins* 50, n 3 (Feb. 2003): 496–506. doi:10.1002/prot.10299.

Lobley, Anna, Michael I Sadowski, e David T Jones. «pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination.» *Bioinformatics (Oxford, England)* 25, n 14 (July 2009): 1761–7. doi:10.1093/bioinformatics/btp302.

Lock, Rowena L, e Elizabeth J Harry. «Cell-division inhibitors: new insights for future antibiotics.» *Nature reviews. Drug discovery* 7, n 4 (Ap. 2008): 324–38. doi:10.1038/nrd2510.

Lovering, Andrew L, e Natalie C J Strynadka. «High-resolution structure of the major periplasmic domain from the cell shape-determining filament MreC.» *Journal of molecular biology* 372, n 4 (Sep. 2007): 1034–44. doi:10.1016/j.jmb.2007.07.022.

Luginbühl, Peter, e Kurt Wüthrich. «Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules». *Progress in Nuclear Magnetic Resonance Spectroscopy* 40, n 3 (Ap. 2002): 199–247.

Markley, John L, Ad Bax, Yoji Arata, C.W Hilbers, Robert Kaptein, Brian D Sykes, Peter E Wright, e Kurt Wüthrich. «Recommendations for the presentation of NMR structures of proteins and nucleic acids». *Journal of Molecular Biology* 280, n 5 (1998): 933–952.

Matteï, Pierre-Jean, David Neves, e Andréa Dessen. «Bridging cell wall biosynthesis and bacterial morphogenesis». *Current Opinion in Structural Biology* 20, n 6 (2010): 749–755.

McGuffin, L J, K Bryson, e D T Jones. «The PSIPRED protein structure prediction server.» *Bioinformatics (Oxford, England)* 16, n 4 (Ap. 2000): 404–5.

McGuffin, L. J., K. Bryson, e D. T. Jones. «The PSIPRED protein structure prediction server». *Bioinformatics* 16, n 4 (Ap. 2000): 404–405. doi:10.1093/bioinformatics/16.4.404.

Meiboom, S., e D. Gill. «Modified Spin-Echo Method for Measuring Nuclear Relaxation Times». *Review of Scientific Instruments* 29, n 8 (Aug. 1958): 688. doi:10.1063/1.1716296.

Metcalfe, Emily E, Jamillah Zamoon, David D Thomas, e Gianluigi Veglia. «(1)H/(15)N heteronuclear NMR spectroscopy shows four dynamic domains for phospholamban reconstituted in dodecylphosphocholine micelles.» *Biophysical journal* 87, n 2 (Aug. 2004): 1205–14. doi:10.1529/biophysj.103.038844.

Mitobe, Jiro, Itaru Yanagihara, Kiyouhisa Ohnishi, Shouji Yamamoto, Makoto Ohnishi, Akira Ishihama, e Haruo Watanabe. «RodZ regulates the post-transcriptional processing of the Shigella sonnei type III secretion system.» *EMBO reports* 12, n 9 (Sep. 2011): 911–6. doi:10.1038/embor.2011.132.

Mohammadi, Tamimount, Aneta Karczmarek, Muriel Crouvoisier, Ahmed Bouhss, Dominique Mengin-Lecreulx, e Tanneke den Blaauwen. «The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in Escherichia coli.» *Molecular microbiology* 65, n 4 (Aug. 2007): 1106–21. doi:10.1111/j.1365-2958.2007.05851.x.

Morris, A L, M W MacArthur, E G Hutchinson, e J M Thornton. «Stereochemical quality of protein structure coordinates.» *Proteins* 12, n 4 (Ap. 1992): 345–64. doi:10.1002/prot.340120407.

Moult, John. «Rigorous performance evaluation in protein structure modelling and implications for computational biology.» *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 361, n 1467 (Mar. 2006): 453–8. doi:10.1098/rstb.2005.1810.

Muhandiram, D.R., e L.E. Kay. «Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity». *Journal of Magnetic Resonance, Series B* 103, n 3 (1994): 203–216.

Mumenthaler, Ch., e W. Braun. «Automated Assignment of Simulated and Experimental NOESY Spectra of Proteins by Feedback Filtering and Self-correcting Distance Geometry». *Journal of Molecular Biology* 254, n 3 (1995): 465–480.

Mumenthaler, Christian, Peter Güntert, Werner Braun, e Kurt Wüthrich. «Automated Combined Assignment of NOESY Spectra and Three-dimensional Protein Structure Determination». *Journal of Biomolecular NMR* 10, n 4 (Dec. 1997): 351–362. doi:10.1023/A:1018383106236.

Nakano, M M, e P Zuber. «Anaerobic Growth of a "Strict Aerobe" (Bacillus Subtilis).» *Annual Review of Microbiology* 52 (Jan. 1998): 165–90. doi:10.1146/annurev.micro.52.1.165.

Nanninga, N. «Cell division and peptidoglycan assembly in Eschenchia coli». *Molecular Microbiology* 5, n 4 (Ap. 1991): 791–795. doi:10.1111/j.1365-2958.1991.tb00751.x.

Nederveen, Aart J, Jurgen F Doreleijers, Wim Vranken, Zachary Miller, Chris a E M Spronk, Sander B Nabuurs, Peter Güntert, et al. «RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank.» *Proteins* 59, n 4 (June 2005): 662–72. doi:10.1002/prot.20408.

Ni, Lisheng, Weijun Xu, Muthiah Kumaraswami, e Maria A Schumacher. «Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition.» *Proceedings of the National Academy of Sciences of the United States of America* 107, n 26 (June 2010): 11763–8. doi:10.1073/pnas.1003817107.

Nilges, Michael, Maria J Macias, Séan I O'Donoghue, e Hartmut Oschkinat. «Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin11Edited by P. E. Wright». *Journal of Molecular Biology* 269, n 3 (1997): 408–422.

Nilges, Michael, e Seán I. O'Donoghue. «Ambiguous NOEs and automated NOE assignment». *Progress in Nuclear Magnetic Resonance Spectroscopy* 32, n 2 (1998): 107–139.

Nilges, Michael, e Seán I. O'Donoghue. «Ambiguous NOEs and automated NOE assignment». *Progress in Nuclear Magnetic Resonance Spectroscopy* 32, n 2 (Ap. 1998): 107–139.

Noggle, Joseph H., e Roger E. Schirmer. *Nuclear Overhauser Effect: Chemical Applications*. Academic Press Inc, 1971.

Nugent, Timothy, e David T Jones. «Transmembrane protein topology prediction using support vector machines.» *BMC bioinformatics* 10, n 1 (Jan. 2009): 159. doi:10.1186/1471-2105-10-159.

Oldfield, Christopher J, Yugong Cheng, Marc S Cortese, Celeste J Brown, Vladimir N Uversky, e A Keith Dunker. «Comparing and combining predictors of mostly disordered proteins.» *Biochemistry* 44, n 6 (Feb. 2005): 1989–2000. doi:10.1021/bi047993o.

Ottiger, Marcel, Frank Delaglio, e Ad Bax. «Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra». *Journal of Magnetic Resonance* 378, n 131 (1998): 373–378.

Palmer, Arthur G., Mark Rance, e Peter E. Wright. «Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy». *Journal of the American Chemical Society* 113, n 12 (June 1991): 4371–4380. doi:10.1021/ja00012a001.

Pearson, Dinshaw J Patel, Eric Westhof / Anna Tramontano, William R, Ken A Dill, S Banu Ozkan, Thomas R Weikl, John D Chodera, e Vincent A Voelz. «The protein folding problem: when will it be solved?» *Current Opinion in Structural Biology* 17, n 3 (2007): 342–346.

Pereira, Ana. *NMR investigation of the Bacillus subtilis morphogenic factor RodZ. Graduation final thesis.* FCT-UNL, 2011.

Petsko, Gregory A. *From Sequence to Consequence*. Vol 1. 1. BioMed Central, 2000. /pmc/articles/PMC138826/

Potapov, Vladimir, Vladimir Sobolev, Marvin Edelman, Alexander Kister, e Israel Gelfand. «Protein–protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins.» *Journal of molecular biology* 342, n 2 (Sep. 2004): 665–79. doi:10.1016/j.jmb.2004.06.072.

Raman, Srivatsan, Oliver F Lange, Paolo Rossi, Michael Tyka, Xu Wang, James Aramini, Gaohua Liu, et al. «NMR structure determination for larger proteins using backbone-only data.» *Science (New York, N.Y.)* 327, n 5968 (Feb. 2010): 1014–8. doi:10.1126/science.1183649.

Reto Koradi, Martin Billeter, Max Engeli, Peter Guntert, Kurt Wuthrich. «Automated Peak Picking and Peak Integration in Macromolecular Nmr Spectra Using AUTOPSY» ( ).

Rhodes, Valerie Daggett, Alan R. Fersht / Elena Conti, Daniela, Peter E Wright, e H Jane Dyson. «Linking folding and binding». *Current Opinion in Structural Biology* 19, n 1 (2009): 31–38.

Rohl, Carol A., e David Baker. «De Novo Determination of Protein Backbone Structure from Residual Dipolar Couplings Using Rosetta». *Journal of the American Chemical Society* 124, n 11 (Mar. 2002): 2723–2729. doi:10.1021/ja016880e.

Romero, P, Z Obradovic, X Li, E C Garner, C J Brown, e A K Dunker. «Sequence complexity of disordered protein.» *Proteins* 42, n 1 (Jan. 2001): 38–48.

Rost, Burkhard. *Protein structures sustain evolutionary drift*. Vol 2, 1997.

Roy, Ambrish, Alper Kucukural, e Yang Zhang. «I-TASSER: a unified platform for automated protein structure and function prediction.» *Nature protocols* 5, n 4 (Ap. 2010): 725–38. doi:10.1038/nprot.2010.5.

Roy, Ambrish, Jianyi Yang, e Yang Zhang. «COFACTOR: an accurate comparative algorithm for structure-based protein function annotation.» *Nucleic acids research* 40, n Web Server issue (July 2012): W471–7. doi:10.1093/nar/gks372.

Sanchez-Torres, Viviana, Toshinari Maeda, e Thomas K Wood. «Global regulator H-NS and lipoprotein NlpI influence production of extracellular DNA in Escherichia coli.» *Biochemical and biophysical research communications* 401, n 2 (Oct. 2010): 197–202. doi:10.1016/j.bbrc.2010.09.026.

Sargent, M G. «Control of cell length in Bacillus subtilis.» *J. Bacteriol.* 123, n 1 (July 1975): 7–19.

Sattler, Michael, Jürgen Schleucher, e Christian Griesinger. «Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients». *Progress in Nuclear Magnetic Resonance Spectroscopy* 34, n 2 (Mar. 1999): 93–158.

Schanda, Paul, Hélène Van Melckebeke, e Bernhard Brutscher. «Speeding up three-dimensional protein NMR experiments to a few minutes.» *Journal of the American Chemical Society* 128, n 28 (July 2006): 9042–3. doi:10.1021/ja062025p.

Scheffers, Dirk-Jan, e Mariana G Pinho. «Bacterial cell wall synthesis: new insights from localization studies.» *Microbiology and molecular biology reviews : MMBR* 69, n 4 (Dec. 2005): 585–607. doi:10.1128/MMBR.69.4.585-607.2005.

Schleucher, J., M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S.J. Glaser, O.W. Sørensen, e C. Griesinger. «A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients». *Journal of Biomolecular NMR* 4, n 2 (Mar. 1994). doi:10.1007/BF00175254.

Schleucher, Jurgen, Michael Sattler, e Christian Griesinger. «Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNCO Experiment». *Angewandte Chemie International Edition in English* 32, n 10 (Oct. 1993): 1489–1491. doi:10.1002/anie.199314891.

Schwede, T. «SWISS-MODEL: an automated protein homology-modeling server». *Nucleic Acids Research* 31, n 13 (July 2003): 3381–3385. doi:10.1093/nar/gkg520.

Schwede, Torsten, Jürgen Kopp, Nicolas Guex, e Manuel C Peitsch. «SWISS-MODEL: An automated protein homology-modeling server.» *Nucleic acids research* 31, n 13 (July 2003): 3381–5.

Serrano, Pedro, Bill Pedrini, Biswaranjan Mohanty, Michael Geralt, Torsten Herrmann, e Kurt Wüthrich. «The J-UNIO protocol for automated protein structure determination by NMR in solution.» *Journal of biomolecular NMR* 53, n 4 (Aug. 2012): 341–54. doi:10.1007/s10858-012-9645-2.

Shen, Yang, e Ad Bax. «Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology.» *Journal of biomolecular NMR* 38, n 4 (Aug. 2007): 289–302. doi:10.1007/s10858-007-9166-6.

Shen, Yang, Oliver Lange, Frank Delaglio, Paolo Rossi, James M Aramini, Gaohua Liu, Alexander Eletsky, et al. «Consistent blind protein structure generation from NMR chemical shift data.» *Proceedings of the National Academy of Sciences of the United States of America* 105, n 12 (Mar. 2008): 4685–90. doi:10.1073/pnas.0800256105.

Shen, Yang, Robert Vernon, David Baker, e Ad Bax. «De novo protein structure generation from incomplete chemical shift assignments.» *Journal of biomolecular NMR* 43, n 2 (Feb. 2009): 63–78. doi:10.1007/s10858-008-9288-5.

Shiomi, Daisuke, Masako Sakai, e Hironori Niki. «Determination of bacterial rod shape by a novel cytoskeletal membrane protein.» *The EMBO journal* 27, n 23 (Dec. 2008): 3081–91. doi:10.1038/emboj.2008.234.

Simons, K T, R Bonneau, I Ruczinski, e D Baker. «Ab initio protein structure prediction of CASP III targets using ROSETTA.» *Proteins* Suppl 3 (Jan. 1999): 171–6.

Simons, Kim T, Charlie Strauss, e David Baker. «Prospects for ab initio protein structural genomics1 1Edited by B. Honig». *Journal of Molecular Biology* 306, n 5 (2001): 1191–1199.

Simons, Kim T., Charles Kooperberg, Enoch Huang, e David Baker. «Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions11Edited by F. E. Cohen». *Journal of Molecular Biology* 268, n 1 (1997): 209–225.

Sippl, Manfred J. «Calculation of conformational ensembles from potentials of mena force». *Journal of Molecular Biology* 213, n 4 (1990): 859–883.

Sitbon, Einat, e Shmuel Pietrokovski. «Occurrence of protein structure elements in conserved sequence regions.» *BMC structural biology* 7, n 1 (Jan. 2007): 3. doi:10.1186/1472-6807-7-3.

Spronk, Christian A.E.M., Jens P. Linge, Cornelis W. Hilbers, e Geerten W. Vuister. «Improving the Quality of Protein Structures Derived by NMR Spectroscopy**». *Journal of Biomolecular NMR* 22, n 3 (Mar. 2002): 281–289. doi:10.1023/A:1014971029663.

Standley, Daron M., Volker A. Eyrich, Anthony K. Felts, Richard A. Friesner, e Ann E. McDermott. «A branch and bound algorithm for protein structure refinement from sparse NMR data sets11Edited by F. Cohen». *Journal of Molecular Biology* 285, n 4 (1999): 1691–1710.

Stoscheck, C M. «Quantitation of protein.» *Methods in enzymology* 182 (Jan. 1990): 50–68.

Teng, Quincy. «Structural Biology, Practical NMR Applications». *Structural Biology* (2005).

Tieleman, D.P., Waldemar Vollmer, e Ute Bertsche. «Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli». *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778, n 9 (2008): 1714–1734.

Van den Ent, Fusinita, Christopher M Johnson, Logan Persons, Piet Boer, e Jan Löwe. «Bacterial actin MreB assembles in complex with cell shape protein RodZ.» *The EMBO journal* 29, n 6 (Mar. 2010): 1081–90. doi:10.1038/emboj.2010.9.

Van den Ent, Fusinita, Mark Leaver, Felipe Bendezu, Jeff Errington, Piet Boer, e Jan Löwe. «Dimeric structure of the cell shape protein MreC and its functional implications.» *Molecular microbiology* 62, n 6 (Dec. 2006): 1631–42.

Van Heijenoort, Jean. «Recent Advances in the Formation of the Bacterial Peptidoglycan Monomer Unit (1985 to 2000)». *Natural Product Reports* 18, n 5 (Jan. 2001): 503–519. doi:10.1039/a804532a.

Tieleman, D.P., Waldemar Vollmer, e Ute Bertsche. «Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli». *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778, n 9 (2008): 1714–1734.

Tompa, Peter. «Intrinsically unstructured proteins». *Trends in Biochemical Sciences* 27, n 10 (2002): 527–533.

Tompa, Peter, Zsuzsanna Dosztanyi, e Istvan Simon. «Prevalent structural disorder in *E. coli* and S. cerevisiae proteomes.» *Journal of proteome research* 5, n 8 (Aug. 2006): 1996–2000. doi:10.1021/pr0600881.

Typas, Athanasios, Manuel Banzhaf, Carol A Gross, e Waldemar Vollmer. «From the regulation of peptidoglycan synthesis to bacterial growth and morphology.» *Nature reviews. Microbiology* 10, n 2 (Feb. 2012): 123–36. doi:10.1038/nrmicro2677.

Ulrich, Eldon L, Hideo Akutsu, Jurgen F Doreleijers, Yoko Harano, Yannis E Ioannidis, Jundong Lin, Miron Livny, et al. «BioMagResBank.» *Nucleic acids research* 36, n Database issue (Jan. 2008): D402–8. doi:10.1093/nar/gkm957.

Varley, A W, e G C Stewart. «The divIVB region of the Bacillus subtilis chromosome encodes homologs of Escherichia coli septum placement (minCD) and cell shape (mreBCD) determinants.» *J. Bacteriol.* 174, n 21 (Nov. 1992): 6729–6742.

Vitkup, D, E Melamud, J Moult, e C Sander. «Completeness in structural genomics.» *Nature structural biology* 8, n 6 (June 2001): 559–66. doi:10.1038/88640.

Vollmer, Waldemar, Didier Blanot, e Miguel A Pedro. «Peptidoglycan structure and architecture.» *FEMS microbiology reviews* 32, n 2 (Mar. 2008): 149–67. doi:10.1111/j.1574-6976.2007.00094.x.

Vucetic, Slobodan, Celeste J Brown, A Keith Dunker, e Zoran Obradovic. «Flavors of protein disorder.» *Proteins* 52, n 4 (Sep. 2003): 573–84. doi:10.1002/prot.10437.

Wagner, Jennifer K, Cheryl D Galvani, e Yves V Brun. «Caulobacter crescentus requires RodA and MreB for stalk synthesis and prevention of ectopic pole formation.» *Journal of bacteriology* 187, n 2 (Jan. 2005): 544–53. doi:10.1128/JB.187.2.544-553.2005.

Wallner, Björn, e Arne Elofsson. «All are not equal: a benchmark of different homology modeling programs.» *Protein science : a publication of the Protein Society* 14, n 5 (May 2005): 1315–27. doi:10.1110/ps.041253405.

Ward, J.J., J.S. Sodhi, L.J. McGuffin, B.F. Buxton, e D.T. Jones. «Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life». *Journal of Molecular Biology* 337, n 3 (2004): 635–645.

Weinreb, P H, W Zhen, A W Poon, K A Conway, e P T Lansbury. «NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded.» *Biochemistry* 35, n 43 (Oct. 1996): 13709–15. doi:10.1021/bi961799n.

White, Courtney L, e James W Gober. «MreB: pilot or passenger of cell wall synthesis?» *Trends in microbiology* 20, n 2 (Feb. 2012): 74–9. doi:10.1016/j.tim.2011.11.004.

White, Courtney L, Aleksandar Kitich, e James W Gober. «Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD.» *Molecular microbiology* 76, n 3 (May 2010): 616–33. doi:10.1111/j.1365-2958.2010.07108.x.

White, Courtney L., e James W. Gober. «MreB: pilot or passenger of cell wall synthesis?» *Trends in Microbiology* 20, n 2 (2012): 74–79.

Wilkins, M R, E Gasteiger, A Bairoch, J C Sanchez, K L Williams, R D Appel, e D F Hochstrasser. «Protein identification and analysis tools in the ExPASy server.» *Methods in molecular biology (Clifton, N.J.)* 112 (Jan. 1999): 531–52.

Williams, A F, e A N Barclay. «The Immunoglobulin Superfamily–domains for Cell Surface Recognition.» *Annual Review of Immunology* 6 (Jan. 1988): 381–405. doi:10.1146/annurev.iy.06.040188.002121.

Wishart, D S, e B D Sykes. «The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data.» *Journal of biomolecular NMR* 4, n 2 (Mar. 1994): 171–80.

Wishart, DavidS., e BrianD. Sykes. «The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data». *Journal of Biomolecular NMR* 4, n 2 (Mar. 1994). doi:10.1007/BF00175245.

Wu, Sitao, Jeffrey Skolnick, e Yang Zhang. «Ab initio modeling of small proteins by iterative TASSER simulations.» *BMC biology* 5, n 1 (Jan. 2007): 17. doi:10.1186/1741-7007-5-17.

Wuthrich, Kurt. «NMR of proteins and nucleic acids». *The George Fisher Baker non-resident lectureship in chemistry at Cornell Unversity* (1986).

Xu L, Sedelnikova SE, Baker PJ, Rice DW. «The structure of ymfm, a putative DNA-binding membrane protein from staphylococcus aureus». (not published yet)

Young, Kevin D. «Bacterial shape: two-dimensional questions and possibilities.» *Annual review of microbiology* 64 (Jan. 2010): 223–40. doi:10.1146/annurev.micro.112408.134102.

Zhang, Haiyan, Stephen Neal, e David S Wishart. «RefDB: a database of uniformly referenced protein chemical shifts.» *Journal of biomolecular NMR* 25, n 3 (Mar. 2003): 173–95.

Zhang, Haiyan, Stephen Neal, e David S. Wishart. «RefDB: A Database of Uniformly Referenced Protein Chemical Shifts». *Journal of Biomolecular NMR* 25, n 3 (Mar. 2003): 173–195. doi:10.1023/A:1022836027055.

Zhang, Yang. «I-TASSER server for protein 3D structure prediction.» *BMC bioinformatics* 9, n 1 (Jan. 2008): 40. doi:10.1186/1471-2105-9-40.

Zhang, Yang, Daisuke Kihara, e Jeffrey Skolnick. «Local energy landscape flattening: parallel hyperbolic Monte Carlo sampling of protein folding.» *Proteins* 48, n 2 (Aug. 2002): 192–201. doi:10.1002/prot.10141.

Zhang, Yang, Andrzej Kolinski, e Jeffrey Skolnick. «TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction». *Biophysical Journal* 85, n 2 (2003): 1145–1164.

Zhang, Yang, e Jeffrey Skolnick. «Automated structure prediction of weakly homologous proteins on a genomic scale.» *Proceedings of the National Academy of Sciences of the United States of America* 101, n 20 (May 2004): 7594–9. doi:10.1073/pnas.0305695101.

Zhang Y, Skolnick J. «Scoring function for automated assessment of protein structure template quality.» *Proteins* 57, n 4 (Dec. 2004): 702–10. doi:10.1002/prot.20264.

Zhang Y, Skolnick J. «TM-align: a protein structure alignment algorithm based on the TM-score.» *Nucleic acids research* 33, n 7 (Jan. 2005): 2302–9. doi:10.1093/nar/gki524.

Bacillus subtilis and its closest relatives : from genes to cells SONENSHEIN Abraham L., HOCH James A., LOSICK Richard: Librairie Lavoisier, .

Bacillus Subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology and Molecular Genetics, 1993.

Biomolecular NMR - Stable Isotopes from Cambridge Isotope Laboratories, .

NanoDrop 2000/2000c Spectrophotometer. V1.0 User Manual, .

NMR Sample Tubes and Accessories from: NEWERA, .

pET vector Expression System Manual, 11th Edition protocols and methods.

Structural Bioinformatics, 2009.